Current Topics in Microbiology and Immunology

Bryan R. Cullen *Editor*

Intrinsic Immunity

Springer

Current Topics in Microbiology and Immunology

Volume 371

For further volumes:
http://www.springer.com/series/82

Bryan R. Cullen
Editor

Intrinsic Immunity

Responsible Series Editor: Peter K. Vogt

Springer

Editor
Prof. Dr. Bryan R. Cullen
Department of Molecular Genetics
and Microbiology and
Center for Virology
Duke University Medical Center
Durham, NC
USA

ISSN 0070-217X
ISBN 978-3-642-37764-8 ISBN 978-3-642-37765-5 (eBook)
DOI 10.1007/978-3-642-37765-5
Springer Heidelberg New York Dordrecht London

Library of Congress Control Number: 2013938747

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

The immune responses that can limit or prevent disease induction by viruses are historically divided into innate and adaptive immune responses. Adaptive immunity refers to the selection and rapid expansion of T cell and B cell clones that have rearranged their T cell receptor and antibody genes, respectively, in ways that allow them to effectively recognize and neutralize most viruses or virus-infected cells. In contrast, innate immunity refers to cellular receptors that recognize Pathogen-Associated Molecular Patterns (PAMPs) and then activate signaling pathways leading to the production of interferons. This in turn induces the expression of a large number of Interferon-Stimulated Genes (ISGs), in both the infected cell and adjacent, as yet uninfected cells, many of which are able to alter the cellular environment in ways that inhibit viral replication. Cellular factors involved in PAMP recognition, so-called Pattern Recognition Receptors (PRRs) include the Toll-Like Receptors (TLRs) as well as cytoplasmic PRRs such as the Retinoic acid Inducible Gene-I (RIG-I) protein family, as well as the NOD-like receptors.

In addition to innate and adaptive immune responses, both of which are only activated after pathogen infection, many cells also constitutively express antiviral factors that can act as potent inhibitors of viral replication. These factors are now generally grouped together as mediators of intrinsic antiviral immunity. While several intrinsic immune factors, including APOBEC3G, TRIM5α, Tetherin, and SAMHD1, were initially discovered by researchers studying the replication of retroviruses, particularly HIV-1, it is now clear that some of these proteins, especially APOBEC3G and Tetherin, are in fact capable of inhibiting a wide range of viral species. In addition, it is becoming increasingly apparent that cells also express intrinsic factors that can limit the replication of other, non-retroviral species, including for example the inhibition of large DNA viruses by DNA damage response proteins. Finally, especially in invertebrate animals and plants, RNA-mediated intrinsic immunity can also play a key role in limiting viral replication and pathogenesis and even in vertebrates, microRNAs can play a major role in either restricting or, in some cases, facilitating viral replication. Because intrinsic immunity is, at least over short time periods, a set of fixed host antiviral mechanisms, it is not surprising that viruses have evolved numerous mechanisms to overcome both protein and RNA-mediated intrinsic immunity in their normal host

species. However, as discussed in the chapter by W. E. Johnson, cellular intrinsic immune factors are also capable of evolution in ways that can circumvent viral countermeasures and this "rapid adversarial co-evolution" is clearly important in defining the host range of viruses.

This volume brings together nine papers reviewing different aspects of antiviral intrinsic immunity from scientists who have made major contributions to this area of research. I believe this field is an important one from several perspectives, including not only the potential design of antiviral drugs but also achieving a better understanding of the coevolution of viral pathogens and their hosts. I hope the reader will find these contributions as interesting as I did.

Bryan R. Cullen

Contents

The APOBEC3 Family of Retroelement Restriction Factors

Eric W. Refsland and Reuben S. Harris

Abstract The ability to regulate and even target mutagenesis is an extremely valuable cellular asset. Enzyme-catalyzed DNA cytosine deamination is a molecular strategy employed by vertebrates to promote antibody diversity and defend against foreign nucleic acids. Ten years ago, a family of cellular enzymes was first described with several proving capable of deaminating DNA and inhibiting HIV-1 replication. Ensuing studies on the apolipoprotein B mRNA-editing enzyme catalytic polypeptide-like 3 (APOBEC3) restriction factors have uncovered a broad-spectrum innate defense network that suppresses the replication of numerous endogenous and exogenous DNA-based parasites. Although many viruses possess equally elaborate counter-defense mechanisms, the APOBEC3 enzymes offer a tantalizing possibility of leveraging innate immunity to fend off viral infection. Here, we focus on mechanisms of retroelement restriction by the APOBEC3 family of restriction enzymes, and we consider the therapeutic benefits, as well as the possible pathological consequences, of arming cells with active DNA deaminases.

Contents

E. W. Refsland · R. S. Harris (✉)
Department of Biochemistry, Molecular Biology, and Biophysics, Institute for Molecular Virology, Center for Genome Engineering, and Masonic Cancer Center,

University of Minnesota, 321 Church Street S.E., 6-155 Jackson Hall, Minneapolis, MN 55455, USA
e-mail: rsh@umn.edu

B. R. Cullen (ed.), *Intrinsic Immunity*, Current Topics in Microbiology and Immunology 371, DOI: 10.1007/978-3-642-37765-5_1,

1 DNA Deaminase Evolution

Central to nucleic acid metabolism is the near-ubiquitous process of enzymatic deamination of adenine and cytosine bases, individually or in the context of larger nucleic acid constituents (Conticello et al. 2005; Grosjean 2009). For instance, in most species, the wobble base in several of the tRNA anti-codons is frequently changed by deamination of adenosine to inosine (A-to-I), which can then base pair with cytosine and thereby increase the flexibility and decoding capacity of the tRNA anti-codon (Gerber and Keller 1999). These enzymes belong to the adenosine deaminase acting on tRNA (ADAT) family. Related proteins in most metazoans from nematodes and flies to humans catalyze A-to-I editing of a variety of RNA targets (Kim et al. 1994; Nishikura 2010). These enzymes are called, appropriately, adenosine deaminases acting on RNA (ADAR). Editing events that occur in the coding region of an mRNA can result in amino acid substitutions in the resulting protein. However, the majority of editing events occur in non-coding regions of mRNA or in non-coding RNAs, and these A-to-I editing events can alter RNA secondary structure, stability, function, and/or capacity to be bound by regulatory RNAs such as siRNAs (Morse et al. 2002; Levanon et al. 2004; Agranat et al. 2008; Li et al. 2009).

Cytosine to uracil (C-to-U) deamination is almost as ancient as A-to-I editing (Conticello et al. 2007b; Grosjean 2009). The pyrimidine salvage pathways of most organisms use cytidine deaminase (CDA) to produce the essential RNA and DNA building blocks of uridine directly and thymidine after additional enzymatic steps (Zrenner et al. 2006). However, at some point near the root of the vertebrate tree, polynucleotide cytosine deaminases emerged, with the lamprey CDA being a present-day example (Rogozin et al. 2007). This enzyme is thought to underpin a unique form of adaptive immunity in which the DNA segments that encode arrays of highly diverse leucine-rich repeats are assembled into mature variable lymphocyte receptor genes by a recombination-mediated process. It is thought that an ancestor of the present-day lamprey enzyme served as the original substrate for expansion of the polynucleotide cytosine deaminase gene family during vertebrate evolution (Fig. 1a) (Rogozin et al. 2007; Conticello 2008). The result in most vertebrates alive today is a much larger repertoire of polynucleotide C-to-U editing enzymes that execute diverse biological functions from lipid metabolism to adaptive and innate immunity (Figs. 1 and 2).

All vertebrate polynucleotide cytosine deaminases belong to the so-called ‘APOBEC’ family. The defining feature of this family is a conserved His-X-Glu-$X_{25–31}$-Pro-Cys-$X_{2–4}$-Cys zinc (Z)-coordinating motif, which is strictly required for deaminase activity (where X can be a variety of amino acids) (Wedekind et al. 2003; Harris and Liddament 2004; Conticello et al. 2005; LaRue et al. 2009). As described in more detail below, key residues within this motif position zinc at the active site of the enzyme (Fig. 1c). The protein sequences within these motifs enable phylogenetic groupings into three subfamilies: APOBEC1, AID, and the APOBEC3s. The APOBEC3s can be further subdivided into three subgroups: Z1, Z2, and Z3 (Fig. 1a) (Conticello 2008; LaRue et al. 2009). Importantly, the number and organization of the A3 Z-domains can vary dramatically from branch

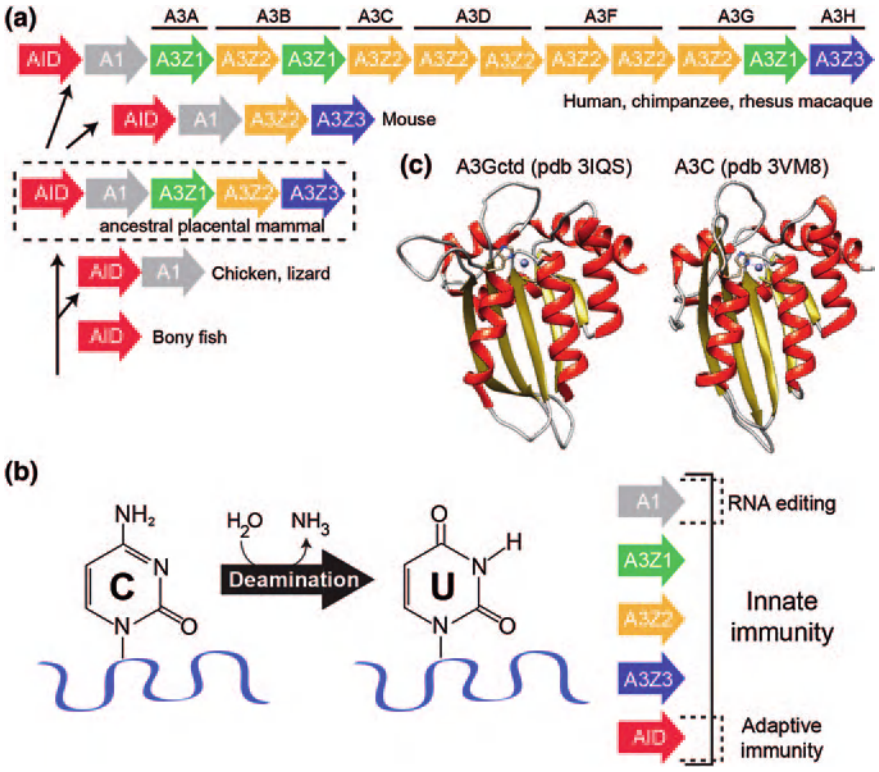

Fig. 1 Evolution and structure–function of APOBEC cytosine deaminases. **a** Expansion of the modern primate APOBEC3 locus encoding seven APOBEC3 genes with eleven zinc-coordinating (Z) domains (reprinted with permission from Lackey et al. 2012). **b** Deamination of C-to-U plays a central role in innate immunity. **c** Three-dimensional structures of the APOBEC3G C-terminal domain and APOBEC3C. A zinc ion (purple) is shown coordinated in both proteins by one histidine and two cysteine residues in the α2-β3-α3 core

to branch throughout the mammalian portion of the vertebrate phylogenetic tree (e.g., human versus mouse loci depicted in Fig. 1a).

Apolipoprotein B mRNA-editing catalytic subunit 1 (APOBEC1) has provided the namesake to the larger family. It was discovered as an enzyme that catalyzes the deamination of a specific cytosine within the APOB mRNA (Figs. 1b and 2a) (Teng et al. 1993). This produces a premature translation stop codon and a smaller secondary gene product. These two APOB proteins (APOB100 and APOB48) differentially regulate the secretion of lipoproteins from the liver (Chan 1992). Many mammalian APOBEC1 enzymes also possess DNA C-to-U deaminase activity (Harris et al. 2002; Ikeda et al. 2008; Petit et al. 2009; Ikeda et al. 2011). Taken together with the fact that earlier vertebrate lineages, such as the one represented by birds and lizards, lack an APOB-like gene, it is probable that the DNA-editing function preceded involvement in RNA editing (Severi et al. 2011).

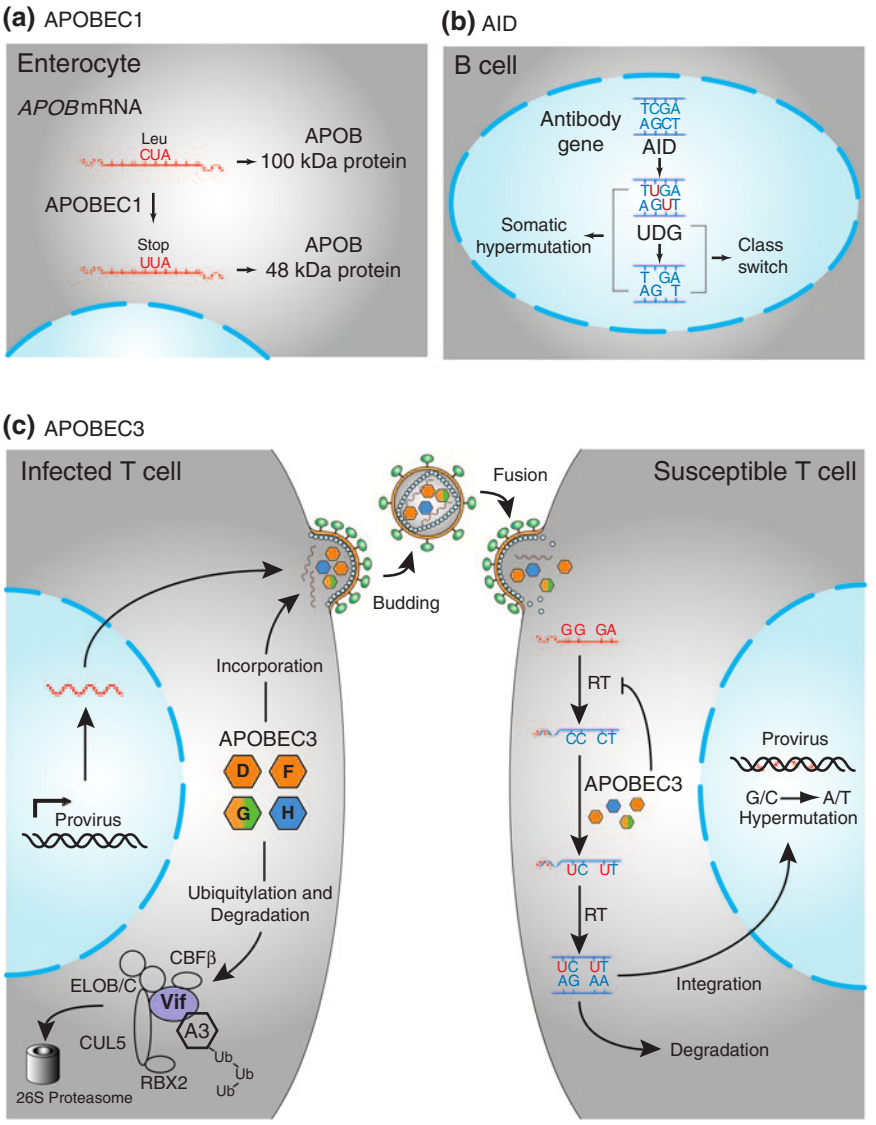

Fig. 2 The physiological functions of the APOBEC family. **a** RNA editing by APOBEC1 generates a truncated APOB protein. **b** AID activity underpins the processes of somatic hypermutation and class switch recombination in B cell germinal centers. **c** Four APOBEC3 proteins restrict HIV through cytosine deamination in the absence of Vif (reprinted with permission from Hultquist et al. 2011)

The most conserved DNA cytosine deaminase in vertebrates is activation-induced deaminase (AID; gene name *AICDA*) (Fig. 1a). AID has a central role in adaptive immunity by seeding somatic hypermutation, gene conversion, and

class switch recombination with its DNA deaminase activity (Figs. 1b and 2b) [(Muramatsu et al. 1999, 2000; Di Noia and Neuberger 2002; Petersen-Mahrt et al. 2002); reviewed by (Longerich et al. 2006; Di Noia and Neuberger 2007; Conticello 2008)]. Interestingly, the genes that encode APOBEC1 and AID are positioned adjacent to each other in the genomes of most vertebrates (an inversion has placed the human gene farther away on the same chromosome). This suggests that an ancestral AID gene duplicated and diverged to produce APOBEC1 (Fig. 1a). It is likely that duplication of an ancestral AID/APOBEC1 locus produced the genetic seeds for the mammal-exclusive APOBEC3 subfamily (Fig. 1a) (Jarmuz et al. 2002; Harris and Liddament 2004).

In humans, the seven APOBEC3 proteins are encoded by a tandemly arranged gene cluster (Fig. 1a) (Jarmuz et al. 2002). These present-day genes are the products of continual evolution, in which an ancestral cluster of three Z-domains is predicted to have undergone a minimum of eight duplication events over the past 100 million years to produce the locus found in most primates (LaRue et al. 2008; Münk et al. 2012). These domains are either expressed singly or one enzyme may consist of two Z-domains (LaRue et al. 2009). In contrast, the ancestral APOBEC3 locus experienced a deletion in the rodent lineage of one of the ancestral Z-domains, leading to the present-day two domain loci, which encodes a single protein quite distinct from any of the primate enzymes (Fig. 1a).

One possible explanation for why some mammalian lineages, like primates, have many APOBEC3s, while other lineages, such as rodents, have few is that these enzymes have overlapping innate immune functions to protect the host from a variety of parasitic elements (e.g., in HIV-1 restriction, Fig. 2c; mechanism elaborated in Sect. 3, below). Because multiple distinct innate immune mechanisms serve to suppress the spread of such parasitic elements, it is reasonable to postulate that some mammals will be fortified at the APOBEC3 locus and weaker at other loci, with each mammalian lineage being distinct. For instance, primates encode a single TRIM5α protein, whereas mice have the capacity to encode a total of eight TRIM5α-like proteins (Sawyer et al. 2007; Tareen et al. 2009; Chap. 13 in Lever et al. 2010). It is likely that each species' present-day innate immune fortifications were independently shaped by past pathogenic pressures, which one can only speculate may have been the ancestors of present-day viruses and transposable elements.

2 Biochemical and Structural Insights

Zinc-dependent deaminases, such as the APOBECs, catalyze the conversion of C-to-U in polynucleotide substrates (Fig. 1b). This reaction requires the activation of water by a zinc ion coordinated by the enzyme (Fig. 1c). A glutamic acid in the active site of the enzyme protonates N3, priming the nucleophilic attack on the C4 position of the pyrimidine ring, followed by the removal and subsequent protonation of an amino group (NH_2) that results in the release of ammonia (NH_3) and

uracil as products. This conversion can theoretically occur within both RNA and single-stranded DNA substrates. However, apart from APOBEC1, which has both RNA and DNA-editing activities, AID and the APOBEC3s have proven specific to DNA substrates in vitro and in vivo.

The extent of amino acid homology to APOBEC1 originally suggested that the APOBEC3 enzymes might be a family of RNA-editing proteins (Jarmuz et al. 2002). Three lines of evidence, however, demonstrated that this view was incorrect and established the APOBEC3 enzymes as single-stranded DNA cytosine deaminases. First, APOBEC3 has a high degree of homology to AID, and experiments in *E. coli* demonstrated AID, APOBEC3C, and APOBEC3G are capable of inducing high levels of mutation in an antibiotic resistance gene (Harris et al. 2002; Petersen-Mahrt et al. 2002). This was clearly due to DNA editing because mutation levels rose synergistically in a bacterial strain deficient for uracil DNA glycosylase (UDG), an enzyme that initiates base excision repair by recognizing and removing uracil exclusively from DNA (Lindahl 2000; Di Noia and Neuberger 2002; Harris et al. 2002). Second, unambiguous evidence for DNA versus RNA editing comes from head-to-head biochemical studies using recombinant enzymes. AID and APOBEC3G have a strong preference for single-stranded DNA substrates, with no detectable RNA-editing activity (Bransteitter et al. 2003; Iwatani et al. 2006). Third, a strong preference for single-stranded DNA substrates is also evident in sequencing studies of retroviruses produced in the presence of a given APOBEC3 protein, such as APOBEC3G (Fig. 2c) (Harris et al. 2003; Lecossier et al. 2003; Mangeat et al. 2003; Zhang et al. 2003). In this experimental system, each APOBEC3 protein presumably has a chance to deaminate viral genomic RNA cytosines before the reverse transcription process converts it to a single-stranded cDNA intermediate and then to the double-stranded DNA required for integration. However, the most common APOBEC3-dependent mutations detected in integrated viral DNA that has survived this process are genomic strand G-to-A mutations, entirely attributable to cDNA minus strand C-to-U deamination events. Genomic strand C-to-T editing events possibly due to RNA editing are rarely detected. Importantly, APOBEC3 DNA-editing activity is required to explain previously reported G-to-A mutation biases in HIV-1 substrates in vivo (Vartanian et al. 1994; Janini et al. 2001).

The solved structures of bacterial and yeast cytidine and cytosine deaminases were used to inform early functional and structural studies of various APOBEC3 family members (Betts et al. 1994; Ireton et al. 2003; Ko et al. 2003; Johansson et al. 2004; Xie et al. 2004). Each of these bacterial and yeast proteins, in monomeric form, is globular with a hydrophobic β-stranded core and several surrounding α-helices. The most conserved structural feature is the active site, which is defined by a histidine and two cysteines in the yeast enzyme and three cysteines in the bacterial enzymes (Xiang et al. 1997; Ireton et al. 2003; Ko et al. 2003). In both instances, these residues are positioned similarly by alpha helices and they serve to coordinate a zinc ion in the active site, which, as described above, is essential for the deamination reaction (Fig. 1b). Although these conserved features have been useful for generating models of APOBEC3 structures, they have

also been misleading because the oligomeric state of each enzyme is variable. For instance, the *E. coli* CDA is homodimeric and the yeast enzyme is homotetrameric (Betts et al. 1994; Johansson et al. 2002). This has fuelled (likely incorrect) speculation that APOBEC3 family members must also function as oligomers.

Generating high-resolution structures of APOBEC3 family members has proved challenging in large part due to insolubility at higher protein concentrations [e.g., (Iwatani et al. 2006)]. However, several NMR and crystal structures have been achieved for the APOBEC3G catalytic domain (representing Z1-type deaminases), and crystal structures were obtained recently for APOBEC3C and the APOBEC3F catalytic domains (representing Z2-type deaminases) (e.g., Fig. 1c) (Chen et al. 2008; Holden et al. 2008; Furukawa et al. 2009; Shandilya et al. 2010; Kitamura et al. 2012b; Li et al. 2012; Bohn et al. 2013).

These structures have several conserved features that provide insight into how these enzymes may function. First, these proteins are all globular with a hydrophobic core consisting of five beta strands surrounded by six alpha helices and the hallmark α2-β3-α3 zinc-coordinating motif that defines the larger cytosine deaminase superfamily. Second, β-strands 3, 4, and 5 are arranged in parallel, similar to the RNA-editing enzyme TadA (an ADAT) but different from the antiparallel arrangement found in bacterial and yeast CDAs. This parallel β3-β4-β5 organization may be a key feature that distinguishes polynucleotide from non-polynucleotide deaminases. Third, although many potential oligomeric interfaces have been captured in the crystal lattices, none have proven critical for enzymatic activity and no common themes have emerged (Furukawa et al. 2009; Shandilya et al. 2010; Kitamura et al. 2012b; Bohn et al. 2013). This is consistent with a number of other studies, indicating that oligomerization may not be essential for binding and deaminating single-stranded DNA substrates (Opi et al. 2006; Nowarski et al. 2008; Shlyakhtenko et al. 2011, 2012). However, more work on this topic is clearly needed to define the role of oligomerization in vivo, because several of the family members, including APOBEC3G, elicit such a property in living cells (Bransteitter et al. 2003; Chiu et al. 2006; Soros et al. 2007; Chen et al. 2013 in preparation). Finally, it is notable that the majority of structural and amino acid differences between APOBEC3 structures are confined to non-catalytic loop regions. Such differences likely relate to substrate targeting and possible cofactor binding, ultimately reflecting physiological function.

A significant remaining question in our understanding of APOBEC3 function is how these enzymes bind single-stranded DNA substrates. A current working model proposes a positively charged brim in the region surrounding the active site consisting of R213, R215, R313, and R320 in APOBEC3G (Chen et al. 2008; Shindo et al. 2012). These residues are predicted to position single-stranded DNA substrates in a manner that allows the target cytosine to enter the active site (Chen et al. 2008). This model also predicts that in order to access to the catalytic glutamic acid, the target C will be flipped out with respect to the phosphodiester backbone. A base-flipping mechanism is in good agreement with the structure of the adenosine deaminase TadA complexed with its RNA substrate (Losey et al. 2006).

Finally, the brim-domain model and TadA structures suggest an explanation for the different local single-stranded DNA deamination preferences among APOBEC family members (Conticello et al. 2007a; Chen et al. 2008). Unlike bacterial restriction enzymes with 4, 6, or 8 base palindromic recognition sequences, APOBEC3 family members have a notable preference for the base immediately 5′ of the target C (Harris et al. 2002, 2003; Mangeat et al. 2003; Zhang et al. 2003; Bishop et al. 2004; Liddament et al. 2004; Wiegand et al. 2004; Yu et al. 2004a, b; Zheng et al. 2004; Doehle et al. 2005a; Langlois et al. 2005; Dang et al. 2006; Aguiar et al. 2008; Harari et al. 2009; Stenglein et al. 2010). Specifically, AID prefers a 5′ purine base (5′-AC or GC), APOBEC3G a 5′ cytosine (5′-CC), and all other family members a 5′ thymine (5′-TC). Several studies have recently mapped this activity to a loop adjacent to the active site, positioned between β4 and α4 secondary structural elements (Conticello et al. 2007b; Chen et al. 2008; Holden et al. 2008; Kohli et al. 2009, 2010; Rathore et al. 2013 *in preparation*). This is most dramatically evidenced by loop grafting experiments, in which this loop in AID can be replaced by the homologous loop from APOBEC3G or APOBEC3F resulting in a complete switch of the preferred base immediately 5′ of the target cytosine (Kohli et al. 2009, 2010; Carpenter et al. 2010; Wang et al. 2010). Moreover, exchanging the same loop (or even a single amino acid) between APOBEC3A and APOBEC3G completely swaps the dinucleotide preference of these enzymes (Rathore et al. 2013 in preparation). Despite this progress, the field still anxiously awaits high-resolution structures of enzyme–substrate complexes that will more precisely define the substrate binding mechanism and advance our understanding of how these enzymes function in vivo.

3 Biological Functions

3.1 APOBEC3 Proteins in HIV Restriction

Pathogens including the retrovirus HIV-1 (hereafter HIV) must both engage and avoid numerous host factors to replicate and cause disease. Genome-wide knockdown and proteomic studies suggest that up to 10 % of human proteins either directly or indirectly impact HIV replication (Brass et al. 2008; Konig et al. 2008; Zhou et al. 2008; Yeung et al. 2009; Jäger et al. 2012a, b). The majority of these proteins are required in some capacity for virus replication (i.e., dependency factors). In contrast, a small number of these cellular proteins are dominant proteins that directly suppress virus replication (i.e., restriction factors). Restriction factor hallmarks include the capacity to potently inhibit virus replication, signatures of rapid evolution (positive selection), responsiveness to interferon, and neutralization by at least one viral counter-restriction strategy (Malim and Emerman 2008; Harris et al. 2012; Malim and Bieniasz 2012). Here, we focus on the mechanism of HIV restriction by APOBEC3 DNA cytosine deaminases, and we encourage readers to see chapters on equally interesting restriction and counter-restriction mechanisms