Development of Artificial Enzymes System for RNA Editing Towards Genetic Restoration and Gene Therapy

March, 2018

Md Thoufic Anam Azad

Japan Advanced Institute of Science and Technology

Doctoral Dissertation

Development of Artificial Enzymes System for RNA Editing Towards Genetic Restoration and Gene Therapy

by

Md Thoufic Anam Azad

1540001

Submitted to

Japan Advanced Institute of Science and Technology

In partial fulfilment of the requirements

For the degree of Doctor of Philosophy

Supervisor: Professor Dr. Toshifumi Tsukahara

School of Materials Science

Japan Advanced Institute of Science and Technology

March, 2018

Doctoral Dissertation

Development of Artificial Enzymes System for RNA Editing Towards Genetic Restoration and Gene Therapy

Supervisor	Professor Dr. Toshifumi Tsukahara
Second Supervisor	Associate Professor Dr. Hidekazu Tsutsui
Advisor for Minor Research	Professor Dr. Takahiro Hohsaka

Judging Committee

Head of Committee	Professor Dr. Toshifumi Tsukahara
Internal Member	Professor Dr. Takahiro Hohsaka
Internal Member	Professor Dr. Kenzo Fujimoto
Internal Member	Associate Professor Dr. Hidekazu Tsutsui
External Member	Associate Professor Dr. Shin-ichi Horike

Md Thoufic Anam Azad

1540001

Japan Advanced Institute of Science and Technology

March, 2018

Abstract

Adenosine deaminase acting on RNA (ADAR) family enzymes consist of double-stranded RNA binding domains (dsRBDs) and a deaminase domain (DD) that convert adenosine (A) into inosine (I) acting as guanosine (G) during translation. Site-directed RNA editing is an important technique for correcting gene sequences and ultimately tuning protein function.

In this study, I engineered the DD of adenosine deaminase acting on RNA (ADAR1) and the MS2 system to target specific adenosines, with the goal of correcting G-to-A mutations at the RNA level. For this purpose, the ADAR1-DD was fused downstream of the RNA-binding protein MS2, which has an affinity for the MS2 RNA. I checked the binding affinity of artificial enzyme system by Biacore ™ X100. To direct editing to specific targets, I designed guide RNAs complementary to target RNAs. The guide RNAs directed the ADAR1 deaminase to the desired editing site, where it converted adenosine to inosine. To provide proof of principle, I used an allele of EGFP bearing a mutation at the 58th amino acid (TGG), encoding Trp, into an amber (TAG) or ochre (TAA) stop codon. In HEK-293 cells, this system could convert stop codons to read-through codons, thereby turning on fluorescence. I confirmed the specificity of editing at the DNA level by restriction fragment length polymorphism (RFLP) analysis and sequencing, and at the protein level by western blotting. The editing efficiency of this enzyme system was ~5%.

Further, I tried to compare the deaminase activity of ADAR1-DD and different isoforms of ADAR2-DD. The guide sequence was fused with MS2 stem-loop. As a target, mutated amber (TAG) stop codon at 58 amino acid (TGG) of EGFP was used. After transfection of the above three factors in HEK 293 cells, varying degree of the fluorescence signal was observed. Regarding ADAR2 isoforms, 120 bp consisting the Alu-cassette present in the middle of the DD. ADAR2-long without Alu-cassette showed much higher fluorescence signal than the ADAR2-long with Alu-cassette. According to the I-TASSER (Iterative Threading Assembly Refinement) data, inserted Alu-cassette result longer coil in the middle of the deaminase domain. Due to insertion of the Alu-cassette, the distance between residues after 203 is increased dramatically. The ligand binding site i.e. nucleic acid binding capacity also largely differ due to the insertion of the Alu-cassette. Another isoform ADAR2-short which is approximately 81 bp shorter at C-terminal, the fluorescence signal was undetectable. A single amino acid substitution of ADAR2-long (E488Q) renders the enzyme more active than the wild types. The fluorescence microscopic image and fluorescence spectra analysis are suggesting that ADAR1-DD is more active than ADAR2-long-DD. In the result of Western blot and sequencing, I found that ADAR1-DD is the most active deaminase than any other DDs. To my knowledge, this is a complete biological approach for the comparative study of ADARs-DD that gives important information on the rational use of DD in the future application for therapeutic purposes.

Regarding guide length, 21 bp guide found more functional and regarding the position, upstream guide is more efficient. The editing efficiency can be increased approximately 16%. Double mutated ochre (TAA) stop codon can be converted to (TGG) with 5' adenosine preference. Off-target editing increase with the increase of efficiency to the targeted adenosine. I believe that this system could be used to treat genetic diseases resulting from G-to-A point mutations.

Keywords- ADARs, MS2 RNA, RNA editing, stop codon, genetic diseases.

CONTENTS

1.1 RNA editing

RNA editing was first defined as the modification of RNA ultimately changing its coding capacity.[1] The term "RNA editing" was first invented in 1986 when it was found in *Trypanosoma* mitochondrial cytochrome oxidase (cox) subunit II gene, four nucleotides were added in the RNA which was not included in the genomic DNA.[2] The report on mammalian RNA editing was reported in 1987.[3] From that time, this term has been used for insertion[2], deletion or substitution[4,5] of nucleotides. But recent scientific advancement and application of next-generation sequencing technology have changed the understanding and effect regarding RNA editing. In higher eukaryotes insertion and deletion has not been reported yet but the most widespread phenomenon is base conversion.[1]

RNA editing is one of the most important epigenetic processes by which post-transcriptional modification of gene occurs. This phenomenon results in diversity of transcript by insertion, deletion or conversion of nucleotides ultimately leading to protein diversity.[6] Previously it was assumed that one gene is responsible for one protein but by RNA editing along with another mechanism like splicing, different functional proteins can be produced from the same gene. Therefore, RNA editing is very important for differentiation, growth and development in life.

1.2 Types of RNA editing

There are several types of RNA editing events in living organisms. Among these A-to-I, C-to-U and U-to-C are the common types. A-to-I and C-to-U RNA editing are generally caused by the ADARs and APOBEC-AID deaminase family respectively. However, the enzymes responsible for U-to-C editing does not discover yet although it is the abundant phenomenon in lower plant species and rare in animals. Different types of RNA editing are described below-

1.2.1 A-to-I RNA editing

The widespread RNA editing in the vertebrate is A-to-I RNA editing[7,8] where the single adenosine is converted to inosine that works as guanosine during translation. One of the important epigenetic mechanism is RNA editing that caused by ADARs resulting A-to-I/G base-modification.[9] In this case the target site forms dsRNA with complementary sequence (ECS) (Figure 1). Only a single nucleotide change produces a wide variety of transcriptome. A-to-I RNA editing is the most prevalent type of RNA editing in mammals caused by ADARs[7,1] in the C6 position of adenosine.[10]

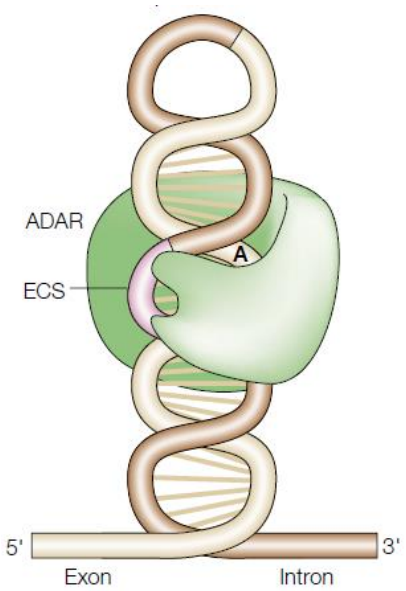


Figure 1. Adenosine deaminases acting on RNA (ADARs) recognize duplex RNA that is formed between the editing site and the editing site complementary sequence (ECS) which is often located in a downstream intron. The enzyme binds to the double-stranded (ds) RNA through their dsRBDs and deaminates a specific adenosine to inosine.[1]

The frequency of editing events is higher in the central nervous system (CNS)[11–16] but reported in almost in every tissues.[10,17,18] During development of brain, ADARs play an important role. ADAR1 knockout mouse was found embryonic death[19] whereas ADAR2 knockout mouse showed epilepsy leading to death.[20] However, when ADARs activity abolished in flies there is neuromuscular disorders and degeneration of nerve tissues observed.[18] In the mammalian brain,

different proteins need to be finely modulated by the ADARs which are required for different electrochemical functions.[20–22] Glutamate mRNA editing is a continuous process in the development of the brain. Approximately 100% editing occurs in the adult brain.[23] Any hindrance in the RNA editing of the GluR2 receptor by the ADAR2 result progressive epilepsy leading to death within a few weeks in mice. Q-to-R RNA editing of subunit 2 of AMP GluR2 results impossible for Ca^{2+} impermiability of the receptor containing glutamate subunit.[23] G protein coupling receptor HTR2C is the only one that undergoes editing in five sites (A, B, E, C, D) which are close to proximity in the second intracellular loop of the receptor. Serotonin signaling is modulated by the RNA editing; RNA editing of the receptor decrease the proficiency of receptor within the G proteins.[12] ADAR1 and ADAR2 found enzymatically active however, ADAR3 activity not reported yet although its deaminase domain characteristic is conserved.

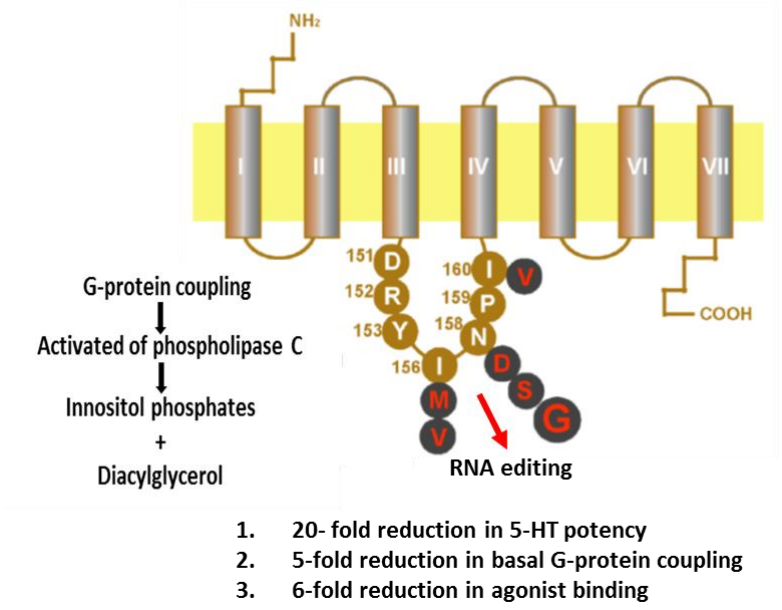


Figure 2. RNA editing in seven transmembrane 5-HT2CR (gray cylinders) shown with the lipid membrane (yellow color). Seven amino acids in loop II (white characters) are presented together with the conversion (red characters). Unedited isoform indicated by INI whereas fully edited isoform indicated by VGV with greatly reduced sensitivity of the receptor. Highest reduction of the sensitivity of the receptor observed when 158 position asparagine (N) is converted to glycine (G) therefore, it is shown in larger font size.[24]

1.2.2 C-to-U RNA editing

Tissue-specific apolipoprotein B (apoB) was first identified RNA editing in mammals.[3,25] Apolipoprotein B can produce two forms of protein–one full-length apoB-100 produced in liver another isoform apoB-48 produced in the small intestine.

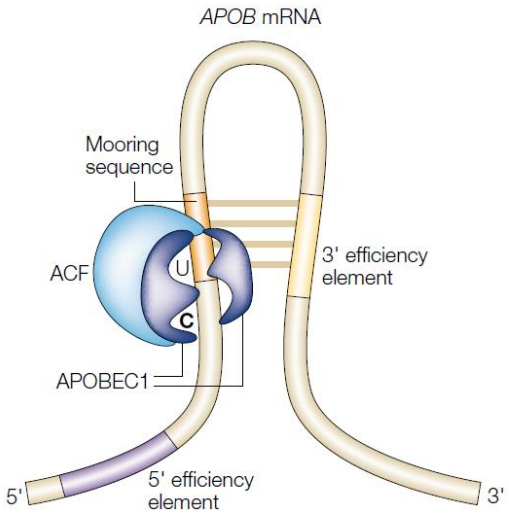


Figure 3. APOBEC1 binds near to the Mooring sequence of APOB mRNA with the help of the associated factor (ACF) and deaminate C-to-U.[1]

Both isoforms are responsible for lipid metabolism and these are produced from the same transcript. The shorter isoform is produced when the 6666 cytidine of glutamine codon (CAA) is converted to uridine stop codon TAA. This conversion and generation of shorter apoB48 (241 kDa) have also required for dietary lipid absorption.[3,25] The longer form (241 kDa) synthesized in liver required for carrying for endogenously synthesized triglyceride and cholesterol in blood circulation. Mooring sequence is important for site selection in C-to-U RNA editing (Figure 3).

1.2.3 U-to-C RNA editing

A-to-I RNA editing in dsRNA and C-to-U in DNA or ssRNA are the common phenomenon in higher eukaryotes whereas U-to-C conversion is very rare [26,27,28] which only reported in the transcript encoding the Wilms' tumour susceptibility gene in mammals.[28] In arthropods, U-to-C RNA editing found in generating tissue-specific calcium channel which is affected by the RNA editing activity.[29] However, U-to-C by amination is a common phenomenon in plants sometimes it may exceed C-to-U events in some cases.[30] In mammals, RNA U-to-C editing reported [27] in the mitochondrial transcript. RNA editing also affect alternative splicing,[31] micro RNA processing[32] and other many important cellular events.[1,33,34,35]

1.3 Types of ADARs and their characteristics

ADARs are modular protein with different domain for individual function (Figure 4). The dsRNA unwinding activity used to the questionable discovery of ADARs in *Xenopous levis* egg and embryos.[18,20] Later on, it was revealed that such activity mainly from the dsRNA specific ADAR.[19,36] However, ADAR1 is the first discovered mammalian gene among the ADARs.[21,22] Soon after its discovery ADAR1 was cloned and characterized in protein level by sequencing. This research leads to the discovery of ADAR2. A-to-I RNA editing causing deaminases genes are three in numbers ADAR1, ADAR2 and ADAR3.[37] Expression of ADAR1 and ADAR2 occurs in many tissues whereas ADAR3 only in the brain. ADAR1 expresses in 2 forms as a result of alternative splicing. One full-length ADAR1 with 150-kDa that transcribed by interferon inducible promoter mainly localize in the cytoplasm. Another one 110-kDa (p110) which transcribed from constitutive promoter started from downstream start codon primarily localize in the nucleus.[38]

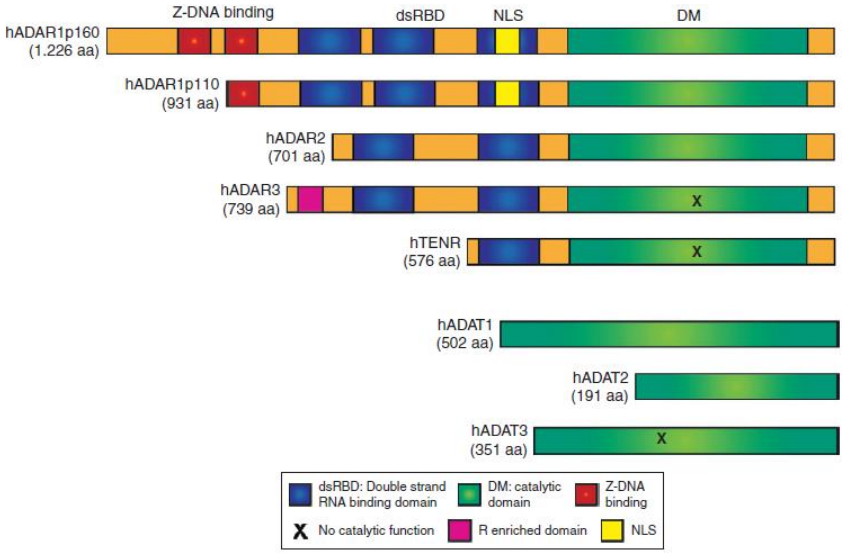


Figure 4. Protein domains of ADARs consist of dsRNA (blue) which bind dsRNA and the catalytic deaminase domain (DM; green box). ADAT1, ADAT2, and ADAT3 have a DM domain which they also use to bind tRNA. ADAR3, TENR, and ADAT3 are inactive and that is represented by an (X) in the DM domain. ADAR1 has Z-DNA binding at the N-terminus domains and ADAR3 contains an R-enriched region at the N-terminus.[39]

1.4 Mechanism of RNA editing

For a better understanding of the mechanism of RNA editing by ADAR2, NMR data has been revealed with the substrate[40] and without substrate.[41] The target is recognized by the dsRBD and the deaminase activity mainly performed by the DD of the enzymes (Figure 5. A and B).

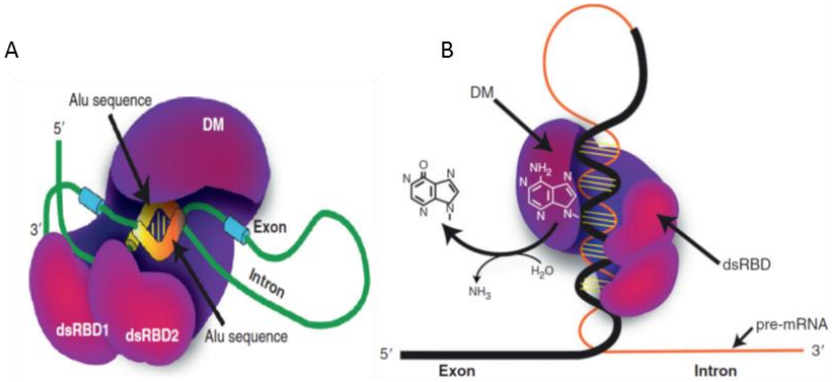


Figure 5. **A**. Transcripts encoding Alu repeat elements are often present in introns of transcripts. If two are present in opposite orientation less than 2 kb apart they can base pair with the other and form a duplex structure that is edited by ADARs.[39] **B**. The dsRNA-binding domains of ADARs bind to a duplex that is formed between an exon and a downstream intron.

ADAR's catalytic deaminase domain (DM) catalyses the conversion of a specific adenosine to inosine by hydrolytic deamination.[39]

Base flipping is the important step before deamination, in case of ADAR2-DD the residue coordinating for base filliping are recognized as 488 approaches to the targeted duplex RNA and form a minor groove. The side chain occupies the space created by the flip out base (Figure 6) ADAR prefer A-C>A-U>A-A>A-G mismatch, it indicates that pyrimidine preferred than purine because of purine conflict with 488 residues. This finding also complies with the other duplex DNA modifying enzymes.[42,43] The ADAR-DD result some conformational changes in the RNA causing kink in the RNA strand opposite to the target adenosine. This activity is also aided by the R510 and S495 side chains. The R510 conserved in ADAR1 and ADAR2 ion-pairs with the flip out base.

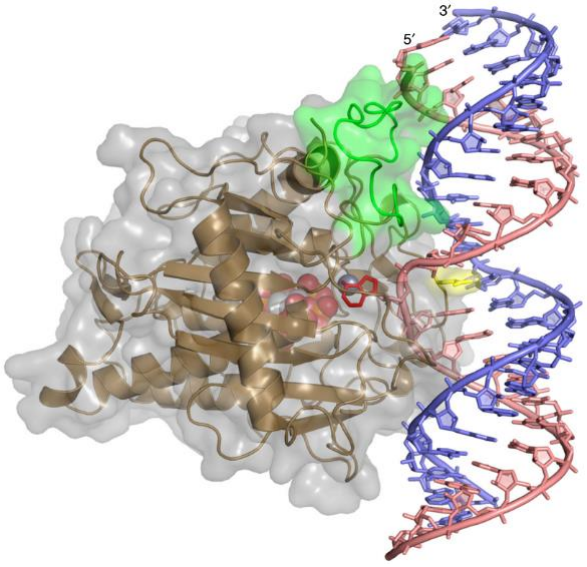


Figure 6. View of structure perpendicular to the dsRNA helical axis. Colors correspond to those in Red, flipped-out base ; gray space-filling sphere, zinc; yellow, Q488; green, previously disordered aa 454–477 loop; space filling, IP6. A transparent surface is shown for the hADAR2d protein.[40]

The C-terminal of the deaminase domain form the flanking region to be buried the inositol hexakisphosphate molecule (IP6). IP6 required for RNA editing by the ADAR2. The IP6 buried inside the catalytic domain and bounded to the domain with many polar residues.[40] This molecule associated with 29 water molecules. Hydrogen bond formed directly with the residue

or indirectly with the water molecules.[41] Thus main functions of IP6 is folding of the DD side by side it may have the function of modulating the DD during deamination.[44]

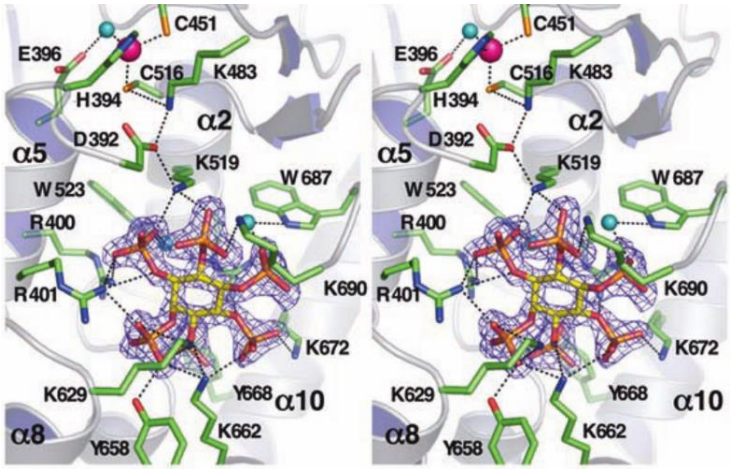


Figure 7. Image of the active site and IP6 binding site in hADAR2-D are shown in different views. Nucleophilic water (aqua sphere) is coordinated by zinc (magenta sphere) and E396 residue. Zinc and IP6 (yellow sticks) are communicated by hydrogen bonds (dashes) same condition between the conserved residues (green sticks) and IP6. W523 and W687 residues also relayed by water (aqua spheres).IP6 is almost embaded.[41]

IP6 is essential for the activity of ADARs *in vitro* and even *in vivo*.[44] IP6 may form during synthesis, external IP6 can also increase the activity of the ADARs. Why the cytidine deaminases, not deaminases the usual ADARs substrate because there are distance and space differences between these two deaminase molecules. Involvement of Zinc in the enzyme may acts as a structural assistance another may work as functional assistance. In case of ADAR2 active site associated to the Zn is principally classified as structural Zn(II) there is a water bound [41] which related to the catalytic assistance Zn (II). In case of catalytic Zn (II) ions the bonds with the ligands are longer, so easier to be inhibited by inhibitors.[45] Milimoller concentration EDTA have no activity on the protein.[46,47] It indicates that Zn is more structural rather than functional involvement.[44] DD activates the water molecule by the help of Zinc ion in their active site.[48,49] Zinc containing active site is present in ADAR2-DD, these are two cytosine 451 and 516 a histidine 394 (Figure 7). Side by side water molecule responsible for hydration of adenosine.[41] However, it is interesting that when mutation of the corresponding

residue in case of ADAR1, render this enzyme non-functional, it indicates that the hydrolytic mechanism is almost similar in both isoforms.[50] ADAR chemical reaction to its substrate is more similar to ADA, however, ADARS are evolved from CDA like APOBEC and AID[51,52] less likely related to ADA.

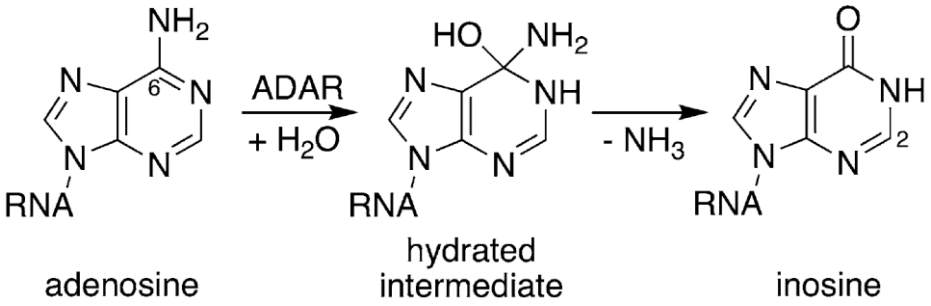


Figure 8. ADARs catalyse the deamination of adenosine to inosine via the addition of water to the 6-position to form a hydrated intermediate. Inosine differs from guanosine only at the 2-position, where inosine lacks guanosine's amino group.[44]

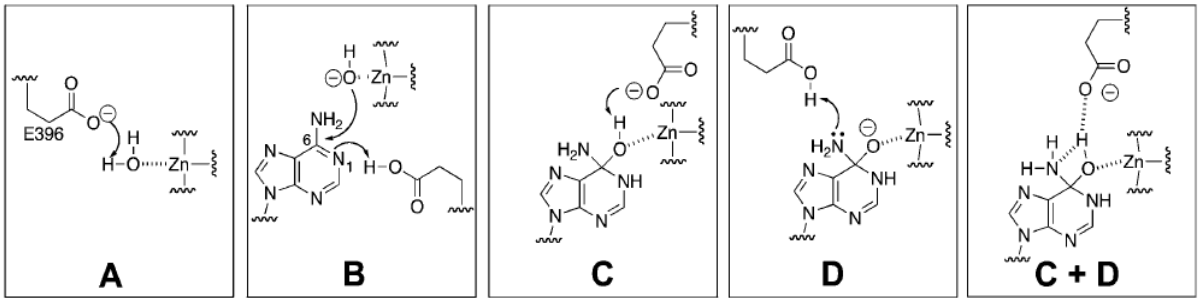


Figure 9. Proton transfers involving E396 during ADAR2-catalyzed adenosine deamination.[44]

Hydrolytic deamination (Figure 8) during adenosine deamination is facilities by zinc in presence of glutamic acid residues (E394) (Figure 9). This residue involved in proton transfer. In the above reaction from the water bound zinc molecule E394 transfer proton and convert zinc as highly reactive zinc-hydroxide which attack to C6 of the adenosine near the active site of deaminase domain. Ultimately removing the proton from the zinc-hydroxide produce ammonia during the reaction.

1.5 Point mutation and genetic diseases

A single base change in genome or mRNA can develop diseases. Some genetic diseases caused by point mutation are shown in Table 1. If we can restore mutated RNAs, these disease can be treated.

Table 1: T>C or G>A substitutions causing human genetic diseases.[53,54]

No	Disease state	Gene Symbol	Base change	Amino acid	Codon	Author	Journal	Vol.	Page	Year
1.	ADA deficiency	ADA	CTG-CCG	Leu-pro	107	Hirschhorn	PNAS	87	6171	1990
2	ADA deficiency	ADA	AAA-AGA	Lys-Arg	80	Valeria	EMBO J	5	113	1986
3	Adrenal hyperplasia	CA21HB	AAC-AGC	Asn-Ser	494	Rodrigues	EMBO J	6	1653	1987
4	APRT Deficiency	ART	ATG-ACG	Met-Thr	136	Hidaka	JCI	81	945	1988
5	Albinism, Ocul (1)	TYR	GAC-GGC	Asp-Gly	42	King	MBM	8	19	1991
6	Albumin Komagome 2	ALB	CAT-CGT	His-Arg	128	Madison	PNAS	88	9853	1991
7	Aldolase A defic.	ALDA	GAT-GGT	Asp-Gly	128	Kishi	PNAS	84	8623	1988
8	Amyloid polyneur	PALB	TAC-TGC	Tyr-Cys	114	Ueno	BBRC	169	143	1990
9	Amyloid prealbumin	PALB	GTG-GCG	Val-Ala	30	Johns	CLIN GENET	41	70	82
10	Androgen insens. syn.	AR	TAC-TGC	Tyr-Cys	761	McPhaul	JCI	87	1413	1991
11	Antithrombin III def.	AT3	TTC-TCC	Phe-Ser	402	Olds	TH	65	670	1991
12	Antitrypsin ∝ 1 def.	PI	CTC-CCG	Leu-Pro	41	Tahahashi	JBC	263	15528	1988
13	Antitrypsin ∝1 def.	PI	CTC-GCG	Val-Ala	213	Nukiwa	JBC	261	15989	1986
14	Chr. Granulomat. dis.	CYBB91	CAT-CGT	His-Arg	101	Bolscher	BlOOD	77	2482	1991
15	Cystic Fibrosi	CFR	TAT-TGT	Tyr-Cys	913	Vidaud	HUM	85	446	1990

16	Elliptocytosis	SPTA	CTG-CCG	Leu-Pro	207	Gallagher	JCI	89	892	1992
17	Elliptocytosis	SPTA	AAG-AGG	Lys-Arg	48	Floyd	BLOOD	78	1364	1991
18	Epidermolysis bull	KRT14	CTG-CCG	Leu-Pro	384	Bonfias	Science	254	1202	1991
19	G6PD deficiency	G6PD	CAC-CGC	His-Arg	32	Chao	NAR	19	6056	1991
20	G6PD deficiency	G6PD	CTG-CCG	Leu-pro	968	Beutler	BLOOD	74	2550	1989
21	Galactosaemia	GALT	CTG-CCG	Leu-Pro	195	Reichardt	GENOMICS	12	596	1992
22	Galactosaemia	GALT	CAG-CGG	Gln-Arg	188	Reichardt	AJHG	49	860	1991
23	Gangliosidosis GM1	GLB1	ATC-ACC	Ile-Thr	51	Yoshida	AJHG	49	435	1991
24	Gangliosidosis GM1	GLB1	TAT-TGT	Tyr-Cys	316	Yoshida	AJHG	49	435	1991
25	Gaucher's disease (1)	GBA	AAC-AGC	Asn-Ser	370	Tsuji	PNAS	85	2349	1988
26	Gaucher's disease (2)	GBA	CTG-CCG	Leu-Pro	444	Tsuji	PNAS	316	570	1987
27	Gyrate atrophy	OAT	CTT-CCT	Leu-Pro	402	Mitchell	PNAS	86	197	1989
28	HPRT deficiency	HPRT	CTA-CCA	Leu-Pro	40	Davidson	JCI	84	342	1989
29	HPRT deficiency	HPRT	ATT-ACT	Ile-Thr	41	Davidson	AJHG	48	951	1991
30	HPRT deficiency	HPRT	ATG-ACG	Met-Thr	56	Skopeck	HUM GENET	85	111	1990
31	HPRT deficiency	HPRT	TTG-TCG	Leu-Ser	130	Gibbs	PNAS	86	1919	1989
32	HPRT deficiency	HPRT	ATT-ACT	Ile-Thr	131	Davidson	AJHG	48	951	1991
33	HPRT deficiency	HPRT	GAT-GGT	Asp-Gly	52	Lightfoot	HUM GENET	88	695	1992
34	HPRT deficiency	HPRT	ATT-ACT	Ile-Thr	182	Tarle	GENOMICS	10	499	1991
35	HPRT deficiency	HPRT	GAT-GGT	Asp-Gly	200	Davidson	JBC	264	20	1989
36	HPRT deficiency	HPRT	CAT-CGT	His-Arg	203	Tarle	GENOMICS	10	499	1991
37	Haemoglobin	HBB	CAT-CGT	His-Arg	117	Kutlar	HUM	86	591	1991
38	Haemolytic anaemia	PGK	CTG-CCG	Leu-Pro	88	Maeda	BLOOD	77	1871	1991
39	Haemophilia A	F8	TTC-TCC	Phe-Ser	293	Higuchi	PNAS	88	7405	1991
40	Haemophilia A	F8	TTG-TCG	Leu-Ser	2166	Levinson	AJHG	46	53	1990
41	Haemophilia A	F8	GAA-CGA	Glu-Gly	272	Youssoufia	AJHG	42	867	1988
42	Haemophilia A	F8	AAA-AGA	Lys-Arg	425	Higuchi	PNAS	88	7405	1991
43	Haemophilia A	F8	TAT-TGT	Tyr-Cys	473	Higuchi	PNAS	88	7405	1991
44	Haemophilia A	F8	GAT-GGT	Asp-Gly	542	Higuchi	PNAS	88	7405	1991
45	Haemophilia A	F8	TAT-TGT	Tyr-Cys	1680	Traystman	GENOMICS	6	293	1990
46	Hepatic lipase def.	HL	AAT-AGT	Asn-Ser	193	Hegele	BBRC	179	78	1991
47	Insulin Resistance	INSR	CTG-CCG	Leu-Pro	233	Klinkham	EMBO J	8	2503	1989
48	Isovaleric acidaemia	IVD	CTA-CCA	Leu-Pro	13	Vockley	AJHG	49	147	1991
49	LDLR deficiency	LDLR	TAT-TGT	Tyr-Cys	807	Davis	CELL	45	15	86

50	Laron dwarfism	GHR	TTT-TCT	Phe-Ser	96	Amselem	NEJM	321	989	1989
51	Leprechaunism	INSR	CAC-CGC	His-Arg	209	Kadowaki	JCI	86	254	1990
52	Leukocyte adhes. Def.	LFA1	CTA-CCA	Leu-Pro	149	Wardlaw	JEM	172	335	1990
53	Lipoprt. lipase def.	LPL	ATT-ACT	Ile-Thr	194	Henderson	JCI	87	2005	1991
54	Lipoprt. lipase def.	LPL	GAT-GGT	Asp-Gly	158	Ma	JBC	267	1918	1992
55	LCAM deficiency	LCAM	AAT-AGT	Asn-Ser	351	Nelson	JBC	267	3351	1992
56	MCAD deficiency	MCAD	ATA-ACA	Ile-thr	375	Yokota	AJHG	49	1280	1991
57	Marfan syndrome	COL1 A2	CAG-CCG	Gln-Arg	618	Phillips	JCI	86	1723	1990
58	Methaemoglobin	DIA1	CTG-CCG	Leu-Pro	148	Katsube	AJHG	48	799	1991
59	Methylmalonicac id	MCM	CAT-CGT	His-Arg	532	Crane	JCI	89	385	1992
60	Neurofibromatos is (1)	NF1	CTC-CCG	Leu-Pro		Cawthon	CELL	62	193	1990
61	OTC deficiency	OTC	CTA-CCA	Leu-Pro	45	Grompe	AJHG	48	212	1991
62	OTC deficiency	OTC	CTT-CCT	Leu Pro	111	Grompe	AJHG	48	212	1991
63	Phenylketonuria	PAH	TTG-TCG	Leu-Ser	48	Konecki	HUM GENET	87	389	1991
64	Phenylketonuria	PAH	TTG-TCG	Leu-Ser	255	Hofman	AJHG	48	791	1991
65	Phenylketonuria	PAH	CTG-CCG	Leu-Pro	311	Licht-k	BIOCHEM	27	2881	1988
66	Phenylketonuria	PAH	TAT-TGT	Tyr-Cys	204	Wang	GENOMICS	10	449	1991
67	Phenylketonuria	PAH	GAA-GGA	Glu-Gly	221	Konecki	HUM GENET	87	389	1991
68	Phenylketonuria	PAH	TAC-TGC	Tyr-Cys	414	Okano	NEJM	324	1232	1991
69	Pompe disease	GAA	ATG-ACG	Met-Thr	318	Zhong	AJHG	49	635	1991
70	Retinitis pigmentosa	RDS	CTG-CCG	Leu-Pro	185	Kajiwara	MATURE	354	480	1991
71	Retinitis pigmentosa	RHO	TAC-TGC	Tyr-Cys	178	Sung	PNAS	88	6481	1991
72	Retinitis pigmentosa	RHO	GAC-GGC	Asp-Gly	190	Sung	PNAS	88	6481	1991
73	Ster.18-hydrox. Def.	CYP18	GTG-GCG	Val-Ala	386	Mitsuuchi	BBRC	182	974	1992
74	Thalassaemia α	HBA2	ATG-ACG	Met-Thr	-1	Piratsu	JBC	259	12315	1984
75	Thalassaemia α	HBA2	CTG-CCG	Leu-Pro	125	Goossens	NATURE	296	854	1982
76	Thalassaemia α	HBB	CTG-CCG	Leu-Pro	110	Kobayshi	BLOOD	70	1688	1987
77	Thalassaemia α	HBD	CTG-CCG	Leu-Pro	141	Trifillis	BLOOD	78	3298	1991
78	Wilm` tumor	WT1	GAC-GGC	Asp-Gly	396	Pelletier	CELL	67	437	1991

1.6 Artificial system for RNA editing and genetic restoration

The initial concept of therapeutic application by site-directed RNA editing activity reported by Woolf and his colleagues in 1995.[55] The designed complementary sequence should not interfere in any cellular process as it contains a mismatch. The DD contains dsRNA base flip-out capacity;[56,41,40] side by side the translating ribosomal protein also have a strong RNA unwinding activity.[57,58]

1.6.1 Chemically modified system

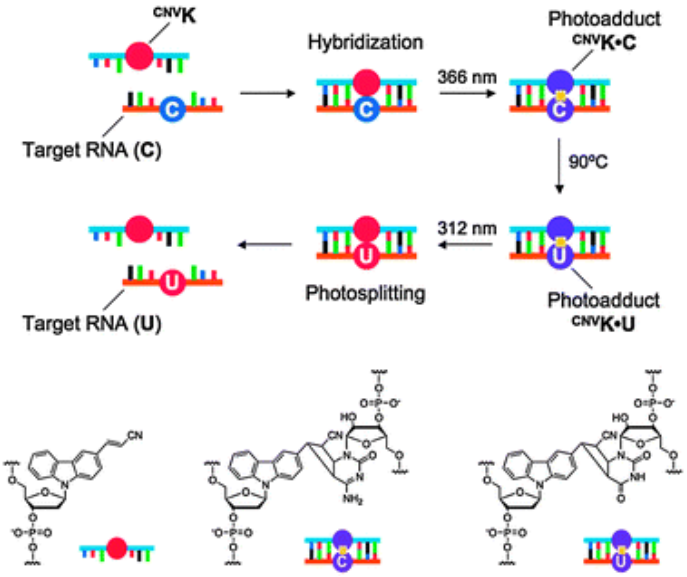


Figure 10. Schematic drawing of the photochemical RNA editing.[59]

Recently photochemical RNA editing is exciting as it has a great response and less expensive than enzymatic RNA editing. Professor Fujimoto group in JAIST has developed ultrafast crosslinking that are able to convert C-to-U for RNA editing. They designed modified ODN ^{CV}U and ^{CNV}K for the editing of RNA and DNA[60,59] (Figure 10). Professor Tsukahara group in JAIST use the above modified ODN and converted BFP to GFP by artificial RNA editing[61,62] for the future purpose of genetic restoration.

Another chemically modified system used the DD of ADAR1 and the ADAAR2 where they replaced the dsRBDs with SNAP-tag. This tag have the binding affinity with a cogent

benzylguanine (BG). BG can be attached to any guide RNA depending on the target (Figure 11). They reported that this modified system is functional in *in vitro*.[63,64] In case of HEK-293 T cell, annelids egg; light source augment the assembly and functionality of this system as well.[65] However, these therapies must continue to supply effector molecules forever.

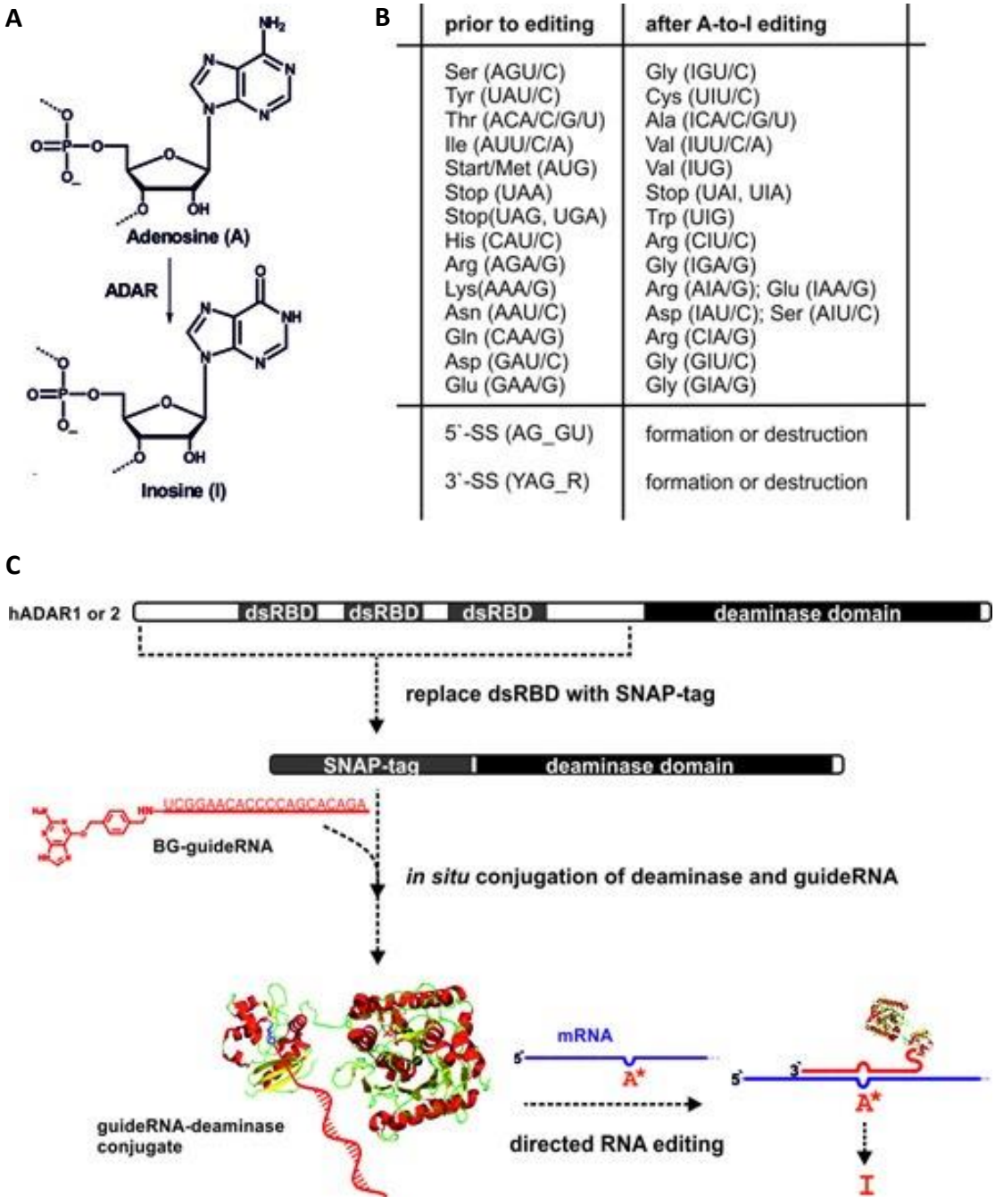


Figure 11. **A.** Hydrolytic deamination of adenosine by ADAR results inosine. **B**. Almost 12 among 20 amino acids can be converted by A-to-I RNA editing. **C.** The dsRBD is replaced by a SNAP-tag (an engineered O6-alkylguanine-DNA-alkyl transferase) which guided by a cogent benzylguanine (BG)-modified guide RNAs towards specific target.[63]

1.6.2 Tethering proteins used to control protein of interest

Basically, in eukaryotic organism two types of tethering system are used originated from bacteriophage, these are MS2 and the lambda N system.[66] The tethering system is used to monitor RNA localization *in vitro* and *in vivo* by cogent RNA, fused with the reporter RNA.[66] MS2 system is the widely used system in molecular biology used for *in vitro* and *in vivo* studies.[67–69] The theme of MS2 system originated from MS2 bacteriophage where the viral RNA is coated with the multimeric MS2 proteins which form the viral capsid. The icosahedral capsid of MS2 bacteriophage consists of 90 dimer protein elements that binds to the bacteriophage RNA to protect it from different adverse environments. The X-ray structure and the crystallographic data[70,71] indicate that the MS2 proteins form and assembled as dimer orientated in antiparallel orientation (Figure 12). The C-terminal of one molecule closely positioned to the N-terminal of another molecule. One MS2-RNA molecule bound to the MS2 protein dimer. The RNA binding sites of the cognate RNA span around the total ten β-sheets (five from each molecule).[66] This system was designed by Singer and colleagues for the monitoring of ASH1 particle localization in yeast cells[72] and was first applied for the mRNA localization of budding yeast.[73] The basic application of this system reported when it was used in yeast three-hybrid system which was used for identification of RNA target for an RNA binding protein.[74] Tethering assay was first reported in 1998. SR (Ser/Arg-rich) protein as a splicing modulator was studied by tethering with MS2 protein whereas the number and position of MS2 protein were changed. In addition, of the above work, the MS2 system was used to monitor mRNA localization of ASH1 (absent, small and homeotic discs 1) containing the 6X MS2-RNA loop in the 3′UTR and GFP to the bud tip in *Saccharomyces cerevisiae*. The most important features of this system are the specificity and sensitivity made in more reliable.[75] To date, more than hundred publication described the use of this system in different work related to RNA tagging, live imaging etc.[75]

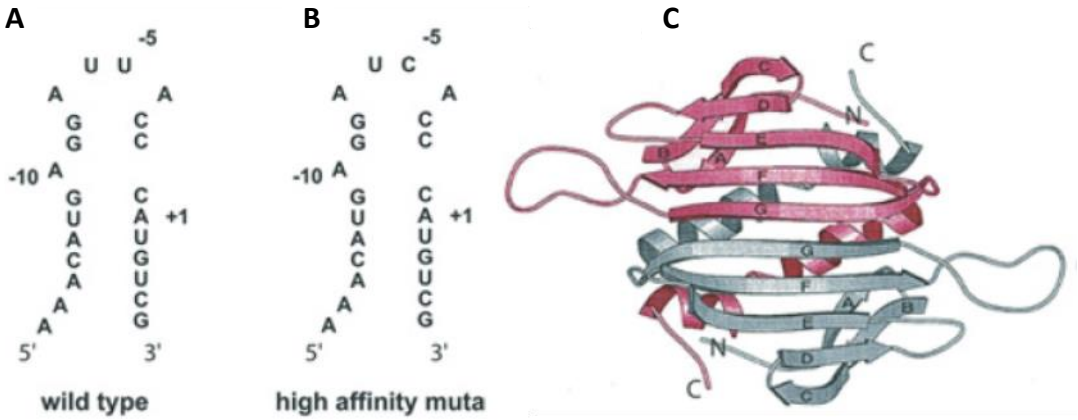


Figure 12. **A.** MS2-RNA stem-loop wild-type, **B.** change in few nucleotide result higher binding mutant, **C.** Antiparallel oriented MS2 protein dimer with 10 β-sheets.[66]

In principle, the MS2 protein is fused to the 3′ end of the reporter protein.[76] The other tethering system of eukaryotic can be used [77] but these have some disadvantages because iron-responsive elements present in 5′ or 3′ UTR of a number of elements may affect the binding efficiency and may also affect the hybrid protein as well.[78] MS2 closely related tether system was first reported in 1990 for protein RNA complex purification.[79]

The MS2 and the lambda N tethering system can produce the similar result. Factors responsible for nonsense-mediated decay was studied by both MS2 and Lambda N system indicate that tethering systems do not affect the functionality of adjacent factors.[80–82] To find out the difference between these two systems require more intensive study. In spite of these, the MS2 system bind with the cognate RNA as a dimer whereas the lambda N system binding is monomeric. Both systems forms a stem-loop structure with its cogent RNA but in MS2 system loop structure sequence is more important whereas in case of lambda N both the stem and loop RNA sequence is important for binding to its specific protein. Both Lambda N System is a very small system (12.2 kDa, 107 amino acids) that binds to box B Lambda N RNA. Montiel-Gonzalez *et al.* found that they can direct the ADAR2 catalytic domain to specific adenosine for site-directed RNA editing.[83] It also proved that this system can restore functionality of CFTR receptors in frog oocytes by eliminating stop codon of mutated CFTR genes.[83]

Both tethering proteins produces a stem-loop structure but MS2 bind with the loop containing nucleotides whereas lambda N binds to all the nucleotides. Therefore, we speculate that there might have more rotational freedom in case of the MS2 system than the lambda N system. Having rotational freedom is more advantageous for the artificial enzyme system

1.6.3 Genetically encoded enzyme system

In this case, the active parts of the deaminase system like deaminase, guide and substrate are encoded in plasmid or encoded in the cellular chromosome (Figure 13 and 14) that are expressed continuously and become assembled in the cell cytoplasm. After that the deaminase exert its function to the specific target.

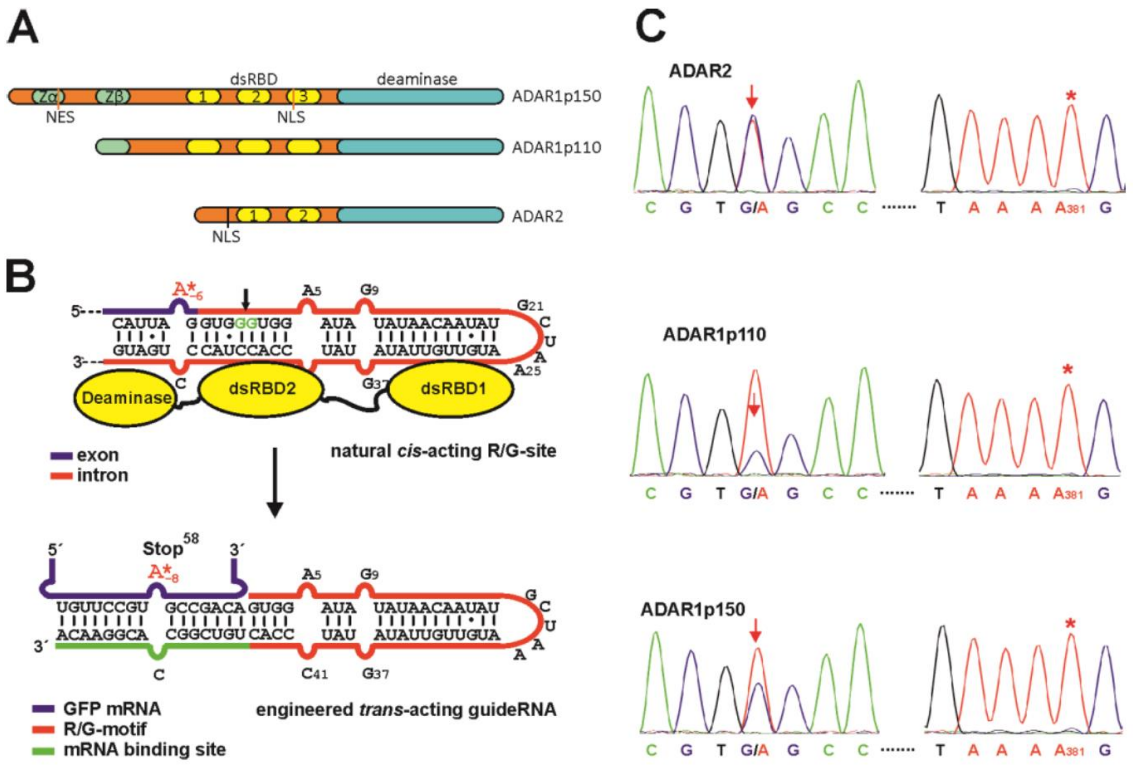


Figure 13. RNA editing with genetically encoded R/G-guide RNAs. **A.** Abovementioned three type of human ADARs used. Scheme of the three human ADARs used in this study. **B.** Assembly of substrate, guide R/G-motif of the GluR2 and deaminase. **C.** Sequencing result indicates that full-length ADAR2 more functional than the ADAR1 and the long form ADAR1 more functional than the short form when W58x codon in eGFP used as substrate. However, off-target editing was evident in all isoforms.[84]

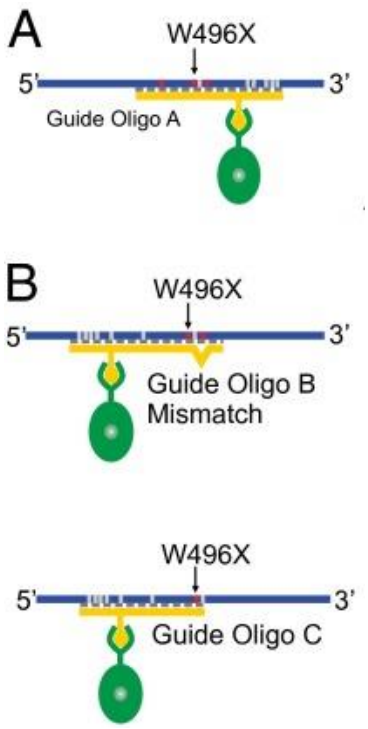


Figure 14. Use of lambda N system for directing the ADAR2 deaminase domain fused with the Lambda N peptide *in vitro*. Figure indicate that the guide can be designed to avoid the off-target W496X is the premature termination codon in CFTR is the target whereas the red marks are the editing site and white are the nonedited adenosine.[83]

Montiel-Gonzalez *et al.* in 2013, showed that the system (Figure 13) can guide the ADAR2-DD in HEK cell and frog oocyte by transient transfection to convert a stop codon of mutated GFP to read through codon that turns on fluorescence. It is found that in serotonin receptor RNA editing occurs leading to adverse physiological disorders. In some site-specific target to serotonin site, it was found overlapping specificities. Therefore, the serotonin receptor parts that bind with the dsRBD can be utilized for engineering artificial enzyme system by not only ADAR1 but also ADAR2.

1.7 Comparative activity of different ADARs towards RNA editing

Comparative RNA editing of different isoforms of ADAR not reported well in cell culture system even in *in vitro* system. In the previous study, most of the cases the full-length ADARs were used for their deaminase activity.[84] However, in my study and in some other study[83] only the deaminase domain is used. Use of the deaminase domain results in deletion of different parts of the enzyme like dsRNA binding domain NLS, NES.[85] These elements render this enzyme as a shuttle protein between nucleus and cytoplasm. The full-length ADARs are shuttle protein, ADAR2 shuttles between nucleoplasm, nucleolus whereas ADAR1 shuttles between nucleolus, nucleoplasm and cytoplasm.[86] Even a region in the N terminal half of the ADAR1-DD interferes with nuclear localization.[86] In spite of that, recently only deaminase domain of ADAR2 is used for cytosolic mRNA correction.[87,83,88] But only deaminase domain use for RNA editing purpose most probably will render the enzyme cytoplasmic. So, I speculate that the use of only deaminase domain will be more reliable for RNA editing in therapeutic purposes. As it is planned to use the target the cytosolic defective mRNAs, therefore, the enzyme system should be localized in cytosol only. The pre-mRNA and mRNAs have a more complex structure with loops, mismatches and bulges, that make it easier for the full-length ADAR1 to recognize it with its dsRBDs irrespective of a specific target.

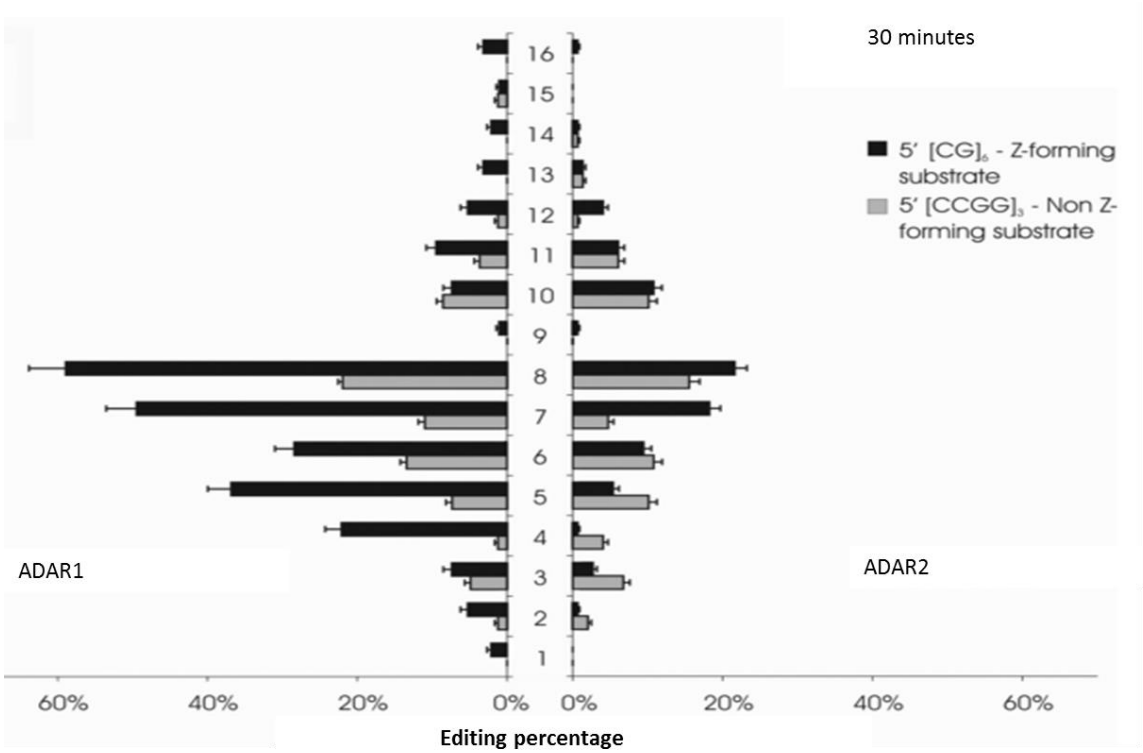


Figure 15. The relative A-to-I editing values after 30 min of incubation are shown. Both substrates were co-incubated (each at a concentration of 0.16 fmol/l) with either ADAR1-L or ADAR2 containing cell extracts [89].

In the above figure, two substrates were incubated with the ADAR1 and ADAR2 separately and the histogram indicates that ADAR1 deaminase activity much higher than the ADAR2 (Figure 15). In another previous study regarding the ADAR2 containing Alu-cassette in the deaminase domain found that, insertion of the Alu-cassette in the deaminase domain does not differ in substrate specificity but differ in their deaminase activity i.e. inserted Alu-cassette decreases the deaminase activity of the full-length enzyme.[90] In my study, I studied the effect of inserted Alu-cassette on the structure and the deaminase activity of the only deaminase domain of ADARs.

1.8 Aim of this study

In case of gene therapy whole gene or transcript is targeted but my aim is to correct point mutations that result stop codons, defective codons ultimately production of defective or non-functional protein and diseases. The defective codons can be altered to the desired codon or another read through codon by base substitution. Considering the above purpose the aims of this study are-

1. Development of artificial enzyme system for correction of the mutated stop codons amber (TAG) and (TAA) to read through codon.
2. To find out the most active ADARs deaminase domain (DD) that can be utilized for the purpose of genetic treatment.
3. Improve the efficiency of this system.

1.9 References

1 Keegan LP, Gallo A, O'Connell M a. The many roles of an RNA editor. *Nat Rev Genet* 2001; **2(11)**: 869–78.

2 Benne R, Van Den Burg J, Brakenhoff JPJ, Sloof P, Van Boom JH, Tromp MC. Major transcript of the frameshifted coxll gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* 1986; **46**: 819–826.

3 Powell LM, Wallis SC, Pease RJ, Edwards YH, Knott TJ, Scott J. A novel form of tissue-specific RNA processing produces apolipoprotein-B48 in intestine. *Cell* 1987; **50**: 831–840.

4 Kim H, Kim J-S. A guide to genome engineering with programmable nucleases. *Nat Rev Genet* 2014; **15**: 321–334.

5 Maas S, Rich A. Changing genetic information through RNA editing. *BioEssays* 2000; **22**: 790–802.

6 Zipeto MA, Jiang Q, Melese E, Jamieson CHM. RNA rewriting, recoding, and rewiring in human disease. *Trends Mol Med* 2015; **21**: 549–559.

7 Maas S, Rich A, Nishikura K. A-to-I RNA editing: Recent news and residual mysteries. *J Biol Chem* 2003; **278**: 1391–1394.

8 Bass BL. RNA editing by adenosine deaminates that act on RNA. *Annu Rev Biochem* 2002; **71**: 817–846.

9 Maydanovych O, Beal PA. Breaking the central dogma by RNA editing. *Chem Rev* 2006; **106**: 3397–3411.

10 Kazuko Nishikura. Functions and regulation of RNA editing by ADAR deaminases. *Annu Rev Biochem* 2010; **79**: 321–349.

11 Hoopengardner B. Nervous System Targets of RNA Editing Identified by Comparative Genomics. *Science (80-)* 2003; **301**: 832–836.

12 Burns CM, Chu H, Rueter SM, Hutchinson LK, Canton H, Sanders-Bush E *et al.* Regulation of serotonin-2C receptor G-protein coupling by RNA editing. *Nature* 1997; **387**: 303–308.

13 Lomeli H, Mosbacher J, Melcher T, Höger T, Jörg RP, Kuner T *et al.* Control of Kinetic Properties of AMPA Receptor Channels by Nuclear RNA Editing. *Science (80-)* 1994; **266**: 1709–1713.

14 Köhler M, Burnashev N, Sakmann B, Seeburg PH. Determinants of ca2+ permeability in both TM1 and TM2 of high affinity kainate receptor channels: Diversity by RNA editing. *Neuron* 1993; **10**: 491–500.

15 Ohlson J, Pedersen JS, Haussler D, Marie O. Editing modifies the GABA A receptor subunit a3. *RNA* 2007; : 698–703.

16 Sommer B, Köhler M, Sprengel R, Seeburg PH. RNA editing in brain controls a determinant of ion flow in glutamate-gated channels. *Cell* 1991; **67**: 11–19.

17 Tariq A, Jantsch MF. Transcript diversification in the nervous system: A to I RNA editing in CNS function and disease development. *Front Neurosci* 2012; **6**: 1–13.

18 Palladino MJ, Keegan LP, O'Connell MA, Reenan R a. A-to-I Pre-mRNA Editing in Drosophila Is Primarily Involved in Adult Nervous System Function and Integrity. *Cell* 2000; **102**: 437–449.

19 Wang Q. Requirement of the RNA Editing Deaminase ADAR1 Gene for Embryonic Erythropoiesis. *Science (80-)* 2000; **290**: 1765–1768.

20 Higuchi M, Maas S, Single FN, Hartner J, Rozov a, Burnashev N *et al.* Point mutation in an AMPA receptor gene rescues lethality in mice deficient in the RNA-editing enzyme ADAR2. *Nature* 2000; **406**: 78–81.

21 Bhalla T, Rosenthal JJC, Holmgren M, Reenan R. Control of human potassium channel inactivation by editing of a small mRNA hairpin. *Nat Struct Mol Biol* 2004; **11**: 950–956.

22 Vissel B, Royle GA, Christie BR, Schiffer HH, Ghetti A, Tritto T *et al.* The Role of RNA Editing of Kainate Receptors in Synaptic Plasticity and Seizures. *Neuron* 2001; **29**: 217–227.

23 Burnashev N, Monyer H, Seeburg PH, Sakmann B. Divalent ion permeability of AMPA receptor channels is dominated by the edited form of a single subunit. *Neuron* 1992; **8**: 189–198.

24 Maas S, Kawahara Y, Tamburro KM, Nishikura K. A-to-I RNA editing and human disease. *RNA Biol* 2006; **3**: 1–9.

25 Chen S-H, Habib G, Yang C-Y, Gu Z-W, Lee BR, Weng S-A *et al.* Apolipoprotein B-48 is the product of a messenger RNA with an organ-specific in-frame stop codon. *Science* 1987; **238**: 363–366.

26 Nagalla SR, Barry BJ, Spindel ER. Cloning of complementary DNAs encoding the amphibian bombesin-like peptides Phe8 and Leu8 phyllolitorin from Phyllomedusa sauvagei: potential role of U to C RNA editing in generating neuropeptide diversity. *Mol Endocrinol* 1994; **8**: 943–951.

27 Villegas J, Müller I, Arredondo J, Pinto R, Burzio LO. A putative RNA editing from U to C in a mouse mitochondrial transcript. *Nucleic Acids Res* 2002; **30**: 1895–901.

28 Mohini Sharma P, Bowman M, Madden SL, Rauscher III FJ, Sukumar S. RNA editing in the Wilms' tumor susceptibility gene, WTl. *Genes Dev* 1994; **8**: 720–731.

29 Song W, Liu Z, Tan J, Nomura Y, Dong K. RNA editing generates tissue-specific sodium channels with distinct gating properties. *J Biol Chem* 2004; **279**: 32554–32561.

30 Knie N, Grewe F, Fischer S, Knoop V. Reverse U-to-C editing exceeds C-to-U RNA editing in some ferns – a monilophyte-wide comparison of chloroplast and mitochondrial RNA editing suggests independent evolution of the two processes in both organelles. *BMC Evol Biol* 2016; **16**: 1–12.

31 Rueter SM, Dawson TR, Emeson RB. Regulation of alternative splicing by RNA editing. *Nature* 1999; **399**: 75–80.

32 Ota H, Sakurai M, Gupta R, Valente L, Wulff BE, Ariyoshi K *et al.* ADAR1 forms a complex with dicer to promote MicroRNA processing and RNA-induced gene silencing. *Cell* 2013; **153**: 575–589.

33 Sakurai M, Shiromoto Y, Ota H, Song C, Kossenkov A V, Wickramasinghe J *et al.* ADAR1 controls apoptosis of stressed cells by inhibiting Staufen1-mediated mRNA decay. *Nat Struct Mol Biol* 2017; **24**: 534–543.

34 Wang IX, So E, Devlin JL, Zhao Y, Wu M, Cheung VG. ADAR Regulates RNA Editing, Transcript Stability, and Gene Expression. *Cell Rep* 2013; **5**: 849–860.

35 Mannion NM, Greenwood SM, Young R, Cox S, Brindle J, Read D *et al.* The RNA-Editing Enzyme ADAR1 Controls Innate Immune Responses to RNA. *Cell Rep* 2014; **9**: 1482–1494.

36 Wahlstedt H, Daniel C, Ensterö M, Res G, Wahlstedt H, Daniel C *et al.* Large-scale mRNA sequencing determines global regulation of RNA editing during brain development Large-scale mRNA sequencing determines global regulation of RNA editing during brain development. *Genome Res* 2009; **19**: 978–986.

37 Slavov D, T. Crnogorac-Jurc ˇevic ´., Clark M, Gardiner K. Comparative analysis of the DRADA A-to-I RNA editing gene from mammals, pufferfish and zebrafish. *Gene* 2000; **250**: 53–60.

38 Patterson JB, Samuel CE. Expression and regulation by interferon of a double-stranded-RNA-specific adenosine deaminase from human cells: evidence for two forms of the deaminase. *Mol Cell Biol* 1995; **15**: 5376–5388.

39 Hogg M, Paro S, Keegan LP, O'Connell MA. *RNA Editing by Mammalian ADARs*. 1st ed. Elsevier Inc., 2011 doi:10.1016/B978-0-12-380860-8.00003-3.

40 Matthews MM, Thomas JM, Zheng Y, Tran K, Phelps KJ, Scott AI *et al.* Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity. *Nat Struct Mol Biol* 2016; **23**: 426–433.

41 Mark R. Macbeth, Heidi L. Schubert, Andrew P. VanDemark, Arunth T. Lingam, Christopher P. Hill BLB. Inositol Hexakisphosphate Is Bound in the ADAR2 Core and Required for RNA Editing. *Science (80-)* 2005; **309**: 2033–2037.

42 Klimasauskas S, Kumar S, Roberts RJ, Cheng X. Hhal methyltransferase flips its target base out of the DNA helix. *Cell* 1994; **76**: 357–369.

43 Daujotyte D, Serva S, Vilkaitis G, Merkiene E, Venclovas Č, Klimašauskas S. HhaI DNA methyltransferase uses the protruding Gln237 for active flipping of its target cytosine. *Structure* 2004; **12**: 1047–1055.

44 Goodman RA, MacBeth MR, Beal PA, Rena A. Goodman MRM and PAB, Goodman RA, MacBeth MR *et al.* ADAR Proteins: Structure and Catalytic Mechanism. *Curr Top Microbiol Immunol* 2012; **353**: 1–33.

45 Lee Y ming, Lim C. Physical Basis of Structural and Catalytic Zn-Binding Sites in Proteins. *J Mol Biol* 2008; **379**: 545–553.

46 Hough RF, Bass BL. Purification of the Xenopus laevis double-stranded RNA adenosine deaminase. *J Biol Chem* 1994; **269**: 9933–9939.

47 Saccomanno L, Bass BL. The cytoplasm of Xenopus oocytes contains a factor that protects double-stranded RNA from adenosine-to-inosine modification. *Mol Cell Biol* 1994; **14**: 5425–5432.

48 Betts L, Xiang S, Short S a, Wolfenden R, Carter CW. Cytidine deaminase. The 2.3 A crystal structure of an enzyme: transition-state analog complex. J. Mol. Biol. 1994; **235**: 635–656.

49 Wilson DK, Rudolph FB, Quiocho FA. Atomic Structure and Understanding Catalysis Immunodeficiency Mutations. *Science (80-)* 1991; **252**: 1278–1284.

50 Fang Lai RD and KN. Mutagenic Analysis of Double-stranded RNA Adenosine Deaminase, a Candidate Enzyme for RNA Editing of Glutamate-gated Ion Channel Transcripts. *J Biol Chem* 1995; **270**: 17098–17105.

51 Gerber AP, Keller W. RNA editing by base deamination: More enzymes, more targets, new mysteries. *Trends Biochem Sci* 2001; **26**: 376–384.

52 Kim U, Wang Y, Sanford T, Zeng Y, Nishikura K. Molecular cloning of cDNA for double-stranded RNA adenosine deaminase, a candidate enzyme for nuclear RNA editing. *Proc Natl Acad Sci U S A* 1994; **91**: 11457–61.

53 Luyen VUTHI. *Study on Conversion from Blue Fluorescent Protein (BFP) to Green Fluorescent Protein (GFP) in vitro by Using Site-Directed Chemical RNA Editing VU*. School of Materials Science, Japan Advanced Institute of Science and Technology, 2015.

54 Miller AD. Human gene therapy comes of age. *Nature* 1992; **357**: 455–460.

55 Woolf TM, Chase JM, Stinchcomb DT. Toward the therapeutic editing of mutated RNA sequences. *Proc Natl Acad Sci* 1995; **92**: 8298–8302.

56 Kuttan A, Bass BL. Mechanistic insights into editing-site specificity of ADARs. *Proc Natl Acad Sci* 2012; **109**: E3295–E3304.

57 Melton DA. Injected anti-sense RNAs specifically block messenger RNA translation in vivo. *Proc Natl Acad Sci U S A* 1985; **82**: 144–148.

58 Louis J. Maher IBJD. Comparative hybrid arrest by tandem antisense oUgodeoxyribonucleotides or oligodeoxy- ribonucleoside methylpbosphonates in a cell-free system. *Nucleic Acids Res* 1988; **16**: 3341–3358.

59 Fujimoto K, Konishi-Hiratsuka K, Sakamoto T, Yoshimura Y. Site-specific photochemical RNA editing. *Chem Commun (Camb)* 2010; **46**: 7545–7547.

60 Fujimoto K, Yamada A, Yoshimura Y, Tsukaguchi T, Sakamoto T. Details of the ultrafast DNA photo-cross-linking reaction of 3-cyanovinylcarbazole nucleoside: Cis-trans isomeric effect and the application for SNP-based genotyping. *J Am Chem Soc* 2013; **135**: 16161–16167.

61 Vu LT, Nguyen TTK, Md Thoufic AA, Suzuki H, Tsukahara T. Chemical RNA Editing for Genetic Restoration: The Relationship between the Structure and Deamination Efficiency of Carboxyvinyldeoxyuridine Oligodeoxynucleotides. *Chem Biol Drug Des* 2016; **87**: 583–593.

62 Vu LT, Nguyen TTK, Alam S, Sakamoto T, Fujimoto K, Suzuki H *et al.* Changing Blue Fluorescent Protein to Green Fluorescent Protein Using Chemical RNA Editing as a Novel Strategy in Genetic Restoration. *Chem Biol Drug Des* 2015; **86**: 1242–1252.

63 Schneider MF, Wettengel J, Hoffmann PC, Stafforst T. Optimal guideRNAs for re-directing deaminase activity of hADAR1 and hADAR2 in trans. *Nucleic Acids Res* 2014; **42**: 1–9.

64 Vogel P, Schneider MF, Wettengel J, Stafforst T. Improving site-directed RNA editing in vitro and in cell culture by chemical modification of the guideRNA. *Angew Chemie - Int Ed* 2014; **53**: 6267–6271.

65 Hanswillemenke A, Kuzdere T, Vogel P, Jékely G, Stafforst T. Site-Directed RNA Editing in Vivo Can Be Triggered by the Light-Driven Assembly of an Artificial Riboprotein. *J Am Chem Soc* 2015; **137**: 15875–15881.

66 Keryer-Bibens C, Barreau C, Osborne HB. Tethering of proteins to RNAs by bacteriophage proteins. *Biol Cell* 2008; **100**: 125–138.

67 Fusco D, Accornero N, Lavoie B, Shenoy SM, Blanchard JM, Singer RH *et al.* Single mRNA molecules demonstrate probabilistic movement in living mammalian cells. *Curr Biol* 2003; **13**: 161–167.

68 Jurica MS, Licklider LJ, Gygi SR, Grigorieff N, Moore MJ. Purification and characterization of native spliceosomes suitable for three-dimensional structural analysis. *Rna* 2002; **8**: 426–439.

69 Hook B, Bernstein D, Zhang B, Wickens M. RNA-protein interactions in the yeast three-hybrid system: affinity, sensitivity, and enhanced library screening. *Rna* 2005; **11**: 227–233.

70 Ni CZ, Syed R, Kodandapani R, Wickersham J, Peabody DS, Ely KR. Crystal structure of the MS2 coat protein dimer: implications for RNA binding and virus assembly. *Structure* 1995; **3**: 255–63.

71 Valegârd K, Murray JB, Stonehouse NJ, van den Worm S, Stockley PG, Liljas L. The three-dimensional structures of two complexes between recombinant MS2 capsids and RNA operator fragments reveal sequence-specific protein-RNA interactions. *J Mol Biol* 1997; **270**: 724–738.

72 Bertrand E, Chartrand P, Schaefer M, Shenoy SM, Singer RH, Long RM. Localization of ASH1 mRNA Particles in Living Yeast. *Mol Cell* 1998; **2**: 437–445.

73 Beach DL, Salmon ED, Bloom K. Localization and anchoring of mRNA in budding yeast. *Curr Biol* 1999; **9**: 569–578.

74 Bardwell VJ WM. Purification of RNA and RNA-protein complexes by an R17 coat protein affinity method. *Nucleic Acids Res* 1991; **93**: 8496–8501.

75 Urbanek MO, Galka-Marciniak P, Olejniczak M, Krzyzosiak WJ. RNA imaging in living cells - methods and applications. *RNA Biol* 2014; **11**: 1083–95.

76 Tyagi S. Imaging intracellular RNA distribution and dynamics in living cells. *Nat Methods* 2009; **6**: 331–338.

77 De Gregorio E, Preiss T, Hentze MW. Translation driven by an eIF4G core domain in vivo. *EMBO J* 1999; **18**: 4865–4874.

78 Pantopoulos K. Iron metabolism and the IRIE/IRP regulatory system - An update. *Ann N Y Acad Sci* 2004; **5**: 1–13.

79 Bardwell VJ, Wickens M. Purification or RNA and RNA - protein complexes by an R17 coat protein affinity method. *Nucleic Acids Res* 1991; **19**: 1980.

80 Lykke-Andersen J, Shu M Di, Steitz JA. Human Upf proteins target an mRNA for nonsense-mediated decay when downstream of a termination codon. *Cell* 2000; **103**: 1121–1131.

81 Lykke-Andersen J. Communication of the Position of Exon-Exon Junctions to the mRNA Surveillance Machinery by the Protein RNPS1. *Science (80-)* 2001; **293**: 1836–1839.

82 Kunz JB, Neu-Yilik G, Hentze MW, Kulozik AE, Gehring NH. Functions of hUpf3a and hUpf3b in nonsense-mediated mRNA decay and translation. *Rna* 2006; **12**: 1015–1022.

83 Montiel-Gonzalez MF, Vallecillo-Viejo I, Yudowski GA, Rosenthal JJC. Correction of mutations within the cystic fibrosis transmembrane conductance regulator by site-directed RNA editing. *Proc Natl Acad Sci* 2013; **110**: 18285–18290.

84 Heep M, Mach P, Reautschnig P, Wettengel J, Stafforst T. Applying human ADAR1p110 and ADAR1p150 for Site-Directed RNA editing—G/C substitution stabilizes guide-RNAs against editing. *Genes (Basel)* 2017; **8**: 1–7.

85 Desterro JMP, Keegan LP, Lafarga M, Berciano MT, O'Connell M, Carmo-Fonseca M. Dynamic association of RNA-editing enzymes with the nucleolus. *J Cell Sci* 2003; **116**: 1805–1818.

86 Strehblow A, Hallegger M, Jantsch MF. Nucleocytoplasmic Distribution of Human RNA- editing Enzyme ADAR1 Is Modulated by Double-stranded RNA-binding Domains, a Leucine-rich Export Signal, and a Putative Dimerization Domain. *Mol Biol Cell* 2002; **13**: 3822–3835.

87 Vogel P, Stafforst T. Site-directed RNA editing with antagomir deaminases--a tool to study protein and RNA function. *ChemMedChem* 2014; **9**: 2021–2025.

88 Montiel-Gonz´alez MF, Vallecillo-Viejo IC, Rosenthal JJC. An efficient system for selectively altering genetic information within mRNAs. *Nucleic Acids Res* 2016; **44**: 1–12.

89 Koeris M, Funke L, Shrestha J, Rich A, Maas S. Modulation of ADAR1 editing activity by Z-RNA in vitro. *Nucleic Acids Res* 2005; **33**: 5362–5370.

90 Gerber A, O'Connell MA, Keller W. Two forms of human double-stranded RNA-specific editase 1 (hRED1) generated by the insertion of an Alu cassette. RNA. 1997; **3**: 453–63.

Chapter II

Site-directed RNA editing by adenosine deaminase acting on RNA (ADAR1) for correction of the genetic code in gene therapy

2.1 Introduction

Programmable nuclease tools for genome editing, including transcription activator-like effector nucleases (TALENs), RNA-guided engineered nucleases (RGENs), zinc-finger nucleases (ZFNs), and the clustered regularly interspaced short palindromic repeat (CRISPR)-Cas9 system,[1] hold promise for use in gene therapy. Recently, notwithstanding the ethical issues related to clinical applications,[2] the use of the CRISPR-Cas9 system in eukaryotic cells has attracted a great deal of attention.[3] Point mutations generally cause single amino acid substitutions, which can generate tumor-specific antigens that ultimately lead to cancer.[4] Genetically engineered T-cells have been used to eliminate cells harboring such point mutations.[4] Other diseases caused by point mutations include cystic fibrosis,[5] chronic granulomatous disease,[6] hepatic lipase deficiency, and hemophilia A.[7]

RNA editing is a type of post-transcriptional modification of the gene-encoded sequence, which allows several forms of a protein to be produced from the same gene.[8] RNA editing can result from insertion, deletion, or substitution of nucleotides.[1, 9] In mammalian cells, adenosine (A) and cytosine (C) are edited to inosine (I) [which base-pairs like guanosine (G)] and uracil (U), respectively, in both coding and noncoding sequences.[10-12] The AID-APOBEC enzyme family is responsible for C-to-U conversion, and numerous targets of APOBEC1 have been reported in 3'UTRs.[12] On the other hand, A-to-I editing is mediated by the adenosine deaminase acting on RNA (ADAR) family. Such editing occurs in a wide variety of RNAs.[13, 14] In a systematic study, A-to-I editing sites were detected in 6% of human pri-miRNAs.[15] In microRNAs, A-to-I editing alters binding affinity and gene silencing activity.[16]

Vertebrates have three ADARs: ADAR1[17] and ADAR2,[18] which are functional, and ADAR3,[19] which is non-functional. All three enzymes have a typical deaminase domain in their C-terminal regions,[10] and N-termini containing double-stranded RNA-binding domains (dsRBDs).[20] ADAR1 contains an additional Z-DNA-binding domain,[21] whereas ADAR3 contains an Arg-rich single-stranded RNA (ssRNA)-binding domain.[19] Homodimerization is important for ADAR activity.[22, 23] ADAR1 and ADAR2 form homodimers, whereas ADAR3 does not, potentially explaining its lack of deaminase activity.[22] The first and third dsRBDs of ADAR1 and ADAR2, respectively, are responsible for their homodimerization.[24] Both elevation[25] and reduction[26] of ADAR editing activity have been associated with carcinoma.[27,28]

Abnormalities in RNA editing can result in various neurological and behavioral anomalies, including epilepsy and Prader–Willi syndrome.[11,29,30] In mice, *Adar1* knockout results in early embryonic death,[31,32] whereas, in humans, an insertion mutation in the ADAR1 gene results in dyschromatosis symmetrica hereditaria.[33] Reprogramming of genetic information at the RNA level has the potential to provide insight into fundamental biology, as well as opportunities for gene therapy applications.[34] Although A-to-I substitution is a single-nucleotide change, it can have a significant physiological or clinical impact. When RNA editing occurs in open reading frames of genes, 12 of the 20 standard amino acids can be altered. Accordingly, this strategy can target amino acids (Gln, Arg, His, Tyr, Ser, Thr, and others) that are important for catalysis, signaling, and post-translational modifications.[35] Start and stop codons, splicing signals, and miRNA recognition sites can likewise be modified.[35] Therefore, applications of site-specific RNA editing hold promise for therapeutic purposes. Reports of correcting mutant RNAs *in vitro*

and in *Xenopus* embryos have attracted a great deal of attention, but several technical challenges remain, including more specific control of deaminase activity.[36]

Recently, different types of guide RNA have been used to make the enzyme target-specific. Earlier studies used synthetic guide RNA, along with fusion of lambda-N peptide to the deaminase domain of human ADAR2 (lambda-N+ADAR2), to correct mutations in CFTR (cystic fibrosis transmembrane conductance regulator),[37] and chemically modified guide RNAs have been used to correct point mutations in human cell lines and annelids.[38] Previously, we demonstrated chemical RNA editing using oligodeoxynucleotides (ODNs) containing 5'-carboxyvinyl-2'-deoxyuridine [(CV)U].[39,40] Very recently, a genetically encoded guide RNA was used to repair point mutations associated with a neuronal disease phenotype.[35]

In this study, I sought to develop more useful tools for restoring the genetic code, using the MS2 system and a genetically encoded novel guide RNA to direct the deaminase domain of wild-type ADAR1 to specific targets. The MS2 system is more frequently used than lambda N/B box systems for tethering of protein to RNAs.[41] This system is a novel system commonly used for *in vivo* label with GFP for RNAs.[42] To our knowledge, this study represents the first example of engineering the MS2 system to design an RNA editing enzyme complex capable of targeting specific point mutations.

2.2 Materials and methods

2.2.1 Plasmid construct preparation

To enable targeting of the enzymes to a TAG stop codon of interest, I cloned the deaminase domain of ADAR1 downstream of MS2 in pCS2+MT using the *Nco*I and *Xba*I (Takara) restriction enzymes to yield pCS2+MT-MS2HB-ADAR1 (Figure 1). The domain was amplified by PCR from cDNA of human intestine using forward and reverse primers harboring the appropriate restriction sites. The primers were designed using the Primer3 online software. ADAR1 expresses well in small intestine, testes, thymus, and uterus (Figure 2). The in-frame and domain structure were confirmed using the ExPASY bioinformatics resource portal and NCBI BLAST search.

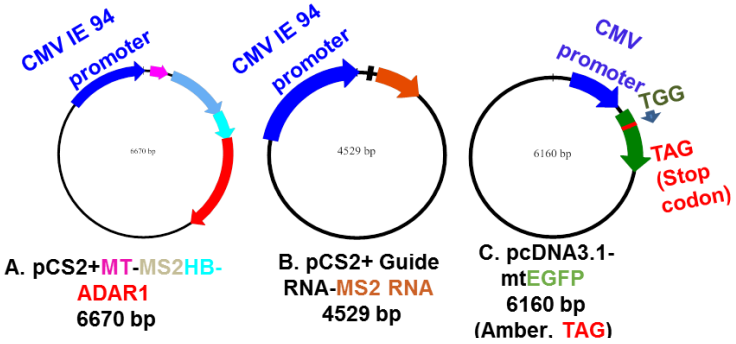


Figure 1. Plasmid constructs used in the study

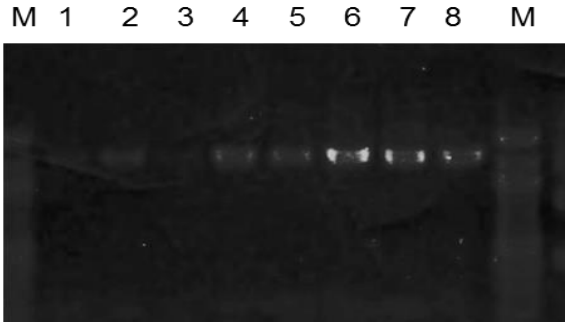


Figure 2. Expression of the human ADAR1 deaminase domain (~1200 bp) in different tissues. 1, bone marrow; 2, lung; 3, skeletal muscle; 4, small intestine; 5, spinal cord ; 6, testes; 7, thymus; 8, uterus; M, 100 bp marker.

pCS2+MT-MS2HB-ADAR1 nucleotide sequence

6x Myc Tag, ATGGAGCAAAAGCTCATTTCTGAAGAGGACTTGAAT; italics after NcoI, MS2; underlined italics, HB; bold underlined, linker between MS2HB and ADAR1; after the linker sequence, ADAR1 deaminase domain.

The restriction enzymes used to prepare the construct were NcoI (CCATGG), XhoI (CTCGAG), and XbaI (TCTAGA). Start and stop codons are highlighted

GGACTTGAATGAAATGGAGCAAAAGCTCATTTCTGAAGAGGACTTGAATGAAATGGAGAGCT
TGGGCGACCTCACCATGGCTGCAGGAATTGATCCGCGGCCGCGC*ATGGCTTCTAACTTTACTC*
AGTTCGTTCTCGTCGACAATGGCGGAACTGGCGACGTGACTGTCGCCCCAAGCAACTTCGCTA
ACGGGGTCGCTGAATGGATCAGCTCTAACTCGCGTTCACAGGCTTACAAAGTAACCTGTAGCG
TTCGTCAGAGCTCTGCGCAGAAGCGCAAATACACCATCAAAGTCGAGGTGCCTAAAGTGGCAA
CCCAGACTGTTGGTGGTGTAGAGCTTCCTGTAGCCGCATGGCGTTCGTACTTAAATATGGAAC
TAACCATTCCAATTTTCGCTACGAATTCCGACTGCGAGCTTATTGTTAAGGCAATGCAAGGTCT
CCTAAAAGATGGAAACCCGATTCCCTCAGCAATCGCAGCAAACTCCGGCCTCTACGGCGCAAT
GGCTTCTAACTTTACTCAGTTCGTTCTCGTCGACAATGGCGGAACTGGCGACGTGACTGTCGCC
CCAAGCAACTTCGCTAACGGGGTCGCTGAATGGATCAGCTCTAACTCGCGTTCACAGGCTTAC
AAAGTAACCTGTAGCGTTCGTCAGAGCTCTGCGCAGAAGCGCAAATACACCATCAAAGTCGAG
GTACCTAAAGTGGCAACCCAGACTGTTGGTGGTGTAGAGCTTCCTGTAGCCGCATGGCGTTCG
TACTTAAATATGGAACTAACCATTCCAATTTTCGCCACGAATTCCGACTGCGAGCTTATTGTTAA
GGCAATGCAAGGTCTCCTAAAAGATGGAAACCCGATTCCCTCAGCAATCGCAGCAAACTCCTT
AATTAAC<u>CATCATCACCACCATCATGACTACGATATACCCACAACCGCCAGCGAGAATCTGTAC</u>
<u>TTTCAAGGAGAGTTAGGGATGAGGGGTTCAGCTGGAAAGGCCGGTGAAGGTGAAATCCCTGC</u>
<u>CCCTCTTGCTGGTACCGTTTCTAAGATACTGGTAAAAGAAGGTGACACTGTTAAAGCTGGTCA</u>
<u>AACAGTTCTGGTGCTGGAGGCTATGAAAATGGAGACAGAAATTAACGCTCCTACTGACGGAA</u>
<u>AAGTTGAAAAGGTGTTAGTTAAGGAAAGAGATGCTGTTCAAGGTGGTCAAGGTCTAATCAAG</u>
*<u>ATCGGCGTTCATCATCACCACCATCAT</u>***<u>CTCGAGGAACGCATGGGTTTCACAGAGCTCCCTCTC</u>**
<u>ACTGGCAGCACCTTCCATGACCAGATAGCC</u>ATGCTGAGCCACCGGTGCTTCAACACTCTGACT
AACAGCTTCCAGCCCTCCTTGCTCGGCCGCAAGATTCTGGCCGCCATCATTATGAAAAAAGACT
CTGAGGACATGGGTGTCGTCGTCAGCTTGGGAACAGGGAATCGCTGTGTGAAAGGAGATTCT
CTCAGCCTAAAAGGAGAAACTGTCAATGACTGCCATGCAGAAATAATCTCCCGGAGAGGCTTC
ATCAGGTTTCTCTACAGTGAGTTAATGAAATACAACTCCCAGACTGCGAAGGATAGTATATTT
GAACCTGCTAAGGGAGGAGAAAAGCTCCAAATAAAAAAGACTGTGTCATTCCATCTGTATATC
AGCACTGCTCCGTGTGGAGATGGCGCCCTCTTTGACAAGTCCTGCAGCGACCGTGCTATGGAA
AGCACAGAATCCCGCCACTACCCTGTCTTCGAGAATCCCAAACAAGGAAAGCTCCGCACCAAG
GTGGAGAACGGAGAAGGCACAATCCCTGTGGAATCCAGTGACATTGTGCCTACGTGGGATGG
CATTCGGCTCGGGGAGAGACTCCGTACCATGTCCTGTAGTGACAAAATCCTACGCTGGAACGT
GCTGGGCCTGCAAGGGGCACTGTTGACCCACTTCCTGCAGCCCATTTATCTCAAATCTGTCACA

TTGGGTTACCTTTTCAGCCAAGGGCATCTGACCCGTGCTATTTGCTGTCGTGTGACAAGAGAT
GGGAGTGCATTTGAGGATGGACTACGACATCCCTTTATTGTCAACCACCCCAAGGTTGGCAGA
GTCAGCATATATGATTCCAAAAGGCAATCCGGGAAGACTAAGGAGACAAGCGTCAACTGGTG
TCTGGCTGATGGCTATGACCTGGAGATCCTGGACGGTACCAGAGGCACTGTGGATGGGCCAC
GGAATGAATTGTCCCGGGTCTCCAAAAAGAACATTTTTCTTCTATTTAAGAAGCTCTGCTCCTT
CCGTTACCGCAGGGATCTACTGAGACTCTCCTATGGTGAGGCCAAGAAAGCTGCCCGTGACTA
CGAGACGGCCAAGAACTACTTCAAAAAAGGCCTGAAGGATATGGGCTATGGGAACTGGATTA
GCAAACCCCAGGAGGAAAAGAACTTTTATCTCTGCCCAGTATAGTATGCTCCAGTCTAGA

2.2.2 Guiding ADAR1 to a specific target

It is important to control and direct the RNA editing enzyme to a specific target. For this purpose, I fused the MS2 protein upstream of the ADAR1 deaminase domain to produce MS2-ADAR1 chimeric protein from the same vector under the control of the CMV-IE 94 promoter. MS2-RNA expressed by another vector contained the upstream guide RNA. This MS2 RNA binds the MS2 protein and recruits the deaminase protein to the specific target. The optimum 21 bp guide RNA sequence complementary to the targeting sequence was inserted upstream (pCS2+-Guide-MS2 RNA) of the MS2 RNA (Figure 1. B) sequence by PCR using a proofreading polymerase (Pfu). The guide RNA sequence was 21 bp long with two mismatches: an A-to-C mismatch to facilitate deamination and an A-to-G mismatch to avoid off-target editing. As a target, I used mutated EGFP (X58, TAG) in pcDNA3-EGFP. Restriction enzymes *Eco*RI and *Xho*I were used for cloning. The resultant clones were confirmed by sequencing (Applied Biosystems 3130*xl* Genetic Analyzer, USA).

pCS2+-Guide-MS2 RNA: Guide RNA 21 base pair and MS2RNA 6x were inserted into pCS2+ vector (addgene)

Bold underlined text, EcoR1 and Xho1 restriction enzyme sites; bold italic underlined text, MS2RNA 6x; highlighted text, 21 base pair guide sequence.

CGCCATTCTGCCTGGGGACGTCGGAGCAAGCTTGATTTAGGTGACACTATA
GAATACAAGCTACTTGTTCTTTTTGCAGGATCCCATCGATTC**GAATTC*GAGG
GGGGGCCAGGGCACGGG*GAATGGCCATGGGACGTCGACCTGAGGTAATT
ATAACCCGGGCCCTATATATGGATCC*TAAGGTACCTAATTGCCTAGAAAAC
ATGAGGATCACCCATGTCTGCAGGTCGACTCTAGAAAACATGAGGATCACCC
ATGTCTGCAGTATTCCCGGGTTCATTAGATCCTAAGGTACCTAATTGCCTAGA
AAACATGAGGATCACCCATGTCTGCAGGTCGACTCTAGAAAACATGAGGATC
ACCCATGTCTGCAGTATTCCCGGGTTCATTAGATCCTAAGGTACCTAATTGCC
TAGAAAACATGAGGATCACCCATGTCTGCAGGTCGACTCCAGAAAACATGAG
GATCACCCATGTCTGCAGTATTCCCGGGTTCATT*AGATCTGCGCGCGATCG
ATATCAGCGCTTTAAATTTGCG**CTCGAG**CCTCTAGAACTATAGTGAGTCG
TATTACGTAGAT

2.2.3. Construction of target and reporter substrate

In this study, mutated EGFP was used as a target of the system. Mutations were introduced at the 58th amino acid (TGG to TAG or TAA, encoding amber or ochre stop codons, respectively) (Figure 1 C) using a site-directed mutagenesis kit. The primers used for this purpose were designed with PrimerX, an online utility for automated design of mutagenic primers: amber forward, AAGCTGCCCGTGCCCTAGCCCACCCTCGT; amber reverse, ACGAGGGTGGGCTAGGGCACGGGCAGCTT; ochre forward, AAGCTGCCCGTGCCCTGGCCCACCCTCGT; ochre reverse, ACGAGGGTGGGCCAGGGCACGGGCAGCTT. The presence of the stop codon at the 58th amino acid position completely abolished EGFP fluorescence.

2.2.4 Cell culture and transfection

About 3×10^5 cells/well were seeded in 12-well culture plates (Costar, Corning, USA), grown for 24 h to 50–70% confluence, and then subjected to transfection. Cell culture medium was D-MEM with high glucose (WAKO) supplemented with 10% fetal bovine serum (Invitrogen). Lipofectamine 3000 (Invitrogen) was used for transfections.

2.2.5 Cell observation

Cells were observed on a JuLi Smart fluorescence cell imager microscope (Figure 4) and a Keyence Biozero-800 fluorescence microscope under standard conditions during observation.

2.2.6 RNA extraction and cDNA synthesis

RNA was extracted from the cells using the Trizol reagent (Invitrogen). Briefly, culture medium was aspirated, and 500 µl of Trizol reagent was added to each well. Cells were lysed directly in the culture dish by pipetting up and down several times, and the lysate was transferred to 1.5 ml tubes. Chloroform (100 µl) was added, and the sample was shaken vigorously by hand for 15 s. After incubation for 2–3 min at room temperature, the sample was centrifuged at 12,000 × g for 15 min at 4°C. The aqueous supernatant was transferred to a new tube and mixed by inversion with 250 µl of 100% isopropanol. To accelerate precipitation of RNA, the mixture was incubated at room temperature for 10 min, and then centrifuged at 12,000 × g for 10 min at 4°C. The supernatant was removed, and the RNA pellet was washed with 0.5 ml of 75% ethanol, followed by centrifugation (7500 × g for 5 min at 4°C). The supernatant was discarded, and the RNA pellet was allowed to air dry for 5–10 min. The pellet was suspended in RNase-free water. After DNase treatment, an equal amount of RNA was used for cDNA synthesis using a gene-

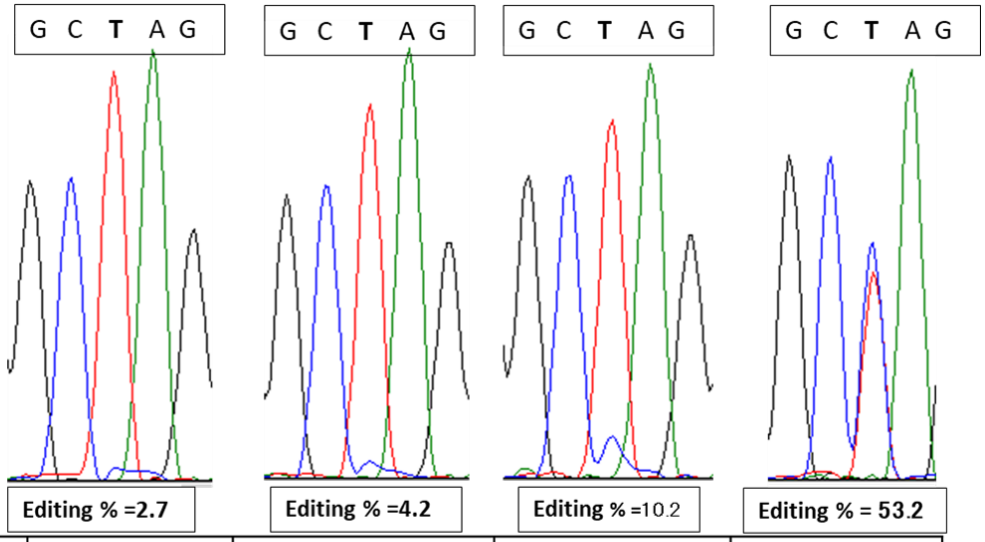
specific primer for EGFP and the Superscript III cDNA synthesis kit (Invitrogen). Every experiment was performed at least three times, yielding similar results.

2.2.7 Confirmation of restoration

PCR products were digested with a restriction enzyme that differentiated between the edited and unedited DNA sequences. PCR was performed using Go Taq polymerase (Promega). Thermal cycling conditions were as follows: 35 cycles of denaturation at 96°C for 2 min, annealing at 60°C for 0.5 min, and extension at 72°C for 0.5 min (GeneAmp PCR system 9700, Applied Biosystems, USA). The PCR products were ethanol-precipitated and run on an agarose gel. PCR bands of ~324 bp were cut out and purified using the Qiagen Gel Extraction kit. DNA was quantitated using an ND-1000 spectrophotometer. Equal amounts of DNA were incubated 37°C for 3 h with *Hae*III restriction enzyme, which digested the 324 bp PCR product into two fragments of 228 and 96 bp. In the mutated allele of EGFP, *Hae*III did not digest (5'…TAGCC…3') due to the presence of the mutated TAG codon at the 58th amino acid in place of the wild-type TGG (5'…TGGCC…3') codon. The digested DNA was run in 6% polyacrylamide gel, stained with SYBR green (Lonza), and imaged on a LAS-3000 Imager (Fujifilm) (Figure 5).

2.2.8 Bulk sequencing of PCR products for editing observation

Direct sequencing of PCR-amplified products was performed on an Applied Biosystems 3130*xl* Genetic Analyzer (Applied Biosystems, USA). PCR-amplified products were run on 6% polyacrylamide gels, and then stained with DNA gel stain solution. The bands were cut out, frozen, and homogenized by crushing. About 10 µl of 0.1× TE was added, and the sample was vortexed well.



EGFP /mutated EGFP%	2.5/97.5	5.0/95.0	10.0/90.0	50.0/50.0
Editing %	2.7	4.2	10.2	53.2

Figure 3. Optimization of sequencing output for the determination of editing efficiency. Optimizing the sequencing experiment. Mutated and wild type EGFP was mixed together and sequenced. Calculative editing percentage approximate to the ratio of plasmid mix.

After centrifugation at 14,000 rpm for 15 min, 2 µl of supernatant was subjected to a sequencing reaction using a reverse primer for EGFP. Raw sequence data were analyzed with the Seq Scanner 2 software (Applied Biosystems). When the edited and unedited products were present together, a dual peak C (edited) and T (unedited) was observed at the target site. Peak height was measured using the ImageJ software. Editing was quantified based on maximum peak height ratio of the edited and unedited products (100% × [C height/ (T height + C height)]).[43, 44] However, for validation by Sanger sequencing, plasmids containing antisense CCA (edited) and CTA (unedited) were mixed in various ratios (2.5%, 5%, 10%, 20%, and 50%) to observe the correlation of peak height with the concentration ratio (Figure 3). In most of the above mentioned treatments, I found that the difference between the calculated and experimental editing results was >10% (Figure 3).

2.2.9 Western blot analysis for detection of full-length functional protein

Western blotting to detect full-length functional EGFP and MS2-ADAR1 chimeric protein was performed as previously described[45] with GFP polyclonal antibody (GTX113617) and the Myc-tag (Cell Signaling), respectively, at a 1:5000 dilution. Cultured HEK-293 cells were washed twice with ice-cold PBS, and total protein was extracted with RIPA cell lysis buffer containing 1× proteinase inhibitor cocktail (Nacalai Tesque). Protein concentration was measured by Bradford assay. Equal amounts (30 µg) of whole protein extract were subjected to 12% SDS-PAGE at 100 V for 1.5 h. For the positive control, the EGFP sample was 30-fold more dilute than the other samples. Precision Plus Protein Dual-Color standards (Bio-Rad) were used as protein markers. Secondary antibody was used at 1:2500 dilution. Amersham ECL western blotting detection reagent (GE Healthcare) was used to develop the blots, and images (Figure 6 A and B) were acquired on an LAS-3000 Imager (Fujifilm).

2.3 Results

2.3.1 RNA restoration using an artificial enzyme system

To develop a genetic restoration tool for therapeutic use, I sought to restore expression of a mutated EGFP in the human cell line HEK-293. The EGFP allele used for this study harbored a mutation in the 58th amino acid of EGFP (TGG, encoding Trp) to a TAG stop codon, and consequently does not encode a protein that generates a fluorescent signal (Figure 4 c–d). When all three factors were provided of the system (amber-mutated EGFP, ADAR1, and guide RNA) to HEK-293 cells, fluorescence was detected, indicating that the TAG stop codon had been converted to a TGG (Trp) codon (Figure 4 i–j). The fluorescence signals were those of cell cytoplasmic EGFP, because the engineered

deaminase domain is expressed in and localized to the cytoplasm. However, when only two of the factors were provided, no detectable fluorescence was observed (Figure 4 e–h). When all three factors were present, the fluorescence signal was comparable to that of the EGFP positive control (Figure 4 K–I).

2.3.2 Confirmation of restoration

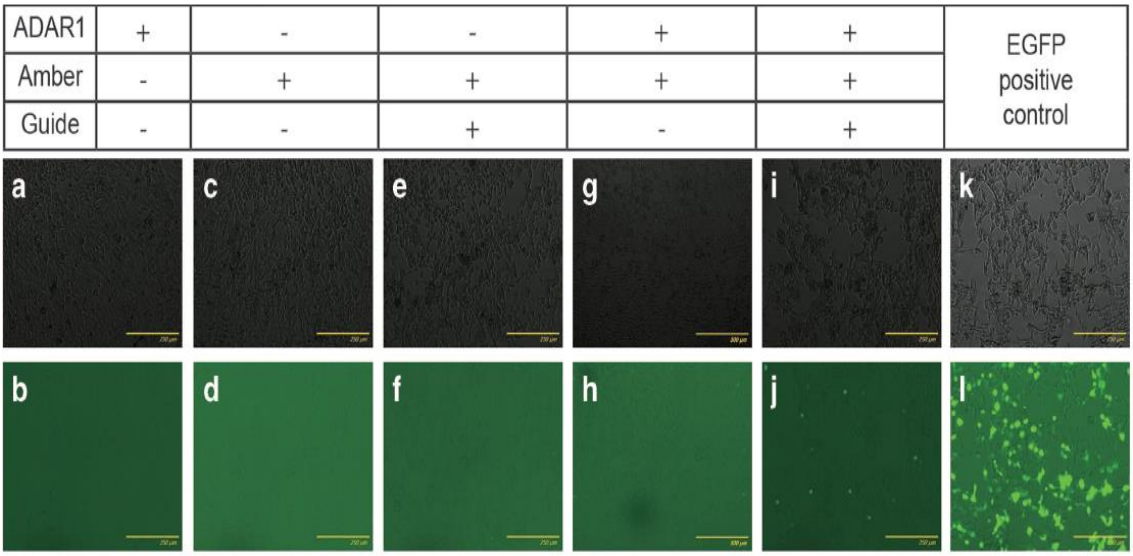


Figure 4. Cell imaging using a JuLi Smart fluorescence cell imager microscope after transfection with different factors. Three different factors are indicated above the figures. "+" indicates the presence of factors and "-" indicates the absence of factors during transfection. The upper row of images are bright-field images and the lower row of images are fluorescence images. Scale bar is 250 µm. Every experiment was performed at least three times with similar results.

Target specificity is an important consideration in the context of genetic restoration. The goal was to convert the amber (TAG) stop codon of EGFP to a TGG (Trp) codon in HEK-293 cells. To test specificity at the sequence level, I performed RT-PCR followed by RFLP analysis of EGFP using the *Hae*III restriction enzyme, which only digests the 5'...TGGCC…3' sequence. I performed restriction enzyme digestion of RT-PCR-amplified products from experimental and control cells to confirm the specificity of the system for the target. Only converted EGFP should be cut by *HaeIII* (Figure 5, lane 3,

positive control) into two bands of 228 and 96 bp, whereas the amber-mutated EGFP sequence (5'…TAGCC…3') should not be cut. When all three factors of the system were

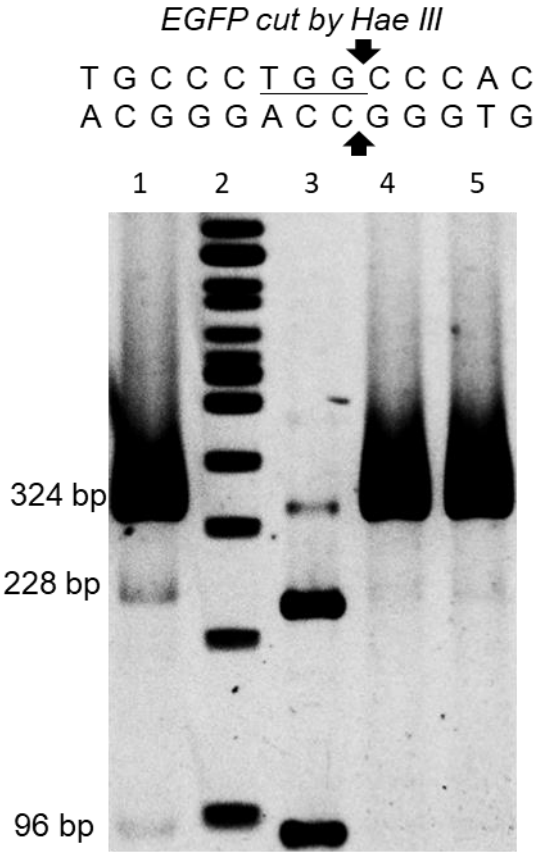


Figure 5. Restriction fragment length polymorphism (RFLP) analysis of amber conversion to EGFP. Purified PCR products (324 bp) were cut with *Hae III* restriction enzyme, which should result in restriction fragments of 228 bp and 96 bp. Lane 1, MS2-ADAR1+Amber +Guide; lane 2, 100 bp marker; lane 3, EGFP; Lane 4, Amber + Guide; lane 5, MS2-ADAR1 + Amber.

present (Figure 5, lane 1, experimental), we observed the 228 and 96 bp bands, reflecting conversion of the amber (TAG) stop codon to the read-through (TGG) codon. In lanes 4 and 5, no *Hae*III-digested bands were observed, confirming that neither guide RNA nor ADAR1 alone can perform the G-to-A conversion.

2.3.3 Determination of RNA editing efficiency

I calculated the editing efficiency at the RNA level by direct Sanger sequencing of PCR-amplified products. To optimize the sequencing assay, I mixed various concentrations of mutated and wild-type EGFP. The proportion of mixing and analytical value were found approximately similar to each other (Figure 3).

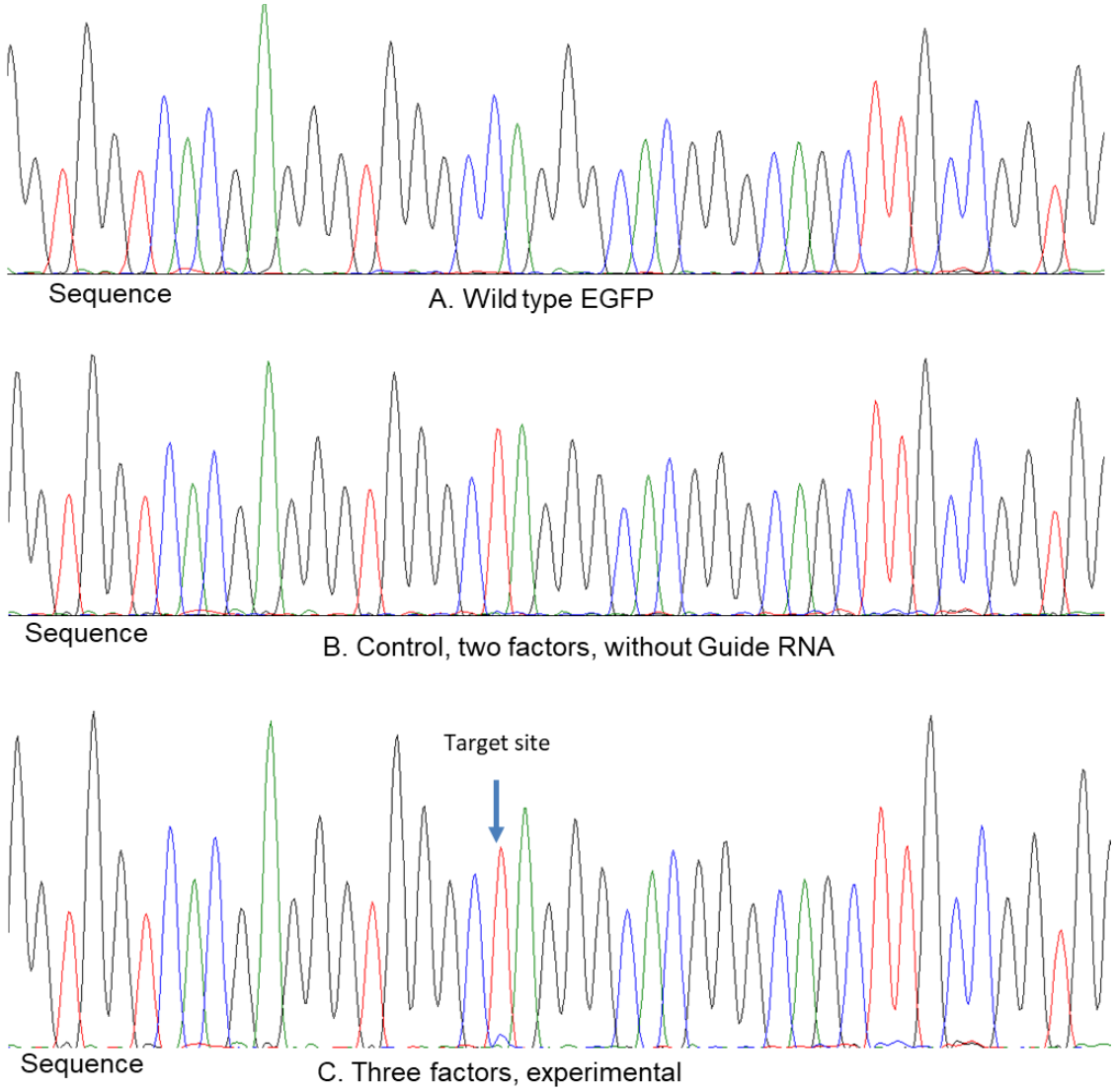


Figure 6. Antisense sequencing result of the targeted EGFP.

To quantify RNA editing, the samples from Figure 4 g–l were sequenced using reverse primers. Dual peaks from unedited and edited RNAs were only detected when all three factors were applied to the cells (Figure 6 C). The editing efficiency of the system was approximately 5% at the RNA level, and we observed no off-target changes within the region complementary to the guide RNA or adjacent adenosine nucleotides (Figure 6 C).

2.3.4 Western blot analysis for detection of full-length functional protein

To determine whether the fluorescence signal was derived from full-length EGFP or another source (i.e., to determine whether the system is capable of producing full-length functional protein), I performed western blot analysis using polyclonal GFP antibody. When all three factors were transfected, I could detect full-length GFP of 29 KDa (Figure 7 A, lane 1). No EGFP band was present when only the mutated EGFP and guide RNA (Figure 7 A, lane 3), mutated EGFP alone, or ADAR1 alone (Figure 7 A, lanes 4 and 5) was present. However, a weak signal was detected when ADAR1 was provided along with the amber mutant allele of EGFP (Figure 7 A, lane 2).

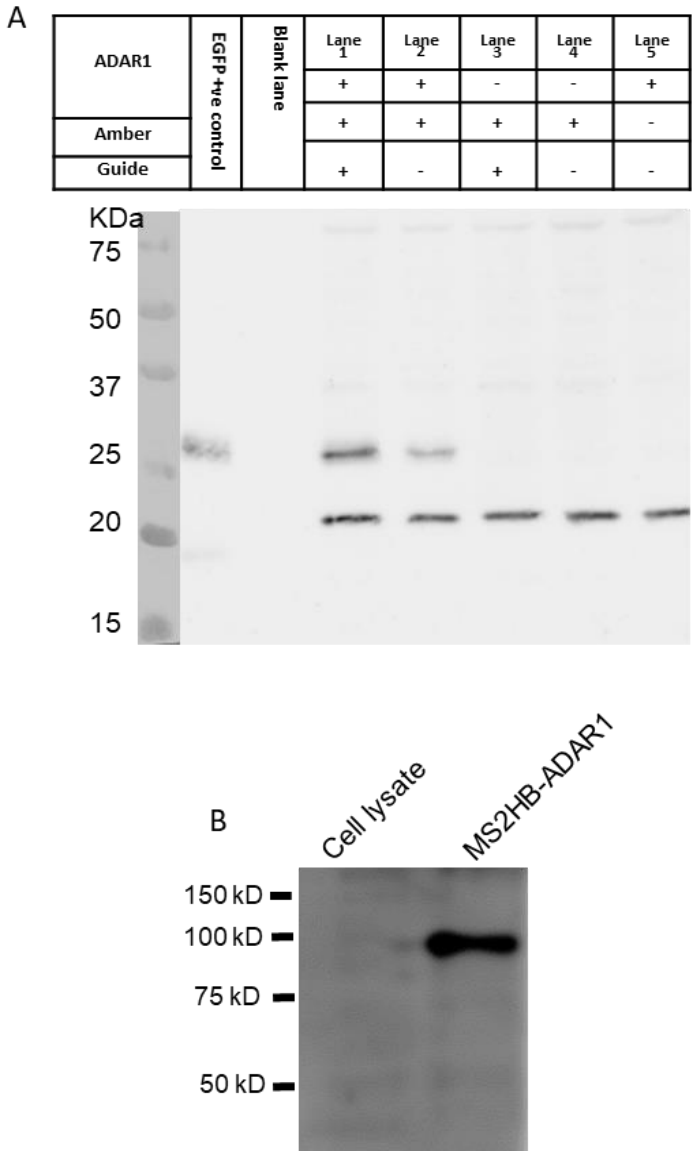


Figure 7. **A.** The genetically engineered MS2-ADAR1 enzyme system can convert a stop codon TAG to TGG and produce a full length functional protein. All lanes contained the same amount of protein except for lane EGFP +ve control containing 29 KDa EGFP, which was diluted 30-fold more than the materials in the other lanes and used as positive control. A protein of 20 KDa was observed as a background signal in all samples even without amber (lane 5). All experiments were performed at least three times with similar result. **B.** Western blot detection of an approximately 93 KDa Myc-tag MS2HB-ADAR1 protein from pCS2+MT- MS2HB-ADAR1 transfected HEK 293 cells.

2.3.5 Restoration of double mutation ochre stop codon

In the experiments described above, we confirmed that the enzyme system could recode an amber (TAG) stop codon to TGG in EGFP. We next sought to determine whether we could recode another stop codon, ochre (TAA). Because of the double A-to-G corrections for the ochre codon, it is more difficult to restore it than the amber codon. According to the fluorescence images, wild-type EGFP generated a high level of fluorescence (Figure 8 k–l), whereas the allele harboring the ochre (TAA) stop codon did not. ADAR1 and guide RNA alone had no effect on the stop codon, and consequently yielded no fluorescence (Figure 8 a–b and c–d). However, upon addition of the guide RNA along with ADAR1, the TAA codon was restored to TGG, as demonstrated by the fluorescence signal (Figure 8 i–j). Further confirmation was provided by PCR followed by RFLP analysis. Based on the fluorescence and PCR-RFLP results, we conclude that the TAA codon was restored to TGG.

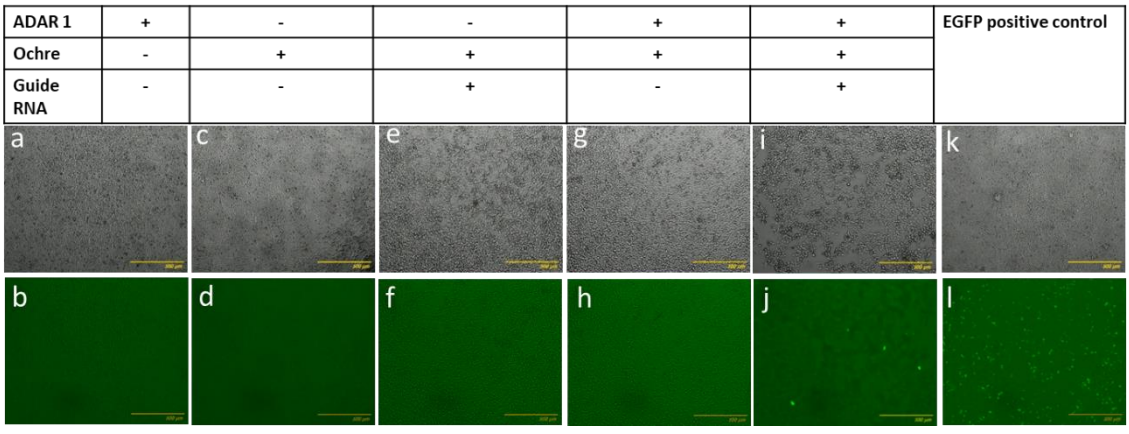


Figure 8. Cell imaging using a JuLi Smart fluorescence cell imager microscope after transfection with different factors. Three different factors are indicated above the figures. "+" indicates the presence of factors and "–" indicates the absence of factors during transfection. The upper row of images are bright-field images and the lower row of images are fluorescence images. Scale bar is 500 µm.

2.3.6 Detection of cytosolic localization of the enzyme system

From Confocal microscopy image (Figure 9) it is found that the artificial enzyme system is localized in the cell cytoplasm and convert mutated EGFP to wild type fluorescence.

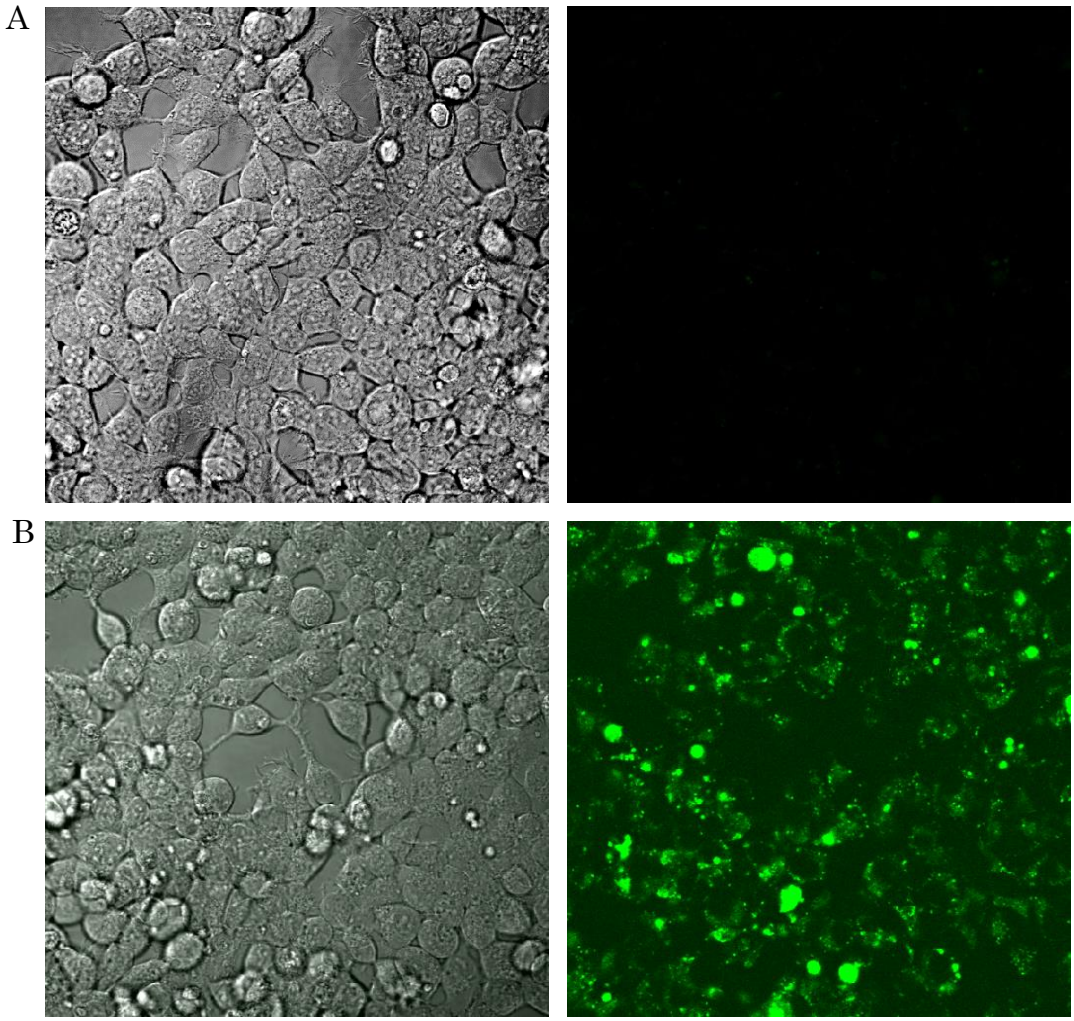


Figure 9. Confocal microscopy (The Olympus FLUOVIEW FV1000 confocal laser scanning microscope images). Row **A.** with two factors, row **B.** EGFP fluorescence signal from cell cytoplasm when three factors applied. Left column bright and right column fluorescence image.

2.4 Discussion

In this study, I investigated whether the deaminase domain of ADAR1 would be functional in a biological approach to A-to-I RNA editing. I also sought to determine whether the MS2 system is compatible with designing an RNA editing enzyme model. The experimental data suggest that only the deaminase domain of ADAR1, lacking the double-stranded RNA-binding domain (dsRBD), can be made competent for RNA editing by adding an artificial guide RNA. Because the deaminase domain of ADAR1 is devoid of a nuclear localization signal (NLS) and a nuclear export signal (NES)[46,47], the engineered enzyme system is localized to the cytoplasm and easily trapped by the targeted mRNA in the cell cytoplasm. Moreover, the guide RNA which is a component of the artificial enzyme and EGFP mRNAs which are substrates of the artificial enzyme should be present in the cytosol (Figure 9). The use of only the deaminase domain decreases the packaging and gene delivery load for adenovirus or other systems that might be used to target this system to specific tissues.

Previous work showed that ADAR2 without the dsRBDs is functional and can perform A-to-I conversion to achieve RNA editing *in vitro* and *in vivo*.[37] The authors of that study used the lambda N system to control the site-specificity of their RNA editing enzyme. In this study, I used the MS2 system instead of lambda N; the former is more frequently used in RNA studies.[41] When guide RNA is chemically attached to the deaminase domain of ADAR1 and ADAR2, the enzymes exhibit deaminase activity *in vitro*, in cell culture, and in annelids.[35,38] Natural cis-acting R/G-site was designed to trans-acting guide RNA for wild-type ADAR2 including dsRBDs. In this case, the deaminase domain was located 3´ of the R/G-site.[35] In the case of a guide RNA with the ADAR1 deaminase domain alone, the protein was also positioned at the 3´ end of the

guide apparatus. In HEK-293 cells, we were able to convert a TAG and TAA stop codon to a TGG (Trp) read-through codon. Conversion was detected by fluorescence microscopy (Figure 3). Further confirmation of conversion of the amber mutation was provided by RT-PCR followed by RFLP analysis (Figure 5). For this purpose, I used the *Hae*III restriction enzyme, which only digested the converted (or wild-type) EGFP (5'...TGGCC…3'). This RFLP assay confirmed the specificity of the editing enzyme system, as well as the editing location.

In addition, I performed Sanger sequencing of the PCR-amplified EGFP products to observe the peak height at the targeted adenosine position. Quantification of editing at the RNA level is performed by antisense sequencing, as it provides a more accurate peak height ratio.[48] In antisense sequencing of the PCR products, I observed a dual peak only at the target site. By contrast, in the adenosine adjacent to the target, I did not observe any off-target editing (Figure 6C), confirming the specificity of the engineered system. The use of an A-to-G mismatch during guide RNA design helps to avoid off-target editing (Figure 9).[49]

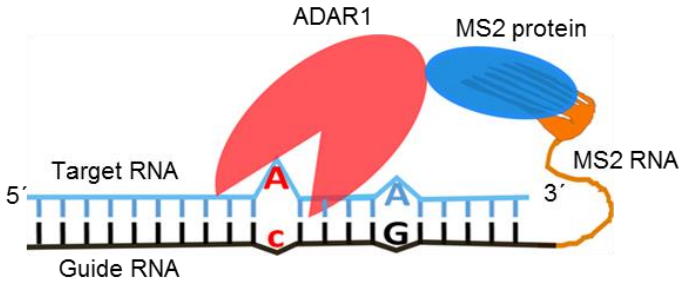


Figure 9. Experimental assembly of engineered ADAR1 (red), MS2 protein (green), and MS2 RNA (yellow) connected to the guide RNA. In the target RNA, the targeted adenosine (red) will be converted to inosine, which will be read as guanosine during translation. The targeted adenosine mismatches with cytosine and is flipped out, resulting in easy deamination.

I also conducted western blot analysis to detect functional EGFP, in order to determine whether the system was capable of producing full-length protein. When all three components of the system were introduced to HEK-293 cells and incubated for 48 h, mutated EGFP was converted to EGFP. By contrast, no EGFP signals were detected when only mutated EGFP or mutated EGFP and guide RNA (but no enzyme) were provided, indicating that other cellular factors were not involved in editing. However, when the cells were treated with ADAR1 and the amber-mutated allele of EGFP, I observed a weak EGFP signal. This effect could be attributed to the effects of complex secondary structure in the EGFP mRNA, or to inappropriate proportions of enzyme and substrate inside the cells. In the future, it will be necessary to perform further tuning of the system in terms of guide RNA, rotational opportunity of MS2 system with the enzyme, and the relative proportions of the factors. Previous work showed that proper concentrations of enzyme and reporter substrate can eliminate minor autoediting.[38]

Efficiency is a critical consideration in biological RNA editing. Several opportunities exist to tune the efficiency of the system reported here. Altering the length and multiplicity of the guide or increasing the MS2 RNA loop from 6× to 12× may make the system more effective and efficient. In addition, optimization of the binding between guide and target could increase the efficiency of RNA editing by this system. Standardization of the concentrations of different factors is also important for achieving the desired output.

In conclusion, ADAR1 can be used for RNA editing with the MS2 system. Our guide RNA is very simple, and I am confident that this type of guide RNA design, in conjunction with the MS2 system, will be applicable to specific targeting of a variety of

other active enzymes. This system could be used to restore point-mutated RNA in patients with various diseases, thereby mitigating disease symptoms.

2.5 References

1. Kim H, Kim J-S. A guide to genome engineering with programmable nucleases. *Nat Rev Genet* 2014; **15**(5)**:** 321-334.

2. Doudna J. Perspective: Embryo editing needs scrutiny. *Nature* 2015; **528**(7580)**:** 230-236.

3. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. *Science* 2012; **337**(6096)**:** 816-821.

4. Leisegang M, Engels B, Schreiber K, Yew PY, Kiyotani K, Idel C *et al.* Eradication of large solid tumors by gene therapy with a T-cell receptor targeting a single cancer-specific point mutation. *Clin Cancer Res* 2016; **22**(11)**:** 2734-2743.

5. Ferec C, Audrezet MP, Mercier B, Guillermit H, Moullier P, Quere I *et al.* Detection of over 98% cystic fibrosis mutations in a Celtic population. *Nat Genet* 1992; **1**(3)**:** 188-1891.

6. Bolscher BG, de Boer M, de Klein A, Weening RS, Roos D. Point mutations in the beta-subunit of cytochrome b558 leading to X-linked chronic granulomatous disease. *Blood* 1991; **77**(11)**:** 2482-2487.

7. Higuchi M, Kazazian HH, Kasch L, Warren TC, McGinniss MJ, Phillips JA *et al.* Molecular characterization of severe hemophilia A suggests that about half the

mutations are not within the coding regions and splice junctions of the factor VIII gene. *Proc Natl Acad Sci* 1991; **88**(16)**:** 7405-7409.

8. Maas S, Rich A. Changing genetic information through RNA editing. *BioEssays* 2000; **22**(9)**:** 790-802.

9. Nishikura K. Functions and regulation of RNA editing by ADAR deaminases. *Annu Rev Biochem* 2010; **79:** 321-349.

10. Nishikura K. A-to-I editing of coding and non-coding RNAs by ADARs. *Nat Rev Mol Cell Biol* 2016; **17**(2)**:** 83-96.

11. Slotkin W, Nishikura K. Adenosine-to-inosine RNA editing and human disease. *Genome Med* 2013; **5**(11)**:** 1-13.

12. Rosenberg BR, Hamilton CE, Mwangi MM, Dewell S, Papavasiliou FN. Transcriptome-wide sequencing reveals numerous APOBEC1 mRNA-editing targets in transcript 3′ UTRs. *Nat Struct Mol Biol* 2011; **18**(2)**:** 230-236.

13. Zabner J, Couture LA, Gregory RJ, Graham SM, Smith AE, Welsh MJ. Adenovirus-mediated gene transfer transiently corrects the chloride transport defect in nasal epithelia of patients with cystic fibrosis. *Cell* 1993; **75**(2)**:** 207-216.

14. Davies JC, Alton EW. Gene therapy for cystic fibrosis. *Proc Am Thorac Soc* 2010; **7**(6)**:** 408-14.

15. Blow MJ, Grocock RJ, van Dongen S, Enright AJ, Dicks E, Futreal PA *et al.* RNA editing of human microRNAs. *Genome Biol* 2006; **7**(4)**:** 27.1-27.8.

16. Kawahara Y, Zinshteyn B, Sethupathy P, Iizasa H, Hatzigeorgiou AG, Nishikura K. Redirection of silencing targets by adenosine-to-inosine editing of miRNAs. *Science* 2007; **315**(5815)**:** 1137-1140.

17. Kim U, Wang Y, Sanford T, Zeng Y, Nishikura K. Molecular cloning of cDNA for double-stranded RNA adenosine deaminase, a candidate enzyme for nuclear RNA editing. *Proc Natl Acad Sci* 1994; **91**(24)**:** 11457-11461.

18. Melcher T, Maas S, Herb A, Sprengel R, Seeburg PH, Higuchi M. A mammalian RNA editing enzyme. *Nature* 1996; **379**(6564)**:** 460-464.

19. Chen CX, Cho DS, Wang Q, Lai F, Carter KC, Nishikura K. A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single and double-stranded RNA binding domains. *RNA* 2000; **6**(5)**:** 755-767.

20. Stefl R, Xu M, Skrisovska L, Emeson RB, Allain FH. Structure and specific RNA binding of ADAR2 double-stranded RNA binding motifs. *Structure* 2006; **14**(2)**:** 345-355.

21. Herbert A, Alfken J, Kim YG, Mian IS, Nishikura K, Rich A. A Z-DNA binding domain present in the human editing enzyme, double-stranded RNA adenosine deaminase. *Proc Natl Acad Sci* 1997; **94**(16)**:** 8421-8426.

22. Cho DSC, Yang WD, Lee JT, Shiekhattar R, Murray JM, Nishikura K. Requirement of dimerization for RNA editing activity of adenosine deaminases acting on RNA. *J Biol Chem* 2003; **278**(19)**:** 17093-17102.

23. Valente L, Nishikura K. RNA binding-independent dimerization of adenosine deaminases acting on RNA and dominant negative effects of nonfunctional subunits on dimer functions. *J Biol Chem* 2007; **282**(22)**:** 16054-16061.

24. Ota H, Sakurai M, Gupta R, Valente L, Wulff BE, Ariyoshi K *et al.* ADAR1 forms a complex with dicer to promote microRNA processing and RNA-induced gene silencing. *Cell* 2013; **153**(3)**:** 575-589.

25. Chen L, Li Y, Lin CH, Chan TH, Chow RK, Song Y *et al.* Recoding RNA editing of AZIN1 predisposes to hepatocellular carcinoma. *Nat Med* 2013; **19**(2)**:** 209-216.

26. Shimokawa T, Rahman MF, Tostar U, Sonkoly E, Stahle M, Pivarcsi A *et al.* RNA editing of the GLI1 transcription factor modulates the output of Hedgehog signaling. *RNA Biol* 2013; **10**(2)**:** 321-333.

27. Paz-Yaacov N, Bazak L, Buchumenski I, Porath HT, Danan-Gotthold M, Knisbacher BA *et al.* Elevated RNA editing activity is a major contributor to transcriptomic diversity in tumors. *Cell Rep* 2015; **13**(2)**:** 267-276.

28. Han L, Diao L, Yu S, Xu X, Li J, Zhang R *et al.* The genomic landscape and clinical relevance of A-to-I RNA editing in human cancers. *Cancer Cell* 2015; **28**(4)**:** 515-528.

29. Morabito MV, Abbas AI, Hood JL, Kesterson RA, Jacobs MM, Kump DS *et al.* Mice with altered serotonin 2C receptor RNA editing display characteristics of Prader–Willi syndrome. *Neurobiol Dis* 2010; **39**(2)**:** 169-180.

30. Silberberg G, Lundin D, Navon R, Ohman M. Deregulation of the A-to-I RNA editing mechanism in psychiatric disorders. *Hum Mol Genet* 2012; **21**(2)**:** 311-321.

31. Hartner JC, Schmittwolf C, Kispert A, Muller AM, Higuchi M, Seeburg PH. Liver disintegration in the mouse embryo caused by deficiency in the RNA-editing enzyme ADAR1. *J Biol Chem* 2004; **279**(6)**:** 4894-4902.

32. Wang Q, Miyakoda M, Yang W, Khillan J, Stachura DL, Weiss MJ *et al.* Stress-induced apoptosis associated with null mutation of ADAR1 RNA editing deaminase gene. *J Biol Chem* 2004; **279**(6)**:** 4952-4961.

33. Zhu C, Zhu K, Zhou Y, Fan Y. A novel insertion mutation in the ADAR1 gene of a Chinese family with dyschromatosis symmetrica hereditaria. *Genet Mol Res* 2013; **12**(3)**:** 2858-2862.

34. Vogel P, Stafforst T. Site-directed RNA editing with antagomir deaminases-a tool to study protein and RNA function. *Chem Med Chem* 2014; **9**(9)**:** 2021-2025.

35. Wettengel J, Reautschnig P, Geisler S, Kahle PJ, Stafforst T. Harnessing human ADAR2 for RNA repair–recoding a PINK1 mutation rescues mitophagy. *Nucleic Acids Res* 2016**; 45:** 1-12.

36. Woolf TM, Chase JM, Stinchcomb DT. Toward the therapeutic editing of mutated RNA sequences. *Proc Natl Acad Sci* 1995; **92**(18)**:** 8298-8302.

37. Montiel-Gonzalez MF, Vallecillo-Viejo I, Yudowski GA, Rosenthal JJ. Correction of mutations within the cystic fibrosis transmembrane conductance

regulator by site-directed RNA editing. *Proc Natl Acad Sci* 2013; **110**(45)**:** 18285-18290.

38. Hanswillemenke A, Kuzdere T, Vogel P, Jekely G, Stafforst T. Site-directed RNA editing *in vivo* can be triggered by the light-driven assembly of an artificial riboprotein. *J Am Chem Soc* 2015; **137**(50)**:** 15875-15881.

39. Vu LT, Nguyen TTK, Thoufic AAM, Suzuki H, Tsukahara T. Chemical RNA editing for genetic restoration: the relationship between the structure and deamination efficiency of carboxyvinyldeoxyuridine oligodeoxynucleotides. *Chem Biol Drug Des* 2016; **87**(4)**:** 583-593.

40. Vu LT, Nguyen TTK, Alam S, Sakamoto T, Fujimoto K, Suzuki H *et al.* Changing blue fluorescent protein to green fluorescent protein using chemical RNA editing as a novel strategy in genetic restoration. *Chem Biol Drug Des* 2015; **86**(5)**:** 1242-1252.

41. Keryer-Bibens C, Barreau C, Osborne HB. Tethering of proteins to RNAs by bacteriophage proteins. *Biol Cell* 2008; **100**(2)**:** 125-138.

42. Buxbaum AR, Haimovich G, Singer RH. In the right place at the right time: visualizing and understanding mRNA localization. *Nat Rev Mol Cell Biol* 2015; **16**(2)**:** 95-109.

43. Eggington JM, Greene T, Bass BL. Predicting sites of ADAR editing in double-stranded RNA. *Nat Commun* 2011; **2:** 1-9.

44. Rinkevich FD, Schweitzer PA, Scott JG. Antisense sequencing improves the accuracy and precision of A-to-I editing measurements using the peak height ratio method. *BMC Res Notes* 2012; **5**(1)**:** 1-6.

45. Hakim NH, Kounishi T, Alam AH, Tsukahara T, Suzuki H. Alternative splicing of Mef2c promoted by Fox-1 during neural differentiation in P19 cells. *Genes to cells : devoted to molecular & cellular mechanisms* 2010; **15:** 255-67.

46. Desterro JMP, Keegan LR, Lafarga M, Berciano MT, O'Connell M, Carmo-Fonseca M. Dynamic association of RNA-editing enzymes with the nucleolus. *J Cell Sci* 2003; **116**(9)**:** 1805-1818.

47. Strehblow A, Hallegger M, Jantsch MF. Nucleocytoplasmic distribution of human RNA-editing enzyme ADAR1 is modulated by double-stranded RNA-binding domains, a leucine-rich export signal, and a putative dimerization domain. *Mol Biol Cell* 2002; **13**(11)**:** 3822-3835.

48. Nurpeisov V, Hurwitz SJ, Sharma PL. Fluorescent dye terminator sequencing methods for quantitative determination of replication fitness of human immunodeficiency virus type 1 containing the codon 74 and 184 mutations in reverse transcriptase. *J Clin Microbiol* 2003; **41**(7)**:** 3306-3311.

49. Schneider MF, Wettengel J, Hoffmann PC, Stafforst T. Optimal guideRNAs for re-directing deaminase activity of hADAR1 and hADAR2 in trans. *Nucleic Acids Res* 2014; **42**(10)**:** 1-9.

Chapter III

Comparative activity of adenosine deaminase acting on RNA (ADARs) isoforms for correction of genetic code in gene therapy

3.1 Introduction

Genetic manipulation mainly encompasses gene knockout and knockdown[1,2] but recently codon level genetic manipulation has been drawing much attention for its application in therapeutic approach. Most of these approaches are focused on DNA level but theoretically, manipulation can be performed at any stage up to translation.[3] RNA editing belongs to the post-transcriptional modification resulting in some distinct changes to targeted nucleotide sequences within the RNA molecule. This may include insertion, deletion[4] or substitution of nucleotides.[5,6] There are several types of RNA editing such as A-to-I, C-to-U, U-to-C and so on. A-to-I RNA editing is the most important and widely observed type of RNA editing in higher eukaryotes[7,8] caused by ADARs. ADARs have three types; ADAR1, ADAR2, and ADAR3 are classified depending on their period of discovery. ADAR1 and ADAR2 are widely expressed in almost all tissues,[9,10] and functional,[11] whereas ADAR3 expresses mainly in the brain.[12] ADAR3 cannot deaminate any targeted adenosine.[13,12,12,14,15] Natural ADAR enzymes consist of two components; the first one is responsible for recognition of the substrate, which includes the ds RNA binding domains, while the other is the functional deaminase domain that has the adenosine modifying activity.[13] Adenosine deaminases mainly deaminate the adenosines within dsRNAs, however, dsRNA substrate are also measles virus genomes become

hypermutated[16] as well as the hepatitis delta virus antigenome in the host cell containing a stop codon turn on to read through codon coding as tryptophan[17] due to the deaminase activity of adenosines.

Site-directed RNA editing has a great impact on tuning protein function as it can edit 12 out of 20 amino acids.[14] Not only stop codons but also splice sites and splice modulating elements can be altered by RNA editing.[18] RNA editing also has a potential therapeutic prospect. ADAR2-DD are organized by a wide variety of technologies to make the system more target specific.[19,14,20] In previous studies, it has been reported that full-length ADAR1 and ADAR2 have several isoforms with varying activity.[21,22] Side by side, only the deaminase domains of the ADARs have also been reported as functional.[19,23,24] Therefore, I choose ADAR1-DD and ADAR2-DD isoforms for their comparative deaminase activity. Several mutations have been corrected by *in vitro* trial and even in *in vivo* studies. CFTR mutated gene was corrected in *Xenopus oocytes* by lambda N system.[19] Parkinson's disease-related cause PINK1/Parkin-mediated mitophagy is also corrected in the cell line.[20] In our previous study, we used a chemical modification approach for RNA editing.[25,26] Recently, I modified the deaminase domain of ADAR1 with the MS2 system for correction of amber (TAG) and ochre (TAA) stop codons to readable codons.[27] The MS2 system is widely used in molecular biology[28] for tagging

GFP and in the study of RNA and protein localization[29] *in vivo*. However, in most of the previous studies, either the ADAR1 or ADAR2 was used but the aim of this work to compare the deaminase activity of only the deaminase domain of these two isoforms. The use of only the DD will probably result in the study is more accurate and conclusive as the deaminase domains of ADARs are more conserved than other parts. Comparatively, the deaminase activity is also affected by the dsRBD. Use of the DD only as a functional unit for genetic restoration will make deaminase compatible for gene delivery in future.

3.2 Materials and Methods

3.2.1 Preparation of the plasmid construct

Preparation of the amber stop codon (TAG) containing plasmid was done by the site-directed mutagenesis in the pcDNA-EGFP (Addgene) having the TGG codon at the 189 nt (nucleotide) position. The mutagenesis was done in the TGG to convert it into the TAG (Amber stop codon). The process of site-directed mutagenesis was done using the Quick Change II Site-Directed Mutagenesis Kit (Agilent Technologies, Catalog No. 200523). Along with the kit, forward and reverse primer were also necessary. Forward primer: AAGCTGCCCGTGCCCTAGCCCACCCTCGT, Reverse primer: ACGAGGGTGGGCTAGGGCACGGGCAGCTT.

3.2.2 Construction of ADARs-DD and guide RNA plasmid

MS2 protein was inserted into pCS2+MT vector plasmid (Addgene) within NcoI-XhoI restriction site. Then, hADARs (Human ADARs-DD) were inserted into the same vector at the XhoI-XbaI restriction site to form fusion protein MS2-DD. The plasmid was extracted from the transformed DH5∝ (Takara) using the QIAGEN spin prep kit. Restriction digestion of the extracted plasmid DNA was done by using corresponding restriction enzymes. For confirmation of restriction digestion, the inserted portion in the plasmid was sequenced using the Applied Biosystem genetic analyzer (3130xl Sequence Analyzer). However, in the case of ADAR2 long-DD, the whole plasmid was sequenced using the 3130xl Sequence analyzer (primer list and sequence given in the supportive information). For the preparation of the guide RNA, 21 nucleotides complementary to the target site was inserted either upstream or downstream of MS2-RNA, with either forward or reverse primer (pCS2+Guide-MS2-RNA-Guide). Then cloned products were transformed into DH5∝ and extraction of the plasmid DNA was done using the QIAGEN spin prep kit and the confirmation of the length was done by performing Gel electrophoresis and the final confirmation was done by the sequence of the bases using the 3130xl Sequence analyzer.

3.2.3 Determination of binding efficiency of MS protein and MS2 RNA

pCS2+-Guide-MS2RNA was linearized by restriction digestion with XhoI (CTCGAG) (Takara). Sp6 promoter (ATTTAGGTGACACTATAGAA) located just upstream of our target was used as a promoter for RNA synthesis. Briefly, restriction digestion of pCS2-Guide-MS2RNA and pCS2-MS2RNA-Guide plasmid by XhoI (Takara) then ethanol precipitation was done to purify the samples. The reaction mixture for RNA synthesis by MEGAscript SP6 was as follows- 20 µl reaction as manufacturer recommendation and incubation at 37°C for 5-6 hours followed by DNase (Promega) treatment. MS2 RNA purification by Phenol-chloroform and MS2 RNA quantification by ND-1000 spectrophotometer. An ssRNA band of Approximately 550 bp was checked by denaturing polyacrylamide gel electrophoresis. MS2 protein preparation from pCS2+MT-MS2HB-ADAR1-DD. MS2 protein and MS2RNA binding efficiency was determination by Biacore ^{TM} X100.

3.2.4 Cell culture transfection with Lipofectamine 3000

HEK (Human Embryonic Kidney) 293 cells was cultured in 24 well plates (Costar, Corning, USA) for up to 80-100% confluency by using D-MEM (Dulbecco's Modified Eagle's medium, WAKO) with high glucose supplemented with 10% fetal bovine serum (Invitrogen) and was kept inside the incubator at 37^0C and 5% CO_2 for 24 hours.

Transfection was done in cultured HEK 293 cell line by using the Lipofectamine 3000 (Invitrogen), following instruction of manufacturer. Approximately $2X10^5$ cells/well counted by hemocytometer was cultured in the 24 well plate. After 24 hours of incubation, the cells were observed under a microscope for transfection. The confluency of the cell was about 70%-80%. All the three factors were transfected (ADAR1+Amber+Guide RNA) using OptiMEM (Gibco) and Lipofectamine 3000 (Invitrogen), according to the protocol of the manufacturer. After 6 hours of transfection and incubation, the medium was changed to growth medium D-MEM with high glucose (WAKO) supplemented with 10% fetal bovine serum (Invitrogen) for maximization of the growth of the cells. Again kept in the incubator for incubation. Following 24 hours of incubation from the transfection, images were taken using the JuLi Smart fluorescence cell imager microscope and keyance biozero 800 fluorescence microscope.

3.2.5 Cellular observation for GFP expression

For evaluating GFP expression JuLi Smart fluorescence cell imager microscope, the Keyance Biozero 800 fluorescence microscope were used under optimized conditions during the observation.

3.2.6 RNA extraction and cDNA synthesis

RNA from the cells was extracted by Trizol (Invitrogen). Briefly, after aspirating the growth media cells were washed gently with ice-cold PBS. For 24 well plates, about 200 µl Trizol reagent (Invitrogen) was added. After mixing properly, the solution was taken to 1.5 ml tube and about 50 µl chloroform was added and mixed gently by inverting 4-5 times. After centrifugation at 12,000 rpm for 10 minutes, the clear supernatant was transferred to new tubes containing 100 µl 100% isopropanol. Again, after centrifugation at 12,000 rpm for 10 minutes, the supernatant was discarded. About 300 µl of 70% ethanol was added and centrifuged for 10 minutes at the same speed. The supernatant was then carefully discarded and the tubes allowed to air dry for 5 minutes. The samples were treated with DNase (Promega, USA) according to manufacturer`s instructions. Total RNA concentration was measured by ND-1000 spectrophotometer and an equal amount of RNA was used for cDNA synthesis by Superscript III (Life Technologies), according to manufacturer`s instruction, by the gene-specific primer of EGFP.

3.2.7 Western blot analysis for detection of full-length functional protein and ADARs-DD deaminase efficiency

Experimental cells were cultured in 24 well plate. For western blot analysis of the targeted EGFP, the specific polyclonal GFP antibody (GTX113617) was used at a 1:5000 dilution. The media was aspirated and the well containing cells were washed gently with ice-cold PBS. After that, the cells were scraped by cell scraper and was taken in corresponding tubes. Cell wells were washed again and transferred to their corresponding tubes as well. Total protein was extracted with RIPA cell lysis buffer containing 1X proteinase inhibitor cocktail (Nacalai Tesque). Protein concentration was measured by Bradford assay and an equal amount of protein was loaded to run in 12% SDS-PAGE at 100 V for 1.5 h. Precision Plus Protein Dual-Color standards (Bio-Rad) were used as protein markers. The semi-dry transfer was conducted to PVDF membrane for 30 minutes at 100 volts. The secondary antibody was used at 1:2500 dilution. Amersham ECL western blotting detection reagent (GE Healthcare) was used to develop the blots, and images were acquired on a LAS-3000 Imager (Fujifilm).

3.2.8 Quantitative analysis of ADARs-DD deaminase efficiency

Sequencing of the extracted plasmid was done by using the Big Dye Terminator Cycle Sequencing Kit following the manufacturer`s guideline. After the thermal cycle, ethanol precipitation was done. The samples were then loaded into the 96 well plate after mixing with HiDi formamide. The sequencing was done using the 3130 Sequence Analyzer. Ultimately, the site-directed mutagenesis was confirmed by the sequencing. Editing was quantified based on maximum peak height ratio of the edited and unedited products (100% × [C height/ (T height + C height)]) according to the described protocol.[30,31]

3.2.9 Prediction of deaminase domain structures affected by Alu-cassette

Alu-cassette affecting the ADAR2 structures were analyzed by the I-TASSER (Iterative Threading Assembly Refinement) bioinformatics tools.[32–34] I-TASSER is the consistently top-ranked bioinformatics method according to CASP (Critical Assessment of protein Structure Prediction).[35,36] The nucleotide sequence was first transcribed by Expassy translation tools then the amino acid sequence was submitted to the I-TASSER tools. Depending on the sequence length, after 3-7 days the analyzed data with different parameters were given in a link that contains different structural models, tertiary structure ligand binding site etc. Based on these values the differences that might affect functionality were determined.

3.3 Results

3.3.1 Determination of binding efficiency of MS2 protein and MS2 RNA

MS2 system is a widely used system in molecular biology to study RNA localization and preparation of chimeric proteins. In this system, I modified the ADARs-DD with an MS2 protein associated with Myc and HB tag protein, therefore, I checked the binding affinity of the modified MS2 system. The purposes of the tagged peptides were to confirm full-length chimeric protein production and their purification (Figure 1 A and B). We found considerable bonding response in Biacore ^{TM} X100. The relative binding response is raised in response to increased analytes (MS2-RNA-Guide RNA) concentration (Figure 1 D). Therefore, we can conclude that the modified system has a considerable binding affinity and can be utilized in *in vivo* studies.

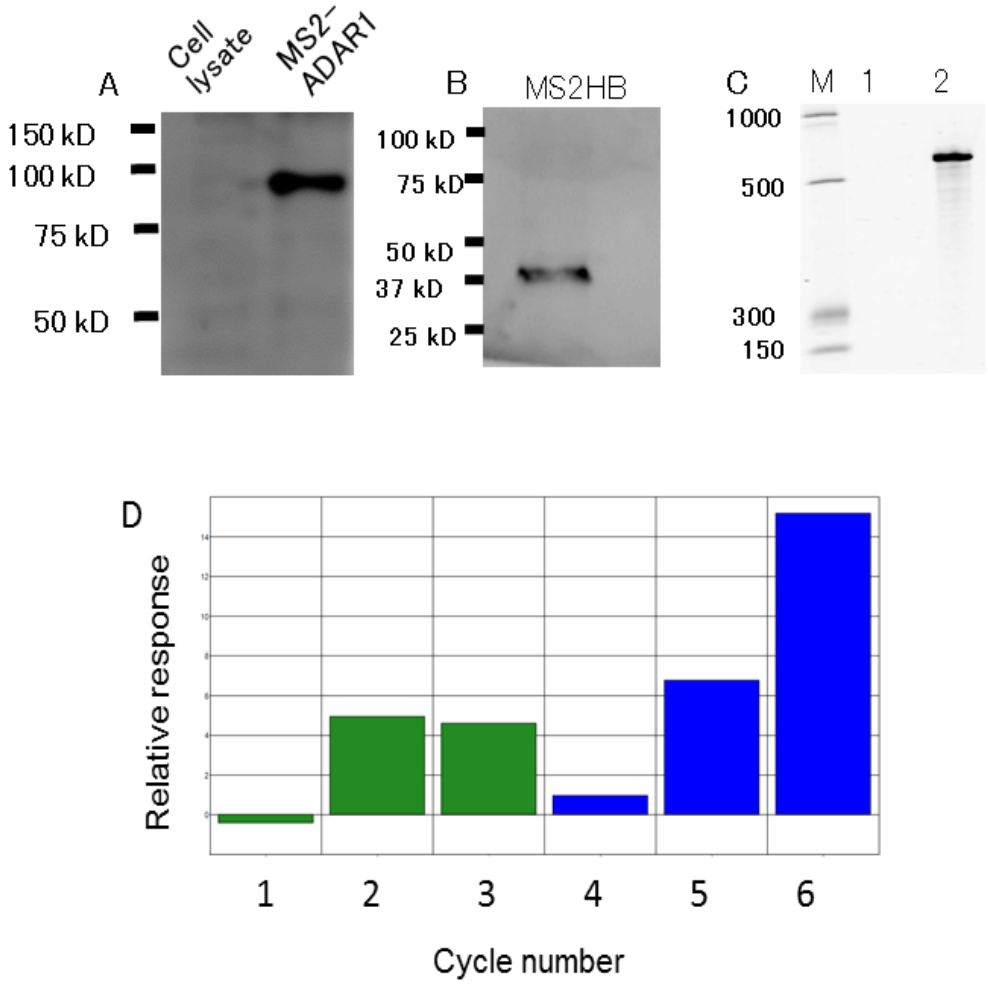


Figure 1. **A**. Western blot analysis for detection of approximately 115 KDa Myc-taged MS2-ADAR1 protein from pCS2+MT- MS2-ADAR1 transfected into HEK 293 cells. Only cell lysates without transfection were used as a negative control. **B.** Western blot analysis for detection of approximately 44 KDa Myc-taged MS2 protein from pCS2+MT-MS2 transfected HEK 293 cells. **C.** In vitro synthesis of Guide RNA analytes. M indicates the ssRNA marker, lane 1 blank, lane 2 in vitro synthesized approximately 550 bp RNA in denatured polyacrylamide gel **D.** Biacore ™ X100 result to analyse the binding affinity of MS2 and MS2 RNA. Cycle 1-3 for start up and cycle 4-6 response for sample where in cycle 1-3 response only from buffer were used, cycle 4-5 blank and cycle 6 indicates binding response of MS2 and MS2 RNA.

3.3.2 Fluorescence microscopic observation

I used the mutated EGFP (TAG) at the 58th amino acid as a reporter substrate to be converted to tryptophan (TGG) by the artificial enzyme system. Desired A-to-I editing produced a full-length fluorescence protein. I used the Keyance Biozero 800 fluorescence microscope where fluorescence signals were found to be brighter than the JuLi smart fluorescence microscopy (Figure 2). The other enzymes like ADAR2-long without Alu SDM, ADAR2-long with Alu SDM, ADAR2-long without Alu-cassette showed fluorescence signals (Figure 2). Roughly the intensity of fluorescence signal like ADAR1> ADAR2-long without Alu SDM> ADAR2-long with Alu SDM > ADAR2-long without Alu (Figure 2). Therefore in summary, from the fluorescence examination, we only found the deaminase activity of the ADAR2-long with Alu, ADAR2-long without Alu, ADAR2 long with Alu SDM, ADAR2 long without Alu SDM, ADAR1 and ADAR2 with short C-terminal did not show any observable fluorescence.

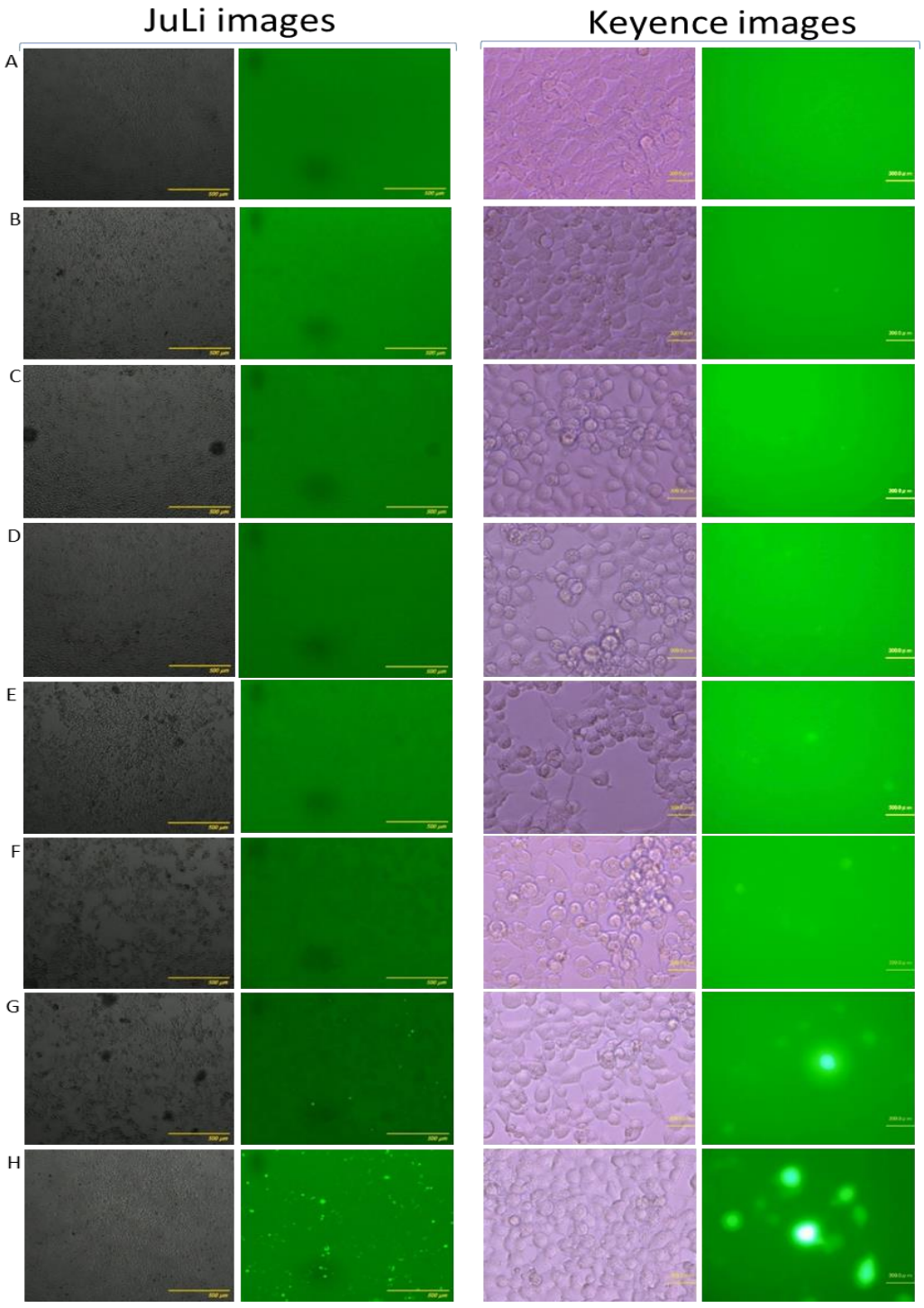


Figure 2. JuLi Smart fluorescence cell imager microscope and Keyence Biozero-800 fluorescence microscope images. Colum **A.** no enzyme used, **B.** ADAR2-short C-terminal, **C.** ADAR2-long with Alu, **D.** ADAR2-long without Alu, **E.** ADAR2-long with Alu SDM, **F.** ADAR2-long without alu SDM, **G.** ADAR1, **H.** EGFP

3.3.3 Western blot analysis to detect full-length functional EGFP protein

As we inserted an amber (TAG) stop codon in the substrate, we detected the comparative functionality of different enzymes by western blot to determine whether this system with all DD can convert mutated EGFP and produce full-length functional proteins or not. To find the answer, I aimed to detect 29 kDa EGFP in all ADARs, even in the case of ADAR2-short. In western blot analysis, almost all ADARs showed signals (Figure 3). In western blot, ADAR1 showed the strongest signal than any other ADARs. In this case, ADAR2-short in C-terminal showed a very little signal but it was still higher than the negative control (Figure 3).

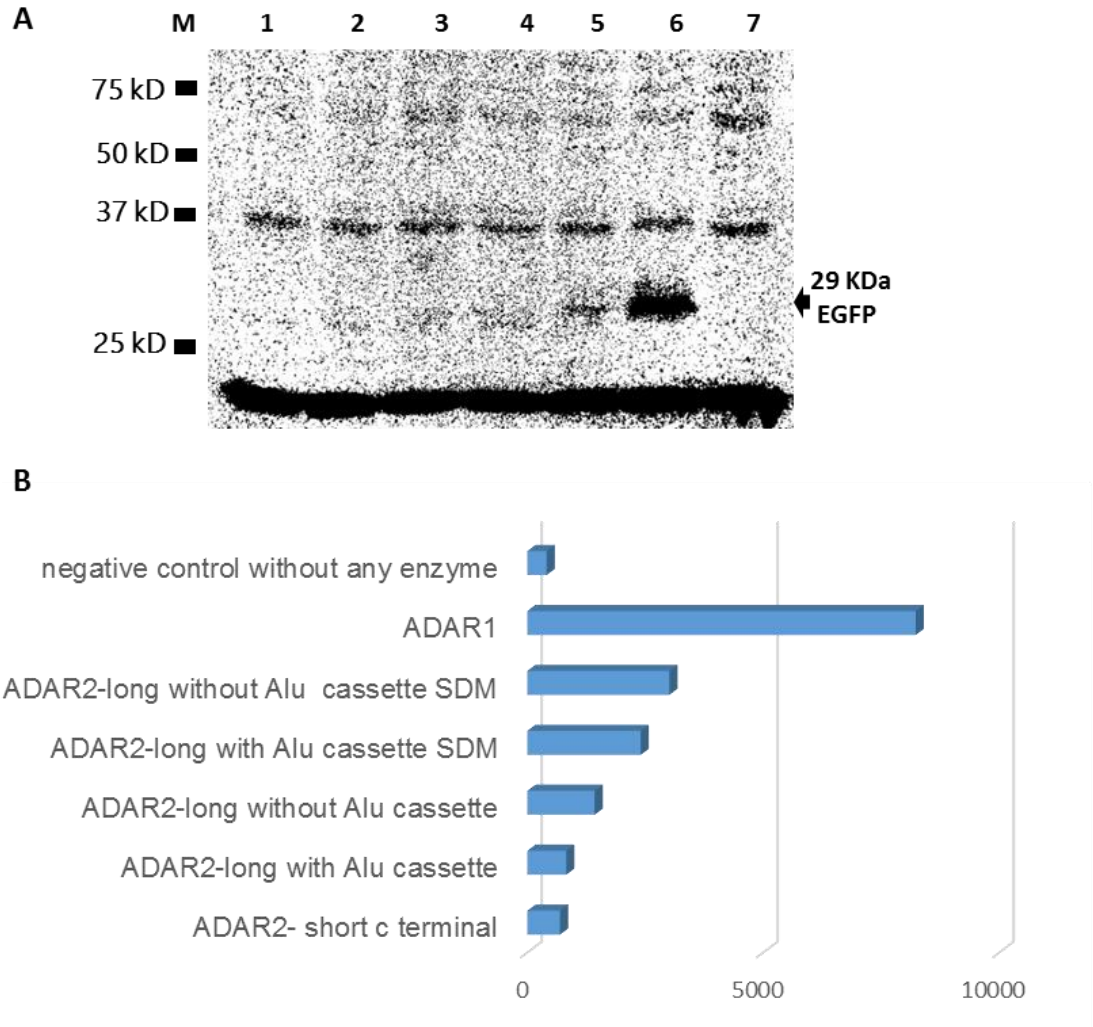


Figure 3. **A.** Western blot analysis to detect comparative efficiency of ADARs isoforms. lane 1. ADAR2- short c terminal; lane 2. ADAR2-long with Alu cassette; lane 3. ADAR2-long without Alu cassette; lane 4. ADAR2-long with Alu cassette SDM; lane 5. ADAR2-long without Alu cassette SDM; lane 6. ADAR1; lane 7. negative control without any enzyme. In all lanes (1-6), 29 kDa EGFP was present in almost all ADAR isoforms **B.** Comparative efficiency of ADAR isoforms. Western blot and densitometry analysis.

3.3.4 Quantitative analysis of ADARs-DD deaminase efficiency

To find out the specificity of the adenosine editing, I sequenced the targeted EGFP mRNA. In the antisense sequencing result, I found that the targeted adenosine was deaminated which was observed as dual C/T peak height in the targeted adenosine. C peak heights were observable in the case of ADAR1 and ADAR2 long without Alu-cassette (Figure 4). In another case the C peak signals were undetectable. This also gives the result that ADAR1 is more active than any other isoform. Editing efficiency for ADAR1-DD is approximately 7.5%, ADAR2-long without Alu 6.5%.

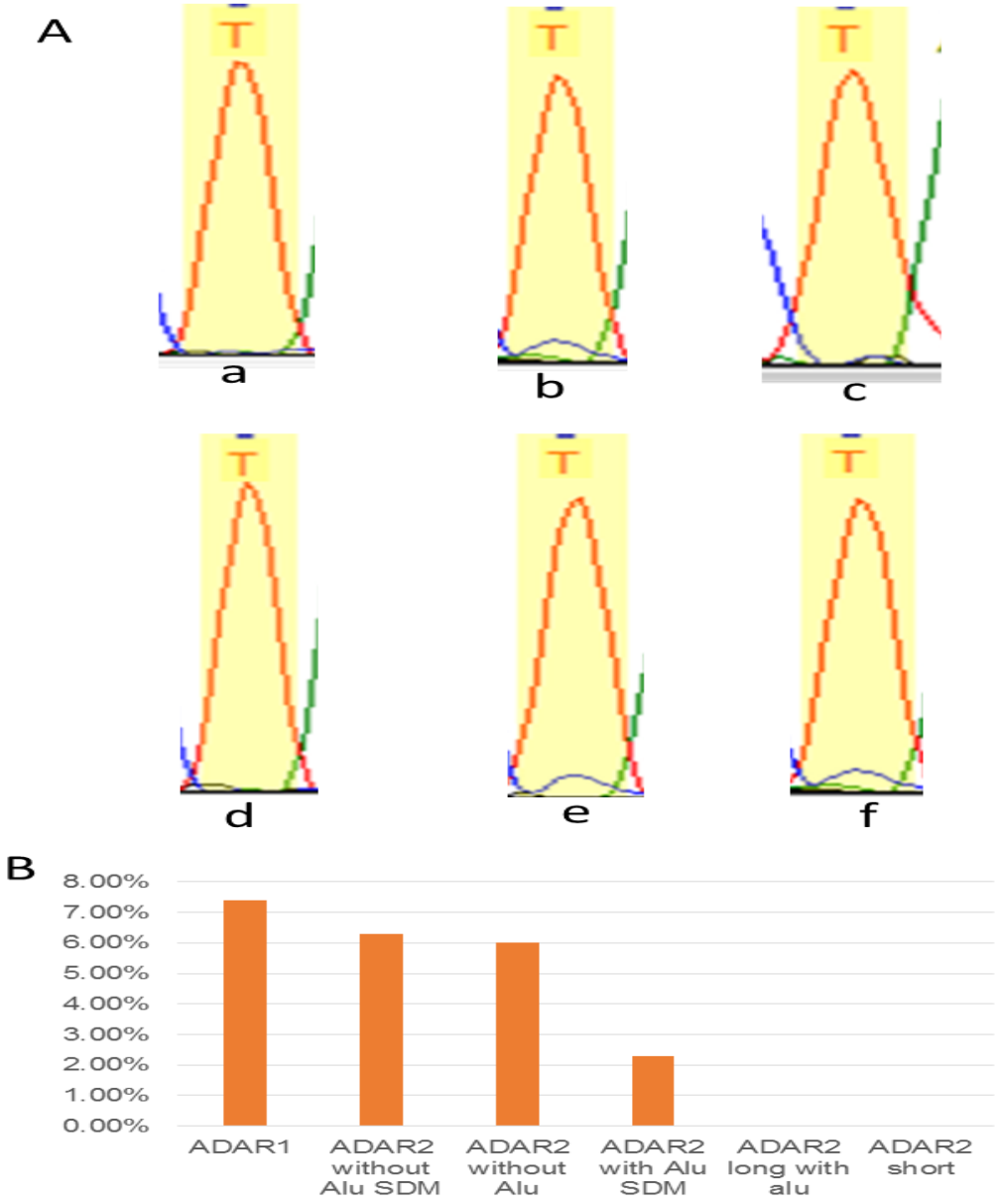


Figure 4. **A.** Sequencing result for different isoforms of ADARs-DD; a. ADAR2-short C-terminal, b. ADAR2-long without Alu cassette SDM, c. ADAR2-long with Alu cassette SDM, d. ADAR2-long with Alu cassette, e. ADAR2-long without Alu cassette SDM, f. ADAR1 **B.** Comparative efficiency of different isoforms of ADARs towards amber stop codon TAG restoration. Data was obtained from sequencing result by peak height ratio calculation.

3.3.5 Deaminase domain structure, estimated distance among residues and ligand binding sites affected by Alu-cassette

According to the I-TASSER (Iterative Threading Assembly Refinement) data, inserted Alu-cassette in the ADAR2 long result in the longer coil (Figure 5 A) in the middle of the deaminase domain than the isoform without Alu-cassette (Figure 5 B). In another finding of the Swiss model and FATCAT superposition, the coil is found larger when Alu-cassette added (Figure 5 C). Due to insertion of the Alu-cassette, the distance among the residues after 203 is increased dramatically (Figure 6 A and B). The ligand binding site i.e. nucleic acid binding capacity also largely differ due to the insertion of the Alu-cassette (Figure 7).

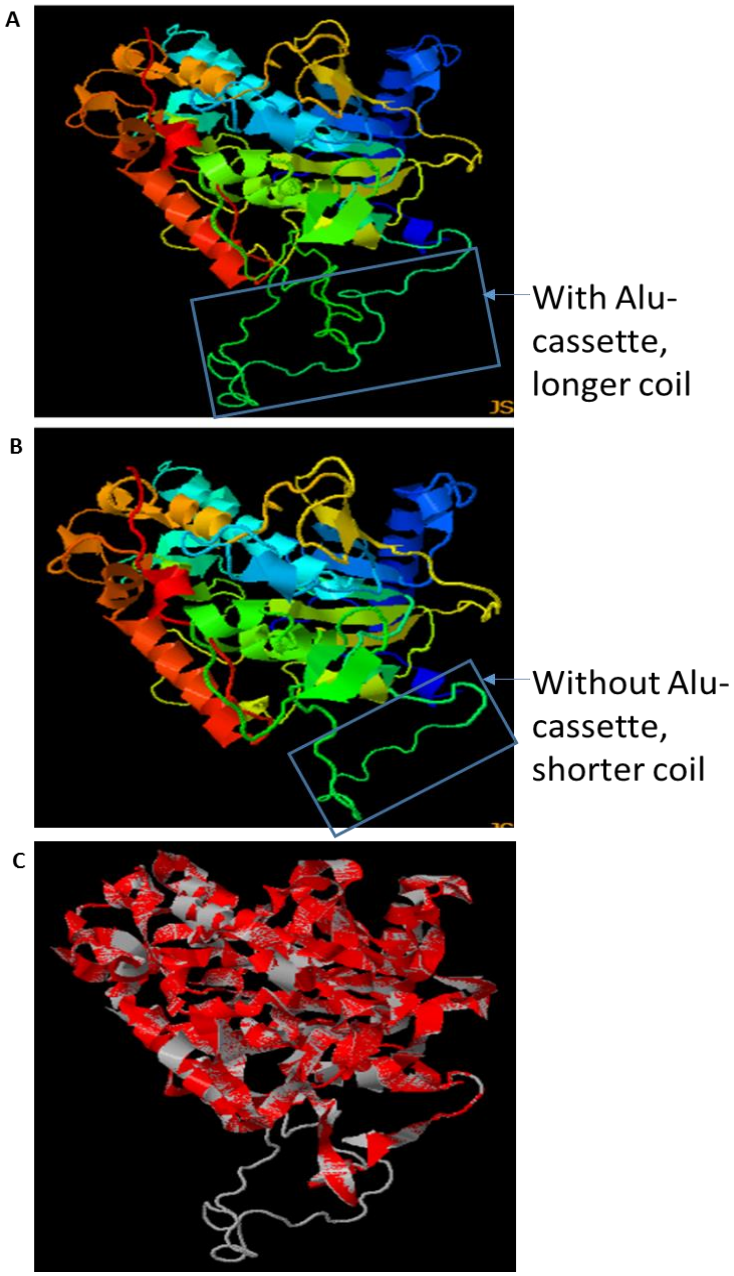


Figure 5. With and without Alu-cassette affecting the structure of ADAR2-long isoform. **A.** Structure with Alu-cassette, **B.** Without Alu-cassette **C.** superposition of two isoforms with different colors generated by FATCAT. Gray and read color indicate figure A and B respectively.

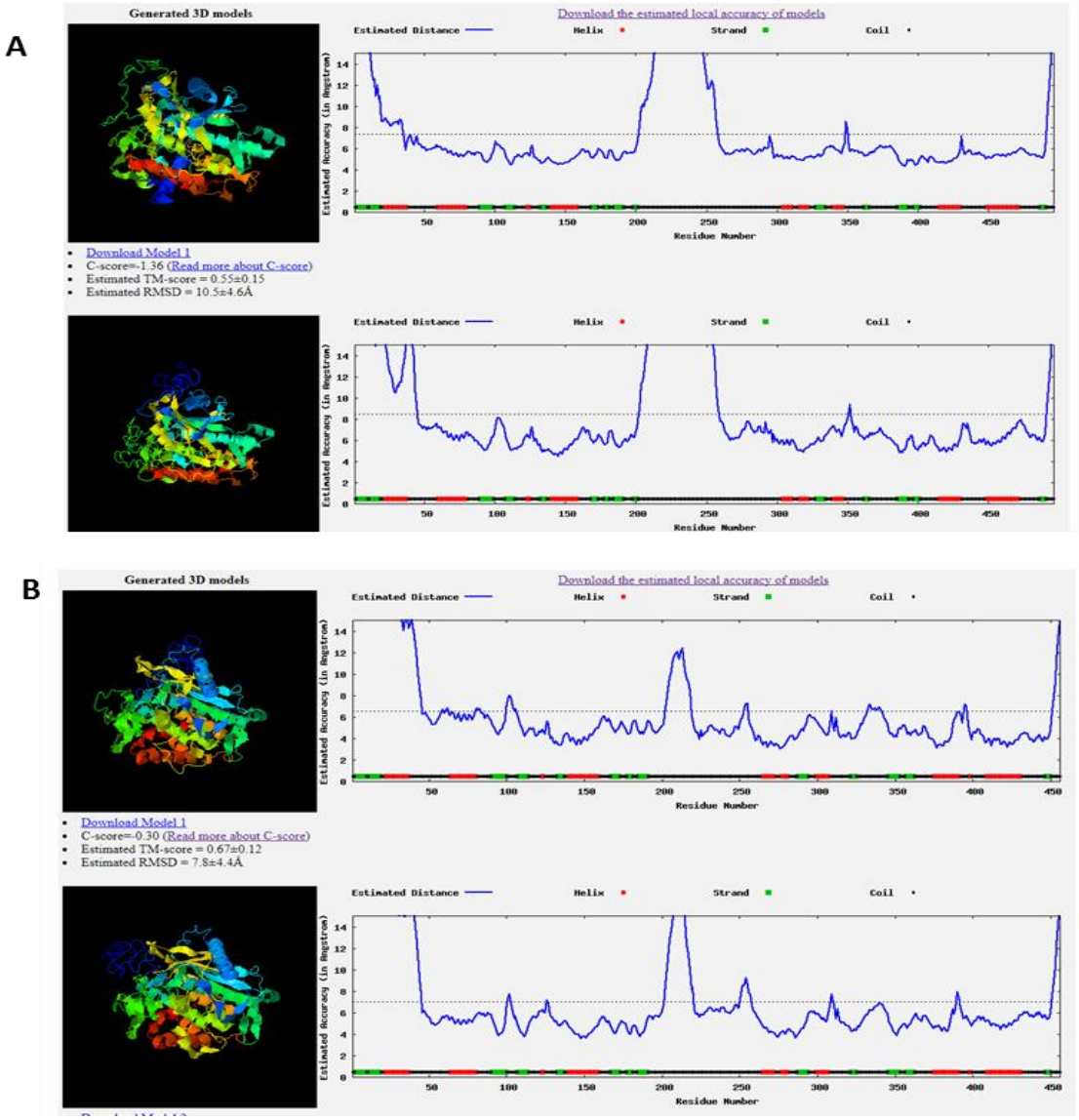


Figure 6. The graph indicates the estimated distance among residues by I-TASSER bioinformatics tools. **A.** Indicates two models regarding ADAR2-long with Alu-cassette, **B.** Indicates two models regarding ADAR2-long without Alu-cassette. Difference of gap after 200 residue is widely visible.

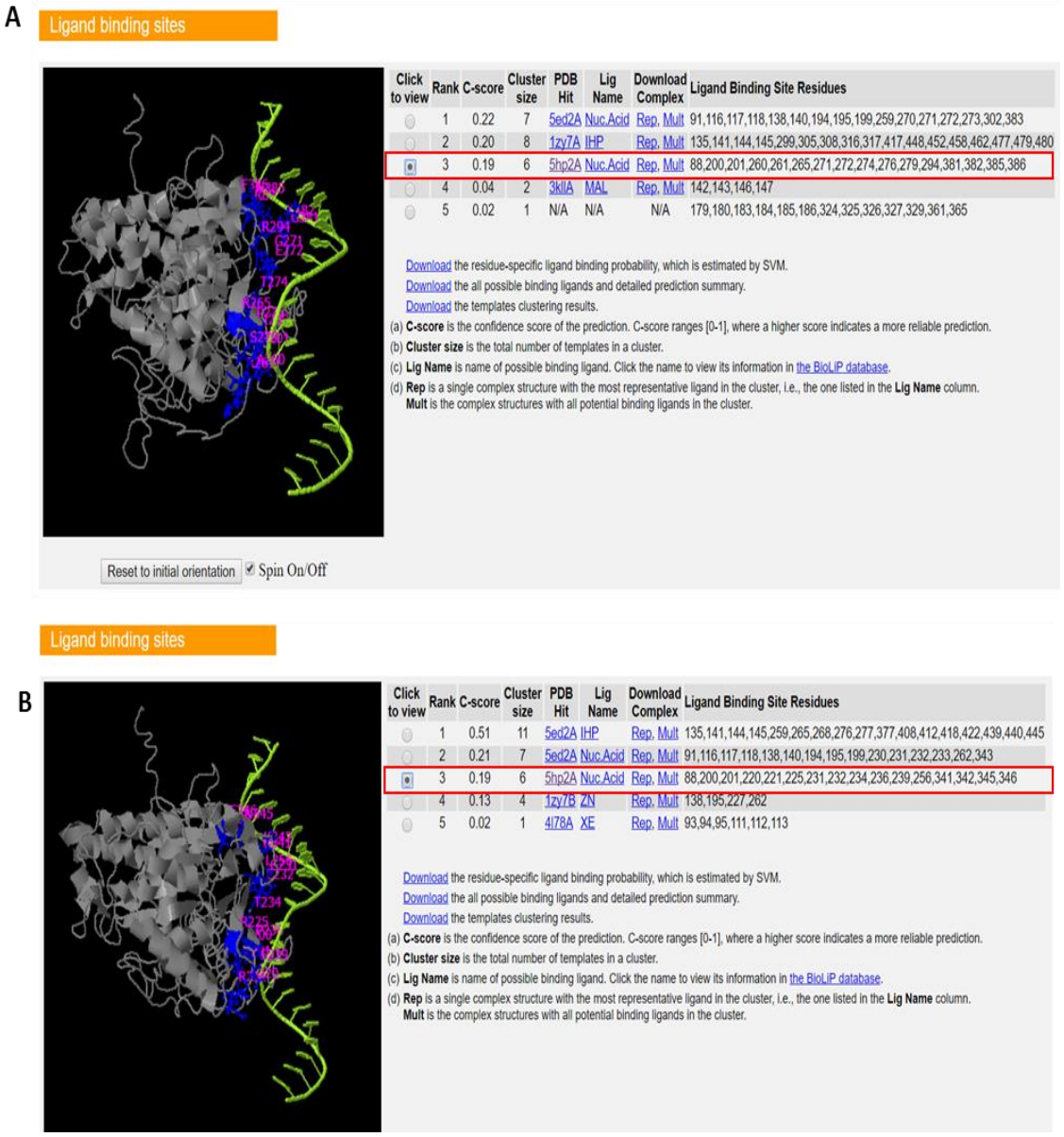


Figure 7. Alu-cassette and alteration of ligand binding sites in deaminase domain by I-TASSER bioinformatics tools. **A.** ADAR2-long with Alu-cassette, **B.** ADAR2-long without Alu-cassette.

3.4 Discussion

In most of the recent site-directed RNA editing approaches recently either ADAR1-DD or ADAR2-DD is used.[3] [20,37] In our newly developed MS2-ADARs-DD systems we also targeted the ADAR2-DD first but in our study, we found lower deaminase activity (Figure 2, 3 and 4). However, it is interesting that the ADRA1-DD was found to be more active than the other isoforms (Figure 2, 3 and 4). ADARs-DD are usually conserved but they have highly variable dsRBDs with varying activity.[4] In this study as only the DD of ADARs are used, the variability is minimized very well. The MS2 system was similar for all ADARs. The target amber (TAG) stop codon of mutated EGFP was selected as it is easy to deaminate. Therefore, the minimum deaminase activity of ADARs isoforms was possible to study by this system.

It was found that the binding affinity of the modified MS2 system was retained well (Figure 1 D). In most of the study, the deaminase domain of ADAR2[14,19,20,37,38] is used. A detailed study on ADAR2 and its substrate may be the cause for this. Structural study on both ADAR2-DD and dsRBDs has been revealed[39,40] whereas the study in the case of ADAR1 is less informative. In sequencing quantitative analysis, it was found that only ADAR1 and ADAR2 showed signal. However, in western blot analysis, all the ADARs-DD showed full-length EGFP converted from mutated EGFP. Antisense sequencing of

the targeted mRNA was performed as it gives more accurate quantification than the sense sequencing.[41] It is also found that, the editing efficiency of this system approximately 7 percent. In the sequencing result, deaminase activity of some isoforms was undetected. It is reported that editing efficiency needs to be more than 5 % at RNA level to be detected by Sanger sequencing methods.[42] In this system, efficiency can be increased by inserting multiple guides and multiple MS2 RNA in the plasmid construct. In the very recent work, multiple guides and chimeric protein complexes greatly increased the efficiency of editing system.[3,20] In the genetic correction where ADAR2-DD was used previously, if ADAR1-DD is used in that case the correction efficiency could be increased more. In an *in-vitro* study, it was observed that full-length ADAR1 is more functional than ADAR2.[43] In another recent study, with the full-length ADARs and the guide system that basically designed on the basis of natural binding strategy of ADARs with the glutamate receptor found that, ADAR2 more functional than the ADAR1.[44] It might be due to the different binding affinity of ADARs dsRBD with the same designed guide and substrate. However, insertion of Alu-cassette almost in the middle of the deaminase domain results in the alternatively spliced isoform of ADAR2.[21] Due to insertion of the Alu-cassette, the deaminase activity of the ADAR2 is decreased but the substrate specificity remains unchanged.[21] This indicates that, substrate recognition mainly executed by the dsRBDs

which we modified by the MS2-system in this study and similar for all studied deaminase domains here. According to I-TASSER result, the inserted Alu-cassette alters the ligand binding capacity of the deaminase as well (Figure 7 A and B). Taken together it can be stated that, Alu-cassette in the coding region of ADARs play a significant role in the alteration of its functionality. The E488Q mutation of the ADAR2 long makes the enzyme more active in *in vitro* study.[47,38] However, in the cellular environment, it was not reported well. I prepared the construct converting E-to-Q amino acid and found that, the converted amino acid makes the deaminase domain more functional in the HEK-293 cell. In the cellular environment, full-length ADARs behave like a shuttle protein.[48,49,50]

ADAR2 shuttles between nucleoplasm, nucleolus whereas ADAR1 shuttles between nucleolus, nucleoplasm and cytoplasm.[48] Even a region in the N-terminal half of the ADAR1-DD interferes with nuclear localization.[48] Our guide RNA system ensures to make dsRNA structure and mismatch with the target RNA and thus easy to be recognized and deaminate. Moreover, the guide RNA which is a component of the artificial enzyme and the EGFP mRNA which is the substrates of the artificial enzyme should be present in the cytosol. To our knowledge, this is a complete comparative biological approach of RNA editing by ADARs-DD that reveals more information for future use of these deaminases for RNA editing purpose.

3.5 References

1 Napoli C, Lemieux C, Jorgensen R. Introduction of a Chimeric Chalcone Synthase Gene into Petunia Results in Reversible Co-Suppression of Homologous Genes in trans. *Plant Cell* 1990; **2**: 279–289.

2 Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A Programmable Dual-RNA – Guided. *Science (80)* 2012; **337**: 816–822.

3 Montiel-Gonz´alez MF, Vallecillo-Viejo IC, Rosenthal JJC. An efficient system for selectively altering genetic information within mRNAs. *Nucleic Acids Res* 2016; **44**: 1–12.

4 Kazuko Nishikura. Funtions and regulation of RNA editing by ADAR deaminases. *Annu Rev Biochem* 2010; **79**: 321–349.

5 Maas S, Rich A. Changing genetic information through RNA editing. *BioEssays* 2000; **22**: 790–802.

6 Sun T, Bentolila S, Hanson MR. The Unexpected Diversity of Plant Organelle RNA Editosomes. *Trends Plant Sci* 2016; **21**: 962–973.

7 Maas S, Rich A, Nishikura K. A-to-I RNA editing: Recent news and residual mysteries. *J Biol Chem* 2003; **278**: 1391–1394.

8 Bass BL. RNA editing by adenosine deaminates that act on RNA. *Annu Rev Biochem* 2002; **71**: 817–846.

9 Keegan LP, Gallo A, O'Connell M a. The many roles of an RNA editor. *Nat Rev Genet* 2001; **2(11)**: 869–78.

10 Kim U, Wang Y, Sanford T, Zeng Y, Nishikura K. Molecular cloning of cDNA for double-stranded RNA adenosine deaminase, a candidate enzyme for nuclear RNA editing. *Proc Natl Acad Sci U S A* 1994; **91**: 11457–61.

11 Bass BL, Nishikura K, Keller W, Seeburg PH, Emeson RB, O'Connell MA, Samuel CE HA. A standardized nomenclature for adenosine deaminases that act on RNA. *RNA* 1997; **3**: 947–949.

12 Chen CX, Cho DS, Wang Q, Lai F, Carter KC, Nishikura K. A third member of the RNA-specific adenosine deaminase gene family, ADAR3, contains both single- and double-stranded RNA binding domains. *RNA* 2000; **6**: 755–67.

13 Woolf TM, Chase JM, Stinchcomb DT. Toward the therapeutic editing of mutated RNA sequences. *Proc Natl Acad Sci* 1995; **92**: 8298–8302.

14 Schneider MF, Wettengel J, Hoffmann PC, Stafforst T. Optimal guideRNAs for re-directing deaminase activity of hADAR1 and hADAR2 in trans. *Nucleic Acids Res* 2014; **42**: 1–9.

15 Melcher T, Maas S, Herb A, Sprengel R, Higuchi M, Seeburg PH. RED2, a Brain-specific Member of the RNA-specific Adenosine Deaminase Family. *J Biol Chem* 1996; **271**: 31795–31798.

16 Cattaneo R, Schmid A, Eschle D, Baczko K, ter Meulen V, Billeter MA. Biased hypermuation and other genetic changes in defective measles viruses in human brain infections. *Cell* 1988; **55**: 255–265.

17 Polson AG, Bass BL, Casey JL. RNA editing of hepatitis delta virus antigenome by dsRNA-adenosine deaminase. *Nature* 1996; **380:** 454-456.

18 Rueter SM, Dawson TR, Emeson RB. Regulation of alternative splicing by RNA editing. *Nature* 1999; **399**: 75–80.

19 Montiel-Gonzalez MF, Vallecillo-Viejo I, Yudowski GA, Rosenthal JJC. Correction of mutations within the cystic fibrosis transmembrane conductance regulator by site-directed RNA editing. *Proc Natl Acad Sci* 2013; **110**: 18285–18290.

20 Wettengel J, Reautschnig P, Geisler S, Kahle PJ, Stafforst T. Harnessing human ADAR2 for RNA repair-recoding a PINK1 mutation rescues mitophagy. *Nucleic Acids Res* 2017; **45**: 2797–2808.

21 Gerber A, O'Connell MA, Keller W. Two forms of human double-stranded RNA-specific editase 1 (hRED1) generated by the insertion of an Alu cassette. *RNA* 1997; **3**: 453–63.

22 Lai F, Chen CX, Carter KC, Nishikura K. Editing of glutamate receptor B subunit ion channel RNAs by four alternatively spliced DRADA2 double-stranded RNA adenosine deaminases. *Mol Cell Biol* 1997; **17**: 2413–2424.

23 Macbeth MR, Lingam AT, Bass BL. Evidence for auto-inhibition by the N terminus of hADAR2 and activation by dsRNA binding. *RNA* 2004; **10**: 1563–1571.

24 Vogel P, Schneider MF, Wettengel J, Stafforst T. Improving site-directed RNA editing in vitro and in cell culture by chemical modification of the guideRNA. *Angew Chemie* 2014; **53**: 6267–6271.

25 Vu LT, Nguyen TTK, Md Thoufic AA, Suzuki H, Tsukahara T. Chemical RNA editing for genetic restoration: the relationship between the structure and deamination efficiency of carboxyvinyldeoxyuridine oligodeoxynucleotides. *Chem Biol Drug Des* 2016; **87**: 583–593.

26 Vu LT, Nguyen TTK, Alam S, Sakamoto T, Fujimoto K, Suzuki H *et al.* Changing blue fluorescent protein to green fluorescent protein using chemical RNA editing

as a novel strategy in genetic restoration. *Chem Biol Drug Des* 2015; **86**: 1242–1252.

27 Azad TA, Bhakta S, Tsukahara T. Site-directed RNA editing by adenosine deaminase acting on RNA (ADAR1) for correction of the genetic code in gene therapy. Nature Publishing Group, *Gene Therapy* 2017 doi:10.1038/gt.2017.90.

28 Keryer-Bibens C, Barreau C, Osborne HB. Tethering of proteins to RNAs by bacteriophage proteins. *Biol Cell* 2008; **100**: 125–138.

29 Buxbaum AR, Singer RH. In the right place at the right time: visualizing and understanding mRNA localization. *Nat Rev Mol Cell Biol* 2015; **16**: 95–109.

30 Rinkevich FD, Schweitzer P, Scott JG. Antisense sequencing improves the accuracy and precision of A-to-I editing measurements using the peak height ratio method. *BMC Res Notes* 2012; **5**: 63.

31 Eggington JM, Greene T, Bass BL. Predicting sites of ADAR editing in double-stranded RNA. *Nat Commun* 2011; **2**: 319.

32 Yang J, Yan R, Roy A, Xu D, Poisson J, Zhang Y. The I-TASSER Suite: protein structure and function prediction. *Nat Methods* 2015; **12**: 7–8.

33 Roy A, Kucukural A, Zhang Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nat Protoc* 2010; **5**: 725–738.

34 Zhang Y. I-TASSER server for protein 3D structure prediction. *BMC Bioinformatics* 2008; **9**: 40.

35 James N. D. Battey J, ¨rgen Kopp, Lorenza Bordoli, Randy J. Read, Neil D. Clarke and TS. Automated server predictions in CASP7 James. *Proteins* 2007; **69**: 68–82.

36 Moult J, Pedersen JT, Judson R, Fidelis K. A large-scale experiment to assess protein structure prediction methods. *Proteins Struct Funct Bioinforma* 1995; **23**: ii–iv.

37 Hanswillemenke A, Kuzdere T, Vogel P, Jékely G, Stafforst T. Site-directed RNA editing *in vivo* can be triggered by the light-driven assembly of an artificial riboprotein. *J Am Chem Soc* 2015; **137**: 15875–15881.

38 Kuttan A, Bass BL. Mechanistic insights into editing-site specificity of ADARs. *Proc Natl Acad Sci* 2012; **109**: E3295-E3304.

39 Mark R. Macbeth, Heidi L. Schubert, Andrew P. VanDemark, Arunth T. Lingam, Christopher P. Hill BLB. Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. *Science (80)* 2005; **309**: 2033–2037.

40 Stefl R, Oberstrass FC, Hood JL, Jourdan M, Zimmermann M, Skrisovska L *et al.* The solution structure of the ADAR2 dsRBM-RNA complex reveals a sequence-specific readout of the minor groove. *Cell* 2010; **143**: 225–237.

41 Nurpeisov V, Hurwitz SJ, Sharma PL. Fluorescent dye terminator sequencing methods for quantitative determination of replication fitness of human immunodeficiency virus type 1 containing the codon 74 and 184 mutations in reverse transcriptase. *J Clin Microbiol* 2003; **41**: 3306–3311.

42 Nakae A, Tanaka T, Miyake K, Hase M, Mashimo T. Comparing methods of detection and quantitation of RNA editing of rat glycine receptor alpha3. *Int J Biol Sci* 2008; **4**: 397–405.

43 Koeris M, Funke L, Shrestha J, Rich A, Maas S. Modulation of ADAR1 editing activity by Z-RNA in vitro. *Nucleic Acids Res* 2005; **33**: 5362–5370.

44 Heep M, Mach P, Reautschnig P, Wettengel J, Stafforst T. Applying human ADAR1p110 and ADAR1p150 for Site-Directed RNA editing-G/C substitution stabilizes guideRNAs against editing. *Genes (Basel)* 2017; **8**: 1–7.

45 Matthews MM, Thomas JM, Zheng Y, Tran K, Phelps KJ, Scott AI *et al.* Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity. *Nat Struct Mol Biol* 2016; **23**: 426–433.

46 Goodman RA, MacBeth MR, Beal PA, Rena A. Goodman MRM and PAB, Goodman RA, MacBeth MR *et al.* ADAR proteins: structure and catalytic mechanism. *Curr Top Microbiol Immunol* 2012; **353**: 1–33.

47 Phelps KJ, Tran K, Eifler T, Erickson AI, Fisher AJ, Beal PA. Recognition of duplex RNA by the deaminase domain of the RNA editing enzyme ADAR2. *Nucleic Acids Res* 2015; **43**: 1123–1132.

48 Strehblow A, Hallegger M, Jantsch MF. Nucleocytoplasmic distribution of human RNA-editing enzyme ADAR1 is modulated by double-stranded RNA-binding domains, a leucine-rich export signal, and a putative dimerization domain. *Mol Biol Cell* 2002; **13**: 3822–3835.

49. Nie Y, Zhao Q, Su Y, Yang JH. Subcellular distribution of ADAR1 isoforms is synergistically determined by three nuclear discrimination signals and a regulatory motif. *J Biol Chem* 2004; **279**: 13249–13255.

50. Desterro JMP, Keegan LP, Lafarga M, Berciano MT, O'Connell M, Carmo-Fonseca M. Dynamic association of RNA-editing enzymes with the nucleolus. *J Cell Sci* 2003; **116**: 1805–1818.

Chapter IV

Efficiency of site-directed RNA editing by MS2-ADAR1-DD system

4.1 Introduction

RNA editing by converting adenosine (A) to inosine (I) which read as guanosine (G) during translation is the most reported RNA editing in mammals caused by the adenosine deaminase acting on RNA (ADAR). Majority of these events occurs in the central nervous system therefore altered RNA editing results in different diseases related to the neuromuscular system. ADARs consist of dsRBDs and a deaminase domain (DD). I programmed the ADAR1-DD to the specific adenosine with the help of MS2 system mediated guide RNA. Previously, it was showed approximately 7 % RNA editing efficiency in amber stop codon (TAG) conversion to tryptophan (TGG) of EGFP and no off-target. The neighboring sequences of the targeted adenosine play an important role regarding the efficiency of A-to-I RNA editing.[1] I tried to correct three type of mutated codons in EGFP. Regarding the sequence context 5′neighbor preferences are U>A>C> G whereas 3′ neighbor preferences are G>C~A>U[2, 3]. The off-target effect is a big concern in genome editing and RNA editing as well. The potentiality of RNA editing by base substitution is much higher as it recoded the stop codons (UGA, UAG, and UAA) to tryptophan (UGG) and thus can alter 12 out of 20 amino acids. This technique has been utilized to target a wide variety of diseased transcript of Cystic Fibrosis,[4] Duchenne muscular dystrophy[5] and Factor V Leiden thrombophilia.[6] In the Sanger sequencing, off-

target editing is difficult to assess in genome-wide. Genome-wide off target asses will provide more information on editing status. According to the above findings and EGFP context, amber (TAG) is the easiest and opal (TGA) is the most difficult codon to be converted to (TGG). Ochere (TAA) has the moderate possible codon for conversion. However, double conversion (TAA>TGG) might be the fact regarding conversion efficiency. The aim of this work is to evaluate the efficiency of different guide RNAs to improve the editing efficiency. Therefore, I have designed the guide RNA upstream and downstream of the MS2 RNA. The guide sequences are 19, 21, 23, 19 2X bases, another 2 guides containing 13 bp upstream but in a different location was designed. Total 9 guides complementary to the targeted RNA were designed where 7 were tested. In this time, I decrease the transfecting reporter target by $1/4^{th}$ and found the editing efficiency up to highest 16.6%, however, I found one off target and without guide RNA the targeted editing was noticeable as well. It indicates that for specificity, there should be a precise calculation of the enzyme, guide and targeted RNA expression. Optimization of the promoter of the constructs and the tagged peptides present adjacent to the MS2-ADAR1-DD might have an effect on the functionality of this system for therapeutic applications.

4.2 Materials and Methods

Regarding the ADAR1-DD construct with MS2 protein was used in this study. ADAR1-DD was checked from UCSC human genome browser[7] described in chapter II. For the comparative study of six guide RNA amber stop codon (TAG) in 58th position of EGFP was used. Even the opal (TGA) and ochre (TAA) also converted by this deaminase system.

Here I designed 7 guide RNA complementary to target RNA. Guide RNA length is 19, 21, 23, 19 2X bp in length. The increased length is to N and C terminal of the target RNA simultaneously figure. In this time the purpose of different guide design was to increase the efficiency of the artificial enzyme system and to find out the most efficient guide. MS2-RNA sequence was amplified with a guide containing forward and reverse primer with reverse (MS2RNARv ATTCCTCGAGCGCAAATTTAAAGCGCTGAT) and (MS2RNAFw GATTACGAATTCGAATGGCCATG) primer respectively Guide sequence was inserted in the pCS2+ only (addgene) plasmid in EcoRI-XhoI restriction sites by cloning. Briefly, the cloning process the pCS2+ only plasmid and the PCR product of MS2 RNA 6 x containing guide sequence and EcoRI and Xhol (Takahara) were digested in a separate reaction. After that ethanol precipitation of the reaction mix was conducted. The digested vector was incubated for 30 minutes with bovine alkaline

phosphatase followed by ethanol precipitation. The purified products were separated by 1.5% agarose gel and the desired band and were cut by short exposure of UV trans illumination. Quiagen gel extraction kit was used to purify the vector and insert. Ligation was performed by mighty mix (Takara) according to manufacturer instruction. HEK-293 cells were cultured in 24 well plate 1.5x105 cells/well. After overnight incubation at 37 with 5% CO2 cells was transfected with lipofectamine -3000 (Invitrogen) according to manufacturer instruction. After 6 hours of transfection media was changed to fresh DMEM (Wako) with 10% fetal bovine serum. After 72 hours of transfection cells were collected for RNA extraction by Trizol (Invitrogen). RNA was treated with DNase (Promega, USA) according to manufacturer instructions. cDNA was synthesized with superscript III (Invitrogen) according to manufacturer instructions with EGFP gene-specific primer. PCR was conducted for 324 bp length with Go Taq polymerase (Promega) at annealing temperature 60 and extension temperature 72 for 30 seconds. Total 40 cycles were run for the desired concentration of the PCR products. PCR products were run on 6% polyacrylamide gel for 20 minutes in presence of 1X TBE buffer. Exact band size gel was cut with help of UV trans illumination. Then the gel pieces were frozen for 30 minutes at -80 degrees. After that, the gel piece was crushed with the help of disposable pellet pestle/tissue grinder (Kimble®, Capitol Scientific, Inc TX 78758, USA, Catalog no.

749520-0090). About 10 up 0.1x TE was added and vortexed for 10 minutes. The supernatant was collected to use in sequencing reaction after centrifugation for 10 minutes at high speed. Antisense sequencing was performed with the PCR reverse primer to assess the C/T peak height at the targeted adenosine. Editing percentage in RNA level was calculated from the C/T peak height[3,8].

Table 1. Guide RNA designed and inserted to upstream and downstream of MS2RNA 6X by PCR and cloning with forward and reverse primer respectively. Red letters, restriction site; green mark and underline, guide sequence; yellow mark, primer for MS2RNA; red mark CCA and TGG, targeting to mutated codon. Red mark C and G mismatch to avoid off-target editing.

No.	Guide Name	Guide sequence
1	Guide 19 bp upstream	atcaGAATTCAGGGGGGGCCAGGGCACGGGAATGGCCATGGGACGTCGAC
2	Guide 21 bp upstream	atcaGAATTCGAGGGGGGGCCAGGGCACGGGGAATGGCCATGGGACGTC
3	Guide 23 bp upstream	atcaGAATTCCGAGGGGGGGCCAGGGCACGGGCGAATGGCCATGGGACGT
4	Guide 19 bp downstream	attcCTCGAGCCGTGCCCTGGCCCCCCCTCGCAAATTTAAAGCGCTGAT
5	Guide 21 bp downstream	attcCTCGAGCCCGTGCCCTGGCCCCCCCTCCGCAAATTTAAAGCGCTG
6	Guide 23 bp downstream	attcCTCGAGGCCCGTGCCCTGGCCCCCCCTCGCGCAAATTTAAAGCGCT
7	Guide 19 2x upstream	atcaGAATTCAGGGGGGGCCAGGGCACGGAGGGGGGGGCCAGGGCACGGGAATGGCCATGGGACGTCGAC
8	Guide 13g+ MSRNA upstream	atcaGAATTCAGGGGGGGCCAGGGAATGGCCATGGGACGTCGAC
9	Guide 13 g- MS2RNA upstream	atcaGAATTCGGCCAGGGCACGGGAATGGCCATGGGACGTCGAC

4.3 Result and discussion

Optimal guide design can increase the efficiency and specificity of RNA editing.[4] I increase the length of guide RNA from 19 to 23 base pairs (Table 1). However, in my study, I have found that 21 bass guide is more active than the 19 bases and 23 bases guide (Figure 1). Upstream located guides are more functional than the downstream guides for all three lengths and 21 bases guides were more functional than the other two forms i.e. functionality 21bp guide>23 bp guide> 19 bp guides (Figure 1). Regarding the position of guide RNA, upstream guide were more active than the downstream guide RNA (Figure 1). However, without guide RNA there is a considerable level of target site editing which may be due to the concentration difference of ADAR1 deaminase and reporter EGFP. I also prepared another guide with 19 2x guide which contains 19 bp repeat in the upstream guide. In a previous study, it was found that repeated guide increases the efficiency of the deaminase system.[9] Off-target is a big challenge for site-directed RNA editing which can be mitigated by changing the complementarity of guide RNA[4]. However, the structure of only the deaminase domain can decrease the off target. Recently it is published that, the addition of nuclear localization signal (NLS) in the deaminase domain decrease off-target

editing.[10]

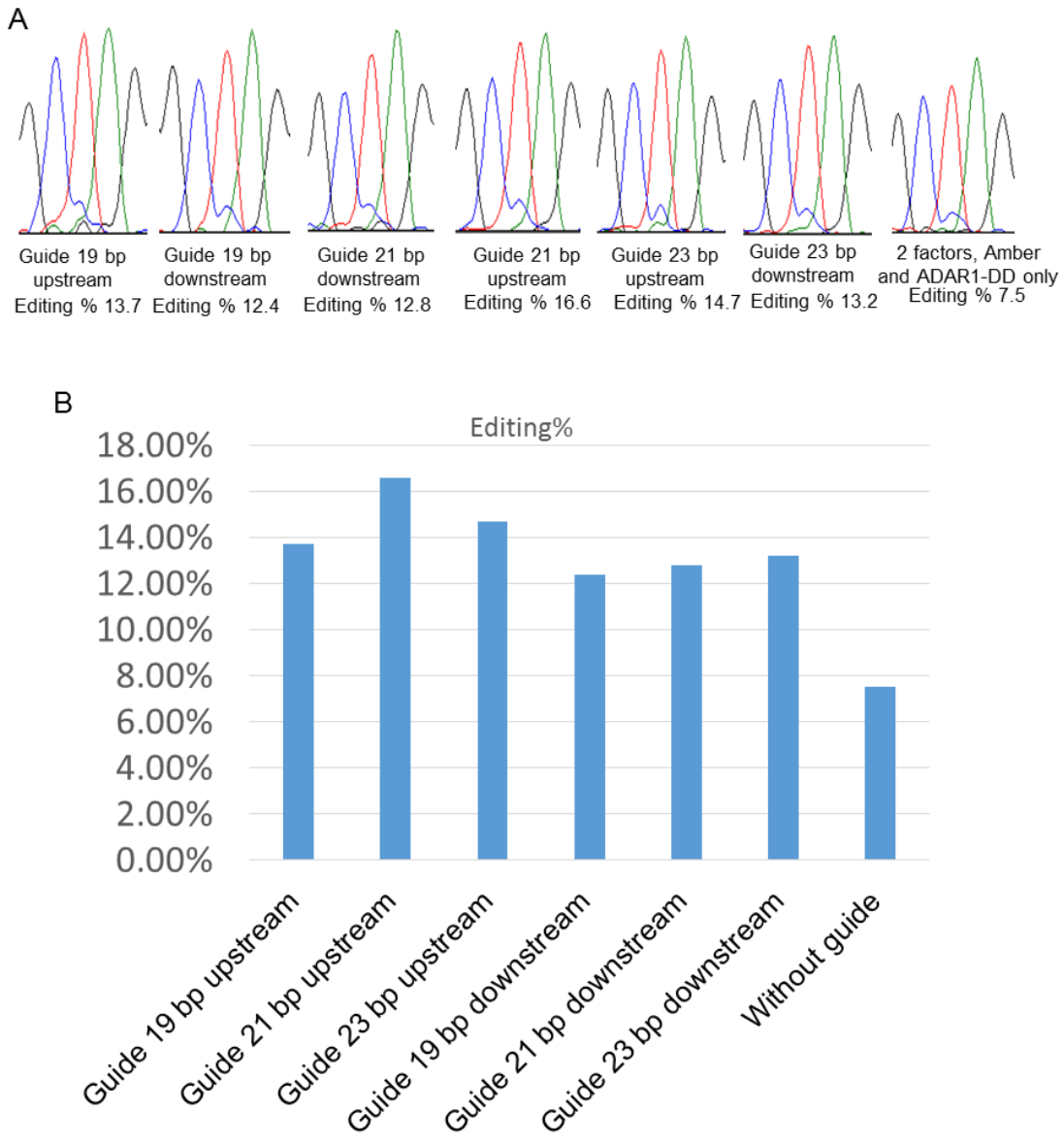


Figure 1 **A.** Sequencing result of the comparative efficiency of different guide RNA for EGFP RNA editing. **B.** Quantification of different guide RNA, It is found that upstream guide more active than the downstream guide and 21 bp guide is the most active one.

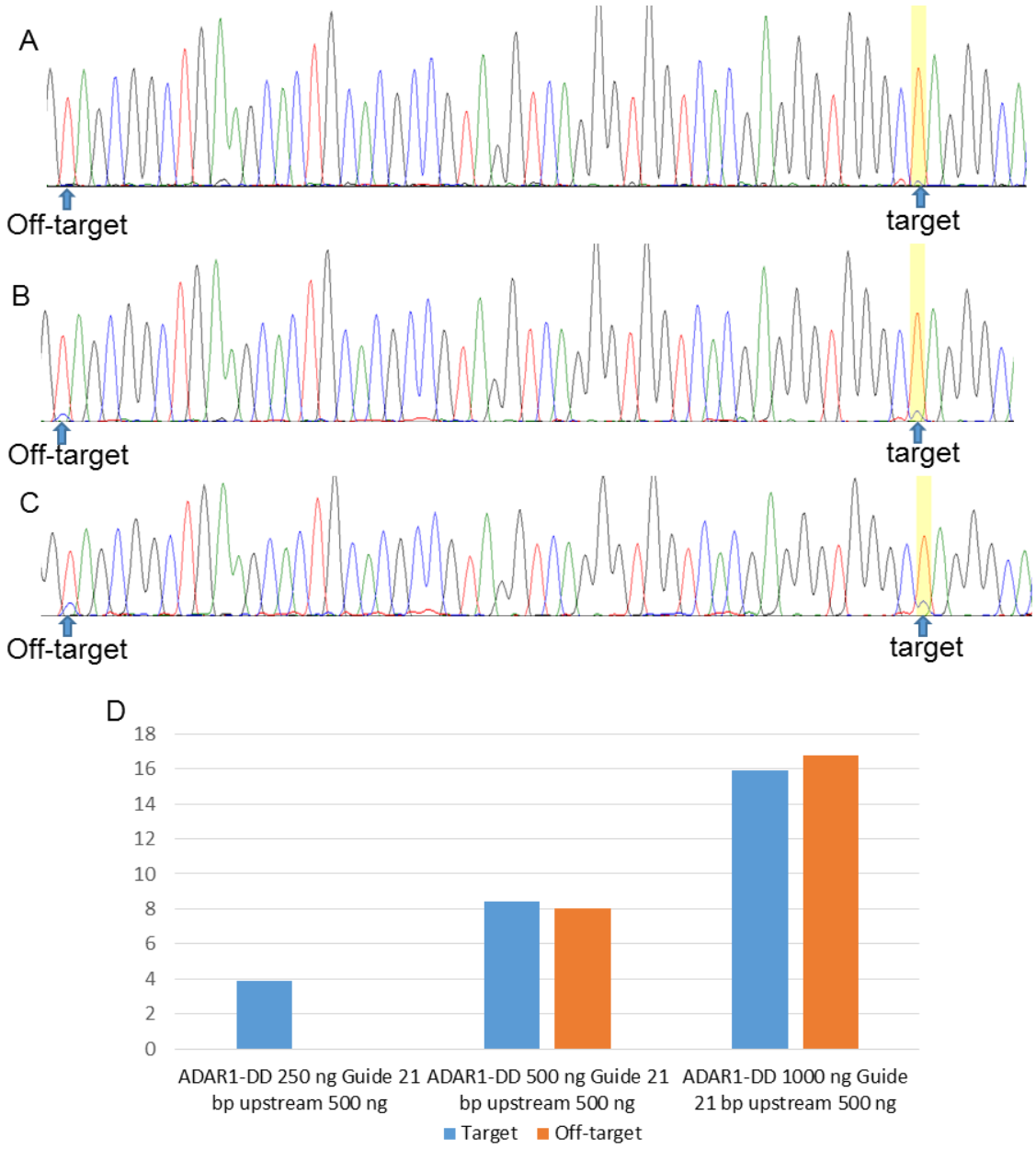


Figure 2. ADADR1-DD concentration affecting editing efficiency with guide 21 bp upstream. **A.** ADADR1-DD 250 ng Guide 21 bp upstream 500 ng, **B.** ADADR1-DD 500 ng Guide 21 bp upstream 500 ng, **C.** ADADR1-DD 1000 ng Guide 21 bp upstream 500 ng. **D.** Quantification of editing efficiency affected by the concentration of ADAR1-DD. Green bar indicates the editing in target site and orange bars indicate editing in off-target site.

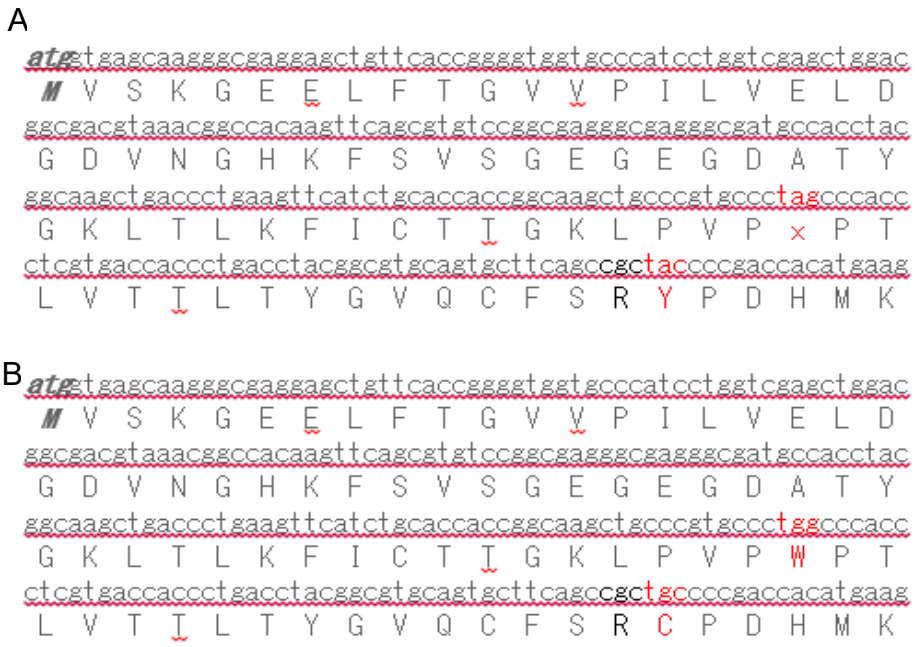


Figure 3. **A.** Upstream sequence of mutated EGFP **B**. Upstream sequence of restored EGFP. Due to off-target editing red mark TAC converted to TGC (TAC>TGC;Y>C).

Off-target editing may be due to the higher similarity of the complementary sequence in the off-target site especially due to higher G-C contents. As the off-target increase in response to the higher enzyme concentration (Figure 2), this may also due to the free enzymes that are helped to bind to the off-target sites by endogenous molecules. Off-target editing (TAC>TGC) found 50 bp downstream of the target (TAG>TGG) (Figure 3). Off-target editing increase when the concentration of ADAR1-DD is increased. In the off-target conversion of an adenosine results missense mutation (TAC>TGC; Y>C) in the EGFP. In complementary to the guide sequence no off-target editing observed.

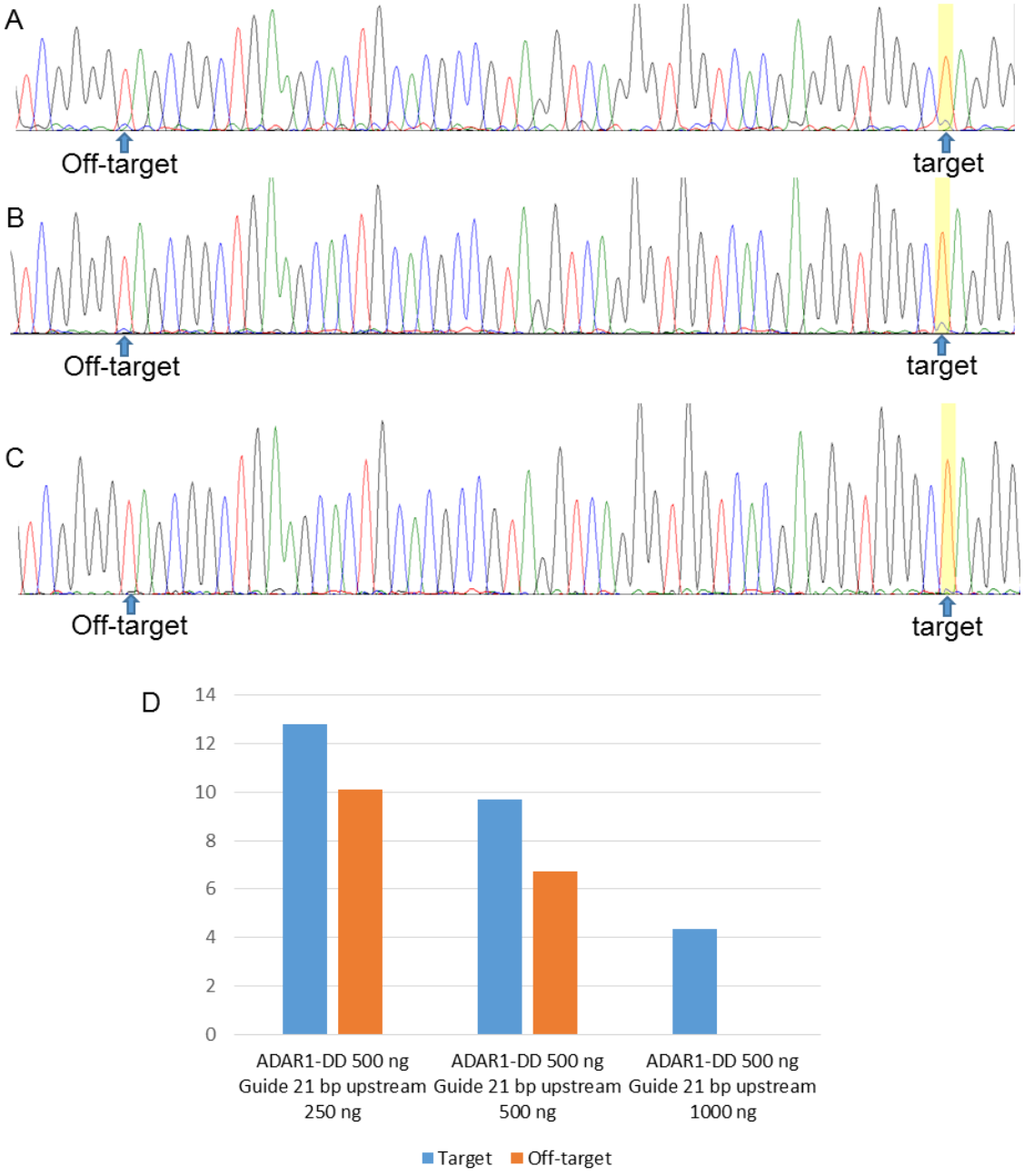


Figure 4 **A.** ADADR1-DD 500 ng Guide 21 bp upstream 250 ng, **B.** ADADR1-DD 500 ng Guide 21 bp upstream 500 ng, **C.** ADADR1-DD 500 ng Guide 21 bp upstream 1000 ng, **D.** Quantification of editing efficiency affected by the concentration of guide RNA. Green bar indicates the editing in target site and orange bars indicate editing in off-target site.

Editing efficiency of the deaminase system increases with increasing the concentration of deaminase concentration. Editing efficiency is proportional to the concentration of deaminase (Figure 2). Regarding ADADR1-DD, 250 ng is applied editing efficiency approximately 4%. When the amount increased 2-4 times the efficiency also increases 2-4 times than the initial editing efficiency (Figure 2). In another observation, I kept the deaminase concentration constant to 500 ng and changed the guide concentration (Figure 4). In this case, the guide concentration and the editing efficiency is inversely proportional (Figure 4). This may be due to higher binding of chimeric deaminase with the guide system and thus make them less available for targeted EGFP. However off-target also found lower comparing to the figure 3. In case of off-target editing site which is found 50 bp downstream of the target-no off-target observed when the ADADR1-DD concentration was 250 ng but it increases as the deaminase concentration is increased. The codon (TAC) is the off-target codon, although there are, more probable off targetable (TAC) sequence near the off-target but in only one (TAC) location found off target editing. This off-target is proportional to the higher concentration of the ADAR1-DD deaminase. When higher proportion amount of deaminases is applied the off-target editing in that site also increased. It indicates that it can be minimised by changing the proportion of different factors. Recently it is reported the addition of NLS in the N terminal of the ADAR2-DD

decrease the off target.[10] Another thing, it may be due to the free deaminase in the cell cytoplasm and its prolonged association with the EGFP transcript. Therefore, off-target can be minimized by precise calculation of the effector molecules.

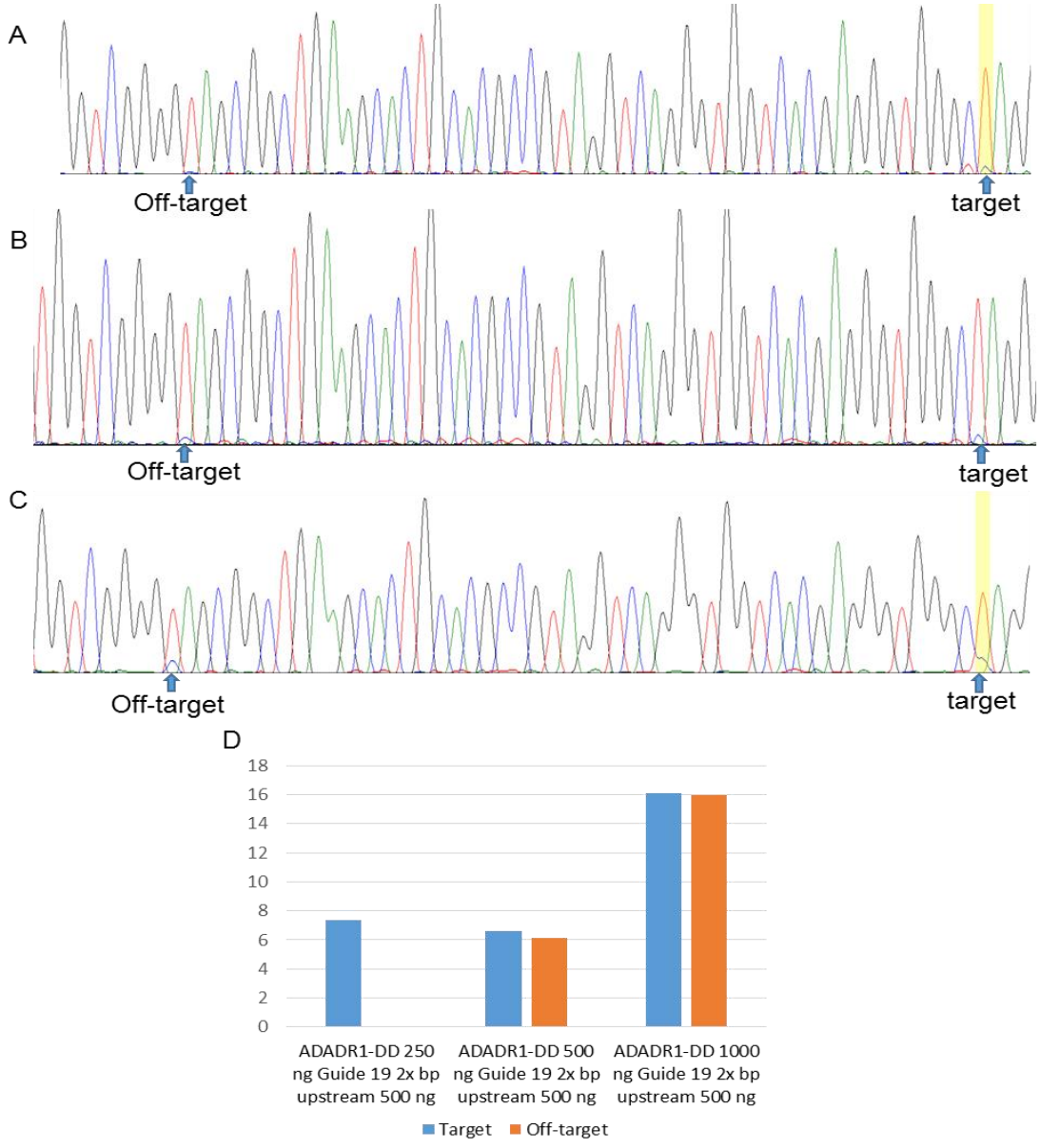


Figure. 5 **A.** ADADR1-DD 250 ng Guide 19 2x bp upstream 500 ng, **B.** ADADR1-DD 500 ng Guide 19 2x bp upstream 500 ng, **C.** ADADR1-DD 1000 ng Guide 19 2x bp upstream 500 ng. **D.** Quantification of editing efficiency affected by the concentration of enzyme concentration. Green bar indicates the editing of target site and orange bars indicate editing in off-target site.

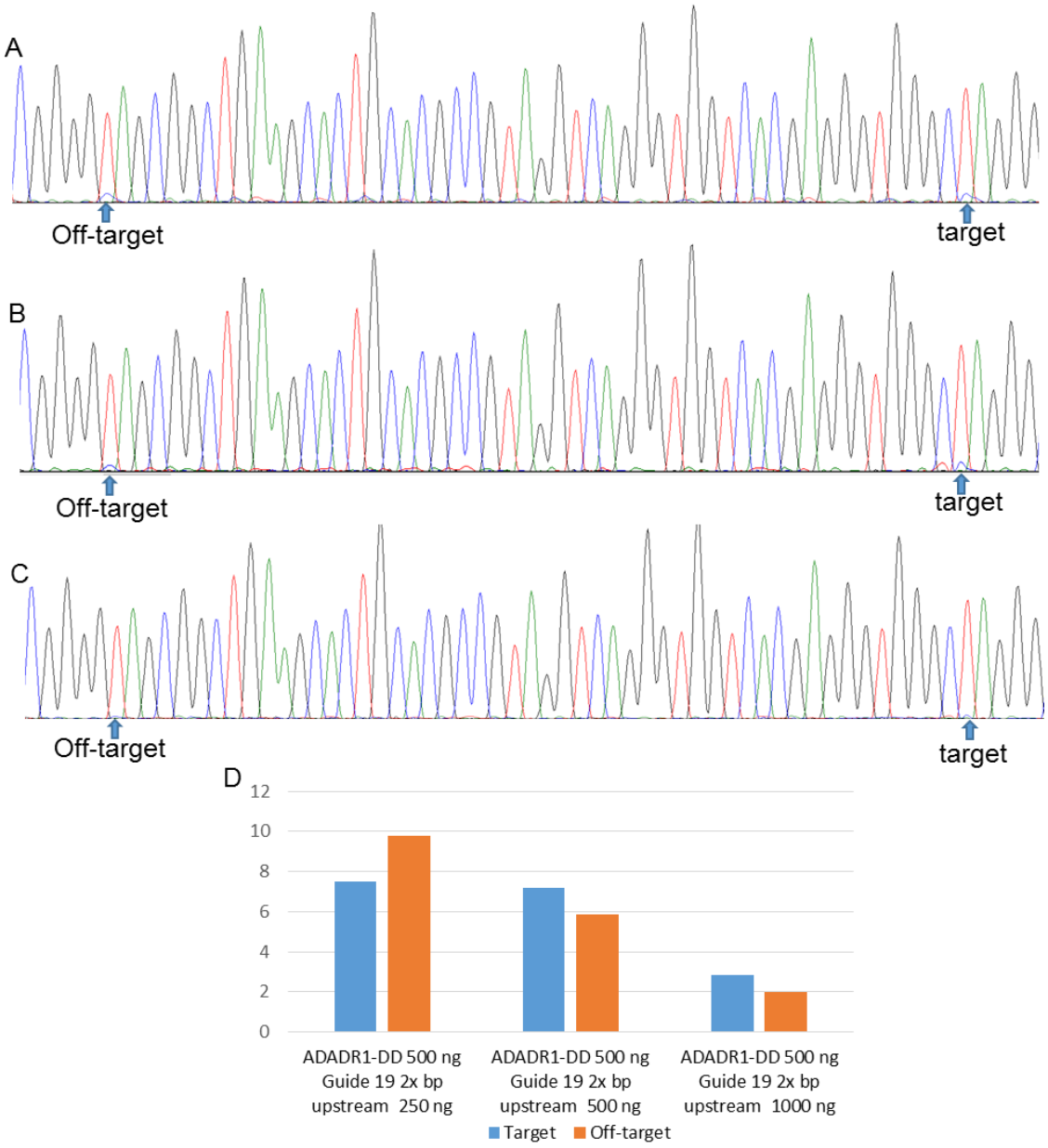


Figure 6 **A.** ADADR1-DD 500 ng Guide 19 2x bp upstream 250 ng, **B.** ADADR1-DD 500 ng Guide 19 2x bp upstream 500 ng, **C.** ADADR1-DD 500 ng Guide 19 2x bp upstream 1000 ng. **D.** Quantification of editing efficiency affected by the concentration of guide RNA. Green bar indicate the editing of target site and orange bars indicate editing in off-target

I used another guide 19 2X bp upstream where 19 bp is repeated twice as a guide. In this case when 250 ng is used the editing efficiency is >6% (Figure 5) and it is higher than the 21 bp upstream guide (Figure 2). On the other hand, when the deaminase concentration is increased twice than the editing efficiency is not increased as twice as the 21 bp upstream guide. But when the concentration of the deaminase is increased to 1000 ng, the deamination efficiency is approximately 16% (Figure 5). However, in case of changing the guide concentration with the 19 2X guide, the editing efficiency decreases with the increase of guide concentration (Figure 6).

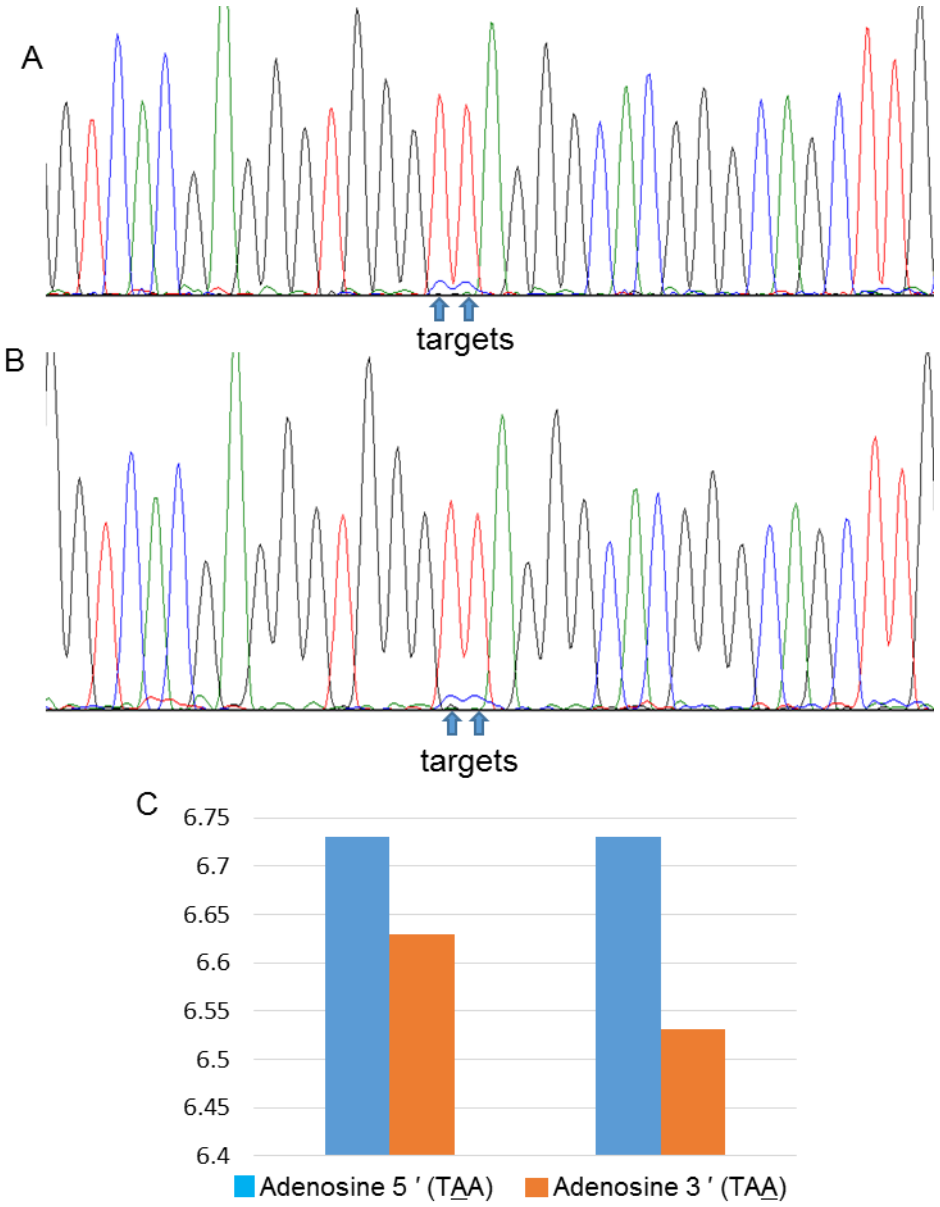


Figure 7. **A** and **B** Replicate of the experiment. **C.** Editing efficiency of conversion of stop codon ochre (TAA) to tryptophan (TGG) codon. Green color bars indicate the conversion of 5′ nearest-neighbor adenosine and orange color bars indicate the 3′ nearest-neighbor adenosine. Replicates of the experimental data are analyzed and presented side by side.

In a previous study, we have observed that ochre (TAA) stop codon can be converted to (TGG) and turn on fluorescence signals in HEK-293 cells.[11] In this study, I have quantified RNA editing efficiency and determined the extent of editing levels regarding 5' and 3' preferences. From the experimental data, it is observed that the 5' adenosine (TAA) is more efficiently edited than the 3' adenosine (TAA) (Figure 7). Editing efficiency in 5 adenosine is almost 7% whereas in 3' adenosine is 6.5%. It may due to the preference of 5' adenosine than the 3' adenosine by the deaminase system. The efficiency of RNA editing depends on several factors. For example in *in vitro* system efficiency is always higher.[4] It may be due to the presence of reactive molecules in close contact and very few interfering molecules. In case of *in vivo* system, RNA editing efficiency is comparatively lower than *in vitro* system. Regarding editing efficiency, in cell culture system is lower than the editing efficiency in egg or zygote. It may be due to the difference in the delivery of genetic material. In case of the cell, transfection is the common methods for gene delivery however in egg or zygote microinjection is a good choice. Transfection may not ensure all kinds of reactive molecules inserted simultaneously into cell cytoplasm. Whereas microinjection may ensure delivery of all reactive molecules at a time. Off-target editing is a challenge for SDRE. This types of off-target editing may found in the complimentary of guide region containing adenosines or in the other portion of targeted

mRNA. It is reported that the off-target editing even may higher than the targeted editing.[12] Off-target editing can be minimized by changing the concentration of guide RNA but this sometimes also decrease the target site editing efficiency.[13] Another recent work reported that addition of nuclear localization signal sequence in N-terminal of the deaminase domain can reduce off-target editing. The wild-type DD of ADAR2 is more specific than the mutated ADAR2 (E488Q).[12] However, the efficiency of wild-type isoforms is lower. In my study, I used the ADAR1-DD which I found more functional then ADAR2. Regarding lambda-N system and deaminase domain of ADAR2 it showed appropriately 20% editing efficiency in frog oocytes. But when they applied this technique in HEK-293 cells the efficiency decreases to approximately 12%.[4] It may be due to the difference in inserting effector molecules inside cells and oocytes. However, the off-target and how the guides were designed was not clear to me from the manuscript.[4] In my finding the editing efficiency up to 16% and the of target sites comparatively lower than the previous report. Even no off-target events in complementary to guide sequence are observed. It is reported that E488Q the more active deaminase[1] then the others but this target many off-target comparing the others.[12] Off-target is strongly related to the concentration of guide RNA expression[14]. RNA editing is a transient process and does not affect the genome level and even off-target not as malignant as genome level

malignancy. It is not found any off-target in the guide complementary region in my result. However, it is common phenomena that off-target editing can be found in case of SDM. Even multiple off-target found within complementary guide sequence in case of ADAR2-DD and ADAR2-DD (E488Q)[10]. Although 50% of the EGFP transcript was analyzed by Sanger sequencing and observable off-target was found only 50 bp downstream of the target adenosine. It might be due to a higher presence of G-C content adjacent to targeted and off targeted adenosine (Figure 2 A and B). Therefore, if NLS is added in this studied construct the off-target editing efficiency can be decreased and the editing efficiency might improve.

4.4 References

1 Nakamura T, Yagi Y, Kobayashi K. Mechanistic insight into pentatricopeptide repeat proteins as sequence-specific RNA-binding proteins for organellar RNAs in plants. Plant Cell Physiol. 2012; **53**. doi:10.1093/pcp/pcs069.

2 Lehmann KA, Bass BL. Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. *Biochemistry* 2000; **39**: 12875–12884.

3 Eggington JM, Greene T, Bass BL. Predicting sites of ADAR editing in double-stranded RNA. *Nat Commun* 2011; **2**: 319.

4 Montiel-Gonzalez MF, Vallecillo-Viejo I, Yudowski GA, Rosenthal JJC.

Correction of mutations within the cystic fibrosis transmembrane conductance regulator by site-directed RNA editing. *Proc Natl Acad Sci* 2013; **110**: 18285–18290.

5 Woolf TM, Chase JM, Stinchcomb DT. Toward the therapeutic editing of mutated RNA sequences. *Proc Natl Acad Sci* 1995; **92**: 8298–8302.

6 Vogel P, Schneider MF, Wettengel J, Stafforst T. Improving site-directed RNA editing in vitro and in cell culture by chemical modification of the guideRNA. *Angew Chemie - Int Ed* 2014; **53**: 6267–6271.

7 Hinrichs AS. The UCSC Genome Browser Database: update 2006. *Nucleic Acids Res* 2006; **34**: D590–D598.

8 Rinkevich FD, Schweitzer P a, Scott JG. Antisense sequencing improves the accuracy and precision of A-to-I editing measurements using the peak height ratio method. *BMC Res Notes* 2012; **5**: 63.

9 Montiel-González MF, Vallecillo-Viejo IC, Rosenthal JJC. An efficient system for selectively altering genetic information within mRNAs. *Nucleic Acids Res* 2016; **44**: gkw738.

10 Vallecillo-viejo IC, Liscovitch-brauer N, Montiel- MF, Eisenberg E, Rosenthal JJC, Liscovitch-brauer N *et al.* Abundant off-target edits from site-directed RNA

editing can be reduced by nuclear localization of the editing enzyme. *RNA Biol* 2017; **6286**. doi:10.1080/15476286.2017.1387711.

11 Azad TA, Bhakta S, Tsukahara T. *Site-directed RNA editing by adenosine deaminase acting on RNA (ADAR1) for correction of the genetic code in gene therapy*. Nature Publishing Group, 2017 doi:10.1038/gt.2017.90.

12 Vallecillo-Viejo IC, Liscovitch-Brauer N, Montiel-Gonzalez MF, Eisenberg E, Rosenthal JJC. Abundant off-target edits from site-directed RNA editing can be reduced by nuclear localization of the editing enzyme. *RNA Biol* 2017. doi:10.1080/15476286.2017.1387711.

13 Vallecillo-Viejo IC, Liscovitch-brauer N, Montiel-Gonzalez MF, Eisenberg E, Rosenthal JJCC, Montiel- MF *et al.* Abundant off-target edits from site-directed RNA editing can be reduced by nuclear localization of the editing enzyme. *RNA Biol* 2017; **6286**: 00–00.

14 McMahon AC, Rahman R, Jin H, Shen JL, Fieldsend A, Luo W *et al.* TRIBE: Hijacking an RNA-Editing Enzyme to Identify Cell-Specific Targets of RNA-Binding Proteins. *Cell* 2016; **165**: 742–753.

Chapter V

Final discussion

5.1 Final discussion

Recently genome editing has gained much attention after the discovery of CRISPR-Cas system although other reprogrammable genome editing tools like TALEN, zinc finger nuclease reported and still being used mainly for developing different genetically modified organism (GMO) for research purposes. Engineering of different gene editing tools largely depend on some algorithms for the design of guide sequence, avoid off target, target specificity calculating etc. Efficiency, specificity and safety are big deal regarding the use of genome editing tools. For CRISPR-Cas system, homology-mediated repair system is needed to repair between cut and inserted DNA which is only present in dividing cells. In case of non-dividing cells like nerve and muscle cells, this system facing challenges. In association with technical challenges[1] and bioethical issue, it's still controversy regarding the clinical application of this technology. Because, in case of genetic disease only one gene or part of gene is defective but when CRISPR system is applied in germ cells or embryo, other genes may have the possibility to be mutated thus ultimately it may affect multi-organ functionality. CRISPR system in association of deaminase is also used for single nucleotide substitution in DNA level.[2] However, China has tasted CRISPR for therapeutic purposes recently[3] but the outcome still under observation and preliminary result of treatment not revealed due to the privacy of the

patients. Side by side, very recently the USA has approved this system for clinical application.[4] New frontier regarding correction of point mutation in DNA has developed by using CRISPR for unwinding the DNA double helix and deamination by deaminase.

RNA editing is one of the important post-transcriptional modification of genetic information encoded in the genome of a living organism. There are different family of deaminase that are responsible for A-to-I, C-to-U and even U-to-C type RNA editing. Artificial RNA editing is the process where only targeted mRNA is edited for the purposes of tuning protein function without dismantling the highly organized and complicated genomics DNA. Naturally, the deaminase binds to the target by sequence-specific manner with help of its associated structures. For example, ADARs binds to the dsRNA formed by the inverted Alu repeats in RNA whereas APOBEC target to the cytidine adjacent to "Mooring sequences" etc. These deaminases are reprogrammable to target specific nucleotide at any sequence. Harnessing and engineering the natural deaminase is an excellent prospect for therapeutic application for the purpose of treatment of point mutated genetic diseases in future. A-to-I editing can alter 12 out of 20 amino acids. My developed system can be successfully used to target any cytosolic RNA and could be used for any developmental stage.

In the chapter II, I have shown that the recognized MS2 system can be utilized to control the deaminase domain of ADAR1 towards site-specific A-to-I RNA editing. Genetically encoded guide RNA and the chimeric protein of MS2-ADAR1-DD express well in HEK-293 cell and becomes active after being expressed. This is the first report for site-directed RNA editing by the MS2 and ADAR1-DD deaminase. Fluorescence microscopic result indicated that the signals in the experimental wells is resembling the wild-type EGFP which proves that this designed system capable of converting mutated amber (TAG) EGFP to wild type (TGG) EGFP. Restriction fragments length polymorphism (RFLP) clearly indicates that the RNA editing site is exactly the targeted site. Sanger sequencing of the target substrate is a good evidence regarding the specificity of editing by duel peak height at desired adenosine site. I performed antisense sequencing as it gives more accurate results according to the previous study. In protein level, western blot was performed to check whether the deaminase system can convert mutated EGFP substrate to full-length functional protein or not. Full-length functional protein was produced in case of the experimental treatment. It indicates that the guide RNA and chimeric protein of MS2-ADAR1-DD are working properly.

In Chapter III, I utilized my developed deaminase system in association of MS2 system for the comparative study of the DDs of different ADARs in HEK-293 cell. In most of the previous study for A-to-I conversion, ADAR2 is used but in HEK-293 cells I found that ADAR1-DD is more functional than other DDs. ADAR1-DD showed highest fluorescence signal than any others DDs. In western blot analysis, I was able to detect the full-length EGFP by polyclonal anti GFP antibody in all isoforms of ADAR-DDs. Densitometry analysis of the western blot image showed that the highest level of EGFP was expressed in case of ADAR1-DD whereas the lowest level of protein was detected in case of ADAR2-with short C-terminal. ADAR1-DD activity showed almost twice than the ADAR2-DD. However, ADAR2-long without Alu-cassette was also showed a considerable level of EGFP expression. In case of quantitative RNA editing analysis from the sequencing result, ADAR1-DD is the most active. However, ADAR2-long with Alu-cassette and ADAR2-with short C-terminal, the editing event was not detected well. It may be due to the limitation of this sequencing technique regarding detection of the low level of editing events. In this comparative study, only the DD of ADARs were used. In the previous reports, the ADAR2-long without Alu-cassette is used for base substitution.[5–9] The detailed structural study may favor that studies.[10–12] In my analysis using I-TASSER bioinformatics tools, it is obvious that inserted Alu creates a larger coil

in the middle of the deaminase domain. The longer coil even creates a wider gap among the residues located upstream and downstream of it. It is reported that adequate folding ability of the ADAR2-DD is required for the proper activity that is aided by IP6 molecule buried in the core of the DD.[10,13] The dsRBD of ADARs have different substrate specificity and affinity. Therefore, use of only deaminase domain has made the study more conclusive.

In Chapter IV, I tried to find out the most appropriate guide structure and position regarding the MS2-RNA loop. Side by side the efficiency of the system was improved with different concentration of the guide and ADAR1-DD. In this chapter, efficiency was improved from 7% to approximately 16%. Another stop codon ochre (TAA) was also converted to tryptophan (TGG). In this case, the 5' adenosine more efficiently edited than the 3' adenosine. However, the off-target editing reported in this case. The probable cause and mitigation measures regarding off-target editing also discussed here.

The recent advancement in gene editing is very exciting. This designed guide system is genetically encoded and can be utilized for any other deaminases as well. To my knowledge, any other sequence containing the adenosine can be targeted by this system. CRISPR-Cas has two components mainly; CRISPR-Cas that cut the DNA and an attached guide RNA molecule that helps to recognize the target DNA. Even CRISPR-Cas

system recently modified in association of modified adenosine deaminases from *E. Coli* [2] for the correction of point mutation in DNA level. In these case use of this human ADAR1-DD could be a good candidate for correction of point mutation in DNA level as well.

In conclusion, in my developed deaminase system, the efficiency is approximately 7% to 16%. However, efficiency and specificity are the big challenges and need consideration of many factors. Still, I am working on design and screening of different kinds of guide RNAs with variation in length and complementary sequence. Another thing, the MS2-deaminase containing plasmid pCS2+MT-MS2HB-ADARs-DD has myc-tag and HB tag peptides for detection purposes; where myc-tag is located upstream of this chimeric protein and HB located between MS2-deaminase. HB-tag is appropriately 300 bp in langth. Therefore, these structures might affect the efficiency of the whole system although such activities not reported in case of ADAR yet but reported in case of cytidine deaminase. Trimming out these tag peptides might increase the efficiency of this system. In the present result, I have utilized MS2-RNA-6X as a cognate RNA for MS2 protein. We already ordered to purchase MS2-RNA-12X plasmid which has more binding freedom with MS2 protein. Non-cutting version of CRISPR-Cas with the ADAR1-DD or full-length ADAR1 could be a good candidate for correction of point mutations in genome

level by base substitution. Lastly, report from the NMR study of ADAR1-DD in a complex with the MS2 system or ADAR1-DD alone will certainly be an important discovery and answer many questions. The above points obviously will be a landmark contribution for correction of point mutation for the therapeutic purpose in near future.

5.2 References

1 Kim D, Bae S, Park J, Kim E, Kim S, Yu HR *et al.* Digenome-seq: genome-wide profiling of CRISPR-Cas9 off-target effects in human cells. *Nat Methods* 2015; **12**: 237–243.

2 Gaudelli NM, Komor AC, Rees HA, Packer MS, Badran AH, Bryson DI *et al.* Programmable base editing of A-T to G-C in genomic DNA without DNA cleavage. *Nature* 2017; : 1–27.

3 Cyranoski D. CRISPR gene editing tested in a person. *Nature* 2016; **539**: 479.

4 Reardon S. First CRISPR clinical trial gets green light from US panel. Nature. 2016. doi:10.1038/nature.2016.20137.

5 Montiel-Gonzalez MF, Vallecillo-Viejo I, Yudowski GA, Rosenthal JJC. Correction of mutations within the cystic fibrosis transmembrane conductance regulator by site-directed RNA editing. *Proc Natl Acad Sci* 2013; **110**: 18285–18290.

6 Montiel-Gonz´alez MF, Vallecillo-Viejo IC, Rosenthal JJC. An efficient system for selectively altering genetic information within mRNAs. *Nucleic Acids Res* 2016; **44**: 1–12.

7 Heep M, Mach P, Reautschnig P, Wettengel J, Stafforst T. Applying human ADAR1p110 and ADAR1p150 for Site-Directed RNA editing-G/C substitution stabilizes guideRNAs against editing. *Genes* 2017; **8**: 1–7.

8 Kuttan A, Bass BL. Mechanistic insights into editing-site specificity of ADARs. *Proc Natl Acad Sci* 2012; **109**: E3295-E3304.

9 Wettengel J, Reautschnig P, Geisler S, Kahle PJ, Stafforst T. Harnessing human ADAR2 for RNA repair–recoding a PINK1 mutation rescues mitophagy. *Nucleic Acids Res* 2016**; 45:** 1-12.

10 Mark R. Macbeth, Heidi L. Schubert, Andrew P. VanDemark, Arunth T. Lingam, Christopher P. Hill BLB. Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. *Science (80)* 2005; **309**: 2033–2037.

11 Kuttan A, Bass BL. Mechanistic insights into editing-site specificity of ADARs. *Proc Natl Acad Sci* 2012; **109**: 3295–3304.

12 Phelps KJ, Tran K, Eifler T, Erickson AI, Fisher AJ, Beal PA. Recognition of duplex RNA by the deaminase domain of the RNA editing enzyme ADAR2.

Nucleic Acids Res 2015; **43**: 1123–1132.

13 Matthews MM, Thomas JM, Zheng Y, Tran K, Phelps KJ, Scott AI *et al.* Structures of human ADAR2 bound to dsRNA reveal base-flipping mechanism and basis for site selectivity. *Nat Struct Mol Biol* 2016; **23**: 426–433.

List of Publication

1. **Azad MTA**, Bhakta S, Tsukahara T. Site-directed RNA editing by adenosine deaminase acting on RNA (ADAR1) for correction of the genetic code in gene therapy. Nature Publishing Group, *Gene Therapy* 2017 doi:10.1038/gt.2017.90. (In press)
2. **Azad MTA**, Qulsum U, Tsukahara T. Comparative activity of adenosine deaminase acting on RNA (ADARs) isoforms for correction of genetic code in gene therapy (To be submitted).
3. **Azad MTA**, Tsukahara T. Efficiency of site-directed RNA editing by MS2-ADAR1-DD system (Manuscript in preparation).
4. Vu LT, Nguyen TTK, **Azad MTA**, Suzuki H, Tsukahara T. Chemical RNA editing for genetic restoration: the relationship between the structure and deamination efficiency of carboxyvinyldeoxyuridine oligodeoxynucleotides. *Chem Biol Drug Des* 2016; **87**(4)**:** 583-593.

Conference/workshop attended

1. Site-directed RNA editing approach by MS2 and adenosine deaminase acting on RNA (ADAR1). **Md Thoufic Anam Azad**, Hitoshi Suzuki, Toshifumi Tsukahara. Molecular Biology Society Japan Annual meeting 2016, Pacifico Yokohama, Japan.
2. Study of adenosine deaminase acting on RNA (ADAR) isoforms towards genetic code restoration. **Md Thoufic Anam Azad**, Hitoshi Suzuki, Toshifumi Tsukahara, American Society for Cell Biology annual meeting, 2106, California, USA.
3. Comparative activity of adenosine deaminase acting on RNA (ADARs) isoforms for correction of genetic code in gene therapy. **Md Thoufic Anam Azad**, Hitoshi Suzuki, Toshifumi Tsukahara. The RNA Society of Japan, Annual meeting 2017, Toyama, Japan.
4. Genetic code restoration by ADAR1 in Ochre (TAA) stop codon. Sonali Bhakta, **Md Thoufic Anam Azad**, Matomo Sakari and Toshifumi Tsukahra. The RNA Society of Japan, Annual meeting 2017, Toyama, Japan.
5. Artificial enzyme system for treatment of genetic disorders. **Md Thoufic Anam Azad**, Sonali Bhakta, Toshifumi Tsukahra. Smart knowledge/smart information/smart materials workshop 2017, Bangkok, Thailand.

ACKNOWLEDGEMENTS

First of all, I would like to express my great gratitude and thanks to my honourable supervisor Professor Dr. Toshifumi Tasukahara for his continuous support with great patience and wisdom throughout my doctoral study. He not only supervise my study with great care through the whole process with creative ideas but also helped me to grant Japanese Government MEXT scholarship during my PhD.

I greatly acknowledge the support and scholastic advices of my second supervisor Associate Professor Hidekazu Tsutsui and advisor for minor research Professor Takahiro Hohsaka for their great support and constructive advices.

I have to acknowledge the department of Veterinary and Animal Sciences, University of Rajshahi, Bangladesh for granting me study leave and understanding during my stay and study in Japan.

I would like to thanks all past members of Professor Tsukahara Lab. Specially Professor Dr. AHM Khurshid Alam, Dr. Shafiul Alam, Dr. Vu Thi Luyen, Mr. Shirakawa Yoshiaki, Mr. Iwata and present member Ms. Umme Qulsum, Dr. Sonali Bhakta, Mr. Saifullah, Mr. John Munene, Dr. Sakari Matomo, Ms. Saitaka Mamikio, Mr. Jiarui Li, Mr. Tanaka Ryuta for their support, advices and ultimate tolerance. I would like to thanks to Professor Hohsaka Lab., Professor Tsutsui Lab., Professor Takamura Lab., Professor Yamaguchi Lab., and Professor Matsumura Lab. members as well.

I would like to thank to all my colleagues and friends mainly Mr. Ehsanul Haque, Mr. Earul Islam, Mr. Md Mosaddik Hasan, Dr. S. M. Nizamuddin, Mr. Eamin Ali Akanda, Dr. Abdus Sattar, Dr. Muhammad Samir Ullah, Mr. Mohammed Khaled Alom, Dr.Tofiq Ahmed, Dr. Redwan Newaz, Dr. Md Anwarul Islam, Mr. Md Abul Kalam Siddiki, Dr. Rashed Mozumder, Dr. Abu Rashed, Dr. H. M. Belal, Mr. Farhan Ferdows, Mr. Mohammad Mesbahul Karim, Ms. Farha Diba Trisha, Dr. Shahiduzzaman Sohel, Ms. Sanchita Afroz, Dr. Reza Kaher, Dr. Asif Ali, Mr. Mohammad Nur Hasan, Mr. Amr Ashmawy, Mr. Md. Arif Hossain, Mr. Shen Qian, Ms. Savvy, Ms. Gargi Joshi, Dr. Trinh Xuan Dai, Mr. Azadul Kabir, Mr. Faiz Al Faisal, Dr. SM Ahasanul Hamid for their motivation and company during my stay in JAIST.

Last but not least, I am really grateful to my beloved wife, daughter, son and family. I am thankful to my wife for her inspiration and continuous support. I would not be able to compensate her suffering and sacrifice during my research work.

AZAD, Md Thoufic Anam
JAIST, Ishikawa, Japan
10/30/2017