SUPPLEMENTAL INFORMATION

Solution Structure and Membrane Interaction of the Cytoplasmic Tail of HIV-1 gp41 Protein

R. Elliot Murphy,[1] **Alexandra B. Samal,**[1] **Jiri Vlach, and Jamil S. Saad***

Department of Microbiology, University of Alabama at Birmingham, Birmingham, AL 35294

Correspondence: saad@uab.edu

[1] Authors contributed equally to this work.

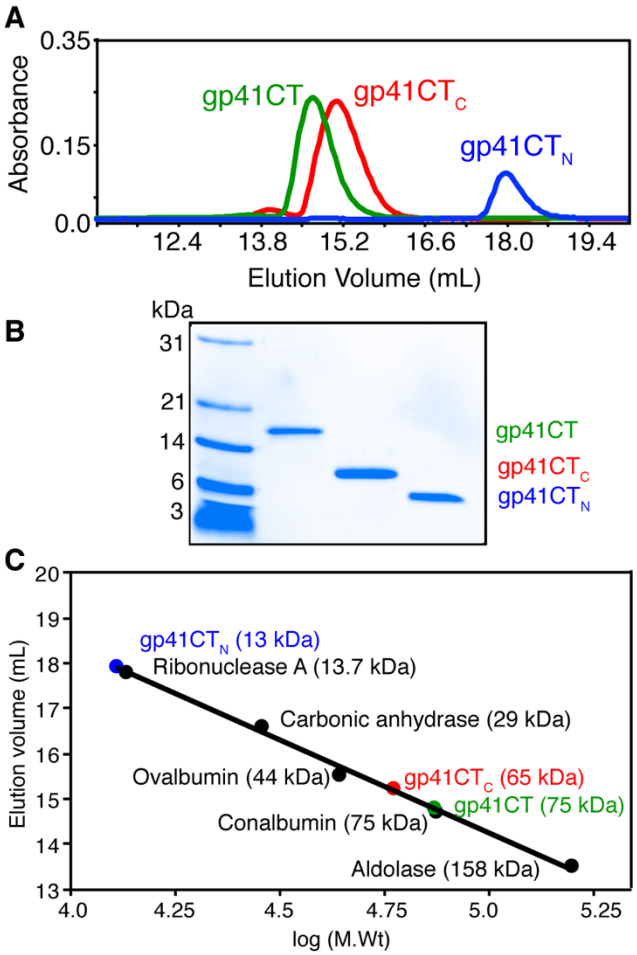

Figure S1. Related to Figure 1. Gel Filtration Assay of gp41CT Proteins. (A) Elution profiles of gp41CT, gp41CT$_N$, and gp41CT$_C$ on a HiLoad Superdex 200 (10/300 GL) column. (B) Protein fractions from the gel filtration assay were checked by SDS-PAGE and stained with Coomassie brilliant blue. (C) A molecular weight calibration kit was used to determine the approximate molecular weight of the gp41CT constructs.

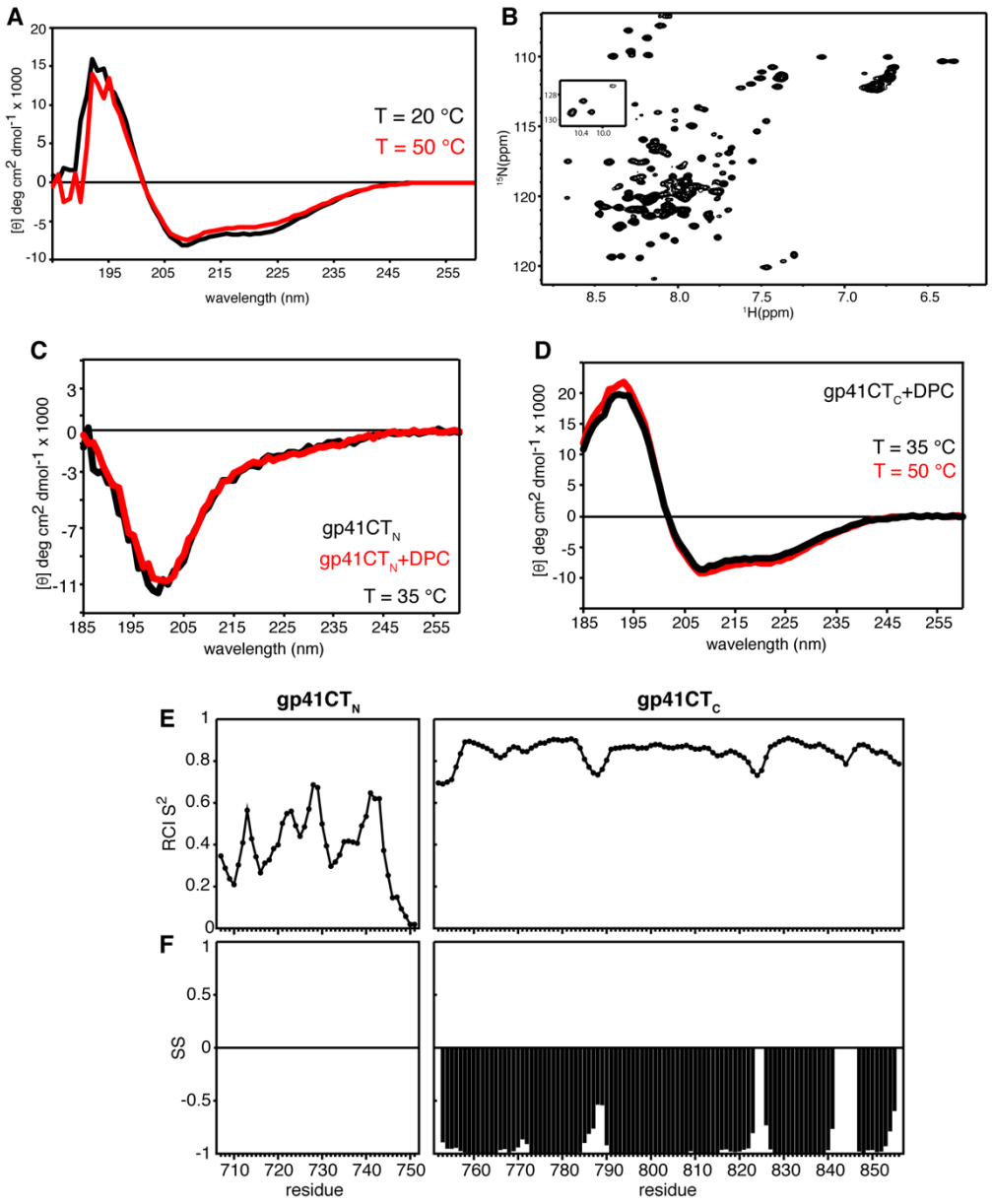

Figure S2. Related to Figure 2. CD and NMR spectra, and Secondary Structure Prediction of gp41CT Proteins. (A) Far-UV CD spectra obtained for gp41CT at 20 and 50 °C. The spectra are almost identical, demonstrating that elevated temperature does not induce any structural changes to the protein. CD features are consistent with an α-helical protein. (B) 2D ^{1}H-^{15}N HSQC spectrum obtained for gp41CT (150 μM) at 50 °C in a buffer containing 50 mM sodium phosphate (pH 6.0), 50 mM NaCl, 1 mM TCEP, 25 mM DPC, and 5% D_2O. Inset shows tryptophan side-chain signals. (C) CD spectra of $gp41CT_N$ collected in the absence or presence of DPC at 35 °C. The spectra are almost identical, demonstrating that DPC does not induce any structural changes to the protein. Both spectra display a negative band at ~200 nm consistent with a random coil. (D) CD spectra of $gp41CT_C$ collected in the presence of 3 mM DPC at 35 and 50 °C. Two minima at 208 and 222 nm are observed, consistent with an α-helical structure. Experimental temperature has no effect on the structure of the protein. (E) Random coil index (RCI)-derived order parameters S^2 and (F) the probability of secondary structure (positive values are obtained for extended structure, negative for α-helix) plotted for $gp41CT_N$ and $gp41CT_C$ residues. Only secondary structure probabilities |SS| > 0.5 are shown.

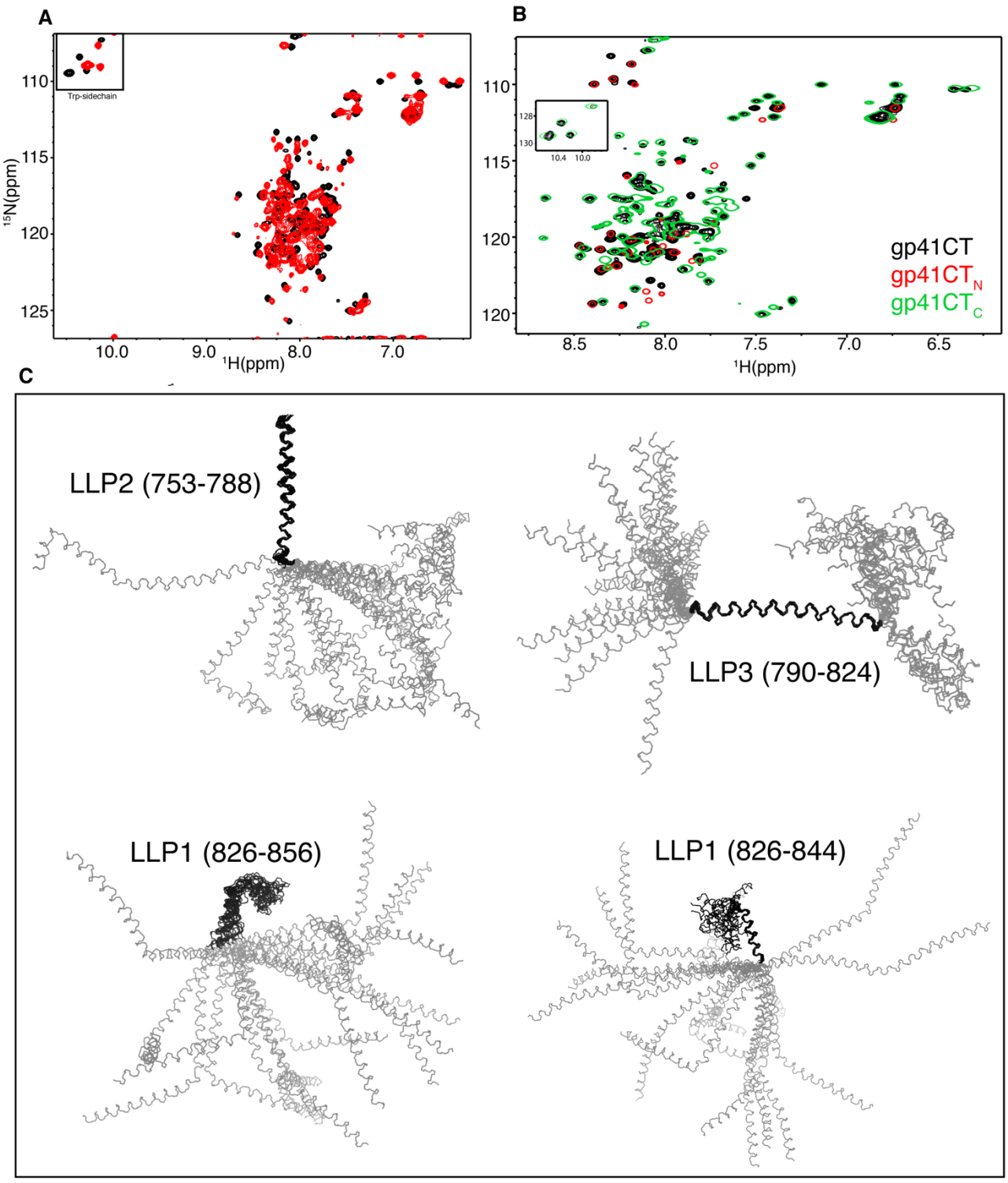

Figure S3. Related to Figures 3 and 4. NMR Spectra in Micelles or Bicelles and Ensemble of gp41CT$_C$ Structures. (A) Overlay of 2D ^{1}H-^{15}N HSQC spectra for gp41CT$_C$ (150 μM) in DPC micelles (black) or DMPC/DHPC bicelles (red). ^{15}N positions of boxed Trp side chain signals are shifted by 20 ppm due to aliasing. (B) Overlay of 2D ^{1}H-^{15}N HSQC spectra for gp41CT, gp41CT$_N$, and gp41CT$_C$ (150 μM) at 50 ºC in the presence of 25 mM DPC. Inset shows tryptophan side chain signals. (C) Best-fit backbone superposition of the 20 refined structures calculated for gp41CT$_C$. Superimposed residues are labeled and backbone bonds represented as dark lines. The calculated ensemble showed a good convergence and correspondingly low positional RMSD values within the structured regions.

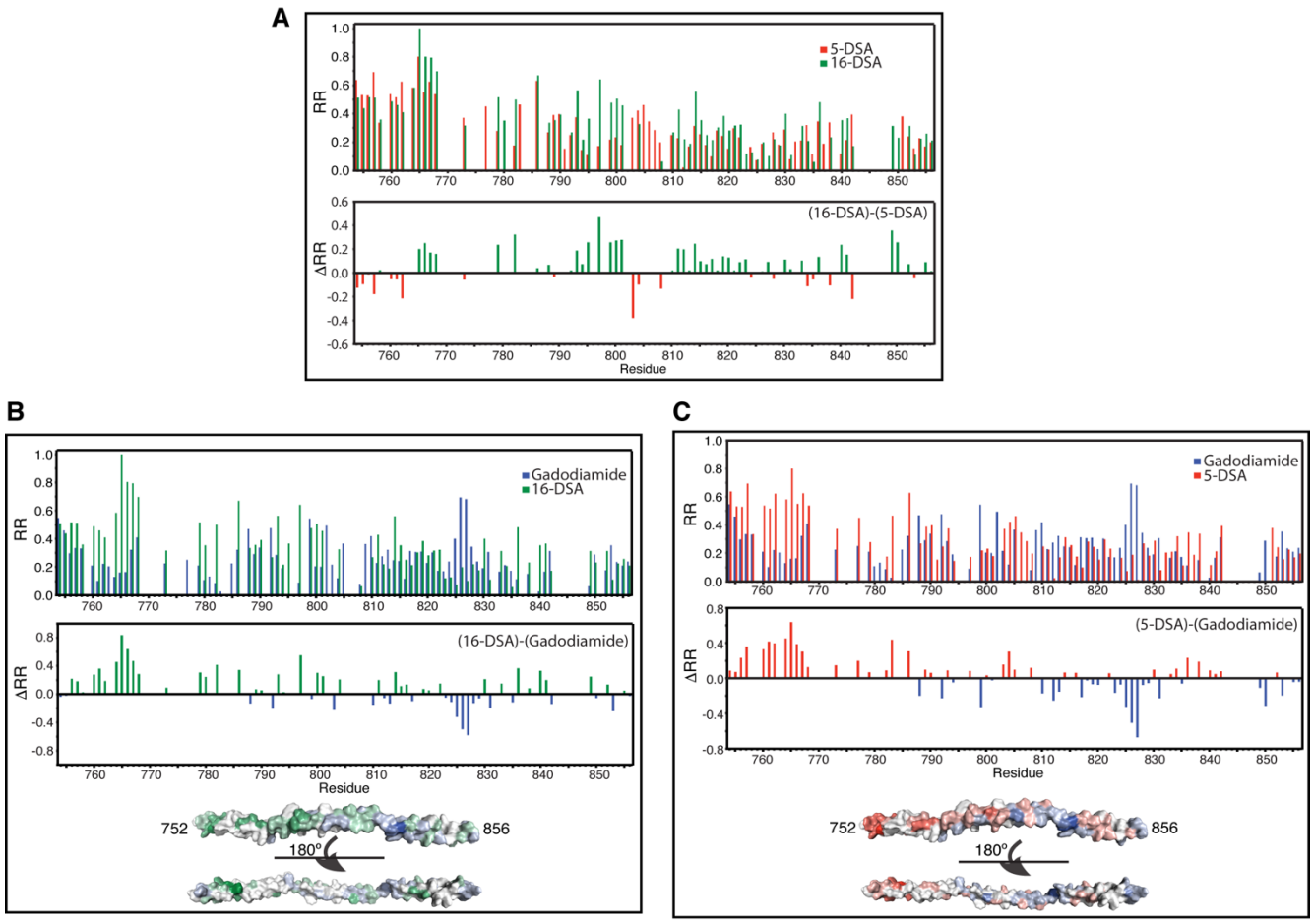

Figure S4. Related to Figure 5. Paramagnetic Relaxation Enhancement Effect on gp41CT$_C$. (A) Histograms depict the relative intensity reduction (RR) for gp41CT$_C$ backbone amide signals determined from ^{1}H,^{15}N-HSQC NMR spectra before and after the addition of 16-DSA or 5-DSA. In the upper chart, RR values are shown for 5-DSA and 16-DSA. In the lower chart, their difference is shown as ΔRR = $RR_{16DSA} - RR_{5DSA}$ only for residues for which both 5-DSA and 16-DSA data were available. The titration data are shown at the point before first amide signals became broadened beyond detection (0.4 and 0.8 mM for 16-DSA and 5-DSA, respectively). (B) Histograms depict RR values for gp41CT$_C$ backbone amide signals before and after the addition of 16-DSA, 5-DSA or gadodiamide. In the upper chart of each panel, RR values are shown for (B) 16-DSA and gadodiamide and (C) 5-DSA and gadodiamide. In the middle panels, the difference between DSA and gadodiamide is shown as ΔRR = $RR_{DSA} - RR_{Gd}$ only for residues for which both DSA and gadodiamide data were available. Titration data are shown at the point before first amide signals became broadened beyond detection (0.4, 0.8 and 9.6 mM for 16-DSA, 5-DSA and gadodiamide, respectively). As indicated by the data, the solvent exposed region of the protein is very minimal (mainly residues 825-827). Lower panels in B and C show surface representation of the representative model of gp41CT$_C$ highlighting residues based on their respective ΔRR values from the middle panels. The residues are colored based on a gradient from (A) green to blue and (B) red to blue. The intensity of the color is proportional to the magnitude of the ΔRR value. White color denotes residues lacking ΔRR values.

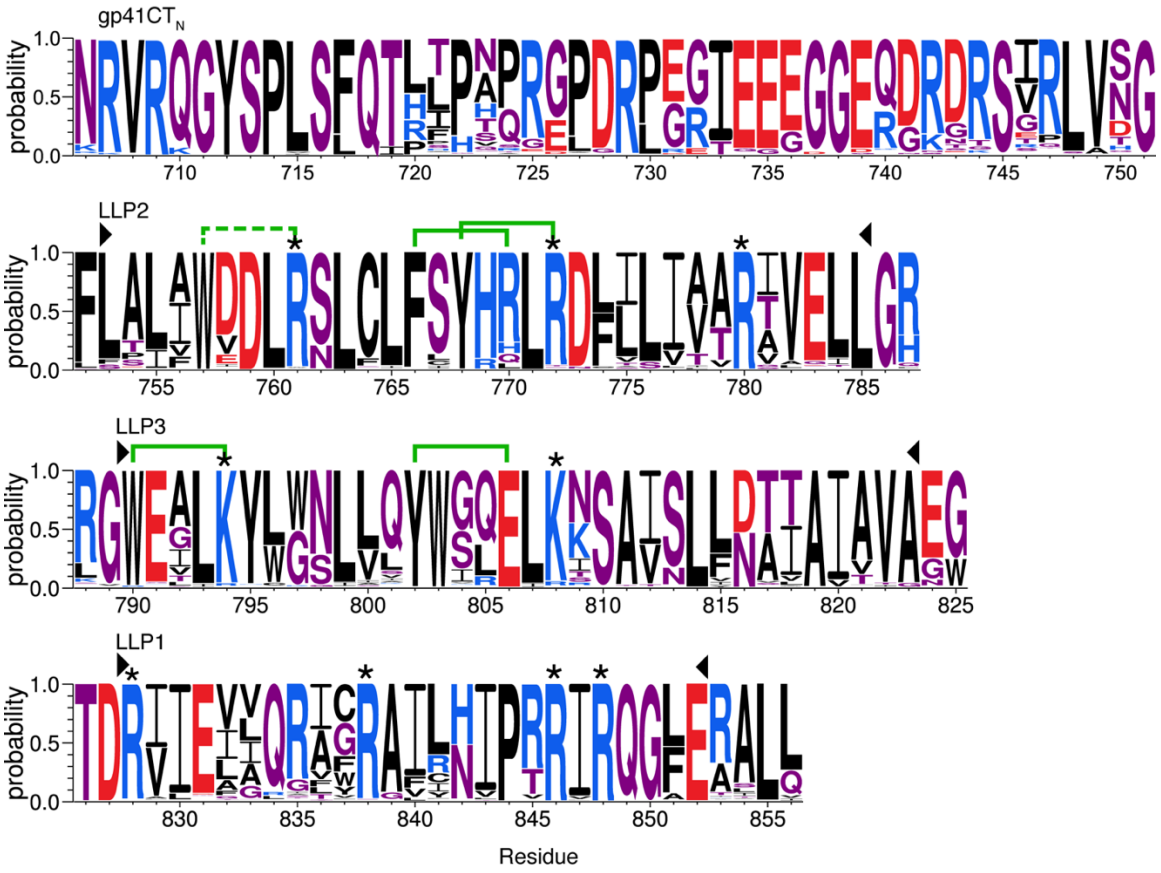

Figure S5. Related to Figure 5. Sequence conservation of the HIV-1 gp41CT protein. Logo diagrams of sequence conservation in HIV-1 gp41CT. Arginine and lysine residues with >90 % conservation are denoted by an asterisk. Green solid lines indicate residues for which cation-π and potential anion-π interactions were observed. Dashed green line indicates a potential cation-π pair for which no NOE contacts were detected. Arrowheads mark helix boundaries. Protein sequence alignments for 5095 Env regions of HIV-1 M group isolates, including recombinants, were obtained from the Los Alamos database as filtered web alignments (https://www.hiv.lanl.gov). Region corresponding to gp41CT residues 752–856 was extracted and gaps with less than 2.5% occurrences removed. Also removed was a 7-residue insertion between LLP2 and LLP3 (between residues 787 and 788) that is found only in subtypes A, C and G. Sequences were edited in Unipro UGENE and logo sequence diagrams created with WebLogo 3 (http://weblogo.threeplusone.com).

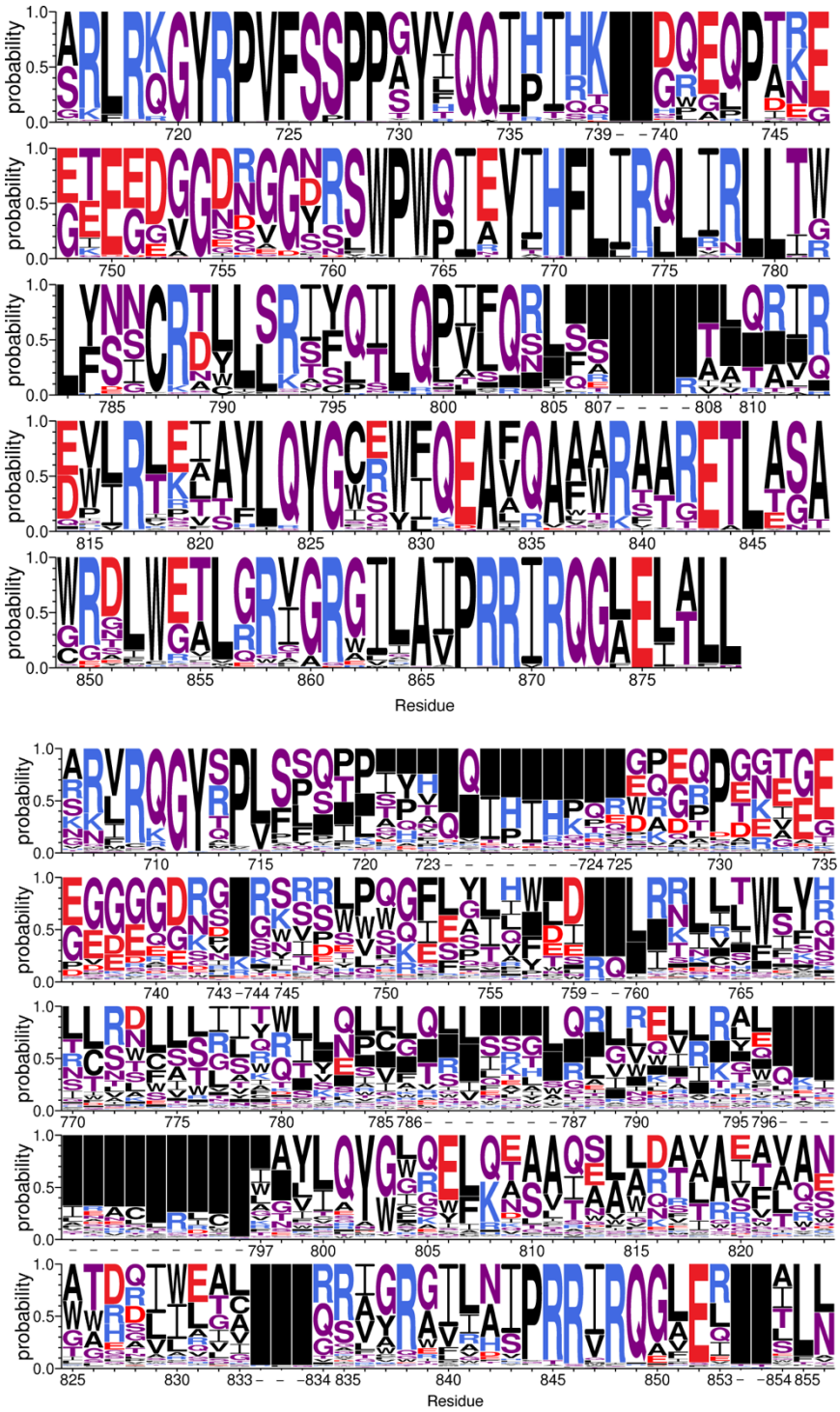

Figure S6. Related to Figure 5. Sequence conservation of the HIV-2 and SIV gp41CT protein. Logo diagrams of gp41CT sequence conservation in HIV-2 (top) and SIV (bottom). Logo diagram was produced using AnalyzeAlign tool available at Los Alamos National Laboratory (New Mexico, USA) web site (http://www.hiv.lanl.gov/content/sequence/ANALYZEALIGN/analyze_align.html). The diagram is based on protein Env sequence alignments of 128 HIV-2 (residues 715–879, HXB2 numbering) and 206 SIV (residues 706–856, HXB2 numbering) available therein. Black boxes represent gaps present in the sequence alignment, with box height proportional to the gap population. Skips in sequence numbering are indicated by dashes.