FEBS Letters 581 (2007) 4761–4766

Extensive mutagenesis experiments corroborate a structural model for the DNA deaminase domain of APOBEC3G

Kuan-Ming Chen[a,b,1], Natalia Martemyanova[a,b,c,1], Yongjian Lu[a,b], Keisuke Shindo[a,b,c], Hiroshi Matsuo[a,b,*], Reuben S. Harris[a,b,c,*]

[a] *University of Minnesota, Department of Biochemistry, Molecular Biology and Biophysics, Minneapolis, MN 55455, United States*
[b] *University of Minnesota, Institute for Molecular Virology, Minneapolis, MN 55455, United States*
[c] *University of Minnesota, Beckman Center for Transposon Research, Minneapolis, MN 55455, United States*

Received 10 July 2007; revised 29 August 2007; accepted 30 August 2007

Available online 7 September 2007

Edited by Hans Eklund

Abstract APOBEC3G is a single-strand DNA cytosine deaminase capable of blocking retrovirus and retrotransposon replication. APOBEC3G has two conserved zinc-coordinating motifs but only one is required for catalysis. Here, deletion analyses revealed that the minimal catalytic domain consists of residues 198-384. Size exclusion assays indicated that this protein is monomeric. Many (31/69) alanine substitution derivatives of APOBEC3G198-384 retained significant to full levels of activity. These data corroborated an APOBEC2-based structural model for the catalytic domain of APOBEC3G indicating that most non-essential residues are solvent accessible and most essential residues cluster within the protein core.

Keywords: APOBEC3G; DNA cytosine deamination; DNA editing; Hypermutation

1. Introduction

Human APOBEC family members are important mediators of adaptive and innate immune responses (reviewed by [1,2]). These proteins are defined by a highly conserved zinc-coordinating motif, HXE-$X_{23–28}$-$CX_{2–4}$C, in which the histidine and the two cysteines position zinc and the glutamate positions water to promote the nucleophilic deamination of cytosines within single-stranded, polynucleotide substrates (usually DNA). One family member, APOBEC3G (A3G), was identified as a cellular protein capable of blocking the replication of virion infectivity factor (Vif)-defective HIV-1 [3]. Independent work showed that A3G was capable of DNA cytosine deamination [4]. Subsequent studies linked these two activities by demonstrating that A3G could inhibit the replication of HIV-1 and other retroviruses by deaminating viral cDNA cytosines to uracils during reverse transcription [5–7]. Uracils template the incorporation of adenines during the synthesis of the complementary viral DNA strand, and subsequent replication (or DNA repair) ultimately produces strand-specific C/G to T/A transition mutations (hypermutations).

The deaminase activity of A3G has been studied genetically and biochemically [5,8–10]. It has been mapped to the conserved C-terminal zinc-coordinating domain [9–12]. The local dinucleotide deamination preference of A3G (5′-CC) is readily observed in AIDS patient-derived HIV-1 DNA sequences as viral genomic strand 5′-GG to 5′-AG hypermutations (cDNA strand 5′-CC to 5′-CT transition mutations; e.g. [13]). Here, an extensive mutational analysis of A3G was performed to delineate the minimal region required for catalysis, to define non-essential (and essential) amino acids and to corroborate structure-based predictions.

2. Materials and methods

2.1. Plasmid constructs

The A3G cDNA used here matches NM_021822. A3G and mutant derivatives were expressed as GST fusion proteins using pGEX6P1 or pGEX6P2 (GE Healthsciences). An EcoRI–SalI DNA fragment from pTrc99A-A3G encoding full-length A3G [9] was sub-cloned directly into pGEX6P1. A3G deletion mutants were constructed by amplifying the relevant A3G coding regions, digesting the resulting PCR products with SmaI and SalI and ligating them into the SmaI and XhoI sites of pGEX6P2. Alanine mutants were constructed using the QuikChange protocol (Stratagene). All constructs were verified by DNA sequencing. An oligonucleotide table is available online.

2.2. Escherichia coli mutation assays

The *E. coli*-based, rifampicin-resistance (Rif^R) mutation assay has been used extensively to monitor the intrinsic DNA cytosine deaminase activity of several A3 proteins including A3G (e.g. [4,9]). Aliquots of saturated overnight cultures (LB plus 200 μg/mL ampicillin) were spread onto LB plates containing 100 μg/mL rifampicin to select for Rif^R mutants and diluted aliquots were spread onto LB plates containing 200 μg/mL ampicillin to determine the number of viable cells. Mutation frequencies were calculated by dividing the number of Rif^R mutants by the number of viable cells in each culture. For the deletion experiments, the mutation frequencies of eight individual cultures were plotted and the median indicated. For the alanine mutant experiments, the Rif^R mutation frequency for each construct was determined by assaying the median Rif^R mutation frequency for 8–12 independent cultures, calculating the fold-difference relative to the median value of the vector control cultures and averaging at least two (and up to 12) independent experiments.

*Corresponding authors. Address: University of Minnesota, Department of Biochemistry, Molecular Biology and Biophysics, Minneapolis, MN 55455, United States. Fax: +1 612 625 2163.
E-mail addresses: chen0775@umn.edu (K.-M. Chen), natasha_mart@yahoo.com (N. Martemyanova), luxxx079@umn.edu (Y. Lu), matsu029@umn.edu (H. Matsuo), rsh@umn.edu (R.S. Harris).

[1]Equal contributions.

Abbreviation: APOBEC3G, apolipoprotein BmRNA editing catalytic subunit-like protein 3G

0014-5793/$32.00
doi:10.1016/j.febslet.2007.08.076

2.3. GST-A3G expression, purification and size exclusion procedures

GST-A3G protein levels were monitored by expressing them in *E. coli* BL21 DE3 RIL (Stratagene), sonicating the cells in lysis buffer (100 mM NaCl, 50 mM Na_2HPO_4/NaH_2PO_4 [pH7.0], protease inhibitor [Roche]), separating the soluble (supernatant) and insoluble (pellet) fractions by centrifugation (12 110 × *g*, 20 min, 4 °C) and fractionating aliquots of the resulting proteins by SDS–PAGE. Proteins were detected by coomassie blue and quantified by Image J software (http://rsb.info.nih.gov/ij/). Immunoblots were done with antibodies from GE Healthcare (anti-GST) and from J. Lingappa (anti-A3G; [11]) For size exclusion experiments, GST-A3G198-384 was bound to glutathione sepharose, washed several times with lysis buffer, eluted with PreScission protease (GE Healthcare) in 1 mM DTT, 50 mM Na_2HPO_4/NaH_2PO_4 [pH 7.0] buffer, quantified and immediately 1 mg (approx. 1 mL) was passed through a Superdex 75 size exclusion column (GE Healthcare) and detected in fractions by UV (absorbance 280). GST and lysozyme were purchased from Sigma.

2.4. Structural modeling

A3G198–384 and APOBEC2 primary amino acid sequences were aligned with the homology modeling module of the InsightII program (Accelrys) (Online Fig. S1). Secondary structural elements of A3G198–384 were predicted using the HNN program [14,15]. A model was generated by fitting these elements (allowing for gaps) to those in the actual structure of APOBEC2 (PDB 2YNT; [16]). No gross differences in secondary structure were observed. The homology modeling module of the InsightII program (Accelrys) was used to construct and refine the model by minimizing energetically unfavorable amino acid side chain interactions.

3. Results and discussion

3.1. Residues 198-384 of APOBEC3G are sufficient for DNA deamination

Chimeric APOBEC3 proteins and site-directed mutants have been used to demonstrate that the intrinsic DNA cytosine deaminase activity of human APOBEC3G resides in the conserved C-terminal zinc-coordinating domain [9–12]. To further delineate the minimal domain required for catalysis, nine A3G deletion constructs were tested for mutability in the *E. coli*-based Rif^R mutation assay (Fig. 1). Bacteria expressing GST showed a median of 2.5 Rif^R mutants per 10^7 viable cells. Expression of full-length A3G fused to GST caused a 4.4-fold increase in mutation frequency. Apart from two notable exceptions, all of the A3G deletion constructs were inactive. The inactivity of constructs lacking 22 or 40 C-terminal residues was consistent with prior work showing that A3G lacking 38 or 89 C-terminal amino acids was unable to inhibit HIVΔvif [17,18]. A more interesting result emerged from analyses of N-terminal deletion set. A3G variants encoding residues 175-384 or 198-384, but not 215-384, were considerably more mutable than full-length A3G. These data demonstrated that the entire N-terminal, zinc-binding domain is dispensable for activity, and only A3G residues 198-384 are required.

3.2. Size exclusion experiments indicate that A3G198–384 is monomeric

Previous studies indicated that an A3G C97A mutant was incapable of co-immunoprecipitating wildtype A3G, but was still capable of DNA deamination [12,19]. Consistent with these studies, A3G198–384 profiled as a 22 kDa monomer in size exclusion experiments, migrating clearly between the elution positions of lysozyme (14 kDa) and GST (25 kDa) (Fig. 1C). It is not likely that stable oligomeric forms of A3G198–384 exist, because fractions 1–27 did not contain protein peaks.

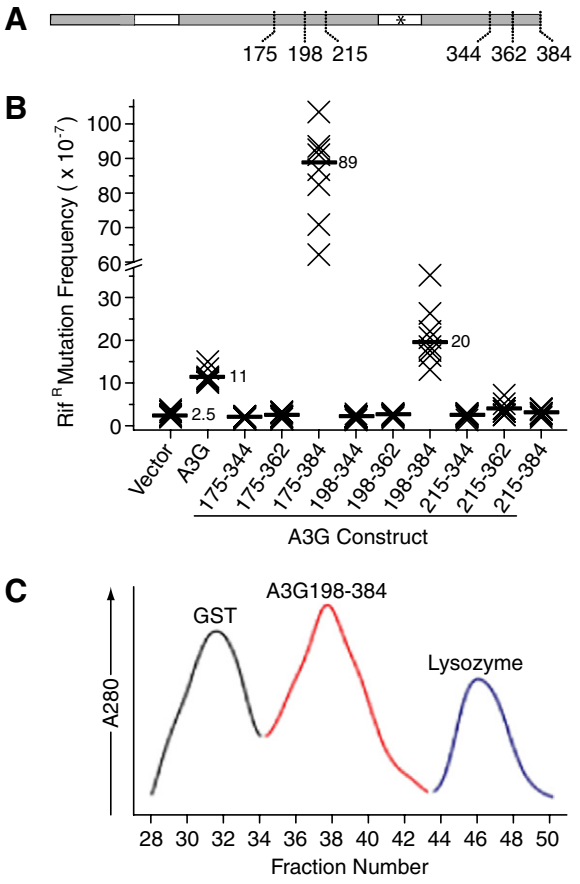

Fig. 1. APOBEC3G deletion mutants delineate a minimal active domain. (A) An illustration showing the amino acid boundaries used for deletion constructs. The HXE-$X_{23–28}$-$CX_{2–4}$C motifs are depicted by open boxes, and the asterisk designates the catalytic domain. (B) Rif^R mutation frequencies of the indicated GST-A3G constructs. Each X represents the mutation frequency of an individual culture ($n = 8$ per construct). The median mutation frequency for cells expressing the vector control, A3G, A3G175–384 and A3G198–384 is indicated. (C) Size exclusion profiles of GST (25 kDa), A3G198–384 (22 kDa) and lysozyme (14 kDa) indicate that A3G198–384 is monomeric.

3.3. Alanine mutations define essential and non-essential residues of A3G198–384

To more precisely delineate the residues and domains required for DNA deamination, the A3G198–384 variant was used to construct a series of 69 alanine mutants (Fig. 2). Mutagenesis was concentrated to hydrophobic residues and cysteines. This strategy was motivated partly by the likelihood that some of the hydrophobic amino acids would likely be important structurally (perhaps forming part of the enzyme core) whereas, more intriguingly, others would be positioned on the surface of the protein (perhaps defining interaction sites). We also envisaged that mutating select hydrophobic residues might help overcome the solubility issues of A3G and other family members [8,10,19]. All 69 mutants were analyzed using the *E. coli*-based Rif^R mutation assay, because *in vitro* experiments were hindered by the fact that A3G198–384 frequently precipitated during biochemical purification and invariably during long-term storage.

Twelve independent Rif^R experiments, each with at least 10 constructs (and 8–12 independent cultures per mutant), were required to analyze 69 alanine mutant derivatives of A3G198–384 (Fig. 2). It was not feasible to simultaneously