

DISSECTING APOBEC3G SUBSTRATE SPECIFICITY BY
NUCLEOSIDE ANALOG INTERFERENCE

Jason W. Rausch¹, Linda Chelico², Myron F. Goodman² and Stuart F. J. Le Grice¹

Retroviral Replication Laboratory, HIV Drug Resistance Program, National Cancer Institute at Frederick,
Frederick, Maryland¹

Department of Biological Sciences, Molecular and Computational Section, University of Southern
California, Los Angeles, California²

Running head: APOBEC3G substrate specificity

Address correspondence to: Stuart F. J. Le Grice, Ph.D., 1050 Boyles Street, Frederick, Maryland 21702.
Tel: 301-846-5256; Fax 301-846-6013; E-mail: slegrice@ncifcrf.gov

The apolipoprotein B mRNA-editing enzyme, catalytic polypeptide-like (APOBEC) cytidine deaminase genes encode a set of enzymes including APOBEC1 (A1), APOBEC2 (A2), APOBEC4 (A4) and APOBEC3A-H (A3A-H). Although each possesses one or more zinc binding motifs conserved among enzymes catalyzing C→U conversion, the functions and substrate specificities of these gene products vary considerably. For example, although two closely related enzymes, A3F and A3G, both restrict HIV-1 infection in strains deficient in virus infectivity factor (*vif*), A3F selectively deaminates cytosine within 5'-TTCA-3' motifs in single stranded DNA, while A3G targets 5'-CCCA-3' sequences. In the present study, we have used nucleoside analog interference mapping to probe A3G-DNA interactions throughout the enzyme-substrate complex, as well as to determine which DNA structural features determine substrate specificity. Our results indicate that multiple components of nucleosides within the consensus sequence are important for substrate recognition by A3G (with base moieties being most critical), whereas deamination interference by analog substitution outside this region is minimal. Furthermore, exocyclic groups in pyrimidines 1-2 nts 5' of the target cytosine were shown to dictate substrate recognition by A3G, with chemical composition at ring positions 3 and 4 found to be more important than at ring position 5. Taken together, these results provide insights into how the enzyme selects A3G hotspot motifs for deamination, as well as which approaches might be best suited for forming a stable, catalytically competent cross-linked A3G-DNA complex for future structural studies.

APOBEC3G (A3G) is a cellular polynucleotide cytosine deaminase that restricts HIV-1 infection in strains deficient in virus infectivity factor (*vif*) (1), a viral accessory protein that targets A3G for ubiquitination and proteosomal degradation (2,3). Expressed in numerous cell types including T-lymphocytes and macrophages, A3G is packaged into Δvif HIV-1 particles without affecting assembly, release, binding or fusion to new target cells (4-8). Reverse transcription, however, is impaired (9-12), and infection typically aborts prior to integration (13-15). As a consequence, infectivity of Δvif HIV-1 assembled in the presence of A3G is reduced by as much as 1000-fold (8).

Sequencing of DNA derived from these viruses reveals numerous G→A mutations throughout the coding strand (4,6,11,16-19). These transitions are not likely to be generated directly, but occur as cytosine deamination products are copied during the course of reverse transcription. A3G selectively deaminates the 3'C within 5'-CCCA-3' "hot spots" in single stranded DNA, which, *in vivo*, only become available following RNase H-mediated hydrolysis of the viral genome during and after minus-strand DNA synthesis (4,6,11,16-19). Neither RNA nor dsDNA are deaminated (20,21). Over the length of the genome, mutation frequency decreases in the 5'→3' direction from maxima immediately 5' to the two plus-strand primers (19). This pattern reflects (a) a direct relationship between the extent of deamination within a particular segment of minus-strand DNA and the time it remains single-stranded during the course of reverse transcription, and (b) the inherent 3'→5' directional bias of A3G-mediated deamination (11,18-20,22).

A3G is a member of the mammalian APOBEC cytidine deaminase family, which in humans includes APOBEC1 (A1), APOBEC2 (A2),

APOBEC4 (A4), APOBEC3A-H (A3A-H), and AID (23,24). These enzymes cover a diverse range of functions and substrate specificities. For example, A1 is an RNA editing enzyme, while the function and preferred substrate of A2 are unknown. AID, like A3G, catalyzes C→U deamination on ssDNA, but is involved in immunoglobulin somatic hypermutation and principally targets 5'-WRC-3' motifs (W = A or T; R = A or G). Specificities also differ among the A3 enzymes. A3F, an enzyme that has also been shown to possess antiviral activity against HIV-1, principally deaminates ssDNA in the context of the 5'-TTCA-3' sequence element (25-28), while A3C targets both 5'-TC-3' and 5'-CC-3' sequence motifs (29,30).

Understanding the basis for substrate recognition by A3G requires knowledge of enzyme structure as well as of the structural features of DNA important for enzyme recognition. Depending on salt conditions and substrate length, A3G is active in monomeric, dimeric, and larger oligomeric states (31-33). Expressed as a ~46kD polypeptide, each A3G monomer/subunit possesses two "catalytic domains", CD1 and CD2, located at the N- and C-termini of the protein, respectively. The two domains are highly homologous, each containing a single (H/C)XE(X)23-28CXXC motif conserved among zinc-binding proteins (34,35). However, the functions of the two elements are quite different: CD1 contributes to substrate binding, A3G encapsidation, and perhaps dimerization, while enzyme activity and substrate specificity are principally attributable to CD2 (21,36,37). Although the structure of full-length A3G is not currently available, its overall folding and subunit organization may resemble that of the recently resolved A2 enzyme (38). In contrast to previously resolved deaminase structures (39-41), A2 crystallizes as a rod-shaped tetramer that can be viewed as two homodimers joined head-to-head. Antiparallel B-strands within the CD1 domains of adjacent subunits interact to form the dimer interface in an arrangement that renders the helices bearing the zinc-coordination residues solvent accessible. Low-resolution small-angle scattering experiments indicate that human A3G also forms a rod-like structure (33), suggesting that the enzyme may contact DNA at multiple sites within, or even throughout, an extended

nucleoprotein complex in which enzyme and nucleic acid are in parallel configurations. This notion is supported by biochemical data indicating that A3G does not effectively bind substrates shorter than 10 nt (20). In addition, multiple binding sites for a single enzyme may contribute to the 3' → 5' processive deamination of ssDNA catalyzed by A3G on longer substrates (20). However, despite these data, the sites and nature of contacts between enzyme and DNA along the length of the complex remain unknown.

Recently, the structure of the CD2 domain of A3G has been solved both by NMR and X-ray crystallography (42,43), although the latter structure reveals features of the domain unresolved by the former, and convincingly maps the substrate binding groove (43). The overall folding of the A3G domain closely resembles that of A2 (38), with a core β-sheet surrounded by six α-helices, and a zinc-coordinating active center nestled near the bottom of the groove. However, active center (AC) loops 1 and 3 – conserved loop regions that flank the groove – are longer and more removed from the active site in the A3G domain. The bottom of the groove is lined with hydrophobic and positively charged residues that likely stack or form hydrogen bonds with the ssDNA substrate. Indeed, mutational analysis reveals most of these residues to be critical to A3G function (43). However, because individual nucleoprotein contact sites were not specifically delineated in this enzyme structure, those DNA structural features that determine deamination specificity remain unclear. Furthermore, although mutational analysis of APOBEC enzymes demonstrates that substrate specificities are determined by amino acid residues within CD2, the effects of point mutations within this domain are subtle and difficult to explain (30).

We demonstrate here the application of nucleoside analog interference mapping to probe A3G-DNA interactions throughout the enzyme-substrate complex, as well as to determine which DNA structural features direct substrate specificity. Previous exploration in this area has been limited primarily to conventional biochemical and molecular biological techniques, in which only the four standard deoxynucleosides (A, C, T, and G) have been used to produce substrate variants. In contrast, the approach presented here permits thorough dissection of the

chemical requirements of substrate recognition at the submolecular level. Our results indicate that multiple features of nucleosides within the 5'-CCCA-3' consensus sequence are important for substrate recognition (with base moieties being most critical), whereas deamination interference by analog substitution outside this region is minimal. Furthermore, exocyclic groups in pyrimidines 1-2 nts 5' to the target cytosine dictate substrate recognition, with chemical composition at ring positions 3 and 4 more important than the presence or absence of a methyl group at ring position 5. Taken together, these results provide insights into how the enzyme specifically selects A3G hotspot motifs for deamination, as well as which approaches might be best suited for forming a stable, catalytically competent, cross-linked A3G-DNA complex for future structural studies.

EXPERIMENTAL PROCEDURES

Expression and Purification of A3G – Recombinant wild type and D316R/D317R A3G were constructed, expressed, and purified as previously described (20,43,44). Sf9 cells were infected with recombinant A3G virus at a multiplicity of infection of 1 and harvested after 72 h. Cells were lysed in the presence of 50 µg/ml RNaseA. Cleared lysates were then incubated with glutathione-Sepharose resin (GE Healthcare) and subjected to a series of salt washes (0.25-1M NaCl) before elution in buffer as previously described (31). A3G was treated with 0.02 units/µl thrombin (GE Healthcare) for 16 h at 21°C to release the GST-tag and subsequently the sample was diluted to 50 mM NaCl and loaded onto a DEAE FF column (GE Healthcare) equilibrated with 50 mM Tris-HCl, pH 8.9, 50 mM NaCl, 1 mM DTT and 10% glycerol. A3G was eluted at ~200 mM NaCl by using a linear gradient from 75 to 1000 mM NaCl. Fractions were collected and stored at -70°C. Both A3G and the mutant derivative are approximately 95% pure.

Design, Synthesis, and Purification of Oligodeoxyribonucleotides – The standard oligodeoxyribonucleotide substrate 5' - AAA GAG AAA GAG AAA CCC AAA GAG GAA AGG TGA GGA GAA -3' was purchased from Integrated DNA Technologies (Coralville, IA), as were variants containing A, T, C, G, deoxyuridine (U), deoxyinosine (I) or 2'-O-methyl substitutions.

Additional analog-substituted oligonucleotides were synthesized according to manufacturers' protocols on a 200 nmole scale using an Expedite Model 8909 DNA synthesizer (PE Biosystems, Waltham, MA). Most DNA synthesis reagents, 2'-deoxy-5-methylcytidine (mC), 2'-deoxy-pseudo-uridine (ψ), 2'-deoxy-5-aza-5,6-dihydrocytidine (C*), 2'-deoxyzebularine (Z), 2'-deoxy-uridine (U), dSpacer (Abasic), UltraMILD and standard phosphoramidites, and methyl phosphoramidites (for methyl phosphotriesters (PTE)) were purchased from Glen Research (Sterling, VA). Isocytidine (iC), 4-ethyl cytidine (4C), and N3-methyl cytidine (3C) phosphoramidites were purchased from ChemGenes Corporation (Wilmington, MA). Deprotection and cleavage from the CPG support were carried out by incubation in 30% aqueous ammonia at 55° for 16 hr, except for oligonucleotides containing iC, C*, or Z (24 hr, 37°C), or methyl phosphoramidites (0.05 M potassium carbonate, 4 hr, 25°C). All oligonucleotides were purified by polyacrylamide gel electrophoresis, quantified spectrophotometrically, and stored in 10 mM Tris (pH 8.0), 1 mM EDTA at -80°C.

Cytosine Deamination Assay – 50 nM wild type or 10 nM D316R/D317R A3G was incubated with 50 nM 5' ³²P-end-labeled DNA oligonucleotide in reaction buffer (50 mM Tris-Cl (pH 8.0), 5 mM EDTA, 5 mM DTT, and 200 µg/ml BSA) for 30 min at 37°C. Reactions were terminated by extraction with 25:24:1 phenol/chloroform/ isoamyl alcohol, and products recovered by ethanol precipitation of the aqueous phase. Reaction products resuspended in reaction buffer were then incubated with 2 units of uracil DNA glycosylase (Fermentas, Glen Burnie, MD) for 1 hr at 37°C to remove uracil generated by A3G-mediated cytosine deamination, or, in the case of the control substrates, chemically incorporated uracil. Previously, DNAs containing nascent alkali-sensitive abasic sites were hydrolyzed by heating at 95°C for 20 min in the presence of 0.2 M NaOH (20,21). However, several of the nucleoside analogs used in the present study were incompatible with this treatment and required milder hydrolysis conditions. Therefore, reaction products were incubated in 30% aqueous ammonia for 16 hr at 37°C, lyophilized to dryness, resuspended in a urea-based gel-loading buffer, and fractionated

over a 15% denaturing polyacrylamide gel. Deamination/cleavage products were quantified by phosphorimaging using an FX model Molecular Imager and QuantityOne software (BioRad, Hercules, CA). Deamination activity was measured as the fraction of total radioactivity present in a specific hydrolysis product, while Relative Deamination is defined as the ratio of deamination activity measured on a modified substrate divided by the deamination activity measured on the standard substrate in reactions catalyzed by a common enzyme.

RESULTS

Experimental System – Oligonucleotides containing single or multiple nucleoside analog substitutions were used to probe substrate binding features and deamination specificity of A3G. Specifically, a series of synthetic 39-mer oligonucleotides containing one or more modifications to their base, deoxyribose, or phosphate moieties were synthesized and utilized in the biochemical assay of A3G activity described in Experimental Procedures. Each oligonucleotide is a variant of the standard substrate, which contains a unique extended A3G-recognition motif (5'-AAACCCAAA-3'; Fig. 1[A]) determined by Chelico *et al.* to be the most receptive target of A3G-mediated deamination among the sequences tested (20). 5' ³²P-end-labeled standard and modified substrates are incubated with A3G, which site-specifically catalyzes C→U deamination. Reaction products are exposed to uracil DNA glycosylase (UDG), which exclusively hydrolyses the glycosidic bond of uracil to create an alkali-labile abasic site that can be quantitatively hydrolyzed by overnight incubation in 30% ammonium hydroxide (producing strand scission at that position). Hence, by fractionation of the processed reaction products over a denaturing polyacrylamide gel, both the site and extent of A3G-mediated deamination in a given reaction can be quantitatively determined.

Two basic approaches were taken to explore substrate recognition by A3G. First, oligonucleotides containing single 2'O-methyl (2'Om), phosphotriester (PTE) or abasic (Ab) substitutions from position -10 to +10 (defined as 10 nt 5' or 3' to the deamination target, respectively; Fig. 1[A]) were used to explore the

effects of each of these modifications on enzyme activity, as well as to probe the A3G binding site (Fig. 1[B]). Once the bases critical to A3G activity were determined (and on the basis of previous findings), pyrimidine analogs (Fig. 1[C]) were introduced into synthetic substrates at positions -2, -1, +1, +2, +3, and combinations thereof, in order to dissect the base composition requirements of A3G-mediated deamination. Due to assay limitations, the effects of structural variations in the deamination target itself (i.e., the nucleoside at position 0) were not examined directly.

Note that in the first set of experiments, substrate recognition by A3G containing D316R and D317R substitutions was also examined. Based on homology with other cytosine deaminases, these mutations were introduced to increase positive charge at the A3G-DNA interface, thereby exploring the interplay among A3G surface charge, increased affinity for substrate (perhaps by virtue of additional interactions with the DNA phosphate backbone), and deamination specificity. Indeed, the recent A3G CD2 crystal structure places the two residues at the base of the substrate binding groove, where they may play a role in positioning ssDNA for deamination (43).

Scanning 2'O-methyl Substitution – A3G deamination profiles derived from substrates containing 2'Om substitutions at positions -10 through +10, and their quantitation, are shown in Fig. 2[A] and [B], respectively. "0" marks the migration of the 5'-labeled hydrolysis product resulting from deamination of the cytosine at position 0. To ensure that the 2'Om analog does not affect the assay in a manner unrelated to A3G activity, control substrates containing deoxyuridine (dU) at position 0, as well as 2'Om at position -2, -1, +1, or +2, were also evaluated (Fig. 2[A], lanes marked "dU"). Substrate containing 2'Om at position 0 (lanes "0") was included in this experiment, but only as a reference. While reactions containing this substrate yield no product, this cannot be attributed to failure of A3G to deaminate the target cytosine, since 2'Om-U is not recognized by UDG.

The results indicate that the effects of 2'Om substitution are greatest when the analog is introduced in the immediate vicinity of the target cytosine, *i.e.* at positions -3 to +1, although some

effect is also observed upon substitution at positions -5, -4, and +2 to +4 (decreasing with distance from the target cytosine). This indicates that only binding of the catalytic domain (most likely CD2) is affected by 2'Om substitution, in a manner consistent with the 3 nt (or bp) nucleic acid footprint typically observed for zinc finger binding motifs (reviewed in (45)). It is possible that on short oligonucleotides, the interaction of other regions of A3G (e.g., CD1, or the opposite CD2 in a dimeric form) with DNA is not required for activity. Alternatively, contacts between other regions of A3G and substrate do occur, but (1) they are not affected by introduction of a 2'Om group, or (2) binding outside the catalytic domain is not restricted to any particular region within the substrate, and is therefore not susceptible to interference by analog substitution at any single position.

Interestingly, while reduced relative to standard substrate, A3G activity on substrate containing a -1, 2'Om substitution is greater than that observed with oligonucleotides containing substitutions at positions -3, -2, or +1, despite equal or greater proximity to the active site. This result is reproducible, and suggests that direct contacts between A3G and the 2' H moiety at position -1 of the substrate may not exist, or are not completely disrupted by the presence of an O-methyl moiety. It also indicates that the ribose 2' position of the adjacent cytosine may be a suitable site for insertion of a cross-linking agent in order to form a covalently linked, yet functional, enzyme-substrate complex for biochemical and/or structural analysis.

Fig. 2[C] depicts the deamination profiles of the A3G D316R/D317R mutant on standard and 2'Om-substituted substrates. Comparing Lanes "Sm" and "S" illustrates the unusual activity of the mutant enzyme, *i.e.*, deamination occurs at positions 0 and -1 with nearly equal efficiency. In addition, deamination of ~30% of the substrate required half as much D316R/D317R mutant relative to wild type. While the precise biochemical mechanism remains unclear, it is likely that the broader specificity and increased activity of this mutant result from increased positive charge at the enzyme-substrate interface, and that the function of the wild type residues may involve electrostatic repulsion of the substrate phosphate backbone. Such a notion is further

supported by fluorescence anisotropy studies, which demonstrate that the affinity of the mutant enzyme is 1.5- to 2-fold higher than that of the wild type enzyme (Supp. Fig. 1).

With respect to deamination of the cytosine at position 0 (Fig. 2[A]-[D]), the same substitutions affect wild type and mutant enzyme activity, but the degree of interference varies significantly between the two enzymes. Specifically, 2'Om substitution at positions -3, -2, or +1 has a lesser effect on deamination catalyzed by the mutant enzyme relative to the wild type (compare equivalent bars in 2[B] and 2[D]), while the mutant appears to be more sensitive to substitution at position -1. This entire pattern appears to be shifted one nucleotide when deamination/hydrolysis at position -1 is quantified (Fig. 1[E]), except for minor differences at the peripheries of the two windows. These data suggest that the entire catalytic domain of the mutant enzyme shifts to deaminate cytosine at position -1, and that the sensitivities to substitution at adjacent positions shift with it.

Methyl Phosphotriester Substitution – The previous experiments illustrated that A3G-mediated deamination is sensitive to 2'Om substitution at positions -3 to +1, suggesting that critical nucleoprotein contacts in this region are disrupted. However, while these experiments explore the interaction between A3G and the sugar moieties of the DNA substrate, they do not address the dependence of A3G activity on phosphate contacts. Indeed, because the trajectories of 2'Om and non-phosphate oxygens emerging from the DNA strand differ by more than 90° (Fig. 1[B]), enzyme contacts with sugar and phosphate moieties may be significantly different. For a more thorough analysis of the dependence of A3G activity on phosphate contacts, substrates in which methyl phosphotriester (PTE) substitutions were introduced at positions -10 to +11 were evaluated (where phosphates immediately 5' and 3' to the target cytidine are defined as -1 and +1, respectively).

A3G-mediated deamination profiles derived from PTE-substituted substrates, and quantitation thereof, are shown in Fig. 3[A] and 3[B], respectively. The results are similar to those observed with 2'Om substitution, with a window of sensitivity spanning positions -3 to +1, further indicating that interference only occurs when

binding of the catalytic domain is disrupted. However, the degree to which A3G activity is reduced upon PTE-substitution is not as pronounced as that observed with the 2'Om substrates (compare Figs. 2[B] and 3[B]). This is likely because (1) the PTE moiety is an enantiomeric mixture (*i.e.*, the methyl group may be on either non-bridging phosphate oxygen at a particular position), and/or (2) basic residues may still effectively form hydrogen bonds with the non-bridging oxygen despite the presence of the methyl PTE.

In the D316R/D317R A3G profiles, the window of sensitivity contracts by a single nucleotide (Fig. 3[C] and [D]), with PTE substitution at positions -2, -1 or +1 resulting in approximately 40% relative deamination of the target cytosine. When deamination at position -1 is quantified (Fig. 3[E]), a shift in the entire window of sensitivity by a single nucleotide toward the 5' end of the substrate is observed, consistent with the corresponding 2'Om-substituted substrate profiles of Fig. 2.

Scanning Abasic Substitutions – Single abasic moieties were introduced stepwise to examine how base removal from nucleotides -10 to +10, and potential stacking, would impact A3G activity. Note that the abasic analog used here is a reduced derivative of the naturally occurring abasic lesion, and is therefore insensitive to alkaline hydrolysis. Thus, as with the equivalent 2'Om substitution, the absence of hydrolysis products from substrates containing an abasic moiety at position 0 (lanes "0") cannot be meaningfully interpreted.

From the profiles depicted in Fig. 4[A] it is apparent that removing an entire base has a greater effect on A3G activity than introducing an O-methyl or methyl group into the sugar phosphate backbone, as abasic substitutions from position -3 to +1 almost completely prevents cytosine deamination. In this experiment, the "edge" of the window of sensitivity 5' to the target cytosine is clearly defined, with abasic substitution at position -4 having minimal effect on A3G activity, while substitution at position -3 almost completely prevents cytosine deamination (Fig. 4[A],[B]). In contrast, the window has no clear border 3' to the deamination target, with a more gradual reduction in sensitivity as the abasic substitution is progressively relocated from position +1 to +5. This indicates that base-dependent A3G contacts

between +1 and +5 are important, but not absolutely essential for deamination activity. The profile also indicates that removal of the base at position -1 almost completely prevents deamination of the target cytosine. This is in contrast to prior experiments in which 2'Om and PTE-substituted substrates were utilized, indicating that A3G activity is highly dependent on base composition at this position.

While D316R/D317R A3G exhibits similar profiles on these abasic substrates with respect to deamination at position 0, this mutant is less sensitive to substitution at position -3, -2, -1, +1 or +2 (Figs. 4[C],[D]). This may be the result of an overall higher affinity of the mutant enzyme for its substrate (43). In addition, the window of sensitivity relative to deamination at position -1 is poorly defined (Fig. 4[C],[E]), as relative deamination increases gradually (in either direction) with the number of nucleotides from the target cytosine. This window can be loosely interpreted as spanning from position -4 to +5, inclusive. It is not clear why the 5' border should be so ill-defined in this case, although variations in sequence context between the two target cytosines may play a role.

Pyrimidine Analog Substitution 5' to the Target Cytosine – In each of the scanning substitution approaches described above, the nucleotide composition 2-3 nt immediately 5' and one nt 3' to the target site are shown to be important for efficient deamination of the target cytosine. In particular, removing nucleobases adjacent to the target site is especially detrimental. These data support the conventional definitions of the A3G consensus sequence as 5'-CCCA-3' (11), with 5'-AAACCCAAA-3' recognized most efficiently (20). To examine the chemical basis for A3G recognition of the consensus sequence, we first introduced a series of pyrimidine analog substitutions at position -1, -2, or both (Fig. 5[A]-[C], respectively; quantitation below each panel). The chemical structure of each analog is shown in Fig. 1[C].

Because some of the pyrimidine analogs are alkaline labile, a weak base (ammonium hydroxide) was used to hydrolyze reaction products following treatment with uracil DNA glycosylase. 16 hr incubation in this reagent was sufficient to hydrolyze all dU-containing control substrates, and did not otherwise interfere with

reaction processing. Despite the mild conditions, a small amount of strand scission at the site(s) of analog substitution is evident (*i.e.*, in products that migrate to positions other than that marked "0"). These products do not affect how deamination is quantified, however, because (a) degradation occurs after the substrate has been incubated with A3G, and (b) the extent of degradation is 8% or less of starting material in all cases. Profiles illustrating products of pyrimidine analog hydrolysis in the absence of A3G are shown in Supplemental Data (Supp. Fig. 2).

Fig. 5[A] illustrates clear preferences among pyrimidine analogs introduced at position -1. Deamination of substrates containing T at position -1 is highly disfavored relative to those containing C at the same position. Unlike C, T is protonated at the 3 position of the pyrimidine ring, and contains carbonyl oxygen and methyl groups at ring positions 4 and 5, respectively. Interestingly, the 5-methyl moiety appears to contribute minimally to the distinction among pyrimidines. Substrate containing 5-methyl C (mC) at position -1 is deaminated with greater than 60% efficiency relative to the control substrate, while relative deamination of DNA containing pseudouridine (ψ) (which lacks a 5-methyl group, but otherwise resembles T) is less than 2%.

Substrates containing 2'-deoxy-5-aza-5,6-dihydrocytidine (C*) or 2'-deoxy-4-ethylcytidine (4C) at position -1 are deaminated by A3G with 25% and 19% efficiency, respectively. This intermediate phenotype is likely the result of competing effects by different structural features of each analog on enzyme recognition. For example, while the pyrimidine ring composition of C and C* are identical at positions 2, 3, and 4, saturation of the 4-5 bond in C* likely affects stacking between this analog and adjacent residues, which may disrupt substrate recognition. Similarly, despite being structurally identical to C in all other respects, the 4-ethyl moiety at the N4 ring position may sterically interfere with A3G substrate recognition.

Substrates containing 3-methyl cytosine (3C) or isocytosine (iC) at position -1 are not appreciably deaminated by A3G. The former analog contains an N3 methyl group, the steric and/or H-bonding effects of which are apparently sufficient to completely disrupt enzymatic function. In contrast, iC not only lacks an N3

alkyl moiety, but, like C, it is most likely not protonated at this position under the conditions of this assay. Therefore, although an unprotonated N3 correlates well with A3G activity, this alone is insufficient to direct deamination of the 3'-adjacent C. Furthermore, because enzyme activity is completely disrupted by exchange of amino and carbonyl groups (relative to C) at ring positions 2 and 4, the chemical determinants of A3G recognition in this context are clearly position-specific. Finally, relative deamination of substrate containing 5-methyl Zebularine (Z) is 8.8%, indicating that the 4-amino group, while not absolutely essential, contributes significantly to substrate recognition by A3G.

Based on both prior results (20) and rotational anisotropy experiments conducted for the present study (Supp. Fig. 1), these results cannot be explained by differences in the affinity of A3G for the various substrates. Indeed, neither varying the base composition at position -1, nor removal of the recognition sequence entirely appears to affect A3G-DNA binding.

Profiles derived from substrates containing pyrimidine analogs at the -2 position are shown in Fig. 5[B]. In general, these profiles are quite similar to those derived from analog substitution at position -1, although A3G is somewhat less sensitive to introduction of T, ψ , iC, and 3C at the -2 position. This is consistent with reports that A3G tolerates T at the -2 position, although C is preferred (22). Remarkably, substituting 4C for C at position -2 results in a nearly 50% increase in deamination relative to the wild type substrate. It may be that while substitution of this analog at position -1 produces a steric clash that inhibits A3G activity, interference does not occur when 4C is removed one base further from the target cytosine. Instead, the hydrophobic ethyl moiety may make favorable, likely hydrophobic, contacts with A3G that stabilize the complex and augment enzyme activity. As was the case with 2'Om substitution at position -1, it may be that the 4-amino group of the cytosine at position -2 is an ideal site for introduction of a chemical crosslinking agent.

In some cases, the effects of analog substitution at positions -1 and -2 appear to be additive, based on dual substitution profiles presented in Fig. 5[C]. As was the case with -1 substitution, deamination of substrates containing

T, ψ , iC, or 3C at both positions is minimal. The same is true for substrates containing C* or Z at both positions, where the effects of tandem substitution appear to be cumulative. Deamination of oligonucleotides containing mC at both positions is ~57%; a lower value than was measured on either individually substituted substrate. However, introducing 4C at position -2 partially counterbalances the inhibitory effects of 4C at position -1, as the relative deamination value obtained from the dually substituted DNA is more than double that derived from substrate containing 4C at position -1 alone.

Nucleoside Analog Substitution 3' to Target Cytosine – Analysis of virus mutation patterns indicate that A3G-mediated deamination is most efficient on sequences where A is immediately 3' to the target C (11). In addition, Chelico *et al.* (2006) have shown that oligonucleotides containing a 5'-CCCAA-3' motif is deaminated more efficiently than 5'-CCCG-3', and much more efficiently than 5'-CCCTT-3', demonstrating that as many as three nt 3' to the target C can influence activity (20). To analyze this in more detail, A \rightarrow X purine and pyrimidine analog substitutions were introduced at positions +1 and +1 through +3, inclusive, with the resulting profiles shown in Fig. 6[A] and [B].

Fig. 6[A] demonstrates substituting +1A of the standard substrate with another purine (G or I) has negligible effect. However, introducing a pyrimidine at this position impairs enzyme activity by 43% or more, with the naturally occurring T being the most tolerated. This suggests that the double-ring purine structure, and not its exocyclic composition, is the principle determinant of deamination efficiency in this context. Interestingly, while A \rightarrow C substitution at position -1 creates a series of 4 consecutive Cs (and thus two overlapping 5'-CCC-3' motifs), neither potential cytosine target is deaminated as efficiently as in the 5'-CCCA-3' motif in the standard substrate.

When nucleosides positions +1 to +3 are triply-substituted, G and I are less well tolerated (Fig. 6[B]). Inosine nucleoside, like A, contains only hydrogen at the purine 2 position, but also contains a 4-carbonyl substitution and possesses an electrostatic potential more closely resembling that of G. Hence, it is likely that latter properties

are somewhat inhibitory to A3G activity in this context, and that this effect is cumulative with consecutive purines. Similarly, substrates containing consecutive pyrimidines at positions +1 through +3 are deaminated much less efficiently than those containing a pyrimidine only at position +1, an effect which otherwise appears to be largely independent of base composition. Again, substrate which contains more than three consecutive Cs (6, in this case) is deaminated inefficiently.

Because the substrates in question contain A \rightarrow C substitutions rather than C insertions, the sequence context 3' to the string of consecutive Cs varies (*i.e.*, Cs are no longer flanked on both sides by 3 As). In addition, the separation between the 3'-terminal C in the motif and the 3' terminus of the oligonucleotide is reduced by 1 or 3 nucleotides (in substrates containing 4 or 6 consecutive Cs, respectively). This may be important, as Chelico *et al.* (2006) demonstrated that both sequence context of an A3G recognition sequence, and its proximity to the substrate 3' terminus, significantly affect deamination activity (20). To address the possibility that poor deamination of overlapping 5'-CCC-3' motifs is due to altered sequence context, a series of substrates were constructed that each contain a unique 5'-AAAC_xAAA-3' deamination target within the sequence context of the standard substrate (*i.e.* 5'...AAAC₃AAA...3'), and where X = 2 to 6. The resulting profiles are depicted in Fig. 6[C]. Note that unlike previous experiments where only deamination at position 0 (or -1, in the case of the mutant enzyme) is considered for quantitation, total relative deamination is defined in Fig. 6[C] as the ratio of the sum of all products in a reaction to the sum of products derived from the standard substrate (*i.e.*, where X = 3).

Consistent with previous definitions of the A3G consensus sequence, substrate containing only two consecutive Cs (flanked by As) is not efficiently deaminated. In addition, it is clear that even when sequence context is considered, motifs containing more than three consecutive cytosines are deaminated only half as efficiently as the optimized 5'-AAACCCAAA-3' motif present in the standard substrate. Also, in motifs containing 5 or 6 consecutive Cs, the 3'- and 5'-terminal Cs are deaminated more efficiently than internal Cs, which likely reflects the importance of the flanking As in substrate recognition.

DISCUSSION

In the present study, the molecular footprint of A3G on a 39-base oligonucleotide substrate was determined by nucleoside analog interference mapping. Deamination profiles derived from substrates containing stepwise 2'Om, PTE, or abasic substitutions revealed overlapping windows of sensitivity spanning positions -3 to +2, indicating that (i) introducing a methyl or O-methyl group into the sugar-phosphate backbone, or (ii) removal of a base within the 5'-CCCA-3' consensus sequence (or one nt to either side) is sufficient to inhibit or prevent deamination of the target cytosine. In general, the extent to which deamination was disrupted was highly analog dependent. Abasic substitution virtually abolished A3G activity over a wider range of positions than other analogs (*i.e.*, from -3 to +2, with decreasing sensitivity from +3 to +5), while relative deamination of substrate containing a methyl PTE linkage at any position was always at least 28%. Although specific sites of contact have not been described, aromatic residues within the substrate binding cleft of the A3G CD2 (Y315, W285) may participate in stacking interactions with nucleoside bases, while positively charged residues R213, R320, R374 and/or **R376** may stabilize the enzyme-substrate complex by hydrogen-bonding to the phosphate backbone (43).

The breadth of the A3G-DNA interface indicates that while analog substitution interferes with substrate recognition by CD2, nucleoprotein contacts outside of this domain either do not exist or are unaffected. Since truncated, bacterially expressed A3G containing only the CD2 domain is enzymatically active (46), contacts outside of this domain may not be essential and/or may not occur. This notion supported by a recent study predicting that under low-salt conditions, short oligonucleotides are most likely deaminated by monomeric A3G (31).

It is also possible, however, that contacts beyond the catalytic CD2 domain are indeed formed, but are not restricted to a single site on the DNA substrate and are therefore not detectable by our methodology. The latter scenario is consistent with a model proposed by Chelico *et al.* to explain the 3'→5' processive nature of A3G-mediated deamination (20). In this model, CD1 domains

interact to form a "head-to-head" homodimer, while CD2 domains either concurrently or alternately bind substrate as the enzyme "jumps" along long ssDNAs. While at least one of these CD2 domains must be located at an A3G consensus sequence for catalysis to occur, the other may bind ssDNA non-specifically. While intriguing, principally applicable to A3G-mediated deamination of long ssDNAs, and are therefore not subject to validation by the present study.

A3G containing D→R substitution mutations at amino acid residues 316 and 317 is not only more active than the wild type enzyme, but exhibits an altered deamination specificity, catalyzing C→U conversion at position -1 as well as at position 0 (Fig. 2[A]). The differing sequence contexts of the two deamination targets (*i.e.*, 5'-ACCC-3' vs. 5'-CCCA-3' indicate that the mutant enzyme is less sensitive to base composition 2 nts 5' and 1 nt 3' to the target cytosine. Similarly, D316R/D317R A3G is generally less sensitive to 2'Om, abasic, or PTE substitution than the WT enzyme, as evidenced by narrower molecular footprints (2'Om, PTE) and/or higher relative deamination values on individual substrates (Figs. 2, 3, and 4, compare panels [B] and [D]). Taken together, and with the fluorescence anisotropy data presented in Supp. Fig. 1, these results are consistent with the notion that the D316R and D317R substitutions increase the affinity of A3G for its substrate by (a) substituting for or complementing existing hydrogen bonds, or (b) by forming of new bonds with the DNA backbone. Conversely, the role of the wild type, negatively charged residues may be to increase the stringency and/or specificity of binding by electrostatic repulsion of the phosphate oxygens.

The data presented in Figs. 5 and 6 demonstrate that the composition of nucleosides at positions proximal to the target cytosine is crucial for A3G activity, with the chemical makeup of the pyrimidine at position -1 being especially critical (summarized in Table 1). T is a poor substitute for C at this position, as has been previously determined using conventional biochemical and molecular biological techniques (11,18,22). Remarkably, however, methyl group occupancy at position 5 of the pyrimidine ring, while constituting the greatest steric difference between C and T bases, does not correlate well with

deaminase function. Furthermore, since mC substitution at position -1 minimally affects deaminase activity, it is unlikely that critical contacts occur between A3G and the 5-position of -1C in the consensus recognition sequence. Note that these results do not conflict with data of Iwatani *et al.*, as the oligonucleotide used that study contained a C→mC substitution at position 0 (as well as at -1) and was therefore not deaminated by A3G (21).

The features of pyrimidine composition at position -1 (and -2) that best correlate with A3G activity relate to the chemical occupancy at ring positions 3 and 4. Of the five -1-substituted DNAs that were deaminated with greatest efficiency, the three best substrates contained a 4-amino moiety, and none were protonated at the N3 position (Table 1). Conversely, -1-substituted DNAs containing a carbonyl group at ring position 4 were universally poor A3G substrates, and three-of-four were protonated at N3. One consequence of the chemical differences between these two groups of substrates is that H-bonding roles at positions 3 and 4 would be inverted. Specifically, while N3 in the former group would be expected to serve as a hydrogen bond acceptor, the protonated nitrogen prevalent in the latter group would act as a donor. The inverse is true of -1-substituted DNAs alternately containing amino or carbonyl moieties at the 4-position of the pyrimidine ring. It is likely, therefore, that hydrogen bonds between A3G amino acid residues and DNA at the 3- and 4- ring positions in -1-substituted DNAs, or disruption thereof, play a critical role in substrate recognition. Potential contributors to this interaction include Y315 and/or one of the charged residues that line the floor of the substrate binding groove, although specific contact sites have not been demonstrated (43). Hydrogen bond and/or partial charge inversion may also explain how product is released from the active site following A3G-mediated conversion of C→U, although this possibility could not be tested here.

Alternatively, or perhaps in addition, changing the exocyclic groups within one or more pyrimidines in the 5'-CCCA-3' consensus sequence may alter stacking and/or electrostatic interactions among adjacent bases on flexible ssDNA substrates, which in turn could affect the efficiency with which substrate is recognized by

A3G. Such interactions (or disruption thereof) may also explain why purine trimers are preferred to pyrimidine trimers 3' (and 5') to the A3G consensus sequence (Fig. 6; (20)). For example, strong stacking interactions among purine trimers flanking the recognition sequence may, by excluding the 3 Cs, render the target sequence more accessible to the enzyme.

Although this study advances our understanding of structural requirements for A3G mediated deamination, a high resolution structure of the A3G-substrate complex will be required before important individual contacts can be identified. However, variability among A3G-DNA complexes may prove problematic, since the enzyme appears to bind and move along ssDNA randomly (20,31). Previously, catalytically competent nucleoprotein complexes have been stabilized by chemical crosslinking to facilitate crystallographic analysis (47-49). Our study illustrates that several sites within the DNA substrate might be suitable for attachment of a crosslinking agent without adversely affecting enzyme activity, including (i) the non-bridging phosphate oxygens (at several positions), (ii) the sugar 2'position (at substrate position -1), (iii) the pyrimidine 4 position (at substrate position -2), and (iv) the pyrimidine 5 position (at substrate positions +1, -1, or -2). While it remains to be seen whether nucleoprotein crosslinking will occur efficiently at any of these locations, or even which crosslinking approach is best suited for the application, this information may prove useful for obtaining a high resolution, catalytically relevant A3G-DNA complex, which would facilitate future structural analysis of this important enzyme.

REFERENCES

1. Sheehy, A. M., Gaddis, N. C., Choi, J. D., and Malim, M. H. (2002) *Nature* **418**(6898), 646-650
2. Conticello, S. G., Harris, R. S., and Neuberger, M. S. (2003) *Curr Biol* **13**(22), 2009-2013
3. Kobayashi, M., Takaori-Kondo, A., Miyauchi, Y., Iwai, K., and Uchiyama, T. (2005) *The Journal of biological chemistry* **280**(19), 18573-18578
4. Harris, R. S., Bishop, K. N., Sheehy, A. M., Craig, H. M., Petersen-Mahrt, S. K., Watt, I. N., Neuberger, M. S., and Malim, M. H. (2003) *Cell* **113**(6), 803-809
5. Kao, S., Khan, M. A., Miyagi, E., Plishka, R., Buckler-White, A., and Strebel, K. (2003) *Journal of virology* **77**(21), 11398-11407
6. Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L., and Trono, D. (2003) *Nature* **424**(6944), 99-103
7. Marin, M., Rose, K. M., Kozak, S. L., and Kabat, D. (2003) *Nature medicine* **9**(11), 1398-1403
8. Strebel, K., Daugherty, D., Clouse, K., Cohen, D., Folks, T., and Martin, M. A. (1987) *Nature* **328**(6132), 728-730
9. Guo, F., Cen, S., Niu, M., Yang, Y., Gorelick, R. J., and Kleiman, L. (2007) *Journal of virology* **81**(20), 11322-11331
10. Li, X. Y., Guo, F., Zhang, L., Kleiman, L., and Cen, S. (2007) *The Journal of biological chemistry* **282**(44), 32065-32074
11. Yu, Q., Konig, R., Pillai, S., Chiles, K., Kearney, M., Palmer, S., Richman, D., Coffin, J. M., and Landau, N. R. (2004) *Nature structural & molecular biology* **11**(5), 435-442
12. Zhang, L., Saadatmand, J., Li, X., Guo, F., Niu, M., Jiang, J., Kleiman, L., and Cen, S. (2008) *Virology* **370**(1), 113-121
13. Goncalves, J., Korin, Y., Zack, J., and Gabuzda, D. (1996) *Journal of virology* **70**(12), 8701-8709
14. Simon, J. H., and Malim, M. H. (1996) *Journal of virology* **70**(8), 5297-5305
15. von Schwedler, U., Song, J., Aiken, C., and Trono, D. (1993) *Journal of virology* **67**(8), 4945-4955
16. Lecossier, D., Bouchonnet, F., Clavel, F., and Hance, A. J. (2003) *Science (New York, N.Y)* **300**(5622), 1112
17. Zhang, H., Yang, B., Pomerantz, R. J., Zhang, C., Arunachalam, S. C., and Gao, L. (2003) *Nature* **424**(6944), 94-98
18. Suspene, R., Sommer, P., Henry, M., Ferris, S., Guetard, D., Pochet, S., Chester, A., Navaratnam, N., Wain-Hobson, S., and Vartanian, J. P. (2004) *Nucleic acids research* **32**(8), 2421-2429
19. Suspene, R., Rusniok, C., Vartanian, J. P., and Wain-Hobson, S. (2006) *Nucleic acids research* **34**(17), 4677-4684
20. Chelico, L., Pham, P., Calabrese, P., and Goodman, M. F. (2006) *Nature structural & molecular biology* **13**(5), 392-399
21. Iwatani, Y., Takeuchi, H., Strebel, K., and Levin, J. G. (2006) *Journal of virology* **80**(12), 5992-6002
22. Beale, R. C., Petersen-Mahrt, S. K., Watt, I. N., Harris, R. S., Rada, C., and Neuberger, M. S. (2004) *Journal of molecular biology* **337**(3), 585-596
23. Huthoff, H., and Malim, M. H. (2005) *Virology* **334**(2), 147-153
24. Harris, R. S., and Liddament, M. T. (2004) *Nature reviews* **4**(11), 868-877
25. Wiegand, H. L., Doehle, B. P., Bogerd, H. P., and Cullen, B. R. (2004) *The EMBO journal* **23**(12), 2451-2458
26. Zheng, Y. H., Irwin, D., Kurosu, T., Tokunaga, K., Sata, T., and Peterlin, B. M. (2004) *Journal of virology* **78**(11), 6073-6076
27. Bishop, K. N., Holmes, R. K., Sheehy, A. M., Davidson, N. O., Cho, S. J., and Malim, M. H. (2004) *Curr Biol* **14**(15), 1392-1396

28. Liddament, M. T., Brown, W. L., Schumacher, A. J., and Harris, R. S. (2004) *Curr Biol* **14**(15), 1385-1391
29. Yu, Q., Chen, D., Konig, R., Mariani, R., Unutmaz, D., and Landau, N. R. (2004) *The Journal of biological chemistry* **279**(51), 53379-53386
30. Langlois, M. A., Beale, R. C., Conticello, S. G., and Neuberger, M. S. (2005) *Nucleic acids research* **33**(6), 1913-1923
31. Chelico, L., Sacho, E. J., Erie, D. A., and Goodman, M. F. (2008) *The Journal of biological chemistry* **283**(20), 13780-13791
32. Opi, S., Takeuchi, H., Kao, S., Khan, M. A., Miyagi, E., Goila-Gaur, R., Iwatani, Y., Levin, J. G., and Strebel, K. (2006) *Journal of virology* **80**(10), 4673-4682
33. Wedekind, J. E., Gillilan, R., Janda, A., Krucinska, J., Salter, J. D., Bennett, R. P., Raina, J., and Smith, H. C. (2006) *The Journal of biological chemistry* **281**(50), 38122-38126
34. Bhattacharya, S., Navaratnam, N., Morrison, J. R., Scott, J., and Taylor, W. R. (1994) *Trends in biochemical sciences* **19**(3), 105-106
35. Wedekind, J. E., Dance, G. S., Sowden, M. P., and Smith, H. C. (2003) *Trends Genet* **19**(4), 207-216
36. Hakata, Y., and Landau, N. R. (2006) *The Journal of biological chemistry* **281**(48), 36624-36631
37. Shindo, K., Takaori-Kondo, A., Kobayashi, M., Abudu, A., Fukunaga, K., and Uchiyama, T. (2003) *The Journal of biological chemistry* **278**(45), 44412-44416
38. Prochnow, C., Bransteitter, R., Klein, M. G., Goodman, M. F., and Chen, X. S. (2007) *Nature* **445**(7126), 447-451
39. Betts, L., Xiang, S., Short, S. A., Wolfenden, R., and Carter, C. W., Jr. (1994) *Journal of molecular biology* **235**(2), 635-656
40. Johansson, E., Mejlhede, N., Neuhard, J., and Larsen, S. (2002) *Biochemistry* **41**(8), 2563-2570
41. Losey, H. C., Ruthenburg, A. J., and Verdine, G. L. (2006) *Nature structural & molecular biology* **13**(2), 153-159
42. Chen, K. M., Harjes, E., Gross, P. J., Fahmy, A., Lu, Y., Shindo, K., Harris, R. S., and Matsuo, H. (2008) *Nature* **452**(7183), 116-119
43. Holden, L. G., Prochnow, C., Chang, Y. P., Bransteitter, R., Chelico, L., Sen, U., Stevens, R. C., Goodman, M. F., and Chen, X. S. (2008) *Nature in press*
44. Bransteitter, R., Pham, P., Calabrese, P., and Goodman, M. F. (2004) *The Journal of biological chemistry* **279**(49), 51612-51621
45. Pabo, C. O., Peisach, E., and Grant, R. A. (2001) *Annual review of biochemistry* **70**, 313-340
46. Chen, K. M., Martemyanova, N., Lu, Y., Shindo, K., Matsuo, H., and Harris, R. S. (2007) *FEBS letters* **581**(24), 4761-4766
47. Huang, H., Chopra, R., Verdine, G. L., and Harrison, S. C. (1998) *Science (New York, N.Y)* **282**(5394), 1669-1675
48. Huang, H., Harrison, S. C., and Verdine, G. L. (2000) *Chemistry & biology* **7**(5), 355-364
49. Sarafianos, S. G., Clark, A. D., Jr., Tuske, S., Squire, C. J., Das, K., Sheng, D., Ilankumaran, P., Ramesha, A. R., Kroth, H., Sayer, J. M., Jerina, D. M., Boyer, P. L., Hughes, S. H., and Arnold, E. (2003) *The Journal of biological chemistry* **278**(18), 16280-16288

FOOTNOTES

The costs of publication were defrayed in part by the payment of page charges. This article must therefore be hereby-marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

FIGURE LEGENDS

Fig. 1. [A] Standard A3G substrate. The A3G consensus sequence is flanked on both sides by 2'-deoxyadenosines (A). The remainder of the sequence is purine rich, contains no cytosines, and was chosen to minimize self-complementarity. Position designations, where the target deamination site is defined as position "0", are indicated. [B], Structural image of a 5'-CCC-3' DNA trimer depicting the structural consequences of substitution with 2' O-methyl, methyl phosphotriester (PTE), or abasic nucleoside analogs. [C], Structures of pyrimidine base analogs within nucleosides introduced at positions adjacent to the deamination target site. These include C (2'-deoxycytosine), T (thymidine), mC (2'-deoxy-5-methyl-cytidine), ψ (2'-deoxypseudouridine), C* (2'-deoxy-5-aza-5,6-dihydro-cytosine), iC (2'-deoxyisocytosine), 4C (2'-deoxy-4-ethyl-cytosine), 3C (2'-deoxy-N3-methyl-cytosine), and Z (2'-deoxyzebarine). U (2'-deoxyuridine) is the product of 2'-deoxycytosine deamination, and is also depicted. Standard positions designations (1-6) common to all pyrimidine rings are indicated for C.

Fig. 2. Deamination of substrates containing a 2'O-methyl modification. [A] 5' ^{32}P end-labeled substrates were incubated with wild type A3G, and hydrolyzed at the site of 2'-deoxycytosine (C) deamination. Reaction products were fractionated over a 15% denaturing polyacrylamide gel and detected by phosphorimaging. Lane designations reflect the position at which each substrate was modified (e.g., lane -10 shows the wild type A3G-mediated deamination/cleavage products obtained from substrate containing a 2'O-methyl moiety at DNA position -10). Lane S shows products obtained from deamination of the unmodified substrate by wild type A3G. Lanes marked dU reflect processing of control substrates containing a deoxyuridine substitution at position 0 as well as a 2'O-methyl moiety at the position indicated. Migration positions of unmodified substrate and substrate hydrolyzed at position 0 are denoted S and 0, respectively. [B] Quantification of deamination products derived from 2'O-methyl-substituted DNA substrates. "Relative Deamination" indicates the extent of wild type A3G-mediated deamination of modified substrate at position 0 as a percentage of that observed on the standard substrate. [C] Deamination profiles of 2'O-methyl-substituted DNAs treated with D316R/D317R A3G, a mutant enzyme which deaminates at positions +1 as well as position 0. Lane S_m indicates the profile of the mutant enzyme on the unmodified substrate. [D] Quantification of D316R/D317R A3G-mediated deamination at position 0. [E] Quantification of D316R/D317R A3G-mediated deamination at position -1. Note that in these cases, the deamination profile of the mutant enzyme on the unmodified substrate is used as the standard for calculating "Relative Deamination".

Fig. 3. Deamination sensitivity of substrates containing a methyl-phosphotriester (MPTE) modification. [A]-[E] Lane, position and deamination/cleavage product designations are identical to those described in Fig. 2[A]-[E], respectively.

Fig. 4. Deamination sensitivity of substrates containing an abasic site. [A]-[E] Lane and deamination/cleavage product designations are identical to those described in Fig. 2[A]-[E], respectively.

Fig. 5. A3G-mediated deamination sensitivity of substrates containing pyrimidine analogs 5' to the deamination target. Substitutions were introduced at position -1, -2, or at both sites (marked by "X"), with profiles depicted in panels [A]-[C], respectively. Lane designations reflect the analog incorporated into the substrate used for each reaction. These include C (cytosine; present in the standard substrate), T (thymidine), mC (2'-deoxy-5-methyl-cytidine), ψ (2'-deoxy-pseudouridine), C* (2'-deoxy-5-aza-5,6-dihydro-cytidine), iC (2'-deoxy-isocytosine), 4C (2'-deoxy-4-ethyl-cytosine), 3C (2'-deoxy-N3-methyl-

cytosine), and Z (2'-Deoxyzebularine). Relative deamination values were calculated from the wild type A3G profiles as previously described (Fig. 2-4, Experimental Procedures) and are shown beneath the respective panels. Note that products of A3G-independent pyrimidine analog hydrolysis (lanes C*, iC, 3C and Z, especially) were identified and quantified by comparison of these profiles with those obtained in the absence of A3G (Supp. Fig. 2).

Fig. 6. A3G-mediated deamination of substrates containing pyrimidine analogs 3' to the target C. Substitutions were introduced at position +1, or at each of positions +1, +2 and +3 (marked by "X"), with resulting profiles depicted in [A] and [B], respectively. Quantification of deamination activity is also shown. Lane designations and analog substitutions were as described in Fig. 5, except that 2'-deoxyadenosine (A) occupies positions -1 to -3 in the standard substrate, and substrates containing 2'-deoxyguanosine (G) or 2'-deoxyinosine (I) were also examined. Panel [C] depicts the deamination profiles derived from substrates containing a unique 5'-AAAC_XAAA-3' motif, where the value of X varies from 2 to 6, as indicated in lane designations. In this case, "Relative Deamination (Total)", is defined as the sum of deamination products at all positions within a particular substrate divided by the total measured for the standard substrate (i.e., where X = 3).

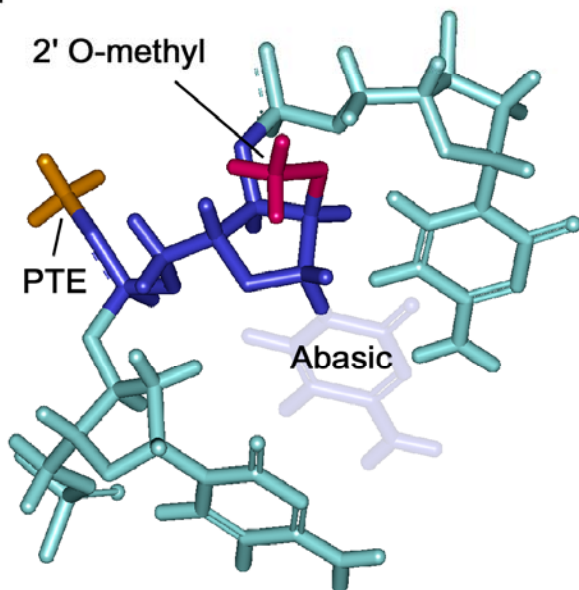
Table 1. Correlation between pyrimidine composition at position -1 and A3G-mediated deamination. Structural distinguishing features of each analog, and those which best correlate with A3G activity, are highlighted by grey and clear boxes, respectively.

Figure 1

[A]



[B]



[C]

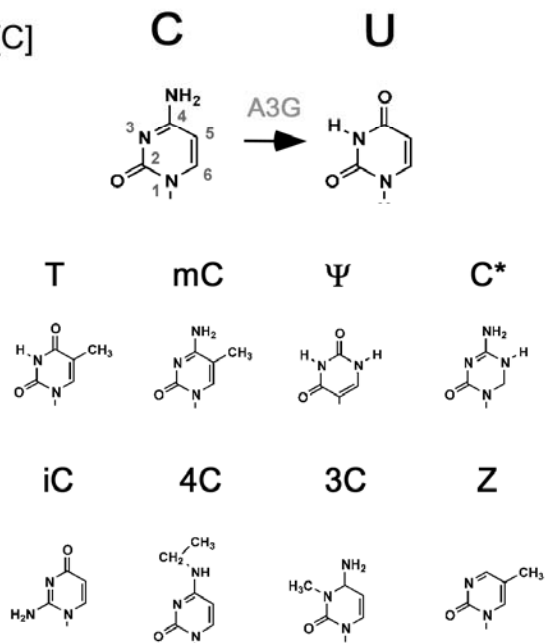


Figure 2

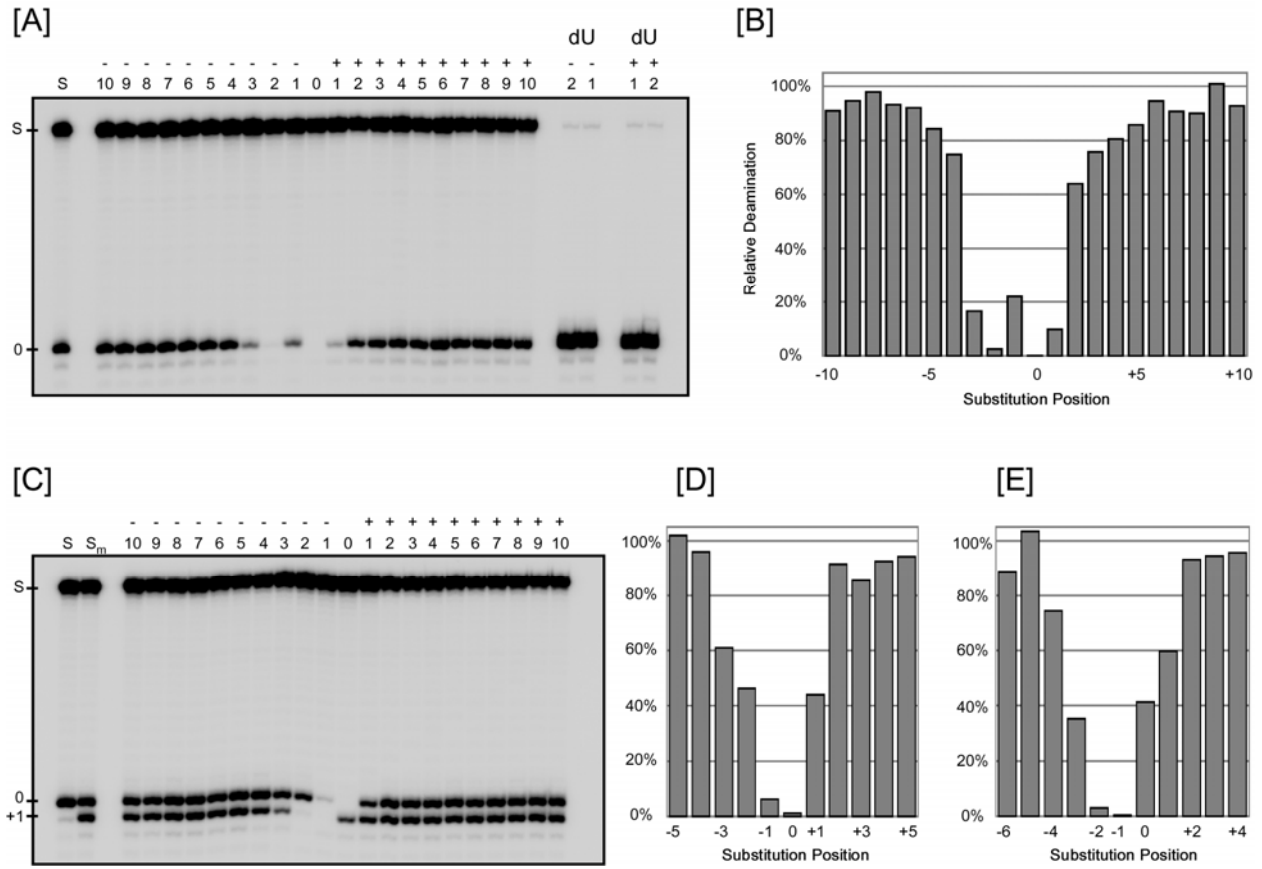


Figure 3

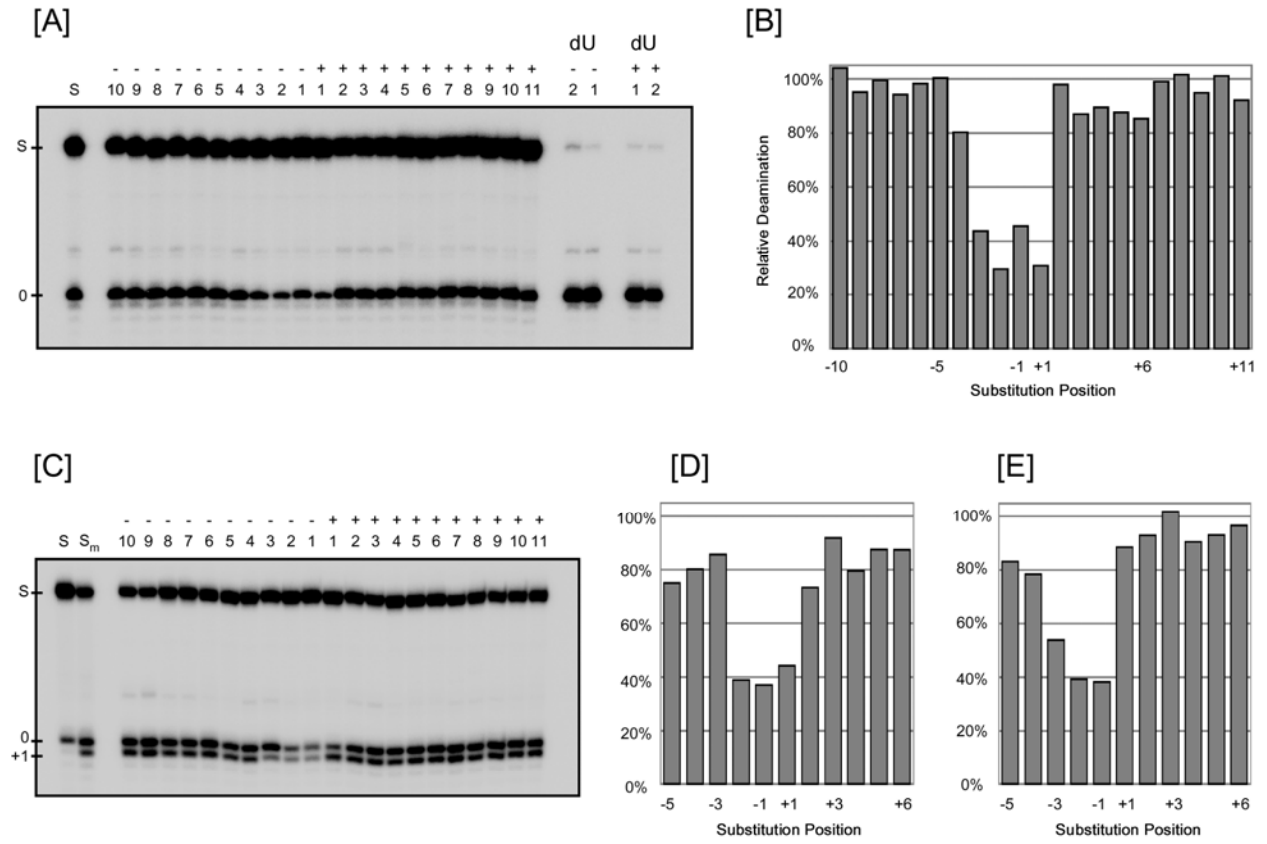


Figure 4

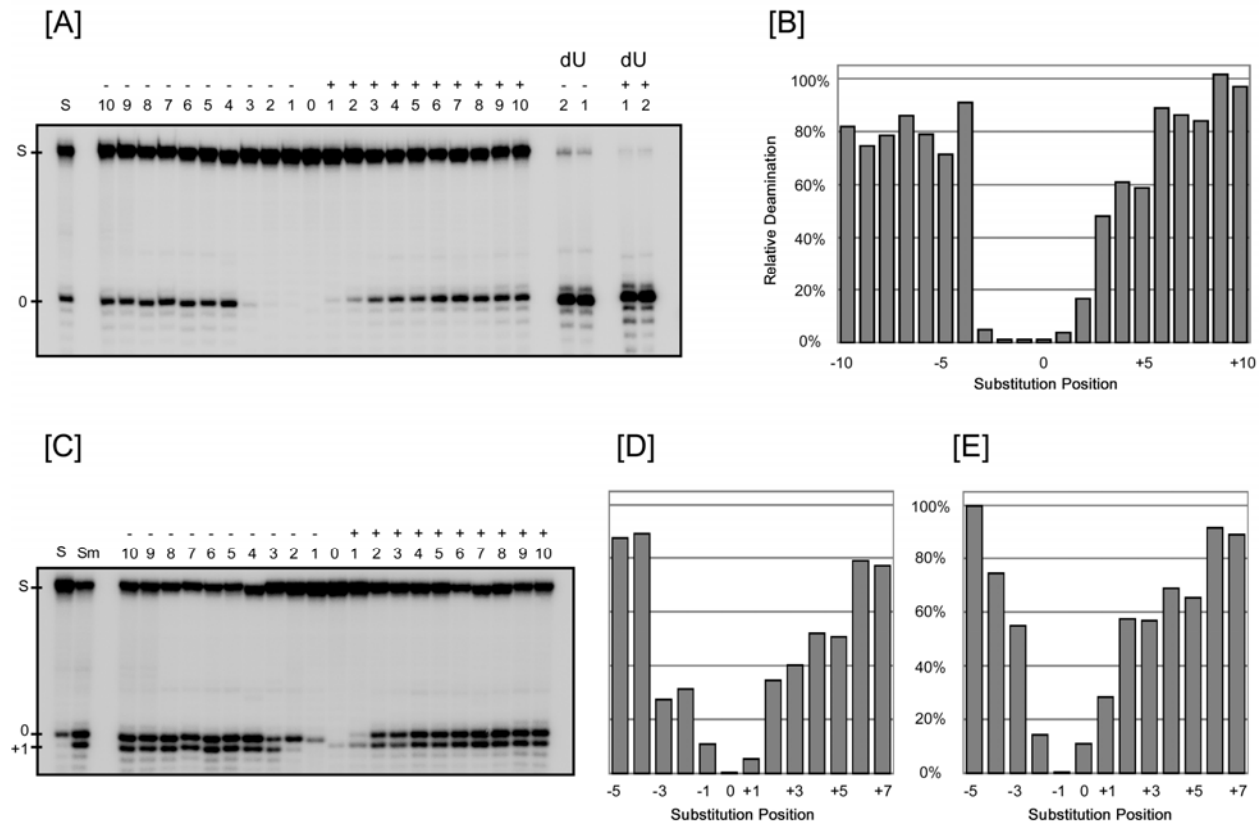


Figure 5

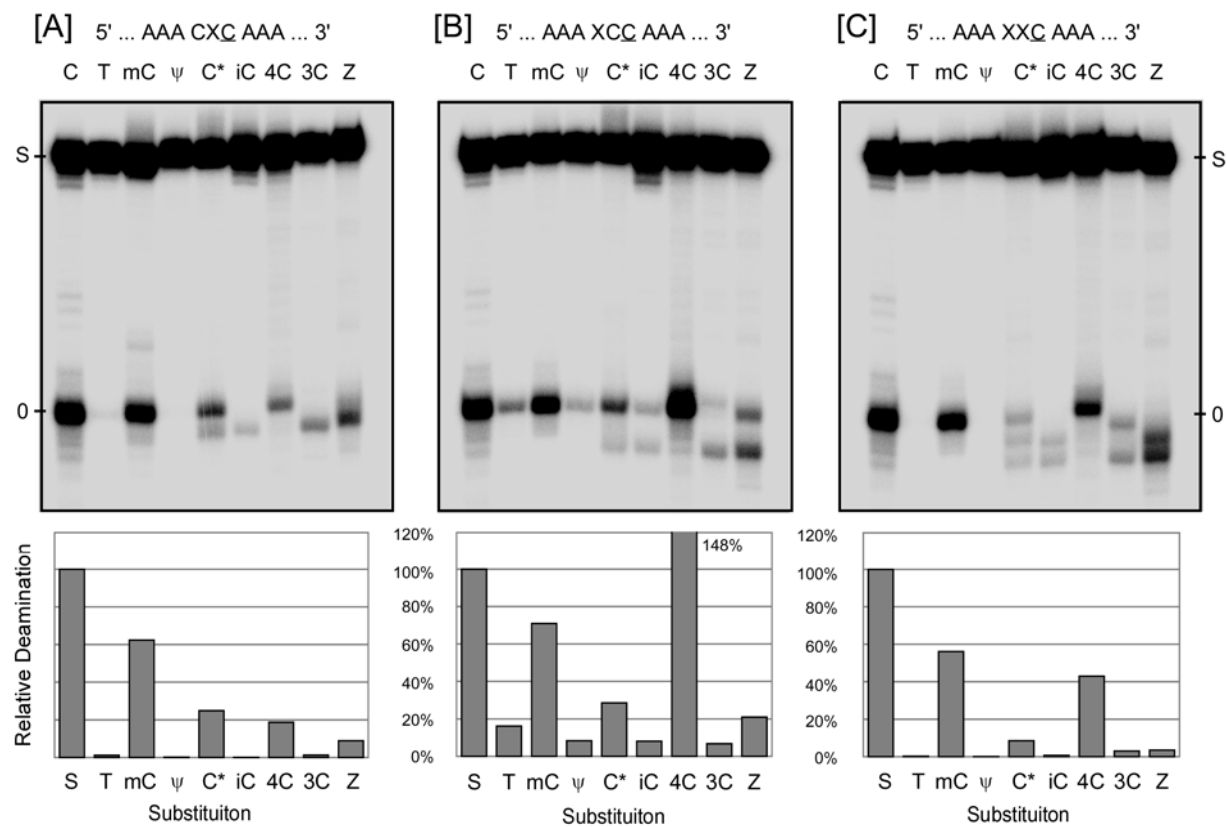


Figure 6

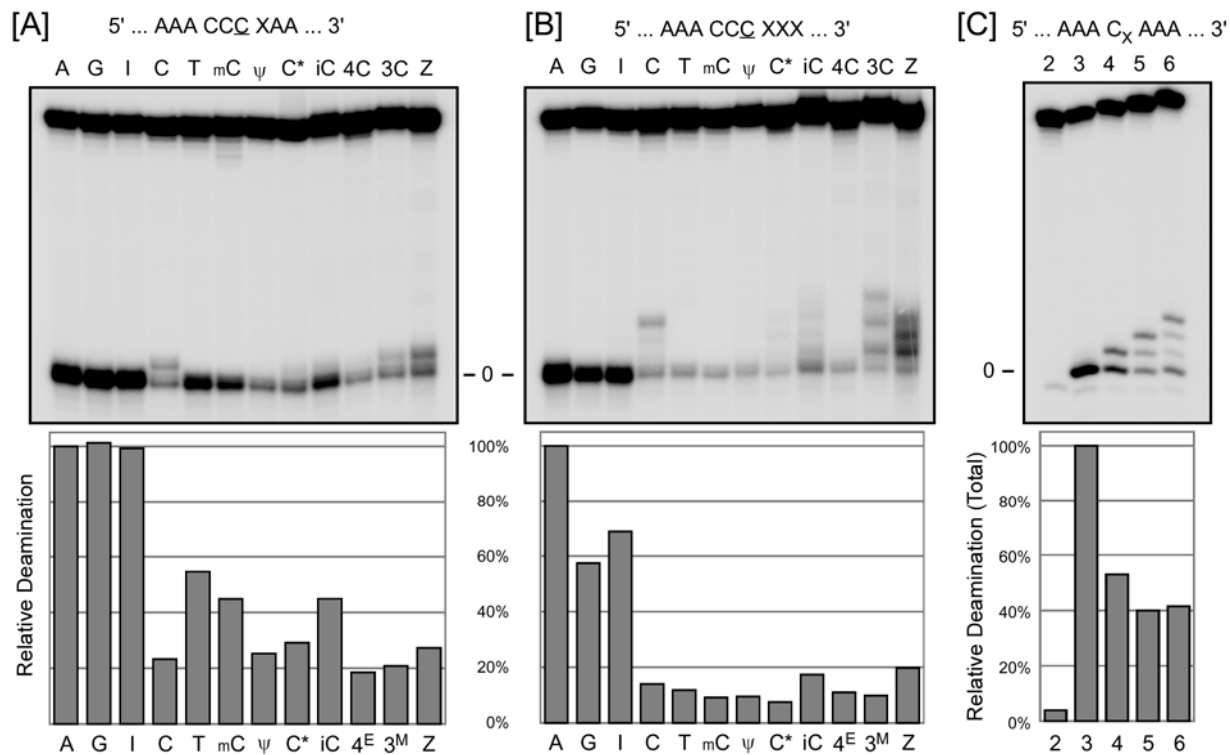


Table 1

Pyrimidine Analog	Ring Position						% Activity (analog at -1)
	1	2	3	4	5	6	
2'-deoxycytidine	$\begin{array}{c} -N- \\ \end{array}$	$O=C-$	$\begin{array}{c} N= \\ \end{array}$	$\begin{array}{c} NH_2 \\ \\ =C- \end{array}$	$\begin{array}{c} -C-H \\ \end{array}$	$\begin{array}{c} \\ -C-H \end{array}$	100%
5'-methyl deoxycytidine	$\begin{array}{c} -N- \\ \end{array}$	$O=C-$	$\begin{array}{c} N= \\ \end{array}$	$\begin{array}{c} NH_2 \\ \\ =C- \end{array}$	$\begin{array}{c} -C-CH_3 \\ \end{array}$	$\begin{array}{c} \\ -C-H \end{array}$	62%
5-aza-5,6-dihydro dC	$\begin{array}{c} -N- \\ \end{array}$	$O=C-$	$\begin{array}{c} N= \\ \end{array}$	$\begin{array}{c} NH_2 \\ \\ =C- \end{array}$	$\begin{array}{c} -N-H \\ \end{array}$	$\begin{array}{c} \\ -C-H \\ \\ H \end{array}$	25%
N4-ethyl 2'-deoxycytidine	$\begin{array}{c} -N- \\ \end{array}$	$O=C-$	$\begin{array}{c} N= \\ \end{array}$	$\begin{array}{c} HNEt \\ \\ =C- \end{array}$	$\begin{array}{c} -C-H \\ \end{array}$	$\begin{array}{c} \\ -C-H \end{array}$	19%
2'-deoxy Zebularine	$\begin{array}{c} -N- \\ \end{array}$	$O=C-$	$\begin{array}{c} N= \\ \end{array}$	$\begin{array}{c} H \\ \\ =C- \end{array}$	$\begin{array}{c} -C-CH_3 \\ \end{array}$	$\begin{array}{c} \\ -C-H \end{array}$	9%
3-methyl 2'-deoxycytidine	$\begin{array}{c} -N- \\ \end{array}$	$O=C-$	H_3C-N-	$\begin{array}{c} NH_2 \\ \\ =C- \end{array}$	$\begin{array}{c} -C-H \\ \end{array}$	$\begin{array}{c} \\ -C-H \end{array}$	<2%
thymidine	$\begin{array}{c} -N- \\ \end{array}$	$O=C-$	$H-N-$	$\begin{array}{c} O \\ \\ -C- \end{array}$	$\begin{array}{c} -C-CH_3 \\ \end{array}$	$\begin{array}{c} \\ -C-H \end{array}$	<2%
2'-deoxypseudouridine	$\begin{array}{c} -C= \\ \end{array}$	$O=C-$	$H-N-$	$\begin{array}{c} O \\ \\ -C- \end{array}$	$\begin{array}{c} -N-H \\ \end{array}$	$\begin{array}{c} \\ =C-H \end{array}$	<2%
2'-deoxyisocytidine	$\begin{array}{c} -N- \\ \end{array}$	$H_2N-C=$	$\begin{array}{c} N= \\ \end{array}$	$\begin{array}{c} O \\ \\ -C- \end{array}$	$\begin{array}{c} -C-H \\ \end{array}$	$\begin{array}{c} \\ -C-H \end{array}$	<2%
2'-deoxyuridine	$\begin{array}{c} -N- \\ \end{array}$	$O=C-$	$H-N-$	$\begin{array}{c} O \\ \\ -C- \end{array}$	$\begin{array}{c} -C-H \\ \end{array}$	$\begin{array}{c} \\ -C-H \end{array}$	ND

