UNIVERSITY OF CALIFORNIA, SAN DIEGO

The Biomechanical Basis of DNA Breakage in Chronic Myelogenous Leukemia

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

in

Doctor of Philosophy in Chemical Engineering

by

Chi-Chiang Tu

Committee in Charge:

Professor Lanping Amy Sung, Chair
Professor Pao Chau, Co-Chair
Professor Charles Cantor
Professor Michael J. Heller
Professor Richard Herz

2010

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

The Dissertation of Chi-Chiang Tu is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

Co-Chair

Chair

University of California, San Diego

2010

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

17-A	17 consecutive deoxyadenosine triphosphate
3C	Chromosome conformation capture
3C-DNA	Chromosome conformation captured DNA
μ-*BCR*	Micro breakpoint cluster region
μg	microgram or 10^{-6} grams
μl	microliter or 10^{-6} liters
AAP	Abridged anchor primer
ATCC	American Type Culture Collection
AUAP	Abridged universal amplification primer
bp	base pair
BP	Breakpoint
CML	Chronic myelogenous leukemia
dCTP	2'-deoxycytidine 5'-triphosphate
dNTP	Deoxyribonucleotide triphosphate
DD	Digested DNA control
DE	Digestion efficiency
DNase	Deoxyribonuclease
DSB	DNA double-stranded breakage
EDTA	Ethylenediaminetetraacetic acid
FCS	Fetal calf serum
FISH	Fluorescence *in situ* hybridization

FRET	Fluorescence resonance energy transfer
GSP1	Primary gene specific primer
GSP2	Secondary gene specific primer
HGNC	HUGO Gene Nomenclature Committee
IgH	Immunoglobulin heavy locus
Kb	kilobase or 10^3 bases
LCR	Locus control region
m-*BCR*	Minor breakpoint cluster region
ml	milliliter or 10^{-3} liters
mM	millimolar or 10^{-3} molars
MBr	Major breakpoint region
M-*BCR*	Major breakpoint cluster region
MC	Monte-Carlo
MNase	Micrococcal nuclease
nm	Nanometer or 10^{-9} meters
ng	nanogram or 10^{-9} grams
NP-40	Nonyl phenoxylpolyethoxylethanol
Oligo-C	Oligo-deoxycytidine triphosphate
pN	piconewton or 10^{-12} newtons
PB	Protein binding site
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction

Ph’	Philadelphia
PP	Positioning protein
rpm	round per minute
RD	Replication-directed
RO	Replication origin
SDS	Sodium dodecyl sulfate
TE	Tris-EDTA
TKM	Tris-potassium-magnesium
TS	Transcription start site
UD	Undigested DNA control

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to first acknowledge my advisor Dr. Amy Sung for guiding me in the study of nucleosome and DNA mechanics. Her wise and inspiring lectures have sparked my interest into the field of molecular biology, and her clear logic and intelligent concepts have facilitated the development of my research. Through her mentorship, I have greatly honed my analytical thinking and problem solving skills. Even though my topic was new to this lab, Dr. Sung has painstakingly built up my knowledge from the ground up. Even though our lab resource was limited, Dr. Sung has generously provided the environment for my laboratory skills to develop and mature. In short, Dr. Sung is not only a great research advisor, but also an exceptional mentor who looks after her students' personal well being.

I would like to express my appreciation to the individual members of my senate committee. I would like to first thank Dr. Herz for critically reviewing my progress. Next I would like to thank both Dr. Chau and Dr. Heller for the key suggestions they provided to my methodology. Finally I would like to thank Dr. Cantor for his many insightful suggestions. My research has been inspired by the members of my committee, and I would like to model my scientific career after each of them.

I am grateful for the help and support I received from the senior members of my lab, my colleagues, and friends here in UCSD. In particular I would like to acknowledge the following people: Dr. Weijuan Yao, who has been my teacher since my first day in lab; Dr. Ian Lian, who is a miracle worker and showed me how to

streamline my experiment; Dr. Carlos Vera, who has been a great consultant and an educator, providing the opportunity for me to refine my teaching skills; Albert Chung, who has been a great friend through both good times and bad; Terrell Green, who has been a good company and our lab spokesman; Dr. Ying-Ja Chen and Dr. Sheng-Hong Chen, who have provided sound advice in my research design and methodologies; Dr. Wilda Helen, who offered her expertise in my cell culture experiments. They and many other friends have brought color to my life, and my doctoral study was better for having known them.

I cannot express in word the gratitude and love I have for my parents. Only with their full support and unconditional love was it possible for me to travel abroad and fulfill my dreams. Finally, I would like to dedicate my doctoral degree to my wife Yi-Chiao Hsieh. She is the pillar of my life. Her love and support is the source of my strength. She has blessed me with the greatest gift any man can ask for, our daughter Sophia. Without her none of this would be possible.

Chapter 1, in part is currently being prepared for submission for publication of the material. Tu, Chi-Chiang; Sung, Lanping A. The dissertation author was the primary investigator and author of this material.

Chapter 2, in part is currently being prepared for submission for publication of the material. Tu, Chi-Chiang; Sung, Lanping A. The dissertation author was the primary investigator and author of this material.

Chapter 3, in part is currently being prepared for submission for publication of the material. Tu, Chi-Chiang; Sung, Lanping A. The dissertation author was the primary investigator and author of this material.

VITA

1995	Bachelor of Science, Chemical Engineering, National Tsing-Hua University, Taiwan
1997	Master of Science, Chemical Engineering, National Tsing-Hua University, Taiwan
1997-1999	Taiwan Army Service
1999-2004	Senior process engineer, Taiwan Semiconductor Manufacturing Company Ltd., Taiwan
2004-2010	Doctor of Philosophy, Chemical Engineering, University of California, San Diego

Publications

Tung, H.L., **Tu, C.C.**, Chang, Y.Y. and Wu, W.T. (1998) "Bubble characteristics and mass transfer in an airlift reactor with multiple net draft tubes." Bioprocess Engineering 18: 323-328.

Tung, H.L., Chion, S.Y., **Tu, C.C.** and Wu, W.T. (1997) "An airlift reactor with double net draft tubes and its application in fermentation." Bioprocess Engineering 17: 1-5.

Tu, C.C. (1997) "Scale up of an airlift reactor with multi-net draft tubes." Master thesis.

Tung, H.L., Chiou, S.Y., **Tu, C.C.** and Wu, W.T. (1996) "Characterization of oxygen transfer in a modified bubble column." APCChE. 547-550, Taipei, Taiwan.

Tu, C.C. (2010) "The biomechanical basis of DNA breakage in chronic myelogenous leukemia." Doctoral dissertation.

Tu, C.C., Sung, L.A. (2010) "The biomechanical basis of DNA breakage in chronic myelogenous leukemia." (Manuscript in preparation)

ABSTRACT OF DISSERTATION

The Biomechanical Basis of DNA Breakage in Chronic Myelogenous Leukemia

by

Chi-Chiang Tu

Doctor of Philosophy in Chemical Engineering

University of California, San Diego, 2010

Professor Lanping Amy Sung, Chair

Professor Pao Chau, Co-Chair

Chronic myelogenous leukemia, a cancer of white blood cells, is characterized by a chromosome translocation between *ABL* and *BCR* due to their close proximity during *BCR* replication. Twenty seven DNA breakpoints are within a major breakpoint cluster region (M-*BCR*), but why these breakpoints are clustered remains unclear. Initially, a Monte-Carlo algorithm calculating the sequence-dependent bending energy

was used to position nucleosomes simultaneously in M-*BCR*. MNase digestion followed by adapter-mediated PCR was then used to map experimentally the nucleosome boundaries. The discrepancy prompted the development of a replication-directed (RD) algorithm predicting nucleosome positions from the replication origin, one after another, each at the local minimum of bending energy. The better agreement confirms the spatial and temporal importance in nucleosome assembly *in vivo*. Therefore, a stepwise model for coupled nucleosome disassembly/reassembly across a replication fork was proposed: Unwinding of DNA by helicases may generate positive superhelical tension to facilitate the disassembly of downstream nucleosomes, which in turn may facilitate the reassembly of upstream nucleosomes. The latter may include a DNA loop formation followed by a sequential closing of two DNA arms around each histone core. Chromatin conformation capturing PCR, involving cross-linking, *Alu*I digestion, re-ligation and PCR, preserved the core DNA conformation in 3-D and confirmed a 292-bp nucleosome-excluded region (A-rich region) containing a stretch of 17 consecutive A (17-A). Nineteen breakpoints are upstream from this region: 6 are clustered immediately upstream to the 17-A, and 10 out of 13 further upstream ones are spread in 4 nucleosome cores. It is likely that (1) the long A-rich region may extrude from the chromatin fiber or form an empty loop or other unusual secondary structures, which may experience greater mechanical stress locally without the "protection" of a histone core; (2) DNA in the newly reassembled nucleosome cores upstream to the A-rich region may lose histone cores under mechanical stress or in a stalled replication. These mechanisms may account for most of the M-*BCR*

breakpoints and at least two other diseases caused by chromosome translocation, suggesting a functional role for the histone core as a "uniform curvature controller" to minimize the DNA breakage.

Introduction

Chronic myelogenous leukemia (CML) is a clonal myeloproliferative expansion of transformed, primitive hematopoietic progenitor cells. It has an annual incidence of 1 to 2 cases per 100,000 people and accounts for 15 percent of leukemia in adults (Faderl *et al.*, 1999). Most CML cases are diagnosed in the chronic phase which presents fatigue, weight loss, abdominal fullness, bleeding, purpura, splenomegaly, leukocytosis, anemia, and thrombocytosis (Faderl *et al.*, 1999). Three to five years after onset, CML progresses from the chronic phase to an accelerated and fatal blast crisis phase where hematopoietic progenitor cells fail to mature (Sawyers, 1999).

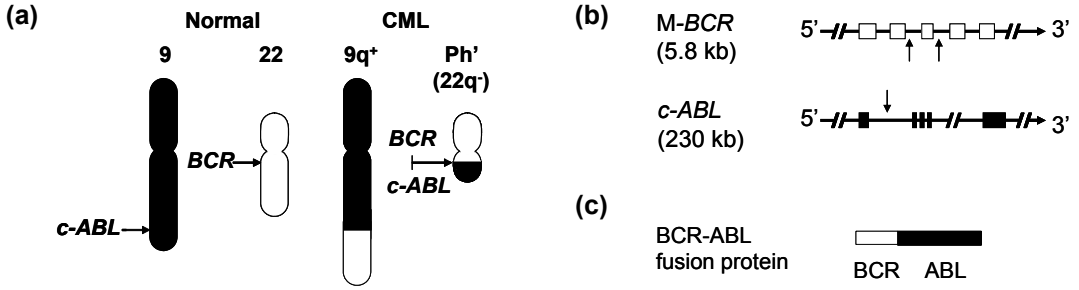


Figure I-1. Ph' chromosome, proto-oncogene, and fusion protein. (a) *c-ABL* gene is translocated from chromosome 9 to *BCR* on chromosome 22 to form Ph'. (b) The major *BCR* region (M-*BCR*, 5 exons) and *c-ABL* region (11 exons) are reciprocally translocated. Boxes refer to exons. Arrows refer to examples of breakpoints. This figure is not to scale. (c) The resulting BCR-ABL fusion protein.

Over 90% CML cases exhibit Philadelphia (Ph') chromosome (Rowley, 1973). It is a fusion chromosome resulting from a reciprocal translocation (de Klein *et al.*, 1982) between 5' *BCR* gene and 3' *c-ABL* gene (Figure I-1a & I-1b). Depending on DNA breakpoint locations, Ph' chromosomes of different sizes produce *BCR-ABL*

fusion mRNA of different lengths (Sawyers, 1999; Faderl *et al.*, 1999; Laurent *et al.*, 2001). The resulting BCR-ABL fusion protein (Figure I-1c), an un-regulable tyrosine kinase, causes massive myeloid expansion in the bone marrow and morphological abnormalities such as hyper/hyposegmentation, abnormal lobulation and ring-shaped nuclei of the leukocytes (Faderl *et al*., 1999; Clarkson *et al.*, 2003).

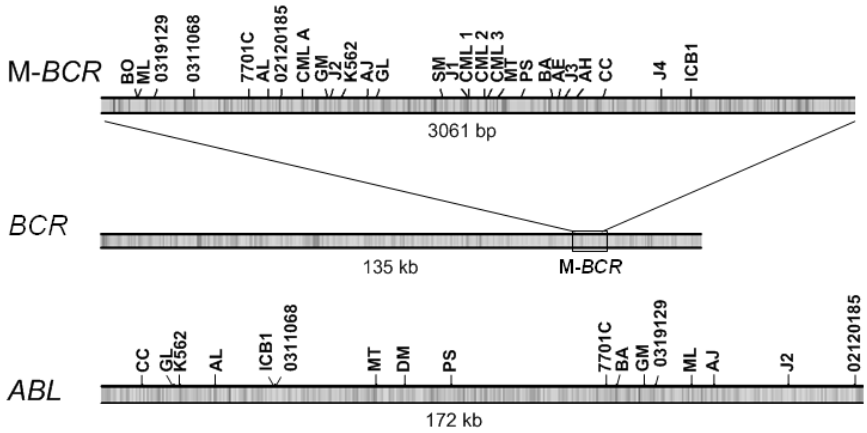


Figure I-2. Precisely mapped breakpoints in *BCR* and *ABL* genes. The gene map is generated by pDRAW32 (Acaclone Software). Twenty seven breakpoints are mapped in M-*BCR* and 17 in *ABL*. Each breakpoint is labeled with the ID from the original articles.

Since the finding of Ph' chromosome, extensive studies have been conducted in attempts to characterize the leukemogenetic mechanisms (Sawyers, 1999; Faderl *et al.*, 1999; Jeffs *et al.*, 1998; Laurent *et al.*, 2001; Abeysinghe *et al.*, 2003; Fisher *et al.*, 2005). The clonal origin of CML is suggested to be the progenitor of blood cells (Fialkow *et al.*, 1977). The sequential organizations of *BCR* and *ABL* have been scanned for clues to the aberrant chromosomal translocation. DNA breakpoints are found to cluster in certain regions of *BCR*, such as the major breakpoint cluster region (M-*BCR*, Figure I-2) and more scattered in their counterparts in *ABL* (Groffen *et al*.,

1984; Heisterkamp *et al.*, 1985; Sowerby *et al.*, 1993; Chissoe *et al.*, 1995; Zhang *et al*., 1995; Jeffs *et al*., 1998). Sequence analysis reveals an *Alu* homology of about 40% for *BCR* and *ABL* genes, respectively (Chissoe *et al.*, 1995). However no consistent breakpoint features are immediately apparent (Chissoe *et al.*, 1995; Jeffs *et al*., 2001). It is suggested that *Alu* sequences have an affinity for the *BCR–ABL* recombination process. Sequence motifs similar to IgH switch pentamers and Translin binding sites presenting on 3' M-*BCR* recombination sites may also be involved in the translocation (Jeffs *et al.*, 1998). Recent investigations show that GC-rich nucleotide sequences are preferred for translocation breakpoints to occur, especially in the high GC content regions of *BCR* (Abeysinghe *et al.*, 2003; Fisher *et al.*, 2005). Therefore, Ph' translocations may not be random but rather favored by certain nucleotide compositions. On the other hand, studies using fluorescence in situ hybridization (FISH) show that in hematopoietic cells, intergenic distances between *BCR* and *ABL* vary periodically during the cell cycle (Kozubek *et al*. 1997; Neves *et al.*, 1999; Anastasi *et al.*, 1999). A significant association of *ABL* with *BCR* is seen at the transition between S and G2 phases (Neves *et al.*, 1999). Further study shows *BCR* and *ABL* even merge at the time when *BCR* seems to be replicating (Anastasi *et al.*, 1999). It is therefore proposed that *BCR*/*ABL* translocation could be a replication error in *BCR*. This is indirectly supported by the fact that a very low level of Ph' chromosome is exhibited in blood of some healthy individuals (Biernaux *et al.*, 1995), meaning the replication error occurs more frequently than just by chance. Our immune surveillance seems to prevent us from the emergence of CML. The current

understanding in molecular biology of CML has led to one effective treatment against the Ph' translocation product, BCR-ABL fusion protein. A specially designed BCR-ABL inhibitor, Gleevac, has shown promising results in patients in chronic phases whose blood cell responses are much higher than those treated with traditional chemotherapeutic agents (Huntly *et al.*, 2003). However, there is a high relapse rate with this agent and it is less impressive for patients in blast crisis phase (Huntly *et al.*, 2003). Although CML is probably one of the most characterized cancerous diseases in molecular biology, the exact mechanism by which Ph' translocations initiate remains unclear.

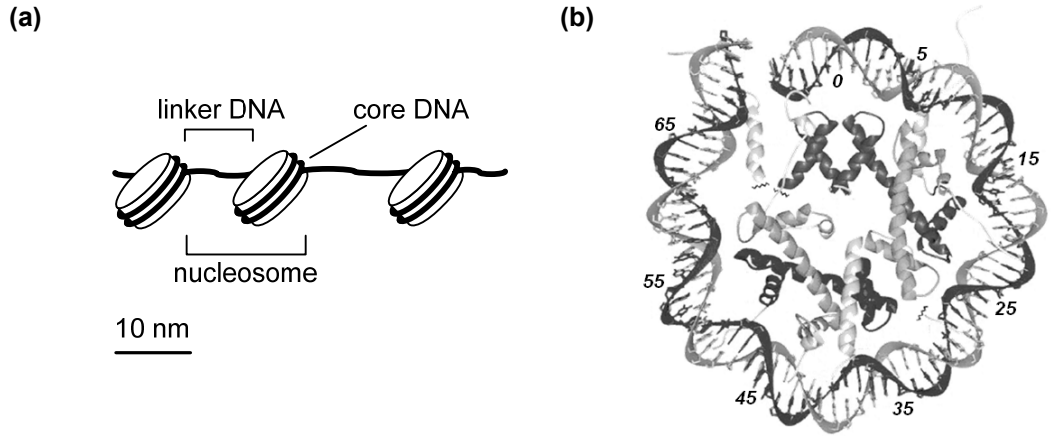


Figure I-3. Structure of nucleosomes. (a) Structural organization of the nucleosome. A nucleosome consists of 147 bp of core DNA wrapping around a disk-like histone core octamer, and averagely 54 bp of linker DNA connecting to the adjacent nucleosome. (b) The core region of a nucleosome by Davey *et al.*, 2002. Molecules in the center are histone core proteins, H2A, H2B, H3, and H4. Numbers along the double helical DNA refer to the nucleotide distances to the dyad (0, the center of a symmetrical wrapping). Only the molecules on the front side are shown for clarity.

It is hypothesized that changes of chromatin structure may increase the probability of a translocation during G1 and S (G2) stages of the cell cycle (Kozubek

et al., 1997). Studies have also revealed the close connection between chromosomal stability and DNA double-stranded breakages (DSB) (Richards, 2001; van Gent *et al*., 2001; Freudenreich, 2007). In eukaryotic cells, chromosomal organization is built up from millions of repeating units – nucleosomes. A nucleosome is composed of a core region and a linker region (Figure I-3a). A nucleosome core is made of averagely 147 base pairs (bp) of double stranded DNA (core DNA) wrapping around histone octamers by 1.7 times. A linker region is averagely 54 bp of linear DNA connecting two adjacent nucleosome cores (Figure I-3b, Davey *et al.*, 2002). Core DNA is subjected to substantial and uneven bending (Luger *et al.*, 1997). DNA binding affinity along the superhelix location (wrapping location along the histone core) is also uneven (Luger and Richmond, 1998), implying an uneven stress distribution along the core DNA. Crystallographic study (Schalch *et al.*, 2005) has resolved the structure of a tetranucleosome array, compacted by four nucleosomes in a zigzag fashion. It implies that each chromosome is compressed from a chromatin fiber of stacked nucleosome arrays (Figure I-4). Therefore, nucleosome positions (histone core binding sites) pinpoint the first feature of chromosome structure. Nucleosome positions have been thought to play important roles in regulating gene expression (Segal *et al.*, 2006; Schones *et al.*, 2008). The preferred nucleosome positions, if any, are thought to be determined by DNA flexibilities and preoccupied DNA binding proteins (Alberts *et al.*, 2003a). Different nucleotide compositions may affect the level of DNA bending, which significantly influence nucleosome positioning (Drew and Travers, 1985; Widom, 2001).

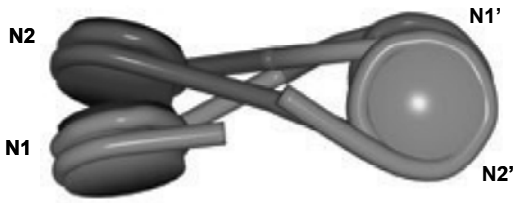


Figure I-4. Tetranucleosome array. Two nucleosome stacks (N1-N2 stack and N1'-N2' stack) are rotated by an angle and adjacent nucleosomes along the nucleotides are connected by a linker section of DNA. N: nucleosome.

Theoretical and experimental methods have been developed to predict or map nucleosome positions *in vitro* and *in vivo*. Predictive tools such as the RECON web server (Levitsky, 2004) and a bioinformatics prediction (Segal *et al*., 2006) are based on the occurrence frequencies of different dinucleotides. Other predictive methods are based on selected aspects of DNA structure and/or deformability (Sivolob and Khrapunov, 1995; Kiyama and Trifonov. 2002). Most predictive methods have not considered the effect of preoccupied DNA binding proteins that determined available DNA space, much less the effect of possible time sequences in which nucleosomes are assembled along the gene. On the other hand, traditional nucleosome mapping methods involve radioactive Southern hybridization and are time-demanding (Thoma, 1996; Flaus *et al.*, 1996; Wellinger *et al.*, 1998). Recent mapping methods with tiling microarrays have higher throughput (Dennis *et al.*, 2007; Schones *et al.*, 2008), but they are pricey and insufficient for nucleotide-level resolution. Together, it is in desire to overcome current weakness in predictive methods, and to balance cost and quality in experimental methods. Nevertheless, nucleosome mapping studies have revealed that nucleosome positions are generally dynamic (Pennings *et al.*, 1991; Luger, 2003; Li and Widom, 2004), but in actively expressed genes nucleosome positions are found

to be highly stable (Schones *et al.*, 2008). A study shows that *BCR* is ubiquitously expressed among many human cell types including hematopoietic cells (Collins *et al.*, 1987). It is thus possible that nucleosome positions in *BCR* are relatively stable when DNA breakages occur. It implies chromosome structure of *BCR* may play a role in the mechanism for DNA breakages in CML.

Major chromatin assembly is coupled to DNA replication (Gasser *et al.*, 1996; Krude, 1999; Lucchini *et al.*, 2001). Nucleosomes are disassembled in front of the replication fork and reassembled behind the fork (Sogo *et al.*, 1986; Gruss *et al.*, 1993; Gasser *et al*., 1996; Krude, 1999; Groth *et al.*, 2007). The unwrapping of nucleosomes gives parental DNA the access to the replication machinery. The duplex parental DNA is then separated by DNA helicases, a leading enzyme in the machinery, into two single stranded structures. Each parental strand is replicated by DNA polymerases into a daughter duplex DNA, which in turn rewrapped into nucleosomes (Tyler, 2002). Biochemical studies have proposed stepwise mechanisms of nucleosome disassembly and assembly, respectively (Levchenko *et al.*, 2005; Mazurkiewicz *et al.*, 2006). How chromatins are assembled on replicating DNA *in vitro* has also been proposed (Almouzni *et al.*, 1990; Smith and Stillman, 1991). It is known that as the replication fork advances to unwind the parental DNA, positive torsional stress is generated in front of the fork and may diffuse to the daughter DNA behind the fork (Sogo *et al*., 1999; Peter *et al.*, 1998; Postow *et al.*, 1999). Also, single-stranded DNA is exposed right behind the fork. Studies have proposed mechanisms to connect defects in DNA replication to chromosomal damages (Lemoine *et al.*, 2005; Freudenreich, 2007).

However, the detail changes of structure and topology on replicating DNA and how this is coordinated with replication and chromatin assembly were not addressed. These changes may cause stresses or damages on replicating chromatins. Therefore, an illustrative model of this process is in desire to find out possible mechanisms for DNA breakage.

This study aims to find connections between replicating chromosome structure and DNA breakage in M-*BCR* since *BCR* replication was suspected the occasion of chromosome translocation. DNA breakpoints are well characterized in CML, so this study takes advantage of this knowledge and plans to resolve chromosomal feature of each breakpoint. A stepwise nucleosome disassembly/reassembly model at the replication fork is then proposed to elaborate detail chromosomal structure changes. By incorporating the chromosomal features of M-*BCR* breakpoints into this model, the goal of this study is to propose detail mechanisms for DNA breakage in CML. Mechanical stress and structure dynamics of the chromosome in breakpoint regions are analyzed in the mechanisms. The knowledge gained from this study should provide better understanding of how chromosomal structure may play a role in DNA breakage and chromosome translocation, and offer more insight in the oncology of all cancers.

Chapter 1 Nucleosomal Organization of M-*BCR*

1.1 Rationale

CML is unique in that many DNA breakpoints in CML were previously mapped to the nucleotide level (Heisterkamp *et al.*, 1985; Chissoe *et al.*, 1995), so this study would like to take advantage of this finding. Both theoretical and experimental methods would be applied to resolve the exact chromosomal feature of each breakpoint, specifically the nucleosomal organization. The results may provide answers to questions such as: Is there a fixed chromosomal feature for each breakpoint? Do these breakpoints share a common feature in their nucleosomal organization, such as in the core or in the linker regions? If they are all distributed in the core region, do they share a particular superhelix position (such as the dyad or inner/out minor groove)? If so, could we find a mechanism from this common feature to account for DNA breakage? The nucleosomal organization of breakpoints also reveals whether nucleosome assembly is random or coupled with DNA replication, which is crucial for DNA breakage in CML. In short, the nucleosomal organization of each breakpoint will pave ways for analysis of the mechanical stress and elucidation of DNA breakage mechanisms.

Current theoretical algorithms predict nucleosome positions in a simultaneous manner. Independent nucleosome positions are predicted all at once for a given sequence. In fact, nucleosome assembly is not an isolated event but tightly coupled to DNA replication (Russev *et al.*, 1982; Fotedar *et al.*, 1989; Gasser *et al.*, 1996; Krude,

1999; Lucchini *et al.*, 2001). Behind the replication fork, $(H3\text{-}H4)_2$ tetramer complexes are able to recognize the same nucleosome positioning signals as the intact nucleosomes (Dong and van Holde, 1991; Hayes *et al.*, 1991), so nucleosome positions are likely to be conserved through replications. As the replication initiates from the origin and as the fork advances, nucleosomes are assembled one after another along the DNA right behind the fork (Figure 1-1). Consequently, not only the direction of nucleosome assembly follows that of replication, the available space for new nucleosome formation is also constrained by both the upstream, newly formed nucleosome cores and the downstream replication fork.

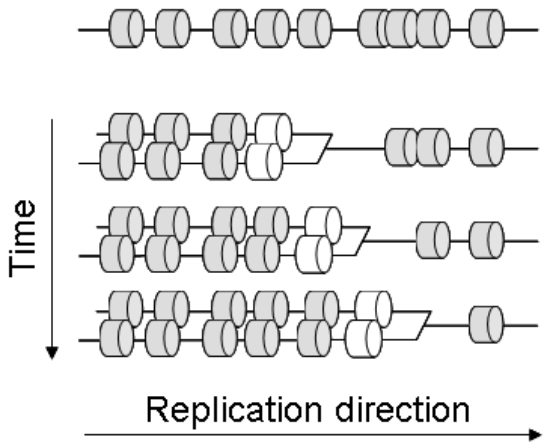


Figure 1-1. Replication-coupled nucleosome assembly. This figure is modified from Lucchini *et al.*, 2001. Nucleosomes are drawn as short cylinders. DNA is drawn as a line. A section of un-replicated chromatin is shown on the top. A time course of this chromatin during replication is shown on the bottom as three time steps. Newly formed nucleosomes in each time step are shown in white. The drawing is not to scale.

To improve current predictive algorithms and gain more insight into nucleosome positions in M-*BCR*, two key factors that take place in DNA replication, the direction of nucleosome assembly and pre-occupied DNA binding proteins

(presumably histones), would be incorporated into a modified algorithm. It is hoped to better mimic the nucleosome assembly *in vivo*. This new algorithm is proposed as "replication-directed" (RD) algorithm since the temporal and spatial sequences of nucleosome assembly are directed by replication. Having nucleosome positions accurately predicted by this algorithm, we not only can verify the coupling of replication in nucleosome assembly, but also can offer more perspectives for the upcoming prediction algorithms in the future. Moreover, the verified RD algorithm connects DNA replication and DNA structural changes, opening up a mechanism to explain previous observations that *BCR* replication coincided with chromosome translocation in CML.

1.2 Specific Aim

The specific aim here is to resolve the nucleosomal organization of previously mapped breakpoints on M-*BCR* in order to obtain the distribution of mechanical stress each breakpoint may experience and to find connection between nucleosomal organization and DNA replication.

1.3 Experimental Design

As mentioned above, nucleosome positioning mostly depends on intrinsic DNA flexibility and DNA binding proteins. It takes energy for DNA to bend and fold into nucleosome cores (Camerini-Otero and Felsenfeld, 1977). Free energy of core DNA bending (bending energy) depends on its sequence composition and has been

used to successfully predict *in vitro* nucleosome positions (Sivolob and Khrapunov, 1995). Therefore, to evaluate the feasibility of nucleosome formation in M-*BCR*, bending energy was chosen as an index to predict preferred nucleosome positions. To verify the prediction, established methods were adapted to experimentally map nucleosome positions in M-*BCR*. For a relatively small scale mapping in this study which required the quality of high resolution, adapter ligation-mediated PCR (Pfeifer *et al.*, 1993; Teng *et al.*, 2001) and automated sequencing was chosen to map nucleosome positions precisely. At the end, to verify the connection between nucleosome positioning and DNA replication, these two events were coupled in a modified algorithm to improve the prediction accuracy of nucleosome positions.

1.4 Materials and Methods

1.4.1 Human Genomic Sequence

The human genomic sequence of M-*BCR* (HGNC ID: 1014) was obtained from Ensembl database website (http://www.ensembl.org/Homo_sapiens/Info/Index).

1.4.2 Calculation of Core DNA Bending Energy in M-*BCR*

The bending energy of DNA around a histone core was calculated based on the algorithm from Sivolob and Khrapunov, 1995. The free energy ΔG is given by:

$$\Delta G = \left(kTd / 2R^2\right)\sum_{i=1}^{N} p_{Li} \qquad (1)$$

where k is the Boltzmann's constant (1.38×10^{-23} J/K), T is the absolute temperature (310 K), d is the distance between neighboring base-pair (0.34 nm), R is the radius of the curvature folding into a nucleosome (4.3 nm from Richmond *et al.*, 1984), and N is the number of base-pair steps in the core DNA ($N = 146$). The subscript i refers to the individual base-pair step in a DNA section with N steps, and p_{Li} is the persistence length (in nm) for a given base-pair step that occurs at the i^{th} position. The experimentally determined persistence lengths correlate well with the melting temperature and thus are sequence-dependent (for more detail, see Sivolob and Khrapunov, 1995). The calculation was programmed in MATLAB (The MathWorks Inc., Natick, MA).

1.4.3 Prediction of Nucleosome Positions with MC algorithm

An established Monte-Carlo (MC) algorithm (Sivolob and Khrapunov, 1995) was adapted in this study for nucleosome position predictions. Briefly, the core DNA bending energy was used as a preference index for nucleosome formation. At the beginning of each prediction, 14 nucleosomes were given to be placed in the 2,800 bp sequence covering M-*BCR* (from the 109,369th bp in *BCR*). Each nucleosome was randomly placed within one of the non-overlapping 200 bp sections. Starting from the initial configuration, a randomly chosen nucleosome was then displaced either up- or downstream by a random distance in each trial. The trials proceeded until equilibrium was reached so that the total bending energy of all positioned nucleosomes stabilized at a minimum. The total number of trials was 20,000 while typical equilibrium was

reached below 10,000 trials (Sivolob and Khrapunov, 1995). The simulation was programmed in MATLAB.

1.4.4 Design of RD Algorithm

Mimicking what happens *in vivo*, the position of the first nucleosome core was predicted immediately downstream from the replication origin (RO) within a limited length of sequence (a nucleosome-free window). The consensus sequence of human replication origin is WAWTTDDWWWDHWGWHMAWTT, where W = A or T; D = A, G or T; H = A, C or T, and M = A or C (Dobbs *et al.*, 1994). It was used to search for replication origins in the predicted genes. From there on, positions of the following nucleosomes are predicted one after another in a 5' to 3' direction, each in a nucleosome-free window that was moving in the same direction. Within each window, the preferred starting position of the nucleosome core was chosen to be the local minimum of bending energy where a nucleosome core was most easily positioned. There was no need for each nucleosome to search for an optimal position, which was the case in the MC algorithm, since there was only one nucleosome to be placed in that window at a given time. Therefore, whichever position was energetically favored would be the predicted position in the window.

The nucleosome-free window was initially set downstream from the replication origin, extending from the first available base pair to a designated distance (see below). After the first and each nucleosome prediction thereafter, the window moved to the first base pair downstream to the newly predicted nucleosome position. For example,

if the first nucleosome core was predicted to occupy the base pair 1 – 147 downstream from the origin, then the window for the second nucleosome prediction would start from the base pair 148. The designated length of the window remained the same throughout the region of predictions. According to previous studies (Sogo *et al.*, 1986; Gasser *et al.*, 1996), the distance from the fork to the first nucleosome core was about 255 bp (225 ± 145 in leading strands and 285 ± 120 in lagging strands). Three nucleosome-free window lengths were therefore chosen: 225 bp, 255 bp, and 285 bp to conduct the predictions. To observe extreme conditions, short and long window lengths (80 bp and 405 bp) were also used for predictions. The predictions were programmed in MATLAB.

1.4.5 Prediction of Nucleosome Positions with RD Algorithm

For a given region of the gene, we first used the consensus sequence to search for replication origin(s) near the nucleosomes. If multiple origins were found near the region of interest, only the one in the shortest distance to nucleosomes on either end of the region was used as the start site of predictions. The prediction results from indicated window lengths were aligned with the mapped nucleosome and other protein binding sites, if any, along the base position defined in the references.

1.4.6 HL-60 Cell Culture

A well-studied leukemogenesis model, HL-60 cell line (Gallagher *et al*., 1979; Breitman *et al.*, 1980; Collins, 1987) was used as the source of leukocyte

chromosomes. Cells (ATCC, Manassas, VA) were maintained as described (Yalcintepe *et al.*, 2005). Briefly, cells were cultured in RPMI 1640 medium (Mediatech, Manassas, VA) containing 10% FBS (Mediatech, Manassas, VA), with 100 U/ml penicillin, 100 μg/ml streptomycin (Invitrogen, Carlsbad, CA) and 2 mM L-glutamine (Invitrogen, Carlsbad, CA). All cultures were grown at 37°C in a humidified incubator (Thermo Series II, Waltham, MA) purged with 5% CO_2 / 95% air.

1.4.7 Isolation of Nuclei

Nuclei of HL-60 cells were isolated as described previously (Meyer and Radsak, 2000) with minor modifications. Cells (1.2×10^7) were washed once with 1 ml of cold PBS (138 mM NaCl, 2.7 mM KCl, 1.76 mM KH_2PO_4, 10.1 mM Na_2HPO_4, pH 7.4), pelleted (150 ×*g* for 5 min at 20°C) and resuspended in 2 ml of cold TKM buffer (50 mM Tris–HCl, pH 7.5, 25 mM KCl, 5 mM $MgCl_2$) at 4°C and kept for 5 min prior to the addition of 2 ml of TKM containing 1% NP-40. The mixture was incubated with occasional vortexing for 5 min on ice. The nuclei pellet was then sedimented at 2100 ×*g* for 20 min through a 2 ml cushion of 1.62 M sucrose in TKM. The integrity of extracted nuclei was examined under an optical microscope (Zeiss Axiovert 25, Thornwood, NY) with 100 × magnifications.

1.4.8 Limited Micrococcal Nuclease Digestion

The micrococcal nuclease (MNase) digestion procedure was modified from

Yalcintepe *et al.*, 2005. Nuclei pellet was gently re-suspended in 392 µl of MNase digestion buffer (0.1 M NaCl, 10 mM Tris-HCl pH 8.0, and 1 mM EDTA) prior to the addition of 8 µl of 0.1 M $CaCl_2$ (final concentration 2 mM). The nuclei was gently resuspended and incubated at 37^{o}C for 5 min. A 100 µl aliquot from the nuclei solution were added to a 1.5 ml microcentrifuge tube containing 1 µl of 1, 2, or 4 units of MNase (USB, Cleveland, OH), or 1 µl of distilled water. A 100 µl aliquot of naked DNA was added to a separate tube containing 4 units of MNase as a control. Each digestion was incubated at 37^{o}C for 5 min. Reaction was then stopped by adding to each sample 100 µl of 25 mM EDTA, 1% SDS, and 2 µl of 20 mg/ml proteinase K (Fermentas, Glen Burnie, MA) immediately prior to an incubation for 16 h at 55^{o}C. Samples were then phenol/chloroform extracted and ethanol precipitated. Finally, each sample was dissolved in 100 µl of TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0) and quantified by a spectrophotometer (Beckman DU640, Fullerton, CA) with a ratio of 260/280 at ~1.8. The final DNA concentration for each sample was usually ~200 ng/µl. To monitor the quality of digestion, digestion products were resolved by a 1.2% agarose gel electrophoresis (Gibco, Carlsbad, CA).

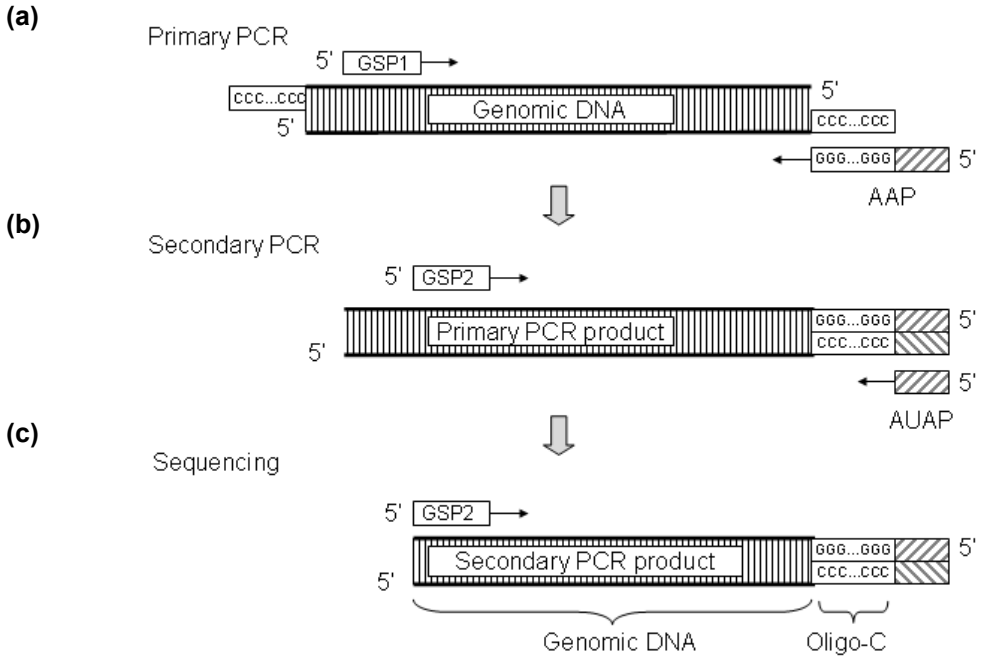


Figure 1-2. The adapter ligation-mediated PCR procedure. (a) Primary PCR. Multiple dCTP are ligated to the 3' end of the genomic DNA as primary PCR templates. A primary gene specific primer (GSP1) and an abridged anchor primer (AAP) are used to amplify the template in the primary PCR. (b) Secondary PCR. A secondary gene specific primer (GSP2) and an abridged universal amplification primer (AUAP) are used to amplify the primary PCR product. (c) Final product for sequencing. The secondary PCR product is sequenced by using GSP2 as a primer. The final sequence includes a 5' genomic DNA and a 3' oligo-C section. Small arrows indicate the direction of amplification.

1.4.9 Adapter Ligation-mediated PCR and Sequencing

After extraction from MNase digestion, about 15 to 20 nucleotides of dCTP (oligo-C) were added to the 3' ends of core DNA fragments using terminal deoxynucleotidyl transferase (USB, Cleveland, OH), according to the manufacturer's manual. Oligo-C attached MNase digested fragments were directly served as templates for a primary PCR (Figure 1-2a), where a primary gene-specific primer (GSP1), an abridged anchor primer (AAP: 5'-GGCCAGGCGTCGACTAGTACGGGIIGGGIIG-GGIIG-3'), Accuprime Taq polymerase (Invitrogen, Carlsbad, CA) and the corresponding buffer were included. A secondary PCR (Figure 1-2b) then followed

with the same setup except: (1) 10% diluted primary PCR products was used as templates; (2) a nested secondary gene-specific primer (GSP2) replaced GSP1, and (3) an abridged universal amplification primer (AUAP: 5'-GGCCACGCGTCGACTAGT-AC-3') replaced AAP. The primary and secondary PCR programs were identical according to the manufacturer's manual (Invitrogen, Carlsbad, CA). The secondary PCR products were resolved on a 1.5% agarose gel by electrophoresis. The smallest sharp band in each reaction was excised from the gel, extracted (Qiagen, Valencia, CA) and sequenced (Fisher SeqWright, Houston, TX). GSP2 was used as the sequencing primer (Figure 1-2c). The results were analyzed by FinchTV (Geospiza Inc., Seattle, WA). A regular sequencing result contained a genomic sequence section at the 5' end and an oligo-C section at the 3' end. The junction of these two sections indicates the cutting site of MNase on the genomic sequence, preferentially the boundary of a core region. To be statistically significant, each reaction was repeated for over 10 times, and a nucleosome boundary was considered significant if its occurrence is 1 standard deviation over the average occurrence. The amplified region covered a 2.6kb section of M-*BCR*, where 27 breakpoints have been reported.

All gene specific primers were designed and optimized using FastPCR (Institute of Biotechnology, Helsinki, Finland), and searched against the public database (http://www.ncbi.nlm.nih. gov) to confirm unique amplification products. GSP2 were designed to be nested to the corresponding GSP1, and the overlapping sequences between the two primers were minimized in order to increase specificity. We intended to design all primers from the core regions which were preserved after

MNase digestion. Since the nucleosome positions in the M-*BCR* has not been fully resolved, the first set of primers were chosen at a location about 200 base pairs (the length of one nucleosome) upstream to the most 5' breakpoint (ID 0319129), in the hope that at least one nucleosome can be mapped upstream to the first breakpoint. The rest of the primers were then designed based on the nucleosome mapped immediately upstream. Upstream forward primers were used to reveal the 3' boundary within the nucleosome core (Figure 1-4b), while the immediate downstream reverse primers to reveal the 5' boundary of the same nucleosome core. The primer sequences are listed in Table 1-1.

Table 1-1. Primers for PCR amplifications in M-*BCR*. Primers are named by its direction of amplification (F: forward; R: reverse) followed by their starting locations downstream from 109369th bp in *BCR*. FP-84 and FP-46 are two exceptions since they are upstream to 109369th bp in *BCR* by the indicated distance.

GSP1	GSP2	5' to 3' sequence
FP-84		TGCACCTCTTTTCCAACCTCCCAGG
	FP-46	TCCCTGACATCCGTGGAGCTGCAGA
FP120		CCTGTTAGCACTTTTGATGGGACT
	FP139	GGACTAGTGGACTTTGGTTCAGAAG
FP502		ACCCCCTCTGCTGTCCTTGGAACC
	FP531	ACACTTCGAGTCACTGGTTTGCCTG
FP703		TGTCACCTGCCTCCCTTTCCCGGG
	FP724	GGGACAACAGAAGCTGACCTC
FP901		CATCGGGCAGGGTGTGGGGAAACAG
	FP928	AGGTTGTTCAGATGACCACGGGAC
FP1206		CTCTGTCGAGCTGGATGGATACTAC
	FP1255	CCTTTCCCTCTAAGTGGGGGTCTCC
FP1755		GGTGAGAGCAGTGTCGTG
	FP1777	GACTGTGGTGCTGTTTGCGCTCAC
FP1976		TGTCTGTGAGCAATACAGCGTGAC
	FP1990	ACAGCGTGACACCCTACGCTGC
FP2184		AAGCCAGAAACCGTGGTCTG
	FP2222	TGCCATTCTCCATCAGTGAGGC
	RP25	CGGAATGCTGTGGACAGTCTGGAGT
RP49		CCCACCTTCCTTATTGATGGTCAGC
	RP178	CAAGCATAGCTCTTCCTTCTGAACC
RP192		ACAAGAGGCCCTAACAAGCATAGC
	RP255	ACCCAGCAGCAGCCACTCAGGGAGA
RP272		CGTGCATCTCCTCAACCACCCAGCA
	RP600	TGATTTCTGTCTTGGGGATCTCAGG
RP640		CAAGGTAAGCCAGTGTCATGGGCCA
	RP786	CCATAGAGCCCCGGAGACTCATC
RP816		GTGGCTGAGTGGACGATGACATT
	RP1023	CCGGGAAAAACAATTCCACACCCAG
RP1047		CTAGGGGAGAGGGCAGAGGCCACTC
	RP1310	CTTCACTGTTCTGACAGCTCCAG
RP1335		ACAGTGCAACTCATGTGTTACCAGC
	RP1897	GCCCAGGAGAAAGATTATGTTATGAAGGTGC
RP1915		AGGCAGCCAGAGACAGGGGCCCA
	RP2038	AAGGAGAGACAAGCCCGGGACCAC
RP2057		AGAAAGGTAACAGGGAGGCAAGG
	RP2259	CAGAGATGACTAAGAAGCCTCACTG
RP2271		TGGCCAGGCAGCCAGAGATG

1.4.10 Precise Mapping of Nucleosome Positions for Breakpoints

To map nucleosome positions from the core boundaries, we followed recent findings such as: (1) regular nucleosome core ranges widely in length, from 90 to 200 bp (Gasser *et al*., 1996); (2) the average core region was 147 bp long, and the linker regions may vary in their lengths (Thoma, 1996; Alberts *et al*., 2003b); (3) about 90 bp around the dyad was the strong binding site of each nucleosome (Luger and Richmond, 1998). We also followed general observations: (1) the core regions were bracketed by a pair of upstream boundary and downstream boundary; (2) if the boundary separation was shorter than 90 bp, we assumed that this nucleosome core had more dynamic behavior than the others, or there might exist multiple positions for this nucleosome; (3) if the separation was between 90 bp and 147 bp, we assumed that there was only one nucleosome core between the boundaries; (4) if the separation was longer than 200 bp, we assumed either more than one set of histone octamers constantly bind to this region, or the binding of other proteins (such as H1/H5) protected the DNA from MNase digestion. Once the nucleosome core regions were mapped, DNA breakpoints from the literatures were aligned and superimposed to the mapping so that their individual positions in the core region were revealed.

1.4.11 Statistic Comparisons of Nucleosome Positions from Different Methods

To verify whether nucleosome positions from either method (MC, RD, or experimental mapping) significantly preferred at the sites with low bending energy, a

binomial test was performed. The 2600-bp M-BCR region was evenly divided to 13 200-bp subregions, in each of which the site of minimal bending energy was determined by direct comparison of the bending energy value at each site. A nucleosome core was considered "close" to the minimal bending energy site if the distance between the 5' boundary of that core region and the closest site of minimal bending energy was within 10 bp. The probability of having a nucleosome randomly start at the site within 10 bp of a site of minimal bending energy (p_r) was 20/200 or 0.1 since an average core region spans 200 bp. The *P* value is the probability that one would see the observed number of nucleosomes close to sites with minimal bending energy, or more given random chance, and is calculated as:

$$P = \sum_{i=j}^{N} \frac{N!}{i!(N-i)!} p_r^i \left(1 - p_r\right)^{N-i} \qquad (2)$$

where *N* is the total number of nucleosomes (either mapped or predicted), *j* is the number of nucleosomes close to the local sites of minimal bending energy, and p_r is the random probability (0.1). The calculation is programmed in MATLAB.

To determine whether the mapped or predicted nucleosome positions were uniformly distributed in M-*BCR*, a t-test was performed. A uniform nucleosome distribution was virtually created by placing the same number of mapped or predicted nucleosomes evenly within M-*BCR*. Then, the 5' boundaries of these uniformly distributed nucleosomes were compared against those of the predicted or mapped nucleosomes in a t-test in EXCEL. The *P* value is the probability that one would see the predicted or mapped nucleosomes distribute uniformly within M-*BCR*.

To determine whether one prediction is significantly more accurate than the

other one, two indices were evaluated: the number and the positions of nucleosomes between the prediction and experimental mapping. The number of nucleosomes was counted by eye after each prediction and the difference in the number of nucleosomes between the prediction and experimental mapping was determined. The distance between one predicted and one mapped nucleosome is calculated by the distance of each predicted nucleosome to that of the closest experimentally mapped nucleosome. A t-test was then performed to compare the distances from one algorithm and those from another algorithm to determine the statistic significance. The *P* value from the t-test is a probability that one would see these two algorithms predict the same in distances to the experimentally mapped nucleosomes.

1.5 Results and Discussion

1.5.1 Nucleosome Positions Predicted by MC Algorithm in *M-BCR*

The nucleosome positions in M-*BCR* were predicted with the MC algorithm (Sivolob and Khrapunov, 1995) for 20 times. More predictions did not alter the result. The bending energy and cumulative prediction result was shown in Figure 1-3. In each prediction, the MC algorithm keeps predicting 14 nucleosomes (from M0 to M13) simultaneously at random sites of M-*BCR* until equilibrium of minimal total bending energy was reached. The minimal energy was usually reached within 10,000 trials (Figure 1-4). The MC algorithm therefore found the equilibrated nucleosome positions in a simultaneous and random fashion.

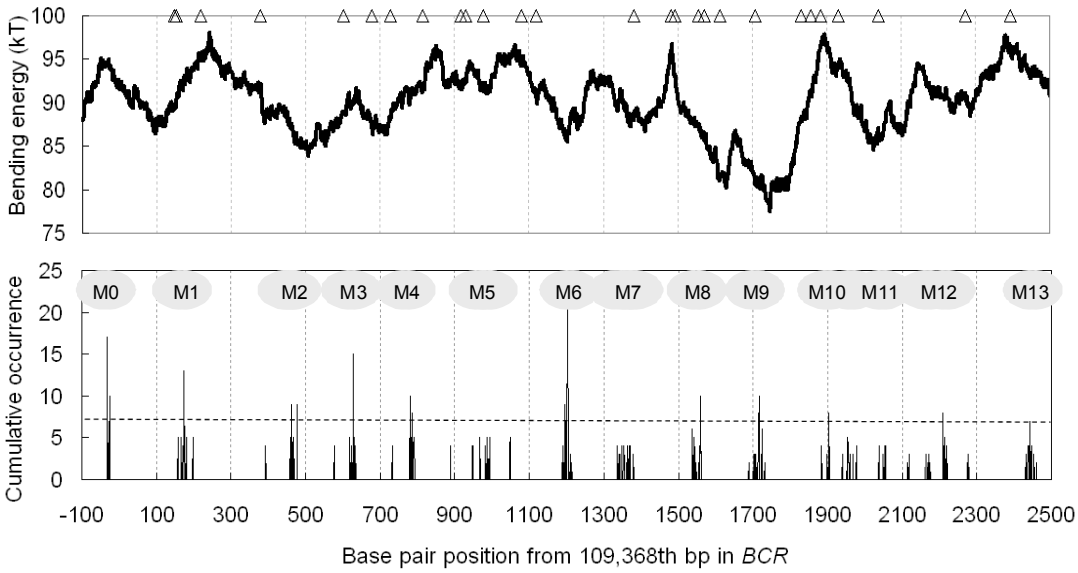


Figure 1-3. Bending energy and predicted nucleosome positions in M-*BCR* using MC algorithm. Bending energy (top) and nucleosome positions from 20 predictions (bottom) are plot against the base pair positions. DNA breakpoints are shown as open triangles on top of the bending energy trend (top). The most abundant nucleosome positions are indicated as shaded ovals (bottom) and named from upstream (M0) to downstream (M13). A dashed line represents one standard deviation from the average non-zero occurrence.

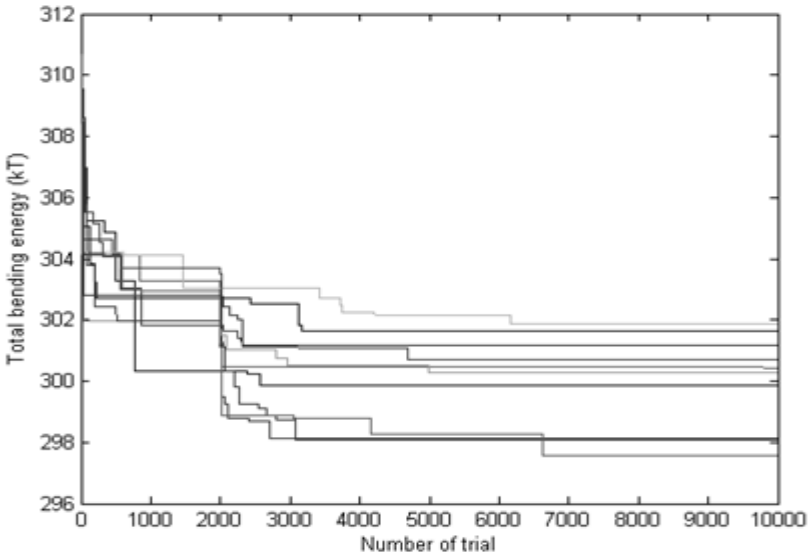


Figure 1-4. Converged total bending energy in MC predictions. Total bending energy from ten independent predictions was plotted against the number of trials. The total bending energy converged within 10,000 trials.

Four of the 14 nucleosomes were predicted within 10 bp to the local sites of minimal bending energy ($P = 0.16$), meaning that bending energy somewhat dictated the positioning of nucleosomes. The expression of bending energy shows obviously that the bending energy depends on the nucleotide composition of a DNA section. When there was only one nucleosome to be placed in a DNA section, the position on the DNA with minimal bending energy was the preferred starting site of a nucleosome (Sivolob and Khrapunov, 1995). However this may not be the case in a multiple nucleosome region. The prediction result showed that most nucleosomes (except M5, M10~M12) had rather isolated positions, and only a few (M0, M1, M2, M9) appeared to start from the local minimum of bending energy. Twenty three out of 27 breakpoints were predicted to be within the nucleosome core regions. This seemed to imply that breakpoints preferred to occur within the regions that were supposed to wrap around histone cores. However, we know that core DNA should be protected by the histone core, and the MC algorithm on the prediction of a long DNA molecule has not been experimentally verified. Therefore we would like to use experimental methods to precisely map and verify nucleosome positions in M-*BCR*.

1.5.2 Experimentally Mapped Nucleosome Positions

1.5.2.1 Limited MNase Digestion on HL-60 Chromosomes

The MNase digestion pattern of HL-60 DNA was shown in Figure 1-5a. While the digestion products of chromosomal DNA were distributed as distinct ladders, those from the naked DNA exhibited a more homogeneous pattern (lane D4). Each ladder pattern had a pitch of about 200 bp in length, consistent with the length of a nucleosome (147 bases of core DNA plus 54 bases of linker DNA). As the amount of MNase increased from 1 to 4 units, longer fragments (multi-nucleosomes) vanished away and shorter fragments (oligo-nucleosomes) stood out, implying that more linker regions have been digested while core regions were still preserved. Obviously, 4-unit MNase digestion enriched the population of oligo-nucleosomes. For the purpose of this study, 4-unit MNase digest should be more suitable than 1-unit MNase digest in terms of PCR template composition since the lengths of the templates were more uniform, so we chose the 4-unit MNase digest as our PCR templates.

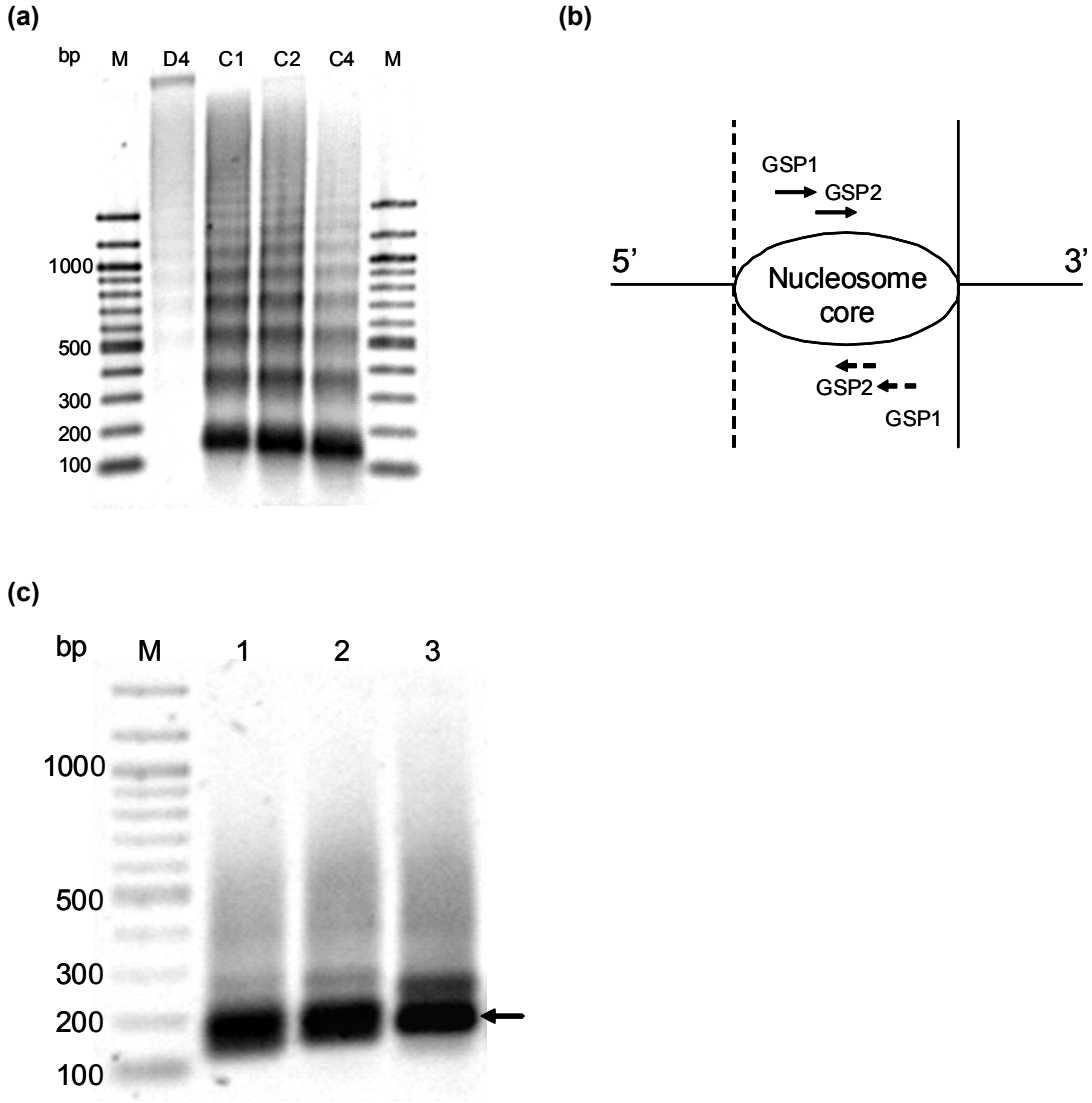


Figure 1-5. Nucleosome mapping in M-*BCR*. (a) Limited MNase digestion of HL-60 naked and chromosomal DNA. Lane M is a 100 bp DNA ladder (NEB, Ipswich, MA). Lane D4 is naked DNA digested by 4 units of MNase. Lanes C1 to C4 are chromosomal DNA digested by 1, 2, and 4 units of MNase, respectively. (b) Schematic design of nested PCR primers. A pair of upstream forward primers (solid arrows) maps the downstream boundary (solid line), and a pair of downstream reverse primers (dashed arrows) maps the upstream boundary (broken line) of the same nucleosome. (c) Nested PCR product from three independent reactions. An arrow indicates the extracted bands.

1.5.2.2 Nucleosomal Organization of M-*BCR*

We designed the first set of primers purposely about 200 base pairs upstream to the most 5' breakpoint (BO). In this way, we hoped at least one nucleosome can be

mapped before the most 5' breakpoint, and we could design the further downstream primers according to the mere upstream one. Two pairs of nested primers were used to map nucleosome boundaries for higher specificity (Figure 1-5b). The secondary PCR products were resolved by gel electrophoresis (Figure 1-5c). The distinct band was excised, extracted, and sequenced to reveal the boundaries of a nucleosome core. A typical sequencing chromatogram of a PCR product is shown in Figure 1-6. Each readout sequence usually consists of a stretch of genomic sequence at the 5' end and an oligo-C section at the 3' end. The boundary of a nucleosome core region was revealed at the junction.

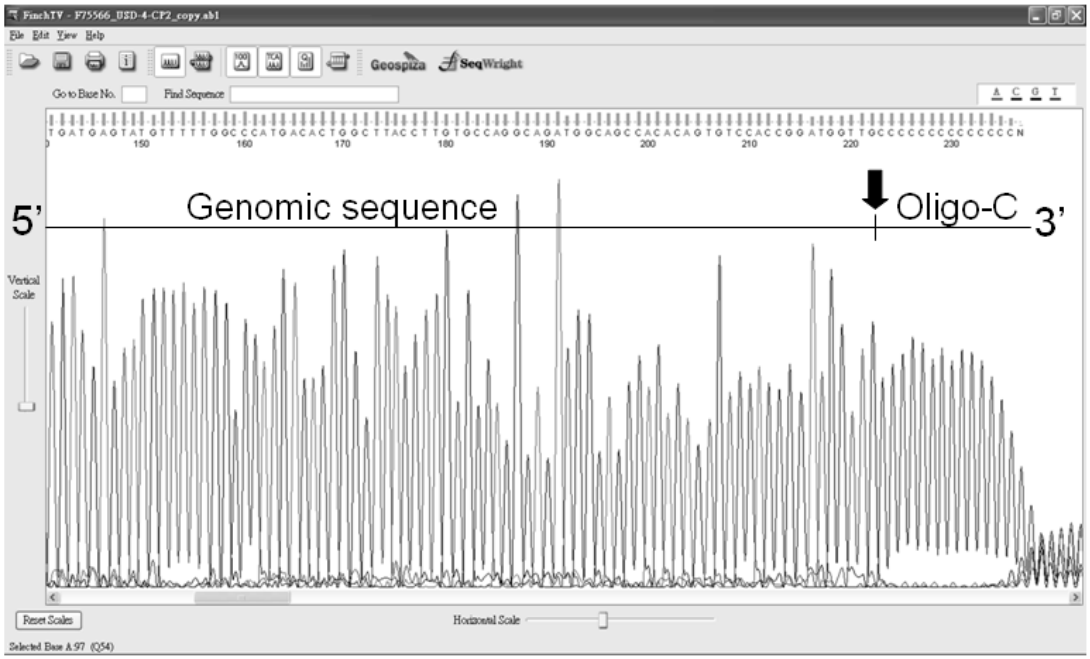


Figure 1-6. A typical sequencing result of nested PCR. This figure is generated by FinchTV. Peaks stand for sequenced nucleotides in a left (5') to right (3') direction. At the 3' end of each sequence is an oligo-C section. The nucleosome core boundary is at the junction between the genomic sequence and the oligo-C section (arrow), such as a G in this figure. (Colors not shown as suggested by the guideline for UCSD thesis.)

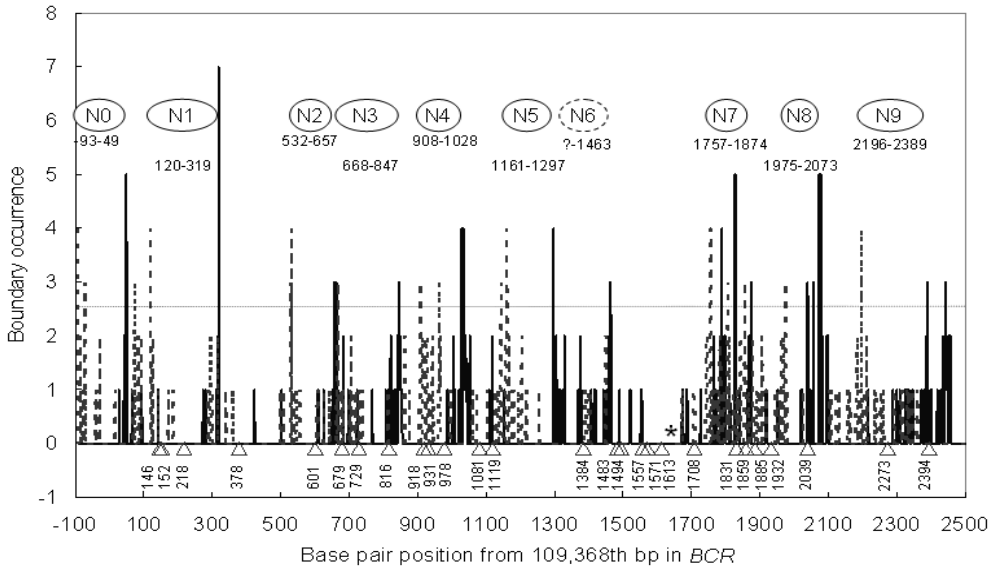


Figure 1-7. Nucleosome organization in M-*BCR*. The accumulated occurrences of core nucleosomal boundaries are plot against the base pair locations in *BCR* (from the 109,368th base pair). Solid vertical lines are downstream boundaries, and dotted vertical lines are upstream boundaries. DNA breakpoints are indicated as open triangles with precise locations. The significance cut-off (two standard deviations above the average) is marked as a horizontal dotted line. Numbered nucleosomes are shown as open ovals with the corresponding boundaries on the bottom (N6: dashed oval). *: 17-A region (see p. 54).

Experiments were repeated and all boundary occurrences were summarized in Figure 1-7. A cut-off was chosen as two standard deviations above the non-zero average occurrence to reveal significant boundaries. Ten nucleosomes were mapped within M-*BCR*. Nucleosomes were named from upstream to downstream as N0 to N9. Six of the 10 nucleosomes (N0, N1, N2, N3, N5, and N9) had unique and significant boundaries, meaning their positions were relatively stable. Other nucleosomes (N4, N6, N7, and N8) had multiple significant boundaries, suggesting they may have multiple preferred positions or their positions were more dynamic than others. Over 20 reactions with various primers have been attempted to map a possible nucleosome between N1 and N2, but only boundaries of N1 and N2 were significantly revealed, suggesting that there may not be a nucleosome core between N1 and N2. N6 has a

significant downstream boundary but no obvious upstream boundary after numerous mapping attempts. It means that N6 is highly dynamic in its position and a nucleosome core may move frequently between N5 and N6. In either case, N6 is not a stable core region and may be exposed from histone binding. By comparing the average lengths of core and linker DNA, one would expect that DNA breakpoints, if occurring purely by chance, should be reported three times more in the core regions than in the linker regions. Here, however, we found 14 of the 27 reported breakpoints in the core regions. This finding suggests that (1) DNA breakage may not be totally by chance or length-dependent; (2) a core region in the same length as a linker region would break less likely than the linker region. It implies that core DNA is protected, presumably by histone cores. Still, we can not rule out the possibility that this conclusion may have been misled by a small sampling number.

Five of the 10 mapped nucleosomes fell within 10 bp to the local sites of minimal bending energy ($P = 0.0016$), confirming that bending energy is a key factor to determine the preferred nucleosome positions. The average length of the mapped core regions was 145 bp, very close to that of a typical nucleosome core (147 bp) (Davey *et al.*, 2002). The lengths of individual core regions spread from 99 bp to 200 bp. The lower limit suggests a dynamic behavior of nucleosomes, so MNase may chew into the exposed DNA. The upper limit suggests the existence of a conformation larger than regular nucleosome, meaning either more than 1 histone core or additional proteins bound to the region and protected the DNA from being digested by MNase. Overall, the mapping revealed a native, first-level chromosomal organization of M-

BCR, and it met our expectation that nucleosomes have relatively stable positions in this region.

1.5.3 Comparison of Nucleosome Position between MC Prediction and Experimental Mapping

To compare the nucleosome position between the prediction and experimental mapping, we chose the dyad location of each mapped/predicted nucleosome core as an index. In those cases that multiple locations were predicted or mapped for particular nucleosomes, the dyad locations of the same nucleosome were averaged by their occurrences for the comparison. The comparison is shown in Figure 1-8. Pearson correlation tests and t-tests were applied to quantify the correlation and difference between the two results. The detail dyad positions and statistical comparison are shown in Table 1-2.

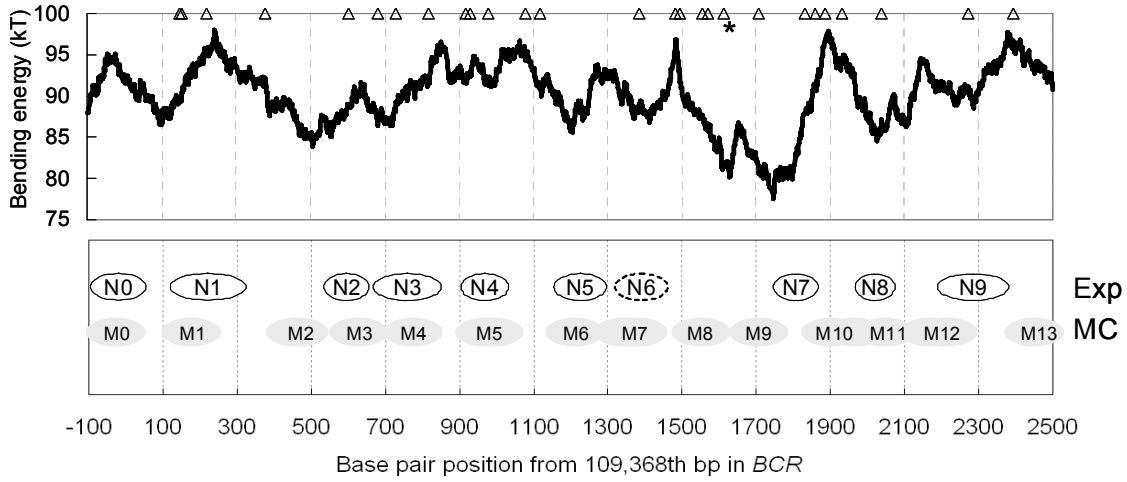


Figure 1-8. Comparison of nucleosome positions in M-*BCR* from MC prediction and experimental mapping. The top panel shows the bending energy trend and breakpoint locations (Δ). The bottom panel shows the experimental mapping (Exp, open ovals) and MC prediction (MC, shaded ovals) results. Nucleosome cores are represented as in Figure 1-3 and Figure 1-7. *: 17-A region (see p. 54).

Table 1-2. Comparison of predicted and experimentally mapped nucleosome positions in M-*BCR*. The dyad positions (in bp, from the 109368th bp in *BCR*) of nucleosomes predicted or mapped by different methods are listed. The Pearson correlation coefficient *R* values and the *t*-test *P* values are shown at the right.

MC prediction		Experimental mapping		*R*	*P*
Nucleosome	Dyad position	Nucleosome	Dyad position		
M0	-29	N0	-22		
M1	175	N1	220		
M2	460				
M3	622	N2	595	0.998	0.980
M4	782	N3	758		
M5	982	N4	968		
M6	1200	N5	1229		
M7	1357	N6	1390		
M8	1551				
M9	1713				
M10	1938	N7	1816		
M11	2066	N8	2024	0.998	0.880
M12	2205	N9	2293		
M13	2442				
Overall				0.997	0.711

The results from the two methods had excellent correlation ($R > 0.997$) and no significant difference ($P > 0.711$). The upstream nucleosome positions from the prediction, M0 to M7 except M2, matched well with those from the mapping, N0 to N6 ($P = 0.980$). The downstream nucleosomes M10 to M12 from the prediction had multiple positions and matched the mapping not as well ($P = 0.880$). While 14 nucleosomes were predicted, only 10 nucleosomes were experimentally mapped in M-*BCR*. M2, M8, M9, and M13 were predicted but not mapped. The prediction of N8 and N9 was obviously due to the low bending energy between 1500 and 1900 bp in M-*BCR*. However, nucleosomes seemed to be excluded between 1500 and 1700 bp according to the mapping. The reason for the missed prediction on M2 and M13 is not

clear, but may come from the principles MC predictions followed: (1) the number of core regions was pre-determined by the length of the whole region. One nucleosome was assigned every 200 bp; (2) the available DNA space for a nucleosome is always 200 bp no matter what level the bending energy is at in that region; (3) the prediction of each nucleosome core was independent from the other cores. After all, the random and simultaneous manner to predict nucleosome positions in MC algorithm did not mimic *in vivo* conditions where a pre-existing nucleosome may act as a boundary for the proceeding nucleosomes to position. The formation of nucleosomes should follow a certain temporal and spatial order along the DNA sequence. This discrimination motivated us to propose a RD algorithm to improve the accuracy.

1.5.4 Nucleosome Positions in M-*BCR* Predicted by RD Algorithm

In RD algorithm, nucleosome positions are coupled to DNA replication. The first nucleosome core was predicted within a limited length of sequence (a nucleosome-free window) immediately downstream from the replication origin (RO). This mimics the starting of nucleosome assembly during replication. From there on, the following nucleosome positions are predicted one after another in a 5' to 3' direction, each in a nucleosome-free window that was moving in the same direction. This mimics both the temporal sequence and spatial constraint of *in vivo* nucleosome assembly. Within each window, the preferred starting position of the nucleosome core was chosen to be the local minimum of bending energy since a nucleosome core was energetically most favored there.

The RD prediction was aligned with the MC prediction and experimentally mapped positions in Figure 1-9. The statistical comparison is shown in Table 1-3. Amazingly, the primary window lengths (225, 255, and 285 bp, see Materials and Methods, under 1.4.4) in the RD algorithm predicted 10 nucleosome cores, same number as that of the experimental results. There was little difference in the predicted positions among three primary window lengths, probably due to the small length differences (30 bp). A shorter window (80 bp) increased the number of predicted nucleosomes and a longer window (405 bp) reduced this number, since obviously the same region could fit fewer windows if the windows were longer.

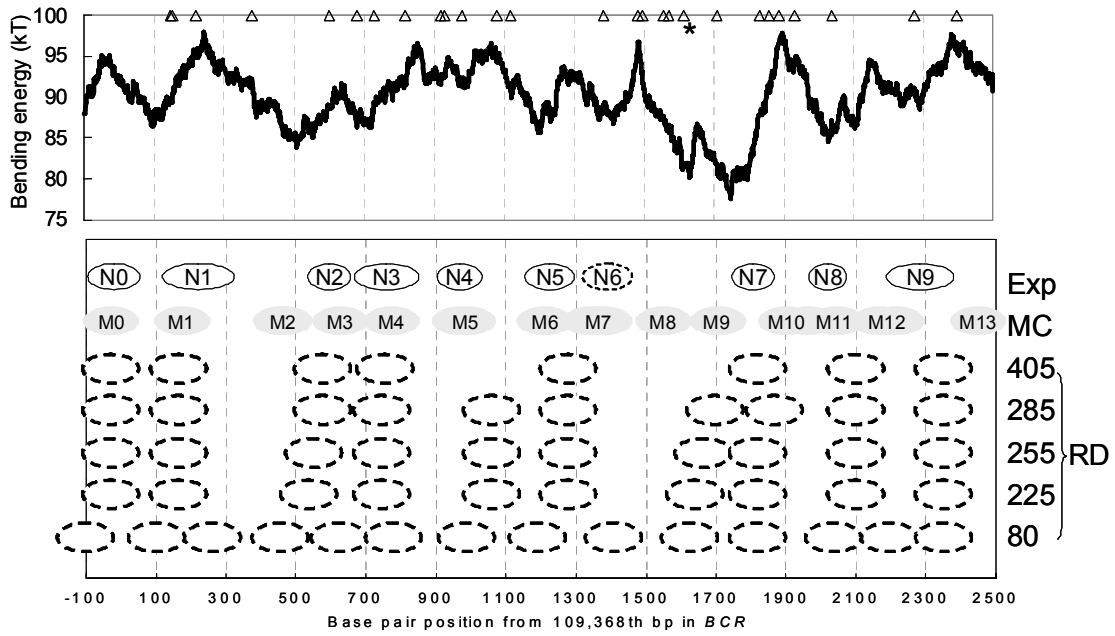


Figure 1-9. Predicted and experimentally mapped nucleosome positions in M-*BCR*. Nucleosome positions are plot against the base pair positions in *BCR*. Nucleosome positions from experiments (Exp: open ovals), MC prediction (MC: shaded ovals), and RD predictions with different window lengths (80 – 405 bp: broken ovals) are indicated to the right. Other legends are the same as in Figure 1-8.

Table 1-3. Comparison of predicted and experimentally mapped nucleosome positions in M-*BCR*. The dyad positions (in bp, from the 109,368th bp in *BCR*) of nucleosomes predicted or mapped by different methods are listed. There are five window lengths (in bp) in RD predictions. The linear regression *R* values and the *t*-test *P* values of prediction results against the experimental data are shown at the bottom.

Experimental mapping	MC prediction	RD prediction				
		80	225	255	285	405
-22	-29	-99	-27	-27	-27	-27
220	175	108	169	169	169	169
	460	262				
595	622	463	540	558	580	580
758	782	626	753	753	753	753
968	982	786	1062	1062	1062	
1229	1200	990	1277	1277	1277	1277
1390	1357	1195				
	1551	1410				
	1713	1635	1644	1668	1703	
1816	1938	1821	1821	1821	1869	1821
2024	2066	2047	2099	2099	2099	2099
2293	2205	2199	2360	2360	2360	2360
	2442	2360				
R	0.997	0.995	0.998	0.999	0.999	0.999
P	0.711	0.996	0.905	0.896	0.873	0.996

At least 8 of the 10 predicted nucleosomes fell within 10 bp to local sites of minimal bending energy ($P = 3.7 \times 10^{-7}$), showing that bending energy strongly dictated nucleosome positions in the RD algorithm. Both MC and RD algorithms showed excellent correlations ($R > 0.99$) with and no significant difference ($P > 0.87$ with RD; $P = 0.71$ with MC) to the experimental data. There is little difference between the MC and RD algorithms in terms of the distance between the predicted nucleosome and the mapped ones ($P \sim 0.4$), but obviously the RD algorithm predicted a more accurate number of nucleosomes than the MC algorithm. The fact that primary window lengths (225 – 285 bp) in RD algorithm predicted the same number of nucleosomes as the experimental data is consistent with previous finding (Gasser *et al.*, 1996; Krude, 1999) that nucleosome-free DNA in this size (225 – 285 bp) are likely

the available space for nucleosome formation right behind the fork. Overall, the difference between the two algorithms suggested that nucleosome assembly should follow spatial and temporal sequences in which DNA bound proteins act as boundaries, and new nucleosomes may be assembled along the sequence accordingly in a one after another fashion.

Even though the RD algorithm improved the accuracy in predicting nucleosome positions in M-*BCR* as compared to the MC algorithm, it is still not perfect for all genes of interest. Here we summarized the defects that remain in the RD algorithm. (1) The predicted positions may be affected by the window lengths we have chose. In fact, the distribution of nucleosome-free window lengths along the gene at replication has not been thoroughly investigated. It is possible that the window lengths are different in one region from another during replication. (2) The DNA binding proteins other than histone cores were not considered in the algorithm. In fact, DNA binding proteins may also occupy the DNA sites before or/and after nucleosome assembly, which may dramatically change the local nucleosome positions. (3) Nucleosome assembly involved not only individual factors such as DNA flexibilities, DNA binding proteins, and replication, but also interactions between these factors during the cell cycle, such as gene regulations and transcriptions. In some cases, nucleosome positions were not interfered by transcriptional activities (Komura and Ono, 2003) or the overlapped binding proteins (Sewack and Hansen, 1997). In other cases, nucleosome positions were directly related to protein binding, gene activities, and even their distances to the transcription start sites (Ozsolak *et al.*, 2007; Schones

et al., 2008). The RD algorithm presented here nevertheless provided a general and primitive way for nucleosome position prediction. To perfect the algorithm, more information regarding how nucleosome positions may be affected by gene-specific regulation and other DNA-protein interactions may be needed.

1.6 Summary

Core DNA bending energy was used as a preference index to predict nucleosome core positions in M-*BCR* with the Monte-Carlo algorithm. To verify the prediction, the nucleosome positions in M-*BCR* were experimentally mapped. Results showed that DNA breakpoints occurred in a same chance between the linker region and the core region, implying that core DNA may be protected by histone proteins. The discrepancy between the prediction and experimental mapping showed that nucleosome assembly along DNA may not be a random and simultaneous event. To improve the accuracy of the prediction, and to verify the temporal and spatial sequences of nucleosome assembly, a modified RD algorithm was proposed. This new algorithm predicted nucleosome positions in a way mimicking *in vivo* condition. The prediction result in M-*BCR* showed that RD algorithm had a higher accuracy than the MC algorithm, suggesting nucleosome assembly may likely follow spatial and temporal sequences. More accurate predictions would involve gene-specific information regarding protein-DNA interaction patterns in the region of interest.

Chapter 1, in part is currently being prepared for submission for publication of the material. Tu, Chi-Chiang; Sung, Lanping A. The dissertation author was the primary investigator and author of this material.

Chapter 2 Modeling Nucleosome Assembly at DNA Replication

2.1 Rationale

Since erroneous DNA recombination may happen during DNA replication where DNA is subjected to great topological and mechanical stress, and there is no detail model to describe the dynamic transitions of the nucleosome disassembly/reassembly at the replication fork, it is necessary to adapt experimental findings from literatures to propose a stepwise nucleosome disassembly/reassembly model at the replication fork. This model would provide opportunities to scrutinize the nature and time course of DNA structural and topological changes, and find out possible mechanisms for DNA breakage therein. Replicating DNA goes through dramatic changes in its topology and structures, but previous studies have not proposed an illustrative model to address this issue (see Introduction). A stepwise model would therefore dissect this process more clearly, offer us scenarios to analyze potential stresses that apply to DNA, and hopefully help us figure out the possible mechanisms by which chromosomal DNA breaks in M-*BCR*.

2.2 Specific Aim

The specific aim here is to propose a stepwise nucleosome disassembly/reassembly model at DNA replication in order to dissect the changes of DNA structure and topology in replication. This model should help us find possible mechanism for DNA breakage happening in CML.

2.3 Key Findings from Literatures

Studies have shown that nucleosome disassembly/reassembly at the replication fork is very efficient (Khrapunov et al., 1997; Mazurkiewicz *et al.*, 2006; Sogo *et al.*, 1986). It is conceivable that the disruption and deposition of nucleosomes during replication are coordinated with the advancing replication fork (Gasser *et al.*, 1996; Lucchini *et al.*, 2001), but how these steps are orchestrated remains unclear. Here we gathered up-to-date findings from literatures to support a nucleosome disassembly/reassembly model we proposed at the replication fork. A biochemical study (Levchenko *et al.*, 2005) has suggested a stepwise nucleosome disassembly mechanism. It was found that when sufficient positive stress is applied to a nucleosome core, one H2A-H2B dimer is readily displaced from the nucleosome core. Further stress leads to the dissociation of the second H2A-H2B dimer followed by the $(H3\text{-}H4)_2$ tetramer dissociation. This mechanism clearly showed that nucleosome disassembly follows a three-step process: (1) first dimer dissociation; (2) second dimer dissociation; (3) tetramer dissociation. Verifying this mechanism in an opposite direction, another biochemical study (Mazurkiewicz *et al.*, 2006) revealed a stepwise nucleosome assembly mechanism that is completely reversed from the disassembly mechanism. After a $(H3\text{-}H4)_2$ tetramer efficiently binds to DNA and form a tetrasome (with a time constant $\sim 10^4$ $M^{-1}s^{-1}$), a H2A-H2B dimer quickly binds to the tetrasome (with a time constant $\sim 10^5$ $M^{-1}s^{-1}$) and form a hexasome. The second H2A-H2B dimer binds to the hexasome efficiently but slower than the first dimer for the formation of a

nucleosome (with a time constant ~10^3 $M^{-1}s^{-1}$). The related studies are summarized in Table 2-1.

Table 2-1. Summary of literatures related to nucleosome disassembly/reassembly process dynamics.

Process and steps	Driving force
Nucleosome disassembly: 1. One H2A-H2B dimer dissociates. 2. Another H2A-H2B dimer dissociates. 3. The $(H3\text{-}H4)_2$ tetramer dissociates.	Transcription induced positive stress (Levchenko *et al.*, 2005). DNA unwinding induced positive supercoiling (Alberts *et al.*, 2003d). Mechanical stretching (Cui and Bustamante, 2000; Bennink *et al.*, 2001; Brower-Toland *et al.*, 2002; Sakaue and Löwen, 2004; Gemmen *et al.*, 2005).
Nucleosome assembly: 1. The $(H3H4)_2$ tetramer binds to DNA. 2. One H2A-H2B dimer binds. 3. Another H2A-H2B dimer binds.	Nucleosome assembly protein NAP1 (Mazurkiewicz *et al.*, 2006). Histone-induced supercoiling (Camerini-Otero and Gary Felsenfeld, 1977).

Recently, many studies have emerged on mechanically stretching or twisting a single DNA molecule or chromatin fiber containing one or multiple nucleosomes (Strick *et al.*, 1999; Cui and Bustamante, 2000; Bennink *et al.*, 2001; Brower-Toland *et al.*, 2002; Bryant *et al.*, 2003; Sakaue and Löwen, 2004; Gemmen *et al.*, 2005; Mosconi *et al.*, 2009). The common finding of these studies is that nucleosome disassembly/assembly is a reversible process before a critical threshold of stretching or twisting force is reached. The torsional dynamics of DNA has also provided valuable information that may be adapted during the DNA unwinding process (see 2.4.2).

The topological problem we are interested here is best described in terms of the linking number (*Lk*, Hagerman, 1988). *Lk* is a quantitative measure of the winding of two strands around each other in closed circular DNA or DNA with both ends

constrained, such as chromosomal DNA (Alberts *et al.*, 2003c). *Lk* can be expressed as the sum of the twist number (*T*) and the writhe number (*W*, the number a strand runs across itself). For a constrained system, *Lk* should remain a constant, so the change in one topological parameter (*e.g.*, *T*) must be accompanied by the change of the other parameter (*e.g.*, *W*) in a compensating direction so that there is no net change in *Lk* of the system.

2.4 Results and Discussion

2.4.1 Nucleosome Disassembly in Front of the Fork

In front of each replication fork, up to two nucleosomes are disturbed (Gasser *et al.*, 1996). Electron microscopic studies revealed that first histones H1 then H2A-H2B dimers are progressively displaced in these nucleosomes (Sogo *et al.*, 1986; Gasser *et al.*, 1996). The dissociation of histones unpacks the chromatin and increases the length of nucleosome-free DNA (Figure 2-1a – c) (Brower-Toland *et al.*, 2002). When H2A-H2B dimers dissociate, up to 56 bp of DNA (28 bp on each side of the core) could be released from the dimer binding region (Luger *et al.*, 1997; Brower-Toland *et al.*, 2002). When the $(H3\text{-}H4)_2$ tetramer dissociates, the rest of the core DNA is further released (Figure 2-1d). The total released DNA is 147 bp long.

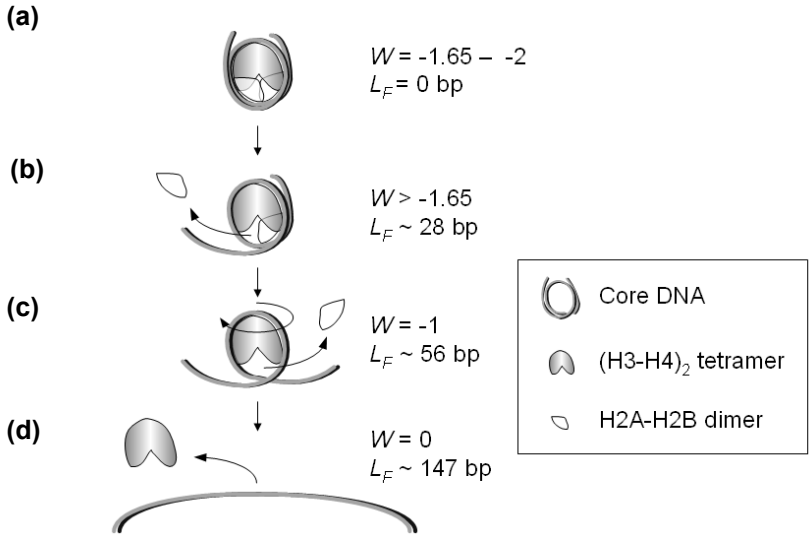


Figure 2-1. Stepwise unwrapping of a nucleosome core. (a) A complete nucleosome core is disturbed to release one H2A-H2B dimer (b) then another dimer (c). Finally a $(H3\text{-}H4)_2$ tetramer is dissociated (d). The writhe number and the free DNA length both increase during this process. W: writhe number. L_F: length of free core DNA. The double helix of DNA is not shown for clarity. The drawing is not to scale.

A mechanical study, in which core DNA is fixed on one end and stretched on the other end, showed that once a critical stretching force is reached, the dissociation of the histones and the release of core DNA are very rapid (Brower-Toland *et al.*, 2002). In a tightly packed chromosome, DNA is subject to significant tension, so it is thought that the sudden release of free DNA in the unwrapping region may facilitate the upstream nucleosome rewrapping process. The topology of each unwrapping core DNA is also affected. DNA originally wraps around a histone core by 1.65 – 2 rounds in a negative supercoiling fashion (Luger *et al.*, 1997; Richmond and Davey, 2003), meaning the writhe number W of the core DNA is -1.65 – -2 compared to linear DNA. When the nucleosome is fully unwrapped, the nucleosomal loop is gone and W increases to 0. The change in writhe ($\triangle W$ = 1.65 – 2) may easily be compensated by the positive supercoiling tension due to DNA unwinding (see below). Overall, the

unwrapping of each nucleosome core increases the length of free DNA by 147 bp and the writhe of the core DNA by 1.65 – 2.

2.4.2 DNA Unwinding and Replicating at the Fork

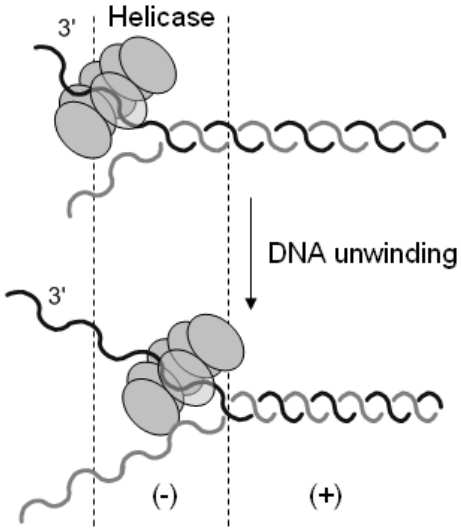


Figure 2-2. Helicase steric exclusion mechanism of DNA unwinding. Top: the leading strand (dark) is bound inside a helicase, which actively translocates along the ssDNA in a 3' to 5' direction and separate it from the lagging strand (gray) by steric exclusion. Bottom: after unwinding of each helical turn, the twist number of the unwound region (between two broken line) is reduced (a bracketed negative sign) and the superhelical tension in front of the fork is increased (a bracketed positive sign). The newly replicated daughter strands and other proteins are not shown for clarity. This drawing is not to scale.

At each replication fork, parental DNA is unwound into two single strands by a ring-shaped hexameric helicase that leads the replication machinery (Matson *et al.*, 1994; Patel and Picha, 2000). Upon sequential NTP hydrolysis (Donmez and Patel, 2006), ssDNA is actively translocated unidirectionally by the helicase such that it circulates *clockwise* within the central cavity of the ring (Singleton *et al.*, 2000). Single stranded DNA is translocated 6 ~ 7 Å or about 2 nucleotides (nt) on each step

of the binding (Singleton *et al.*, 2000), so about 12 nt of ssDNA is translocated after a 360^{o} turn hexameric binding. Different helicase models of DNA unwinding have been proposed (Patel and Picha, 2000), and the current consensus is that only one parental strand is specifically bound within the helicase ring. While the bound parental strand is actively translocated, the other parental strand outside the ring is physically separated by steric exclusion (Figure 2-2) (Kaplan *et al.*, 2002; McGeoch *et al.*, 2005). Human helicases bind and translocate along ssDNA specifically from 3' to 5' (the leading strand, Kaplan *et al.*, 2002).

Helicases unwind the double helix DNA and thus decrease its twist number (*T*). As mentioned above, for every 10 bp DNA unwound, *T* is reduced by 1. The decrease in *T* is compensated by an increase in positive supercoiling tension (either in *T* or in *W*) outside the unwound region to maintain the overall linking number (Figure 2-7, Peter *et al.*, 1998; Wang, 2002; Postow *et al.*, 2004). This positive supercoiling tension, if applied to the nucleosome in front of the fork, may energetically facilitate nucleosome unwrapping. At a first glance, DNA unwinding produces much more positive supercoiling tension than required for nucleosome unwrapping. In a region of same length, DNA unwinding produces 20 compensatory positive supercoils every 200 bp (a typical nucleosome length) and nucleosome core unwrapping in the same DNA length requires only 2 positive supercoils (Figure 2-2). However, a simple mechanical analysis as follows shows that unwinding of 200 bp DNA produces compensatory tension that is just as much as needed to unwrap a nucleosome core.

Compared to the dissociation of H2A-H2B dimers, an energy barrier of 36 ~ 38 *kT* exists for the dissociation of one $(H3\text{-}H4)_2$ tetramer from the core DNA (Brower-Toland *et al.*, 2002). In a tightly coordinated system such as the replication fork, it is immediately assumed that this energy may be provided by the compensatory supercoiling tension (Alberts *et al.*, 2003d). The torque (Γ, in *kT*) generated by twisting a section of dsDNA (Hagerman, 1988; Mosconi *et al.*, 2009) can be expressed as:

$$\Gamma = \frac{C_T}{L_D}(\Delta T) \qquad (3)$$

, where C_T is the torsional modulus of dsDNA (C_T ~ 100 *kT*·nm, Bryant *et al.*, 2003), L_D is the length of the twisted DNA, and ΔT is the change in *T* in this DNA section. For 200 bp of unwound DNA, L_D = 68 nm, ΔT = 20 (positive for compensatory tension), so we have Γ = 29 *kT*, very close to the energy barrier level. This result proved that the assumption is reasonable. The positive supercoiling torque created by DNA unwinding of a nucleosome core region is coupled to the energy required to unwrap the negative supercoiling of the next nucleosome, so the system could self-sustain in both topology and energy. Of course, the spacing between nucleosomes is not always regular, so residual positive supercoiling tension may still cause intertwining nodes (Peter *et al.*, 1998; Wang, 2002; Postow *et al.*, 2004; Bermejo *et al.*, 2007). Topoisomerases I and II act to remove the excessive tension (Bermejo *et al.*, 2007). This mechanical analysis also explains why sometimes the helicase has to move into the immediate prefork nucleosome core (Gasser *et al.*, 1996), presumably because there is not enough DNA spaced between two nucleosome cores to be

unwound and to generate a sufficient torque for unwrapping. A biochemical study showed that the active translocation of DNA helicase produces sufficient force to drive DNA branch migration of Holliday junctions (Kaplan *et al.*, 2002), so it is conceivable that a force-generating molecular motor such as the helicase may itself peel histones off the core DNA if necessary (Brower-Toland *et al.*, 2002; Shin *et al.*, 2007). Besides NTP hydrolysis, DNA unwinding is also facilitated by at least two mechanisms: the release of unwrapped core DNA (see above) and the advancing DNA synthesis. The helicase and polymerase activities are coupled in which DNA synthesis by the DNA polymerase depends on the unwinding activity of the helicase, which provides ssDNA template, and the rapid trapping of the ssDNA bases by the DNA synthesis activity of the polymerase in turn drives the helicase to move forward (Stano *et al.*, 2005).

2.4.3 Nucleosome reassembly after the Fork

After replication, nucleosomes reassemble almost as soon as enough dsDNA is available. The average distance from the fork to the first daughter nucleosome core is shorter than 300 bp (Sogo *et al.*, 1986; Gasser *et al.*, 1996), just enough for a couple of Okazaki fragments to mature (Ogawa and Okazaki, 1980). Studies i*n vitro* showed the nucleosome reassembly follows steps reversed to the disassembly: the $(H3\text{-}H4)_2$ tetramer binds to daughter DNA prior to the binding of H2A-H2B dimers (Almouzni *et al.*, 1990; Smith and Stillman, 1991). *In vivo*, the time lapse between tetramer binding and dimers binding is too short to differentiate (Gasser *et al.*, 1996). The rapid

reassembly of nucleosomes is facilitated by at least four mechanisms: partial attachment of the tetramer to parental strand, coupled pushing of the unwrapped DNA in front of the fork, high affinity of H2A-H2B dimers to the DNA-tetramer complex, and ATP-dependent chromosome assembly proteins. The stepwise nucleosome reassembly is shown in Figure 2-3.

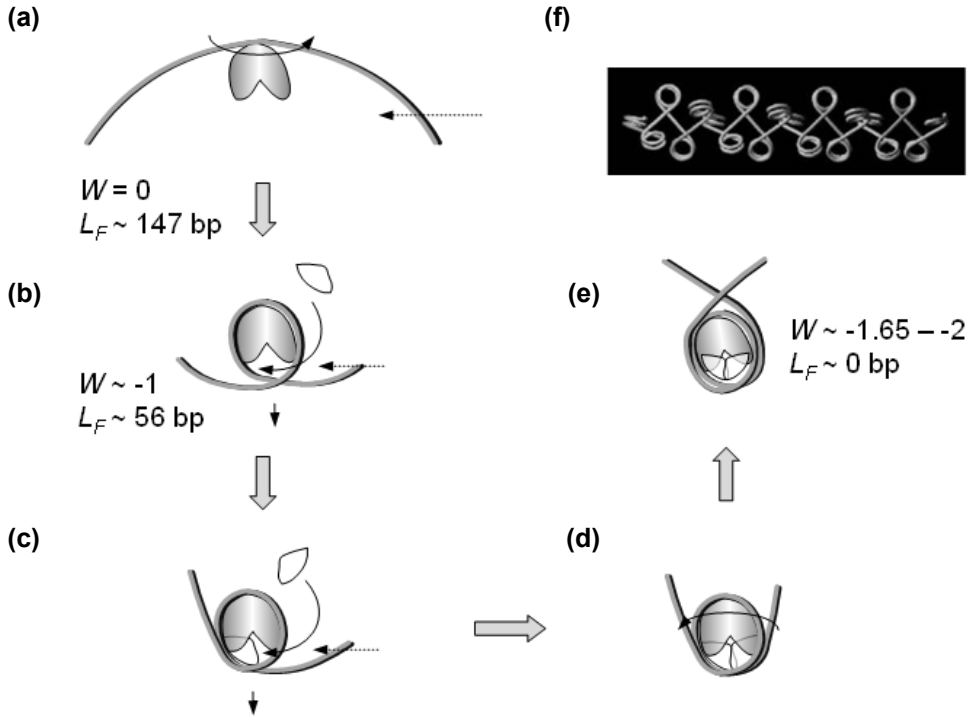


Figure 2-3. Stepwise rewrapping of nucleosomes. (a) Partially attached $(H3\text{-}H4)_2$ tetramers rebind to the daughter DNA. Supercoiling induced by the tetramer (curved arrow) facilitates the rewrapping. (b – c) The pushing of DNA (broken arrow), high affinity of the dimers to DNA-tetramer complex (curved arrow), and chromatin assembly proteins (not shown) help deposit the dimers one after another onto the tetramer-DNA complex. (d) Upon dimer binding, the arms of each nucleosome core closes to compact the structure. (e) After a final twist by H1 (not shown), the rewrapped nucleosome restores its topology in compaction. (f) Nucleosome organizations from Bednar *et al.*, 1998. The legends are as in Figure 2-1. The drawing is not to scale.

Parental $(H3\text{-}H4)_2$ tetramers were shown to remain loose contact with DNA during the passage of the replication fork (Gruss *et al.*, 1993; Lucchini *et al.*, 2001).

The attachment may help the tetramer quickly recognize the same nucleosome positioning signals on the daughter strands as the intact nucleosome (Dong and van Holde, 1991; Hayes *et al.*, 1991). The tetramer itself can induce negative supercoiling required for rewrapping (Figure 2-3a), and thus no tension or topoisomerase is needed for tetramer rewrapping (Camerini-Otero and Felsenfeld, 1977; Annunziato, 1989). After tetramer binding, abrupt release of the unwrapped DNA in front of the fork provides pulses of pushing behind the fork, resembling a reversed stretching of DNA that rewrap the nucleosomes (Cui and Bustamante, 2000; Brower-Toland *et al.*, 2002; Leuba et al., 2003; Gemmen *et al.*, 2005). The high affinity of H2A-H2B dimers to the DNA-tetramer complex (Mazurkiewicz *et al.*, 2006) further speeds the binding of dimers. Upon dimer binding, the two DNA arms of each nucleosome core close one after another and compact the organization (Figure 2-3b and c). According to the chromatin organization observed under an electromicroscope (Bednar *et al.*, 1998), another twist on each nucleosome core must exist to further compact the structure (Figure 2-3d). This final twist was found to be accomplished by H1 histones (Bednar *et al.*, 1998). Adjacent nucleosome cores are arranged in a zigzag fashion (Figure 2-3f). Alternatively, several ATP-dependent chromatin assembly proteins (CAF-1, RCAF, NAP1…etc) couple with DNA synthesis to actively deposit newly synthesized histones onto the daughter strands (Tyler, 2002). Overall, the reassembly of each nucleosome recovers the topology and the compaction of chromosomal DNA (Figure 2-3e and f).

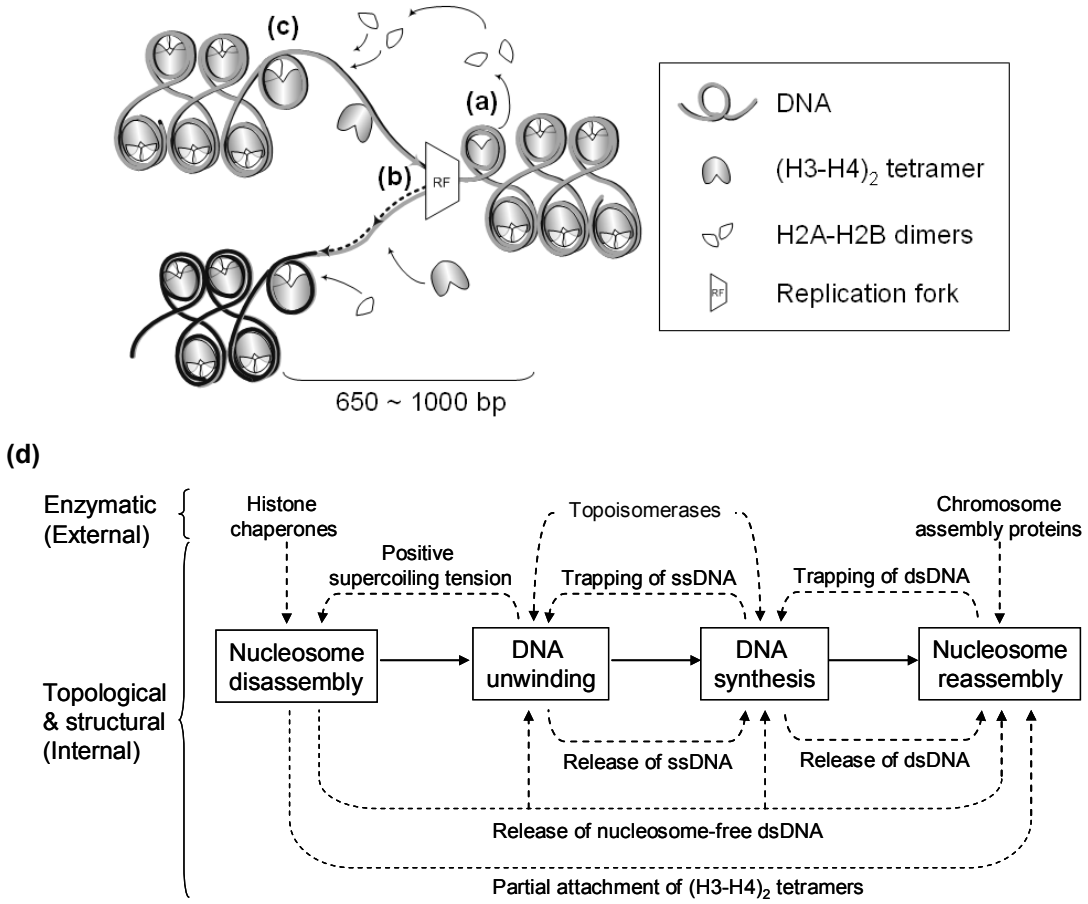


Figure 2-4. Coupled nucleosome disassembly/reassembly model at the replication fork. Nucleosome disassembly (a) is facilitated by DNA unwinding at the fork (b). DNA from unwrapped nucleosomes in turn facilitates downstream DNA unwinding and nucleosome reassembly (c). The disturbed DNA region spans 650 – 1000 bp. (d) Mechanisms to assure the efficiency and robustness of the process. For each step (box) of the system, several mechanisms exist to facilitate (broken arrow) the processes. These mechanisms come from internal (mutual facilitation) as well as external (ATP-dependent enzymes) sources.

To sum up, nucleosome disassembly, DNA unwinding/synthesis, and nucleosome reassembly across the fork are tightly coupled to each other. A summary model is shown in Figure 2-4. During replication, the disrupted chromatin region spans 650 – 1000 bp across the fork (Gasser *et al.*, 1996). DNA unwinding at the fork produces compensatory positive supercoiling tension that helps disassemble

nucleosomes in front of the fork. Nucleosome disassembly abruptly releases dsDNA to the helicase which in turn facilitates the DNA unwinding. DNA unwinding is also coupled to DNA synthesis in that these two processes are mutually facilitated. Finally, several mechanisms assure rapid nucleosome reassembly behind the fork. Each individual step in nucleosome disassembly/reassembly is supported by multiple factors so that the process is coordinated, efficient, and robust (Figure 2-4d). With this general model to elaborate DNA structural and topological changes at replication fork, we can scrutinize these changes in specific chromosome structure of M-*BCR* and uncover possible mechanisms for DNA breakages.

2.5 Summary

To scrutinize topological and mechanical stress on DNA during replication and in hope to find possible mechanisms for DNA breakage therein, a stepwise nucleosome disassembly/reassembly model at the replication fork was proposed. The model elaborated tight coordination among nucleosome disassembly, DNA unwinding, synthesis, and nucleosome reassembly. DNA unwinding topologically and energetically facilitates nucleosome disassembly, which in turn helps the progressing nucleosome reassembly. DNA unwinding is also coupled with and mutually facilitated by DNA synthesis. Nucleosome reassembly behind the fork is facilitated by the nucleosome disassembly in front of the fork, partial attachment of parental histone tetramers, and high affinity of dimers to the tetramer. Using this model as the basis, the possible mechanisms for DNA breakage in M-*BCR* replication can be revealed.

Chapter 2, in part is currently being prepared for submission for publication of the material. Tu, Chi-Chiang; Sung, Lanping A. The dissertation author was the primary investigator and author of this material.

Chapter 3 Mechanisms for DNA Breakage during M-*BCR* Replication

3.1 Rationale

The nucleosome mapping of M-*BCR* in Chapter 1 revealed that 14 breakpoints were in the core regions, and 13 were in the linker regions (Figure 3-1). Moreover, 19 of the 27 breakpoints were found upstream to a 292-bp nucleosome-excluded region (from the downstream boundary of N6 to upstream boundary of N7), of which 5 of the 19 clustered within the nucleosome-excluded region. A notable feature found in this region is an A-tract composed of 17 consecutive deoxyadenine nucleotides (17-A), 168 bp downstream to N6 and 110 bp upstream to N7.

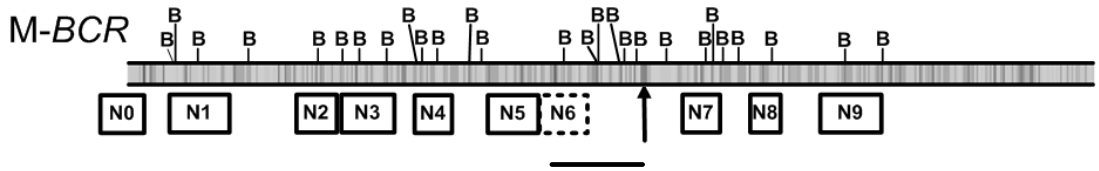


Figure 3-1. Nucleosome-excluded region in M-*BCR*. Mapped nucleosome positions (open box), an *Alu* homology element (horizontal line), and 17-A (arrow) are aligned with the M-*BCR* sequence. Twenty seven breakpoints (B) are indicated. Note there is a 292-bp nucleosome-free region between N6 and N7.

A-tracts are common in most eukaryotes and are generally located in the A-rich regions at the 3' end of *Alu* repeats within longer repetitive sequences (Jelinek and Schmid, 1982). The mean length of 3' A-tract in *Alu* elements (J and S subfamily) is 21 ± 8 bp (Roy-Engel *et al.*, 2002). In fact, scanning the M-*BCR* sequence with

CENSOR (Kohany *et al.*, 2006) revealed an *Alu* S element at the 5' end of the 17-A. AT-rich nucleotides are predicted to be flexible (Drew and Dickerson, 1981; Travers and Klug, 1987). Average DNA is able to bend ~90^{o} smoothly in about 200 bp (Alberts *et al.*, 2003c), but an 8-bp TATA box is able to bend ~80^{o} upon protein binding (Kim *et al.*, 1993). In a nucleosome core, AT-rich nucleotides are often found in the sharp bent inner minor grooves while GC-rich nucleotides in the outer minor grooves (Satchwell *et al.*, 1986; Travers, 1987). A 219-bp A-rich DNA fragment formed a loop by itself (Griffith *et al.*, 1986). On the other hand, poly-A fragments are predicted to be more rigid than average sequences (Nelson *et al.*, 1987; Coll *et al.*, 1987). Short poly-A tracts (4 or 5 bp) were found to occupy the least bent positions within nucleosomes (Satchwell *et al.*, 1986; Travers and Klug, 1987) and long poly-A tracts (80-100 bp) are nucleosome-exclusive (Kunkel and Martinson, 1981). It is thus conceivable that the 292-bp A-rich region in M-*BCR* may adapt a conformation somehow excluding nucleosome formations. Chromosome conformation is the organization of chromosomes in a cell's natural state. Studying the structural properties and spatial organization of chromosomes is important for the understanding and evaluation of the regulation of gene expression, DNA replication, repair, and recombination (Dekker *et al.*, 2002). In Chapter 2, a stepwise nucleosome disassembly/reassembly model at the replication fork was proposed. This general model elaborated changes in topology and structure of replicating DNA and nucleosomes. The next step in this study is to incorporate the distinct chromosome

conformation of M-*BCR*, if any, into this model so as to understand a potential mechanism for DNA breakage.

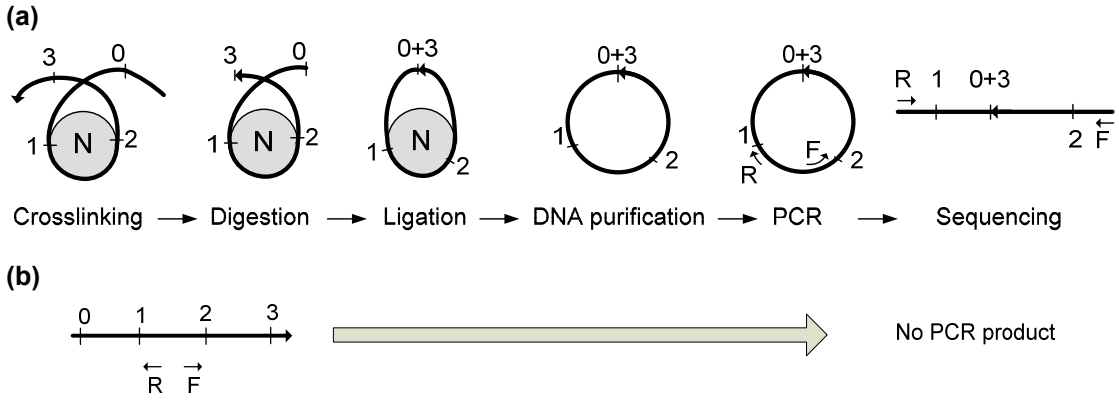


Figure 3-2. Principle of chromosome conformation capture. (a) The flow of the procedure. DNA is shown as an arrowed line (pointing 3' end) and the histone is shown as a shaded particle (N). Restriction sites are numbered (0 to 3) in a 5' to 3' direction. The region of interest is flanked by two restriction sites (*e.g.*, 1 and 2). Primers (R: reverse; F: forward) are shown as small arrows. The procedure goes from left to right. Chromosome DNA is crosslinked, digested, ligated, purified, PCR amplified, and finally sequenced. (b) Same procedure with protein-free DNA results in no PCR product.

To see how chromosome conformation may contribute to the DNA breakage at the breakpoints in M-*BCR*, an up-to-date technology would be applied to experimentally resolve the conformation around the A-rich region. "Chromosome Conformation Capture" (3C) is a pioneering methodology that allows *in vivo* genomic organization to be explored (Figure 3-2a) based on Dekker *et al.*, 2002 and Hagège *et al.*, 2007. It involves formaldehyde fixation of cells (Orlando *et al.*, 1997), thus freezing the dynamic chromosome conformation. The following digestion and ligation trim unprotected DNA away and bring together two DNA ends in close proximity. After purification, the ligated DNA serves as a template and can be amplified by PCR (3C-PCR) with primers designed specifically for recombined but not original DNA.

The PCR product is then sequenced to reveal the exact recombination sites. A section of protein-free DNA would result in no PCR product from the uniquely designed primers (Figure 3-2b).

The resolved conformation of the A-rich region in M-*BCR* would provide answers to key questions in this study such as: why were 5 breakpoints clustered within the A-rich region? Why 19 of the 27 breakpoints in M-*BCR* were found upstream to the A-rich region? How did DNA breakage occur within the core regions? The result would also contribute more insight to the DNA breakage mechanisms in *BCR*.

3.2 Specific Aim

The specific aim here is to experimentally resolve the chromosomal conformation in the A-rich region of M-*BCR*. The stepwise nucleosome disassembly/reassembly model and the nucleosome organization of M-*BCR* should be incorporated to reveal mechanisms for DNA breakage in CML.

3.3 Experimental Design

The choice of restriction enzymes is critical for the quality of 3C procedures (Hagège *et al.*, 2007). Since a relative small region (<10 kb) is analyzed, frequent cutting enzymes (4-base cutters) are of the choice. Restriction enzymes sensitive to DNA methylation were not considered. The digestion temperature should not be higher than 65^{o}C to avoid any premature decrosslinking. The digestion product should

also be easy to re-ligate (*i.e.*, sticky end if possible). An optimal enzyme should also have restriction sites flanking the A-rich region and one or more nucleosome core region. After all, to allow sufficient resolution and sensitivity of the procedure, *Alu*I was chosen as the restriction enzyme. There are 19 *Alu*I sites in M-*BCR*. The 17-A is flanked by two *Alu*I sites separated by 99 bp (smallest among all enzymes considered), and several *Alu*I sites are protected by nucleosome cores (Figure 3-3).

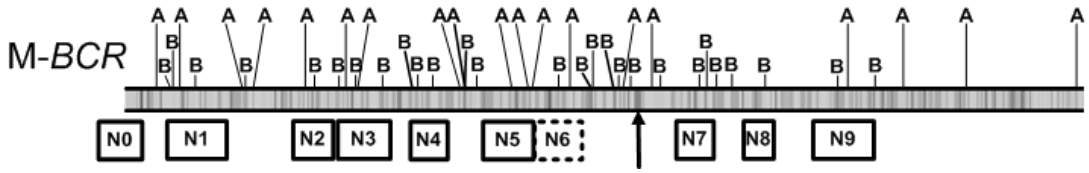


Figure 3-3. *Alu*I restriction map of M-*BCR*. Nineteen *Alu*I sites (A) are found in the M-*BCR* region. The 17-A (arrow) is flanked by two *Alu*I sites separated by 99 bp. Several *Alu*I sites are expected to be protected by nucleosome cores (*e.g.*, N5, N9). Other legends are as in Figure 3-1.

To design suitable primers, all possible conformations were sketched and the corresponding 3C-PCR products in this region were predicted (not shown). It is noteworthy that even though only one protein is required to maintain specific conformation, multiple proteins may exist in the same region and generate the same 3C-PCR product as long as the same restriction site(s) are protected and the same DNA curvature is maintained. It turned out that each set of primers was capable of amplifying multiple templates and detecting several conformations. Because more than a single product will be generated, traditional PCR had to be used instead of quantitative real-time PCR. Primers were then designed to exclusively work on the recombined DNA. Each set of primers contained an upstream reverse primer and a downstream forward primer (Figure 3-4), so only the recombined DNA but not the

normal genomic DNA would be amplified. Typically, three sets of primers were designed to detect all possible conformations in one region.

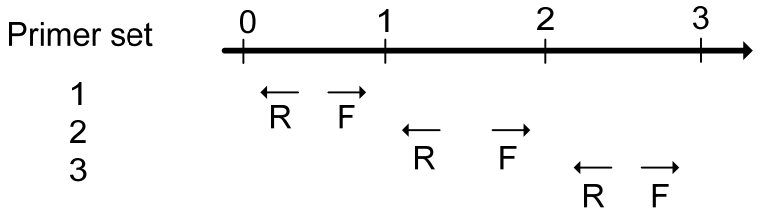


Figure 3-4. Design of 3C primers. DNA is shown as a thick arrowed line (pointing 3' end). Primer sets (arrow) are aligned with the restriction sites (numbered vertical line) on the DNA. Each set of primers includes an upstream reverse primer (R) and a downstream forward primer (F) that only amplifies recombined DNA. One set of primer is able to detect several conformations in the same region.

3.4 Materials and Methods

3.4.1 Formaldehyde Crosslinking with Single Cells

HL-60 cells (ATCC, Manassas, VA) were maintained as described (Yalcintepe *et al.*, 2005). Approximately 10^7 cells were centrifuged for 1 min at 400 ×g at room temperature. After discarding the supernatant, the cell pellet was resuspended in 500 µl of 10% FCS/PBS. Then 9.5 ml of 1% formaldehyde in 10% FCS/PBS was added to the suspension. The solution was tumble-mixed at room temperature for 10 min to crosslink the chromosome conformations. The reaction tube was immediately transferred to ice and 1.425 ml of ice-cold 1 M glycine was added to quench the reaction. The tube was centrifuged for 8 min at 225 ×g at 4^{o}C and the supernatant was carefully removed. The cell pellet was then frozen in liquid nitrogen and stored at -80^{o}C until use.

3.4.2 Cell Lysis

The cell pellet was resuspended in 500 µl cold lysis buffer (10 mM Tris-HCl pH 7.5, 10 mM NaCl, 1% NP-40, 0.5% Triton-X 100, 1× complete Mini protease inhibitor)(Roche, Mannheim, Germany). The cell suspension was then delivered drop by drop into a tube of 4.5 ml cold lysis buffer with a micropipette and incubated for 10 min on ice with constant shaking and pippeting. The quality of nuclei was verified by methyl-green pyronin staining (Sigma, St. Louis, MO) under microscope. The lysis reaction may last longer if necessary. The nuclei pellet was spun down at 2,100 ×g for 20 min through a 2 ml cushion of 1.62 M sucrose. The supernatant was carefully removed. For storage, the nuclei were quickly frozen in liquid nitrogen and stored at -80^{o}C until use.

3.4.3 *Alu*I Digestion

The nuclei pellet was resuspended in 500 µl of 1.2× *Alu*I restriction buffer (Fermentas, Glen Burnie, MA) and transferred to a 1.5 ml tube prior to the addition of 8.5 µl of 20% (w/v) SDS (final concentration 0.3%). The sample was then incubated for 1 h at 37^{o}C followed by the addition of 56.5 µl of 20% (v/v) Triton X-100 (final concentration 2%) and a second incubation for 1 h at 37^{o}C while shaking at 250 rpm in the entire procedure. A 50 µl aliquot of the sample was taken and labeled as undigested DNA control (UD) and stored at -20^{o}C until use. Four hundred units of *Alu*I (Fermentas, Glen Burnie, MA) were added to the remaining sample which was then incubated overnight at 37^{o}C while shaking at 250 rpm. Another 50 µl aliquot of

the digested sample was taken, labeled as digested DNA control (DD), and stored at -20^{o}C until use.

3.4.4 Digestion Efficiency Check

The procedure was as described (Hagège *et al.*, 2007) except that traditional PCR was used instead of quantitative PCR. Five hundred microliters of Proteinase K buffer (5 mM EDTA pH 8.0, 10 mM Tris-HCl pH 8.0, 0.5% SDS) and 1 µl of Proteinase K (Fermentas, Glen Burnie, MA) were added to each control aliquot saved previously and the samples were incubated at 55^{o}C overnight. After equilibrating at 37^{o}C, 1 µl of 1 mg/ml RNase A was added to each sample which was then incubated for 2 h at 37^{o}C. Each control was phenol-chloroform extracted twice before the addition of 50 µl of 3 M sodium acetate pH 7.0 and 1 ml of ethanol, mixed respectively, and placed at -80^{o}C to precipitate DNA. The control was then centrifuged for 20 min at 16,100 ×g at 4^{o}C, supernatant removed, and washed with 500 µl of 70% ethanol. The control was centrifuged again for 4 min at 16,100 ×g at 4^{o}C. The supernatant was removed and the pellet was dried at room temperature and resuspended in 60 µl of TE buffer. The DNA concentration was quantified with a UV spectrophotometer (Beckman DU640, Fullerton, CA) with a ratio of 260/280 at ~1.8. The typical concentration of each control is ~120 ng/µl.

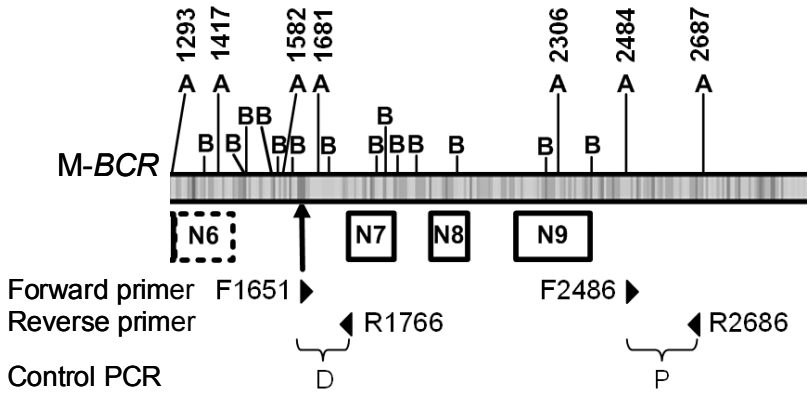


Figure 3-5. The location of primers to check digestion efficiency. Primers (arrowhead) are aligned with M-*BCR* sequence. The specific locations of *Alu*I sites (A) are indicated. Primers are named by the direction of amplification (F: forward; R: reverse) and the 5' end location (in bp) from the 109,369th bp in *BCR*. The primer pair for PCR D includes one *Alu*I site (at 1681 bp). The primer pair for PCR P includes no *Alu*I site. The *Alu*I digestion efficiency is determined by the concentrations of these PCR products before and after *Alu*I digestion (See text for detail).

The purified DNA from UD and DD served as control PCR templates to determine the digestion efficiency as detailed below. Primers were designed to amplify either the region preserved from the digestion (PCR product P) or the region including an *Alu*I site (PCR product D). The primer design is shown in Figure 3-5 and the sequences are listed in Table 3-1. Both PCR were performed for each control, 25 µl each with 7 ng template, 200 µM dNTP mix, 1 µM of either primer mix, 0.25 µl iProof DNA polymerase (Biorad, Hercules, CA), and a suitable buffer. The program was 98^{o}C for 30 sec, 28 cycles of 98^{o}C for 10 sec, 63^{o}C for 15 sec, 72^{o}C for 30 sec, and a final 72^{o}C for 7 min (The cycle number was reduced from 35 to 28 to assure the amplification was in a linear range.) The PCR products were resolved on a 1.3% agarose gel by electrophoresis (Gibco, Carlsbad, CA). The gel picture was analyzed by ImageJ 1.41 (Wayne Rasband, NIH, US) to determine the intensity of the PCR products.

For each control (UD or DD), the intensity of PCR product P reflected the initial template concentration, and that of the PCR product D reflected the template concentration after digestion. The digestion efficiency was determined by the ratio of digested template concentration to preserved template concentration, both of which normalized by the initial template concentration. The efficiency was calculated by the formula:

$$DE = 1 - \frac{D_{DD} / P_{DD}}{D_{UD} / P_{UD}} \qquad (4)$$

where D_{DD} and D_{UD} are the respective product intensity of PCR D from DD and UD; P_{DD} and P_{UD} are the respective product intensity of PCR P from DD and UD. Only the samples with digestion efficiency higher than 80% were used for the following procedures to avoid unwanted artifacts (Hagège *et al.*, 2007).

Table 3-1. Primer sets for digestion efficiency check in M-*BCR*. Primers are named by the direction of amplification (F: forward; R: reverse) and the 5' end location (in bp) from the 109,369th bp in *BCR*.

PCR product	Primer name	5' to 3' sequence
D	F1651	TGCTTCCTGTGCCCCACAGTG
	R1766	TAAAGCCCCTACGATGAGAAGGGCC
P	F2486	CCTAGAAACAGCAAAATGTGGAGACAG
	R2686	ACTGCTCTCACCTAAAAAGTCCCCC

3.4.5 Ligation of Digested Chromosomes

Forty nine micro liters of 20% (w/v) SDS (final 1.6%) was added to the remaining sample from the digestion and the sample were transferred to a 50 ml falcon tube prior to the addition of 6 ml of 1.16× ligation buffer (Promega, Madison, WI)

and 375 µl of 20% (v/v) Triton X-100 (final 1%). The sample was incubated for 1 h at 37^{o}C while shaking at 160 rpm prior to the addition of 100 U ligase (Promega, Madison, WI), and was incubated overnight at 16^{o}C. The ligation was stopped by addition of 300 µg Proteinase K (Fermentas, Glen Burnie, MA) and incubation at 65^{o}C overnight.

3.4.6 DNA Purification

After adding 30 µl of 10 mg/ml RNase A, the ligated sample was immediately incubated for 45 min at 37^{o}C. Seven milliliters of phenol-chloroform was added to the sample and was mixed vigorously. The sample was centrifuged for 15 min at 2,200 ×g at room temperature. The supernatant was transferred into a new 50 ml tube. If necessary, the supernatant was extracted again with phenol-chloroform. Seven milliliters of distilled water was added to the supernatant prior to the addition of 1.5 ml of 3 M sodium acetate (pH 7.0) and 30 ml of 100% ethanol. The sample was then mixed and placed at -80^{o}C for 3 h prior to centrifugation for 45 min at 2200 ×g at 4^{o}C. The supernatant was removed and the DNA pellet was washed with 10 ml of 70% (v/v) ethanol. The sample was then centrifuged for 15 min at 2200 ×g at 4^{o}C, supernatant removed, and dried at room temperature. The DNA pellet was dissolved in 500 µl of TE buffer, labeled as 3C-DNA (chromosome conformation captured DNA), and stored at -20^{o}C until use. Since the optical density (OD_{260}) of 3C-DNA is not reliable (Hagège *et al.*, 2007), the concentration of each 3C-DNA sample was estimated from the control DNA concentration. The typical 3C-DNA concentration is ~ 130 ng/µl.

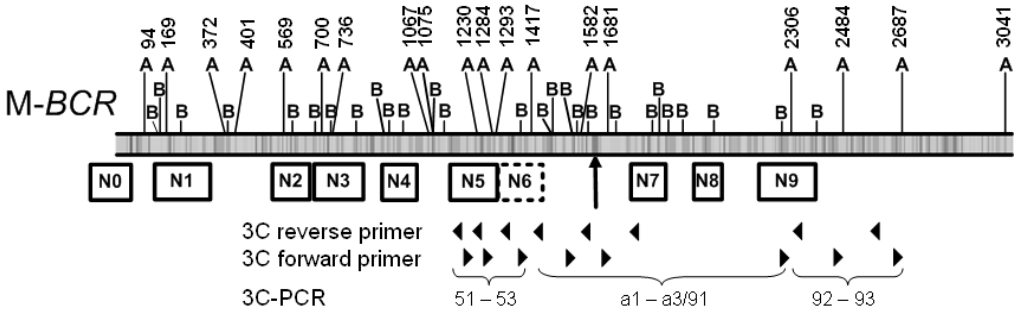


Figure 3-6. Design of 3C-primers in M-*BCR*. 3C-primers (arrowhead) are aligned with M-*BCR* sequence. 3C-PCR are named by their chromosome feature (5: N5; a: 17-A; 9: N9) and sequential order (1 – 3: upstream to downstream). Other legends are as in Figure 3-5.

3.4.7 3C-PCR Amplification and Sequence Analysis

The design of 3C-primers in this experiment is illustrated in Figure 3-6 and detail sequences listed in Table 3-2. A 50 µl 3C-PCR was performed for each primer set with 200 ng 3C-DNA template, 200 µM dNTP mix, 1 µM of either primer mix, 0.5 µl iProof DNA polymerase (Biorad, Hercules, CA), and a suitable buffer. The program was 98^{o}C for 30 sec, 35 cycles of 98^{o}C for 10 sec, 59 ~ 63^{o}C for 15 sec, 72^{o}C for 30 sec, and a final 72^{o}C for 7 min. The PCR products were resolved on a 1.3% agarose gel by electrophoresis (Gibco, Carlsbad, CA), extracted (Qiagen, Valencia, CA) and sequenced (Fisher, SeqWright, Houston, TX). The sequencing results were analyzed by FinchTV (Geospiza Inc., Seattle, WA). From the designed primer locations and predicted sequence of PCR products, the size of each 3C-PCR product from various conformations were predicted and calculated.

Table 3-2. 3C-Primers in M-*BCR*. Primers are named by their direction of amplification (F: forward; R: reverse) and their 5' end locations (in bp) downstream from the 109369th bp in *BCR*.

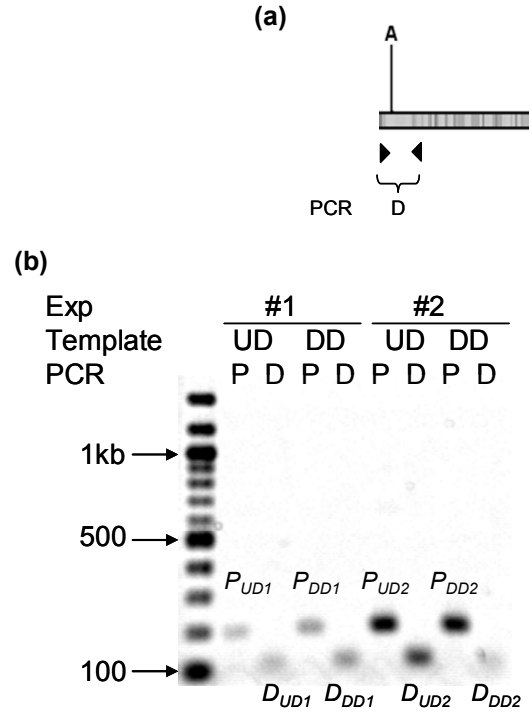
PCR product	Primer name	5' to 3' sequence
51	R1105	GCAATTACTCGATCCACCTGCA
	F1206	CTCTGTCGAGCTGGATGGATACTAC
52	R1259	AAAGGAAAAAAAAAGTAGTATCCATCCAG
	F1261	CCTCTAAGTGGGGGTCTCC
53	R1320	TGTTACCAGCCTTCACTGTTCTGACA
	F1394	GCTGAGGCAGGTGGATCGCTT
a1	R1463	ACACAGGGTTTCACCATGTTGGTC
	F1507	TGTAATCACAACTGCTTGGGAGGCT
a2	R1604	GGCTGGAATGCAGTGGCACA
	F1651	CCTAGAAACAGCAAAATGTGGAGACAG
a3 or 91	R1766	ACTGCTCTCACCTAAAAAGTCCCCC
	F2222	TGCCATTCTCCATCAGTGAGGC
92	R2333	GGACCCTTTCTGCAGGGATAT
	F2463	CCACGACTTCTCCAGCAC
93	R2559	GCAGACCCCGCTGTACTAT
	F2662	GGCCCTTCTCATCGTAGGGGCTTTA

3.5 Results and Discussion

3.5.1 Digestion Efficiency of 3C-DNA

A typical digestion efficiency result is shown in Figure 3-7. Two sets of primers were designed such that one set (D) included an *Alu*I site while the other set (P) excluded (Figure 3-7a). The D/P value (product intensity of PCR D normalized to that of PCR P) of a sample reflected its *Alu*I digestion level without the bias from initial sample concentration. The ratio of D/P from the digested sample (DD) to D/P from the undigested sample (UD) determined an unbiased *Alu*I digestion efficiency (DE). The expected size was 201 bp for PCR product P and 116 bp for PCR product D. In Experiment#1, the cell lysis incubation step was purposely skipped, thus the nuclei

were not fully exposed before digestion. The DE value was 37% (Figure 3-7c). In Experiment#2, the cells were completely lysed and nuclei isolation was verified by methyl-green pyronin staining. The digestion efficiency was 84%. It shows that cell lysis was crucial to digestion efficiency of chromosomal DNA and traditional PCR was able to differentiate the digestion level. Only samples with digestion efficiency higher than 80% were used as 3C-DNA templates to avoid unwanted artifacts.



(c)

Exp	PCR product	Intensity	D/P	DE
#1	P_{UD1}	53	2.21	37%
	D_{UD1}	117		
	P_{DD1}	140	1.39	
	D_{DD1}	194		
#2	P_{UD2}	306	1.12	84%
	D_{UD2}	343		
	P_{DD2}	307	0.18	
	D_{DD2}	54		

Figure 3-7. Typical digestion efficiency result. (a) A simplified *Alu*I restriction map of M-*BCR* showing the design of primer set D and primer set P. Legends are as in Figure 3-5. (b) PCR products from two independent batches of samples (Exp#1 and #2) were resolved on a 1.3% agarose gel. Each batch had 2 controls (UD and DD) and each control had 2 PCR products (P and D). The expected PCR product sizes are indicated (P: 201 bp; D: 116 bp). The left lane is a 100 bp DNA ladder (NEB, Ipswich, MA). The major sizes are indicated. (c) Calculation of the digestion efficiency (DE). The DE for Exp#1 is 37% and that for Exp#2 is 84%.

3.5.2 Chromosome Conformations around Nucleosome Core N5 Region

Qualified 3C-DNA was amplified by 3C-primers in the N5 region. As negative controls, UD and DD were also amplified separately with the same 3C-primers. All templates were also amplified by primer set C as positive controls. The PCR products were resolved on a 1.3% agarose gel. All distinct bands, including PCR product C from all templates and distinct bands from UD, DD, and 3C-DNA were extracted and sequenced. The 3C-PCR products around N5 are shown in Figure 3-8a. No distinct product was found from the UD and DD, meaning the primers exclusively amplified the recombined DNA, not the original DNA. The sequence of the smear in PCR 53 from the DD turned out unspecific. Two distinct bands were found from 3C-DNA: 242 bp in PCR 51 (product 1) and 341 bp in PCR 52 (product 2). The sequential illustration and analysis of these PCR products are shown in Figure 3-8b and c. The sequence of PCR 51 product revealed a conformation where the exposed *Alu*I site at 1075 and 1417 bp recombined, and three *Alu*I sites in between (1230, 1284, 1293 bp) were protected. The sequence of PCR 52 product revealed an alternate conformation where the *Alu*I sites at 1075 and 1293 bp recombined, suggesting *Alu*I sites at 1075 and 1293 bp were exposed and those in between (1230 and 1284 bp) were protected.

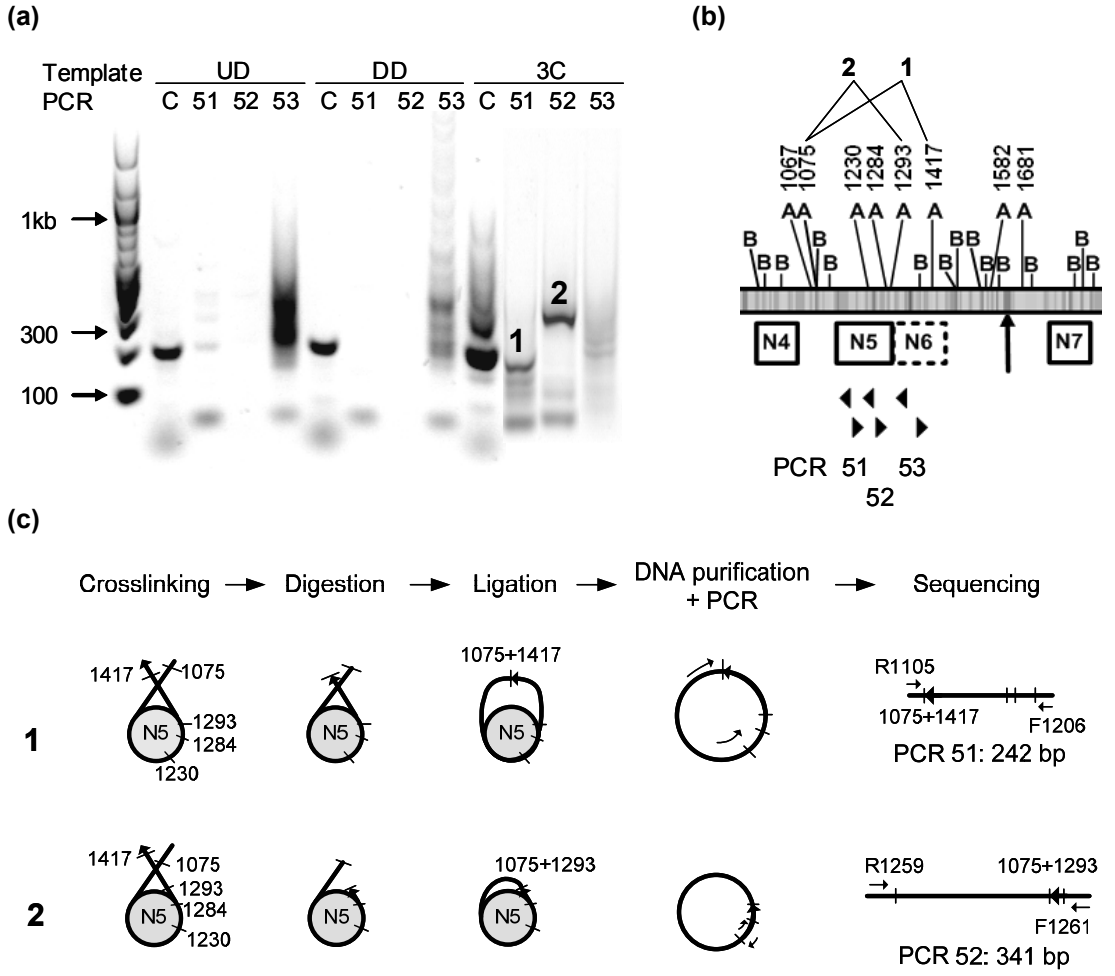


Figure 3-8. 3C-PCR products around N5. (a) The gel pattern. The left lane is a 100 bp DNA ladder. Templates include UD, DD, and 3C-DNA. PCR C serves as a positive control. Both PCR 51 and 52 from 3C-DNA generate a distinct band (marked 1 and 2). (b) Sequential illustration of 3C-PCR products in the N5 region. The ligated *Alu*I sites in indicated PCR products are connected by solid and dotted lines, respectively. The legends are as in Figure 3-7. (c) Sequence analysis of PCR products. The legends are as in Figure 3-2. This figure is not to scale.

Both PCR products agreed with the experimentally mapped nucleosome positions in that the *Alu*I site at 1075 and 1417 bp were in the linker regions. Notably, although the *Alu*I sites at 1075 and 1417 bp are 342 bp away from each other along the sequence, they were recombined in the ligation step due to their close proximity in a chromatin compaction and also the preserved DNA curvature by a histone core. The

fact that the *Alu*I site at 1293 bp was protected in PCR 51 but exposed in PCR 52 may be viewed as a dynamic behavior of nucleosome positions. Since the 1293 bp was near the boundary of N5, when the histone core moves along the DNA during remodeling, this *Alu*I site may be exposed from the core boundary. The 3C-PCR was able to detect the alternate populations of nucleosome positions. The 3C-PCR result also suggested that there may not be a stable nucleosome core (N6) to protect the *Alu*I site at 1417 bp. This finding is consistent with the previous mapped positions where N6 was not a stable nucleosome core. Therefore, the nucleosome-free region may actually be longer than 292 bp (459 bp from the N5 boundary to the N7 boundary).

3.5.3 Chromosome Conformations around Nucleosome Core N9 Region

The 3C-PCR products around N9 region are shown in Figure 3-9a. Only PCR 92 and 93 results are shown as PCR 91 is the same as PCR a3, presented in the next section. No unspecific product was found from UD and DD, as expected. From 3C-DNA, two distinct products were found in PCR 92 (product 1: 674 bp; product 4: 252 bp). The sizes of these products matched the predictions for nucleosome conformations. Two products were also found in PCR 93 (product 2: 556 bp; product 3: 279 bp) where one matched a nucleosome conformation (product 3) and the other does not belong to the prediction (product 2).

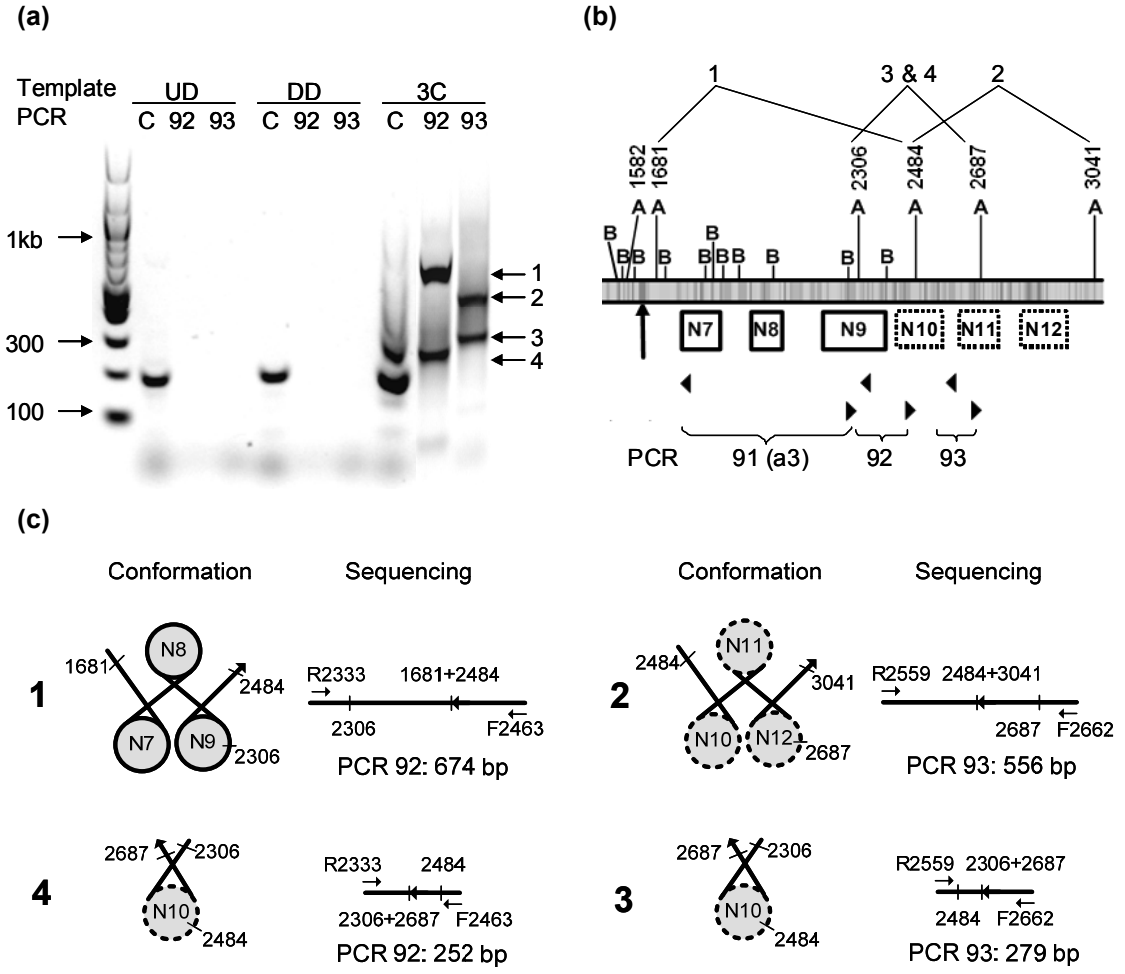


Figure 3-9. 3C-PCR products around N9. (a) The gel pattern. Each of PCR 92 and 93 from 3C-DNA generates 2 distinct products (marked 1 – 4). (b) Sequential illustration of 3C-PCR products in the N9 region. The ligated *Alu*I sites in indicated PCR products are connected by solid lines respectively. The unmapped nucleosome cores (marked N10 – N12) are shown in dashed particles. The legends are as in Figure 3-7. (c) Sequence analysis of PCR products 1 – 4. The legends are as in Figure 3-2. The drawing is not to scale.

The sequential illustration and analysis of these PCR products are shown in Figure 3-9b and c. The sequence of product 1 revealed a conformation where the *Alu*I site at 2306 bp was protected and those at 1681 and 2464 bp recombined to each other. Again, the recombination of *Alu*I sites between 1681 and 2464 bp was facilitated by the zigzag formation of nucleosomes as shown in Figure 3-9c, even though these two

sites are about 800 bp separated. The sequence of product 2 revealed an unpredicted conformation where the *Alu*I site at 2687 bp was protected and those at 2484 and 3041 bp were exposed. This suggests a protein binding at 2687 bp, presumably by a histone core which was not mapped before (Figure 3-9b). The sequence of product 3 and 4 both revealed a conformation where the *Alu*I site at 2484 bp was protected, suggesting another possible nucleosome core. The fact that the *Alu*I site at 2484 bp was protected in product 3 and 4 but exposed in product 2 is again due to the dynamic behavior of nucleosome positions.

The combined result from PCR 92 and 93 shows there were proteins to cover each of the 3 *Alu*I sites at 2306, 2484, and 2687 bp (Figure 3-9b). This is in part consistent with the mapped nucleosome positions where a nucleosome N9 covers the *Alu*I site at 2306 bp. Two possible nucleosome cores at 2484 and 2687 bp were not mapped previously since there was no breakpoint reported. The 3C-PCR also revealed the dynamic behavior of nucleosomes in this region where these three *Alu*I sites may not be always covered by histones. In contrast, the *Alu*I site at 1681 bp was always found exposed, confirming a nucleosome-free conformation within the A-rich region.

3.5.4 Chromosome Conformation in A-rich region

The PCR products are shown in Figure 3-10a. From 3C-DNA, only PCR a3 (and 91) had two distinct products (product 1: 348 bp; product 2: 170 bp), matching the predicted conformations. PCR a1 and a2 showed smeared patterns and no distinct product.

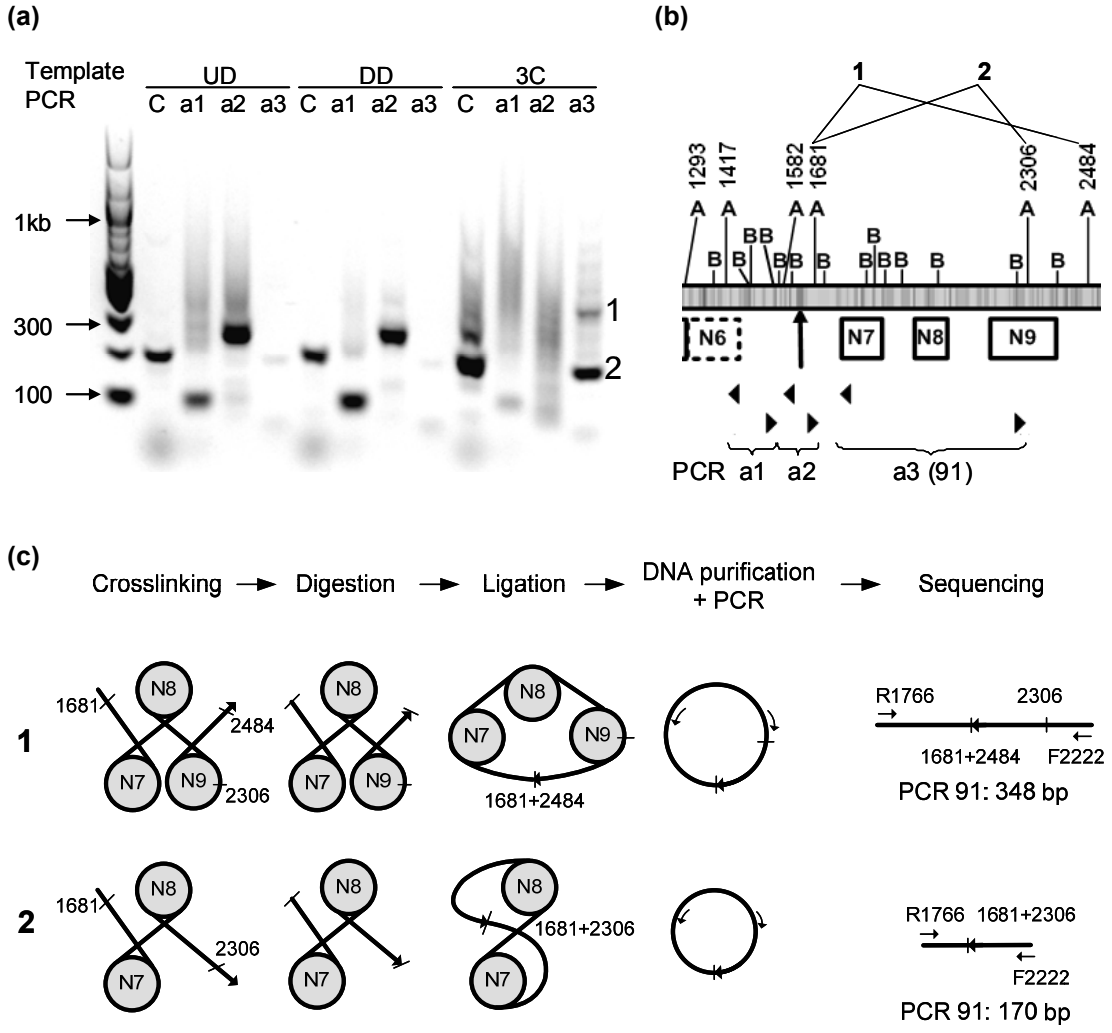


Figure 3-10. 3C-PCR products in the A-rich region. (a) The gel pattern. PCR a3 from 3C-DNA generates 2 distinct bands (marked * and **). (b) Sequential illustration of 3C-PCR products in the A-rich region. The ligated *Alu*I sites in indicated PCR products are connected by solid lines, respectively. The legends are as in Figure 3-7. (c) Sequence analysis of PCR products 1 and 2. The legends are as in Figure 3-2. This figure is not to scale.

The sequential illustration and analysis of the distinct PCR products are shown in Figure 3-10b and c. The sequence of product 1 revealed a conformation where the *Alu*I site at 2306 bp was protected and those at 1681 and 2464 bp exposed and recombined. This result is consistent with the product 1 from PCR 92 (Figure 3-9c), suggesting a nucleosome N9 protecting 2306 bp. The sequence of product 2 revealed a

conformation in PCR 91 (or L2 in PCR a3) where the *Alu*I site at 1681 and 2306 bp were exposed and recombined. It is consistent with PCR 92 and 93 that N9 may have dynamic positions where occasionally the *Alu*I site at 2306 bp was exposed (Figure 3-10b). No distinct product in PCR a1 and a2 was found, suggesting that the *Alu*I sites at 1417, 1582, and 1681 bp were not protected by proteins and not preserved after digestion. The *Alu*I site at 1681 bp was reconfirmed as an exposed site.

Surprisingly, some PCR products were found from both UD and DD where a 100-bp product was generated in PCR a1 and a 300-bp product in PCR a2. The lengths of these products did not match that of any predicted products, and the corresponding sequences revealed unspecific results. This finding was repetitively confirmed, so it was thought that these PCR products may come from unspecific primer annealing to chromosomal DNA in this region.

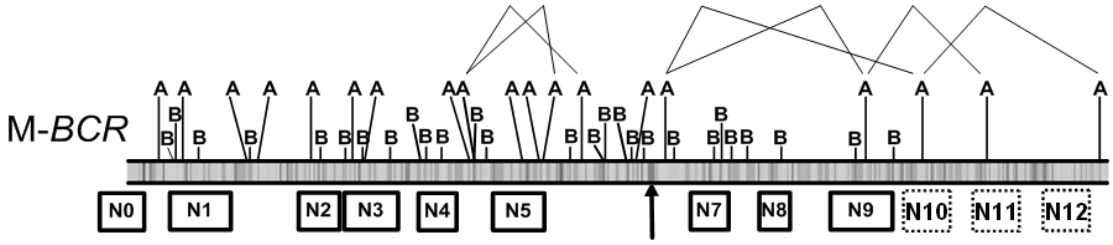


Figure 3-11. Chromosome conformation around the A-rich region in M-*BCR*. Ligated *Alu*I sites are connected with lines. Mapped nucleosome cores (open box), potential nucleosome cores (dotted box), 17-A (arrow) are indicated.

Overall, the chromosome conformation around the A-rich region revealed by the 3C-PCR is shown in Figure 3-11. The M-*BCR* sequence with the mapping of nucleosomes, breakpoints, *Alu* homologous sequence, and exons is shown in Figure 3-

12. The *Alu*I sites within the A-rich region were always digested, meaning that this region was not protected by histones. The *Alu*I sites immediately upstream or downstream to the A-rich regions were protected at least once in multiple PCR tests, indicating the existence of proteins (presumably histones) protecting the *Alu*I sites.

The fact that 3C-PCR was able to detect a loop formation in the N9 region but not a loop in the A-rich region may come from the following possibilities: (1) there was actually no loop in the A-rich region. Instead, some unknown secondary structure may have existed to prevent nucleosome formations. In fact, studies have shown that a DNA triplex may be formed between poly-A·poly-T and oligo-T (Beal and Dervan, 1991; Chen, 1991). The triplex does not bend toward the surface of histones within a nucleosome and thus may exclude nucleosome formations (Brown and Fox, 1998); (2) there was a loop within the A-rich region, but it was not detectable by the current method due to its small length. The loop formed by the A-rich region, if any, may have bounced back to be linear after the *Alu*I digestion separated the loop from the chromatin and released the tension. Since no protein was holding the curvature of the loop, it easily recovered to an energetically most favored state, a linear shape, for a short DNA fragment. It is not revealed with the current method if the 17-A naturally possessed a curvature. Even if it did, the curvature alone was not sufficient to facilitate the re-ligation of the *Alu*I sites within the A-rich region. In either possibility, the A-rich region has a unique conformational feature and with its nucleosome-excluding properties, it separated two packs of nucleosomes in M-*BCR*. According to the replication model proposed earlier and the unique M-*BCR* conformation revealed here,

mechanisms by which DNA breakage may occur during M-*BCR* replication were proposed and discussed.

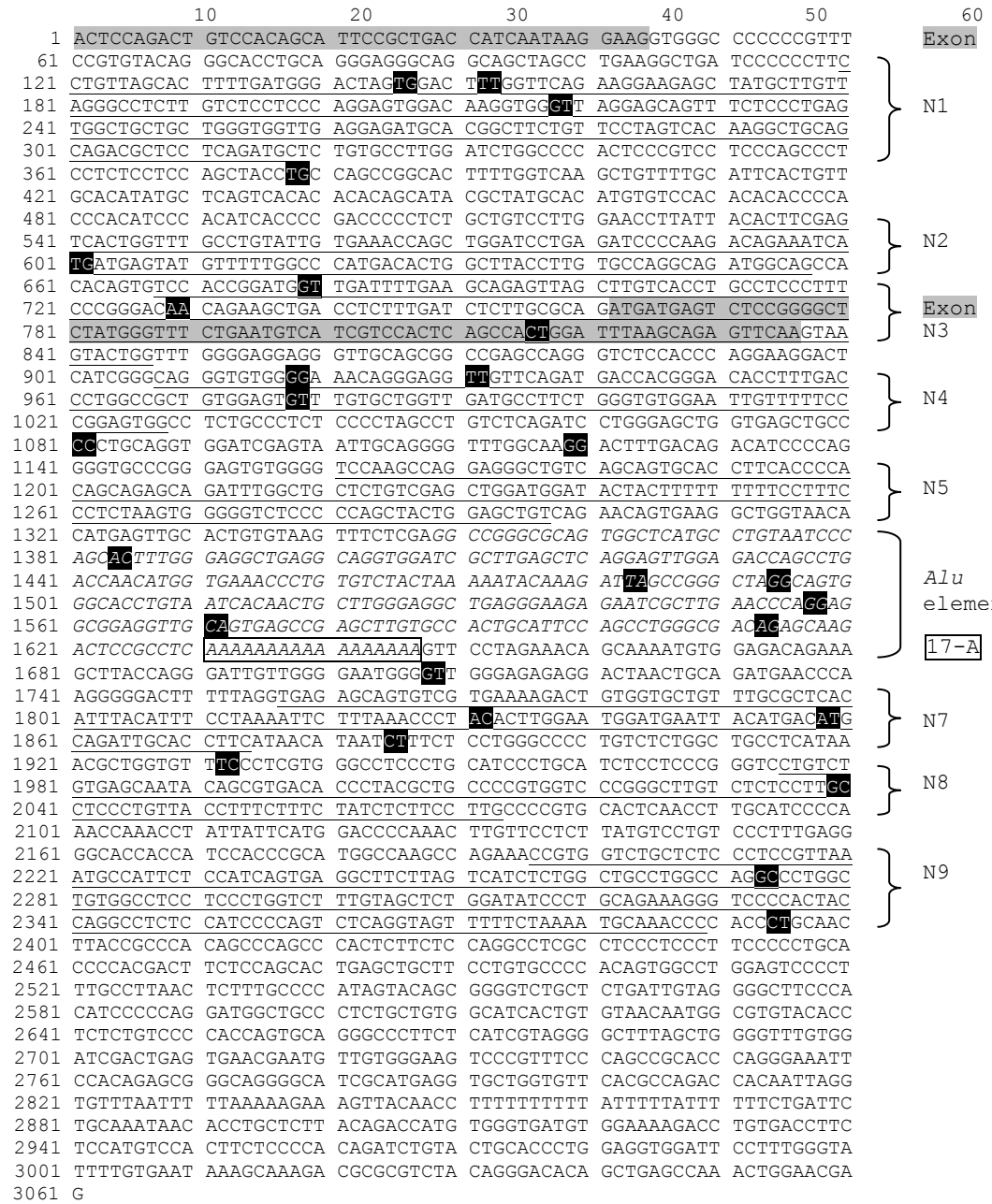


Figure 3-12. Sequence-level breakpoint diagram for M-*BCR*. Breakpoints are negatively inverted. Nucleosome core regions are underlined. The *Alu* sequence is Italic. Exons are shaded. The 17-A is boxed. The mapped regions are also indicated to the right.

3.5.5 Mechanisms for DNA Breakage during M-*BCR* Replication

Numerous mechanisms have been proposed for DNA breakage (van Gent *et al.*, 2001; Freudenreich, 2007). DNA double stranded breakage may be caused by exogenous agents (ionizing radiation, virus or carcinogens) or endogenous agents (free radicals or enzymes of DNA metabolism) (Raghavan and Lieber, 2006). Here, only the mechanisms that relate DNA breakage of M-*BCR* to its replication are considered. According to the nucleosome mapping in M-*BCR*, linker breakpoints are about as many as core breakpoints. Therefore, the mechanisms for DNA breakage in exposed and core regions, as well as the role of the A-rich region in either mechanism are discussed in detail as below.

3.5.5.1 DNA Breakage in the Nucleosome-excluded Regions

Compared to histone-protected core regions, nucleosome-excluded regions such as the A-rich region in M-*BCR* are prone to DSB (Freudenreich, 2007). A few chromosome fragile sites were found related to nucleosome-excluded regions (Wang and Griffith, 1996; Wang *et al.*, 1996; Hsu and Wang, 2002). Ideally, the compacted chromatin fiber arranges all linker DNA within the center of the fiber and covers them with nucleosome cores (see Figure I-4 in Introduction and Figure 3-13). However, in certain occasions the nucleosome-free DNA may be exposed to DSB-inducing agents or higher mechanical stress. These occasions are related to the specific M-*BCR* conformations and discussed as below.

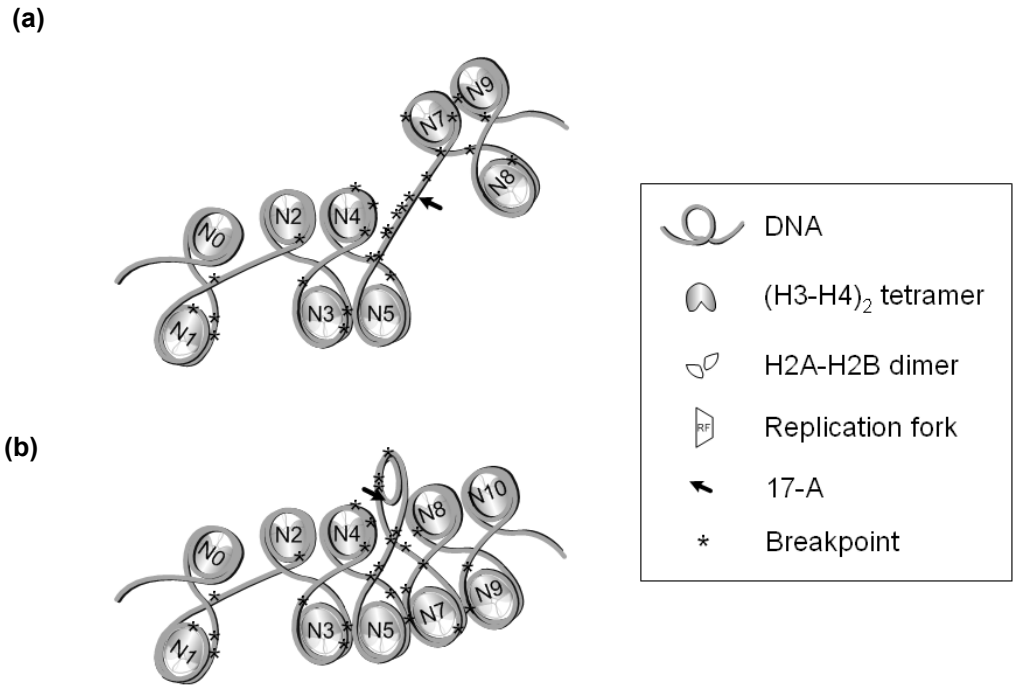


Figure 3-13. Possible configurations of the M-*BCR* chromatin fiber. The legends are shown in the box. (a) The A-rich region is straight between 2 packs of nucleosomes. (b) The A-rich region forms a loop and is packed in the chromatin fiber. Mapped nucleosomes are indicated. The drawing is not to scale.

First, in M-*BCR*, some nucleosome-free DNA regions are exposed from the chromatin fiber due to their extraordinary lengths (Figure 3-13). Among 13 linker breakpoints, 8 were in long linker regions (7 in the A-rich region, 1 in a 213-bp linker region between N1 and N2). The exposed DNA in these regions is much longer than that in an average linker region. When a long nucleosome-free region was packed into a chromatin fiber, it is possible that DNA in this region in part or as a whole extruded from the compacted chromatin. Even if the A-rich region may have folded into a loop (Figure 3-13b), the diameter of a 292 bp DNA circle (~32 nm) was still longer than the radius of a chromatin fiber (~15 nm), easily extruding the DNA from the packed chromatin. Bare DNA extruding from the chromatin is prone to attack by both

exogenous and endogenous agents any time in the cell cycle, including replication (Raghavan and Lieber, 2006). Therefore the possibility of DSB in these long nucleosome-excluded regions is higher than that in the average linker DNA.

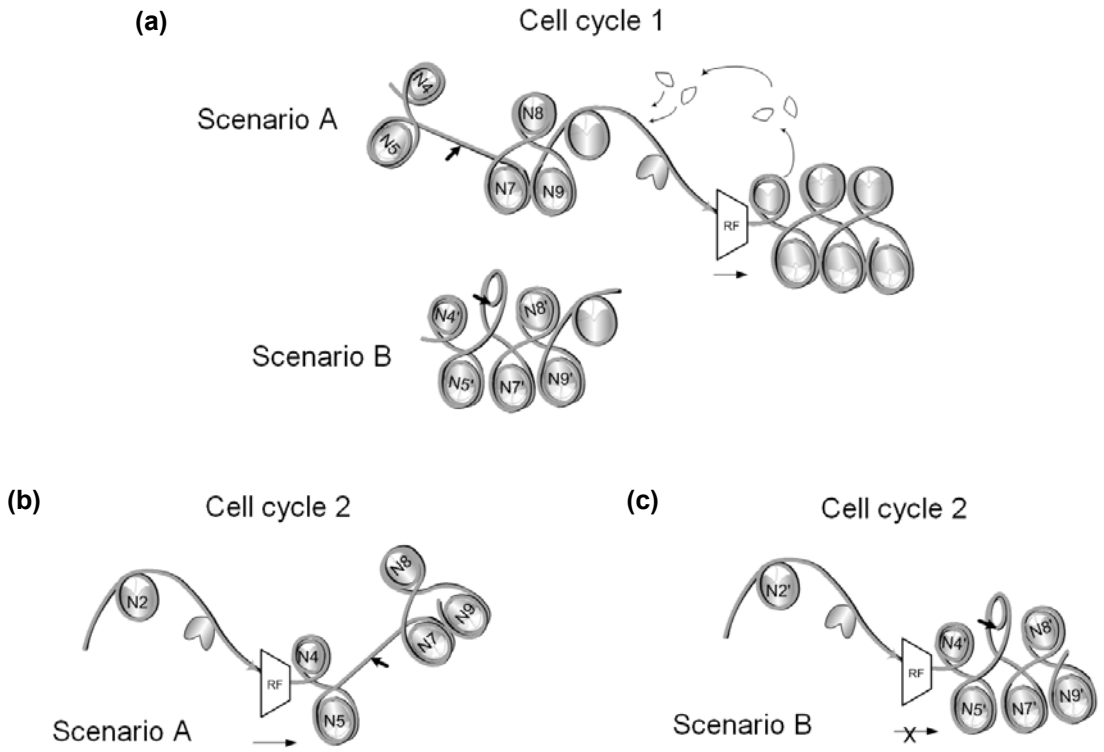


Figure 3-14. DSB in the A-rich region during nucleosome disassembly. The legends are as in Figure 3-12. (a) In scenario A, the A-rich region is replicated and packed as a straight linker between two nucleosome arrays (top). In scenario B, the A-rich region is replicated and packed as an entangled loop in the chromatin fiber (bottom). (b) In the next replication, the superhelical tension built up by DNA unwinding can be resolved in the straight A-rich region. (c) In scenario B, superhelical tension by DNA unwinding builds up in the tangled A-rich region and causes DSB upstream to or within the A-rich region. Only the leading strands are shown for clarity.

Second, DSB may preferentially occur in nucleosome-excluded regions within or upstream to the A-rich region where superhelical tension accumulates during replication (Figure 3-14). In the case when the A-rich region is packed without unusual conformations (Figure 3-14a, scenario A), the positive superhelical tension

generated by DNA unwinding at replication can be resolved in the A-rich region (Figure 3-14b). In other cases when the A-rich region is packed into a secondary structure (*e.g.*, a loop or a triplex, Figure 3-14a, scenario B), the positive superhelical tension accumulates upstream to the tangled A-rich region. This superhelical tension, if not resolved, may stall the advancing replication fork (Figure 3-14c). A stalled replication fork leads to DSB (Freudenreich, 2007; Clemente-Ruiz and Prado, 2009).

The DSB occurs preferentially upstream but not downstream to the A-rich region since (1) the superhelical tension is generated upstream to the A-rich region; (2) the downstream chromatin structure is relatively stable compared to the upstream one. Only one nucleosome immediately downstream to the fork is disturbed while up to two newly formed nucleosomes behind the fork are not stabilized (Gasser *et al.*, 1996). Therefore, both the source of the mechanical tension and the uneven stability of chromatin structure across the fork favored DSB within or upstream to the A-rich region. One may argue that topoisomerase II should have acted to untangle the tension, but the consequence remained the same that a DSB takes place upstream to or within the A-rich region.

The discussion above provides explanations to why 8 breakpoints were found in the nucleosome-free regions, and why they prefer to occur within or upstream to the A-rich region. Comparable to the linker regions, 14 breakpoints were also reported in the core regions upstream to the A-rich region. If DNA was protected by histones, why were 14 breakpoints found in the core regions? It is known that DNA bends significantly when it wraps around the histone core and the double helix is subjected

to a bending stress (Sivolob and Khrapunov, 1995; Segal *et al*., 2006). Is it possible that the bending stress may break the double helix? To answer this question, the position of core breakpoints was first analyzed to see whether there was a common breakpoint position around the histone core. The maximal bending stress on core DNA was then estimated to see if the level could break the phosphodiester bonds. Finally, mechanisms for core DNA breakage in M-*BCR* were proposed and discussed.

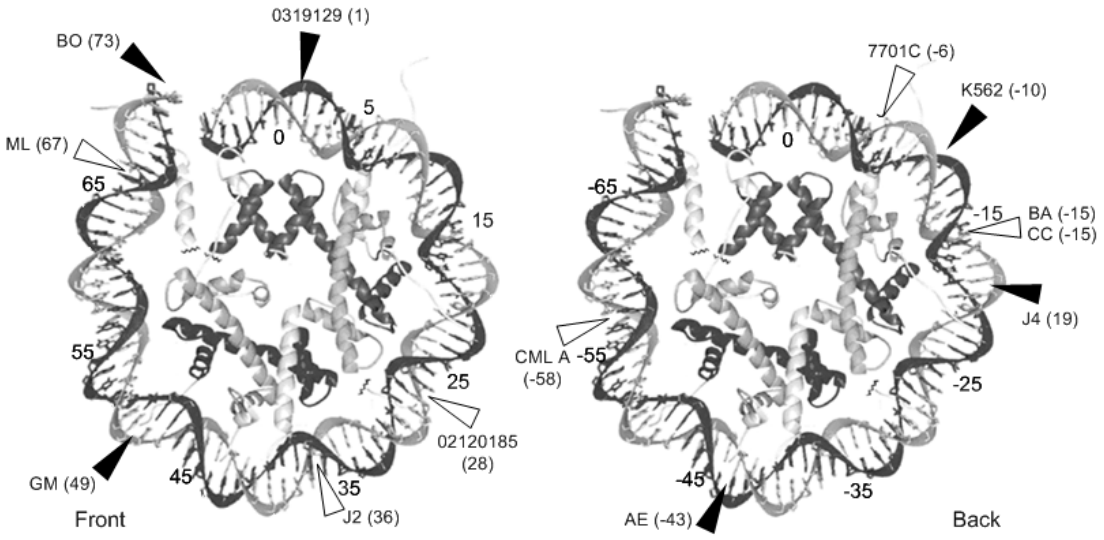


Figure 3-15. Positions of DNA breakpoints around the core. The front half of the nucleosome is on the left, and the back half is flipped horizontally on the right. Nucleotides wrapping around the histones are numbered for clarity (positive on the front; negative on the back). Six breakpoints are found near the peak of outer minor grooves (dark arrowheads) and seven are near the inner minor grooves (white arrowheads).

3.5.5.2 Positions of DNA Breakpoints around the Histone Core

The symmetry of nucleosome wrapping was utilized and the dyad (center of symmetrical wrapping, designated 0 in Figure 3-15) was placed in the middle of each core DNA. The positions of 13 core breakpoints (BO, ML, 0319129, 7701C, 02120185, CML A, GM, J2, K562, BA, AE, CC, and J4) were resolved from their

distances to the corresponding dyads and superimposed onto one virtual nucleosome core in Figure 3-15. The detail breakpoint positions of core DNA are listed in Table 3-3.

Table 3-3. Breakpoint positions around the histone core in M-*BCR*. The positions are indicated by the distance to the dyad (in bp) of a virtual nucleosome core.

Breakpoint ID	Nucleosome	Near a inner or outer minor groove?
BO	N1	Outer minor groove (73)
ML		Inner minor groove (67)
0319129	N2	Outer minor groove (1)
7701C		Inner minor groove (-6)
02120185	N3	Inner minor groove (28)
CML A		Inner minor groove (-58)
GM	N4	Outer minor groove (49)
J2		Inner minor groove (36)
K562		Outer minor groove (-10)
BA	N7	Inner minor groove (-15)
AE		Outer minor groove (-43)
CC	N8	Inner minor groove (-15)
J4	N9	Outer minor groove (19)

Seven breakpoints (ML, 7701C, 02120185, CML A, J2, BA, and CC) were near the peak of inner minor grooves, in which DNA binds histone proteins by electrostatic interaction (Luger and Richmond, 1998) and the DNA bending was facilitated. Six breakpoints (BO, 0319129, GM, K562, AE, and J4) were near the peak of outer minor grooves, in which DNA had little interaction with histone proteins and bending stress on DNA was more significant. The next step is therefore to analyze the level of bending stress in the outer minor grooves.

3.5.5.3 Maximal Bending Stress on DNA around the Histone Core

An average DNA sequence makes a smooth 90^o bend (a curved turn) about once every 200 base pairs (Alberts *et al.*, 2003c). However, a DNA sequence of only 147 bp long bends 1.65 turns to wrap onto a nucleosome. It implies that even with the electrostatic balance between the negatively charged DNA backbone and positively charged histones, core DNA is subjected to a level of tension in bending. Therefore, the topology of the DNA wrapping around the histone core was looked closely, and the local maximal stress (σ_{max}) induced by bending on the double helix was analyzed. The maximal stress was then converted to a force applied on the surface of the double helix to investigate its significance. The crystallographic study has provided an in-scale picture of the core DNA wrapping on the histone core (Figure 3-16a).

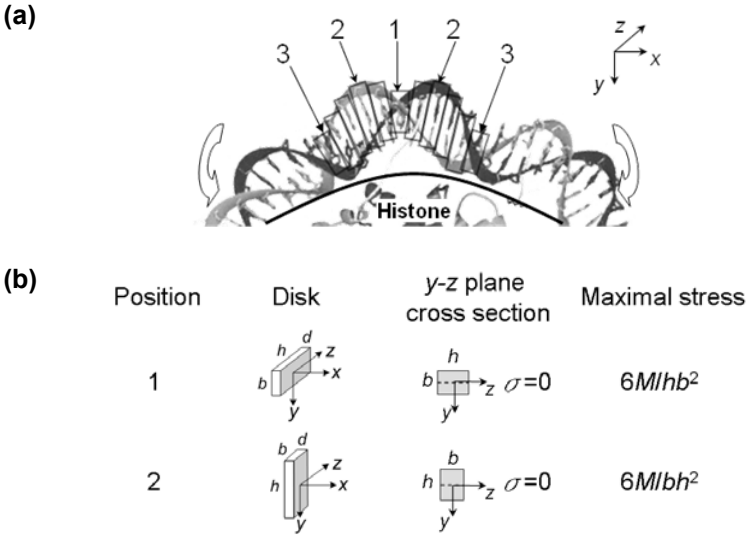


Figure 3-16. Bending of DNA and maximal stress on the double helix. (a) The x-y plane bending of DNA toward histones. Open arrows indicate the bending direction about z-axis, which points inside the page, on the x-y plane. DNA is modeled as stacked rectangular disks each of which rotate along the helix. Disks subjected to different levels of stresses are labeled in position 1 to 3. (b) Dimensions of the disks in position 1 and 2, the corresponding y-z plane cross sections and maximal stresses. The stress-free axis ($\sigma = 0$) is b/2 from the surface in position 1 and h/2 from the surface in position 2.

The double helix bends toward the histone core as it wraps around, and for every 10 bp the DNA makes contact at the inner minor groove with the histone proteins (position 3). At each inner minor groove, the attachment between the DNA and the histones is made via hydrogen bonds and salt bridges mostly between the peptide main chains of the histones and the phosphodiester bonds of the DNA (Luger and Richmond, 1998). Phosphodiester bonds in the inner minor grooves are more protected than those in the other positions based on the studies of DNase I digestion (Thoma, 1996), which shows this endonuclease preferentially cut the phosphodiester bonds at the outer minor grooves. To model the mechanical stress of bending on the core DNA, the focus was placed on the outer minor grooves (positions 1 and 2), where there is no direct attachment between DNA and the histone core. The DNA double helix was assumed to be a helicoidal beam. The helicoidal beam was set to bend in the *x*-*y* plane toward a histone core (Figure 3-16a). On the helical axis of the DNA, there was no stress of deformation. From this stress-free axis, one half of the beam facing outward was subjected to a stress of expansion, while the other half of the beam facing the histone core was subjected to a stress of compression. The beam was composed of stacked rectangular disks, each of which was defined by one base pair and the phosphodiester bonds connecting the neighboring nucleotides. Since base pairs rotate along the double helix, the stack rotates along its longitudinal axis and makes a turn every 10 disks. The stress σ (in pN/nm^2) along *x*-axis in the beam can be related to the bending moment *M* by $\sigma = My / I$, where *I* is the moment of inertia (hb^3/12 at position

1 and $bh^3/12$ at position 2, in nm^4) and y (in nm) is the distance from the stress-free axis (Gere and Timoshenko, 1984). When y is measured to the top or bottom surface of the cross section, the stress reaches its local maximum and can be calculated as:

$$\sigma_{\max,1} = 6M/b^2h \text{ at position 1} \qquad (5)$$

and

$$\sigma_{\max,2} = 6M/bh^2 \text{ at position 2} \qquad (6)$$

where h is the long dimension (2 nm) and b the short dimension (about 1 nm) of the cross section of double helix (Figure 3-16b). Now, consider the force acting on the top or bottom surface of outer minor grooves (position 1 and 2) due to bending. The surface area on which the force is acting is dh in position 1 and db in position 2 (d is the distance between adjacent nucleotides, see Figure 3-16b), thus the bending force acting on the surface is:

$$F_1 = \sigma_{\max,1} \cdot dh = \frac{6Md}{b^2} \text{ at position 1} \qquad (7)$$

And

$$F_2 = \sigma_{\max,2} \cdot db = \frac{6Md}{h^2} \text{ at position 2} \qquad (8)$$

Since the dimension b is twice the dimension h, F_1 is likely to be four times greater than F_2. However, since at position 1 both strands of the DNA share the same surface while at position 2 only one strand is on either top or bottom surface, the force acting on phosphodiester bonds at position 1 is likely to be twice as that at position 2 instead of four times.

According to Megson (2005), strain energy in bending (bending energy) is related to the bending angle θ and the bending moment M by:

$$dG = \frac{1}{2}\int M d\theta \qquad (9)$$

Since $d\theta$ is also related to the unit length of the bending DNA dx and the radius of the curvature R by $dx = Rd\theta$, the total bending energy of a core DNA section was:

$$\Delta G = \int dG = \frac{1}{2}\int M d\theta = \frac{M}{2}\int \frac{dx}{R} = \frac{ML}{2R} \qquad (10)$$

Take the nucleosome core bending energy in the M-*BCR* as an example. From bending energy estimation, it was learned that the average $\Delta G \approx 90\, kT$ in M-*BCR*. Knowing the value of the Boltzmann's constant (k = 1.38 × 10^{-23} J/K), and the body temperature (T = 310 K), the value of ΔG can be calculated. Further, given the length of a core DNA (L is 147 bp, about 49.6 nm long) and R = 4.3 nm, by using the equation (9) the typical bending moment of the core DNA in the *BCR* was calculated: M = 67.2 pN·nm. Substituting the value of M into equation (6) obtained: F_1 = 137 pN. The bending stress estimated here may be exaggerated due to the simplified small deformation model (Camerini-Otero and Felsenfeld, 1977), so the actual value may be smaller.

Is this force great enough to break the double helix? Studies showed that to unravel nucleosomes only 5 – 65 pN of stretching force is needed (Cui and Bustamante, 2000; Bennink *et al.*, 2001; Gemmen *et al.*, 2005). On the other hand, a wide range of forces were reported to break the double helix in different conditions (Bustamante *et al.*, 2000; Bensimon *et al.*, 1995). By bond potential theory, about

5000 pN is needed to break the double helix. Experiments have shown only 100 – 300 pN was needed to break the bulk DNA sheared in a flowing buffer (Bustamante *et al.*, 2000). Therefore the bending stress at certain positions around the histone core may be sufficient to cause instability of nucleosomes, but definitely not enough to break the double helix. This conclusion is expected since chromosomes are supposed to be stable structures. It implies that DNA breakage would not likely to occur simply by wrapping into a nucleosome. However, with the unique conformation in M-*BCR*, breakage of the core DNA may occur in rare occasions when histones are absent and the DNA is exposed. These occasions are detailed as below.

3.5.5.4 DNA Breakage in the Nucleosome Core Regions

First, a remote occasion exists when DNA is newly replicated and has not been rewrapped into a nucleosome core. This is a short period of time when core DNA is exposed to all attack, if exposed. When the cells, at various stages of differentiation including stem cells (Whetton and Graham, 1999), circulate in the blood stream, they may have to squeeze through narrow spaces, such as cell junctions. When the cells are compressed under these circumstances, mechanical stress applied to the nucleus may contribute to the curvature of the newly replicated DNA before it can be rewrapped by the histone core. If the curvature is not suitable for rewrapping, the nucleosome reassembly process would be stalled (Figure 3-17).

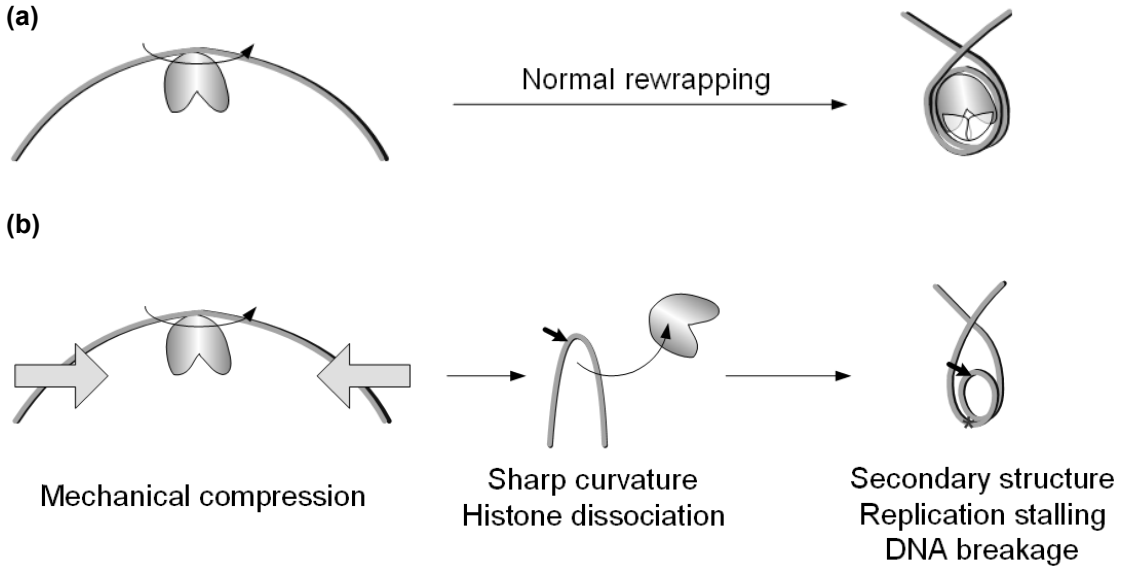


Figure 3-17. Core DNA breakage by mechanical compression. (a) A normal nucleosome rewrapping process. (b) During the rewrapping process, DNA has been subjected to mechanical compression. The sharp DNA curvature may not accommodate histones and lead to dissociation of the latter. The superhelical tension and mechanical stress leads to its own breakage and/or a secondary structure which may stall the replication. The legends are as in Figure 3-12.

The absence of histone protection and the increase of mechanical tension on DNA both elevate the possibility of DNA damage. Although this mechanism does not require specific chromosome conformation of M-*BCR*, and the magnitude of mechanical stress requires further verification, it remains a plausible possibility toward core DNA breakage during *BCR* replication.

Second, in rare occasions, the A-rich region may form a secondary structure. In the next replication (Figure 3-18a), the advancing fork upstream to the A-rich region may be stalled if the secondary structure is not untangled. The stalled fork may further disturb the upstream nucleosomes and cause dissociation of the histones from nucleosome formation (Figure 3-18b and c). A stalled fork was shown under electronic microscopy to regress for more than 400 bp (the length for 2 nucleosomes)

in vitro and formed a four-way junction of parental and daughter strands (Postow *et al.*, 2001; Sogo *et al.*, 2002; Michel *et al.*, 2004). The nucleosomes upstream to the fork (and the A-rich region) may therefore be disassembled to release DNA from the histones. The exposed core DNA is prone to attack by exogenous or endogenous agents that leads to DSB (Seigneur *et al.*, 1998; Freudenreich, 2007).

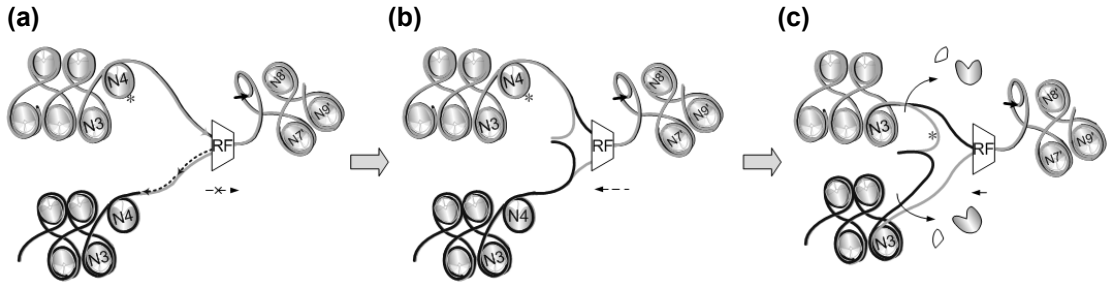


Figure 3-18. Regressing fork at the A-rich region. The legends are as in Figure 3-12. The replication direction is from left to right. A DNA breakpoint in N4 is indicated (*). (a) The replication fork is stalled (crossed arrow) in front of the A-rich region. (b) The fork regresses (broken arrow to the left), two parental strands reanneal, and two daughter strands anneal to form a four-way junction. (c) The fork further regresses and a four-way junction forms behind the fork. Nucleosomes upstream to the A-rich region (N4) are disturbed and disassembled. The breakpoint originally in the core region is now exposed to DSB attack.

This mechanism is thought plausible because (1) It is shown experimentally that a four-way junction may form behind a stalled fork (Postow *et al.*, 2001); (2) the A-rich region is known to form secondary structures (*e.g.*, a triplex or a loop) that has high potential to stall the fork *in vivo* (Brown and Fox, 1998; Baran *et al.*, 1991); (3) The core DNA breakpoints almost exclusively occurred upstream to the A-rich region, supporting that newly reassembled nucleosomes behind the fork are less stable; (4) DNA breakage in M-*BCR* at replication matches the time point when *BCR* is in close

proximity to *ABL*. Nevertheless, other occasions where core DNA may be exposed to DSB attack can not be excluded.

Ideally, DNA breakages are either repaired or lead to cell death. In rare cases, two independent DSBs at two different locations can join and result in chromosomal translocation (Raghavan and Lieber, 2004). In CML, it is thought that DNA repair mechanisms recombine the damaged *BCR* with nearby *ABL* by mistake and lead to aberrant translocation.

3.5.6 Implications of DNA Breakage Mechanisms in Other Cancerous Diseases

By studying the mechanisms for DNA breakage in M-*BCR*, it is learned that the A-rich region, which excludes nucleosomes and forms a secondary structure, plays a crucial role in inducing DNA damages in either nucleosome core or linker regions. In fact, this is consistent with the finding that chromosomal fragile sites are associated with expanded repeats capable of adopting unusual DNA structures that can perturb DNA replication (Zlotorynski *et al.*, 2003). Nucleosome-excluded sequences, such as triplex-forming poly-A tracts, cruciform DNA, quadruplex DNA, are all found associated with chromosome translocation sites (Raghavan and Lieber, 2006).

A recently studied example is a major breakpoint region (MBr) in *BCL2* gene, accounting for >95% follicular lymphomas (Raghavan and Lieber, 2006). Although the gene is over 200-kb long, about 75% of the translocations in this gene occur within the 150-bp MBr which forms secondary structures. Notably, triplex conformations are

found in two 17-bp GA-rich regions within MBr (Raghavan *et al.*, 2005). Interestingly, the RD algorithm predicted a 225-bp long nucleosome-free region right at MBr (Figure 3-19).

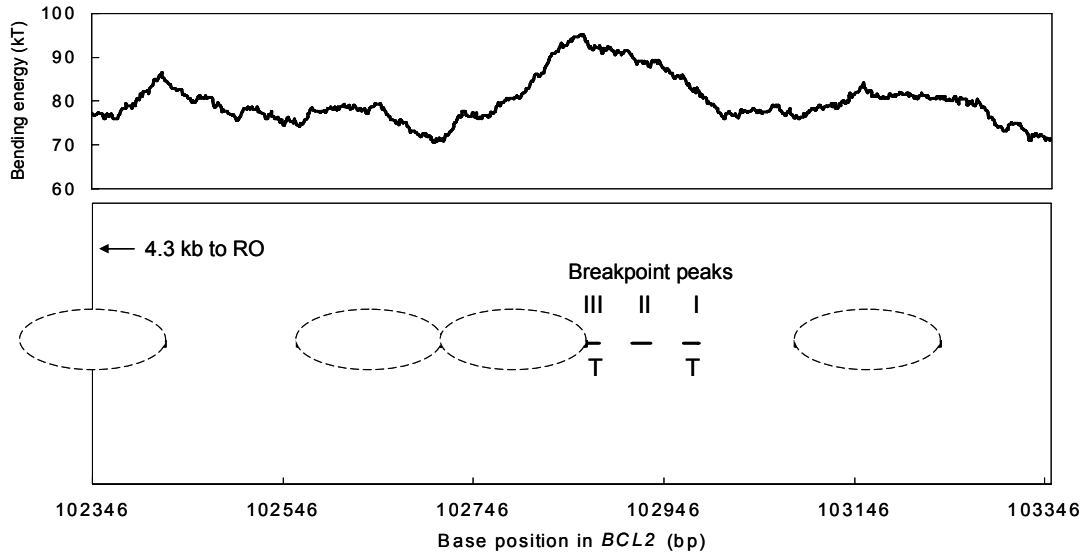


Figure 3-19. RD prediction of nucleosome positions in the major breakpoint region of *BCL2*. Bending energy is shown as a trend in the top panel. Predicted nucleosome cores (dotted oval) and breakpoint peaks (short horizontal bar) are shown in the bottom panel. The window length is 255 bp in the prediction. The peaks are numbered (I, II, III) in Raghavan *et al.*, 2005. The peak forming a triplex is indicated by a "T" underneath the bar.

Another example is *CCND1* gene, accounting for >90% mantle cell lymphoma (Raghavan and Lieber, 2006). Two DNA breakpoints were reported upstream to this gene (Tsujimoto *et al.*, 1985). Although the chromatin conformation in the breakpoint region has not been experimentally mapped, the RD algorithm again predicted a 207-bp long nucleosome-free region where the breakpoints have been identified (Figure 3-20).

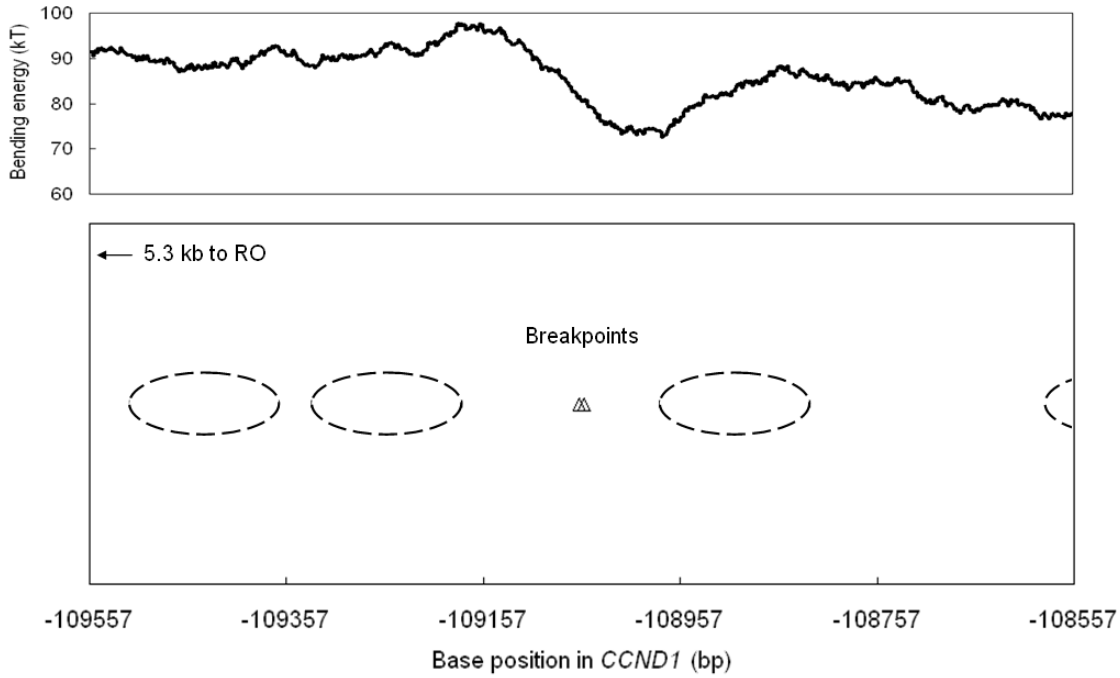


Figure 3-20. RD prediction of nucleosome positions in the breakpoint region of *CCND1*. Bending energy is shown as a trend in the top panel. Predicted nucleosome cores (dotted oval) and breakpoints (open triangle) are shown in the bottom panel. The window length is 255 bp in the prediction.

Indeed, if the mechanisms for DNA breakage in M-*BCR* can be further validated, these mechanisms may be common to some of the DNA breakages in many chromosomal fragile sites. Overall, the mechanisms revealed for DNA breakage in CML as well as other diseases show the importance of nucleosome formations in the integrity of chromosomes. Any sequence that prevents itself from nucleosome formations and forms secondary structures instead may inevitably be prone to DSB attack, either from biomechanical, biochemical, or viral stresses. In this sense, a histone core may actually be viewed as a "uniform curvature controllers" that functions to maintain a suitable and uniform curvature of DNA in chromatin compaction, and minimizes the possibility of DNA breakage.

3.6 Summary

In order to understand the mechanisms for DNA breakage in M-*BCR*, the distinct M-*BCR* chromosome conformation was incorporated into the stepwise nucleosome assembly model at DNA replication. A 3C-PCR method was used to resolve the M-*BCR* conformation. The result revealed a unique feature that a 292 bp A-rich region was nucleosome-excluded between N5 and N7. This result was consistent with our previous nucleosome mapping. Combining the M-*BCR* conformation and the stepwise replication model, mechanisms were proposed to explain how DNA breakages occurred in the nucleosome-excluded regions as well as in the nucleosome core regions. For nucleosome-excluded regions, bare DNA in extraordinary lengths extruded from the chromatin fiber and is exposed to attack by exogenous or endogenous agents. The replication fork stalled by the secondary structure combined with superhelical tension upstream to the A-rich region favored DNA breakage within or upstream to the A-rich region. Core DNA may also break due to the exposure from dissociated nucleosome cores upstream to the A-rich region. Overall, the finding in this study is consistent with current knowledge that genomic sequences capable of forming unusual DNA structures that affect replication are highly correlated with chromosomal fragile sites.

Chapter 3, in part is currently being prepared for submission for publication of the material. Tu, Chi-Chiang; Sung, Lanping A. The dissertation author was the primary investigator and author of this material.

Chapter 4 Future Work

This study focused on the nucleosome organization and structural change in the M-*BCR* region of *BCR* where DNA breakpoints mainly clustered. In the future, it would be interesting to explore the chromosome structure in the other regions of *BCR* (such as m-*BCR* and μ-*BCR*) if precise breakpoint locations are available, as well as the counterpart *ABL* gene in the Ph' translocation. It is known that *ABL* gene has a much lower GC ratio (44.3%), and sequence analysis has shown a distribution of A-rich loci more frequent than in M-*BCR*. Whether a same conformational feature is shared among breakpoints in different loci of *BCR* would enable a global view to DNA breakage mechanisms in *BCR*; how multiple A-rich regions are accommodated in the nucleosome organization of the low GC *ABL* and what structural features are shared by DNA breakpoints in *ABL* would provide fundamental information to understanding the DNA breakage mechanisms in *ABL*.

In the future, it would also be interesting to explore the chromosome structure of *BCR*, *ABL*, and *BCR*/*ABL* fusion chromosome of CML patients. It is expected that the product of aberrant chromosome translocation would not survive the DNA damage and replication checkpoint mechanisms. However, Ph' chromosomes survive in CML patients through generations of cell cycles. The chromosome structure in *BCR*/*ABL* fusion region would provide a possible explanation of how this illegitimate recombination avoided detection and deletion by the DNA damage checkpoint

mechanisms, and help determining the temporal order of Ph' translocation from other genetic disorders in CML.

A human leukemia cell line (HL-60) was used in this study to mimic the clonal origin of CML. To complete the research in terms of leukemogenesis origin, it would be necessary to continue the exploration in chromosome structure upstream to human hematopoietic stem cells and downstream to differentiated human myeloid cells. The change in chromosome structure of *BCR* and *ABL* in the course of differentiation may bring us one step closer in finding the exact differentiation stage of myeloid cells when Ph' translocation most likely to occur and CML to initiate.

We proposed a RD algorithm to improve the accuracy in nucleosome position predictions. This algorithm is far from maturation. In the future, more indexes than only DNA bending energy (for example, consensus protein-binding sequences or secondary structure formation other than a nucleosome core) may be included to consider and determine a preferred nucleosome core position. The completeness of this algorithm would depend a lot on our understanding of specific structural features in the gene of our interest. If in the future these gene-specific features can be collected in a database, then the RD algorithm can adapt them to predict nucleosome organizations even more accurately.

This study laid foundation in a new direction to studying the generation of cancerous diseases. As far as we know, many cancers and inherited diseases originate from illegitimate chromosome translocations just like CML, and studies have gathered the precise DNA breakpoint locations in these diseases (Abeysinghe *et al.*, 2003;

Chuzhanova *et al.*, 2003). It would be of much significance if we could extend our exploration to the genes involved in these diseases. We may be able to find a common mechanism to trigger aberrant chromosome translocation due to the specific structural weakness of the chromosomes, and to that, a possible preventive or therapeutic treatment.

Finally, this study proposed stepwise nucleosome disassembly/reassembly model at the replication fork based on findings from the literatures. In the future, with permission of equipment and financial resources, we may be able to monitor the real-time stepwise nucleosome assembly mechanism by applying fluorescence resonance energy transfer (FRET) methods. Studies using stopped-flow FRET (Tims and Widom, 2007) have shown the capability to reveal stepwise nucleosome remodeling dynamics. If we fluorescently differentiate the labels on DNA, histone tetramers and dimers, we can apply the same technique to reveal the steps of nucleosome assembly so as to answer fundamental questions such as: does the nucleosome loop complete before or after dimer binding? Does the tetramer bind to DNA by pure bending or a flipping of the DNA into a loop? The same technique may also be used to investigate and characterize the helicase DNA unwinding mechanism proposed in this study, which has not been resolved by traditional technologies.

References

Abeysinghe, S.S., Chuzhanova, N., Krawczak, M., Ball, E.V., Cooper D.N. (2003) "Translocation and gross deletion breakpoints in human inherited disease and cancer I: nucleotide composition and recombination-associated motifs." Hum. Mutat. 22: 229–244.

Alberts, B., Johnson, A., Lewis, J., Raff, M., Roberts, K., Walter, P. (2003) Molecular Biology of the Cell. Garland Science, London, UK. 4th Ed: a: pp.211; b: pp.400; c: pp.314; d: pp.315.

Almouzni, G., Clark, D.J., Méchali, M., Wolffe, A.P. (1990) "Chromatin assembly on replicating DNA *in vitro*." Nucl. Acids Res. 18: 5767–5774.

Anastasi, J., Moinuddin, R., Daugherty, C. (1999) "The juxtaposition of *ABL* with *BCR* and risk for fusion may come at the time of *BCR* replication in late S-phase." Blood 94: 1137–1138.

Annunziato, A.T. (1989) "Inhibitors of topoisomerases I and II arrest DNA replication, but do not prevent nucleosome assembly *in vivo*." J. Cell Sci. 93: 593–603.

Baran, N., Lapidot, A., Manor, H. (1991) "Formation of DNA triplexes account for arrests of DNA synthesis at d(TC)n and d(GA)n tracts." Proc. Natl. Acad. Sci. USA 88: 507–511.

Beal, P.A., Dervan, P.B. (1991) "Second structural motif for recognition of DNA by oligonucleotide-directed triple-helix formation." Science 251: 1360–1363.

Bednar, J., Horowitz, R.A., Grigoryev, S.A., Carruthers, L.M., Hansen, J.C., Koster, A.J., Woodcock, C.L. (1998) "Nucleosomes, linker DNA, and linker histone form a unique structural motif that directs the higher-order folding and compaction of chromatin." Proc. Natl. Acad. Sci. USA 95: 14173–14178.

Bennink, M.L., Leuba, S.H., Leno, G.H., Zlatanova, J., de Grooth, B.G., Greve, J. (2001) "Unfolding individual nucleosomes by stretching single chromatin fibers with optical tweezers." Nat. Struct. Biol. 8: 606–610.

Bensimon, D., Simon, A.J., Croquette, V.V., Bensimon, A. (1995) "Stretching DNA with a receding meniscus: Experiments and models." Phys. Rev. Lett. 74: 4754–4757.

Bermejo, R. Doksani,Y., Capra, T., Katou, Y., Tanaka, H., Shirahige, K., Foiani, M. (2007) "Top1- and Top2-mediated topological transitions at replication forks ensure fork progression and stability and prevent DNA damage checkpoint activation." Genes & Dev. 21: 1921–1936.

Biernaux, C., Loos, M., Seis, A., Huez, G., Stryckmans, P. (1995) "Detection of major *bcr-ab1* gene expression at a very low level in blood cells of some healthy individuals." Blood 86: 3118–3122.

Bloomfield, V.A., Crothers, D.M., Tinoco, I. (1974) Physical Chemistry of Nucleic Acids, Harper and Row Publishers, New York.

Boyes, J., Omichinski, J., Clark, D., 2, Pikaart, M., Felsenfeld, G. (1998) "Perturbation of nucleosome structure by the erythroid transcription factor GATA-1." J. Mol. Biol. 279: 529–544.

Bryant, Z., Stone, M.D., Gore, J., Smith, S.B., Cozzarelli, N.R., Bustamante, C. (2003) "Structural transitions and elasticity from torque measurements on DNA." Nature 424: 338–341.

Breitman, T.R., Selonick, S.E., Collins, S.J. (1980) "Induction of differentiation of the human promyelocytic leukemia cell line (HL-60) by retinoic acid." Proc. Natl. Acad. Sci. USA 77: 2936–2940.

Brower-Toland, B.D., Smith, C.L., Yeh, R.C., Lis, J.T., Peterson, C.L., Wang, M.D. (2002) "Mechanical disruption of individual nucleosomes reveals a reversible multistage release of DNA." Proc. Natl. Acad. Sci. USA 99: 1960–1965.

Brown, P.M., Fox, K.R. (1998) "DNA triple-helix formation on nucleosome-bound poly(dA)·poly(dT) tracts." Biochem. J. 333: 259–267.

Bustamante, C., Smith, S.B., Liphard, J., Smith, D. (2000) "Single-molecule studies of DNA mechanics." Curr. Opin. Struct. Biol. 10: 279–285.

Camerini-Otero, R.D., Felsenfeld, G. (1977) "Supercoiling energy and nucleosome formation: the role of the arginine-rich histone kernel." Nucl. Acids Res. 4: 1159–1182.

Chen, F.M. (1991) "Intramolecular triplex formation of the purine-purine-pyrimidine type." Biochemistry 30: 4472–4479.

Chissoe, S.L., Bodenteich, A., Wang, Y.F., Wang, Y.P., Burian, D., Clifton, S.W., Crabtree, J., Freeman, A., Iyer, K., Jian, L., Ma, Y., Mclaury, H.J., Pan, H.Q., Sarhan, O., Toth, S., Wang, Z., Zhang, G., Heisterkamp, N., Groffen, J. and Roe, B.A. (1995) "Sequence and analysis of the human *ABL* gene, the *BCR* gene, and regions involved in the Philadelphia chromosomal translocation." Genomics 27: 67–82.

Chuzhanova, N., Abeysinghe, S.S., Krawczak, M., Cooper, D.N. (2003) "Translocation and gross deletion breakpoints in human inherited disease and cancer II: potential involvement of repetitive sequence elements in secondary structure formation between DNA ends." Hum. Mutat. 22: 245–251.

Clarkson, B., Strife, A., Wisniewski, D., Lambeck, C.L., Liu, C. (2003) "Chronic myelogenous leukemia as a paradigm of early cancer and possible curative strategies." Leukemia 17: 1211–1262.

Clemente-Ruiz, M., Prado, F. (2009) "Chromatin assembly controls replication fork stability." EMBO Rep. 10: 790–796.

Coll, C.M., Frederick, C.A., Wang, A.H.J., Rich, A. (1987) "A bifurcated hydrogen-bonded conformation in the d(A·T) base pairs of the DNA dodecamer d(CGCAAATTTGCG) and its complex with distamycin." Proc. Natl. Acad. Sci. U.S.A. 84: 8385–8389.

Collins, S.J. (1987) "The HL-60 promyelocytic leukemia cell line: proliferation, differentiation, and cellular oncogene expression." Blood 70: 1233–1244.

Collins, S., Coleman, H., Goundine, M. (1987) "Expression of bcr and bcr-abl fusion transcripts in normal and leukemic cells." Mol. Cell. Biol. 7: 2870–2876.

Cui, Y., Bustamante, C. (2000) "Pulling a single chromatin fiber reveals the forces that maintain its higher-order structure." Proc. Natl. Acad. Sci. USA 97: 127–132.

Cusick, M.E., DePamphilis, M.L., Wassarman, P.M. (1984) "Dispersive segregation of nucleosomes during replication of SV40 chromosomes." J. Mol. Biol. 178: 249–271.

Davey, C.A., Sargent, D.F., Luger, K., Maeder A.W., Richmond, T.J. (2002) "Solvent mediated interactions in the structure of the nucleosome core particle at 1.9A resolution." J. Mol. Bio. 319: 1097–1113.

Dekker, J., Rippe, K., Dekker, M., Kleckner, N. (2002) "Capturing chromosome conformation." Science 295: 1306–1311.

de Klein, A., Geurts van Kessel, A., Grosveld, G., Bartram, C., Hagemeijer, A., Bootsma, D., Spurr N.K., Heisterkamp, N., Groffen, J., and Stephenson, J.R. (1982) "A cellular oncogene is translocated to the Philadelphia chromosome in chronic myelocytic leukemia." Nature 300: 765–767.

Dennis, J.H., Fan, H.Y., Reynolds, S.M., Yuan, G., Meldrim, J.C., Richter, D.J., Peterson, D.G., Rando, O.J., Noble, W.S., Kingston, R.E. (2007) "Independent and complementary methods for large-scale structural analysis of mammalian chromatin." Genome Res. 17: 928–939.

Dobbs, D.L., Shaiu, W.L., Benbow, R.M. (1994) "Modular sequence elements associated with origin regions in eukaryotic chromosomal DNA." Nucl. Acids Res. 22: 2479–2489.

Donmez, I., Patel, S.S. (2006) "Mechanisms of a ring shaped helicase." Nucl. Acid Res. 34: 4216–4224.

Dong, F., van Holde, K.E. (1991) "Nucleosome positioning is determined by the $(H3\text{-}H4)_2$ tetramer." Proc. Natl. Acad. Sci. USA. 88: 10596–10600.

Drew, H.R., Dickerson, R.E. (1981) "Structure of a B-DNA dodecamer III. Geometry of hydration" J. Mol. Biol. 151: 535–556.

Drew, H.R., Travers, A.A. (1985) "DNA bending and its relation to nucleosome positioning." J. Mol. Bio. 186: 773–790.

Enemark, E.J., Joshua-Tor, L. (2006) "Mechanism of DNA translocation in a replicative hexameric helicase." Nature 442: 270–275.

Faderl, S., Talpaz, M., Estrov, Z., O'Brien, S., Kurzrock, R., Kantarjian, H.M. (1999) "The biology of chronic myeloid leukemia." N. Engl. J. Med. 341: 164 – 172.

Fialkow, P.J., Jacobson, R.J., Papayannopoulou, T. (1977) "Chronic myelocytic leukemia: clonal origin in a stem cell common to the granulocyte, erythrocyte, platelet and monocyte/macrophage." Am. J. Med. 63: 125–130.

Fisher, A.M., Strike, P., Scott, C., Moorman, A.V. (2005) "Breakpoints of variant 9;22 translocations in chronic myeloid leukemia locate preferentially in the CG-richest regions of the gene." Genes, Chromosomes & Cancer 43: 383–389.

Flaus, A., Luger, K., Tan, S., Richmond, T.J. (1996) "Mapping nucleosome position at single base-pair resolution by using site-directed hydroxyl radicals." Proc. Natl. Acad. Sci. USA 93: 1370–1375.

Fotedar, R., Roberts, J.M. (1989) "Multistep pathway for replication-dependent nucleosome assembly." Proc. Natl. Acad. Sci. USA 86: 6459–6463.

Freudenreich, C.H. (2007) "Chromosome fragility: molecular mechanisms and cellular consequences." Front. Biosci. 12: 4911–4924.

Gallagher, R., Collins, S., Trujillo, J., McCredie, K., Ahearn, M., Tsai, S., Metzgar, R., Aulakh, G., Ting, R., Ruscetti, F., Gallo, R. (1979) "Characterization of the continuous, differentiating myeloid cell line (HL-60) from a patient with acute promyelocytic leukemia." Blood 54: 713–733.

Gambus, A., Jones, R.C., Sanchez-Diaz, A., Kanemaki, M., van Deursen, F., Edmondson, R.D., Labib, K. (2006) "GINS maintains association of Cdc45 with MCM in replisome progression complexes at eukaryotic DNA replication forks." Nat. Cell Biol. 8:358–366.

Gasser, R., Koller, T., Sogo, J.M. (1996) "The stability of nucleosomes at the replication fork." J. Mol. Bio. 258: 224–239.

Gemmen, G.J., Sim, R., Haushalter, K.A., Ke, P.C., Kadonaga, J.T., Smith, D.E. (2005) J. Mol. Biol. 351: 89–99.

Gere, J.M., Timoshenko, S.P. (1984) Mechanics of materials. PWS-Kent, Boston, MA. 2nd Ed: 214.

Gotoh, O., Tagashira, Y. (1981) "Stabilities of nearest neighbor doublets in double-helical DNA determined by fitting of calculated melting profiles to observed profiles." Biopolymers 20: 1033–1042.

Griffith, J., Bleyman, M., Rauch, C.A., Kitchin, P.A., Englund, P.T. (1986) "Visualization of the bent helix in kinetoplast DNA by electron microscopy." Cell 46: 717–724.

Groffen, J., Stephenson, J.R., Heisterkamp, N., de Klein, A., Bartram, C.R., Grosveld, G. (1984) "Philadelphia chromosomal breakpoints are clustered within a limited region, bcr, on chromosome 22." Cell 36: 93–99.

Groth, A., Rocha, W., Verreault, A., Almouzni, G. (2007) "Chromatin challenges during DNA replication and repair." Cell 128: 721–733.

Gruss, C., Wu, J., Koller, T., Sogo, J.M. (1993) "Disruption of the nucleosomes at the replication fork." EMBO J. 12: 4533–4545.

Hagège, H., Klous, P., Braem, C., Splinter, E., Dekker, J., Cathala, G., de Laat, W., Forne, T. (2007) "Quantitative analysis of chromosome conformation capture assays (3C-qPCR)." Nat. Protoc. 2: 1722–1733.

Hagerman, P.J. (1988) "Flexibility of DNA." Ann. Rev. Biophys. Biophys. Chem. 17: 265–286.

Hayes, J.J., Clark, D.J., Wolffe, A.P. (1991) "Histone contributions to the structure of DNA in the nucleosome." Proc. Natl. Acad. Sci. USA. 88: 6829–6833.

Heisterkamp, N., Stephenson, J.R., Groffen, J., Hansen, P.F., de Klein, A., Bartram, C.R., Grosveld, G. (1983) "Localization of the c-abl oncogene adjacent to a translocation break point in chronic myelocytic leukemia." Nature 306: 239–242.

Heisterkamp, N., Stam, K. and Groffen, J. (1985) "Structural organization of the bcr gene and its role in the Ph' translocation." Nature 315: 758–761.

Hsu, Y.Y., Wang, Y.H. (2002) "Human fragile site FRA16B DNA excludes nucleosomes in the presence of distamycin." J. Biol. Chem. 277:17315–17319.

Huntly, B.J.P., Guilhot, F., Reid, A.G., Vassiliou, G., Hennig, E., Franke, C., Byrne, J., Brizard, A., Niederwieser, D., Freeman-Edward, J., Cuthbert, G., Bown, N., Clark, R. E., Nacheva, E.P., Green, A.R., Deininger, M.W.N. (2003) "Imatinib improves but may not fully reverse the poor prognosis of patients with CML with derivative chromosome 9 deletions." Blood 102: 2205–2212.

Jeffs, A.R., Benjes, S.M., Smith, T.L., Sowerby, S.J., Morris, C.M. (1998) "The *BCR* gene recombines preferentially with Alu elements in complex *BCR–ABL* translocations of chronic myeloid leukaemia." Hum. Mol. Genet. 7: 767–776.

Jeffs, A.R., Wells, E., Morris, C.M. (2001) "Nonrandom distribution of interspersed repeat elements in the *BCR* and *ABL1* genes and its relation to breakpoint cluster regions." Genes, Chromosomes & Cancer 32: 144–154.

Jelinek, W.R., Schmid, C.W. (1982) "Repetitive sequences in eukaryotic DNA and their expression." Ann. Rev. Biochem. 51: 813–844.

Kaplan, D.L., Davey, M.J., O'Donnell, M. (2003) "Mcm4,6,7 uses a 'pump in ring' mechanism to unwind DNA by steric exclusion and actively translocate along a duplex." J. Bio. Chem. 278: 49171–49182.

Kaplan, D.L., O'Donnell, M. (2002) "DnaB drives DNA branch migration and dislodges proteins while encircling two DNA strands." Mol. Cell 10: 647–657.

Khrapunov, S.N., Dragan, A.I., Sivolob, A.V., Zagariya, A.M. (1997) "Mechanisms of stabilizing nucleosome structure. Study of dissociation of histone octamer from DNA." Biochim. Biophys. Acta. 1351: 213–222.

Kim, J.L., Nikolov, D.B., Burley, S.K. (1993) "Co-crystal structure of TBP recognizing the minor groove of a TATA element." Nature 365: 520–527.

Kim, A.R., Murray, V. (2001) "Chromatin structure at the 3'-boundary of the human beta-globin locus control region hypersensitive site-2." Int. J. Biochem. 33: 1183–1192.

Kiyama, R., Trifonov, E.N. (2002) "What positions nucleosomes? – a model." FEBS Letter 523: 7–11.

Kohany O., Gentles A.J., Hankus L., Jurka J. (2006) "Annotation, submission and screening of repetitive elements in Repbase: RepbaseSubmitter and Censor." BMC Bioinformatics 7:474.

Komura, J., Ono, T. (2003) "Nucleosome positioning in the human c-FOS promoter analyzed by in vivo footprinting with psoralen and ionizing radiation." Biochemistry 42: 15084–5091.

Kozubek, S., Lukášová, E., Rýznar, L., Kozubek, M., Lišková, A., Govorun, R.D., Krasavin, E.A., Horneck, G. (1997) "Distribution of *ABL* and *BCR* genes in cell nuclei of normal and irradiated lymphocytes." Blood, 89: 4537-4545.

Krude, T. (1999) "Chromatin assembly during DNA replication in somatic cells." Eur. J. Biochem. 263: 1–5.

Kunkel, G.R., Martinson, H.G. (1981) "Nucleosomes will not form on double-stranded RNA or over poly(dA)·poly(dT) tracts in recombinant DNA." Nucleic Acids Res. 9: 6869–6888.

Laurent, E., Talpaz, M., Kantarjian, H., Kurzrock. R. (2001) "The BCR gene and Philadelphia chromosome-positive leukemogenesis." Cancer Res. 61: 2343–2355.

Lemoine, F.J., Degtyareva, N.P., Lobachev, K., Petes, T.D. (2005) "Chromosomal translocations in yeast induced by low levels of DNA polymerase: a model for chromosome fragile sites." Cell 120: 587–598.

Leuba, S.H., Karymov, M.A., Tomschik, M., Ramjit, R., Smith, P., Zlatanova, J. (2003) "Assembly of single chromatin fibers depends on the tension in the DNA molecule: magnetic tweezers study." Proc. Natl. Acad. Sci. USA. 100: 495–500.

Levchenko, V., Jackson, B., and Jackson, V. (2005) "Histone release during transcription: displacement of the two H2A-H2B dimers in the nucleosome is dependent on different levels of transcription-induced positive stress." Biochemistry 44: 5357–5372.

Levitsky, V.G. (2004) "RECON: a program for prediction of nucleosome formation potential." Nucl. Acids Res. 32: W346–W349.

Li, G., Widom, J. (2004) "Nucleosomes facilitate their own invasion." Nat. Struct. Mol. Biol. 11: 763–769.

Lucchini, R., Wellinger, R.E., Sogo, J.M. (2001) "Nucleosome positioning at the replication fork." EMBO J. 20: 7294–7302.

Luger, K. (2003) "Structure and dynamic behavior of nucleosomes." Curr. Opin. Genet. Dev. 13: 127–135.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., Richmond, T.J. (1997) "Crystal structure of the nucleosome core particle at 2.8 A resolution." Nature 389: 251–260.

Luger, K., Richmond, T.J. (1998) "DNA binding within the nucleosome core." Curr. Opin. Struct. Biol. 8: 33–40.

Matson, S.W., Bean, D.W., George, J.W. (1994) "DNA helicases: enzymes with essential roles in all aspects of DNA metabolism." Bioessays 16: 13–22.

Mazurkiewicz, J., Kepert, J.F., Rippe, K. (2006) "On the mechanism of nucleosome assembly by histone chaperone NAP1." J. Bio. Chem. 281: 16462–16472.

McGeoch, A.T., Trakselis, M.A., Laskey, R.A., Bell, S.D. (2005) "Organization of the archaeal MCM complex on DNA and implications for the helicase mechanism." Nat. Struct. Mol. Biol. 12: 756–762.

Megson, T.H.G. (2005) Structural and Stress Analysis. Elsevier, MA. 2nd Ed: 226.

Meyer, G.A., Radsak, K.D. (2000) "Identification of a novel signal sequence that targets transmembrane proteins to the nuclear envelope inner membrane." J. Bio. Chem. 275: 3857–3866.

Michel, B.D., Grompone, G., Florès, M., Bidnenko, V. (2004) "Multiple pathways process stalled replication forks." Proc. Natl. Acad. Sci. USA 101: 12783–12788.

Mosconi, F., Allemand, J.F., Bensimon, D., Croquette, V. (2009) "Measurement of the torque on a single stretched and twisted DNA using magnetic tweezers." Phys. Rev. Lett. 102: 078301.

Nedelcheva-Veleva, M.N., Krastev, D.B., Stoynov, S.S. (2006) "Coordination of DNA synthesis and replicative unwinding by the S-phase checkpoint pathways." Nucl. Acid Res. 34: 4138–4146.

Nelson, H.C.M., Finch, J.T., Luisi, B.F., Klug, A. (1987) "The structure of an oligo(dA)·oligo(dT) tract and its biological implications." Nature 330: 221–226.

Neves, H., Ramos, C., Gomes, da Silva, M., Parreira, A., Parreira, L. (1999) "The nuclear topography of ABL, BCR, PML, RARa genes: Evidence for gene proximity in specific phases of the cell cycle and stages of hematopoietic differentiation." Blood 93: 1197–1207.

Orlando, V., Strutt, H., Paro, R. (1997) "Analysis of chromatin structure by *in vivo* formaldehyde cross-linking." Methods 11: 205–214.

Ozsolak, F., Song, J.S., Liu, X.S., Fisher, D.E. (2007) "High-throughput mapping of the chromatin structure of human promoters." Nat. Biotechnol. 25: 244–248.

Patel, S.S., Picha, K.M. (2000) "Structure and function of hexameric helicases." Annu. Rev. Biochem. 69: 651–697.

Pennings, S., Meersseman G., Bradbury, E.M. (1991) "Mobility of positioned nucleosomes on 5 S rDNA. J. Mol. Bio. 220:101–110.

Peter, B.J., Ullsperger, C., Hiasa, H., Marians, K.J., Cozzarelli, N.R. (1998) "The structure of supercoiled intermediates in DNA replication." Cell 94: 819–827.

Pfeifer, G.P., Singer-Sam, J., Riggs, A.D. (1993) "Analysis of methylation and chromatin structure." Meth. Enzymol. 225: 567–583.

Postow, L., Hardy, C.D., Arsuaga, J., Cozzarelli, N.R. (2004) "Topological domain structure of the *Escherichia coli* chromosome." Genes & Dev. 18: 1766–1779.

Postow, L., Peter, B.J., Cozzarelli, N.R. (1999) "Knot what we thought before: the twisted story of replication." Bioessays 21: 805–808.

Postow, L., Ullsperger, C., Keller, R.W., Bustamante, C., Vologodskiii, A.V., Cozzarelli, N.R. (2001) "Positive torsional strain causes the formation of a four-way junction at replication forks." J. Biol. Chem. 276: 2790–2796.

Raghavan, S.C., Lieber, M.R. (2004) "Chromosomal translocations and non-B DNA structures in the human genome." Cell Cycle 3: 762–768.

Raghavan S.C., Chastain P., Lee J.S., Hegde B.G., Houston S. Langen, R., Hsieh, C.L., Haworth, I.S., Lieber, M.R. (2005) "Evidence for a triplex DNA conformation at the bcl-2 major breakpoint region of the t(14;18) translocation." J. Biol. Chem. 280: 22749–22760.

Raghavan, S.C., Lieber, M.R. (2006) "DNA structures at chromosomal translocation sites." Bioessays 28: 480–494.

Richards, R.I. (2001) "Fragile and unstable chromosomes in cancer: causes and consequences." Trends in Genet. 17: 339–345.

Richmond, T.J., Davey, C.A. (2003) "The structure of DNA in the nucleosome core." Nature 423: 145–150.

Richmond, T.J., Finch, J.T., Rushton, B., Rhodes, D., Klug, A. (1984) "Structure of the nucleosome core particle at 7 A° resolution." Nature 311: 532–537.

Rowley, J.D. (1973) "A new consistent chromosomal abnormality in chronic myelogenous leukemia identified by quinacrine fluorescence and Giemsa staining." Nature 243: 290–293.

Russev, G., Hancock, R. (1982) "Assembly of new histones into nucleosomes and their distribution in replicating chromatin." Proc. Natl. Acad. Sci. USA 79: 3143–3147.

Satchwell, S.C., Drew, H.R., Travers, A.A. (1986) "Sequence periodicities in chicken nucleosome core DNA." J. Mol. Biol. 191: 659–675.

Sato, M., Gotow, T., You, Z., Komamura-Kohno, Y., Uchiyama, Y., Yabuta, N., Nojima, H., Ishimi, Y. (2000) "Electron microscopic observation and single stranded DNA binding activity of the Mcm4,6,7 complex." J. Mol. Bol. 300: 421–431.

Sawyers, C.L. (1999) "Chronic myeloid leukemia." N. Engl. J. Med. 340: 1330–1340.

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

Schones, D.E., Cui, K., Cuddapah, S., Roh, T., Barski, A., Wang, Z., Wei, G., Zhao, K. (2008) "Dynamic regulation of nucleosome positioning in the human genome." Cell 132: 887–898.

Segal, E., Fondufe-Mittendorf, Y., Chen, L., Thåström, A., Field Y., Moore I.K., Wang, J.Z., Widom J. (2006) "A genomic code for nucleosomes positioning." Nature 442: 772–778.

Seigneur, M., Bidnenko, V., Ehrlich, S.D., Michel, B. (1998) "RuvAB acts at arrested replication forks." Cell 95: 419–430.

Sewack, G.F., Hansen, U. (1997) "Nucleosome positioning and transcription-associated chromatin alterations on the human estrogen-responsive pS2 promoter." J. Bio. Chem. 272: 31118–31129.

Shin, J.H., Santangelo, T.J., Xie, Y., Reeve, J.N., Kelman, Z. (2007) "Archaeal minichromosome maintenance (MCM) helicase can unwind DNA bound by archaeal histones and transcription factors." J. Bio. Chem. 282: 4908–4915.

Singleton, M.R., Sawaya, M.R., Ellenberger, T., Wigley, D.B. (2000) "Crystal structure of T7 gene 4 ring helicase indicates a mechanism for sequential hydrolysis of nucleotides." Cell 101: 589–600.

Sivolob, A.V., Khrapunov, S.N. (1995) "Translational positioning of nucleosomes on DNA: The role of sequence-dependent isotropic DNA bending stiffness." J. Mol. Bio. 247: 918–931.

Smith, S., Stillman, B. (1991) "Stepwise assembly of chromatin during DNA replication *in vitro*." EMBO J. 10: 971–980.

Sogo, J.M., Lopes, M., Foiani, M. (2002) "Fork reversal and ssDNA accumulation at stalled replication forks owing to checkpoint defects." Science 297: 599–602.

Sogo, J.M., Stahl, H., Koller, T., Knippers, R. (1986) "Structure of replicating simian virus 40 minichromosomes. The replication fork, core histone segregation and terminal structures." J. Mol. Bio. 189: 189–204.

Sogo, J.M., Stasiak, A., Martínez-Robles, M.L., Krimer, D.B., Hernández, P., Schvartzman, J.B. (1999) "Formation of knots in partially replicated DNA molecules." J. Mol. Bio. 286: 637–643.

Sowerby, S.J., Kennedy, M.A., Fitzgerald, P.H. and Morris, C.M. (1993) "DNA sequence analysis of the major breakpoint cluster region of the *BCR* gene rearranged in Philadelphia-positive human leukemias." Oncogene 8: 1679–1683.

Smith, S., Stillman, B. (1991) "Stepwise assembly of chromatin during DNA replication *in vitro*." EMBO J. 10: 971–980.

Stano, N.M., Jeong, Y., Donmez, I., Tummalapalli, P., Levin, M.K., Patel, S.S. (2005) "DNA synthesis provides the driving force to accelerate DNA unwinding by a helicase." Nature 435: 370–373.

Strick, T.R., Bensimon, D., Croquette, V. (1999) "Micro-mechanical measurement of the torsional modulus of DNA." Genetica 106: 57–62.

Teng, Y., Yu, S., Waters, R. (2001) "The mapping of nucleosomes and regulatory protein binding sites at the *Saccharomyces cerevisiae MFA2* gene: a high resolution approach." Nucl. Acids Res. 29: e64.

Thoma, F. (1996) "Mapping of nucleosome positions." Meth. Enzymol. 274: 197–213.

Thömmes, P., Hübscher, U. (1990) "Eukaryotic DNA replication." Eur. J. Biochem. 194: 699–712.

Tims, H., Widom, J. (2007) "Stopped-flow fluorescence resonance energy transfer for analysis of nucleosome dynamics." Methods 41: 296–303.

Travers, A.A. (1987) "DNA bending and nucleosome positioning." Trends Biochem. Sci. 12: 108–112.

Travers, A.A., Klug, A. (1987) "The bending of DNA in nucleosomes and its wider implications." Phil. Trans. R. Soc. Lond. B 317: 537–561.

Tsujimoto, Y., Jaffe, E., Cossman, J., Gorham, J., Nowell, P.C., Croce, C.M. (1985) "Clustering of breakpoints on chromosome 11 in human B-cell neoplasms with the t(11;14) chromosome translocation." Nature 315: 340–343.

Tyler, J.K. (2002) "Chromatin assembly: Cooperation between histone chaperones and ATP-dependent nucleosome remodeling machines." Eur. J. Biochem. 269: 2268–2274.

Van Gent, D.C., Hoeijmakers J.H.J., Kanaar, R. (2001) "Chromosomal stability and the DNA double-stranded break connection." Nat. Rev. Genet. 2: 196–206.

Wang, J.C. (1996) "DNA topoisomerases." Annu. Rev. Biochem. 65: 635–692.

Wang, J.C. (2002) "Cellular roles of DNA topoisomerases: A molecular perspective." Nat. Rev. Mol. Cell Biol. 3: 430–440.

Wang, Y.H., Gellibolian, R., Shimizu, M., Wells, R.D., Griffith, J. (1996) "Long CCG triplet repeat blocks exclude nucleosomes: a possible mechanism for the nature of fragile sites in chromosomes." J. Mol. Biol. 263: 511–516.

Wang, Y.H., Griffith, J. (1996) "Methylation of expanded CCG triplet repeat DNA from fragile X syndrome patients enhances nucleosome exclusion." J. Biol. Chem. 271: 22937–22940.

Wellinger, R.E., Sogo, J.M. (1998) "In vivo mapping of nucleosomes using psoralen–DNA crosslinking and primer extension." Nucl. Acids Res. 26: 1544–1545.

Whetton, A.D., Graham, G.J. (1999) "Homing and mobilization in the stem cell niche." Trends Cell Biol. 9: 233–238.

Widom J. (2001) "Role of DNA sequence in nucleosome stability and dynamics." Quart. Rev. Biophys. 34: 269–324.

Yalcintepe L., Albeniz I., Adin-Cinar S., Tiryaki D., Bermek E., Graeff R.M., Lee H.C. (2005) "Nuclear CD38 in retinoic acid-induced HL-60 cells." Exp. Cell Res. 303: 14–21.

Zhang, J.G., Goldman, J.M., Cross, N.C.P. (1995) "Characterization of genomic BCR-ABL breakpoints in chronic myeloid leukemia by PCR." Br. J. Haematol. 90: 138–146.

Zlotorynski, E., Rahat, A., Skaug, J., Ben-Porat, N., Ozeri, E., Hershberg, R., Levi, A., Scherer, S.W., Margalit, H., Kerem, B. (2003) "Molecular basis for expression of common and rare fragile sites." Mol. Cell. Biol. 23: 7143–7151.