Use of the Z-DNA binding domain of ADAR1 as a probe for nucleic acids in left-handed conformation

Li, Heng

2008

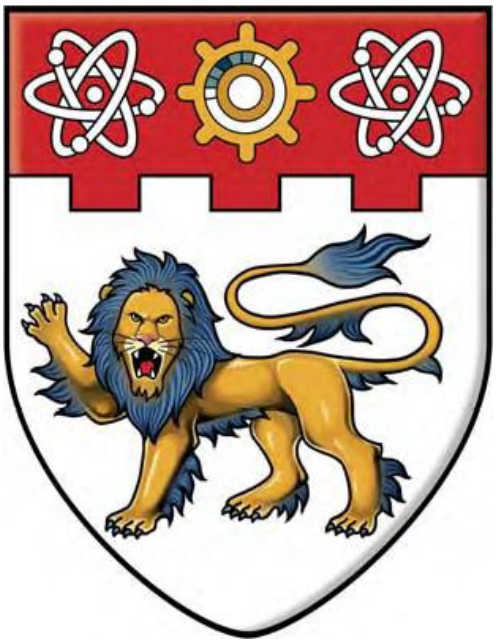
Li, H. (2008). Use of the Z-DNA binding domain of ADAR1 as a probe for nucleic acids in left-handed conformation. Doctoral thesis, Nanyang Technological University, Singapore.

https://hdl.handle.net/10356/15657

https://doi.org/10.32657/10356/15657

Use of the Z-DNA Binding Domain of ADAR1 as a Probe for Nucleic Acids in Left-handed Conformation

Li Heng



SCHOOL OF BIOLOGICAL SCIENCES

NANYANG TECHNOLOGICAL UNIVERSITY

2008

Use of the Z-DNA Binding Domain of ADAR1 as a Probe for Nucleic Acids in Left-handed Conformation

Li Heng

School of Biological Sciences

Nanyang Technological University

A thesis submitted to the Nanyang Technological University
in fulfilment of the requirement for the degree of
Doctor of Philosophy

2008

ABSTRACT

Both DNA and RNA can be in Z-conformation. Compared to classical B-DNA conformation, Z-DNA is in a higher energy state and studied extensively, so far mainly *in vitro*. It has been recognized that Z-DNA plays roles in many biological activities, including regulation of transcription, chromatin remodeling, genetic instability, etc. Bioinformatics analyses and experimental studies on particular genomic loci suggest that the Z-DNA forming sequences are enriched in gene regulatory regions, where the formation of Z-DNA is drived by transcription-induced negative supercoiling. However, the global distribution of Z-DNA in the human genome *in vivo* is still experimentally elusive, due to instable nature of Z-DNA and lack of Z-DNA specific probe. Here, we show for the first time the distribution of Z-DNA hotspots in the genome of cultured human cancer cells A549 by exploiting the Zα domain from human ADAR1 ($Z\alpha_{ADAR1}$, simplified as Zα for convenience in the thesis) as a Z-DNA specific probe with a novel experimental strategy. A protocol for *in vitro* chromatin affinity precipitation (ChAP) was developed and a Z-DNA library was constructed and sequenced. In total, 186 Z-DNA hotspots were identified. Enrichment of hotspots in gene regulation regions was not found. Unexpectedly, 46 hotspots localize to the centromeric regions. Further examination on hotspots therein suggests that Z-DNA hotspots are strongly correlated with high occurrences of SNPs (single nucleotide polymorphism). Bioinformatics analysis using a Z-DNA prediction program shows that most of these Z-DNA forming hotspots (143 out of 186) could be transited to

Z-conformation only at $\sigma < -0.08$, suggesting particular regions in the centromeres contain high negative superhelical torsional strain. Genetic instability and gene conversion promoted by Z-DNA provide a new clue for the rapid evolution of the centromere.

Z-RNA shares many similar structural features with Z-DNA. But so far its existence *in vivo* and potential biological roles are still elusive. In the second part of this study, we found that Zα could bind to RNA *in vivo* which was identified as rRNA. Further studies suggested that Zα could bind to mammalian ribosomes in a conformation-dependent manner. The effect of the binding of Zα to ribosomes was examined and Zα could inhibit *in vitro* translation efficiently.

ACKNOWLEDGEMENT

I would like to give my special and deepest appreciation to my supervisor Prof. Peter Dröge for his valuable instruction and encouragement on my PhD study and his sincere care and help on my life. His serious attitude and persistence on research will lead me to the future.

I deeply thank Prof. Jinming Li for his instruction and cooperation on the bioinformatics analyses. Many thanks go to Prof. Alexander Rich and Prof. Thomas Schwartz in MIT who provided the original expression vectors. Thanks also go to Prof. Yu Chin-Wen, Kenneth for his instruction on confocal microscope. I would like to thank Prof. Yoon Ho Sup for his kind help and encouragement.

I would present my special thanks to my cooperators: Miss Feng Shu and Miss Xiao Jie. Miss Feng Shu did a lot of bench work on this project, especially on RNA part. Miss Xiao Jie analyzed all sequencing data using bioinformatics tools. Cooperation with them tells me how to play as a team member.

Thanks to all the members in our group: Feng Shu, Bao Qiuye, Tan Shen Mynn, Zhou Ruijie, Li Ou, Sabrina Peter, Ren Yudan, Sathiyanathan Padmapriya, Dr. Klaus Neef and Dr. Heike Summer, especially those gave comments on my thesis.

I am grateful for the financial support of predoctoral scholarship from Nanyang Technology University during my study.

Finally, I will give sincere thanks to my parents, my wife and our cute daughter, my sister and brother. They gave me selfless care and encouragement to finish my Ph.D. study.

ABBREVIATIONS

aa	amino acids
ADAR	adenosine deaminase on double-stranded RNA
ADEAM	adenosine deaminase motif
ALR	alpha satellite repeat
AP	ammonium persulfate
bp	basepair
BSA	bovine serum albumin
BSR	human beta satellite
CD	circular dichroism
ChAP	chromatin affinity precipitation
ChIP	chromatin immunoprecipitation
CIP/CIAP	alkaline phosphatase, calf intestinal
CV	column volume
DAPI	4'-6-Diamidino-2-phenylindole
DOC	deoxycholate
DNA	deoxynucleotide
DSB	double-strands break
DSRM	double-strand RNA binding domain
dsRNA	double-strand RNA
DMEM	Dulbecco's Modified Eagle's Medium
DTT	dithiothreitol
EDTA	ethylene diamine tetra-acetic acid
EGTA	ethyleneglycol bis(2-aminoethylether)-N,N,N'N'-tetraacetic acid
eGFP	enhanced green fluorescent protein
EtBr	ethidium bromide
FBS	fetal bovine serum
FPLC	fast protein liquid chromatography
HAPA	hydroxy-azophenyl-benzoic acid
HSATII	human satellite II
IPTG	isopropyl-beta-D-thiogalactopyranoside
kDa	kilodalton
Kb	kilobase
LB	Luria-Bertani
LMB	leptomycin B
LTR	long terminal repeat
MER	Medium reiteration frequency repeat
NEB	New England Biolabs
NES	nuclear export signal

NLoS	nucleolus localization signal
NLS	nuclear localization signal
ORF	open reading frame
PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PMSF	phenylmethanesulfonyl fluoride
PVDF	polyvinylidnene fluoride
RIPA	radioimmunoprecipitation
RNA	ribonucleotide
RRL	rabbit reticulocyte lysate
S/MAR	scaffold/matrix attachment site
SDS	sodium dodecyl sulfate
SNP	single nucleotide polymorphism
SPR	surface plasmon resonance
SSTI	a moderately tandem repeated DNA sequence
TBS	Tris-buffered saline
TEMED	N, N, N', N'-tetramethylethylenediamine
TLS	translation start site
Tris	Tris (hydroxmethyl)-aminomethane
v/v	volume per volume
wt	wild type
w/v	weight per volume
ZDR	Z-DNA forming region

CONTENT

LIST OF FIGURES

I. INTRODUCTION

I.1. Left-handed Z-conformation and its relevant biological functions

I.1.1. Characteristics of Z-DNA structure

The model of classical right-handed double helical DNA structure was first proposed by Watson and Crick in 1953 (Watson and Crick, 1953). In the late 1970s, when chemically synthesized DNA was available, theoretical analysis and several experiments had shown that different DNA conformation might exist (Narasimhan and Bryan, 1976; Yathindra and Sundaralingam, 1976; Neifakh Iu and Tumanian, 1979; Wang et al., 1979). In 1979, when the first single-crystal X-ray structure of a DNA fragment, a self-complementary DNA hexamer $d(CG)_3$, was resolved, it was not in the right-handed conformation as expected; instead, it was a left-handed double helix with two anti-parallel strands held together by Watson-Crick base pairing. This unusual DNA conformation was designated as Z-DNA, named after its zigzag arrangement of the sugar-phosphate backbone of the DNA molecule (Crawford et al., 1980).

Naturally DNA exists mainly in 3 conformations: A-，B-, and Z-conformation (molecular models in **Figure I-1** and the parameters listed in **Table I-1**). In a Z-DNA structure, purines are inverted and in the *syn* conformation, while the pyrimidines remain in the *anti* conformation such that the pyrimidine nucleotides must rotate to maintain the Watson-Crick base-pairing. Such an alternating *syn-anti* conformation drives the backbone into a zigzag shape.

Because purines adopt the *syn* conformation more readily than pyrimidines, Z-DNA conformation occurs most favorably in alternating purine/pyrimidine repeat sequences, especially alternating $d(GC)_n$ repeats (Malfoy et al., 1986; Johnston, 1992) and $d(GT)_n$ repeats. It has been reported that GT repeats are the most frequent simple repeating elements in the human genome, accounting for more than 0.25% of the total genome (Lander et al., 2001). Additionally, other sequences with high GC content, such as $d(GGGC)_n$, have also been detected in Z-DNA conformation (Feigon et al., 1985; Eichman et al., 1999).

Now it is clear that Z-DNA conformation has the following physical features (Brown and Rich, 2001): 1) left-handed sense; 2) requiring negative DNA superhelical tension and alternating purine-pyrimidine sequences; 3) narrower minor groove, more elongated helix than A- and B- conformation; 4) favored by high salt concentrations; 5) base pairs nearly perpendicular to helix axis; 6) purine-pyrimidine repeat as structural unit, instead of single base-pair; 7) zigzag backbone. In order to define the slight difference between various crystal forms in the Z-DNA family, Z-DNA has been classified as Z_I, Z_{II}, and Z' (Wang et al., 1981; Fujii et al., 1983).

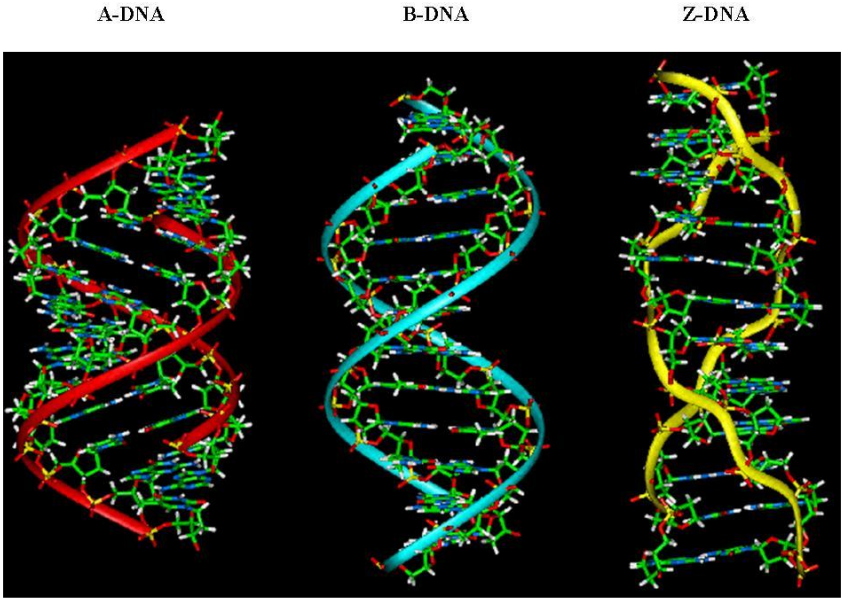


Figure I-1. The structures of three distinct DNA conformations.

(Cited from http://upload.wikimedia.org/wikipedia/en/b/b9/A-B-Z-DNA_Side_View.png).

Parameter	A form	B form	Z form
Helix sense	Right	Right	Left
Diameter (Å)	23	20	18
Rotation per residue (°)	33	34.3	-30
Base pairs per helical turn	11	10.5	12
Helix pitch (°)	28	34	45
Helix rise per base pair (Å)	2.55	3.4	3.7
Base tilt normal to the helix axis (°)	20	-6	7
Sugar pucker conformation			
dA, dT, dC	C3' endo	C2'endo	C2'endo
dG	C3' endo	C2'endo	C3'endo
Glycosidic bond configuration			
dA, dT, dC	anti	anti	anti
dG	anti	anti	syn

Table I-1. The helix parameters of A-, B-, and Z- DNA.

All the values are taken from X-ray diffraction data. This table is modified from (Sinden, 1994).

I.1.2. Z-RNA conformation

It was postulated that Z-RNA could be formed when Z-DNA forming sequence was transcribed into RNA. Experimental evidence showed that chemical bromination could stabilize synthesized poly[r(C-G)] in Z-conformation under physiological conditions, which could be recognized by anti-Z-DNA antibody (Hardin et al., 1987). By using an anti Z-RNA monoclonal antibody which did not recognize Z-DNA, it has been found that some RNAs in both the cytoplasm and nuclei in fixed cells were in Z-RNA conformation (Zarling et al., 1987). Z-RNA can also be recognized by the Z-DNA binding domain Zα (Brown et al., 2000) and recently Z-RNA structure has been determined in co-crystallized Zα-RNA complex (Placido et al., 2007). These results suggest that Z-RNA shares many similarities with Z-DNA in respect to 3-D structure and physical features, except that the Z-RNA helix is associated with a unique solvent pattern in crystals.

I.1.3. Formation of Z-DNA *in vitro*

In 1972, it was observed (Pohl and Jovin, 1972) that with addition of large amounts of salt (up to 4 M salt), the circular dichroism (CD) spectrum between 240 nm and 310 nm of poly $(GC)_3$ was completely inverted, a characteristic of Z-DNA formation (as shown in **Figure I-2**). It is recognized that the dominant contribution to the free energy difference between B-DNA and Z-DNA is due to electrostatic interactions (Gueron and Demaret, 1992; Gueron et al., 2000). Because Z-DNA (12 bp per turn) is more compact than B-DNA (10.5 bp per

turn) along the backbone axis, the successive negatively-charged phosphate groups along the backbone reside on average closer to each other. So at lower ionic strength, the resulting increase in electrostatic repulsion significantly destabilizes the Z form relative to the B form, although the stacking interactions are expected to be more favorable in the Z-DNA conformation (Sponer et al., 1997). With increasing salt concentration, counterions provide more efficient electrostatic screening and the relative stabilities of the two forms are inverted (Pohl, 1983). High salt (6 M $NaClO_4$) conditions also can promote Z-RNA formation (Popenda et al., 2004).

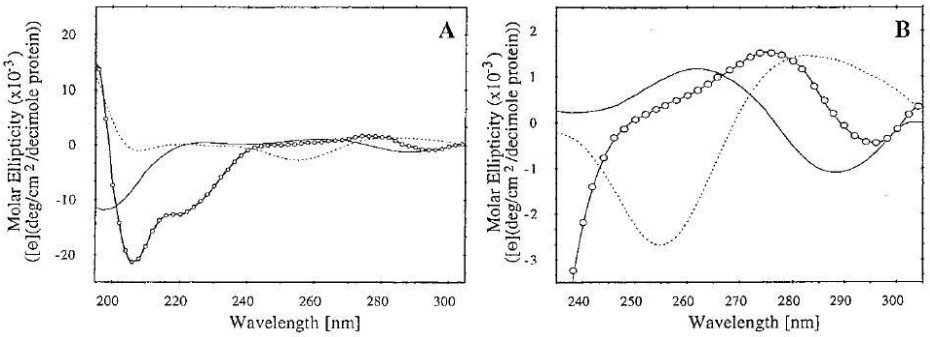


Figure I-2. CD spectra of B- and Z-DNA.

A. the CD spectra of poly(dGdC) in 20 mM NaCl buffer (dashed line) and in 4 M NaCl buffer (solid line) and poly(dGdC) with Zα in 20 mM NaCl (-o-) were showed. The characteristic B-to-Z inversion of CD spectrum is showed. Cited from (Berger et al., 1998). B. Zooming in the CD spectra between 240 nm and 310 nm. The negative peak at 258 nm in low salt buffer was shifted to 285 nm in high salt buffer and to 295 nm in polymer-Zα mixture. The difference between spectra of polymer in high salt and in the mixture indicated that Z-DNA conformation induced by high salt is slightly different from the one in DNA-Zα complex, which is consistent with the finding with Z-DNA antibody (Lafer et al., 1986).

Additional studies using circular dichroism (CD) demonstrated that C-5 methylation of deoxycytosine can flip poly(dG-m^5dC).poly(dG-m^5dC) from B-conformation to Z-conformation even at physiological salt conditions (Behe and Felsenfeld, 1981; Fujii et al., 1982). N-7 or C-8 methylation in poly(dG-dC).poly(dG-dC) can also facilitate Z-DNA formation at physiological salt conditions (Moller et al., 1981; Kawai et al., 1995). Other chemical modifications such as bromination can also induce and stabilize poly(dG-dC) in the Z-DNA form under low-salt conditions (Moller et al., 1984).

Besides ionic strength, nucleotide modifications and nucleotide compositions of DNA sequence, the stability of Z-DNA is also influenced by the temperature. As Z-DNA has lower conformational entropy compared to B-DNA, the relative stability of the Z-DNA conformation generally decreases upon raising the temperature (Irikura et al., 1985). This property, together with the intrinsic differences in stacking patterns between the two DNA forms (Sponer et al., 1997), has recently been exploited in the design of a nanothermometer (Tashiro and Sugiyama, 2003). Based on these intrinsic properties of Z-DNA, it has been proposed to design a nanomechanical motor device (Mao et al., 1999) and a temperature-dependent reversible biomolecular switch (Tashiro and Sugiyama, 2005).

I.1.4. Formation of Z-DNA *in vivo*

The discovery that suggests Z-DNA's biological relevance is the induction of Z-DNA formation by negative superhelical stress (Klysik et al., 1981; Singleton

et al., 1982). The free energy necessary to induce the B to Z transition can be quantified by analyzing the plasmid topology on 2-D agarose gel (Peck and Wang, 1983; Ellison et al., 1985). The required energy is proportional to the square of the number of negative supercoils lost from a covalently closed circular plasmid when a sequence flips into the Z-DNA conformation. Approximately two supercoils are lost in order to stabilize one turn of Z-DNA.

From the 1980s' some notable studies supported that Z-DNA played roles in transcription (Nordheim and Rich, 1983; Lancillotti et al., 1987; Wittig et al., 1991; Wittig et al., 1992). Z-DNA can be induced transiently behind the moving RNA polymerases and be stabilized by the negative supercoiling generated by DNA transcription as demonstrated in **Figure I-3** (Liu and Wang, 1987; Droge and Nordheim, 1991; Droge and Pohl, 1991). After the polymerase stops moving, topoisomerase is able to catch up and release torsional strain and the Z-DNA conformation reverts to the lower energy B-DNA conformation (Wang and Droge, 1996). Thus the formation of Z-DNA *in vivo* is very dynamic.

In prokaryotic cells, one approach to detect Z-DNA is using chemicals to detect the unusual DNA structures (Kochel and Sinden, 1988; Palecek et al., 1988; Jiang et al., 1991). Another approach exploited a construct containing one *EcoR*I site embedded in a Z-DNA forming sequence (Jaworski et al., 1987; Rahmouni and Wells, 1989). Since DNA methylation can only occur on B-DNA, but not on Z-DNA, the susceptibility to methylation of the *EcoR*I site thus provides an indicator of *in vivo* Z-DNA formation. In *E. coli*, the formation of

Z-DNA is increased by transcription and enhanced by mutations inactivating topoisomerase (Jaworski et al., 1991). Methylation assay suggested that methylation of cytosine could promote the formation of Z-DNA in *E.coli* (Zacharias et al., 1990).

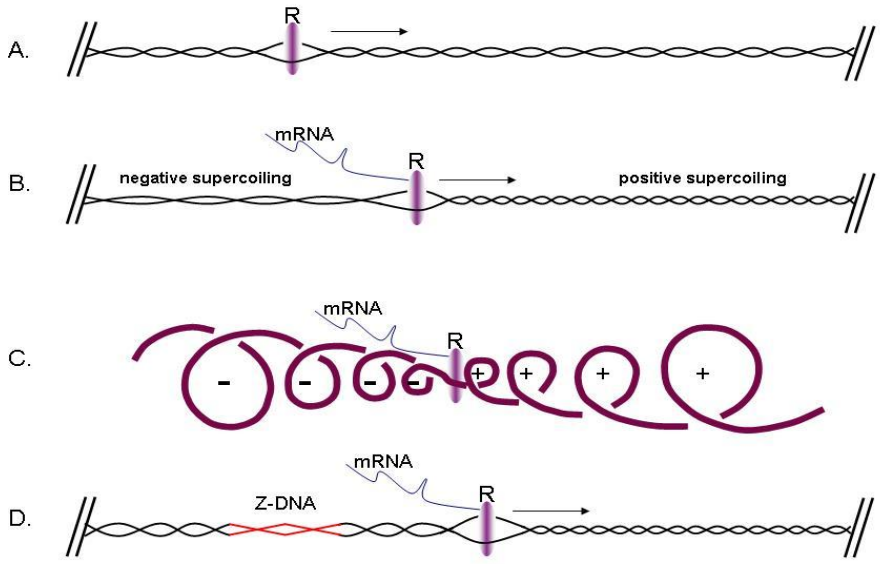


Figure I-3. The proposed model of Z-DNA formation induced by transcription.
R stands for RNA polymerase. When RNA polymerase moves along the DNA template, it would not rotate around the template; otherwise the nascent mRNA will tangle with DNA template. In this model, RNA polymerase ploughs through the DNA double-stranded helix, and produces negative supercoiling behind and positive supercoiling ahead of it, as shown in B. C is a diagram of B, showing the supercoiling state of DNA. D. Formation of Z-DNA in negative region would buffer the negative supercoiling torsion. In prokaryotes, topoisomerase I and gyrase would release the negative and positive supercoiling produced by transcription, respectively. This figure is modified from (Liu and Wang, 1987).

Detection of *in vivo* Z-DNA in mammalian cells has been difficult due to its instability; however, several early observations suggested its existence. Since

Z-DNA is strongly immunogenic, the detection of its highly specific antibodies found in the sera from patients with auto-immune diseases, especially lupus erythematosus, suggests that Z-DNA exists *in vivo* (Lafer et al., 1983). Antibodies raised against Z-DNA in rabbit and sheep were used as powerful tools to detect Z-DNA in both fixed (Nordheim et al., 1981) and unfixed polytene chromosomes of *Drosophila* (Lancillotti et al., 1987). The interband regions, especially in the puffs which are associated with high level of transcriptional activity, were intensely stained by Z-DNA antibodies, but the bands were not. Another experiment was performed using Z-DNA antibody to stain ciliated protozoa, which have one macronucleus where transcription occurred and one micronucleus where was the site for genetic reproduction (Lipps et al., 1983). The macronucleus was stained exclusively. These early observations suggest the existence of Z-DNA *in vivo* and its link to transcriptional activity.

Due to the absence of a phenotype associated with Z-DNA, it is impossible to use genetic manipulation to investigate the function of Z-DNA in mammalian cells. Further analyses of presence of Z-DNA in mammalian cells were performed using metabolically active permeabilized mammalian cell nuclei (Jackson et al., 1988) which were transcription competent. In these experiments, the amount of Z-DNA present in the nuclei was measured by diffusing biotin-labeled anti-Z-DNA monoclonal antibodies in the agarose beads in which the nuclei were embedded (Wittig et al., 1989). Then radioactive streptavidin was used to bind biotin and quantified the amount of Z-DNA in the nuclei. The

results suggest that Z-DNA is present *de novo* in the prepared nuclei, rather than being induced by antibodies. In addition, the amount of Z-DNA present was increased prominently with active transcription, but was not affected by DNA replication (Wittig et al., 1991).

Using metabolically active mammalian nuclei, Z-DNA segments have been detected in several individual genes by crosslinking Z-DNA antibody to its binding sites (Wittig et al., 1992). The crosslinked DNA fragments were released by *in situ* digestion and identified by PCR or hybridization. It has been identified that there are three Z-DNA forming segments in the 5' regions of *c-myc* gene, with two of them near the promoter (Wittig et al., 1992; Wolfl et al., 1995). When the expression of *c-myc* was down-regulated, the amount of Z-DNA segments was quickly reduced. In another study using a primary liver cell line, induction of Z-DNA was discovered in a corticotrophin hormone-releasing gene (Wolfl et al., 1996) and the amount of Z-DNA was positively related to the expression level of the gene.

I.1.5. Proposed models of B to Z transition

The formation of Z-DNA *in vitro* has been studied in details. However, the process of the transition from B- to Z-conformation is still unknown. Apart from the inverted helicity, Z-DNA, containing 12 bp per turn, is more compact along the double-helical axis, compared to B-DNA with 10.5 bp per turn. The interconversion can not be a simple matter of twisting the helix ends. One hypothesis is known as the Harvey model, in which the base pairs flip over one

at a time while keeping the Watson-Crick hydrogen bonds intact (Harvey, 1983). Another hypothesis is based on a molecular model of poly d(G-C).poly d(G-C) built with structural characteristics in the A and A2 form, which are metastable in B-to-Z transition without disruption of the Watson-Crick hydrogen bonds and without steric interference (Saenger and Heinemann, 1989). Recently, based on the first crystallographic structure of a B-Z DNA junction (Ha et al., 2005), a new "zipper" model was proposed (Kastenholz et al., 2006), in which after an initial high-barrier nucleation step resulting in a short Z-DNA segment enclosed between two B-Z junction, the junctions migrate in opposing directions as the intervening basepairs flip and the backbone winds in the left-handed direction to form Z-DNA.

I.1.6. The B-Z junction

One puzzling question in past years was what the structure of B-Z junction was. It had been realized that the B-Z junction could not be recognized efficiently by restriction enzymes (Singleton et al., 1983) and could be specifically cleaved by S1 nuclease, which degrades single-stranded nucleic acids (Singleton et al., 1984). And it had been proved that in B-Z junction there were unpaired bases or non-Watson-Crick pairing (Palecek et al., 1987). However, it was not until recently that the B-Z junction was resolved (Ha et al., 2005). In this study, oligonucleotides carrying Z-DNA forming sequence at one end was used. Upon binding with Zα, the permissive sequence was converted into Z-conformation and a B-to-Z junction formed in the middle of the polymer. As shown in Figure I-4, there is successive stacking of bases between B-DNA and Z-DNA

segments, with breaking the pairing of one base at the junction and extrusion of the base on each side. This finding is in agreement with previous observations and more importantly, the structure of B-Z junction minimizes the energetic cost through maximizing base pairing and preserving the helical integrity. The economical energy cost makes it possible that Z-DNA functions widely in the genome. Another point of this finding is that these extruded bases in the B-Z junctions in the genome may be sites for DNA modification since they are more accessible.

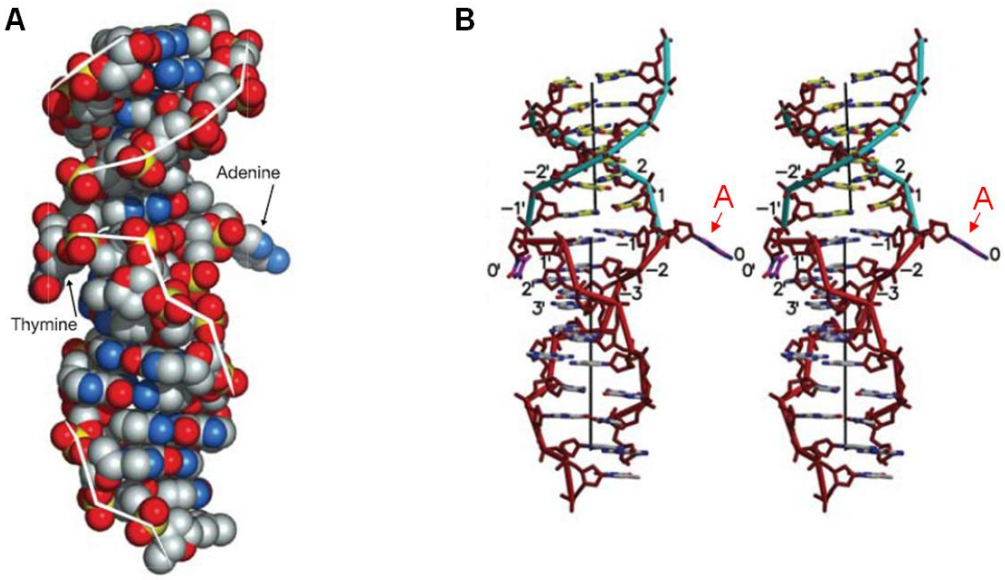


Figure I-4. The structure of B-Z DNA junction.

A. A van der Waals view of the structure. Two bases have been extruded from base stacking at the junction. The white line shows the backbone. O is shown in red, N in blue, P in yellow and C in grey. B. Skeletal stereo views of the structure. Front view shows unpaired A (right) and T (left), which are indicated by arrows. Red and blue lines connect phosphate groups in Z-DNA and B-DNA, respectively. Cited from (Ha et al., 2005).

I.2. Biological functions of Z-DNA

I.2.1. Z-DNA and transcription

It had been noticed in 1983 that Z-DNA is involved in the regulation of specific gene transcription *in vivo* (Rich et al., 1983). And with more evidence accumulated, the association between Z-DNA and gene transcription became more solid (Santoro and Costanzo, 1983; Hamada et al., 1984; Santoro et al., 1984; Lancillotti et al., 1987). The model proposed by Liu and Wang (Liu and Wang, 1987) in Figure I-3 suggested that the transcription on a DNA template produced negative supercoiling behind the polymerase, thus facilitating Z-DNA formation at permissive regions. Formation of Z-DNA would reduce the superhelical tension and keep the rest of the DNA in stable B-conformation. However, the relationship between transcriptional activity and Z-DNA formation is still in dispute as the influence of Z-DNA on transcription depends on the individual gene that was examined (Rich and Zhang, 2003). Recently it was proposed that Z-DNA could be used as a new *in situ* marker for transcription (Cerna et al., 2004).

Although Z-DNA is induced and stabilized by transcription, formation of Z-DNA, in turn, also can up- or down- regulate the transcriptional activity in *cis*. Z-DNA formation stabilizes the open chromatin structure at the *CSF1* promoter and promotes its transcription (Liu et al., 2001); however, Z-DNA segment in the promoter region of rat *Ncl* gene has shown inhibitory effect on the promoter activity (Rothenburg et al., 2001). Z-DNA can determine the allelic expression

pattern as well. Study on rat nucleolin gene which has varied Z-DNA forming sequences at the 5' end of its promoter region suggested that Z-DNA might exert quantitative effects on the expression of alleles (Rothenburg et al., 2001), which is another mechanism of mono-allele expression, other than determined stochastically or by parental origin. Thus, Z-DNA may regulate gene expression via different mechanisms, perhaps dependent on its location and stability.

I.2.2. Z-DNA and chromatin remodeling

Most eukaryotic DNA sequences within the nucleus are packaged in nucleosomes; however, it has been found that Z-DNA could not be incorporated in the nucleosome core particles (Garner and Felsenfeld, 1987). So Z-DNA formation *in vivo* could affect significantly nucleosome packaging and positioning. On the other hand, prior to transcription, unwrapping DNA from nucleosomes produces negative supercoiling, which may induce Z-DNA formation at permissive sequences. One example of the influence of Z-DNA formation on nucleosome positioning and chromatin remodeling was identified (Liu et al., 2001; Liu et al., 2006). In SW-13 cells when colony-stimulating factor 1 (*CSF-1*) was silent, the TG repeat in its promoter region was in B-conformation and packaged in nucleosomes. Upon activation of *CSF1* expression by the chromatin remodeling complex BAF, the nucleosome particle became loose and resulted in partial chromatin remodeling, and the TG stretch was still in Z-like intermediate conformation. Active transcription of *CSF1* did not require the formation of Z-conformation in the TG repeats, but enhanced the B- to Z- transition of TG repeats and stabilized it. As a consequence,

Z-conformation of TG repeats promoted the expression of *CSF1* by disrupting the nucleosome structure extensively and leading to complete chromatin remodeling. This experiment suggests that Z-DNA formation is an important chromatin modulating mechanism, as presented in Figure I-5. In this model of chromatin remodeling by BAF complex, NFI (nuclear factor I) or CTF (CAAT-box transcription factor) was the initiator. The co-occurrence of NFI binding sites and Z-DNA permissive sequences on chromosome 22 was analyzed (Champ et al., 2004). Bioinformatics analysis suggested that there was a positive correlation between NFI binding sites and Z-DNA permissive sequences in known and predicted genes on chromosome 22, consistent with the model proposed by Liu. However, it should be noticed that in the *in silico* analysis, the distance between NFI binding sites and Z-DNA forming sequences are longer than the experimental model, in which Z-DNA forming sequence are immediately upstream of NFI binding site. Another experimental evidence of the involvement of Z-DNA in chromatin remodeling was found in *S. cerevisiae* (Wong et al., 2007); a $d(GC)_9$ repeat in Z-conformation was found to stimulate gene activity by blocking nucleosome formation and producing transcriptionally favorable locations for general transcription machinery.

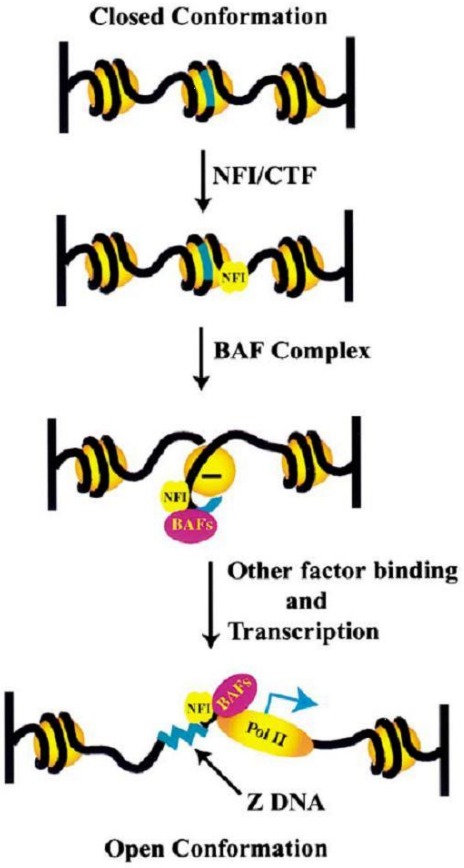


Figure I-5. The postulated mechanism of chromatin remodeling promoted by Z-DNA.
The promoter region of CSF1 gene was packaged in closed conformation when in silent state. Upon recognized by NFI (nuclear factor I) or CTF (CAAT-box transcription factor), BAF complex can bind its binding site more effectively, resulting in the partial remodeling of nucleosome particle. In this loose nucleosome, TG stretch is in an intermediate state between B- and Z- conformation due to negative superhelical tension released from nucleosome structure dismiss. Partial remodeling allows transcription machinery to start transcription, which produced stronger negative superhelical torsion and stabilized the Z-conformation in TG repeats. The formation of Z-conformation promotes complete chromatin remodeling. Cited from (Liu et al., 2001)

I.2.3. Z-DNA and genetic instability

I.2.3.1. Z-DNA induces deletion in bacteria and mammalian cells

It is well known that plasmids containing Z-conformation forming sequences were unstable in *E. coli*. When plasmids carrying (CG) repeats ranging from

6-29 dinucleotides were replicated in *E. coli*, deletions in the $(CG)_n$ region occurred more frequently than in plasmids without CG repeats (Klysik et al., 1981; Klysik et al., 1982). Further experiment using frameshift mutation assay showed that both the number of repeat unit and the propensity to form Z-conformation were crucial factors for the deletion events (Freund et al., 1989). In this assay, various lengths of alternating purine-pyrimidine bases (GC, GT or AT) repeats were cloned in upstream of the *lac*Z gene in plasmid pUC8 and the β-galactosidase α compensation assay was used to monitor the frameshift mutation frequency within the inserted regions. When the plasmids containing $(GC)_{13}$ repeat were transformed into *E.coli*, spontaneous mutation frequency was about 3%, significantly greater than that of plasmids containing $(GC)_6$, in which the mutation frequency was only about 0.3×10^{-4}. And in the 3% mutations caused by $(GC)_{13}$ insert, 90% were long deletion events and 10% were small addition/deletion events. Interestingly, the deletions involved an even number of base pairs such as 4, 10 and 16 basepairs. It was proposed and proved that the small addition/deletion events were due to slippage events during replication. Experiments with plasmids containing $(GT)_n$ or $(AT)_n$ suggested that cruciform structure was not involved in the generation of long deletions. Comparison between results from $(GT)_n$ and from $(GC)_{12}$ plasmids showed that spontaneous mutation frequencies was related to the potential of the inserts to adopt a Z-DNA conformation. It is possible that the Z-DNA structure can halt or suspend the DNA polymerase, giving rise to larger or multiple rounds of deletions until the GC repeat is too short to form a stable Z-conformation.

Deletions induced by Z-DNA structure occurred not only in prokaryotic cells, but also in mammalian cells (Wang et al., 2006). In this experiment, mutation reporter shuttle plasmids carrying various inserts which could adopt Z-conformation were used. Firstly, these shuttle plasmids were tested in bacteria, and small deletions were induced by Z-DNA conformation, consistent with previous report (Freund et al., 1989). The interesting finding was that about 95% mutations induced by Z-DNA forming sequences detected in COS-7 cells were large deletions or rearrangements. Most deletions spanned the regions from the Z-DNA forming sequences to another S1 nuclease sensitive site in SV40 replication origin. Furthermore, more than 85% analyzed deletion junctions had 1-6 bp of homology, indicating that these mutants were generated by double-strands break (DSB) repair system. This hypothesis was confirmed by the linker-mediated PCR (LM-PCR). The mechanism of large scale deletion induced by Z-DNA structure in mammalian cells is different from that in bacteria, since no DSBs were induced by Z-DNA in bacteria. Moreover, the influence of transcription on genetic instability was checked. By using the inducible mouse mammary tumor virus-LTR promoter on the plasmids, when transcription was induced by addition of dexamethasone, the mutation frequencies for the Z-DNA containing plasmids were increased to a greater extent than those of the control plasmids.

I.2.3.2. Z-DNA and genetic recombination

It has been demonstrated that Z-DNA forming sequences correlate with

recombination events in both prokaryotic and eukaryotic cells. It has been reported that $d(GT)_{30}$ cloned in a supercoiled plasmid could stimulate recombination in bacteria (Murphy and Stringer, 1986). These recombination events usually occurred between $d(GT)_n$ tracks and also between sequences bordering the dinucleotide repeats. The recombination in bacteria was enhanced when two $d(GT)_n$ tracks were clustered within 250 bp of each other, but not when the repeats were separated by 3 kb.

In eukaryotes, the concurrence of Z-DNA forming sequences with recombination hotspots was observed (Blaho and Wells, 1989; Wahls et al., 1990; Majewski and Ott, 2000). When two non-replicating plasmids co-introduced into human cells in culture, the sequences $d(GT)_{30}$ enhanced homologous recombination in human cells up to 20-fold (Wahls et al., 1990). Crosses assay indicated that the substrate containing the Z-DNA motif preferentially acted as the recipient of genetic information during gene conversion events. The presence of a 34 bp GT track was shown to enhance reciprocal meiotic recombination in yeast (Treco and Arnheim, 1986). Moreover, in mice a hotspot of 1,000 bp in the major histocompatibility complex (MHC) containing several copies of long GT repeats account for 2% of recombination on the entire chromosome (Kobori et al., 1986; Wahls, 1998). A bioinformatics analysis suggested that GT repeats were associated with recombination events on human chromosome 22 (Majewski and Ott, 2000).

I.2.3.3. Susceptibility of Z-DNA to DNA damage

In B-conformation, the base-paring between strands is protected within the helix. Different from B-DNA, the guanosine nucleotide in a Z-DNA conformation is in *syn* conformation, with the purine base located over the sugar without protection, thus possibly being more accessible to DNA damaging factors (Zimmerman, 1982). Also the fact that Z-DNA cannot be incorporated in nucleosome particle increases its accessibility to genotoxic agents. UV irradiation could induce significant chromatin fragmentation at GT repeats and a H-DNA forming sequence (Luokkamaki et al., 1993). As identified in the crystal structure, the B-Z junction has two extrusions of nucleotides (Ha et al., 2005), which are sensitive to certain agents (Rodolfo et al., 1994). The unusual structure of Z-DNA also increased oxidative damage on guanine induced by Ni^{2+}, which did not occur on B-DNA (Tang et al., 1999). In addition, DNA damage on Z-DNA conformation was more resistant to processing by DNA repair enzymes (Lagravere et al., 1984). Together, it is possible that endogenous DNA damage accumulates at Z-DNA regions, contributing to their induction of genetic instability *in vivo*.

I.2.4. Other possible roles of Z-DNA

Beside the biological functions of Z-DNA discussed above, it may also play other roles. One role is cell memory. In an *in vitro* model system, Z-DNA formation driven by RNA polymerase, in conjunction with Z-DNA binding proteins that exhibit a high affinity for Z-DNA and form very stable protein-DNA complex, can constitute a memory system that enables cells to store information about the expression history of specific segments in the

genome (Pohl, 1987; Droge and Pohl, 1991). Z-DNA may also be involved in cell terminal differentiation (Gagna et al., 1999).

I.3. Proteins recognizing Z-conformation

I.3.1. Antibodies against Z-conformation

Compared to B-DNA, Z-DNA is a strong antigen. Anti-Z-DNA antibodies induced in rabbits and mice were highly specific for the Z-DNA structure (Lafer et al., 1981). Brominated and unbrominated poly (dG-dC)·poly(dC-dG) mixed with methylated bovine serum albumin were used to immunize rabbits and mice. Antibodies specific for Z-DNA were found not only in the sera of animals immunized with brominated poly (dG-dC)·poly(dC-dG) which was in Z-form but also in sera from animals immunized with the unbrominated polymer in B-DNA. The antibodies raised against Z-DNA are highly specific for Z-conformation and can recognize unmodified poly (dG-dC)·poly(dC-dG) in 4 M NaCl, in which the polymer assumes the Z-conformation (Thamann et al., 1981). Naturally occurring Z-DNA antibodies were found in the sera from patients of systemic lupus erythematosus (Lafer et al., 1983; Sehgal and Ali, 1990) but not in other rheumatic diseases. Monoclonal antibodies against Z-DNA have been exploited to detect Z-DNA conformation in plasmids (Lang et al., 1982), to isolate plasmids containing Z-conformation (Thomae et al., 1983), to detect the occurrence of Z-DNA *in vivo* (Hill and Stollar, 1983; Gagna et al., 1991; Wittig et al., 1992; Wolfl et al., 1996) and to stabilize the formation of Z-DNA *in vitro* (Lafer et al., 1985). The limitation of Z-DNA antibodies is that it cannot be used to detect Z-DNA in intact cells.

Besides antibodies against Z-DNA, antibodies raised against Z-RNA also had been produced and used to detect RNA in Z-conformation (Zarling et al., 1987). By using antibodies specific for Z-RNA which did not cross-react with Z-DNA, some cytoplasmic RNA in fixed cells existed in the left-handed conformation (Zarling et al., 1987). Cytoplasmic microinjection of antibodies recognizing Z-RNA could inhibit human cell growth and, interestingly, the antibodies microinjected in the nuclei bound nucleolar RNA and these complexes appeared to be eliminated from the nucleus within min (Zarling et al., 1990).

I.3.2. Z-DNA binding proteins

It is postulated that Z-DNA functions as a potential *cis* element, although the exact role of Z-DNA in biological processes is still unknown. In principle, the event of Z-DNA formation could have a functional role that does not need recognition by certain proteins. However, identification of proteins that bind to Z-DNA in a structure-specific fashion could help to understand the biological role of Z-DNA *in vivo*. In the past decades, several Z-DNA binding proteins have been reported in vaccinia virus (Liu et al., 2001; Kim et al., 2003), Epstein-Barr virus (Bhende et al., 2005), *E. coli* (Blaho and Wells, 1987; Lafer et al., 1988; Krishna et al., 1990), *Drosophila* (Nordheim et al., 1982; Arndt-Jovin et al., 1993), *S. cerevisiae* (Zhang et al., 1992), wheat germ (Lafer et al., 1985), zebrafish (Rothenburg et al., 2005), chicken (Herbert et al., 1993), rat (Rothenburg et al., 2002), bovine (Rohner et al., 1990) and human cell nuclei (Leith et al., 1988; Pham et al., 2006). Although these proteins have not

been characterized in detail, numerous evidence suggest that they play important roles in gene expression and regulation, DNA recombination, RNA editing, viral pathogenicity, and tumor development (Rich and Zhang, 2003). Thanks to Dr. Alexander Rich, three Z-DNA specific binding proteins, ADAR1 (Herbert et al., 1997), E3L (Liu et al., 2001) and DLM1 (Schwartz et al., 2001), have been characterized in detail and the Z-DNA binding domains from these three proteins show similar Z-DNA binding capability and specificity.

I.3.2.1. Structures of *Zα* domains and their interactions with Z-DNA

Upon identification of Z-DNA binding ability and specificity of the Z-DNA binding proteins, the interactions between Z-DNA and various Zα domains have been investigated extensively by crystallography, NMR, etc. The binding ability of Zα to Z-DNA is very tight with a K_D of 4 nM, based on BIAcore SPR (surface plasmon resonance) analysis using brominated $d(GC)_3$ as the substrate (Herbert et al., 1998). The binding abilities of Zα to poly(dGdC) with and without bromination or methylation under different salt concentrations were checked by various spectroscopy techniques, including circular dichroism (CD), Laman spectroscopy and fluorescence spectroscopy (Berger et al., 1998). The specific interaction between Zα and Z-DNA segment in negatively supercoiled plasmid was visualized by atomic force microscopy (Lushnikov et al., 2004). The classical inversion of CD spectrum by B-to-Z transition was observed when Zα bound to unmodified d(GC) polymer at low salt concentration (**Figure I-2**), suggesting that Zα was able to induce and stabilize the B- to Z-conformation transition at permissive sequences. Studies using Laman

spectroscopy confirmed the specificity of Zα binding to Z-DNA. The solution structure resolved by NMR (Schade et al., 1999) and crystallography (Schwartz et al., 1999) suggested that Zα consisted of three α-helices and three β-sheets with an α1β1α2α3β2β3 topology, as shown in **Figure I-6**. The three α-helices are perpendicular to each other and the hydrophobic core is buried inside. This structure is also called as helix-turn-helix (HTH, in α2α3), winged (β2β3 form the wing) structure. The superposition of solution structure of free Zα with crystal structure shows that free Zα has almost an identical structure as Zα in complex. Both NMR and crystallography studies show that two Zα bind one Z-DNA segment, with each monomer recognizing one DNA strand. However, two Zα monomers do not interact with each other in the complex, suggesting formation of dimer is not a pre-requisite for binding. On the binding surface, extensive hydrogen bonds are formed between amino acid residues in α3 and β3 and five consecutive phosphates in the backbone of DNA with and without mediation of water, as shown in **Figure I-8**. In addition to polar interaction, close van der Waals contact is observed between the face of the aromatic ring of Tyr^{177} with the exposed carbon 8 of G4 base, the latter is in *syn* conformation and is a characteristic of Z-DNA. Pro^{192} and Pro^{193} in β3 form another important van der Waals contact with the DNA backbone. In contrast to the predominant *trans* peptide configuration, Pro^{192} forms a *cis* peptide bond, rendering its uniqueness and importance in the binding.

After the discovery of Zα, more Z-DNA binding proteins have been discovered and confirmed by blasting protein database with the sequences of amino acids

crucial for Z-DNA binding, including E3L from vaccinia virus (Kahmann et al., 2004) and its orthologue in Yaba-like disease virus (Ha et al., 2004), tumor-associated protein DLM-1(Schwartz et al., 2001) and PKR-like PKZ in zebrafish (Rothenburg et al., 2005). All of them possess the conserved Z-DNA binding domain, as shown in **Figure I-7**. E3L has lower binding affinity, compared to Zα. Solution structure analysis by NMR showed that the side chain of Tyr takes a different conformation in free $Zα_{E3L}$ (Kahmann et al., 2004), although the overall structure of protein-DNA complex is very similar with that of Zα. Upon binding to Z-DNA, a rearrangement of the side chain of Tyr occurs in order to stabilize the binding, resulting in the lower binding affinity of $Zα_{E3L}$ to Z-DNA. Superposition of structures of Zα, $vvZα_{E3L}$, $yabZα_{E3L}$, $Zα_{DLM-1}$ complexed with Z-DNA suggested that the structure of HTH core is conserved and prominent variations are found in the wing structure (Ha et al., 2004).

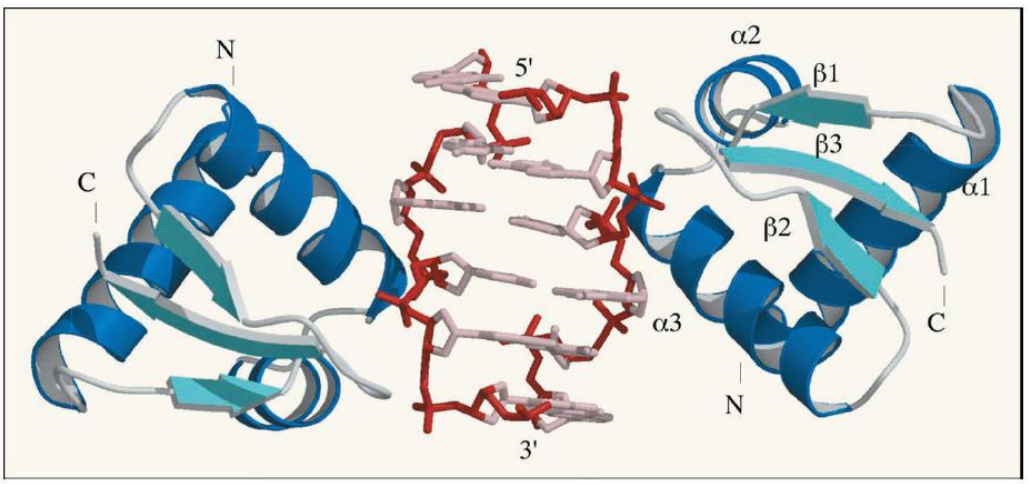


Figure I-6. Overview of Zα binding to d(GC)3 oligomer in Z-conformation.
The topology of Zα is shown (Schwartz et al., 1999) and C- and N- terminal of Zα are indicated. Backbone of Z-DNA is shown in red.

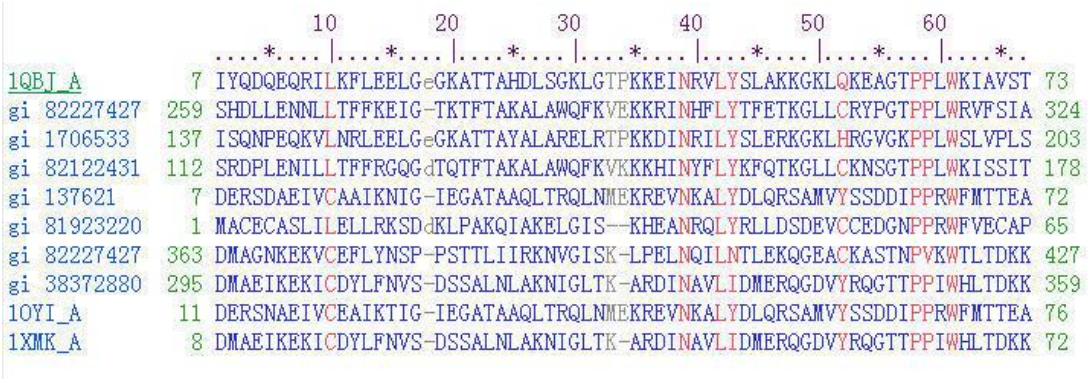


Figure I-7. The conserved domain of Z-DNA binding proteins

The multiple alignments show the conserved amino acid residues in red which are pivotal for Z-DNA binding. (CDD id: pfam02295.13) (Marchler-Bauer et al., 2005). **1QBJ_A**: chain A, crystal structure of the Zα/Z-DNA complex; **gi 82227427**: DsRNA adenosine deaminase (Xenopus laevis); **gi 1706533**: ADAR (Rattus norvegicus); **gi 82122431**: DsRNA adenosine deaminase (Xenopus laevis); **gi 137621**: Protein E3 (p25) (Vaccinia virus); **gi 81923220**: Interferon resistance determinant protein (Orf virus); **gi 38372880**: ADAR (P136); **1OYI-A**: sequence A in the solution structure of the Z-DNA binding domain of the Vaccinia virus gene E3L; **1XMK_A**: sequence A in the crystal structure of the Zβ domain from ADAR1.

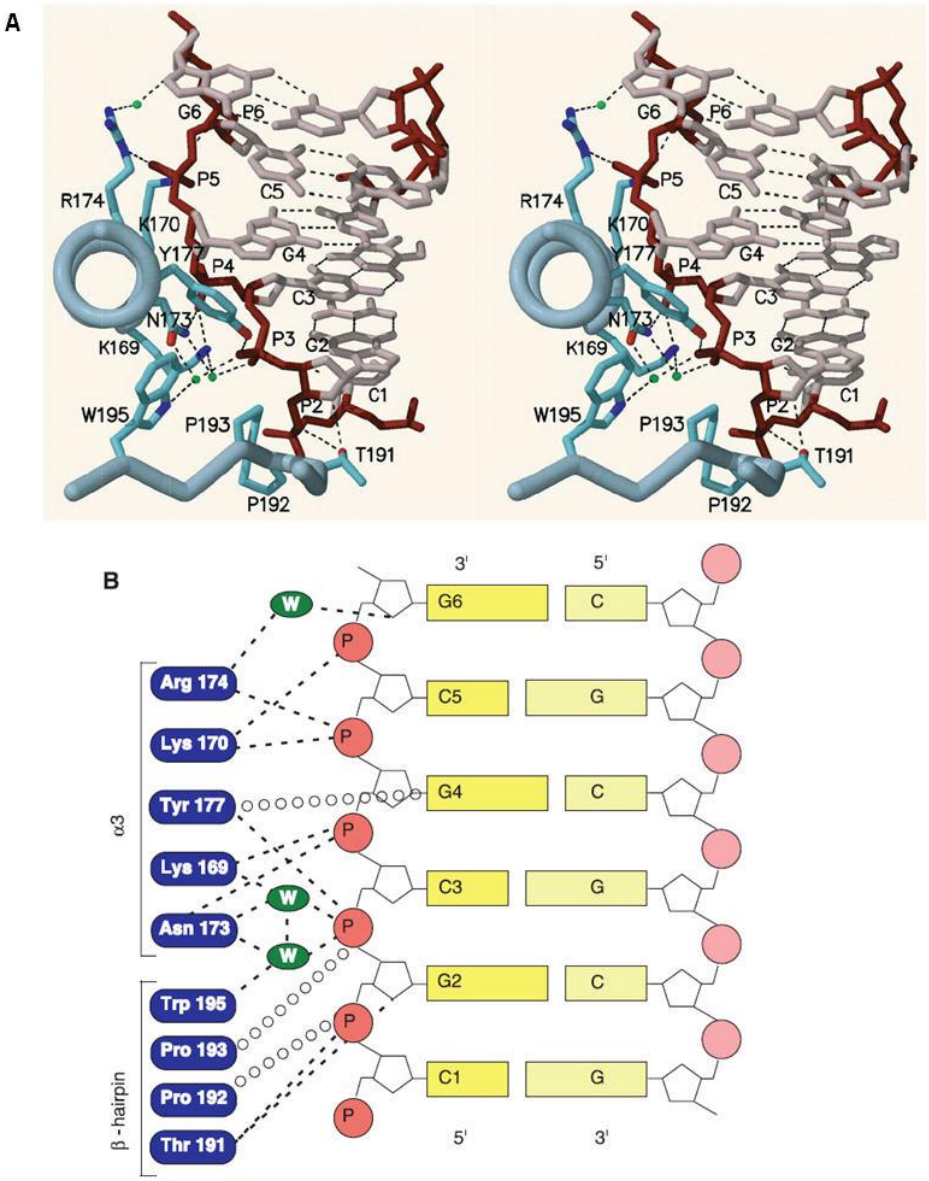


Figure I-8. The binding surface between Zα and Z-DNA.

A. Stereoview from the axis orientation of helix α3. Extensive hydrogen bonds are formed in the interaction surface with or without mediation of water (green). Five consecutive phosphates in the DNA backbone are recognized directly. The aromatic ring of Tyr177 is perpendicular with G4 and involved in close van der Waals interaction. B. Schema of the binding surface. The residues critical to Z-DNA specific binding are listed and the interactions are showed. Dashed line stands for hydrogen bond and open circle for van der Waals contact. Cited from (Schwartz et al., 1999)

I.3.2.2. Interactions between Z-RNA and Zα

Previous studies had indicated that Z-RNA was energetically more unfavorable than Z-DNA (Hall et al., 1984). Recently, Z-RNA structure was resolved by crystallizing Z-RNA-Zα complex (Placido et al., 2007). In this study, dUr(CG)3 duplex RNA was complexed to Zα and transited to Z-RNA conformation by the binding of Zα under physiological salt condition and low temperature. The co-crystal of Z-RNA showed a much conserved binding mode with Zα/Z-DNA. The protein maintains its winged helix-turn-helix (HTH) folding pattern and extensive hydrogen bonds are formed between residues in α3 and β3 and five consecutive phosphates in the backbone of the RNA segment. Most residues crucial for Z-DNA binding are involved in recognition of Z-RNA, except for Arg_{174}.

Despite much conserved binding mode between Zα/Z-DNA and Zα/Z-RNA, several differences are observed. The first difference is the packing of the segments in the crystal. Z-DNA co-crystals showed a pseudo continuous helix along the crystal; however, Z-RNA duplexes in the co-crystals are arranged with a 90^{o} angle between each segment. The second difference is Z-RNA possesses a unique solvent pattern, largely due to the presence of the cytosine 2'-OH groups, distinguishing it from Z-DNA. In Zα/Z-DNA complex, water molecules are involved in the interactions between Trp_{195} and Asn_{173} and the phosphate oxygen (**Figure I-8**), whereas in Zα/Z-RNA complex, a Na^{+} ion occupies the position of water molecule, suggesting the three conserved amino acid residues form a highly polarized pocket.

I.3.2.3. Adenosine deaminase on double-stranded RNA 1, ADAR1

ADAR is a protein family which converts adenosine to inosine within double-stranded regions of RNA (Maas et al., 1997). The postulated mechanism of A-to-I editing is shown in **Figure I-9**. ADARs are involved in both non-specific and site-specific A-to-I editing. Non-specific A-to-I hyper-editing by ADARs is observed on long uninterrupted dsRNA. Since extended dsRNA duplexes are relatively uncommon in cells and are frequently associated with infection by DNA or RNA viruses, the hyper-editing by ADAR is one possible antiviral defense mechanism by disrupting viral open reading frames (ORF). Extensive adenosine to inosine editing has been reported for various viruses, including measles virus (Bass et al., 1989; Cattaneo et al., 1989), human parainfluenza virus (Murphy et al., 1991), vesicular stomatitis virus (O'Hara et al., 1984), avian leukosis virus (Hajjar and Linial, 1995), respiratory syncitial virus (Rueda et al., 1994), and polyoma virus (Kumar and Carmichael, 1997). Site-specific A-to-I editing occurs on transcribed mRNA, in which inosine is read as guanosine by the translation machinery, resulting in codon changes and increased protein diversity. The examples of site-specific editing are editing of glutamate receptor subunit B (gluR-B) mRNA(Wang et al., 2000; Wang et al., 2004), 5-HT_{2C} serotonin receptor mRNA (Burns et al., 1997; Fitzgerald et al., 1999) and hepatitis delta virus (HDV) RNA (Luo et al., 1990; Zheng et al., 1992). The biological activities of ADARs include antiviral defense, involvement in apoptosis and embryogenesis, increase of neurotransmitter receptors diversity, and interference with RNAi pathway, as reviewed by Toth,

etc.(Toth et al., 2006).

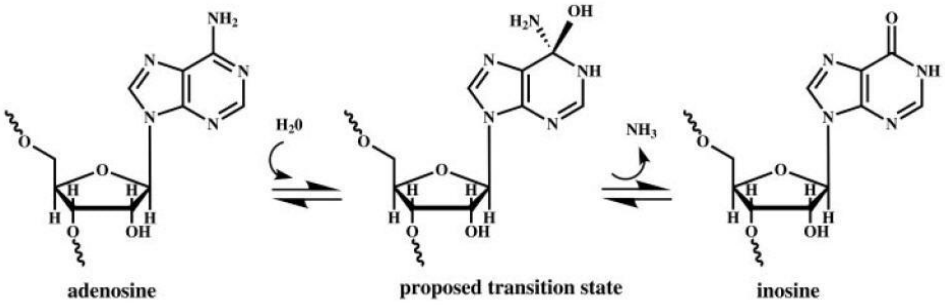


Figure I-9. The postulated mechanism of A-to-I conversion by ADARs.

The transition state is observed with cytidine deaminases (CDAs), which share similarity with ADARs. Cited from (Bass, 2002).

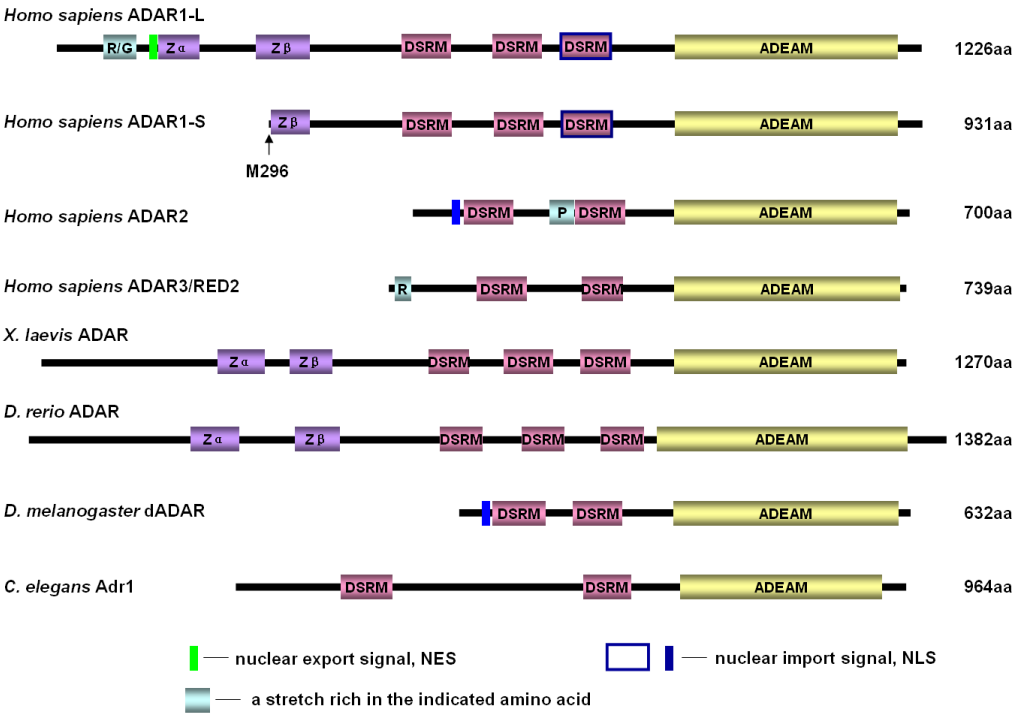


Figure I-10. The primary structures of ADARs in several species.

The Z-DNA binding domains (Zα and Zβ), double-stranded RNA binding motif (DSRM) and adenosine deaminase motif (ADEAM) are shown. The hADAR1-S was translated from M296 in hADAR1-L. The lengths of ADARs are indicated at the right side. aa: amino acid.

ADARs are widely conserved in their adenosine deaminase domains, but differ

in dsRNA binding motifs and Z-DNA binding motifs, as shown in **Figure I-10**. ADARs have been identified in multiple species including mammals (Kim et al., 1994; Melcher et al., 1996; Chen et al., 2000), *Xenopus* (Bass and Weintraub, 1988), *Drosophila* (Palladino et al., 2000; Palladino et al., 2000), and Zebrafish (Slavov et al., 2000). Two different forms of catalytically active ADAR1 proteins are expressed from the *ADAR1* gene (Patterson and Samuel, 1995): an interferon (IFN)-inducible p150 form ADAR1-L and a constitutively expressed p110 form ADAR1-S; the latter's translation was initiated from M296 in exon 2 in ADAR1-L, as shown in **Figure I-10**. In addition to the single copy ADAR1 gene, two more mammalian ADAR genes and proteins have been identified (**Figure I-10**). ADAR2 is constitutively expressed and catalytically active (Melcher et al., 1996; O'Connell et al., 1997), while ADAR3, also termed as RED2, does not have deaminase activity (Chen et al., 2000), although it has all the key conserved residues that are known to be required for catalysis.

In ADAR family, only ADAR1-L contains two Z-DNA specific binding domains, named Zα and Zβ. Z-DNA binding domains are conserved features of human, rat, bovine, chicken and *xenopus* ADAR1 (Herbert et al., 1997), suggesting the importance of Z-DNA binding domains for the biological function of ADAR1. Comparison between ADAR1-L with other human ADARs suggests that the presence of Z-DNA binding domains clearly is not an obligate requirement for dsRNA binding or deaminase activity of ADAR1. However, mutations in Zα in ADAR1-L decreased the efficiency of editing short dsRNA substrates (Herbert and Rich, 2001), and the *in vitro* ability of ADAR1-L to edit

a 50-bp segment of dsRNA was greatly enhanced when a sequence favoring Z-RNA form was used as substrate (Koeris et al., 2005). The inducibility of expression of ADAR1-L by interferon and inflammation, but not other ADARs, implies the special involvement of Z-DNA binding domains in antiviral defense and inflammation.

A novel role for ADAR1 in the regulation of gene expression has been reported, whereby ADAR1 interacts with nuclear factor 90 (NF90) proteins, known regulators that bind the antigen response recognition element (Nie et al., 2005). Notably, the regulation of both nuclear translation and regulation of NF90-mediated gene expression occurs independently of its RNA-editing activity.

The subcellular localization of ADAR1-L has been studied in detail. Immunostaining and biochemical fractionation studies initially revealed that ADAR1-L was primarily a nuclear protein (Kim et al., 1994; O'Connell et al., 1995). Further immunofluorescence microscopy and Western blotting assay demonstrated that constitutively expressed ADAR1-S (p110) was localized predominantly in nucleus, especially in or on nucleolus, whereas the IFN-inducible ADAR1-L (p150) was localized to both cytoplasm and the nucleus (Patterson and Samuel, 1995). But there are some arguments that ADAR1-L is predominantly in cytoplasm, but not in the nucleus (Yang et al., 2003). ADAR1-L is initially transported into the nucleus and then re-exported by a nuclear export signal (NES) near the N-terminus (Poulsen et al., 2001).

Leptomycin B (LMB), a drug that inhibits the nuclear export protein Crm1, stimulates nuclear accumulation of ADAR1-L. In ADAR1-L, at least two, and possibly three, functional nuclear localization signals (NLSs) exist. An atypical nuclear/nucleolus localization signal (NLS/NLoS) resides within DSRM3 (dsRNA binding motif 3), so the full length ADAR1-L shows the characteristics of a shuttling protein (Eckmann et al., 2001; Strehblow et al., 2002). Another NLS was reported in the C-terminal 39 amino acid residues of mouse ADAR1 (Nie et al., 2004), and an N-terminal NES might also exist in the first 269 amino acid region of ADAR1 (Poulsen et al., 2001). Shuttling of ADAR1-L between the cytoplasm and the nucleus may affect which dsRNA substrates are edited. Wild type ADAR1-L localized in the cytoplasm has poor editing activity of GluR-B mRNA, but both mutated ADAR1-L in NES region and N-terminus truncated ADAR1-L can edit GluR-B mRNA efficiently (Poulsen et al., 2001). Localization in the nucleus and aggregation on the surface of nucleoli of ADAR1-S may involve in the regulation of nuclear translation and protein synthesis (Herbert et al., 2002).

I.3.2.4. E3L from vaccinia virus

E3L, similar to ADAR1, contains a Z-DNA binding motif (Herbert et al., 1997) in the N-terminal half and a dsRNA binding motif in the C-terminal half (Chang et al., 1992; Chang and Jacobs, 1993; Yuwen et al., 1993). It has been reported that E3L localized to both cytoplasm and nucleus, with nuclear localization being independent on the N-terminal region (Yuwen et al., 1993; Chang et al., 1995). Upon infection, vaccinia virus encodes several proteins to antagonize the

antiviral defense of the interferon (IFN) system. E3L protein can rescue the replication of vaccinia virus in the cells treated with IFN-α (Davies et al., 1993; Chang et al., 1995). Resistance to IFN is restored by transient transfection of E3L (Chang et al., 1995), through introducing recombinant E3L into the virus itself (Shors et al., 1997), or by introducing another dsRNA binding protein such as the reovirus σ3 protein (Beattie et al., 1996). It has been found that functional dsRNA binding ability is necessary for IFN-resistance of vaccinia virus. Mutations that destroy the dsRNA binding ability impair the ability of E3L to rescue the replication of the virus in the treatment of IFN. N-terminus of E3L is dispensable for the rescue, since the deletion of N-terminus of E3L did not affect the ability to rescue viral replication in cell culture (Chang et al., 1995; Shors et al., 1997). However, the pathogenesis of vaccinia virus in mouse model requires both the C-terminus and the N-terminus, in conflict with the cell culture study (Brandt and Jacobs, 2001). Vaccinia viruses containing an N-terminal deletion of E3L are attenuated for neuropathogensis and fail to infect the lungs or brains in mice.

The Z-DNA binding ability of Z-DNA binding domain of E3L ($Z\alpha_{E3L}$) has been analyzed *in vitro* and it is reported that the binding affinity of $Z\alpha_{E3L}$ to Z-DNA is lower than that of $Z\alpha_{ADAR1,}$ due to residue Y48 in the Zα domain that adopts a different side chain conformation and requires conformation change for binding to Z-DNA (Kahmann et al., 2004). When $Z\alpha_{E3L}$ was replaced with its homologues in ADAR1 or DLM1, the virus retains its lethality after intracranial inoculation (Kim et al., 2003). Mutation assays suggested that Z-DNA specific

binding ability of E3L was essential to its activity. It has been reported recently that E3L or $Z\alpha_{E3L}$ alone can selectively increase reporter gene expression in Hela cells from a basic promoter with TATA box when different elements were inserted upstream of TATA box (Kwon and Rich, 2005). Beside the gene transactivation, both E3L and $Z\alpha_{E3L}$ show strong anti-apoptosis capability in a dose- and time- dependent manner. Mutation analysis suggested Z-DNA binding ability is vital for this activity (Kwon and Rich, 2005).

Because of the high homology between $Z\alpha_{ADAR1}$ and $Z\alpha_{E3L}$, and the opposite correlation with IFN, which means that ADAR1 is induced by IFN while E3L provides resistance to IFN, the interaction between E3L and ADAR1 is very interesting. It is found that E3L can antagonize the deaminase activity of ADAR1-L and ADAR1-S in co-transfected cells, which requires the intact dsRNA binding motif in E3L (Liu et al., 2001). Surprisingly, disruption of the Z-DNA binding domain of E3L by double substitutions of two highly conserved residues abolished its antagonistic activity, whereas deletion of the entire Z domain had little influence, suggesting that protein conformation may affect biochemical function of E3L in a manner that is mediated by the C-terminal region including the dsRNA binding motif (Liu et al., 2001). Binding of either RNA or DNA by $Z\alpha_{E3L}$ has the potential to induce conformational changes that affect protein function by altering subsequent protein-protein interaction, protein-dsRNA interaction involving the C-terminal dsRNA binding motif, or both.

I.3.2.5. Tumor-associated protein DLM-1

The Z-DNA binding capability of DLM1 has been confirmed by its crystal structure (Schwartz et al., 2001) and circular dichroism (CD) analysis. *DLM-1* was firstly detected as a novel gene in the peritoneal lining tissue of mouse inoculated with ascites tumors and tumor stromal cells by using RNA differential display (Fu et al., 1999). Expression of DLM-1 in macrophages was greatly up-regulated by stimulation with IFN-γ and lipopolysaccharide (LPS). The biological function of DLM-1 is still unclear, although it is proposed that DLM-1 plays roles in defense against tumor by involving in neoplasia and angiogenesis. Further analysis showed that about 2000 different mRNAs of DLM-1 can be produced due to alternative splicing and different 5' and 3' ends, and only the N-terminal region contains two conserved Z-DNA binding domain, e.g. Zα and Zβ, similar to ADAR1-L (Rothenburg et al., 2002). Recently, cellular localization of DLM-1 was studied and found that full length DLM-1 localized in the cytoplasm with a punctate distribution pattern (Deigendesch et al., 2006). Interestingly, under environmental stress, DLM-1 accumulated in stress granules (SGs) which are composed of stalled 48S pre-initiation complexes, translation initiation factors and many RNA binding proteins (Kedersha et al., 2005). Deletion of Zα eliminated the association with stress granules, suggesting that DLM-1 may participate in the regulation of translation and mRNA sorting. Recent study shows that DLM-1 (named as DAI) can function as a cytosolic DNA sensor to activate innate immune response (Takaoka et al., 2007). In this study, poly (dT-dA)/poly (dA-dT) in B-conformation and DNA from different sources (including bacteria, virus,

mammalian DNA) were found to be able to evoke the expression of DLM-1 and the subsequent up-regulation of expression of IFN-1. poly d(GC)/d(CG) which may take Z-conformation has the similar effect with B-form poly (dT-dA)/poly (dA-dT). Thus, the function of DLM-1 in cytoplasm relies on B-DNA. However, the possibility of the involvement of Z-DNA can not be excluded.

I.4. The scope of the thesis

Chemical, physical, biochemical and biological properties and functions of Z-DNA have been characterized in many aspects. Another important question yet to be answered is where Z-DNA segments are in the human genome. The distribution information of Z-DNA along the human genome will be greatly helpful to further understand its biological functions. In this study, Zα was used as the Z-DNA specific probe to map the distribution of Z-DNA fragments in human genome and all the experiments were controlled with Zα mutant in which two amino acids residues essential for Z-DNA binding were mutated. In order to simplify the experimental process, FLAG and StrepII epitope tags were added to the C-terminus of Zα with or without NLS (nuclear localization signal). The binding ability of the designed Zα probe to Z-DNA was examined. The subcellular localization of Zα and Zα mut were examined.

In this study, initially direct *in vivo* chromatin affinity precipitation (ChAP) was exploited. However, *in vivo* ChAP showed that in mammalian cells the majority of material precipitated by Zα was RNA. Then a novel *in vitro* ChAP protocol

was proposed based on the observation that crosslinking of mammalian cells by formaldehyde could stabilize Z-DNA in the genome and the formaldehyde crosslinking did not affect the recognition by Zα (Liu et al., 2006). *In vitro* ChAP DNA fragments were used to construct a Z-DNA library and sequenced. The first Z-DNA map of the human genome was made and the biological significance of Z-DNA in mammalian cells was discussed.

In our study several lines of evidence pointed out that RNA is the major *in vivo* substrates for Zα, rendering us an opportunity to study the naturally occurring Z-RNA in mammalian cells. The precipitated RNA fragments were identified as rRNA in functional ribosomes. The binding sites of Zα on mammalian ribosome were cloned and identified; furthermore, the influence of Zα binding of ribosome on *in vitro* translation was checked.

II. MATERIALS AND METHODS

II.1. Mapping Z-DNA in the human genome using Zα as a specific probe

II.1.1. Materials

Plasmids

plasmid	from
pET17b	Invitrogen
pET28a-Za77	From (Herbert et al., 1998)
pET17b-Zα-NLS-FS	this study
pET17b-Zαmut-NLS-FS	this study
GT-pPGKss-Puro	this study
pTZ18R	Pharmacia
p101	Lab stock
pPGKss-Zα-NLS-FS	this study
pPGKss-Zαmut-NLS-FS	this study

Primers and oligonucleotides sequences

pET17b-Za-NLS-FS	
Za-U:	5'AGATATACATATGGGGCTGAGTATCTACCAAG3'
Za-L1:	5'CTCTTATCGTCGTCATCCTTGTAATCAACCTTCCTCTTCTTCTTAGGACCGCCGCTTCCACCTCC3'
Za-L2:	5'CAGAATTCTCATTATTTTTCGAACTGCGGGTGGCTCCAAGCGCTCTTATCGTCGTCATCCTTGTA3'
pET17b-Zamut-NLS-FS	

U-mut-Za-FS	5'ATCGCCCGAGTTTTAGCCTCCCTGGCAAAGAAGGGCAAG3'
L-mut-Za-FS	5'GGAGGCTAAAACTCGGGCGATTTCTTTCTTCGGAGTCC3'
GT-pPGKss-Puro	
EcoR.1-U	5'CGGTATCGATAAGCTTGATAT3'
EcoR.1-L	5'GTGCCAGCGGGGCTGCTAAAG3'
GT-N-U:	5'GGG CTG GCA GAA TAT CAA GT 3'
GT-N-L:	5'TCA TTG TGC ACT CAA CCG TGT GA3'
adaptor for ChIP DNA	
Adaptor-L	5'aaacGAATTCGAGGAGATTATGGATCCGAC3'
Adaptor-S	5'pGTCGGATCCATAATCTCCTCGAATTCGT3'
primers for pTZ18R	
pTZ-U	5'GCCTGCAGGTCGACTCTAGAGGA3'
pTZ-L	5'GATTACGAATTTAATACGACTCACTA3'
pPGKss-Za-NLS-FS	
U (Pst I)	5'CCAACTGCAGATGGGGCTGAGTATCTACCAAG3'
Internal-U (XbaI)	5'GAAAAATAATGATCTAGAGCTCGCTGATCAGC3'
internal-L (XbaI)	5'CGAGCTCTAGATCATTATTTTTCGAACTGC3'
L (Not I)	5'AGCTCCACCGCGGTGGCGGCCGCTCTA3'

Bacteria strains

DH5α

F$^-$, φ80d*lacZ*ΔM15, Δ(*lacZYA-argF*)U169, *deoR*, *recA*1, *endA*1, *hsdR*17(rk$^-$, mk$^+$), *phoA*, *supE*44, λ$^-$, *thi*-1, *gyrA*96, *relA*1 (Hanahan, 1983).

BL21(DE3)

F-, *omp*T, *hsd*S$_\beta$(r$_\beta$-m$_\beta$-), *dcm*, *gal*, (DE3) (from Stratagen).

Stbl4

*mcr*A Δ(*mcr*BC-*hsd*RMS-*mrr*) *rec*A1 *end*A1 *gyr*A96 *gal*$^-$ *thi*-1 *sup*E44 λ$^-$ *rel*A1 Δ(*lac-pro*AB)/F'*pro*AB$^+$ *lac*I^{q}Z ΔM15 Tn*10* Tet^R (from Invitrogen).

Media for bacteria culture

LB media

1% (w/v) bacto-tryptone,

0.5% (w/v) yeast extract,

1% (w/v) NaCl

LB agar plates

LB media plus 1.5% (w/v) bacto-agar.

LB agar plates with Amp selection

LB agar plates with 300 μg/ml ampicillin

LB agar plates for BL21(DE3)

LB agar plates with 300 μg/ml ampicillin, 34 μg/ml chloramphenicol, 15 μg/ml tetracycline.

* Ampicillin was prepared at 100 mg/ml in ddH_2O and sterilized by filtration through 0.22 μm filter.

Chloramphenicol and Tetracycline were prepared at 34 mg/ml and 15 mg/ml in ethanol, respectively.

SDS-PAGE

Stock solutions for SDS-PAGE gel:

30% acrylamide/Bis solution (29:1); 1.5 M Tris (pH 8.8); 10% SDS; 10% AP; TEMED; 1 M Tris (pH 6.8);

SDS-PAGE

15% resolving gel

15% (v/v) acrylamide/Bis solution (29:1), 25% (v/v) 1.5 M Tris-HCl (pH 8.8), 0.1% (w/v) SDS, 0.1% (w/v) AP, 0.04% TEMED.

Stacking gel

5% (v/v) acrylamide/Bis solution (29:1), 12.5% (v/v) 1.0 M Tris-HCl, (pH 6.8), 0.1% (w/v) SDS, 0.1% (w/v) AP, 0.1% TEMED.

2x SDS gel loading buffer

100 mM Tris-HCl (pH 6.8), 4% (w/v) SDS, 0.2% (w/v) bromophenol blue, 20% (v/v) glycerol, 200 mM dithiothreitol (DTT)

5x Tris-glycine electrophoresis buffer

25 mM Tris, 250 mM glycine, 0.1% (w/v) SDS, pH 8.3

Staining reagent

2.5% (w/v) coomassie brilliant blue R-250, 45% (v/v) methanol, 10% (v/v) glacial acetic acid

Destaining reagent

45% methanol, 10% glacial acetic acid, 45% dH_2O.

Enzymes

Restriction enzymes	NEB
T4 DNA ligase	NEB
T4 polynucleotide kinase	NEB
Alkaline phosphatase, Calf intestinal (CIP)	NEB
T4 DNA polymerase	NEB
T4 RNA ligase I	NEB

Kits

Qiaprep Spin miniprep kit	Qiagen
Qiaquick Gel extraction kit	Qiagen
Qiaquick PCR purification kit	Qiagen
Qiagen Plasmid Midi Kit	Qiagen
Endofree Plasmid Maxi Kit	Qiagen
DyeEXTM 2.0 spin kit	Qiagen
Lumi-lightplus Western blotting kit	Roche

II.1.2. Methods

II.1.2.1. Common techniques

II.1.2.1.1. Preparation of competent cells

DH5α competent cells

DH5α was plated on LB agar plate and incubated at 37 ^{o}C for 12-14 h. A single colony was inoculated into 50 ml of LB media. Cultures were incubated at 37 ^{o}C, 250-300 rpm until the OD_{600nm} reached 0.375-0.4. Bacteria were harvested by centrifugation at 4 ^{o}C, 4000 x g for 5 min. The pellet was resuspended in 30 ml of ice-cold $MgCl_2$-$CaCl_2$ solution (80 mM $MgCl_2$, 20 mM $CaCl_2$) and incubated on ice for 20 min. Then the cells were centrifuged at 4 ^{o}C, 4000 x g for 5 min. The supernatant was discarded and the tube was put in an inverted position on a pad of paper towels for 1 min. Finally the pellet was resuspended in 2 ml of cold 0.1 M $CaCl_2$ containing 17% glycerol and aliquoted into 100 µl/tube and stored at -80 ^{o}C.

BL21(DE3)

LB agar plates and LB media used for BL21-DE3 competent cells contain 34 µg/ml Chloramphenicol and 15 µg/ml tetracycline. The preparation of BL21(DE3) competent cells is the same as that of DH5α.

II.1.2.1.2. Transformation of competent cells

Heat shock transformation

100 µl of competent cells from -80 ^{o}C storage were put on ice to thaw slowly and were mixed with ligation products or plasmid DNA. The mixture was kept on ice for 30 min, followed by heat shock at 42 ^{o}C for 90 seconds, then incubated on ice for 5 min; cells were recovered by adding 880 µl of LB media and shaking at 37 ^{o}C, 230 rpm for 40 min. After incubation, the cells were harvested by centrifuge at 16,000 x g for 1 min. The supernatant was discarded, and the cell pellet was resuspended in 100 µl of fresh LB media and plated onto LB agar plate with proper antibiotics. LB agar plate was incubated at 37 ^{o}C overnight.

Quick transformation

Quick transformation was used when intact plasmids were introduced into competent cells. 100 µl of competent cells were mixed with 50 ng of plasmids and incubated on ice for 5 min, followed by heat shock incubation at 42 ^{o}C for 90 second. After incubated on ice for 5 min, the cells were spread directly on LB plate with proper antibiotics and incubated in 37 ^{o}C overnight.

Electroporation of ElectroMAX Stbl4 cells

DNA ligation reaction was purified with phenol/chloroform extraction and precipitated by ethanol in the presence of 20 μg of glycogen (Roche). The DNA sample was resuspended in 6 μl of dH_2O and mixed with 40 μl of Stbl4 competent cells. The mixture was transferred into 1 mm electroporation cuvette (Bio-Rad) and electroporated by Bio-Rad GenePulser II electroporator. The settings for electroporation were 1.2 kV, 25 μF, 200Ω, as recommended (Invitrogen). After electroporation, Stbl4 cells were inoculated into 2 ml of SOC media without antibiotics and recovered by shaking at 30 ^{o}C for 90 min; whereafter 200 μl of SOC culture medium was plated on LB agar plate with antibiotics and 50 μg/ml X-gal and incubated over night in 30 ^{o}C incubator. The remained 1.8 ml of transformed Stbl4 cells were mixed with 1.8 ml of fresh SOC medium containing 40% glycerol and stored in -80 ^{o}C freezer.

II.1.2.1.3. Phenol/chloroform extraction

Phenol/chloroform extraction is a common technique to get rid of protein contamination in DNA samples. An equal volume of Tris-saturated phenol was added to the DNA sample and mixed by vortex. After separation by centrifugation for 3 min at 13 100 x g in a bench-top microcentrifuge, the upper aqueous phase was transferred carefully to a new micro-centrifuge tube and mixed with equal volume of chloroform (chloroform: isoamyl alcohol=24:1) by vortex. If there was a visible interface, the phenol extraction was repeated. After centrifugation at 16,000 x g for 3 min, the upper aqueous phase was transfer to a new tube.

II.1.2.1.4. Ethanol precipitation

Ethanol precipitation is an efficient method to concentrate a DNA or RNA sample. 1/9 sample volume of 3 M KOAc was added into the sample and 2.5 sample volumes of ethanol was added. If the amount of a sample was less than 100 ng, 20 µg of glycogen was added to facilitate DNA precipitation. After gentle mixing, the DNA sample was cooled at -80 ^{o}C for 20 min and then spun down at 16 000 x g for 15 min at 4 ^{o}C. The supernatant was discarded carefully and 1 ml of 70% ethanol was added and the sampe was spun down again for 10 min. Subsequently, the supernatant was removed carefully and the sample was air-dried for 10 min.

II.1.2.2. Design of *Zα* probe and its mutant version

In this study, $Z\alpha_{ADAR1}$ (amino acid 133-209) from human ADAR1 (Genbank file NP_001102) was exploited as a specific probe to detect nucleic acids in Z-conformation in cultured mammalian cells. The $Z\alpha_{ADAR1}$ probe and its mutant were designed as shown in **Figure II-1**. For convenience, the $Z\alpha_{ADAR1}$ probe and $Z\alpha_{ADAR1}$ mut probe was termed as Zα and Zα mut in this thesis. The mutated amino acids were N173A and Y177A (amino acid numbering is given according to Genbank file NP_001102) since these two amino acids are crucial for specific Z-DNA binding (Schade et al., 1999; Kahmann et al., 2004). The FLAG epitope tag can be recognized by its monoclonal antibody (M2 FLAG antibody, Sigma) and the StrepII tag is an excellent candidate for affinity purification since the interaction between StrepII tag and specially engineered

streptavidin called Strep-Tactin is highly selective and easily controlled.

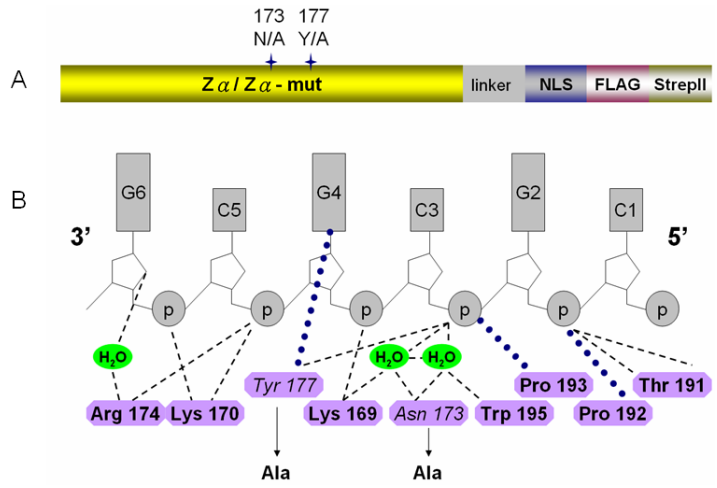


Figure II-1. The schema of Zα probe and its mutant.

A. The schema of Zα and Zα mut probe. There is a linker region (GGSGG) between Zα/mutant domain and SV40 Nuclear Localization Signal (NLS; PKKKRKV) which is followed by FLAG (DYKDDDDK) and Strep II (WSHPQFEK) epitope tags. Between FLAG tag and StrepII tag there is one short linker (SA). B. The amino acids residues crucial for specific Z-DNA binding. Modified from (Schwartz et al., 1999). The Z-DNA binding specificity of Zα mut was eliminated due to the indicated mutations.

II.1.2.3. Construction of expression vectors

II.1.2.3.1. Prokaryotic expression vectors for Zα probe and Zα mut

The process of construction of Zα prokaryotic expression vector was shown in **Figure II-2**. Zα (GI:2795789) was amplified by primer Za-U and primer Za-L1 from pET28a-Za77 (Herbert et al., 1998), which was a generous gift from Prof. Alexander Rich in MIT. The PCR product was purified from 1.5% agarose gel in 0.5 X TBE by using Gel Extraction Kit (Qiagen) and further amplified by Primer Za-U and primer Za-L2 to add the sequence encoding the epitope tags

(FLAG and Strep II) to Zα probe. After gel extraction, the PCR products were sequentially digested by *Nde*I and *EcoR*I and ligated into pET17b vector which was treated by the same endorestriction enzymes and CIP (Alkaline Phosphatase, Calf Intestinal; NEB). The stoichiometry of insert fragments to backbones was 3:1 in 20 µl of ligation reactions with 1 µl of T4 DNA ligase (NEB). The reactions were incubated at 16 ^{o}C overnight and then transformed into 100 µl of DH5α competent cells. Isolated colonies were inoculated into LB media with 300 µg/ml ampicilin and incubated at 37 ^{o}C, 250-300 rpm shaking overnight. The plasmids were prepared by using mini-prep kit (Qiagen) and validated by digestion and sequencing (by Research Biolabs Pte Ltd).

The mutations of Zα probe were introduced at N173A and Y177A by assembly PCR with mutation-specific primers (depicted in **Figure II-3**). Assembly PCR is the process of PCR amplification of a pool of oligonucleotides with short overlapping segments. The oligonucleotides with short overlapping segments are PCR templates. The upstream fragment was amplified by primer Za-U and primer L-mut-Za-FS and the downstream fragment was amplified by primer U-mut-Za-FS and primer Za-L2, both using plasmid pET17b-Za-NLS-FS as the templates. After separation on 1.2% agarose gel and extraction, equal molar amount of each PCR product was mixed and used as assembly PCR template which was amplified with primers Za-U and Za-L2 and inserted into pET17b vector as described above. The expression vector for Zα mut probe was named as pET17b-Zamut-NLS-FS.

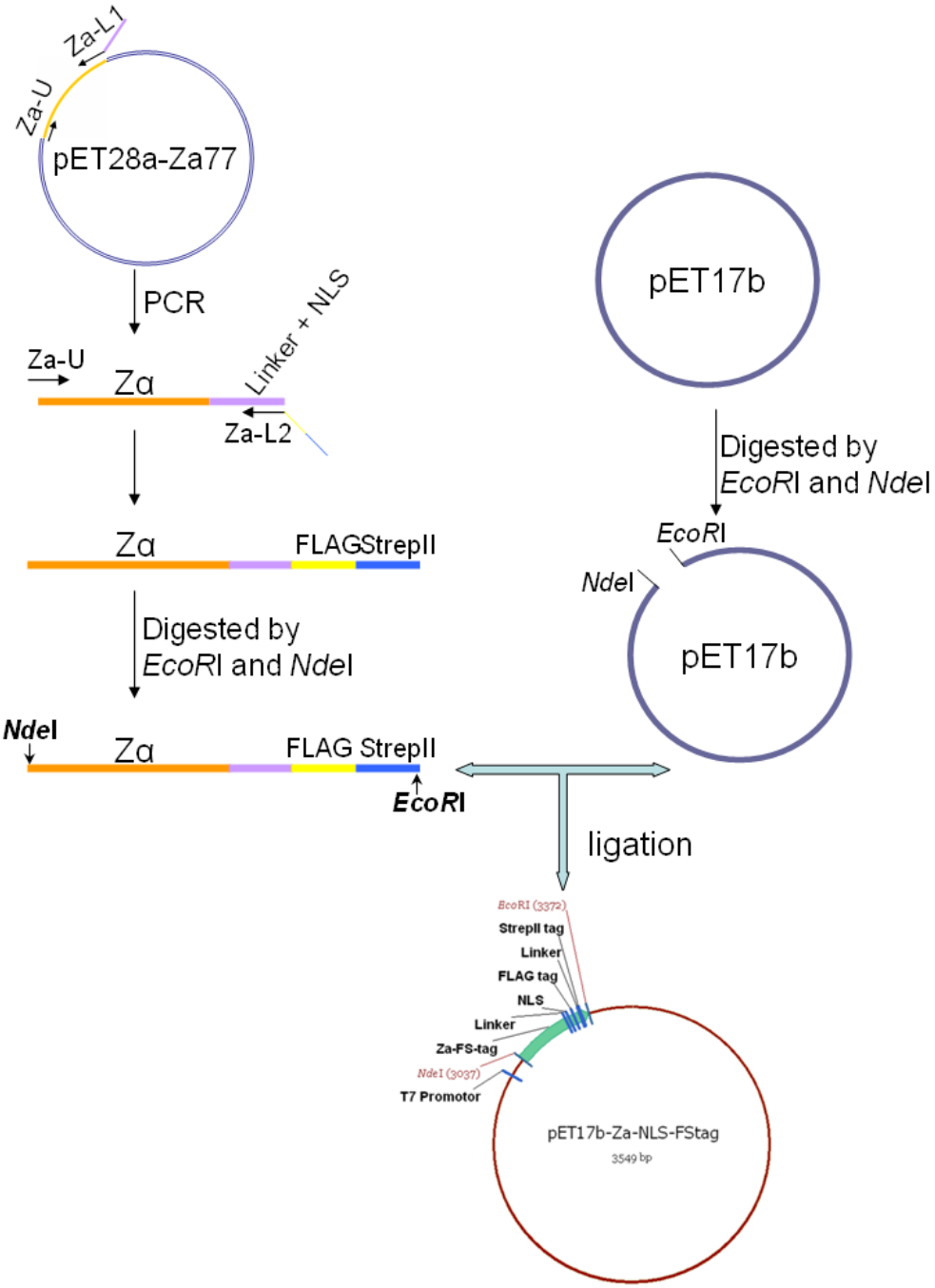


Figure II-2. Construction of Zα prokaryotic expression vector pET17b-Za-NLS-FS.
An SV40 NLS, a linker and two epitope tags (FLAG and StrepII) were added to the C-terminal of Zα by PCR. The construction of pET17b-Za-NLS-FS was described in the text.

II.1.2.3.2. Eukaryotic expression vector of Zα probe and Zα mut probe

pPGKss-Za-NLS-FS and pPGKss-Zamut-NLS-FS

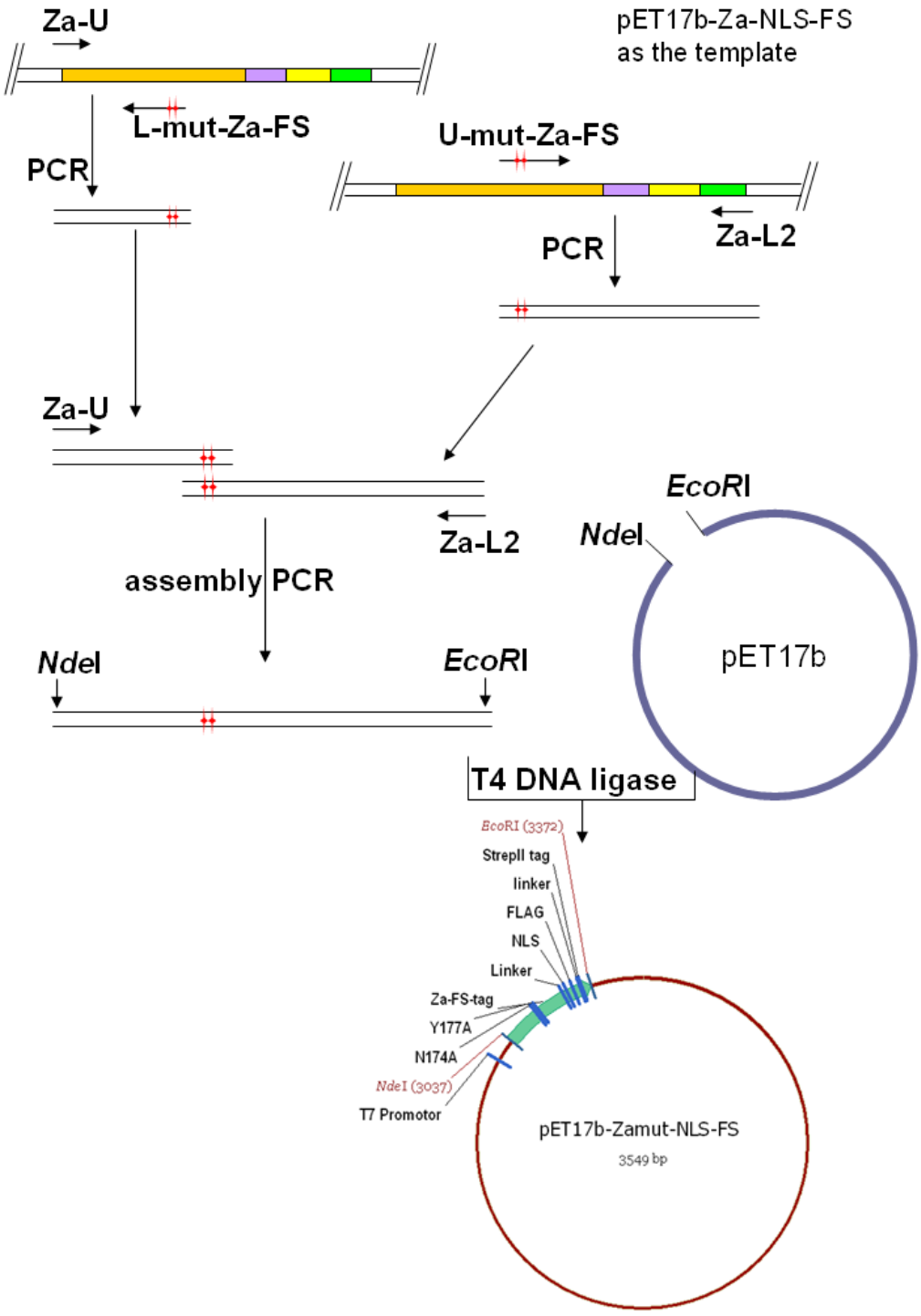


Figure II-3. The process of cloning pET17b-Zamut-NLS-FS.
The mutations in pET17b-Zamut-NLS-FS were introduced by assembly PCR. The experimental process was described in the text.

The sequences of Zα probe and Zα mut probe were amplified from

corresponding prokaryotic expression vectors by primers U (PstI) and internal-L (XbaI). The polyA signal sequence was amplified from pPGKss-Puro by primers Internal-U (XbaI) and L (NotI). The PCR products were isolated on agarose gel and recovered using a Gel Extraction Kit (Qiagen) and used as templates of assembly PCR, in which primers U (PstI) and L (NotI) were used to amplify the fragments. The assembly PCR product was digested by *Pst*I and *Not*I and inserted into pPGKss-puro vector which was treated with the same restriction enzymes. The ligation and transformation processes were the same as described above. The maps were shown in **Figure II-4**.

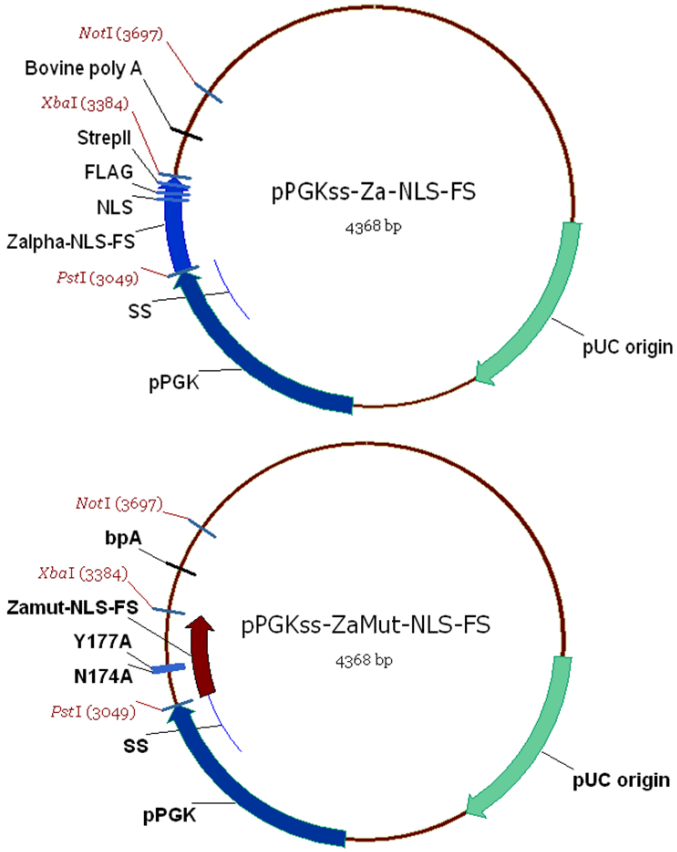


Figure II-4. Plasmid maps of pPGKss-Za-NLS-FS and pPGKss-Zamut-NLS-FS
The eukaryotic expression vectors of Zα and Zα mut were constructed as described in the text.

II.1.2.3.3. Purification of *Zα* and *Zα* mut

Transformation and induction of protein expression in BL21(DE3)

BL21(DE3) competent cells were transformed by prokaryotic expression vectors of Zα or Zα mut. After recovery, 100 µl of bacteria was plated on LB agar plate with 300 µg/ml ampicillin, 34 µg/ml chloramphenicol and 15 µg/ml tetracycline and were incubated at 37 ^{o}C overnight. A colony was inoculated into 50 ml of LB media with antibotics and cultured at 37 ^{o}C overnight with shaking at 225 rpm. Subsequently, 20 ml of culture were inoculated into 1L of LB media with antibiotics and cultured at 37 ^{o}C until OD_{600} reached 0.6. 1 mM (final concentration) IPTG was added to induce the expression of Zα for 3 h. Cells were harvested by centrifugation at 5500 rpm for 15 min at 4 ^{o}C. The pellet was stored at -80 ^{o}C. Expression of Zα mut was induced under the same conditions.

Purification of *Zα* and *Zα* mut with Strep-tactin resin

Purification using Strep-tactin resin was performed following the guidelines from manufacturer (IBA GmbH). The cell pellet was resuspended in 10 ml of cold buffer W (100 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA) and the suspension was sonicated at amplitude 40, pulse 4 s for 20 min (ultrasonic processor VCX 130; Sonics & Materials Inc). The lysate was cleared by centrifugation at 18,000 rpm for 20 min at 4 ^{o}C and the supernatant was collected. The clear lysate was applied to Strep-Tactin resin column (Column Volume (CV) = 1 ml) which was packed as recommended by the manufacturer. After the lysate completely entered the column by gravity flow, the column was

washed 5 times with 1 CV of buffer W. Proteins were eluted with 0.5 CV of buffer E (100 mM Tris, pH 8.0, 150 mM NaCl, 1 mM EDTA, 2.5 mM Desthiobiotin) for 6 times and the eluate was collected in 0.5 CV fractions. Each fraction was checked on 15% SDS-PAGE. The column was regenerated by washing 3 times with 5 CVs Buffer R and stored with 2 ml overlaid buffer R at 4 ^{o}C.

Purification of *Zα* and *Zα* mut by size-exclusion column

The fractions from Strep-Tactin affinity column were checked on SDS-PAGE and the fractions containing Zα or Zα mut were combined and concentrated to 2 ml by Amicon ultra-15 centrifugal filter devices (Millipore). Then the protein solutions were loaded on S-200 (Bio-Rad) size exclusion column and eluted out by buffer W. The fractions were collected and UV absorbance at 280 nm was monitored. The fractions containing Zα or Zα mut were examined on SDS-PAGE and combined for further purification.

Further purification of *Zα* and *Zα* mut by Uno-S ion-exchange column

The fractions from size-exclusion column containing Zα or Zα mut were combined and concentrated to 2.5 ml by Amicon ultra-15 centrifugal filter devices (Millipore). For further purification through ion-exchange column, the buffer of Zα and Zα mut were changed through disposable PD-10 desalting column (Amersham) to HEPES cation chromatography buffer (50 mM HEPES, pH 7.4, 50 mM NaCl, 1 mM DTT, 0.125 mM PMSF) and the protein solutions were loaded onto a cation exchange chromatography column (UNO S1,

Bio-Rad). The proteins were eluted with a 50-1000 mM NaCl gradient.

Dialysis of proteins

Dialysis tubing was prepared as follows. First, dialysis tubings were boiled in 2 L of buffer containing 2% (w/v) sodium bicarbonate and 1 mM EDTA (pH 8.0), and then rinsed thoroughly with distilled water. The tubing was further boiled in distilled water containing 1 mM EDTA (pH 8.0) for 10 min and then stored in dH_2O with 2 mM EDTA at 4 ^oC with tubing submerged. Prior to use, the tubings were washed extensively with dH_2O and sealed with a clamp at one end. Protein sample was put into the tubing with long glass pipette and sealed by a clamp. The samples were dialyzed in 800 ml of dialysis buffer containing 50% glycerol with gentle stirring in 4 ^oC overnight (50 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1 mM EDTA, 50% glycerol). For the proteins used in *in vitro* translation assay, EDTA in the dialysis buffer was omitted and the concentration of NaCl was reduced to 50 mM.

Determination of protein concentration

Protein concentration was determined by Bio-Rad protein assay kit, using BSA as the standard. BSA was prepared in following concentrations: 2 μg/ml, 4 μg/ml, 6 μg/ml, 8 μg/ml and 10 μg/ml. 800 μl of each standard BSA solution was mixed with 200 μl of dye reagent concentrate and vortexed, while in the control tube 800 μl of dH_2O was used. After incubation on bench for 7 min, the spectrophotometer at OD_{595nm} was zeroed with the control sample and the readings of BSA standards were read out and plotted on the Y-axis, against BSA

concentration on X-axis, then the standard curve was made based on these data. The following figure (**Figure II-5**) is one standard curve for protein dialyzed in 50% glycerol buffer.

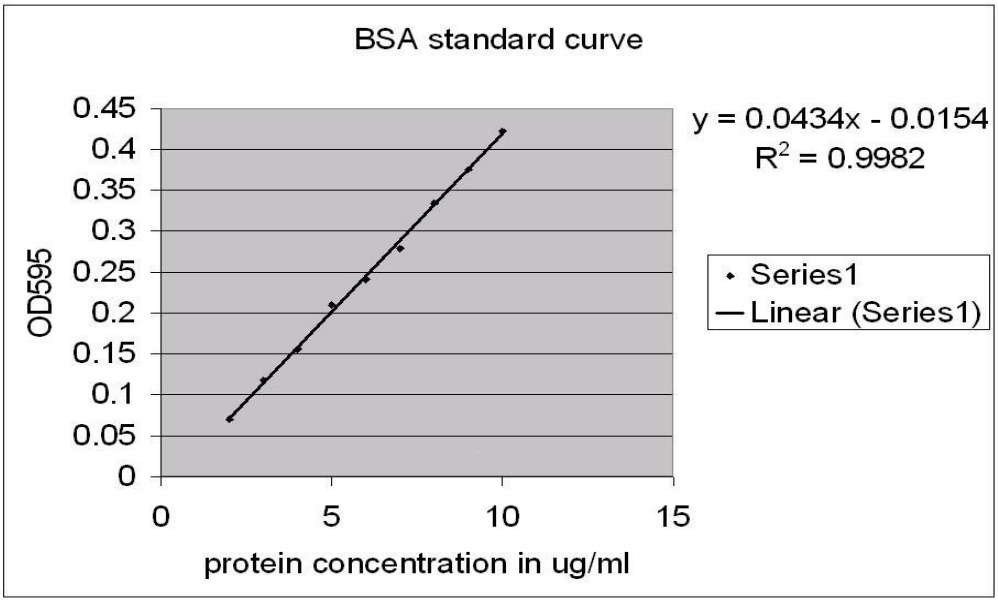


Figure II-5. Standard curve for Bio-Rad protein assay, BSA as the standard protein.

II.1.2.4. Characterization of *in vitro* Z-DNA binding specificity of Zα

II.1.2.4.1. Pull-down assay of Z-DNA containing plasmids

Plasmid pTZ18R and p101, which was derived from plasmid pTZ18R by inserting $(GC)_{16}$ stretch (Droge and Nordheim, 1991), were used as the substrate of Zα. 1 μg of supercoiled pTZ18R or p101 were mixed with 2 μg of purified Zα in 400 μl of buffer containing 20 mM Tris-HCl, pH 8.0, 1 mM EDTA with various concentrations of NaCl (100 mM, 150 mM and 200 mM). After 20 min incubation at room temperature, 15 μl of Strep-Tactin beads was

added to each tube and incubated for 20 min at 4 ^{o}C on a roller. After 3 washs with respective binding buffers, the beads were resuspended in 20 µl of DNA gel-loading buffer containing 1% SDS and boiled for 5 min. Precipitated materials were checked on 0.8% agarose gel.

II.1.2.4.2. *In vitro* specific binding and crosslinking of *Zα* to Z-DNA

Construction of GT-pPGKss-puro

DNA fragments containing $(GT)_{46}$ were amplified from promoter region of mouse mast cell protease-6 (MMCP-6) (Reynolds et al., 1991) by PCR with primers EcoR.1-U and EcoR.1-L. The fragments were digested with *EcoR*I and ligated into pPGKss-puro vector backbone which was treated with *EcoR*I and CIP. The ligation and transformation were the same as described above. The diagram of GT-pPGKss-puro is shown in **Figure II-6**.

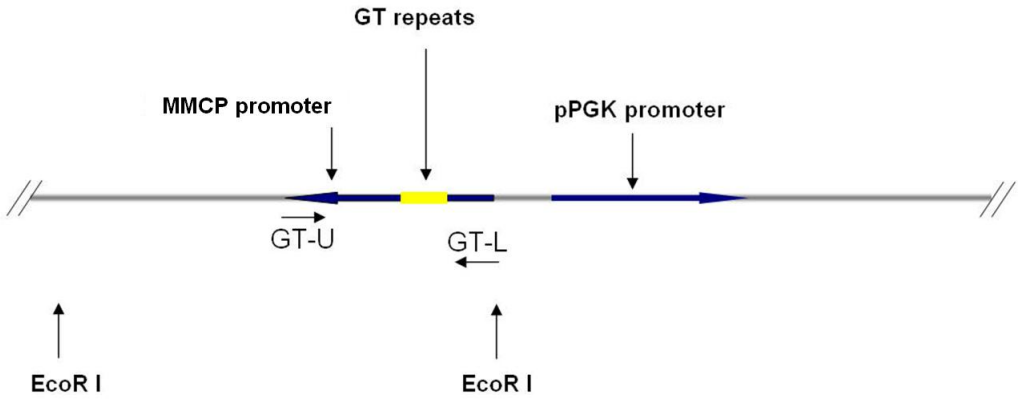


Figure II-6. The diagram of plasmid GT-pPGKss-puro

Binding and crosslinking assay

1.5 μg of supercoiled plasmid GT-pPGKss-puro was mixed with 200 ng of GT fragments which was 360 bp and amplified from plasmid GT-pPGKss-puro by primers GT-U and GT-L; the molar ratio between supercoiled plasmid and GT fragment was about 1:1. The DNA mixture was diluted into 350 μl of HEPES binding buffer (50 mM HEPES-KOH, pH 7.4, 150 mM NaCl, 1 mM EDTA, 1 mM DTT) and incubated with 1 μg of purified Zα, Zα mut or BSA at room temperature for 15 min. After incubation, 0.5% (w/v) formaldehyde was added into each tube and incubated on bench for 5 min to crosslink the DNA-protein complex. Then 125 mM glycine was added into each tube and incubated for 5 min to quench the free formaldehyde. The DNA samples were recovered by ethanol precipitation and digested by *EcoR*I, followed by analysis on 1.4% agarose gel in 0.5 x TBE.

II.1.2.5. Expression and subcellular localization of *Zα* in transfected cells

II.1.2.5.1. Expression of *Zα* assayed by Western blotting

The expression level of Zα in transient transfected cells was tested by Western blotting. A cell lysate from 1 x 10^6 transfected cells was prepared by sonication (ultrasonic processor VCX 130; Sonics & Materials Inc; sonicator probe: 2 mm microtip) at 2 Watts and 3 seconds/pulse for 30 seconds and boiled in 1 x SDS sample loading buffer for 5 min. Equal amounts of proteins were separated on a denaturing 17% (w/v) SDS-PAGE gel and electro-transferred onto a PVDF membrane (Immobilon P, Millipore) in 1 x transfer buffer (25 mM Tris-base,

192 mM glycine) over night at 35 volts at 4 ^{o}C. The PVDF membrane was blocked in TBST (10 mM Tris-HCl, pH 7.4, 150 mM NaCl, 0.05% Tween-20) with 5% milk for 30 min and incubated with M2 mouse monoclonal anti-FLAG antibodies (1:5000 dilution, Sigma) for 90 min. Bound antibodies were detected with anti-mouse horseradish peroxidase-conjugated secondary antibody (1:5000 dilution, Pierce).

II.1.2.5.2. Subcellular localization of *Zα* and *Zα* mut

8 μg of pPGKss-Za-NLS-FS, pPGKss-Zamut-NLS-FS and pCMVss-eGFP were transfected into A549 cells grown on coverslips in a 6-well plate by using Lipofectamine 2000 (Invitrogen), respectively. At 24 h posttransfection, the transfected cells grown on coverslips were washed with PBS, fixed in 4% formaldehyde/PBS for 15 min, washed twice with PBS, and incubated for 10 min in 50 mM ammonium chloride/PBS. Fixed cells were then permeabilized with 0.5% Triton X-100/PBS for 5 min and blocked in 2% FBS (fetal bovine serum)/PBS for 30 min. The cells were incubated for 1.5 hr with FITC-conjugated mouse monoclonal anti-FLAG M2 antibody diluted in PBS (1:100 dilution, Sigma), followed by 4 washs with PBS. In order to stain DNA, 0.5 μg/ml 4'-6-diamidino-2-phenylindole (DAPI; Sigma) was added into PBS and incubated for 1 hr in the dark. After 3 washs with PBS, the coverslips were air-dried and mounted onto glass slides (with cell side downward) with anti-fade oil. Stained cells were analyzed with a Ceiss laser scan confocal microscope (LSM 500 Meta, Zeiss).

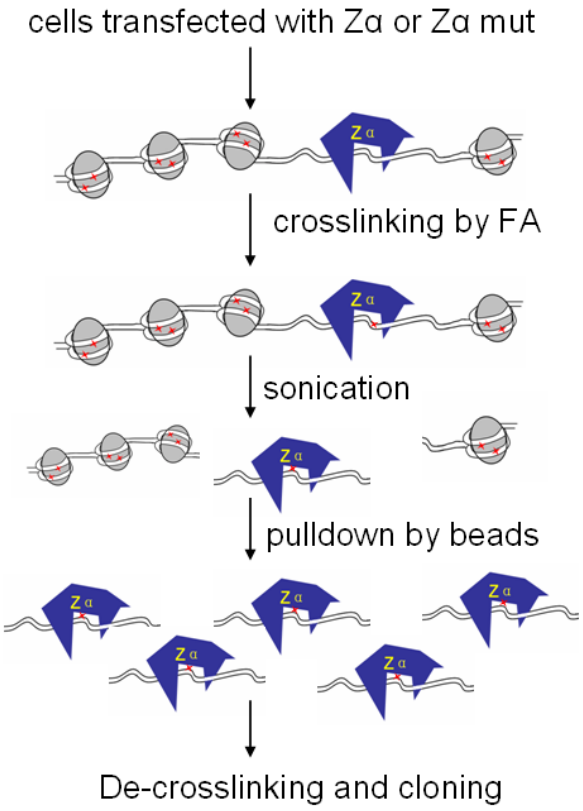


Figure II-7. **The mechanism of *in vivo* chromatin affinity precipitation (ChAP)**.
FA stands for formaldehyde; Red dots stand for formaldehyde crosslinking.

II.1.2.6. *In vivo* chromatin affinity precipitation (ChAP)

A549 cells grown in 175 cm^2 flasks were transfected with pPGKss-Za-NLS-FS or pPGKss-Zamut-NLS-FS by Lipofectamine 2000, as described above. At 24 h posttransfection, the transfected cells were crosslinked with 1% formaldehyde in culture media at 37 ^oC. After 10 min incubation with formaldehyde, 125 mM glycine was added to each flask and incubated for 5 min to stop crosslinking. Then the cells were washed with ice-cold PBS 3 times and scraped down using a rubber policeman and pelleted by centrifugation at 600 x g for 5 min. The cell pellets were resuspended in chromatin precipitation lysis buffer (50 mM

HEPES, pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 140 mM NaCl, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100), and the nuclei were harvested by centrifugation at 600 x g for 10 min and washed once with nuclei washing buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 200 mM NaCl). The nuclei pellets were resuspended in 200 μl of 1% SDS lysis buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 1% SDS) and sonicated 6 times at 20 amplitude, 22 seconds per pulse. Between pulses, the samples were placed in ice-water for 2 min. The sonicated nuclei lysates were cleared by centrifugation at 16,100 x g for 10 min at 4 ^{o}C and the clear supernatants were transferred to 2 ml tubes and diluted with 1.8 ml of dilution buffer (16.7 mM Tris-HCl, pH 8.1, 0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 167 mM NaCl). Then the diluted supernatants were incubated with 40 μl of BSA-blocked Strep-Tactin beads overnight at 4 ^{o}C on a roller. Subsequently, the Strep-Tactin beads were spun down at 1,000 x g for 2 min and washed 4 times with washing buffer (10 mM Tris-HCl, 250 mM LiCl, 1 mM EDTA, 1% Triton X-100, 0.5% Na-DOC, 0.5% NP-40), 10 min for each wash. The precipitated materials were eluted with 100 μl of buffer E (100 mM Tris-HCl, pH 8.0, 150 mM NaCl, 1 mM EDTA, 2.5 mM Desthiobiotin). The concentrations of NaCl in the eluates were adjusted to 200 mM by addition of appropriate amount of 4 M NaCl solution. De-crosslinking was performed by incubating the eluates at 65 ^{o}C for 8 h. After de-crosslinking, the precipitated DNA samples were recovered by ethanol precipitation in the presence of 20 μg of glycogen. The nucleic acids precipitated by Zα, Zα mut and mock control were examined on 12% PAGE gel after radiolabelled with γ^{32}P-ATP by T4 polynucleotide kinase (NEB). In order

to characterize whether the precipitated material is RNA or DNA, RNase sensitivity assays were performed by incubating the radio-labeled precipitated materials with 20 units of DNase-free RNase (Roche) at 37 ^{o}C for 30 min. After RNase treatment, the pull-down materials were analyzed on 12% PAGE.

II.1.2.7. *In vitro* chromatin affinity precipitation (ChAP)

The process of *in vitro* chromatin affinity precipitation is shown in **Figure II-8**.

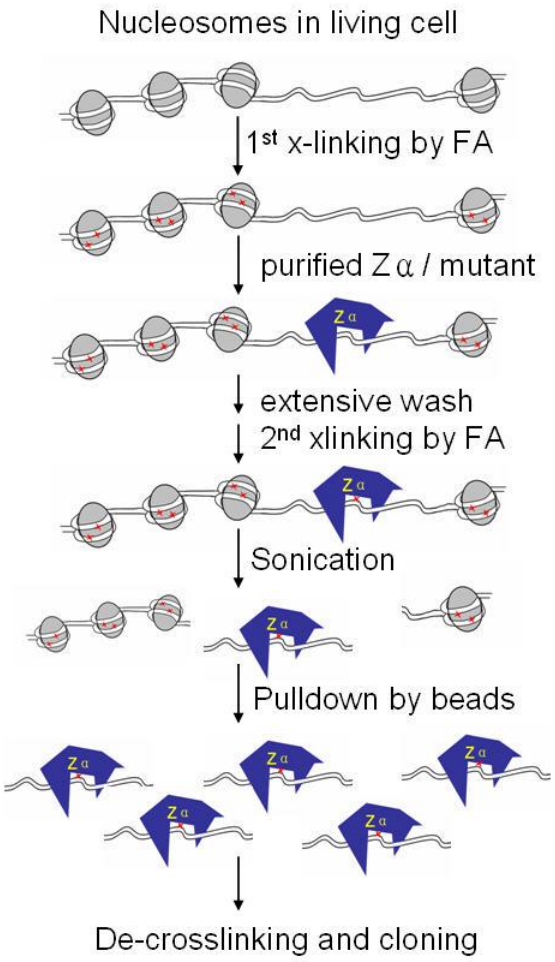


Figure II-8. The proposed mechanism of *in vitro* ChAP.

FA stands for formaldehyde. The experimental process was described in text.

A549 cells were passaged into three 75 cm^2 flasks one day before the experiments. The cells were crosslinked with 1% formaldehyde at 37 ^oC for 5 min (the 1st crosslinking), followed by addition of 125 mM glycine to the media and incubation on shaker at room temperature for 5 min. Then the cells were washed 5 times with cold PBS, followed by incubation with PBS containing 1% Triton X-100 for 2.5 min at room temperature. After permeabilization with Triton X-100, the cells were washed twice with cold PBS and twice with cold HEPES binding buffer (50 mM HEPES-KOH, pH 8.0, 150 mM NaCl, 1 mM EDTA). 30 μg of purified Zα, Zα mut and BSA were diluted into 5 ml of HEPES binding buffer and added into the flasks, respectively. After incubation at 4 ^oC for 5 h, the cells were washed with cold HEPES binding buffer for 4 times at 4 ^oC, 15 min for each wash. The 2nd crosslinking was performed by adding 0.5% (w/v) formaldehyde in HEPES binding buffer. After incubation on bench for 5 min, the 2nd crosslinking was stopped by addition of 125 mM glycine and incubation on shaker at room temperature for 5 min. The cells in each flask were washed with cold PBS for 5 times and scraped down with a rubber policeman in 4 ml of cold ChAP lysis buffer (50 mM HEPES-KOH, pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 140 mM NaCl, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100). The cells were pelleted by centrifugation at 1,500 x g for 5 min at 4 ^oC and were resuspended in nuclei wash buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 200 mM NaCl) and spun down again to harvest the nuclei. After resuspended in 200 μl of 1% SDS lysis buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 1% SDS), the

nuclei pellets were sonicated for 6 times, and each sonication pulse lasted for 22 seconds with amplitude 20. After each pulse, the sample was put in ice-water for 2 min. The cell lysates were cleared by centrifuge at 16,100 x g for 10 min at 4 ^{o}C and the clear supernatants were transferred to new tubes and diluted with ChAP dilution buffer (16.7 mM Tris-HCl, pH 8.1, 0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 167 mM NaCl) to 2 ml. Then 30 µl of BSA-blocked Strep-Tactin beads were added to each sample and incubated on a roller at 4 ^{o}C overnight. After spun down at 1,000 rpm for 2 min at 4 ^{o}C, the beads were washed 5 times with cold RIPA buffer containing 1.5 M urea (50 mM Tris-HCl, pH 7.5, 1% NP-40, 1% sodium DOC, 0.1% SDS, 1 mM EDTA, 1 M NaCl, 1.5 M urea, 0.2 mM PMSF), 15 min for each wash, followed by 3 washs with cold TBS (20 mM Tris-HCl, pH 7.4, 150 mM NaCl, 1mM EDTA), 10 min for each. After eluted with elution buffer E (100 mM Tris, pH 8.0, 150 mM NaCl, 1 mM EDTA, 2.5 mM dethiobiotin), the concentrations of NaCl in eluates were adjusted to 200 mM by addition of appropriate amount of 4 M NaCl. After de-crosslinking by incubation at 65 ^{o}C for 8 h, the precipitated DNA samples were cleaned by phenol/chloroform extraction and recovered by ethanol precipitation in the presence of 20 µg of glycogen. The ChAP DNA samples were resuspended in 50 µl of TE (pH 8.0) and stored at -20 ^{o}C.

II.1.2.8. Construction of the Z-DNA library

The cloning process of *in vitro* ChAP DNA fragments is depicted in **Figure II-9**.

II.1.2.8.1. T4 DNA polymerase treatment

After de-crosslinking, the precipitated DNA fragments were blunt-ended by T4 DNA polymerase treatment. The reactions were composed of 10 units of T4 DNA polymerase (NEB), 1 µl of 100 µg/ml BSA, 1 µl of 10 mM dNTP mixture, 50 µl of ChAP DNA sample, 10 µl of buffer 2 (NEB), and 35 µl of dH_2O and occurred at 12 ^oC for 20 min. Then the reactions were stopped by addition of 1 µl of 0.5 M EDTA (pH 8.0). DNA samples were recovered by ethanol precipitation.

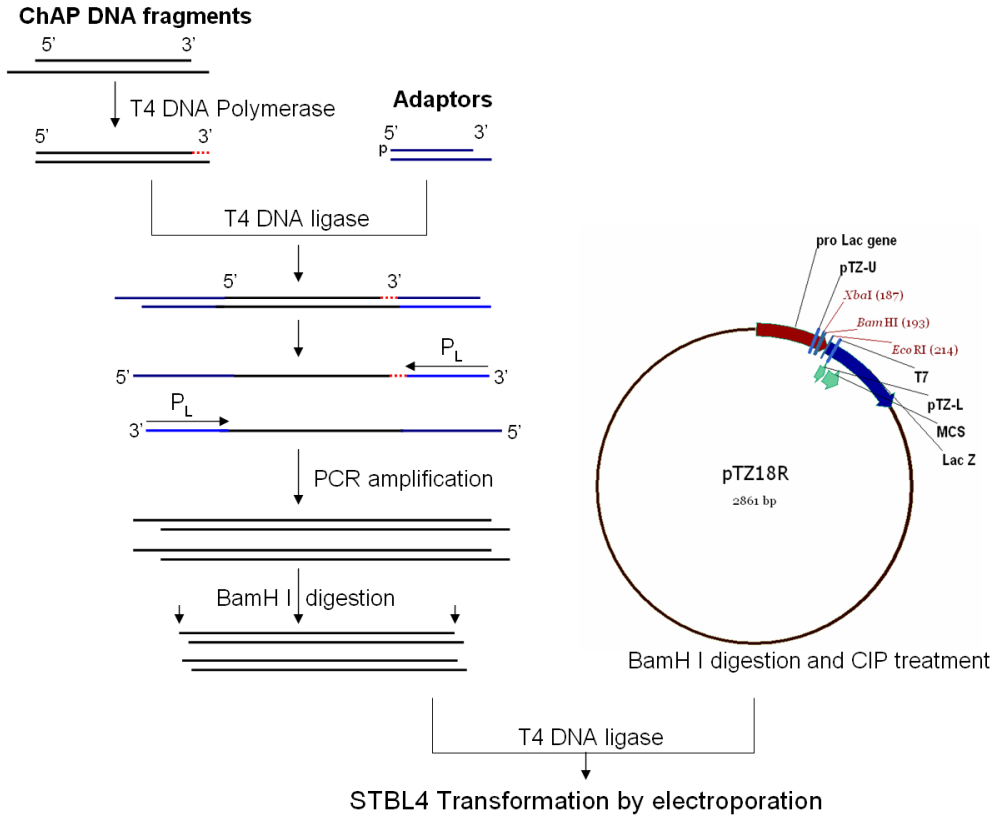


Figure II-9. Depiction of cloning process of *in vitro* ChAP DNA fragments.

II.1.2.8.2. Annealing of adaptor

The sequences of adaptor are shown in **Figure II-10**. These two oligos were

mixed in equal molar amount in annealing buffer (10 mM Tris-HCl, pH 8.0, 1 mM EDTA, 100 mM NaCl) in a 1.5 ml centrifuge tube and were put in 1L of boiling water until the water cooled down to room temperature on bench. The concentration of the annealed adaptor was 50 mM.

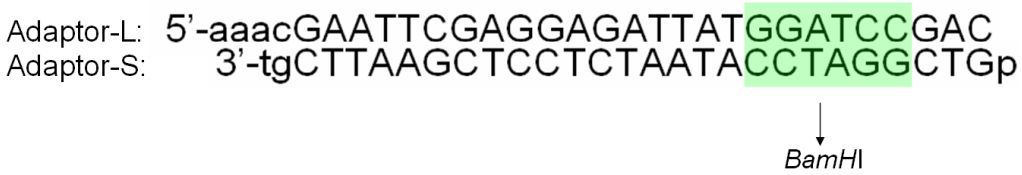


Figure II-10. The Schema of adaptor.

The recognition site of *Bam*HI was labeled. One end of the adaptor is blunt-ended, with a 5' phosphate group at the Adaptor-S; another end is stiky.

II.1.2.8.3. Ligation of ChAP DNA fragments with adaptors

The ligation reactions were composed of 30 µl of ChAP DNA, 0.5 µl of 1 mM adaptor, 4 µl of T4 DNA ligase buffer, 2 µl of T4 DNA ligase and 3.5 µl of dH_2O and were incubated at 16 ^oC overnight. After ligation, the DNA samples were recovered by ethanol precipitation.

II.1.2.8.4. Amplification of ChAP DNA fragments

After ligation, DNA samples were amplified by Adaptor-L with 25 PCR cycles. PCR conditions were: 95 ^oC for 5 min; 95 ^oC for 30 seconds, 65 ^oC for 30 seconds, 72 ^oC for 2 min; 72 ^oC for 7 min; The PCR products were separated on 1.5% agarose gel in 0.5 x TBE and recovered by using gel extraction kit (Qiagen). Then the recovered PCR products were digested by *Bam*HI and separated again on 1.5% agarose gel. After extracted from the gel, the DNA fragments were ligated into pTZ18R vector which was digested by *Bam*HI and

treated with CIP. The ligation conditions were the same as described above.

II.1.2.8.5. Transformation of Stbl4

The ligation products were purified and electroporated into Stbl4 competent cells. After electroporation, Stbl4 cells were inoculated into 2 ml of SOC media without antibiotics and recovered by incubation at 30 ^{o}C for 90 min with shaking; whereafter 200 µl of SOC culture medium was plated on LB agar plates with 200 µg/ml ampicillin and 50 µg/ml X-gal and incubated over night in 30 ^{o}C incubator. The remained 1.8 ml transformed Stbl4 cells were mixed with 1.8 ml of fresh SOC medium containing 40% glycerol and stored in -80 ^{o}C freezer.

II.1.2.8.6. Quality of Z-DNA library tested by colony PCR

The Stbl4 transformants on LB plates were incubated in 30 ^{o}C incubator until the blue colonies were recognized clearly. Then colony PCR was used to check the lengths of inserts. 24 white colonies were inoculated and resuspended in 30 µl of dH_2O; 10 µl of the suspension was inoculated into 5 ml of LB media containing 200 µg/ml ampicillin. The rest 20 µl was boiled at 100 ^{o}C for 5 min and cleared by spin at 16,100 x g for 1 min. 1 µl of the boiled supernatant was used as template of colony PCR and amplified with primers pTZ-U and pTZ-L. The inserts longer than 16 bp (which is from adaptor dimer) were from the ChAP precipitated DNA fragments.

II.1.2.9. Large scale sequencing and data mining

II.1.2.9.1. Large scale sequencing

The constructed Z-DNA library was sequenced by Beijing Genomics Institute. Briefly, the library was plated on LB plates with 200 μg/ml ampicillin and 50 μg/ml X-gal. 10,500 white colonies were inoculated into 96-well plates in LB media with 200 μg/ml ampicillin. Mini-prepped plasmids were amplified by primer pTZ-L in GeneAmp PCR System 9700 and the sequences were read out on ABI Prism 3730xl DNA Analyzer.

II.1.2.9.2. Data mining and analyses

Mapping the sequences on the human genome build 36.2

Initially, sequences of vector and adaptor were removed from original sequencing reads by multiple alignments with BLASTN against vector sequence and adaptor sequence. The insert sequences between adaptor regions or vector sequences were isolated. Totally, 10,563 fragments were isolated and aligned to the reference genome (the Human Genome Build 36.2 provided by NCBI) by using BLASTN or Blat (http://genome.ucsc.edu). 7,321 reads were able to give alignments with corresponding human genome fragments which were equal or longer than 30 bps with the sequence identity >= 90%, and no more than 1% gap.

Definition of Z-DNA forming hotspots

Some fragments were mapped closely on the human genome. Z-DNA forming hotspots were defined as the minimum genome regions covering more than one ChAP sequences which were overlapped or less than 100 bp apart. When a

sequencing fragment was aligned against the reference genome, the positions of the first and the last nucleotides on the reference genome were recorded. After all the sequencing fragments were mapped, the numbers were sorted and compared and hotspots were marked according to its definition.

Refinement of hotspots with high-confidence

Some hotspots marked based on its definition were redundant or dubious. So the hotspots were classified and refined. The criteria of refinements of hotspots were described in detail under Results. After selection, the hotspots on Giemsa-stained chromosomes were mapped based on the mapping results of the UCSC genome browser.

Comparison of Z-DNA forming hotspots with SNPs densities

Correlation of Z-DNA forming hotspots and the occurrences of Single Nucleotide Polymorphism (SNP) collected in NCBI dbSNP build 128 were checked. RefSNPs in hotspot regions and in the flanking regions at both sides (500 bp, 1000 bp, and 2000 bp) were counted and generalized as SNP density (number of RefSNPs/bp). The comparisons of the densities of RefSNPs in the hotspots and the corresponding flanking regions were analyzed statistically by pair-wised Wilcoxon test. *p* values were calculated and listed.

II.1.2.10. ChAP-PCR

II.1.2.10.1. Comparison between ChAP materials of Zα and Zα mut

16 hotspots consisting of at least 4 ChAP sequences were subjected for further

analysis using ChAP-PCR. PCR primers on the hotspots were designed by Primer3 (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi) or Vector NTI Suite® 8 (Invitrogen). A new batch ChAP DNA was used as ChAP-PCR templates and each PCR was optimized individually. The sequences of primers and PCR parameters were listed in **Table II-1**. PCR products from Zα, Zα mut and BSA ChAP samples were checked on agarose gel and visualized by EtBr staining. PCR using genomic DNA as the template were used as the control to check the quality of primers.

HOTSPOT5:	10% DMSO, 1.8 mM MgCl2, Ta = 61 °C, 30 cycles;
Sense Primer:	5’ CTGCACGTAGATATTTTGACCACTTAGA 3’
Antisense Primer:	5’ CGAAGGCCACAAGATGTCAGA 3’
HOTSPOT7:	10% DMSO, 1.8 mM MgCl2, Ta = 62 °C, 30 cycles;
Sense Primer:	5’ GTGCTGGTTCCAAGCCTGAGT 3’
Antisense Primer:	5’ GGAAGAGCTTGGGGAAGGGA 3’
HOTSPOT17:	10% DMSO, 2 mM MgCl2, Ta = 62 °C, 45 cycles;
Sense Primer:	5’ TGCGCCCCCAACTAACAGTGTT 3’
Antisense Primer:	5’ GGCCTAATATCCACTTGCATATTCCCCA 3’
HOTSPOT19:	10% DMSO, 2 mM MgCl2, Ta = 61 °C, 45 cycles;
Sense Primer:	5’ ACACGCCTTTTGTCATATCTGGAAGTTG 3’
Antisense Primer:	5’ CGCTCCAAATGTCCACATACAGA 3’
HOTSPOT21:	10% DMSO, 2 mM MgCl2, Ta = 58 °C, 35 cycles;
Sense primer:	5’ TCATATCTGGAAGTTGTCCATTT 3’
Antisense primer:	5’ GCGCTCCAAATGTCCACATA 3’
HOTSPOT65:	10% DMSO, 2 mM MgCl2, Ta = 58 °C, 45 cycles;
Sense Primer:	5’ TGATGCCAACAGTAGAAAGGGA 3’
Antisense Primer:	5’ AGGCCTCAATGCGGTCCATA 3’
HOTSPOT85:	10% DMSO, 2 mM MgCl2, Ta = 61 °C, 45 cycles;
Sense Primer:	5’ CAGATGGAAATGAATGGAATCATCACT 3’
Antisense Primer:	5’ CCATTCGAGTTCATTGACTATTCCATT 3’
HOTSPOT86:	10% DMSO, 2 mM MgCl2, Ta = 58 °C, 35 cycles;

Sense primer:	5’ TCAAAAATAACCATCATCAGTTGG 3’
Antisense primer:	5’ TCATTCCATTCCATTAGAGGATTC 3’
HOTSPOT87:	10% DMSO, 2 mM MgCl2, Ta = 58 °C, 35 cycles;
Sense primer:	5’ GAATAATCCATGGTCCCGAAT 3’
Antisense primer:	5’ CTCCTTTCACTTCCATTCGAT 3’
HOTSPOT88:	10% DMSO, 2 mM MgCl2, Ta = 55 °C, 35 cycles;
Sense Primer:	5’ TTTCAAGAGATTTGAGACCAATGGT 3’
Antisense Primer:	5’ GCCAACCCAAAGAAGTTACTAAGA 3’
HOTSPOT89:	10% DMSO, 2 mM MgCl2, Ta = 58 °C, 35 cycles;
Sense primer:	5’ AACACTCTTTTTTGGAAAT 3’
Antisense primer:	5’ AAATTCCACAAAGAGTGTCT 3’
HOTSPOT131:	10% DMSO, 2 mM MgCl2, Ta = 58 °C, 45 cycles;
Sense Primer:	5’ GAATTCTTAGTAACTTCTTCGTGTTGTG 3’
Antisense Primer:	5’ TTTCCTTTTCTGCCATTGGC 3’
HOTSPOT139:	10% DMSO, 2 mM MgCl2, Ta = 61 °C, 45 cycles;
Sense primer:	5’ AGCACTTTGGAAACTCTCCTTGT 3’
Antisense Primer:	5’ TATCAAAGAAAGGTTCAACACTGTG 3’
HOTSPOT150:	10% DMSO, 2 mM MgCl2, Ta = 58 °C, 35 cycles;
Sense primer:	5’ TTTTCCCGTTTCCAATGA 3’
Antisense primer:	5’ CGCTTTGATGCCTACAGTGA 3
HOTSPOT158:	10% DMSO, 1.8 mM MgCl2, Ta = 62 °C, 35 cycles;
Sense Primer:	5’ GTGGTGCATGGCCTTAGTTGGTG 3’
Antisense Primer:	5’ CACGCTGAGCCAGTCAGTGT 3’
HOTSPOT183:	10% DMSO, 2 mM MgCl2, Ta = 58 °C, 35 cycles;
Sense primer:	5’ CCGGACATGGACAGGATTGA 3’
Antisense primer:	5’ CACACGCTGAGCCAGTCACT 3’

Table II-1. PCR conditions and primers used in the first round ChAP-PCR.

II.1.2.10.2. Comparison between hotspot regions and non-hotspot regions

In order to confirm whether the hotspot regions were pulled down specifically by Zα during *in vitro* ChAP, 7 hotspots and 7 non-hotspots regions were

randomly picked and analyzed. 7 pairs of primers priming at hotspot regions or at non-hotspot regions were designed (**Table II-2**). Totally, 14 PCRs using Zα ChAP DNA as the templates were performed. Parameters of PCR are 2 mM $MgCl_2$, Ta=60 ^oC and 35 cycles. To standardize PCR products from different pairs of primers, control PCRs were performed at the same PCR conditions and all control PCRs were amplified from the same amount of genomic DNA. All the PCRs were performed in triplicate. PCR products were analyzed on agarose gel with EtBr staining and quantified with Bio-Rad Quantity One software. After measurement, the ratios of signals between PCR products from Zα ChAP DNA and from genomic DNA were calculated. 7 ratios from PCRs in hotspot regions and 7 ratios from PCRs in non-hotspot regions were sorted with increasing values and plotted and compared.

HS9-S:	5'- CACAGCTGTATGTGGAAGGAAGTGA-3'
HS9-AS:	5'-TCCAAGGCAGCCAGGAAGAA-3'
HS34-S:	5'-CCACTCTGCTCTTCCTCCTTCTCTGT-3'
HS34-AS:	5'-ACGCGGCAAGATGGCTGACT-3'
HS41-S:	5'-TGTACAGCTCAGGTTGTGGCT-3'
HS41-AS:	5'-GGATGGTGGCACTCTTCTCA-3'
HS49-S:	5'-AGAAAGCATGTCTGAGGTAGTAGAGG-3'
HS49-AS:	5'-TTGCCTTTTCCCACTGGACT-3'
HS52-S:	5'-GAAGACCAGGCACCTGAGGACT-3'
HS52-AS:	5'-AGGAGCTGGGCACTCGGTTA-3'
HS63-S:	5'-GAGCGAGCATTGCTTCCTGA-3'
HS63-AS:	5'-GCAGAGATCACCAGCCTTTTTG-3'
HS156-S:	5'-CTCAGTAACTTCCTTTTGTTGTGTGTAT-3'
HS156-AS:	5'-TTTCCTTTTCTGCCTTTGGC-3'

Non-HS1-S:	5’-TGAATGTGATTTAAAGTCACCTGCT-3’
Non-HS1-AS:	5’-CAATCTTGTGGAAGGCTGATGA-3’
Non-HS2-S:	5’-CTTTCTTTTTTGACAGCTTGTCACCTTA-3’
Non-HS2-AS:	5’-CTCCACATAAAAGAAATCTGAACTCCAA-3’
Non-HS3-S:	5’-CAGCCAGGGCAACTAAGTGA-3’
Non-HS3-AS:	5’-TTCCCCTCACTCTCACTCTCTCT-3’
Non-HS4-S:	5’-TGCAGCCATGTCTGTCCTTG-3’
Non-HS4-AS:	5’-GAGCAGCAATGCCAGGATTC-3’
Non-HS5-S:	5’-TTCCCATTTTTTCATTAGAGCCTTG-3’
Non-HS5-AS:	5’-TGGCGTATGCCTCTGGTCCT-3’
Non-HS6-S:	5’-GGAACCTATGTCCTTTTATCAGTTATG-3’
Non-HS6-AS:	5’-CACAAATAGAAAAGTCCAATCAAGC-3’
Non-HS7-S:	5’-GCTGAATGGGAGGAGTCCTGTA-3’
Non-HS7-AS:	5’-CCGGCCACCAGTGCTATTAT-3’

Table II-2. Sequences of primers used in the second round ChAP-PCR.

II.2. Detection of natural Z-RNA in mammalian cells using Zα as a probe

II.2.1. Materials

II.2.1.1. Reagents and kits

Treated RRL	Promega
Untreated RRL	Promega
RNase A	Roche
DEPC	Sigma
RNasin	Promega
Heparin	Sigma

II.2.1.2. Plasmids

pPGKss-Za-FS wo NLS	This study
pPGKss-Zamut-FS wo NLS	This study
pPGKss-FS	This study
pCMVss-Za-eGFP	This study
pCMVss-eGFP	Lab stock

II.2.1.3. Primers

D-NLS-L:	5'TCGTCGTCATCCTTGTAATCACCGCCGCTTCCACCTCCG3'
D-NLS-U:	5'CGGAGGTGGAAGCGGCGGTGATTACAAGGATGACGACGA3'
U-FS:	5'CCAACTGCAGATGGGTGGAAGCGGCGGTCCTAAG3'
L(GFP)	5'AGCTCCTCGCCCTTGCTCACCATGGAACCGCCGCTTCCACCTCCG3'
U(GFP)	5'CATG GTGAGCAAGG GCGAGGAGCT3'

II.2.2. Methods

II.2.2.1. Construction of plasmids

pPGKss-Za-FS wo NLS and pPGKss-Zamut-FS wo NLS

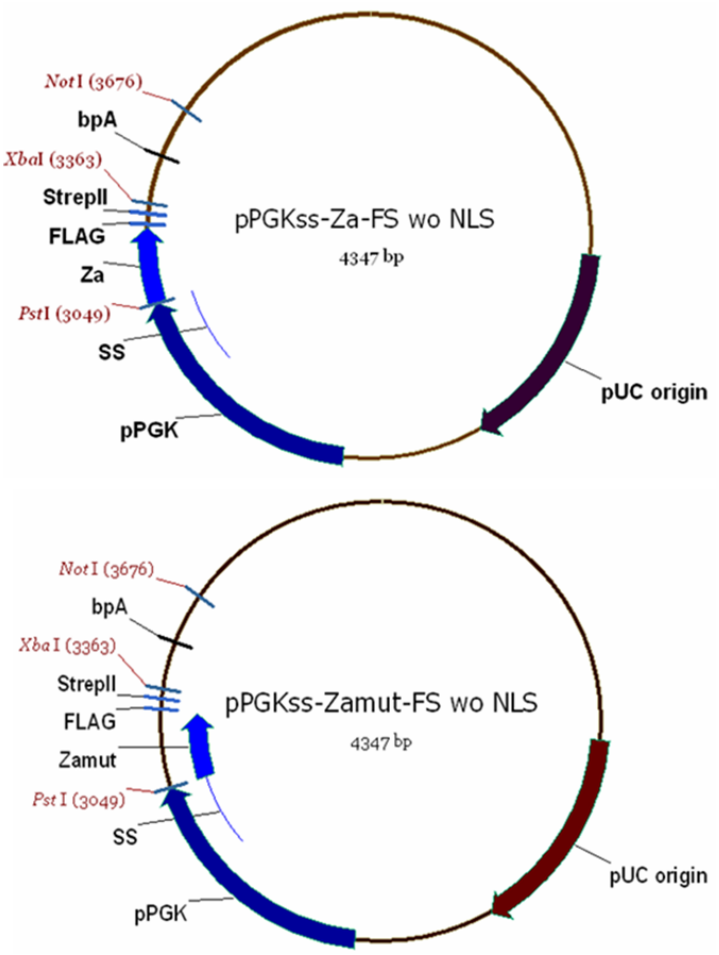


Figure II-11. Plasmid map of pPGKss-Za-FS wo NLS and pPGKss-Zamut-FS wo NLS. NLS was removed from pPGKss-Za-NLS-FS and pPGKss-Zamut-NLS-FS by assembly PCR. The cloning process was described in the text.

Eukaryotic expression vectors of Zα/mut probe without a NLS in the coding region were constructed (**Figure II-11**). The deletion of the NLS was done by assembly PCR. DNA fragments coding Zα/mut probe were amplified from the corresponding expression vectors by U(Pst I) and D-NLS-L; the lower fragments coding FLAG and StrepII tags and poly(A) signal were amplified by

D-NLS-U and L(Not I). The two fragments were gel-extracted and used as the templates of assembly PCR. The enzymatic digestions and ligation reactions were performed as described for pPGKss-Za-NLS-FS.

pPGKss-NLS-FS

The control vector carrying the NLS and epitope tags was constructed (**Figure II-12**). The coding sequence was amplified from pPGKss-Za-NLS-FS by U-FS and L(NotI) and digested by *Pst*I and *Not*I, followed by ligation with T4 DNA ligase into vector backbone pPGKss-Za-NLS-FS which was treated with the same enzymes and CIP.

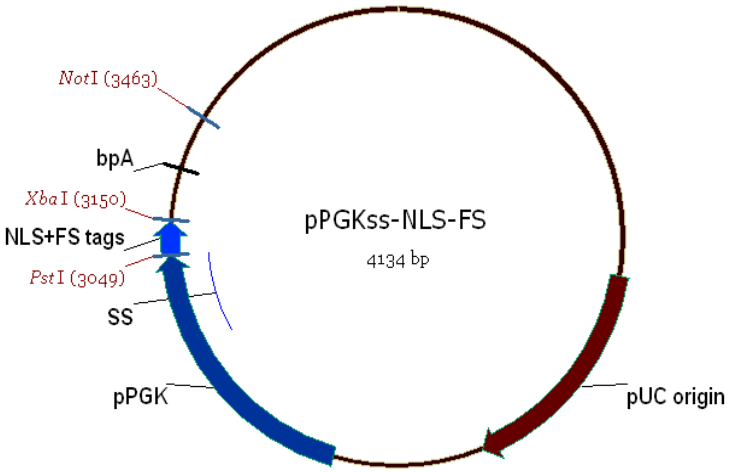


Figure II-12. Plasmid map of pPGKss-NLS-FS

pCMVss-Za-eGFP

The Za coding sequence was amplified from plasmid pET28-Za77 with primer U(PstI) and L(GFP). eGFP coding DNA sequence was amplified from plasmid pCMVss-eGFP with primer U(GFP) and L(NotI). After separation on an agarose gel and gel extraction, two DNA fragments were mixed and used as the

template for assembly PCR using primers U(PstI) and L(NotI). The assembled PCR product was gel-extracted and digested with *Pst*I and *Not*I and inserted into vector backbone pCMVss-eGFP which was treated with the same restriction enzymes and CIP.

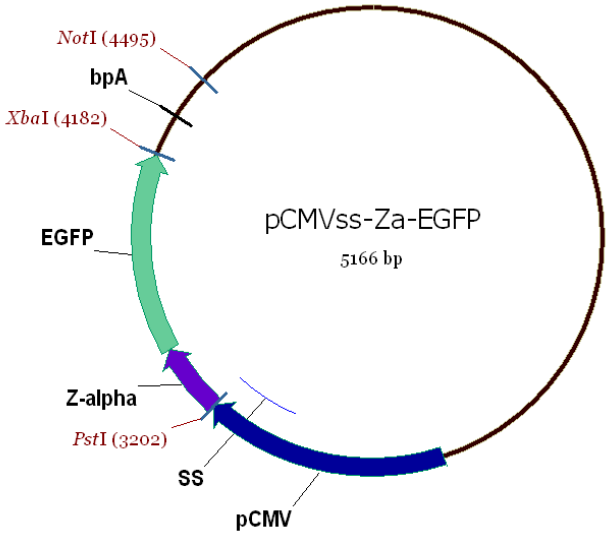


Figure II-13. Plasmid map of pCMVss-Za-eGFP

II.2.2.2. Subcellular localization of *Zα* without NLS

Hela cells were cultured routinely and passaged on coverslips in 6-well plates before transfection. Then Hela cells were transfected with pPGKss-Za-FS wo NLS and pPGKss-Zamut-FS wo NLS, respectively. At 24 h posttransfection, localization of Zα and Zα mut without NLS were analyzed by immunostaining with FITC-conjugated anti-FLAG antibody M2, as previously described protocol for NLS containing constructs.

II.2.2.3. Pull-down assays

II.2.2.3.1. Pull-down assay with untreated rabbit reticulocyte lysate
(This part was contributed by Miss Feng Shu)

In this pull-down experiment, human genomic B-DNA fragments were used as non-specific competitors. Firstly, 0, 2, 6, 20 μg of B-DNA fragments were mixed with 10 μg of purified Zα or Zα mut in 200 μl of binding buffer (10 mM Tris-HCl, pH 7.4, 100 mM KCl, 1 mM MgCl2, 1 mM DTT) and incubated on ice for 20 min. After pre-incubation with B-DNA competitors, 30 μl of untreated RRL was added to each tube and incubated for 45 min on ice. 5 μl of BSA-blocked Strep-Tactin beads was added into each tube and incubated for another 45 min on a roller at 4 ^{o}C. The beads were washed 3 times with binding buffer, 10 min for each, and eluted by elution buffer E containing Desthiobiotin. After boiled with SDS loading buffer containing 10 mM EDTA, 10 μl of eluate of each sample was analyzed on an agarose gel visualized by EtBr staining. 1 μl of eluate was separated on SDS-PAGE and analyzed by a Western blotting assay to detect the presence of Zα or Zα mut.

II.2.2.3.2. Pull-down assay with separated ribosomal subunits

Purification of ribosomal 40S and 60S subunits
(This part was done with assistance from Miss Feng Shu)

Ribosomal subunits were purified from Hela cells as described (Blobel and Sabatini, 1971; Jan and Sarnow, 2002). Hela cells were cultured routinely on 175 cm^2 flasks and collected by trypsin treatment. Hela cells (about 10^7 cells) pellet was resuspended in 1 ml of ice-cold 0.25 M sucrose-TMK buffer (50 mM Tris-HCl, pH 7.5 at 25 ^{o}C, 25 mM KCl, 5 mM $MgCl_2$, 1 mM DTT) with 200 units of RNasin (Promega). All subsequent operations were performed on ice or at 4 ^{o}C. The resuspension of cells was homogenized with a motor-driven

Heidolph DIAX 900 homogenizer. The settings of homogenizer were Level 5, 30 seconds per pulse, 3 pulses in total. A postmitochondrial supernatant fraction was prepared by centrifugation at 16,100 x g for 20 min. Then the supernatant was layered on equal volume of 50% (w/v) sucrose-TMK cushion and centrifuged in a Beckman SW41 rotor at 200,000 x g for 24 h at 4 ^{o}C to yield a pellet of free polysomes, which was rinsed once with DEPC-treated water and resuspended in buffer B (20 mM Tris-HCl, pH 7.5, 6 mM $Mg(OAc)_2$, 150 mM KCl, 6.8% sucrose, 1 mM DTT, 200 u/ml RNasin). OD_{260} and OD_{280} were measured and the concentration of polysome solution was adjusted to 50 OD_{260}/ml with buffer B. To 0.2 ml of this suspension, 25 µl of 0.02 M puromycin and 45 µl of compensating buffer (20 mM Tris-HCl, pH 7.5, 6 mM $Mg(OAc)_2$, 3000 mM KCl) were added. The final concentrations were 20 mM Tris-HCl, pH 7.5, 6 mM $Mg(OAc)_2$, 500 mM KCl. The suspension was put on ice for 15 min, followed by incubation in 37 ^{o}C water for 10 min and then an additional 5 min incubation on ice. The dissociated ribosomes were then separated over a 10-30% (w/w) sucrose gradient by centrifuge in a Beckman SW41 rotor at 200,000 x g for 3.5 h at 4 ^{o}C. The separated 40S and 60S subunits were fractionated drop-wisely from the bottom to the top of the centrifuge tube and detected on an agarose gel by EtBr staining. The 40S and 60S subunits were concentrated using Amicon ultra-15 centrifugal filter devices (Millipore) in buffer C (20 mM Tris-HCl, pH 7.5, 0.2 mM EDTA, 10 mM KCl, 1 mM $MgCl_2$, 6.8% sucrose) and stored at -80 ^{o}C. The concentrations of 40S and 60S subunits were determined by spectrophotometry, using the conversions 1 $A_{260\ nm}$= 50 nM for 40S and 1 $A_{260\ nm}$ =25 nM for 60S subunits.

Pull-down experiments

40 µl of purified 40S and 60S subunits were thawed on ice slowly and diluted into 200 µl of binding buffer (20 mM Tris-HCl pH 7.5, 150 mM KCl, 6 mM MgCl2, 1 mM DTT) and mixed with 8 µg of purified Zα and incubated on ice for 40 min, while 8 µg of BSA was added in the control tube. Then 15 µl of BSA-blocked Strep-Tactin beads was added into each tube and incubated on a roller at 4 ^{o}C for 40 min. The beads were washed with binding buffer for 3 times, 10 min for each wash, and then were eluted with elution buffer E. The eluates were analyzed on an agarose gel and visualized by EtBr staining.

II.2.2.4. Effect of *Zα* on *in vitro* mammalian translation

Initially, RNase contamination in purified Zα and Zα mut was checked by incubation with purified RNA and no obvious RNase activity was observed. These purified proteins and Flexi rabbit reticulocyte lysate (RRL) *in vitro* translation system (Promega) was employed. First, master mix was prepared as indicated in Table II-3. Then 60 µl of a master mix was aliquoted into 3 tubes, labeled as C1 (for control), Z1 (for Zα) and M1 (for Zα mut). Then 12 µl of protein dialysis buffer was added into the remained master mix, which was then divided into 12 tubes, 31 µl per tube and labeled as C2-5, Z2-5 and M2-5, respectively. 2 µl of protein dialysis buffer, 2 µl of Zα (4 µg/µl) and 1 µl of mutant (8 µg/µl) + 1 µl dialysis buffer were added into C1, Z1 and M1, respectively. After gentle mixing, 31 µl from C1, Z1 and M1 was transferred into C2, Z2 and M2, respectively. And then C2, Z2 and M2 were serially diluted into the following tubes. All the reactions were incubated in a 30 ^{o}C incubator

for 90 min. After incubation, the reactions were put on ice to stop the translation. The amount of luciferase in each tube was quantified as arbitrary Light Unit in a Turner Designs TD 20/20 luminometer after reaction with 40 µl of luciferase substrate.

treated RRL	500 µl
amino acids - M	6 µl
amino acids - L	6 µl
KCl	15.2 µl
Rnasin	12 µl
Luciferase mRNA	8 µl

Table II-3. Preparation of master mix for *in vitro* translation reaction.

II.2.2.5. Cell proliferation assay

Hela cells grown on 6-cm dishes at 95% confluency were transfected by Lipofectamine 2000 (Invitrogen) with plasmids pPGKss-Za-NLS-FS, pPGKss-Za-FS wo NLS, pPGKss-Zamut-NLS-FS, pPGKss-Zamut-FS wo NLS and pPGKss-NLS-FS, individually. At 6 h posttransfection, the transfected cells were trypsinized and equal number of cells was aliquoted into 96-well plates. Each transfection had five replicates per time-point and was monitored at 6 h, 24 h, 48 h, 72 h and 96 h posttransfection. To quantify the cell number, 300 µl of WST-1 reagent (Roche) was mixed with 3 ml fresh DMEM media and 110 µl of the mixture was added to each well. WST-1, a tetrozolium salt, is cleaved by mitochondrial dehydrogenases in metabolically active cells to formazan, which

is quantified by measuring the absorbance at 440 nm via a scanning multiwell spectrophotometer (ELISA reader). After 30 min incubation in a cell incubator (37 ^{o}C, 5% CO_2), the absorbance of samples at 440 nm against a background control as blank (600 nm) were recorded (Bio-Rad Benchmark Plus microplate reader). The proliferation curves of transfected cells within 4 days were plotted.

II.2.2.6. Subcellular localization of Zα-eGFP fusion protein

Hela cells were transfected with pCMVss-Za-eGFP or pCMVss-eGFP by using Lipofectamine 2000. At 6 h and 24 h posttransfection, Hela cells were fixed with 4% formaldehyde for 10 min. To neutralize free formaldehyde, 50 mM NH_4Cl was added and incubated for 5 min. After permeabilization with 1% Triton X-100 in PBS for 10 min, the cells were incubated with 0.5 μg/ml DAPI in PBS for 40 min, followed by 3 10-min washs with PBS. Subsequently, the cells were incubated with 0.5 μg/ml R6 (rhodamine B, hexyl ester, perchlorate; Invitrogen) in PBS for 5 min to stain the endoplasmic reticulum (ER), followed by 3 10-min washs with PBS. After being air-dried for 20 min, the coverslips were mounted with anti-fade oil on glass slides and recorded with Zeiss LSM 500 Meta confocal microscope.

III. RESULTS

III.1. Mapping Z-DNA in the human genome using Zα as a probe

III.1.1. Zα is a Z-DNA specific binding domain

III.1.1.1. Protein purification

Both Zα and Zα mut were over-expressed in BL21(DE3) cells with 1 mM IPTG induction. The first round protein purification was achieved through affinity chromatography of Strep-Tactin resins which is engineered from streptavidin and recognizes StrepII tag in the C-terminus of Zα or Zα mut. The fractions of affinity chromatography were checked on SDS-PAGE and those containing Zα or Zα mut were combined and applied to an S-100 (Amersham) size-exclusion column. Two peaks were observed. The first peak was collected and analyzed on agarose gel and SDS-PAGE, suggesting that abundant RNA fragments and proteins were in the first peak. However, the second peak contained purified Zα / Zα mut and the buffer was changed to HEPES cation chromatography buffer. After further purification by an ion-exchange column Uno-S (Bio-Rad), proteins were checked on SDS-PAGE (**Figure III-1**) and dialyzed against dialysis buffer containing 50% glycerol and stored in -20 ^{o}C. Qualities of the purified proteins were checked on SDS-PAGE (**Figure III-2**) and protein concentrations were determined using a Bio-Rad protein assay kit, taking BSA as the standard.

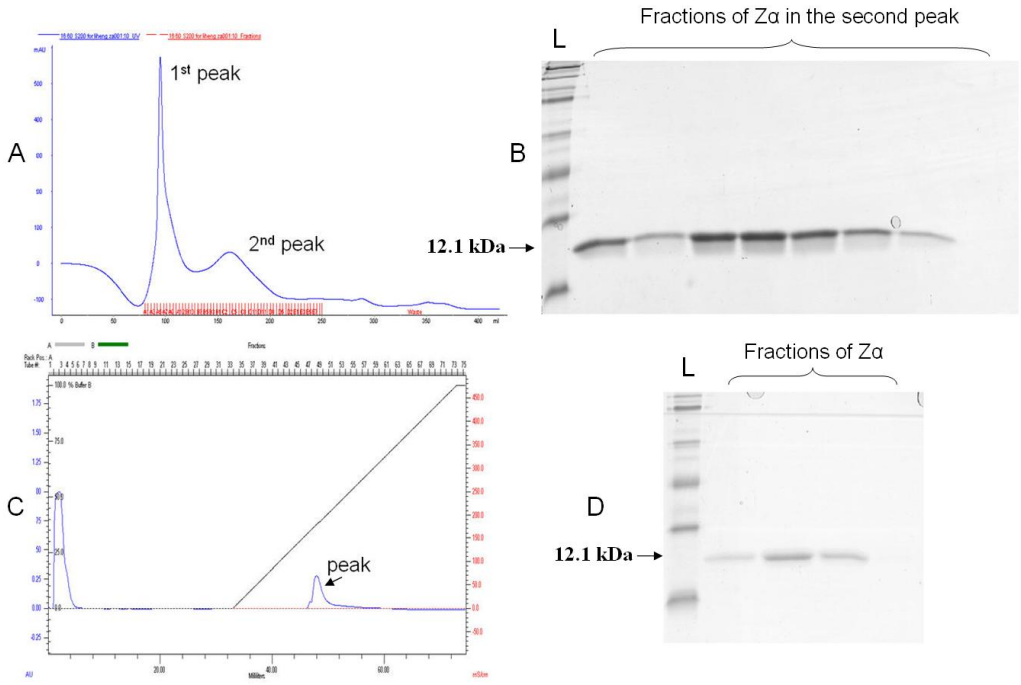


Figure III-1. Purification of *Zα* by size-exclusion and ion-exchange chromatography.

A. Purification of Zα through size-exclusion column. Two peaks were observed and the fractions were collected. B. The fractions in the 2nd peak from size-exclusion column purification were checked on SDS-PAGE. C. Further purification of Zα through Uno-S ion-exchange chromatography. D. The fractions in the indicated peak of Uno-S purification were checked on SDS-PAGE. L: protein molecular weight ladder. The predicted molecular weight of Zα was about 12.1 kDa.

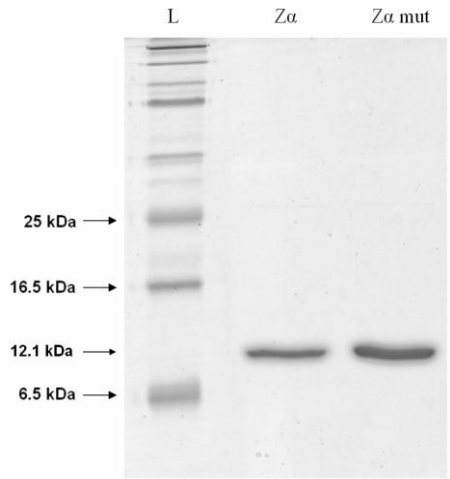


Figure III-2. Purity of *Zα* and *Zα* mut was checked on SDS-PAGE after dialysis.

III.1.1.2. Zα bound and was crosslinkable to Z-DNA specifically

In order to investigate the binding activity and specificity of Zα probe designed in this study, plasmid p101 containing $(GC)_{16}$ repeat (Droge and Nordheim, 1991) and plasmid GT-pPGKss-puro containing $(GT)_{46}$ repeat (as shown in **Figure II-6**) were used. Previous studies had shown that $(GC)_n$ and $(GT)_n$ repeats in negative supercoiled plasmids were in Z-conformation, rather than in B-conformation (Santoro and Costanzo, 1983; Droge and Nordheim, 1991; Rothenburg et al., 2001).

Plasmid p101 is derived from pTZ18R by insertion of $(GC)_{16}$ stretch. In a pull-down assay using plasmid pTZ18R and p101 as the substrates, as shown in **Figure III-3**, both pTZ18R and p101 were pulled down by Zα in Tris-HCl buffer containing 100 mM NaCl. In 150 mM salt buffer, only p101, but not pTZ18R, could be pulled down by Zα; and in 200 mM NaCl, Zα failed to pull down both plasmids, suggesting that the binding of Zα to Z-DNA is very sensitive to ionic environment.

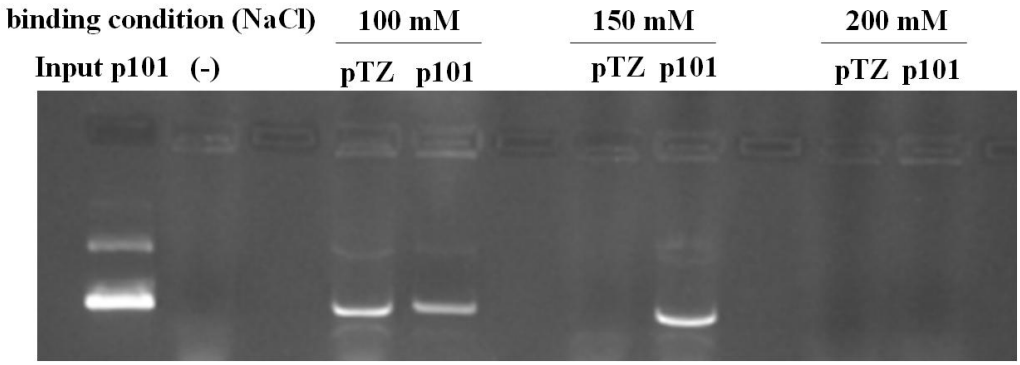


Figure III-3. Pull-down assay of supercoiled pTZ18R and p101 by Zα.
In this pull-down assay, plasmid p101 containing Z-DNA fragment was used. Input: plasmid p101; (-) is negative control, in which no protein was added. pTZ stands for pTZ18R. In 100 mM NaCl binding and wash condition, both pTZ18R and p101 were pulled down by Zα; however, in 150 mM salt condition, only p101 was pulled down by Zα. And in 200 mM salt condition, both plasmids were failed to be pulled down, suggesting the binding of Zα to Z-DNA is sensitive to ionic strength.

To determine whether the Z-DNA binding activity of Zα was conformation-dependent and sequence-independent, supercoiled plasmid GT-pPGKss-puro was constructed and used as the substrate. GT-containing fragments amplified from this plasmid by PCR were mixed as non-specific competitors as these fragments were in B-conformation. As shown in **Figure III-4**, the 760 bp fragments released by *EcoR*I digestion from supercoiled plasmid GT-pPGKss-puro were up-shifted by crosslinked Zα, while the migration of shorter GT-fragments from PCR was not affected. The smear pattern in Zα sample suggested that varied copies of Zα were crosslinked to each 760 bp DNA fragment. Since Z-conformation was stabilized by negative supercoiling, the covalent crosslinking of Zα to its binding sites by formaldehyde "freezed" the binding event even after the binding sites were

released from the negative supercoiled plasmids. This method had been used to identify Z-conformation in supercoiled plasmids by using anti-Z-DNA antibody (Nordheim et al., 1982). On the other hand, this result suggested that the Zα mut constructed in this study could not be crosslinked to either B- or Z-DNA by formaldehyde. Together, Zα is a crosslinkable Z-DNA specific binding domain recognizing different permissive sequences and Zα mut can be used as a negative control.

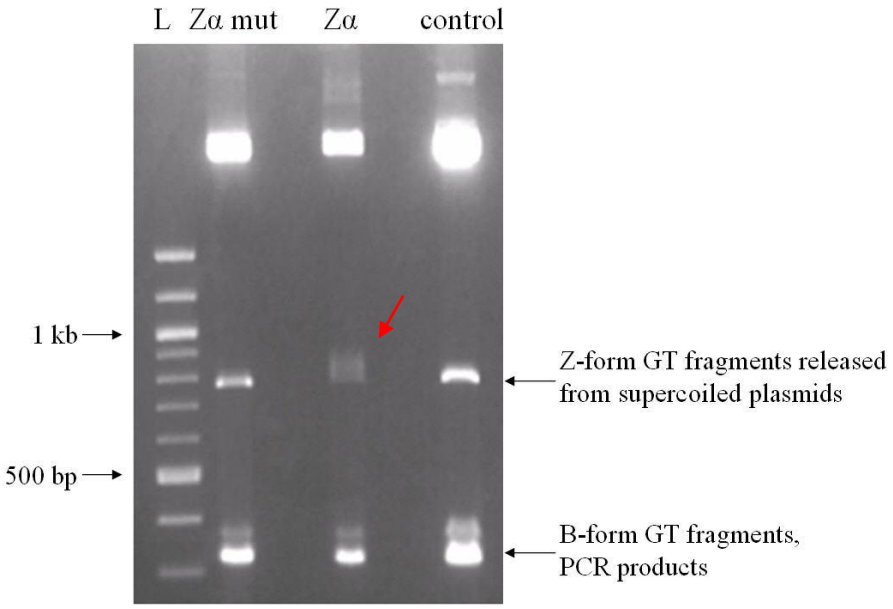


Figure III-4. Zα bound and was crosslinked to Z-DNA specifically.

The 760 bp fragments released from the supercoiled plasmids were up-shifted (as indicated by red arrow) by crosslinked Zα, compared with Zα mut and the negative control in which the same amount of BSA was used. The migration of GT fragments from PCR products was not affected by Zα or Zα mut. DNA fragments were analyzed on 1.4% agarose gel in 0.5 X TBE.

III.1.2. Expression and localization of Zα in mammalian cells

Since in this study Zα was used to probe Z-DNA fragments in mammalian cells and nuclear localization was desired, a nuclear localization signal (NLS) from

Simian virus 40 was inserted between Zα domain and epitope tags. The expression level of Zα in transiently transfected cells were analyzed by Western blotting assay (**Figure III-5**). The estimated copy number of Zα in each transfected cell was about 10^5, by comparing with the positive control. Since Zα could induce B-to-Z transition and stabilize the Z-conformation (Berger et al., 1998), Zα over-expressed at high level may induce Z-conformation in living cells, thus the expression vectors used here did not contain a Kozak sequence and the expression level was moderate.

Figure III-5. Expression of Zα in transfected Hela cells checked by Western blotting.
Untransfected Hela cells were used as the mock control and 10 ng of purified Zα was used as the positive control. The copy number of Zα in each transfected Hela cell was estimated as about 10^5, compared with the positive control, based on the measurement with a Bio-Rad densitometer.

The subcellular distribution patterns of Zα and Zα mut in mammalian cells were examined by immunostaining using FITC-conjugated anti-FLAG M2 monoclonal antibody (Sigma). As shown in **Figure III-6**, both Zα and Zα mut localized not only in the nucleus but also in the cytoplasm. Surprisingly, Zα accumulated in the nucleoli or on the surfaces of nucleoli. This novel finding may contribute to understanding the role the Zα domain plays as a part of

ADAR1, which was reported to localize in both nucleus and cytoplasm (Patterson and Samuel, 1995).

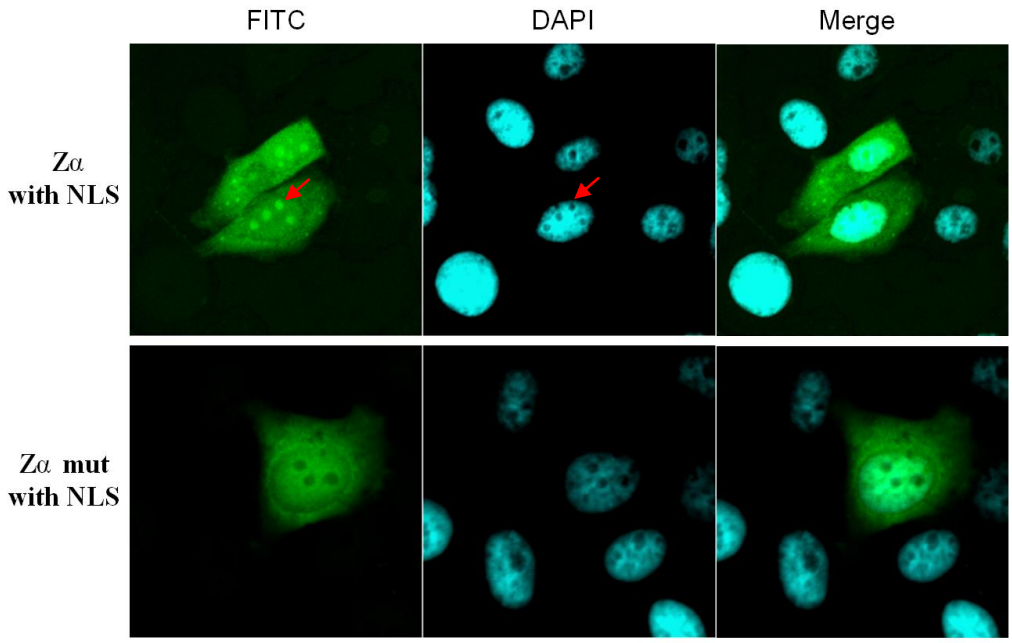


Figure III-6. Immunostaining of *Zα* and *Zα* mut after 24 h posttransfection.
Zα accumulated to form green granules (as indicated by red arrow) which overlapped perfectly with the nucleoli (as indicated by red arrow in DAPI stained nuclei). But Zα mut did not accumulate as granules at all. This picture is one example from at least 3 independent experiments.

III.1.3. *In vivo* chromatin affinity precipitation (ChAP)

Initially *in vivo* ChAP by Strep-Tactin beads was tested since the StrepII tag was added to the C-terminus of Zα. Prior to being eluted from the beads, the pull-down material was washed stringently with buffer containing 250 mM LiCl and then eluted by elution buffer. The eluted materials were radio-labeled with γ^{32}P-ATP by T4 polynucleotide kinase and analyzed on 12% native PAGE, as shown in **Figure III-7A**. It was shown that more nuclei acids were precipitated by Zα than by Zα mut.

In order to characterize the nature of the nucleic acids precipitated by Zα and Zα mut, RNase sensitivity tests were performed. The precipitated materials by Zα and Zα mut were radio-labeled and treated with DNase-free RNase (Roche), then checked on 12% native PAGE. As shown in **Figure III-7B**, the nucleic acids precipitated by both Zα and Zα mut were RNase sensitive, suggesting that the precipitated nucleic acids were RNA fragments. Since Zα is a domain from a RNA editing enzyme and it has been reported that Zα could bind to Z-RNA *in vitro* (Brown et al., 2000; Koeris et al., 2005), it is reasonable to postulate that Zα binds to RNA *in vivo*, probably Z-RNA. Most importantly, so far no naturally occurring Z-RNA has been reported and this novel finding provided an opportunity to characterize Z-RNA *in vivo*.

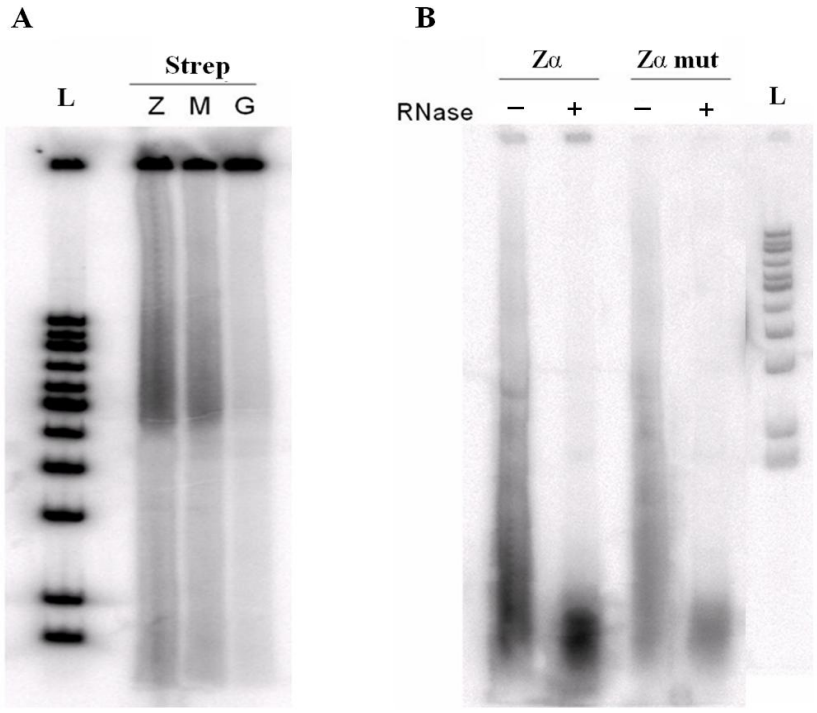


Figure III-7. *In vivo* chromatin precipitation.

Z, M and G stand for Zα, Zα mut and eGFP, respectively. L stands for 100 bp DNA ladder. A, ^{32}P radiolabelled nucleic acids precipitated by Zα, Zα mut and eGFP with Strep-Tactin beads. It

shows that more nuclei acids were precipitated by Zα than by Zα mut; B, RNase sensitivity test. In the presence of RNase, the nucleic acids precipitated by both Zα and Zα mut were degraded, suggesting that the precipitated materials were RNA fragments.

III.1.4.*In vitro* ChAP

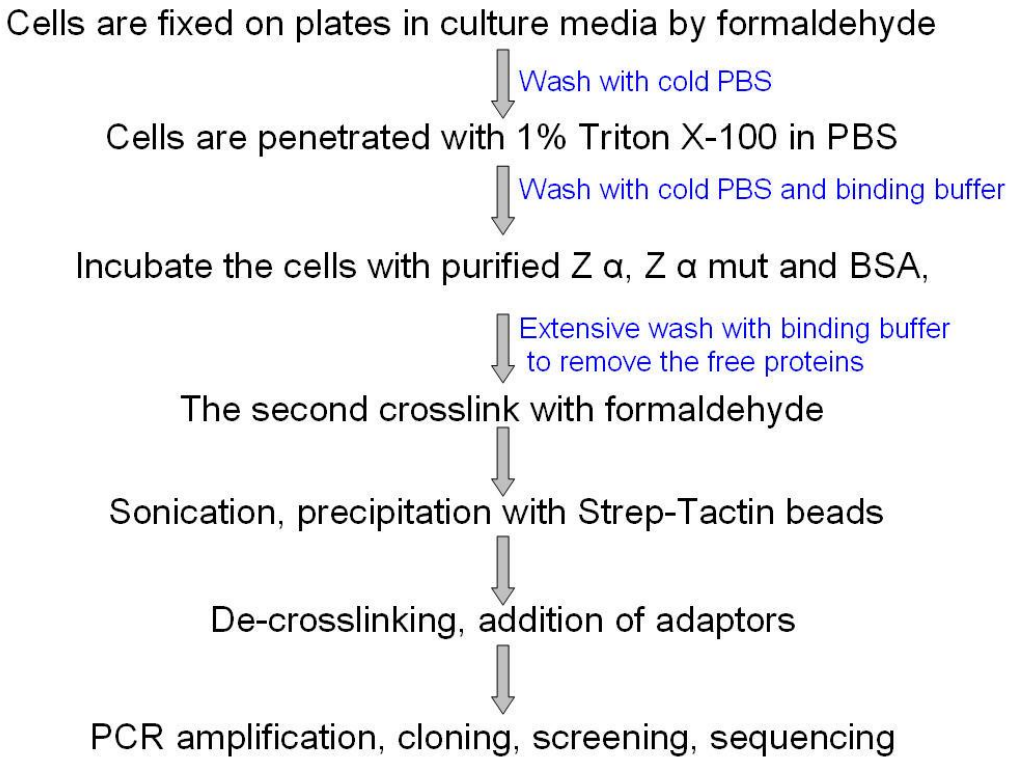


Figure III-8. The process of *in vitro* ChAP.

Since the material precipitated by Zα was RNA molecules when using *in vivo* ChAP, a novel experimental strategy modified from *in vivo* ChAP was proposed and exploited. It has been reported that Zα could recognize Z-DNA structure in cells fixed by formaldehyde (Liu et al., 2006). Based on this finding, the proposed *in vitro* ChAP exploited two steps of crosslinking with formaldehyde (**Figure III-8**). Initially, A549 cells grown in flasks were fixed with 1% formaldehyde to 'freeze' the cells in the living state. Because Z-DNA is in

higher energy state and unfavorable, the first crosslinking would prevent the flip from Z- to B-conformation in the cells, as evidenced by the study on the CSF1 promoter (Liu et al., 2006). Then 1% Triton X-100 in PBS was used to permeabilize the cellular membranes. After a wash, equal amounts of purified Zα and Zα mut were added to the cells to let the proteins recognize their binding sites, followed by extensive washs to remove unbound proteins. The second crosslinking was applied to fix the binding between Zα and its binding sites. Then the cells were collected and sonicated. After precipitation by Strep-Tactin beads and a stringent wash with 1 M NaCl and 1.5 M urea, the precipitated DNA fragments were ligated to adaptors and amplified to check the differences in quantity of DNA precipitated by Zα, Zα mut and BSA control. After addition of adaptors, the *in vitro* ChAP samples were amplified by PCR with primers priming adaptor regions and checked on agarose gel, as shown in **Figure III-9**. The PCR products from Zα ChAP sample was significantly more than those from Zα mut ChAP sample. Since the primers and PCR conditions were the same, the difference in the final PCR products reflected more DNA fragments were precipitated by Zα than by Zα mutant.

III.1.5. Construction of Z-DNA library and quality test

After de-crosslinking and generating blunt-ends, the precipitated DNA molecules were linked to adaptors at both ends and amplified by PCR with Adaptor-L for 25 cycles. The PCR products were separated on agarose gel and extracted, followed by digestion with *Bam*HI. And then the DNA samples were separated on agarose gel again and extracted to get rid of short fragments

(**Figure III-10**). After *Bam*HI digestion, the sizes of DNA fragments ranged from 100 bp to 1000 bp, mainly in 200-400 bp. The gel-purified DNA fragments were inserted into linearized pTZ-18R vectors and electroporated into Stbl4 competent cells. After recovery, 10% transformed Stbl4 cells were plated on LB plates with blue/white screening. When blue colonies were recognized clearly, the number of white colonies was counted. Totally, there were about 1,200 white colonies and 350 blue colonies. It was possible that the dimer of self-ligated adaptors, which was 16 bp after *Bam*HI digestion, could be cloned and formed the white colonies. In order to investigate the sizes of inserts, colony PCRs using primers priming the flanking regions of inserts in pTZ18R were performed. Colony PCRs showed that 24/24 white colonies contained inserts longer than 100 bp, which could not be the dimer of adaptors (**Figure III-11**), suggesting that almost 100% white colonies contained the precipitated DNA fragments, other than dimers of the adaptors.

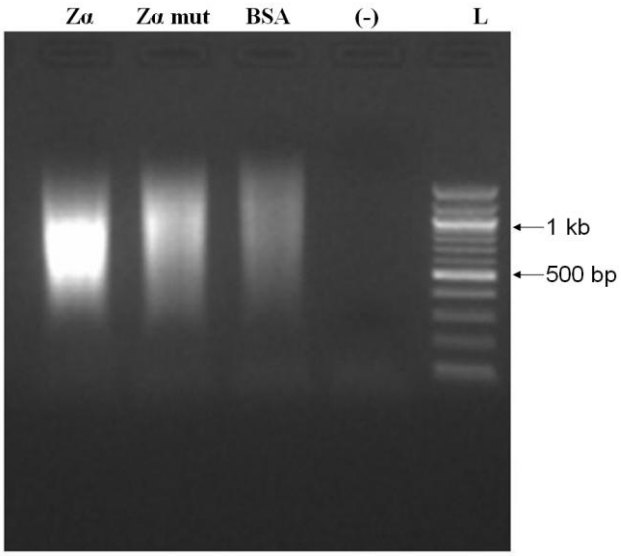


Figure III-9. PCR amplification of *in vitro* ChAP DNA.

(-) stands for PCR negative control (no template added). L is 100 bp DNA ladder. Zα precipitated much more DNA fragments than Zα mut and BSA.

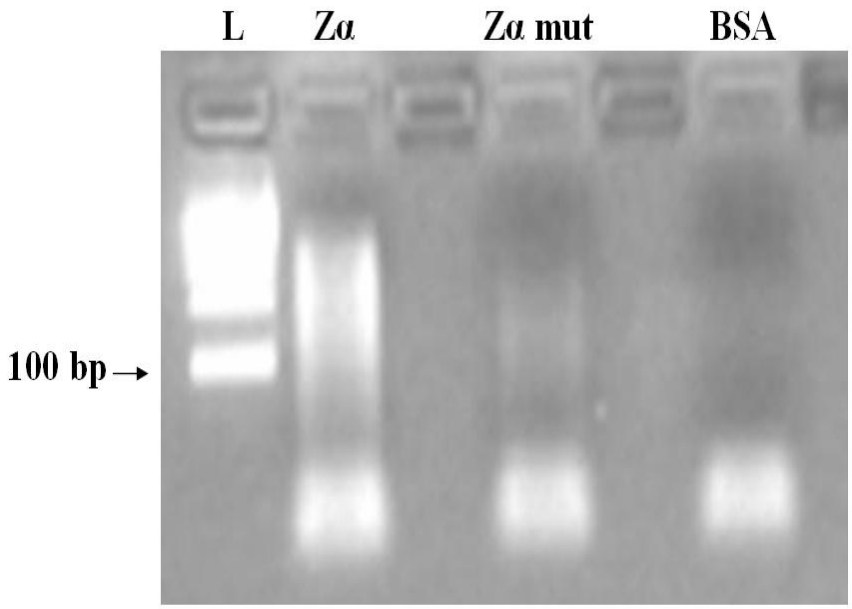


Figure III-10. PCR products from *in vitro* ChAP digested by *Bam*HI.
L stands for 100 bp DNA ladder. The upper smear region from Zα sample was recovered by gel extraction. The lower smear was released from the adaptors and existed in all 3 samples.

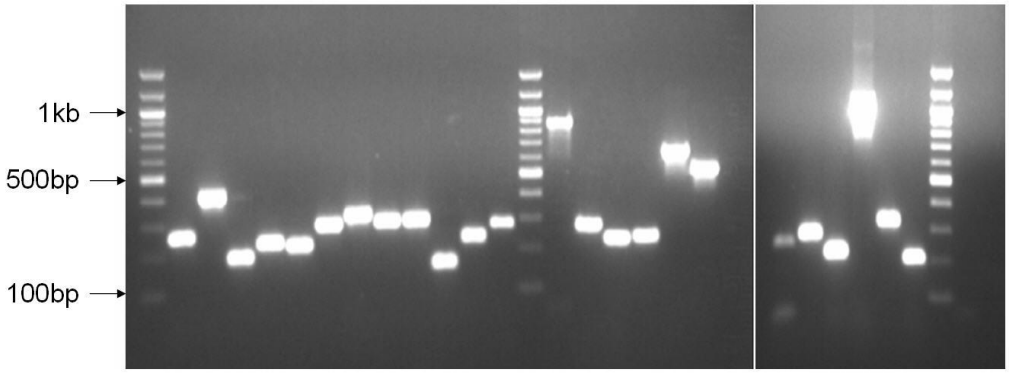


Figure III-11. Colony PCR showed that all white colonies had inserts larger than 100 bp.

III.1.6. Large scale sequencing and data analyses

(Performed with assistance from Miss Xiao Jie)

III.1.6.1. Summary of large scale sequencing data

All white colonies in the Z-DNA library were inoculated and cultured and plasmids were prepared and sequenced using primer pTZ-L. In total, 10,563 plasmids were sequenced and the distribution pattern of the lengths of the raw sequencing data are depicted in **Figure III-12**. The main peak of the lengths of original sequencing readings was in the range of 450-630 bp.

Initially, pTZ18R vector sequences and adaptor sequences in the raw sequencing data were removed by multiple alignments. Some sequencing readings contained more than one DNA fragments which were interrupted by adaptors, due to concatenation during cloning process. These DNA fragments were isolated and treated as independent ChAP fragments. In order to gain precise alignments on the human genome, those fragments less than 30 bp were discarded. The total number of analyzed ChAP fragments was 10,005. The length distribution pattern of these isolated ChAP fragments was depicted in **Figure III-13** and the main peak of length distribution was from 100 bp to 300 bp.

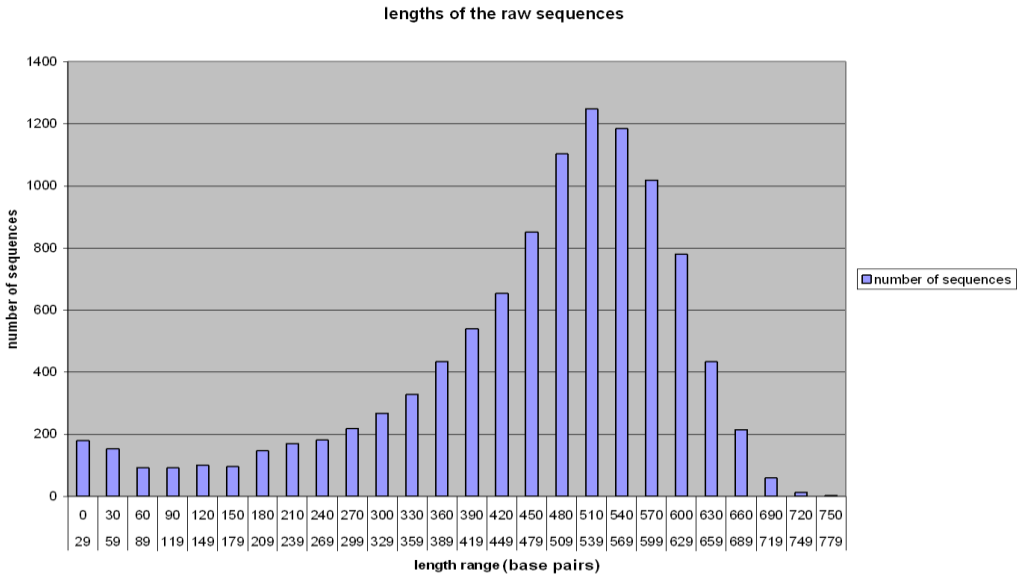


Figure III-12. Length distribution pattern of raw sequencing readings.

The main peak was at 450-630 bps which was the reasonable reading reach for dye-terminator sequencing strategy.

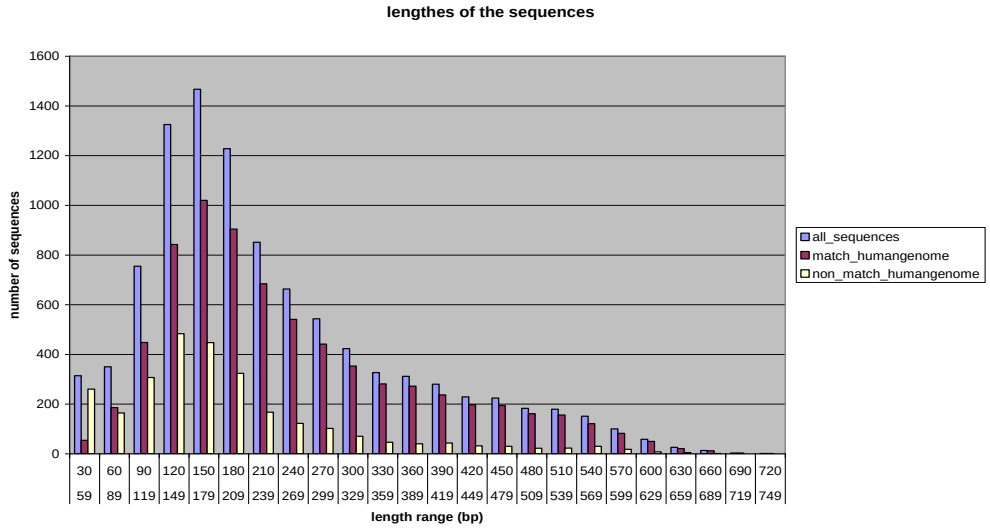


Figure III-13. Length distribution patterns of isolated ChAP fragments.

The sequences of vectors and adaptors were removed by multiple alignments. The fragments matched and unmatched with the human genome were also listed individually.

III.1.6.2. Mapping ChAP fragments on the human genome

The 10,005 segments were subjected to BLASTN (http://blast.ncbi.nlm.nih.gov) or Blat (http://genome.ucsc.edu) for alignments against the human genome build 36.2 (NCBI). 7,321 reads gave an alignment with a corresponding human genome fragment with sequence identity >= 90%, and no more than 1% gap; 1,715 reads were mapped to the *E. coli* genome and 969 were not found in NCBI nr (non-redundant) database. Interestingly, these 969 unidentifiable sequences are GC-rich and are prone to form Z-DNA, based on predictions via Z-Hunt (http://gac-web.cgrb.oregonstate.edu/zDNA/), which is an online server for Z-DNA prediction (Schroth et al., 1992; Champ et al., 2004). However, since these sequences could not be identified in the database, we did not analyze them further.

III.1.6.3. Refinement of hotspots

When the ChAP fragments were mapped to the human genome, some fragments were close to each other and even overlapped. Z-DNA hotspot is defined as the minimum genome region covering at least two ChAP fragments which are overlapped or less than 100 bp apart. Hotspots consisted of two identical DNA sequences were eliminated since the identical fragments might be generated redundantly from PCR amplification. After removal of these hotspots, 350 hotspots were marked based on the definition of hotspots. However, there were several possibilities to produce **redundant or dubious hotspots** if we just follow the definition of hotspots, as shown in **Figure III-15**. Hotspots with identical sequences were classified as redundant hotspots and removed. Besides

these removed hotspots, there was another situation to produce some Z-DNA hotspots, named as **high potential Z-DNA hotspots** (**Figure III-14**). The ChAP fragments consisted of these high potential Z-DNA hotspots were in repeating regions in the genome, however, they were found to accumulate only in one specific region, i.e. to form one hotspot. These hotspots were different from the dubious hotspots listed in **Figure III-15**. If ChAP fragments in one hotspot were mapped to unique loci in the genome, the hotspot was termed as **unique hotspot**. Totally, 122 unique and 64 high potential hotspots were defined and mapped on the human chromosomes (as shown in **Figure III-16**). The detailed information about Z-DNA hotspots was listed in Supplementary **Table VII-1**.

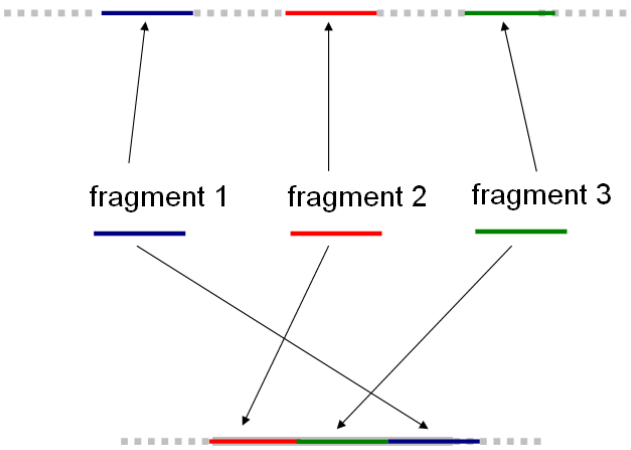


Figure III-14. The mechanism of marking high potential Z-DNA forming hotspots.
ChAP fragments 1, 2, 3 could be mapped to more than one region; however, they were accumulated only in one specific region to form one Z-DNA hotspots. If binding of Zα to every locus was considered as one independent event during ChAP process (4 independent binding events), then the chance for that all 3 ChAP fragments came from 3 different regions was 1/4 x 1/4 x 1/4=1/64; while the chance for that all 3 ChAP fragments came from one hotspot region was 1/4, much higher than the chance of coming from difference loci. So these hotspots were regarded as high potential Z-DNA forming hotspots.

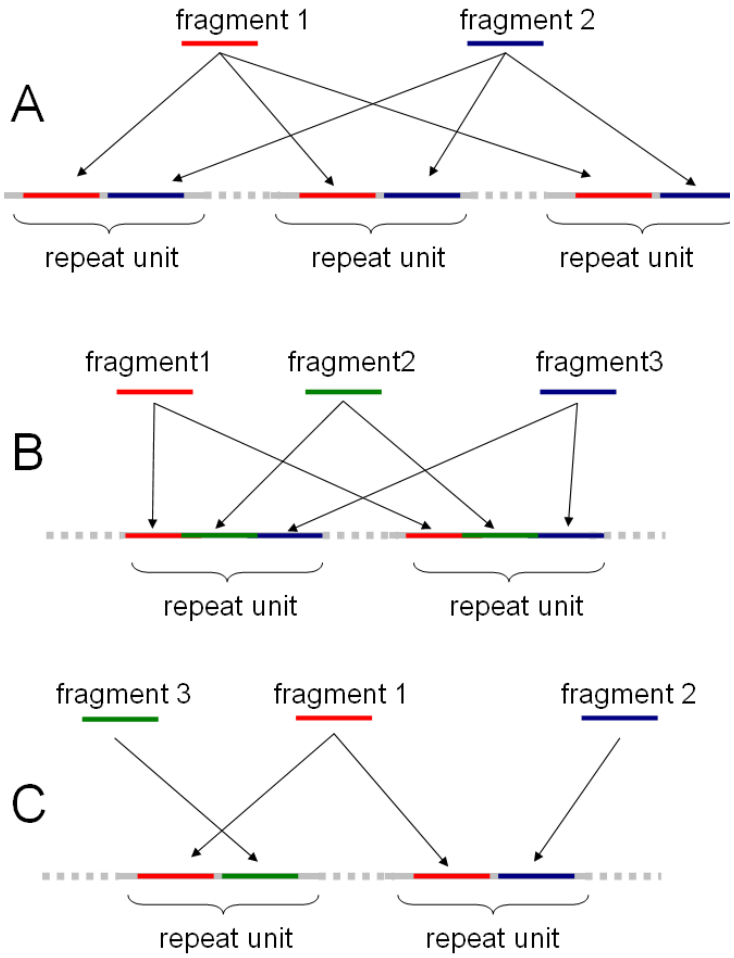


Figure III-15. The possible reasons for marking redundant or dubious hotspots.

A. Both ChAP fragment 1 and 2 could be mapped to 3 genome regions which were identically repeated in the genome. B. 3 ChAP fragments could be mapped to 2 genome regions which were identical. In this case, these 3 ChAP fragments would form at least one hotspot, however, since it was impossible to tell precisely which region was in Z-conformation, these two hotspots were classified as dubious hotspots. C. ChAP fragment 3 and 2 were highly homologous and mapped to different genome regions; however, ChAP fragment 1 could be found at two genome regions, thus forming two dubious hotspots and at least one of them was in Z-conformation.

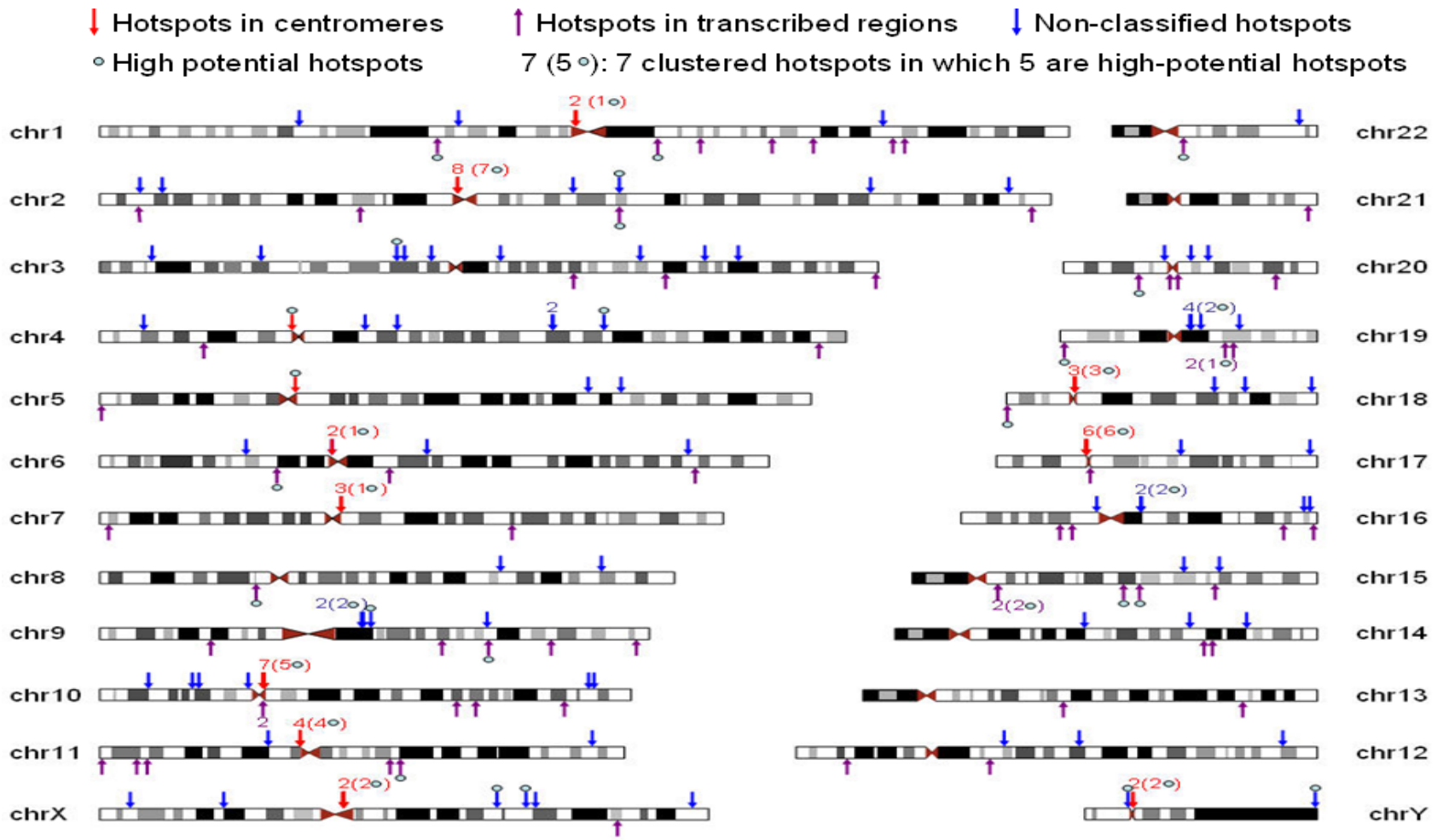


Figure III-16. The map of Z-DNA hotspots on the human genome.

III.1.6.4. Z-DNA hotspots were confirmed by ChAP-PCR

In order to confirm that these hotspots are *bona fide* binding sites of Zα, two rounds of ChAP-PCRs were performed. The first round ChAP-PCR was used to compare the enrichments of hotspots in ChAP DNA samples precipitated by Zα, Zα mut and BSA. Since the start number of cells for ChAP by Zα, Zα mut and BSA were the same, the input DNA in these three ChAP experiments should be the same. Then the difference of PCR products reflected the difference of copy number of hotspots in precipitated materials. PCR results showed that 14 out of 16 PCRs gave more PCR products from Zα ChAP sample than from both Zα-mut sample and BSA sample (**Figure III-17**).

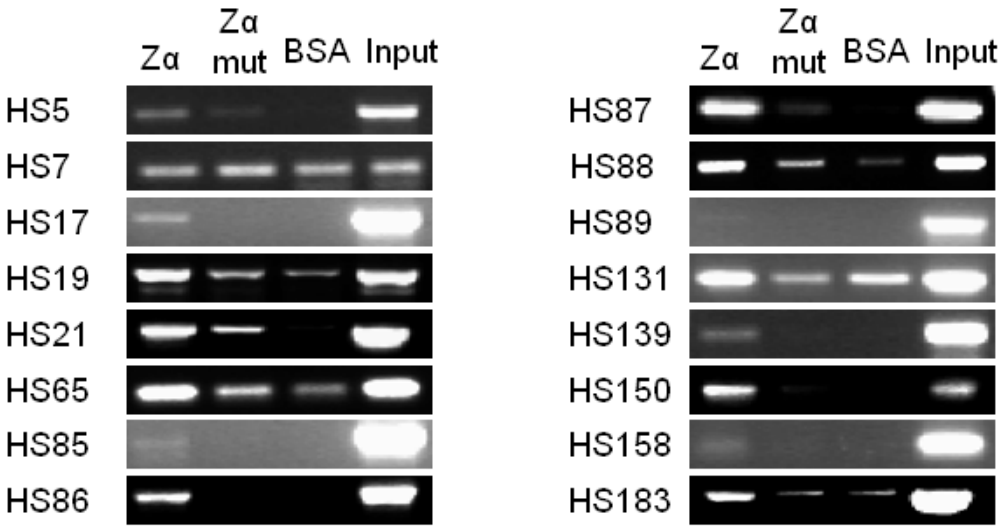


Figure III-17. ChAP-PCRs using ChAP materials of Zα and Zα mut as the templates.
HS stands for hotspot. Input is PCR product using input DNA as template. 14 out 16 (except HS 7 and HS 89) PCRs produce more products from Zα ChAP sample than from both Zα mut and BSA sample. HS85 and HS158 give quite weak signals in Zα ChAP sample.

It has been confirmed that Zα could pull down more Z-DNA hotspot fragments than Zα mut; however, it was still questioned whether these DNA fragments were pulled down specifically by Zα or not. To clarify the question, the second round ChAP-PCR was performed to compare the enrichments of hotspot regions and non-hotspots regions in the Zα ChAP sample. In order to compare the enrichments, PCR products from Zα ChAP DNA were standardized to its corresponding control PCR products using genomic DNA as the template. Since the same amounts of genomic DNA in all control PCRs were used, and all these PCR regions were unique in the genome, the start copy number of template for each control PCR should be the same. Each PCR was performed in triplicate to minimize experimental variations. Totally, 7 pairs of primers in hotspot regions and 7 pairs in non-hotspot regions were checked (**Figure III-18**A). The signal of each PCR product was quantified and the ratio between the ChAP-PCR and the control PCR was calculated. The enrichments were expressed as ratios after generalized against their corresponding control PCRs. In each series, 7 ratios were sorted with increasing values and plotted, as shown in **Figure III-18**B. 6 out of 7 hotspot PCRs showed higher enrichments than non-hotspot PCRs, suggesting that these Z-DNA hotspots defined in this study are preferentially precipitated, compared to non-hotspot regions.

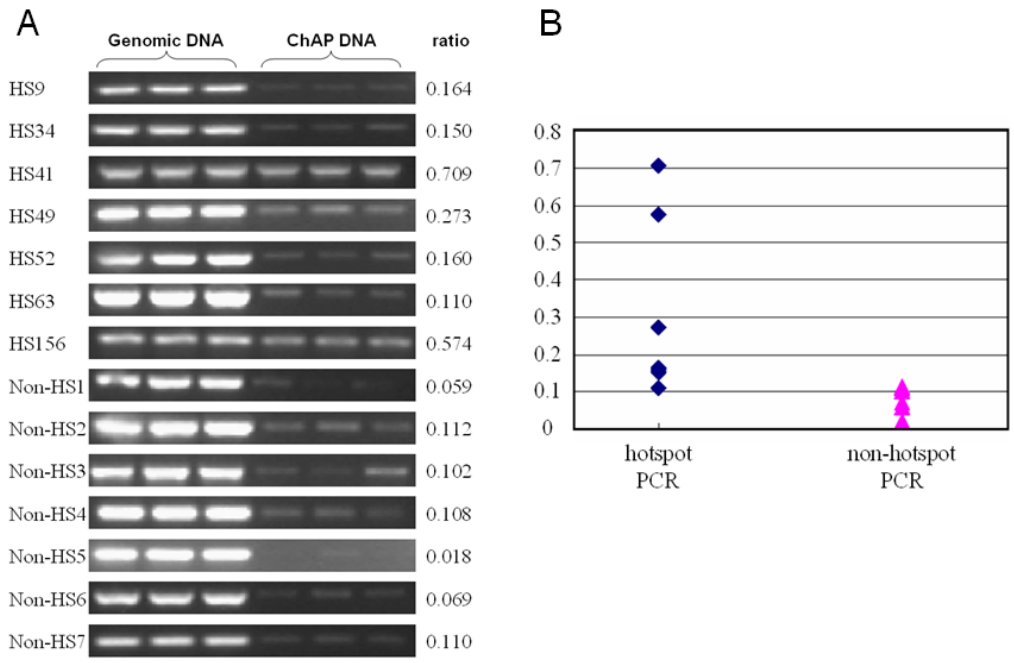


Figure III-18. ChAP-PCRs of hotspots regions and non-hotspot regions.

A. ChAP-PCR to quantify the enrichments of hotspot DNA and non-hotspot DNA in Zα ChAP sample. All PCRs were performed in triplicate to minimize experimental variations. The signal of each band was quantified using Bio-Rad Quantity One software and the ratios between ChAP-PCRs and Genomic DNA PCRs were calculated and listed at the right side. B. The ratios of hotspot regions and non-hotspot regions were sorted with increasing values and plotted. The difference between these two series was significant, suggesting that these Z-DNA hotspots were preferentially precipitated by Zα.

III.1.6.5. Analyses of Z-DNA hotspots

(With assistance from Miss Xiao Jie)

III.1.6.5.1. No correlation between Z-DNA hotspots and transcription

Intuitively, it was expected that Z-DNA hotspots were enriched in gene regulatory regions according to previous *in sicilo* analyses (Schroth et al., 1992; Champ et al., 2004; Khuu et al., 2007). The analysis performed via Z-catcher in our laboratory (a Z-DNA prediction software written by our laboratory) also suggests that Z-DNA forming regions (ZDRs) are encriched in the

transcriptional regulatory regions in the human genome (**Figure III-19**).

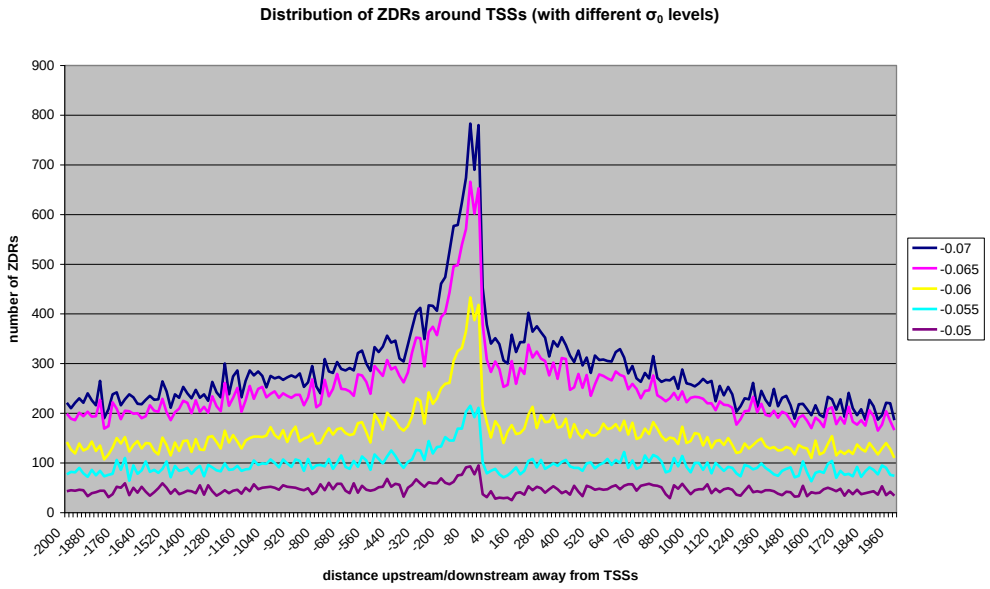


Figure III-19. Distribution of predicted ZDRs with varied σ_0 values around TSSs in the human genome.
The distribution of ZDRs flanking the transcription start sites (TSS) is enriched in the region [-600, 100] around TSSs, and there is a dramatic drop of ZDR abundance in the [100, 140] region downstream of TSSs. The obvious peak at σ_0= -0.055 suggests that these transcriptional regulatory regions have strong Z-DNA forming potentials. The figure was produced based on the prediction of Z-catcher, a Z-DNA prediction software, by Miss Xiao Jie.

However, no correlation between Z-DNA hotspots and transcription start sites (TSS) was found. Interestingly, we found that 66 hotspots were located in transcribed regions if we took all EST and mRNA in the UCSC Genome Browser as evidence of transcription; Closer inspection revealed that the majority (49 out of 66, 74.2%) of these hotspots was in introns; the rest of the hotspots were in exons or splicing regions. Z-DNA forming hotspots in introns might be linked with gene transcription as reported (Wittig et al., 1992; Wolfl et al., 1996). The transcription levels in A549 cells of these 66 hotspots were

checked in NCBI GEO profile GSM94306 (gene expression profile of A549) and only 31 hotspots were found in the database. Since 10 hotspots are transcribed at high level and the rest 21 are at low level, no correlation is established between transcriptional status and Z-DNA hotspots.

III.1.6.5.2. Most hotspots are tandem repeats

136 out of 186 Z-DNA hotspots belonged to varied tandem repeat families; 49 belonged to ALR/alpha satellite and 22 belonged to various Alu subfamilies; the remaining hotspots were in different repeat classes, including LTR/L1, MER, etc (Supplementary **Table VII-1**). Interestingly, 34 hotspots in ALR/alpha satellite family were mapped to the centromeric regions on 13 chromosomes. In addition, 12 non-alpha satellite hotspots were also mapped to centromeric regions. Analyses via RepeatMasker Web Server (http://www.repeatmasker.org/) suggested that 6 of these 12 hotspots were HSATII satellites, 1 was SST1 satellite and 2 were BSR/beta satellites, 3 were in non-repeat regions. In all 46 hotspots in centromeres, 35 were high potential Z-DNA hotspots. Since centromeric regions are not GC rich, formation of Z-DNA in these regions is not preferred. Analyses of hotspots sequences via Z-catcher (a Z-DNA prediction program produced by this lab) demonstrated that only at $\sigma < -0.08$ most of these hotspots (143 out of 186) could be flipped into Z-conformation, suggesting that, in centromeric regions, negative supercoiling in the surrounding micro-environments is strong enough to drive these hotspots to flip into Z-conformation.

III.1.6.5.3. Hotspots co-localized with high SNP density regions in centromeres but not in arms

syn conformation of purines in Z-DNA stretches makes them more accessible to modification or genome toxins (Zimmerman, 1982), and two extruded nucleotides at B-Z junction are potential modification sites (Ha et al., 2005), both of which led us to investigate the correlation of Z-DNA hotspots and the occurrences of single nucleotide polymorphism (SNP) collected in NCBI dbSNP build 128, compared with their flanking regions on both sides (depicted in **Figure III-20**). Both validated RefSNPs and non-validated RefSNPs (all entries of RefSNPs in the database) were used in this study. RefSNPs in hotspots and in the flanking regions at both sides (500 bp, 1 kb, and 2 kb) were counted and generalized as SNP density. Pair-wised Wilcoxon test was used to compare SNP densities in hotspots and in flanking regions. Interestingly, it revealed that 46 Z-DNA hotspots in centromeres were correlated with high occurrences of SNPs very tightly ($p<0.05$) when only validated RefSNPs were counted. When all entries of RefSNPs in dbSNP128 were counted, this correlation was even more significant ($p<0.001$) (examples were presented in **Figure III-20**B). If only the 11 unique hotspots in centromeres were calculated using all RefSNPs entries, the correlation was still significant ($p<0.05$). However, similar analysis of hotspots on arm regions ($n = 140$) did not show significant enrichment of SNPs in hotspots regions ($p > 0.1$) whenever either validated or all RefSNPs were used (p values are listed in **Table III-1**).

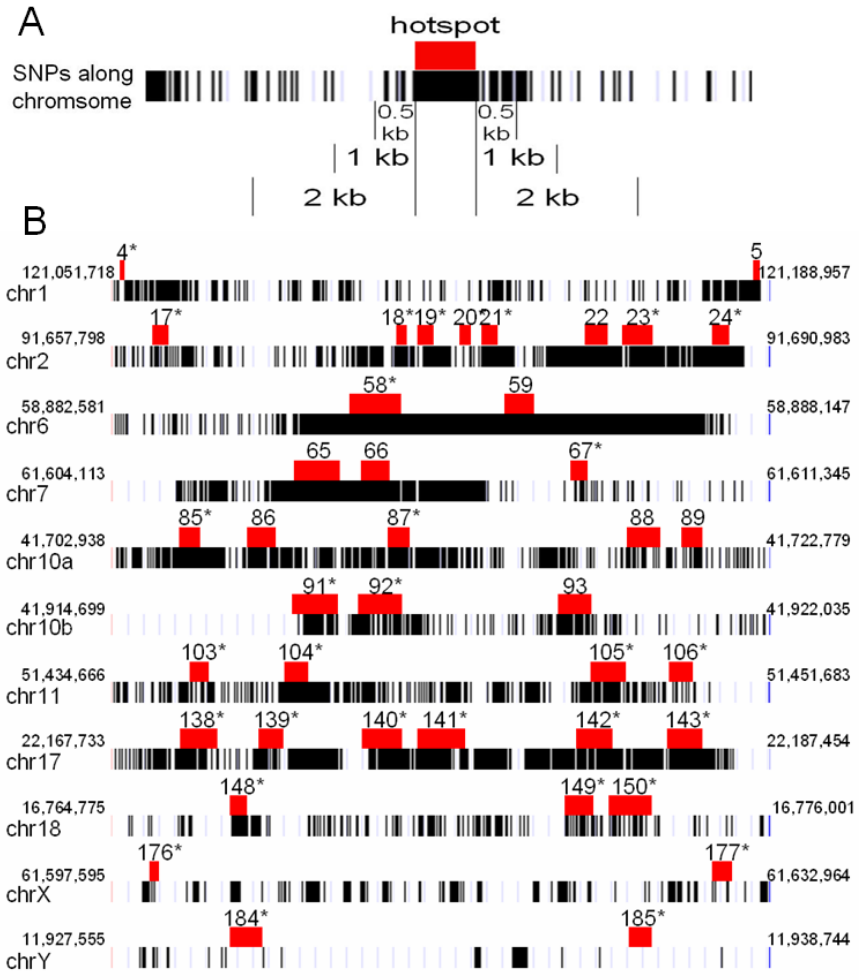


Figure III-20. Z-DNA hotspots are correlated with high occurrences of SNPs in centromeric regions.

A, Schema of the comparison of SNP density in hotspot region and its flanking regions. Distributions of RefSNPs were generalized as densities and compared between hotspots and flanking regions at two sides, including 0.5 kb, 1 kb and 2 kb. **B**, Hotspots mapped against the distribution of RefSNPs showed the correlation between RefSNPs and hotspots. High potential Z-DNA hotspots were labeled by asterisks.

	46 hotspots in centromeres			11 unique hotspots in centromeres		
	0.5 kb	1 kb	2 kb	0.5 kb	1 kb	2 kb
All_RefSNPs	<0.001	<0.001	<0.001	0.021	0.026	0.013
Validated_RefSNPs	0.014	0.045	0.042	0.075	0.041	0.05
	140 hotspots in arms			all 186 hotspots		
	0.5 kb	1 kb	2 kb	0.5 kb	1 kb	2 kb
All_RefSNPs	0.283	0.211	0.388	0.11	0.103	0.089
Validated_RefSNPs	0.865	0.816	0.942	0.238	0.325	0.219

Table III-1. *p* values of pair-wised Wilcoxon tests of SNP densities.

The statistical analyses suggested that only the Z-DNA hotspots in centromeric regions were correlated with high occurrences of SNPs.

III.2. Detection of natural Z-RNA in mammalian cells by Zα

III.2.1. *In vivo* RNA binding activity of Zα

During the progress of our experiments, there was some evidence suggesting that Zα was able to bind to RNA *in vivo*. Firstly, when Zα was overexpressed in *E. coli* BL21(DE3) and purified with Strep-Tactin beads, lots of RNA fragments were co-purified. When further purified with a size-exclusion column, there was one great peak before the protein was flowed out. Nuclease sensitivity test and analysis on SDS-PAGE suggested that the great peak was caused by RNA-protein complex (**Figure III-21**). The second hint came from *in vivo* ChAP in which the majority of precipitated material of Zα was RNA (**Figure III-7**). So Zα might function as a RNA binding domain to recognize RNA in a special conformation. Both structural analysis and *in vitro* experiments have shown that Zα could bind to Z-RNA specifically (Brown et al., 2000; Koeris et al., 2005; Placido et al., 2007). However, in all these experiments, the substrates were chemically synthesized or artificially designed RNA molecules, not naturally occuring RNA molecules in cells. Here we tried to identify the natural Z-RNA in mammalian cells using Zα as a conformation-specific probe.

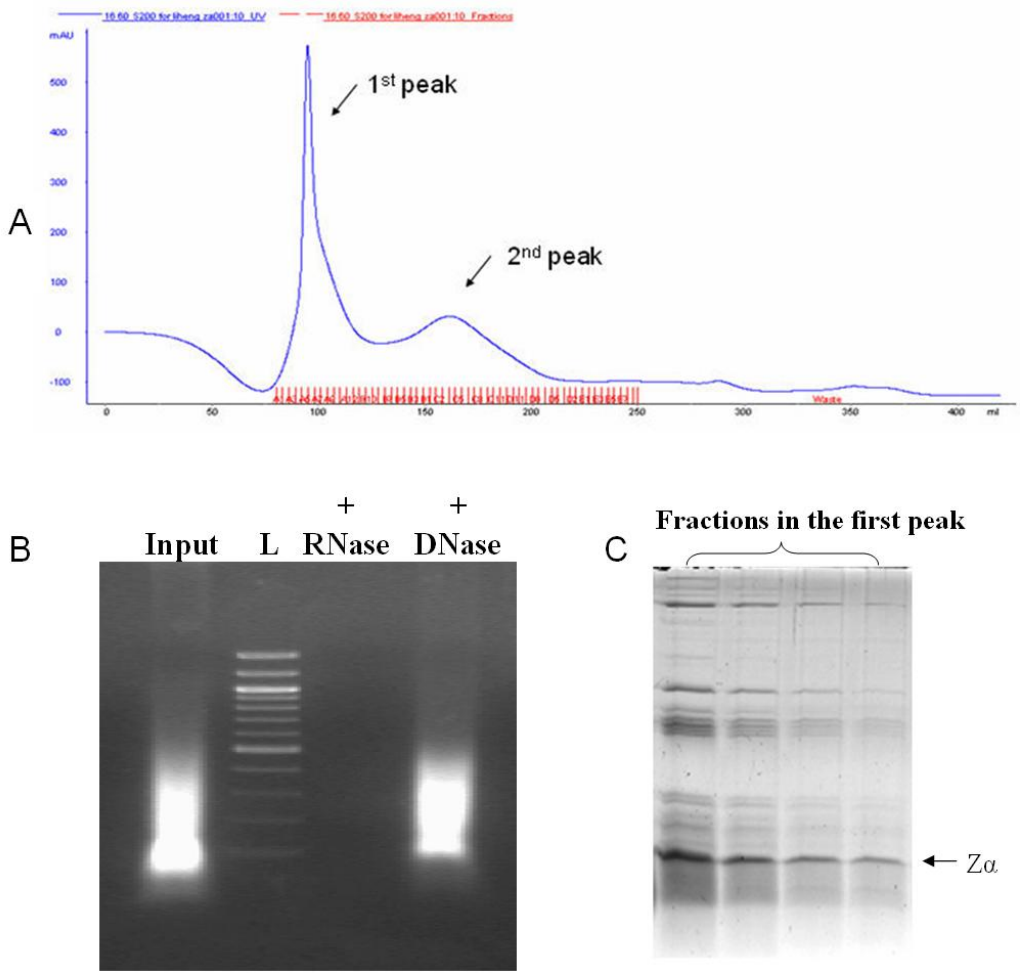


Figure III-21. Abundant RNAs and proteins were co-purified with Zα.

A. Purification profile of Zα through a size-exclusion column. The second peak was purified Zα. B. Nuclease sensitivity test of a fraction from the first peak. It suggested that abundant RNA fragments were co-purified with Zα during the first round affinity purification. C. The fractions from the first peak were checked on 15% SDS-PAGE and visualized with Coomassie blue staining. It showed that abundant proteins were co-purified with Zα.

III.2.2. Subcellular localization of *Zα* without NLS

In order to investigate the natural binding sites of Zα, the NLS was removed from the expression vector. Subcellular distributions of Zα and Zα mut without NLS were examined by immunostaining. The localization patterns of Zα and Zα mut without NLS (**Figure III-22**) were comparable to their counterparts

with NLS (**Figure III-6**). Interestingly, even when there was no NLS in the Zα construct, Zα still entered the nucleoli or accumulated on the surfaces of the nucleoli; however, Zα mut did not. This finding strengthens the possibility that Zα domain, as a part of ADAR1, plays an important role in nucleolus. Since the only differences between Zα and Zα mut are mutations of two amino acids essential for Z-conformation binding, the binding and localization of Zα in nucleoli and even its potential biological functions should be Z-conformation dependent.

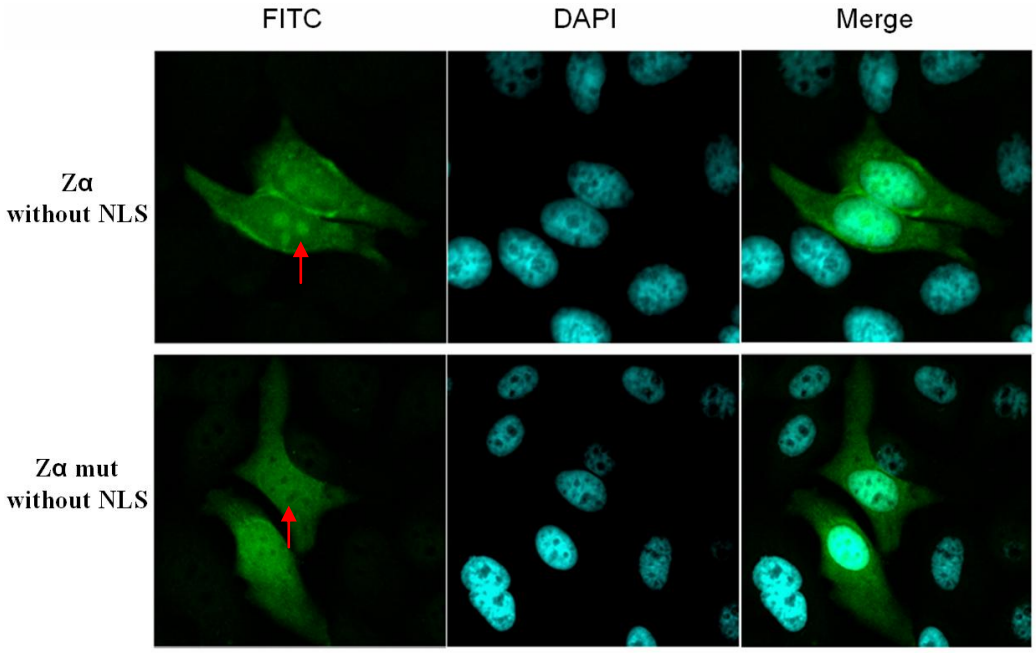


Figure III-22. Subcellular localizations of Zα and Zα mut without NLS.
Zα aggregated prominently in nucleoli or on the surfaces of nucleoli, but Zα mut did not (indicated by red arrows). These observations were in agreement with those of Zα and Zα mut with NLS, suggesting that the accumulation of Zα in the nucleoli was not caused by the NLS.

III.2.3. *Zα* binds to mammalian ribosomes specifically

III.2.3.1. *Zα* binds to ribosomes in a conformation-dependent manner

(Contributed by Miss Feng Shu)

In order to identify the binding specificity of Zα and Zα mut to the mammalian ribosome, untreated rabbit reticulocyte lysate (RRL), which did not contain DNA, was used and various amounts of B-DNA fragments were added as competitors. As shown in **Figure III-23**B, with increasing amounts of B-DNA competitors, the amounts of ribosome precipitated by Zα mut decreased and almost nothing was pulled down in the presence of 20 μg of B-DNA competitors; in contrast, pull-down of ribosomes by Zα was not affected by B-DNA competition at all, indicating that the binding of Zα on ribosome was B-conformation independent, and probably Z-conformation dependent.

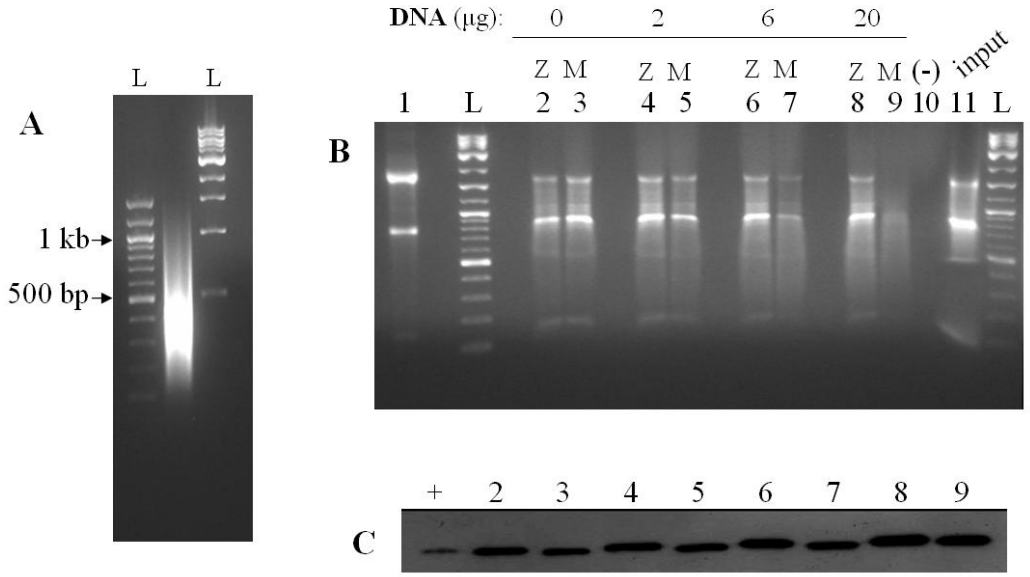


Figure III-23. Pull-down of untreated RRL with B-DNA competitors.

A. Hela genomic DNA fragments were used as competitors in the pull-down assay. L: DNA

ladder. B. Materials pulled down by Zα (Z) and Zα mut (M) with increasing amounts of B-DNA competitors. Lane 1 was purified total RNA from Hela cells. Lane 2 and 3 were materials pulled down by Zα and Zα mut without B-DNA competitors. Lane 4 and 5, lane 6 and 7, and lane 8 and 9 were materials pulled down by Zα and Zα mut, respectively, with increasing amount of competitor B-DNA, from 2 μg to 6 μg and 20 μg. Lane 10 was the negative control in which no protein was added. Lane 11 was input RRL. With the increasing amount of B-DNA competitor, ribosomes precipitated by Zα mut were decreased; however, precipitation of ribosomes by Zα was not affected. C. Comparable amounts of Zα and Zα mut in corresponding pull-down materials checked by Western blotting assay. + stands for purified Zα.

III.2.3.2. Zα binds to both 40S and 60S ribosomal subunits

In order to make clear which ribosomal subunit was recognized by Zα, separated 40S and 60S subunits from Hela cells were prepared and used as the binding substrates of Zα. As shown in **Figure III-24**, both 40S and 60S ribosomal subunits could be recognized and pulled down by Zα.

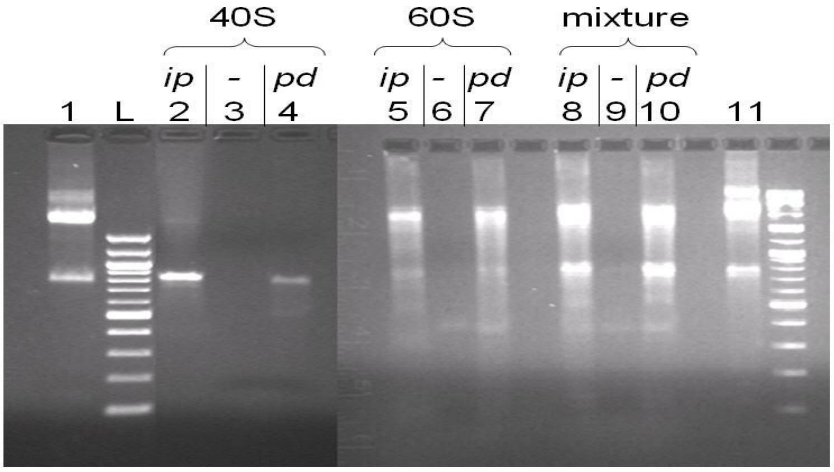


Figure III-24. Pull-down assay of separated mammalian ribosomal subunits by Zα.
ip: input material; -: negative control without protein; *pd*: pull-down material; Lane 1 and 11 were purified total RNA from Hela cells; lane 2 to 4 were pull-down assays of 40S subunits; lane 5-7 were pull-down assays of 60S subunits; lane 8-10 were pull-down assays of the mixture of separated 40S and 60S. These results demonstrated that both 40S and 60S subunits could be recognized and precipitated by Zα.

III.2.4. Inhibition of *in vitro* mammalian translation by *Zα*

The influence of Zα on translation was investigated using Flexi® RRL translation system (Promega), taking luciferase as the reporter. The amount of translated luciferase was quantified by reaction with its substrates, expressed as arbitrary Light Unit and plotted on Y axis, against various protein concentrations on X-axis. As shown in **Figure III-25**, the inhibition effect of Zα on translation was dose-dependent and, at about 17 µM of Zα, there was only about 10% active luciferase, i.e. 90% translation was inhibited by Zα at this concentration. However, at the same concentration, only 30% translation activity was impaired by Zα mut.

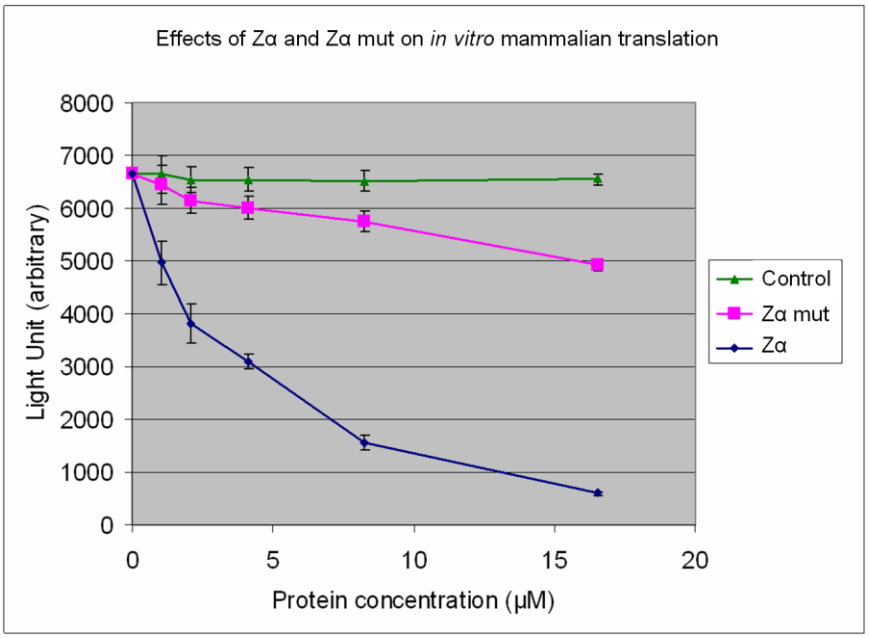


Figure III-25. Effects of *Zα* and *Zα* mut on *in vitro* translation assay.

The reporter luciferase was quantified as Light Unit, plotted on Y-axis. The inhibition effect on translation exhibited by Zα was in a dose-dependent manner and 90% translation was inhibited in the presence of about 17 µM of Zα. However, at this concentration, only 30% translation was impaired by Zα mut. This experiment was repeated successfully with different batch of purified Zα and Zα mut.

III.2.5. Zα has no effect on cell proliferation

Although Zα is able to inhibit *in vitro* mammalian translation efficiently, it was unclear about its influence on cell growth. In this study, the proliferation rates of Hela cells transfected with Zα or Zα mut were examined. Surprisingly, the proliferation rate of cells transfected with Zα had not much difference compared with that of cells transfected with Zα mut and the blank vector, suggesting that Zα did not influence cell proliferation.

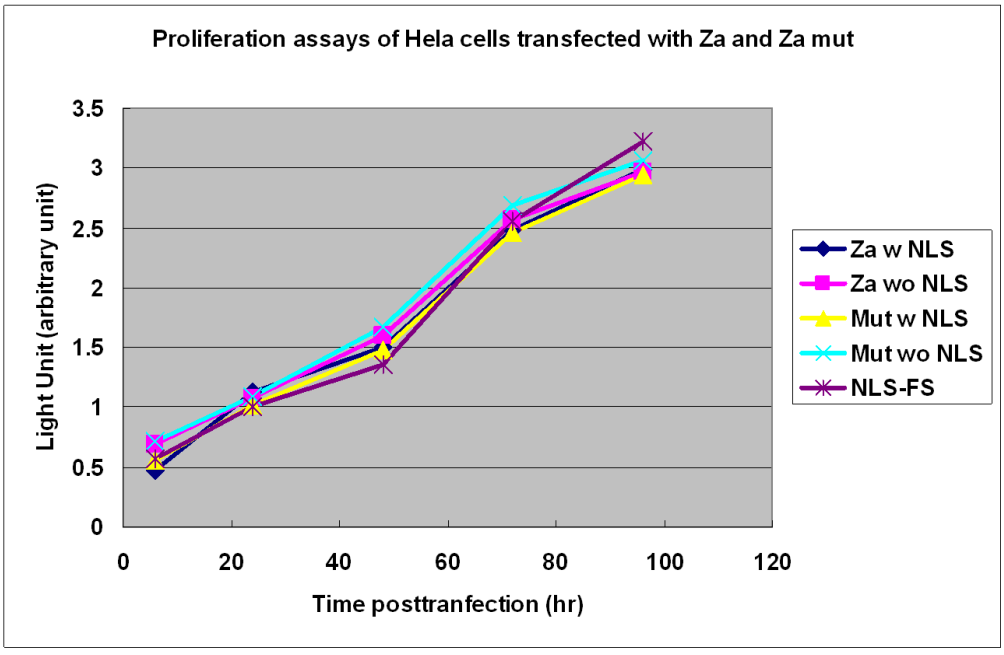


Figure III-26. Cell proliferation assay of transiently transfected Hela cells.

Hela cells were transfected with Zα and Zα mut expression vectors with and without NLS. Blank vector carrying NLS-FS was used as the negative control. In order to compare the effect of different protein constructs on cell proliferation, the readings of each assay were normalized to the first reading. In this 96-hr assay, proliferation curves of Zα and Zα mut transfected cells were very similar.

III.2.6. Zα was translocated from the ER to the nucleoli

In order to trace the transport route of Zα to the nucleoli to understand the discordant results from the *in vitro* translation assay and the cell proliferation assay, Hela cells transfected with Zα-eGFP fusion protein was examined at 6 h and 24 h posttransfection, using eGFP as a negative control. R6 (rhodamine B, hexyl ester, perchlorate) was used to stain ER to clarify whether Zα binds to ribosomes on ER. Surprisingly, at 6 h posttransfection, Zα-eGFP accumulated around the nuclei and in the cytoplasm, overlapping with R6 stained ER (**Figure III-27**). However, after further incubation to 24 h posttransfection, Zα-eGFP was found to accumulate in the nucleoli, in agreement with the finding with immunostaining of Hela cells transfected with Zα (**Figure III-6**). This observation indicated that Zα was removed from the functional ribosomes on ER by certain unknown mechanism.

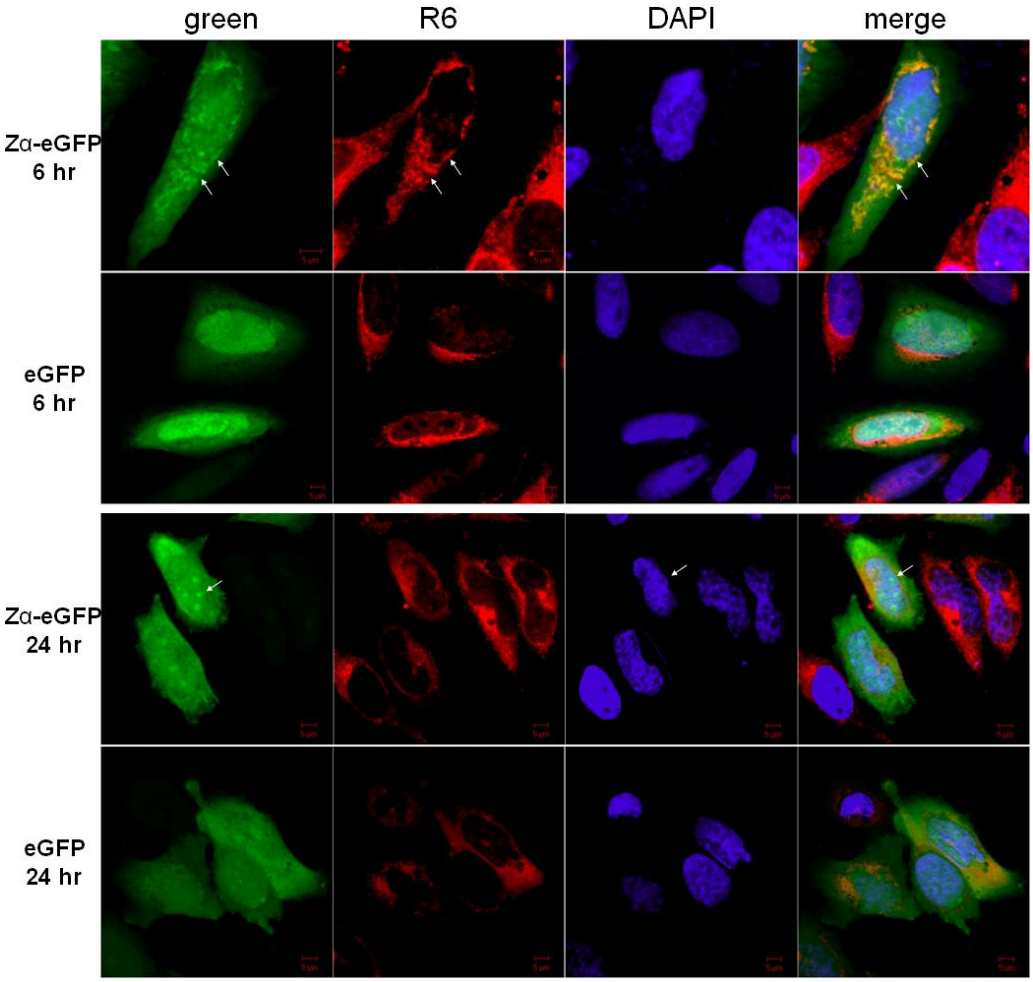


Figure III-27. Zα-eGFP was translocated from the ER to the nuclei.

At 6 h posttransfection, Zα-eGFP was found to accumulate around the nuclei and in the cytoplasm (as indicated by white arrows), overlapping with R6 stained ER. The overlapped signals are in yellow in merged picture. However, at 24 h posttransfection, Zα-eGFP was found to be accumulated in the nucleoli (indicated by arrow) and no accumulation of Zα-eGFP was found around the nuclei. These localization patterns were not found in eGFP transfected cells.

IV. DISCUSSION

IV.1. Mapping Z-DNA in the human genome by *Zα*

IV.1.1. *Zα* is a specific probe for Z-conformation

In this project, *Z*α was used as a Z-DNA specific probe to detect Z-DNA and Z-RNA in mammalian cells. Compared to other known Z-DNA binding proteins, including Z-DNA antibodies (Nordheim and Meese, 1988), E3L (Kahmann et al., 2004; Kwon and Rich, 2005), DLM-1(Schwartz et al., 2001), and others (Krishna et al., 1988; Leith et al., 1988), *Z*α is the best characterized Z-DNA binding domain with high affinity to Z-conformation, independent of the sequence, rendering it an excellent Z-conformation probe. Furthermore, our experimental results suggest that the addition of FLAG and StrepII epitope tags to the C-terminus of *Z*α does not compromise its Z-conformation binding affinity and specificity. *Z*α mut is a good control since it does not bind to Z-DNA, due to mutations of only two amino acid residues which are crucial for Z-DNA specific binding.

IV.1.2. *In vitro* chromatin affinity precipitation (ChAP)

Mapping the Z-DNA forming segments over the genomes of *Drosophila*, human being, etc. has attracted much interest (Herbert et al., 1997; Phan et al., 2006). When *Z*α was identified as a Z-DNA binding domain by Prof. Alexander Rich group at MIT, it was proposed that *Z*α could be used as a Z-DNA probe to detect the Z-DNA fragments in the human genome. However, when using *in*

vivo chromatin affinity precipitation (ChAP), it was very astonishing to find that the nucleic acids precipitated by transiently over-expressed Zα were RNA but not DNA fragments, as shown in **Figure III-7**. We postulated that Zα expressed in living cells were exhausted by binding to RNA since the expression level of Zα was moderate and the copy number was estimated as 10^5 copies per cell.

The novel experimental strategy for *in vitro* ChAP proposed in this thesis was based on the fact that formaldehyde treatment did not affect the recognition of Zα to Z-DNA. It has been reported (Liu et al., 2006) that the promoter region of CSF1 gene containing TG repeat could be recognized and digested in cells fixed with 1% formaldehyde by ZaaFOK which is a recombinant restriction enzyme created by fusing the nuclease domain from *Fok* I endonuclease to two copies of Zα (Kim et al., 1997). In the proposed *in vitro* ChAP protocol, A549 cells were treated in a similar way and incubated with equal amounts of purified Zα and Zα mut, thus overcoming competitive binding to RNA. The advantage of *in vitro* ChAP is that it can be exploited to detect Z-DNA sequences in the cells or tissues which do not express Zα. It has been proposed that Z-DNA might be considered as *in situ* transcription markers for certain genes (Cerna et al., 2004), thus *in vitro* ChAP might be used to check the formation of Z-DNA at specific locus and the expression of certain gene in biopsy specimen and even in fixed tissues. This protocol may be applicable to study the distribution of other non-classical DNA conformations in genomes as well, such as H-DNA, G-quadruplex, etc., using the corresponding specific probes.

It might be questioned whether fixation by formaldehyde and treatment with Triton X-100 would induce or reverse the B-to-Z transition in mammalian cells. So far we do not have certain approach to clarify this question. However, there is evidence to support that these treatments did not affect the formation of Z-DNA. Firstly, previous study on *Drosophila* suggested that fixation with formaldehyde did not change the immunostaining pattern detected by anti-Z-DNA antibody (Nordheim et al., 1981), suggesting that fixation by formaldehyde did not induce or reverse B-to-Z transition significantly. Secondly, DNA sequences retrieved from our ChAP experiment contained only very few GT or GC repeats. Because GT-containing microsatellites are abundant in the human genome and undergo transition to Z-conformation upon negative supercoiling, they would be recognized and precipitated by Zα if negative supercoiling was produced by removal of histone with Triton X-100. Thirdly, reproducibility of the ChAP results (according to ChAP-PCR) suggests that binding and crosslinking of *Zα* to genomic DNA in *in vitro* ChAP were occurred selectively but not randomly.

IV.1.3. Cloning and sequencing strategy

Initially, *in vitro* ChAP DNA fragments were cloned directly to TOPO vector (Invitrogen) after adding overhang adenosine triphosphate residues at 3' ends by Taq DNA polymerase, resulting in a very low number of colonies after transformation. The possible reasons include low copy number of ChAP DNA fragments and potential chemical modifications on ChAP DNA fragments

which might be obstacles for plasmid replication in *E. coli*. Subsequently, the ChAP DNA fragments were amplified in the presence of Q solution (Qiagen) before cloning, to improve amplification of GC-rich segments. Since all fragments were amplified under the same conditions with the same primer, it was supposed that they were amplified at the same rate. Afterwards, routine cloning and sequencing strategies were exploited. In later bioinformatics analyses, DNA fragments with identical sequences were removed since they might be produced by this PCR amplification.

It has been realized for a long time that Z-DNA forming sequences in plasmids are unstable during replication in *E. coli* (Klysik et al., 1981; Freund et al., 1989). Although Stbl4 competent cells are modified genetically to accommodate instable and repeating sequences, it is still possible that ChAP DNA fragments could be mutated or deleted during replication, which would affect later alignments and analyses.

In this study, only about 10,000 colonies were sequenced and analyzed, which is not enough for a in-depth genomic study. Further studies with more samples are needed; thus traditional cloning and sequencing strategies will not be the best choice. An up-to-date sequencing technology, i.e. pyrosequencing technology, would be helpful (Margulies et al., 2005). In this novel sequencing strategy, there is no need to clone the DNA fragments into vectors and transform *E.coli*, which would be a great advantage especially for this project.

IV.1.4. Confirmation of Z-DNA hotspots by ChAP-PCR

In order to confirm that these Z-DNA hotspots are authentic binding sites of Zα, two rounds of ChAP-PCR were performed. The first round ChAP-PCRs were used to compare the enrichment of hotspot regions in ChAP materials precipitated by Zα and Zα mut. The results showed that in most cases Zα precipitated more copies of hotspots DNA than Zα mut did. The different precipitation efficiency may result from Y177A mutation in Zα mut since the tyrosine residue is very essential for the specific Z-DNA binding of Zα and is a potential reaction site with formaldehyde. The second round ChAP-PCRs were performed to compare the enrichments of hotspot regions and non-hotspot regions in Zα *in vitro* ChAP material. After standardization, the enrichments of hotspot regions were compared with those of non-hotspot regions. The results showed that Zα has precipitated more copies of hotspots DNA than non-hotspots DNA. Together, these results suggest that the *in vitro* ChAP experiment was repeatable and reliable, and the marked hotspots in A549 cells were in Z-conformation consistently in two independent experiments. These PCR results strongly testified the specificity of Zα in *in vitro* ChAP and the authenticity of these Z-DNA hotspots.

IV.1.5. The first map of Z-DNA hotspots in the human genome

In this study, we tried to map the Z-DNA fragments in the human genome using an experimental approach for the first time since previous studis of genome-wide distribution of Z-DNA were mainly performed using

bioinformatics analyses. Totally 350 hotspots were initially identified. In order to make a precise map of Z-DNA hotspots over the human genome, 164 redundant or dubious hotspots (**Figure III-15**) were removed from the final list and 122 unique and 64 high potential Z-DNA hotspots were presented in the preliminary draft. However, it should be kept in mind that these so-called dubious hotspots are still potential Z-DNA forming sites in the human genome, as explained in **Figure III-15**. The map of Z-DNA hotspots presented in this study is definitely incomplete. In future studies, if high-throughput technologies were exploited, such as genome-wide DNA microarray or pyrosequencing technology, more Z-DNA hotspots would be marked.

Formation of Z-DNA is driven by negative supercoiling which can be generated by the active transcription machinery or removal of histone core particles. Moreover, unrestrained torsional tension exists in living cells as well (Zheng et al., 1991). It implies that formation of Z-DNA at a specific genome locus is probably cell-type specific, depending on its transcriptional status and local chromatin structure. So this preliminary map might be limited to the A549 cells used here. Additionally, cell cycle phases determine the transcriptional activities and chromatin arrangement, thus affecting the formation of Z-DNA. In our study, bulks of non-synchronized A549 cells were used to cover all phases of the cell cycle, since there is no data available so far about cell cycle specific Z-DNA formation. Future studies of Z-DNA formation in specific phases will be necessary.

IV.1.6. Z-DNA hotspots and the regulation of transcription

Although previous bioinformatics analyses point towards that Z-DNA forming sequences are enriched in transcription start sites (TSSs) (Khuu et al., 2007), in this study no correlation was identified between Z-DNA forming hotspots and transcription start sites (TSS) or translation start sites (TLS). 66 Z-DNA hotspots were located in transcribed regions and most of them (49 out of 66) were located in introns. It is possible that formation of Z-DNA in introns is involved in regulation of transcription as reported (Wittig et al., 1992; Wolfl et al., 1996). It was postulated that the formation of Z-DNA in the first intron in c-*myc* gene (Wittig et al., 1992) and corticotrophin-releasing hormone (*CRH*) gene (Wolfl et al., 1996) were induced by negative supercoiling generated behind the transcription machinery and, in turn, it could prevent subsequent transcription event since RNA polymerase could not pass through the unusual DNA conformation (Peck and Wang, 1985). Further analysis suggested that these Z-DNA hotspots were not enriched in the first intron of each transcript. Investigation on the expression level of 31 transcripts containing Z-DNA forming hotspots reveals that formation of Z-DNA is not correlated with the high expression level (3 out of 31 are highly transcribed, 7 are moderately transcribed, and 21 are transcribed at low level; based on NCBI GEO profile GSM94306). It might be possible that the influence of Z-DNA formation on transcription depends on the surrounding chromatin structure. Since most (21 out 31) hotspots are transcribed at a low level, energy required for flipping from B- to Z- conformation might not be provided by transcription machinery,

probably by unwrapping the nucleosomes (Rich and Zhang, 2003).

IV.1.7. Z-DNA may contribute to centromere evolution

In this study we found that Z-DNA hotspots were enriched in centromeres. Analyses of sequences of hotspots via our own Z-DNA forming prediction program (Z-catcher) showed that only at $\sigma < -0.08$ most of these hotspots (37 out of 46) could be flipped into Z-conformation, suggesting that, in centromeric regions, there must be strong negative supercoiling in surrounding micro-environments which drives these hotspots to flip into Z-conformation. The mechanism of producing strong negative supercoiling in centromere is unknown, probably due to DNA unwrapping of nucleosomal core particles. Formation of Z-DNA may contribute to the irregular nucleosome positioning in functional centromeres (Marschall and Clarke, 1995) since Z-DNA can not be compacted into nucleosomal core particles (Garner and Felsenfeld, 1987; Wong et al., 2007) and CENP-A-containing (a histone H3 variant) nucleosomes which are in centromere exclusively are in a more rigid conformation (Black et al., 2007).

Centromeres are spindle attachment sites at mitosis and meiosis and mediate proper separation and segregation of sister chromatids; however, centromeric DNA is extremely diverse from 125 bp “point” centromere in *Saccharomyces cerevisiae* to highly repetitive satellite DNA in vertebrates. Centromeric DNA repeats are most rapidly evolving sequences in eukaryotic genomes (Henikoff et al., 2001), i.e. these sequence variants are fixed by expansion and contraction

and can occur *de novo* at new sites. It has been proposed (Henikoff et al., 2001) that expansion or contraction of alpha-satellite repeats might be an advantage in female asymmetric meiosis and thus be fixed. Here we proposed that genetic instabilities induced by Z-DNA (Bacolla et al., 2004; Wang and Vasquez, 2006), such as gross deletions and double strand breaks, and gene conversion promoted by Z-DNA (Wahls et al., 1990) might provide a driving force for rapid evolution of centromeres.

IV.1.8. Clusters of special chromatin structures may be epigenetic marks of centromeres

In primates, all centromeres are characterized by an accumulation of repetitive alpha satellite DNA, and in human, there is a consensus sequence of alpha satellite DNA in all chromosomes. Therefore, a tight correlation between DNA sequences and centromeres was expected. However, the discovery of neocentromeres (Marshall et al., 2008) on arm regions, which contains all known critical centromeric proteins and functions identically in mitosis and meiosis as its satellite-based counterpart, suggests that alpha satellite DNA is not necessary for a functional centromere and the position of centromere on a chromosome is determined epigenetically. CENP-A protein, which is a histone H3 variant (Earnshaw and Rothfield, 1985; Sullivan et al., 1994) and present in centromere nucleosomes exclusively (Smith, 2002), is thought to be an epigenetic marker for centromere position. Here we found that Z-DNA hotspots were enriched in several centromeres, and more significantly, these hotspots in centromeres are clustered in several particular regions, suggesting that in these

regions strong negative strain stresses are maintained. It has been reported that S/MARs (scaffold/matrix attachment site) are more abundant in centromeres or neocentromeres than in arm regions (Sumer et al., 2003), it is possible that these S/MARs demark these regions in strong strain stress and prevents expansion of the negative supercoiling which would drive these regions into Z-conformation. The clustering of the special chromatin structures might be an epigenetic marker of a centromere. Thus, there are two different epigenetic markers for centromere: one is protein CENP-A which could be incorporated exclusively into centromeric nucleosomes; another one is the cluster of special chromatin conformations with strong negative superhelical stress.

IV.1.9. Z-DNA forming hotspots and high occurrences of SNPs

syn conformation of purines in Z-DNA and extrusion of nucleotides at B-Z junction make the nucleotides in Z-DNA and at B-Z junctions are more accessible to modifications and genetic toxins, which may introduce nucleotide mutations in the genome. There was a strong correlation between high occurrences of SNPs and Z-DNA forming hotspots in centromeres, but not in arm regions, implying that Z-DNA plays a crucial role in the accumulation of SNPs in centromeres during evolution. Since centromeres are dedicated to the attachment of kinetochore to separate sister chromatids properly, its structure in both mitosis and meiosis should be conserved (Sullivan et al., 2001), leading to constitutive Z-DNA forming at specific loci where nucleotides are more prone to be mutated and SNPs are accumulated over generations. However, this explanation is not applicable for Z-DNA hotspots in arm regions where the

chromatin conformations are unnecessarily conserved between germline cells and somatic cells. So the Z-DNA hotspots detected in arm regions here (in A549 cells) are facultative, thus the mutations caused by Z-DNA or negative superhelical stress are not propagated and accumulated over generations.

IV.1.10. Comparison with known Z-DNA forming sites

Z-DNA fragments have been detected in regulatory regions of several genes in the human genome (Wittig et al., 1992; Muller et al., 1996; Wolfl et al., 1996; Kwon and Rich, 2005). These Z-DNA fragments were mapped in particular cell lines (c-*myc* in U937, β-globin gene cluster in K562, and corticotrophin-releasing hormone gene in NPLC) and involved in regulation of gene transcription. It was not surprising to find that these Z-DNA fragments were not included in our map, since these three genes were transcribed at high level in a cell-type specific manner.

IV.2. Detection of natural Z-RNA in mammalian cells by Zα

IV.2.1. Binding of Zα to the ribosome is conformation-dependent

We showed that Zα can bind to Z-RNA *in vivo* in *E.coli* and in mammalian cells. So far, most studies of Z-RNA are focused on its structure and no naturally occurring Z-RNA is identified. Its biological significance has yet to be understood. It has been reported that Z-RNA could be detected by Z-RNA specific antibodies in the cytoplasm of fixed cells (Zarling et al., 1987). Since Zα can bind to both Z-DNA and Z-RNA *in vitro* (Brown et al., 2000) and the binding patterns of Zα to Z-DNA and Z-RNA in crystals are quite similar (Placido et al., 2007), it seemed to be a good candidate to detect native Z-RNA in mammalian cells by using Zα as a specific probe, with Zα mut as a negative control.

Zα could bind to functional mammalian ribosomes in RRL specifically, which could not be competed by non-specific B-DNA competitors, whereas Zα mut behaved in an opposite way, suggesting that the binding of Zα to mammalian ribosomes is non-B conformation-dependent, probably Z-conformation dependent. A further pull-down assay of separated ribosomal subunits revealed that Zα could bind to both the 40S and the 60S ribosomal subunits. Consistent with the results of *in vitro* pull-down assays, immunostaining of Zα and Zα mut showed that at 24 h posttransfection Zα was, but Zα mut was not, accumulated in the nucleoli, where ribosomal subunits are produced and assembled. This

finding is also in agreement with a previous report that anti-Z-RNA antibody can recognize nucleolar rRNA specifically (Zarling et al., 1990). More interestingly, Zα-eGFP localized on ER in transfected cells at 6 h posttransfection. Together, these results suggest that Zα binds to ribosomes specifically *in vitro* and *in vivo*.

IV.2.2. Inhibition of *in vitro* translation by Zα and the *in vivo* consequence

The specific binding of Zα to ribosomes has strong inhibition effect on *in vitro* eukaryotic translation (**Figure III-25**), probably due to the specific binding to both the large and the small ribosomal subunits and the resulting spatial hindrances. Parallel studies on *E.coli in vitro* translation suggested that Zα could also inhibit the translation efficiently (data not shown). One implication of the inhibition mechanism came from mapping the binding sites of Zα on both mammalian and *E.coli* ribosomes (unpublished data, done by Miss Feng Shu). One Zα binding site was found on each mammalian and *E.coli* ribosomal subunit, and interestingly, the binding site on the large subunit of *E.coli* is close to the exit tunnel, thus the translation activity is impaired probably by interfering with the exit of nascent polypeptides. Since the crystal structure of mammalian ribosome is still unavailable, the spatial positions of Zα binding sites on mammalian ribosome are unknown and the mechanism(s) of inhibition of Zα on mammalian translation is still unclear.

We expected to observe an inhibitory effect of Zα on cell proliferation since

protein synthesis is critical for cell growth and division. And it was reported that cytoplasmic microinjection of specific antibodies against Z-RNA could inhibit cell growth (Zarling et al., 1990). However, transient expression of Zα in mammalian cells did not show impact on cell proliferation (**Figure III-26**). It was noticed that after 24 h posttransfection, Zα accumulated in the nucleoli, rather than on the ER or in the cytoplasm where proteins were translated by ribosomes. In order to trace the transport route of Zα, the localization of Zα-eGFP at 6 h and 24 h posttransfection was examined. Interestingly, Zα-eGFP was found on the ER at 6 h and in the nucleoli at 24 h, suggesting that Zα bound to ribosomes on the ER first and then was removed very quickly and imported to the nucleoli. Thus the potential influence of Zα on *in vivo* translation is transient, which is a reasonale explanation of the discrepancy between results from *in vitro* translation assay and *in vivo* proliferation assay. Since in this study only R6 was used to detect ER in the cytoplasm, and no special probe was used to detect the free ribosomes in the cytosol, the possibility that Zα-eGFP is localized on the free ribosomes in the cytosol can not be excluded. Higher expression level of Zα (using expression vectors carrying Kozak sequences) did also not influence proliferation rates of transfected cells (data not shown), indicating that the removal of Zα from ribosomes on the ER is quite efficient and complete. The mechanism of Zα migration from ER to nucleoli is unknown. It is possible that there is one NLoS (nucleolus localization signal) in Zα.

Although Zα did not exhibit any influence on cell proliferation, our efforts to

establish stable cell lines expressing Zα, either constantly or regulated, failed. We tried several strategies to establish stable cell lines expressing Zα, including RheoSwitch Mammalian Inducible Expression System (NEB) without success. Analysis of genomic DNA from isolated colonies by PCR suggested that the coding sequence of Zα was integrated into the genome, but the expression of Zα was inactivated by unknown mechanism, perhaps via gene silencing by DNA methlylation.

Because $Z\alpha_{ADAR1}$, $Z\alpha_{E3L}$ and $Z\alpha_{DAI}$ (previously known as $Z\alpha_{DLM1}$) share similar structural features and can bind to Z-DNA specifically, it is interesting to know whether $Z\alpha_{E3L}$ and $Z\alpha_{DAI}$ can recognize ribosomes. It is known that DAI is associated with stress granules and processing bodies, which are both cytoplasmic aggregates of RNA and proteins (Deigendesch et al., 2006). Stress granules contain 48S translational preinitiation complexes. More interestingly, $Z\alpha_{DAI}$ and $Z\alpha_{ADAR1}$ have a common structure-specific recognition core. So it is reasonable to postulate that $Z\alpha_{DAI}$ can recognize ribosomes as well. If it was true, DAI may be involved in translation regulation (inhibition) by interacting with ribosomes through $Z\alpha_{DAI}$, thus protecting macrophages and tumor lining tissues (Fu et al., 1999) (which highly express DAI) from tumor evasion. However, so far there is no report that $Z\alpha_{E3L}$ is associated with ribosome or ribosomal subunits, and Y48 in unbound E3L, which is critical for specific Z-DNA binding, takes a different side chain conformation with $Z\alpha_{ADAR1}$ (Kahmann et al., 2004). Thus, the binding ability of $Z\alpha_{E3L}$ to Z-DNA/Z-RNA is lower than $Z\alpha_{ADAR1}$ and $Z\alpha_{DAI}$. So it is possible that $Z\alpha_{E3L}$ binds to ribosomes

with lower affinity, or even fails to bind ribosomes.

IV.2.3. Implications on understanding the biological function of ADAR1

ADAR1 contains two Z-DNA binding domains at its N-terminus, followed by three double-stranded RNA binding motifs (DSRM), and one catalytic adenosine deaminase motif (ADEAM) at C-terminus (**Figure I-10**). As a unique feature in ADAR1-L, Zα is thought to play roles in specifying the substrates of A-to-I editing since Zα could increase *in vivo* editing efficiency on short RNA hairpin (15 bp) (Herbert and Rich, 2001) and *in vitro* editing on a 50-bp dsRNA substrate which contains a Z-DNA permissive sequence (Koeris et al., 2005). Besides this, a potential nuclear export signal (NES) in Z-DNA binding domain may be involved in the regulation of cellular localization of ADAR1 (Poulsen et al., 2001), thus modulating its editing substrate and activity. It has been observed that ADAR1 is a shuttling protein and can be found in the cytoplasm and the nucleus. Both immunostaining and binding assays of ribosomal subunits of Zα in this study suggest that Zα binds to specific RNA motifs on ribosomes and may specify the substrates of ADAR1.

It is noteworthy that ADAR1-S, which does not contain Z-DNA binding domain, is found in nucleoli and can induce nuclear translation which is independent of deaminase activity (Herbert et al., 2002); Also the translation induction is dependent on its nucleoli localization and is thought to be associated with interactions with translational machinery. However, ADAR1-L with Z-DNA

binding domain is not observed in nucleoli and has no induction activity of nuclear translation. In addition, it has been reported that ADAR2 (Sansam et al., 2003), which contains two dsRNA binding motifs and a similar catalytic domain but does not have Z-DNA binding domains, also localizes to the nucleoli but has less catalytic activity there. It was postulated that the sequestration of ADAR2 from the nucleoplasm by enrichment in the nucleoli was an efficient means to modulate its editing activity from transcribed pre-mRNA. It is of interest that both induction of nuclear translation by ADAR1-S and sequestration of ADAR2 in the nucleoli are dependent on the interaction of ADAR proteins with ribosomes. Taking these reported findings together with our results from this study, it is possible that in ADAR1-L, Zα functions as an auto-inhibitor to separate ADARl-L away from nucleoli by spatial hindrance or competition, similar to the function of N-terminus of ADAR2 (Macbeth et al., 2004). Besides this potential auto-inhibitory effect, one nuclear export signal (which is not included in the Zα probe used in this study) of Zα mediates the export of ADAR1-L to the cytoplasm (Poulsen et al., 2001). So Zα can specify the editing substrates of ADAR1-L by modulating its subcellular localization to the cytoplasm, since it was reported that editing activity of ADAR1-L was more efficient in cytoplasm (Wong et al., 2003).

In this study, we found that Zα could bind specifically to mammalian ribosomes and inhibit translation *in vitro*; however, whether ADAR1-L binds to ribosomes in mammalian cells *in vivo* is unclear. The fact that ADAR1-L can bind to Z-DNA specifically (Herbert et al., 1995) suggests that the Zα domain in

ADAR1-L has the same or very similar 3-D structure and recognition pattern as the Zα in Zα/Z-DNA crystals, so it is reasonable to postulate that ADAR1-L can also bind to ribosomes in living cells if there is no spatial hindrance, just like the Zα probe investigated in this study. So far, there is no report about the binding ability of ADAR1-L to ribosomes. If this assumption is true, ADAR1-L could function as a quick response to IFN response system against viral infection by inhibiting general translation activity and delaying the replication of the virus. When other viral defense mechanisms start to function, such as translation inhibition by phosphorylation of eIF-2α (mediated by protein kinase PKR), the translation inhibition exerted by ADAR1-L will be released by removing it from the ribosomes on ER, similar to the findings with Zα.

V. SUMMARY

With the completion of the human genome project, the nucleotide sequence of the human genome is known and accessible. However, not all DNA in a living cell is in B-conformation and simply packaged in nucleosomes. As the best characterized non-B conformation, Z-DNA was extensively studied, so far mainly *in vitro*. However, the distribution and function(s) of Z-DNA in the human genome are still unknown. In the first part of this study, we exploited Zα from ADAR1 as a specific probe to detect Z-DNA in the human genome. Initially, Z-DNA binding specificity of the designed Zα probe was characterized *in vitro*. Zα probe could bind and be crosslinked to Z-DNA segments in supercoiled plasmids without bias on the sequences. A novel protocol for *in vitro* ChAP was proposed and developed. Using *in vitro* ChAP, DNA fragments were precipitated by Zα, a Z-DNA library was constructed and about 10,000 colonies were sequenced. 122 unique and 64 high potential Z-DNA hotspots were identified by bioinformatics analyses of the sequencing results. The first and preliminary map of Z-DNA hotspots was made based on these data. The potential functions of Z-DNA in mammalian cells were examined. No correlation was found between Z-DNA hotspots and transcription regulation sites. However, Z-DNA hotspots clustered in centromeres. Surprisingly, we found that Z-DNA hotspots co-localized with high occurrences of SNPs in centromeric regions, but not in arms of chromosomes.

In the second part of this thesis, Zα was exploited to detect naturally occurring Z-RNA in mammalian cells. The subcellular localization of Zα was examined and the results showed that Zα localized to the endoplasmic reticulum at 6 h posttransfection and then transported to the nucleoli at 24 h. Pull-down assays suggested that Zα bound to mammalian ribosomes in a conformation-dependent manner. Zα could inhibit *in vitro* translation efficiently by binding to both the 40S and the 60S ribosomal subunits. However, Zα did not show an impact on cell proliferation.

As a conclusion, we made the first preliminary map of Z-DNA on the human genome and detected the existence of Z-RNA on functional mammalian ribosomes for the first time by using the Zα domain from human ADAR1, which has important functional implications.

VI. BIBLIOGRAPHY

Arndt-Jovin, D. J., A. Udvardy, M. M. Garner, S. Ritter and T. M. Jovin (1993). "Z-DNA binding and inhibition by GTP of Drosophila topoisomerase II." Biochemistry **32**(18): 4862-72.353

Bacolla, A., A. Jaworski, J. E. Larson, et al. (2004). "Breakpoints of gross deletions coincide with non-B DNA conformations." Proc Natl Acad Sci U S A **101**(39): 14162-7.2

Bass, B. and H. Weintraub (1988). "An Unwinding Activity That Covalently Modifies Its Double-Stranded RNA Substrate." Cell **55**: 1089-98.488

Bass, B. L. (2002). "RNA editing by adenosine deaminases that act on RNA." Annu Rev Biochem **71**: 817-46.145

Bass, B. L., H. Weintraub, R. Cattaneo and M. A. Billeter (1989). "Biased hypermutation of viral RNA genomes could be due to unwinding/modification of double-stranded RNA." Cell **56**(3): 331.260

Beattie, E., E. B. Kauffman, H. Martinez, M. E. Perkus, B. L. Jacobs, E. Paoletti and J. Tartaglia (1996). "Host-range restriction of vaccinia virus E3L-specific deletion mutants." Virus Genes **12**(1): 89-94.86

Behe, M. and G. Felsenfeld (1981). "Effects of methylation on a synthetic polynucleotide: the B--Z transition in poly(dG-m5dC).poly(dG-m5dC)." Proc Natl Acad Sci U S A **78**(3): 1619-23.984

Berger, I., W. Winston, R. Manoharan, et al. (1998). "Spectroscopic characterization of a DNA-binding domain, Z alpha, from the editing enzyme, dsRNA adenosine deaminase: evidence for left-handed Z-DNA in the Z alpha-DNA complex." Biochemistry **37**(38): 13313-21.202

Bhende, P. M., W. T. Seaman, H. J. Delecluse and S. C. Kenney (2005). "BZLF1 activation of the methylated form of the BRLF1 immediate-early promoter is regulated by BZLF1 residue 186." J Virol **79**(12): 7338-48.52

Black, B. E., M. A. Brock, S. Bedard, V. L. Woods, Jr. and D. W. Cleveland (2007). "An epigenetic mark generated by the incorporation of CENP-A into centromeric nucleosomes." Proc Natl Acad Sci U S A **104**(12): 5008-13.38

Blaho, J. A. and R. D. Wells (1987). "Left-handed Z-DNA binding by the recA protein of Escherichia coli." J Biol Chem **262**(13): 6082-8.677

Blaho, J. A. and R. D. Wells (1989). "Left-handed Z-DNA and genetic recombination." Prog Nucleic Acid Res Mol Biol **37**: 107-26.580

Blobel, G. and D. Sabatini (1971). "Dissociation of mammalian polyribosomes into subunits by puromycin." Proc Natl Acad Sci U S A **68**(2): 390-4.3

Brandt, T. A. and B. L. Jacobs (2001). "Both carboxy- and amino-terminal domains of the vaccinia virus interferon resistance gene, E3L, are required for pathogenesis in a mouse model." J Virol **75**(2): 850-6.62

Brown, B. A., 2nd, K. Lowenhaupt, C. M. Wilbert, E. B. Hanlon and A. Rich (2000). "The zalpha domain of the editing enzyme dsRNA adenosine deaminase binds left-handed Z-RNA as well as Z-DNA." Proc Natl Acad Sci U S A **97**(25): 13532-6.8

Brown, B. A., 2nd and A. Rich (2001). "The left-handed double helical nucleic acids." Acta Biochim Pol **48**(2): 295-312.90

Burns, C. M., H. Chu, S. M. Rueter, L. K. Hutchinson, H. Canton, E. Sanders-Bush and R. B. Emeson (1997). "Regulation of serotonin-2C receptor G-protein coupling by RNA editing." Nature **387**(6630): 303-8.223

Cattaneo, R., K. Kaelin, K. Baczko and M. A. Billeter (1989). "Measles virus editing provides an additional cysteine-rich protein." Cell **56**(5): 759-64.261

Cerna, A., A. Cuadrado, N. Jouve, S. M. Diaz de la Espina and C. De la Torre (2004). "Z-DNA, a new in situ marker for transcription." Eur J Histochem **48**(1): 49-56.80

Champ, P. C., S. Maurice, J. M. Vargason, T. Camp and P. S. Ho (2004). "Distributions of Z-DNA and nuclear factor I in human chromosome 22: a model for coupled transcriptional regulation." Nucleic Acids Res **32**(22): 6501-10.10

Chang, H. W. and B. L. Jacobs (1993). "Identification of a conserved motif that is necessary for binding of the vaccinia virus E3L gene products to double-stranded RNA." Virology **194**(2): 537-47.96

Chang, H. W., L. H. Uribe and B. L. Jacobs (1995). "Rescue of vaccinia virus lacking the E3L gene by mutants of E3L." J Virol **69**(10): 6605-8.87

Chang, H. W., J. C. Watson and B. L. Jacobs (1992). "The E3L gene of vaccinia virus encodes an inhibitor of the interferon-induced, double-stranded RNA-dependent protein kinase." Proc Natl Acad Sci U S A **89**(11): 4825-9.99

Chen, C. X., D. S. Cho, Q. Wang, F. Lai, K. C. Carter and K. Nishikura (2000). "A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains." Rna **6**(5): 755-67.180

Crawford, J. L., F. J. Kolpak, A. H. Wang, G. J. Quigley, J. H. van Boom, G. van der Marel and A. Rich (1980). "The tetramer d(CpGpCpG) crystallizes as a left-handed double helix." Proc Natl Acad Sci U S A **77**(7): 4016-20.673

Davies, M. V., H. W. Chang, B. L. Jacobs and R. J. Kaufman (1993). "The E3L and K3L vaccinia virus gene products stimulate translation through inhibition of the double-stranded RNA-dependent protein kinase by different mechanisms." J Virol **67**(3): 1688-92.97

Deigendesch, N., F. Koch-Nolte and S. Rothenburg (2006). "ZBP1 subcellular localization and association with stress granules is controlled by its Z-DNA binding domains." Nucleic Acids Res **34**(18): 5007-20.51

Droge, P. and A. Nordheim (1991). "Transcription-induced conformational change in a topologically closed DNA domain." Nucleic Acids Res **19**(11): 2941-6.2986

Droge, P. and F. M. Pohl (1991). "The influence of an alternate template conformation on elongating phage T7 RNA polymerase." Nucleic Acids Res **19**(19): 5301-6.430

Earnshaw, W. C. and N. Rothfield (1985). "Identification of a family of human centromere proteins using autoimmune sera from patients with scleroderma." Chromosoma **91**(3-4):

313-21.17

Eckmann, C. R., A. Neunteufl, L. Pfaffstetter and M. F. Jantsch (2001). "The human but not the Xenopus RNA-editing enzyme ADAR1 has an atypical nuclear localization signal and displays the characteristics of a shuttling protein." Mol Biol Cell **12**(7): 1911-24.159

Eichman, B. F., G. P. Schroth, B. E. Basham and P. S. Ho (1999). "The intrinsic structure and stability of out-of-alternation base pairs in Z-DNA." Nucleic Acids Res **27**(2): 543-50.198

Ellison, M. J., R. J. Kelleher, 3rd, A. H. Wang, J. F. Habener and A. Rich (1985). "Sequence-dependent energetics of the B-Z transition in supercoiled DNA containing nonalternating purine-pyrimidine sequences." Proc Natl Acad Sci U S A **82**(24): 8320-4.748

Feigon, J., A. H. Wang, G. A. van der Marel, J. H. van Boom and A. Rich (1985). "Z-DNA forms without an alternating purine-pyrimidine sequence in solution." Science **230**(4721): 82-4.756

Fitzgerald, L. W., G. Iyer, D. S. Conklin, et al. (1999). "Messenger RNA editing of the human serotonin 5-HT2C receptor." Neuropsychopharmacology **21**(2 Suppl): 82S-90S.477

Freund, A. M., M. Bichara and R. P. Fuchs (1989). "Z-DNA-forming sequences are spontaneous deletion hot spots." Proc Natl Acad Sci U S A **86**(19): 7465-9.4807

Fu, Y., N. Comella, K. Tognazzi, L. F. Brown, H. F. Dvorak and O. Kocher (1999). "Cloning of DLM-1, a novel gene that is up-regulated in activated macrophages, using RNA differential display." Gene **240**(1): 157-63.6

Fujii, S., K. Matsui, K. Tomita, S. Uesugi and M. Ikehara (1983). "Flexibility and rigidity of left-handed Z-DNA." Nucleic Acids Symp Ser(12): 209-12.614

Fujii, S., A. H. Wang, G. van der Marel, J. H. van Boom and A. Rich (1982). "Molecular structure of (m5 dC-dG)3: the role of the methyl group on 5-methyl cytosine in stabilizing Z-DNA." Nucleic Acids Res **10**(23): 7879-92.944

Gagna, C. E., H. Kuo and W. C. Lambert (1999). "Terminal differentiation and left-handed Z-DNA: a review." Cell Biol Int **23**(1): 1-5.2824

Gagna, C. E., O. G. Mitchell and J. H. Chen (1991). "Fixation and immunolocalization of left-handed Z-DNA sequences in the calf lens." Lens Eye Toxic Res **8**(4): 489-509.468

Garner, M. M. and G. Felsenfeld (1987). "Effect of Z-DNA on nucleosome placement." J Mol Biol **196**(3): 581-90.664

Gueron, M., J. Demaret and M. Filoche (2000). "A unified theory of the B-Z transition of DNA in high and low concentrations of multivalent ions." Biophys J **78**(2): 1070-83.165

Gueron, M. and J. P. Demaret (1992). "A simple explanation of the electrostatics of the B-to-Z transition of DNA." Proc Natl Acad Sci U S A **89**(13): 5740-3.391

Ha, S. C., N. K. Lokanath, D. Van Quyen, et al. (2004). "A poxvirus protein forms a complex with left-handed Z-DNA: crystal structure of a Yatapoxvirus Zalpha bound to DNA." Proc Natl Acad Sci U S A **101**(40): 14367-72.72

Ha, S. C., K. Lowenhaupt, A. Rich, Y. G. Kim and K. K. Kim (2005). "Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases." Nature **437**(7062): 1183-6.31

Hajjar, A. M. and M. L. Linial (1995). "Modification of retroviral RNA by double-stranded RNA adenosine deaminase." J Virol **69**(9): 5878-82.263

Hall, K., P. Cruz, I. Tinoco, Jr., T. M. Jovin and J. H. van de Sande (1984). "'Z-RNA'--a left-handed RNA double helix." Nature **311**(5986): 584-6.536

Hamada, H., M. Seidman, B. H. Howard and C. M. Gorman (1984). "Enhanced gene expression by the poly(dT-dG).poly(dC-dA) sequence." Mol Cell Biol **4**(12): 2622-30.823

Hanahan, D. (1983). "Studies on transformation of Escherichia coli with plasmids." J Mol Biol **166**(4): 557-80.1

Hardin, C. C., D. A. Zarling, J. D. Puglisi, M. O. Trulson, P. W. Davis and I. Tinoco, Jr. (1987). "Stabilization of Z-RNA by chemical bromination and its recognition by anti-Z-DNA antibodies." Biochemistry **26**(16): 5191-9.25

Harvey, S. C. (1983). "DNA structural dynamics: longitudinal breathing as a possible mechanism for the B in equilibrium Z transition." Nucleic Acids Res **11**(14): 4867-78.588

Henikoff, S., K. Ahmad and H. S. Malik (2001). "The centromere paradox: stable inheritance with rapidly evolving DNA." Science **293**(5532): 1098-102.15

Herbert, A., J. Alfken, Y. G. Kim, I. S. Mian, K. Nishikura and A. Rich (1997). "A Z-DNA binding domain present in the human editing enzyme, double-stranded RNA adenosine deaminase." Proc Natl Acad Sci U S A **94**(16): 8421-6.218

Herbert, A., K. Lowenhaupt, J. Spitzner and A. Rich (1995). "Chicken double-stranded RNA adenosine deaminase has apparent specificity for Z-DNA." Proc Natl Acad Sci U S A **92**(16): 7550-4.245

Herbert, A. and A. Rich (2001). "The role of binding domains for dsRNA and Z-DNA in the in vivo editing of minimal substrates by ADAR1." Proc Natl Acad Sci U S A **98**(21): 12132-7.136

Herbert, A., M. Schade, K. Lowenhaupt, et al. (1998). "The Zalpha domain from human ADAR1 binds to the Z-DNA conformer of many different sequences." Nucleic Acids Res **26**(15): 3486-93.205

Herbert, A., S. Wagner and J. A. Nickerson (2002). "Induction of protein translation by ADAR1 within living cell nuclei is not dependent on RNA editing." Mol Cell **10**(5): 1235-46.133

Herbert, A. G., J. R. Spitzner, K. Lowenhaupt and A. Rich (1993). "Z-DNA binding protein from chicken blood nuclei." Proc Natl Acad Sci U S A **90**(8): 3339-42.356

Hill, R. J. and B. D. Stollar (1983). "Dependence of Z-DNA antibody binding to polytene chromosomes on acid fixation and DNA torsional strain." Nature **305**(5932): 338-40.896

Irikura, K. K., B. Tidor, B. R. Brooks and M. Karplus (1985). "Transition from B to Z DNA: contribution of internal fluctuations to the configurational entropy difference." Science **229**(4713): 571-2.768

Jackson, D. A., J. Yuan and P. R. Cook (1988). "A gentle method for preparing cyto- and nucleo-skeletons and associated chromatin." J Cell Sci **90 (Pt 3)**: 365-78.1012

Jan, E. and P. Sarnow (2002). "Factorless ribosome assembly on the internal ribosome entry site of cricket paralysis virus." J Mol Biol **324**(5): 889-902.2

Jaworski, A., N. P. Higgins, R. D. Wells and W. Zacharias (1991). "Topoisomerase mutants and physiological conditions control supercoiling and Z-DNA formation in vivo." J Biol Chem **266**(4): 2576-81.462

Jaworski, A., W. T. Hsieh, J. A. Blaho, J. E. Larson and R. D. Wells (1987). "Left-handed DNA in vivo." Science **238**(4828): 773-7.416

Jiang, H., W. Zacharias and S. Amirhaeri (1991). "Potassium permanganate as an in situ probe for B-Z and Z-Z junctions." Nucleic Acids Res **19**(24): 6943-8.415

Johnston, B. H. (1992). "Generation and detection of Z-DNA." Methods Enzymol **211**: 127-58.414

Kahmann, J. D., D. A. Wecking, V. Putter, et al. (2004). "The solution structure of the N-terminal domain of E3L shows a tyrosine conformation that may explain its reduced affinity to Z-DNA in vitro." Proc Natl Acad Sci U S A **101**(9): 2712-7.661

Kastenholz, M. A., T. U. Schwartz and P. H. Hunenberger (2006). "The transition between the B and Z conformations of DNA investigated by targeted molecular dynamics simulations with explicit solvation." Biophys J **91**(8): 2976-90.13

Kawai, K., H. Sugiyama, K. Fujimoto and I. Saito (1995). "8-Methylguanine-containing oligonucleotides: useful structural constraint for Z form DNA." Nucleic Acids Symp Ser(34): 33-4.287

Kedersha, N., G. Stoecklin, M. Ayodele, et al. (2005). "Stress granules and processing bodies are dynamically linked sites of mRNP remodeling." J Cell Biol **169**(6): 871-84.10

Khuu, P., M. Sandor, J. DeYoung and P. S. Ho (2007). "Phylogenomic analysis of the emergence of GC-rich transcription elements." Proc Natl Acad Sci U S A **104**(42): 16528-33.8

Kim, U., T. L. Garner, T. Sanford, D. Speicher, J. M. Murray and K. Nishikura (1994). "Purification and characterization of double-stranded RNA adenosine deaminase from bovine nuclear extracts." J Biol Chem **269**(18): 13480-9.258

Kim, U., Y. Wang, T. Sanford, Y. Zeng and K. Nishikura (1994). "Molecular cloning of cDNA for double-stranded RNA adenosine deaminase, a candidate enzyme for nuclear RNA editing." Proc Natl Acad Sci U S A **91**(24): 11457-61.482

Kim, Y. G., P. S. Kim, A. Herbert and A. Rich (1997). "Construction of a Z-DNA-specific restriction endonuclease." Proc Natl Acad Sci U S A **94**(24): 12875-9.4457

Kim, Y. G., M. Muralinath, T. Brandt, et al. (2003). "A role for Z-DNA binding in vaccinia virus pathogenesis." Proc Natl Acad Sci U S A **100**(12): 6974-9.108

Klysik, J., S. M. Stirdivant, J. E. Larson, P. A. Hart and R. D. Wells (1981). "Left-handed DNA in restriction fragments and a recombinant plasmid." Nature **290**(5808): 672-7.662

Klysik, J., S. M. Stirdivant and R. D. Wells (1982). "Left-handed DNA. Cloning, characterization, and instability of inserts containing different lengths of (dC-dG) in Escherichia coli." J Biol Chem **257**(17): 10152-8.12

Kobori, J. A., E. Strauss, K. Minard and L. Hood (1986). "Molecular analysis of the hotspot of recombination in the murine major histocompatibility complex." Science **234**(4773): 173-9.1

Kochel, T. J. and R. R. Sinden (1988). "Analysis of trimethylpsoralen photoreactivity to Z-DNA provides a general in vivo assay for Z-DNA: analysis of the hypersensitivity of (GT)n B-Z junctions." Biotechniques **6**(6): 532-43.616

Koeris, M., L. Funke, J. Shrestha, A. Rich and S. Maas (2005). "Modulation of ADAR1 editing activity by Z-RNA in vitro." Nucleic Acids Res **33**(16): 5362-70.4

Krishna, P., B. P. Kennedy, J. H. van de Sande and J. D. McGhee (1988). "Yolk proteins from nematodes, chickens, and frogs bind strongly and preferentially to left-handed Z-DNA." J Biol Chem **263**(35): 19066-70.3078

Krishna, P., B. P. Kennedy, D. M. Waisman, J. H. van de Sande and J. D. McGhee (1990). "Are many Z-DNA binding proteins actually phospholipid-binding proteins?" Proc Natl Acad Sci U S A **87**(4): 1292-5.509

Kumar, M. and G. G. Carmichael (1997). "Nuclear antisense RNA induces extensive adenosine modifications and nuclear retention of target transcripts." Proc Natl Acad Sci U S A **94**(8): 3542-7.265

Kwon, J. A. and A. Rich (2005). "Biological function of the vaccinia virus Z-DNA-binding protein E3L: gene transactivation and antiapoptotic activity in HeLa cells." Proc Natl Acad Sci U S A **102**(36): 12759-64.78

Lafer, E. M., A. Moller, A. Nordheim, B. D. Stollar and A. Rich (1981). "Antibodies specific for left-handed Z-DNA." Proc Natl Acad Sci U S A **78**(6): 3546-50.981

Lafer, E. M., R. Sousa, R. Ali, A. Rich and B. D. Stollar (1986). "The effect of anti-Z-DNA antibodies on the B-DNA-Z-DNA equilibrium." J Biol Chem **261**(14): 6438-43.4955

Lafer, E. M., R. Sousa and A. Rich (1985). "Anti-Z-DNA antibody binding can stabilize Z-DNA in relaxed and linear plasmids under physiological conditions." Embo J **4**(13B): 3655-60.743

Lafer, E. M., R. Sousa, B. Rosen, A. Hsu and A. Rich (1985). "Isolation and characterization of Z-DNA binding proteins from wheat germ." Biochemistry **24**(19): 5070-6.763

Lafer, E. M., R. J. Sousa and A. Rich (1988). "Z-DNA-binding proteins in Escherichia coli purification, generation of monoclonal antibodies and gene isolation." J Mol Biol **203**(2): 511-6.599

Lafer, E. M., R. P. Valle, A. Moller, A. Nordheim, P. H. Schur, A. Rich and B. D. Stollar (1983). "Z-DNA-specific antibodies in human systemic lupus erythematosus." J Clin Invest **71**(2): 314-21.927

Lagravere, C., B. Malfoy, M. Leng and J. Laval (1984). "Ring-opened alkylated guanine is not repaired in Z-DNA." Nature **310**(5980): 798-800.836

Lancillotti, F., M. C. Lopez, P. Arias and C. Alonso (1987). "Z-DNA in transcriptionally active chromosomes." Proc Natl Acad Sci U S A **84**(6): 1560-4.686

Lander, E. S., L. M. Linton, B. Birren, et al. (2001). "Initial sequencing and analysis of the human genome." Nature **409**(6822): 860-921.3

Lang, M. C., B. Malfoy, A. M. Freund, M. Daune and M. Leng (1982). "Visualization of Z sequences in form V of pBR322 by immuno-electron microscopy." Embo J **1**(10): 1149-53.964

Leith, I. R., R. T. Hay and W. C. Russell (1988). "Detection of Z DNA binding proteins in tissue

culture cells." Nucleic Acids Res **16**(17): 8277-89.602

Lipps, H. J., A. Nordheim, E. M. Lafer, D. Ammermann, B. D. Stollar and A. Rich (1983). "Antibodies against Z DNA react with the macronucleus but not the micronucleus of the hypotrichous ciliate stylonychia mytilus." Cell **32**(2): 435-41.928

Liu, H., N. Mulholland, H. Fu and K. Zhao (2006). "Cooperative activity of BRG1 and Z-DNA formation in chromatin remodeling." Mol Cell Biol **26**(7): 2550-9.6

Liu, L. F. and J. C. Wang (1987). "Supercoiling of the DNA template during transcription." Proc Natl Acad Sci U S A **84**(20): 7024-7.4903

Liu, R., H. Liu, X. Chen, M. Kirby, P. O. Brown and K. Zhao (2001). "Regulation of CSF1 promoter by the SWI/SNF-like BAF complex." Cell **106**(3): 309-18.140

Liu, Y., K. C. Wolff, B. L. Jacobs and C. E. Samuel (2001). "Vaccinia virus E3L interferon resistance protein inhibits the interferon-induced adenosine deaminase A-to-I editing activity." Virology **289**(2): 378-87.59

Luo, G. X., M. Chao, S. Y. Hsieh, C. Sureau, K. Nishikura and J. Taylor (1990). "A specific base transition occurs on replicating hepatitis delta virus RNA." J Virol **64**(3): 1021-7.478

Luokkamaki, M., K. Servomaa and T. Rytaomaa (1993). "Onset of chromatin fragmentation in chloroma cell apoptosis is highly sensitive to UV and begins at non-B DNA conformation." Int J Radiat Biol **63**(2): 207-13.595

Lushnikov, A. Y., B. A. Brown, 2nd, E. A. Oussatcheva, V. N. Potaman, R. R. Sinden and Y. L. Lyubchenko (2004). "Interaction of the Zalpha domain of human ADAR1 with a negatively supercoiled plasmid visualized by atomic force microscopy." Nucleic Acids Res **32**(15): 4704-12.74

Maas, S., T. Melcher and P. H. Seeburg (1997). "Mammalian RNA-dependent deaminases and edited mRNAs." Curr Opin Cell Biol **9**(3): 343-9.1

Macbeth, M. R., A. T. Lingam and B. L. Bass (2004). "Evidence for auto-inhibition by the N terminus of hADAR2 and activation by dsRNA binding." Rna **10**(10): 1563-71.91

Majewski, J. and J. Ott (2000). "GT repeats are associated with recombination on human chromosome 22." Genome Res **10**(8): 1108-14.155

Malfoy, B., N. Rousseau, N. Vogt, E. Viegas-Pequignot, B. Dutrillaux and M. Leng (1986). "Nucleotide sequence of an heterochromatic segment recognized by the antibodies to Z-DNA in fixed metaphase chromosomes." Nucleic Acids Res **14**(8): 3197-214.728

Mao, C., W. Sun, Z. Shen and N. C. Seeman (1999). "A nanomechanical device based on the B-Z transition of DNA." Nature **397**(6715): 144-6.192

Marchler-Bauer, A., J. B. Anderson, P. F. Cherukuri, et al. (2005). "CDD: a Conserved Domain Database for protein classification." Nucleic Acids Res **33**(Database issue): D192-6.1021

Margulies, M., M. Egholm, W. E. Altman, et al. (2005). "Genome sequencing in microfabricated high-density picolitre reactors." Nature **437**(7057): 376-80.5268

Marschall, L. G. and L. Clarke (1995). "A novel cis-acting centromeric DNA element affects S. pombe centromeric chromatin structure at a distance." J Cell Biol **128**(4): 445-54.25

Marshall, O. J., A. C. Chueh, L. H. Wong and K. H. Choo (2008). "Neocentromeres: new insights into centromere structure, disease development, and karyotype evolution." Am J Hum Genet **82**(2): 261-82.16

Melcher, T., S. Maas, A. Herb, R. Sprengel, M. Higuchi and P. H. Seeburg (1996). "RED2, a brain-specific member of the RNA-specific adenosine deaminase family." J Biol Chem **271**(50): 31795-8.484

Melcher, T., S. Maas, A. Herb, R. Sprengel, P. H. Seeburg and M. Higuchi (1996). "A mammalian RNA editing enzyme." Nature **379**(6564): 460-4.236

Moller, A., A. Nordheim, S. A. Kozlowski, D. J. Patel and A. Rich (1984). "Bromination stabilizes poly(dG-dC) in the Z-DNA form under low-salt conditions." Biochemistry **23**(1): 54-62.874

Moller, A., A. Nordheim, S. R. Nichols and A. Rich (1981). "7-Methylguanine in poly(dG-dC).poly(dG-dC) facilitates z-DNA formation." Proc Natl Acad Sci U S A **78**(8): 4777-81.979

Muller, V., M. Takeya, S. Brendel, B. Wittig and A. Rich (1996). "Z-DNA-forming sites within the human beta-globin gene cluster." Proc Natl Acad Sci U S A **93**(2): 780-4.8

Murphy, D. G., K. Dimock and C. Y. Kang (1991). "Numerous transitions in human parainfluenza virus 3 RNA recovered from persistently infected cells." Virology **181**(2): 760-3.11

Murphy, K. E. and J. R. Stringer (1986). "RecA independent recombination of poly[d(GT)-d(CA)] in pBR322." Nucleic Acids Res **14**(18): 7325-40.706

Narasimhan, V. and A. M. Bryan (1976). "Temperature-induced perturbations in the circular dichroic spectrum of the synthetic polymer poly[d(G-C)]." Biochim Biophys Acta **435**(4): 433-7.684

Neifakh Iu, A. and V. G. Tumanian (1979). "[Possible DNA conformations]." Biofizika **24**(3): 556-63.678

Nie, Y., L. Ding, P. N. Kao, R. Braun and J. H. Yang (2005). "ADAR1 interacts with NF90 through double-stranded RNA and regulates NF90-mediated gene expression independently of RNA editing." Mol Cell Biol **25**(16): 6956-63.65

Nie, Y., Q. Zhao, Y. Su and J. H. Yang (2004). "Subcellular distribution of ADAR1 isoforms is synergistically determined by three nuclear discrimination signals and a regulatory motif." J Biol Chem **279**(13): 13249-55.104

Nordheim, A., E. M. Lafer, L. J. Peck, J. C. Wang, B. D. Stollar and A. Rich (1982). "Negatively supercoiled plasmids contain left-handed Z-DNA segments as detected by specific antibody binding." Cell **31**(2 Pt 1): 309-18.946

Nordheim, A. and K. Meese (1988). "Topoisomer gel retardation: detection of anti-Z-DNA antibodies bound to Z-DNA within supercoiled DNA minicircles." Nucleic Acids Res **16**(1): 21-37.3072

Nordheim, A., M. L. Pardue, E. M. Lafer, A. Moller, B. D. Stollar and A. Rich (1981). "Antibodies to left-handed Z-DNA bind to interband regions of Drosophila polytene chromosomes." Nature **294**(5840): 417-22.971

Nordheim, A. and A. Rich (1983). "Negatively supercoiled simian virus 40 DNA contains Z-DNA segments within transcriptional enhancer sequences." Nature **303**(5919): 674-9.908

Nordheim, A., P. Tesser, F. Azorin, Y. H. Kwon, A. Moller and A. Rich (1982). "Isolation of Drosophila proteins that bind selectively to left-handed Z-DNA." Proc Natl Acad Sci U S A **79**(24): 7729-33.947

O'Connell, M. A., A. Gerber and W. Keller (1997). "Purification of human double-stranded RNA-specific editase 1 (hRED1) involved in editing of brain glutamate receptor B pre-mRNA." J Biol Chem **272**(1): 473-8.231

O'Connell, M. A., S. Krause, M. Higuchi, J. J. Hsuan, N. F. Totty, A. Jenny and W. Keller (1995). "Cloning of cDNAs encoding mammalian double-stranded RNA-specific adenosine deaminase." Mol Cell Biol **15**(3): 1389-97.250

O'Hara, P. J., S. T. Nichol, F. M. Horodyski and J. J. Holland (1984). "Vesicular stomatitis virus defective interfering particles can contain extensive genomic sequence rearrangements and base substitutions." Cell **36**(4): 915-24.262

Palecek, E., P. Boublikova, K. Nejedly, G. Galazka and J. Klysik (1987). "B-Z junctions in supercoiled pRW751 DNA contain unpaired bases or non-Watson-Crick base pairs." J Biomol Struct Dyn **5**(2): 297-306.420

Palecek, E., E. Rasovska and P. Boublikova (1988). "Probing of DNA polymorphic structure in the cell with osmium tetroxide." Biochem Biophys Res Commun **150**(2): 731-8.640

Palladino, M. J., L. P. Keegan, M. A. O'Connell and R. A. Reenan (2000). "dADAR, a Drosophila double-stranded RNA-specific adenosine deaminase is highly developmentally regulated and is itself a target for RNA editing." Rna **6**(7): 1004-18.138

Palladino, M. J., L. P. Keegan, M. A. O'Connell and R. A. Reenan (2000). "A-to-I pre-mRNA editing in Drosophila is primarily involved in adult nervous system function and integrity." Cell **102**(4): 437-49.134

Patterson, J. B. and C. E. Samuel (1995). "Expression and regulation by interferon of a double-stranded-RNA-specific adenosine deaminase from human cells: evidence for two forms of the deaminase." Mol Cell Biol **15**(10): 5376-88.272

Peck, L. J. and J. C. Wang (1983). "Energetics of B-to-Z transition in DNA." Proc Natl Acad Sci U S A **80**(20): 6206-10.1005

Peck, L. J. and J. C. Wang (1985). "Transcriptional block caused by a negative supercoiling induced structural change in an alternating CG sequence." Cell **40**(1): 129-37.14

Pham, H. T., M. Y. Park, K. K. Kim, Y. G. Kim and J. H. Ahn (2006). "Intracellular localization of human ZBP1: Differential regulation by the Z-DNA binding domain, Zalpha, in splice variants." Biochem Biophys Res Commun **348**(1): 145-52.19

Phan, A. T., V. Kuryavyi and D. J. Patel (2006). "DNA architecture: from G to Z." Curr Opin Struct Biol **16**(3): 288-98.24

Placido, D., B. A. Brown, 2nd, K. Lowenhaupt, A. Rich and A. Athanasiadis (2007). "A left-handed RNA double helix bound by the Zalpha domain of the RNA-editing enzyme ADAR1." Structure **15**(4): 395-404.1

Pohl, F. M. (1983). "Salt-induced transition between two double-helical forms of oligo

(dC-dG)." Cold Spring Harb Symp Quant Biol **47 Pt 1**: 113-7.937

Pohl, F. M. (1987). "Hysteretic behaviour of a Z-DNA-antibody complex." Biophys Chem **26**(2-3): 385-90.676

Pohl, F. M. and T. M. Jovin (1972). "Salt-induced co-operative conformational change of a synthetic DNA: equilibrium and kinetic studies with poly (dG-dC)." J Mol Biol **67**(3): 375-96.7

Popenda, M., J. Milecki and R. W. Adamiak (2004). "High salt solution structure of a left-handed RNA double helix." Nucleic Acids Res **32**(13): 4044-54.76

Poulsen, H., J. Nilsson, C. K. Damgaard, J. Egebjerg and J. Kjems (2001). "CRM1 mediates the export of ADAR1 through a nuclear export signal within the Z-DNA binding domain." Mol Cell Biol **21**(22): 7862-71.153

Rahmouni, A. R. and R. D. Wells (1989). "Stabilization of Z DNA in vivo by localized supercoiling." Science **246**(4928): 358-63.535

Reynolds, D. S., D. S. Gurley, K. F. Austen and W. E. Serafin (1991). "Cloning of the cDNA and gene of mouse mast cell protease-6. Transcription by progenitor mast cells and mast cells of the connective tissue subclass." J Biol Chem **266**(6): 3847-53.1

Rich, A., A. Nordheim and F. Azorin (1983). "Stabilization and detection of natural left-handed Z-DNA." J Biomol Struct Dyn **1**(1): 1-19.895

Rich, A. and S. Zhang (2003). "Timeline: Z-DNA: the long road to biological function." Nat Rev Genet **4**(7): 566-72.103

Rodolfo, C., A. Lanza, S. Tornaletti, G. Fronza and A. M. Pedrini (1994). "The ultimate carcinogen of 4-nitroquinoline 1-oxide does not react with Z-DNA and hyperreacts with B-Z junctions." Nucleic Acids Res **22**(3): 314-20.227

Rohner, K. J., R. Hobi and C. C. Kuenzle (1990). "Z-DNA-binding proteins. Identification critically depends on the proper choice of ligands." J Biol Chem **265**(31): 19112-5.475

Rothenburg, S., N. Deigendesch, K. Dittmar, F. Koch-Nolte, F. Haag, K. Lowenhaupt and A. Rich (2005). "A PKR-like eukaryotic initiation factor 2alpha kinase from zebrafish contains Z-DNA binding domains instead of dsRNA binding domains." Proc Natl Acad Sci U S A **102**(5): 1602-7.63

Rothenburg, S., F. Koch-Nolte and F. Haag (2001). "DNA methylation and Z-DNA formation as mediators of quantitative differences in the expression of alleles." Immunol Rev **184**: 286-98.124

Rothenburg, S., F. Koch-Nolte, A. Rich and F. Haag (2001). "A polymorphic dinucleotide repeat in the rat nucleolin gene forms Z-DNA and inhibits promoter activity." Proc Natl Acad Sci U S A **98**(16): 8985-90.142

Rothenburg, S., T. Schwartz, F. Koch-Nolte and F. Haag (2002). "Complex regulation of the human gene for the Z-DNA binding protein DLM-1." Nucleic Acids Res **30**(4): 993-1000.3

Rueda, P., B. Garcia-Barreno and J. A. Melero (1994). "Loss of conserved cysteine residues in the attachment (G) glycoprotein of two human respiratory syncytial virus escape mutants that contain multiple A-G substitutions (hypermutations)." Virology **198**(2): 653-62.264

Saenger, W. and U. Heinemann (1989). "Raison d'etre and structural model for the B-Z

transition of poly d(G-C).poly d(G-C)." FEBS Lett **257**(2): 223-7.341

Sansam, C. L., K. S. Wells and R. B. Emeson (2003). "Modulation of RNA editing by functional nucleolar sequestration of ADAR2." Proc Natl Acad Sci U S A **100**(24): 14018-23.31

Santoro, C. and F. Costanzo (1983). "Stretches of alternating poly(T-dG), with the capacity to form Z-DNA, are present in human liver transcripts." FEBS Lett **155**(1): 69-72.913

Santoro, C., F. Costanzo and G. Ciliberto (1984). "Inhibition of eukaryotic tRNA transcription by potential Z-DNA sequences." Embo J **3**(7): 1553-9.842

Schade, M., C. J. Turner, R. Kuhne, et al. (1999). "The solution structure of the Zalpha domain of the human RNA editing enzyme ADAR1 reveals a prepositioned binding surface for Z-DNA." Proc Natl Acad Sci U S A **96**(22): 12465-70.173

Schroth, G. P., P. J. Chou and P. S. Ho (1992). "Mapping Z-DNA in the human genome. Computer-aided mapping reveals a nonrandom distribution of potential Z-DNA-forming sequences in human genes." J Biol Chem **267**(17): 11846-55.393

Schwartz, T., J. Behlke, K. Lowenhaupt, U. Heinemann and A. Rich (2001). "Structure of the DLM-1-Z-DNA complex reveals a conserved family of Z-DNA-binding proteins." Nat Struct Biol **8**(9): 761-5.4

Schwartz, T., M. A. Rould, K. Lowenhaupt, A. Herbert and A. Rich (1999). "Crystal structure of the Zalpha domain of the human editing enzyme ADAR1 bound to left-handed Z-DNA." Science **284**(5421): 1841-5.180

Sehgal, V. and R. Ali (1990). "Naturally occurring anti-native DNA antibodies in SLE binding Z-DNA." Indian J Med Res **92**: 158-62.495

Shors, T., K. V. Kibler, K. B. Perkins, R. Seidler-Wulff, M. P. Banaszak and B. L. Jacobs (1997). "Complementation of vaccinia virus deleted of the E3L gene by mutants of E3L." Virology **239**(2): 269-76.79

Sinden, R. R. (1994). "DNA Structure and Function." Academic Press.1013

Singleton, C. K., M. W. Kilpatrick and R. D. Wells (1984). "S1 nuclease recognizes DNA conformational junctions between left-handed helical (dT-dG n. dC-dA)n and contiguous right-handed sequences." J Biol Chem **259**(3): 1963-7.560

Singleton, C. K., J. Klysik, S. M. Stirdivant and R. D. Wells (1982). "Left-handed Z-DNA is induced by supercoiling in physiological ionic conditions." Nature **299**(5881): 312-6.634

Singleton, C. K., J. Klysik and R. D. Wells (1983). "Conformational flexibility of junctions between contiguous B- and Z-DNAs in supercoiled plasmids." Proc Natl Acad Sci U S A **80**(9): 2447-51.597

Slavov, D., T. Crnogorac-Jurcevic, M. Clark and K. Gardiner (2000). "Comparative analysis of the DRADA A-to-I RNA editing gene from mammals, pufferfish and zebrafish." Gene **250**(1-2): 53-60.266

Smith, M. M. (2002). "Centromeres and variant histones: what, where, when and why?" Curr Opin Cell Biol **14**(3): 279-85.19

Sponer, J., H. A. Gabb, J. Leszczynski and P. Hobza (1997). "Base-base and deoxyribose-base stacking interactions in B-DNA and Z-DNA: a quantum-chemical study." Biophys J **73**(1):

76-87.223

Strehblow, A., M. Hallegger and M. F. Jantsch (2002). "Nucleocytoplasmic distribution of human RNA-editing enzyme ADAR1 is modulated by double-stranded RNA-binding domains, a leucine-rich export signal, and a putative dimerization domain." Mol Biol Cell **13**(11): 3822-35.135

Sullivan, B. A., M. D. Blower and G. H. Karpen (2001). "Determining centromere identity: cyclical stories and forking paths." Nat Rev Genet **2**(8): 584-96.39

Sullivan, K. F., M. Hechenberger and K. Masri (1994). "Human CENP-A contains a histone H3 related histone fold domain that is required for targeting to the centromere." J Cell Biol **127**(3): 581-92.18

Sumer, H., J. M. Craig, M. Sibson and K. H. Choo (2003). "A rapid method of genomic array analysis of scaffold/matrix attachment regions (S/MARs) identifies a 2.5-Mb region of enhanced scaffold/matrix attachment at a human neocentromere." Genome Res **13**(7): 1737-43.829

Takaoka, A., Z. Wang, M. K. Choi, et al. (2007). "DAI (DLM-1/ZBP1) is a cytosolic DNA sensor and an activator of innate immune response." Nature **448**(7152): 501-5.1

Tang, N., J. G. Muller, C. J. Burrows and S. E. Rokita (1999). "Nickel and cobalt reagents promote selective oxidation of Z-DNA." Biochemistry **38**(50): 16648-54.167

Tashiro, R. and H. Sugiyama (2003). "A nanothermometer based on the different pi stackings of B- and Z-DNA." Angew Chem Int Ed Engl **42**(48): 6018-20.88

Tashiro, R. and H. Sugiyama (2005). "Biomolecule-based switching devices that respond inversely to thermal stimuli." J Am Chem Soc **127**(7): 2094-7.1009

Thamann, T. J., R. C. Lord, A. H. Wang and A. Rich (1981). "The high salt form of poly(dG-dC).poly(dG-dC) is left-handed Z-DNA: Raman spectra of crystals and solutions." Nucleic Acids Res **9**(20): 5443-57.975

Thomae, R., S. Beck and F. M. Pohl (1983). "Isolation of Z-DNA-containing plasmids." Proc Natl Acad Sci U S A **80**(18): 5550-3.897

Toth, A. M., P. Zhang, S. Das, C. X. George and C. E. Samuel (2006). "Interferon action and the double-stranded RNA-dependent enzymes ADAR1 adenosine deaminase and PKR protein kinase." Prog Nucleic Acid Res Mol Biol **81**: 369-434.27

Treco, D. and N. Arnheim (1986). "The evolutionarily conserved repetitive sequence d(TG.AC)n promotes reciprocal exchange and generates unusual recombinant tetrads during yeast meiosis." Mol Cell Biol **6**(11): 3934-47.1018

Wahls, W. P. (1998). "Meiotic recombination hotspots: shaping the genome and insights into hypervariable minisatellite DNA change." Curr Top Dev Biol **37**: 37-75.1019

Wahls, W. P., L. J. Wallace and P. D. Moore (1990). "The Z-DNA motif d(TG)30 promotes reception of information during gene conversion events while stimulating homologous recombination in human cells in culture." Mol Cell Biol **10**(2): 785-93.510

Wang, A. H., G. J. Quigley, F. J. Kolpak, J. L. Crawford, J. H. van Boom, G. van der Marel and A. Rich (1979). "Molecular structure of a left-handed double helical DNA fragment at atomic resolution." Nature **282**(5740): 680-6.676

Wang, A. J., G. J. Quigley, F. J. Kolpak, G. van der Marel, J. H. van Boom and A. Rich (1981). "Left-handed double helical DNA: variations in the backbone conformation." Science **211**(4478): 171-6.666

Wang, G., L. A. Christensen and K. M. Vasquez (2006). "Z-DNA-forming sequences generate large-scale deletions in mammalian cells." Proc Natl Acad Sci U S A **103**(8): 2677-82.31

Wang, G. and K. M. Vasquez (2006). "Non-B DNA structure-induced genetic instability." Mutat Res **598**(1-2): 103-19.22

Wang, Q., J. Khillan, P. Gadue and K. Nishikura (2000). "Requirement of the RNA editing deaminase ADAR1 gene for embryonic erythropoiesis." Science **290**(5497): 1765-8.169

Wang, Q., M. Miyakoda, W. Yang, J. Khillan, D. L. Stachura, M. J. Weiss and K. Nishikura (2004). "Stress-induced apoptosis associated with null mutation of ADAR1 RNA editing deaminase gene." J Biol Chem **279**(6): 4952-61.108

Wang, Z. and P. Droge (1996). "Differential control of transcription-induced and overall DNA supercoiling by eukaryotic topoisomerases in vitro." Embo J **15**(3): 581-9.1011

Watson, J. D. and F. H. C. Crick (1953). "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid." Nature **171**(4356): 737-738.1

Wittig, B., T. Dorbic and A. Rich (1989). "The level of Z-DNA in metabolically active, permeabilized mammalian cell nuclei is regulated by torsional strain." J Cell Biol **108**(3): 755-64.567

Wittig, B., T. Dorbic and A. Rich (1991). "Transcription is associated with Z-DNA formation in metabolically active permeabilized mammalian cell nuclei." Proc Natl Acad Sci U S A **88**(6): 2259-63.457

Wittig, B., S. Wolfl, T. Dorbic, W. Vahrson and A. Rich (1992). "Transcription of human c-myc in permeabilized nuclei is associated with formation of Z-DNA in three discrete regions of the gene." Embo J **11**(12): 4653-63.376

Wolfl, S., C. Martinez, A. Rich and J. A. Majzoub (1996). "Transcription of the human corticotropin-releasing hormone gene in NPLC cells is correlated with Z-DNA formation." Proc Natl Acad Sci U S A **93**(8): 3664-8.2

Wolfl, S., B. Wittig and A. Rich (1995). "Identification of transcriptionally induced Z-DNA segments in the human c-myc gene." Biochim Biophys Acta **1264**(3): 294-302.262

Wong, B., S. Chen, J. A. Kwon and A. Rich (2007). "Characterization of Z-DNA as a nucleosome-boundary element in yeast Saccharomyces cerevisiae." Proc Natl Acad Sci U S A **104**(7): 2229-34.2

Wong, S. K., S. Sato and D. W. Lazinski (2003). "Elevated activity of the large form of ADAR1 in vivo: very efficient RNA editing occurs in the cytoplasm." Rna **9**(5): 586-98.32

Yang, J. H., Y. Nie, Q. Zhao, Y. Su, M. Pypaert, H. Su and R. Rabinovici (2003). "Intracellular localization of differentially regulated RNA-specific adenosine deaminase isoforms in inflammation." J Biol Chem **278**(46): 45833-42.112

Yathindra, N. and M. Sundaralingam (1976). "Analysis of the possible helical structures of nucleic acids and polynucleotides. Application of (n-h) plots." Nucleic Acids Res **3**(3):

729-47.685

Yuwen, H., J. H. Cox, J. W. Yewdell, J. R. Bennink and B. Moss (1993). "Nuclear localization of a double-stranded RNA-binding protein encoded by the vaccinia virus E3L gene." Virology **195**(2): 732-44.95

Zacharias, W., A. Jaworski and R. D. Wells (1990). "Cytosine methylation enhances Z-DNA formation in vivo." J Bacteriol **172**(6): 3278-83.498

Zarling, D. A., C. J. Calhoun, B. G. Feuerstein and E. P. Sena (1990). "Cytoplasmic microinjection of immunoglobulin Gs recognizing RNA helices inhibits human cell growth." J Mol Biol **211**(1): 147-60.516

Zarling, D. A., C. J. Calhoun, C. C. Hardin and A. H. Zarling (1987). "Cytoplasmic Z-RNA." Proc Natl Acad Sci U S A **84**(17): 6117-21.23

Zhang, S., C. Lockshin, A. Herbert, E. Winter and A. Rich (1992). "Zuotin, a putative Z-DNA binding protein in Saccharomyces cerevisiae." Embo J **11**(10): 3787-96.385

Zheng, G. X., T. Kochel, R. W. Hoepfner, S. E. Timmons and R. R. Sinden (1991). "Torsionally tuned cruciform and Z-DNA probes for measuring unrestrained supercoiling at specific sites in DNA of living cells." J Mol Biol **221**(1): 107-22.454

Zheng, H., T. B. Fu, D. Lazinski and J. Taylor (1992). "Editing on the genomic RNA of human hepatitis delta virus." J Virol **66**(8): 4693-7.1

Zimmerman, S. B. (1982). "The three-dimensional structure of DNA." Annu Rev Biochem **51**: 395-427.1020

VII. SUPPLEMENTARY

Table VII-1. Detailed information of Z-DNA hotspots.

Hotspot	No. of fragments	Unique/high potential	Length (bp)	Chromosome	transcription	Start	End	centromere region	Repeat_type
1	2	Unique	204	chr1		50499670	50499873		MER106A;MIR3
2	2	High potential	375	chr1	Yes	85986886	85987260		L1HS
3	2	Unique	476	chr1		91056181	91056656		MLT2B1;THE1B
4	3	High potential	229	chr1		121053718	121053946	yes	ALR/Alpha
5	27	Unique	1351	chr1		121185607	121186957	yes	ALR/Alpha
6	3	High potential	796	chr1	Yes	142124545	142125340		
7	4	Unique	232	chr1	yes	152841114	152841345		
8	2	Unique	541	chr1	yes	171672245	171672785		AluSg;MER77
9	2	Unique	631	chr1	yes	181945114	181945744		
10	2	Unique	219	chr1		198074634	198074852		L1PA13
11	2	Unique	203	chr1	yes	200540037	200540239		
12	3	Unique	333	chr1	yes	203300649	203300981		
13	2	Unique	535	chr2		9764854	9765388		AluSx
14	2	Unique	535	chr2	yes	10039771	10040305		MER5B;MER5A
15	2	Unique	197	chr2		15902179	15902375		
16	2	Unique	161	chr2	yes	66068688	66068848		MLT1B
17	4	High potential	848	chr2		91659798	91660645	yes	ALR/Alpha
18	5	High potential	462	chr2		91672172	91672633	yes	ALR/Alpha
19	7	High potential	851	chr2		91673188	91674038	yes	ALR/Alpha
20	2	High potential	537	chr2		91675372	91675908	yes	ALR/Alpha
21	7	High potential	800	chr2		91676464	91677263	yes	ALR/Alpha
22	12	Unique	1163	chr2		91681655	91682817	yes	ALR/Alpha
23	13	High potential	1556	chr2		91683535	91685090	yes	ALR/Alpha
24	12	High potential	871	chr2		91688113	91688983	yes	ALR/Alpha
25	2	Unique	170	chr2		120908123	120908292		MIR3;MIR

Hotspot	No. of fragments	Unique/high potential	Length (bp)	Chromosome	transcription	Start	End	centromere region	Repeat_type
26	2	High potential	364	chr2	yes	132728904	132729267		SSU-rRNA_Hsa
27	2	High potential	301	chr2		132755377	132755677		LSU-rRNA_Hsa
28	2	Unique	202	chr2		196774706	196774907		
29	2	Unique	138	chr2		232294584	232294721		
30	2	Unique	338	chr2	yes	238127412	238127749		
31	2	Unique	160	chr3		13177205	13177364		
32	2	Unique	153	chr3		40924858	40925010		L1PA3;LTR17
33	3	High potential	665	chr3		75800691	75801355		LSAU
34	2	Unique	353	chr3		78135525	78135877		L1PA10;Tigger1
35	2	Unique	153	chr3		84705565	84705717		THE1A
36	2	Unique	185	chr3		102732952	102733136		AluSx;LTR34
37	2	Unique	249	chr3	yes	121312848	121313096		L3_Mars;AluSq
38	2	Unique	258	chr3		138609236	138609493		
39	2	Unique	197	chr3	yes	144835445	144835641		L1MC4a;HAL1b
40	2	Unique	79	chr3		154904324	154904402		HERVL40
41	2	Unique	555	chr3		163847925	163848479		THE1A-int
42	2	Unique	121	chr3	yes	199070582	199070702		MER58A
43	2	Unique	309	chr4		11227935	11228243		
44	2	Unique	218	chr4	yes	26344932	26345149		L3
45	2	High potential	192	chr4		49008521	49008712	yes	
46	2	Unique	211	chr4		67947262	67947472		ALR/Alpha
47	2	Unique	247	chr4		76201605	76201851		MIR;MER52A
48	2	Unique	218	chr4		115921399	115921616		L1M1
49	2	Unique	381	chr4		116145911	116146291		L2
50	2	High potential	220	chr4		129002021	129002240		L1PA3

Hotspot	No. of fragments	Unique/high potential	Length (bp)	Chromosome	transcription	Start	End	centromere region	Repeat_type
51	2	Unique	118	chr4	yes	184415042	184415159		
52	2	Unique	586	chr5	yes	489741	490326		AluJo
53	2	High potential	729	chr5		49476097	49476825	yes	ALR/Alpha
54	2	Unique	117	chr5		124276929	124277045		LTR1
55	2	Unique	392	chr5		132619797	132620188		
56	2	Unique	609	chr6		37077364	37077972		MIR;MLT2F;MLT2E
57	2	High potential	455	chr6	yes	44950607	44951061		L1PA2
58	2	High potential	451	chr6		58884581	58885031	yes	ALR/Alpha
59	2	Unique	244	chr6		58885904	58886147	yes	ALR/Alpha
60	2	Unique	200	chr6	yes	73932436	73932635		MIR;AluY
61	2	Unique	396	chr6		83402631	83403026		L1PA3
62	2	Unique	162	chr6		150289643	150289804		MER41E
63	2	Unique	548	chr6	yes	151940394	151940941		MER1B
64	2	Unique	243	chr7	yes	2100911	2101153		L2
65	6	Unique	514	chr7		61606113	61606626	yes	ALR/Alpha
66	2	Unique	323	chr7		61606845	61607167	yes	ALR/Alpha
67	2	High potential	185	chr7		61609161	61609345	yes	ALR/Alpha
68	2	Unique	157	chr7	yes	104973277	104973433		AluSx
69	2	High potential	235	chr8	yes	39566526	39566760		L1PA3
70	2	Unique	396	chr8		101766284	101766679		AluSg;L1MC4;AluSg
71	2	Unique	136	chr8		127677697	127677832		MLT1F1;AluY
72	2	Unique	157	chr9	yes	28142379	28142535		MSTB1
73	20	High potential	630	chr9		66710805	66711434		ALR/Alpha
74	12	High potential	630	chr9		67010137	67010766		ALR/Alpha
75	21	High potential	630	chr9		69000548	69001177		ALR/Alpha

Hotspot	No. of fragments	Unique/high potential	Length (bp)	Chromosome	transcription	Start	End	centromere region	Repeat_type
76	2	Unique	283	chr9	yes	87407930	87408212		L1MA4A
77	2	Unique	473	chr9		98875115	98875587		L1MB4;MER68-int
78	2	High potential	145	chr9	yes	99002086	99002230		
79	2	Unique	173	chr9	yes	115090340	115090512		
80	2	Unique	123	chr9	yes	136757676	136757798		L1MB3
81	2	Unique	136	chr10		12369283	12369418		
82	2	Unique	106	chr10		23756467	23756572		LTR67
83	2	Unique	194	chr10		25425840	25426033		L1MA2
84	2	Unique	223	chr10		38010728	38010950		
85	5	High potential	635	chr10		41704938	41705572	yes	HSATII
86	4	Unique	883	chr10	yes	41706997	41707879	yes	
87	6	High potential	663	chr10		41711251	41711913	yes	
88	9	Unique	1013	chr10		41718470	41719482	yes	ALR/Alpha
89	7	Unique	657	chr10		41720123	41720779	yes	ALR/Alpha
90	3	High potential	220	chr10		41847917	41848136	yes	ALR/Alpha
91	2	High potential	518	chr10		41916699	41917216	yes	
92	2	High potential	471	chr10		41917450	41917920	yes	
93	4	Unique	366	chr10	yes	41919670	41920035	yes	
94	2	Unique	194	chr10	yes	91136961	91137154		AluY
95	2	Unique	124	chr10	yes	95691602	95691725		
96	2	Unique	205	chr10	yes	118370872	118371076		
97	2	Unique	190	chr10		124439626	124439815		MER20;MIR
98	2	Unique	186	chr10		125921770	125921955		MIR;MLT1D
99	2	Unique	299	chr11	yes	599094	599392		
100	2	Unique	517	chr11	yes	9765302	9765818		AluSc

Hotspot	No. of fragments	Unique/high potential	Length (bp)	Chromosome	transcription	Start	End	centromere region	Repeat_type
101	2	Unique	227	chr11	yes	12147950	12148176		L1MC4a
102	2	Unique	308	chr11		43108944	43109251		MIR3;AluJb
103	2	High potential	507	chr11		51436666	51437172	yes	ALR/Alpha
104	3	High potential	629	chr11		51439131	51439759	yes	ALR/Alpha
105	4	High potential	911	chr11		51447036	51447946	yes	ALR/Alpha
106	2	High potential	600	chr11		51449084	51449683	yes	ALR/Alpha
107	2	Unique	205	chr11	yes	74581756	74581960		
108	2	High potential	199	chr11	yes	77275122	77275320		
109	2	Unique	384	chr11		126010413	126010796		MIRb
110	2	Unique	558	chr12	yes	49427456	49428013		
111	2	Unique	481	chr12	yes	52917660	52918140		AluSc
112	2	Unique	161	chr12		59301756	59301916		L1PA7
113	2	Unique	466	chr12		72705142	72705607		THE1D
114	2	Unique	142	chr12		123781421	123781562		AluSx
115	2	Unique	159	chr13	yes	50268841	50268999		MLT1J2
116	2	Unique	139	chr13	yes	95296479	95296617		SVA
117	2	Unique	274	chr14		47785172	47785445		AluSg
118	2	Unique	393	chr14		74185271	74185663		L2;MIRb
119	2	Unique	351	chr14	yes	77798023	77798373		L2
120	2	Unique	441	chr14	yes	79939577	79940017		LTR17
121	2	Unique	153	chr14		88536577	88536729		L2
122	2	High potential	151	chr15	yes	21236623	21236773		
123	3	High potential	250	chr15	yes	21237185	21237434		
124	2	High potential	116	chr15	yes	52306218	52306333		MER52A
125	3	High potential	329	chr15	yes	56426735	56427063		L1PA5

Hotspot	No. of fragments	Unique/high potential	Length (bp)	Chromosome	transcription	Start	End	centromere region	Repeat_type
126	2	Unique	119	chr15		67218580	67218698		
127	2	Unique	446	chr15	yes	75025378	75025823		AluJo
128	2	Unique	308	chr15		76015277	76015584		AluY
129	2	Unique	385	chr16	yes	24796809	24797193		AluSg;AluSx;L2;L2
130	2	Unique	309	chr16	yes	27758938	27759246		AluJo;AluJ/FLAM
131	5	Unique	534	chr16		33871285	33871818		
132	3	High potential	582	chr16		44948604	44949185		
133	3	High potential	416	chr16		44963745	44964160		
134	2	Unique	188	chr16	yes	80524610	80524797		Alu
135	2	Unique	239	chr16		85546275	85546513		
136	2	Unique	112	chr16		87000125	87000236		
137	2	Unique	150	chr16	yes	88027574	88027723		
138	8	High potential	1117	chr17		22169733	22170849	yes	ALR/Alpha
139	6	High potential	741	chr17		22172112	22172852	yes	ALR/Alpha
140	7	High potential	1208	chr17		22175202	22176409	yes	ALR/Alpha
141	9	High potential	1452	chr17		22176867	22178318	yes	ALR/Alpha
142	6	High potential	1117	chr17		22181624	22182740	yes	ALR/Alpha
143	10	High potential	1071	chr17		22184384	22185454	yes	ALR/Alpha
144	2	Unique	163	chr17	yes	22998352	22998514	yes	
145	2	Unique	123	chr17		45370783	45370905		AluSx
146	2	Unique	123	chr17		77017153	77017275		
147	3	High potential	360	chr18	yes	100802	101161		BSR/Beta
148	2	High potential	287	chr18		16766775	16767061	yes	ALR/Alpha
149	2	High potential	515	chr18		16772490	16773004	yes	ALR/Alpha
150	5	High potential	740	chr18		16773262	16774001	yes	ALR/Alpha

Hotspot	No. of fragments	Unique/high potential	Length (bp)	Chromosome	transcription	Start	End	centromere region	Repeat_type
151	2	Unique	213	chr18		50905039	50905251		
152	2	Unique	166	chr18		58425200	58425365		AluJ/FLAM
153	2	Unique	175	chr18		74875513	74875687		
154	2	High potential	441	chr19	yes	1154649	1155089		AluSq
155	10	Unique	616	chr19		32423682	32424297		ALR/Alpha
156	3	Unique	263	chr19		32424908	32425170		ALR/Alpha
157	2	High potential	311	chr19		32426275	32426585		ALR/Alpha
158	4	High potential	345	chr19		32427976	32428320		ALR/Alpha
159	2	Unique	144	chr19		35015042	35015185		AluY
160	2	Unique	271	chr19	yes	40929077	40929347		
161	4	High potential	503	chr19	yes	41491474	41491976		SST1
162	2	Unique	142	chr19	yes	43112435	43112576		AluJo/AluY
163	2	Unique	413	chr19		44471198	44471610		L1MD2;AluSx;AluSq/x
164	2	High potential	723	chr20	yes	18530809	18531531		L1PA2
165	2	Unique	170	chr20		24825266	24825435		L1MC5
166	2	Unique	367	chr20	yes	26137749	26138115	yes	
167	3	Unique	567	chr20	yes	28131880	28132446		SST1
168	2	Unique	191	chr20		31389206	31389396		
169	2	Unique	212	chr20		35646855	35647066		
170	2	Unique	390	chr20	yes	52291256	52291645		L1PA4
171	2	Unique	183	chr21	yes	44799139	44799321		Charlie5
172	2	High potential	195	chr22	yes	17261872	17262066		HSATI
173	2	Unique	195	chr22		45346722	45346916		L2
174	2	Unique	283	chrX		7474644	7474926		LTR1D
175	2	Unique	313	chrX		31376079	31376391		

Hotspot	No. of fragments	Unique/high potential	Length (bp)	Chromosome	transcription	Start	End	centromere region	Repeat_type
176	3	High potential	505	chrX		61599595	61600099	yes	ALR/Alpha
177	4	High potential	1114	chrX		61629851	61630964	yes	ALR/Alpha
178	2	High potential	539	chrX		100722197	100722735		L1PA3
179	3	High potential	198	chrX		108184005	108184202		L1
180	2	Unique	219	chrX		110407421	110407639		L1PA3
181	2	Unique	495	chrX	yes	131681586	131682080		
182	2	Unique	164	chrX		150410395	150410558		MIRb;MIRb
183	4	High potential	608	chrY		10646059	10646666		
184	2	High potential	543	chrY		11929555	11930097	yes	BSR/Beta
185	2	High potential	390	chrY		11936355	11936744	yes	BSR/Beta
186	2	High potential	316	chrY		57387229	57387544		AluYc3

VIII. PUBLICATIONS

1. Vladimir V. Kulish, Li Heng, Peter Dröge. Z-DNA-induced super-transport of energy within genomes. *Physica A: Statistical Mechanics and its Application.* 2007 Oct; 384(2): 733-738

2. Heng Li, Jie Xiao, Jinming Li & Peter Dröge. A Z-DNA map of the human genome using the Zα domain of ADAR1 as a probe. Manuscript in preparation

3. Shu Feng, Heng Li, Peter Dröge. The Zα domain of ADAR1 is a potent translation inhibitor through specific interactions with pro- and eukaryotic ribosomes. Manuscript in preparation.