RNA EDITING BY CYTIDINE DEAMINATION IN MAMMALS


5 Adenosine deaminases that act on RNA

RONALD F. HOUCH AND BRENDA L. BASS

1. Introduction to a family of A→I RNA editing enzymes

The RNA editing enzymes now known as adenosine deaminases that act on RNA (ADARs) comprise a unique protein family with members throughout the metazoa. Discovered in 1987 in Xenopus laevis, the enzymes were first characterized as double-stranded RNA (dsRNA) unwinding activities (1, 2) and later shown to be dsRNA adenosine-to-inosine modifying activities (3, 4). The observed ‘unwinding’ of the RNA turned out to be a consequence of the modification which, in the context of dsRNA, converts AU base-pairs to the less stable IU mismatch.

Two different names were originally given to the first member of the family, ADAR1, after its purification from X. laevis eggs and bovine liver: dsRAD (dsRNA adenosine deaminase), 5) and DRADA (double-stranded RNA adenosine deaminase); 6) Following the identification of the second member of the family, RNA editase L (RED-1); 7) now known as ADAR2, a standardized nomenclature was adopted. The acronym ADAR, an extension of the three-letter acronym designating adenosine deaminases that act on nucleosides (ADA), was recommended by the Human Genome Organization and subsequently adopted by researchers in the field (8).

ADARs target RNAs that are largely, or completely, double-stranded and convert adenosines to inosines by catalyzing a hydrolytic deamination at Cs of the adenine base (see Fig. 1). 9) In some substrates, editing occurs at specific sites to produce codon changes within open reading frames (ORFs), while in others, particularly those with long regions of uninterrupted duplex RNA, deamination is more promiscuous and results in hypermutations (see Chapter 6: 10). In vitro studies of the best-characterized ADARs, ADAR1 and ADAR2, show that these enzymes have overlapping specificities and produce similar, but not identical, editing patterns. Because some adenosines can be targeted by both enzymes, which enzyme is responsible for editing each of the endogenous substrates in vivo is still, in most cases, ambiguous (see section 5.3). In fact, it is possible that some sites are targeted by both enzymes. Chapter 6 of this volume reviews the natural substrates of ADARs with an emphasis
on biological function, while here we focus on the properties of the enzymes themselves.

1.1 General features of the primary structure

There are now at least 21 unique DNA sequences deposited in DNA and genomic databases that encode members of the ADAR family (including putative homologs in different organisms but excluding alternatively spliced variants). Based on alignments of the amino acid sequences predicted from these DNA sequences, ADAR topology can be divided into three distinct regions (Fig. 2):

- amino-termini that vary significantly in length and sequence;
- central regions with variable numbers of dsRNA-binding motifs (dsRBMs) separated by spacers that vary in length and sequence; and
- highly conserved carboxyl-termini containing the deaminase catalytic domain.

As shown in Fig. 2, ADARs containing one to three copies of the dsRBMs have been identified. ADARs containing three dsRBMs have been found in human (11, 12), cow (13), rodents (11), chicken (14) and X. laevis (5, 15); because these were the first members of the family identified, they are referred to as ADAR1. ADARs with two dsRBMs have been identified in mammals (ADAR2) and Caenorhabditis elegans (H5SN14.1a/b). ADARs containing only a single motif have been identified in C. elegans (T20H4.4; 16) and mouse (mouse nuclear RNA-binding protein, or Tenr; 17) although the latter has not yet been proven to have deaminase activity. In this regard, studies of worms containing deletions in H5SN14.1a/b or T20H4.4 suggest that the C. elegans ORFs encode active ADARs (L. Tonkin, D. Morse and E. Bais, unpublished data), but in vivo activity with recombinant proteins has not yet been demonstrated.

Importantly, it is not yet clear if ADARs containing the same number of dsRBMs will always be homologs. Although the sequences and biochemical properties of ADAR1 from different organisms are very similar, suggesting that enzymes with three copies of the dsRBMs are homologs, this may be the exception to the rule.

Interestingly, enzymes lacking any dsRBMs and containing only the catalytic domain have been shown to be responsible for converting adenosines to inosines in tRNAs (18-20). These enzymes are referred to as ADATs (adenosine deaminases that act on RNA), and their existence suggests that ADARs evolved from tRNA-modifying enzymes (see section 4.4).

1.2 Intracellular localization

Some of the natural targets for ADARs are pre-mRNAs in which the requisite dsRNA structure is created by base-pairing between exons and nearby introns (see Chapter 6). In these cases, RNA editing must occur before splicing. Thus, it is not surprising that ADARs appear to be predominantly nuclear proteins. However, in certain interesting cases ADARs are found in the cytoplasm. For example, in X. laevis oocytes, ADAR activity is in the nucleus until it is released when the nucleus breaks down as a natural process of oocyte maturation. After the mature oocyte, or egg, has been fertilized, embryogenesis begins and ADAR activity is again sequenced to the nucleus, sometime between the eight-cell stage and mid-blastula transition (1, 2). At present it is not clear if this change in localization is merely a consequence of nuclear breakdown, or is functionally important, for example, to regulate an early develop-
2. ADAR1

ADAR1, the first protein in the family to be isolated, was purified by assaying chromatographic fractions for deaminase activity, using a completely base-paired dsRNA unrelated to natural ADAR substrates. Proteins of similar size (116-122 kDa by SDS-PAGE) were purified from X. laevis (122 kDa; 5), calf thymus (116 kDa; 13) and bovine brain (120 kDa; 26). Smaller proteins (93-83 kDa) purified from bovine liver (6) were later shown to be proteolytic fragments lacking the amino-terminus (12, 27). Surprisingly, the subsequent cloning and sequencing of ADAR1 cDNAs revealed ORFs that encoded much larger proteins (Fig. 2). Proteins synthesized from these long cDNAs, identified for X. laevis (~139 kDa; 15, 28), human (~136 kDa; 11, 12, 24) and rat (~130 kDa; 11), migrate during SDS-PAGE with apparent molecular masses of >150 kDa.

It now seems clear that the smaller, ~120 kDa, forms of ADAR1 were not artifacts generated during purification, but indicative of normal pathways that operate in vivo. Several mechanisms for producing the smaller proteins have been considered, including the processing of the RNA or protein, and internal initiation at a downstream AUG codon (15). The latter idea has focused on a metionine that is conserved in all ADAR1s (human Me396, X. laevis Me384; see section 2.2.1 and, if serving as the initiation codon, would yield a protein of the size observed for the truncated, short ADAR1 form.

Although the story is far from complete, recent experiments provide clues about the relationship of the short and long forms of ADAR1. Studies in human cells show that a mRNA containing an alternative 5' exon, that excludes the methionine (Met1) required to initiate the long protein, encodes a short form of ADAR1 (see section 2.1.1). Experiments in X. laevis show that in addition to mRNAs that encode long and short forms, smaller, amino-terminally truncated proteins can be produced by proteolytic processing of the larger precurso polypeptide (see section 2.1.2). Whether internal initiation on full-length ADAR1 mRNAs represents a third mechanism for generating the smaller protein is not yet known.

2.1 Expression of the ADAR1 gene

The human ADAR1 gene (lq21.1-21.2) consists of at least 17 exons, spanning at least 39 kb (29-31 and see GenBank accession numbers X79448 and X79449). In mammals, the gene is expressed in a wide variety of tissues. Although some differences exist between species, high expression is always observed in brain, while low levels exist in skeletal muscle (12, 32). The latter correlates with the amount of inosine found in mRNA of various rat tissues, which is high in brain and low in skeletal muscle (32).

With few exceptions, all ADARs studied to date, in all organisms, are alternatively processed to yield a heterogeneous set of mRNAs. For ADAR1, we currently have the most information about processing of human and X. laevis transcripts, and these systems are described in the following sections. Among these organisms, three types of processing have been documented for ADAR1 transcripts:

- alternative splicing to allow inclusion of different 5'-most exons;
- alternative splicing to alter the distance between dsRBMs or dsRBMs and the catalytic domain; and
- alternative processing to alter the length of the 3' untranslated region (UTR).

2.1.1 mRNA processing in mammals

Northern analyses show a single band >7.5 kb in calf thymus poly(A)+ mRNA (11) and a single band from human sources in the range of approximately 6.7-7.1 kb (11, 12, 24, 31). As mentioned, processing of ADAR1 transcripts is very common, so neither of these bands is likely to be homogeneous. In fact, in human tissues, ADAR1 mRNAs with different 5' exons have been identified (31, 34; see also GenBank accession numbers X79448 and X79449) as well as several other variants showing alternative splicing in the central region of the mRNA (35).

At the 5' end of human ADAR1 mRNAs, exon 1A or exon 1B is alternatively spliced to exon 2 (Fig. 3). Synthesis of these alternative 5'-most exons requires the use of at least two promoters and can explain the existence of the long and short forms of the ADAR1 protein. The long form of the protein is encoded by an mRNA containing exon 1A, which is located ~5 kb upstream of exon 2, downstream of an interferon-α-inducible promoter (Pα) that contains a consensus interferon-α/β-galactosidase response element (ISRE) and a kinase conserved sequence (KCS). Interestingly, the KCS is also found at the promoter of another protein that contains dsRBMs, PKR (01; see section 6.3): mRNAs containing exon 1A include an initiation codon (Met1) that allows synthesis of the larger, 1226 amino acid, interferon-inducible form of human ADAR1 (p130). A second constitutive promoter (Pγ) produces mRNAs containing exon 1B, which encodes a short form of ADAR1. Exon 1B is located ~14.5 kb upstream of exon 2 (31). Exon 1B lacks an initiation codon equivalent to Met1 and encodes an amino-terminally truncated ADAR1 of 931 amino acids (p110) which presumably initiates at Met296 (Fig. 3).

The alternative exons observed in human ADAR1 transcripts appear to satisfactorily address the discrepancy between the predicted long ORFs encoded by the
cloned cDNAs, and the predominant, shorter protein products found in cellular extracts. However, other observations suggest there is more to this story. For example, despite the fact that the truncated ADAR1 encoded by the exon 1B-spliced mRNA appears to be the constitutive, more abundant protein, cDNAs previously isolated from human libraries all encode the long (i.e. exon 1A spliced) ORF of ADAR1. These include cDNAs prepared from interferon-ω-stimulated U cells (24) as well as cells predicted not to have an interferon-inducible form of the protein, such as unstimulated human NK cells (12) and HeLa cells (11). Future studies will be required to resolve these apparent inconsistencies.

In addition to artificial splicing of exon 1A and exon 1B, three internal splice variants of human ADAR1 have been detected by screening cDNA libraries by PCR (ADAR1a-c: Fig. 3; 35). These variants result from utilization of alternative splice sites for exons 6 and 7, and yield isoforms with different linker regions between dsRBM2 and dsRBM3, or dsRBM3 and the carboxy-terminal catalytic domain. (By convention, dsRBMS are numbered according to their order, amino- to carboxy-terminal.) As shown in Fig. 3, the ADAR1a isoform contains the longest linker regions. In the ADAR1b isoform the ORF is missing 26 amino acids between dsRBM3 and the catalytic domain. ADAR1c contains the same deletion as ADAR1b, as well as a 19 amino acid deletion that shortens the distance between dsRBM2 and dsRBM3. The only mouse ADARI cDNA identified so far corresponds to splice-form ADAR1b (see GenBank accession number AF052506), but at least two splice-forms are represented in a mouse macrophage-like library (35).

The activities of the various isoforms have been compared in vitro, using partially purified, overexpressed human recombinant ADAR1a-c (36, 37). The isoforms have been evaluated in the context of both exon 1A and 1B, although how these alternative splicing events are coupled in vivo is not clear. No major differences are observed between the in vitro activities of the various isoforms, although there is a hint that the isoforms target endogenous substrates with slightly different efficiencies. Any differences are subtle, and future experiments will be required to understand the relationship of the in vitro studies to differences that may exist between the isoforms in vivo.

Some data suggest that mutations in the dsRBMS have different effects in the context of the different splice variants (35).

2.1.2 mRNA and protein processing in X. laevis
cDNA analyses indicate there are two ADARI genes in X. laevis (ADAR1.1 and ADAR1.2: Fig. 3; 15, 28); this is not surprising, since duplicate genes are common in this pseudotetraploid species (38). X. laevis ADAR1.1 and ADAR1.2 both encode active deaminases (15) and show ~50% identity to mammalian ADAR1, although the carboxy-terminal domains are even more conserved (see Fig. 2).

cDNA sequences from the different X. laevis genes are 90% identical at the nucleotide level, and for the most part these differences are distributed throughout their length. However, the 5' sequences differ in that ADAR1.1 has 14 copies of a repeat that is present only once in ADAR1.2 (see corriged box, Fig. 3). In fact, this single repeat may not even be translated in ADAR1.2, and is shown as a 5' UTR in Fig. 3. A cDNA analogous to Met1 of ADAR1.1 has not been identified in ADAR1.2, despite extensive screening with 5' cDNA probes (15, 28) and rapid amplification of cDNA ends (RACE; 28). Thus, the ADAR1.2 cDNAs probably encode a short form of the protein that initiates downstream at the conserved tripeptide sequence Asp-Met-Ala (DMA; Fig. 3) corresponding to the nucleotide sequence GAU/C/AUGGC (15), which fits the consensus for a eukaryotic translation initiation signal (39).

To determine if the conserved methionine codon could serve as the initiation codon for the native ADAR1 proteins, recombinant, full-length ADAR1.1 (1270 amino acids; ~139 kDa), a long form of ADAR1.2 synthesized from a construct engineered to contain an initiation codon at the very 5' end of the cDNA (1125 amino acids; ~124 kDa) and truncated versions of each protein designed to initiate at the methionine of
the conserved DNA (907 amino acids; ~100 kDa) were synthesized in vitro and compared directly with the endogenous enzyme purified from egg homogenates (5). The truncated forms of ADAR1.1 and ADAR1.2 were indistinguishable from one another and from the native enzyme, each displaying an apparent molecular mass of ~126 kDa after electrophoresis in 6-12% SDS-polyacrylamide gels (R. Hough and B. Bass, unpublished results). The longer versions migrated much more slowly, displaying apparent molecular masses of >150 kDa for ADAR1.1 (~167 kDa) and ADAR1.2 (~133 kDa).

Although ADAR1.2 may encode a short form of the protein, peptide sequence analyses performed on the smaller, native protein purified from X. laevis eggs show that at least a portion of this native protein is derived from the ADAR1.1 gene, which encodes the long form (15). Further, isoelectric focusing of the native protein, purified from X. laevis eggs, resolved four co-migrating species with pIs of 5.50, consistent with the predicted pI of truncated ADAR1.1 (5.63) and ADAR1.2 (5.52). Possibly ADAR1.1 could initiate from Met634 to generate a truncated protein like that proposed to be synthesized from ADAR1.2. Recent results offer another possible resolution to this paradox. In particular, when mRNAs encoding full-length ADAR1.1 are injected into oocytes, a full-length protein as well as a protein similar in size to the short ADAR1 are observed (Fig. 3; 40). The migration of the truncated protein by SDS-PAGE, and the observation of low molecular weight amino-terminal fragments, suggest that the short protein results from proteolytic cleavage of the long protein. The size of the truncated protein suggests that the specific proteolytic cleavage site(s) reside downstream of the Z-DNA-binding subdomain, Z6 (27; see section 2.3, Fig. 3). Consistent with the latter, the truncated ADAR1 proteins have mobilities similar to that of X. laevis ADAR1.1 expressed from the conserved Met634, which is analogous to human Met296 (40). Similar to the results of studies in human cell lines, the longer form of the protein preferentially accumulates in the cytoplasms, while the truncated protein is most abundant in the nucleus.

These studies also show that the endogenous ADAR1 in oocytes and X. laevis cells migrates as a single band during SDS-PAGE (40), consistent with the size of the purified native enzyme (5). Although these experiments focused primarily on proteolytic processing of ADAR1.1, these authors’ data also show that an ADAR1.2 construct engineered to encode a long form (1125 amino acid ORF) also gives rise to a band similar to, but slightly different from, ADAR1.1 protein bands produced by proteolytic cleavage or by translation from Met634. Thus, proteolytic processing of pre-cursor polypeptides generates polypeptides nearly indistinguishable from truncated proteins translated from the conserved, internal AUG codon.

Whether this pathway operates in mammalian cells remains to be determined. Assigning a particular protein product to a specific mRNA may be significantly more challenging in human cells, since the sequences downstream of Met296 are not conserved. Note that, although the long form of ADAR1 is predicted by abundant ADAR1 cDNAs, normal endogenous expression of the large protein in Xenopus has not yet been demonstrated. Possibly, this large, high molecular weight protein was isolated previously as a co-migrating, dsRNA-binding protein, in highly purified deaminase fractions from egg 5100 extract (5, 41). Given the differences in the primary sequences of the two X. laevis genes, which can be exploited for both mRNA- and protein-specific probes, experiments in X. laevis to resolve these ambiguities may prove more tractable.

In contrast to the single band observed on northern blots of human tissues, mRNAs of ~5.3, ~4.3 and ~3.8 kb are observed in poly(A+) RNA from X. laevis oocytes (15). Northern analyses using probes that distinguish the ADAR1.1 and 1.2 sequences show that each gene produces a long and short mRNA (ADAR1.1, ~5.3 and ~4.4 kb; ADAR1.2, ~5.3 and ~3.8 kb). The cDNAs characterized so far correspond to the long mRNA for ADAR1.1 (5.3 kb) and the short mRNA for ADAR1.2 (3.8 kb) and analyses of multiple cDNAs indicate that the additional bands observed on northern blots can be accounted for by 3’ UTRs of different lengths (Fig. 3; 15, 28). Northern analyses of rat tissues also show a three-band pattern of ~7.0, ~5.5 and ~4.4 kb mRNAs (32), suggesting the rat mRNAs may also be variably processed at their 3’ ends.

Although many cDNAs have been sequenced for X. laevis ADAR1.1 and 1.2 (15, 28), alternative splicing in the central region, like that observed for human mRNAs, has not been observed. Rather, the relationship of the dsRBMs to each other, and the start of the carboxy-terminal domain, correlates most closely with the human ADAR1b isoform.

2.2 ADAR1 dsRNA-binding motifs
dsRBMs were first identified in 1992 (42) and are now recognized as a motif shared by many dsRNA-binding proteins (dsRBPs) (43-44). While some motifs allow tight binding to dsRNA in the complete absence of other sequences, other motifs work best in the context of additional motifs. Along the same lines, when multiple motifs exist in a single protein, they are usually not equivalent. For example, studies of the dsRNA-activated kinase, PKR, show that mutations in motif I are more deleterious to binding than mutations in motif II (45).

Similarly, in vitro assays of ADAR1 enzymes containing point mutations within each of the three dsRBMs suggest that dsRBMs, the motif closest to the catalytic domain, is most important for activity (35-37). However, the latter mutations have not been evaluated for their effects on binding, and studies of proteins deleted for various dsRBMs suggest that proteins that lack activity may still bind well (46). Solution binding studies of individual dsRBMs from X. laevis ADAR1 suggest that dsRBMs is also most important for binding, but different results are obtained when binding is assayed by northwestern analyses (28). Assays of enzymes containing truncations or deletions that remove dsRBMs indicate that two or more motifs are required to modify perfect duplex RNA (15, 46, 47), while a protein containing only dsRBMs and the catalytic domain can edit the glutamate receptor-β (GluR-B) K+ site as efficiently as the full-length protein (47). (GluR-B pre-mRNA is a natural substrate for ADARs, as described in Chapter 6.)
2.3 ADAR1 amino-terminus: repeats and the Z-DNA-binding domain

The most distinctive feature of ADAR1 is its amino-terminal domain, which is much longer than that of other ADARs, particularly in the long form of the enzyme. In contrast to the sequence conservation observed for the dRNAs and carboxy-terminal catalytic domain (Fig. 2), the amino-termini of ADAR1 from different animal species are quite variable. Two outstanding features of the amino-termini of ADAR1 are not present in the amino-termini of ADAR2: a repeated sequence (11, 15) and a Z-DNA-binding domain consisting of two subdomains, Zv and Zb (27). Between the Zv and Zb subdomains is a variable region containing the conserved tripeptide (DNA; Fig. 3). As discussed in previous sections, the conserved methionine may function to initiate translation in certain cases. Interestingly, the Zv and Zb subdomains show significant similarities to amino-terminal residues of the vaccinia virus E3L protein (24; see section 6.3).

Despite earlier speculations, a conserved sequence in the amino-termini of ADAR1, with similarities to a bipartite nuclear localization signal (NLS, KxKxNxxLxxxGK; K11, 12, 15), is clearly not required for import since recombinant proteins lacking this sequence are found exclusively in nuclear extracts from human (24) and X. laevis (40). Rather, these conserved residues appear to be part of the α-helix of the Zv subdomain (48). Alanine mutational analyses of this region show that only the conserved asparagine and leucine (residues 173 and 176, respectively, of human ADAR1) and none of the conserved lysines are critical for Z-DNA binding. Whether a second conserved stretch of basic residues, KKRRxKxK (15), immediately downstream of Zb, is a nuclear targeting signal is not yet known.

The Z-DNA-binding domain (Zb) of ADAR1 contains a core subdomain of 77 amino acids, Zv, that is sufficient to bind left-handed Z-DNA (40), and a downstream secondary subdomain, Zb, related to Zv (27; Fig. 3). The NMR structure of Zv predicts an α/β topology (a1β1β2α2β3β3γαβ; 48) that has been confirmed by a co-crystal structure of Zv bound to Z-DNA (50). Zv binds as a dimer with high affinity and is capable of stabilizing a range of DNA sequences in the Z-DNA conformation (51).

From sequence alignments Zb is predicted to adopt a structure similar to that of Zv (27).

Z-DNA is thought to form transiently in the wake of transcribing polymerases, so it has been proposed that the ADAR1 Z-DNA-binding domain serves to target the enzyme close to the site of transcription (27). In cases where transcripts are involved in forming the dsRNA structure targeted by ADARs, this would ensure that editing had time to occur before splicing. Arguments against this proposal include the fact that ADAR2 does not have the Zp domain and that the constitutive, active, short forms of ADAR1 lack the Zp subdomain (see Figs 3 and 4). Of course, given the very different rates of splicing, it is easy to imagine that some ADARs would require localization to the site of transcription and others would not.

Herbert and colleagues also consider the possibility that the Z-DNA-binding domain may itself regulate transcription, since Zv belongs to the (α+β) helix-turn-helix family of DNA-binding proteins (27, 48). Along these same lines, an amino-terminal fragment of ADAR1 could function independently as a transcription factor if released by proteolysis. An interesting precedent for this speculation is the NF-κB1 gene, where two functionally distinct proteins are synthesized from the same gene by cotranslational proteolytic cleavage (52).

Regardless of whether the ADAR1 amino-termini bind to Z-DNA, it seems probable that this region of the protein plays a role in regulation (see section 6.3). Another possible function for the extended amino-terminus would be to promote nuclear-cytoplasmic shuttling or facilitate retention of ADAR1 within the cytoplasm. The latter would be consistent with the observation that the long form of human ADAR1, which accumulates after interferon induction, is observed in the cytoplasm as well as the nucleus (24).

3. ADAR2

ADAR2 was discovered during the search for the activity responsible for editing the Q/R site in GluR-B pre-mRNAs (see Chapter 6). The search began when it became clear that ADAR1 could not efficiently edit this site in vitro (83, 84), despite the fact that the site is edited very efficiently in vivo (99% of mRNAs); 7). dRNA clones were isolated by a low stringency screen of a rat hippocampal library with a cDNA probe to a portion of the rat ADAR1 carboxyl-terminus. Positive clones were then subcloned into expression vectors to assay for specific editing of the Q/R site (7). Similarly, by tracking editing at the Q/R site, human ADAR2 was resolved from ADAR1 and purified from HeLa cell nuclear extracts (55, 56).

So far, ADAR2 dRNAs have only been isolated from human and rat. Although enzymes with two dsRNAs have been detected in other organisms, it is not clear if they are ADAR2 homologs. The mammalian ADAR2 dRNAs encode a protein of 76-81 kDa, consisting of a short, nonconserved amino-terminus followed by two dsRNAs and a carboxyl-terminal catalytic domain that is ~50% identical to the carboxyl-terminal domain of mammalian ADAR1 (Fig. 2).

3.1 Expression of the ADAR2 gene

The human ADAR2 gene is located on chromosome 21q22.3 and is composed of 10 exons spanning ~25 kb (57, 58). Northern analyses show two bands for human ADAR2 (~8.6-8.8 and ~3.5-5.0 kb), which are broadly expressed, although reports differ considerably in regard to the levels observed in various tissues (57-60). Like some ADAR1 mRNAs, the major difference in size is probably due to differences in processing of the 3′ UTR (57). Northern analyses suggest that ADAR2 mRNAs are not transcribed in adult human liver or kidney (58, 59), but western analyses have detected ADAR2 protein in these tissues (59), suggesting that ADAR2 mRNAs may be inherently unstable in liver and kidney or in extracts from these organs. Conversely, although ADAR2 mRNAs are abundant in HEK293 cells, a human embryonic kidney
3.2 Autoregulation: creating a splice site by RNA editing

Although ADARs are suspected to have multiple functions in vivo, their only proven function, until recently, was to alter codons. However, ADAR2 was recently shown to act on its own pre-mRNA in rat brain, to create a new splice site (61). In this example, editing occurs on an AA dinucleotide to generate an AI dinucleotide which is recognized as AG and treated as a new 5' splice site. The alternative splicing results in a 47 nucleotide insertion 28 nucleotides downstream of the rat ADAR2 start codon (Met1); the insertion has been observed in the absence (ADAR2e) and presence (ADAR2o) of the 10 amino acid insertion found in rat ADAR2b (see Fig. 4). The 47 nucleotide insertion is predicted to result in a frameshift that would lead to the production of a short 88 amino acid polypeptide, lacking both dsRBMs and the carboxyl-terminal domain necessary for catalysis (Fig. 4). The short polypeptide has not been detected, and its sequence does not suggest a function.

As noted by Rueter et al. (61), an obvious interpretation of these data is that the alternative splicing is part of an autoregulatory loop designed to regulate the endogenous levels of ADAR2. When levels of ADAR2 were high, the enzyme would modify its own pre-mRNA, altering the splicing pattern so that the mature mRNA would encode a non-functional protein and result in lower levels of active enzyme. However, not all data are consistent with this interpretation. For example, despite the fact that the frameshift is predicted to destroy the functional ORF, some active ADAR2 is synthesized from the edited mRNA; this seems to result by translation from an AUG codon downstream of the frameshift (Met25; Fig. 4). In addition, while 80% of the mRNA in rat brain contains the 47 nucleotide insertion created by editing, only 5% of the protein appears to derive from this mRNA. Thus, 95% of the active ADAR2 is synthesized from the 20% of the mRNA that is not edited and thus spliced normally. At present it is not clear if editing of the ADAR2 mRNA can change the overall enzyme activity in a cell; transfection of isoforms with or without the 47 nucleotide insertion into HeK293 cells leads to similar levels of editing activity (61). Further, although the 47 nucleotide insertion appears in 70–80% of the mRNAs from rat brain and lungs, the proportion identified in other tissues is much lower, and that in human brain is lower still (~11%). In a large number of different tissue culture cells, no editing and alternative splicing of the 5' end can be detected.

Regardless of these caveats, it appears that, in rat brain and lungs, a primary substrate for ADAR2 is its own message (61), indicating that much of the mRNA previously detected in these tissues by northern analyses (7) is probably non-functional. Whether the system represents an important way in which the endogenous levels of active ADAR2 can be regulated remains to be proven. However, the mere observation that an ADAR can alter splicing patterns is significant progress in understanding the various roles of ADARs in vivo.
4. Other ADARs and ADAR-like proteins

New ADAR and ADAR-like sequences continue to appear in various databases; in the sections below we briefly describe some of these. Given the large amount of effort devoted to characterizing ADARs in mammalian cells, most of the ADAR-related sequences in mammals may have been found. However, as described below, two of these sequences, mouse Tenr (Fig. 2; 17) and RED2, are still mysterious in regard to function.

ADARs are now being characterized in more primitive metazoans, such as Drosophila melanogaster and C. elegans, and such studies will be helpful in determining whether enzymes with the same number of dsRBMs are true homologs and whether ADARs play similar roles in these more distantly related organisms. Below we also discuss the ADATs, which are particularly interesting for the insight they provide in regard to the evolution of the ADAR family.

4.1 Mouse testes nuclear RNA-binding protein (Tenr)

Expression of the Tenr protein is tightly regulated during spermatogenesis and is localized to the nuclei of round and early elongating spermmatids (17). Like T20H4.4, this protein contains a single dsRBM at the amino-terminus (Fig. 2). The function of Tenr is currently unknown.

4.2 Mammalian RNA editase 2 (RED2)

Shortly after the identification of ADAR2, a similar protein, RED2, was cloned (62) and mapped to a different chromosome (10p15; 63). mRNAs for RED-2 are found only in mammalian brain (62, 64). With the exception of a 54 amino acid aminoterminal extension rich in arginine residues, RED-2 is very similar to ADAR2, sharing 60% overall amino acid identity (62). However, despite extensive analyses of recombinant RED-2, the protein appears to be inactive using dsRNA as a substrate, as well as several characterized endogenous substrates. Perhaps RED-2 acts on different RNAs that remain to be discovered (62). Alternatively, RED-2 may modulate the activities of ADAR1 and ADAR2 by competing with the active enzymes for binding sites, as proposed for the ADARs alternatively spliced variant (59, 60; see section 3.1).

4.3 C. elegans H15N14.1a/b and T20H4.4

The completion of the C. elegans genome project revealed two ORFs as candidates for the worm ADARs: H15N14.1a/b and T20H4.4. H15N14.1a/b mRNAs have several alternatively spliced forms (L. Tenk, R. Hough and B. Bass, unpublished data) and all predicted ORFs encode proteins of ~105 kDa, with two dsRBMs and catalytic domains that are most similar to the mammalian proteins ADAR2 and Tenr.

The predicted ORF of T20H4.4 encodes a protein with a single dsRBM at the amino-terminus and a carboxyl-terminus that is ~40% identical to that of X. laevis and mammalian ADARs (Fig. 2). T20H4.4 is the second gene in a six-gene operon, downstream of a gene that encodes a 24 kDa subunit of mitochondrial NADH oxidoreductase complex I (16); the significance, if any, of clustering the six genes in the operon, is currently unclear.

A number of inosine-containing RNAs have recently been identified in C. elegans (65); these RNAs are presumably deaminated by proteins expressed from the H15N14.1a/b and T20H4.4 ORFs. Interestingly, in the C. elegans ADAR substrates, inosines are found in non-coding regions of mRNAs, predominantly 5’ and 3’ UTRs, suggesting new functions for ADARs (see Chapter 6).

4.4 tRNA-specific deaminases (ADATs)

It has long been known that tRNAs contain inosine at multiple positions (66), including the wobble position (1uG, 67), where the promiscuous base-pairing properties of inosine allow tRNAs to translate codons ending in U, C or A. In some cases these inosines are known to derive from modification of an encoded adenosine, so it seemed possible that the enzymes that catalyze these tRNA modifications might be related to the ADARs. Indeed, several years ago, during the sequencing of the Saccharomyces cerevisiae genome, an ORF with sequence similarity to the ADARs was identified in the carboxyl-terminal catalytic domain, but lacking any dsRBMs. The function of this ORF, identified as the 69 kDa protein encoded by the ttd1p ORF (68), has not been defined, and it is not clear whether it is related to the ADARs. However, its sequence similarity to the ADARs is striking.

Further clues as to the evolution of the ADAR family come with the recent identification of the ADAT responsible for deaminating A5′ in yeast, ttd1p, and the other tRNA-modifying enzymes.

The activity of the ADARs and ADATs is thought to be encoded by a single gene, with the activity responsible for deaminating A5′ in yeast, ttd1p, evolved from Tad2p (20) and subsequently acquired dsRBMs to create the metazoan ADAR family.
5. ADAR catalysis
As shown in Fig. 1, and described below (see section 5.1) the ADAR reaction occurs by a simple hydrolytic deamination mechanism. However, surprisingly little is known about the actual details of catalysis. Clearly the carboxyl-terminal domain contains the catalytic active site, as emphasized by the fact that this is the only domain common to all members of the ADAR/ADAT family; further, substantial amounts of the amino-terminus can be deleted with little consequence to activity (15, 47). Sequence similarities provide clues as to which amino acids in the carboxyl-terminal domain are involved in catalysis (see section 5.4), but these predictions have not yet been substantiated by biochemical or structural studies.

5.1 Reaction mechanism
When ADAR1 reactions are performed in 18O-labeled water, the heavy oxygen becomes incorporated at C6 of the inosine product (9). This suggests that water is the nucleophile in the deamination reaction, and that the reaction proceeds through a tetravalent intermediate (see Fig. 1). This mechanism is also used by adenosine and AMP deaminases and cytidine and CMP deaminases, enzymes that act on mononucleosides and nucleotides (see Fig. 1). Since the latter enzymes are known to utilize a zinc for catalysis, it seems likely that ADARs also have a catalytic zinc, although this has not yet been directly proven. The metal chelator 1,10-phenanthroline inhibits the activity of ADAR1 (6) and ADAR2 (33), but the inhibition may be due to interference with other aspects of ADAR catalysis. Consistent with this idea, nonchelating 1,7-phenanthroline and, to a lesser extent, 4,7-phenanthroline, inhibit ADAR1 and ADAR2 activity nonspecifically (R. Hough and B. Bass, unpublished data). ADAR2 activity was recovered by removal of 1,10-phenanthroline by gel filtration from preincubation mixtures before mixing with substrates, or by 20-fold dilution into the substrate reaction mixtures. Similar findings have been reported for reversible inhibition of the partially purified hRNA A34 deaminase, suggesting that the inhibition is due to intercalation into the substrate rather than irreversible inactivation of the enzyme active site (70). Further, active chicken ADAR1 has been purified in the presence of 2.5 mM 1,10-phenanthroline (14), and ADAR activity assays are routinely carried out in high concentrations of EDTA (5–50 mM). Nevertheless, it is possible that ADARs do use a zinc for catalysis, but that, as in marine ADA, it is tightly bound and difficult to chelate (71). Although studies of the ADAR reaction mechanism have been performed mainly with ADAR1, given the high conservation of the catalytic domain, all ADARs probably employ the same catalytic strategy.

5.2 Biochemical properties
If ADARs do require a catalytic zinc, the metal is presumably tightly bound since in vitro reactions proceed without the addition of any cofactors. An exception to this is the reaction of the ADAT, Tad2p, which requires magnesium (2.5 mM), probably to stabilize the conformation of its substrate, tRNA80 (18). Although not yet reported, the Tad2p/Tad3 reaction presumably requires magnesium for the same reason. ADARs are inhibited by thiol reagents, organomercurials and vanadyl ribonucleoside complexes (5, 6, 13, 26, 35).

The absolute specific activities of ADARs are not known; where estimates have been made for ADAR1, the assay conditions and forms of ADAR1 used were different. Initial experiments using native 92 kDa X. laevis ADAR1 (9) and native 116 kDa bovine thymus ADAR1 (13) reported similar estimates of approximately 2 nmmol Inosine min⁻¹ mg⁻¹ when assayed with a 794 bp dsRNA substrate in alkaline reaction media (pH 7.9) of slightly less than physiological ionic strength (0.125 M). In contrast, the purified bovine liver ADAR1 proteins (93, 88 and 83 kDa) were more active by an order of magnitude (22 nmmol Inosine min⁻¹ mg⁻¹) when assayed using a 575 bp dsRNA substrate and incubated at neutral pH with 0.25 M KCl (6). However, under these same 'high salt' reaction conditions, both native bovine thymus and X. laevis enzymes, and the 120 kDa ADAR1 isolated from bovine brain (26), are completely inhibited.

5.3 Substrate specificity
Like other dsRBPs, ADARs have very little sequence specificity and will bind to any sequence as long as it has sufficient double-stranded character. However, once bound, the enzymes deaminate certain adenosines more efficiently than others, although these preferences are not absolute. For example, ADAR1 has a preference for adenosines with a 5' A, U or C over those with a 5' G (22). ADAR2 has a similar 5' neighbor preference (A = U > C > G) but, unlike ADAR1, also has a 3' neighbor preference (U > C > G > A; 109).

Comparisons of the ability of ADAR1 and ADAR2 to deaminate the three best-characterized editing sites in GlnR-B RNA, emphasize that the enzymes have overlapping but distinct preferences (7, 47, 59, 60). All studies consistently show very efficient editing of the Q/R site by ADAR2 (~80% maximum) and very little editing of the same site by ADAR1 (~10% maximum). Further, all studies show that both enzymes efficiently edit the R/G site of GlnR-B. However, the actual amounts of editing observed for each site, and conclusions as to whether both enzymes, or only ADAR1, can edit the +6 hotspot, differ between laboratories. These differences probably result because different laboratories use different assay conditions, and different concentrations of recombinant enzymes that sometimes correspond to different isoforms from different organisms. In addition, although the GlnR pre-mRNA substrates used in these studies were the same or similar, all studies used very large fragments of the GlnR-B sequence. Given that such long RNAs would have the potential for many alternative conformations, observed differences may derive from the fact that the different conditions used by each laboratory favor different conformations.

The most important point to be drawn from the in vitro comparisons of ADAR1 and ADAR2 is that the two enzymes have distinct but overlapping specificities. In some cases an identical amount of editing can be achieved with either enzyme
5.5 The carboxy-terminal domain: evolutionary relationships

Importantly, the regions of the carboxy-terminal domain thought to be significant for catalysis correlate with conserved regions identified with the various search tools. For example, motifs 1-3 contain amino acids thought to be important for catalysis based on sequence similarities to CDAs (12, 46, 76; Fig. 5, residues in bold above consensus). Specifically, the ADAR sequences HAE1(xa)sPCG(xa)s15SCDK contain conserved residues (underlined) thought to be analogous to sequences in CDAs: C/HAE1(s)PCG(C77, 78). Based on studies of CDAs, the cysteine and histidine residues are proposed to coordinate zinc, while the glutamate is proposed to serve a proton transfer function. Indeed, mutagenesis experiments in which the putative catalytic residues (His910, Cys912, Cys966, Cys1006) in human ADAR1 were targeted demonstrated that each was required for enzymatic activity (46).

The recent cloning of two polypeptides, Tad2p and Tad3p, that act together to deaminate A34 of certain tRNAs provides the most compelling evidence that ADARs are related to CDAs and putative cytidine deaminases (55). In contrast to the ADARs, these new cytidine deaminases proposed to act in the cytoplasmic compartment are separated by only two amino acids, exactly the spacing observed in the cytidine deaminases. However, these proteins cannot deaminate cytidine or cytosine, emphasizing that, while they provide clues to the ancestry of the ADARs, they belong to the ADAR/ADAT family.

Interestingly, neither Tad2p nor Tad3p has deaminase activity on their own; furthermore, Tad3p lacks the glutamate residue of the HAE motif that is proposed to serve a proton-transferring function in the CDAs and adenosine deaminases. This residue is also missing in murine Ten and C. elegans H15N14.1/a. In addition, in H15N14.1/b the conserved histidine is swapped with the conserved aspartate in the DCDAE element (H15D1, Fig. 5). If either of these two deviant members of the ADAR family proves to be a deaminase, the strict CDA-like hypothesis will require revision. However, such results would not rule out the presence of a catalytic zinc ion, nor would findings of negative activity exclude the possibility that these two proteins are components of active, ADAR complexes.

In addition to sequence similarities with the deaminases, ADARs show similarities to conserved motifs in the N'As Mases (79); these similarities are within the most conserved regions of the ADARs (15; see Fig. 5, motifs (m) 1-3). The N'As Mases seem to be involved in the specific active sites by extracting them from the DNA helix, leaving the helical structure relatively unperturbed (reviewed in 80). Based on the sequence similarities, it has been proposed that ADARs may also use such a 'base-flipping' mechanism, which would solve the mystery of how the enzymes gain access to the site of hydrolytic deamination (adenine C6) which is buried deep within the major groove of the A-form helix (15, 72).

N'As Mases have two separate nucleotide-binding sites, defined primarily by motifs I-III and motifs IV-VIII, which accommodate the methyl-donating cofactor, 5-adenozymethionine (SAM) and the methyl-accepting, extrahelical deoxyadenosine, respectively. If ADARs adopt an α/β topology like the N'As Mases (81) and
6. Regulation of ADARs in vivo

Very little is known about how ADARs are regulated, and this will certainly be one of the next aspects of ADARs that will be investigated. The sections below note observations indicating that ADARs are regulated and provide clues as to avenues for future research.

6.1 ADAR1 complexes

Some evidence suggests that ADAR1 exists as a multimer or that it forms complexes with other proteins in vivo. For example, a high M, complex of X. laevis ADAR1 (200 ± 20 kDa, 9–10S), consistent with dimerization, was isolated from egg S100 extracts (3). Similarly, ADAR1 was found in ribonucleoprotein particles (INPs; ≤ 500 kDa, 1.35-1.42 g/ml) isolated from nuclear extracts prepared from bovine brain and human glioma HTB-14 and U373 cell lines (26). These RNP s can be disrupted by ribonuclease A treatment to yield more slowly sedimenting components similar in size to the >95 kDa complexes. For both X. laevis and bovine brain ADAR1, active monomers of ~120kDa are recovered from the high M, species by dissociation with monovalent salts (>0.3 M NaCl or KCl) and exchange by dsRNA affinity chromatography (5, 26).

A caveat in all protein purifications is that the proteins of interest can appear to be part of a complex merely because of adventitious associations that occur during the isolation procedure. This may be a particular problem with dsRBPs, since they are not sequence-specific and will bind to any RNA that has sufficient double-stranded character. Thus, a single dsRNA molecule can bind multiple dsRBPs which may have nothing to do with each other in vivo. Future studies aimed at determining if macromolecular complexes of ADARs exist in vivo will be essential for understanding the regulation of ADARs.

Interestingly, immunological studies show that X. laevis ADAR1 is genetically associated with transcriptionally active chromosomal loops and that it is particularly enriched on special RNA loops thought to be transcriptionally silent (40). Neither the 14 amino-terminal amino acid repeats of ADAR1.1 nor Xa (Fig. 3) are required for this chromosomal localization. The assembly is also independent of splicing formation, but deletion of a single dsRBM prevents chromosomal localization. Interestingly, another dsRBP has been shown to have a chromosomal localization. In particular, the D. melanogaster protein MLE, which is involved in dosage compensation, localizes to the X chromosome in an RNA-dependent manner (83).

6.2 Conditions that alter ADAR activity

One of the most significant observations in regard to how ADARs might be regulated is the demonstration that ADAR1 is induced by interferons. Specifically, interferon-α and -β increase the steady-state levels of ADAR1 mRNA approximately five- to 10-fold in human amniotic U cells and neuroblastoma SH-SYSY cells (24). This results in an increase in the proportion of the long, ~150 kDa form of ADAR1 compared with the more abundant, constitutively expressed ~110 kDa protein.

A number of caveats in regard to the above need to be resolved. For example, although the amount of human p150 ADAR1 protein in U and SH-SYSY cells increases after interferon-α treatment, a corresponding increase in ADAR activity is only observed in SH-SYSY cell extracts (24). Similarly, no apparent increase in ADAR activity is observed in extracts from several different cell lines after interferon stimulation (84). Further, activation of the interferon system in cultured HeLa cells by addition of exogenous dsRNA, or viral infection, previously correlated with loss of ADAR activity (85). Similarly, ADAR activity in cellular extracts decreased during interferon-β-induced differentiation of pluripotent P19 embryonal carcinoma cells (86).

There are other hints about how ADARs are regulated beyond their relationship to the interferon pathway, but at present their significance is unclear. For example, Hek293 cells have negligible endogenous ADAR activity compared with HeLa cells, but contain similar levels of ADAR1 and ADAR2 transcripts (35). For this reason, Hek293 cells are a favorite mammalian expression vehicle for recombinant ADAR proteins (47, 54, 60, 61). Other studies show that serum-induced release of growth-arrested 3T3 fibroblasts is accompanied by increased deaminase activity (84, 67). Finally, editing of Gub-R mRNAs increases during terminal differentiation of human neurons, following a significant lag period between transcription of ADAR1 and ADAR2 mRNAs and the appearance of edited receptor mRNAs (33).

6.3 Relationships with other dsRNA-binding proteins

The narrow major groove of the A-form dsRNA helix makes sequence-specific contacts difficult, so it is not surprising that dsRBPs lack sequence specificity. It follows that a substrate for any dsRBP has the potential to interact with all dsRBPs, raising the possibility that dsRBPs may be coordinately regulated. In fact, as discussed below, there is already evidence that there are similarities between the regulation of PKR, a well-characterized dsRBP, and that of the ADARs.

PKR is a serine/threonine kinase that contains two dsRBMs (88). dsRNA leads to
activation of the kinase and, once activated, PKR phosphorylates the e subunit of the translation initiation factor eIF2, which leads to inhibition of protein synthesis. PKR is thought to be activated during a variety of stress responses, including virus infection; in the latter case the activating dsRNA is thought to derive from viral sequences during replication.

One of the most obvious parallels between ADARs and PKR is that both enzymes are induced by interferon (89). Further, a hallmark of PKR is concentration-dependent activation and inhibition by dsRNA (90-94); such substrate inhibition has also been observed in vitro with native X. laevis ADAR1 (5). One model for how dsRNA leads to inhibition of PKR is based on the observation that activation of PKR by dsRNA requires PKR dimerization (95). The model assumes that activation requires binding of a PKR dimer to a single dsRNA molecule. Since each PKR monomer has dsRBMs, high concentrations of dsRNA result in binding of each monomer to a separate dsRNA molecule, precluding the formation of dimers and activation. The observation that ADARs are also subject to substrate inhibition suggests that these proteins may dimerize, but evidence for this is, as yet, only indirect (see section 6.1).

A number of viruses have evolved strategies to prevent activation of PKR during infection, including expression of RNAs that bind to PKR to occlude its dsRNA-binding site, and expression of dsRBPs that sequester the activating viral dsRNA. Studies of how the dsRBP E3L expressed by vaccinia virus eludes the activation of PKR provide further hints at how ADARs might interact with regulatory proteins. For many years, E3L was thought to prevent activation of PKR by sequestering dsRNA using its single dsRBM. Although this mechanism is still possible, recent results suggest that E3L can also interact directly with PKR using its dsRBM. The amphipathic α-helical regions within the dsRBMs of PKR contain both a dsRNA recognition face and a hydrophobic face required for dimerization (95). It is postulated that PKR homodimerization increases its affinity for dsRNA and leads to activation, while heterodimerization with E3L sequesters PKR by forming non-functional heterodimers. If the dsRBMs in ADARs also promote dimerization, various combinations of the enzymes with themselves, other active and inactive ADARs and other dsRBPs could lead to complex and diverse regulatory mechanisms.

The amino-terminal region of E3L, which does not contain the dsRBM, also contributes to antagonizing PKR by binding to its kinase domain (96). The latter is relevant since the regulatory domain shows significant sequence similarity to the ADAR1 Z-DNA subdomains, Zx and Zβ. In particular, the Zx and Zβ terminal regulatory domain is essential for interaction with the catalytic domain of PKR (97), and a tryptophan residue is conserved in both Zx and Zβ of all known ADAR1 proteins (27). Thus, the Zx and Zβ regions are good candidates for mediating interactions of ADAR1 with other proteins.

7. Future perspectives

RNA editing by adenosine deamination is now firmly established as another type of RNA processing which, like RNA splicing, lends greater diversity and complexity to gene expression in metazoans. The amounts of inosine detected within mammalian mRNA suggest that many ADAR substrates remain to be identified (32), and future studies will no doubt include efforts to identify these inosine-containing RNAs (65).

Many additional functions of ADARs have been proposed, and characterization of these new substrates will, it is hoped, offer clues about other functions that ADARs serve in vivo. ADARs can change the sequence (A→I) as well as the structure (AU→IU) of an RNA. Thus, ADARs have the ability to affect any process that involves sequence- or structure-specific interactions, and effects on RNA processing, stability and translatability are all possible.

As additional substrates and functions are identified, it will become more imperative to understand how ADARs are regulated. As apparent from the discussion in section 6.1, at present it is hard to put the observations made so far into a simple and clear model. This is not surprising, given the pleiotropic, and often opposing, effects exerted by dsRNA, interferons and other cytokines in different cell lines and biochemical contexts (89, 98, 99; see section 6.3). Future studies should clarify these issues.

Finally, the next phase in ADAR/ADAT research will undoubtedly offer insight into the catalytic mechanism used by the enzyme family. The predominant reason for
the lack of detailed biochemical and structural studies since ADARs were discovered is that the enzymes are difficult to overexpress in large quantities. Fortunately, several laboratories have now overcome these difficulties, setting the stage for many long-anticipated characterizations.

Acknowledgments

We are grateful to Danny Morse and Herbert L. Levy III for critically reading the manuscript, and to the entire Bass laboratory for helpful discussions and sharing of unpublished data. We thank Jana Brubaker for helping to format and edit this manuscript.

References


32. Paul, M. S. and Biss, B. L. (1998) Xrn1iexonin is mRNA at tissue-specific levels and is most abundant in brain mRNA. EMBO J., 17, 1120.


34. George, C. X. and Samuel, C. E. (1999) Human RNA-specific adenosine deaminase ADAR1 transcripts possess alternative exon 1 structures that initiate from different promoters, one constitutively active and the other interferon inducible. Proc Natl Acad Sci USA, 96, 18451-18455.


6. ADENOSINE DEAMINASES THAT ACT ON RNA


REFERENCES: 1-107


109. Lehmann, K. A. and Bass, B. L. Double-stranded RNA adenosine deaminases ADAR1 and ADAR2 have overlapping specificities. Biochemistry, in press.

6 | Adenosine-to-inosine RNA editing: substrates and consequences

RONALD B. EMESON AND MINATI SINGH

1. Introduction

The conversion of adenosine to inosine (A→I) by RNA editing is an increasingly recognized RNA processing event by which multiple RNA and protein isoforms can be generated from a single genomic locus. This type of modification was first observed in yeast tRNAs (1, 2), but has since been identified in multiple viral RNA transcripts and cellular mRNA species (3–5). Since inosine preferentially base-pairs with cytidine, an inosine within RNA transcripts is read as guanosine during translation, often resulting in specific change(s) in the amino acid coding potential of the mRNA, which can dramatically alter the functional properties of the encoded protein product. In several instances, A→I editing events have also been described in untranslated RNA species (e.g., rRNAs) (6, 7) and non-coding regions of mRNA transcripts (8–11), suggesting that such modifications may also affect other aspects of RNA function including translation, splicing and stability.

A→I editing is most often identified as an adenosine to guanosine (A→G) discrepancy between genomic and cDNA sequences. Due to the similar base-pairing properties of inosine and guanosine, the edited inosine moiety directs the incorporation of cytidine by reverse transcriptase during cDNA synthesis, thereby appearing as a guanosine in double-stranded cDNA or reverse transcription–polymerase chain reaction (RT–PCR) products. A→I conversion is catalyzed by hydrolytic deamination at the C6 position of the purine ring. Studies showing that A→I editing in mammalian mRNAs and yeast tRNAs requires the presence of higher-order RNA structures (6, 10, 12) suggested that adenosine deaminases(s) directed toward double-stranded RNA (dsRNA), or specifically toward tRNA structures, could catalyze such post-transcriptional modifications. The recent purification, cloning and characterization of a family of these adenosine deaminases (13–18), referred to as ADARs (adenosine deaminases that act on RNA) (19) or ADATs (adenosine deaminases that act on RNA) (6, 7, 20), has indicated that such proteins can selectively modify