# SUPPORTING INFORMATION

## A Hydrodynamic Analysis of APOBEC3G Reveals a Monomer-Dimer-Tetramer Self-Association that has Implications for Anti-HIV Function[†]

Jason D. Salter[‡], Jolanta Krucinska[‡], Jay Raina[#], Harold C. Smith[‡,*] and Joseph E. Wedekind[‡,*]

[‡]Department of Biochemistry & Biophysics and Center for RNA Biology, 601 Elmwood Avenue Box 712, Rochester New York 14642, [#]ImmunoDiagnostics Inc., 1 Presidential Way Suite 104, Woburn Massachusetts 01801

*To whom correspondence should be addressed. E-mail: harold.smith@rochester.edu. Phone: (585) 275-4267. Fax: (585) 275-6007. E-mail: joseph.wedekind@rochester.edu Phone: (585) 273-4516. Fax: (585) 275-6007.

[†]Support from the National Institutes of Health (NIH) grant AI076085 to JEW is acknowledged. JDS was supported by pre-doctoral graduate training grant NIH T32 AI49815.

## Supporting Methods

**Immobilized metal-affinity chromatography of A3G**: APOBEC3G was prepared as described (*1*). Briefly, A3G was expressed with a four-His, C-terminal tag in baculovirus-infected Sf9 cells at Immunodiagnostics, Inc (Woburn, MA). Frozen cell pellets were suspended in Buffer A, comprising 50 mM Tris (pH 7.2), 500 mM NaCl, 10 mM $MgCl_2$, 5% (v/v) glycerol, 5 mM β-mercaptoethanol and EDTA-free complete protease inhibitor (Roche Applied Science, Indianapolis, IN). Cells were lysed by successive freezing and thawing cycles followed with shearing by passage through 22 and 26-gauge needles. Nuclease digestion was performed with 0.125 mg $mL^{-1}$ DNase I (Sigma, St. Louis, MO), 0.25 mg $mL^{-1}$ RNase A (Sigma) and 0.4% (v/v) Triton-X 100, at 37 $^oC$ for 30 min. The cell lysate was cleared by centrifugation at 4$^o$ C. NiNTA IMAC was used to purify A3G from cell lysate. NiNTA-immobilized A3G was washed sequentially with buffer A + 1.0 M urea, buffer A + 0.5 M urea, buffer A, buffer A + 0.01 M imidazole, and then buffer A + 0.070 M imidazole. The protein was eluted with buffer A + 0.5 M imidazole.

**Size-exclusion chromatography of A3G**: NiNTA purified A3G was then passed through a Sephacryl S300 (16/60) size exclusion chromatography column (GE Healthcare, Piscataway, NJ) using a Beckman (Fullerton, CA) System Gold 126 HPLC pump equilibrated with Buffer B, which comprised 20 mM Tris pH 8.0, 300 mM KCl, 5% (v/v) glycerol, 10 mM $MgCl_2$ and 1 mM DTT. Protein elution was monitored with a Beckman System Gold 168 UV/visible diode array spectrophotometer (Figure S1). The predominant protein peak eluted with an apparent molecular mass between 98 and 78.5 kDa, which is close to the calculated mass of dimeric A3G-4His (93.9 kD). Only the peak material above 40 mAU was used for AUC. In order to obtain sufficient protein for the AUC investigation, two separate preparations were conducted. Protein was concentrated to 7.35 μM in Buffer B using a centrifugal filtration device (Millipore, Billerica, MA and PALL, Ann Arbor, MI). Protein purity was assessed by SDS-PAGE stained with coomassie blue dye (Figure S1B).

**Analytical Ultracentrifugation of A3G**: Sedimentation velocity experiments were conducted at the University of Connecticut Analytical Ultracentrifugation Facility using a Beckman-Coulter XL-I analytical ultracentrifuge. Experiments were performed simultaneously for three concentrations (1.38 μM, 2.63 μM, 7.35 μM) at 10 $^oC$ and 50,000 RPM in double sector cells equipped with quartz windows. Absorbance at 280 nm was monitored every 4.5 min for a 4.75 hr period. The lower concentration samples were prepared by dilution of a 7.35 μM A3G stock in Buffer B. Physical constants for data analysis were calculated using the known amino acid composition for the protein in the program SEDNERP (*2*). The molecular mass based on the 4His-tagged sequences is 46,957 Da. The partial specific volume, $\nu$ (bar) at 10 $^oC$ was 0.7215 mL $g^{-1}$. The Buffer B density and viscosity were calculated to be 1.02661 g $mL^{-1}$ and 0.01493 poise at 10 $^oC$. The raw sedimentation data were analyzed by use of *dcdt*+ (*3*) and SEDFIT (*4*). Individual data sets were analyzed in *dcdt*+ to obtain the respective normalized concentration profile, g(*s**), and weight-averaged $s_{20,w}$ values including errors reported in Figure 2A, inset. The resulting profiles for each concentration were superimposed in *dcdt*+. SEDFIT was used subsequently to analyze the continuous sedimentation coefficient distribution, c(*s*). Here the broadening effects of diffusion are removed by use of an average frictional coefficient. Data were modeled in continuous sedimentation coefficient mode using the Marquardt-Levenberg fitting option. The

size distribution was analyzed using regularization by maximum entropy. The c(*s*) analysis was run using a point spacing of 0.05 S giving 201 points between 0.1 S and 10.1 S. Typical rmsd values for fitting were below 0.006 with runs testZ values below 20.0. Data were subsequently exported into SEDPHAT (*5*) in order to fit the dimer-tetramer and monomer-dimer equilibrium constants. Similar equilibrium constants were obtained independently by use of SEDANAL (*6*).

**A3G shape modeling for the dimeric and tetrameric species**: Shape analysis of A3G was conducted in SEDNTERP using the experimental $s_{20,w}$ values derived from SEDFIT at the 7.4 μM protein concentration (Figure 2B, inset). Values for the subunit molecular mass, protein partial specific volume, and buffer density were identical to those above. It should be noted that the dimensions of the A3G dimer change as much as 4 to 8% when the $s_{20,w}$ is varied between 5.6 S and 5.5 S, which is the result of the concentration dependence of self-association (Figure 2). This has the effect of making the subunit structure longer and narrower for the prolate and cylindrical models, with the narrowest species appearing at the highest concentration. A similar effect is observed for the tetramer models in which a change of the input $s_{20,w}$ from 7.6 to 8.0 alters the dimensions by as much as 7 to 13%. The oblate models remain considerably flat and do not appear to represent viable shape models. Importantly, the volumes of the prolate and cylindrical models do not change as a function of concentration, which provides an important restraint on the shapes of the quaternary structures (as described below). Overall, the results strongly support elongated shapes for both the dimer and tetramer, although the absolute dimensions of each quaternary species will require additional, high-resolution analysis.

**Rigid-body docking of the A3G C-terminal Domain into the SAXS envelope**: Previously we reported docking of a fundamental cytidine deaminase subunit into the dimeric SAXS molecular envelope (*1*). To demonstrate the compatibility of the restored A3G molecular envelope from our prior studies with recent crystallographic observations, we docked the known A3G C-terminal domain (CD2, residues 197-380) into the restored SAXS molecular envelope from ref. (*1*) using SUPCOMB (*7*). The results reveal that the molecular envelope can accommodate up to four CD2 subunits along the length of the restored shape (Figure 1B). The result is consistent with the assignment of A3G as a dimer by SAXS, but it cannot reveal the molecular orientation of each subunit due to a lack of resolution in the *ab initio* molecular envelope.

**Volumetric analysis of the C-terminal cytidine deaminase of A3G**: To calculate the hydration shell for the known C-terminal A3G subunit (PDB entry 3EIU) reported in ref. (*8*), we utilized the program CRYSOL (*9*). The output value of 30.8 $nm^3$ represents the volume occupied by the protein plus its theoretical hydration shell. All residues of the PDB file were included, which extended from amino acids 197-380 of A3G-CD2 (Figure 1B). The structure of the N-terminal half of A3G (A3G-CD1) is unknown. However, A3G-CD1 (residues 1-196) has a molecular weight (24.3 kDa) similar to the molecular weight of A3G-CD2 (22.1 kDa). In addition, the N-terminal CD1 is 38% identical (53% homologous) to CD2 based on primary structure. As such, the tertiary fold of CD2 is likely to be a reasonable approximation for CD1, which implies similarity in their molecular volumes. Given these considerations, the dimeric A3G model would occupy approximately four times the volume of a single A3G-CD2 domain (i.e. PDB file 3EIU) plus its hydration shell. As such, we calculated that a dimeric A3G species would displace a volume of 123 $nm^3$ whereas a tetramer would displace 246 $nm^3$.

**Calculation of the Hydrodynamic Properties for the A3G SAXS Dimer**: In order to assess the agreement between the prior SAXS dimer (*1*) and the current, experimentally derived hydrodynamic proprieties of the A3G dimer (Figure 2B, inset), we subjected our *ab initio* SAXS bead model to a hydrodynamic analysis. We assumed that the bead model was in water at 20 $^{o}$C, which represents standard conditions. The SAXS-based bead model for the dimer comprises 331 dummy atoms and was analyzed by HYDROPRO (*10*). The radius of each bead was taken to be 0.38 nm as defined by the original *ab initio* shape restoration (*1*). The result provided the $s_{20,w}$ value. The output bead model from HYDROPRO was used as input for SOLPRO (*11*). This program provided the Perrin parameter, $f/f_0$. Unless otherwise mentioned, default values were employed for these programs.

**References**:


1. Wedekind, J. E., Gillilan, R., Janda, A., Krucinska, J., Salter, J. D., Bennett, R. P., Raina, J., and Smith, H. C. (2006) *J. Biol. Chem. 281*, 38122-38126.
2. Hayes, D. B., Laue, T. M., and Philo, J. (2006) Sedimentation Interpretation Program, 1.09 ed., University of New Hampshire.
3. Philo, J. S. (2006) *Anal. Biochem. 354*, 238-246.
4. Schuck, P. (2000) *Biophys. J. 78*, 1606-1619.
5. Schuck, P. (2003) *Anal. Biochem. 320*, 104-124.
6. Stafford, W. F., 3rd. (1992) *Anal. Biochem. 203*, 295-301.
7. Kozin, M. B., and Svergun, D. I. (2001) *J. Appl. Crystallogr. 34*, 33-41.
8. 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 456*, 121-124.
9. Svergun, D. I., Barberato, C., and Koch, M. H. J. (1995) *J. Appl. Crystallogr. 28*, 768-773.
10. Garcia de la Torre, J., Huertas, M. L., and Carrasco, B. (2000) *Biophys. J. 78*, 719-730.
11. Garcia de la Torre, J., Carrasco, B., and Harding, S. E. (1997) *Eur. Biophys J. 25*, 361-372.

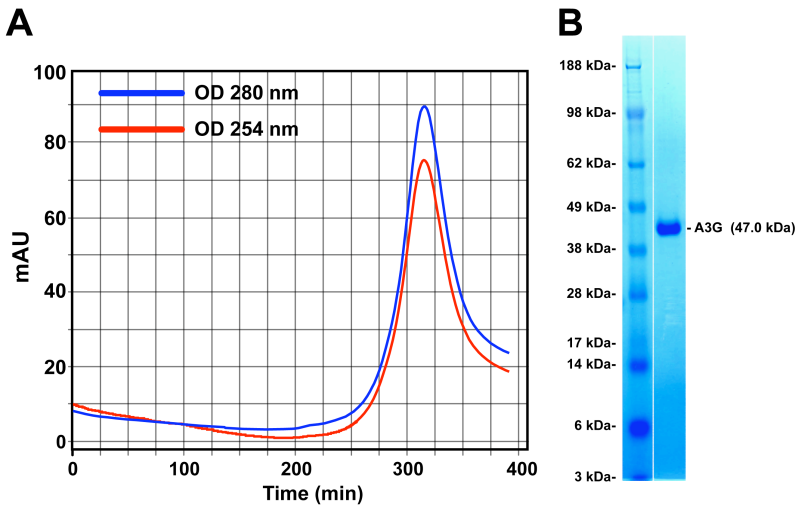


**Figure S1**: Representative A3G protein used for analytical ultracentrifugation. (A) The elution profile of A3G separated by Sephacryl S-300 gel filtration chromatography, and monitored at 280 nm and 254 nm to demonstrate RNA depletion. (B) SDS-PAGE of pooled A3G used for AUC analysis. Purity is estimated to be >99%.