University of Alberta

STRUCTURAL INVESTIGATION OF ADAR PROTEINS WITH THE USE OF SMALL ANGLE X-RAY SCATTERING

by

Edyta Agnieszka Sieminska

A thesis submitted to the Faculty of Graduate Studies and Research
in partial fulfillment of the requirements for the degree of

Master of Science

Department of Biochemistry

Edmonton, Alberta
Fall 2008

Your file Votre référence
ISBN: 978-0-494-47412-9
Our file Notre référence
ISBN: 978-0-494-47412-9

ABSTRACT

One example of chemical modification of a nucleotide is the hydrolytic deamination of adenosine to inosine which is then read as guanosine in the mature messenger-RNA during translation. The family of enzymes responsible for this deamination reaction are adenosine deaminases that act on RNA (ADARs). Three ADAR proteins (ADAR1-3) have been identified in mammals. The apparent requirement for proper function of ADARs in multiple regulatory pathways in mammals necessitates the need to understand the structure and biochemistry of these enzymes. The experiments presented in this thesis investigate the overall structure of ADAR2 and the structures of catalytic domains from ADAR1 and ADAR3. Small angle X-ray scattering studies in solution have resulted in generation of low resolution structures for full length ADAR2 and deaminase domain of ADAR2 and ADAR3. Work within provides insight into structural organization of ADARs and possible modes of interaction between these proteins and their substrates.

ACKNOWLEDGEMENTS

I would like to acknowledge all the people who have helped me, supported me and provided guidance.

I would like to thank my husband Jordan for giving me his unconditional love and support. Thank you for your patience and understanding and for being there for me when I had a bad day in the lab. Jesteś miłością mojego życia.

Dziękuję moim rodzicom Jackowi i Lidii za wychowanie mnie na osobę którą teraz jestem. Dziękuję Wam za waszą opiekę, za zapoznanie mnie ze światem nauk ścisłych, za wiarę we mnie i cierpliwość. Dziękuję dziadkowi i babci, Bronisławowi i Genowefie, za umożliwienie mi pobytu w Kanadzie i pomoc w zdobyciu wykształcenia. Dziękuję Wam za to że wierzyliście w lepsze jutro dla swojej rodziny. Dziękuję moim siostrom Klaudii, Sylwii i Gabrieli za to że jesteście moimi siostrami i najlepszymi przyjaciółkami. Thank you to my parents-in-law, Jim and Alice, for your love, support and belief in me. You have been there for me for eleven years now, providing me with the reassurance and confidence I needed.

Thank you to Dr. Andrew MacMillan for giving me an opportunity to work in his lab. Thank you for his support, guidance and patience.

I would like to thank all members of the MacMillan lab, both past and present. Dr. Oliver Kent, Emily Gesner, Matt Schellenberg, Dustin Ritchie, Tao Wu and Chao Liang; thank you for your comments and suggestions to my research, for proof reading my thesis, and attending my practice talks.

I would also like to thank Dr. Ross Edwards for his help with the SAXS data analysis and model building. His patience and generous help is what made this thesis possible.

I would like to acknowledge the professors who participated in my defence. Thanks to Dr. Howard Young, Dr. Richard Fahlman and Dr. Joe Casey for their encouragement and helpful comments. Thank you to Dr. Joanne Lemieux for being the Chairperson for the defence and for the encouraging words for the future.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

A	adenosine
Å	angstrom
aa	amino acid
ADAR	adenosine deaminase that acts on RNA
ADAT	adenosine deaminase that acts on tRNA
AMP	adenosine monophosphate
AMPA	α-amino-3-hydroxy-5-methyl-4-isoxazolepropionic acid
BME	β-mercaptoethanol
C	cytidine
CDA	cytidine deaminase
CDAR	cytidine deaminase that acts on RNA
Ci	Curie
D_{max}	maximum particle dimension
DR	dummy residue
DRADA	double stranded adenosine deaminase
dsRBM	double stranded RNA binding motif
dsRBP	double stranded RNA binding protein
dsRNA	double stranded RNA
DTT	dithiolthreitol
EDTA	ethylene diamine tetraacetic acid
F-hADAR	Flag tagged human ADAR
FRET	fluorescence resonance energy transfer
G	guanosine
GluR	glutamate receptor subunit
hADAR	human ADAR
HRP	horseradish peroxidase
I	inosine
IP_6	inositol hexaphosphate
MALDI	matrix-assisted laser desorption/ionization
MWCO	molecular weight cut off
NSD	normalized spatial discrepancy
PAGE	polyacrylamide gel electrophoresis
PCR	polymerase chain reaction
PDB	protein data bank
PEI	polyethyleneimine
PKR	RNA dependent protein kinase
pre-mRNA	precursor-messenger RNA
R_g	radius of gyration
RNAi	RNA interference
rt	room temperature
SAXS	small angle X-ray scattering
SDS	sodium dodecyl sulfate
TBE	tris borate EDTA
TFA	trifluoroacetic acid
U	uridine
UTR	untranslated region
Xlrbpa2	*Xenopus laevis* RNA binding protein

Chapter 1: pre-mRNA EDITING BY ADENOSINE DEAMINASES

1.1. Introduction. pre-mRNA editing by adenosine deaminases.

1.1.1. RNA editing

The Central Dogma of molecular biology, "DNA codes for RNA codes for protein," states that RNA is a mandatory intermediate in the flow of genetic information from DNA to proteins. However, since this notion was first introduced in the late 1950's it has become evident that in additon to being a carrier of genetic information, many cellular processes are regulated at the level of messenger RNA (mRNA). The process of transcription within a cell nucleus produces precursor mRNAs (pre-mRNAs) which must undergo a number of specific modifications to yield fully functional mature RNAs (Figure 1-1). The post-transcriptional processing events not only ensure that an accurate and complete message is produced, but also expand the capacity of the genome by generating divergent mRNAs with different coding potential from a single gene.[1] These processing events include 5' cap formation, polyadenylation, splicing, and RNA editing. The 5' cap formation refers to the addition of a methylguanosine base to the 5' end of the message, and is thought to play a role in defining the eukaryotic translation start site.[2] Polyadenylation is the addition of a long chain of adenosine (A) residues to the 3' end of mRNA transcripts, which protects them from RNase activity so the message can be translated into a functional protein.[3] Splicing involves a mechanism whereby the non-coding regions are excised from the message followed by ligation of the flanking sequences.[4] During RNA editing, a subtle change is introduced into the message by insertion, deletion, or chemical modification of a single base which affects the content of the message.[5]

Much more research is needed to fully understand the molecular process of RNA editing.

The process of RNA editing results in modification of the primary transcript through insertion, deletion, or base modification of a nucleotide to create a message that is different from the one encoded by the genome.[6] The first example of RNA editing was reported by Rob Benne in the 1980's as an activity that inserted or deleted uridine (U) nucleotides in the mitochondrial pre-mRNA of trypanosomes.[5] At first it was belived that such editing events are limited only to this organism. Editing by insertion or deletion is in fact prevalent in prokaryotes, however RNA editing by base modification is limited to higher eukaryotes with a developed tissue system.[6]

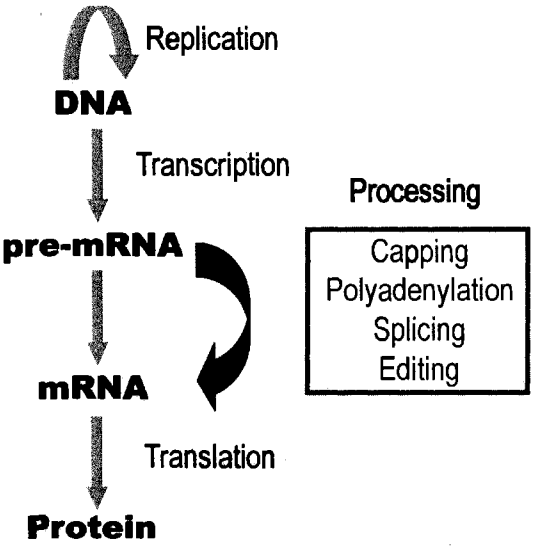

Figure 1-1. Central dogma of molecular biology. The flow of genetic information is unidirectional from DNA to proteins. Pre-mRNA must undergo a number of post-transcriptional processing steps before it can be translated into proteins.

Examples of base modifications include hydrolytic deamination at C6 of adenosine (A) or C4 of cytidine (C) which result in a change from adenosine to inosine (A to I) or cytidine to uridine (C to U) respectively.[7] Since the discovery of RNA editing, it has been shown that the base modification reactions observed in eukaryotes are catalyzed by a family of RNA-dependent deaminases.[8-11]

1.1.2. Adenosine deaminases that act on RNA (ADARs)

The activity responsible for A to I modification within RNA was discovered accidentally in 1987 by researchers working on anti-sense gene silencing in *Xenopus laevis*.[12] Upon injection of anti-sense RNA, no stable RNA:RNA duplexes with the complementary mRNA could be detected. Instead, an activity which appeared to be responsible for unwinding the RNA duplexes in *Xenopus* eggs was isolated. Subseqently, it was realized that the activity did not actually unwind the RNA:RNA duplexes but rather was responsible for introducing multiple A to I substitutions in one of the strands (Figure 1-2 A).[13] Inosine base pairs with cytidine similar to guanosine (Figure 1-2 B). Modification of multiple adenosines to inosines creates a number of I:U mismatches which replace the A:U Watson-Crick base pairs thus resulting in the loss of base pairing properties and destabilization of the duplex. Initially the activity responsible for introducing these covalent modifications was named double-stranded adenosine deaminase or DRADA but the nomenclature was subsequently changed to adenosine deaminases that act on RNA (ADARs).[8] ADAR homologues have been identified in eukaryotes, including worms, flies, fish, frogs,

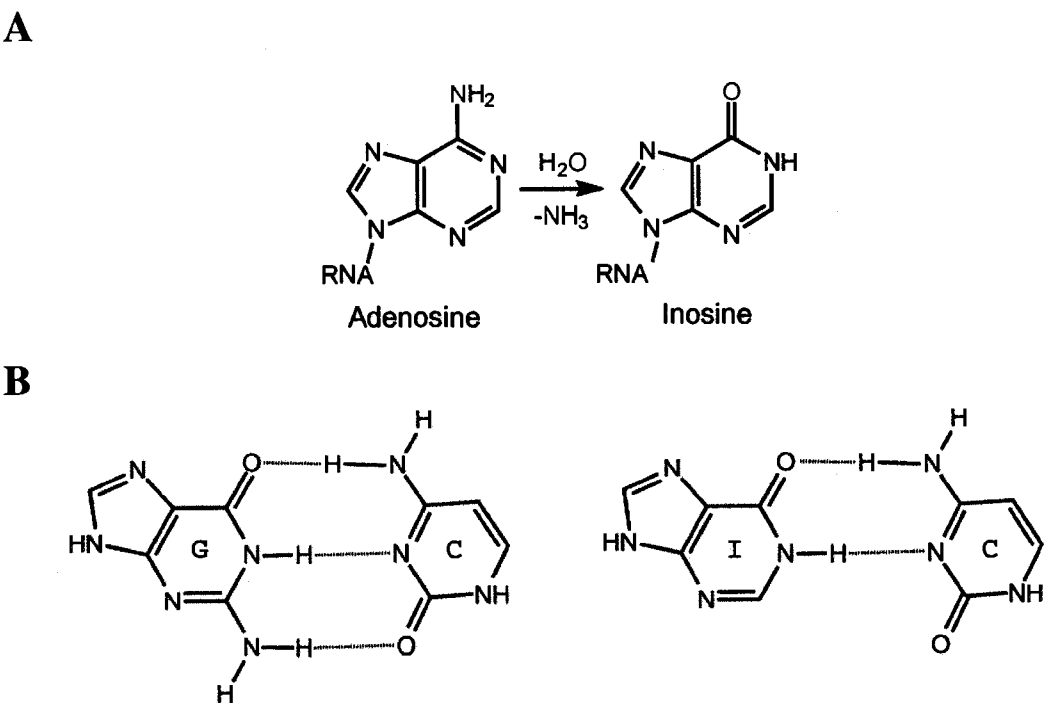

Figure 1-2. Adenosine deamination by ADARs. **(A)** ADARs catalyze hydrolytic deamination of adenosine to inosine. **(B)** Inosine forms Watson-Crick base pairs with cytidine, similar to guanosine. (Adapted from ref. 6)

birds and mammals.[14-25] ADARs from all organisms share a common modular domain structure which includes one or more double-stranded RNA binding motifs (dsRBMs) in the N-terminus and a deaminase domain in the C-terminus (Figure 1-3).[6] Organisms differ in the number of ADAR genes they express. In *C. elegans* ADR-1 and ADR-2 are expressed whereas *Drosophila* has only one dADAR.[14, 15] In mammals three different ADAR genes coding for ADAR1, ADAR2 and ADAR3 have been identified.[17, 20-25]

ADAR1 is expressed as two different splice variants: a long interferon induced protein (150 kDa) found in the nucleus and the cytoplasm and a short form (110 kDa), which is localized to the nucleus.[26] The long form of ADAR1 is thought

to participate in the anitiviral response in the host cell launched by the presence of high levels of dsRNA. ADAR1 has been shown to nonselectively deaminate the viral dsRNAs at multiple positions which in turn affects the production of fully functional viral proteins. ADAR1 is also unique among the ADAR family in that its N-terminus in addition to three dsRBMs contains putative Z-DNA binding domains.[27] The short protein isoform contains one whereas the long form contains two Z-DNA binding domains. Shortly after the discovery of ADAR1 it became apparent that some of the substrates containing A to I modification *in vivo* were not edited by ADAR1 *in vitro*. A search for other adenosine deaminases led to the discovery of ADAR2 and more recently ADAR3.[23, 25] ADAR2 is a smaller protein (80 kDa) which has four different splice variants (ADAR2a-d).[22] Isoforms ADAR2b and -2c contain a 40 amino acid insertion in the deaminase domain which has a high degree of homology to the *Alu* cassette insert. The in-frame *Alu* sequence element is found in the protein-coding exons of numerous human genes.[28] Variant -2c has a shorter C-terminus than -2b. ADAR2d is similar to -2a in that it does not contain the *Alu* cassette but it it also has a shorter C-terminus similar to variant -2c. ADAR1 and ADAR2 are expressed ubiquitously in all mammalian tissues although ADAR2 has elevated expression in the brain and heart.[24] ADAR3 has only been found in the brain and no known substrates have been identified to date.[25] The physiological significance of ADAR proteins is not fully understood as relatively few mammalian edited gene transcripts have been identified.

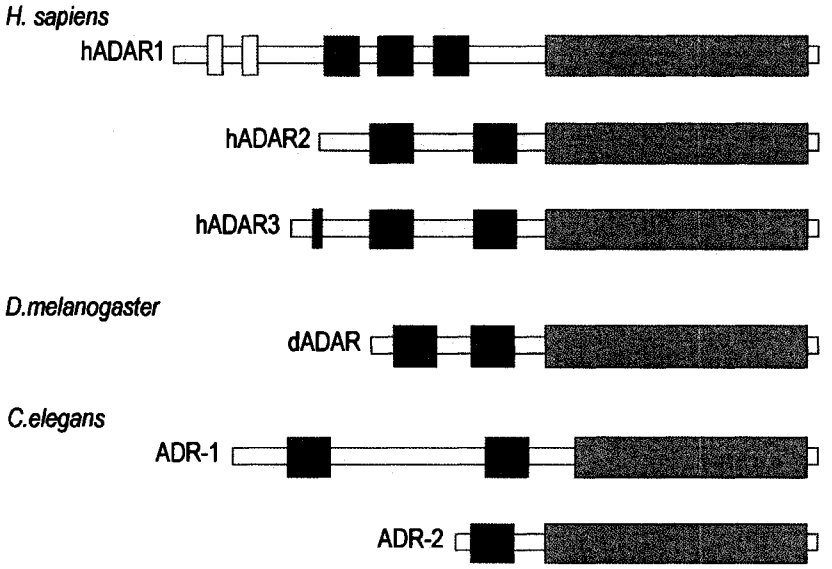

Figure 1-3. Domain maps of ADARs from selected species. Green boxes indicate the catalytic domain. Grey boxes represent approximate location of dsRBMs within the protein. Yellow boxes in the N-terminal of hADAR1 indicate location of the Z-DNA binding domains. Red box in the N-terminus of hADAR3 indicates double and single stranded RNA binding region. Major difference among ADARs from different species arise from the number and location of the dsRBMs. (Adapted from ref. 6)

1.1.3. Functional consequences of A to I editing by ADARs.

The enzymatic activity of ADARs has been studied extensively, revealing a number of important characteristics of the editing reaction as well as a small number of substrates with biological relevance (Table 1-1).[6, 29] ADARs are double-stranded RNA binding proteins (dsRBPs), therefore they bind RNA duplexes that are at least 15 base pairs in length and mostly double-stranded in nature.[13] Several of the double-stranded substrates which ADARs edit are formed by folding back of long RNAs on themselves to create large hairpin structures containing intramolecular base pairing.[30] These large structures are often imperfect dsRNA containing bulges, loops or

mismatches and are often edited at a single or limited number of adenosines.[31] Thus mismatches and loops have been suggested to be important factors which allow ADAR1 and ADAR2 to specifically deaminate some adenosines over others.[32] Perfectly dsRNA can be generated by repetitive elements, viral transcription or from sense and anti-sense RNA interactions. In a perfectly dsRNA substrate longer than 50 base pairs, ADARs indiscriminately deaminate up to 50% of adenosines.[30, 31] Regardless of whether A to I editing is specific or promiscuous it can occur in cellular pre-mRNAs including the 5' and 3' untranslated regions (UTRs), viral RNAs, and non-coding RNAs.[32-36]

Editing events within coding regions can be identified by comparison of cDNA sequences to their corresponding genomic counterparts. Such comparisons have led to the discovery of several substrates in which site-selective editing gives rise to mutiple and biologically important protein isoforms from a single gene.[33] Since inosine is recognized as guanosine by the translational machinery, editing results in a functional A to G replacement wich has the potential to alter amino acid sequence within protein coding regions of RNA.[37] In this respect, the editing may be a mechanism to enhance the protein diversity offered by the genome. In mammals, pre-mRNAs coding for receptor proteins involved in neurotransmission including the glutamate receptors and seretonin receptors are edited at multiple positions.[38]

One of the best-studied examples of A to I editing by ADARs within neuronal mRNA are the ionotropic glutamate receptor (GluR) subunits.[38] GluR subunits make up the glutamate-gated ion channels, which in turn allow the passage of monovalent or divalent cations through the plasma membrane upon activation with glutamate.

Table 1-1. Known substrates for ADAR1 and ADAR2 editing.

Edited pre-mRNA	Protein product	Editing enzyme	Physiological effect
GluR-B	Glutamate receptor subunit B	ADAR1 ADAR2	Ion permeability of the ion channel Desensitization recovery rate after ligand binding
GluR-5	Glutamate receptor subunit 5	ADAR2	Ion permeability of the ion channel
GluR-6	Glutamate receptor subunit 6	ADAR2	Ion permeability of the ion channel
5-HT_{2C}R	Seretonin receptor 2C subtype	ADAR1 ADAR2	G protein interactions
Kv1.1	Kv1.1 potassium channel	ADAR2	Channel inactivation kinetics
HDV antigenome	Hepatitis delta virus antigen	ADAR1	Expression of large or small antigen
ADAR2	ADAR2	ADAR2	Expression of inactive protein splice variant

(adapted from ref. 6)

Among the different families of GluR receptor subunits, the AMPA/kainate family is responsible for mediation of the fast excitatory component of neurotransmission. Malfunction of the glutamate-gated ion channels has been associated with a number of neurodegenerative diseases such as stroke, epilepsy, amyothrophic lateral sclerosis and Parkinson's.[39-41] The AMPA/kainate receptor family includes six distinct GluR subunits GluR-A, -B, -C, -D, -5 and -6. All GluR subunits contain four hydrophobic segments M1-M4 (Figure 1-4 A).[38] Segment M2 forms a pore-loop structure, entering and exiting the cell membrane on the intracellular side (Figure 1-4 B).[1, 38] The specific amino acid residues within this segment, which line the inner channel pore, determine the distinct ion selectivity for each of the GluR subunits.[33]

ADARs edit the pre-mRNA of GluR-B, -C, -D, -5 and -6, generating a unique pattern of editing sites for each subunit. All of the editing sites within GluR pre-mRNA occur within regions of extensive base pairing between the exon sequence

and its complementary sequence of an adjacent intron and are interrupted by secondary structure features such as single stranded bulges and loops (Figure 1-4 C).[31] Each editing event within the receptor subunit pre-mRNA generates a novel protein isoform with unique properties resulting in an altered physiological function of the channel.[41] GluR-B, -5 and -6 are edited at a critical nucleotide within the M2 segment where a neutral glutamine (Q) is changed to a positively charged argenine (R, Q/R site).[33] This single amino acid change renders the ion channels containing one of the edited subunits impermeable to calcium ions. In contrast GluR-A, C and D which are not edited at this site exhibit high calcium permeability. Interestingly, the Q/R site within GluR-B is edited up to 99% *in vivo* only by ADAR2 and not ADAR1.[31, 38] GluR-5 and -6 are edited at this site to lower extent (40% and 80% respectively).[31, 42] Moreover, editing at the GluR-B Q/R site is essential in mammals. Mice lacking editing at this site develop epilepsy and die shortly after birth.[41] The region between segments M3 and M4 of GluR-B, -C and -D has been shown to be a ligand interacting domain.[33, 43] In GluR-A to -D subunits, segment M4 is preceded by a "flip-flop" region which varies among different splice variants of these four subunits (Figure 1-4 A, B). An editing event in the pre-mRNA of this region results in an arginine (R) to glycine (G) change (R/G site) in a flexible loop which affects the desensitization recovery rate of the receptor after ligand binding.[41] Unlike the editing of the Q/R site, editing at the R/G site is not critical and it can be catalyzed by either ADAR2 or ADAR1. The extent of R/G site editing for the three GluR subunits ranges from 40 to 95%. In the rat brain, the extent of editing appears

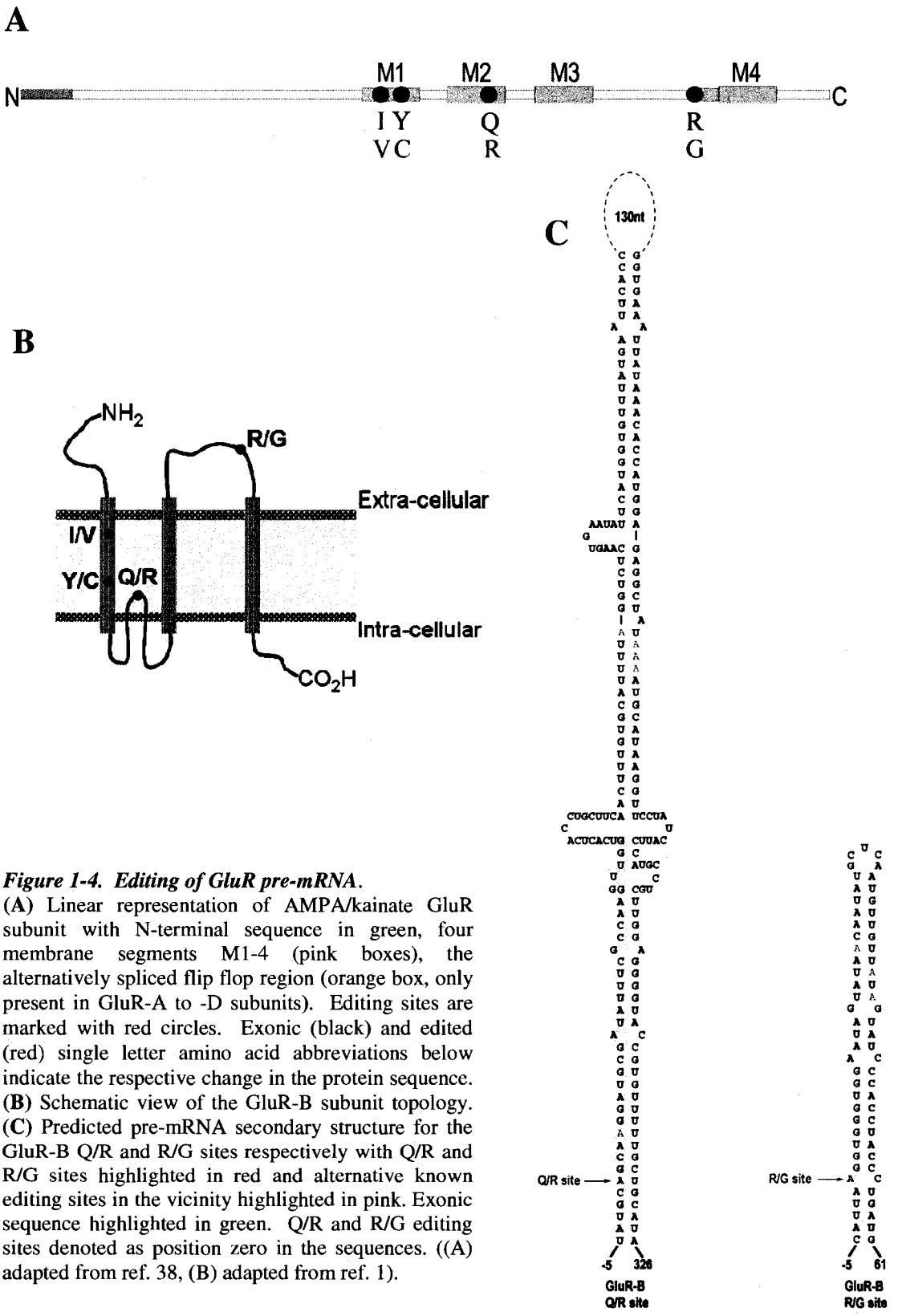

Figure 1-4. Editing of GluR pre-mRNA.
(A) Linear representation of AMPA/kainate GluR subunit with N-terminal sequence in green, four membrane segments M1-4 (pink boxes), the alternatively spliced flip flop region (orange box, only present in GluR-A to -D subunits). Editing sites are marked with red circles. Exonic (black) and edited (red) single letter amino acid abbreviations below indicate the respective change in the protein sequence. **(B)** Schematic view of the GluR-B subunit topology. **(C)** Predicted pre-mRNA secondary structure for the GluR-B Q/R and R/G sites respectively with Q/R and R/G sites highlighted in red and alternative known editing sites in the vicinity highlighted in pink. Exonic sequence highlighted in green. Q/R and R/G editing sites denoted as position zero in the sequences. ((A) adapted from ref. 38, (B) adapted from ref. 1).

to change with the different developmental stages. The levels of editing are low at the early embryonic stages but increase rapidly until postnatal day 42.[33] Furthermore, two editing sites are found in the M1 segment of GluR-6 which result in either isoleucine (I) to valine (V) or tyrosine (Y) to cysteine changes (C) (Figure 1-4 A, B).[42-44] Each of the three changes possible (Q/R, I/V and Y/C) within the pre-mRNA of GluR-6 results in fine-tuning of the calcium permeability properties of the ion channels.

However, all of the known selectively edited pre-mRNA substrates taken together do not account for the levels of inosine detected in various mammalian tissues.[45] Recently, novel computational analysis methods were developed to establish a systematic and global approach for identification of new editing sites within double-stranded substrates.[46] The mounting evidence collected with the use of these methods suggests that most of the editing actually occurs in the noncoding sequences and 5' and 3' untranslated regions (UTRs). The function of these editing events, for the most part, is poorly understood. In one instance editing within an intron is known to create an alternate splice site.[47] ADAR2 is known to edit a specific adenosine residue within intron 4 of its own pre-mRNA to generate a more proximal 3' splice acceptor with an AI dinucleotide replacing the conserved AG found in canonical 3' splice junctions. Splicing to this novel 3' site produces a catalytically inactive variant of ADAR2 and has been proposed as a self-regulation mechanism for the protein. Long, perfectly double-stranded RNAs resulting from sense-antisense RNA interactions are often promiscuously edited and contain large number of inosines. One proposed role of this extensive editing is to retain these hyperedited

duplexes in the nucleus and prevent their export to the cytoplasm where they could pose a serious threat to the cell.[48] RNA editing has also been linked to an antiviral response launched by extensive modification of the viral transcripts.[49] The interferon-induced long form of ADAR1 has been implicated in editing of a number of dsRNA viral genomes including the measles and hepatitis C viruses.[50-52] ADAR mediated deamination within noncoding regions may also affect the structure of RNA species and thus their fate in the subsequent processing events as seen in the microRNA (miRNA) pathway.[53]

The double-stranded nature of the RNA species encountered in the miRNA pathway makes them a plausible target for RNA editing. miRNAs are short non-coding RNAs which make up a class of small regulatory molecules.[53] miRNAs influence the role of various protein coding genes in multicellular organisms by directing either degradation or translational repression of their target mRNAs. These short RNAs are processed sequentially by two RNase III - type ribonucleases called Drosha and Dicer. Drosha cleaves the long double-stranded precursors in the nucleus into shorter 60-70 nt hairpins, which are then exported to the cytoplasm where Dicer generates the mature 20-22 nt miRNAs. A number of studies provide evidence for editing of endogenous miRNAs by ADARs, although they do not firmly establish their biological relevance.[54-57] In one study, precursor miRNA was shown to be edited by both ADAR2 and the short form of ADAR1 which prevented it from being cleaved by Drosha.[55] Moreover, this inosine containing pri-miRNA was rendered sensitive to cleavage by another ribonuclease Tudor-SN. Effectively, editing of the pri-miRNA resulted in lower levels of mature miRNA in the cytoplasm.

More recently, the first example of editing within a mature miRNA having a biological relevance for the organism has been reported.[57] Expression of an endogenous protein in mouse was found to be dependent on ADAR2 mediated editing of a region of a mature miRNA responsible for its target specificity. Thus editing of mature miRNAs can generate diversity in a miRNA population and modulate the number of possible targets. The growing evidence supporting the importance of the microRNA pathway as a regulatory pathway in development suggests that ADARs may also be implicated. For example, ADAR1 *null* mice are found to have severe developmental defects including rapidly disintegrating liver structure, defects in hematopoiesis and extensive apoptosis which culminate in the death between embryonic days 11 and 12.[58] One intriguing possibility is that these defects are a result of the lack of editing of some critical miRNA. Further studies are however necessary to firmly establish the role of RNA editing in this novel regulatory pathway.

1.1.4. ADAR structure

Cloning and expression of ADARs has allowed for the identification of several essential sequences as well as independently folded domains deemed necessary for the proper function of these proteins. The N-terminus of ADARs contains a nuclear localization signal (NLS), multiple copies of dsRBMs and in the case of ADAR1, Z-DNA binding domains. The NLS consists of one or more clusters of basic amino acids which target the protein to the nucleus.[22] Substrate recognition and binding is mediated by the dsRBMs.[20, 23, 25] The function of the Z-domains in

ADAR1 is not known although evidence suggests that this domain can bind both Z-DNA and Z-RNA.[59] Alternatively, the ADAR3 N-terminus includes an arginine rich region thought to be capable of binding not only double but also single-stranded RNA.[25] Proteins in the ADAR family share a common C-terminal catalytic domain evolutionarily related to cytidine deaminases from *E. coli* and ADATs (adenosine deaminases that act on tRNA).[9-11] Structural work with the ADAR family of proteins has been challenging and only recently high resolution structures have been obtained independently for the Z-DNA binding domain in complex with Z-RNA, the rat ADAR2 dsRBMs and the human ADAR2 deaminase domain.[60-62]

The interferon-induced form of ADAR1 has two Z-DNA binding domains within its N-terminus. Z-DNA domain is a nucleic acid binding domain often found in proteins which either participate in the antiviral interferon response pathway or are viral inhibitors of that pathway.[63, 64] Shortly after the discovery of the Z-DNA domain in ADAR1, it was proposed that since Z-DNA is generated by the movement of RNA polymerase during transcription, Z-domains could sequester this protein to sites of active transcription.[19] The Z-domain has Zα and Zβ subdomains. More recently it has been shown that the Zα subdomain binds both Z-DNA and Z-RNA and its presence in ADAR1 influences this protein's ability to preferentially target dsRNA substrates containing a $(CG)_6$ repeat which can exist in the Z-conformation.[59] The presence of a Z-domain within ADAR1 may also play a role in the recognition of double-stranded viral transcripts which can also adopt the Z-conformation. However, how Z-RNA binding domains affect the editing of natural stubstrates by ADAR1 has not yet been reported.

Multiple copies of dsRBMs are found in the N-terminal amino acid sequence of ADAR proteins. The known structures of other dsRBPs in complex with their substrates including the second dsRBM from *Xenopus laevis* RNA binding protein A (Xlrbpa 2), the third dsRBM from *Drosophila* Staufen protein and dsRBM from the Rnt1p RNase III homologue from budding yeast indicate that the interaction is primarily structural and not sequence specific.[65-67] In 2006 a high resolution structure was determined for both dsRBMs from rat ADAR2.[61] Both domains exhibit the canonical αββββα topology found in other dsRBM structures in which the two α-helices are packed against the surface of a three-stranded antiparallel β-sheet (Figure 1-5 A, B). The interaction between each of the dsRBMs and RNA covers a region of approximately 10 base-pairs on an A-form RNA helix. The α-helix1 and β1-β2 flexible loop in both dsRBM1 and dsRBM2 comprise the RNA binding surface. In addition, the β3-α2 loop and N-terminus of α-helix2 in dsRBM1 also play roles in the interaction (Figure 1-5 C). The major difference between the two dsRBMs comes from the positioning of α-helix1 in the structure. This is in agreement with the common view that helix1 is the least conserved region of the dsRBMs among the family of dsRBPs.[68, 69] The subtle differences between α-helix1 as well as the number and spacing of dsRBMs within the protein may be responsible for specific recognition patterns leading to different specificities and affinities for dsRNA. The model for the interaction between the the two dsRBMs and a 71 nt GluR-B R/G substrate shows that the domains bind the substrate independently and with preferences for distinct regions (Figure 1-5 C).[61] dsRBM1 binds near and

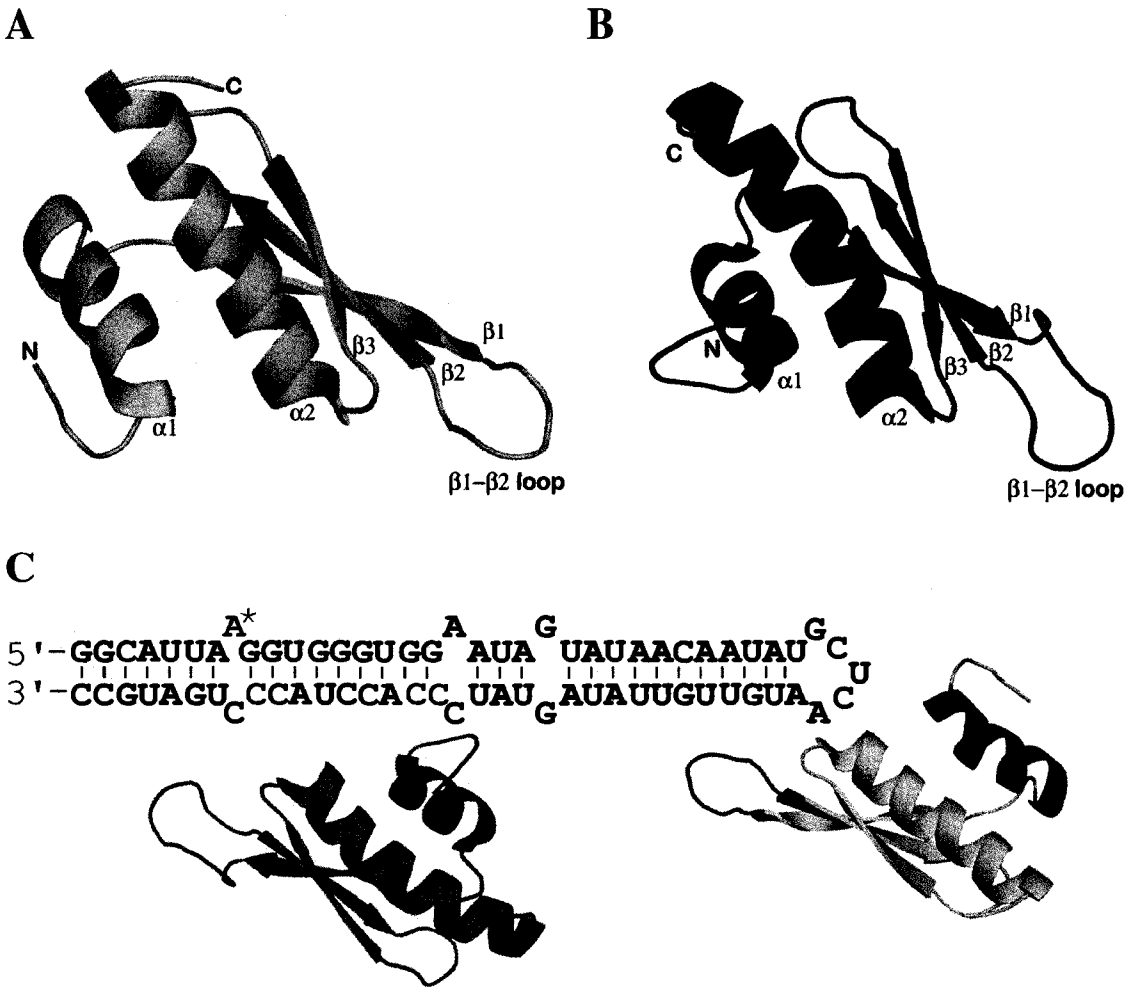

Figure 1-5. Structure of dsRBM1 and dsRBM2 from rat ADAR2. Both domains exhibit the canonical αββββα topology. (**A**) dsRBM1. (**B**) dsRBM2. (**C**) GluR-BR/G pre-mRNA with edited adenosine in green and the regions which interact with dsRBM1 in blue and dsRBM2 in red. The regions of dsRBMs which interact with the R/G substrate are highlighted in dark blue and dark red respectively. (Adapted from ref. 61)

interacts with the pentaloop structure whereas dsRBM2 binds to duplex RNA opposite to the edited adenosine and appears to recognize the two A-C mismatches in the substrate. The preference of dsRBM1 for the pentaloop is in agreement with dsRBMs from Rnt1p and Staufen proteins which bind tetraloop structures.[68, 69] The N-terminal interaction with substrate is mediated solely by the two dsRBMs in ADAR2 and not the linker between them and the dsRBMs do not dimerize on the

substrate.[61] Moreover, the deletion of either dsRBM1 or dsRBM2 results in significantly decreased editing efficiency at the edited position whereas deletion of both domains abolishes editing efficiency at the R/G site completely. However, this is somewhat in disagreement to other reports in the literature which indicate that although removal of the dsRBMs from ADAR2 diminished the efficiency and possibly specificity of editing, the catalytic domain of ADAR2 is capable of editing the GluR-B R/G RNA.[62, 70]

Until recently, most of the information about the catalytic domain of ADARs was inferred from the studies on evolutionarily related proteins such as *E. coli* cytidine deaminases (CDAs) or cytidine deaminases that act on RNA (CDARs).[11, 71] Cytidine deaminases catalyze the hydrolytic deamination of cytidine. Sequence alignments between ADARs and CDAs show that ADARs possess the highly conserved residues of two cysteines, one histidine and a glutamic acid residue essential for the catalysis reaction in CDAs. These predictions of the active site of ADARs were confirmed in 2005 by the high resolution structure of the human ADAR2 deaminase domain expressed in *S. cerevisiae*.[62] The ADAR2 deaminase domain contains the conserved "deamination motif" consisting of two α-helices and four β-strands (Figure 1-6 A). The structure shows the four conserved residues (H394, E396, C451 and C516) as well as the Zinc (Zn) ion and a water molecule all being present at the active site (Figure 1-6 B). The two cysteines, histidine and the water act as ligands for the Zn ion, and the glutamic acid which is hydrogen-bonded to the water molecule is critical for proton transfer during the deamination

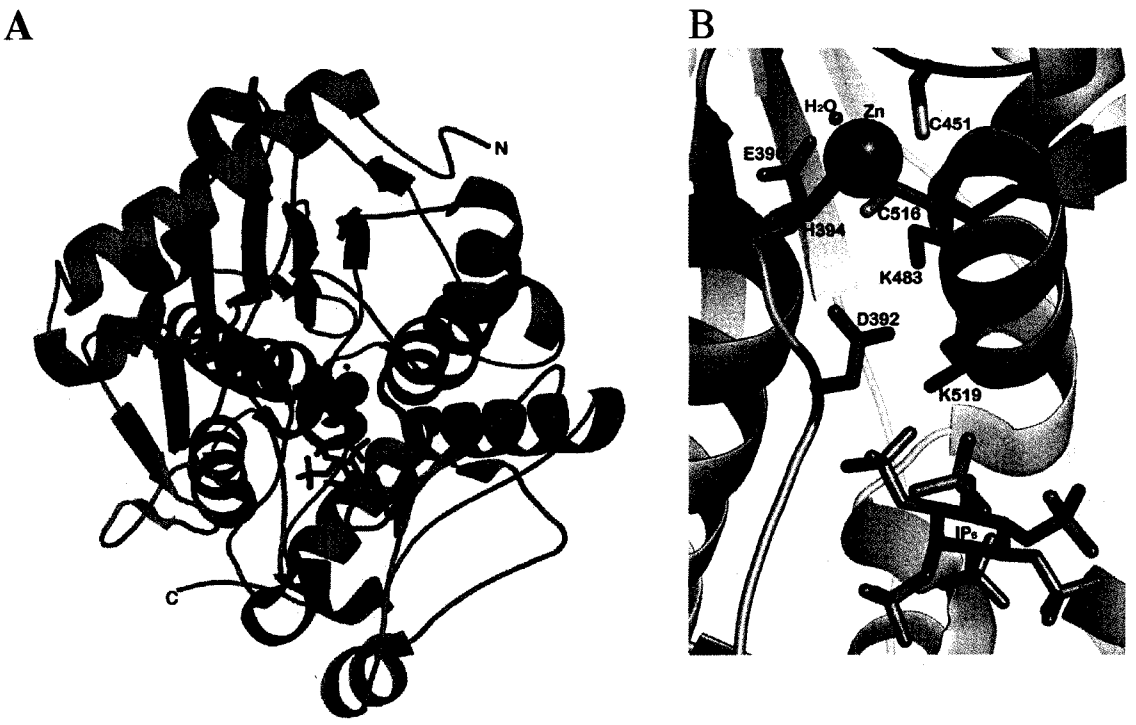

Figure 1-6. Structure of the deaminase domain of human ADAR2. **(A)** The deaminase domain of ADAR2 contains a highly conserved "deamination motif" (blue) comprising two helices and four strands. A Zn^{2+} ion located at the active site is shown in purple. Also found within the protein fold is a molecule of IP_6. **(B)** At the active site the Zn ion is coordinated by H394, C516 and C451 residues and a water molecule shown in red. The IP_6 forms a hydrogen bond network with the active site Zn through residues K519 and D392. (Adapted from ref. 62)

reaction. One major difference between the active site of CDAs and ADARs is the presence of threonine (T375) side chain positioned to interact with the 2'OH of the edited ribonucleotide in the active site.[62, 71] Models of either AMP or a known inhibitor of CDAs, zebularine, generated by applying the geometry of the active site of CDAs illustrate why ADARs catalytic domain can only accomodate adenosine and not cytidine. The side chain of T375 appears to be in position to accommodate the 2'OH of adenosine but sterically clashes with the ribose ring of cytidine's mimic zebularine. The structure of the ADAR2 catalytic domain also provided a suprising and an unusual structural feature never seen before for other members of the

deaminase family of proteins. Namely a molecule of inositol hexakisphosphate (IP_6) was discovered burried deep within the core of the protein with an extensive network of hydrogen bonds to the polar residues. IP_6 is an important signaling molecule which must have been sequestered from the internal pools in *S. cerevisiae*.[72-74] The location of IP_6 within the protein indicates that this molecule is essential for proper protein folding. The signaling properties of IP_6 suggest that it may play a role in the regulation of ADAR activity although the underlying mechanism must still be elucidated.

1.1.5. Deamination mechanism

The high degree of conservation of the active site residues between the deaminase domains of CDAs and ADARs suggests a similar catalytic mechanism for both proteins. Mutation of the residues involved in zinc binding (H394, C451 and C516) in ADARs results in a loss of activity.[75] As a consequence of the double-stranded nature of the substrate the adenosine destined for editing must be removed from the helix before it can access the active site of ADARs. It has been proposed that ADARs use a "base-flipping" mechanism also found in CDAs, where the active nucleotide is flipped out of the dsRNA before it is accomodated into the active site.[62] The conformational changes in the RNA upon binding to ADARs have been investigated by incorporation of fluorescent nucleotide 2-aminopurine (2-AP) into the substrate.[76] The fluorescence of 2-AP is quenched when the nucleotide is incorporated into the helix. Studies with 2-AP support the "base-flipping"

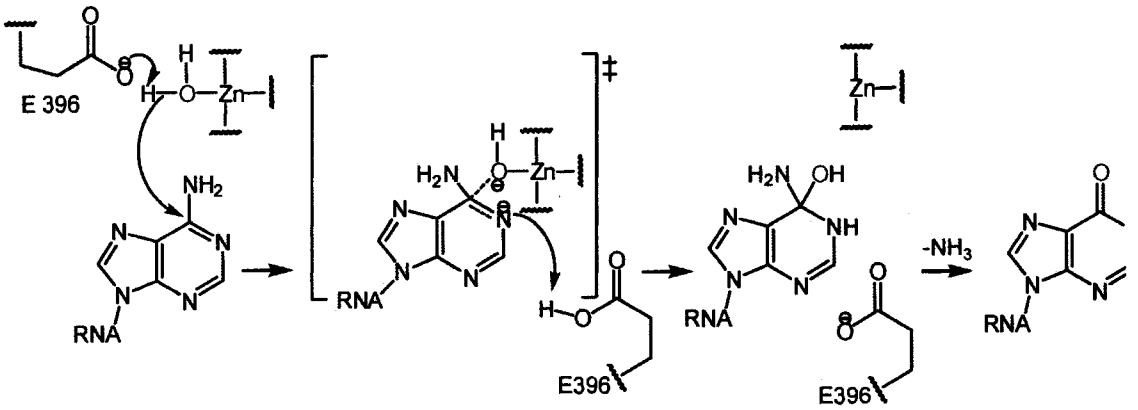

Figure 1-7. Proposed model for catalytic mechanism of adenosine deamination in RNA. (Adapted from ref. 29)

mechanism for ADAR proteins. Much less is known about the conformational changes that occur in the protein itself upon RNA binding. In a proposed model for the catalytic mechanism of ADARs the deaminase domain exists in an auto-inhibited state.[29, 70] The relief of the auto-inhibition is possible only when both of the dsRBMs can bind to the double-stranded substrate of sufficient length. Upon binding of the dsRBMs, the deaminase domain is activated and can contact the RNA to extract the reactive adenosine into the active site. In the rate limiting step, the water molecule bound by zinc is deprotonated by E396 to form a hydroxide which can then attack the C6 of the purine ring. The protonation at N1 of purine ring generates a high energy tetrahedral intermediate. Subsequent proton transfer to N6 and leaving of a molecule of ammonia results in the inosine product (Figure 1-7). The model speculates that the inhibitory sequences which keep the catalytic domain in check are located in the N-terminus of ADARs. However, no specific sequences have been identified so far. Overall, the structural details of the interaction between the catalytic domain and its substrates remain unknown.

1.1.6. Substrate recognition

Although ADARs bind tightly to any dsRNA they show preferences and selectivity for certain adenosines over others even within the context of perfectly dsRNA.[77] There are three distinct aspects of substrate recognition by ADARs which include selectivity, specificity, and the requirement for ADAR dimerization. There is an ongoing debate in the literature on the extent of contributions made by the dsRBMs and the deaminase domain to the overall selectivity and specificity of the enzyme for its substrates. The dsRBMs are associated with the enzyme's selectivity for the substrate and the C-terminal catalytic domain is linked to the enzyme's specificity.[61, 62] There is also evidence which suggests that dimerization is required for the optimal activity among the ADAR family members.[78-81]

Selectivity is dictated by the secondary structure of RNA and refers to the ability of ADARs to distinguish not only between double-stranded and single-stranded RNA but also detect secondary structural features such as loops and bulges. The dsRBMs influence the affinity of the ADARs for their substrates by determining whether the substrate is double-stranded and of sufficient length with 15 base pairs being the minimum requirement. Perfectly double-stranded substrates are edited promiscuously with up to 50% of adenosines converted to inosines, whereas substrates containing mismatches, bulges and loops are edited at specific sites.[30, 32] Thus selectivity is controlled in part by a combination of duplex length and the degree of double-stranded character of the substrate. The structural studies of ADAR dsRBMs show that within the context of specific substrates, these domains exhibit binding preferences for distinct sites on the duplex. Furthermore, disruption of these

preferred sites for binding from the substrate often adversly affects the editing at the specific adenosine.[83] What is not well understood is how, if at all, the selectivity of dsRBMs affects the selection of the editing site.

Although ADARs are generally thought of as structure specific rather than sequence specific enzymes, certain lines of evidence suggests that the deaminase domain shows sequence specificity. *In vitro* studies with ADAR1 show that it selectively deaminates adenosines positioned 3' to U=A>C>G. ADAR2 also prefers U=A>C=G as its 5' neighbours and U=G>C=A for its 3' neighbours.[84] The bases for these nearest neighbour preferences are not clearly understood but it is not likely that they arise from the favorable interactions between nucleotides adjacent to the edited adenosine and structural features around the active site in the deaminase domain of ADARs. It is also possible that base-flipping is promoted by certain combinations of the nucleotides flanking the edited adenosine.[85] Moreover, not only nearest neighbours but also base-pairing partners in the complementary strand affect the preferences of ADARs. For example, it has been shown that an A-C mismatch at the edited site is preferred over A-A or A-G mismatches or A-U base pairs for efficient editing in most cases. Further support for sequence recognition by ADARs comes from mutational analysis studies surrounding the edited adenosines.[76] Introduction of various mutations around the edited site has a dramatic effect on the site selective adenosine modification without affecting the affinity for dsRNA. In addition, a protein chimera containing N-terminal dsRBMs of ADAR1 but the C-terminal deaminase domain of ADAR2 retains specificity for substrates preferentially edited

by ADAR2, which suggests that the deaminase domain makes an important contribution to the overall specificity of the enzyme.[86]

Several groups have reported the necessity of ADARs to associate either as homo- or as heterodimers for optimal activity.[78-81] Initially the possibility of dimerization in the ADAR family originated from observation that other dsRBPs such as the RNA-dependent protein kinase (PKR) require dimerization for activity.[78] Consistent with this observation early reports suggested that two molecules of the enzyme must dimerize on a single molecule of RNA substrate for efficient catalysis.[79] In addition, both ADAR1 and ADAR2 eluted as dimeric species during gel filtration purification.[75] Also, the N-terminus of dADAR, a *Drosophila* homologue to the human ADAR2, was found to be responsible for enzyme dimerization in solution.[80] Removal of this region rendered dADAR inactive. More recently, *in vivo* fluorescence resonance energy transfer (FRET) experiments in mammalian cells found that both ADAR1 and ADAR2 form hetero- and homodimers and that this association is not dependent on the presence of RNA.[81, 82] One intriguing possibility might be that dimerization is a mechanism for controlling the efficiency and specificity of the editing reaction. In contrast to these findings, the structures of dsRBMs and deaminase domain do not show that these domains dimerize.[61, 62] It is possible that other regions of ADARs harbour the sequences responsible for dimerization but so far, no evidence of such regions exists. Thus although it is clear that ADARs have the ability to oligomerize, the conditions under which this occurs and the necessity of this event for activity are still unclear.

1.2. Thesis overview

The apparent requirement for proper function of ADARs in multiple regulatory pathways in mammals necessitates the need to understand the biochemistry of these enzymes. Although high resolution structures of the individual domains of ADAR2 are available, the overall architecture of the protein and mechanism by which it achieves its specificity remain unknown. The work in this thesis addresses structural studies of the full length ADAR2 and determination of structural features in the deaminase domain responsible for achieving substrate specificity. Chapter two presents structural studies in solution with the use of small angle X-ray scattering (SAXS) on the full length human ADAR2. Chapter three describes the expression, purification and activity assays of the deaminase domains from hADAR1, hADAR2 and hADAR3. The activity assays with the deaminase domain of ADAR3 reveal surprising information about catalytic activity of this protein. Furthermore, small angle X-ray scattering studies in solution on deaminase domains of ADAR2 and ADAR3 provide information on features not present in the high resolution structure of ADAR2's domain. The work contained within provides insight into structural organization of ADARs and possible modes of interaction between these proteins and their substrates.

1.3. References

1. Rueter, S.M., Emeson, R.B. Adenosine to Inosine Conversion in mRNA. In H. Grosjean and R. Benne (ed.) *Modification and editing of RNA*. 1998, American Society for Microbiology, Washington, D.C.
2. Banerjee, A.K. 5'-terminal cap structure in eukaryotic messenger ribonucleic acids. *Microbiol Rev* 1980, **44**:175-205.
3. Birnsteil, M.L., Busslinger, M., Strub, K. Transcription termination and 3'processing: the end is in site! *Cell* 1985, **41**:349-359.
4. Sharp, P.A. Splicing of messenger RNA precursors. *Science* 1987, **235**:766-771.
5. Benne, R., Van den Burg, J., Brakenhoff, J.P., Sloof, P., Van Boom, J.H., Tromp, M.C. Major transcript of the frame shifted Cox II gene from trypanosome mitochondria contains four nucleotides that are not encoded in the DNA. *Cell* 1986, **46**: 819-826.
6. Bass, B.L. RNA editing by adenosine deaminases that act on RNA. *Annu Rev Biochem* 2002, **71**:817-846.
7. Gerber, A.P., Keller, W. RNA editing by base deamination: more enzymes, more targets, new mysteries. *TiBS* 2001, **26**:376-384.
8. Bass, B.L., Nishikura, K., Keller, W., Seeburg, P.H., Emeson, R.B., O'Connell, M.A., Samuel, C.E., Herbert, A. A standardized nomenclature for adenosine deaminases that act on RNA. *RNA* 1997, **3**:947–949.
9. Gerber, A., Grosjean, H., Melcher, T., Keller, W. Tad1p, a yeast tRNA-specific adenosine deaminase, is related to the mammalian pre-mRNA editing enzymes ADAR1 and ADAR2. *EMBO J* 1998, **17**:4780–4789.
10. Gerber, A.P., Keller, W. An adenosine deaminase that generates inosine at the wobble position of tRNAs. *Science* 1999, **286**:1146-1149.
11. Chester, A., Scott, J., Anant, S., Navaratnam, N. RNA editing: cytidine to uridine conversion in apolipoprotein B mRNA. *Biochim Biophys Acta* 2000, **1494**:1-13.
12. Bass, B.L., Weintraub, H. A developmentally regulated activity that unwinds RNA duplexes. *Cell* 1987, **48**:607-613.
13. Bass, B.L., Weintraub, H. An unwinding activity that covalently modifies its double-stranded RNA substrate. *Cell* 1988, **55**:1089-1098.
14. Tonkin, L.A., Saccomanno, L., Morse, D.P., Brodigan, T., Krause, M., Bass, B.L. RNA editing by ADARs is important for normal behavior in *Caenorhabditis elegans*. *EMBO J* 2002, **21**:6025-6035.

15. Palladino, M.J., Keegan, L.P., O'Connell, M.A., Reenan, R.A. dADAR, a *Drosophila* double-stranded RNA-specific adenosine deaminase is highly developmentally regulated and is itself a target for RNA editing. *RNA* 2000, **6**:1004-1018.
16. Slavov, D., Crnogorac-Jurcevic, T., 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.
17. Slavov, D., Clark, M., Gardiner, K. Comparative analysis of the RED1 and RED2 A-to-I RNA editing genes from mammals, pufferfish and zebrafish. *Gene* 2000, **250**:41-51.
18. Hough, R.F., Bass, B.L. Analysis of *Xenopus* dsRNA adenosine deaminase cDNAs reveals similarities to DNA methyltransferases. *RNA* 1997, **3**:356-370.
19. Herbert, A., Lowenhaupt, K., Spitzner, J., Rich, A. Chicken double-stranded RNA adenosine deaminase has apparent specificity for Z-DNA. *Proc Natl Acad Sci USA* 1995, **92**:7550-7554.
20. 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 USA* 1994, **91**:11457-11461.
21. O'Connell, M.A., Krause, S., Higuchi, M., Husan, J.J., Totty, N.F., Jenny, A., Keller, W. Cloning of cDNAs encoding mammalian double-stranded RNA-specific adenosine deaminase. *Mol Cell Biol* 1995, **15**:1389-1397.
22. Lai, F., Chen, C.X., Carter, K.C., 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**:2413-24.
23. Melcher, T., Maas, S., Herb, A., Sprengel, R., Seeburg, P.H., Higuchi, M. A mammalian RNA editing enzyme. *Nature* 1996, **379**:460-64.
24. Gerber, A., O'Connell, M.A., Keller, W. Two forms of human double-stranded RNA-specific editase 1(hRED1) generated by the insertion of an Alu cassette. *RNA* 1997, **3**:453-463.
25. Chen, C-X., Cho, D.-S., Wang, Q., Lai, F., Carter, K.C., 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**:755-767.
26. Patterson, J.B., Samuel, C.E. 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.

27. Herbert, A., Rich, A. The role of binding domains for dsRNA and Z-DNA in the *in vivo* editing of minimal substrates by ADAR1. *Proc Natl Acad Sci USA* 2001, **98**:12132-12137.

28. Makalowski, W., Mitchell, G.A., Labuda, D. *Alu* sequences in the coding regions of mRNA: a source of protein variability. *Trends Genet* 1994, **10**:188-193.

29. Maydanovych, O., Beal, P.A. Braking the central dogma by RNA editing. *Chem Rev* 2006, **106**:3397-3411.

30. Mass, S., Melcher, T., Herb, A., Seeburg, P.H., Keller, W., Krause, S., Higuchi, M., O'Connell, M.A. Structural requirements for RNA editing in glutamate receptor pre-mRNAs by recombinant double-stranded RNA adenosine deaminase. *J Biol Chem* 1996, **271**:12221-12226.

31. Higuchi, M., Single, F.N., Kohler, M., Sommer, B., Sprengel, R., Seeburg, P.H. RNA editing of AMPA receptor subunit GluR-B: a base-paired intron-exon structure determines position and efficiency. *Cell* 1993, **75**:1361-1370.

32. Lehmann, K.A., Bass, B.L. Importance of internal loops within RNA substrates of ADAR1. *J Mol Biol* 1999, **291**:1-13.

33. Lomeli, H., Mosbacher, J., Melcher, T., Hoger, T., Geiger, J.R., Kuner, T., Monyer, H., Higuchi, M., Bach, A., Seeburg, P.H. Control of kinetic properties of AMPA receptor channels by nuclear RNA editing. *Science* 1994, **266**:1709-1713.

34. Morse, D.P., Aruscavage, P.J., Bass, B.L. RNA hairpins in noncoding regions of human brain and Caenorhabditis elegans mRNA are edited by adenosine deaminases that act on RNA. *Proc Natl Acad Sci USA* 2002, **99**:7906-7911.

35. Kumar, M., Carmichael, G.G. Nuclear antisense RNA induces extensive adenosine modifications and nuclear retention of target transcripts. *Proc Natl Acad Sci USA* 1997, **94**:3542-3547.

36. Wong, S.K., Lazinski, D.W. Replicating hepatitis delta virus RNA is edited in the nucleus by the small form of ADAR1. *Proc Natl Acad Sci USA* 2002, **99**:15118-15123.

37. Basilio, C., Wahba, A.J., Lengyel, P., Speyer, J.F., Ochoa, S. Synthetic polynucleotides and the amino acid code, V. *Proc Natl Acad Sci USA* 1962, **48**:613-616.

38. Seeburg, P.H., Higuchi, M., Sprengel, R. RNA editing of brain glutamate receptor channels: Mechanism and physiology. *Brain Res Rev* 1998, **26**:217-229.

39. Seeburg, P.H. The role of RNA editing in controlling glutamate receptor channel properties. *J Neurochem* 1996, **66**:1-5.

40. Vissel, B., Royle, G.A., Christie, B.R., Schiffer, H.H., Ghetti, A., Tritto, T., Perez-Otano, I., Radcliffe, R.A., Seamans, J., Sejnowski, T., et al. The role of RNA editing of kainate receptors in synaptic plasticity and seizures. *Neuron* 2001, **29**:217-227.
41. Seeburg, P.H., Single, F., Kuner, T., Higuchi, M., Sprengel, R. Genetic manipulation of key determinants of ion flow in glutamate receptor channels in the mouse. *Brain Res* 2001, **907**:233-243.
42. Bernard, A., Khrestchatisky, M. Assessing the extent of RNA editing in the TmII regions of GluR-5 and GluR-6 kainate receptors during rat brain development. *J Neurochem* 1994, **62**:2057-2060.
43. Stern-Bach, Y., Bettler, B., Hartley, M., Sheppard, P.O., O'Hara, P.J., Heinemann, S.F. Agonist selectivity of glutamate receptors is specified by two domains structurally related to bacterial amino acid-binding proteins. *Neuron* 1994, **13**:1345-1357.
44. Egebjerg, J., Heinemann, S.F. Ca^{2+} permeability of unedited and edited versions of the kainate selective glutamate receptor GluR6. *Proc Natl Acad Sci USA* 1993, **90**:755-759.
45. Paul, M.S., Bass, B.L. Inosine exists in mRNA at tissue-specific levels and is most abundant in brain mRNA. *EMBO J* 1998, **17**:1120-1127.
46. Sixsmith, J., Reenan, R.A. Comparative genomic and bioinformatics approaches for the identification of new adenosine-to-inosine substrates. *Methods Enzymol* 2007, **424**:245-264.
47. Dawson, T.R., Sansam, C.L., Emeson, R.B. Structure and sequence determinants required for the RNA editing of ADAR2 substrates. *J Biol Chem* 2004, **279**:4941-4951.
48. Zhang, Z., Carmichael, G.G. The fate of dsRNA in the nucleus: a p54(nrb)-containing complex mediates the nuclear retention of promiscuously A-to-I edited RNAs. *Cell* 2001 **106**:465-475.
49. Wong, S.K, Lazinski, D.W. Replicating hepatitis delta virus RNA is edited in the nucleus by the small form of ADAR1. *Proc Natl Acad Sci* 2002, **99**:15118-15123.
50. Cattaneo, R. Biased (A/I) hypermutation of animal RNA virus genomes. *Curr Opin Genet Dev* 1994, **4**:895-900.
51. Horikami, S.M., Moyer, S.A. Double-stranded RNA adenosine deaminase activity during measles virus infection. *Virus Res* 1995, **36**:87-96.
52. Polson, A.G., Bass, B.L., Casey, J.L. RNA editing of hepatitis delta virus antigenome by dsRNA-adenosine deaminase. *Nature* 1996, **380**:454-456.

53. Bartel, D.P. MicroRNAs: Genomics, Review, Biogenesis, Mechanism and Function. *Cell* 2004, **116**:281–297.

54. Pfeffer, S., Sewer, A., Lagos-Quintana, M., Sheridan, R., Sander, M., Russo, J.J., Ju, J., Randall, G., Lindnbach, B.D., Rice, C.M., Simon, V., Ho, D.D., Zavolan, M., Tuschl, T. Identification of microRNAs of the herpesvirus family. *Nat Methods* 2005, **2**:269-276.

55. Yang, W., Chendrimada, T.P., Wang, Q., Higuchi, M., Seeburg, P.H., Shiekhattar, R., Nishikura, K. Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nat Struct Mol Biol* 2006, **13**:13-21.

56. Luciano, D.J., Mirsky, H., Vendetti, N.J., Maas, S. RNA editing of microRNA precursor. *RNA* 2004, **10**:1174-1177.

57. Kawahara, Y., Zinshteyn, B., Sethupathy, P., Iizasa, H., Hatzigeorgiou, A.G., Nishikura, K. Redirection of silencing targets by adenosine-to-inosine editing of microRNAs. *Science* 2007, **315**:1137-1140.

58. Wang, Q., Khillan, J., Gadue, P., Nishikura, K. Requirement of the RNA editing deaminase ADAR1 gene for embryonic erythropoiesis. *Science* 2000, **290**: 765-1768.

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

60. Placido, D., Brown, B.A., Lowenhaupt, K., Rich, A., Athanasiadis, A. A left-handed RNA double helix bound by the Z-alpha domain of the RNA editing enzyme ADAR1. *Structure* 2007, **15**:395-404.

61. Stefl, R., Xu, M., Skrisovska, L., Emeson, R.B., Allain, F.H. Structure and specific RNA binding of ADAR2 double-stranded RNA binding motifs. *Structure* 2006, **14**:345-55.

62. Macbeth, M.R., Schubert, H.L., Vandemark, A.P., Lingam, A.T., Hill, C.P., Bass, B.L. Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. *Science* 2005, **309**:1534-1539.

63. Rothenburg, S., Deigendesch, N., Dittmar, K., Koch-Nolte, F., Haag, F., Lowenhaupt, K., Rich, A. A PKR-like eukaryotic initiation factor 2a kinase from zebrafish contains Z-DNA binding domains instead of dsRNA binding domains. *Proc Natl Acad Sci USA* 2005, **102**:1602-1607.

64. Kahmann, J.D., Wecking, D.A., Putter, V., Lowenhaupt, K., Kim, Y.G., Schmieder, P., Oschkinat, H., Rich, A., Schade, M. The solution structure of the N-terminal domain of E3L shows a tyrosine conformation that may explain its reduced affinity to Z-DNA in vitro. *Pro Natl Acad Sci USA* 2004, **101**:2712-2717.

65. Ryter, J.M., Schultz, S.C. Molecular basis of double-stranded RNA-protein interactions: structure of a dsRNA-binding domain complexed with dsRNA. *EMBO J* 1998, **17**:7505-7513.

66. Ramos, A., Grunert, S., Adams, J., Micklem, D.R., Proctor, M.R., Freund, S., Bycroft, M., St Johnston, D., Varani, G. RNA recognition by a Staufen double-stranded RNA-binding domain. *EMBO J* 2000, **19**:997-1009.

67. Wu, H., Henras, A., Chanfreau,G., Feigon, J. Structural basis for recognition of the AGNN tetraloop RNA fold by the double-stranded RNA-binding domain of Rnt1p RNase III. *Proc Natl Acad Sci USA* 2004, **101**:8307-8312.

68. Stefl, R., Skrisovska, L., Allain, F.H. RNA sequence- and shape-dependent recognition by protein in the ribonucleoprotein particle. *EMBO Rep* 2005, **6**:33-38.

69. Fierro-Monti, I., Mathews, M.B. Protein binding to duplex RNA one motif, multiple functions. *TiBS* 2000, **25**:241-246.

70. Macbeth, M.R., Lingam, A.T. Bass, B.L. Evidence for auto-inhibition by the N- terminus of hADAR2 and activation by dsRNA binding. *RNA* 2004, **10**:1563-1571.

71. Betts, L., Xiang, S.A., Short, R., Wolfenden, C.W., Carter, Jr. Cytidine deaminase: The 2·3 Å crystal structure of an enzyme: transition-state analog complex. *J Mol Biol* 1994, **235**:635-656.

72. York, J.D., Odom, A.R., Murphy, R., Ives, E.B., Wente, S.R. A Phospholipase C-dependent inositol polyphosphate kinase pathway required for efficient messenger RNA export. *Science* 1999, **285**:96-100.

73. Shen, X., Xiao, H., Ranallo, R., Wu, W.H., Wu, C. Modulation of ATP-dependent chromatin-remodeling complexes by inositol polyphosphates. *Science* 2003, **299**:112-114.

74. Steger, D.J., Haswell, E.S., Miller, A.L., Wente, S.R., O'Shea, E.K. Regulation of chromatin remodeling by inositol polyphosphates. *Science* 2003, **299**:114-116.

75. Cho, D.-S.C., Yang, W., Lee, J.T., Shiekhattar, R., Murray, J.M., Nishikura, K. Requirement of dimerization for RNA editing activity of adenosine deaminases acting on RNA. *J Biol Chem* 2003, **278**:17093-17102.

76. Yi-Brunozzi, H.Y., Stephens, O.M., Beal, P.A. Conformational changes that occur during an RNA editing adenosine deamination reaction. *J Biol Chem* 2001, **276**:37827-37833.

77. Klaue, Y., Kallman, A.M., Bonin, M., Nellen, W., Ohman, M. Biochemical analysis and scanning force microscopy reveal productive and nonproductive ADAR2 binding to RNA substrates. *RNA* 2003, **9**:839-846.

78. Dar, A.C., Dever, T.E., Sicheri, F. Higher-order substrate recognition of eIF2alpha by the RNA- dependent protein kinase PKR. *Cell* 2005, **122**:887-900.

79. Jaikaran, D.C., Collins, C.H., MacMillan, A.M. Adenosine to inosine editing by ADAR2 requires formation of a ternary complex on the GluR-B R/G site. *J Biol Chem* 2002, **277**:37624-37629.

80. Gallo, A., Keegan, L.P., Ring, G.M., O'Connell, M.A. An ADAR that edits transcripts encoding ion channel subunits functions as a dimmer. *EMBO J* 2003, **22**:3421-3430.

81. Chilibeck, K.A., Wu, T., Liang, C., Shellenberg, M.J., Gesner, E.M., Lynch, J.M., MacMillan, A.M. FRET analysis of in vivo dimerization by RNA-editing enzymes. *J Biol Chem* 2006, **281**:16530-16535.

82. Poulsen, H., Jorgensen, R., Heding, A., Nielsen, F.C., Bonven, B., Egebjerg, J. Dimerization of ADAR2 is mediated by double-stranded RNA binding domain. *RNA* 2006, **12**:1350-1360.

83. Stephens, O.M., Yi-Brunozzi, H.-Y., Beal, P.A. *Biochemistry* 2000, **39**:12243.

84. Lehmann, K.A., Bass, B.L. Double-stranded RNA Adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. *Biochemistry* 2000, **39**:12875-12884.

85. Hart, K., Nystrom, B., Ohman, M., Nilsson, L. Molecular dynamics simulations and free energy calculations of base flipping in dsRNA. *RNA* 2005, **11**:609-618.

86. Wong, S.K., Sato, S., Lazinski, D.W. Substrate recognition by ADAR1 and ADAR2. *RNA* 2001, **7**:846–858.

Chapter 2: SMALL ANGLE X-RAY SCATTERING STUDIES OF HUMAN ADAR2 IN SOLUTION

2.1. Introduction

Adenosine deaminases that act on RNA (ADARs) catalyze the hydrolytic deamination of adenosine (A) to inosine (I) within double-stranded RNA (dsRNA).[1] Inosine forms stable base pairs with cytidine (C) and consequently is recognized by the translational machinery as guanosine (G).[2] Thus editing within pre-mRNA can lead to the alteration of a codon, creation of a splice site, or affect the secondary structure of RNA. Three homologues of the protein (ADAR1-3) have been identified in mammals.[3-6] ADAR1 and ADAR2 are important for proper neuronal function and have been implicated in the microRNA and RNAi pathways.[7-12] The apparent requirement for proper function of ADARs in multiple regulatory pathways in mammals necessitates the need to understand the biochemistry of these enzymes. However, structural studies of ADARs have been limited by the availability of large quantities of pure and biologically active protein.[13]

Various expression systems have been utilized in order to produce pure and active protein. The bacterial systems have only been useful in the expression of the N-terminal portion of the ADAR proteins. Full length ADARs expressed in *E. coli* are insoluble and do not retain their biological activity.[13] The insect and mammalian expression systems require tissue culture facilities and expensive media while the protein yields are relatively low. The most successful expression to date has been achieved in yeast systems where relatively good yields of pure and active ADAR2 have been obtained.[14] The expression of ADAR1 is also possible in yeast although yields of pure protein are much lower than for ADAR2. For thus far unknown

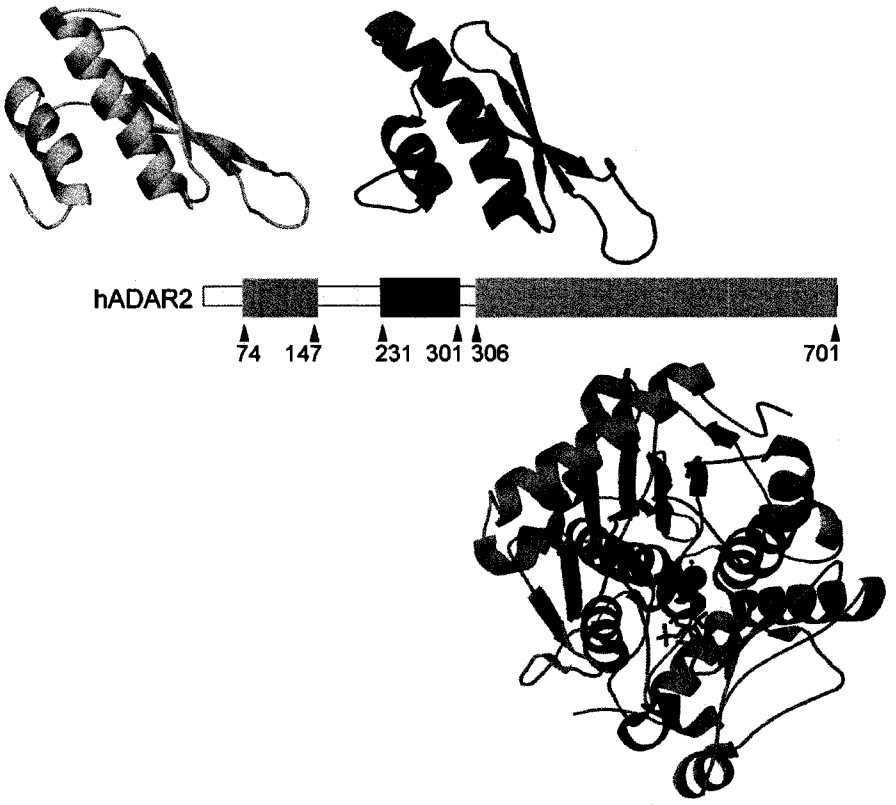

Figure 2-1. Domain organization of hADAR2. High resolution structures for rat dsRBM1 (blue) and dsRBM2 (red) and human deaminase domain (green). (Adapted from ref. 3, 14, 19)

reasons, the expression of full length ADAR3 has only been possible in the mammalian expression systems.[6]

The human ADAR2 (hADAR2) is a 80 kDa protein (701 amino acids) with a well defined modular domain organization (Figure 2-1).[3] In its N-terminus (aa 1-298), ADAR2 contains two double-stranded RNA binding motifs (dsRBMs) and the C-terminus (aa 299-701) contains the catalytic deaminase domain. A high resolution X-ray crystal structure of the deaminase domain of human ADAR2 has been determined from protein expressed in *S. cerevisiae*.[14] The structure confirmed the presence of a highly conserved "deamination motif" and a zinc atom at the active site of the enzyme as predicted previously by the alignments with closely related cytidine

deaminases (CDAs) and adenosine deaminases that act on tRNA (ADATs).[15-18] High resolution structures have also been determined for the dsRBM1 and dsRBM2 from rat ADAR2 and show the common αββββα motif found in other dsRBMs.[19-20]

In a proposed model for the catalytic mechanism of hADAR2 the deaminase domain exists in an auto-inhibited state.[21, 22] The relief of the auto-inhibition is possible only when both of the dsRBMs can bind to a double-stranded substrate of sufficient length. Upon binding of the dsRBMs, the deaminase domain is activated and can contact the RNA to extract an adenosine into the active site. The evidence for this model comes solely from biochemical analysis and no structural data in support of this model is available. The high resolution structures of the individual domains do not provide information on their spatial arrangement and how this arrangement affects ADAR2's ability to interact with its substrates. To gain a better insight into the overall structure of ADAR2 we have decided to study this protein in solution by the method of small-angle X-ray scattering (SAXS).[23] Although SAXS is a low resolution technique (~ 10-20 Å), in the case of ADAR2 it can complement the high resolution atomic structures already available by providing information on the overall shape of the molecule and possibly the spatial arrangement of the domains contained within.

2.1.1. Small-angle X-ray scattering (SAXS) data collection

Several characteristics render SAXS as an attractive technique for studying ADARs.[23, 24] The SAXS experiments require modest sample preparation and offer rapid data collection in solution near physiological conditions. The standard data

collection times vary from 5 to 100 seconds on a standard sample size of 15 μL.[25] Furthermore, the quality of the data can be evaluated at the time of collection which is useful when dealing with limited amounts of material. SAXS methodology takes advantage of the elastic scattering of X-rays by electrons when a sample is irradiated by a collimated, monochromatic beam of X-rays (Figure 2-2 A).[24] Depending on the spatial arrangement of electron pairs within the sample the X-rays scattered can interfere either constructively or destructively in accordance with Bragg's Law ($n\lambda = 2d\ sin\theta$).[25] Thus, SAXS is a contrast method in which the scattering signal is derived from the difference in the average electron density between the macromolecule of interest and the solvent surrounding it. Unlike X-ray crystallography which conveys a single snapshot of the macromolecule in a distinct conformation imposed by the ordered crystal structure, during SAXS all configurations of the macromolecule are sampled as it freely tumbles in the solution.

2.1.2. SAXS data analysis

In a SAXS experiment, information about the quality of the data and its content can be obtained directly from the scattering curve. The scattering curve ($I(s)$) is a function of the momentum transfer and is defined as $s = (4\pi \sin\theta)/\lambda$ where 2θ is the scattering angle and λ is the wavelength of the incident beam.[23] The curve is radially symmetric around the primary beam and since it depends on the size and shape of the molecule it is unique for each macromolecule (Figure 2-2 B, C).[23, 24] At low values of the parameter s and thus at low resolution, the scattering can be

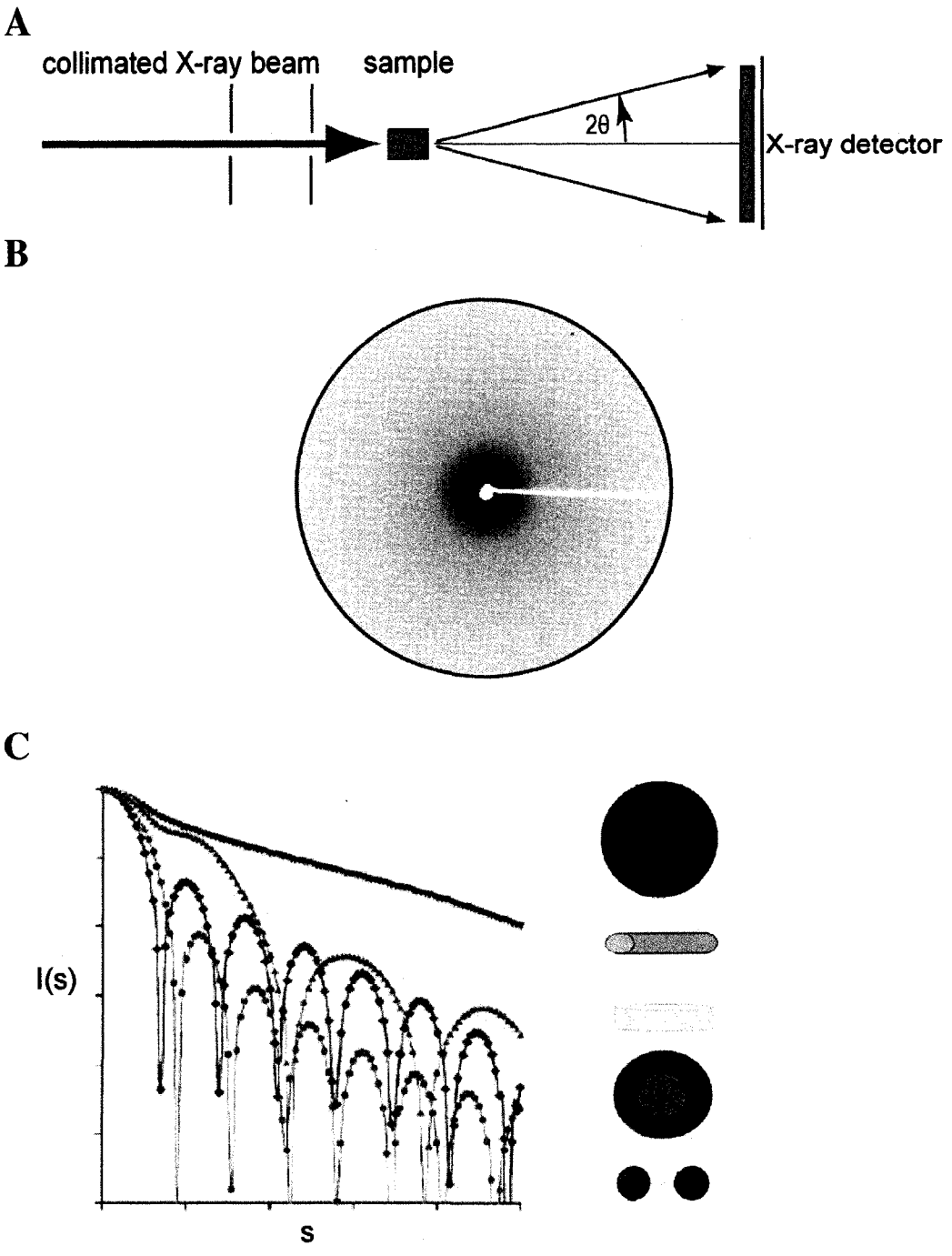

Figure 2-2. SAXS experimental setup. **(A)** The instrumental setup for SAXS involves irradiating a sample with a collimated X-ray beam and measuring the scattered X-rays with a detector. The scattering angle is equal to 2θ. **(B)** Solution scattering is continuous and radially symmetric with a typical resolution limit of 20Å. **(C)** The scattering curve is a function of the momentum transfer s and is dependent on the shape of the molecule. The overall appearance of the scattering curve gives an indication of the shape of the macromolecule (Adapted from ref. 23, 24)

described by the Guinier equation $I(s) = I(0)\exp[-(s^2 R_G{}^2)/3)]$.[26] The first step in data validation is the analysis of the Guinier plot of *log (I(s))* versus s^2 which should be a straight line and can be used to calculate two important parameters for the macromolecule of interest namely radius of gyration (R_G) and the forward scattering intensity (*I(0)*) .[24] The R_G is a size parameter and is described as the square root of the average squared distance of each scatterer from the particle center. The value of intensity of the scattered X-ray at angle s equal to zero (*I(0)*) is defined as the square number of electrons in the scatterer and must be obtained by extrapolation from the plot since it coincides with the direct beam. Unlike R_G, *I(0)* is not dependent on the shape of the macromolecule and thus it is useful in determination of its molecular weight. The Guinier approximation however, is only valid in the small angle range $s<1.3/R_G$. In addition to the Guinier plot several other parameters can be used in the data evaluation.

Another graphical representation of the scattering data is the Kratky plot which can also serve as a predictor of inherently flexible or unfolded regions within the macromolecule.[24] The Kratky plot is a graph of $s^2 I(s)$ values plotted against s where I is the intensity and s is the momentum transfer. For compactly folded proteins of globular shape the Kratky plot is a parabola that approaches zero at higher angles (Figure 2-3). However, for unfolded proteins, the $s^2 I(s)$ approaches a constant maximum value. For partially unfolded proteins the Kratky plot may initially have a parabolic shape but tends to tail-off at higher angles.

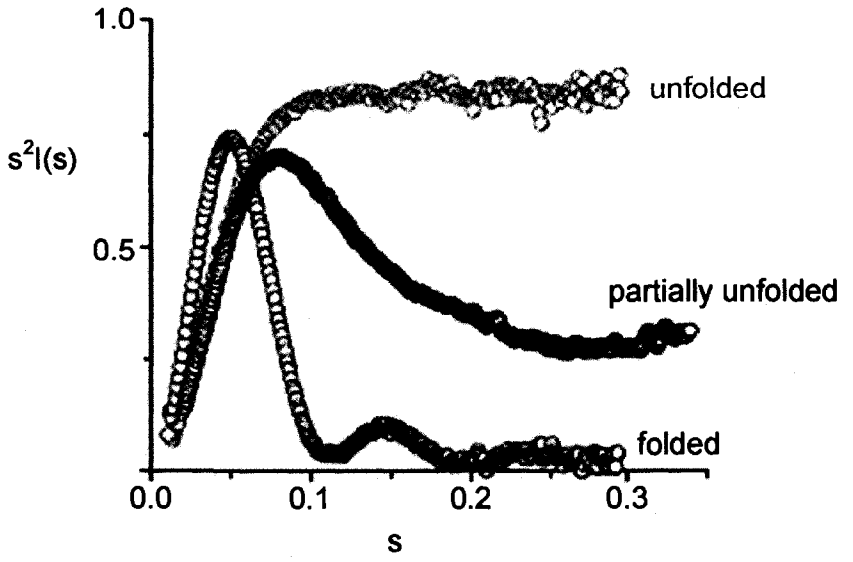

Figure 2-3. The Kratky plot. For folded globular proteins the Kratky plot is a parabola (blue graph). For unfolded proteins the Kratky plot tends to tail-off at higher angles (red graph). (Adapted from ref. 24)

The pair-density distribution function ($P(r)$) allows for calculation of the maximum dimensions of the macromolecule and it gives an indication of the overall shape.[27] Since X-ray scattering from macromolecules is a function of their electron density the scattering curve is a Fourier transform of the auto-correlation function of the electron density of the form $I(s) = 4\pi \int_0^{D_{max}} \left\langle \rho(\vec{r}) * \rho(-\vec{r}) \right\rangle \frac{\sin(sr)}{sr} r^2 dr$ where I describes the intensity, $\rho(r)$ is the electron density, D_{max} is the maximum dimension of the macromolecule and $\vec{r}$ is a spatial position vector. The pair-density distribution function is described by the equation $P(r) = r^2 \left\langle \rho(\vec{r}) * \rho(-\vec{r}) \right\rangle$ and its graphical representation illustrates the number of electrons in a macromolecule that are a distance r away from one another. The program GNOM constructs the distribution function and allows for the calculation of the parameters D_{max} and R_G.[28] It also

provides a total estimate (Q-score) value describing the goodness of fit of the raw data. Moreover, visual inspection of the *P(r)* plot also gives an indication of the overall shape of the macromolecule (Figure 2-4).

2.1.3. Model building approaches in SAXS

Once the initial SAXS data analysis is complete and the sample is deemed as a monodisperse system not dependent on interparticle interferences, the data can be used for *ab initio* shape prediction.[24] Several programs have been developed for *ab initio* structure determination (Figure 2-5). The program DAMMIN attempts to generate a bead model to fill a volume in agreement with the experimental scattering.[29] Initially, a sphere with diameter D_{max} is filled with densely packed

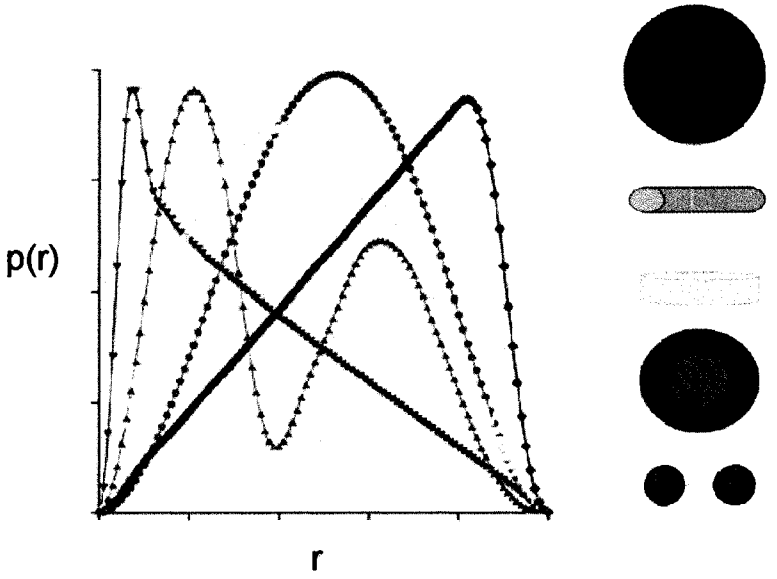

Figure 2-4. Pair-density distribution function P(r). The visual inspection of the *P(r)* plot gives an indication of the overall shape of the macromolecule. The color of the curve corresponds to the respective shape. For example, symmetrical bell shaped curves (red) are indicative of globular proteins whereas curves with the maximum skewed to the extreme left and decaying fast are indicative of elongated particles such as DNA or RNA. (Adapted from ref. 23)

spheres of much smaller radii called dummy residues (DR). Each DR is initially randomly assigned as either belonging to the macromolecule or to the solvent. The DRs are randomly reassigned during multiple rounds of simulated annealing that maximize the agreement between the DR model and the transformed data. In GASBOR, the number of beads in the DR model is based loosely on the number of residues in the macromolecule of interest. The DR are then placed in a spherical search volume and randomly reorganized during the simulated annealing steps to fit the transformed data.[30] Both methods use constraints such as smoothness, connectivity and particle symmetry to decrease the search volume, computational time and the data-to-parameter ratio. In addition GASBOR constrains the beads to form a chain-compatible model in which each bead must have two neighbours within 3.8 Å.[30] This constraint alone improves the resolution and reliability of the model and is especially useful when working with proteins containing flexible regions. The model building programs must be run multiple times (10-15 runs) on the same scattering data. Each run starts with a different random arrangement of beads and tries to build a model that fits the scattering data. A program package called DAMAVER is subsequently used to align, average, and filter the generated models to produce the most reliable model.[31] The *ab initio* envelopes generated by programs such as GASBOR and DAMAVER can then be used as a starting point for further model building.

Since SAXS provides only low resolution data, on its own it has limited applications. However, it can be a powerful technique when used in combination with

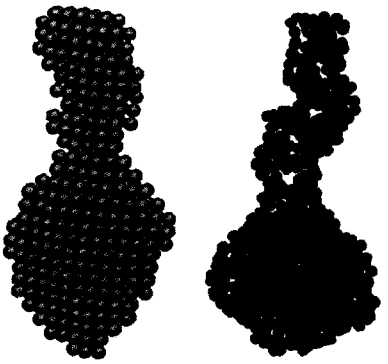

Figure 2-5. Ab initio models generated from SAXS data. Both DAMMIN and GASBOR are used to generate low resolution envelopes from the scattering data. DAMMIN uses a "dummy residue model"(red model) whereas GASBOR uses a "chain-compatible model" (blue model) (Adapted from ref. 24)

other structure determination techniques. A number of methods have been developed which can be used to accurately build models for multidomain proteins and large complexes by incorporating SAXS data with information from other methods such as crystallography or NMR. The choice of the method depends on the complexity of the macromolecule or complex being studied and on the question one is trying to answer. If a high resolution structure for the known macromolecule of interest is known it can be docked into the envelope using programs such us SUPCOMB or SITUS (Figure 2-6 A).[32, 33] If only the atomic structures for the individual domains are known, rigid body refinement programs can be applied to perform a global search of the optimum configuration of subunits that will fit the scattering data. For example, the program SASREF can simultaneously fit scattering from the individual domains with the experimental data and thus correlate their spatial arrangement (Figure 2-6 B).[34] The program CREDO can build missing domains in

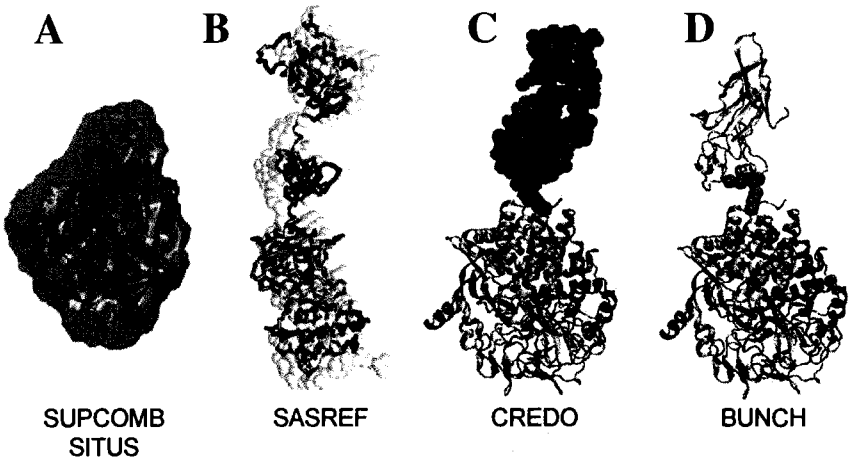

Figure 2-6. Rigid body modeling with the use of SAXS data and atomic resolution structures. **(A)** If available the atomic resolution structures can be docked into the low resolution envelopes with the use of programs such as SUBCOMB or SITUS. **(B)** The program SASREF determines the spatial arrangement of different subdomains within a macromolecule and attempts to fit the arrangement to the scattering data. **(C)** and **(D)** Programs CREDO and BUNCH are used to build in missing domains and flexible regions based on the information provided by SAXS and atomic structures. (Adapted from ref. 24)

macromolecules where the high resolution structure of one or more domains is already known (Figure 2-6 C).[35] On the other hand, the program BUNCH can account for highly flexible regions such as long linkers and loops (Figure 2-6 D).[34] SAXS can also provide insight into dynamic processes such as complex formation or ligand-induced conformational changes by observing changes in the values of R_G and D_{max} parameters.[36] Thus a major strength of SAXS is its ability to provide an overall structure including both architectural and conformational characteristics at a resolution range of 50 to 10 Å which can then be used for docking of high resolution structures and filling in the missing links, loops, or subdomains.

This chapter describes expression and purification of full length human ADAR2 for the purpose of structural studies in solution by small-angle X-ray scattering. Although hADAR2 protein could be produced at sufficient levels from a

mammalian cell line, the protein was aggregated and did not produce good quality data in SAXS experiments. Nevertheless, hADAR2 expression from *S. cerevisiae* yielded significant amounts of pure and active protein. The quality of the SAXS data collected on this sample was significantly improved, meeting the criteria for a monodisperse sample. The analysis of the data provided parameters required for the *ab initio* reconstruction of models for hADAR2. The models generated were used for docking of the high resolution structures of individual domains. The SAXS data was also used to perform rigid body refinement with the use of available high resolution structures in hope of gaining insight into the spatial arrangement of these domains in hADAR2.

2.2. Results and Discussion

2.2.1. Protein expression and purification

Expression of hADAR2 in yeast has several advantages over expression in mammalian cells. Genetic manipulation is easy in yeast systems, and the induction times are relatively short (~6-12 h) which is beneficial in expression of proteins which are toxic or unstable. In addition, the yeast genome does not encode for ADAR proteins and thus there is no contamination from native editing activity. Moreover, the growth media is relatively inexpensive and large scale expression is easy to achieve. A number of protocols for hADAR2 expression in yeast have been studied. We were particularly interested in the *S. cerevisiae* expression system, although the previously reported system required the use of a bioreactor vessel which was not

readily available to us.[13, 14] Consequently, we have cloned N-terminal histidine (10X HIS) tagged hADAR2 into the pYeDP60 vector which maintains the insert of interest in rich media and at high cell densities. The pYeDP60 vector contains a galactose inducible promoter GAL10-CYC1 which is fully repressed in the presence of glucose during the growth phase. When the glucose supply is exhausted and cells rely on ethanol utilization for growth; galactose is added, triggering full induction during the subsequent stationary phase. The strain DSY 904 is also deficient in several proteases making it a plausible expression system for hADAR2 as this protein has been shown to be extensively proteolyzed in the *P. pastoris* yeast expression system, leading to poor protein yields. This expression system has allowed us to achieve cell densities (10-12 g of wet cell pellet per 1 L of culture) comparable to those obtained during growth in the bioreactor vessel.[13]

The detailed purification protocol is described in the Materials and Methods part of this chapter (Section 2-4). The theoretical estimated molecular weight for a monomeric hADAR2 is 76.6 kDa. During size exclusion chromatography analysis on a Superdex 200 the elution volume of the hADAR2 peak was determined to be 154 mL (Figure 2-7). Relative to the elution of standard molecular weight markers hADAR2 appears to elute as a protein with molecular weight of 162 kDa which would suggest that the protein exists in solution as a dimer. There are several conflicting reports on the oligomeric state of ADAR proteins during purification. Nishikura and coworkers reported that both hADAR1 and hADAR2 produced in a Baculovirus Sf9 expression system elute as a dimeric species during size exclusion chromatography, while hADAR3 is monomeric.[38] However, hADAR2 expressed in

S. cerevisiae appears to elute as a monomer during size exclusion chromatography.[13] *In vivo* fluorescence resonance energy transfer experiments in mammalian cells provide convincing evidence for ADAR1 and ADAR2 hetero- and homodimerization. These studies also indicated that the dimerization is mediated by the protein's N-terminus and is not dependent on the presence of dsRNA.[39, 40] Thus it is tempting to speculate that dimerization can also occur *in vitro* under favorable conditions. The yields of protein varied among different preparations ranging from 0.1 -1.0 mg of protein from a liter of original yeast culture.

2.2.2. Protein activity assays

In order to test the activity of the full length hADAR2 we used two well established methods, namely Nuclease P1/thin layer chromatography (TLC) and electrophoretic mobility shift assay (EMSA). Both assays utilized 78 nucleotide GluR-B R/G pre-mRNA substrate uniformly labeled with α-^{32}P (Figure 2-8 A). Briefly, the editing assay involved incubation of a constant amount of the R/G RNA (5 nM) substrate with varying amounts of hADAR2 (1-100 μM), followed by digestion with Nuclease P1 and analysis of the resulting nucleotide 5'-monophosphates by TLC (Figure 2-8 B). The detection of inosine in the edited versus unedited substrate indicated that hADAR2 from *S. cerevisiae* has catalytic activity similar to proteins produced from other previously studied expression systems.[13, 38] In addition, EMSA was performed in order to visualize the complexes which form between the R/G RNA substrate and hADAR2. The incubation of

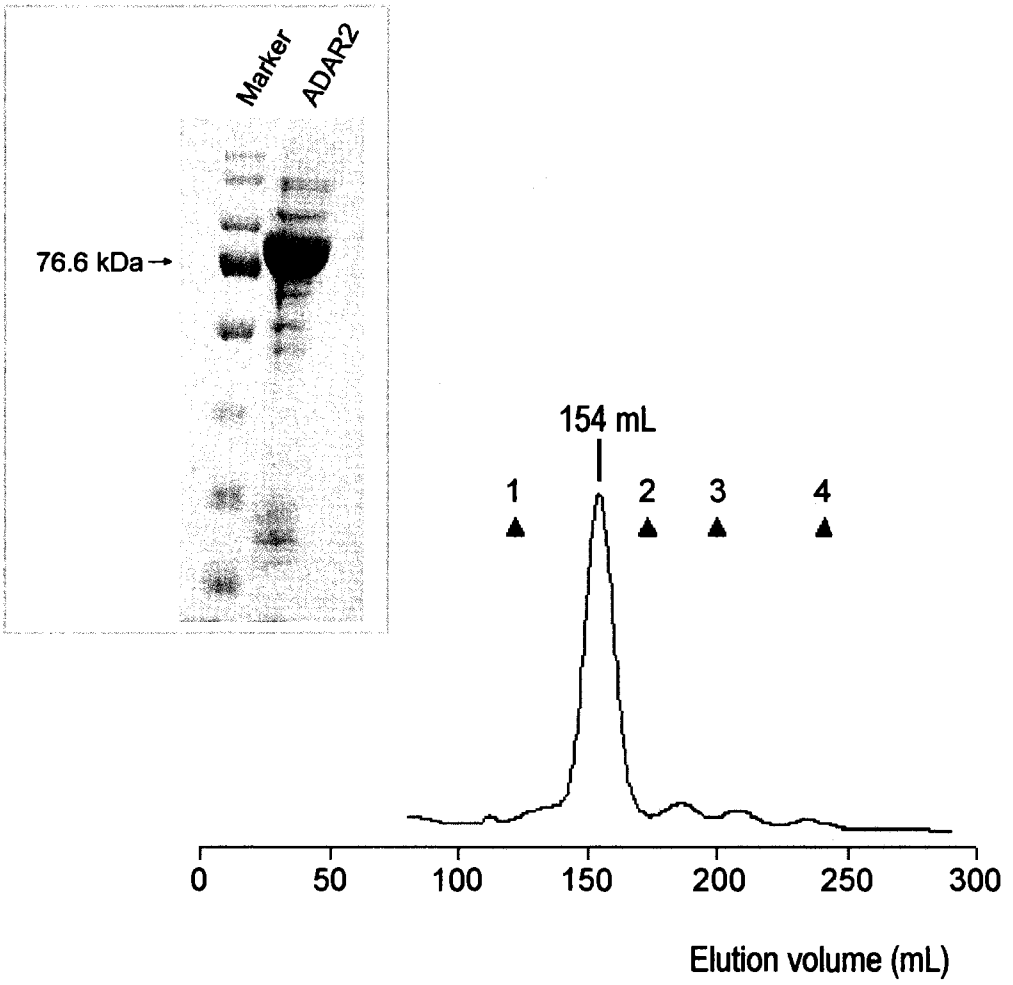

Figure 2-7. ADAR2 purification. The UV trace of hADAR2 eluted from a Superdex 200 26/60 size exclusion column. The elution volume of hADAR2 is 154 mL suggesting hADAR2 exists as a dimer under these conditions. (Size exclusion buffer contained 20 mM Tris, pH 8.0, 200 mM NaCl, 1 mM BME, 5% (v/v) glycerol). The elution peaks of molecular weight standards are also indicated (black triangles) and numbered 1-4 . (1) catalyse 232 kDa, 122.1 mL; (2) aldolase 158 kDa, 173.1 mL; (3) bovine serum albumin 67 kDa, 200.6 mL; (4) chymotrypsinogen 25 kDa, 241.0 mL. Inset: 12% SDS PAGE analysis of the pure hADAR2 with estimated molecular weight of 76.6 kDa.

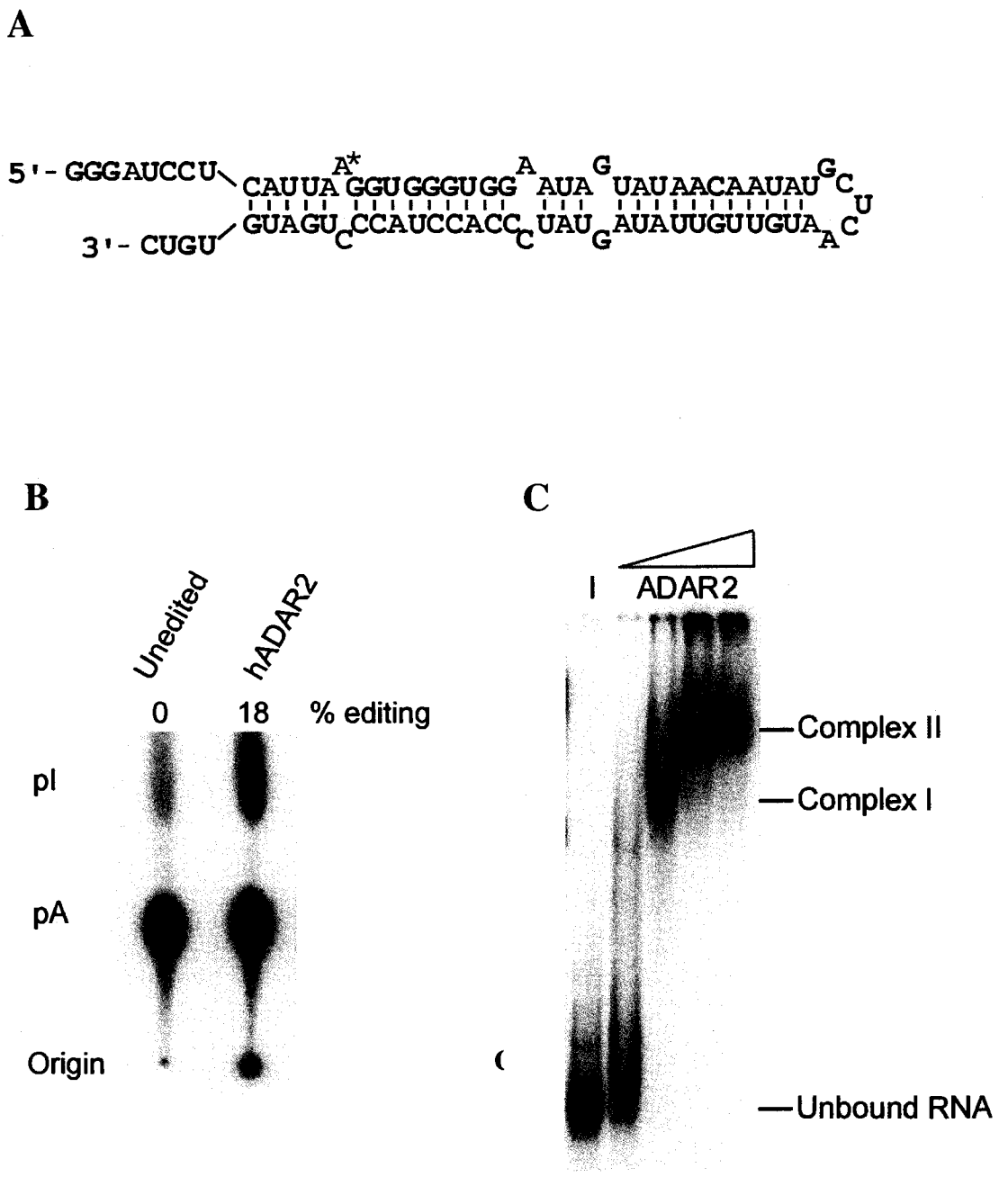

Figure 2-8. Activity assays with hADAR2 from S. cerevisiae. **(A)** Model 78 nucleotide GluR-B R/G RNA substrate. α-^{32}P labeled adenosines are shown in blue and R/G site is indicated by the red asterisk. **(B)** Editing Nuclease P1/TLC assay. Presence of α-^{32}P labeled inosine is visible in the edited versus the unedited sample. **(C)** EMSA. Input lane is indicated by I. Lanes 2-5 show increasing concentration of hADAR2 (1-150 nM) incubated with constant amount of substrate (5 nM). First and second complexes are indicated by the roman numeral I and II respectively.

hADAR2 (1-150 nM) with its substrate RNA (5 nM) and subsequent non-denaturing complexes are formed (Figure 2-8 C). Our lab has previously reported that in the first complex, one monomer of hADAR2 is bound to one molecule of dsRNA whereas in the second complex two molecules of hADAR2 dimerize on one molecule of RNA.[37] It has also been suggested that this dimerization is not RNA dependent, and is required for full activity of the protein.[38, 39] With the use of methodology established in our laboratory the equilibrium dissociation constants were calculated to be 6 and 23 nM for complexes I and II respectively.[37]

2.2.3. Low resolution reconstruction of hADAR2 structure from SAXS [41]

Data collection and analysis

Our lab has established a protocol for stable expression and purification of hADAR2 containing an N-terminal FLAG (F-hADAR2) tag in the HEK 293 mammalian cell line (Wu. T., MacMillan, A.M. Unpublished results). Since SAXS experiments do not require large amounts of material we attempted to collect data on recombinant protein purified from these cells. However, analysis of the scattering curve generated during a 20 second exposure indicated that the F-hADAR2 sample was up to 35% aggregated (Figure 2-9).[42] Since aggregation can adversely affect the entire data set no further processing was performed and we focused our efforts on expression and purification of hADAR2 from yeast.[24, 42]

The low-angle scattering data was collected on a sample (15 μL) of hADAR2 (1-701) at 2.7 mg/mL concentration in SAXS buffer (20 mM Tris-HCl, pH 8.0,

200 mM KCl, 5% glycerol and 1 mM BME). Multiple exposures of varying lengths of time (6 sec and 60 sec) collected on the same protein sample solution superimposed well (Figure 2-10). Furthermore, scattering curves collected at concentrations of 2.7 and 1.4 mg/mL superimposed well in the low angle region indicating that the samples were monodisperse, ie. the scattering was not dependent on the interparticle interactions in the solution. The Guinier plot had good linear fit in the range $s.R_G < 1.3$ indicating that the sample was not aggregated (Figure 2-11). The radius of gyration (R_G), determined from the Guinier plot in the program PRIMUS for data points 5 to 26 was 44.2 Å.[42] The program GNOM was used to calculate the pair-density distribution function, *P(r)*. The function has a slightly asymmetric bell shaped curve with a maximum at 50 Å and D_{max} of 145 Å, characteristic of a globular protein (Figure 2-12). The *P(r)* curve had a slight shoulder at 70 Å and a short tail at higher values of r which may be indicative of the presence of a distinct, slightly flexible domain. We have evaluated D_{max} multiple times by testing a range of values ($50 < D_{max} < 300$) and observing the changes in the total estimate of goodness of fit provided by the GNOM program. Consistently, D_{max} of 145 Å gave the best total estimate score of 0.867. The Kratky plot of the scattering data is a parabola providing further evidence of a globular, well folded protein.

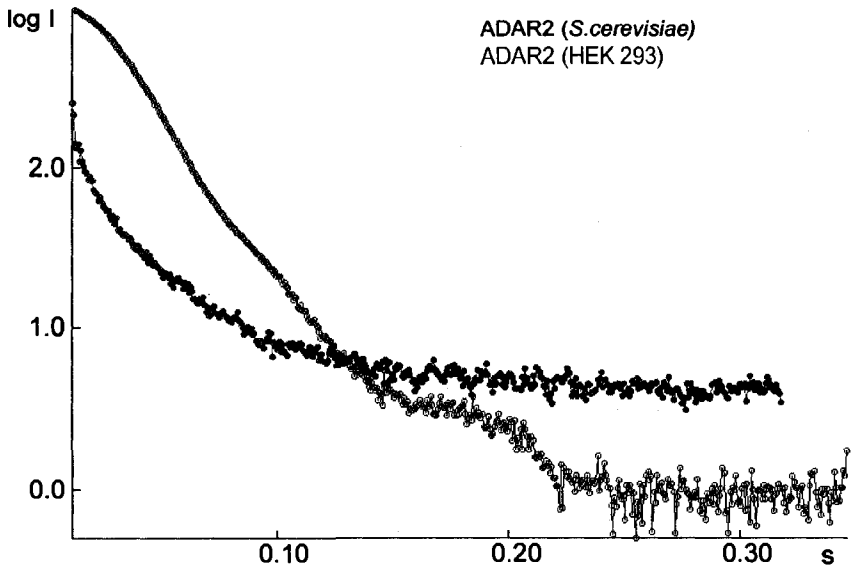

Figure 2-9. Scattering curves for hADAR2 from S. cerevisiae and HEK 293 cells. F-ADAR2 (red) is featureless and approaches infinity at low values of *s*. The scattering curve from ADAR2 expressed in yeast (blue) has discrete values for log I in the low angle region and shows features in the higher angle regions.

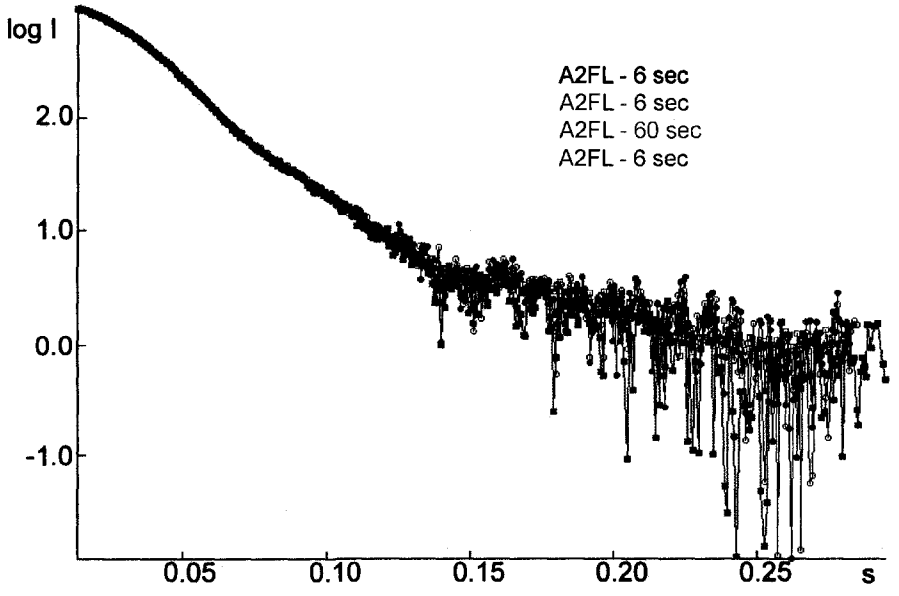

Figure 2-10. Scattering curves for hADAR2. Overlay of four scattering curves collected on hADAR2 sample (15 μL) at concentration of 2.7 mg/mL. The four scattering curves overlap well over the whole range of *s* indicating that the sample is not susceptible to radiation damage.

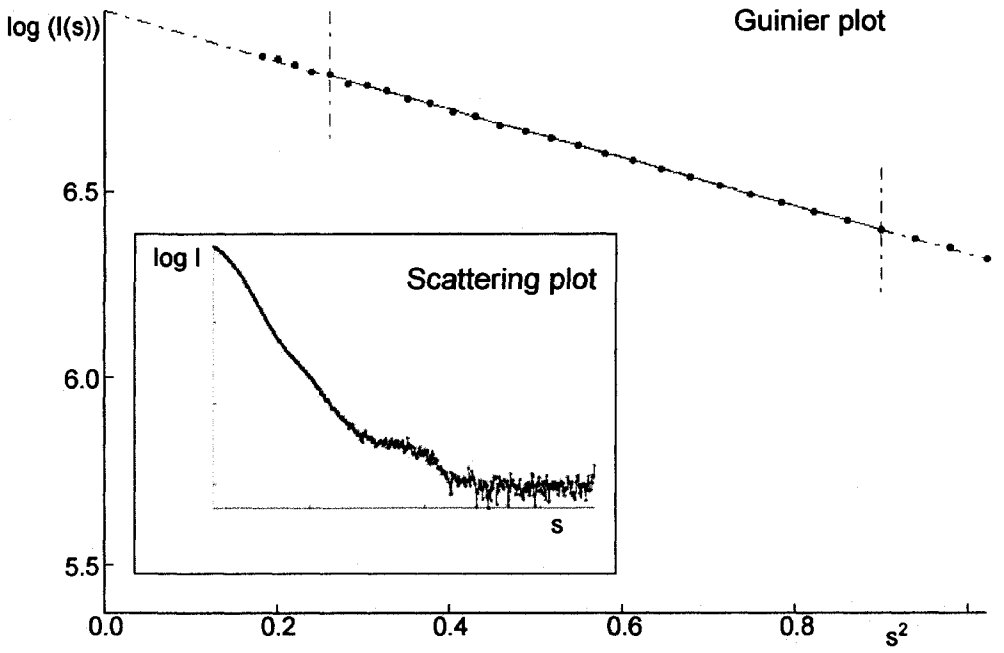

Figure 2-11. Guinier approximation for ADAR2. The data shows good fit to linearity for points 5-26 in the low angle range ($s.R_G < 1.3$). R_G was determined to be 44.2 Å with the Guinier approximation in PRIMUS.

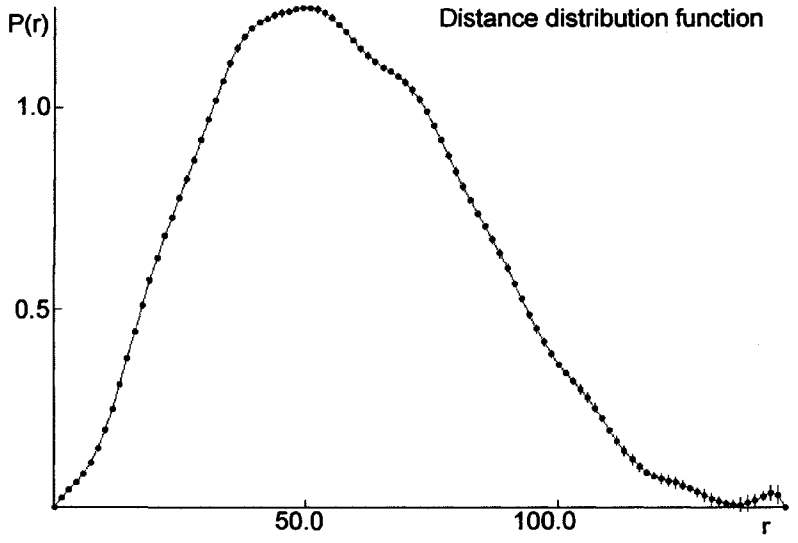

Figure 2-12. The pair-density distribution function for ADAR2. The P(r) curve is bell-shaped indicative of globular protein.

To compare the experimentally derived parameters R_G and D_{max} for hADAR2 to theoretically calculated values of proteins with similar molecular weight we have performed an exhaustive search of the Protein Data Bank (PDB) for proteins containing between 680 to 720 amino acids.[43] The search returned 137 PDB files which were imported into the program CRYSOL.[44] CRYSOL generated theoretical scattering curves from the atomic resolution structures from which values for R_G and D_{max} were calculated for each file. The average values for R_G and D_{max} were found to be 28 and 94 Å respectively for this subset of proteins, compared to those obtained experimentally for hADAR2 of 44 and 145 Å. As determined by the $P(r)$ function and the Kratky plot, hADAR2 is a folded globular protein. Thus the larger than average D_{max} value determined for hADAR2 suggests that the macromolecule is a dimer in solution.

Ab initio structure reconstruction

The *ab initio* reconstruction program GASBOR was used to generate a model of hADAR2 by fitting the experimental scattering data in reciprocal space.[30] The same model building approach was used for the short exposure (6 sec) and long exposure (60 sec) data sets. The overall model for each data set was generated by performing 15 individual GASBOR runs with each run starting the molecular dynamics calculation at a different random arrangement of dummy residues (DRs) in the search volume. Three subsets, each containing 15 individual models, were generated from the same set of data by imposing the following symmetry restraints: P1 with 701 DRs in the search volume (subset 1); P2 with 1402 DRs (subset 2); or P1

with 1402 DRs (subset 3). Subsequently, all models within each subset were aligned, averaged and filtered using the software package DAMAVER. The program SUPCOMB was used to calculate pair-wise normalized spatial discrepancy (NSD) values among all models within each subset. The NSD value is a measure of dissimilarity between the models. Therefore, the model with the smallest NSD was designated as the reference model. The remaining models within each subset were then aligned against the reference model. During this alignment process, the models with an NSD value significantly different from the average NSD were discarded. The superimposed models were averaged. This average was mapped onto a grid of densely packed beads. The program DAMFILT was used to remove regions where the bead density did not meet the threshold limits necessary to generate a particle with the expected excluded particle volume. The quality of each of the three final models reconstructed for each data set was evaluated by comparing the goodness of fit parameter χ^2. Both the short and the long exposure data sets produced comparable models with agreeable χ^2 for each model subset. The worst average χ^2 values (5.5) were observed for the model with P1 symmetry and 701 DRs. However, the χ^2 values for both P1 and P2 symmetry restrained models containing 1402 DRs fit the scattering data, with average χ^2 values of 1.9 and 1.8 respectively. Given that the gel filtration analysis and the experimentally determined D_{max} value for hADAR2 suggest that this protein is a dimer, the model with P2 symmetry restraint was chosen for further model building (Figure 2-13).[31]

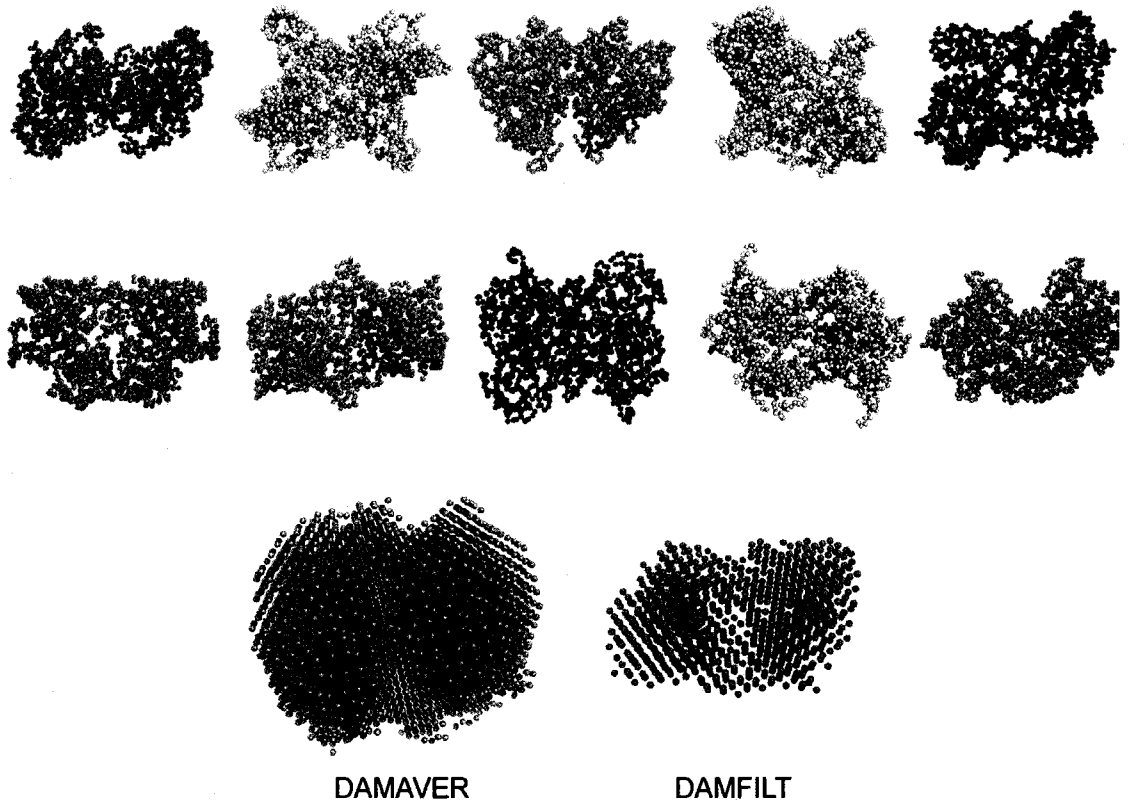

Figure 2-13. Ab initio reconstruction in GASBOR. The first two rows represent individual models for 10 out of the 15 runs conducted with GASBOR with P2 (1402 DR) symmetry restraint. The third row shows the space filling model of all 10 runs aligned (left) and filtered (right) with DAMAVER and DAMFILT respectively.

The program SITUS was used to build an envelope representation of the averaged and filtered GASBOR generated bead model (Figure 2-15 A). The overall shape of the envelope is globular with two distinct lobes oriented with respect to each other in a head-to-tail fashion. We have arbitrarily labeled the head and tail portion of each lobe to aid in the orientation (Figure 2-15 B). The tail has a maximum width of 56 Å and it narrows towards the head which has maximum dimension of 32 Å. The maximum dimension of each lobe from head-to-tail is 85 Å. The two lobes are

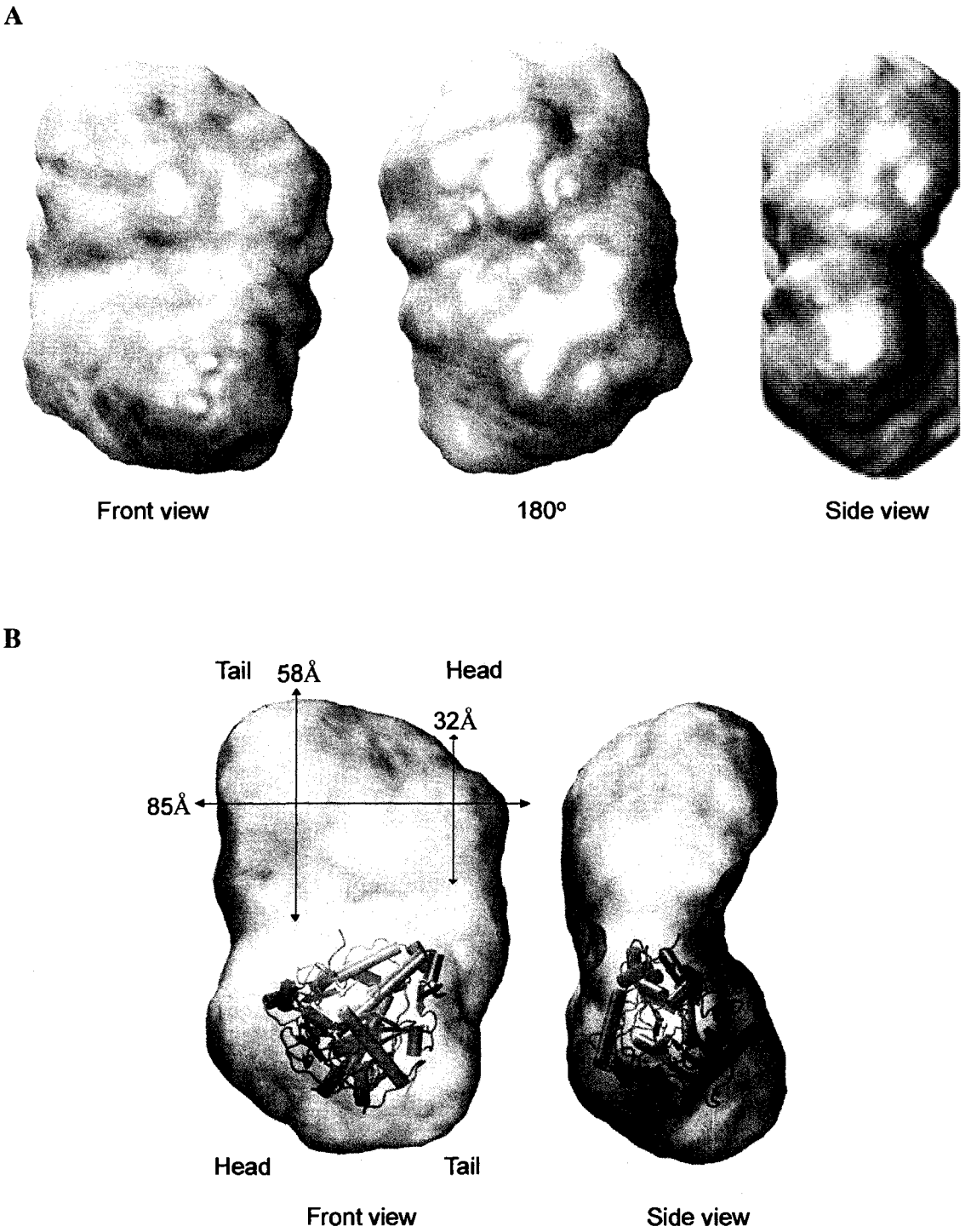

Figure 2-14. Envelope generation and rigid body docking. **A.** Low resolution envelope generated by the program SITUS from the GASBOR bead model. **B.** The high resolution structure of hADAR2 deaminase domain was docked into the low resolution envelope using SITUS. The tail and head of each lobe was assigned arbitrarily to aid with the orientation.

separated by a groove spanning the molecule. This groove is approximately 50 Å long and 28 Å wide. A-form RNA has approximately 11 base pairs per turn and a pitch of 30 Å.[45] Therefore, the dimensions of the groove could possibly accommodate up to 16 base pairs of an A-form RNA duplex. SITUS was also used to dock the high resolution structure of the deaminase domain into the envelope to compare the relative size of the full length protein (Figure 2-15 B). The maximum dimension of the globular deaminase domain structure is 56 Å which fits well with the dimensions of the tail portion of the envelope model.

Rigid body modeling

The program BUNCH was used to model the two dsRBMS and the deaminase domain by a combination of rigid body modeling and *ab initio* model building against the X-ray scattering data.[35, 46] In this model building approach the high resolution structures of the two dsRBMs and the deaminase domains were fixed and moved as rigid bodies while searching for the possible conformations of the interconnecting flexible linker sequences against the X-ray scattering data. Each linker sequence was represented as an interconnected chain composed of dummy residues with a spacing constraint of 3.8 Å. Penalties were imposed on the overlap between the domains, appropriate distribution of bonds and dihedral angles and fit to R_G values. Similarly to the *ab initio* structure reconstruction performed with GASBOR, the model building was repeated for the short exposure (6 sec) and long exposure (60 sec) scattering data (Figure 2-15 A, B). In addition, a model was also generated for a scattering data set

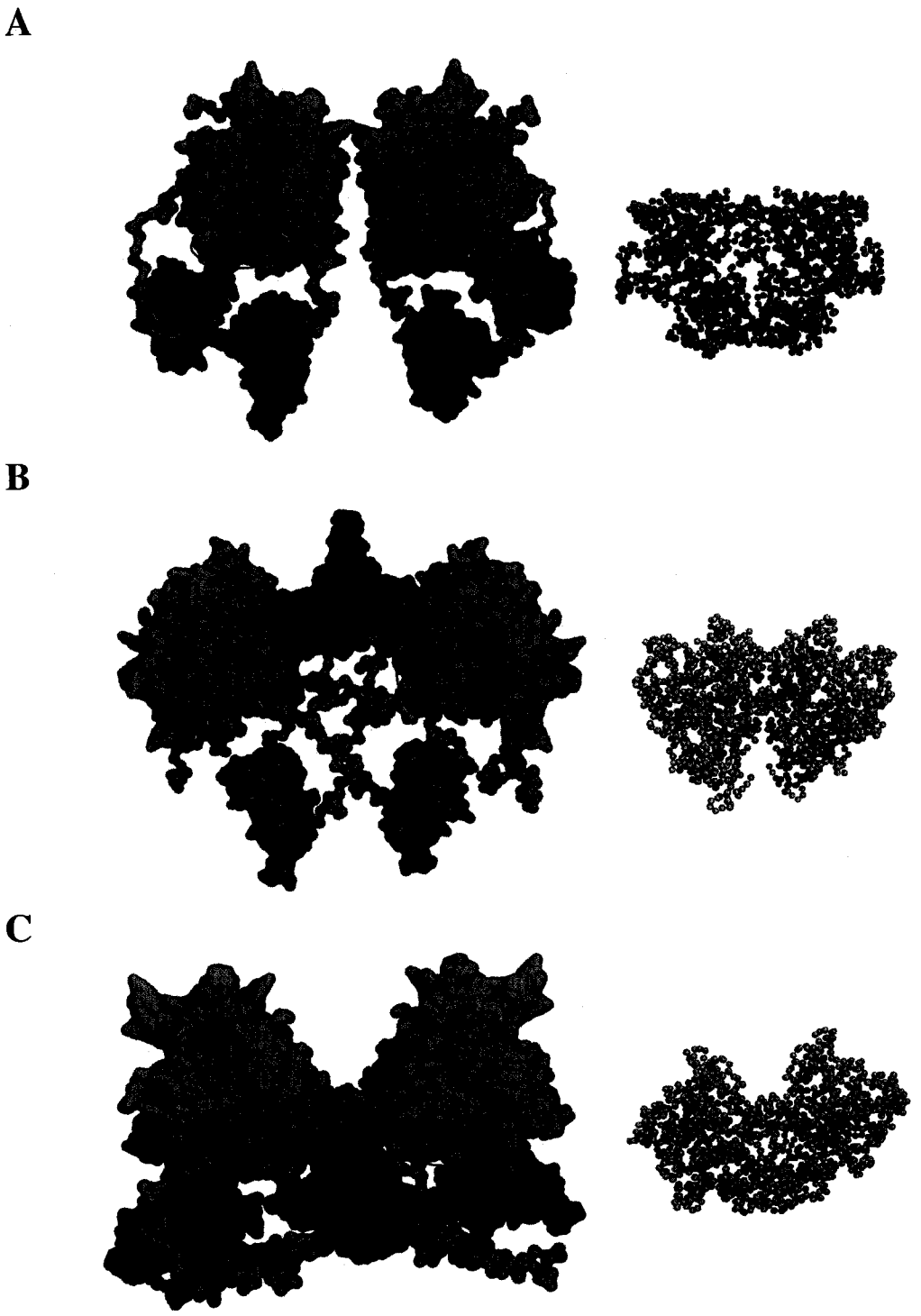

Figure 2-15. Rigid body modeling of hADAR2. Three individual models generated with the program BUNCH against the scattering data from **(A)** short exposure (6 sec), **(B)** long exposure (60 sec) and **(C)** short and long exposures merged. The deaminase domains are indicated in green, dsRBM1 in blue and dsRBM2 in pink. The approximate location of the active site and RNA binding surface on the deaminase domains is represented by the red ovals. The bead space filling models on the right in orange, pink and grey are individual GASBOR generated *ab initio* models which closely resemble the BUNCH generated models.

in which the short exposure and the long exposure data sets were merged (Figure 2-15 C). The configuration of the domains with respect to each other and their interconnecting linkers provided by BUNCH is representative of the average conformation that best fits the data.

Although the GASBOR generated *ab initio* models for the hADAR2 are the same for the short and long exposure data sets, the BUNCH models are considerably different (Figrue 2-15). The model generated against the short exposure scattering data has the deaminase domains from each monomer placed side by side with the dsRBMs obstructing the potential RNA binding surface on the deaminase domain (Figure 2-15 A). In the model generated on the long exposure data set the dimer interface is formed by the dsRBM2 and the linker regions from each monomer (Figure 2-15 B). In this configuration the active sites are on the inside of the dimer interface and the RNA binding surface is obstructed by the presence of the dsRBM2 and the linker regions. The merged data set yielded yet a different model, in which the dimer interface is formed by the dsRBM1 and intervening linker region from each monomer (Figure 2-15 C). The average χ^2 values were 2.4, 2.0 and 1.6 for the models generated against the short exposure, long exposure and the merged data respectively. In each case the RNA binding surface is obstructed by the presence of the dsRBMs and the intervening linker regions. Each of the three models closely resembles one of the individual models generated during GASBOR *ab initio* modeling. Interestingly, the average model generated by GASBOR has two lobes with distinctive head and tail portions. The models produced by BUNCH consistently place the deaminase domains in a head to head rather than head to tail fashion.

Furthermore, the attempts to dock the BUNCH generated models into the *ab initio* envelopes have not been successful. Although the individual domains can be accommodated within the envelope the flexible linkers interconnecting the domains cannot. The low resolution of SAXS combined with the presence of inherently flexible regions within the structure most likely limit the ability of BUNCH to accurately determine the configuration of the domains in three dimensions. According to the *in vivo* and *in vitro* dimerization studies on ADAR2 the models in Figure 2-15 B and C are more probable.[39, 40] Yeast-two hybrid system and fluorescence energy transfer experiments support a protein model in which the N-terminus containing the two dsRBMs and a flexible linker is responsible for protein dimerization.[40] The BUNCH generated models thus can be used as a starting point for mutant design and protein crosslinking experiments to further our understanding on the spatial arrangement of the domains within ADAR proteins.

2.3. Conclusions

A new variation of the *S. cerevisiae* expression system for hADAR2 has been designed which does not require the use of a bioreactor vessel for cell growth. The implementation of pYeDP60 vector which maintains selection at high cell densities and in rich medium has allowed for achieving cell densities similar to those obtained in the bioreactor expression. Consequently, good yields of pure and active ADAR2 were obtained for the purpose of structural studies with SAXS. Although scattering data was collected on both the F-hADAR2 expressed in mammalian cells and hADAR2 purified from the *S. cerevisiae* the latter yielded much better quality

SAXS data. The high levels of aggregation of the F-hADAR2 may be due to the presence of the FLAG tag. Thus one possibility that remains to be explored is to introduce a cleavage site in our current construct of F-hADAR2 that would allow for the removal of the tag prior to SAXS experiments. The scattering data obtained from the yeast purified hADAR2 was evaluated as monodisperse and did not show sensitivity to radiation or concentration dependence. The Guinier plot showed a good fit to linearity. The R_G and D_{max} were determined to be 44.2 Å and 145 Å respectively. The pair-distance distribution function was a bell-shaped curve indicative of a globular protein. The curve had a slight shoulder at approximately 70 Å which may imply the presence of a distinct domain with certain degree of flexibility. The Kratky plot was a hyperbola which further suggests that the protein is mostly folded in solution. The high quality data has allowed for generation of a low resolution solution structure for hADAR2. The final *ab initio* model with imposed P2 symmetry constraints fits the data with average χ^2 values of 1.8. The SITUS generated envelope for the model shows two distinct domains aligned in a head-to-tail fashion and separated by a distinct groove. As determined, the dimensions of the groove could accommodate up to 16 base pairs of A-form helix. The high resolution structure of the deaminase domain fits well into the tail portion of the envelope model. The rigid body modeling with BUNCH yielded three different models each with different spatial arrangement of the domains within the envelope. The models fit the data with average χ^2 of 2.4, 2.0 and 1.6 for the short exposure, long exposure and merged data sets respectively. Moreover, although BUNCH was able to build in the missing interconnecting linkers they did not fit well into the *ab initio* generated

envelope. Further experiments are required to accurately determine the spatial arrangement of the domains with respect to each other. The BUNCH generated models provide a starting point for mutant design of hADAR2 for protein crosslinking experiments.

2.4. Materials and Methods

General considerations

All reagents used were purchased from Sigma Aldrich unless otherwise stated. Primers for the generation of hADAR2 construct and RNA substrates were purchased from IDT. All media used in the yeast transformations and culture growth were autoclaved. Glucose, galactose and sorbitol solutions were sterile filtered. Protein purification steps were performed at 4 °C or on ice and after each step the proteins were analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and when necessary western blot analysis with Penta-His HRP® conjugate antibody (Qiagen).

2.4.1. FLAG-hADAR2 (F-hADAR) expression and purification from HEK 293 cells

The human ADAR2 coding gene was cloned into the p3XFLAG-CMV™-10 vector for over-expression. HEK 239 cells were plated in 10 cm tissue culture dishes and transiently transfected with 8 μg plasmid DNA using Lipofectamine 2000® transfection reagent (Invitrogen) according to the manufacturer's protocol and harvested 24 hours post-transfection. Cells were incubated in 500 μL lysis buffer

(50 mM Tris-HCl, 150 mM NaCl, 1 mM EDTA, 1% Triton-X100, pH 7.4) for 30 min and centrifuged for 10 min at 150,000 *g*. The supernatant was incubated with 20 µL ANTI-FLAG® M2 affinity gel (Invitrogen) prepared according to the manufacturer's protocol for 2 hours and centrifuged for 30 sec at 8,000 *g*. The supernatant was removed and the gel was incubated with 40 µL wash buffer (50 mM Tris, 200 mM NaCl, pH 7.4) containing 150 ng/µL 3X FLAG peptide for 30 min and centrifuged for 1 min at 8,000 *g* to elute F-hADAR2 from the gel. The protein was dialyzed against 50 mL of buffer containing 20 mM Tris-HCl, pH 8.0, 200 mM KCl, 20% glycerol and 1 mM β-mercaptoethanol (BME, Biorad) overnight at 4 °C with the use of 50 µL dialysis buttons (Hampton Research) and 3, 500 Da M.W.C.O. regenerated cellulose dialysis tubing and stored at -80 °C.

2.4.2. Cloning and yeast transformation of hADAR2

The hADAR2 construct was generated by the polymerase chain reaction (PCR) with the following oligonucleotides: N-terminal primer: 5'-GCGCGCGGTACCCAAATG(CACCAT)$_5$GAAAATTTGTATTTTCAAGGT(X)$_{15}$-3' and C-terminal primer: 5'-GCGCGCGAATTCTCATCA(X)$_{15}$-3'. N-terminal primer contains the KpnI restriction site followed by 3 nt spacer (CAA) preceding the initiator methionine (ATG). The 10-histidine tag followed by a tobacco etch virus (TEV) protease recognition site proceed the 15 nucleotides which code for the N-terminal sequence of the hADAR2 gene (X_{15}). The C-terminal primer contains an EcoRI site and two stop codons followed by 15 nucleotides encoding for the C-terminal sequence of the hADAR2 gene (X_{15}). The PCR product was digested with KpnI and

EcoRI (Invitrogen) and ligated into the pYeDP60 yeast expression vector (generous gift from Dr. Howard Young's lab, University of Alberta) linearized with the same two restriction enzymes. pYeDP60 contains a galactose inducible promoter (GAL10-CYC1), polylinker region with the following restriction sites BamHI/SmaI/KpnI/SacI/EcoRI, URA3 and ADE2 (uracil and adenine auxotrophies) markers and ampicillin resistance marker. The BamHI restriction site immediately downstream of the promoter must be as close as possible to the cDNA initiation codon in the the construct. pYeDP60 was chosen for its stability in rich media at high cell densities. The cloned sequences of the hADAR2 gene were sequenced using primers complementary to 5' and 3' ends as well as from within the sequence. The expression vector was transformed by electroporation into several yeast strains. The haploid strain DSY904 with genotype: *bar/ade/his2ura3Δleu2trp1pep4::LEU2* which is deficient in several proteases and allows selection of URA3 auxotrophs was selected as yielding best expression pattern for hADAR2.

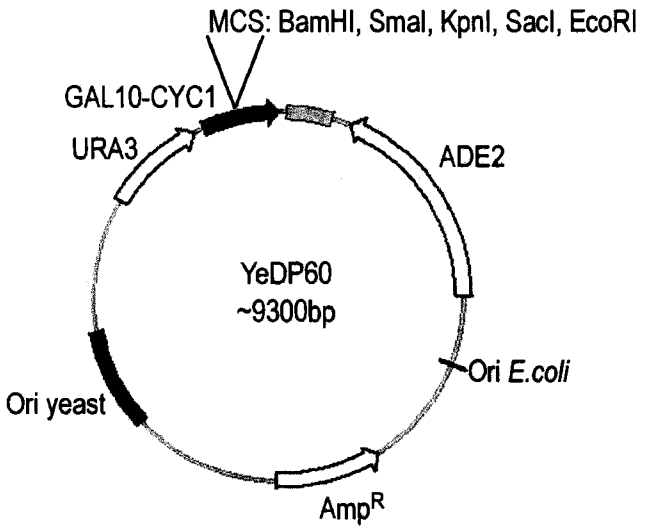

Figure 2-16. pYeDP60 vector.

2.4.3. hADAR2 expression from *S. cerevisiae*

After transformation, cells were plated on SDC URA$^-$ plates (1.7 g/L yeast nitrogen base (Fisher), 5 g/L ammonium sulfate, 2 g/L casamino acids (Fisher), 10.0 g/L succininc acid, 6 g/L ammonium hydroxide and 20 g/L agar supplemented with 1 M sorbitol and 2% (w/v) glucose) and incubated at 30 °C. Colonies appeared after 2-3 days. Expression from several different clones of the same hADAR2 construct was tested on a small scale prior to a large scale expression. We found that the expression levels of the hADAR2 varied from clone to clone. Small scale expression involved picking of 3-5 individual colonies to inoculate 10 mL of SDC URA$^-$ (1.7 g/L yeast nitrogen base (Fisher), 5 g/L ammonium sulfate, 2 g/L casamino acids (Fisher), 10.0 g/L succinic acid, 6 g/L ammonium hydroxide and supplemented with 2% glucose) for 24 h at 28 °C. A sample of this growth (2mL) was used to inoculate 50 mL of YPGE media (10 g/L of yeast extract, 10 g/L of peptone, 5 g/L glucose, 3% (v/v) EtOH). The cells were grown for 16 h at 28 °C upon which the culture was induced with 2% (w/v) galactose and cells were grown for 6 more hours. The cultures were pelleted at 4000 rpm for 30 min at 4 °C. The supernatant was removed and the pellet resuspended in Y-PER® yeast lysis reagent (125 μL/ 50 mg of wet pellet, Pierce) in the presence of Complete® protease inhibitor cocktail (Roche). The lysis was carried out on ice for 20 min and then the cells were pelleted at 4000 rpm for 30 min at 4 °C. Western blot analysis with Penta-His HRP® conjugate antibody (Qiagen) was carried out on all test colonies

and a no-galactose control. Colonies which produced best expression levels of hADAR2 were selected for large scale expression.

For the large scale expression of hADAR2, 10 μL of glycerol stocks were used to inoculate 10 mL of minimal media SDC URA$^-$. The cultures were grown for 24 h at 30 °C while shaking at 250 rpm. The next morning, 4 mL of the starter culture was used to inoculate 200 mL of minimal media and grown for 24 h in the same conditions as the starter culture. A sample (40 mL) of the secondary culture was used to inoculate 2 L of YPGE media (A_{600}=0.05) supplemented with adenine (5 mg/mL) and grown at 28 °C until cell density has reached 8 x 10^7 cells/mL (24-30 h, 1 x 10^6 cells/mL gives an A_{600} = 0.1). The 2 L culture was induced with 2% (w/v) galactose for 8-15 h and grown until cell density reached 2 x 10^8 cells/mL (typically 30-36 h). The cultures were pelleted at 7000 rpm for 15 min at 4 °C in a Sorval Evolution RC centrifuge equipped with a SLC 6000 4.8 L fixed angle rotor (Mandel). The cells were washed with buffer containing 20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 5% glycerol, 0.02% (w/v) sodium azide)and centrifuged again. The wet mass of the pellet was determined prior to storage at -80 °C. The typical yield from 1 L of culture grown in the YPGE rich media is 10-12 g of wet pellet.

2.4.4. hADAR2 purification

Cells from 8 L of culture (80-88 g of pellet) were resuspended in Buffer A (1 mL of buffer/ 0.5 g of pellet; 20 mM Tris-HCl, pH 8.0, 5% glycerol, 1 mM BME) containing 160 mM NaCl, 0.01% NP-40 and Complete® protease inhibitors. The cells were lysed by 5 passes through a Gaulin homogenizer at 20,000 psi[2] and the

extract was purified by centrifugation at 100,000x *g* for 1 h. The lysate was filtered through an Amicon steri-cup 0.22 μM filter device (Millipore). hADAR2 protein was precipitated by addition of 40% (w/v) $(NH_4)_2SO_4$ with stirring for 30 min followed by centrifugation at 17,000 rpm for 30 min. After removal of the supernatant the pellet was resuspended in 50 mL of Buffer A containing 80 mM NaCl and dialyzed overnight against the same buffer with the use of 3,500 Da nominal molecular weight cut off (M.W.C.O.) regenerated cellulose dialysis tubing (Fisher). Next morning the protein solution was loaded onto a 75 mL gravity packed HiTrap® Q-sepharose fast flow column (GE Healthcare) with Buffer A containing 80 mM NaCl and 0.5 mM ethylenediaminetetraacetic acid (EDTA). The loaded column was washed with 3 column volumes of Buffer A containing 80 mM NaCl and no EDTA. hADAR2 was eluted from the Q-column with a salt gradient from 80 mM to 1 M NaCl. The presence of hADAR2 was confirmed by western blotting with Penta-His antibody. Typically the protein eluted from the Q-column at 250-350 mM NaCl. Fractions containing hADAR2 were pooled and the imidazole was added to a final concentration of 35 mM. hADAR2 was incubated with 2.5 mL of Ni-NTA agarose (GE Healthcare) equilibrated with Buffer A containing 350 mM NaCl and 35 mM imidazole for 2 h with gentle rocking. The slurry was poured into 2.5 x 20 cm gravity drip column (Biorad) while collecting the flow through by gravity. The resin was washed in three consecutive 40 mL steps with Buffer A containing 35 mM imidazole and 750 mM, 300 mM and 200 mM NaCl respectively. The protein was eluted in eight 5 mL fractions with Buffer A containing 200 mM NaCl and 400 mM imidazole. The flow-through, washes, and elution fractions were

analyzed by SDS-PAGE and the proteins visualized by staining with Coomassie blue. The fractions containing hADAR2 were pooled together, concentrated to 2 mL with an Amicon-Ultra 15 centrifugal concentrator tube with a 5000 Da M.W.C.O. (Millipore) and diluted up to a final volume of 20 mL in buffer containing 10 mM Hepes, pH 8.0, 1 M NaCl, 500 mM $(NH_4)_2SO_4$ and 2 mM BME. The AKTA FPLC system (GE Healthcare) was used to load the eluted protein onto a 5 mL HiTrap Phenyl column (GE Healthcare). The protein was eluted with a salt gradient from 1 M to 100 mM NaCl. At this point the protein was 90-95% pure. Fractions containing protein were pooled together and concentrated to 2 mL with an Amicon-Ultra 15 centrifugal concentrator tube with a 5000 Da M.W.C.O. (Millipore). TEV protease (Source: *E. coli*; 20 U per 1mg of hADAR2) was added and the TEV cleavage reaction was incubated at 4 °C overnight. The efficiency of the TEV reaction was evaluated by western blotting with Penta-His antibody which has shown the reaction to be 95% complete after overnight incubation. The whole TEV-hADAR2 cleavage reaction was loaded onto a Superdex 200 26/60 gel filtration column (GE Healthcare) and eluted with Buffer A containing 200 mM NaCl. hADAR2 elutes as a symmetrical peak detected at 154 mL. The main peak fractions were pooled, concentrated and dialyzed with the use of 3,500 Da MWCO Slide-A-Lyzer® dialysis cassettes (Pierce) against storage buffer containing 20 mM Tris, pH 8.0, 200 mM NaCl, 20% glycerol and 1 mM BME. The protein was stored in aliquots at -80 °C. The protein was quantified by two methods: with the Bradford method using Protein Concentration Reagent (Biorad) and by UV absorbance at 280 nm. The typical yield of hADAR2 was 0.5-1.0 mg per 1 L of the original culture.

2.4.5. Protein activity assay

Templates for the 78 nucleotide R/G pre-mRNA were generated by PCR from a substrate cloned into pBS vector (Stratagene) with primers 5': 5'-TAATACGACTC ACTATAGGGGTCCTCATTAAGG-3' and 3': 5'-ACACATCAGGGTAG-3'. [α-^{32}P] body labeled 78 nucleotide R/G pre-mRNA was generated in a 25 µL transcription carried out in transcription buffer containing 40 mM Tris, pH 8.0, 6 mM $MgCl_2$, 100 mM NaCl, 10 mM DTT, 2 mM spermidine, 500 µM of each UTP, CTP and GTP, 15 µM of ATP, 1 U/mL of RNase Out (Invitrogen), 100 µCi α-^{32}ATP (Perkin Elmer), 0.1 µM DNA template and 2 U of T7 RNA polymerase (Invitrogen). Transcriptions were incubated at 37 °C for 4 hours. The oligonucleotides were purified by denaturing PAGE (8%, 19:1).

Editing reactions.

hADAR2 was dialyzed against editing buffer containing 20 mM Hepes, pH 8.0, 100 mM KCl, 0.5 mM DTT, 20% glycerol, and 0.01% NP-40. The body-labeled substrate was resuspended in ddH_2O, denatured at 80 °C for 2 min and allowed to renature at rt for 30 min just prior to its addition to the editing reaction. Typical editing reactions were performed in 20 µL total volume containing editing buffer, 0.5 nM RNA substrate and varying amounts of enzyme (1-100 µM) and were incubated at either rt or 30 °C for 4 h. The editing reactions were quenched by extraction with phenol/chloroform/isoamyl alcohol (25:24:1) and then with chloroform followed by ethanol precipitation.

Nuclease P1/TLC cleavage assay

The pellets resulting from the editing assays were resuspended in 9.5 μL of ddH_2O and digested with Nuclease P1 (0.5 U/reaction) overnight at 37 °C. The products of the digestion were resolved by thin layer chromatography (saturated $(NH_4)_2SO_4$:0.1 M NaOAc:isopropanol; 79:19:2 (v/v)) using cellulose-PEI chromatography plates pre-run in ddH_2O. The TLC plates were exposed to a Molecular Dynamics Phosphor screen which was scanned using a Molecular Dynamics Storm 860 PhosphorImager. Data was quantified using Image Quant 5.0 software.

Electrophoretic mobility shift assay (EMSA) with hADAR2

hADAR2 was dialyzed against EMSA buffer containing 10 mM Tris-HCl, pH 8.0, 30 mM NaCl, 1 mM $MgCl_2$, 1 mM DTT, 10% glycerol. The RNA was prepared as described above for the editing assays. The reactions were carried out in 10 μL total volume containing EMSA buffer, 0.5 nM RNA substrate with varying hADAR2 concentrations (1-150 nM). The reactions were incubated at 30 °C for 10 min before being resolved by native 6% 0.5X TBE (45 mM Tris-base, 45 mM boric acid, 0.1 mM EDTA) for 2.5 h at 220 V. Gels were dried and exposed to a Molecular Dynamics Storm 860 PhosphorImager. Data was quantified using Image Quant 5.0 software.

2.4.6. Sample preparation for SAXS

hADAR2 (25 μL) at a concentration of 2.7 mg/mL was dialyzed against 50 mL of buffer containing 20 mM Tris-HCl, pH 8.0, 200 mM NaCl, 5% glycerol

and 1 mM BME overnight at 4 °C with the use of 50 μL dialysis buttons (Hampton Research) and 3,500 Da M.W.C.O. regenerated cellulose dialysis tubing.

2.4.7. SAXS data collection.

SAXS X-ray scattering data were collected at SIBYLS beamline, Advanced Light Source, Lawrence Berkeley National Laboratory. Data were collected at 12 KeV on a MAR165 CCD detector with sample to detector distance of 1.5 m at 15 °C. Immediately prior to data measurement the sample was centrifuged at 10,000 rpm for 10 min and divided into two equal volumes. Data collection was carried out on the first half of the sample as it was. The second half of the sample was diluted in half before the measurement. For each sample a series of four exposures including 6 s, 6 s, 60 s and 6 s were measured to assess sample's sensitivity to radiation. The software program OGRE was used to normalize the data to the intensity of the transmitted beam and subtract the scattering of the buffer from the scattering of the hADAR2 containing solution. Data were analyzed using the programs PRIMUS and GNOM.[31, 32, 42]

2.5. Notes and References

1. Bass, B.L. RNA editing by adenosine deaminases that act on RNA. *Annu Rev Biochem* 2002, **71**:817-846.
2. Basilio C, Wahba AJ, Lengyel P, Speyer JF, Ochoa S. Synthetic polynucleotides and the amino acid code,V. *Proc Natl Acad Sci USA* 1962, **48**:613-616.
3. Lai, F., Chen, C.X., Carter, K.C., 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**:2413-24.
4. Melcher, T., Maas, S., Herb, A., Sprengel, R., Seeburg, P.H., Higuchi, M. A mammalian RNA editing enzyme. *Nature* 1996, **379**:460-64.
5. Gerber, A., O'Connell, M.A., Keller, W. Two forms of human double-stranded RNA-specific editase 1(hRED1) generated by the insertion of an *Alu* cassette. *RNA* 1997, **3**:453-463.
6. Chen, C-X., Cho, D.-S., Wang, Q., Lai, F., Carter, K.C., 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**:755-767.
7. Lomeli, H., Mosbacher, J., Melcher, T., Hoger, T., Geiger, J.R., Kuner, T., Monyer, H., Higuchi, M., Bach, A., Seeburg, P.H. Control of kinetic properties of AMPA receptor channels by nuclear RNA editing. *Science* 1994, **266**:1709–1713.
8. Vissel, B., Royle, G.A., Christie, B.R., Schiffer, H.H., Ghetti, A., Tritto, T., Perez-Otano, I., Radcliffe, R.A., Seamans, J., Sejnowski, T., et al. The role of RNA editing of kainate receptors in synaptic plasticity and seizures. *Neuron* 2001, **29**:217-227.
9. Seeburg, P.H., Single, F., Kuner, T., Higuchi, M., Sprengel, R. Genetic manipulation of key determinants of ion flow in glutamate receptor channels in the mouse. *Brain Res* 2001, **907**:233-243.
10. Yang, W., Chendrimada, T.P., Wang, Q., Higuchi, M., Seeburg, P.H., Shiekhattar, R., Nishikura, K. Modulation of microRNA processing and expression through RNA editing by ADAR deaminases. *Nat Struct Mol Biol* 2006, **13**:13-21.
11. Luciano, D.J., Mirsky, H., Vendetti, N.J., Maas, S. RNA editing of microRNA precursor. *RNA* 2004, **10**:1174-1177.

12. Kawahara, Y., Zinshteyn, B., Sethupathy, P.,Iizasa, H., Hatzigeorgiou, A.G., Nishikura, K. Redirection of silencing targets by adenosine-to-inosine editing of microRNAs. *Science* 2007, **315**:1137–1140.
13. Macbeth, M.R., Bass, B., L. Large scale overexpression and purification of ADARs from *Saccharomyces cerevisiae* for biophysical and biochemical studies. *Methods Enzymol* 2007, **424**:319-331.
14. Macbeth, M.R., Schubert, H.L., Vandemark, A.P., Lingam, A.T., Hill, C.P., Bass, B.L. Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. *Science* 2005, **309**:1534-1539.
15. Betts, L., Xiang, S.A., Short, R., Wolfenden, C.W., Carter, Jr. Cytidine Deaminase. The 2·3 Å Crystal Structure of an Enzyme: Transition-state Analog Complex. *J Mol Biol* 1994, **235**:635-656.
16. Gerber, A., Grosjean, H., Melcher, T., Keller, W. Tad1p, a yeast tRNA-specific adenosine deaminase, is related to the mammalian pre-mRNA editing enzymes ADAR1 and ADAR2. *EMBO J* 1998, **17**:4780-4789.
17. Gerber, A.P., Keller, W. An adenosine deaminase that generates inosine at the wobble position of tRNAs. *Science* 1999, **286**:1146-1149.
18. Chester, A., Scott, J., Anant, S., Navaratnam, N. RNA editing: cytidine to uridine conversion in apolipoprotein B mRNA. *Biochim. Biophys. Acta* 2000, **1494**:1-13.
19. Stefl, R., Xu, M., Skrisovska, L., Emeson, R.B., Allain, F.H. Structure and specific RNA binding of ADAR2 double-stranded RNA binding motifs. *Structure* 2006, **14**:345-55.
20. Ryter, J.M., Schultz, S.C. Molecular basis of double-stranded RNA-protein interactions: strucutre of a dsRNA-binding domain complexed with dsRNA. *EMBO J* 1998, **17**:7505-7513.
21. Macbeth, M.R., Lingam, A.T. Bass, B.L. Evidence for auto-inhibition by the N-terminus of hADAR2 and activation by dsRNA binding. *RNA* 2004, **10**:1563-1571.
22. Maydanovych, O., Beal, P.A. Braking the central dogma by RNA editing. *Chem Rev* 2006, **106**:3397-3411.
23. Svergun, D.I., Koch, M.J. Small-angle scattering studies of biological macromolecules in solution. *Rep Prog Phys* 2003, **66**:1735-1782.
24. Putnam, C.D., Hammel, M., Hura, G.L., Tainer, J.A. Solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q Rev Biophys* 2007, **40**:191-285.

25. Tsutakawa, S.E., Hura, G.L., Frankel, K.A., Cooper, P.K., Tainer, J.A. Structural analysis of flexible protein in solution by small angle X-ray scattering combined with crystallography. *J Struct Biol* 2007, **158**:214-223.
26. Guinier, A., Fournet, F. Small angle scattering of X-rays. 1955. Wiley-Interscience, New York.
27. Glatter, O. A new method for the evaluation of small angle X-ray scattering dat. *J Appl Cryst*, 1977, **10**:415-421.
28. Semenyuk, A.V., Svergun, D.I. GNOM – a program package for small-angle scattering data processing. *J App Cryst* 1991, **24**:537-540.
29. Svergun, D.I. Restoring low resolution structure of biological macromolecules from solution scattering using simulated annealing. *Biophys J* 1999, **76**:2879-2886.
30. Svergun, D.I., Petoukhov, M.V., Koch, M.H. Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* 2001, **76**:2879-2886.
31. Volkov, V.V., Svergun, D.I. Uniqueness of *ab initio* shape determination in small angle scattering. *J Appl Cryst* 2003, **36**:860-864.
32. Kozin, M.B., Svergun, D.I. Automated matching of high- and low- resolution structural models. *J Appl Cryst* 2001, **34**:33-41.
33. Wriggers, W., Chacón, P. Using Situs for the Registration of protein structures with low-resolution bead models from X-ray solution scattering. *J Appl Cryst* 2001, **34**:773-776.
34. Petoukhov, M.V., Svergun, D.I. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys J* 2005, **89**:1237-1250.
35. Petoukhov, M.V., Eady, N.A., Brown, K.A., Svergun, D.I. Addition of missing loops and domains to protein models by X-ray solution scattering. *Biophys J* 2002, **83**:3113-3125.
36. Davies, J.M., Tsuruta, H., May, A.P., Weis, W.I. conformational changes of p97 during nucleotide hydrolysis determined by small-angle X-ray scattering. *Structure* 2005, **13**:183-19.
37. Jaikaran, D.C., Collins, C.H., MacMillan, A.M. Adenosine to inosine editing by ADAR2 requires formation of a ternary complex on the GluR-B R/G site. *J Biol Chem* 2002, **277**:37624-37629.
38. Cho, D.-S.C., Yang, W., Lee, J.T., Shiekhattar, R., Murray, J.M., Nishikura, K. Requirement of dimerization for RNA editing activity of adenosine deaminases acting on RNA. *J Biol Chem* 2003, **278**:17093-17102.

39. Chilibeck, K.A., Wu, T., Liang, C., Shellenberg, M.J., Gesner, E.M., Lynch, J.M., MacMillan,A.M. FRET analysis of in vivo dimerization by RNA-editing enzymes. *J Biol Chem* 2006, **281**:16530-16535.

40. Poulsen, H., Jorgensen, R., Heding, A., Nielsen, F.C., Bonven, B., Egebjerg, J. Dimerization of ADAR2 is mediated by double-stranded RNA binding domain. *RNA* 2006, **12**:1350-1360.

41. SAXS data analysis and model building studies were performed with the use of methodology courtesy of Dr. Ross Edwards.

42. Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H., Svergun, D.I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J Appl Cryst* 2003, **36**:1277-1282

43. Berman, H.M., Westbrook, J., Feng, Z., Gilliland, G., Bhat, T.N., Weissig, H., Shindyalov, I.N., Bourne, P.E. The Protein Data Bank. *Nuc Acids Res* 2000, **28**:235-242.

44. Svergun, D.I., Baraberato, C., Koch, M.H. CRYSOL – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Cryst* 1995, **28**:768-773

45. Burkard, M.E., Turner, D.H., Tinoco, I. "The interactions that shape RNA structure." in *The RNA World*, ed. Gesteland, R.F., Cech, T.R. and Atkins, J.F. New York: Cold Spring Harbor Laboratory Press, 1999.

46. Petoukhov, M.V., Svergun, D.I. Global rigid body modeling of macromolecular complexes against small-angle scattering data. *Biophys J* 2005, **89**:1237-1250.

Chapter 3: STRUCTURAL INVESTIGATION OF THE DEAMINASE DOMAINS OF hADAR PROTEINS

3.1. Introduction

Deamination of adenosines within coding regions of mRNA catalyzed by the ADAR family of enzymes converts adenosines to inosines at specific sites. These editing events can change the coding properties of the message and thus introduce amino acids into the gene product which are not encoded by the gene. Indeed, there are several known ADAR substrates for which site specific editing events result in functionally important isoforms of proteins.[1] What is truly remarkable, but at the same time not well understood is how ADAR enzymes achieve the accurate and specific editing which is necessary for the production of functional proteins.

3.1.1. The specificity of editing GluR-B pre-mRNA by ADAR1 and ADAR2

The best studied examples of A to I editing in mammals are for the pre-mRNA of receptor proteins involved in neurotransmission such as the glutamate receptors. The glutamate receptor subunit B (GluR-B) pre-mRNA is edited at several positions including two sites for which the specific physiological consequences of the editing event are well understood.[2] The long (~100 base pair) double-stranded region formed between exon 11 and its adjacent intron contains over fifty adenosines (Figure 3-1 A). However, only one adenosine located within the exonic sequence is edited with 99% frequency *in vivo*, exclusively by ADAR2.[2-4] The editing at this site (Q/R) changes a genome-encoded CAG (glutamine, Q) to CIG (arginine, R) and this single amino acid change renders the glutamate ion channel impermeable to calcium ions. Editing at the Q/R site is essential in mammals as supported by studies on

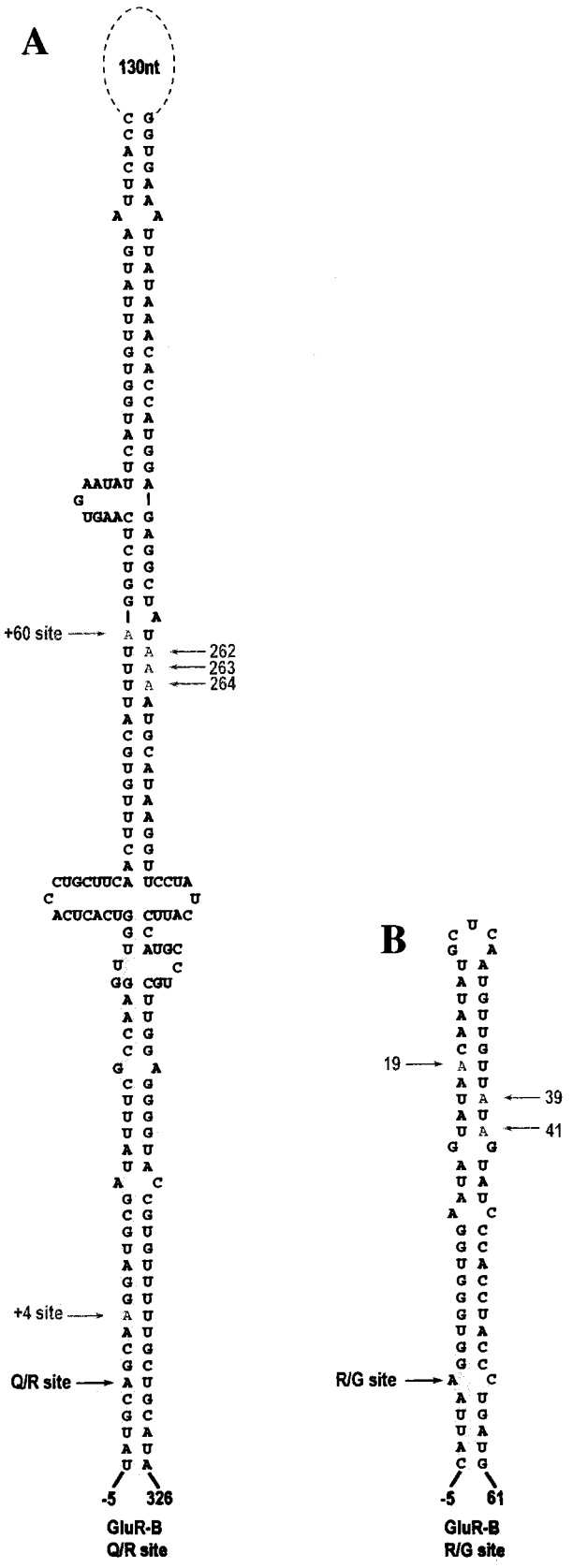

Figure 3-1. Q/R and R/G editing sites within GluR-B RNA. Exonic sequences are highlighted in green. The Q/R and R/G sites are indicated in red. Other adenosines that are also susceptible to editing by ADARs are indicated in pink. **(A)** The predicted secondary structure for GluR-B pre-mRNA containing the Q/R site which is formed by exon 11 and its adjacent intron. **(B)** The predicted stem loop structure for GluR-B pre-mRNA containing the R/G site which is formed by exon 13 with its adjacent intron. (Adapted from ref. 3)

homozygous ADAR2$^{-/-}$ mice. In the absence of editing at the Q/R site, the animals develop epilepsy and die shortly after birth.[5] Sixty nucleotides downstream of the Q/R site in the intronic sequence another adenosine often referred to as the "hot spot" is edited with up to 80% *in vivo* by ADAR1 (Figure 3-1 A). The biological consequences of the editing event at the "hot spot" are unknown. Moreover, an adenosine located within the exon 13 of GluR-B pre-mRNA is edited with up to 70% frequency *in vivo* equally well by both ADAR1 and ADAR2.[6] At this site (R/G) a genome-encoded AGA codon for arginine (R) is changed to GIA coding for glycine (G). This change from a positively charged to neutral amino acid affects the structural features of the protein and allows the glutamate receptors to recover faster from desensitization. Furthermore, the extent of editing at the R/G site appears to change with the different developmental stages. In the rat brain, the levels of editing at the R/G site are low at the early embryonic stages but increase rapidly until postnatal day 42.[7] The above examples for the GluR-B pre-mRNA illustrate the ability of ADAR1 and ADAR2 to accurately distinguish among hundreds of adenosines within their substrates. Remarkably, both proteins exhibit distinct and yet overlapping specificities. It is not clear however, which structural features of the ADAR proteins are responsible for achieving this specificity.

3.1.2. Selectivity and specificity of the editing reaction

Several lines of evidence suggest that the selectivity of the ADAR proteins for the dsRNA substrates is mediated by the dsRBMs. The N-terminus of hADAR2 containing the dsRBMs recognizes and binds dsRNA with minimum length of 15 bp.[8]

The affinity of the N-terminus of hADAR2 for the 78 nt GluR-B R/G substrate has been investigated by several different laboratories including ours. Beal and coworkers reported the affinity of the N-terminus of hADAR2 for dsRNA to be similar (K_d=12±2 nM) to the full length protein (K_d=22±4 nM).[9] Members of our lab have found that the affinity of the N-terminal hADAR2 (79-316) for the 78 nt GluR-B R/G is lower (K_d=50±4 nM) than for the full length hADAR2 (K_d=21±8 nM, Unpublished results: Dul, E., Schellenberg, M. and MacMillan, A.). The discrepancies in the reported values for K_d are likely due to the different conditions used for the electrophoretic mobility shift assays (EMSAs). On the contrary, the C-terminus of hADAR2 (299-701) does not bind the R/G substrate tightly under similar conditions as determined by the EMSA experiments (personal communication with Brenda Bass). Moreover, footprinting experiments and chemical perturbation studies with the dsRBMs indicate that these domains not only recognize the double-stranded character of RNA but also have preferences for binding sites on their substrates.[8, 10] In a perfectly dsRNA, ADARs are non-selective and deaminate up to 50% of adenosines, whereas the presence of secondary structures such as loops, mismatches and bulges confers some selectivity upon the ADAR enzyme. The introduction of secondary structures into the duplex substrate most likely effectively shortens the double-stranded substrate and imposes restrictions on the protein-RNA interaction. However, analysis of the editing sites within GluR-B pre-mRNA substrate suggests that dsRBMs alone cannot be responsible for the specificity of the editing reaction.

Although the interaction between the C-terminal deaminase domain of ADARs and their substrates is poorly understood, some evidence suggests that it may play a role in selecting the adenosine destined for modification. Even though the N-terminus of ADARs has been shown to bind any dsRNA of sufficient length with high affinity (nM range), not all of these potential substrates are edited. Lazinski and coworkers have shown that a protein chimera containing the N-terminal dsRBMs of ADAR1 but the C-terminal deaminase domain of ADAR2 retains specificity for substrates preferentially edited by ADAR2.[11] Considerable evidence also comes from the studies in which 2-aminopurine (2-AP) replaces the edited adenosine in the model R/G substrate (Figure 3-2).[9] Upon interaction of the substrate with ADAR2, a detectable increase in fluorescence is observed. Since stacking of 2-AP within the duplex RNA quenches its fluorescence, the changes observed are indicative of the base being flipped out of the helix. When 2-AP is substituted at positions other than the edited adenosine no changes in fluorescence can be detected. Furthermore, binding of the truncated ADAR2 N-terminus alone cannot induce the adenosine extraction from the dsRNA. This evidence suggests that the base-flipping during the deamination reaction is not dependent on the binding of dsRBMs but rather on the interaction with the deaminase domain. The studies with 2-AP also support mechanism for ADARs in which the enzyme samples the sequence until the target base is encountered and inserted into the active site rather than sampling every base as it moves along the sequence.

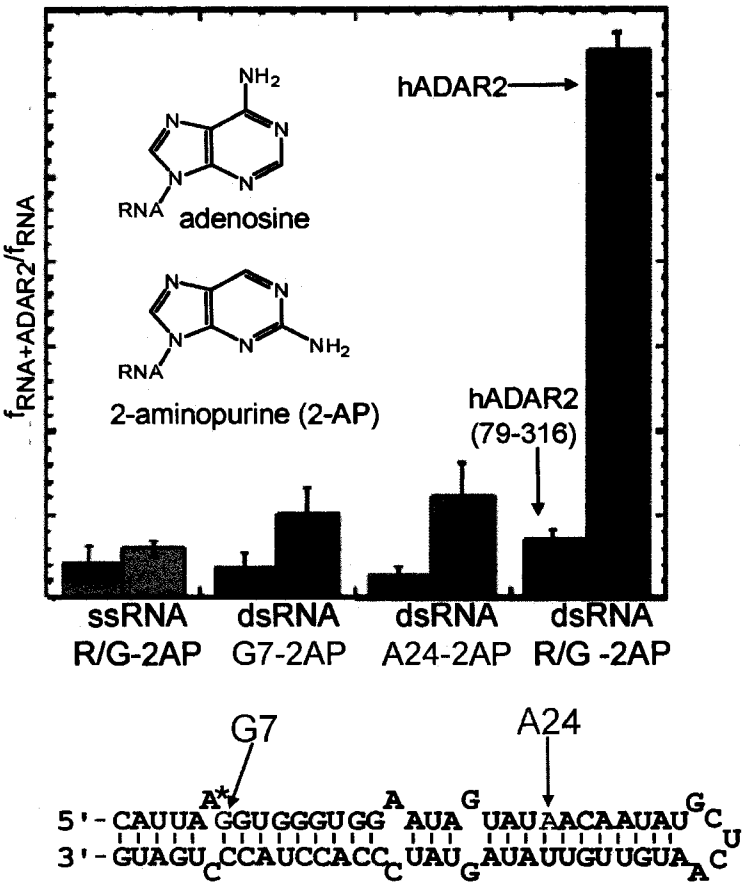

Figure 3-2. Studies of the hADAR2-RNA interaction. Binding of the substrate by hADAR2 induces an increase in fluorescence of 2-AP which replaces the edited adenosine. The increase in fluorescence is not observed for single-stranded RNA (ssRNA) substrate or substrates with 2-AP at positions different from the edited adenosine. Also the N-terminal hADAR2 (79-316) alone cannot induce the increase in fluorescence seen for the full length protein. (Adapted from ref. 9)

Several *in vitro* experiments indicate that ADARs do have preferences for certain sequences over others. ADAR1 selectively deaminates adenosines positioned 3' to U=A>C>G. ADAR2 also prefers U=A>C=G as its 5' neighbours and U=G>C=A for its 3' neighbours.[12] Thus ADAR2 has a number of preferred triplets including UAU, AAG, UAG and AAU. Moreover, the identity of the base-pairing partner to the edited adenosine affects the preferences of ADARs.[12-14] For example, it has been shown that an A-C mismatch at the edited site is preferred over A-A or A-G

mismatches or A-U base pairs for efficient editing in most cases.[12] The basis for these nearest neighbour and base-pairing partner preferences are not clearly understood. One possibility is that the structural features around the active site in the deaminase domain of ADARs favor specific nucleotides. It is also possible that base-flipping is promoted by certain combinations of the nucleotides either base-pairing to or flanking the edited adenosine.[13] These preferences are not always followed by ADARs as it is in the case of the Q/R site where the edited adenosine is in the CAG codon and it base pairs with U. Therefore, other unknown factors may contribute to the specificity of the deaminase domain and their elucidation requires further structural and biochemical data.

3.1.3. High resolution X-ray structure of the deaminase domain from hADAR2

The high resolution structure of the C-terminus (aa 299-701) of hADAR2 expressed and purified from *S. cerevisiae* has been solved by Macbeth and coworkers.[15] The X-ray structure reveals a well-folded globular domain with average radius of 40 Å (Figure 3-3 A). As previously predicted from sequence alignments with *E. coli* cytidine deaminases (CDAs) and adenosine deaminases that act on tRNA (ADATs), the structure contains a well conserved "deamination motif" consisting of 2 α-helices and 4 β-sheets with a molecule of water and a Zn^{2+} ion at the active site. Three conserved residues including two cysteines, histidine and the water act as ligands for the Zn^{2+} ion, and the glutamic acid which is hydrogen-bonded to the water molecule appears to be positioned for proton transfer during the deamination reaction

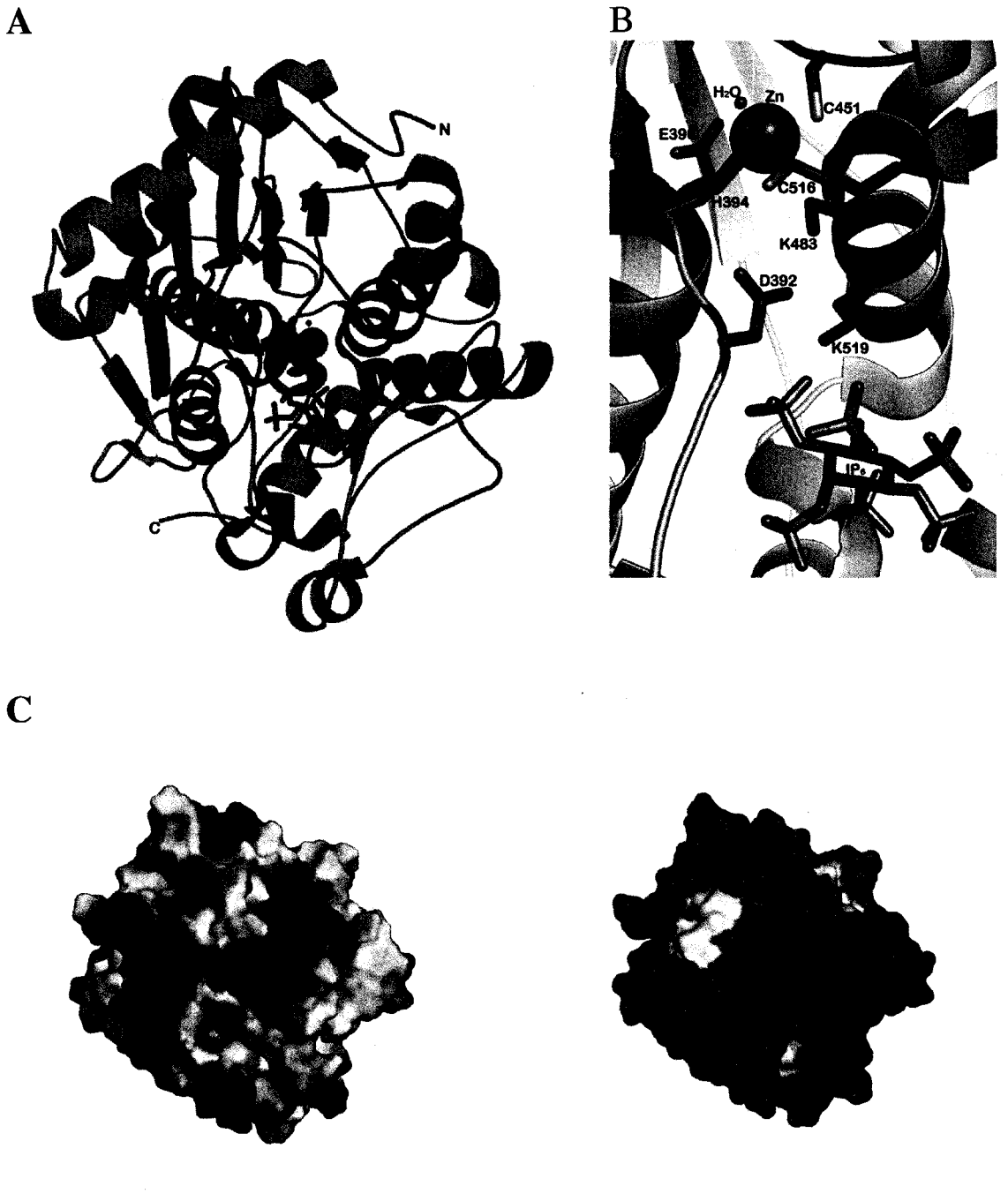

Figure 3-3. Structure of the deaminase domain of human ADAR2. **(A)** The deaminase domain of ADAR2 contains a highly conserved "deamination motif" (blue) comprising two helices and four strands. A Zn^{2+} ion located at the active site is shown in purple. Also found within the protein fold is a molecule of IP_6. **(B)** At the active site the Zn^{2+} ion is coordinated by H394, C516 and C451 residues and a water molecule shown in red. The IP_6 forms a hydrogen bond network with the active site Zn through residues K519 and D392. **(C)** Top view of the RNA binding surface on the deaminase domain. Structure on the left shows electrostatic surface potential for the area surrounding the catalytic pocket. Structure on the right is colored according to the secondary structure features: α-helices are colored in red, β-sheets are yellow and random coils are green. (Adapted from ref. 15)

(Figure 3-3 B). The catalytic center marked by the Zn^{2+} is located deep in a pocket surrounded by a large surface area of positive electrostatic potential. The positively charged surface is made up of a number of flexible loops and provides a putative binding site for the RNA substrate (Figure 3-3 C). The location of the active site within the deep pocket in the structure also supports the previously predicted necessity for the base to be flipped out of the helix during the catalysis. A possible explanation of why ADARs do not deaminate cytidine is found by close examination of one of the flexible loops flanking the enzyme's active site. This loop contains a threonine (T375) side chain positioned as to interact with the 2'OH of the edited ribonucleotide in the active site.[15,16] Models of either AMP or a known inhibitor of CDAs, zebularine, generated by applying the geometry of the active site show that the side chain of T375 appears to be in position to accommodate the 2'OH of adenosine but sterically clashes with the ribose ring of cytidine's analog zebularine.

The high resolution structure of the hADAR2 deaminase domain also reveals an unexpected feature in the form of a molecule of inositol hexakisphosphate (IP_6) buried deep within the core of the protein with an extensive network of hydrogen bonds to the polar residues. IP_6 is an important signaling molecule which must have been sequestered from the internal pools in *S. cerevisiae*.[17-19] The location of IP_6 within a pocket lined with mostly basic residues indicates that this molecule is likely essential for proper protein folding and protein stability. Moreover, the signaling properties of IP_6 suggest that it may be involved in the regulation of ADAR activity. When hADAR2 was expressed in mutant yeast strain with defects in the IP_6 synthetic pathway, the hADAR2 expression was significantly reduced and the protein was

inactive. Over 10 amino acid residues are directly involved in the extensive hydrogen bonding network with the IP_6. which extends to the active site Zn^{2+} ion hrough the K519, D392 and K483 residues. The authors suggest that the presence of IP_6 may not only play a role in the protein folding but may also provide means of regulating the editing activity of ADARs by modulating the pK_a of amino acids at the active site although furhter studies are necessary in this area.

The high resolution structure of the hADAR2 deaminase domain provides structural information necessary for further investigation of factors contributing to the specificity of A to I editing reactions. This chapter describes the expression and purification of deaminase domains of hADAR1-3 (hADAR1-D, hADAR2-D and hADAR3-D) for the purpose of structural and biochemical studies in hope of identifying regions responsible for the specificity of the editing reaction. Using the same expression system as for the full length hADAR2 protein we were able to produce pure and active protein samples. The levels of expression for all three protein constructs were comparable, however during purification hADAR1 deaminase domain was found to be mostly aggregated which resulted in much lower yields of the protein. All three deaminase domains displayed deamination activity although at levels lower than for the full length protein. Our attempts to crystallize hADAR3-D did not produce crystals suitable for X-ray crystallography. However, the protein samples for hADAR2-D and hADAR3-D did produce good quality small angle X-ray scattering data. The analysis of the data provided parameters for *ab initio* reconstruction and rigid body modeling of hADAR2-D and hADAR3-D constructs.

3.2. Results and Discussion

3.2.1. Sequence alignment of deaminase domains from hADAR1, hADAR2 and hADAR3.

The high degree of sequence similarity among the three members of the hADAR family of proteins allows for drawing comparisons between hADAR2-D, for which the high resolution structure has been solved, and the deaminase domains of hADAR1 and hADAR3. Biochemical work with both the full length hADAR1 and hADAR3 has been challenging as their expression is significantly lower than the expression of hADAR2 and these proteins appear to be less stable during protein preparations. Sequence alignments among all three deaminase domains from hADAR proteins show a 35% sequence identity between hADAR1-D and both hADAR2-D and hADAR3-D (Figure 3-4). Interestingly, the sequence identity between hADAR2 and hADAR3 is 54% but the expression and purification of hADAR3 is problematic and no known substrates for this family member have been identified. The conserved residues among the three deaminase domains are those contained within the "deamination motif" as well as those which coordinate the IP_6 molecule in the structure of hADAR2 deaminase domain (Figure 3-4). The greatest sequence variability is found in the N- and C-terminal ends of the domains and in the flexible loop structures, which may be involved in protein-RNA interactions. The X-ray structure for hADAR2-D does not account for the amino acid residues 462 to 473 for which electron density could not be observed (Figure 3-3 C). The presence

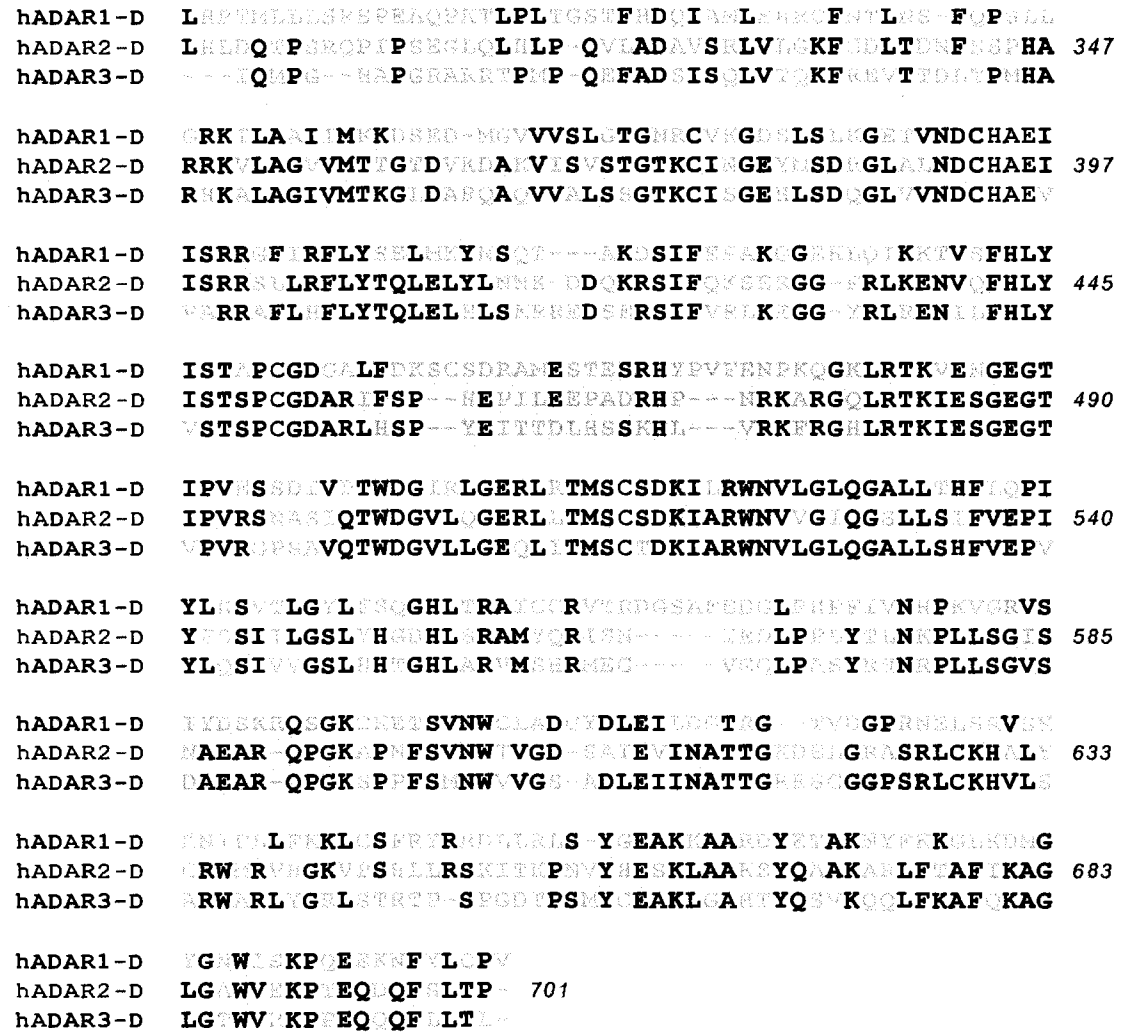

Figure 3-4. Sequence alignment for deaminase domains of hADAR1 (520-931), hADAR2 (299-701) and hADAR3 (342-739). The residues at the catalytic center are indicated in red text. The residues which directly coordinate IP_6 are highlighted in yellow. The flexible loop structures flanking the deep pocket in the surface of the protein containing the active site are highlighted in grey.

of highly flexible loops and their low conservation among the deaminase domains of hADARs suggests the possibility of their involvement in the specificity aspect of the editing reactions. It is tempting to speculate that certain conformational rearrangements within these loop regions may render the active site more or less accessible based on their interaction with the different substrates. Evidence of conformational changes within ADAR proteins comes from tryptophan fluorescence studies.[9] hADAR2 and hADAR3 contain five conserved tryptophan residues, while hADAR1 has four.In all three proteins, the tryptophans are located within the deaminase domain region of the sequence (Figure 3-5). The analysis of tryptophan fluorescence in the presence and absence of the RNA substrate indicates that the protein undergoes conformational changes upon binding of its substrate, which results in at least one of the tryptophans becoming more solvent exposed.[9] The high resolution structure reveals that three out of the five tryptophan residues are surface exposed and two of these residues are located directly below the putative RNA binding surface (Figure 3-5). Alternatively, the loop structures on the surface of the deaminase domain may interact with the regions in the N-terminus (dsRBMs) in the absence of dsRNA substrates.[20] In order to investigate the structural differences among the three deaminase domains, the proteins were expressed in *S. cerevisiae* and their activity was measured.

3.2.2. Protein expression and purification

The N-terminal 10x HIS tagged hADAR1-D (520-931), hADAR2-D (299-701) and hADAR3-D (342-739) constructs were cloned into the pYeDP60 vector and

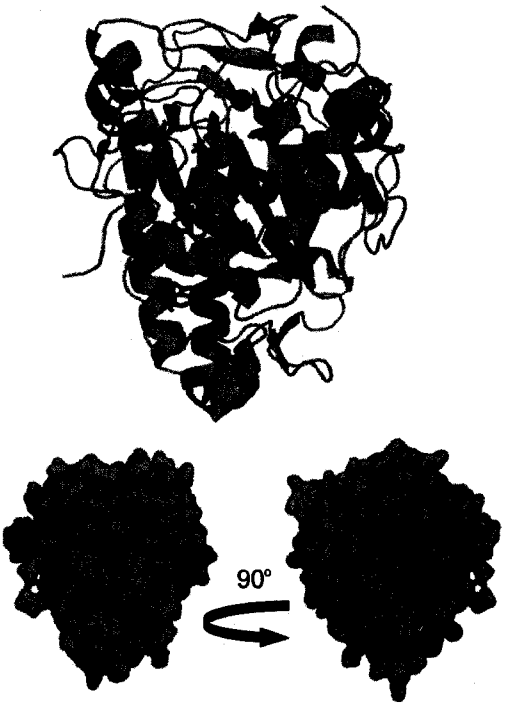

Figure 3-5. Location of tryptophans within the high resolution structure of hADAR2 deaminase domain. The deaminase domain contains five tryptophan residues including two which are buried and three which are surface exposed.

expressed according to the high density protocol previously used for the full length hADAR2 (Table 3-1). The detailed purification protocol is described in the Materials and Methods part of this chapter (Section 3.4). The estimated molecular weights were determined to be 46.5, 44.9 and 44.2 kDa for hADAR1-D, hADAR2-D and hADAR3-D respectively. During a size exclusion chromatography analysis on a Superdex 200 26/60 column, the deaminase domain constructs eluted in a volume of 225 mL (Figure 3-6). Relative to the elution of standard molecular weight markers, the deaminase domain constructs appear to elute as monomers (data not shown). Although the expression levels of the three constructs were similar as indicated by the

Table 3-1. hADAR deaminase domain constructs.

	Sequence	MW 10X HIS protein (kDa)	MW protein (kDa)	pI
hADAR1-D	520-931	48.7	46.5	9.46
hADAR2-D	299-701	47.2	44.9	9.38
hADAR3-D	342-739	46.4	44.2	9.84

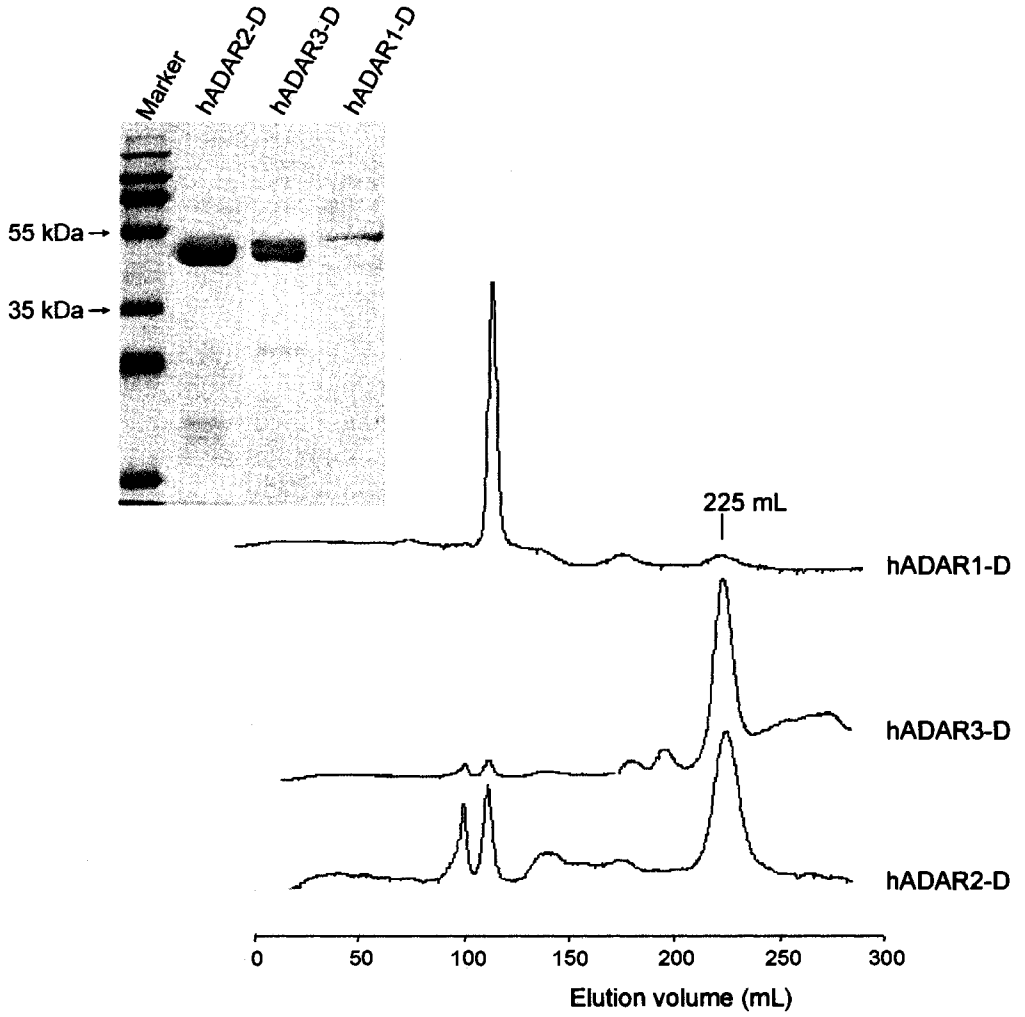

Figure 3-6. Purification of deaminase domains by size exclusion chromatography. The UV trace of hADAR1-3 deaminase domains as eluted from a Superdex 200 26/60 size exclusion column. The elution volume for all three proteins was 225 mL which suggests that the proteins exist as monomers under these purification conditions (Size exclusion buffer contained 20 mM Tris, pH 8.0, 200 mM NaCl, 1 mM BME, 5% (v/v) glycerol). Molecular weight standards used in the calibration (not shown) were as follows: (1) catalyse 232 kDa, 122.1 mL; (2) aldolase 158 kDa, 173.1 mL; (3) bovine serum albumin 67 kDa, 200.6 mL; (4) chymotrypsinogen 25 kDa, 241.0 mL. Inset: 12% SDS PAGE analysis of the pure hADAR1-3 constructs.

western blot analysis of cell lysates, the final yield of pure hADAR1-D was significantly lower than that for hADAR2-D or hADAR3-D. The typical yields of protein for hADAR2-D and hADAR3-D varied between 0.5-1.5 mg of protein per liter of original yeast culture. The preparations of hADAR1-D resulted in significantly lower yields ranging, between 0.05-0.1 mg of protein per liter of yeast culture. The size exclusion chromatography analysis indicated that most of the hADAR1-D was aggregated and only a small portion eluted as a monomer. In an effort to improve hADAR1-D yields of pure protein, various induction times and lysis conditions were tested. However, a significant improvement in protein yields for any of the conditions tested could not be observed.

The biological activity of the deaminase domain constructs was tested by Nuclease P1/thin layer chromatography (TLC). The assay utilized the 78 nt GluR-B R/G pre-mRNA substrate uniformly labeled with α-^{32}P ATP. A constant amount of the R/G RNA substrate was incubated with varying amounts of each of the constructs (1-100 μM), followed by digestion with Nuclease P1 and analysis of the resulting nucleotide 5'-monophosphates by TLC (Figure 3-7). For all three deaminase domains, the presence of inosine could be detected in the edited versus unedited substrate although the efficiency of editing was lower than for the full length hADAR2. For hADAR2-D this result is consistent with previous reports.[15] Since hADAR1 and hADAR2 edit the GluR-B R/G site with equal efficiency *in vitro,* it was not surprising that hADAR1-D also edited the 78 nt R/G substrate. However, low levels of radiolabeled inosine could also be detected in the editing assay with hADAR3-D. Nishikura and coworkers have tested the activity of the full-length

hADAR3 on various synthetic and naturally occurring dsRNA substrates including the GluR-B R/G pre-mRNA but were not able to observe A to I editing.[21] Detectable levels of inosine in the substrate edited by hADAR3-D suggests that in the absence of its N-terminal region this protein has a catalytic activity similar to hADAR1-D and hADAR2-D.

The purified hADAR3-D construct was used in crystallization trials, however none of the conditions screened produced crystals. We turned to small angle X-ray scattering (SAXS) experiments to generate low resolution structures which could be compared to the hADAR2-D high resolution structure.

3.2.3. Low resolution reconstruction of deaminase domains from SAXS[22]

Data collection and analysis

The low-angle scattering data was collected on hADAR2-D (299-701) at 1.5 mg/mL and hADAR3-D (342-739) at 2.6 mg/mL in SAXS buffer (20 mM Tris-HCl, pH 8.0, 400 mM NaCl, 5% glycerol and 1mM BME; Figure 3-9). Multiple exposures at varying lengths of time on the same protein solution indicated no changes in the scattering and thus no measurable radiation damage to the sample. Furthermore, scattering curves collected at half of the original concentration for both constructs superimposed well in the low angle region indicating that the samples were monodisperse. A set of SAXS data was also collected on hADAR1-D monomer isolated during the size exclusion chromatography at 0.5 mg/mL concentration in the

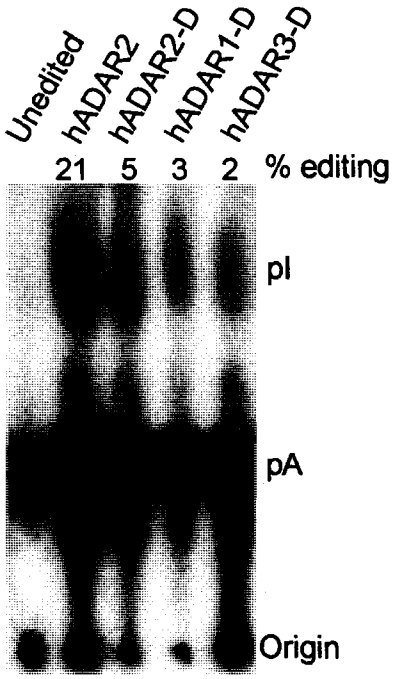

Figure 3-7. Activity assays. Editing Nuclease P1/TLC assay. Presence of α-^{32}P labeled inosine is visible in the edited samples. The efficiency of editing is lower for the deaminase domain constructs as compared to the full length hADAR2 protein.

SAXS buffer. The scattering curves for both short exposures (6 sec) and long exposures (60 sec) had higher variation in data points in the high angle region. In addition, the long exposure on hADAR1-D protein indicated that the sample was aggregating (Figure 3-8). Since aggregation can adversely affect the entire data range, no further processing was carried out on the hADAR1-D data and the following analysis refers only to hADAR2-D and hADAR3-D proteins.

For both hADAR2 and hADAR3 constructs, the Guinier plot displayed a good linear fit in the range $s.R_G < 1.3$, indicating that the sample was not aggregated (Figure 3-8).[23, 24] The radius of gyration (R_G), determined from the Guinier plot had similar values for both proteins. For hADAR2-D R_G calculated for data points 13 to

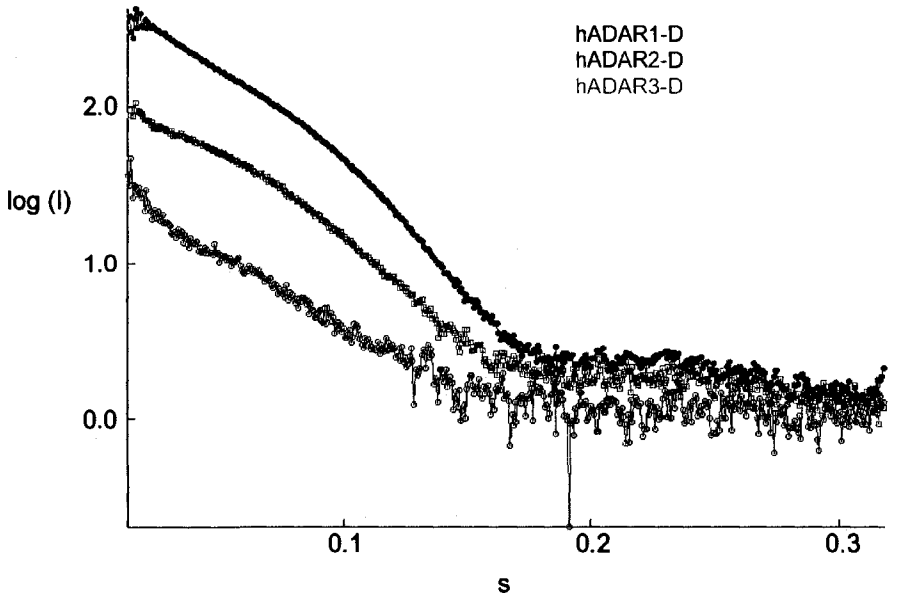

Figure 3-8. Scattering curves for the deaminase domains from human ADAR proteins. The scattering curves (60 sec exposure) for hADAR2-D and hADAR3-D have discrete values in the low angle regions. The scattering curve for hADAR1-D is featureless and approaches infinity at low values of *s* indicative of protein aggregation.

43 was 24.4 Å whereas for hADAR3-D R_G calculated for data points 44 to 72 was 23.3 Å. The program GNOM was used to calculate the pair-density distribution function, *P(r)*.[25] For both constructs the function has a symmetric bell shaped curve with a maximum at 27 Å and D_{max} of 70 Å, characteristic of a globular protein (Figure 3-9). Also, in both cases, the *P(r)* curve had a slight shoulder spanning the 50 to 70 Å region which may be indicative of the presence of a distinct and flexible domain. D_{max} was evaluated multiple times by testing a range of values ($50 < D_{max} < 300$) and observing the changes in the total estimate of goodness of fit (Q-score) provided by the program GNOM. For hADAR2-D, D_{max} of 70 Å gave the highest Q-score of 0.633, whereas for hADAR3-D, the same value of D_{max} gave a Q-score of 0.968. The Kratky plot of the scattering data for both proteins is a

parabola providing further evidence of a globular and folded protein. The high resolution structure of hADAR2-D indicates that the average diameter of the deaminase domain is 40 Å. Thus the value of 70 Å determined for the maximum dimension of the deaminase domain from the SAXS studies was higher than expected. The program CRYSOL was used to compare the experimental scattering data collected on hADAR2-D and hADAR3-D protein samples to the theoretical scattering curve of the hADAR2-D high resolution structure.[26] The theoretical and experimental scattering curves did not overlap well indicating that the X-ray structure is different from the protein structure in solution.

Ab initio structure reconstruction

The *ab initio* reconstruction program GASBOR was used to generate models of hADAR2-D and hADAR3-D by fitting the experimental scattering data in reciprocal space (Figure 3-10 and Figure 3-11; GASBOR methodology can be found in Section 2.2.3). [27] For each of the deaminase constructs, the overall model was generated by performing 15 individual GASBOR runs with each run starting the molecular dynamics calculation at a different random arrangement of beads in the search volume. The model building runs were performed with P1 symmetry constraint as supported by the X-ray crystallography structure and size exclusion chromatography results, which are indicative of the monomeric state for the deaminase constructs. Subsequently, all models within each subset were aligned, averaged and filtered using the software package DAMAVER The program

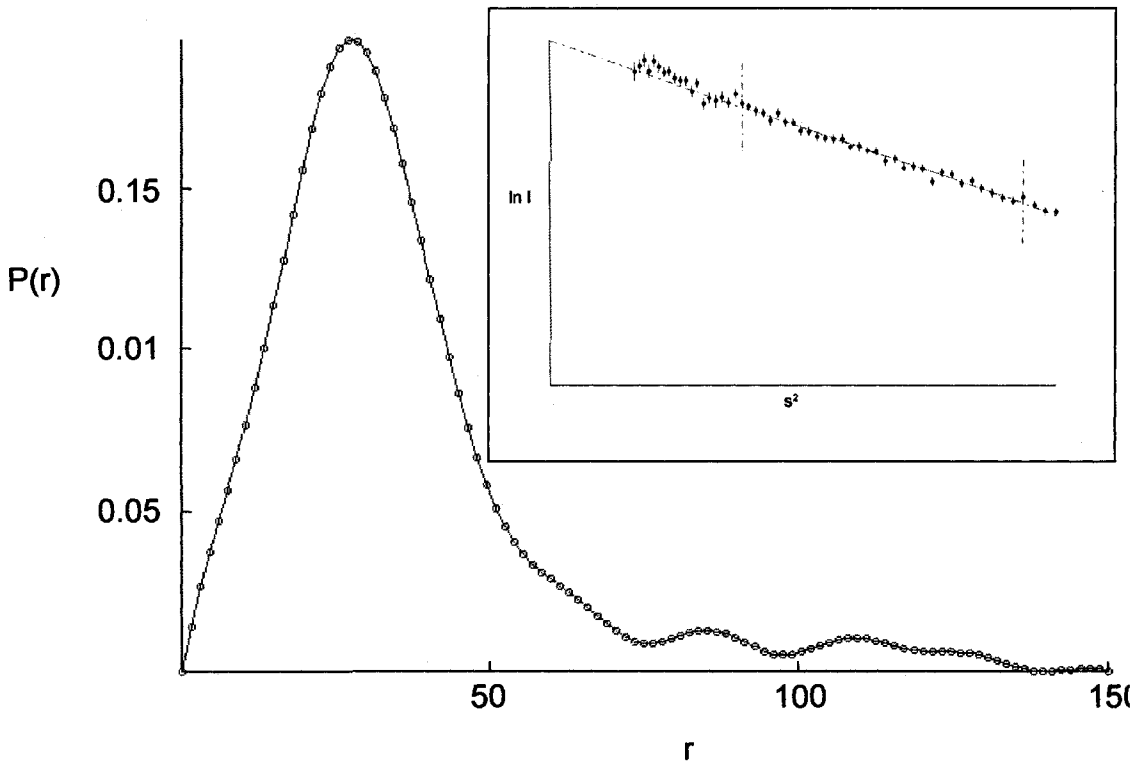

Figure 3-9. Data analysis for hADAR3-D. Inset: The data shows good fit to linearity for points 44-72 in the low angle range ($sR_G < 1.3$). R_G was determined to be 23.3 Å with the Guinier approximation in PRIMUS. Distance distribution function for hADAR3-D is a bell-shaped curve with a maximum at 27 Å and D_{max} of 70 Å. The curve has a slight shoulder in the region of 50-70 Å which may indicate the presence of a distinct domain.

DAMFILT was used to remove regions where the bead density did not meet the threshold limits necessary to generate a particle with the expected excluded particle volume. The quality of the final models reconstructed was evaluated by comparing the goodness of fit parameter χ^2. The χ^2 was determined to be 4.3 and 2.3 for the hADAR2 and hADAR3 models respectively (Figure 3-10 and Figure 3-11).

The *ab initio* low resolution models for both constructs revealed a mostly globular domain with a diameter of approximately 50 Å. However, an unexpected feature was found protruding out of the globular domain. This feature was not found in the X-ray structure, but could possibly correspond to the missing portion of the

unstructured loop. For the hADAR3-D protein, this protrusion shows little variation among the individual models and has a length of 30 Å (Figure 3-10). The structure present in the hADAR2-D shows a larger number of conformations which results in a larger volume occupied by the dummy atoms (Figure 3-11). The length of the protrusion for hADAR2-D is 35 Å. The apparent similarity among the GASBOR generated models for both hADAR2-D and hADAR3-D suggests that the deaminase domains of the hADAR family of proteins share a similar overall structure. This can be further supported by models for hADAR1-D and hADAR3-D generated with 3D structure prediction software Geno3D (Figure 3-12).[28] Geno3D generated models for hADAR1-D and hADAR3-D are also globular and show the conserved deamination motif present in the X-ray structure of hADAR2-D. Moreover, Geno3D has also built in the missing loop structure which distinctively protrudes out of the globular domain in both models.

Rigid body modeling with addition of missing loop structure

The program BUNCH was used to model the missing loop structure in the deaminase domain by a combination of rigid body modeling and *ab initio* model building against the x-ray scattering data (Figure 3-13 A, B).[29, 30] In this model building approach the flexible loop region was represented as an interconnected chain composed of dummy residues with a spacing constraint of 3.8 Å. Each deaminase domain was divided into two segments and restricted by the interconnecting linker sequence. The two domain segments were moved as rigid bodies while searching for

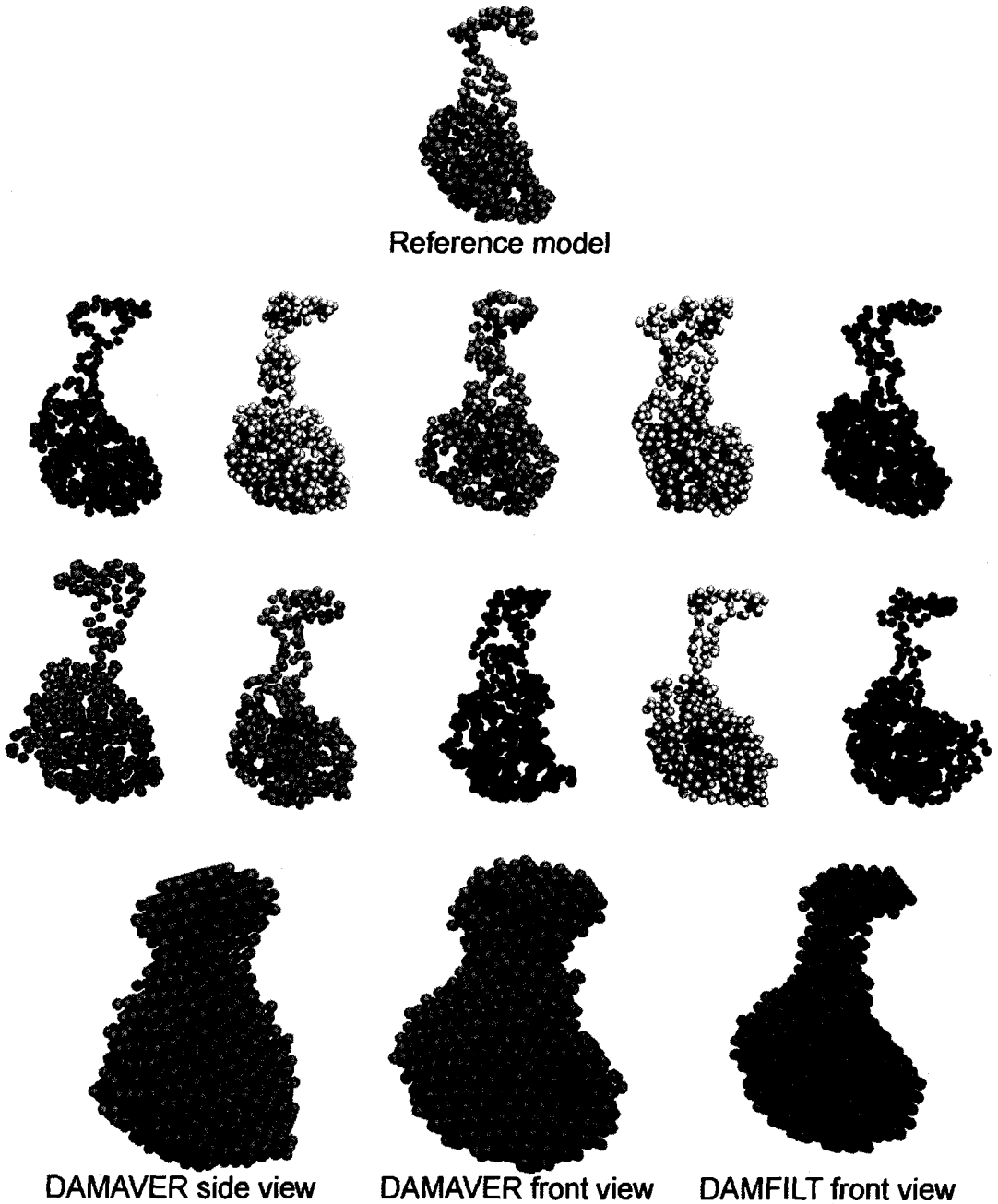

Figure 3-10. Ab initio structure reconstruction using GASBOR for hADAR2-D.

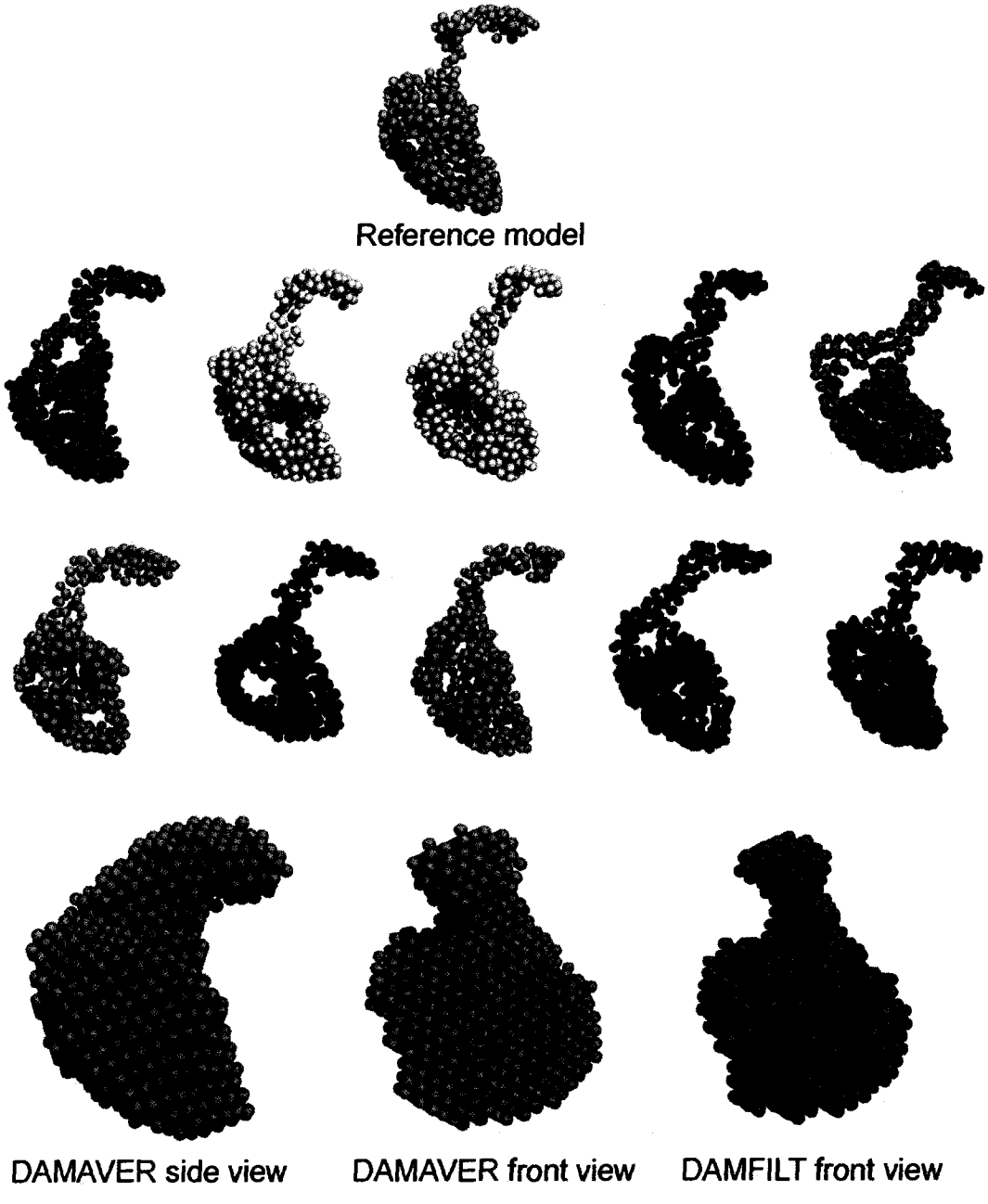

Figure 3-11. Ab initio structure reconstruction using GASBOR for hADAR3-D.

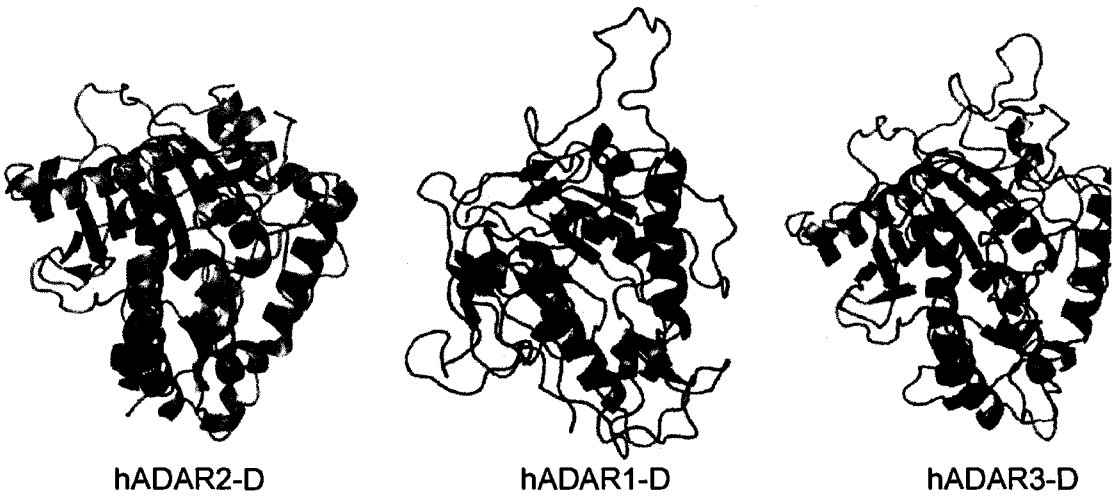

Figure 3-12. Geno3D structure prediction for hADAR1-D and hADAR3-D.

the possible conformations of the flexible linker against the x-ray scattering data. Penalties were imposed on the overlap of the two domains, appropriate distribution of bonds and dihedral angles and fit to R_G values. The configuration of the loop provided by BUNCH is representative of the average conformation that best fits the scattering data. The χ^2 value for the BUNCH generated model was 5.1 which is higher than the χ^2 value of 2.3 for the *ab initio* model. This higher χ^2 may reflect the problem of attempting to find a fixed conformation for a flexible loop. Docking of the BUNCH generated model into the *ab initio* envelope from GASBOR shows that the two models superimpose relatively well, although the volume surrounding the loop structure is much larger than expected (Figure 3-13 A, B). Limited trypsinolysis of hADAR2-D and hADAR3-D constructs generates two fragments whose molecular weights (17.2 and 24.7 kDa) correspond to cleavage at amino residues R455 and K475 located within the flexible loop structure. These results suggest that in the absence of N-terminal sequence of hADAR2 the loop sequence

contained within the deaminase domain is solvent exposed and easily accessible to the proteolytic activity. Trypsinolysis of the full length hADAR2 performed under similar conditions produced only one prominent fragment of 61 kDa which corresponds to the deaminase domain itself, while the 17 and 24 kDa fragments cannot be observed.[20] Furthermore, the protrusion present in the models of the deaminase domain is not visible in the *ab initio* envelope generated for the full length hADAR2, suggesting that this loop is on the interior and possibly interacts with other regions of the protein (Figure 3-14).

3.3. Conclusions

Deaminase domains from hADAR1-3 were expressed and purified from *S. cerevisiae* in order to compare their activity and structural characteristics. The alignment of the three proteins shows high conservation especially among the catalytic residues and residues which interact with the IP_6. The 3D structure prediction based on the high resolution structure of hADAR2-D construct also indicates that the proteins are structurally similar. The purification of hADAR2-D and hADAR3-D produced good yields of pure and active proteins. The yields of hADAR1 were significantly lower due to high levels of aggregation during purification. The editing assays performed indicated that all three deaminase domains are capable of editing adenosine to inosine within the context of a dsRNA substrate, although with lower efficiency than the full length hADAR2. The attempts to crystallize hADAR3-D for X-ray crystallography were not successful. However, the

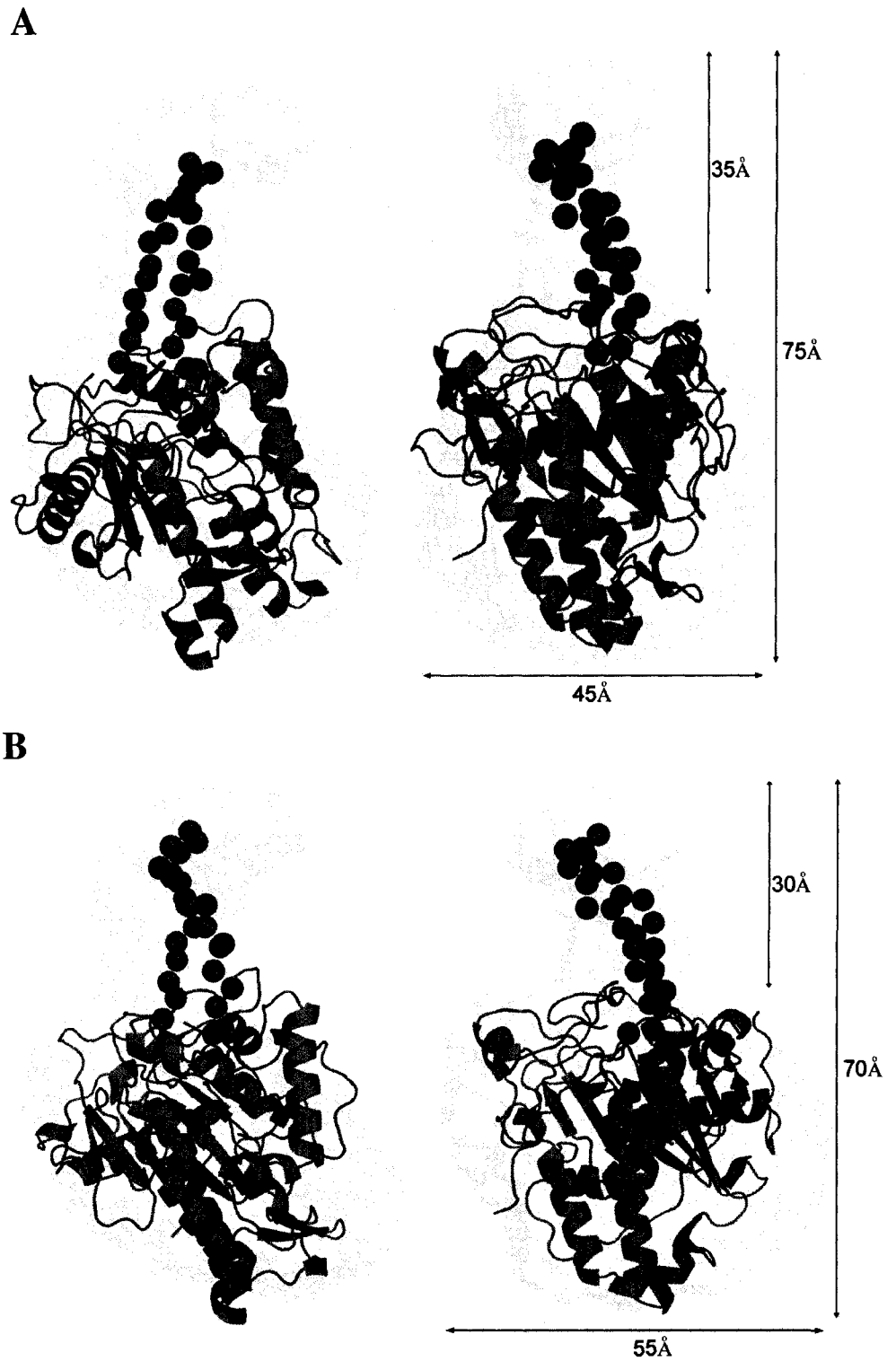

Figure 3-13. Rigid body modeling with the program BUNCH. The program BUNCH was used to build in the missing loop on the dsRNA binding surface of the hADAR2-D protein. The model obtained from BUNCH was docked into the *ab initio* model generated by GASBOR for **(A)** hADAR2-D and **(B)** hADAR3-D.

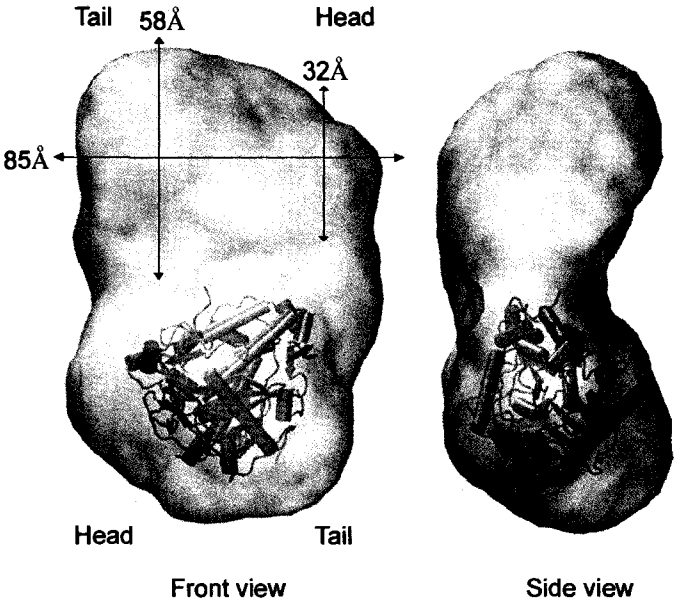

Figure 3-14. Ab initio envelope for the full length hADAR2. The protrusion seen in the *ab initio* model generated for the deaminase domains for hADAR2-D and hADAR3-D cannot be observed in the model of the full length hADAR2.

small angle X-ray scattering data obtained from the purified samples for hADAR2-D and hADAR3-D was evaluated as monodisperse and did not show sensitivity to radiation or concentration dependence. The R_G values were determined to be 24.4 and 23.3 Å for hADAR2-D and hADAR3-D respectively while the D_{max} was found to be 70 Å for both constructs. The pair-distance distribution function was a bell-shaped curve indicative of a globular domain. *Ab initio* structure reconstruction of both hADAR2-D and hADAR3-D yielded similar, mostly globular proteins with a distinctive protrusion. The length of this protrusion was determined to be 35 Å and 30 Å for hADAR2-D and hADAR3-D respectively. Since this feature was not found in the X-ray structure of hADAR2-D it could likely correspond to the missing portion of the unstructured loop. Rigid body modeling program BUNCH was used to built in the missing loop into the X-ray structure of hADAR2-D. The docking of the rigid

body model into the *ab initio* envelope shows that the two models overlap relatively well. The χ^2 value for the BUNCH generated model was 5.1 which is higher than the χ^2 values of 4.3 and 2.3 for the *ab initio* model for hADAR2-D and hADAR3-D respectively. The presence of the distinctive protrusion indicates that the loop does have a preferred range of conformations. The higher χ^2 values may result from attempting to build a model for a structure with inherently flexible elements. Limited trypsinolysis of the hADAR2-D and hADAR3-D constructs produced two fragments with molecular weight of 17.2 and 24.7 kDa. These fragments could be mapped to conserved arginine and lysine residues within the flexible loop in the two deaminase constructs. In contrast, trypsinolysis of the full length hADAR2 produced a single fragment with molecular weight of 61 kDa. These results suggest that the protrusion in the SAXS model represents the missing loop structure and that in the absence of the N-terminal protein sequences the loop is solvent exposed and sensitive to proteolytic activity. Absence of a distinctive protrusion in the *ab initio* envelope for the full length hADAR2 suggests that in the presence of N-terminus the loop interacts with other parts of the protein. Taking into account the auto-inhibition model for the hADAR2 it is tempting to speculate that the flexible loop region interacts with sequences in the N-terminus in the absence of dsRNA. However, when both of the dsRBMs are engaged in binding of the substrate the loop is free to accommodate the substrate for the catalysis. The low resolution models combined with the X-ray structure for hADAR2-D provide basis for experimental design of protein-protein and protein-RNA crosslinking experiments in order to elucidate the interaction between the deaminase domains and their substrates.

3.4. Materials and Methods

General considerations

All reagents used were purchased from Sigma Aldrich unless otherwise stated. Primers for the generation of hADAR1-3 deaminase constructs and RNA substrates were purchased from IDT. All media used in the yeast transformations and culture growth were autoclaved. Glucose, galactose and sorbitol solutions were sterile filtered. Protein purification steps were performed at 4 °C or on ice and after each step the proteins were analyzed by sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and when necessary western blot analysis with Penta-His HRP conjugate antibody (Qiagen).

3.4.1. Cloning and yeast transformation of hADAR2 deaminase domains.

The hADAR deaminase constructs (hADAR1-D [520-931], hADAR2-D [299-701], hADAR3-D [342-739]) were generated by the polymerase chain reaction (PCR) with the following oligonucleotides: N-terminal primers: 5'-GCGCGCYCAAAT G(CACCAT)$_5$GAAAATTTGTATTTTCAAGGT(X)$_{15}$-3' and C-terminal primers: 5'-GCGCGCYTCATCA(X)$_{15}$-3'. N-terminal primer contains a restriction site (Y), followed by the CAA spacer and initiator methionine (ATG) which allow for efficient translation initiation. The 10-histidine tag followed by a tobacco etch virus (TEV) protease recognition site proceed the 15 nucleotides which code for the N-terminal sequence of the hADAR gene (X_{15}) of interest. The C-terminal primer contains a restriction site and two stop codons followed by 15 nucleotides encoding for the C-terminal sequence of the corresponding hADAR gene (X_{15}). The PCR product for

each deaminase domain construct was digested with the following sets of restriction enzymes: hADAR1-D with BamHI and EcoRI, hADAR2-D with KpnI and EcoRI and hADAR3-D with BamHI and KpnI. The PCR products were ligated into the pYeDP60 yeast expression vector (generous gift from Dr. Howard Young's lab, University of Alberta), linearized with restriction enzymes corresponding to the set used for the respective gene of interest. pYeDP60 contains a galactose inducible promoter (GAL10-CYC1), polylinker region with the following restriction sites BamHI/SmaI/KpnI/SacI/EcoRI, URA3 and ADE2 (uracil and adenine auxotrophies) markers and ampicillin resistance marker. The restriction site immediately downstream of the promoter must be as close as possible to the cDNA initiation codon in the construct for efficient translation. pYeDP60 was chosen for its stability in rich media at high cell densities. The cloned sequences of the deaminase domains genes were sequenced using primers complementary to 5' and 3' ends as well as from within the sequence. The haploid strain DSY904 with genotype: *bar/ade/his2ura3Δleu2trp1pep4::LEU2* which is deficient in several proteases and allows selection of URA3 auxotrophs was selected for having best expression pattern for hADAR deaminase constructs.

3.4.2. Protein expression from *S. cerevisiae*

The expression checks as well as large scale protein expression were performed according to the same standard protocol for all three deaminase domain constructs. After transformation, cells were plated on SDC URA$^-$ plates (1.7 g/L yeast nitrogen base (Fisher), 5 g/L ammonium sulfate, 2 g/L casamino acids (Fisher),

10.0 g/L succininc acid, 6 g/L ammonium hydroxide and 20 g/L agar supplemented with 1 M sorbitol and 2% (w/v) glucose) and incubated at 30 °C. Colonies appeared after 2-3 days. Expression from several different clones for each construct was tested on a small scale prior to a large scale expression. Small scale expression involved picking of 3-5 individual colonies to inoculate 10 mL of SDC URA$^-$ (1.7 g/L yeast nitrogen base (Fisher), 5 g/L ammonium sulfate, 2 g/L casamino acids (Fisher), 10.0 g/L succinic acid, 6 g/L ammonium hydroxide and supplemented with 2% glucose) for 24 h at 28 °C. A sample of this growth (1 mL) was used to inoculate 50 mL of YPGE media (10 g/L of yeast extract, 10 g/L of peptone, 5 g/mL of glucose, 3% (v/v) EtOH). The cells were grown for 16 h at 28 °C upon which the culture was induced with 2% (w/v) galactose and cells were grown for 6 more hours. The cultures were pelleted at 4000 rpm for 30 min at 4 °C. The supernatant was removed and the pellet resuspended in Y-PER® yeast lysis reagent (125 µL/ 50 mg of wet pellet, Pierce) in the presence of Complete® protease inhibitor cocktail (Roche). The lysis was carried out on ice for 20 min and then the cells were pelleted at 4000 rpm for 30 min at 4 °C. Western blot analysis with Penta-His HRP® conjugate antibody (Qiagen) was carried out on all test colonies and a no-galactose control. Colonies which produced best expression levels of the protein were selected for large scale expression.

For the large scale expression, 10 µL of glycerol stocks were used to inoculate 10 mL of minimal media SDC URA$^-$. The cultures were grown for 24 h at 30 °C while shaking at 250 rpm. The next morning, 4 mL of the starter culture was used to inoculate 200 mL of minimal media and grown for 24 h in the same conditions as the

starter culture. A portion of the secondary culture (40 mL) was used to inoculate 2 L of YPGE media (A_{600}=0.05) supplemented with adenine (5 mg/mL) and grown at 28 °C until cell density has reached 8 x 10^7 cells/mL (24-30 h, 1 x 10^6 cells/mL gives an A_{600} = 0.1). The 2 L culture was induced with 2% (w/v) galactose for 8-15 h and grown until cell density reached 2 x 10^8 cells/mL (8-12 h) The cultures were pelleted at 7000 rpm for 15 min at 4 °C in a Sorval Evolution RC centrifuge equipped with a SLC 6000 4.8 L fixed angle rotor (Mandel). The cells were washed with buffer containing 20 mM Tris-HCl, pH 8.0, 100 mM NaCl, 5% glycerol, 0.02% (w/v) sodium azide and centrifuged again. The wet mass of the pellet was determined prior to storage at -80 °C. The typical yield from 1 L of culture grown in the YPGE rich media is 10-12 g of wet pellet.

3.4.3. Protein purification for ADAR deaminase domains

Cells from 4 L of culture (40-44 g of pellet) were resuspended in Buffer A containing 750 mM NaCl, 0.01% NP-40 and Complete® protease inhibitors. The cells were lysed by 5 passes through a Gaulin homogenizer at 20,000 psi[2] and the extract was purified by centrifugation at 100,000x *g* for 1 h. The lysate was filtered through an Amicon steri-cup 0.22 μM filter device (Millipore). Ni-NTA agarose (2 mL; GE Healthcare) equilibrated with the lysis buffer was added directly to the filtered cell lysate and batch bound for 1.5 h with gentle rocking. The slurry was poured into 2.5 x 20 cm gravity drip column (Biorad) while collecting the flow through by gravity. The resin was washed in three consecutive 40 mL steps with Buffer A containing 35 mM imidazole and 750 mM, 300 mM, and 200 mM NaCl

respectively. The protein was eluted in eight 5 mL fractions with Buffer A containing 200 mM NaCl and 400 mM imidazole. The flow-through, washes, and elution fractions were analyzed by SDS-PAGE and the proteins visualized by staining with Coomassie blue. Most of the protein typically eluted in Fractions 3 to 6 (15 mL) for all three deaminase domains. These fractions were pooled and diluted up to a total volume of 50 mL with Buffer A (final concentration of NaCl ~100 mM). The AKTA FPLC system was used to load the eluted protein onto a 5 mL HiTrap Heparin column (GE Healthcare). The protein was eluted with a salt gradient from 100 mM to 1 M NaCl. Fractions containing protein were pooled together and concentrated to 2 mL with an Amicon-Ultra 15 centrifugal concentrator tube with a 5,000 Da nominal molecular weight cut off (M.W.C.O.). Fractions containing protein were pooled together and concentrated to 2 mL. TEV protease (Source: *E. coli*, 20 U per 1 mg of protein) was added and the TEV cleavage reaction was incubated at 4 °C overnight. The efficiency of the TEV reaction was evaluated by Western blotting with penta-His antibody which has shown the reaction to be 95% complete after overnight incubation. The whole TEV cleavage reaction was injected onto a Superdex 200 26/60 gel filtration column and eluted with Buffer A containing 200 mM NaCl. The deaminase domain constructs eluted as a symmetrical peak detected at 225mL. The main peak fractions were pooled, concentrated and dialyzed against storage buffer containing 20 mM Tris, pH 8.0, 400 mM NaCl, 20% glycerol and 1 mM BME. At salt concentrations below 200 mM the deaminase constructs were stable for short periods of time (1-3 days). Thus the proteins were stored in buffer containing 400 mM NaCl at -80 °C and dialyzed into appropriate buffer

immediately prior to the experiment. The protein were quantified by the Bradford method using Protein Concentration Reagent (Biorad). The yields for hADAR2-D and hADAR3-D were significantly higher than for the full length hADAR2 on average ranging from 0.5 mg – 1.5 mg of protein from a liter of original yeast culture. The typical yields for hADAR1-D ranged between 0.05-0.1 mg of protein per liter of culture.

3.4.4. Protein activity assay

Templates for the 78 nucleotide R/G pre-mRNA were generated by PCR from a substrate cloned into pBS vector (Stratagene) with primers: **5'**: 5'-TAATACGACTCACTATAGGGGTCCTCATTAAGG-3' and **3'**: 5'-ACACATCAGGGT AG-3'. [α-^{32}P] body-labeled 78 nucleotide R/G pre-mRNA was generated in a 25 µL transcription carried out in transcription buffer containing 40 mM Tris, pH 8.0, 6 mM $MgCl_2$, 10 mM NaCl, 10 mM DTT, 2 mM spermidine, 500 µM of each UTP, CTP and GTP, 15 µM of ATP, 1 U/mL of RNase Out (Invitrogen), 100 µCi α-^{32}ATP (Perkin Elmer), 0.1 µM DNA template and 2 U of T7 RNA polymerase (Invitrogen). Transcriptions were incubated at 37 °C for 4 h. The oligonucleotides were purified by denaturing PAGE (8%, 19:1).

Editing reactions.

hADAR1, hADAR2-D and hADAR3-D were dialyzed against editing buffer containing 20 mM Hepes pH 8.0, 100 mM KCl, 0.5 mM DTT, 20% glycerol and 0.01% NP-40. The body-labeled substrate was resuspended in ddH_2O, denatured at 80 °C for 2 min and allowed to renature at rt for 30 min just prior to its addition to the

editing reaction. Typical editing reactions were performed in 20 μL total volume containing editing buffer, 0.5 nM RNA substrate and varying amounts of enzyme (1-100 μM) and were incubated at either rt or 30 °C for 4 h. The editing reactions were quenched by extraction with phenol/chloroform/isoamyl alcohol (25:24:1) and then with chloroform followed by ethanol precipitation.

Nuclease P1/TLC cleavage assay

The pellets resulting from the editing assays were resuspended in 9.5 μL of ddH_2O and digested with Nuclease P1 (0.5 U/reaction) overnight at 37 °C. The products of the digestion were resolved by thin layer chromatography (saturated $(NH_4)_2SO_4$:0.1 M NaOAc:isopropanol; 79:19:2 (v/v)) using cellulose-PEI chromatography plates pre-run in ddH_2O. The TLC plates were exposed to a Molecular Dynamics Phosphor screen which was scanned using a Molecular Dynamics Storm 860 PhosphorImager. Data was quantified using Image Quant 5.0 software.

3.4.5. Sample preparation for SAXS

A sample of each construct hADAR1-D at 0.5 mg/mL, hADAR2-D (299-701) at 1.5 mg/mL, and hADAR3-D (342-739) at 2.6 mg/mL was dialyzed against 50 mL of buffer containing 20 mM Tris-HCl, pH 8.0, 400 mM KCl, 5% glycerol and 1 mM BME overnight at 4 °C with the use of 50 μL dialysis buttons (Hampton Research) and 3,500 Da M.W.C.O. regenerated cellulose dialysis tubing.

3.4.6. SAXS data collection.

SAXS X-ray scattering data were collected at SIBYLS beamline, Advanced Light Source, Lawrence Berkeley National Laboratory. Samples of each deaminase construct (15 μL) were cooled to 15°C and data was collected at 12 KeV on a MAR165 CCD detector with sample to detector distance of 1.5 m. Immediately prior to data measurement the sample was centrifuged at 10,000 rpm for 10 min and divided into two equal volumes. Data collection was carried out on the first half of the sample as it was. The second half of the sample was diluted in half before the measurement. For each sample a series of four exposures including 6 s, 6 s, 60 s and 6 s were measured to assess sample's sensitivity to radiation. The software program OGRE was used to normalize the data to the intensity of the transmitted beam and subtract the scattering of the buffer from the scattering of the hADAR2 containing solution. Data were analyzed using the programs PRIMUS and GNOM.[24, 25]

3.4.7. Limited proteolysis of hADAR2-D and hADAR3-D

Lyophilized trypsin (Pierce) was resuspended in Buffer A and 200 mM NaCl. For a 10 μL reaction 8 μg of purified hADAR2-D or hADAR3-D was mixed with 0.01 μg of trypsin. The reactions proceeded at room temperature for varying times (0-60 min) and were quenched by addition of 10 μL of SDS-PAGE loading buffer. Samples were boiled at rt for 30 sec then electrophoresed in 12% polyacrylamide minigels. Peptide fragments were visualized with Coomassie blue. Alternatively, the reactions were quenched by addition of 10% (v/v) of trifluoroacetic acid (TFA) and submitted for MALDI mass spectroscopy analysis.

3.5. Notes and References

1. Seeburg, P.H. The role of RNA editing in controlling glutamate receptor channel properties. *J Neurochem* 1996, **66**:1-5.
2. Seeburg, P.H., Higuchi, M., Sprengel, R. RNA editing of brain glutamate receptor channels: Mechanism and physiology. *Brain Res Rev* 1998, **26**:217-229.
3. Rueter, S.M., Emeson, R.B. Adenosine to Inosine Conversion in mRNA. In H.Grosjean and R.Benne (ed.), Modification and editing of RNA. 1998, American Society for Microbiology, Washington, D.C.
4. Higuchi, M., Single, F.N., Kohler, M., Sommer, B., Sprengel, R., Seeburg, P.H. RNA editing of AMPA receptor subunit GluR-B: a base-paired intron-exon structure determines position and efficiency. *Cell* 1993, **75**:1361-1370.
5. Seeburg, P.H., Single, F., Kuner, T., Higuchi, M., Sprengel, R. Genetic manipulation of key determinants of ion flow in glutamate receptor channels in the mouse. *Brain Res* 2001, **907**:233-243.
6. Lai, F., Chen, C.X., Carter, K.C., 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**:2413-24.
7. Lomeli, H., Mosbacher, J., Melcher, T., Hoger, T., Geiger, J.R., Kuner, T., Monyer, H., Higuchi, M., Bach, A., Seeburg, P.H. Control of kinetic properties of AMPA receptor channels by nuclear RNA editing. *Science* 1994, **266**:1709–1713.
8. Stefl, R., Xu, M., Skrisovska, L., Emeson, R.B., Allain, F.H. Structure and specific RNA binding of ADAR2 double-stranded RNA binding motifs. *Structure* 2006, **14**:345-55.
9. Yi-Brunozzi, H.Y., Stephens, O.M., Beal, P.A. Conformational changes that occur during an RNA editing adenosine deamination reaction. *J Biol Chem* 2001, **276**:37827-37833.
10. Stephens, O.M., Haudenschild, B.L., Beal, P.A. The binding selectivity of ADAR2's dsRBMs contributes to RNA-editing selectivity. *Chem Biol* 2004, **11**:1239-1250.
11. Wong, S.K., Sato, S., Lazinski, D.W. Substrate recognition by ADAR1 and ADAR2. *RNA* 2001, **7**:846–858.
12. Lehmann, K.A., Bass, B.L. Double-Stranded RNA Adenosine Deaminases ADAR1 and ADAR2 Have Overlapping Specificities. *Biochemistry* 2000, **39**:12875-12884.
13. Hart, K., Nystrom, B., Ohman, M., Nilsson, L. Molecular dynamics simulations and free energy calculations of base flipping in dsRNA. *RNA* 2005, **11**:609-618.

14. Kallman, A., Sahlin, M., Ohman, M. ADAR2 A to I editing: site selectivity and editing efficiency are separate events. *Nucleic Acids Res* 2003, **31**:4874-4881.
15. Macbeth, M.R., Schubert, H.L., Vandemark, A.P., Lingam, A.T., Hill, C.P., Bass, B.L. Inositol hexakisphosphate is bound in the ADAR2 core and required for RNA editing. *Science* 2005, **309**:1534-1539.
16. Betts, L., Xiang, S.A., Short, R., Wolfenden, C.W., Carter, Jr. Cytidine Deaminase. The 2·3 Å Crystal Structure of an Enzyme: Transition-state Analog Complex. *J Mol Biol* 1994, **235**:635-656.
17. York, J.D., Odom, A.R., Murphy, R., Ives, E.B., Wente, S.R. A Phospholipase C-dependent inositol polyphosphate kinase pathway required for efficient messenger RNA export. *Science* 1999, **285**:96-100.
18. Shen, X., Xiao, H., Ranallo, R., Wu, W.H., Wu, C. Modulation of ATP-dependent chromatin-remodeling complexes by inositol polyphosphates. *Science* 2003, **299**:112-114.
19. Steger, D.J., Haswell, E.S., Miller, A.L., Wente, S.R., O'Shea, E.K. Regulation of chromatin remodeling by inositol polyphosphates. *Science* 2003, **299**:114-116.
20. Macbeth, M.R., Lingam, A.T. Bass, B.L. Evidence for auto-inhibition by the N-terminus of hADAR2 and activation by dsRNA binding. *RNA* 2004, **10**:1563-1571.
21. Chen, C-X., Cho, D.-S., Wang, Q., Lai, F., Carter, K.C., 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**:755-767.
22. SAXS data analysis and model building studies were performed with the use of Linux scripts courtesy of Dr. Ross Edwards. We thank Dr. Edwards for his help on the SAXS methodology.
23. Guinier, A., Fournet, F. Small angle scattering of X-rays. 1955. Wiley-Interscience, New York.
24. Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H., Svergun, D.I. PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J Appl Cryst* 2003, **36**:1277-1282.
25. Semenyuk, A.V., Svergun, D.I. GNOM – a program package for small-angle scattering data processing. *J App Cryst* 1991, **24**:537-540.
26. Svergun, D.I., Baraberato, C., Koch, M.H. CRYSOL – a program to evaluate X-ray solution scattering of biological macromolecules from atomic coordinates. *J Appl Cryst* 1995, **28**:768-773.

27. Svergun, D.I., Petoukhov, M.V., Koch, M.H. Determination of domain structure of proteins from X-ray solution scattering. *Biophys J* 2001, **76**:2879-2886.
28. Combet, C., Jambon, M., Deleage, G., Geourjon, C. Geno3D: Automatic comparative molecular modeling of proteins. *Bioinformatics* 2002, **18**:212-214.
29. Petoukhov, M.V., Svergun, D.I. Global rigid body modeling of macromolecular complexes
against small-angle scattering data. *Biophys J* 2005, **89**:1237-1250.
30. Petoukhov, M.V., Eady, N.A., Brown, K.A., Svergun, D.I. Addition of missing loops and domains to protein models by X-ray solution scattering. *Biophys J* 2002, **83**:3113-3125.