Supporting Information

Structural Characterization of Semen Coagulum-derived SEM1(86–107) Amyloid Fibrils that Enhance HIV-1 Infection

Kinsley C. French, Nadia R. Roan and George I. Makhatadze

Supplementary Materials and Methods

HDX Sample Proteolytic Cleavage and Analysis by LC-MS

Fungal protease from *Aspergillus saitoi* (non-specific cleavage sites, Type XIII, P2143, ≥ 0.6 units/mg, Sigma-Aldrich, St. Louis, MO) was dissolved in 0.1% formic acid at 10 mg/mL and frozen until use. Following quenching, 2.0 mg/mL of fungal protease was added to the peptide solutions to induce peptide fragmentation. Next, 8 µL of sample was loaded onto a reverse-phase C18 HPLC column (BioBasic-C18, Thermo Electron Corporation, Waltham, MA) using a NanoFlow auto-sampler (Agilent, Santa Clara, CA). The samples were eluted at 200 µL/min in a 20% to 34% organic phase gradient over seven minutes. The aqueous phase solvent was 0.1% formic acid in water pH 2.25 and the organic phase solvent was 0.2% formic acid in acetonitrile. These elution conditions were chosen because, at low salt concentrations, HDX is minimized at pH 2.25 (*1*). The eluent was directly injected into an electrospray ionization mass spectrometer Thermo Scientific LTQ Orbitrap XL (Thermo Fisher Scientific Inc., Waltham, MA). To reduce back-exchange, the capillary temperature was set to 200^{o}C based on Walters et al. (*1*). All other parameters were automatically optimized by the LTQ Orbitrap XL Xcalibur 2.0.7 software for the m/z range of SEM1(86-107).

Hydroxyl Radical-Mediated Modification Analysis by LC-MS

Following lyophilization, the samples were dissolved in 10 µL of 20 mM NaPB pH 2.5 (2.5 mg/mL final peptide concentration) and analyzed by ESI-LC-MS both before and after proteolytic cleavage. For the uncleaved samples, 1 µL of the 2.5 mg/mL hydroxyl radical-modified peptide solution was diluted in 9 µL of 20 mM NaPB pH 2.5 (final protein concentration of 0.25 mg/mL). For the cleaved samples, 2 µL of the 2.5 mg/mL hydroxyl radical-modified peptide solution was mixed with 3 µL of 20 mM NaPB pH 2.5 and 5 µL of 10 mg/mL of the *Aspergillus saitoi* protease described above. Samples (8 µL) were loaded on to a

C-18 reverse phase HPLC column (BioBasic-C18, Thermo Electron Corporation, Waltham, MA) using a NanoFlow auto-sampler (Agilent, Santa Clara, CA). Uncleaved peptide samples were eluted at 200 µL/min in a 20% to 27% organic phase gradient over 7 minutes. Cleaved peptide samples were eluted at 200 µL/min in a 0% to 40% organic phase gradient over 20 minutes. The aqueous and organic phases were 0.2% formic acid in water and 0.2% formic acid in acetonitrile, respectively. The eluent was directly injected into the electrospray ionization mass spectrometer Thermo Scientific LTQ Orbitrap XL (Thermo Fisher Scientific Inc., Waltham, MA) and the mass spectra were measured using a capillary temperature of 275^{o}C.

Hydroxyl Radical-Mediated Modification Analysis Calculations

Following hydroxyl-radical mediated modification of SEM1(86-107), the percent exposures of residues in each peptide fragment were calculated using equations 4 and 5 provided in the main body of the paper. To determine the best method to sequence average the data, three averaging algorithms were tested on fictional data sets. In the fictional data sets, the percent modification of each residue was known and used to calculate the percent exposure of a given peptide fragment. Each of the three averaging algorithms was used on the fictional data sets and the root of the residual sum of squares was calculated between the known percent exposure and the calculated percent exposure. The averaging algorithm given in equation 6 was chosen to analyze the experimental data because it yielded the most accurate representation of the fictional test data. Other algorithms that were tested include a modified version of equation 6, where m_j is the number of residues in peptide *j*; and the algorithm given in the following equation:

$$E_{OH_i} = \left(\sum_{j=1}^{n_i} \frac{E_j}{n_i}\right), \quad \textbf{(S1)}$$

where E_{OHi} is the average percent exposure for residue *i*, n_i is the number of peptide fragments containing residue *i*, and E_j is the percent exposure for a peptide fragment *j* containing residue *i*.

Sequence-Based Prediction of SEM1(86-107) Amyloidogenic Regions

The following programs were used to predict the amyloidogenic regions of SEM1(86-107): AGGRESCAN (*7*), AmyloidMutants (*8*), Pafig (*9*), PASTA (*10*), Waltz (*11*), and ZipperDB (*12*). TANGO (*13*) was also used, but did not identify any amyloidogenic regions in SEM1(86-107). For AmyloidMutants, residues predicted to contain beta-sheet in the most abundant cluster were considered to be amyloidogenic. For Pafig, residues were considered to be amyloidogenic

if the program predicted them to be fibrillar. For ZipperDB, residues were considered to be amyloidogenic if the Rosetta energy score was below -23 kCal/mol. For all other programs, no residues in SEM1(86-107) were identified as amyloidogenic using predefined program thresholds. Therefore, the residues that were considered to be amyloidogenic in the results section of this paper are residues that obtained the most amyloid-like scores compared to other residues within the SEM1(86-107) sequence. For AGGRESCAN, residues were considered to be amyloidogenic if the average over the sliding window (a^4v) was greater than -0.4 (the predefined cutoff is -0.02). For PASTA, the residues contained in the lowest energy pair were considered to be amyloidogenic (energy of ~ -1.6; the predefined threshold is < -4.0). For Waltz, residues identified as fibrillar by the program at pH 7 using a user-defined sensitivity threshold of 50 were considered to be amyloidogenic.

Supplementary Figures

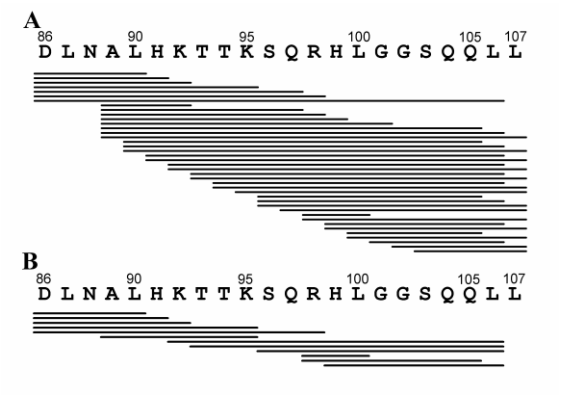

Figure S1. SEM1(86-107) peptide sequences used to calculate HDX (A) and hydroxyl radical-mediated modification (B) of the SEM1(86-107) fibrils. The peptide sequences indicated by the black lines below the SEM1(86-107) peptide sequence indicate the peptide sequences used in HDsite to calculate the HDX profile reported in Figure 3 (A) or equation 6 to calculate the percent hydroxyl radical-mediated modification profile reported in Figure 4 (B).

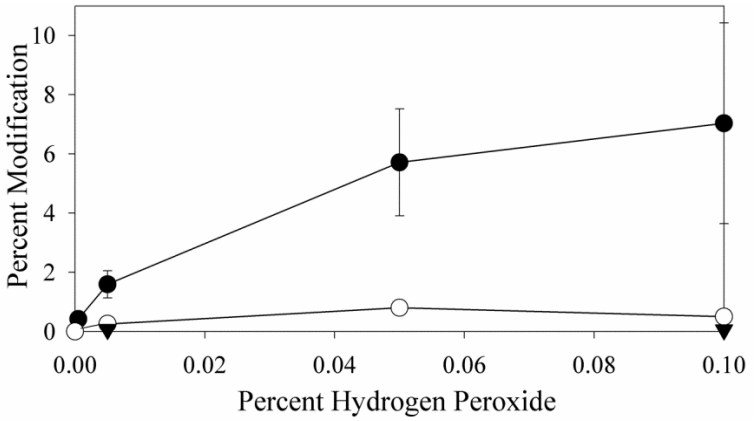

Figure S2. Percent hydroxyl radical-mediated modification as a function of hydrogen peroxide concentration. SEM1(86-107) monomers (black circles) or fibrils (white circles) were exposed to varying concentrations of hydrogen peroxide and UV radiation. The percentage of modified peptide was then determined using liquid chromatography-mass spectrometry and plotted as a function of hydrogen peroxide concentration. SEM1(86-107) monomer was also placed under quenching conditions prior to hydrogen peroxide and UV radiation exposure (black triangles) to measure the amount of modification that occurs after quenching (pre-quenched monomer sample). Error bars reflect the standard deviation of at least three independent experiments.

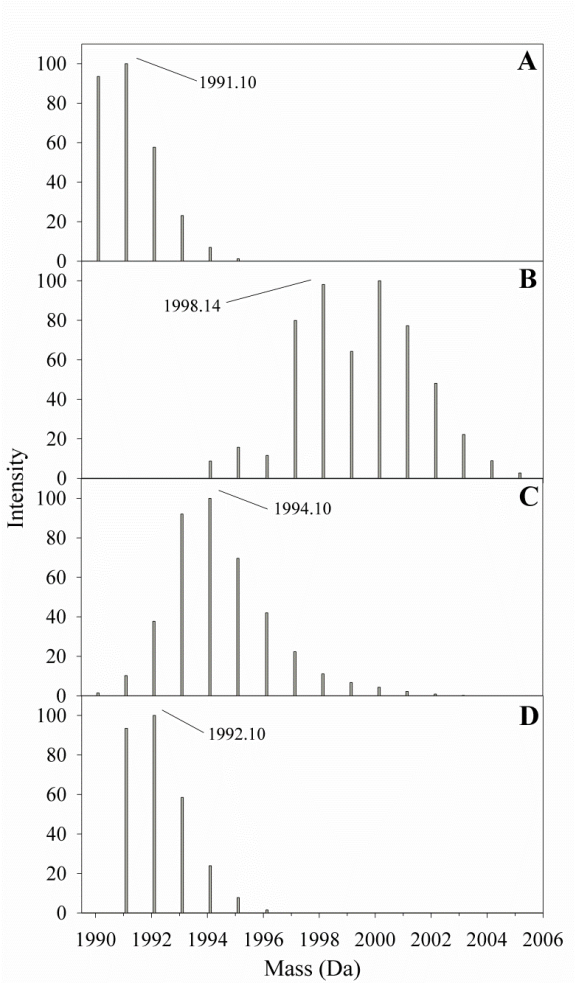

Figure S3. Mass spectra of hydrogen-deuterium exchanged SEM1(86-107) peptide. Mass spectra isotopic peaks corresponding to residues 89-106 of SEM1(86-107) for the unexchanged monomer (A), exchanged monomer (B), exchanged fibril (C), and in-exchange monomer (D), respectively. These peaks show that the fibril is protected from HDX relative to the monomer (the fibril has ~ 19% HDX, while the monomer has ~ 48% HDX), and that there is little HDX in the in-exchange sample (~ 1%).

Figure S4. Aggregation-prone regions of SEM1(86-107) calculated by the indicated prediction method. Residues highlighted in black are predicated to be aggregation-prone as determined by the indicated method (*7-12*).

References

1. Walters, B. T., Ricciuti, A., Mayne, L., and Englander, S. W. (2012) Minimizing back exchange in the hydrogen exchange-mass spectrometry experiment, *J. Am. Soc. Mass Spectrom. 23*, 2132-2139.
2. Kan, Z.-Y., Mayne, L., Chetty, P. S., and Englander, S. W. (2011) ExMS: Data analysis for HX-MS experiments, *J. Am. Soc. Mass Spectrom. 22*, 1906-1915.
3. Lu, X., Wintrode, P. L., and Surewicz, W. K. (2007) Beta-sheet core of human prion protein amyloid fibrils as determined by hydrogen/deuterium exchange, *Proc. Natl. Acad. Sci. U.S.A. 104*, 1510-1515.
4. Zhang, Z., and Smith, D. L. (1993) Determination of amide hydrogen exchange by mass spectrometry: A new tool for protein structure elucidation, *Protein Sci. 2*, 522-531.
5. Damo, S. M., Phillips, A. H., Young, A. L., Li, S., Woods, V. L., and Wemmer, D. E. (2010) Probing the conformation of a prion protein fibril with hydrogen exchange, *J. Biol. Chem. 285*, 32303-32311.
6. French, K. C., and Makhatadze, G. I. (2012) Core sequence of PAPf39 amyloid fibrils and mechanism of pH-dependent fibril formation: The role of monomer conformation, *Biochemistry 51*, 10127-10136.
7. Conchillo-Sole, O., de Groot, N., Aviles, F., Vendrell, J., Daura, X., and Ventura, S. (2007) AGGRESCAN: A server for the prediction and evaluation of "hot spots" of aggregation in polypeptides, *BMC Bioinf. 8*, 65.
8. O'Donnell, C. W., Waldispühl, J., Lis, M., Halfmann, R., Devadas, S., Lindquist, S., and Berger, B. (2011) A method for probing the mutational landscape of amyloid structure, *Bioinformatics 27*, i34-i42.
9. Tian, J., Wu, N., Guo, J., and Fan, Y. (2009) Prediction of amyloid fibril-forming segments based on a support vector machine, *BMC Bioinf. 10*, S45.
10. Trovato, A., Seno, F., and Tosatto, S. C. E. (2007) The PASTA server for protein aggregation prediction, *Protein Eng. Des. Sel. 20*, 521-523.
11. Maurer-Stroh, S., Debulpaep, M., Kuemmerer, N., de la Paz, M. L., Martins, I. C., Reumers, J., Morris, K. L., Copland, A., Serpell, L., Serrano, L., Schymkowitz, J. W. H., and Rousseau, F. (2010) Exploring the sequence determinants of amyloid structure using position-specific scoring matrices, *Nat. Methods 7*, 237-242.
12. Goldschmidt, L., Teng, P. K., Riek, R., and Eisenberg, D. (2010) Identifying the amylome, proteins capable of forming amyloid-like fibrils, *Proc. Natl. Acad. Sci. U.S.A. 107*, 3487-3492.
13. Fernandez-Escamilla, A.-M., Rousseau, F., Schymkowitz, J., and Serrano, L. (2004) Prediction of sequence-dependent and mutational effects on the aggregation of peptides and proteins, *Nat. Biotechnol. 22*, 1302-1306.