Nanostructures of APOBEC3G Support a Hierarchical Assembly Model of High Molecular Mass Ribonucleoprotein Particles from Dimeric Subunits*§

Received for publication, September 19, 2006, and in revised form, October 23, 2006
Published, JBC Papers in Press, October 31, 2006, DOI 10.1074/jbc.C600253200
Joseph E. Wedekind 1, Richard Gillilan 1, Alena Janda 1, Jolanta Krucinska 1, Jason D. Salter 1, Ryan P. Bennett 1 2, Jay Raina 3, and Harold C. Smith 1 2

From the 1Department of Biochemistry and Biophysics, University of Rochester School of Medicine and Dentistry, Rochester, New York 14642, the 2Macromolecular Structure Facility at the Cornell High-Energy Synchrotron Source (MacCHESS), Cornell University, Ithaca, New York 14853, and 3Immunodiagnostics Inc., Woburn, Massachusetts 01801

Human APOBEC3G (hA3G) is a cytidine deaminase that restricts human immunodeficiency virus (HIV)-1 infection in a virion (the virion infectivity factor from HIV)-dependent manner. hA3G from HIV-permissive activated CD4 + T-cells exists as an inactive, high molecular mass (HMM) complex that can be transformed in vitro into an active, low molecular mass (LMM) variant comparable with that of HIV-non-permissive CD4 + T-cells. Here we present low resolution structures of hA3G in HMM and LMM forms determined by small angle x-ray scattering and advanced shape reconstruction methods. The results show that LMM particles have an extended shape, dissimilar to known cytidine deaminases, featuring novel tail-to-tail dimerization. Shape analysis of LMM and HMM structures revealed how symmetric association of dimers could lead to minimal HMM variants. These observations imply that the disruption of cellular HMM particles may require regulation of protein-RNA, as well as protein-protein interactions, which has implications for therapeutic development.

hA3G 5 is an anti-retroviral host defense factor that restricts HIV infection by vif (the virion infectivity factor from HIV)-deficient viral strains (1). hA3G is packaged into HIV-1 virions (2, 3) and causes extensive 2′-deoxyctydine to 2′-deoxyuridine mutations of minus polarity viral DNA during reverse transcription (4). Such “DNA editing” results in extensive 2′-deoxycytosine to 2′-deoxyadenosine changes in the viral cDNA that contribute to reduced HIV infectivity (5–8). However, a deaminase-independent anti-viral mechanism exists as well (9) that may entail RNA binding (10). Although hA3G does not edit RNA, it exhibits general RNA binding properties (11–13). The principal form of hA3G in HIV infection-permissive CD4 + cells of lymphoid tissues is an HMM ribonucleoprotein complex with little or no deaminase activity (14, 15). In contrast, an enzymatically active, LMM form of hA3G predominates in peripheral blood CD4 + cells and serves as a potent post-entry HIV restriction factor (14, 15). Activation of such cells recruits the LMM enzyme into HMM complexes rendering the cell permissive to infection (15). In vitro treatment of HMM hA3G with RNase or in vivo exposure to interferon produces the enzymatically active LMM form suggesting anti-viral activity involves a delicate interplay governed by RNA-protein interactions (14, 16, 17).

hA3G belongs to the family of APOBEC-1 related proteins characterized by a ZDD fold featuring the consensus sequence (Cys/His)-X-Glu-X 25–30-Pro-Cys-XX-Cys, where “X” is any amino acid (18). Although homology models have been generated for some APOBEC-1 family members (19, 20), and the hA3G secondary structure has been predicted (18, 21), no empirical structural information exists for it or any other member of the APOBEC family. Modeling of the hA3G structure based upon known dimeric or tetrameric CDAs (20, 22–24) is complicated by the fact that the protein arose from a novel gene duplication of the fundamental ZDD motif such that tandem active sites are present in each subunit (18). To provide insight into the fundamental physical properties of hA3G in relation to known cytidine deaminase structures, as well as how hA3G oligomerization contributes to retroviral restriction, we undertook a solution SAXS analysis of the recombinant enzyme in its HMM and LMM forms.

EXPERIMENTAL PROCEDURES

Preparation of hA3G—Full-length hA3G cDNA was amplified from oligo(dT)-primed H9 cell RNA and a four His tag (His 4) was added to the C terminus by PCR. This construct was subcloned into pFastbac™ (Invitrogen, CA). Baculovirus production and infection of Sf9 cell cultures for expression were carried out by Immunodiagnosticics, Inc.

Frozen cells (4 g) were lysed in 20 ml of 0.5× hA3G buffer (1× = 50 mm HEPPS, pH 8.8, 75 mm NaCl, 10 mm MgCl 2, 5% (v/v) glycerol, 0.2 mm β-mercaptoethanol, and EDTA-free complete protease inhibitor (Roche Applied Science)) by freezing in N 2 (f) and thawing followed by shearing via successive
passes through 22- and 26-gauge needles. The lysis solution was brought to 1% (v/v) Triton X-100 and made 0.1 mM in CaCl₂. Nuclease digestion ensued with either 0.125 mg ml⁻¹ RNase-free DNase I (Sigma) (hereafter this protein is referred to as hA₃G-D) or 0.125 mg ml⁻¹ DNase I and 0.25 mg ml⁻¹ RNase A (Sigma) (hA₃G-DR) at 37 °C for 30 min. The sample was brought to 1 M urea final concentration, incubated at 24 °C for 20 min, and centrifuged (10,000 × g for 10 min at 24 °C). Cleared supernatants were adsorbed onto 2 ml nickel-nitrilotriacetic acid-agarose (Qiagen) and mixed for 2 h at 24 °C. Contaminants were removed by centrifugation of resin (500 × g for 20 min, and centrifuged (10,000 × g for 10 min at 24 °C). Elution was monitored at 280 nm. Pure fractions were identified and pooled based on SDS-PAGE gels stained with Coomassie Blue dye; estimated purity was >99%.

Nuclease digestion ensued with either 0.125 mg ml⁻¹ RNase A (Sigma) (hA₃G-DR) at 37 °C for 30 min. The sample was brought to 1% (v/v) Triton X-100 and made 0.1 mM in CaCl₂. Protein concentrations were 0.9 mg ml⁻¹ for hA₃G-D, whereas hA₃G-DR produced a specific activity of 30 pmol µg⁻¹ min⁻¹. These activity trends are consistent with those reported (16).

SAXS Experiments—Scattering experiments were performed at beamline G1 of CHESS (Ithaca, NY). Scattered X-rays were recorded on a custom 1024 × 1024 (69.78 µm) pixel CCD detector fabricated by the Gruner group (Cornell University, Ithaca, NY). Scattering was performed at 20 °C at a sample-to-detector distance of 138.0 cm. The wavelength, λ, was 1.249 Å, which produced an accessible q-range from 0.012 to 0.215 Å⁻¹, where q = 4 sinθ/λ (θ is the scattering angle). Samples of hA₃G-DR were prepared at various concentrations in 1 × hA₃G buffer containing 0.25 M imidazole. Protein concentrations were 0.9 mg ml⁻¹ and 1.8 mg ml⁻¹ for hA₃G-D, and 0.55 mg ml⁻¹ and 1.1 mg ml⁻¹ for hA₃G-DR; lower concentrations were examined as well to assure there was no aggregation. Samples were centrifuged at 14,000 × g and immediately transferred to a homemade cuvette composed of a plastic micromachined disk (A Line Inc., Redondo Beach, CA) fitted with 25 µm mica walls. This cell had a capacity of 12 µl and was loaded through an inlet port with a 25 µl blunt-end syringe (Hamilton Corp., Reno, NV). The x-ray beam size was 0.5 × 0.5 mm², which was significantly smaller than the sample cell window. Exposure times were 2–80 s to assess radiation damage; each exposure was recorded in triplicate. Two-dimensional scattering data were corrected for buffer scatter, CCD dark current, and detector non-uniformity. Ag-Behenate powder (The Gem Dugout, State College, PA) was used to calibrate the beam center and sample-to-detector distances. Two-dimensional scattering data were integrated by Data Squeeze 2.07 (25) yielding a one-dimensional intensity profile as a function of scattering vector q.

Analysis of Reduced Scattering Data—The radius of gyration (Rg) was calculated using the indirect Fourier transform package GNOM (26). The result is a pair-distance distribution function, p(r), in real space that represents an alternative means to calculate Rg compared with traditional Guinier approximations that are produced from low angle q values in which qRg < 1.3 (27). In contrast, GNOM produces an Rg calculated from the full experimental scattering curve and generates a maximum particle dimension (Dmax) as the distance where p(r) reaches zero, which is generally superior to the Guinier approximation (28). The GNOM method relies upon perceptual criteria (26) such that a solution for a compact, globular molecule obeys a smooth, monomodal Gaussian centered at Rg. Goodness-of-fit scores were 0.92 for hA₃G-D (an “excellent” score) and 0.894 (a “good” score) for hA₃G-DR. The molecular mass for each sample was obtained from the respective pair-distance distribution functions by extrapolating to I(q = 0) using GNOM (26).

Ab Initio Structural Modeling—The low resolution molecular envelopes of hA₃G-D and hA₃G-DR were restored from their respective SAXS profiles using DAMMIN (29). In this method, simulated annealing is employed for global minimization, whereby random movements in a multiphase dummy atom model minimize the discrepancy χ between observed and calculated scattering curves. No symmetry constraints were applied to the hA₃G-D restorations. Scattering curves with a q range between 0.021 and 0.17 Å⁻¹ and 0.016 to 0.18 Å⁻¹ were used for hA₃G-D and hA₃G-DR, respectively, corresponding to a resolution range between 300 and 35 Å (2π/qmax). A sphere was chosen as the initial starting model for each molecule, with Dmax derived from the corresponding p(r). For hA₃G-D, a dummy atom packing radius of 8.6 Å was assigned by the program; this radius was 3.75 Å for hA₃G-DR. All calculations were run in “slow” annealing mode. DAMMIN calculations were performed on a 64 node dual processor cluster at MacCHESS (Ithaca, NY). Each restoration required ~20 h of CPU time on a 2.0-GHz 32-bit AMD processor. Ten independent DA models were calculated for hA₃G-D and hA₃G-DR. The 10 models of each class were subjected to automated envelope averaging using DAMAVER (30). Here, each model was compared in a pairwise manner to other models of its class, resulting in a series of NSD values. The model with the lowest NSD was chosen as a reference onto which all other models were fit using SUPCOMB (31). Neither ensemble included outliers based on the NSD criterion. As such, each group of 10 models was included in the calculation of the average envelope. Each of the 10 individual envelopes of a given class (hA₃G-D or hA₃G-DR) was mapped onto a densely packed grid of atoms with each position marked by its own occupancy value. Positions with significant, non-zero occupancies were chosen to produce a final model whose volume was equivalent to the average excluded volume derived from each independent model. It has been noted that final averaged structures from small angle scattering should not be considered a single unique macromolecular conformation in solution (32, 33).

Shape Analysis—to determine whether multiple hA₃G-DR envelopes could fit inside the hA₃G-D particle, the hA₃G-D envelope was moved to the origin and its principal axis of inertia oriented along the z-direction using ALPRAXIN. The hA₃G-DR dimer was then subjected to a six-dimensional
search of the oriented envelope using SUPMON (31). Other volumetric calculations were performed with CRY SOL (34).

Relating Dimeric hA3G-DR to Cytidine Deaminase Crystal Structures—A single CDA domain of yeast CDD1 (Protein Data Bank entry 1R5T) was subjected to a six-dimensional search against the hA3G-DR envelope using COLORES in the SITUS suite (35, 36). Several similar solutions were obtained that differed only by the rotational placement of the CDA monomer into the hA3G-DR envelope. With the first CDA subunit fixed, a second search was conducted to fit the remaining hA3G-DR envelope.

RESULTS AND DISCUSSION

Interpretation of the SAXS Data and Distance Distribution Functions—The SAXS data reveal important physical properties of hA3G that define its global morphology in solution on a nanometer scale. The experimental scattering profiles of pure recombinant hA3G-D (no RNase treatment) and hA3G-DR (RNase treated) are depicted in Fig. 1, A and B. Respective distance distribution functions (Fig. 1, C and D) were calculated by GNOM (26). Both are skewed from an ideal bell-shaped curve characteristic of elongated particles (37, 38). The $p(r)$ for hA3G-D indicates an $R_G$ of 72.4 Å and a maximum molecular dimension ($D_{\text{Max}}$) of 210 Å. The forward scattering $I(0)$ was also calculated by GNOM and corresponds to a molecular mass of 292 kDa. RNase-treated hA3G-DR exhibits a smaller $R_G$ of 45.8 Å with a $D_{\text{Max}}$ of 140 Å; its $I(0)$ corresponds to a molecular mass of 100.6 kDa, consistent with a dimer of hA3G subunits. These values agree with those obtained by dynamic light scattering and/or gel filtration chromatography (supplemental data and supplemental Fig. S2).

Quality of ab Initio Models—Bead models for hA3G-D and hA3G-DR were reconstructed from the experimental SAXS curves in DAMMIN (29). The agreement between an individual ab initio model and the experimental data is indicated by the fit of the model scattering curve with actual data (Fig. 1, A and B). Ten ab initio models each were calculated for hA3G-D and hA3G-DR. The final models exhibited $\chi$ values of 1.2 for hA3G-D and 2.8 for hA3G-DR. The observation that the hA3G-DR MW was consistent with a dimer prompted the use of a P2 symmetry constraint in model calculations; no significant difference in $\chi$ was observed using P1 symmetry. The average shape of each molecule was calculated by superposition of all 10 independent models. An NSD value close to unity indicates good agreement between models, whereas ideally superimposed objects tend toward zero (30, 31).

Descriptions of Average hA3G Models—The hA3G-D shape is an elongated cylinder (Fig. 2A). Three principal domains are
The Global Fold of the hA3G-DR Dimer Is a Novel Structure in Comparison with Known Cytidine Deaminases—hA3G is a ZDD enzyme based on its catalytic activity and amino acid sequence alignment with known CDAs (18). However, its secondary structure content and fold classification have not been analyzed experimentally. Using CD spectroscopy, we demonstrated that (i) hA3G-D and hA3G-DR belong to the α/β fold class, consistent with the CDA family (20), and (ii) the secondary structure content of hA3G-D does not change significantly upon RNase treatment (supplemental Table S1 and Fig. S3). These structural and functional similarities prompted a comparison of the LMM hA3G-DR dimer to the fold of a representative CDA crystal structure, i.e. yeast CDD1 (20). The CDD1 tetramer cannot superpose with either monomeric or dimeric hA3G-DR (Fig. 3B). CDD1, like other CDAs (such as the dimeric enzyme from Escherichia coli), is much more compact than the elongated hA3G-DR structure. These observations support a novel tertiary and quaternary organization for hA3G with implications for other APOBEC3 family members such as 3B and 3F (reviewed in Ref. 18).

Docking of a Minimal CDA Domain into the hA3G-DR Envelope Supports Tail-to-Tail Dimerization—The presence of deaminase activity, α/β secondary structure, and two ZDD signature motifs per polypeptide suggested that the hA3G-DR envelope should accommodate at least two minimal CDA structures per subunit. An automated rigid body search of the hA3G-DR envelope was conducted using a single CDD1 subunit (Fig. 3B, oval inset). A CDD1 monomer was chosen because it exhibits the minimal deaminase fold (~132 amino acids) and is structurally homologous to numerous other deaminases with the ZDD signature sequence (20). The results revealed that two CDA monomers could be accommodated per hA3G-DR subunit with an average correlation coefficient of 0.76 per subunit. The top solutions differed only by rotational placement in the hA3G-DR envelope. For practical considerations, solutions were chosen (Fig. 3B) to orient the C terminus of one CDA domain in proximity to the N terminus of another. As such, the spatial relationship of the domains follows a "large-small-large-small" pattern with envelope volumes of ~18,700 Å³, 7480 Å³, 15,180 Å³, and 6600 Å³. This pattern correlates with the domain organization of the hA3G amino acid sequence, i.e. an N-terminal ZDD motif, a smaller non-catalytic domain, a C-terminal ZDD motif, and a short non-catalytic C-terminal domain (18). The volume of a single CDD1 CDA domain is 17,690 Å³, which agrees well with the larger volumes of the LMM hA3G subunit. Although no high resolution structure exists for the smaller ~55-amino acid non-catalytic domains, these segments occupy volumes of ~7200 Å³ based on amino acid van der Waals radii (40), which closely agrees with the envelope volumes observed here. This result supports a tail-to-tail dimerization model for hA3G (Fig. 3C) rather than a head-to-head (Fig. 3D) or head-to-tail configuration (16). We cannot dismiss the possibility that DNA or RNA binding
induces a conformational change that juxtaposes the N- and C-terminal CDA domains as in CDD1 (Fig. 3B) or other trans-acting CDAs (20). However, the extended tail-to-tail topology explains why each hA3G active site functioned as a monomer, devoid of dominant negative effects characteristic of trans subunit complementation (41). Tail-to-tail organization would also confer unique bidentate substrate affinity and deamination properties. Each solvent exposed N-terminal domain of a subunit would exhibit its established nucleic acid binding properties. Each solvent exposed N-terminal domain of a subunit complementation (41). Tail-to-tail organization would also confer unique bidentate substrate affinity and deamination properties. Each solvent exposed N-terminal domain of a subunit would exhibit its established nucleic acid binding properties. Each solvent exposed N-terminal domain of a subunit would exhibit its established nucleic acid binding properties. Each solvent exposed N-terminal domain of a subunit would exhibit its established nucleic acid binding properties.

Acknowledgments—We thank M. Petoukhov for assistance with ALPRAXIN and SLIPMON and the staff at CHESS beamline G1. We are grateful to the Gruner laboratory for access to the SAXS instrumentation used in this study.

REFERENCES

1. Sheehy, A. M., Gaddis, N. C., Choi, J. D., and Malim, M. H. (2002) Nature 418, 646–650
2. Mariani, R., Chen, D., Schrofelbauer, B., Navarro, F., Konig, R., Bollman, B., Munk, C., Nymark-McMahon, H., and Landau, N. R. (2003) Cell 114, 21–31
3. Stopak, K., de Noronha, C., Yonemoto, W., and Greene, W. C. (2003) Mol. Cell 12, 591–601
4. Yu, Q., Konig, R., Pillai, S., Chiles, K., Kearney, M., Palmer, S., Richman, D., Coffin, J. M., and Landau, N. R. (2004) Nat. Struct. Mol. Biol. 11, 435–442
5. 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) J. Mol. Biol. 27, 695–714
6. Lecossier, D., Bouchonnet, F., Clavel, F., and Hance, A. J. (2003) Science 300, 1112
7. Mangeat, B., Turelli, P., Caron, G., Friedli, M., Perrin, L., and Trono, D. (2003) Nature 424, 99–103
8. Zhang, H., Yang, B., Pomerantz, R. J., Zhang, C., Arunachalam, S. C., and Gao, L. (2003) Nature 424, 94–98
9. Newman, E. N., Holmes, R. K., Craig, H. M., Klein, K. C., Lingappa, J. R., Malim, M. H., and Sheehy, A. M. (2005) Curr. Biol. 15, 166–170
10. Navarro, F., Bollman, B., Chen, H., Konig, R., Yu, Q., Chiles, K., and Landau, N. R. (2005) Virology 333, 374–386
11. Jarmuz, A., Chester, A., Bayliss, J., Gisbourne, J., Dunham, I., Scott, J., and Navaratnam, N. (2002) Genomics 79, 285–296
12. Iwatani, Y., Takeuchi, H., Strebel, K., and Levin, J. G. (2006) J. Virol. 80, 5992–6002
13. Kozak, S. L., Marin, M., Rose, K. M., Bystrom, C., and Kabat, D. (2006) J. Biol. Chem. 281, 29105–29119
14. Chiu, Y., Soros, V. B., Kreisberg, J. F., Stopak, K., Yonemoto, W., and Greene, W. C. (2005) Nature 435, 108–114
15. Kreisberg, J. F., Yonemoto, W., and Greene, W. C. (2006) J. Exp. Med. 203, 865–870
16. Chelico, L., Pham, P., Calabrese, P., and Goodman, M. F. (2006) Nat. Struct. Mol. Biol. 13, 392–399
17. Chen, K., Huang, J., Zhang, C., Huang, S., Nunnari, G., Wang, X. F., Tong, X., Gao, L., Nikisher, K., and Zhang, H. (2006) J. Virol. 80, 7645–7657
18. Wedekind, J. E., Dance, G. S., Sowden, M. P., and Smith, H. C. (2003) Trends Genet. 19, 207–216
19. Navaratnam, N., Fujino, T., Bayliss, J., Jarmuz, A., How, A., Richardson, N., Somasekaram, A., Bhattacharya, S., Carter, C., and Scott, J. (1998) J. Mol. Biol. 275, 695–714
20. Xie, K., Sowden, M. P., Dance, G. S., Torelli, A. T., Smith, H. C., and Wedekind, J. E. (2004) Proc. Natl. Acad. Sci. U. S. A. 101, 8114–8119
21. Huthoff, H., and Malim, M. H. (2005) Virology 334, 147–153
22. Betts, L., Xiang, S., Short, S. A., Wolfenden, R., and Carter, C. W., Jr. (1994) J. Mol. Biol. 235, 635–656
23. Johansson, E., Mejhlvede, N., Neuhard, J., and Larsen, S. (2002) Biochemistry 41, 2563–2570
24. Chung, S. J., Fromme, J. C., and Verdin, G. (2005) Small Angle Scattering of X-rays, John-Wiley & Sons, New York
25. Glatter, O. (1977) J. Appl. Crystallogr. 10, 415–421
26. Svergun, D. I. (1999) Biophys. J. 76, 2879–2886
27. Volkov, V. V., and Svergun, D. I. (2003) J. Appl. Crystallogr. 36, 860–864
28. Kozin, M. B., and Svergun, D. I. (2001) J. Appl. Crystallogr. 34, 33–41
29. Hammel, M., Fierobe, H. P., Czjzek, M., Furuya, K., Smith, J. C., Bayer, E. A., Finet, S., and Reche-Precht, V. (2005) J. Biol. Chem. 280, 38562–38568
30. Johs, A., Hammel, M., Waldner, I., May, R. P., Lagonig, P., and Prassl, R. (2006) J. Biol. Chem. 281, 19732–19739
31. Svergun, D. I., Barberato, C., and Koch, M. H. I. (1995) J. Appl. Crystallogr. 28, 768–773
32. Chacon, P., and Wriggers, W. (2002) J. Mol. Biol. 317, 375–384
33. Wriggers, W., and Chacon, P. (2001) Structure (Camb.) 9, 779–788
34. Feigin, L. A., and Svergun, D. I. (1987) Structure Analysis by Small Angle X-ray and Neutron Scattering, 1st Ed., Plenum Press, New York
35. Petoukhov, M. V., Mouie, T. P., Allain, F. H., Matthews, S., Curry, S., and Svergun, D. I. (2006) Structure (Camb.) 14, 1021–1027
36. Smith, H. C., Wedekind, J. E., Xie, K., and Sowden, M. P. (2005) in Mammalian C to U Editing: Topics in Current Genetics (Grosjean, H., ed.) pp. 1610–2096, Springer-Verlag, Berlin
37. Creighton, T. E. (1993) Proteins: Structures and Molecular Properties, 2nd Ed., W. H. Freeman and Co., New York
38. Opi, S., Takeuchi, H., Kao, S., Khan, M. A., Miyagi, E., Goila-Gaur, R., Iwatani, Y., Levin, J. G., and Strebel, K. (2006) J. Virol. 80, 4673–4682
39. Sowden, M. P., Ballatori, N., Jensen, K. L., Reed, L. H., and Smith, H. C. (2002) J. Cell Sci. 115, 1027–1039
40. Delano, W. L. (2002) PyMol, DeLano Scientific, San Carlos, CA