Natural-occurring Z-conformations in nucleic acids: probed by the Zα domain of human RNA editing enzyme

Feng, Shu

2010

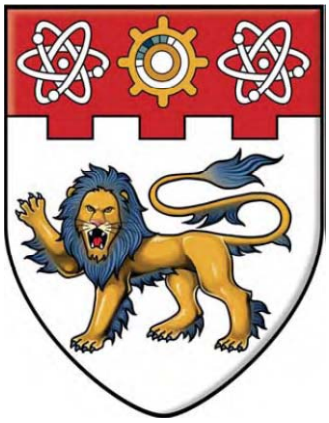
Feng, S. (2010). Natural-occurring Z-conformations in nucleic acids: probed by the Zα domain of human RNA editing enzyme. Doctoral thesis, Nanyang Technological University, Singapore.

https://hdl.handle.net/10356/47999

https://doi.org/10.32657/10356/47999

Natural-occurring Z-conformations in nucleic acids: probed by the Zα domain of human RNA editing enzyme ADAR1

Feng Shu



SCHOOL OF BIOLOGICAL SCIENCES
NANYANG TECHNOLOGICAL UNIVERSITY

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

2010

ABSTRACT

Besides the common right-handed B- or A- structures, the alternate Z-conformation, which is left-handed, can be formed in both DNA and RNA molecules. It is in a higher energy state and well-studied mostly *in vitro*. However, formation of Z-conformation in living organisms and its biological significance remains largely elusive. Here, a protein probe that could specifically bind to Z-conformation was used to map the Z-DNA distribution in the human genome. The probe is derived from the Z-DNA binding domain of the human double-stranded RNA adenosine deaminase 1 (ADAR1), named $Z\alpha_{ADAR1}$. During our experiments, evidence suggested that the primary target of $Z\alpha_{ADAR1}$ *in vivo* is RNA rather than DNA. Further studies revealed that $Z\alpha_{ADAR1}$ could bind to ribosomes and inhibit translation in both *E. coli* and mammalian systems in a Z-conformation dependent manner. Potential $Z\alpha_{ADAR1}$ binding sites on ribosomes were identified. Several binding sites were conserved between *E. coli* and human ribosomes, revealing that formation of Z-like RNA conformations might be a conserved property of the dynamic ribosome structure during translation. The implications for understanding biological functions of the full length ADAR1 were discussed.

In vitro Chromatin Affinity Precipitation (ChAP) experiments using $Z\alpha_{ADAR1}$ were also performed to map the Z-DNA distribution in the human genome, using human embryonic stem (huES) cells as the cell source. The data revealed that the $Z\alpha_{ADAR1}$ binding sites were found to be enriched in centromere regions and G-bands, suggesting that formation of Z-DNA might contribute to specific chromatin structures. The

enrichment of Z-DNA in transcription start sites then suggested that formation of Z-DNA is associated with transcription processes. Enrichment of SNPs in $Z\alpha_{ADAR1}$ binding sites, especially in the centromere region, further suggested that formation of Z-DNA might be mechanistically linked to genetic instability and thus contributes to the evolution of centromere regions. Surprisingly, the $Z\alpha_{ADAR1}$ binding sites did not coincide with Z-DNA forming sequences predicted based on *in silico* analyzes. This indicates that formation of Z-DNA in living cells is not purely depended on DNA sequence contexts, but is strongly affected by on the local physical environment.

ACKNOWLEDGMENTS

I would like to express special gratitude to my supervisor Prof. Peter Dröge for his direction, encouragement and support during the progress of the project.

Thanks go to Dr. Li Heng, who designed and initially performed the basic analysis of $Z\alpha_{ADAR1}$. Thanks for his valuable cooperation and discussion during the project.

Thanks also go to Prof. Curt Davey for the help on analyzing the structures of the ribosome binding sites. Thanks to Prof. Jinming Li for his instruction on the bioinformatics analyses.

I would present my special thanks to my partner Li Bin, for his help on the data analysis of the Z-DNA libraries and for his love and support in the past few years.

I also extend my thanks to all members in our groups, who provided assistance, convenience and a friendly atmosphere in the lab.

I deeply thank Prof. Alex Law for his kind support during my PhD study. I am grateful for the financial support of scholarship from Nanyang Technology University during my study.

ABBREVIATIONS

ADAR	adenosine deaminase acting on RNA
BSA	bovine serum albumin
bp	base pair
DMEM	Dulbecco's Modified Eagle's Medium
DNA	Deoxyribonucleic acid
Dox	Doxcyclin
DTT	dithiothreitol
ds/ss	double strand/single strand
EDTA	ethylene glycol-bis (β-aminoethylether)-N, N' tetraacetic acid
EGTA	Ethylenediaminetetraacetic acid
FBS	Fetal Bovine Serum
FPLC	Fast Protein Liquid Chromatography
hcRBC	high confidence Ribosome Binding site
huES	human Embryonic Stem Cells
IPTG	Isopropyl β-D-1-Thiogalactopyranoside
kDa	kilo Daltons
LB	Luria-Bertani
NLS	nuclear localization signal
nt	nucleotide
OD	optical density

PAGE	polyacrylamide gel electrophoresis
PBS	phosphate buffered saline
PCR	polymerase chain reaction
PMSF	phenylmethanesulfonylfluoride
Puro	puromycin
rcf	Relative centrifugal force
RIPA	Radio Immuno-Precipitation Assay
rpm	revolutions per minute
SDS	sodium dodecyl sulfate
TBST	Tris-Buffered Saline Tween-20
TEMED	N,N,N',N'-tetramethylethylenediamine
TSS	Transcription Start Site
UV	ultraviolet
v/v	volume per volume
w/v	weight per volume

TABLE OF CONTENTS

LIST OF FIGURES

I INTRODUCTION

I.1 Z-DNA and Z-RNA conformations

Double-stranded DNA is known to occur mainly in three conformations, named A-，B-, and Z-DNA (Figure 1). DNA in living cells is found mainly in the right-handed B-conformation (Richmond and Davey 2003). In 1979, the first single-crystal X-ray structure of a DNA fragment was solved (Wang, Quigley et al. 1979). Surprisingly, it revealed a left-handed double helix with two anti-parallel strands held together by Watson-Crick base pairing. It was named Z-DNA, after its zigzag arrangement of the sugar-phosphate backbone. Compared to A- and B- conformations, Z-DNA is more elongated, has a narrower minor groove and the base pairs are nearly perpendicular to the helix axis.

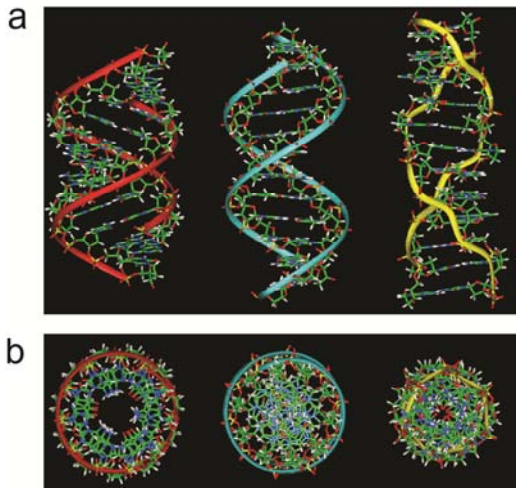


Figure 1 The structures of three distinct DNA conformations: A-, B-, and Z-DNA. a. Side view. b. Top view.

(Cited from http://en.wikipedia.org/wiki/File:A-B-Z-DNA_Side_View.png and http://en.wikipedia.org/wiki/File:A-B-Z-DNA_Top_View.png).

The zigzag arrangement of the backbone is formed by alternating purines and pyrimidines in alternating *syn-anti* conformation. Z-DNA occurs most readily in alternating purine/pyrimidine repeat sequences, as purines adopt the *syn* conformation more readily than pyrimidines, especially in alternating $d(GC)_n$ and $d(GT)_n$ repeats (Malfoy, Rousseau et al. 1986; Johnston 1992). Z-DNA could also form in non-alternating purine/pyrimidine repeat sequences (Ha, Choi et al. 2009), especially with high GC content, such as in $d(GGGC)_n$ repeats (Feigon, Wang et al. 1985). Z-DNA is more favored to be formed under negative DNA superhelical torsion (Feigon, Wang et al. 1985; Brown and Rich 2001) and could be triggered by the presence of low DNA tension (Lee, Kim et al. 2010).

Z-DNA structures were also studied by circular dichroism (CD) measurement (Pohl and Jovin 1972) . Compared to B-DNA, Z-DNA shows a high value at around 260 nm and lower value at around 290 nm in the CD spectrum. A shift of the negative peak from 260 nm to around 290 nm could be used to identify B- to Z- transitions (Berger, Winston et al. 1998). Related experiments showed that formation of Z-DNA could be promoted by high salt (Pohl 1983), or DNA modifications such as C-5 methylation of deoxycytosine (Behe and Felsenfeld 1981; Fujii, Wang et al. 1982), N-7 or C-8 methylation in poly(dG-dC) (Moller, Nordheim et al. 1981; Kawai, Sugiyama et al. 1995) or bromination (Moller, Nordheim et al. 1984). Temperature also plays a role in Z-DNA formation *in vitro.* (Irikura, Tidor et al. 1985).

Z-RNA can also form within the same sequences that are favored to form Z-DNA. Z-RNA is recognized by the Z-DNA binding domain $Z\alpha_{ADAR1}$ *in vitro* (Brown, Lowenhaupt et al. 2000), and a Z-RNA structure has been solved for a co-crystallized $Z\alpha_{ADAR1}$-RNA complex (Placido, Brown et al. 2007). The structures showed that Z-RNA shares many features with Z-DNA (Figure 2). Similar to the formation of Z-DNA, high salt conditions can promote Z-RNA formation (Popenda, Milecki et al. 2004).

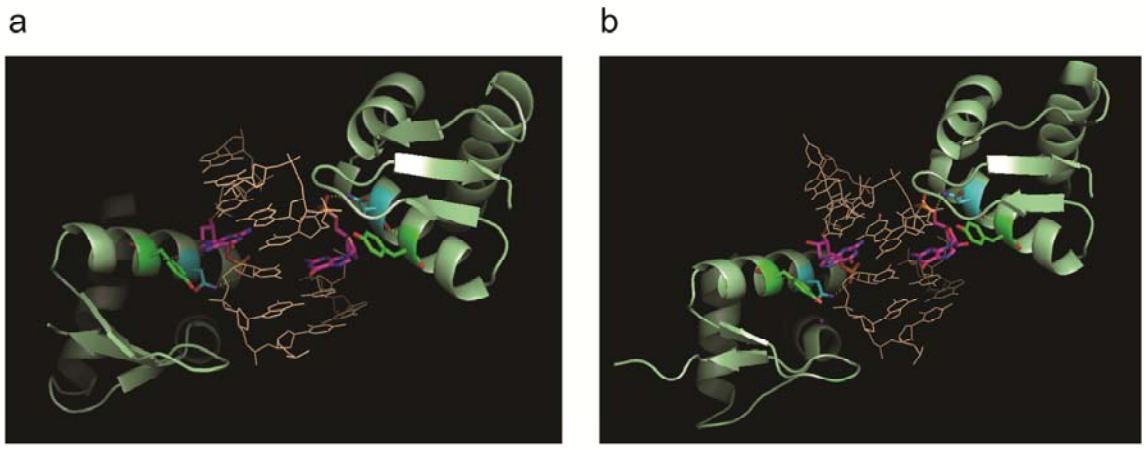


Figure 2 $Z\alpha_{ADAR1}$ binds to Z-DNA and Z-RNA

$Z\alpha_{ADAR1}$ was labeled in light-green and showed as cartoon, while Z-DNA and Z-RNA were labeled in wheat and displayed as line. **a.** $Z\alpha_{ADAR1}$ binds to Z-DNA (PDB:1QBJ) **b**. $Z\alpha_{ADAR1}$ binds to Z-RNA (PDB: 2GXB) Two of the essential amino acids(Asn^{173} and Tyr^{177}) for Z-DNA binding were shown in blue and green, while the G4 base in *syn*-conformation from the Z-DNA and Z-RNA was shown in magentas (See also, Schwartz, Rould et al. 1999; Placido, Brown et al. 2007).

I.2 Evidence for the existence of Z-DNA *in vivo*

The structures solved for Z-DNA occur under non-physiological conditions like high salt, high pressure or by chemical modifications. This raised skepticism about the existence

of Z-DNA in living organisms (Rich and Zhang 2003; Mirkin 2008). However, Z-DNA can form under conditions of negative superhelical stress (Klysik, Stirdivant et al. 1981; Singleton, Klysik et al. 1982), and negative supercoiling can form in DNA behind moving RNA polymerases during transcription. Hence, Z-DNA can be induced transiently behind a moving RNA polymerase and be stabilized by negative supercoiling generated by DNA transcription, as demonstrated in Figure 3. (Liu and Wang 1987; Droge and Nordheim 1991; Droge and Pohl 1991). The negative supercoiling can be released by topoisomerases, and Z-DNA will revert back to the lower energy B-DNA conformation (Wang and Droge 1996), which results in Z-DNA as a transient conformation *in vivo*. *In vivo* experiments in mammalian cells also supported the idea that Z-DNA forming sequences can be detected in several genes by Z-DNA antibodies (Wittig, Wolfl et al. 1992), and in 5' regions of the *c-myc* gene (Wittig, Wolfl et al. 1992; Wolfl, Wittig et al. 1995). Furthermore, Z-DNA formation correlated with the *c-myc* expression level (Wolfl, Martinez et al. 1996).

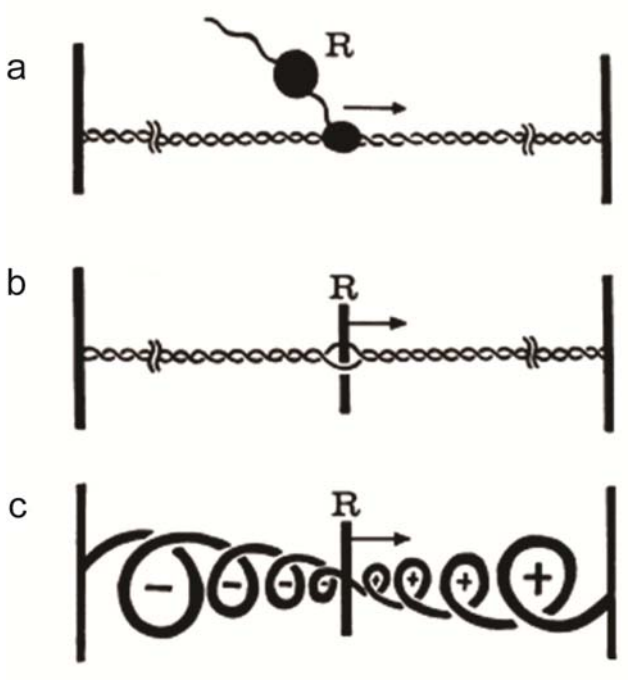


Figure 3 A graphical illustration of the relationship between transcription and DNA supercoiling formation.

a. A transcription ensemble R is moving in the direction of the arrow along a DNA segment; the ends of the DNA segment are anchored on a large structure represented by the solid bars. **b.** The transcription ensemble can be viewed as a divider separating the helical DNA into two parts. **c.** If R is moving from left to right without turning around the DNA, the DNA in front of the polymerase becomes positively supercoiled, while the DNA behind the polymerase becomes negatively supercoiled. (Liu and Wang 1987)

As Z-DNA is likely to be transient *in vivo* and linked to genetic activities such as transcription, detection of Z-DNA in mammalian cells is difficult. Despite which, several early observations still suggested the existence of Z-DNA *in vivo*. One example is the detection of Z-DNA specific antibodies in sera from patients with several auto-immune diseases (Lafer, Valle et al. 1983). Studies using Z-DNA antibodies showed that Z-DNA is associated with chromosome inter-band region (Nordheim, Pardue et al. 1981;

Lancillotti, Lopez et al. 1987), and formation of Z-DNA is mainly related to transcription rather than DNA replication (Wittig, Dorbic et al. 1991).

I.3 Evidence for the existence of Z-RNA *in vivo*

Experimental evidences showed that chemical bromination of poly[r(C-G)] could stabilize synthetic Z-RNA under physiological conditions (Hall, Cruz et al. 1984; Hardin, Zarling et al. 1987). Z-RNA was detected in the cytoplasm of fixed protozoan cells by anti-Z-RNA antibodies (Zarling, Calhoun et al. 1987). Z-RNA existed in both cytoplasm and nucleoli in fixed cells, as revealed by cytoplasm microinjection of antibodies recognizing Z-RNA (Zarling, Calhoun et al. 1990). Interestingly, the Z-DNA binding protein 1 (ZBP1) was found to be accumulated in stress granules, where stalled translation pre-initiation complexes were stored during stress (Deigendesch, Koch-Nolte et al. 2006).

I.4 Biological functions of Z-DNA

I.4.1 Z-DNA and transcription activity

Z-DNA is reported to be involved in transcriptional regulation *in vivo* (Rich, Nordheim et al. 1983; Santoro and Costanzo 1983; Hamada, Seidman et al. 1984; Santoro, Costanzo et al. 1984; Lancillotti, Lopez et al. 1987). However, the relationship between transcriptional activity and Z-DNA formation is still disputed. Formation of Z-DNA in different genes results in different regulatory effects; some promote (Liu, Liu et al. 2001)

while others inhibit transcription (Rothenburg, Koch-Nolte et al. 2001) probability depends on its location and stability.

I.4.2 Z-DNA and chromatin remodeling

Z-DNA is difficult to be incorporated into the nucleosome core particles (Garner and Felsenfeld 1987), and formation of Z-DNA could therefore affect nucleosome packaging and positioning (Liu, Liu et al. 2001; Liu, Mulholland et al. 2006). On the other hand, as nucleosomal protein complexes wrap DNA in a left-handed orientation, dissembled nucleosomes will result in negative DNA supercoiling and facilitates Z-DNA formation.

I.4.3 Z-DNA and genetic instability

Z-DNA is reported to associate with genetic instabilities in many aspects (Wang and Vasquez 2006; Zhao, Bacolla et al. 2010). Firstly, the Z-conformation structure itself is more susceptible to DNA damage than B-conformation. The purines are in *syn*-conformation in Z-DNA, which leave the bases out of protection by the sugar-phosphate backbone. Crystal structures revealed extruded bases at the Z- and B-DNA junction (Ha, Lowenhaupt et al. 2005), which is susceptible to DNA modification or nucleolytic attack. Also, Z-DNA is not likely to be formed in the nucleosome particle (Wong, Chen et al. 2007), which leaves it physically more exposed to DNA damage. These theories were also supported by the fact that Z-DNA forming sequences or ss DNA are more susceptible to DNA damages such as ionizing radiation (Tartier, Michalik et al. 1994) or oxidative damage (Schulz, Mahler et al. 2000). Furthermore, the unusual structure of Z-

DNA may prevent binding and accessibility of DNA repair enzymes to the damaged or modified sites. The repair efficiency will thus be reduced (Lagravere, Malfoy et al. 1984).

Secondly, Z-DNA forming sequences are demonstrated to be correlated with recombination events. It has been shown that Z-DNA forming sequences in supercoiled plasmid could stimulate DNA recombination in prokaryotic cells (Murphy and Stringer 1986).The cellular recombination events on plasmids are also reported to be enhanced at Z-DNA forming sequences in eukaryotes (Wahls, Wallace et al. 1990). Studies of the major histocompatibility complex (MHC) in mice revealed a hot spot of 1000 bp, which may account for up to 2% of recombinants on the entire chromosome, containing several copies of long GT repeats (Kobori, Strauss et al. 1986; Wahls 1998). Furthermore, the association of recombination events with GT repeats were revealed by bioinformatics analysis on human chromosome 22 (Majewski and Ott 2000).

Thirdly, Z-DNA forming sequences are demonstrated to be associated with genetic deletions or genetic rearrangement. Z-DNA sequences are shown to be hotspots for spontaneous deletions in *E. coli* (Klysik, Stirdivant et al. 1981; Klysik, Stirdivant et al. 1982; Freund, Bichara et al. 1989). In mammalian cells, Z-DNA causes DNA double-strand breaks and can result in large-scale deletions and rearrangements (Wang, Christensen et al. 2006; Wang, Carbajal et al. 2008). This seems more severe than in prokaryotic cells, probably due to different repair mechanisms at work in prokaryotes and eukaryotes (Kha, Wang et al. 2010). Chromosomal translocations were reported to

be promoted by Z-DNA forming sequences in variety of human lymphoid tumors (Boehm, Mengle-Gaw et al. 1989).

I.5 Biological functions of Z-RNA

As stated above, some studies indicated the existence of Z-RNA *in vivo*. However, there is no report even speculating on possible biological functions for this alternate RNA secondary structure.

I.6 The ADAR family and the $Z\alpha_{ADAR1}$ domain

Central to the questions highlighted above, i.e. possible biological function(s) of Z-conformations, is the quest for proteins which bind specifically to this alternate nucleic acid structure. Rich and co-worker identified in the late 90's the by now well-known Z-DNA binding domain, Zα, present in the RNA-specific adenosine deaminase 1 (ADAR1).(Schwartz, Rould et al. 1999)

I.6.1 The ADAR family

Adenosine deaminase acting on RNA (ADAR) is a protein family that converts adenosine to inosine in or near double-stranded regions of RNA (Maas, Melcher et al. 1997). ADARs are involved in both non-specific and site-specific A-to-I editing. Extensive adenosine to inosine editing has been reported for various viruses and is one of the possible antiviral defense mechanisms by disrupting viral open reading frames (ORF). The ADARs have antiviral activities, and are involved in apoptosis and

embryogenesis, as well as in increasing neurotransmitter receptors diversity, and interfering with RNAi pathways (Toth, Zhang et al. 2006).

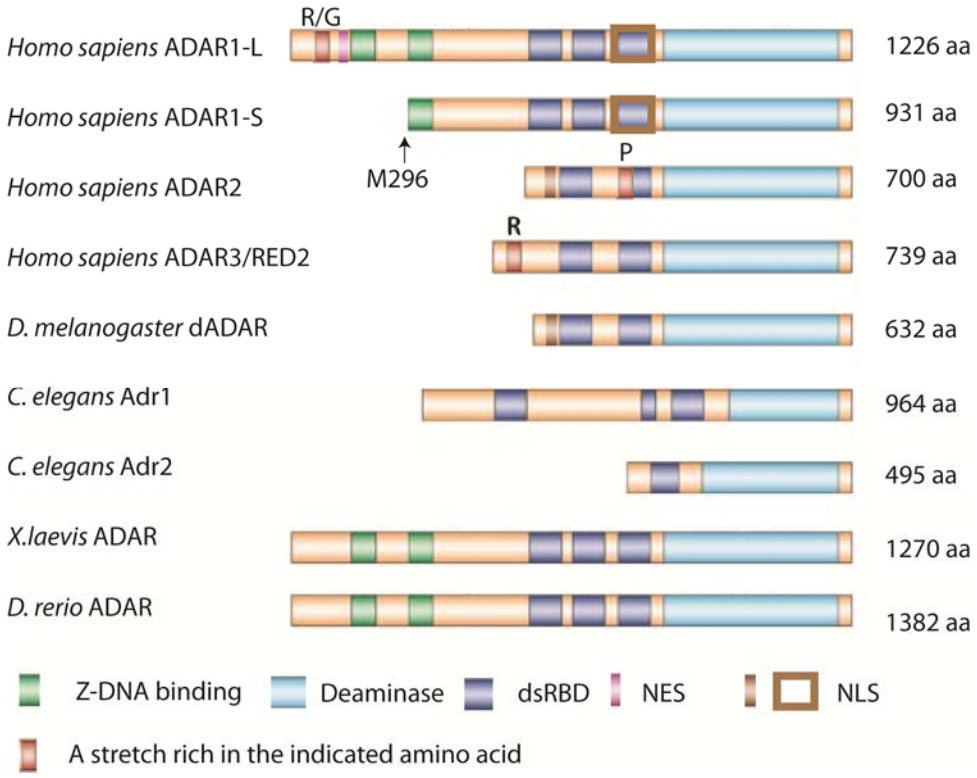


Figure 4 The primary structures of ADARs several species.

The Z-DNA binding domains (Zα and Zβ), double-stranded RNA binding domain (dsRBD) and adenosine deaminase domain are shown. The hADAR1-S is translated start from M296 in hADAR1-L. The lengths of ADARs are indicated at the right side. aa: amino acid. The position of nuclear localization signal (NLS) and the nuclear export signal (NES) are also indicated.(Nishikura 2006)

ADARs are widely conserved in their adenosine deaminase domains, but differ in dsRNA binding and Z-DNA binding motifs (Figure 4). Two forms of ADAR1 proteins are expressed from the human *ADAR1* gene (Patterson and Samuel 1995): an interferon (IFN)-inducible p150 ADAR1-L and a constitutively expressed p110 ADAR1-S. In

addition to ADAR1, two mammalian ADAR genes and proteins have been identified (ADAR2 and ADAR3). ADAR2 is constitutively expressed and catalytically active (Melcher, Maas et al. 1996; O'Connell, Gerber et al. 1997), while ADAR3, also termed as RED2, lacks deaminase activity (Chen, Cho et al. 2000).

Within the ADAR family, only ADAR1-L contains the Z-DNA specific binding domains, named Zα and Zβ. The *Z*α domain has a remarkable ability to bind specifically to nucleic acids in Z-conformation. However, the *Z*β domain, which is homologous to *Z*α, lacks this ability. (Athanasiadis, Placido et al. 2005) Z-DNA binding domains are conserved in human, rat, bovine, chicken and *xenopus* ADAR1 (Herbert, Alfken et al. 1997), implying the importance of these domains for the biological function(s) of ADAR1 (Athanasiadis, Placido et al. 2005). Mutations in Zα decrease the efficiency of editing of short dsRNA substrates (Herbert and Rich 2001), and the editing effect is greatly enhanced when a Z-RNA forming sequence is used as substrate (Koeris, Funke et al. 2005). The induction of ADAR1-L expression by interferon and inflammation, but none of the other ADARs, implies the special involvement of Z-DNA binding domains in antiviral defense. The ADAR1 could suppress virus-induced apoptosis. It is reported that ADAR1 could suppress measles virus-induced apoptosis, through suppression of activation of proapoptotic and double-stranded RNA-dependent activities, as exemplified by PKR and IRF-3.(Toth, Li et al. 2009) A recent report showed that, the double-stranded RNAs containing multiple IU pairs, which are generated by ADAR1, are sufficient to suppress interferon induction and apoptosis.(Vitali and Scadden 2010)

It has been demonstrated that the constitutively expressed ADAR1-S (p110) was localized predominantly in the nucleus, especially in or on the nucleolus, whereas the IFN-inducible ADAR1-L (p150) was localized to both cytoplasm and the nucleus (Patterson and Samuel 1995). ADAR1-L has functional nuclear localization signals (NLSs), nuclear/nucleolus localization signal (NLS/NLoS) and a nuclear export signal (NES) (Nie, Zhao et al. 2004). Hence, the full length ADAR1-L shows characteristics of a shuffling protein (Eckmann, Neunteufl et al. 2001; Strehblow, Hallegger et al. 2002).

I.6.2 Complex formation between the $Z\alpha_{ADAR1}$ domain and Z-DNA/ Z-RNA

The interaction between $Z\alpha_{ADAR1}$ with Z-DNA was analyzed and confirmed by CD, NMR, crystallography, and biochemical assays. $Z\alpha_{ADAR1}$ binds to Z-DNA very tightly with a K_D of 4 nM, based on BIAcore surface plasmon resonance using brominated $d(GC)_3$ as the substrate (Herbert, Schade et al. 1998). A B-to-Z transition in CD spectra was observed when $Z\alpha_{ADAR1}$ bound to $d(GC)_n$ polymer at low salt concentration (Berger, Winston et al. 1998), suggesting that $Z\alpha_{ADAR1}$ was able to induce or stabilize the Z-conformation (Lee, Seo et al. 2010). The same effect was observed with dsRNA. $Z\alpha_{ADAR1}$ promotes A-to-Z transitions in dsRNA at near physiological conditions (Brown, Lowenhaupt et al. 2000). A recent study showed that $Z\alpha_{ADAR1}$ could bind and stabilize the Z-DNA via a “conformational selection” mechanism rather than induce the B to Z transition (Bae, Kim et al. 2010). $Z\alpha_{ADAR1}$ could stabilize the Z-conformation up to few hours (Bae, Kim et al. 2010).

The high-resolution structure was solved by NMR (Schade, Turner et al. 1999) and crystallography (Schwartz, Rould et al. 1999). $Z\alpha_{ADAR1}$ consisted of three α-helices and three β-sheets in a winged helix-turn-helix conformation, which is a novel class for DNA binding domains (Figure 2). The $Z\alpha_{ADAR1}$ structure in solution exhibits very high similarities with $Z\alpha_{ADAR1}$ in the crystal complex, suggesting that $Z\alpha_{ADAR1}$ is already primed for Z-DNA binding in the un-bound form. Both NMR and crystallography studies show that two $Z\alpha_{ADAR1}$ molecules bind one Z-DNA segment. Each monomer recognizes one strand without interactions with each other, suggesting that protein dimerization is not required for binding. On the Z-DNA binding surface, direct hydrogen bonds are formed between amino acid residues in α3 and β3 (Tyr^{177}, Asn^{173}, Lys^{169}, Lys^{170}, Arg^{174} and Thr^{191}) and five consecutive phosphates in the backbone of DNA (Figure 5). In addition to polar interactions, the aromatic ring of Tyr^{177} interacted with G4 base via van der Waals interaction. The G4 base was in *syn* conformation, which is the typical conformation in Z-DNA. Pro^{192} and Pro^{193} in β3 also interacted with the Z-DNA backbone via van der Waals interaction.

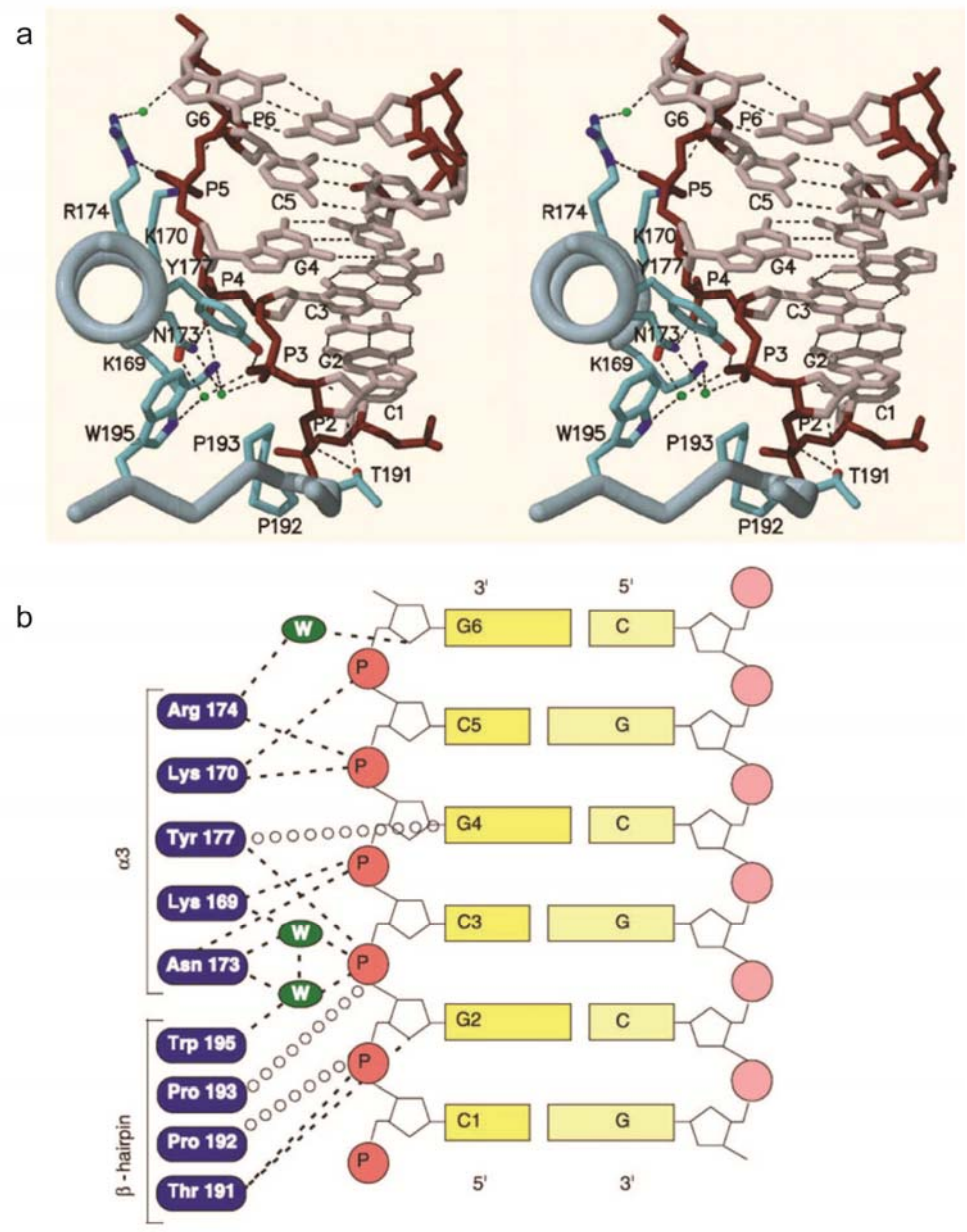


Figure 5 The binding surface between $Z\alpha_{ADAR1}$ and Z-DNA.

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 (See also Figure 2, labeled in magenta) 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 were showed. Dashed line stands for hydrogen bond and open circle for van der Waals contact. (Schwartz, Rould et al. 1999)

The fine structure for $Z\alpha_{ADAR1}$ in complex with Z-RNA was also solved by crystallography (Placido, Brown et al. 2007). $Z\alpha_{ADAR1}$ monomers interact with one RNA strand in a very similar manner as in Z-DNA. Direct hydrogen bonding was also formed between almost identical amino acid residues in α3 and β3 (Tyr^{177}, Asn^{173}, Lys^{169}, Lys^{170} and Thr^{191}) with the RNA backbone. Arg^{174} which interacted with Z-DNA backbone pointed away from RNA in the Z-RNA crystal structure. The Tyr^{177} also forms an edge-to-face π bond with the Z-specific *syn* guanosine (G4), suggested that Tyr^{177} plays an important role in Z-conformation recognition and binding. (Figure 6)

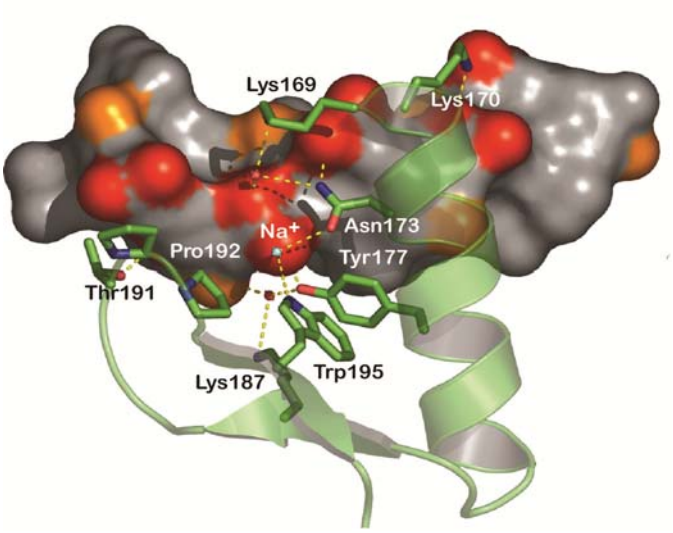


Figure 6 The binding surface between $Z\alpha_{ADAR1}$ and Z-RNA

The surface of a single RNA strand is shown in gray. Ribose 2′ hydroxyl groups are shown in orange, and phosphate oxygen atoms are shown in red. The Zα domain is shown in green. The residues interacting with Z-RNA are labeled and shown as sticks. Hydrogen bonds are indicated as dash lines (Placido, Brown et al. 2007)

I.7 Other Z-DNA binding domains

After the discovery of $Z\alpha_{ADAR1}$, more Z-DNA binding proteins have been discovered and confirmed by computational protein database analyses using the amino acids crucial for Z-DNA binding in Z-alpha. Some examples are shown in Figure 7. Z-DNA binding proteins were found in a variety of species, from viruses to mammals. Some examples are: E3L from vaccinia virus (Liu, Wolff et al. 2001), DLM1 from mammalian species(Ha, Kim et al. 2008), PKZ from zebra fish (Rothenburg, Deigendesch et al. 2005), ADAR family from Drosophila (Nordheim, Tesser et al. 1982), wheat germ(Lafer, Sousa et al. 1985) rat (Rothenburg, Schwartz et al. 2002), chicken (Herbert, Spitzner et al. 1993), and human (Leith, Hay et al. 1988; Pham, Park et al. 2006). Although some of these proteins have not been characterized in details, evidences suggest that they play important roles in DNA recombination, gene expression and regulation, RNA editing, viral pathogenicity, and tumor development (Rich and Zhang 2003).

Figure 7 Domains with conserved sequences for Z-DNA binding

Z-DNA binding domains from different proteins are aligned based on their amino acid sequences, the conserved sequences that essential for Z-DNA binding is highlighted in red. (Placido, Brown et al. 2007)

I.7.1 E3L from vaccinia virus

The Z-DNA binding domain from E3L ($Z\alpha_{E3L}$) shows *in vitro* a comparatively lower affinity to Z-DNA than $Z\alpha_{ADAR1}$ (Kim, Lowenhaupt et al. 2004). NMR analyses showed that the global structure and the Z-DNA interaction surface of $Z\alpha_{E3L}$ are very similar to $Z\alpha_{ADAR1}$. However, the Tyr^{48} residue (conserved as Tyr^{177} in $Z\alpha_{ADAR1}$) which is important for Z-DNA binding, adopts a different side chain conformation in unbound $Z\alpha_{E3L}$ (Figure 8). Thus, binding to Z-DNA requires a rearrangement of this residue, which could explain the lower affinity of $Z\alpha_{E3L}$ to Z-DNA. (Kahmann, Wecking et al. 2004)

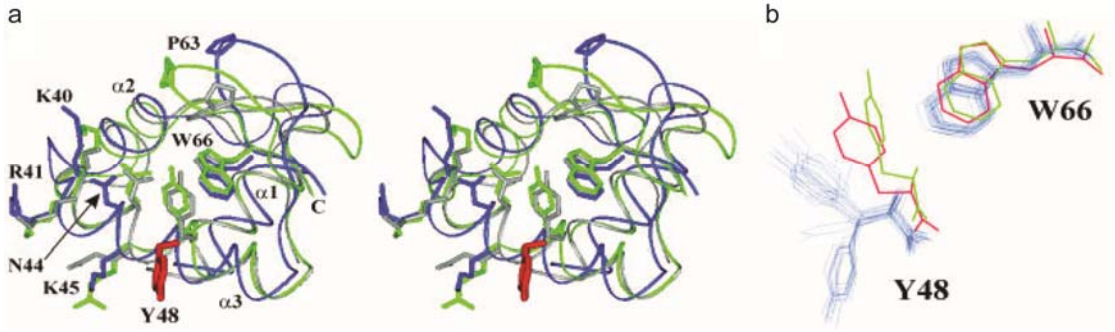


Figure 8 The Tyr^{48} residue adopts a distinct position in $Z\alpha_{E3L}$ as compared with $Z\alpha_{ADAR1}$ and $Z\alpha_{DLM1}$ in solution structure

The 3D structures of $Z\alpha_{E3L}$ (blue), $Z\alpha_{ADAR1}$ (light green), and $Z\alpha_{DLM1}$ (gray) are superimposed. **a.** The side chains of the residues that contact the Z-DNA in the co-crystal structures of $Z\alpha_{ADAR1}$ and $Z\alpha_{DLM1}$ show very similar positions, while the side chain of Tyr^{48} adopts a distinct position in $Z\alpha_{E3L}$ (red). **b.** Distinct Tyr^{48}-Pro^{66} distance between $Z\alpha_{E3L}$ and $Z\alpha_{ADAR1}$. Superposition of Tyr^{48} and Pro^{66} between the 20 lowest energy structures of E3L (blue), unbound $Z\alpha_{ADAR1}$ (red), and bound $Z\alpha_{ADAR1}$ (light green). (Kahmann, Wecking et al. 2004)

The E3L contains a Z-DNA binding motif at the N-terminal and a dsRNA binding motif at the C-terminal, which is similar to the make-up of full length ADAR1. It has been reported that E3L localized to both the cytoplasm and nucleus (Yuwen, Cox et al. 1993; Chang, Uribe et al. 1995). E3L could help virus survival during infection by antagonizing the host antiviral defense (IFN) systems (Davies, Chang et al. 1993; Chang, Uribe et al. 1995; Shors, Kibler et al. 1997). A functional dsRNA domain is required for IFN-resistance of vaccinia. Whether the Z-DNA binding domain could affect the pathogenesis of the virus is still under debate. (Chang, Uribe et al. 1995; Shors, Kibler et al. 1997; Brandt and Jacobs 2001). E3L or $Z\alpha_{E3L}$ alone can selectively increase reporter gene expression in HeLa cells. They have a strong anti-apoptotic capability in a dose- and time-dependent manner. Mutational analysis suggested that the Z-DNA binding ability is vital for gene trans-activation and anti-apoptosis (Kwon and Rich 2005).

I.7.2 DLM-1

DLM-1 was also identified first as Z-DNA binding protein by sequence comparison. $Z\alpha_{DLM1}$ shows a similar Z-DNA binding affinity as $Z\alpha_{ADAR1}$ (Kim, Lowenhaupt et al. 2004). The Z-DNA binding capability has been confirmed by crystal structure (Schwartz, Behlke et al. 2001) and CD analysis.

DLM-1 was firstly detected as a novel gene in peritoneal lining tissues of mouse inoculated with ascites tumors and tumor stromal cells by using RNA differential display (Fu, Comella et al. 1999). Expression of DLM-1 in macrophages was greatly up-regulated by stimulating 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 the defense against tumor formation during neoplasia and angiogenesis. Full length DLM-1 was localized in the cytoplasm with a punctate distribution pattern (Deigendesch, Koch-Nolte et al. 2006). Interestingly, under environmental stresses, 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, Stoecklin et al. 2005), suggesting that DLM-1 may participate in regulation of translation and mRNA sorting. The Z-DNA binding domain appears crucial for the accumulation of DLM in stress granules(Deigendesch, Koch-Nolte et al. 2006).

I.8 Objective of this study

Based on the current knowledge about the occurrence of nucleic acids in Z conformations *in vivo*, as summarized above, we decided to investigate this question in an unbiased manner. We therefore designed a Z-conformation-specific probe and a mutant version that can be used as control.

This probe was previously designed in our lab and is based on the above-mentioned Z-DNA binding domain (amino acid 133-209) from human ADAR1 (Genbank file NP_001102), referred to as $Z\alpha_{ADAR1}$ (Li, Xiao et al. 2009). The $Z\alpha_{ADAR1}$ probe and its mutant were designed as illustrated in Figure 9. The amino acids Asn^{173} and Tyr^{177}, which are crucial for specific Z-DNA binding (Schade, Turner et al. 1999; Kahmann,

Wecking et al. 2004), were substituted by Ala. The control probe was named $Z\alpha_{ADAR1}$mut. The FLAG tag at the C-terminus can be recognized by a monoclonal antibody and the StrepII tag is used for affinity purification using Strep-Tactin beads.

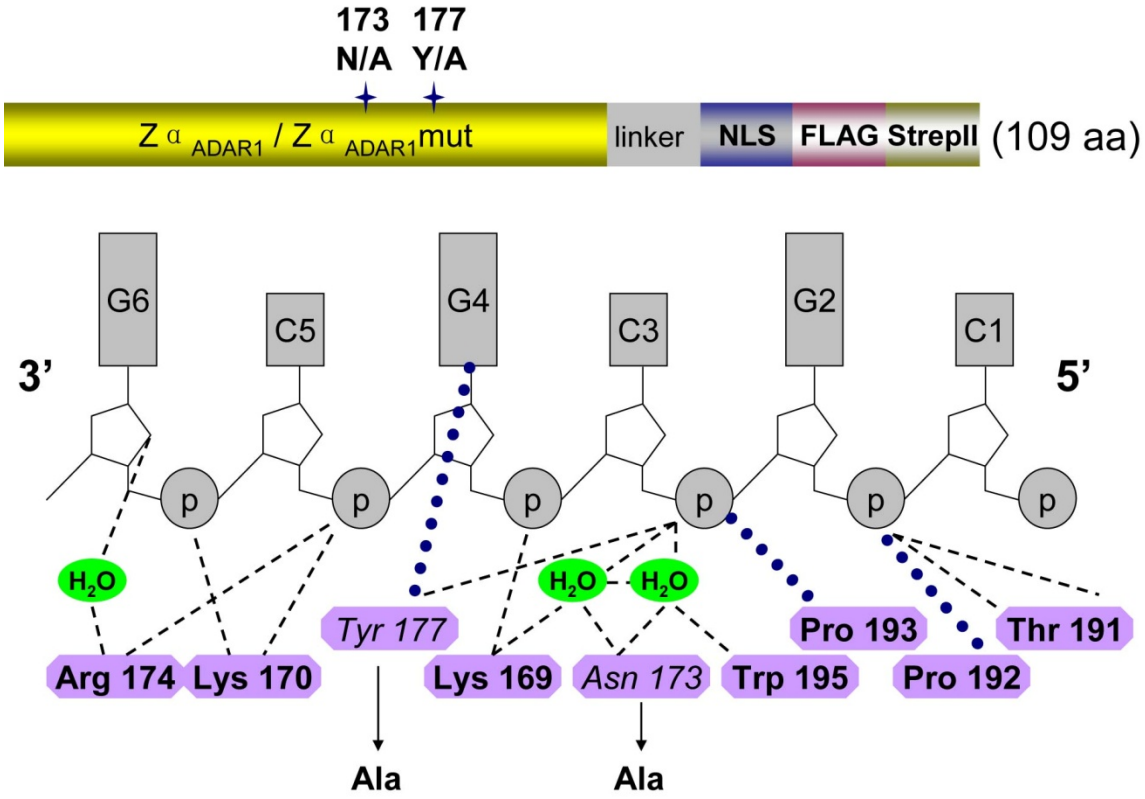


Figure 9 Schematic representation of $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut.

There is a short linker region (GGSGG) between $Z\alpha_{ADAR1}$/mutant domain and the SV40 Nuclear Localization Signal (NLS, PKKKRKV), which is followed by FLAG (DYKDDDDK) and Strep II (WSHPQFEK) epitope tags. Between FLAG and StrepII tags is a short linker composed of SA. Critical interactions between $Z\alpha_{ADAR1}$ and Z-DNA are shown (Schwartz, Rould et al. 1999; Li, Xiao et al. 2009),.

The overall structures of the two probes were almost identical form the NMR study (Feng, Li et al. 2011). The binding specificity to Z-DNA was analyzed in our previous study (Li, Xiao et al. 2009). The $Z\alpha_{ADAR1}$ probe could specifically bind and be cross-linked to Z-DNA, while the mutant form ($Z\alpha_{ADAR1}$mut) lost its binding specificity to Z-DNA,

and more importantly, could not be cross-linked to the Z-DNA. The binding affinity and specificity to Z-DNA is ionic strength-depended. $Z\alpha_{ADAR1}$mut is used as a negative control in this study to identify Z-conformation-specific interactions between nucleic acids and $Z\alpha_{ADAR1}$.

The intracellular localizations of our probes were analyzed in A549 human lung cancer cells (Figure 10). Both $Z\alpha_{ADAR1}$ and the $Z\alpha_{ADAR1}$mut localized not only in the nucleus but also in the cytoplasm. However, it is clear that $Z\alpha_{ADAR1}$ accumulated in the nucleoli regardless of the presence of NLS, while $Z\alpha_{ADAR1}$mut was not found associated with nucleoli. This finding correlated with the observed localization of Z-RNA-specific antibodies after micro-injection into human cells (Zarling, Calhoun et al. 1990). The cellular localization may help to understand the role of $Z\alpha_{ADAR1}$ domain as part of ADAR1-L, which is also reported to localize in both nucleolus and cytoplasm (Eckmann, Neunteufl et al. 2001; Strehblow, Hallegger et al. 2002).

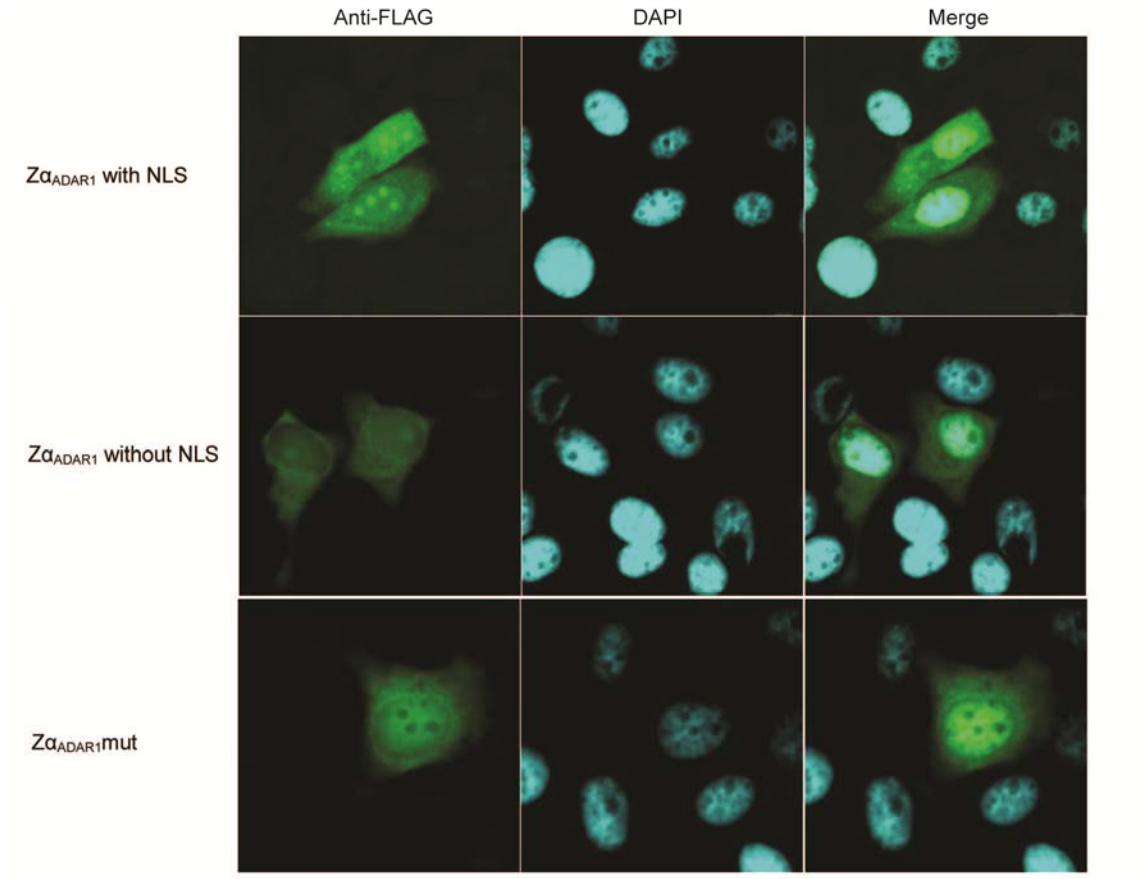


Figure 10 Immunostaining of Zα$_{ADAR1}$(mut) at 24 hours post transfection.

The distribution patterns of Zα$_{ADAR1}$ and Zα$_{ADAR1}$mut were examined by immunostaining using FITC-conjugated FLAG monoclonal antibodies M2 (Sigma). The immunostaining is done 24 hours after A549 cells were transfected with pCMV-Zα(mut)-(NLS)-FS plasmids. The nucleolus accumulation of Zα$_{ADAR1}$ is clearly shown regardless of the presence of NLS. Zα$_{ADAR1}$mut exists in both cytoplasm and nucleus, but not in the nucleolus. This picture is an example from at least 3 independent experiments. (Contributed by Dr. Li Heng)

Equipped with these well-characterized probes, we set out to isolate nucleic acids which formed specific complexes with Z-alpha in prokaryotic and eukaryotic cells. Our results revealed interesting insights into the dynamic nature of RNA and DNA during biological events, such as translation, transcription and formation of special chromatin structures. These results suggest that, besides the plain sequence contexts, the nucleic acid

conformations play an important role in functions and regulations of biological processes.

II MATERIALS AND METHODS

II.1 Plasmids used

II.1.1 Plasmids list

The plasmids are generally constructed by inserting the coding region for $Z\alpha_{ADAR1}$ or $Z\alpha_{ADAR1}$mut into existing expression vector for either *E. coli* or mammalian systems. They are labeled using the expression /antibiotics resistant gene after the respective promoter.

Plasmid	From
pET17b	invitrogen
pET17b-Zα-FS	Our lab's stock (Li, Xiao et al. 2009)
pET17b-Zα-mut-FS	Our lab's stock (Li, Xiao et al. 2009)
pET17b-ZαE3L-FS	This study
pET17b- ZαDLM1-FS	This study
pET17b-Zαα-FS	This study
pET17b-Zαβ-FS	This study
pTRIPZ-Zα-NLS-FS	This study
pTRIPZ-Zα-FS	This study
pTRIPZ-Zαmut-NLS-FS	This study
pPGKss-Zα-NLS-FS-IRES-Puro	This study
pPGKss-Zαmut-NLS-FS-IRES-Puro	This study
pCMVss-KZ-Zα-NLS-FS	This study
pCMVss-KZ-Zαmut-NLS-FS	This study
pCMVss-KZ-Zα-FS	This study
pCMVss-KZ-Zαmut-FS	This study
pEF1-Zα-SV40-Neo	This study
pEF1-Zαmut-SV40-Neo	This study
pEXFH-huADAR1	Stefan Maas (Maas, Gerber et al. 1999)
pEXFH-huADAR1-mut	This study

II.1.2 Construction of plasmids in this study

II.1.2.1 pET17b-ZαE3L-FS, pET17b-ZαDLM1-FS, pET17b-Zαα-FS and pET17b-Zαβ-FS (Figure 11)

The $Zα_{E3L}$, $Zα_{DLM1}$, Zαα or Zαβ genes were sub-cloned from pET-28a-hDLM, pET-28a-vvE3L and pACT- Zαα and pACT- Zαβ plasmid [a gift from A. Rich (MIT)] into pET17b-Zα-FS (Li, Xiao et al. 2009) vector. Zαβ is the Zα and Zβ domains from human ADAR1 linked via the original linker region between the two domains. Zαα is generated from Zαβ by replacing the Zβ with the Zα domain. $Zα_{E3L}$, $Zα_{DLM1}$ are the Z-DNA binding domains from vaccinia E3L and human DLM1, respectively. (Figure 11)

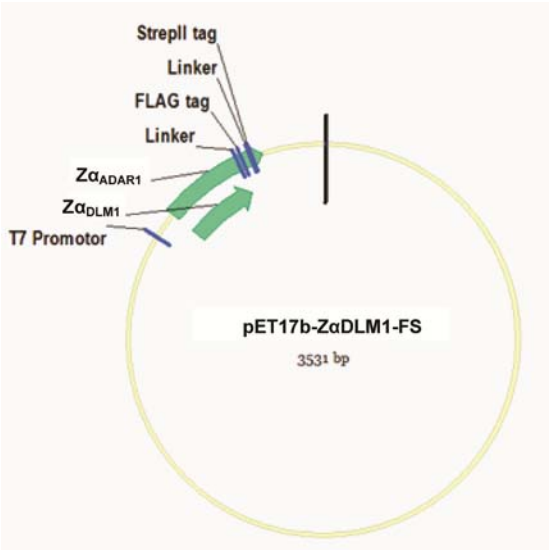


Figure 11 Map of pET17b serial vectors as an example

II.1.2.2 pCMVss-KZ-Zα-NLS-FS,pCMVss-KZ-Zαmut-NLS-FS, pCMVss-KZ-Zα-FS and pCMVss-KZ-Zαmut –FS (Figure 12)

These plasmids were generated by replacing the eGFP gene from pCMVss-eGFP (Clontech) with either $Z\alpha_{ADAR1}$ or $Z\alpha_{ADAR1}$mut. The Kozak sequence was included at the begining of the gene to enhance gene expression. (Figure 12)

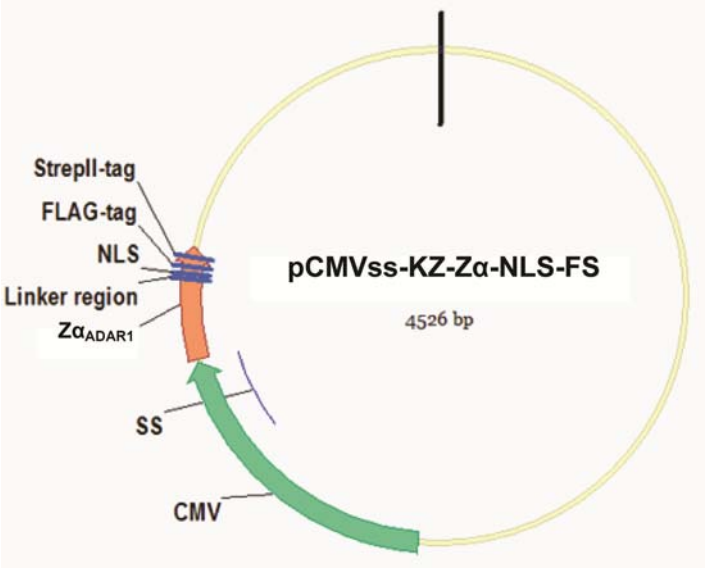


Figure 12 Map of pCMV serial vectors

SS designates the position of a hybrid intron which enhances mRNA stability and, thus, increases protein expression levels.

II.1.2.3 pPGKss-Zα-NLS-FS-IRES-Puro and pPGKss-Zαmut-NLS-FS-IRES-Puro (Figure 13)

These two plasmids were generated by inserting the $Z\alpha_{ADAR1}$ or $Z\alpha_{ADAR1}$mut gene into pPGK-Puro (Invitrogen), the puro gene is replaced by the transcription cassette Zα-IRES-puro. The Zα transcription casette is under the PGK promotor, while the Puromycin resistant gene was linked to Zα via IRES. (Figure 13)

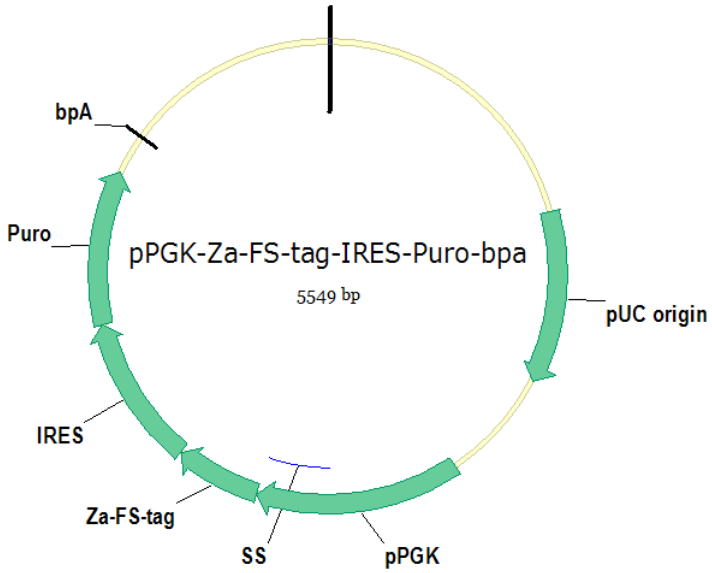


Figure 13 Map of pPGKss- serial vectors

The IRES (internal ribosome entry site), is a nucleotide sequence that allows for translation initiation in the middle of a messenger RNA (mRNA). Thus, the $Z\alpha_{ADAR1}$ and puromycin resistant gene share the same mRNA during expression.

II.1.2.4 pEF1-Zα-SV40-Neo and pEF1-Zαmut-SV40-Neo (Figure 14)

These two plasmids were generated by inserting the $Z\alpha_{ADAR1}$ or $Z\alpha_{ADAR1}$mut gene into the multicloning site in the pEF1-Myc (Invitrogen). The Zα transcription casette is driven by the pEF1 promotor, while the Neomycin resistance gene is under control of the SV40 promotor. (Figure 14)

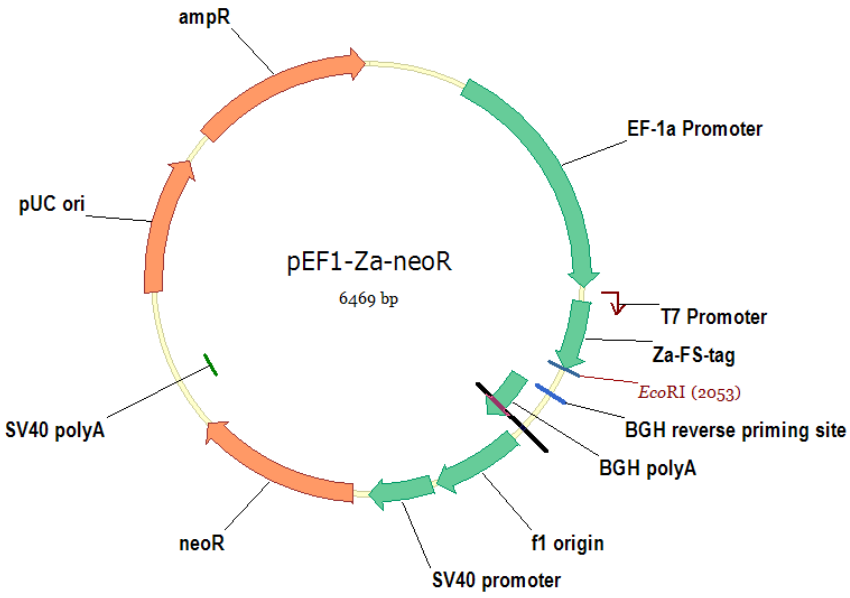


Figure 14 Map of pEF1-SV40-Neo vectors

II.1.2.5 pEXFH-huADAR1 and pEXFH-huADAR1-mut (Figure 15)

pEXFH-huADAR1 is gift from A. Rich (Maas, Gerber et al. 1999). The pEXFH-huADAR1-mut was generated by mutating the Asn^{173} and Tyr^{177} into Ala, in order to introduce the same mutations as in $Z\alpha_{ADAR1}$mut. The mutations were introduced by assembly PCR, and cloned into the *Nde*I restriction sites. The clone with correct orientation was retained after being verified by sequencing. (Figure 15)

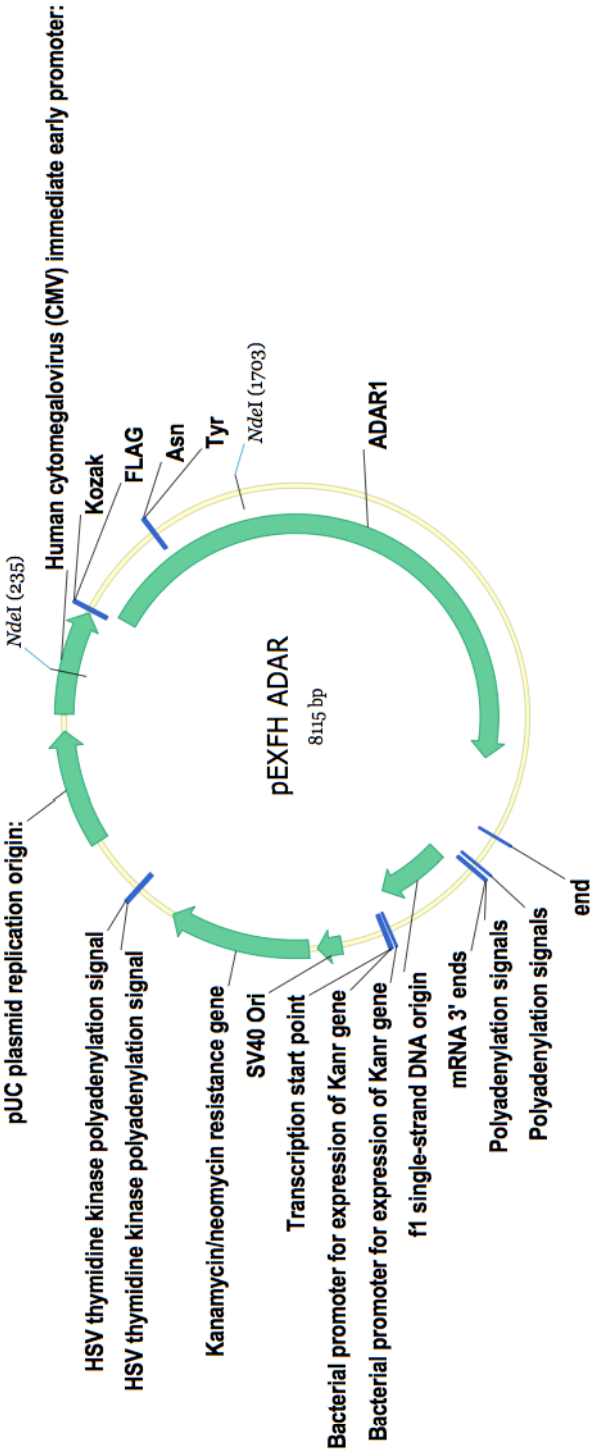


Figure 15 Map of pEXFH-huADAR1 vectors

II.1.2.6 pTRIPZ-Zα-NLS-FS, pTRIPZ-Zα-FS and pTRIPZ-Zαmut-NLS-FS (Figure 16)

The plasmids were constructed by replacing the TurboRFP gene of the pTRIPZ-empty (Open Biosystems) vector with the Zα-NLS-Flag-Strep serial genes sub-cloned from the pPGK vectors. (Figure 16)

The expression of the Zα domains is controlled by the tetracycline response element (TRE), which is controlled by the Reverse tet-transactivator (rtTA3). The rtTA3 is linked to the Puromycin resistant gene with the IRES element, and constitutively expressed. In the absence of Dox, rtTA3 tightly binds Tet operator sequences (tetO), and represses gene expression. That is, the expression of Zα domains is induced by Dox.

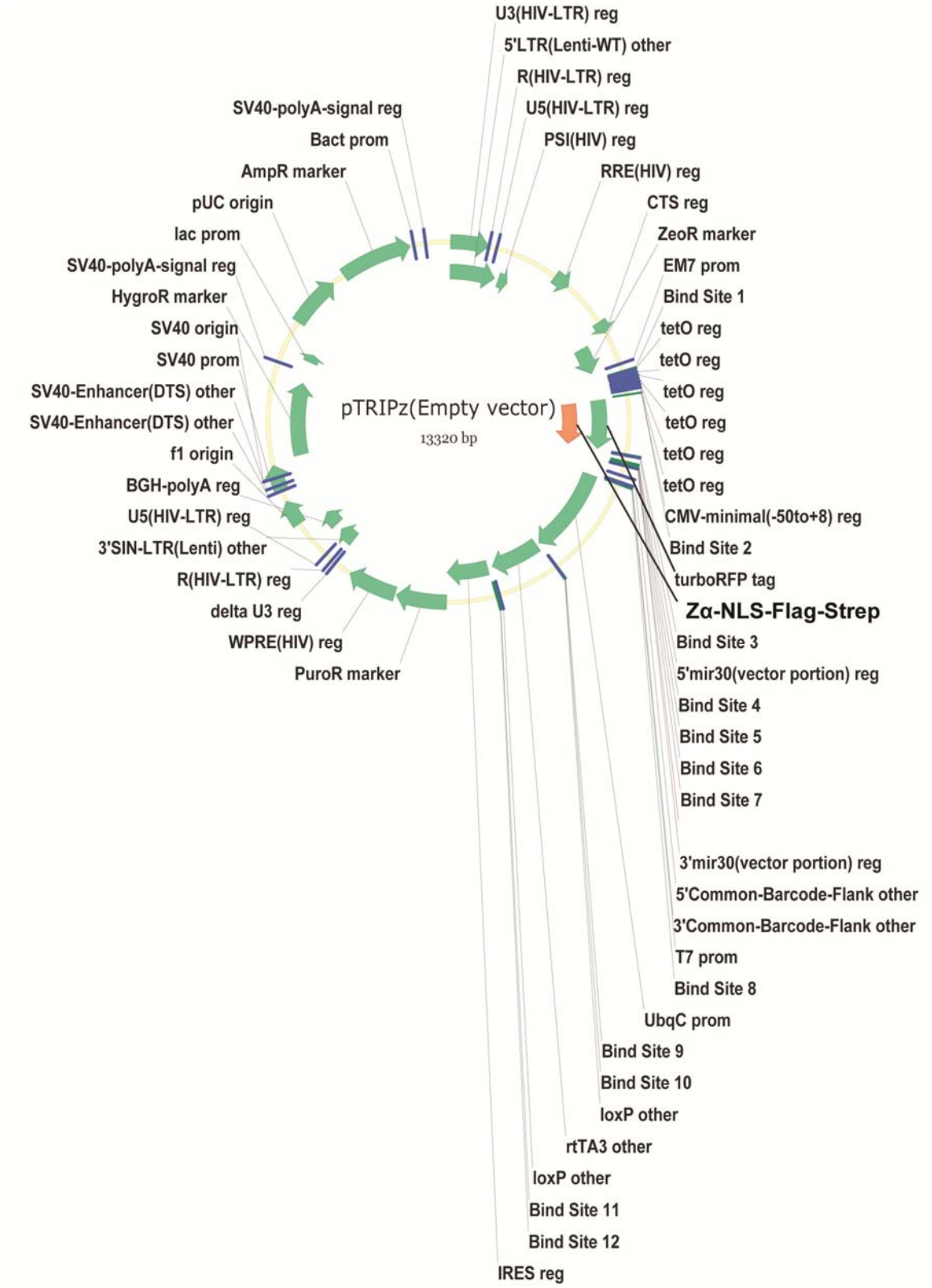


Figure 16 Map of pTRIPZ vectors.

II.2 Kit list

Method	Kit name	Company
Miniprep	Qiaprep Spin miniprep kit	Qiagen
Gel extraction	Qiaquick Gel extraction kit	Qiagen
PCR purification	Qiaquick PCR purification kit	Qiagen
Genomic DNA extraction	Dneasy Tissue Kit	Qiagen
Maxiprep	Endofree Plasmid Maxi Kit	Qiagen
Remove of incorporated ATP	DyeEXTM 2.0 spin kit	Qiagen
Total RNA extraction	RNeasy Mini Kit	Qiagen
Transfection in mammalian cells	Lipofectamine 2000	Invitrogen

II.3 Bacteria strains list

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_β(r_β-m_β-), *dcm*, *gal*, (DE3) (from Stratagen).

STBL4

mcrA □(mcrBC-hsdRMS-mrr) recA1 endA1 gyrA96 gal^- thi-1 $supE44^-$ relA1 (lac-proAB)/F'$proAB^+$ $lacI^qZ$M15 Tn10 Tet^R (from Invitrogen).

II.4 Western Blot analysis

Proteins were resolved on a 8% [for ADAR1(mut)] or 17% (others) SDS-PAGE and transferred onto a PVDF membrane (Immobilon P, Millipore) by wet transfer (Mini Trans-Blot® Electrophoretic Transfer Cell， BioRad) at 30V 4 °C overnight. The membrane was blocked with 5% non-fat dry milk (Bio-Rad) in Tris-Buffered Saline Tween-20 (TBST) for 1 hr and subsequently incubated for 1.5 hrs with anti-FLAG M2 monoclonal antibodies (Sigma) at room temperature. After 3 X 10 mins washes in TBST, the membrane was incubated with anti-mouse peroxidise-conjugated secondary antibody (Pierce) for 1 hour, washed three times in TBST before the secondary antibody was detected using Lumi-Light^{PLUS} Western blotting Substrate (Roche). The signal was either detected by X-Ray film (Kodak) or by Image Quant LAS 4000 imager.

II.5 Cell culture

II.5.1 Human embryonic stem cells (huES)

The culture media for huES were prepared using the following formula: 385 ml Knockout DMEM (Invitrogen Gibco), 5 ml Non-essential amino acid (Invitrogen Gibco), 5 ml Penicillin-Streptomycin (Invitrogen Gibco), 5 ml GlutaMAX-I (Invitrogen Gibco), 50 ml KO-Serum replacement (Invitrogen Gibco), 50 ml plasmanate (Bayer), 4μg Basic Fibroblast Growth Factor (Invitrogen Gibco), 500 μl β-Mercaptoethanol (Invitrogen Gibco). The media was passed through 0.22 μM filter and kept a 4 °C for up to two weeks.

Mouse embryonic fibroblasts cells (MEF) were used as feeder cells. MEFs were inactivated by Mitomycin C (10 μg/ml) for 2 hrs. The inactivated MEFs could be kept at -80 °C or in liquid nitrogen for months.

6-well plates were firstly coated with 0.1% gelatin for a minimum of 30 minutes at 37 °C, and then the MEFs were plated on to the wells at 3.5 X 10^5 cells per well. huES cells were seeded onto the MEFs 24 hrs later. The media was cultured with MEFs for 24 hrs, collected and passed through 0.22 μM filter for later use, and named as MEF media.

MatrixGel (BD Biosciences) was used according to the manufacturer's instruction. When matrix gel was used, MEFs media was used as the culture media for culturing huES.

II.5.2 Other cell types

HeLa, F3, HT1080 or A549 cells were cultured in D3 media including DMEM (Invitrogen Gibco). 10% (v/v) fetal bovine serum (Invitrogen Gibco), 2 mM L-glutamine, 100 units/ml penicillin (Invitrogen Gibco), 100 μg/ml streptomycin (Invitrogen Gibco).

II.6 Transformation and induction of protein expression in BL21(DE3)

BL21(DE3) competent cells were transformed with pET17b-Za-NLS-FS by heat shock. Cells were recovered 40 mins at 37 °C. Bacteria were plated on LB agar with 200 µg/ml ampicillin, 34 µg/ml chloramphenicol and 15 µg/ml tetracycline and incubated at 37 ^{o}C overnight. A single colony was inoculated into 50 ml LB media (200 µg/ml ampicillin, 34 µg/ml chloramphenicol and 15 µg/ml tetracycline) and cultured at 37 °C overnight. Subsequently, bacterial cultures were 1:50 diluted in LB media (200 µg/ml ampicillin, 34 µg/ml chloramphenicol, 15 µg/ml tetracycline) and cultured at 37 °C until OD_{600} reached 0.6. 1 mM (final concentration) IPTG was added to induce expression of $Z\alpha_{ADAR1}$ for 3 hrs. Cells were harvested by centrifugation at 5,000 rcf for 15 min at 4 °C. The pellet was stored at -80 °C. $Z\alpha_{ADAR1}$mut was expressed under the same conditions.

II.7 Purification of $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut

$Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut are purified as previously described (Li, Xiao et al. 2009). Basically, pET17b-Zα-FS and pET17b-Zα-mut-FS were transformed into BL21(DE3). The proteins were induced at OD_{600}=0.6 for three hours at 37 °C. Cells were resuspended in lysis buffer (100 mM Tris, pH = 8.0, 150 mM NaCl, 1 mM EDTA) and lyzed by sonication. The lysates were cleared by centrifugation (30,000 rcf for 20 min at 4 °C) and applied to a Strep-Tactin resin column, which was packed as recommended by the manufacturer (IBA), and washed 5 times with lysis buffer. Proteins were eluted with 1Xbuffer E (IBA)

The eluted material was collected and concentrated (Amicon ultra-15 centrifugal filter devices, Millipore). Proteins were passed through a size filtration column (S-200, Bio-Rad) for further purification. Fractions with $Z\alpha_{ADAR1}$(mut) were collected, and buffer was changed into HEPES Cation chromatography buffer (50 mM HEPES pH = 7.4, 50 mM NaCl, 1 mM DTT, 0.125 mM PMSF) using a concentrator (Amicon ultra-15 centrifugal filter devices, Millipore), applied to the ion-exchange column and developed in a 50-100 mM NaCl gradient.

Fractions with the $Z\alpha_{ADAR1}$(mut) proteins were collected and dialyzed against 50% glycerol (50 mM Tris-HCl, pH = 7.4, 150 mM NaCl, 50% glycerol) at 4 $^{\circ}$C over night. Protein concentration was determined by Bio-Rad protein assay kit following manufacturer's instruction, using BSA as the standard. Purified proteins were stored at -20 $^{\circ}$C. The dialysis buffer was stored in -20 $^{\circ}$C and used as control in the *in vitro* translation assays.

II.8 Stable cell lines expressing $Z\alpha_{ADAR1}$ or $Z\alpha_{ADAR1}$mut

II.8.1 Transfection

Lipofectamine 2000 (Invitrogen) was used for transfection with cells from one 6-well plate as starting material.

II.8.2 Constitutive expression under Puromycin and Neomycin selection

HeLa and F3 cells were transfected with plasmids pPGK-Zα-NLS-FS-IRES-Puro, pEF1-Zα-Neo and their respective mutant plasmids. Puromycin and Neomycin selection started 3 days after transfection. The final concentration for puromycin was 1 μg/ml and for neomycin selection was 500 μg/ml Geneticin (Gibco). When isolated colonies were formed and control cells (without transfection) died (usually two weeks after transfection), single colonies were picked and inoculated into 96-well plates. Protein expression was checked by immunostaining and/or Western blotting.

II.8.3 Induced expression after Puromycin selection

pTRIPZ serial plasmids were introduced in to HT1080 (human fibrosarcoma cells) and F3 (human primary foreskin) cells. 3 μg/ml puromycin selection was applied to cells three days after transfection. Single colonies were picked and inoculated into 96-well plates. Single cell lines were cultured, and protein expression was checked by Western blotting three days after Dox treatment.

II.9 *In vivo* Pull-down of ribosomes from *E. coli* by $Zα_{ADAR1}$

pET17b, pET17b-Zα-NLS-FS, pET17b-Zαmut-NLS-FS, pET17b-E3L-NLS-FS, pET17b-DLM-NLS-FS, pET17b-Zαα-NLS-FS and pET17b-Zαα-NLS-FS were transformed into BL21(DE3) and cultured in LB media. Cell pellets were collected at OD_{600} 0.9 without

IPTG induction. The pellet from a 25 ml culture was dissolved in 1 ml Lysis Buffer (30 mM NaCl, 10 mM Tris-HCl, pH=7.5, 1mM DTT and 10 mM MgCl), and sonicated at amplitude 40, pulse 4 s, for 20 min on ice. The lysate was cleared by centrifugation at 30, 000 rcf for 20 min at 4 ^{o}C and the supernatant was transferred to a new tube. All subsequent operations were performed in the cold (on ice or 4 ^{o}C). 15 µl of Strep-Tactin sepharose beads were added to 200 µl lysate in each tube and incubated on a roller at 4 ^{o}C for 30 min. After washing the beads with Lysis Buffer for 3 times, 10 min each, bound complexes were eluted with 30 µl of Buffer E or by boiling with 30 µl 0.5% SDS for 2 min. 10 µl of eluate were checked on agarose gels and SDS-PAGE.

II.10 *In vitro* Pull-down of ribosomes from *E. coli* by $Z\alpha_{ADAR1}$

The lysate of BL21(DE3) cells was prepared as for *in vivo* pull-down experiments in *E. coli*. 10 µg of purified $Z\alpha_{ADAR1}$, $Z\alpha_{ADAR1}$mut or BSA were added into 200 µl lysate each, and incubated on ice for 30 mins before the pull down experiments. The pulled-down experiments follow the same procedure of the *in vivo* experiment.

II.11 *In vitro* Pull-down of ribosomes from HeLa by $Z\alpha_{ADAR1}$

HeLa cells grown in 175 cm^2 flask at about 90% confluency were collected by trypsin treatment and the cell pellet was resuspended in 1 ml buffer (200 mM NaCl, 10 mM Tris-HCl pH = 7.5, 1 mM DTT and 5 mM $MgCl_2$). All subsequent operations were performed on ice or at 4 ^{o}C. Cells were homogenized with a motor-driven Heidolph

DIAX 900 homogenizer. The settings of homogenizer were Level 5, 30 seconds per pulse, 3 pulses. The homogenate was cleared at 16,000 x g for 10 min at 4 °C, and the supernatant was transferred and divided equally into 3 microtubes. 10 µg of $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut were added; no protein was added to the control. After incubation for 30 min on ice, 15 µl of Strep-Tactin sepharose beads were added to each tube and incubated on a roller at 4 °C for 30 min. After washing the beads with binding buffer 3 times, for 10 mins each, the bound complexes were eluted with 30 µl of Buffer E containing Desthiobiotin. 10 µl of each eluate was checked on agarose gel and 1 µl was analyzed by Western blotting.

II.12 *In vivo* Pull-down of ribosomes from HeLa by $Z\alpha_{ADAR1}$

HeLa cells grown on 10 cm dishes at about 90% confluency was transfected with pCMVss-KZ-Zα-FS, pCMVss-KZ-Zαmut-FS and pCMV-mock respectively. Cell lysates were harvested 24 hrs after transfections. The cell lysate preparation and pull-down procedure follows the protocol of the *in vitro* experiment. The eluates were used for mapping the $Z\alpha_{ADAR1}$ binding sites on HeLa ribosomes *in vivo* after RNase A treatment on beads.

II.13 Co-precipitation of $Z\alpha_{ADAR1}$ with HeLa polysomes

pPGKss-Zα-FS and pPGKss-Zαmut–FS were transfected into HeLa cells on a 10 cm dish each, one dish of un-transfected HeLa was use as “non-protein control”. One day after transfection, cells were UV crosslinked on the plate. Then the cells were collected by scraping, and lysed by homogenizer in lysis buffer (150 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT and 10 mM Tris-HCl pH 7.5). The lysates were cleaned by centrifugation at 30 k rcf for 30 mins on ice. The supernatants were applied to ultracentrifugation tubes containing a 30% (w/v) sucrose cushion (30% sucrose, 500 mM KCl, 10 mM MgCl, 1 mM DTT and 10 mM Tris-HCl pH 7.5), and centrifuged for 2 hrs at 400 k rcf. The polysomes and individual ribosomal subunits will be pelleted at the bottom, and other components like free $Z\alpha_{ADAR1}$ are in the supernatant. (Figure 17) The pellet was washed once with lysis buffer before they were dissolved in water. The pellets were treated with 1 mg/ml RNAse for 30 mins at 37 $^{\circ}$C. The supernatant after ultracentrifugation and the dissolved pellets were analysed for $Z\alpha_{ADAR1}$(mut) by Western blotting.

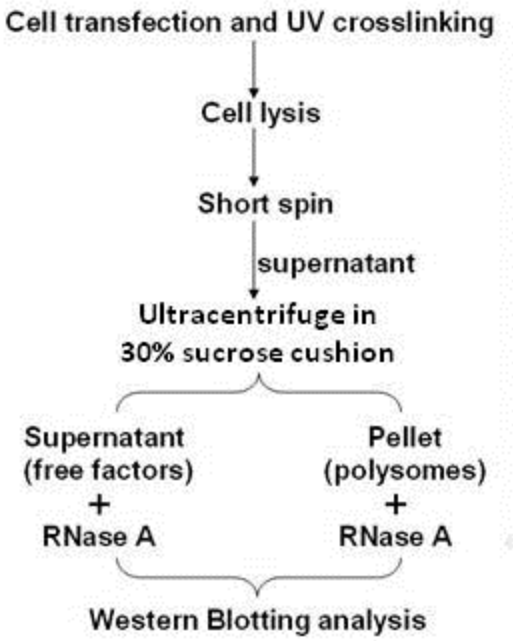


Figure 17 Co-precipitation of $Z\alpha_{ADAR1}$ with HeLa polysome pellet.

II.14 Co-precipitation of ADAR1 with HeLa polysomes

pEXFH–hADAR1 and pEXFH–hADAR1mut were transfected into 10^7 cells. At 48 hrs post transfection, cells were homogenized (150 mM NaCl, 10 mM $MgCl_2$, 1 mM DTT, 10 mM Tris–HCl pH = 7.5) and cleared by centrifugation. Aliquots of the supernatant were kept for Western blotting. The rest was loaded on a 30% (w/v) sucrose cushion in 200 mM NaCl, 5 mM $MgCl_2$, 1 mM DTT and 10 mM Tris–HCl pH 7.5, and centrifuged at 400,000 x *g* for 2 h at 4 ^oC. Pellets were washed once with lysis buffer and dissolved in water. The aliquots of the lysates before centrifugation and the dissolved pellets were test for ADAR1 or ADAR1mut by Western blotting.

II.15 Identification of binding site(s) for $Z\alpha_{ADAR1}$ on *E. coli* and HeLa ribosomes

II.15.1 RNase treatment on beads

After the above pull-down protocol was established, RNase treatment was performed when the ribosomes were still linked to the beads via $Z\alpha_{ADAR1}$. The pull-down process is the same as described above without elution. After 3 times wash with binding buffer, beads were resuspended in 200 µl of binding buffer with 0.5 mg/ml RNase A (Roche). RNase digestion occurred by incubation at 37 °C for 30 min. Then, the supernatant containing RNase A was removed and the beads were washed for three times with binding/lysis buffer to remove RNase A, followed by elution with Buffer E. Half of each eluted material was checked on SDS-PAGE for the existence of other proteins. The rest was purified by phenol/chloroform extraction, followed by ethanol precipitation. Subsequently, the pulled-down RNA fragments were resuspended in 20 µl of dH_2O, and 1 µl was labeled with ^{32}P by T4 polynucleotide kinase (NEB), and resolved on a 12% TBE-PAGE.

Two types of RNase were tested, RNAse A and T1 (Figure 18). The pattern of RNase A and T1 digestion is very different (the second and forth lanes). After further digestion by RNase A, the sample digested by RNase T1 is almost completely digested. This indicates that the pull-downed material is RNA rather than DNA. As the fragments' sizes from RNase A digested sample is smaller and more concentrated, RNAse A was chosen as nuclease used in the cloning method.

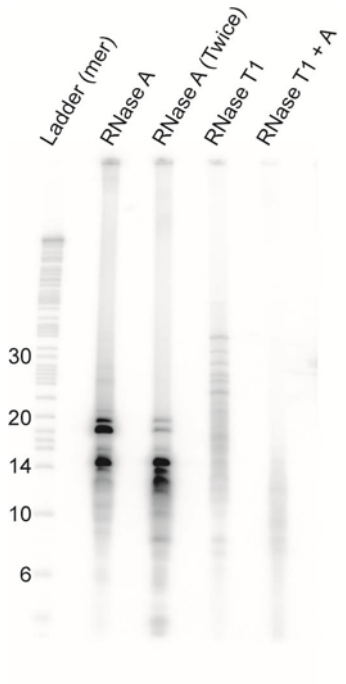


Figure 18 RNase digestion test of pull-down sample.

The pull-downed sample was digested on beads with RNase A or T1. After three times' wash, the samples were eluted and purified, half of the purified samples were labeled (the second and forth lanes). The other half of the purified samples were further digested by RNase A, and labeled (the third and fifth lane)

II.15.2 Cloning of small RNA fragments

The strategy is summarized in Figure 19 below

II.15.2.1 Oligos used in this method

DNA is labeled in upper case, while RNA is labeled in lower case.

3’ end DNA adaptor: 5’-appCTGTAGGCACCATCAAT-NH_2 3’ (NEB)

5’ end RNA adaptor: 5’OH-atcgtaggcacctgaaa-OH 3’ (Damarcon)

3’ end DNA adaptor PCR primer: 5’OH-ATTGATGGTGCCTACAG-OH 3’

5’ end RNA adaptor PCR primer: 5’OH-ATCGTAGGCACCTGAAA-OH 3’

II.15.2.2 Adding adaptors to both ends of the RNA fragments

The pulled-down RNA fragments were treated with CIP to remove 3’ phosphates, followed by phenol/chloroform extraction and ethanol precipitation. The purified RNA fragments were ligated to 3’ end DNA adaptor. The ligation reaction included the pulled-down RNA fragments in 20 µl volume, 0.3 µl of 3’ end DNA adaptor (83 nM), 1 µl RNasin, 2 µl 10 x T4 RNA ligase I buffer and 2 µl of T4 RNA ligase I (NEB). After incubation at room temperature for 4 hrs, reactions were purified by phenol/chloroform extraction and ethanol precipitation.

The first round ligation products were treated with PNK to add the 5’ phosphate, and purified by phenol/chloroform extraction and ethanol precipitation.

Then 5’ end RNA adaptor was added in a similar way. The second ligation (in 20 µl system) was composed by all of the first round ligation products, 1 µl of 5’ end RNA adaptors (50 nM) and 2 µl of T4 RNA ligase I. The ligation products were purified by ethanol precipitation after phenol/chloroform extraction.

II.15.2.3 Reverse transcription and PCR amplification

Reverse transcription was performed following the manufacturer's instruction for the Superscript First-Strand synthesis system kit (Invitrogen). The 3' end DNA adaptor PCR primer was used to initiate reverse transcription to obtain the first strand DNA.

The second strand DNA was synthesized by Taq polymerase, which was primed by the 5' end RNA adaptor PCR. The conditions were: 95 °C for 5 min; 95 °C for 1 min, 50 °C for 1 min, 72 °C for 1 min, 72 °C for 10 min. Each second strand DNA synthesis product was then used in 200 µl PCR reaction volume for amplification. The PCR were performed with the 3' end DNA adaptor PCR primer and the 5' end RNA adaptor PCR primer for 25 cycles under the same condition (95 °C for 5 min; 95 °C for 1 min, 50 °C for 1 min, 72 °C for 1 min; 72 °C for 10 min). PCR products were run on 0.5 × TBE 12% PAGE, and the regions above adaptor dimers were extracted, according to the nearby marker. The gel slices were grounded and submerged in 10 volumes of 300 mM NaCl in TE buffer (pH = 8) overnight, followed by ethanol precipitation.

II.15.2.4 Concatenation and Cloning

The PCR fragments were digested by Ban I for 1 hr. After purification through phenol/chloroform extraction and ethanol precipitation, the fragments were concatenated by ligation with T4 DNA ligase, which occurred at 25 °C for 30 mins. The concatenated fragments were purified by QIAquick® PCR Purification Kit (Qiagen), and CIP treatment was performed to remove 5' phosphates. After purification by the PCR purification kit, one 3' over-hang adenine was added by Taq polymerase. The reactions

were performed by incubation at 72 °C for 30 min with 1 µl of Taq DNA polymerase (Promega). Afterwards, DNA fragments with one over-hang A were ligated into Topo vectors using TOPO TA Cloning® Kits (Invitrogen). After ligation, plasmids were transformed into ElectroMAX™ Stbl4 Cells (Invitrogen) by electroporation. All transformed cells were plated out on LB plates with 150 µg/ml ampicillin, 1 mM IPTG and 50 mg/L X-gal. After incubation at 30 °C overnight, white colonies were inoculated and plasmids were isolated for sequencing using M13 forward primer.

The concatenated DNA fragment sequences were identified from the sequencing results and matched to rRNA sequences, either 18s, 28s and 5.8s rRNA sequences for HeLa or 16s, 26s and 5s rRNA for *E. coli*. The results were plotted in such a way that the X-axis represents the positions on the respective rRNA, while the Y-axis represents the numbers of cloned fragments appeared in that position.

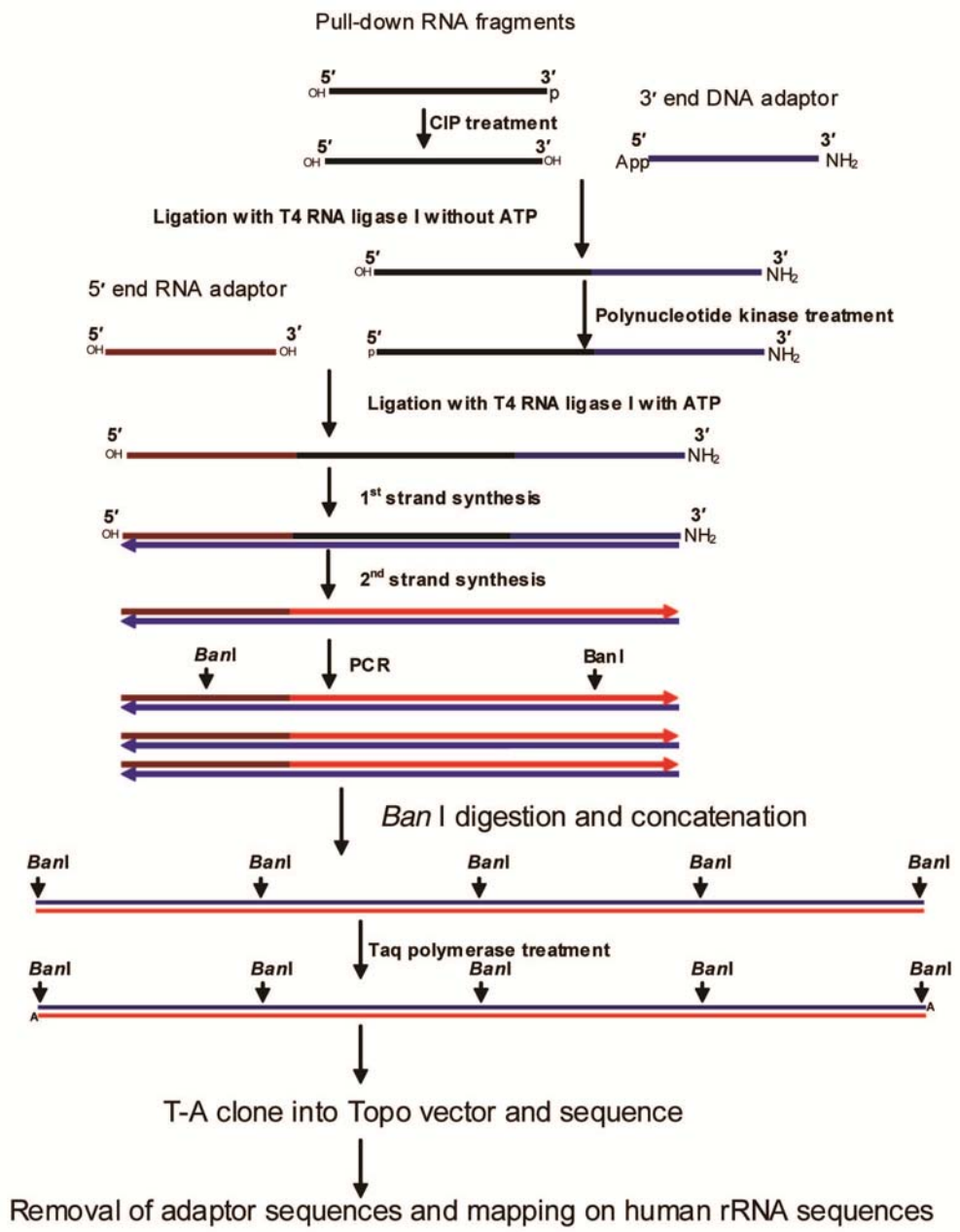


Figure 19 Cloning of small RNA fragments

II.16 Effect of $Z\alpha_{ADAR1}$ on *in vitro* translation

II.16.1 RNase contamination test in purified proteins

Before I tested the effect of $Z\alpha_{ADAR1}$ and the $Z\alpha_{ADAR1}$mut on *in vitro* translation, possible RNase contaminations in purified proteins were checked. 4 µg and 8 µg of either $Z\alpha_{ADAR1}$ or $Z\alpha_{ADAR1}$mut were mixed with 500 ng purified total RNA, with and without 2 µl RNasin in protein dialysis buffer. After 90 mins incubation at 30 °C, samples were analyzed on 1.2% agarose gel.

II.16.2 Effect of $Z\alpha_{ADAR1}$ and its $Z\alpha_{ADAR1}$mut on *in vitro* translation in rabbit reticulocyte lysate (RRL)

Flexi rabbit reticulocyte lysate (RRL) *in vitro* translation system (Promega) was used following the manufacturer's instructions. First, a master mix was prepared, as indicated in Table 1. Then 60 µl of master mix was aliquoted into 3 tubes, labeled as C1 (for control), Z1 (for $Z\alpha_{ADAR1}$) and M1 (for $Z\alpha_{ADAR1}$ mutant). Then 12 µl of protein dialysis buffer was added into the remaining master mix, which was then divided into 12 tubes, 31 µl per tube and labeled as C2 to C5, Z2 to Z5 and M2 to M5. 2 µl of protein dialysis buffer, 2 µl of $Z\alpha_{ADAR1}$ (8 µg) or 2 µl of $Z\alpha_{ADAR1}$mut (8 µg) were added into C1, Z1 and M1, individually. After gentle mixing, 31 µl from C1, Z1 and M1 was transferred into C2, Z2 and M2, individually. C2, Z2 and M2 were then serially diluted into the corresponding tubes. Materials in each tube were then divided into three tubes (10 µl each). Then all tubes were incubated at 30 °C for 90 minutes. After incubation, tubes were put on ice to

stop translation. The amount of luciferase reporter in each tube was quantified by luminance unit, which was obtained out by the reaction with 40 µl of luciferase substrate in a Turner Designs TD 20/20 luminometer.

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 1 preparation of master mix of *in vitro* translation reaction

II.16.3 *In vitro* translation assay with B-DNA competitor

In order to clarify that the inhibitory effect of *in vitro* translation by $Z\alpha_{ADAR1}$ is due to its specific Z-conformation binding, B-DNA competitors were used. The translation assay was prepared as describe above, except that different amount of B-DNA fragments (0 µg, 0.18 µg, 0.36 µg, 0.54 µg, 0.72 µg) were pre-bound to $Z\alpha_{ADAR1}$(mut) before adding to the RRL. The B-DNA fragments were generated by sonication of purified human genomic DNA.

II.16.4 Pull-down of RRL with B-DNA competitor

10µg $Z\alpha_{ADAR1}$ or $Z\alpha_{ADAR1}$mut were pre-incubated with different amount of B-DNA fragment before addition to the RRL. The same protocol was applied as before for the pull-down experiment containing the following salts (10 mM Tris-HCl, pH = 7.4, 100 mM KCl, 1 mM MgCl, 1 mM DTT), which is the same as in translation assays. The pulled-

down material was analyzed on agarose gels and by Western blotting to determine the relative amount of ribosome and $Z\alpha_{ADAR1}$(mut) protein.

II.17 Effect of $Z\alpha_{ADAR1}$, $Z\alpha_{ADAR1}$mut, ADAR1 and ADAR1mut on *in vivo* translation in HeLa cells

0.5 µg pPGK–luciferase was co–transfected with 1 µg of corresponding expression vectors or pCMV-mock into 1.5 x 10^4 HeLa cells using Lipofectamin 2000 (Invitrogen) on 24-well plates. 24 h post transfection, cells were lysed in Lysis Buffer and analyzed by Luciferase Assays (Promega).

II.18 Effect of $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut on *in vitro* translation in *E. coli*

E. coli S30 Extract System for Linear Templates (Promega) was used for testing the effect of $Z\alpha_{ADAR1}$ on *in vitro* translation in *E. coli*. The master mix is prepared following the manufacturer's instructions, while the latter preparation steps are identical as described for the RRL system. The *E. coli in vitro* translation system used is a translation/transcription coupled system, where the DNA template is added.

II.19 Effect of $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut on *in vivo* translation in *E. coli*

Overnight cultured pET17b-Zα-FS, pET17b-Zαmut-FS or pET17b transformed BL21(DE3) were inoculated into fresh LB medium (containing 200 μg/ml ampicilin, 34.5 μg/ml tetracyclin and 15 μg/ml chloramphenicol), to give an initial OD_{600} at about 0.02. OD_{600} were measured at different time points. If induction is applied, the IPTG concentration was 1 mM.

II.20 Measurement of ribosome content in *E. coli* during growth phases

BL21(DE3) was transformed with pET17b-Zαmut-NLS-FS, mock include or pET17b-Zα-NLS-FS. Total RNA was extracted from bacteria harvested at the beginning (OD_{600} = 0.2) or end (OD_{600} = 2.0) of the log phase. The amount of the total RNA was measured at OD_{260}. Samples were also analyzed on agarose gels.

II.21 *In vitro* Chromatin Affinity Precipitation (ChAP) experiment using huES

Three wells of a 6-well plate with 70% confluence huES cells were transferred to three wells of Matrix-Gel w/o MEFs, and cells were cultured for 2 days before the ChAP experiments.

huES cells were crosslinked with 1% formaldehyde (FA) in culture at 37 °C for 5 mins. The crosslinking was stopped by the addition of glycine to a final concentration of 125 mM to inactivate the FA for 5 min. The cells were washed three times with cold PBS. Cells were incubated with 1% triton-X100 in cold PBS for 5 mins at 4 °C, and washed twice with cold PBS and twice with cold HEPES binding buffer (50 mM HEPES-KOH, pH = 8, 150 mM NaCl, 1 mM EDTA, protease inhibitor cocktail). 5 µg of purified $Z\alpha_{ADAR1}$, $Z\alpha_{ADAR1}$mut and BSA were added into each well, 1 ml/well in HEPES binding buffer. After incubation for 5 hrs at 4°C, the cells were washed 15 min X 4 with HEPES binding buffer at 4 °C. The second crosslinking was performed with 0.5% FA for 5 mins and stopped with 125 mM glycine for 5 mins. Cells were washed with ice-cold PBS 3 times, scraped and collected by centrifugation at 600 rcf for 5 mins. The cell pellets were resuspended in chromatin precipitation lysis buffer (50 mM Tris-HCl pH = 8.0, 0.1 mM EDTA, 0.5 mM EGTA, 140 mM NaCl, 10% glycerol, 0.5% NP-40, 0.25% Triton X-100). Cells were pelleted again by centrifugation at 1500 rcf for 5 mins at 4 °C, and then resuspended in nuclei wash buffer (10 mM Tris-HCl pH = 8.0, 1 mM EDTA. 0.5 mM EGTA, 150 mM NaCl). The nuclei were collected at 1500 rcf for 5 mins at 4 °C. Pellets were resuspended in 200 µl of 1% SDS lysis buffer (50 mM Tris-HCl, pH = 8.0, 10 mM 0f EDTA, 1% SDS), and sonicated for 2 mins (3 seconds per pulse) on ice at 1W output.

The cell lysates were cleared by centrifugation at 16,200 rcf for 10 mins at 4 °C and cleared supernatants were collected and diluted with ChAP dilution buffer (16.7 mM Tris-HCl, pH = 8, 0.01% SDS, 1.1% Triton X-100, 1.2 mM EDTA, 167 mM NaCl,

protease inhibitor) to 2 ml. 10 µl of BSA blocked Strep II beads (IBA) were added into each sample and incubated on a roller for overnight.

Beads were collected at 100 rcf for 2 min at 4 °C, and washed 5 times with cold RIPA buffer containing urea (50 mM Tris-HCl, pH = 8, 1% NP-40, 1% sodium DOC, 0.1% SDS, 1 mM EDTA, 1 M NaCl, 1.5 M Urea, protease inhibitor cocktail), each time for 15 mins. Then the beads were washed 3 times with cold TBS (20 mM Tris-HCl pH = 8, 150 mM NaCl, 1 mM EDTA, protease inhibitor cocktail) each time for 15 mins.

The ChAP material were eluted with 90 µl of 1x Buffer E (IBA), Protease K (Roche) were added into each sample and incubation at 65°C overnight to reverse crosslinking. 10 µl of 3 M KAcO, and 1 µl of 20 mg/ml glycogen were added into each sample, and mixed thoroughly. DNA was purified by phenol/chloroform extraction and ethanol precipitation. Samples were resuspended in H_2O and stored in -80 °C.

II.21.1 Quality-check of ChAP material

The ChAP material were blunt-ended by T4 DNA polymerase (10 units of T4 DNA polymerase for NEB, 1 µg/ml BSA, 200 µM dNTP, 10 µl buffer from NEB in 100 µl reaction system) at 12 °C for 15 mins. The reaction was stopped by adding 1 µl of 0.5 M EDTA pH = 8.0. DNA samples were purified via phenol/chloroform extraction and ethanol precipitation.

The purified samples were ligated to 50 pmol ChAP cloning adaptor (Figure 17) with T4 DNA ligase (NEB) at 16°C overnight in 50 µl reaction volume.

Figure 17 ChAP cloning adaptor

This adaptor is phosphorated at the 5' of the ligation end, thus they can self-ligated to form dimers at this end. The two restriction sites were indicated in different colors. The upper oligo is named as Adaptor-L-ChAP, and the lower one is named as Adaptor-S-ChAP.

One percent of the ligated material was amplified by the longer Adaptor primer with 30 PCR cycles (95 °C for 5 mins; 95 °C for 30 seconds, 65 °C for 30 seconds, 72 °C for 1 min; 72 °C for 7 mins). The PCR products were analyzed on 1.5% agarose gels to determine the length distribution of fragments and their relative quantities.

Products above the adaptor dimer region were gel extracted, digested by *Eco*RI and cloned into linearized plasmid pTZ18R (Pharmacia). The cloned library was transformed into Stbl4 cells, and plated on LB with IPTG and X-Gal. Ten white colonies of each sample were picked for initial sequencing to analyze the quality of the ChAP material.

II.21.2 Preparation of ChAP samples for Chip-Sequencing

Because the amount of the ChAP material was not sufficient for Chip-Sequencing, the rest (99%) of the ligated material was amplified by PCR using the Adaptor-L-ChAP as previously described (Figure 17). 23 cycles were applied for $Z\alpha_{ADAR1}$mut and 20 cycles

were applied for $Z\alpha_{ADAR1}$, to adjust for the different amounts of ChAP material. PCR products were purified and eluted by water. The yield was about 240 ng for $Z\alpha_{ADAR1}$mut and 180ng for $Z\alpha_{ADAR1}$. Purified PCR products were digested with *Bse*RI, purified and sent out for sequencing to Beijing Genomics Institute (BGI) by using the "Illumina ChIP–Seq" technique (Section II.21.3).

II.21.3 Illumina ChIP-Seq

The "Illumina ChIP-Seq" method was developed recently. By using the Genome Analyzer and Solexa Sequencing technology, a single sequencing run can scan for genome-wide associations with high resolution. It requires lower DNA input (as low as 20 ng input DNA) and can give up to million-scale sequencing result.

Generally, the ends of samples are modified and two adapters are ligated to the fragments. The ligated samples were purified and applied to a flow cell surface coated with ss oligonucleotides that correspond to the sequences of the adaptors. Single-stranded, adapter-ligated fragments are bound to the surface of the flow cell exposed to reagents for polymerase-based extension. The sequence is amplified locally to give a cluster at each channel. The flow cell containing millions of unique clusters is loaded into the sequencer for automated cycles of extension and imaging. The DNA polymerase adds one fluorescent labeled nucleotide per time, generating a series of images each representing a single base extension at a specific cluster. Reports of useful reads range from 26-50 bases are generated. The sequencing results are then ready for bioinformatics analysis. (Figure 20)

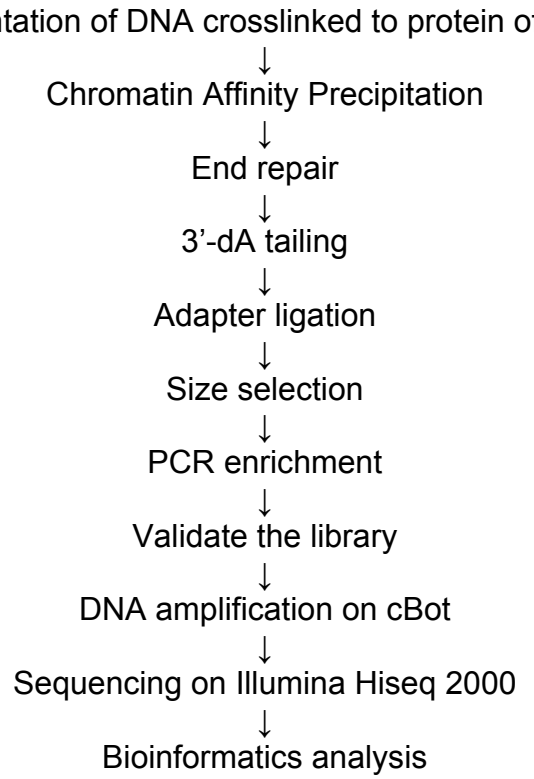


Figure 20 Flow chart of Illumina ChIP Seq steps

II.21.4 Data collection and analyses

The ChAP samples were sequenced by the BGI. Here, 35 sequencing cycles were performed, that is each sequencing result had a length of 35 nt. As 4 nucleotides (CGAC) were left from the adaptor region, the informative sequences had a length of 31 nt.

II.21.5 Selection on raw data

One third of the returned sequences came from the full adaptor. This may come from the inefficient removal of adaptor dimers by the PCR column. These sequences were removed from the sequence library.

In principle, all sequences should have a CGAC left from the adaptor at the 5'. However, about one third of them did not have the CGAC at 5'. One explanation could be that the restriction enzyme cut at a site within the ChAP sequences. Another explanation could be there are contaminations from other sources. In order to eliminate these ambiguities, only sequences which began with CGAC were retained.

When sequences were examined more carefully, a particular sequence could be found up to thousands of times. This may come from the repeated pull down of the same sequence by the probe, or from PCR amplification. There is no way to distinguish between the two possibilities. Considering different sequences may have different PCR amplification efficiencies, the enrichment for identical sequences cannot be considered as true enrichment of binding sites; all identical sequences were therefore combined as one in our library for downstream analyses. (Figure 21)

II.21.6 Map to the human genome

The human genome database HG19 were retrieved from the UCSC Genome Brower webpage (http://genome.ucsc.edu/), and used as a template for mapping. The 5' CGAC

was removed from all selected sequences, and the remaining 31 nucleotides were blasted against the HG19 database. As fragments with only 31 nucleotides in length were used, the mapping criteria were set in such a way that the sequencing result must be 100% identical to the genome without gaps. About one third of the sequences could be mapped to the human genome using these criteria. Some of them could be mapped to more than one position and were excluded from the library.

When the ChAP samples were prepared for sequencing, only fragments that had a length of about 200 bp were selected. The length of identified genomic fragments derived from mapping the obtained sequences was assigned to be 200 bp by adding 169 nt to the respective identified 3' ends. The resulting sequence library is named "Original Library". (Figure 21)

II.21.7 Other sorting criteria

Here, $Z\alpha_{ADAR1}$mut sample was included as negative control. As the mutation in $Z\alpha_{ADAR1}$mut leads to the loss of Z-DNA binding specificity, it could be used as the reference to distinguish the Z-specific sites from background. However, $Z\alpha_{ADAR1}$mut exhibits a substantial B-DNA binding affinity, which could lead to binding to chromosomal positions that are easy to be accessed. Z-DNA is predicted to form in less compacted regions in the chromosome, and less compacted regions are also regions that are easy to be accessed for protein binding in general. That is, if a region is enriched in both $Z\alpha_{ADAR1}$mut and $Z\alpha_{ADAR1}$ ChAP sequences, it doesn't necessarily mean that this region is not specifically bound by $Z\alpha_{ADAR1}$. However, if the region is enriched

for $Z\alpha_{ADAR1}$, but not $Z\alpha_{ADAR1}$mut, it should be Z-conformation specific. A library was generated by removing overlapping sequences between $Z\alpha_{ADAR1}$mut and $Z\alpha_{ADAR1}$ from both libraries. This library was named as “Library with Remaining Sequences”. (Figure 21)

A further classification was generated by assembling the sequencing results into clusters. If sequences overlap, they were retained; while, if they were standing alone (solo), they were removed. This is based on the idea that the pulled-down fragments should be more enriched in binding sites rather than background sites. This library was named “Library with Clustered Sequences”. (Figure 18)

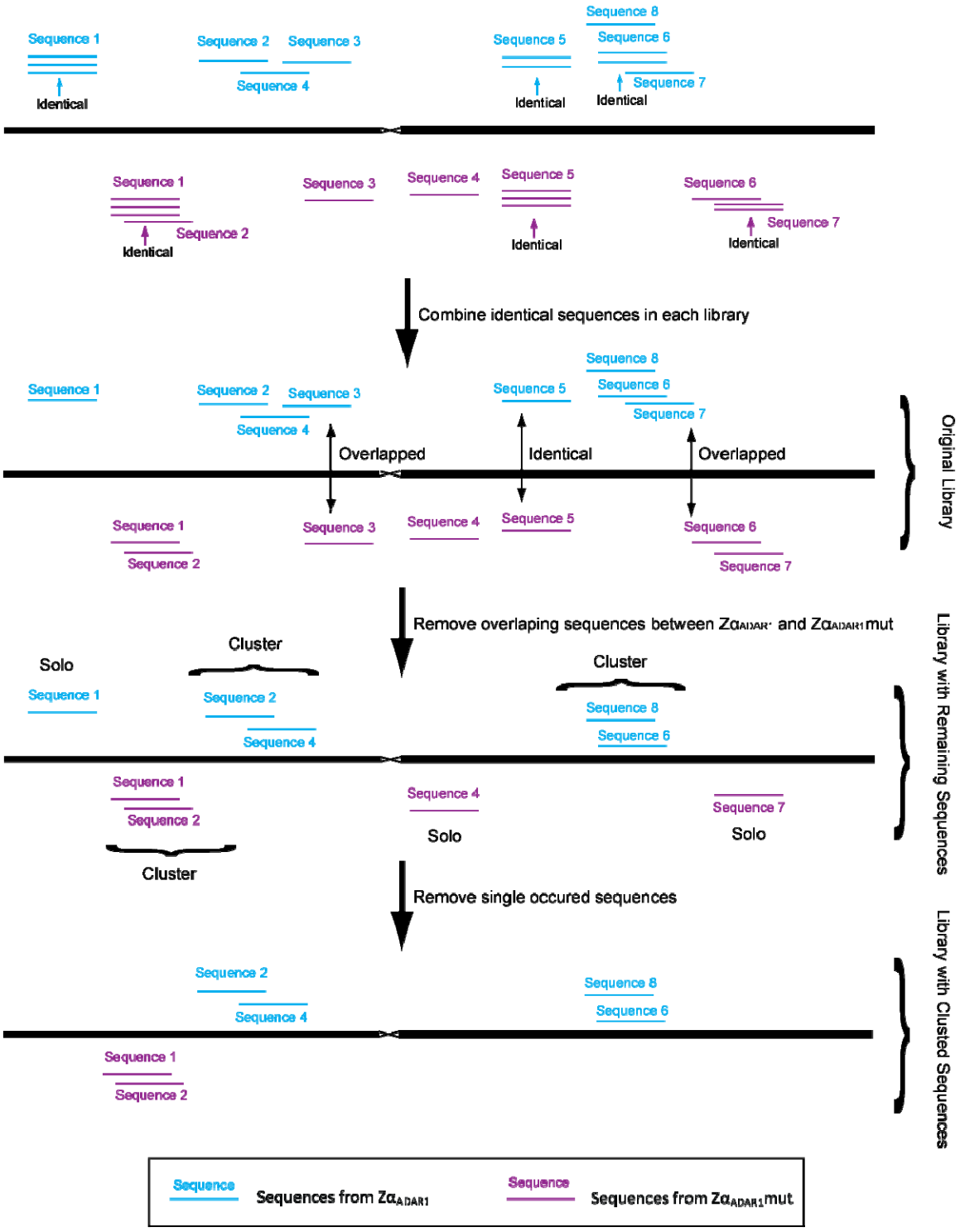


Figure 21 criteria for built up Z-DNA libraries

Different criteria were used to create different libraries from the ChAP sequencing results. The graph shows examples for the selection criteria. The sequences from $Z\alpha_{ADAR1}$ ChAP results are highlighted in cyan, while sequences from $Z\alpha_{ADAR1}$mut ChAP results are in magentas.

II.21.8 General map of Z-DNA distribution on chromosome

Sequences were mapped to each chromosome. The total number of hits within each 10 kb genomic region was counted (Y axis) and plotted against the position on the chromosome (X axis). The 10 kb windows are counted from the centromere region towards the two chromosomal ends. Since a 10 kb region compared to 200 bp individual fragment size is rather large, only the starting position of each sequence was used. Both the “Original Library” and the “Library with Remaining Sequences” were used in this analysis.

II.21.9 Z-DNA distribution around the Transcription start site (TSS)

The “UCSC knownGenes” is downloaded from the UCSC Table Browser (assembly: “Feb. 2009 GRCh37/HG19”; group: “Genes and Gene Prediction Tracks”; track: “UCSC Genes”; table: “knownGene”). The TSS positions were retrieved and the sequencing results were mapped to it. Nearby ±2000 bp regions were divided into four hundred individual 10 bp-long segments. ChAP sequences that overlapped with each 10 bp region were counted. The map is plotted as counting each 10 bp region (Y axis) against the distance from the TSS (X axis). The “Library with Clustered Sequences” and the “Library with Remaining Sequences” were used for this analyzed.

II.21.10 Z-DNA distribution on G-band

The file “cytoBand.txt.gz” is downloaded from the database UCSC HG 19 database (http://hgdownload.cse.ucsc.edu/downloads.html). Sequences were mapped to each

chromosome band, and the number of hits that overlapped with each band was determined. If the fragment is at the edge of the two bands, it was counted twice (each band once). In order to see whether there is an enrichment of sequences toward the band type, the total hits for each band type was divided by the total length of the band type. As the sequences that are overlapped with two bands were counted twice, the reference length for the total band length is calculated as: reference length = total band length + Number of bands X 200 bp. Our “Library with Clustered Sequences” was used for this analysis.

II.21.11 Z-DNA distribution on CpG islands

The file “CpG islands” is downloaded from the UCSC Table Browser (assembly: “Feb. 2009 GRCh37/HG19”; 2009 group: “All Tracks”; track: “CpG Islands”). Sequences that overlapped with CpG islands were counted. The total length of the overlap region was also calculated. The “Library with Clustered Sequences” was used in this analyzed.

II.21.12 SNP density on Z-DNA

The file “SNP(131).gz” is downloaded from the database UCSC HG 19 database (http://hgdownload.cse.ucsc.edu/downloads.html). The number of sequences that contained SNP was determined. The number of SNP on each sequence was also counted. The sequences were grouped based on the distance from the centromere region. The “Library with Clustered Sequences” was used for this analysis.

III RESULTS

III.1 Protein purification

Proteins were first purified through affinity chromatography by Strep-tactin coated beads which recognize StrepII tag at the C-terminus of $Z\alpha_{ADAR1}$. The fractions from affinity chromatography were checked on SDS-PAGE (Figure 22 **a**). The fractions containing $Z\alpha_{ADAR1}$ showed many contaminating proteins.

The fractions containing $Z\alpha_{ADAR1}$ were combined and applied to size-exclusion chromatography. Two protein peaks appeared (Figure 22 **b**). According to SDS-PAGE, $Z\alpha_{ADAR1}$ existed mainly in the second peak (Figure 22 **c**), and the impurities were mainly in the first peak. As the impurities were not fractionated according to their molecular weight when passed through the size-exclusion column, they may exist in complexes.

The fractions in the second peak from size-exclusion column were combined and further purified by ion-exchange chromatography. Purified proteins were checked on SDS-PAGE (Figure 22 **d**). No visible impurity was observed in the purified proteins.

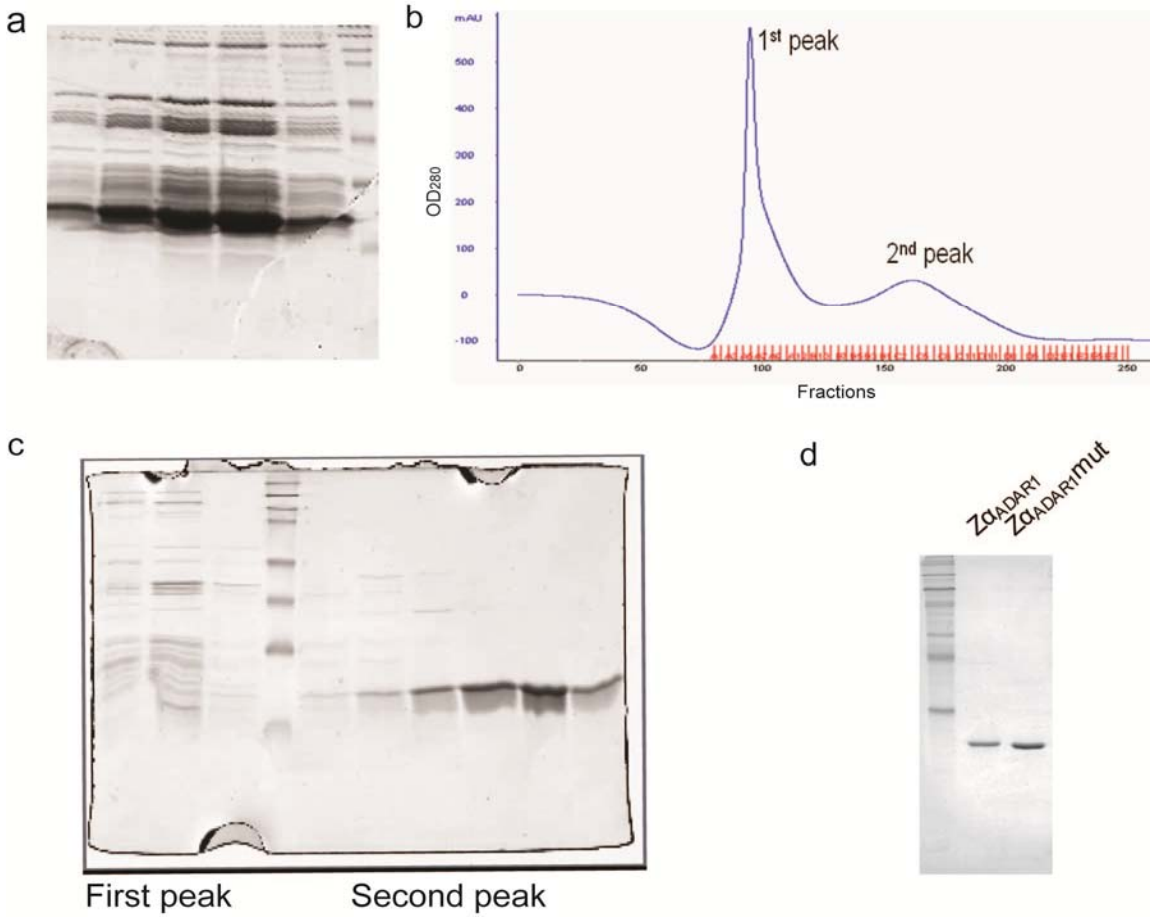


Figure 22 Protein purification

a. Eluates of $Z\alpha_{ADAR1}$ after StrepII beads purification **b.** Purification of $Z\alpha_{ADAR1}$ through a size-exclusion column. **c.** The fractions from the first peak and second peak were analyzed on 17% SDS-PAGE and visualized with Coomassie blue staining. **d.** Purity of proteins was checked on SDS-PAGE. The molecular weight of $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut was about 12.1 kDa, as predicted.

III.2 Evidence for RNA binding activity of $Z\alpha_{ADAR1}$ *in vivo*

During our experiments, there was some evidence suggesting that $Z\alpha_{ADAR1}$ had RNA-binding ability. Firstly, when $Z\alpha_{ADAR1}$ was over-expressed in *E. coli* BL21(DE3) and purified via Strep-Tactin column, RNA contamination was detectable. When size filtration purification was performed, there was one major protein peak before the peak of $Z\alpha_{ADAR1}$ protein. Nuclease sensitivity tests and SDS-PAGE analysis suggested that this peak contains RNA-protein complexes (Figure 23, **a** and **d** [right]). Secondly, when *in vivo* ChAP experiment was performed in mammalian cells, the majority of precipitated material is RNA and not DNA (data not shown). Since it is well characterized that $Z\alpha_{ADAR1}$ is a Z-DNA specific binding protein, DNA molecules were expected to be the major precipitated material. Considering $Z\alpha_{ADAR1}$ is one part from ADAR1, which is a RNA editing enzyme and contains three dsRNA binding domain, it is possible that $Z\alpha_{ADAR1}$ can also bind and recognize RNA.

Both structural analysis and *in vitro* experiments have proven that $Z\alpha_{ADAR1}$ binds Z-RNA specifically (Brown, Lowenhaupt et al. 2000; Koeris, Funke et al. 2005; Placido, Brown et al. 2007). However, in these experiments the RNA substrates were chemically synthesized or artificially designed RNA molecules, and not naturally occured. Therefore, identification of naturally occurring RNA molecules as binding targets for $Z\alpha_{ADAR1}$ and an investigation of the biological significance became interesting at this point.

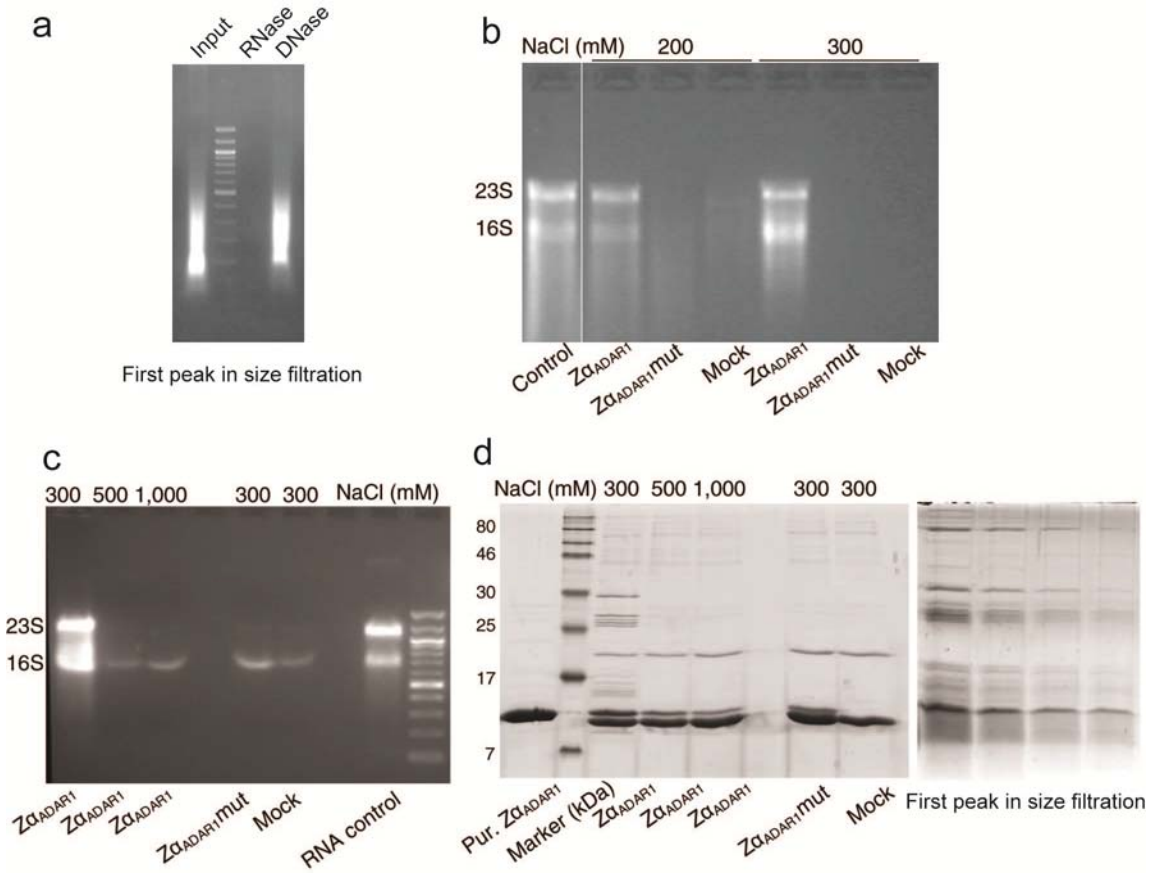


Figure 23 ***In vivo*** **Pull-down experiments of Zα$_{ADAR1}$ in *E. coli*.**

In **a.** Analysis of the first peak after size filtration on agarose gel, sample was digested with either RNase A or DNase I. **b.** Pull down experiment with pET17b-Zα$_{ADAR1}$, pET17b-Zα$_{ADAR1}$mut or pET17b transformed in BL21(DE3) without induction. Cells were lysed in 200 mM NaCl condition, and washed with 200 mM or 300 mM NaCl, as indicated. Purified total RNA from BL21(DE3) was loaded as control. **c.** Pull-down experiment with cells lysed in 300 mM NaCl condition. Zα$_{ADAR1}$_ was washed with 300 mM to 1 M NaCl, while Zα$_{ADAR1}$mut and Mock was washed with 300 mM NaCl. **d.** Left panel, the pull-down sample in **c** was analyzed on 17% SDS-PAGE. The two extra protein bands in the mock sample came from boiled beads. Right panel, SDS-PAGE analysis of the first peak after size filtration.

III.3 Binding of $Z\alpha_{ADAR1}$ to ribosomes in *E. coli*.

III.3.1 *In vivo* pull down of ribosomes by $Z\alpha_{ADAR1}$

In order to test whether $Z\alpha_{ADAR1}$ could bind to RNA in *E. coli*, pull-down experiments were performed in BL21(DE3) with $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut expression. Considering the protein level after induction in this over-expression system is very high, which will compromise *E. coli* growth and reduce the specificity of $Z\alpha_{ADAR1}$ binding, expression of the probes was not inducted.

Different wash conditions were applied in the pull-down experiment. As shown in (Figure 23 **b**), $Z\alpha_{ADAR1}$ can pull down ribosomes, including both large (50S) and small (30S) subunits, in the presence of either 200 and 300 mM NaCl; whereas, there was almost nothing pulled-down by $Z\alpha_{ADAR1}$mut or mock at the same ionic strength. Successful pull-down of ribosomes with 300 mM NaCl wash suggested that the binding of $Z\alpha_{ADAR1}$ to ribosome is quite stable. With increased ionic strength, binding of $Z\alpha_{ADAR1}$ to ribosomal RNA is lost. This is similar to binding of Z-DNA to $Z\alpha_{ADAR1}$, which is also ionic strength-dependent. The presence of $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut in pull-down materials was verified by SDS-PAGE (Figure 23 **d** Left panel). The comparable amount of $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut proteins indicates that the different amount of rRNA pull-down was not caused by different amounts of our probes. Intriguingly, the pattern of proteins that co-precipitated by Za_{ADAR1} in this experiment was very similar to that seen in the first peak after size-exclusion chromatography during Za_{ADAR1} protein purification

(Figure 23 **d**, compare Left and Right panel), suggesting that the first peak contained *E. coli* ribosomes (or ribosomal fragments) that co-purified with $Z\alpha_{ADAR1}$.

.

III.3.2 *In vitro* pull down of ribosomes by $Z\alpha_{ADAR1}$ in *E. coli*

Whole lysates of *E. coli* [BL21(DE3)] was used as the reservoir of RNA molecules for $Z\alpha_{ADAR1}$ binding *in vitro*. The same amount of purified $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut was added to the same volume of *E. coli* lysates. The binding and wash condition was 300 mM NaCl. As shown in Figure 24, $Z\alpha_{ADAR1}$ can pull down only the large 23S rRNA *in vitro*; whereas, there were almost no complexes formed with $Z\alpha_{ADAR1}$mut.

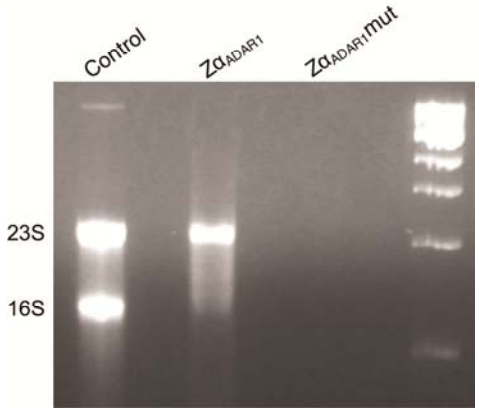


Figure 24 *In vitro* Pull-down experiments of $Z\alpha_{ADAR1}$ in *E. coli*.

Purified total rRNA from *E. coli* was loaded as control. 10 µg $Z\alpha_{ADAR1}$ or $Z\alpha_{ADAR1}$mut were incubated with *E. coli* lysates at 300 mM NaCl. Pull-down experiments were performed at the same salt condition.

III.3.3 $Z\alpha_{ADAR1}$ binds to the large subunit of ribosomes in *E. coli*

Both the *in vitro* and *in vivo* pull down experiments showed that $Z\alpha_{ADAR1}$ could bind to *E. coli* ribosomes. However, whether $Z\alpha_{ADAR1}$ binds to one or both subunits *in vivo* was still unknown. In order to investigate which ribosome subunit $Z\alpha_{ADAR1}$ could bind to, individual subunits from pET17b-Zα-NLS-FS transfected BL21(DE3) were separated by sucrose gradients. The fractions containing large or small subunits were analyzed by western blotting to identify $Z\alpha_{ADAR1}$. As shown in Figure 25, $Z\alpha_{ADAR1}$ co-exists with both large and small subunit. The signal of $Z\alpha_{ADAR1}$ was enriched in fraction 19-23 (Lower panel), which coincides with the signal of the large subunit (Upper panel). The signal of Za_{ADAR1} almost disappeared in fractions 27-31 and appeared again after fraction 31 (Lower panel), which coincides with signal of small subunit (Upper panel, fraction 27-31). However, the signal of $Z\alpha_{ADAR1}$ becomes stronger after fraction 43, most likely resulting from free $Z\alpha_{ADAR1}$ protein at the top of the gradient.

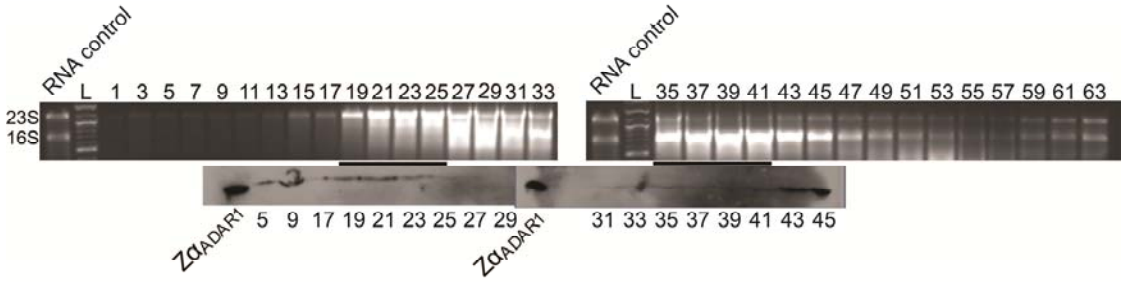


Figure 25 Co-existence of Za_{ADAR1} and ribosomal subunits in sucrose gradient.

Each fraction was loaded onto agarose gels for RNA analysis (upper panel), and Western blotting (lower panel) Purified RNA and probe was used as controls, as indicated

From this result, we can conclude that $Z\alpha_{ADAR1}$ binds to the large subunit of ribosomes in *E. coli*. The probe appears to be also associated with the small subunit, but this conclusion warrants further confirmation.

III.3.4 $Z\alpha_{ADAR1}$ binds to ribosomes through interactions with rRNA

$Z\alpha_{ADAR1}$ binds to *E. coli* ribosome. However, it is still unclear whether the association is due to $Z\alpha_{ADAR1}$ binding to ribosomal proteins or via rRNAs. It is also still unclear whether binding is sequence- or structure-specific.

III.3.4.1 Binding of $Z\alpha_{ADAR1}$ to *E. coli* ribosome is not depended on the sequences of rRNA

If binding of $Z\alpha_{ADAR1}$ to ribosome is depended on the sequence of the rRNA, $Z\alpha_{ADAR1}$ could also pull down purified rRNA. Total RNA from *E. coli* was used as starting material and the pull-down experiment followed the same protocol as developed for *in vitro* pull-down. The result (Figure 26) showed that there was almost nothing pulled down by either $Z\alpha_{ADAR1}$ or $Z\alpha_{ADAR1}$mut. As the Za_{ADAR1} could not bind to purified RNA, the binding of $Z\alpha_{ADAR1}$ to *E. coli* ribosome appears not purely depend on rRNA sequences.

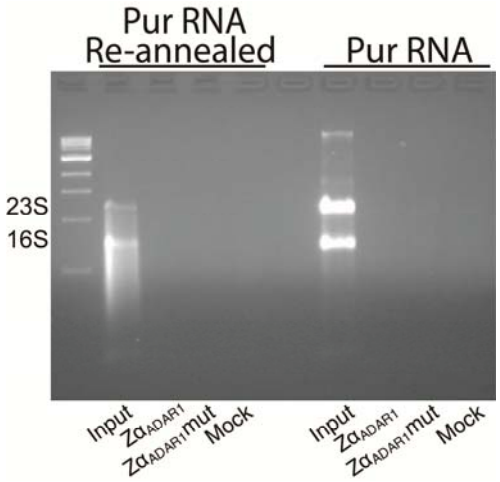


Figure 26 Pull down of purified RNA by $Z\alpha_{ADAR1}$(mut).

Total RNA was purified from *E. coli*, and half of the material was heat re-annealed by heating to 100°C and slowly cooling to room temperature. Purified $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut were incubated with purified RNA and pulled down with 300 mM wash steps. 10% of the input material was loaded as control.

III.3.4.2 Binding of $Z\alpha_{ADAR1}$ to *E. coli* ribosomes does not depend on interactions with ribosomal proteins

The *in vivo* pulled-down material was digested with RNase A on beads and the unbound material was washed away. The remaining material was analyzed by SDS-PAGE and ^{32}P labeling to examine the existence of RNA and protein, respectively. The ^{32}P labeling showed that $Z\alpha_{ADAR1}$ pulled-down nucleic acid fragments from 10-40 nt after RNase A digestion, while $Z\alpha_{ADAR1}$mut and mock control samples showed no substantial signals (Figure 27 **a**).

For both $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut, SDS-PAGE showed no signals for other proteins, if compared with mock control (Figure 27 **b**). The amount of RNA fragments could be

calculated according to the known loaded amount of the nearby marker (16 mer: 100 pmol, 24 mer: 30 pmol). Assuming each fragment was co-precipitated by one protein with MW 10 kDa, potential co-precipitated proteins must have been in the µg amount range, which should have be easily detectable after Coomassie blue staining. $Z\alpha_{ADAR1}$ could still pull down RNA fragments after RNAse A digestion. Thus, the pull-down of RNA by $Z\alpha_{ADAR1}$ was mediated by an interaction with rRNA and not via a second protein.

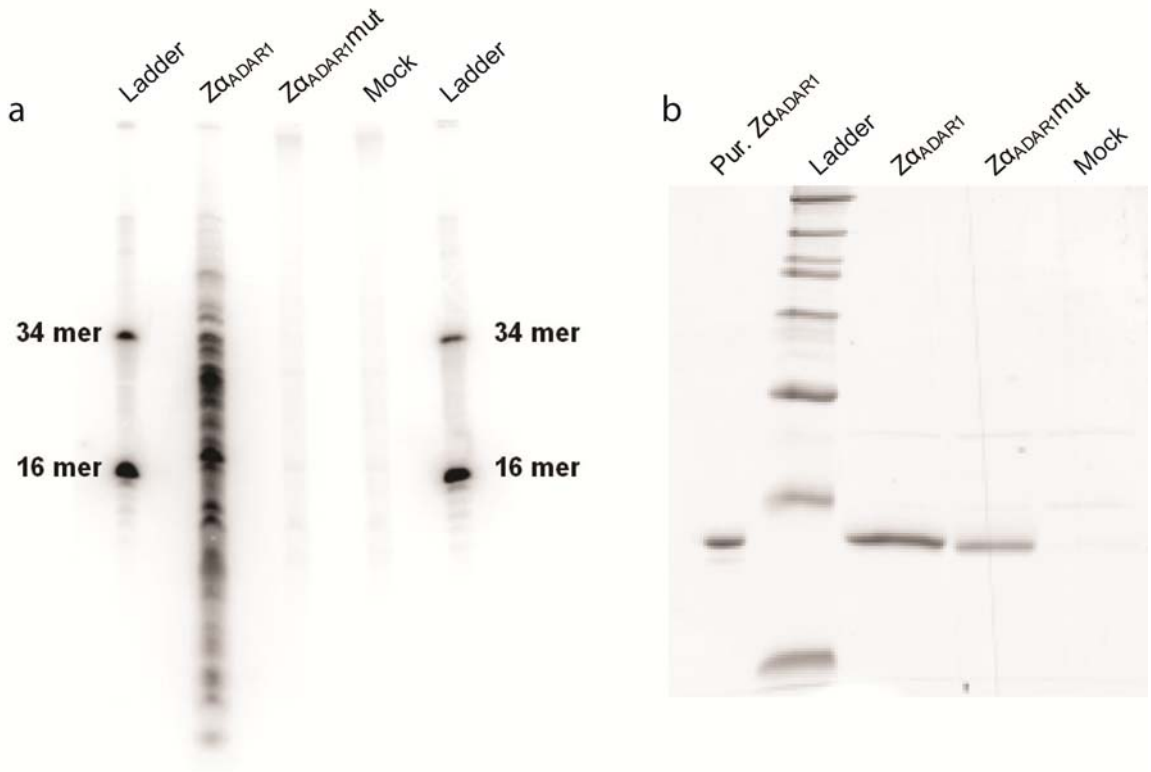


Figure 27 Protein and RNA fragments pulled-down by $Z\alpha_{ADAR1}$ after RNase digestion.

a. Material left in $Z\alpha_{ADAR1}$, $Z\alpha_{ADAR1}$mut or mock pull down after RNase A digestion was labeled with ^{32}P; **b**. corresponding material was analyzed on 17% SDS-PAGE and stained with Coomassie blue. 3 µg purified $Z\alpha_{ADAR1}$ was used as marker.

III.3.5 Binding of other Zα domains to *E. coli* ribosomes

In order to confirm that binding of $Z\alpha_{ADAR1}$ to ribosomes depends on rRNA structure rather than sequence, other known Z-DNA binding domains were expressed in *E. coli* for pull-down assays. The result (Figure 28) showed that all Zα domains, except $Z\alpha_{E3L}$, which is known to exhibit a reduced binding affinity to Z-DNA (Kahmann, Wecking et al. 2004), can precipitate ribosomes *in vivo*. The protein pattern of the pulled down material is strikingly similar to the protein pattern of purified polysomes, which further confirmed that the pulled-down material are ribosomes.

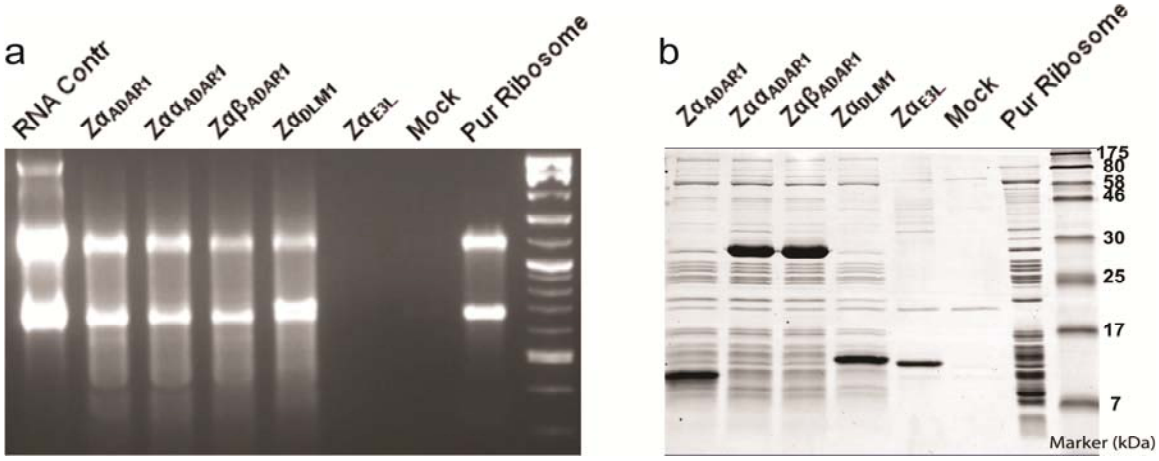


Figure 28 Various Zα domains bind to ribosome in *E. coli*.

Different Zα domains were expressed in BL21(DE3) without induction. Lysis and wash conditions employed 300 mM NaCl. The pulled-down material was analyzed on a 1.2% (w/v) agarose gel (**a**) and aliquots were subjected to SDS-PAGE (17% w/v) for protein analysis (**b**) Purified polysomes from *E. coli* were loaded as control for rRNA and protein patterns of ribosome.

III.3.6 Identification of rRNA binding sites for $Z\alpha_{ADAR1}$ on *E. coli* ribosomes.

Binding of $Z\alpha_{ADAR1}$ to rRNA on ribosomes was confirmed. However, the exact binding sites were still unknown. Previous experiments (Figure 26) showed that after RNase A digestion, RNA fragments were still bound to $Z\alpha_{ADAR1}$ which can tolerate 300 mM NaCl during wash steps. At the same time, $Z\alpha_{ADAR1}$mut or mock could not precipitate RNA fragments. Therefore, RNA fragments binding to $Z\alpha_{ADAR1}$ with high specificity and affinity most likely represent the binding sites on ribosome.

Three independent experiments were performed with different levels of RNase A digestions. The fragments pulled down after RNase A treatment were cloned and sequenced. The sequences were then mapped to *E. coli* rRNA (compare Figure 16 in Section II). For each subunit, there is one position enriched in all three experiments; these are positions 1360-1376 in the small subunit with the sequence 5´-AUGGGGUGACUGCGUAC-3´and the position 548-564 in the large subunit with the sequence 5´-AGAAUGCCACGGUGAAU-3´. (Figure 29, peaks labeled in red). These two sites were dubbed “high confidence Ribosome Binding Sites” (hcRBS). In the sample with less RNase digestion (experiment 2), more and longer fragments were enriched, representing other sites than the hcRBSs; while in samples with more complete digestion, the peaks representing hcRBSs were clearly standing out from the backgroud. This suggested that the other peaks in the less digested samples might represent rRNA fragments that co-purifed with hcRBCs, possibly mediated by interactions with proteins or RNA fragments nearby.

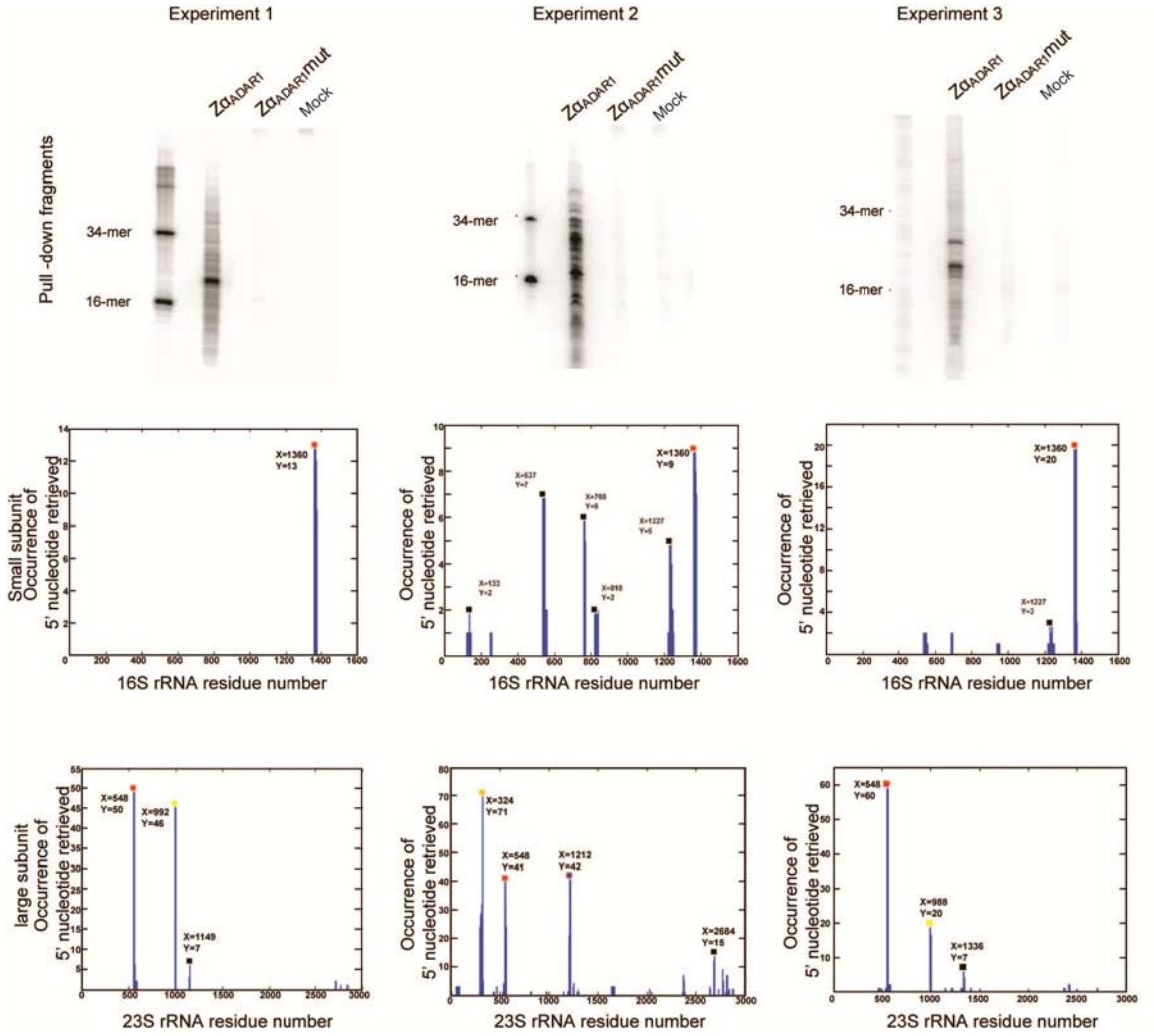


Figure 29 Distribution maps of $Z\alpha_{ADAR1}$ binding sites on *E. coli* rRNA

Three independent pull-down experiments were performed with different levels of RNase A digestions. Aliquots of the pull-down samples were labeled with γP^{32}-ATP, and analyzed on 12% PAGE (Upper panel). The resulting sequences were mapped to the rRNA sequences of *E. coli*, with the X-axis indicating the position on the rRNA, and the Y-axis indicating the copy number of identified sequences at that position returned from the sequencing results (Middle and Lower panel). The hcRBSs are labeled in red color, while the other peaks were labeled in different colors. The same color code was applied to the same mapping position in different experiments.

III.3.7 *In vitro* translation inhibition by $Z\alpha_{ADAR1}$ *E. coli*

III.3.7.1 RNase contamination test

Before the *in vitro* translation experiment was performed, purified probes were tested for RNase contamination. As shown in Figure 30, there is no obvious RNase activity detectable in both purified protein probes.

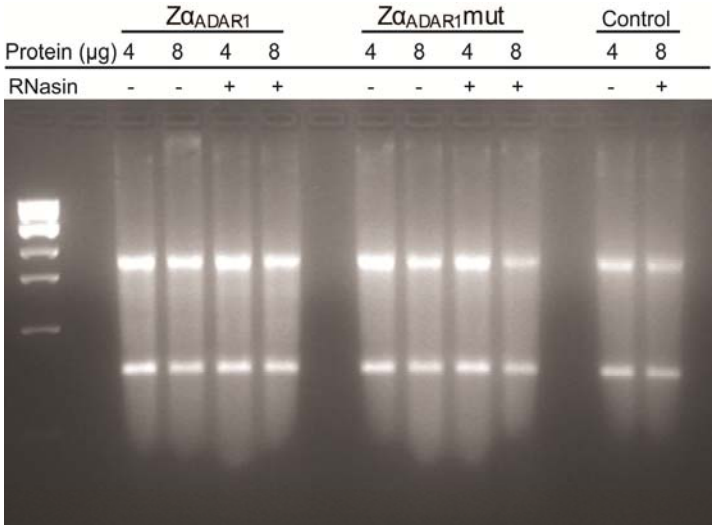


Figure 30 RNase contamination test.

Different amount of $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut were incubated with 500 ng purified total RNA from *E. coli*. RNasin, which is a recombinant RNase inhibitor and inhibits wide range of RNases, were added to half of the samples, as indicated, before incubation.

III.3.7.2 $Z\alpha_{ADAR1}$ inhibits *in vitro* translation

I have demonstrated that $Z\alpha_{ADAR1}$ binds to large subunits *in vitro*, and it was therefore interesting to test the influence of this association on translation. For this, I employed the *E. coli* T7 S30 Extract System, using luciferase as reporter for translational activity. The amount of translated luciferases was quantified by addition of substrate and

expressed in Luminance units. As shown in Figure 31, translation is inhibited by $Z\alpha_{ADAR1}$ in a dose-dependent manner. At about 2 molecules of $Z\alpha_{ADAR1}$ per ribosome, there is only about 20% luciferase activity left. That is, 80% of translational activity was inhibited by $Z\alpha_{ADAR1}$ at this stoichiometry. Addition of $Z\alpha_{ADAR1}$mut, however, gave almost the same reading as mock control.

Thus, $Z\alpha_{ADAR1}$ inhibits *E. coli* translation *in vitro*, and the inhibitory effect is most likely related to its Z-DNA/RNA binding specificity.

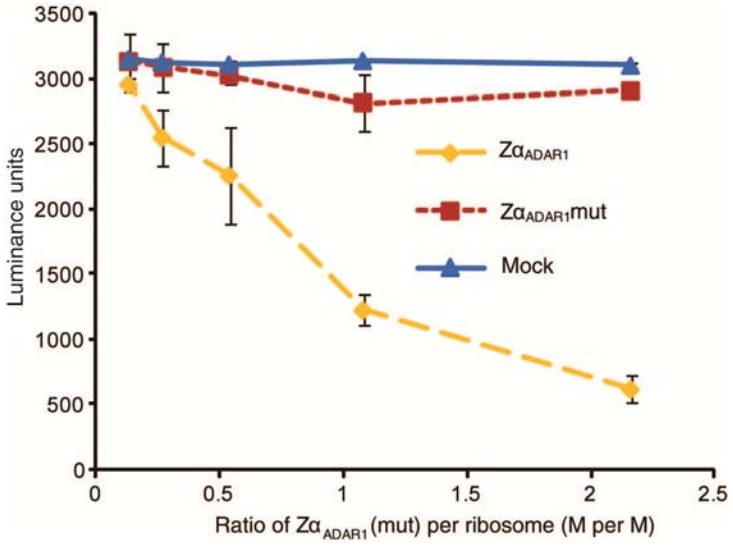


Figure 31 *In vitro* translation assay in *E. coli* extracts

Translated luciferase was quantified as luminance, plotted on the Y axis. The relative $Z\alpha_{ADAR1}$(mut) protein amount per ribosome is plotted on the X axis.

III.3.8 Protein synthesis is affected by $Z\alpha_{ADAR1}$ inside *E. coli*

In order to determine whether translation could be inhibited by $Z\alpha_{ADAR1}$ *in vivo* in *E. coli*, the protein level of $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut in transformed *E. coli* cells were compared

after induction. As a consequence of induction, T7 RNA polymerase is highly expressed and generates massive amounts of $Z\alpha_{ADAR1}$(mut) mRNA. Hence, the translational machinery will be devoted mainly to translate $Z\alpha_{ADAR1}$(mut) proteins. If $Z\alpha_{ADAR1}$ inhibits translation *in vivo*, its own synthesis should therefore be affected if compared with the $Z\alpha_{ADAR1}$mut. The results showed that within one hour of induction, the relative amount of translated $Z\alpha_{ADAR1}$ compared to an internal control protein is less than half of that determined for $Z\alpha_{ADAR1}$mut (Figure 32). Thus, the inhibitory effect on translation by $Z\alpha_{ADAR1}$ could also be observed *in vivo* in *E. coli*.

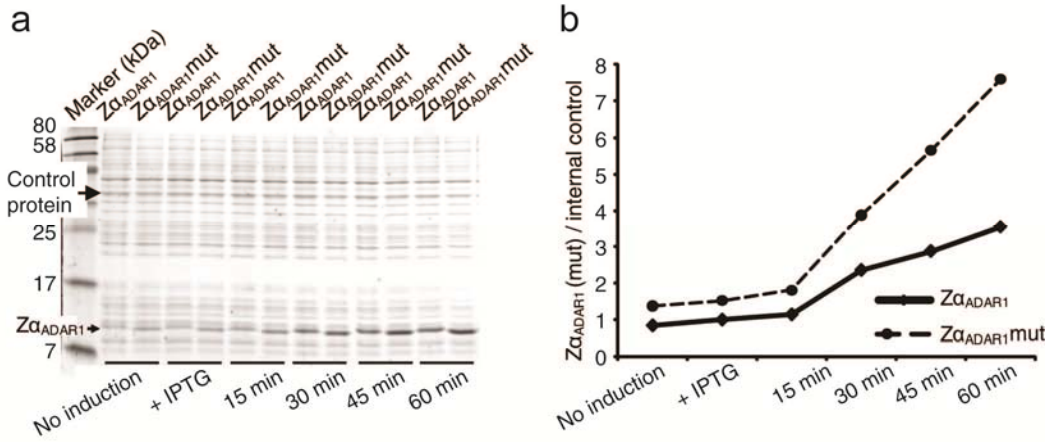


Figure 32 Translational inhibition by $Z\alpha_{ADAR1}$ *in vivo*.

a. Analysis of total protein from *E. coli* after IPTG induction. Both $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut expression is increased after induction; however, $Z\alpha_{ADAR1}$mut has a higher induction level than $Z\alpha_{ADAR1}$. The position of the control protein used for quantification is indicated on the left. **b.** The relative amount of $Z\alpha_{ADAR1}$(mut) to the internal control protein was determined (quantified by Quantity One) and plotted over induction time.

III.3.9 $Z\alpha_{ADAR1}$ expression inhibits *E. coli* growth

The plasmids pET17b-Zα-NLS-FS, pET17b-Zαmut-NLS-FS were transformed into BL21(DE3) and used to express $Z\alpha_{ADAR1}$, $Z\alpha_{ADAR1}$mut in *E. coli*. pET17b transformed BL21(DE3) was used as mock control.

Growth rates were measured at OD_{600}. As shown in Figure 33 **a**, *E. coli* expressing $Z\alpha_{ADAR1}$ showed a substantial delay in exit from stationary phase after cells were transferred to fresh medium. Also, during log phase, the growth rate of $Z\alpha_{ADAR1}$, $Z\alpha_{ADAR1}$mut and control were 0.302 OD/hour, 0.39 OD/hour and 0.452 OD/hour, respectively. That is, the growth rate for *E. coli* expressing $Z\alpha_{ADAR1}$mut is 30% faster than that determined for $Z\alpha_{ADAR1}$.

After induction of probe expression by IPTG at OD 0.55, $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut showed a significant decrease in growth rates.(Figure 33 **b**) Both could not complete one cycle after induction, and $Z\alpha_{ADAR1}$ had a much lower OD in stationary phase than $Z\alpha_{ADAR1}$mut. Hence, both $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut showed an inhibitory effect on *E. coli* growth after induction, as expected, but the impact on growth by $Z\alpha_{ADAR1}$ is much more pronounced than that observed for $Z\alpha_{ADAR1}$mut.

The delay in exit from stationary phase and the growth arrest during log phase was not due to an excess amount of $Z\alpha_{ADAR1}$ over $Z\alpha_{ADAR1}$mut. In fact, $Z\alpha_{ADAR1}$mut had a higher protein level than $Z\alpha_{ADAR1}$ even before induction (compare Figure 32, lane of “No induction”).

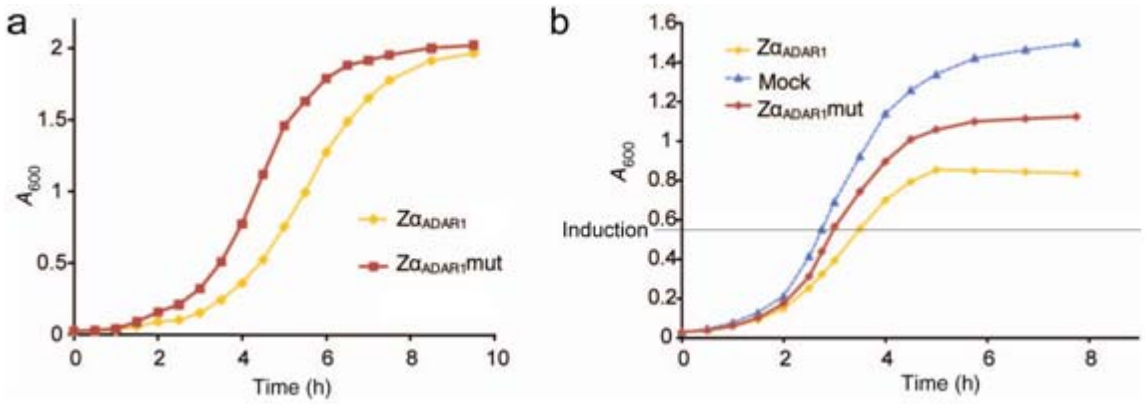


Figure 33 Growth curves of *E. coli* expressing $Zα_{ADAR1}$, $Zα_{ADAR1}$mut or Mock control.

The growth curves of pET17b-Zα-NLS-FS, pET17b-Zαmut-NLS-FS and pET17b transformed BL21(DE3) cells, **a.** Growth curves without induction. **b.** Growth curves with induction. *E. coli* was induced at OD 0.55 for all three cultures using 1mM IPTG.

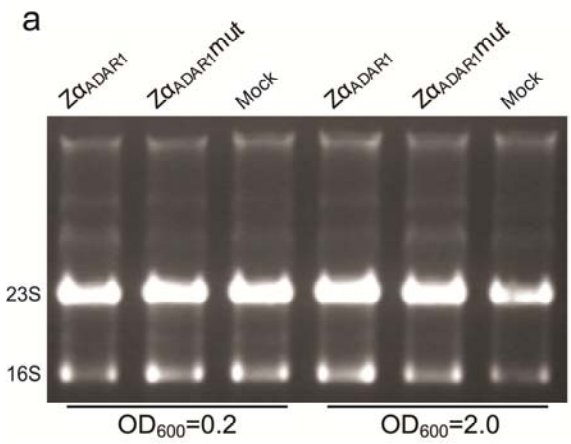
III.3.10 rRNA content in *E. coli* at different phases during growth with $Zα_{ADAR1}$(mut) expression

It has been reported that, when an unneeded protein is produced in *E. coli*, a significant fraction of ribosomes and other cellular systems will be devoted to its synthesis and cell growth will be inhibited. After several cell generations, however, while still in exponential growth, the amount of functional ribosomes increased to its normal steady-state level as a compensatory adjustment, and the cost for the synthesis of unneeded protein is reduced. The growth rates are then less affected (Shachrai, Zaslaver et al. 2010). Based on this new model, as $Zα_{ADAR1}$ could bind and inactivate ribosomes, more ribosomes would be needed to reach the steady-state level in $Zα_{ADAR1}$ expressing cells.

Therefore, a prediction is that cells expressing $Z\alpha_{ADAR1}$ should exhibit higher ribosome content than cells expressing $Z\alpha_{ADAR1}$mut due to the additional inhibitory effect on translation, at least at the end of the log phase. As $Z\alpha_{ADAR1}$mut is an unneeded protein, the ribosome content of $Z\alpha_{ADAR1}$mut expression cells should also be higher than that of the mock control.

At the beginning of log phase, the transformed bacteria do not show much difference in rRNA content. At the end of log phase, however, the rRNA content in $Z\alpha_{ADAR1}$ transformed *E. coli* was about 50% higher than mock and showed a 20% increase compared to $Z\alpha_{ADAR1}$mut transformed *E. coli*. (Figure 34) This agrees well with the recently proposed model (Shachrai, Zaslaver et al. 2010) that more ribosomes needed to be synthesized to overcome the costs that result from both $Z\alpha_{ADAR1}$ synthesis as an unneeded protein and its negative effect on translation.

a



b

	Total RNA extracted (ug)	Relative amount compared to Mock
OD=0.2 4ml		
$Zα_{ADAR1}$	56.4	96.91%
$Zα_{ADAR1}$mut	57	97.94%
Mock	58.2	100.00%
OD=2.0 0.8ml		
$Zα_{ADAR1}$	72.96	146.86%
$Zα_{ADAR1}$mut	63.36	127.54%
Mock	49.68	100.00%

Figure 34 The rRNA amount of $Zα_{ADAR1}$(mut) transformed *E. coli* at beginning and end of log phase.

a. 1/25 of total RNA extracted from $Zα_{ADAR1}$(mut) and mock transformed BL21-De3 at OD_{600}=0.2 and 2.0. **b.** Quantification of total RNA extracted from $Zα_{ADAR1}$(mut) transformed *E. coli* at beginning and end of log phase.

III.4 Does $Z\alpha_{ADAR1}$ also interact with mammalian ribosomes?

Our results revealed that $Z\alpha_{ADAR1}$ binds to *E. coli* ribosomes both *in vivo* and *in vitro*, and inhibits translation. As $Z\alpha_{ADAR1}$ is a human protein, it is reasonable to also analyze its effect on mammalian ribosomes. The RNA contents in the *in vivo* pull-down material in ChAP from A549 cells and the cellular localization of $Z\alpha_{ADAR1}$ in the nucleolus provided already first clues that $Z\alpha_{ADAR1}$ could also bind to rRNA in mammalian cells.

III.4.1 $Z\alpha_{ADAR1}$ interacts with ribosomes in HeLa cell lysates

Firstly, whole lysates from HeLa cells were used as a reservoir of ribosomes for $Z\alpha_{ADAR1}$ binding, and different washing conditions were tested. As shown in Figure 35 **a**, $Z\alpha_{ADAR1}$ pulled-down rRNA at both 150 and 200 mM NaCl; whereas, almost no rRNA was pulled down by $Z\alpha_{ADAR1}$mut or mock at 150 mM NaCl. Successful pull-down of ribosomes at 200 mM NaCl suggested that binding of $Z\alpha_{ADAR1}$ to human ribosomes is also quite stable. The presence of $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut in pulled-down samples was checked by Western blotting, as shown in Figure 35 **b**. The amount of $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut after pull-down was very similar. Thus, these results suggested that $Z\alpha_{ADAR1}$ binds specifically also to HeLa ribosomes *in vitro*.

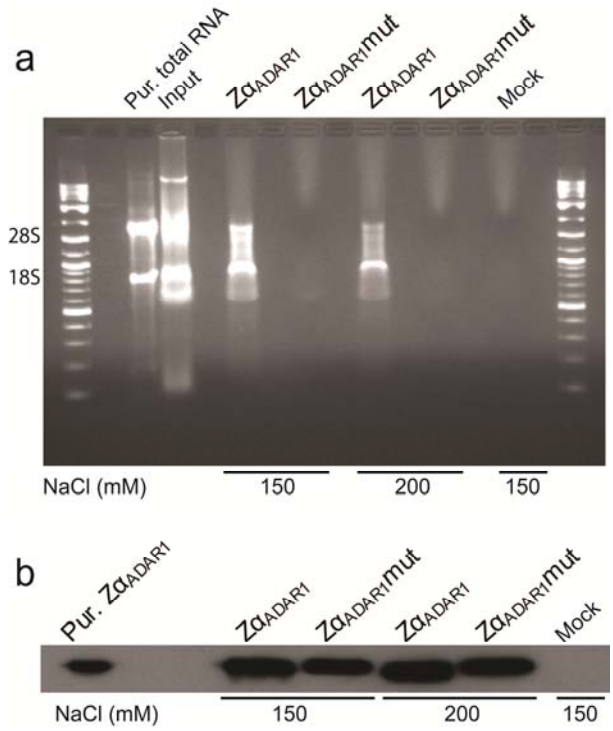


Figure 35 Pull-down experiments of $Z\alpha_{ADAR1}$ from HeLa cell lysate.

Purified $Z\alpha_{ADAR1}$(mut) was incubated with HeLa lysates and the pull-down assay was performed with different wash conditions. **a.** Half of the pulled-down material was analyzed on agarose gels. Purified total HeLa RNA and the input lysate were loaded as control. The wash conditions for $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut were 150 mM or 200 mM NaCl, while only 150 mM NaCl was applied for Mock **b**. The amount of $Z\alpha_{ADAR1}$ or $Z\alpha_{ADAR1}$mut in the other half of the pull-down material was determined by Western blotting.

III.4.2 $Z\alpha_{ADAR1}$ binds to both ribosomal subunits *in vitro*

In order to determine to which ribosomal subunit $Z\alpha_{ADAR1}$ could bind, individual subunits were first purified by ultracentrifugation in sucrose gradients, and *in vitro* pull-down by $Z\alpha_{ADAR1}$ were subsequently performed using purified subunits as substrate. The results showed that $Z\alpha_{ADAR1}$ pulled-down each subunit *in vitro* (Figure 36).

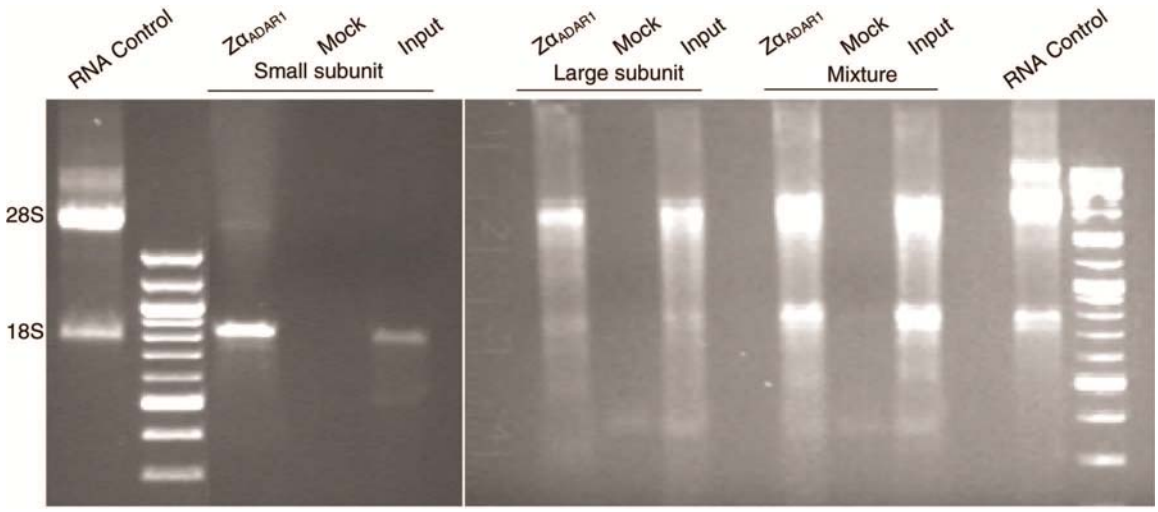


Figure 36 pull-down of separated ribosomal subunits by $Z\alpha_{ADAR1}$.

Individual subunits were isolated from HeLa extracts and incubated with purified $Z\alpha_{ADAR1}$, $Z\alpha_{ADAR1}$mut or BSA as control. *In vitro* pull down assays were performed with 150 mM KCl. After pull–down, the eluted material was analyzed on agarose gels. The extracted total RNA from HeLa cells were loaded as RNA control.

III.4.3 Association of $Z\alpha_{ADAR1}$ and full-length ADAR1 to polysomes *in vivo.*

It is clear that $Z\alpha_{ADAR1}$ binds to ribosomes *in vitro* in HeLa cells. However, it is still unclear whether $Z\alpha_{ADAR1}$ also binds to ribosomes *in vivo*. In order to examine whether $Z\alpha_{ADAR1}$ is associated with ribosomes *in vivo*, polysomes from the $Z\alpha_{ADAR1}$, $Z\alpha_{ADAR1}$mut or Mock transfected HeLa cells were prepared, and the presence of $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut protein in polysomes was analyzed by Western blotting. The result (Figure 37) showed that $Z\alpha_{ADAR1}$ co-purified with polysomes, while the signal for $Z\alpha_{ADAR1}$mut in the polysome fraction was very weak. The $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut level in the supernatant was similar. Therefore, the different protein levels in the polysome fraction were not due to different *in vivo* expression levels for the two probes.

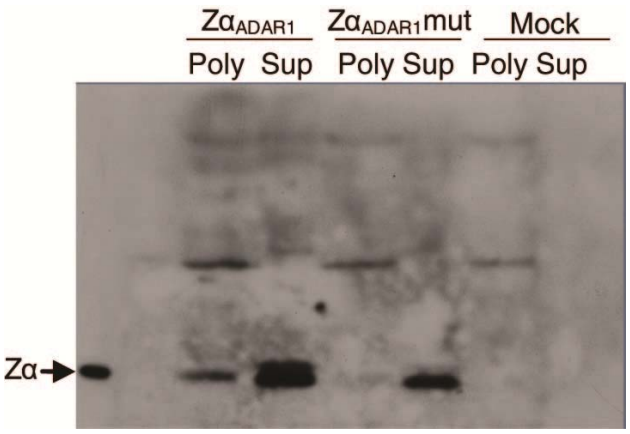


Figure 37 Association of $Z\alpha_{ADAR1}$ domain with polysome *in vivo*

pCMV-Zα-NLS-FS and pCMV-Zαmut-NLS-FS transfected cells were crosslinked by UV, and polysomes were harvested by ultracentrifugation through a 30% sucrose cushion. Both the supernatant (labeled as Sup) and the polysome fraction (labeled as Poly) were analyzed by Western blotting.

In order to examine whether the full-length ADAR1 protein is associated with polysomes, ADAR1, ADAR1-mut or Mock transfected HeLa cells were prepared and analyzed accordingly using Western blotting.

The results (Figure 38, left panel) showed that ADAR1 co-purified with polysomes, while the signal for ADAR1-mut is substantially weaker. The expression level of ADAR1 and ADAR1-mut in transfected HeLa cells was very similar (Figure 38, right panel).

Thus, full-length ADAR1 could associate with the polysomes, while the mutant which lacks the ability to stably bind to Z-DNA/RNA has much reduced affinity.

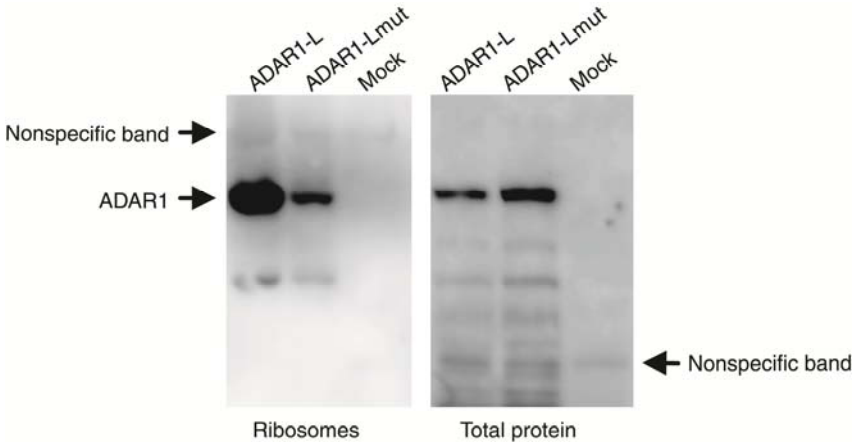


Figure 38 Association of ADAR1 with polysome *in vivo*

Cells were transfected with pEXFH-huADAR1, pEXFH-huADAR1-mut, and polysomes harvested as described above, but in 200 mM NaCl. Aliquots of the lysates were kept for analyzing protein expression levels (right panel) before centrifugation, while the polysome fractions were analyzed by Western blotting after centrifugation to separate ribosome-associated proteins from unbound (left panel). The nonspecific bands can serve as internal loading controls for each sample.

III.4.4 Identification of binding sites for $Z\alpha_{ADAR1}$ on HeLa ribosomes

The same protocol for identifying rRNA binding sites for $Z\alpha_{ADAR1}$ on *E. coli* ribosomes was applied for HeLa ribosomes. Besides identifying binding site *in vivo*, *in vitro* binding sites were also identified by using material from the *in vitro* pull-down experiments.

Two independent experiments were performed for both *in vivo* and *in vitro* pull down. The fragments obtained after RNase A digestions were cloned and sequenced, and the resulting sequences were mapped to human rRNA. (Figure 39 and Figure 40)

For the *in vitro* experiment (Figure 39), there were two peaks in the small subunit in both pull-down experiments. The first and peak was much higher than the second one, and were defined as the hcRBS. It is at position 429–450 with the sequence 5´-CCGGAGAGGGAGCCUGAGAAAC-3´. There were three peaks identified in the large subunit. The first one was at the position 2269–2289 with the sequence 5´-AGGAGGGCCGCUGCGGUGAGC-3´. The second one was at position 3630–3672 with the sequence 5´-AAAGCAUCGCGAAGGCCCGCGGCGGGUGUUGACGCGAUGUGAU-3´. The third peak was at the position 4213–4222 with the sequence 5´-CAGGGAGGAC-3´. They were all defined as hcRBS.

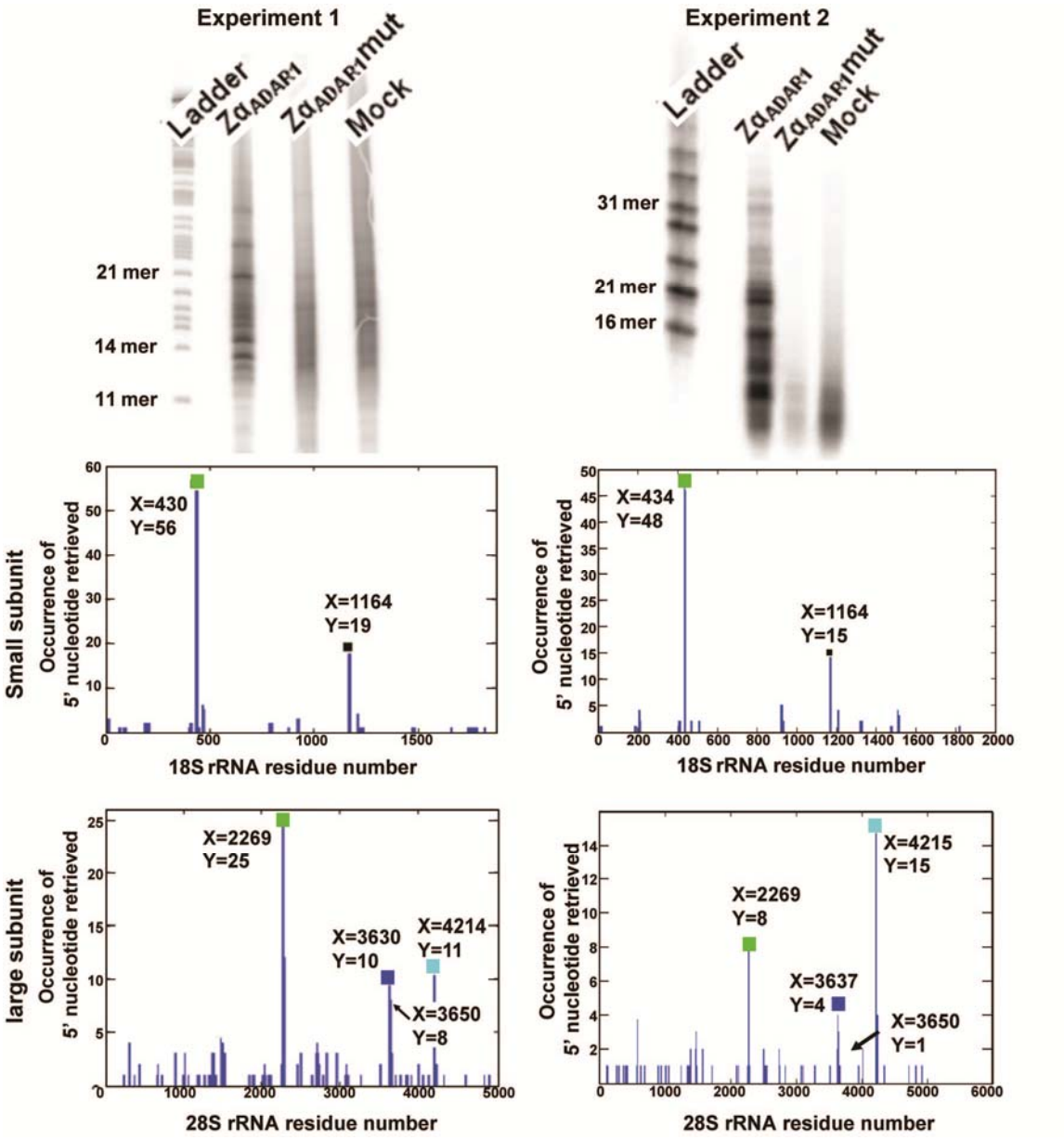


Figure 39 Identification of *in vitro* hcRBS on human ribosome

Two independent pull-down experiments were performed. Aliquots of the pull-down samples were labeled with γP^{32}-ATP, and analyzed on 12% PAGE (Upper panel). The sequences were mapped to human rRNA, with the X-axis indicating the position on the rRNA. The Y-axis indicates the copy number of the sequences at a particular position (Middle and Lower panel). Peaks were labeled with different colors; the same color code was used through the whole thesis.

The *in vivo* experiment (Figure 40) revealed two peaks in the small subunit in both pull-down experiments, they were defined as hcRBS. The first one was at position 1556–1572 with the sequence 5´-ACCCUACGCCGGCAGGC-3´. The second one was at the position 1670-1678 with the sequence 5´-UGCGUUGAU-3´. In the second experiment, there was one peak at the position 1652. The corresponding fragment is predicted to be located very close to that represented by the peak near 1670. There was one peak in the large subunit in both experiments, and defined as hcRBS. It is at position 3635–3672 and has the rRNA sequence 5´-AUCGCGAAGGCCCGCGGCGGGUGUUGACGCGAUGU-3´.

Taken together, the peak on the large subunit near position 3650 could be found in both *in vitro* and *in vivo* experiment. However, the peaks were different in the *in vivo* and *in vitro* experiment for the small subunit.

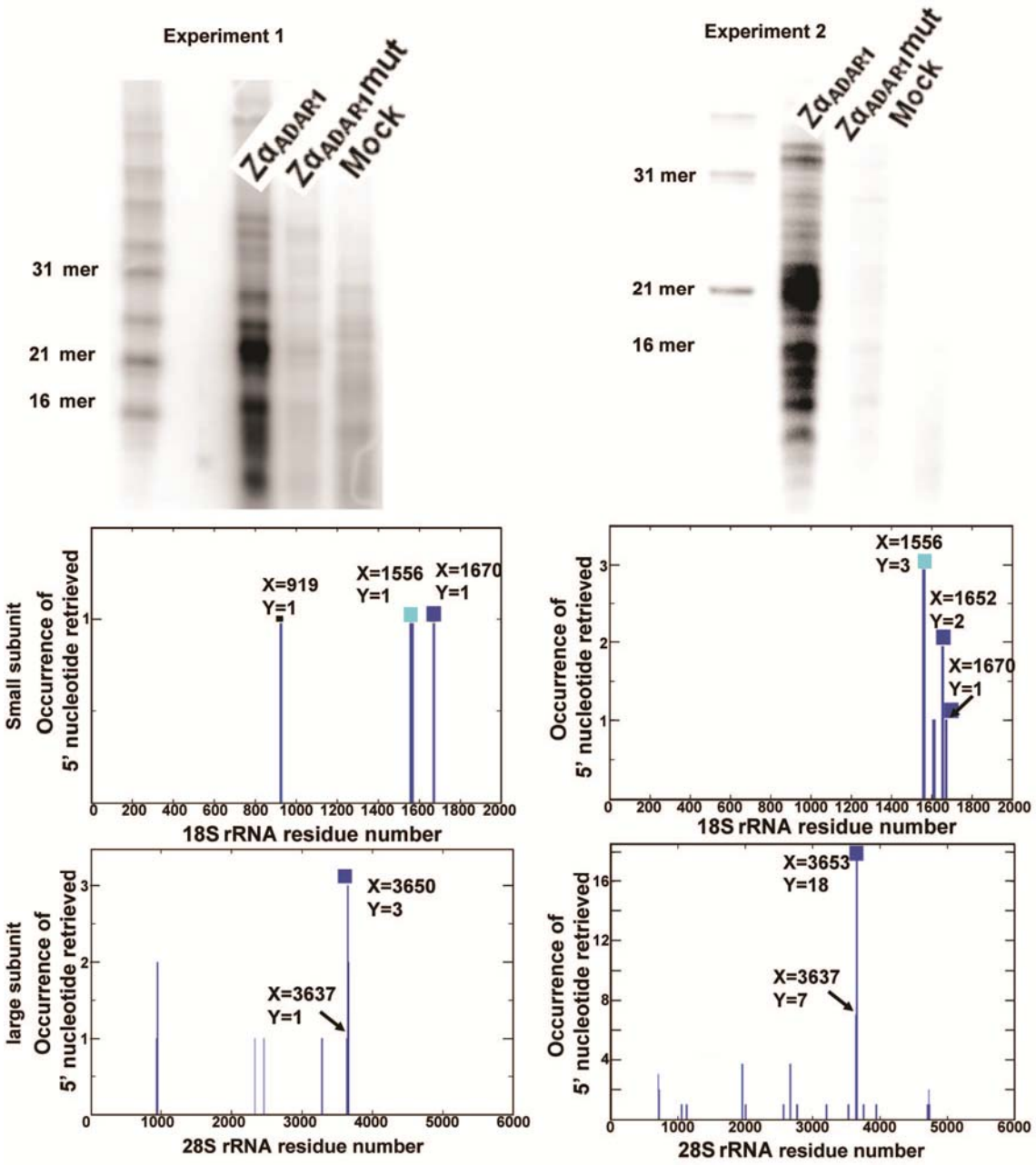


Figure 40 Identification of *in vivo* hcRBS on human ribosome

Two independent pull-down experiments were performed. Aliquots of the pulled-down samples were labeled with γP^{32}-ATP, and analyzed on 12% PAGE (Upper panel). The sequences were mapped to human rRNA sequences, with X-axis indicating the position on the rRNA. The Y-axis indicates the copy number of individual sequences (Middle and Lower panel). Peaks were labeled with different colors. The same color code was used through the thesis.

III.4.5 Translation inhibition by $Z\alpha_{ADAR1}$ in mammalian systems

III.4.5.1 $Z\alpha_{ADAR1}$ inhibits *in vitro* translation

I have shown above that $Z\alpha_{ADAR1}$ binds to the large and small human subunits *in vitro*. Next, I investigated the influence of binding on eukaryotic translation using the Flexi rabbit reticulocyte lysate translation system (RRL) and luciferase activity as reporter. As shown in Figure 41, a dose-dependent manner of translational inhibition by $Z\alpha_{ADAR1}$ is obvious and at about two $Z\alpha_{ADAR1}$ molecules per ribosome, there is only about 10% activity of luciferase left, i.e. 90% of translational activity is inhibited by $Z\alpha_{ADAR1}$. At the same stoichiometry, only 30% of translation was impaired by $Z\alpha_{ADAR1}$mut.

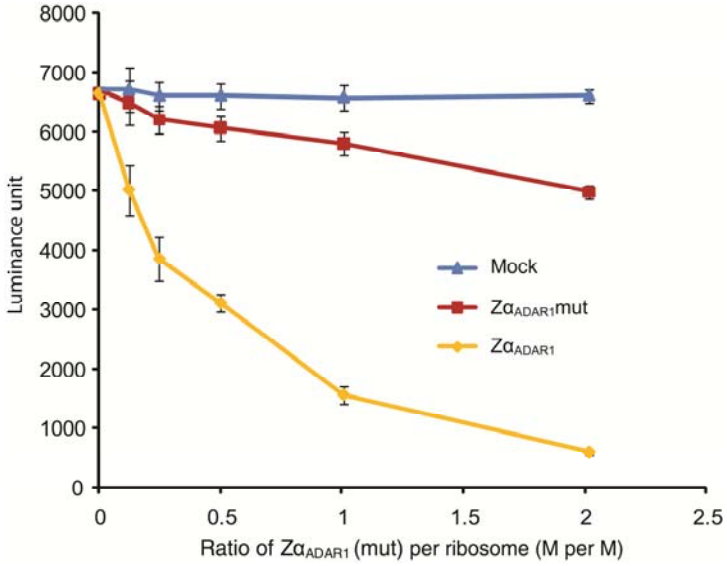


Figure 41 ***In vitro*** **translation assay.**

The amount of translated luciferase was quantified by addition of substrates and expressed as luminance unit, (Y axis), against varied $Z\alpha_{ADAR1}$(mut) protein concentration (X-axis).

III.4.5.2 Inhibition of *in vitro* translation by $Z\alpha_{ADAR1}$ is not affected by the presence of B-DNA

To elucidate whether inhibition of *in vitro* translation by $Z\alpha_{ADAR1}$ is dependent on specific binding to Z-conformation(s), a B-DNA competition assay was performed. As shown in Figure 42, with increasing amount of B-DNA (fragmented human genomic DNA, see also Figure 43 **a**) added, translation in mock and $Z\alpha_{ADAR1}$ samples was not affected; however, the weak inhibitory effect seen for $Z\alpha_{ADAR1}$mut in Figure 41 was eliminated. These results suggested that the strong translational inhibition induced by $Z\alpha_{ADAR1}$ is indeed caused by specific Z-RNA binding to ribosomes, while the weak inhibitory effect by the $Z\alpha_{ADAR1}$mut is most likely due to its unspecific nucleic acids binding activities.

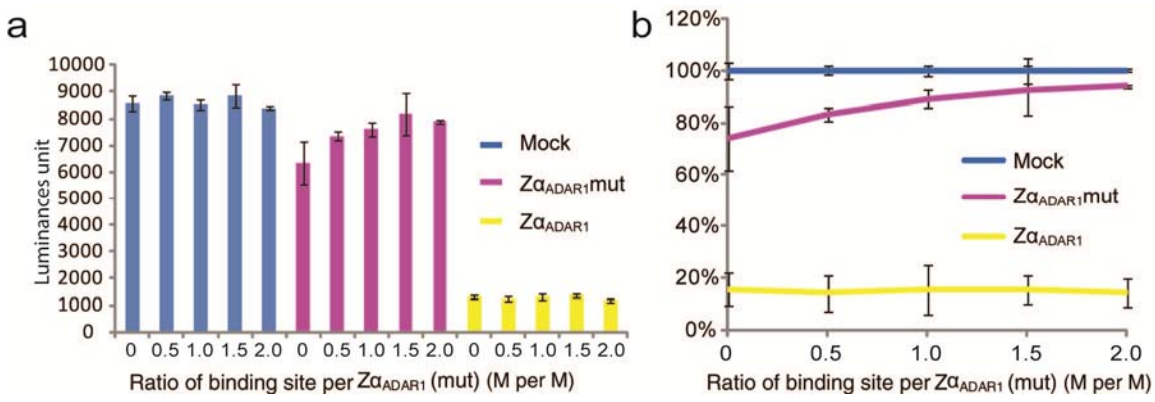


Figure 42 Translation assays with B-DNA competitor.

a. Actual luminance reading after translation assays. **b.** Luminance readings were normalized to mock as 100% at each B-DNA concentration. 10bp of DNA is considered as one binding site for a pair of $Z\alpha_{ADAR1}$(mut) proteins.

III.4.5.3 Binding of $Z\alpha_{ADAR1}$ to ribosomes in untreated rabbit reticulocyte lysates

In order to find out whether DNA competitors interfere with binding of $Z\alpha_{ADAR1}$ or $Z\alpha_{ADAR1}$mut to ribosomes, untreated RRL which did not contain substantial amounts of DNA was used as ribosome source, and B-DNA (fragmented human genomic DNA, see also Figure 43 **a**) were added as competitor. As shown in Figure 43 **b**, with increasing competitor, the amount of ribosomes pulled down by the $Z\alpha_{ADAR1}$mut decreased substantially. Meanwhile, pull-down of $Z\alpha_{ADAR1}$ was not affected, further indicating that binding of $Z\alpha_{ADAR1}$ to ribosome is dependent on Z-conformation(s). Binding of $Z\alpha_{ADAR1}$mut to ribosomes appears, therefore, to be non-specific, and can be easily competed by B-DNA. The salt concentration used in this pull-down experiment is 100 mM KCl, which is same as the salt condition in the translation inhibition assay but lower than the previous pull-down experiments.

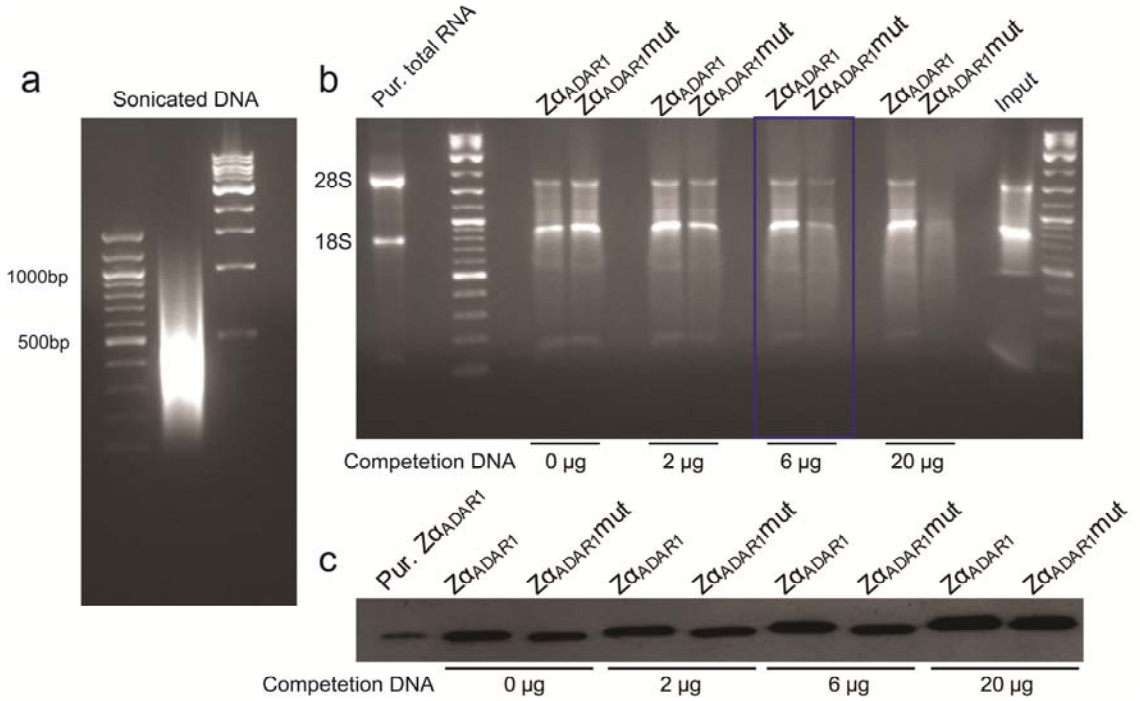


Figure 43 Pull-down of $Z\alpha_{ADAR1}$(mut) from untreated RRL.

In each pull down experiment, 10 μg of probe was added, and the amount of B-DNA competitor was 0 μg, 2 μg, 6 μg and 20 μg. Assuming all DNA molecules are occupied by $Z\alpha_{ADAR1}$ or $Z\alpha_{ADAR1}$mut, each 10 bp DNA would be bound by one protein when about 6 μg B-DNA was used. **a.** shows the size range of B-DNA fragments from sonicated HeLa genomic DNA, which was used as competitor in this pull-down experiment. **b.** shows the results of pull-down assay by $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut, with different amounts of B-DNA competitor. The 1:1 molar ration of $Z\alpha_{ADAR1}$(mut) to B-DNA is reached at 6μg DNA (shown within blue bracket) **c.** shows the amounts of proteins in corresponding pull down sample, determined by Western blotting.

III.4.5.4 $Z\alpha_{ADAR1}$ inhibits *in vivo* translation in HeLa cells

In order to investigate translation inhibition *in vivo* in mammalian cells, pCMV-$Z\alpha_{ADAR1}$-FS, pCMV-$Z\alpha_{ADAR1}$mut-FS, pEXFH-huADAR1, pEXFH-huADAR1-mut or pCMV-mock were co-transfected with a reporter plasmid (pPGK-luciferase), and the luciferase activity was tested 24 hrs after the transfections.

The result (Figure 44 **a**) showed that both $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut had an inhibitory effect on the expression of the reporter gene. However, the effect due to $Z\alpha_{ADAR1}$ was significantly stronger. Western blot analysis confirmed that probes were expressed at a similar level (Figure 44 **b**, left panel).

For the full length protein, ADAR1–L also significantly inhibited expression of the reporter, while ADAR1–Lmut had no effect (Figure 44 **a**). Western analysis showed that both proteins were expressed at very similar levels (Figure 44 **b**, right panel).

The results therefore showed that $Z\alpha_{ADAR1}$ has an inhibitory effect in HeLa cells, functioning alone or within the full length protein. This inhibitory effect is reduced or not detectable with the mutated $Z\alpha_{ADAR1}$, which has no Z conformation-binding specificity. Therefore, the inhibitory effect on translation is most likely directly related to Z-conformation binding.

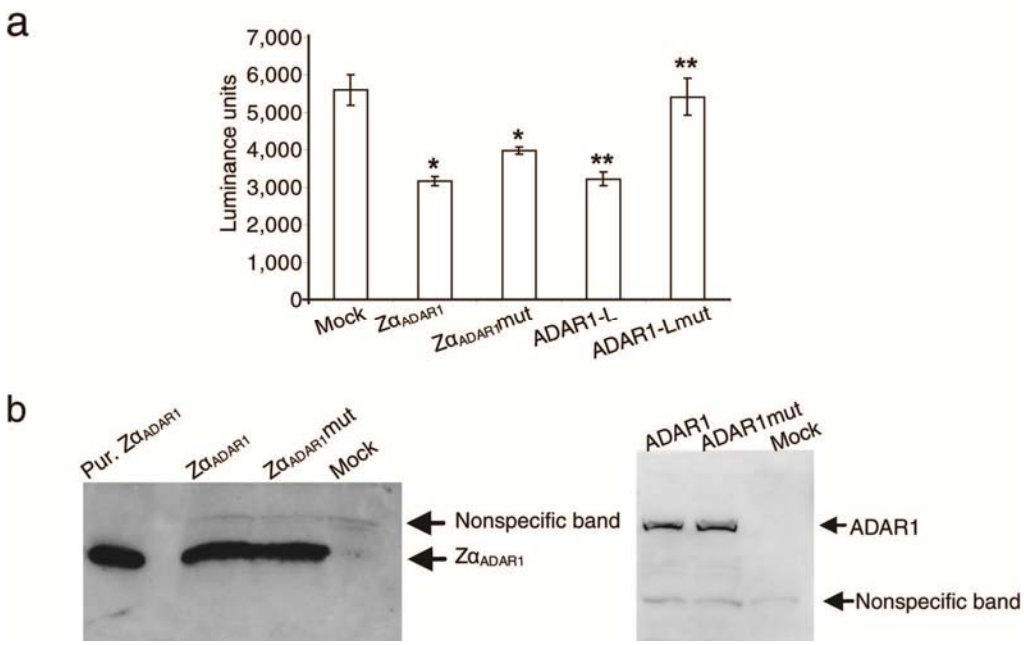


Figure 44 ***In vivo*** **translation inhibition in HeLa cells.**

HeLa cells were co–transfected with luciferase reporter and expression vectors or mock vector, as indicated, and luciferase activities were determined as Luminance units. Assays were performed as triplicates. **a**. Mean values from one experiment with error bars indicating standard deviation. Data were analyzed with Student's t–test revealing significant differences between $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut (*, $P < 0.05$), and between ADAR1–L and ADAR1–Lmut (**, $P < 0.001$). **b.** Western blot analyses confirmed similar levels of $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut expression (Left panel) or ADAR1–L and ADAR1–Lmut expression (Right panel) in cell extracts. The cross–reactive protein, labeled "nonspecific", can serve as internal protein loading control.

III.5 Attempts to establish stable cell lines for $Z\alpha_{ADAR1}$ expression

III.5.1 Failure to generate stable cell lines with constitutive $Z\alpha_{ADAR1}$ expression

Two types of constitutive expression vectors were constructed, and three cell lines were used as source. However, none of them could result in a successful constitutive expression cell line.

III.5.1.1 Construction of pPGK-$Z\alpha_{ADAR1}$(mut)-IRES-puro

For this vector (Figure 13), the selection marker was linked to the expression cassette with the Internal Ribosome Entry Site (IRES) and shared the same promoter as the $Z\alpha_{ADAR1}$(mut) genes. Thus, the $Z\alpha_{ADAR1}$(mut) protein share the same mRNA with the puromycin selection marker. In theory, as long as the cell is resistant to puromycin, the $Z\alpha_{ADAR1}$(mut) protein should be expressed.

This construct was applied to both HeLa and F3 (human primary fibroblast) cell lines. The selection pressure of 1μg/ml puromycin was applied to cells three days after transfection. However, after two weeks' selection, none of the transfected cells survived.

III.5.2 Construction of pEF1-Zα$_{ADAR1}$ (mut)-SV40-Neo

For this vector, the selection marker and Zα$_{ADAR1}$ (mut) expression cassette were under two different promoters (Figure 14). This construct was applied to both HeLa and A549 cell lines. Selection of 500 µg/ml of G418 was applied to cells three days after vector transfection. After two weeks, some colonies formed. Single colonies were picked, expanded and analyzed by Western blotting. More than 20 colonies were picked for each original cell line. However, none of them exhibited detectable Zα$_{ADAR1}$(mut) expression (data not shown). The bulk culture before colonies were picked was also analyzed by Western blotting; yet no detectable Zα$_{ADAR1}$ (mut) expression could be observed (data not shown).

III.5.2.1 Generation of stable cell lines with an inducible expression vector

The Doxycycline-inducible construct pTRIPZ-Zα$_{ADAR1}$(mut)-puro (Figure 16) was used to generate stable cell lines. Expression of Zα$_{ADAR1}$(mut) could be turned on by Doxycycline addition to cell media. This construct was introduced into HT1080 (human fibrosarcoma) cells. The selection of 1µg/ml of puromycin was applied three days after transduction. After two weeks', both Zα$_{ADAR1}$ and Zα$_{ADAR1}$mut vectors resulted in colony formation. However, the number of colonies obtained with pTRIPZ-Zα$_{ADAR1}$ was much lower than with pTRIPZ-Zα$_{ADAR1}$mut. The bulk culture after selection was analyed by Western blotting three days after Doxycycline treatment. Both Zα$_{ADAR1}$ and Zα$_{ADAR1}$mut showed protein expression after induction.(Figure 45 **a**)

Different induction times were also tested. With increasing induction time, expression levels of both $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut were reduced. However, the degree of reduction of $Z\alpha_{ADAR1}$ was much more pronounced than that seen with $Z\alpha_{ADAR1}$mut (Figure 45 **b**, left and right panels). Interestingly, this is very similar to the *in vivo* translation assay results shown in Figure 32, and could be taken as an indication that $Z\alpha_{ADAR1}$ inhibits its own expression in human cells as well.

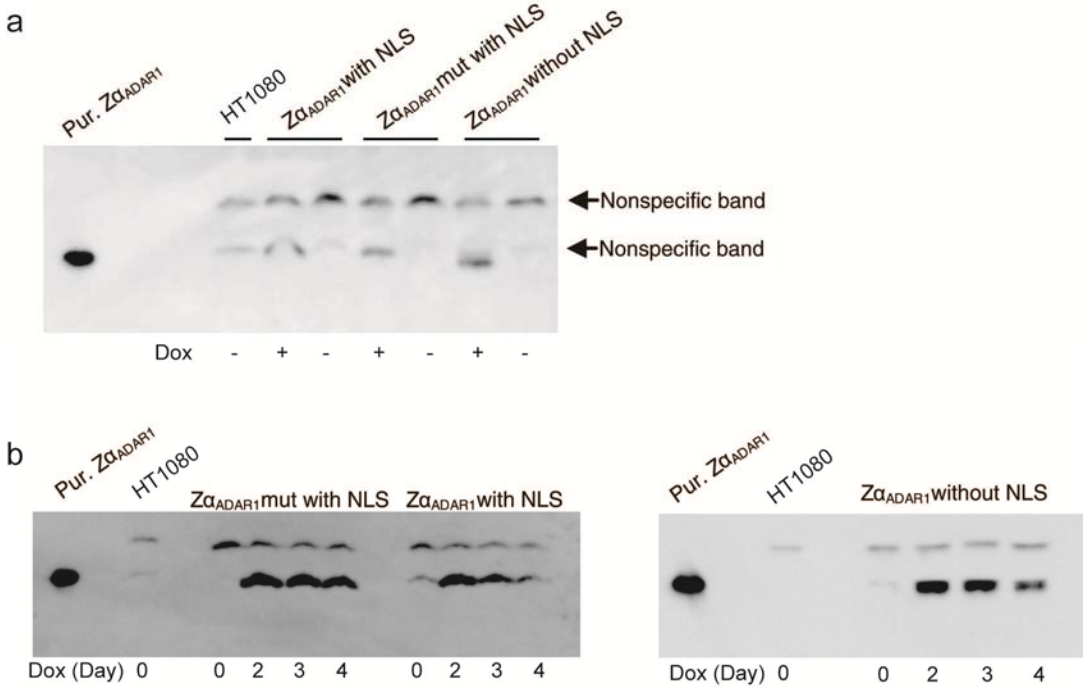


Figure 45 Time course of probe expression in stable cell lines

a. Western blot analyses of pTRIPZ-$Z\alpha_{ADAR1}$(mut)-puro stable cell lines, three days after Doxycycline treatment. The parental HT1080 was included as negative control. Note that there are two cross-reacting nonsepecific protein bands, which can serve as loading controls. However, one of them is vey close to the position of $Z\alpha_{ADAR1}$. Thus, the band seen in the sample without induction is probably the cross-reacting protein and not the probe. **b.** Western blot analyses of pTRIPZ-$Z\alpha_{ADAR1}$(mut)-puro HT1080 stable cell lines (bulk), after different times (days) of Doxycycline treatment.

III.6 Mapping Z-DNA distributions in Human ES cells

III.6.1 ChAP material from Human ES cells

After the ChAP material was ligated to the adaptor as described under materials and Methods (section 0), one percent of each sample was amplified with 30 PCR cycles and analyzed on a 1% agarose gel (Figure 46). Substantially more PCR products from $Z\alpha_{ADAR1}$ ChAP can be detected than for $Z\alpha_{ADAR1}$mut or mock, which indicated that the initial amount of DNA pulled-down templates by $Z\alpha_{ADAR1}$ was higher than for the controls. The PCR product from $Z\alpha_{ADAR1}$mut ChAP is slightly more than mock.

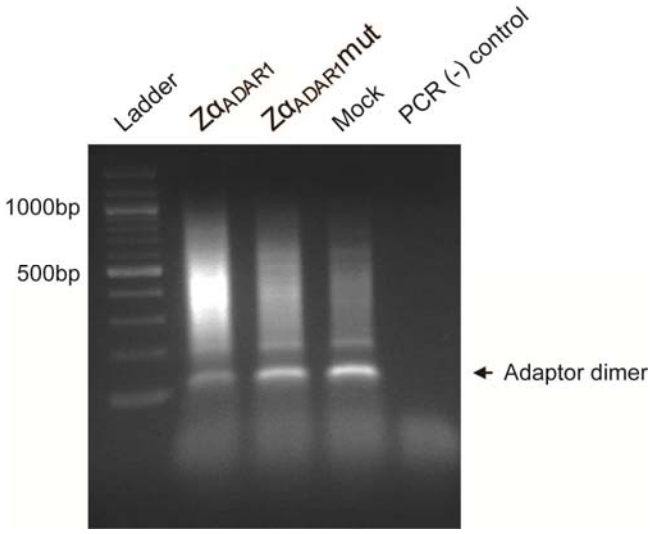


Figure 46 ChAP samples after PCR amplification

After ChAP products were ligated to the adaptor, one percent of each sample was amplified with 30 PCR cycles and checked on agarose gel. One sample without any added template DNA was included as PCR negative control. As the adaptor is 5'-phosphoralated, it has a chance to self ligate. The position of ligated adaptor dimers is indicated by an arrow.

The PCR products were extracted from the gel and cloned into pTZ18R. Ten colonies of $Z\alpha_{ADAR1}$ or $Z\alpha_{ADAR1}$mut were sequenced and blasted against the human genome sequence. Nine of them gave a match, while one of each showed a match to the mouse genome. As the feeder cells used for culturing huES cells were mouse embryonic fibroblasts, it is not surprising that DNA from feeder cells could be co-pulled-down.

III.6.2 General distribution map of $Z\alpha_{ADAR1}$-associated sequences on human chromosomes

Before any selection criteria were assigned, the general distribution of the $Z\alpha_{ADAR1}$ pull-down result was mapped to human chromosomes. When the "Library of original sequences" was used for mapping, the sequences were more or less evenly distributed over the chromosomes, for most of the regions there are not more than 5 hints per 10kb (Figure 47). However, for some of the chromosomes, there was an obvious enrichment at regions adjacent to centromeres (Figure 47). By using the "Library of remaining sequences", where the overlapping sequences between $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut were removed, most (nine out of twelve) of the peaks were still retained (Figure 48; see also Table 2, labeled in red). This observation actually coincided with our previous findings using A549 cells (Li, Xiao et al. 2009), i.e. a rather high fraction (25%) of Z-DNA "hotspots" was located near centromeres. Comparing the A549 results with ES data, there is a high similarity for positions near centromere regions (Table 2) which shows enrichment for ChAP sequences. Nine chromosomes have enrichment at or near the

centromere region (using the “Library with Remaining Sequences”), and six out of the nine were identical in hues and A549 cells (Table 2, highlighted in yellow).

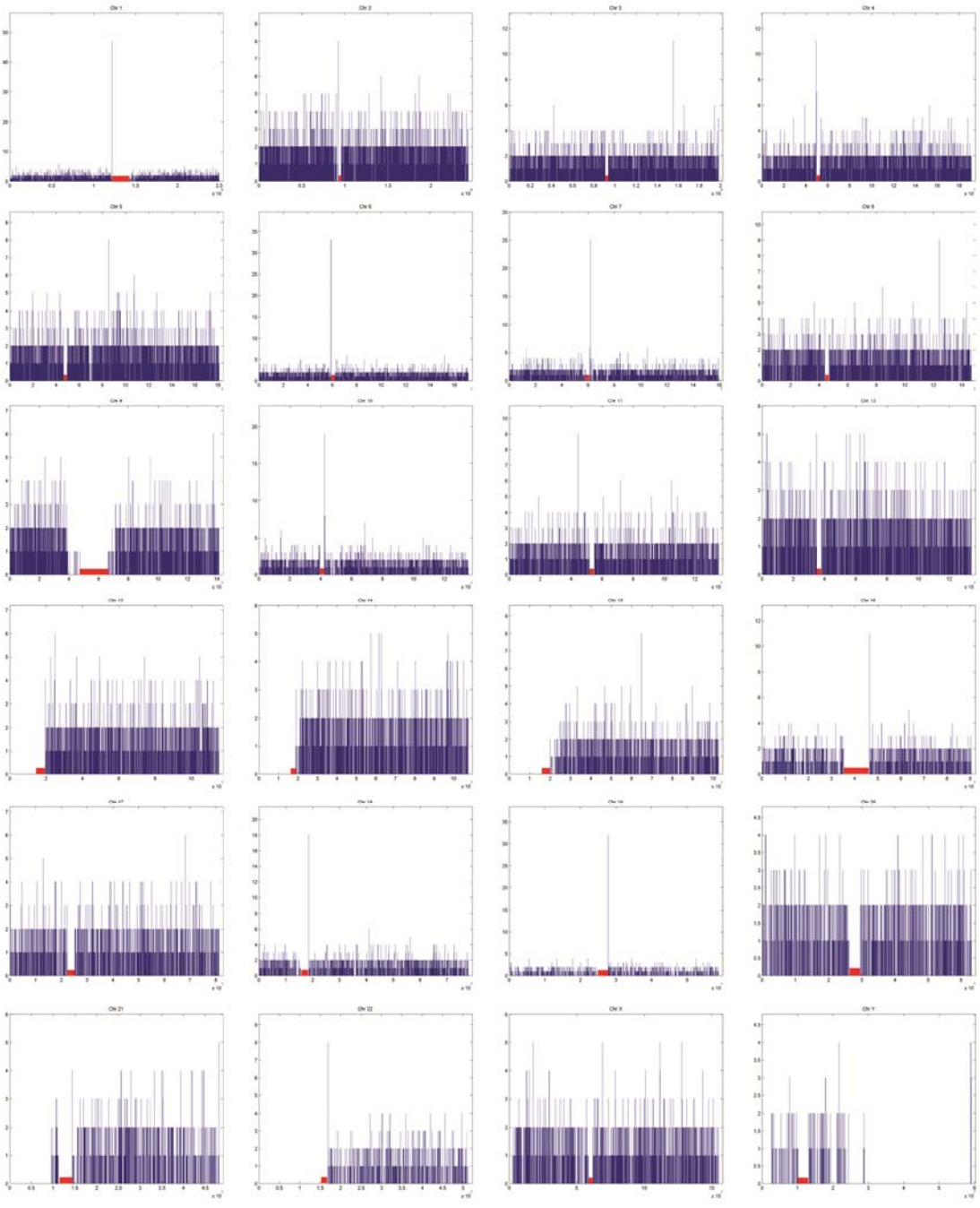


Figure 47 General distribution of $Z\alpha_{ADAR1}$ ChAP result on chromosomes, by using the “Library of original sequences”. Centromere regions are labeled in red.

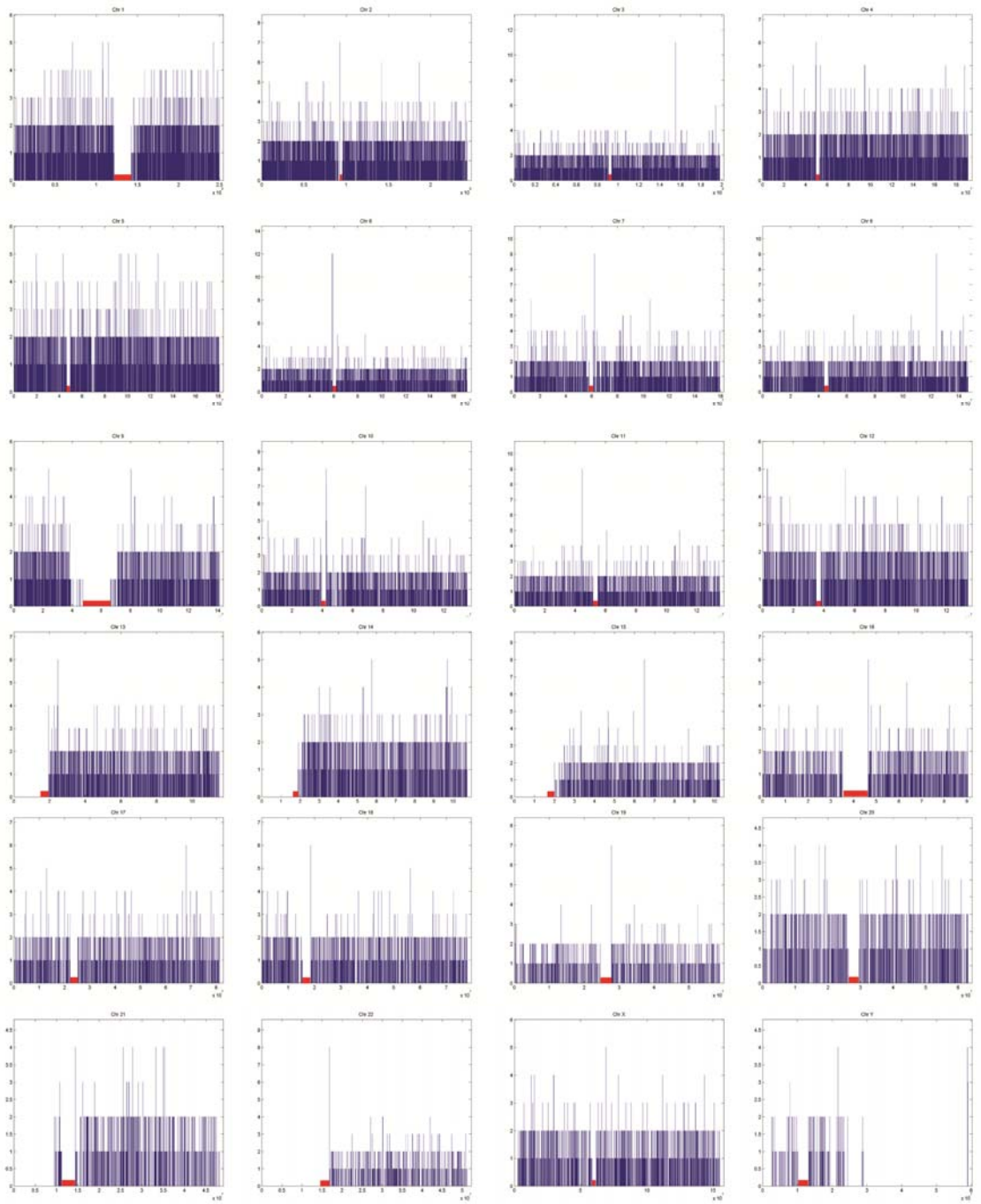


Figure 48 General distribution of $Z\alpha_{ADAR1}$ ChAP result on chromosomes, by using the "Library of Remaining libraries". Centromere regions are labeled in red.

Chr No	Centromere Left			Centromere Right		
	A549	huES		A549	huES	
		The original library	Library with remaining sequences		The original library	Library with remaining sequences
1	√	√				
2	√	√	√			
3						
4	√	√	√			
5				√		
6	√	√	√			
7				√	√	√
8						
9						
10				√	√	√
11	√					
12		√				
13						
14						
15						
16					√	√
17	√					
18				√	√	√
19					√	√
20						
21					√	
22					√	√
X				√		
Y				√		

Table 2 **Comparison of near centromere enrichment of $Z\alpha_{ADAR1}$ ChAP sequences between A549 and huES cell.**

The near centromere regions were sorted according to the chromosome number and the left or right side. The regions that exhibited enriched sequences among A549 and both of the huES libraries are highlighted in yellow. The regions that exhibited enriched sequences from both A549 and the huES original library are highlighted in blue. The regions that exhibited enriched sequences from both of the huES libraries are labeled in red.

III.6.3 Enrichment of $Z\alpha_{ADAR1}$ binding sites near centromere regions with high SNP densities

SNP densities on the $Z\alpha_{ADAR1}$mut ChAP sequences were calculated based on their distance from the centromeres (Figure 49 **a**). As SNPs are enriched near centromeres, the corresponding $Z\alpha_{ADAR1}$mut sequences were included as reference for each region.

By comparing the relative SNP densities on $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut, an obvious enrichment on $Z\alpha_{ADAR1}$ binding sites at the near centromere region could be observed. This enrichment is reduced when the segment length near centromeres is expanded. It reaches almost the same level as $Z\alpha_{ADAR1}$mut when the whole genome was analyzed (Figure 49 **b**). Thus, SNPs are enriched at or near $Z\alpha_{ADAR1}$ binding sites, and this trend is much more pronounced closer to centromere regions.

a

Distance from centromere	Hit	Total	Match	Match/Total	Normalized to $Z\alpha_{ADAR1}$mut
$Z\alpha_{ADAR1}$/ 50kb	43	43	4589	106.7209302	3.213103
$Z\alpha_{ADAR1}$mut/ 50Kb	14	14	465	33.21428571	1
$Z\alpha_{ADAR1}$/ 500Kb	103	140	4772	34.08571429	1.693957
$Z\alpha_{ADAR1}$mut/500Kb	34	41	825	20.12195122	1
$Z\alpha_{ADAR1}$/5Mb	1116	1638	7823	4.775946276	1.3444
$Z\alpha_{ADAR1}$mut/ 5Mb	360	505	1794	3.552475248	1
$Z\alpha_{ADAR1}$ Entire chromosome	18850	26786	46020	1.718061674	1.03199
$Z\alpha_{ADAR1}$mut Entire chromosome	6188	8771	14602	1.664804469	1

b

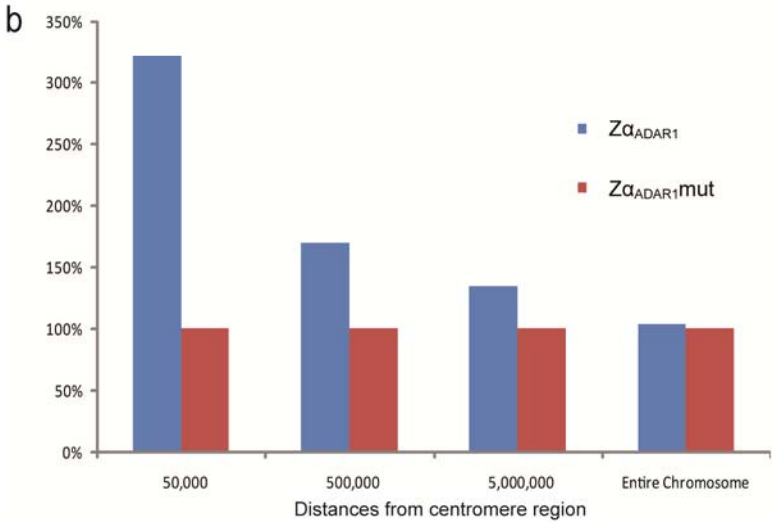


Figure 49 Enrichment of SNP densities at $Z\alpha_{ADAR1}$ binding site near centromere regions.

SNP densities in $Z\alpha_{ADAR1}$(mut) ChAP sequences were analyzed. The "Library with Clustered Sequences" is used in this analyze. (see Figure 21) **a.** Hit: the number of $Z\alpha_{ADAR1}$(mut) sequences that have at least one SNP inside; Total, total number of $Z\alpha_{ADAR1}$(mut) sequences in this region; Match: total number of SNP in $Z\alpha_{ADAR1}$(mut) sequences, i.e. if two SNP fell in one $Z\alpha_{ADAR1}$(mut) sequences , the number is two instead of one. Match/total: average SNP per sequence. Normalized to $Z\alpha_{ADAR1}$mut: the average SNP number per sequences of $Z\alpha_{ADAR1}$ is normalized to $Z\alpha_{ADAR1}$mut as 100%, and plotted in **b**.

III.6.4 Distribution of $Z\alpha_{ADAR1}$ ChAP sequences at transcriptional start sites (TSS) in the human genome.

As Z-DNA is predicted to have a higher chance to form at or before TSS (Rich and Zhang 2003), ChAP sequences were mapped to these regions. Here, both the "Library of Clustered sequences" and the "Library with Remaining Sequences" were used for mapping (see Section II.21.5 Figure 21). For both libraries, $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut ChAP sequences showed very different trends at the point of transcriptional start: the $Z\alpha_{ADAR1}$ sequences were clearly enriched within the surrounding 400 bp, while the opposite was found for $Z\alpha_{ADAR1}$mut sequences (Figure 50 **a** and **b**). The number of $Z\alpha_{ADAR1}$ ChAP sequences was normalized to the $Z\alpha_{ADAR1}$mut. An obvious enrichment at the TSS is still revealed, and the trend was more pronounced when the "Library of Clustered Sequences" was used (Figure 50 **c**). The "Library of Clustered Sequences" gives a higher confidence for Z-DNA formation at the locations (see also, Section II.21.5 Figure 21), therefore, after this comparison we can conclude with more confidence that the enrichment of Z-DNA formation at the TSS is substantial.

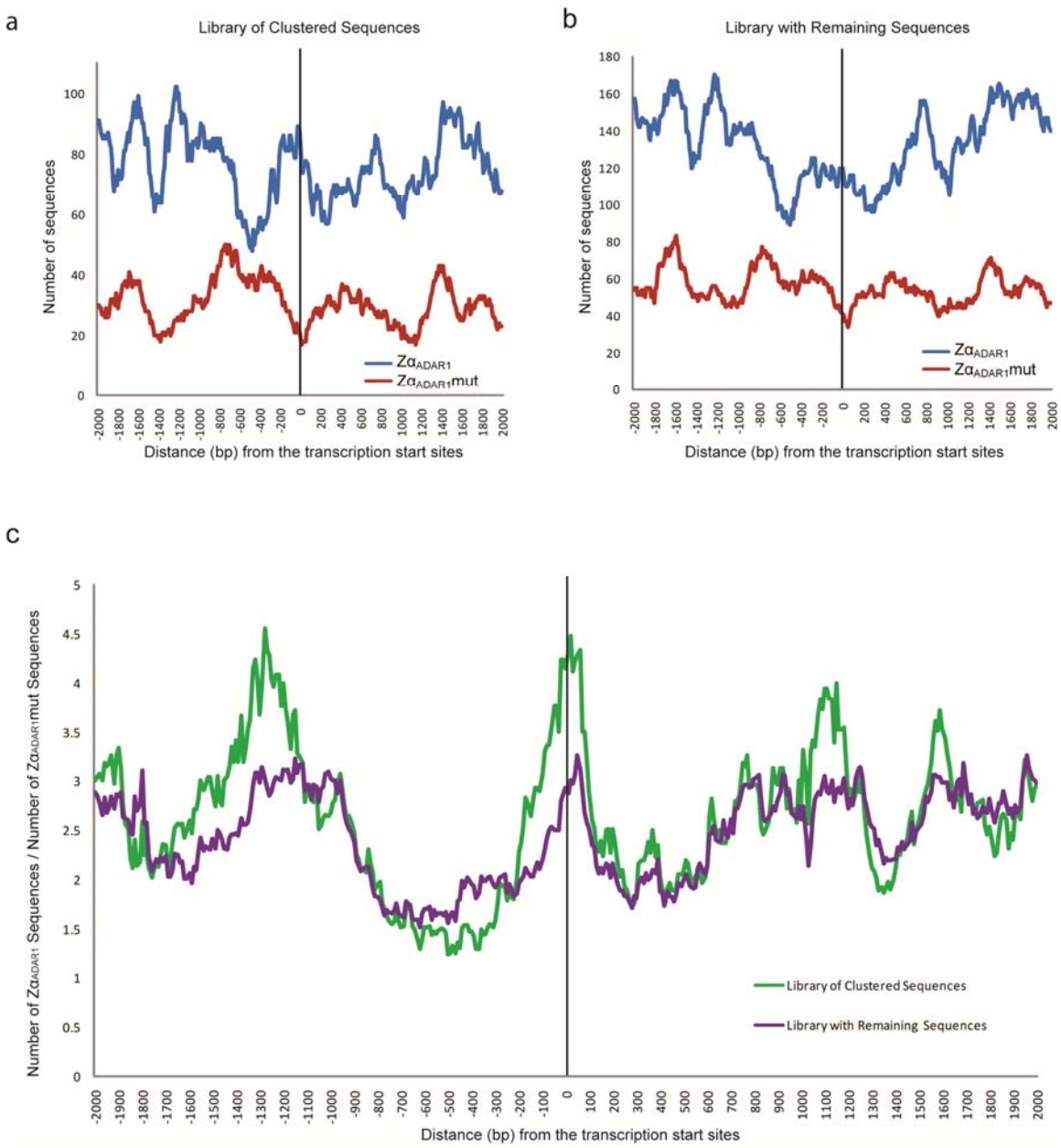


Figure 50 Distribution of ChAP result at the transcription starting site

The $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut result were mapped to transcription starting sites throughout the human genome, and the nearby 2000 bp regions were analyzed using the "Library of Clustered sequences" (**a**) and the "Library with Remaining Sequences" (**b**).The $Z\alpha_{ADAR1}$ result was normalized to $Z\alpha_{ADAR1}$mut in both libraries and compared in **c**.

III.6.5 No enrichment of $Z\alpha_{ADAR1}$ ChAP sequences at CpG islands

CpG islands exhibit high GC contents with high frequency of GC di-nucleotides. They usually appear at the 5' regions of house-keeping genes (Sharif, Endo et al. 2010). Since both properties facilitate Z-DNA formation, the latter could theoretically be related to the presence of CpG islands.

By using our "Library with Clustered sequences" (see Figure 21), I found a total of 128 $Z\alpha_{ADAR1}$ binding sites overlapping with CpG islands in a region spanning 17551 bp. This should be compared with 69 $Z\alpha_{ADAR1}$mut binding sites overlapping with CpG islands over a region of 10223 bp. The total number of binding sites in this library is 26782 for $Z\alpha_{ADAR1}$ and 8762 for $Z\alpha_{ADAR1}$mut. As each sequence is assigned a length of 200bp, it follows that 17551/(26782 X 200) = 0.328% of the $Z\alpha_{ADAR1}$ bound sequences and 10223/(8762 X 200) = 0.583% of the $Z\alpha_{ADAR1}$mut sequences overlapped with CpG islands. The value of $Z\alpha_{ADAR1}$mut is therefore higher than that determined for $Z\alpha_{ADAR1}$. By using the "Library with Remaining Sequences" as template, a similar trend was obtained (0.264% for $Z\alpha_{ADAR1}$ and 0.428% for $Z\alpha_{ADAR1}$mut), suggesting that there is no enrichment of $Z\alpha_{ADAR1}$ in CpG islands in this study.

III.6.6 Enrichment of $Z\alpha_{ADAR1}$ binding sites in G-bands

The genome can be divided into five groups according to the G-band intensity: gneg, gpos 25, gpos50. gpos 75 and gpos 100. Gneg refers to G-band negative, which are the palest bands during Giemsa staining, while the gpos refers to the G-band positive and is further divided into four groups based on the intensity level.

The binding site density was calculated via the number of $Z\alpha_{ADAR1}$ or $Z\alpha_{ADAR1}$mut ChAP sequences that fell in each group divided by the total length (bp) of the corresponding group. For $Z\alpha_{ADAR1}$, the density continuously increased with the increment of G-band density, while the sequence density for $Z\alpha_{ADAR1}$mut probe did not change drastically (Table 3). Furthermore, for all chromosomes that have a gpos100 band, the binding site density for $Z\alpha_{ADAR1}$ in gpos100 is always higher than the corresponding one for gneg. This is not observed for $Z\alpha_{ADAR1}$mut (Supplementary Table 1)

Therefore, $Z\alpha_{ADAR1}$ preferentially binds to Giemsa stained positive regions, while the $Z\alpha_{ADAR1}$ does not, suggesting that Z-DNA is more favor to be formed in the G-band intense region than in the G-band pale region.

G-band intensity	Total length from all Chromosomes	$Z\alpha_{ADAR1}$		$Z\alpha_{ADAR1}$mut	
		Hit	Hit/totalX10^5	Hit	Hit/totalX10^5
gneg	1377861883	12174	0. 88354	4041	0.29328
gpos 25	434995999	3939	0.90553	1261	0.28989
gpos 50	447426600	4274	0.95524	1489	0.33279
gpos 75	491421800	4868	0.9906	1581	0.32172
gpos 100	596920200	6300	1.05542	1994	0.33405

Table 3 General distribution of $Z\alpha_{ADAR1}$(mut) ChAP result on G-band

The G-bands were divided into 5 groups according to the G-band intensity: gneg which refers to G-negative while gpos which refers to G-positive is divided into groups based on the staining intensity level. The darker the bands after staining, the higher the number is. The total length in bp of each group was determined, labeled as “Total length from all chromosomes”. The number of binding sites that fell into each group was counted and labeled as “Hit”. The ratio of $Z\alpha_{ADAR1}$ or $Z\alpha_{ADAR1}$mut hits per total length was calculated accordingly, and labeled as “Hit/Total”. The “Library with Clustered Sequences” is used in this analyze. (see Figure 21)

IV DISCUSSION

IV.1 Binding of $Z\alpha_{ADAR1}$ to ribosomes and its biological implications

IV.1.1 The reliability of identifying $Z\alpha_{ADAR1}$ binding sites on ribosomes

I employed a “native” pull down approach to find binding sites on ribosomes, i.e. no crosslinking or other methods were used to stabilize the $Z\alpha_{ADAR1}$-ribosome interaction. It is known that the Z-conformation is transient in DNA and RNA, and it may therefore be surprising that RNA fragments maintain this conformation in a complex with $Z\alpha_{ADAR1}$ during the pull-down process. However the method to identify binding sites is still feasible in my opinion.

Firstly, $Z\alpha_{ADAR1}$ could bind and stabilize the Z-conformations up to a few hours (Kang, Bang et al. 2009; Bae, Kim et al. 2010). It is possible that $Z\alpha_{ADAR1}$ recognizes and binds to the Z-conformation on ribosome. Once it binds, it could stabilize the binding region in Z-conformation even after RNase A digestion of nearby accessible RNA segments and the resulting disassembly of the ribosome subunits.

Secondly, during the whole process, the same buffer condition was used. i.e. the general environment around the binding region was not changed. In addition, several

pull-down experiments were performed, and the same binding sites could be identified in all experiments, implying a high degree of reproducibility.

Thirdly, one may argue that $Z\alpha_{ADAR1}$, which is a DNA binding domain, could randomly bind to any Z-forming sequences. As described previously, Zα domains interact in a similar way with five nucleotides in one strand of a left–handed Z–DNA or Z–RNA duplex (Schwartz, Rould et al. 1999; Placido, Brown et al. 2007). The Z-conformation structures represent higher energy states, and occur preferentially in stretches of alternating purine and pyrimidine. There are total more than 60 RNA segments that have at least three continuing pairs of alternative purines and pyrimidines, which are on the surface of the ribosome subunits based on the crystal structure (Figure 51). These represent potential Z-forming sequences and should be fully accessible for $Z\alpha_{ADAR1}$ binding in solution. During the pull-down experiment, they were under the same conditions as the "true binding sites" identified as hcRBS. If $Z\alpha_{ADAR1}$ just randomly binds to any Z-forming sequences, these sites should be pulled-down preferentially. However, there was no enrichment of them in the pulled-down sequences. Therefore, the rRNA fragments that were pulled-down by $Z\alpha_{ADAR1}$ are not purely dependent on sequence context. The conformation of the RNA and the local ribosomal environment must also play an important role.

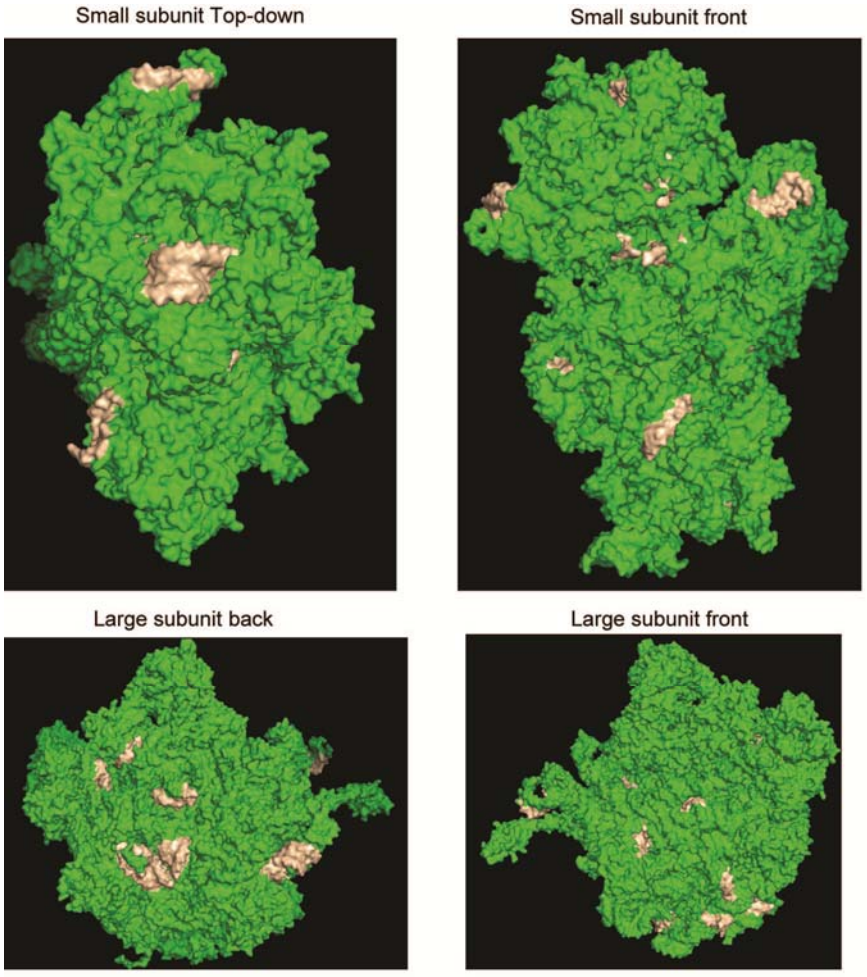


Figure 51 Stretches of alternative purine and pyrimidine residues on the ribosome surface

Surface view of the *E. coli* ribosome shown at different angles (PDB ID: 2AVY and 2AWB). Fragments that have at least three continuous pairs of alternative purine and pyrimidine were labeled in wheat color, while the rest of the ribosomal subunits were labeled in green.

IV.1.2 Binding of $Z\alpha_{ADAR1}$ to ribosomes depends on the presence of Z-like conformations

The results showed that binding of $Z\alpha_{ADAR1}$ to ribosomes occurs most likely via direct interactions with rRNA, and is structure dependent. The Z-conformation specificity of $Z\alpha_{ADAR1}$ and the loss of Z-conformation binding specificity of $Z\alpha_{ADAR1}$mut were fully demonstrated *in vitro* (Li, Xiao et al. 2009). The difference in the ability to bind ribosomal RNA between $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut is very obvious. At the same time, the DNA-binding domains ($Z\alpha_{ADAR1}$, $Z\alpha\alpha_{ADAR1}$, $Z\alpha\beta_{ADAR1}$ and $Z\alpha_{DLM1}$) that possess Z-DNA binding ability could bind to ribosome, while domains that are mutated ($Z\alpha_{ADAR1}$mut) or naturally impaired ($Z\alpha_{E3L}$) in Z–DNA/RNA recognition lacked the ability for ribosome association. Combined, these results indicated that the binding of $Z\alpha_{ADAR1}$ to ribosomes is Z- or Z-like-conformation dependent.

IV.1.3 $Z\alpha_{ADAR1}$ binding sites on *E. coli* ribosome mapped to crystal structure

In order to find the association of ribosome binding site of $Z\alpha_{ADAR1}$ with the effects on biological functions, the potential binding sites of $Z\alpha_{ADAR1}$ on ribosome were mapped to crystal structure.

IV.1.3.1 $Z\alpha_{ADAR1}$ binding sites on the small *E. coli* ribosomal subunit

The binding site was mapped at the “head” region of the small ribosome subunit (Figure 52 **a**). A surface view (Figure 52 **b**) showed that the site is not that exposed in the crystal structure. The exposed area does not have enough space for $Z\alpha_{ADAR1}$ occupation**.** One explanation could be $Z\alpha_{ADAR1}$ binds to rRNA precursor before the rRNAs assembles into the full ribosome. This interpretation is supported by the fact that $Z\alpha_{ADAR1}$ pulled-down both large and small subunits *in vivo*, while it can only pull-down the large subunit under static *in vitro* conditions. There, the binding site in the small subunit of fully assembled ribosome is not accessible for the $Z\alpha_{ADAR1}$ binding.

However, as $Z\alpha_{ADAR1}$ could inhibit *in vitro* translation in *E. coli* lysate, $Z\alpha_{ADAR1}$ should be able to bind to the fully assembled ribosome subunit to carry out its translation inhibition activities, and be able to access the binding site of the intact ribosome subunits. One possibility is that the active binding site for inhibition of *in vitro* translation is on the large subunit. Alternatively, as the head region of ribosomes undergoes dramatic conformational changes during translation (Ratje, Loerke et al. 2010), while in the crystal structure only one conformation is presented, it is possible that at certain stages during translation, the binding site on the head region is exposed at the surface and bound by $Z\alpha_{ADAR1}$. Binding of $Z\alpha_{ADAR1}$ to that region could stabilize the conformation, and interfere with the translation process. An interesting future experiment therefore would be to test whether the small subunit might be pulled-down under *in vitro* translation conditions.

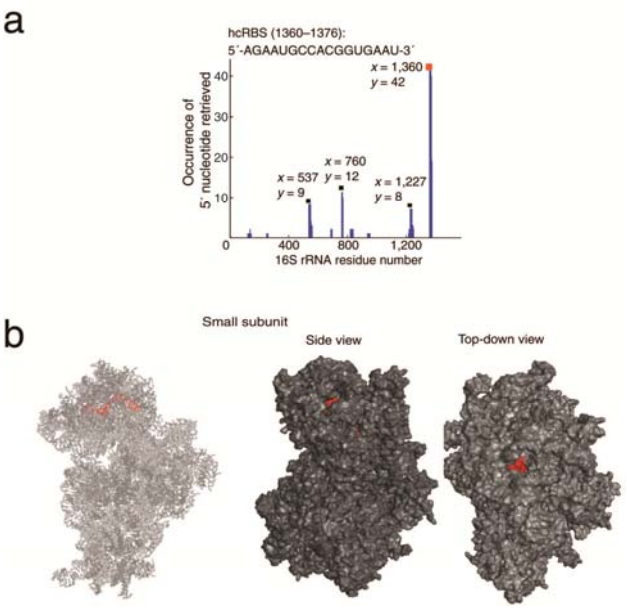


Figure 52 General view of Zα$_{ADAR1}$ binding sites on the small ribosomal subunit of *E. coli*

a. The sequences from the three experiments were combined and mapped to the rRNA position. **b**. Left panel, the ribosome structure is shown as line drawing (PDB ID: 2AVY). Right panel, the structures are displayed in surface view. The overall small subunit is labeled in gray, and the hcRBS is labeled in red (1360-1377).

IV.1.3.2 Zα$_{ADAR1}$ binding sites on the large subunit of *E. coli* ribosomes

As shown in Figure 53 **a**, rRNA fragments (548-564) that appeared in all pull-down experiments was highlighted in red and defined as hcRBS, while the other potential binding sites are labeled in orange (294-333), yellow (988-1003 and 1142-1158) and magenta (1206-1223). All of them are exposed to the surface (Figure 53 **b**) and near to each other, with the hcRBS is in a central position between the others. As all potential binding sites were linked to each other through either RNA or protein interactions, it is very likely that the other potential binding sites were co-precipitated due to incomplete RNase digestion. This is also supported by the fact that, in the complete digestion

samples, only the hcRBS was significantly represented, while the other sites were only observed in the less-digested sample (please compare Experiments 1 to 3 in Figure 29).

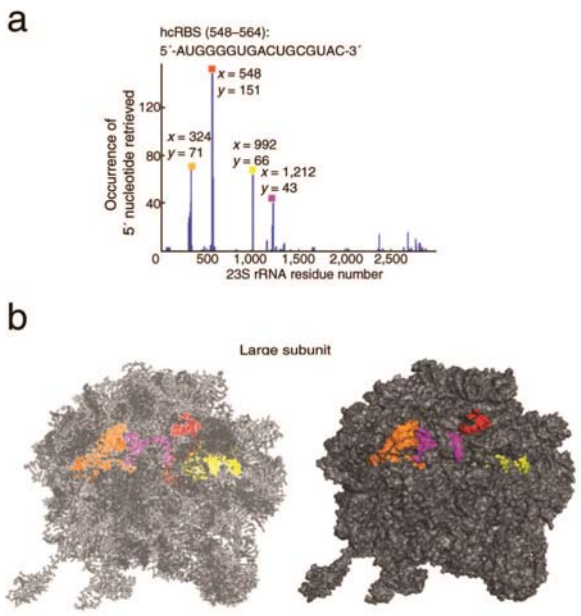


Figure 53 General view of Zα$_{ADAR1}$ binding sites on the large ribosomal subunit of *E. coli*

a. The sequences from the three pull-down experiments were combined and mapped to the rRNA position. **b**. Left panel, the structure is shown as line drawing. Right panel, the structures are displayed in surface view (PDB ID: 2AWB). The overall large subunit is labeled in gray, and the hcRBS is labeled in red (548-564). The other peaks from the pull down experiment were also mapped and labeled in the corresponding color code.

In order to understand the possible mode of Zα$_{ADAR1}$ interaction/recognition with hcRBS, the two hcRBS segments (from large and small subunits) as present in the crystal structure were highlighted (Figure 54). The hcRBS on the large subunit was mapped to only one strand of a hairpin composed of a right–handed helical stem containing several non–canonical base pairs and an unstructured loop with extruded bases (Figure 54, left

panel). The loop is the most accessible portion of the hairpin. Due to different *E. coli* strains, the loop sequence in the crystal structure deviates slightly from that in our hcRBS (compare below Figure 55 **a** and **b**). There were no structural features resembling Z–RNA, but we noted a stretch of seven alternating purines and pyrimidines at the base of the stem that was present only in the strand bound by $Z\alpha_{ADAR1}$ (Figure 54, left panel labed in red).

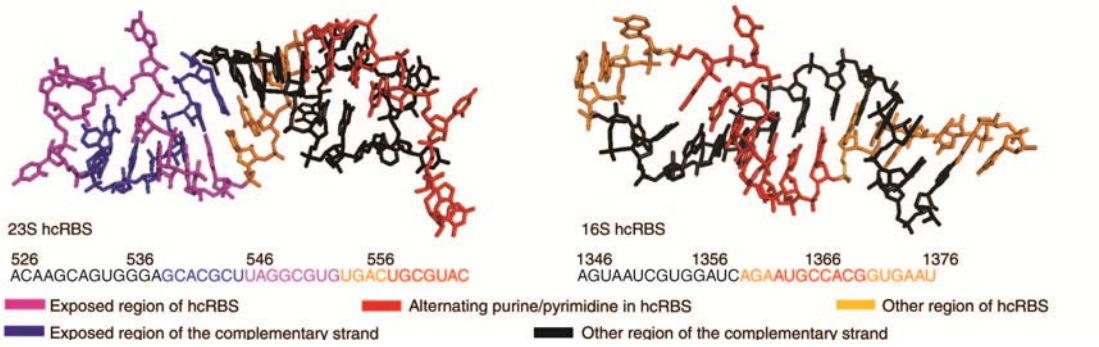


Figure 54 Structure of the *E. coli* hcRBS based on the crystal mode

Secondary structures of both *E. coli* hcRBS based on the *E. coli* ribosome crystal structure (PDB ID: 2AWB, 2AVY). The hcRBS and its nearby sequences were analyzed based on exposure and sequence composition. The hcRBS from the large and small subunit were showed in left and right panel, respectively.

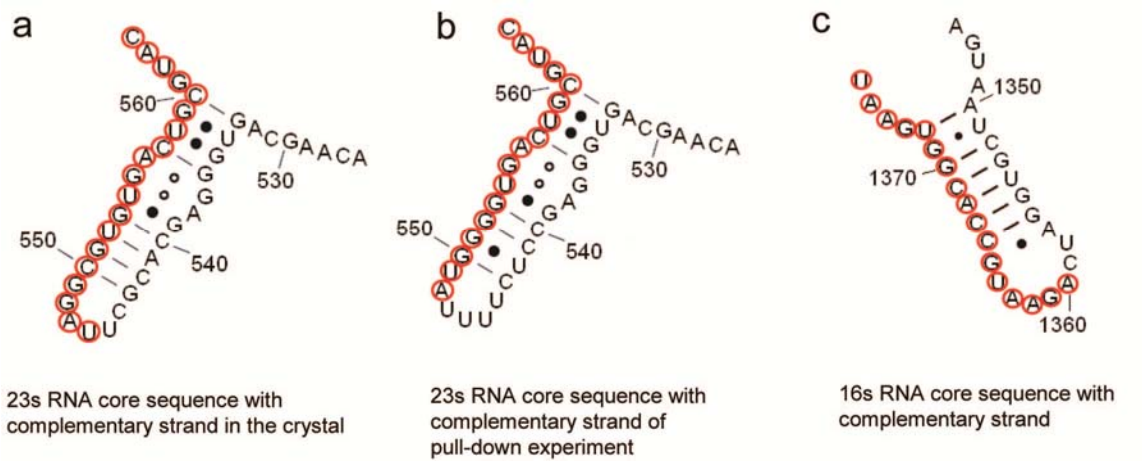


Figure 55 Secondary structure comparison of hcRBS from *E. coli*

The 16S hcRBS was also mapped to one strand of a hairpin with an unstructured loop that contains extruded bases. The stem is composed of mostly canonical base pairs (Figure 55 **c**). No Z–RNA feature was observed, but two adjacent stretches composed of alternating purines and pyrimidines were detected in the strand bound by $Z\alpha_{ADAR1}$ (Figure 54, right panel labeled in red).

Hence, a common feature for the *E. coli* hcRBS is that both represent one strand of a hairpin that harbors alternating purine and pyrimidine residues. However, the predicted binding sites do not adopt Z-conformation in the crystal. As the bacterial strain use in the crystal structure is different from the strain we used in pull-down experiment, the sequence is different at the potential binding site on the large subunit. This may lead to a different conformation at that position. Furthermore, ribosomes undergo dramatic conformation changes during translation, while the crystal structure represents only one of the most stable conformation snapshots. The rRNAs bound to $Z\alpha_{ADAR1}$ in solution might thus adopt a different structure from the one in the crystal.

IV.1.4 $Z\alpha_{ADAR1}$ binding sites on human ribosomes

The sequencing results from either *in vitro* or *in vivo* pull-down experiments were combined (Figure 56). For the *in vitro* experiment, we identified three hcRBS in the large subunit and one in the small subunit (Figure 56 **a**, left and right panel, respectively). While for the *in vivo* experiment, one hcRBS was found on the large and two hcRBS on the small subunit (Figure 56 **b**, left and right panel respectively). Importantly, the sole hcRBS on the large subunit is identical to one of the three *in vitro* hcRBS. The two hcRBS on the small human subunit were predicted to be very close to each other (Figure 56 **b**, left panel in dark blue).

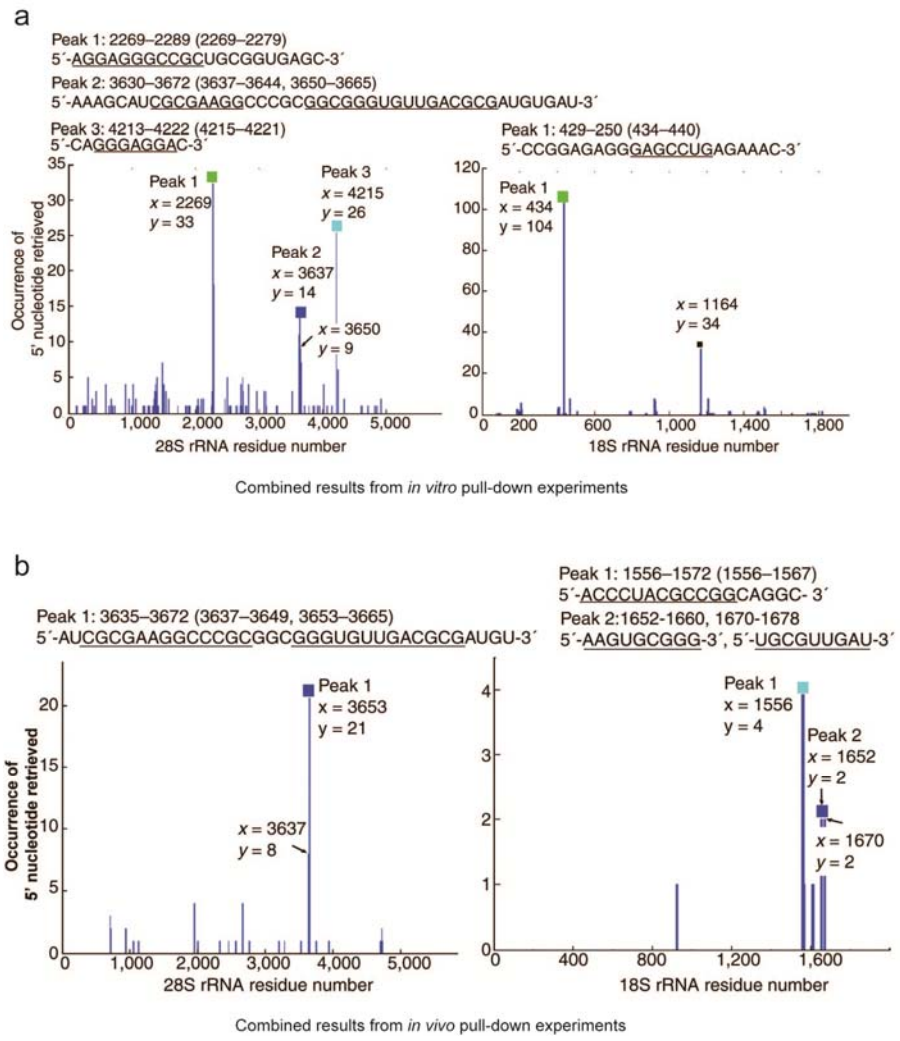


Figure 56 Combined results from HeLa pull-down experiments

Results from two independent experiments were combined (see also Section III.4.4). The peaks considered as hcRBS were labeled in different colors. The same color code was applied for identical mappings.

IV.1.4.1 RNA secondary structure comparisons of binding sites on *E. coli* and HeLa ribosomes.

As high resolution crystal structures of human ribosome are not yet available, the identified hcRBS on HeLa ribosomes were mapped to *E. coli* rRNA on the basis of predicted secondary rRNA structures (Figure 57). In addition, the positions of human hcRBC were also mapped to a recently published low resolution ribosome structure obtained with Cryo-EM (Chandramouli, Topf et al. 2008). The predicted positions of *in vitro* and *in vivo* human hcRBS on *E. coli* ribosomes are very similar to the actual position of hcRBS on the *E. coli* ribosome structure (Figure 59 and Figure 60, compare upper and lower panels).

For the *in vitro* experiment, as shown in Figure 57, the corresponding human hcRBSs are far away from the binding site in *E. coli* in the secondary structure.

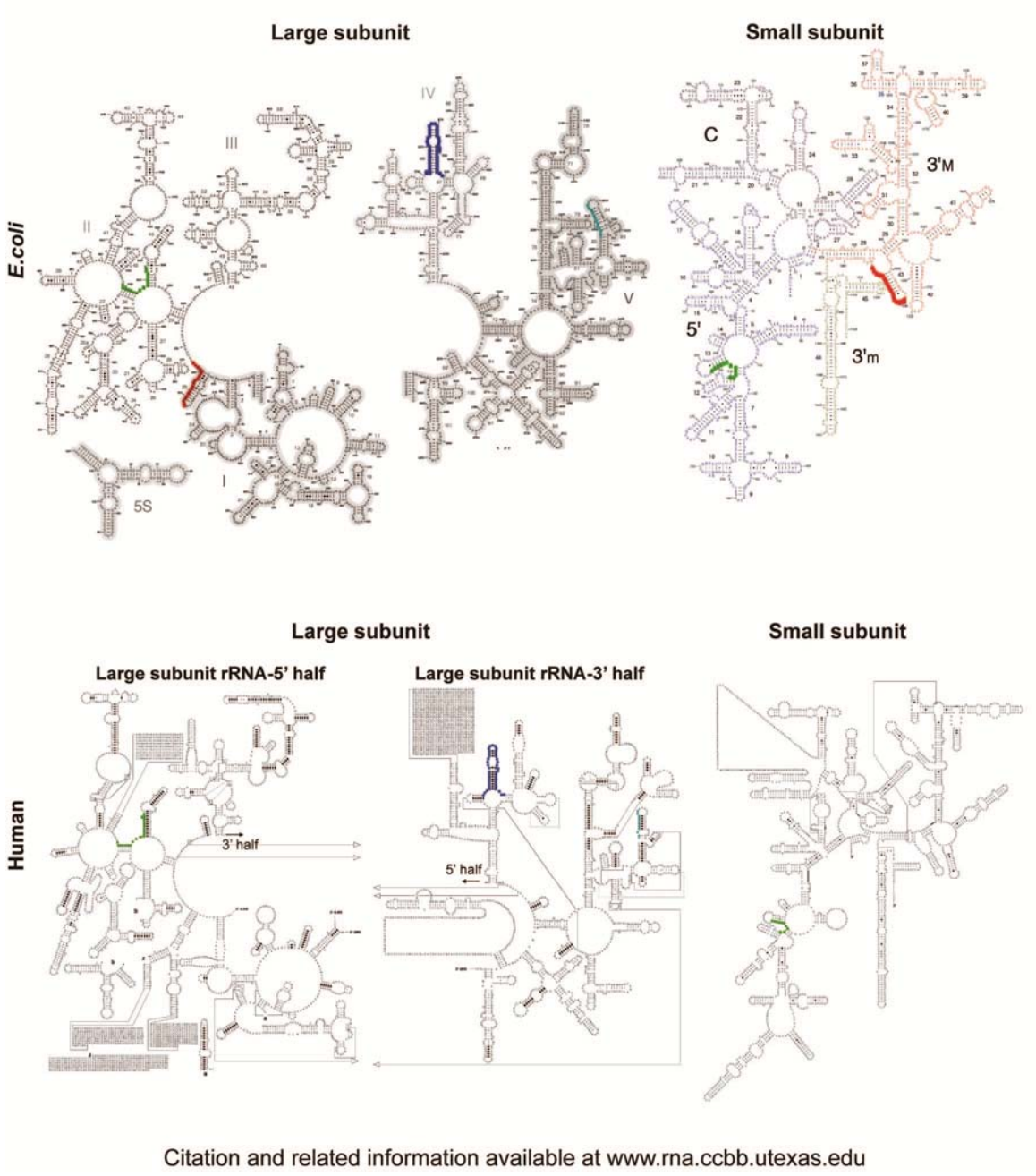


Figure 57 Secondary structure prediction of the *in vitro* hcRBS in human ribosomes, based on *E. coli* secondary structures.

The identified human hcRBS are labeled in the same color code as in Figure 39 and Figure 56. As reference, the *in vivo E. coli* hcRBSs were also included and labeled in red. The positions of the hcRBSs were first located in the human rRNA secondary structure map (bottom) and then mapped them to *E. coli* (top).

For the *in vivo* experiment, the predicted position of the hcRBS in the HeLa large subunit is far away from the hcRBS in *E. coli*. One of the two hcRBS on the human small subunit, which are predicted to be on the two strands from the same loop (

Figure 58 dark blue), were mapped precisely to the position of the *in vivo E. coli* hcRBS (

Figure 58 red). In the human subunit, this hcRBS appeared to comprise both strands of a hairpin with an alternative purine and pyrimidine sequence (GUGCG·UGCGU).

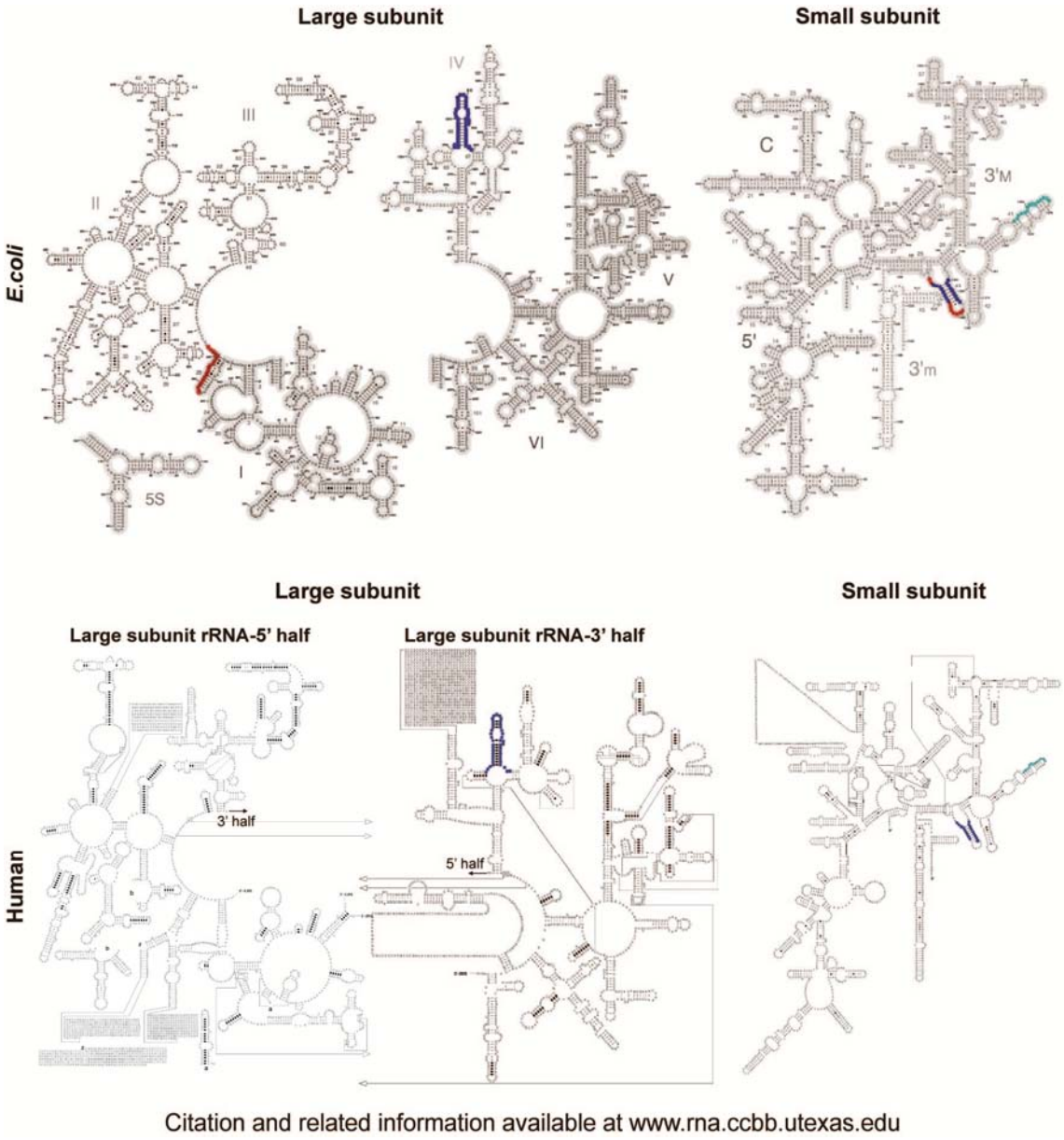


Figure 58 Secondary structure prediction of the *in vivo* hcRBS in human ribosomes, based on *E. coli* secondary structures.

The identified human hcRBS are labeled in the same color code as applied in Figure 40 and Figure 56. As reference, the *in vivo E. coli* hcRBSs were also included and labeled in red. The positions of the hcRBSs were first located in the human rRNA secondary structure map (bottom) and then mapped to *E. coli* (top).

The corresponding positions of HeLa hcRBS were also mapped to the *E. coli* ribosome crystal structures.

For the *in vitro* experiment, one of the corresponding human hcRBS on the back of the large subunit was very close to the potential binding sites in *E. coli* (labeled in green in Figure 59, left panels) in the 3D structure. The other two *in vitro* hcRBSs were at the interface with the small subunit (labeled in light and dark blue in Figure 59, middle panels).

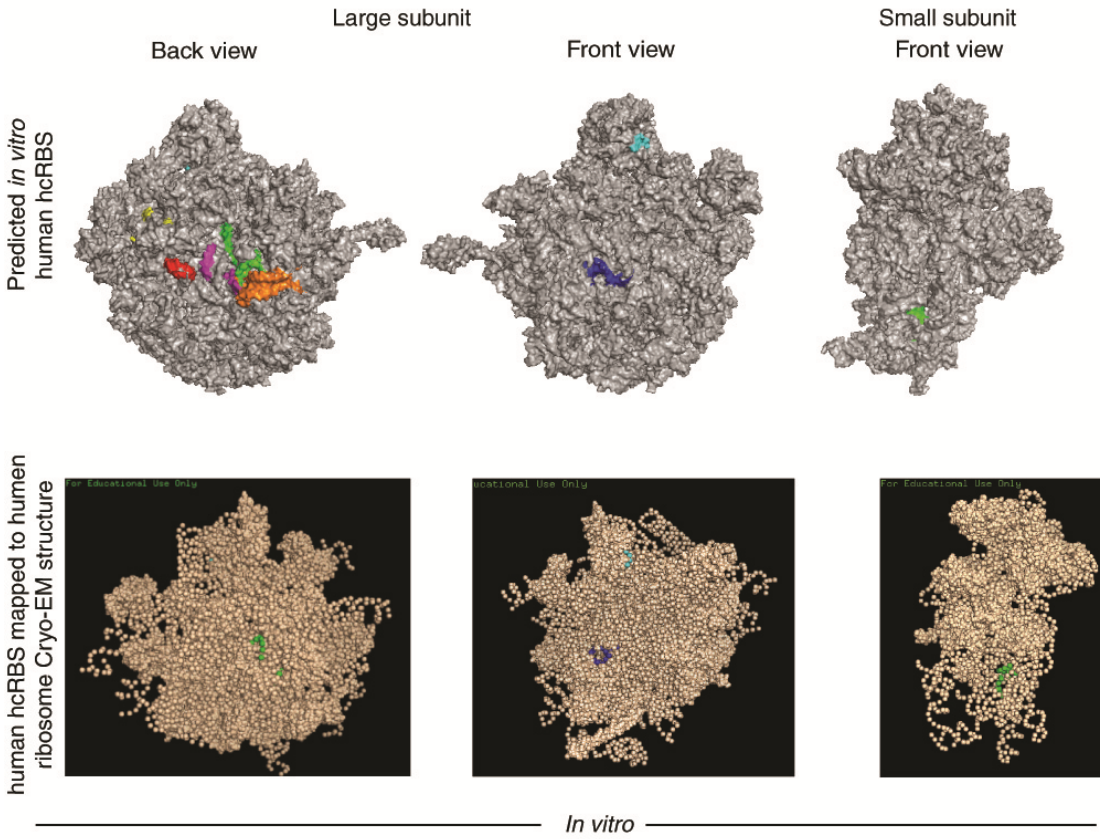


Figure 59 The *in vitro* human hcRBS mapped to *E. coli* crystal structure based on secondary structure predictions and human ribosome Cryo-EM structures.

The hcRBS in *E. coli* that was identified in all experiments is labeled in red (548-566) while the other "potential binding sites" are labeled in orange, yellow and magenta. The hcRBS on human ribosomes were mapped to *E. coli* rRNA according to predicted secondary structures, and are labeled in green, cyan and blue. The other part of the *E. coli* ribosome is labeled in gray (Upper panel). The positions on human ribosomes determined by Cryo-EM structures were also included as a reference. The same color code was applied to hcRBS, while the remaining part of the human ribosome is labeled in wheat color (Lower panel). (Chandramouli, Topf et al. 2008)

For the *in vivo* experiment (Figure 60), one of the corresponding human hcRBS (dark blue) in the small subunit overlapped with the *in vivo* hcRBS in *E. coli*, while the other one (cyan) is located in the immediate vicinity, perhaps being co-precipitated with the dark blue hcRBS (Figure 60, middle and right panels). The *in vivo* binding sites that appeared, therefore, to be conserved between *E. coli* and HeLa ribosomes were relatively buried within the head region of the small subunit. However, as the head region of the small subunit is very dynamic during translation, it may be exposed during translation process (Ratje, Loerke et al. 2010). This is supported by the fact that in both *E. coli* and HeLa systems, this binding site was only observed in *in vivo* pull-down samples, and not in translation-inactive *in vitro* samples. Binding of $Z\alpha_{ADAR1}$ or ADAR1 on this site during translation might stabilize a local rRNA structure in Z conformation, and block further ratchet-like movements of the head domain involved in translational elongation.

The human hcRBS which was mapped on the large subunit and overlapped with the *in vitro* result was not correlated with any *E. coli* hcRBS (Figure 60, left panel). It is on the interface with the small subunit, and binding of $Z\alpha_{ADAR1}$ or ADAR1 to that hcRBS might affect assembly of 80S ribosomes and inhibit translation.

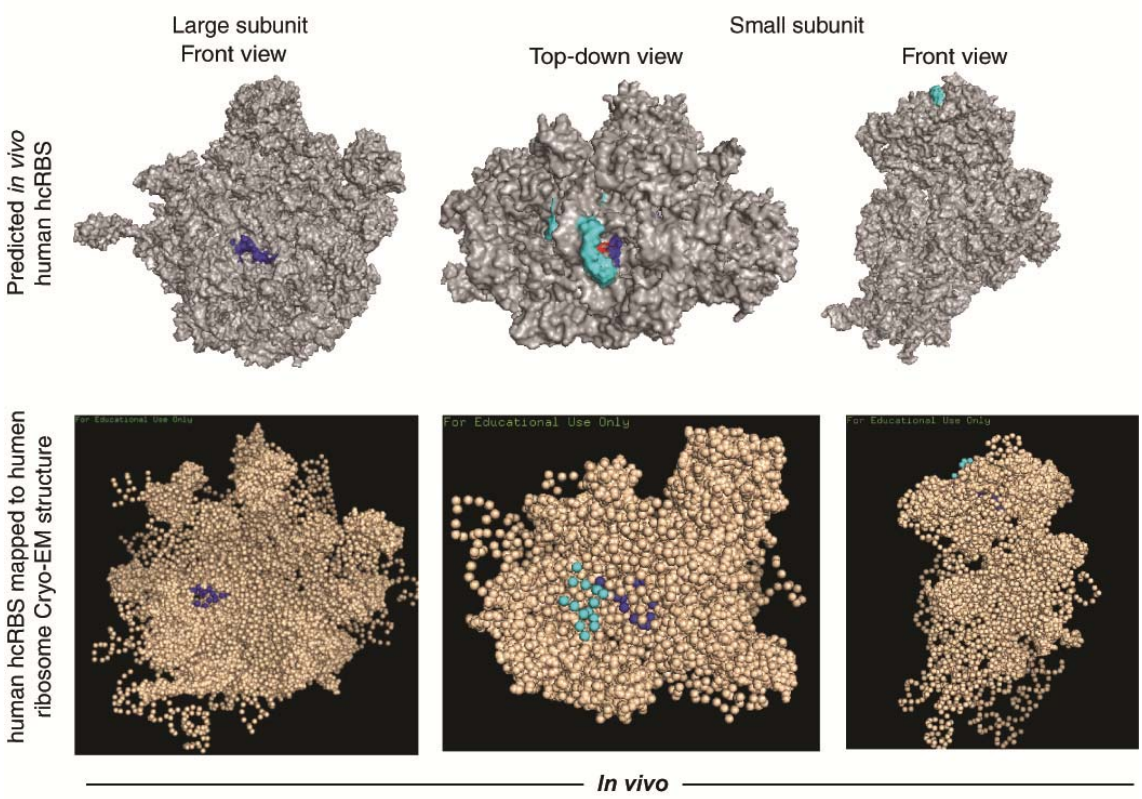


Figure 60 The *in vivo* human hcRBS mapped to *E. coli* crystal structures according to secondary rRNA structure predictions and human ribosome Cryo-EM structures.

The hcRBS in *E. coli* that appeared in all the experiment is labeled in red. The hcRBS on human ribosomes were mapped to *E. coli* rRNA according to predicted secondary structures is labeled in cyan and blue. The other part of the *E. coli* ribosome is labeled in gray (Upper panel). The positions on human ribosomes Cryo-EM structures were also included as a reference. The same color code was applied to the hcRBS, while the rest part of the human ribosome is labeled in wheat color (Lower panel). (Chandramouli, Topf et al. 2008)

IV.1.5 The potential biological significance of an association of $Z\alpha_{ADAR1}$ with ribosomes

Here, we showed that $Z\alpha_{ADAR1}$ could bind to ribosomes and inhibit translation in both *E. coli* and mammalian systems. The translational inhibition effect was also observed for the full-length ADAR1 *in vivo* in HeLa cells.

The induction of ADAR1-L expression by interferon and inflammation, but not other ADARs, implies a special involvement of Z-DNA binding domains in antiviral defense and inflammation. It is possible that $Z\alpha_{ADAR1}$ could turn off the translation machinery to carry out an anti-viral function. As the inhibitory effect is in stoichiometric manner, questions might rise on whether the expression level of the full-length protein could match the ribosome level in the mammalian system to carry out translational inhibition functions. However, ADAR1–L could bind to microbial RNAs via dsRNA binding domains. One proposed function there was editing of viral RNA as a defense against infection. The translation inhibition effect could provide a new layer of antiviral defenses that, concurrent binding of ADAR1–L to ribosomes via Zα could prevent translation of viral RNA. This would make translational inhibition by $Z\alpha_{ADAR1}$ RNA–specific and more effective.

On the other hand, regardless of the translation inhibition effect, binding of $Z\alpha_{ADAR1}$ itself could help the host defenses on virus infection. ADAR-L could bind to and standby at the ribosome via the Zα domain, and when a viral RNA is coming in for translation,

editing could be carried out specifically on the translated RNA by the editing domain. This makes the editing effect more specific and more effective.

Our result showed that $Z\alpha_{ADAR1}$ could inhibit *E. coli* growth. This is similar to the previous observation that cytoplasmic microinjection of antibodies recognizing Z-RNA could inhibit human cell growth (Zarling, Calhoun et al. 1990). In addition, our failure to establish $Z\alpha_{ADAR1}$ constitutive expression cell lines in different cell types also suggested that $Z\alpha_{ADAR1}$ is potential toxic to mammalian cells. It has been reported that $Z\alpha_{E3L}$ can block apoptosis induced by hygromycin B (Kwon and Rich 2005), which is opposite to our finding. However, if the origin of the two domains was considered, it is not surprising that these two Z-DNA binding domains could carry out opposite functions. $Z\alpha_{E3L}$ is from vaccinia virus and expressed to antagonize the host antiviral defense of IFN system, while $Z\alpha_{ADAR1}$ is a mammalian protein and its expression is induced by IFN to fight off infection.

IV.1.6 Future works on binding of $Z\alpha_{ADAR1}$ to ribosomes

At this point, the precise mechanism of translational inhibition by $Z\alpha_{ADAR1}$ is still not understood. Several possibilities, such as binding of $Z\alpha_{ADAR1}$ will interfere with 80S ribosome assembly or affect the movement of the head domain of the small subunit were proposed. Further experiments could focus on the mechanism of the translation inhibition effect.

On the other hand, binding of $Z\alpha_{ADAR1}$ to ribosomes in a Z-conformation-dependent way is well established. However, none of the binding sites adopts Z-conformation in existing crystal structures. This may due to the dynamic nature of the ribosome architecture, especially *in vivo*. A co-crystal structure of $Z\alpha_{ADAR1}$ bond to ribosomes could be obtained and is, in fact, under investigation. More details about the interaction between $Z\alpha_{ADAR1}$ and ribosomes could then be recovered. The co-crystal structure could also help us to understand the mechanism of translational inhibition. A transient state of ribosome structure during translation might be captured as well.

IV.2 Mapping of Z-DNA in huES cells

IV.2.1 *In vitro* ChAP VS. *in vivo* ChAP

Z-DNA-binding by $Z\alpha_{ADAR1}$ was well characterized. When $Z\alpha_{ADAR1}$ was first identified as a Z-DNA binding domain in A. Rich's laboratory at the MIT, it became obvious that it could be used as a highly specific probe to detect Z-DNA segments in any genome. Our results showed that the primary target for $Z\alpha_{ADAR1}$ is most likely the ribosome rather than genomic DNA. The *in vivo* transiently expressed $Z\alpha_{ADAR1}$ will easily be exhausted by the huge number of ribosomes (the expression level of $Z\alpha_{ADAR1}$ was estimated to be about 10^5 copies per cell). Furthermore, $Z\alpha_{ADAR1}$ is reported to have the ability to stabilize the B-to-Z transitions up to few hours (Kang, Bang et al. 2009; Bae, Kim et al. 2010), this may not represent the native topological enviroment. Hence, *in vivo* expression of $Z\alpha_{ADAR1}$ in cells is not a favorable approach when the native Z-DNA map in genomes is investigated. Also, as the $Z\alpha_{ADAR1}$ domain is found to have a biological effect on living

cells, *in vivo* expression might change the cellular environment as well as DNA conformations. The *in vivo* ChAP results may therefore not reflect the true Z-DNA occurrence for the cells. Another advantage of *in vitro* ChAP over *in vivo* ChAP is that the *in vitro* ChAP experiment can be performed on many different biological samples, such as fixed tissues; while *in vivo* ChAP experiment is limited by the availability for $Z\alpha_{ADAR1}$ expression.

The feasibility of using the *in vitro* ChAP method is based on previous findings that formaldehyde cross-linking will not affect binding of $Z\alpha_{ADAR1}$ to Z-DNA (Liu, Mulholland et al. 2006). Here, the cells were pre-crosslinked to stabilize the chromatin structure and Z-DNA before $Z\alpha_{ADAR1}$ proteins was added. Adding of $Z\alpha_{ADAR1}$ protein is unlikely to induce a B-Z transition in the fixed DNA. Excess protein could then be added to the cells to overcome the depletion by ribosomes.

IV.2.2 Choice of mapping and sequencing techniques

The “Illumina Chip-Seq” technique was used in this study to generate a Z-DNA library. Each identified pull-down fragment was used as source to locate the binding site. If compared to the “ChIP-on-Chip” method, it is a genome-wide scope of the interaction, which is not limited by pre-determined criteria (which is still not clear for Z-DNA mapping). Compared to the traditional cloning-based sequencing method, it requires less input material, provides higher output and is more cost effective.

IV.2.3 Is $Z\alpha_{ADAR1}$mut a good negative control?

The loss of Z-DNA binding specificity of $Z\alpha_{ADAR1}$mut is well characterized (Schwartz, Rould et al. 1999; Placido, Brown et al. 2007; Li, Xiao et al. 2009). However, $Z\alpha_{ADAR1}$mut still exhibits a rather high affinity to B-DNA. This could explain why $Z\alpha_{ADAR1}$mut has more ChAP material than the mock control, although one needs to consider the fact that three more amplification cycles were applied. Compared to purely *in silico* analyses, the samples from $Z\alpha_{ADAR1}$mut could in fact provide more information on the local environment of the genome, such as accessibility and mapping-ability. As the $Z\alpha_{ADAR1}$mut and $Z\alpha_{ADAR1}$ samples were treated identically during sample preparation, a comparison between $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut results could help to identify biases resulting from sample preparation.

However, as we could not get enough material for sequencing, a PCR amplification step is involved. Since the material from $Z\alpha_{ADAR1}$mut ChAP experiment is much less than the material from $Z\alpha_{ADAR1}$, three more cycles were applied to the $Z\alpha_{ADAR1}$mut. As the PCR amplification efficiency for different sequences might be different, the number of occurrences of each fragment is not representing the enrichment level of that fragment. In order to remove this amplification ambiguity, all identical sequences were combined as one. However, a possible true enrichment for each sequence was thereby also removed.

Therefore, whether to use of $Z\alpha_{ADAR1}$mut as a negative control should be considered carefully for each case. More than one comparison may be applied to each analysis.

IV.2.4 $Z\alpha_{ADAR1}$ ChAP sequences are enriched near centromere regions

When the general distribution of possible Z-DNA segments was mapped to the human genome, an enrichment of $Z\alpha_{ADAR1}$ ChAP sequences near the centromere regions was observed. This finding corroborates our previous results reported for A549 lung cancer cells (Li, Xiao et al. 2009)

It is interesting to note in this context that the centromere region contains apparently nucleosomes where the DNA winds about the histone octamer in a right-handed, rather than left-handed orientation (Furuyama and Henikoff 2009). These nucleosomes are assembled with histone CENPH3, which can only be found in active centromere regions. Thus, formation of nucleosomes with CENPH3 could lead the local nearby linker DNA to become substantially negatively supercoiled, which in turn, can induce Z-DNA formation.

It is reported that B-Z transitions can be observed in single DNA molecule studies at very low (-) DNA superhelical densities when low tension is applied to DNA (Lee, Kim et al. 2010). Obviously, tension will be generated in chromosomal DNA at or near centromeres during mitosis, when sister-chromatids are being pulled apart by the spindles and separated to each end of the dividing cell.

A possible model could be described like this: during metaphase, when chromosomes were dragged by the spindles to each end of the cell during mitosis, DNA tension will be

generated, which should be largest at or near centromere regions. This tension, combined with local negative superhelical torsion that is generated by dissembling of the canonical H3 nucleosomes or by formation of the alternate CENPH3 nucleosome, will facilitate the B-to-Z transition at nearby regions. huES cultures, which are highly proliferative, will have a substantial number of cells at any time undergoing mitosis. Thus, such a model could very well explain our observed ChAP results.

It may be questioned why the observed enrichment near centromere regions does not apply for all chromosomes. As we known, the centromere regions contain a high degree of repeated sequences within the human genome. The lengths and sequences were hardly defined. Thus, a general length of 3 million bp is assigned to each chromosome in the UCSC HG19 database. It is possible that the Z-DNA forming feature in centromere regions is applied to all the chromosomes. However, as different chromosomes have different level of mapping-ability in the database; the enrichment at/near centromere regions could only observed in the well sequenced chromosomes. For less characterized centromeric regions, the ChAP sequences were hidden in the undefined region in the centromere, and cannot be detected. The high similarity of the enriched positions from the A549 result and the huES results is a good support for this theory. As the similar BLAST template is used, the availability of the sequence information for each centromere region is similar for both analyses.

IV.2.5 Enrichment of $Z\alpha_{ADAR1}$ binding sites with SNP near centromere regions

$Z\alpha_{ADAR1}$ binding sites were shown to be enriched with SNPs. The enrichment of SNP is much stronger at/near centromere regions than in arm regions. The relationship between Z-conformation and genetic instability is well established (Wang and Vasquez 2006; Zhao, Bacolla et al. 2010). The enrichment of SNP is very obvious only at the centromere region. It is not measurable if the entire chromosome is scanned. Considering the Z-DNA segments were more enriched in the centromere region compared to the arm region, we can explain the finding in the following way: the SNP or genetic instability in centromere regions is mainly due to the Z-DNA conformation, while SNP at the arm regions is mainly caused by other DNA structures and/or different mechanisms. Mechanistically, this could be explained by the recent crystal structure of a B-to-Z junction in which the two bases of one base pair are extruded from the helix and thus become prone to genetic modifications. (Ha, Lowenhaupt et al. 2005)

Therefore, the higher density of SNP in the genome at the centromere region could be due to the fact that Z-DNA is easier formed in that region than elsewhere in the genome.

IV.2.6 Enrichment of $Z\alpha_{ADAR1}$ binding sites in chromosomal G-Bands

$Z\alpha_{ADAR1}$ binding sites are enriched in the Giemsa intense staining bands (G-bands). The G-bands are reported to represent highly condensed DNA regions (Burkholder and Weaver 1977) and are related to the heterochromatin (Comings 1978). This finding correlated with immunostaining of *Chironomus* chromosomes with Z-DNA antibodies, showing that the Z-DNA antibodies are enriched in heterochromatin (Arndt-Jovin, Robert-Nicoud et al. 1985). The immunostaining result could be explained by the fact

that heterochromatin itself is DNA-enriched. The enrichment of the Z-DNA antibodies might thus result from more DNA sources to be bound. Here, our result is normalized to the actual DNA length (bp), and the enrichment was still substantial. Therefore, the enrichment in the G-band is not only due to the high DNA content. The immunostaining results were also questioned because the acidic treatment itself would trigger formation of different DNA structures. Here, no additional treatment is performed in our studies except the fixation by formaldehyde, which is more reliable.

The G-band pale region are reported to be the GC rich region on the chromosome, and GT repetitive sequences are reported to be much more frequent in euchromatin than in heterochromatin (Stallings, Ford et al. 1991). However, Z-DNA is reported to be more favored to form in the GC rich sequences and in GT repeats. The enrichment of the $Z\alpha_{ADAR1}$ on the G-band intense region seems to be contradictory to the predictions based on sequence content alone. This, in a way, proved that Z-DNA formation in the living cell is not just depended on the presence of particular sequences. Besides histone proteins, several other non-histone proteins are involved in formation and condensation of G-bands (Vogel, Faust et al. 1974). DNA in G-band intense region might have a different physical environment and, perhaps even more important, topology that might trigger Z-DNA formation. Or, in another word, Z-DNA formation may contribute to the DNA condensation in G-bands.

IV.2.7 Enrichment of Z-segments at transcription start site (TSS) but not in CpG islands

Accumulation of negative supercoiling in the wake of RNA polymerase during transcription is well demonstrated in theories and experiments in plasmids *in vivo* (Droge 1994)(Rich and Zhang 2003). An enrichment of Z-DNA forming sequences at TSS was predicted by *in silico* computational analyses (Li, Xiao et al. 2009). In my study, I observed for the first time an enrichment of Z-DNA forming region at TSS within the surrounding ±400bp, this trend could not be expanded to the nearby ±2000bp regions.

CpG islands are high GC content regions with high frequency of GC di-nucleotides which usually appear at the 5' regions of house-keeping genes (Gardiner-Garden and Frommer 1987). Theoretically, formation of Z-DNA should be related to CpG islands. However, there is no enrichment of $Z\alpha_{ADAR1}$ binding sites in the CpG islands. One explanation could be that although the sequences in CpG islands favor Z-DNA formation, the formation of Z-DNA at TSS is not that dependent on sequence content but perhaps on the level of transcription and the resulting transient wave of negative supercoiling.

IV.3 Future works on mapping of Z-DNA on human genome

The $Z\alpha_{ADAR1}$ binding sites were used to plot out the Z-DNA map on human genome. Several predictions on Z-DNA function were drawn out from the map. However, as the bulk culture was used in this study, which is not a good source when individual properties are examined. In order to further prove the observations in this study, different systems might be set up for each analysis, such as synchronizing the cells in the mitosis phase to analyze the enrichment of Z-DNA in centromere regions, and halt transcription in transcription starting state to analyze the Z-DNA formation at the TSS region.

In this study, as lack of ChAP material, a PCR amplification step was performed. The samples from $Z\alpha_{ADAR1}$ and $Z\alpha_{ADAR1}$mut had been subjected to different PCR cycles. Therefore, the actual copy number of each sequence cannot be considered as the actual occurrences in the ChAP experiment and was neglected. In future studies, the sample size could be increased to eliminate PCR steps. Then, more informative libraries could be built from sequencing results.

V SUMMARY

V.1 Binding of $Z\alpha_{ADAR1}$ to ribosomes and its potential biological significances

This study showed that the Z-conformation-specific binding probe $Z\alpha_{ADAR1}$ binds ribosomes and inhibits translation in both *E. coli* and mammalian systems, most likely in a secondary structure-dependent way. Potential $Z\alpha_{ADAR1}$ binding sites on ribosomes were identified by pull-down experiments. Several binding sites were conserved between *E. coli* and mammalian ribosome, revealing that formation of Z- or Z-like conformations might be a conserved property of the dynamic ribosome structure during translation. A similar inhibitory effect on translation was also observed with full-length ADAR1, which is naturally only expressed after INF-α induction. This suggested that the inhibitory effect on translation by Z-conformation-specific binding might be related to the innate immune response, and facilitates the host defense upon viral infection.

V.2 Mapping of Z-DNA in human embryonic stem cells

The Z-DNA was mapped to the human genome, using $Z\alpha_{ADAR1}$ as a probe and huES cells as substrates. The potential functions of Z-DNA were illustrated. The enrichment of Z-DNA in TSS and correlation of Z-DNA with SNP densities was observed, suggesting that formation of Z-DNA is indeed to some extent related to transcription processes and contributes to DNA instabilities. Enrichment of $Z\alpha_{ADAR1}$ binding sites in GC-enriched G-bands and non-correlation with CpG islands are contradictory to previous *in silico*

predictions based on the distribution of Z-DNA forming sequences. This, in a way, reveals that formation of Z-DNA in living cells is not purely depended on DNA sequences; instead the local physical and/or chemical environment also plays an important role. Other findings such as enrichment of $Z\alpha_{ADAR1}$ binding sites in G-bands suggested that Z-DNA may play some roles in DNA condensation of these regions of chromatin. The enrichment of $Z\alpha_{ADAR1}$ binding sites at or near centromere regions and the enrichment of SNP in these $Z\alpha_{ADAR1}$ binding sites suggest that formation of Z-DNA could mechanistically be a major driving force for the evolution of centromere regions.

Supplementary Data

Supplementary Table 1

$Z\alpha_{ADAR1}$					
Chromosome	gneg	gpos25	gpos50	gpos75	gpos100
chr1	0.090221	0.087764	0.098123	0.088426	0.117632
chr2	0.098337	0.095	0.110746	0.101985	0.118826
chr3	0.105201	0.092715	0.103922	0.09562	0.113659
chr4	0.094163	0.095975	0.094196	0.107718	0.106813
chr5	0.106547	0.120741	0.083772	0.110156	0.110323
chr6	0.092376	0.106667	0.094737	0.105394	0.103409
chr7	0.08116	0.075926	0.1	0.097849	0.118182
chr8	0.08768	0.1	0.09625	0.103279	0.104405
chr9	0.080858	0.068952	0.062937	0.096	0.105455
chr10	0.099094	0.080208	0.075843	0.088043	0.117803
chr11	0.087286	0.071429	0.085774	0.076852	0.095393
chr12	0.096579	0.070588	0.108333	0.10578	0.11049
chr13	0.098771	0.09	0.113889	0.08	0.104467
chr14	0.085715	0.080952	0.116547	0.095	0.106593
chr15	0.080039	0.10084	0.101351	0.117816	
chr16	0.071723	0.0375	0.093119	0.078431	0.115054
chr17	0.091617	0.090909	0.09009	0.091192	
chr18	0.095882	0.141176	0.117143	0.109244	0.109302
chr19	0.047059	0.077551			
chr20	0.088778	0.081579	0.105263	0.115104	
chr21	0.087794		0.106122	0.142553	0.115789
chr22	0.073918	0.076923	0.091729		
chrX	0.043575	0.045122	0.055191	0.052381	0.052533
chrY	0.026554		0.030108		

$Z\alpha_{ADAR1}mut$					
Chromosome	gneg	gpos25	gpos50	gpos75	gpos100
chr1	0.031363	0.034597	0.037531	0.034721	0.03073
chr2	0.030121	0.041246	0.034625	0.024565	0.037407
chr3	0.029006	0.030461	0.033985	0.039657	0.03317
chr4	0.032423	0.030759	0.03482	0.032885	0.037712
chr5	0.02775	0.025184	0.038594	0.032811	0.030106
chr6	0.033706	0.017776	0.034735	0.040662	0.040624
chr7	0.026096	0.014813	0.034284	0.036558	0.031817
chr8	0.026559	0.014999	0.040831	0.033605	0.035241
chr9	0.024729	0.025804	0.02867	0.038998	0.032726
chr10	0.040821	0.016665	0.024156	0.024999	0.041287
chr11	0.031392	0.014284	0.046023	0.021295	0.03008
chr12	0.027166	0.047053	0.036456	0.03468	0.036362
chr13	0.031053	0.017997	0.041663	0.027499	0.034363
chr14	0.03243	0.028569	0.040285	0.024999	0.03553
chr15	0.018816	0.025208	0.02162	0.028735	
chr16	0.026147	0.024997	0.030273	0.043136	0.040859
chr17	0.027819	0.021816	0.018017	0.022797	
chr18	0.036233	0.044115	0.039998	0.02857	0.034883
chr19	0.023075	0.023548			
chr20	0.029482	0.043418	0.052626	0.022916	
chr21	0.034274		0.028569	0.025531	0.042104
chr22	0.040945	0.015382	0.030073		
chrX	0.017043	0.007317	0.0153	0.014814	0.015466
chrY	0.007344		0.00215		

Supplementary Table 1 Distribution of $Z\alpha_{ADAR1}$(mut) ChAP result on G-band on each chromosome

The G-band was grouped into 5 groups according to the G-band intensity gneg which refer to G-negative is the most pale band up on Giesma staining; while gpos which refer to G-positive is grouped into 4 intensity level. The total length of each group of band on each chromosome was added together, the ratio of Z-site or M-site per total length was calculated accordingly

REFERENCES

Arndt-Jovin, D. J., M. Robert-Nicoud, et al. (1985). "Immunofluorescence localization of Z-DNA in chromosomes: quantitation by scanning microphotometry and computer-assisted image analysis." J Cell Biol **101**(4): 1422-1433.

Athanasiadis, A., D. Placido, et al. (2005). "The crystal structure of the Zbeta domain of the RNA-editing enzyme ADAR1 reveals distinct conserved surfaces among Z-domains." J Mol Biol **351**(3): 496-507.

Bae, S., D. Kim, et al. (2010). "Intrinsic Z-DNA Is Stabilized by the Conformational Selection Mechanism of Z-DNA-Binding Proteins." J Am Chem Soc.

Boehm, T., L. Mengle-Gaw, et al. (1989). "Alternating purine-pyrimidine tracts may promote chromosomal translocations seen in a variety of human lymphoid tumours." Embo J **8**(9): 2621-2631.

Brown, B. A., 2nd, K. Lowenhaupt, et al. (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): 13532-13536.

Burkholder, G. D. and M. G. Weaver (1977). "DNA-protein interactions and chromosome banding." Exp Cell Res **110**(2): 251-262.

Chandramouli, P., M. Topf, et al. (2008). "Structure of the mammalian 80S ribosome at 8.7 A resolution." Structure **16**(4): 535-548.

Comings, D. E. (1978). "Mechanisms of chromosome banding and implications for chromosome structure." Annu Rev Genet **12**: 25-46.

Deigendesch, N., F. Koch-Nolte, et al. (2006). "ZBP1 subcellular localization and association with stress granules is controlled by its Z-DNA binding domains." Nucleic Acids Res **34**(18): 5007-5020.

Droge, P. (1994). "Protein tracking-induced supercoiling of DNA: a tool to regulate DNA transactions in vivo?" Bioessays **16**(2): 91-99.

Feigon, J., A. H. Wang, et al. (1985). "Z-DNA forms without an alternating purine-pyrimidine sequence in solution." Science **230**(4721): 82-84.

Feng, S., H. Li, et al. (2011). "Alternate rRNA secondary structures as regulators of translation." Nat Struct Mol Biol **advance online publication**.

Furuyama, T. and S. Henikoff (2009). "Centromeric nucleosomes induce positive DNA supercoils." Cell **138**(1): 104-113.

Gardiner-Garden, M. and M. Frommer (1987). "CpG islands in vertebrate genomes." J Mol Biol **196**(2): 261-282.

Garner, M. M. and G. Felsenfeld (1987). "Effect of Z-DNA on nucleosome placement." J Mol Biol **196**(3): 581-590.

Ha, S. C., J. Choi, et al. (2009). "The structures of non-CG-repeat Z-DNAs co-crystallized with the Z-DNA-binding domain, hZ alpha(ADAR1)." Nucleic Acids Res **37**(2): 629-637.

Ha, S. C., D. Kim, et al. (2008). "The crystal structure of the second Z-DNA binding domain of human DAI (ZBP1) in complex with Z-DNA reveals an unusual binding mode to Z-DNA." Proc Natl Acad Sci U S A **105**(52): 20671-20676.

Ha, S. C., K. Lowenhaupt, et al. (2005). "Crystal structure of a junction between B-DNA and Z-DNA reveals two extruded bases." Nature **437**(7062): 1183-1186.

Herbert, A., J. Alfken, et al. (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): 8421-8426.

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): 12132-12137.

Herbert, A., M. Schade, et al. (1998). "The Zalpha domain from human ADAR1 binds to the Z-DNA conformer of many different sequences." Nucleic Acids Res **26**(15): 3486-3493.

Herbert, A. G., J. R. Spitzner, et al. (1993). "Z-DNA binding protein from chicken blood nuclei." Proc Natl Acad Sci U S A **90**(8): 3339-3342.

Irikura, K. K., B. Tidor, et al. (1985). "Transition from B to Z DNA: contribution of internal fluctuations to the configurational entropy difference." Science **229**(4713): 571-572.

Kang, Y. M., J. Bang, et al. (2009). "NMR spectroscopic elucidation of the B-Z transition of a DNA double helix induced by the Z alpha domain of human ADAR1." J Am Chem Soc **131**(32): 11485-11491.

Kha, D. T., G. Wang, et al. (2010). "Pathways for double-strand break repair in genetically unstable Z-DNA-forming sequences." J Mol Biol **398**(4): 471-480.

Kim, Y. G., K. Lowenhaupt, et al. (2004). "Evidence that vaccinia virulence factor E3L binds to Z-DNA in vivo: Implications for development of a therapy for poxvirus infection." Proc Natl Acad Sci U S A **101**(6): 1514-1518.

Koeris, M., L. Funke, et al. (2005). "Modulation of ADAR1 editing activity by Z-RNA in vitro." Nucleic Acids Res **33**(16): 5362-5370.

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): 12759-12764.

Lafer, E. M., R. Sousa, et al. (1985). "Isolation and characterization of Z-DNA binding proteins from wheat germ." Biochemistry **24**(19): 5070-5076.

Lafer, E. M., R. P. Valle, et al. (1983). "Z-DNA-specific antibodies in human systemic lupus erythematosus." J Clin Invest **71**(2): 314-321.

Lagravere, C., B. Malfoy, et al. (1984). "Ring-opened alkylated guanine is not repaired in Z-DNA." Nature **310**(5980): 798-800.

Lee, E. H., Y. J. Seo, et al. (2010). "NMR study of hydrogen exchange during the B-Z transition of a DNA duplex induced by the Zalpha domains of yatapoxvirus E3L." FEBS Lett **584**(21): 4453-4457.

Lee, J., Y. G. Kim, et al. (2010). "Transition between B-DNA and Z-DNA: free energy landscape for the B-Z junction propagation." J Phys Chem B **114**(30): 9872-9881.

Lee, M., S. H. Kim, et al. (2010). "Minute negative superhelicity is sufficient to induce the B-Z transition in the presence of low tension." Proc Natl Acad Sci U S A **107**(11): 4985-4990.

Li, H., J. Xiao, et al. (2009). "Human genomic Z-DNA segments probed by the Z alpha domain of ADAR1." Nucleic Acids Res **37**(8): 2737-2746.

Liu, H., N. Mulholland, et al. (2006). "Cooperative activity of BRG1 and Z-DNA formation in chromatin remodeling." Mol Cell Biol **26**(7): 2550-2559.

Liu, L. F. and J. C. Wang (1987). "Supercoiling of the DNA template during transcription." Proc Natl Acad Sci U S A **84**(20): 7024-7027.

Liu, Y., K. C. Wolff, et al. (2001). "Vaccinia virus E3L interferon resistance protein inhibits the interferon-induced adenosine deaminase A-to-I editing activity." Virology **289**(2): 378-387.

Maas, S., T. Melcher, et al. (1997). "Mammalian RNA-dependent deaminases and edited mRNAs." Curr Opin Cell Biol **9**(3): 343-349.

Majewski, J. and J. Ott (2000). "GT repeats are associated with recombination on human chromosome 22." Genome Res **10**(8): 1108-1114.

Moller, A., A. Nordheim, et al. (1984). "Bromination stabilizes poly(dG-dC) in the Z-DNA form under low-salt conditions." Biochemistry **23**(1): 54-62.

Murphy, K. E. and J. R. Stringer (1986). "RecA independent recombination of poly[d(GT)-d(CA)] in pBR322." Nucleic Acids Res **14**(18): 7325-7340.

Nishikura, K. (2006). "Editor meets silencer: crosstalk between RNA editing and RNA interference." Nat Rev Mol Cell Biol **7**(12): 919-931.

Nordheim, A., P. Tesser, et al. (1982). "Isolation of Drosophila proteins that bind selectively to left-handed Z-DNA." Proc Natl Acad Sci U S A **79**(24): 7729-7733.

Placido, D., B. A. Brown, 2nd, et al. (2007). "A left-handed RNA double helix bound by the Z alpha domain of the RNA-editing enzyme ADAR1." Structure **15**(4): 395-404.

Placido, D., B. A. Brown, 2nd, et al. (2007). "A left-handed RNA double helix bound by the Zalpha domain of the RNA-editing enzyme ADAR1." Structure **15**(4): 395-404.

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**: 113-117.

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): 375-396.

Popenda, M., J. Milecki, et al. (2004). "High salt solution structure of a left-handed RNA double helix." Nucleic Acids Res **32**(13): 4044-4054.

Ratje, A. H., J. Loerke, et al. (2010). "Head swivel on the ribosome facilitates translocation by means of intra-subunit tRNA hybrid sites." Nature **468**(7324): 713-716.

Rich, A. and S. Zhang (2003). "Timeline: Z-DNA: the long road to biological function." Nat Rev Genet **4**(7): 566-572.

Richmond, T. J. and C. A. Davey (2003). "The structure of DNA in the nucleosome core." Nature **423**(6936): 145-150.

Rothenburg, S., N. Deigendesch, et al. (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): 1602-1607.

Rothenburg, S., F. Koch-Nolte, et al. (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): 8985-8990.

Rothenburg, S., T. Schwartz, et al. (2002). "Complex regulation of the human gene for the Z-DNA binding protein DLM-1." Nucleic Acids Res **30**(4): 993-1000.

Schade, M., C. J. Turner, 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): 12465-12470.

Schwartz, T., J. Behlke, et al. (2001). "Structure of the DLM-1-Z-DNA complex reveals a conserved family of Z-DNA-binding proteins." Nat Struct Biol **8**(9): 761-765.

Schwartz, T., M. A. Rould, et al. (1999). "Crystal structure of the Zalpha domain of the human editing enzyme ADAR1 bound to left-handed Z-DNA." Science **284**(5421): 1841-1845.

Shachrai, I., A. Zaslaver, et al. (2010). "Cost of unneeded proteins in E. coli is reduced after several generations in exponential growth." Mol Cell **38**(5): 758-767.

Sharif, J., T. A. Endo, et al. (2010). "Divergence of CpG island promoters: a consequence or cause of evolution?" Dev Growth Differ **52**(6): 545-554.

Stallings, R. L., A. F. Ford, et al. (1991). "Evolution and distribution of (GT)n repetitive sequences in mammalian genomes." Genomics **10**(3): 807-815.

Tartier, L., V. Michalik, et al. (1994). "Radiolytic signature of Z-DNA." Nucleic Acids Res **22**(25): 5565-5570.

Toth, A. M., Z. Li, et al. (2009). "RNA-specific adenosine deaminase ADAR1 suppresses measles virus-induced apoptosis and activation of protein kinase PKR." J Biol Chem **284**(43): 29350-29356.

Toth, A. M., P. Zhang, et al. (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.

Vitali, P. and A. D. Scadden (2010). "Double-stranded RNAs containing multiple IU pairs are sufficient to suppress interferon induction and apoptosis." Nat Struct Mol Biol **17**(9): 1043-1050.

Vogel, W., J. Faust, et al. (1974). "Othe relevance of non-histone proteins to the production of Giemsa banding patterns on chromosomes." Humangenetik **21**(3): 227-236.

Wahls, W. P., L. J. Wallace, et al. (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): 785-793.

Wang, A. H., G. J. Quigley, et al. (1979). "Molecular structure of a left-handed double helical DNA fragment at atomic resolution." Nature **282**(5740): 680-686.

Wang, Z. and P. Droge (1996). "Differential control of transcription-induced and overall DNA supercoiling by eukaryotic topoisomerases in vitro." Embo J **15**(3): 581-589.

Wittig, B., T. Dorbic, et al. (1991). "Transcription is associated with Z-DNA formation in metabolically active permeabilized mammalian cell nuclei." Proc Natl Acad Sci U S A **88**(6): 2259-2263.

Wittig, B., S. Wolfl, et al. (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): 4653-4663.

Wolfl, S., C. Martinez, et al. (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): 3664-3668.

Wong, B., S. Chen, et al. (2007). "Characterization of Z-DNA as a nucleosome-boundary element in yeast Saccharomyces cerevisiae." Proc Natl Acad Sci U S A **104**(7): 2229-2234.

Zarling, D. A., C. J. Calhoun, et al. (1990). "Cytoplasmic microinjection of immunoglobulin Gs recognizing RNA helices inhibits human cell growth." J Mol Biol **211**(1): 147-160.

Zarling, D. A., C. J. Calhoun, et al. (1987). "Cytoplasmic Z-RNA." Proc Natl Acad Sci U S A **84**(17): 6117-6121.

PUBLICATIONS

Feng, S., H. Li, et al. (2011) "Alternate rRNA secondary structures as regulators of translation." Nat. Struct. Mol. Biol. doi: 10.1038/nsmb.1962.

Li, H., J. Xiao, et al. (2009). "Human genomic Z-DNA segments probed by the Z alpha domain of ADAR1." Nucleic Acids Res **37**(8): 2737-2746.