doi:10.1016/j.jmb.2009.04.031 J. Mol. Biol. (2009) 389, 819–832

JMB

Available online at www.sciencedirect.com

ScienceDirect

ELSEVIER

An Extended Structure of the APOBEC3G Catalytic Domain Suggests a Unique Holoenzyme Model

Elena Harjes[1,2], Phillip J. Gross[1,2,3], Kuan-Ming Chen[1,2], Yongjian Lu[1], Keisuke Shindo[1,2,3], Roni Nowarski[4], John D. Gross[5], Moshe Kotler[4], Reuben S. Harris[1,2,3]* and Hiroshi Matsuo[1,2]*

[1]*Department of Biochemistry, Molecular Biology, and Biophysics, University of Minnesota, Minneapolis, MN 55455, USA*

[2]*Institute for Molecular Virology, University of Minnesota, Minneapolis, MN 55455, USA*

[3]*Center for Genome Engineering, University of Minnesota, Minneapolis, MN 55455, USA*

[4]*Department of Pathology and the Lautenberg Center for General and Tumor Immunology, Hebrew University Hadassah Medical School, Jerusalem 91120, Israel*

[5]*Department of Pharmaceutical Chemistry, University of California San Francisco, 600 16th Street, San Francisco, CA 94107, USA*

Received 30 January 2009; received in revised form 15 April 2009; accepted 16 April 2009
Available online 21 April 2009

Edited by M. F. Summers

Human APOBEC3G (A3G) belongs to a family of polynucleotide cytidine deaminases. This family includes APOBEC1 and AID, which edit *APOB* mRNA and antibody gene DNA, respectively. A3G deaminates cytidines to uridines in single-strand DNA and inhibits the replication of human immunodeficiency virus-1, other retroviruses, and retrotransposons. Although the mechanism of A3G-catalyzed DNA deamination has been investigated genetically and biochemically, atomic details are just starting to emerge. Here, we compare the DNA cytidine deaminase activities and NMR structures of two A3G catalytic domain constructs. The longer A3G191-384 protein is considerably more active than the shorter A3G198-384 variant. The longer structure has an α1-helix (residues 201–206) that was not apparent in the shorter protein, and it contributes to catalytic activity through interactions with hydrophobic core structures (β1, β3, α5, and α6). Both A3G catalytic domain solution structures have a discontinuous β2 region that is clearly different from the continuous β2 strand of another family member, APOBEC2. In addition, the longer A3G191-384 structure revealed part of the N-terminal pseudo-catalytic domain, including the interdomain linker and some of the last α-helix. These structured residues (residues 191–196) enabled a novel full-length A3G model by providing physical overlap between the N-terminal pseudo-catalytic domain and the new C-terminal catalytic domain structure. Contrary to predictions, this structurally constrained model suggested that the two domains are tethered by structured residues and that the N- and C-terminal β2 regions are too distant from each other to participate in this interaction.

Keywords: APOBEC3G; DNA deamination; DNA editing; NMR structure; retrovirus restriction

**Corresponding authors.* Department of Biochemistry, Molecular Biology, and Biophysics, University of Minnesota, 321 Church Street, SE, 6-155 Jackson Hall, Minneapolis, MN 55455, USA. E-mail addresses: rsh@umn.edu; matsu029@umn.edu.

Abbreviations used: A3G, APOBEC3G; A2, APOBEC2; HIV-1, human immunodeficiency virus-1; Rif^R, rifampicin resistant; *rpoB*, RNA polymerase B; GST, glutathione *S*-transferase; ssDNA, single-strand DNA; NOE, nuclear Overhauser enhancement; GFP, green fluorescent protein; PVDF, polyvinylidene fluoride; NOESY, NOE spectroscopy; HSQC, heteronuclear single quantum coherence; PDB, Protein Data Bank.

Introduction

Human APOBEC3G (A3G) is a prominent member of a multifunctional family of Zn^{2+}-dependent polynucleotide cytidine deaminases (reviewed recently by Chiu and Greene,[1] Goila-Gaur and Strebel,[2] and Malim and Emerman[3]). Members include the founder APOBEC1, which edits *APOB* mRNA C6666 to generate an early stop codon and a shorter polypeptide, and AID, which edits immunoglobulin gene DNA cytidines to trigger somatic hypermutation and class switch recombination. Physiological roles have yet to be revealed for other members such as APOBEC2 (A2) and APOBEC4. However, A3G, six other human A3s, and a multitude of A3s from other mammals have elicited DNA cytidine deaminase activity and have been shown to inhibit the replication of a variety of retrotransposons and retroviruses. The most prominent human pathogen that can be restricted by A3G is the AIDS virus human immunodeficiency virus-1 (HIV-1).

A defining feature of DNA cytidine deaminases is a Zn^{2+}-coordinating motif, H-x_1-E-$x_{25–31}$-C-$x_{2–4}$-C.[4–6] This motif is required for catalytic activity (e.g., Refs. 7–14). Based on the structures of free-base and nucleoside deaminases, histidine, two cysteines, and water are thought to directly coordinate Zn^{2+}, and glutamate exchanges hydrogens during catalysis.[15–17] Some APOBEC3 proteins have two Zn^{2+}-binding motifs, whereas others have only one. In A3G, the C-terminal Zn^{2+}-binding motif is catalytically active, whereas the N-terminal Zn^{2+}-binding motif is not.[8,14,18,19]

High-resolution structural information was reported recently for the C-terminal catalytic domain of A3G[20–22] and for residues 41–224 of the single Zn^{2+}-domain protein A2,[23] which lacks deaminase activity (e.g., Refs. 23 and 24). Both are globular proteins with mostly similar secondary structures and Zn^{2+}-coordinating motifs. However, structural and biochemical studies of A2 have shown that it exists as a homodimer mediated by extensive antiparallel β2–β2 contacts.[23] In contrast, the catalytic domain of A3G profiled as a monomer in solution and showed no obvious dimeric points of contact in the crystal,[20–22,25] despite the fact that the A3G holoenzyme is capable of dimerizing and forming higher-order oligomers.[14,26–29] The A2 data therefore suggested that the analogous β2 regions of the N- and C-terminal domains of A3G might mediate interdomain connection.[21–23,25,30] However, the feasibility of such an A3G β2–β2 interaction has been debated because the A3G catalytic domain NMR structures have shown a discontinuous β2–bulge–β2′ region,[20,21] whereas the A3G197-380 crystal structure has been reported to have an A2-like continuous β2 strand.[22]

Here, we report the activity and solution structure of a longer A3G catalytic domain variant that spans residues 191–384. These data provide several novel insights and confirm aspects of prior NMR and crystallographic studies. First, the new structure substantiated the discontinuous β2–bulge–β2′ structure observed by NMR, but not by crystallography. Second, the new structure confirmed the presence and demonstrated the importance of the α1-helix, which was apparent in the crystal, but not previously in solution. Third, a comparison of the A3G191-384 and A3G198-384 NMR structures yielded a robust explanation for the increased activity of the longer construct (substantive α1 to the catalytic core structure interactions). Fourth, the new A3G191-384 solution structure revealed a region of physical overlap—one helical turn (residues 191–194)—between the predicted N-terminal pseudo-catalytic domain (A3G1-196) and the actual C-terminal catalytic domain (A3G191-384). This overlap enabled the construction of a novel full-length A3G model. Fifth, the physical constraints imposed by the new structural elements orientated the N- and C-terminal β2 regions too far away from each other to facilitate interdomain interactions. In this report, we present the structure of the longest fragment of A3G to date, and we address some of the controversies raised by previously published structures. All together, our data provide several important steps toward a robust structural understanding of full-length A3G and its central role in retroelement restriction, and they help broaden our structural and mechanistic knowledge of DNA deaminases.

Results

A longer and more active A3G catalytic domain variant

As outlined in the Introduction, there are clear differences between the NMR and the crystal structures of the catalytic domain of A3G, and none of these has shed light on A3G interdomain connections.

To better understand the A3G catalytic domain and how it is oriented with respect to the N-terminal pseudo-catalytic domain, we revisited prior data showing that A3G175-384 was approximately threefold more active than A3G198-384 in an *Escherichia-coli*-based rifampicin-resistant (Rif^R) mutation assay.[25] The Rif^R mutation assay provides a quantitative but indirect read-out of DNA deaminase activity, with each Rif^R colony reflecting an RNA polymerase B (*rpoB*) gene mutation that occurred during the outgrowth of a single cell into a saturated culture. To better define the residues responsible for this elevated mutation frequency, we examined a series of single-alanine mutants and observed that residues 175–190 were dispensable (data not shown).

We next generated untagged and glutathione *S*-transferase (GST)-tagged A3G191-384 and A3G198-384 constructs for extensive head-to-head genetic, biochemical, and structural characterizations. Derivatives with five amino acid substitutions, L234K, C243A, F310K, C321A, and C356A (2K3A), were also made to enable concentrated