Title: Experimental evaluation of coevolution in a self-assembling particle

Authors: Emily C. Hartman,[a] Marco J. Lobba,[a] Andrew H. Favor,[a] Stephanie A. Robinson,[a] Matthew B. Francis*[a,b], Danielle Tullman-Ercek*[c]

[a]Department of Chemistry, University of California, Berkeley, California 94720-1460, [b]Materials Sciences Division, Lawrence Berkeley National Laboratories, Berkeley, California 94720-1460, and [c]Department of Chemical and Biological Engineering, Northwestern University, 2145 Sheridan Road, Technological Institute E136, Evanston, Illinois 60208-3120

Email: ercek@northwestern.edu; mbfrancis@berkeley.edu

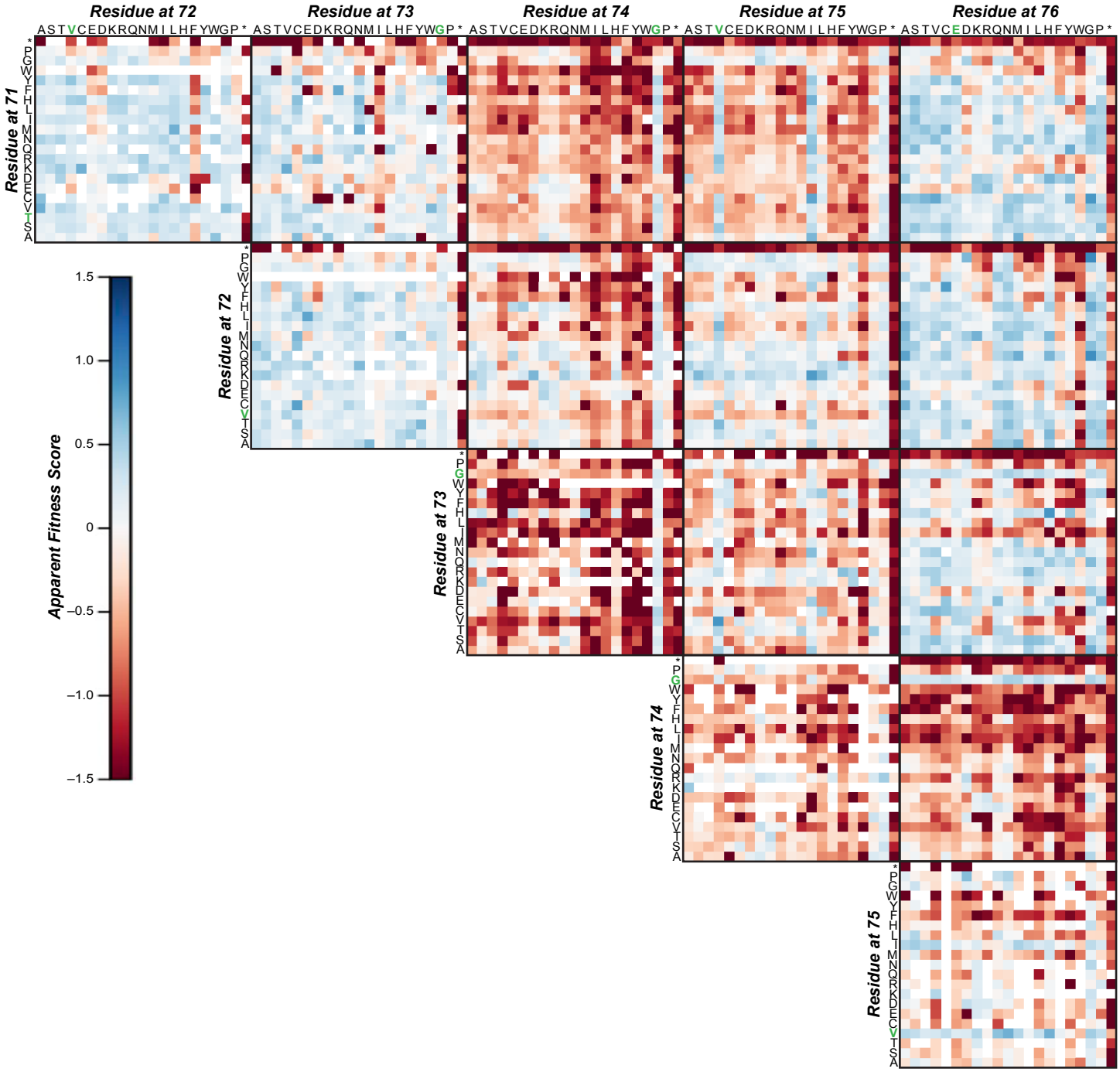

Figure S1. Assembly-selected epistatic landscape of the MS2 CP FG loop. All possible two amino acid variants were characterized for assembly competency. Blue indicates variants that were enriched following the assembly selection, and red indicates variants that were less abundant following the selection. Dark red indicates variants that were sequenced in the plasmid library but absent in the VLP library.

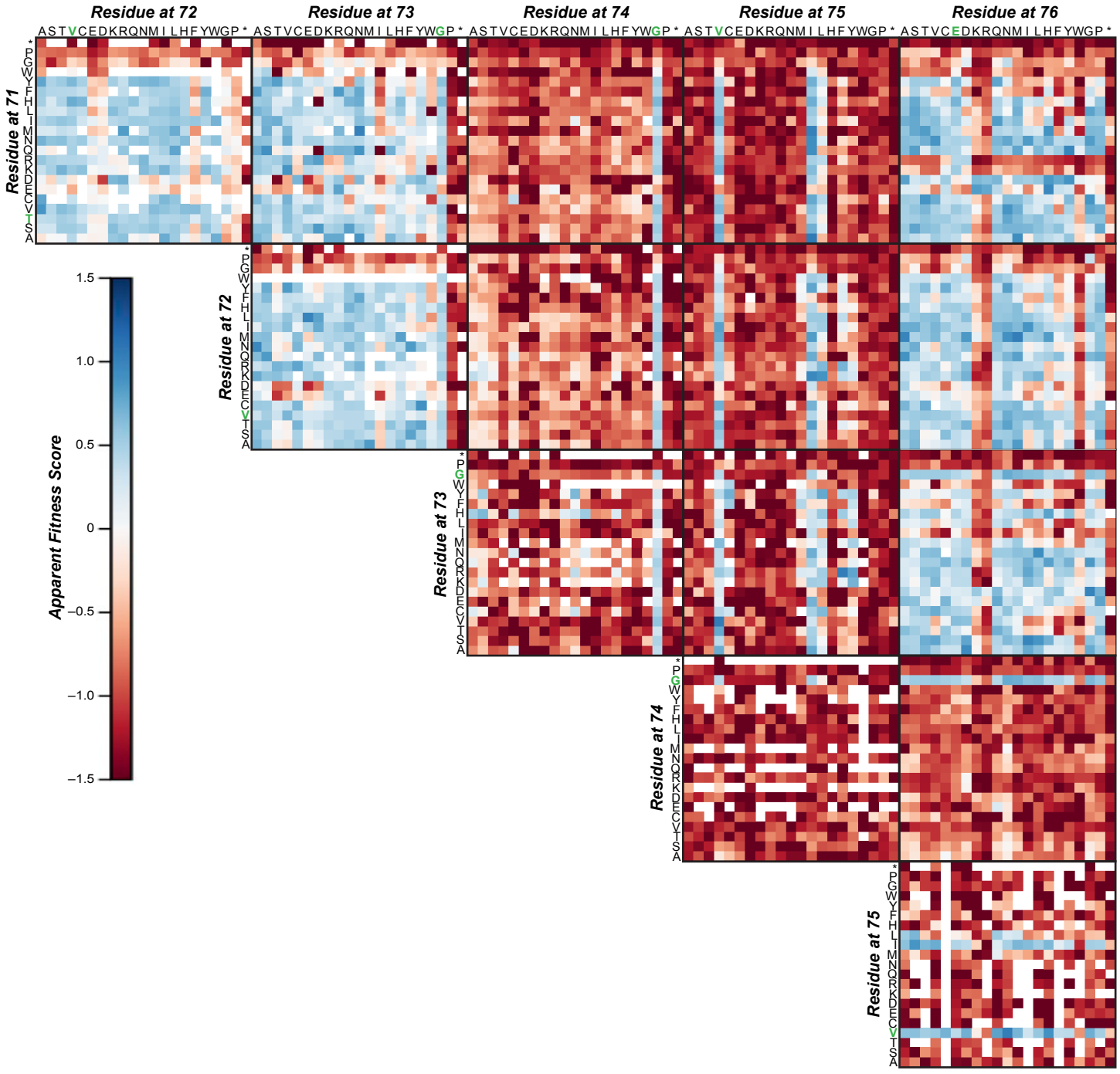

Figure S2. Epistatic landscape of the FG loop following a heat challenge at 50 ºC for 10 min. Blue indicates variants that were enriched following the heat selection, and red indicates variants that were less abundant following the selection. Dark red indicates variants that were sequenced in the plasmid library but absent in the heat-challenged library. Wild-type amino acids are indicated in green for the one-letter codes.

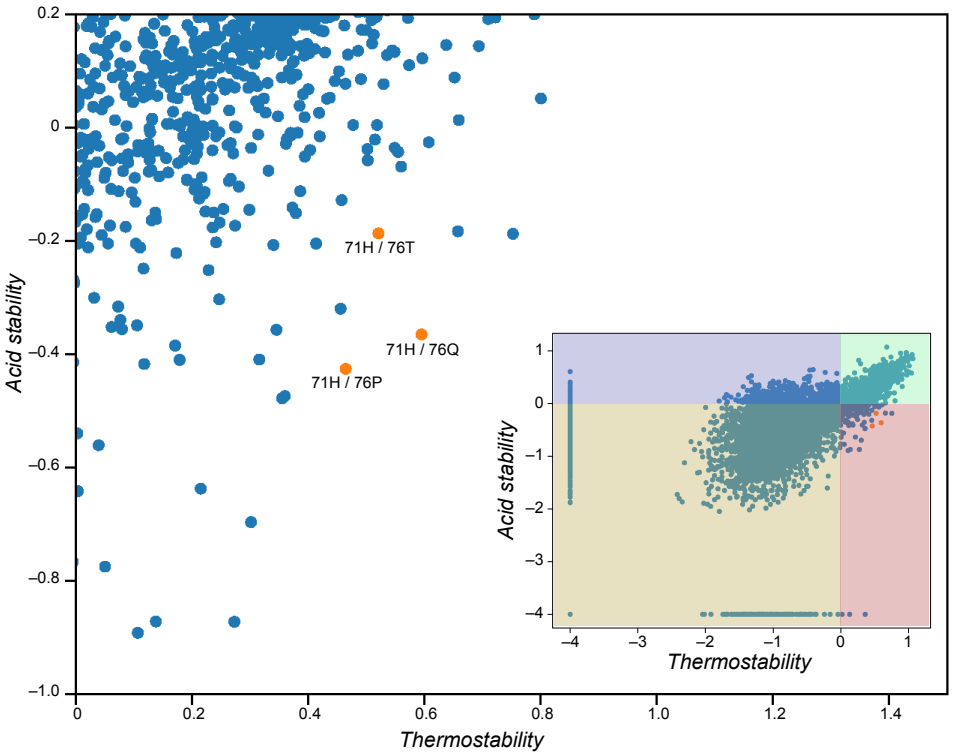

Figure S3. Thermostable VLPs were compared to acid-stable VLPs to identify candidate variants with lowered acid stability and uncompromised thermostability. Variants of interest have high thermostability and low acid stability scores, which corresponds to the region highlighted in red. High acid stability and high thermostability (green), high acid stability and low thermostability (purple), and low acid stability and low thermostability (beige) are all less desirable variants. Several variants, indicated in orange, had strong positive scores in the heat-selected 2D-AFL with far reduced abundances following acidic pressure.

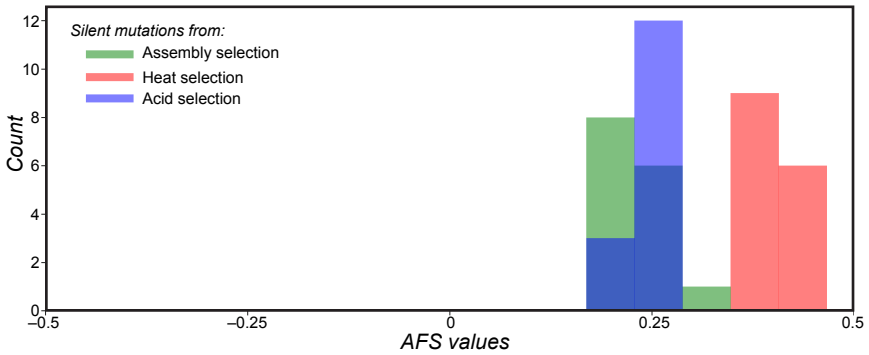

Figure S4. Silent mutations from the assembly, heat, and acid selections. In each selection, all 15 instances of unmuated VLPs score higher than 0.2. In the heat selection, CP[WT] scores increase, likely because many mutants are not tolerant to the heat selection, resulting in increased relative percent abundances.

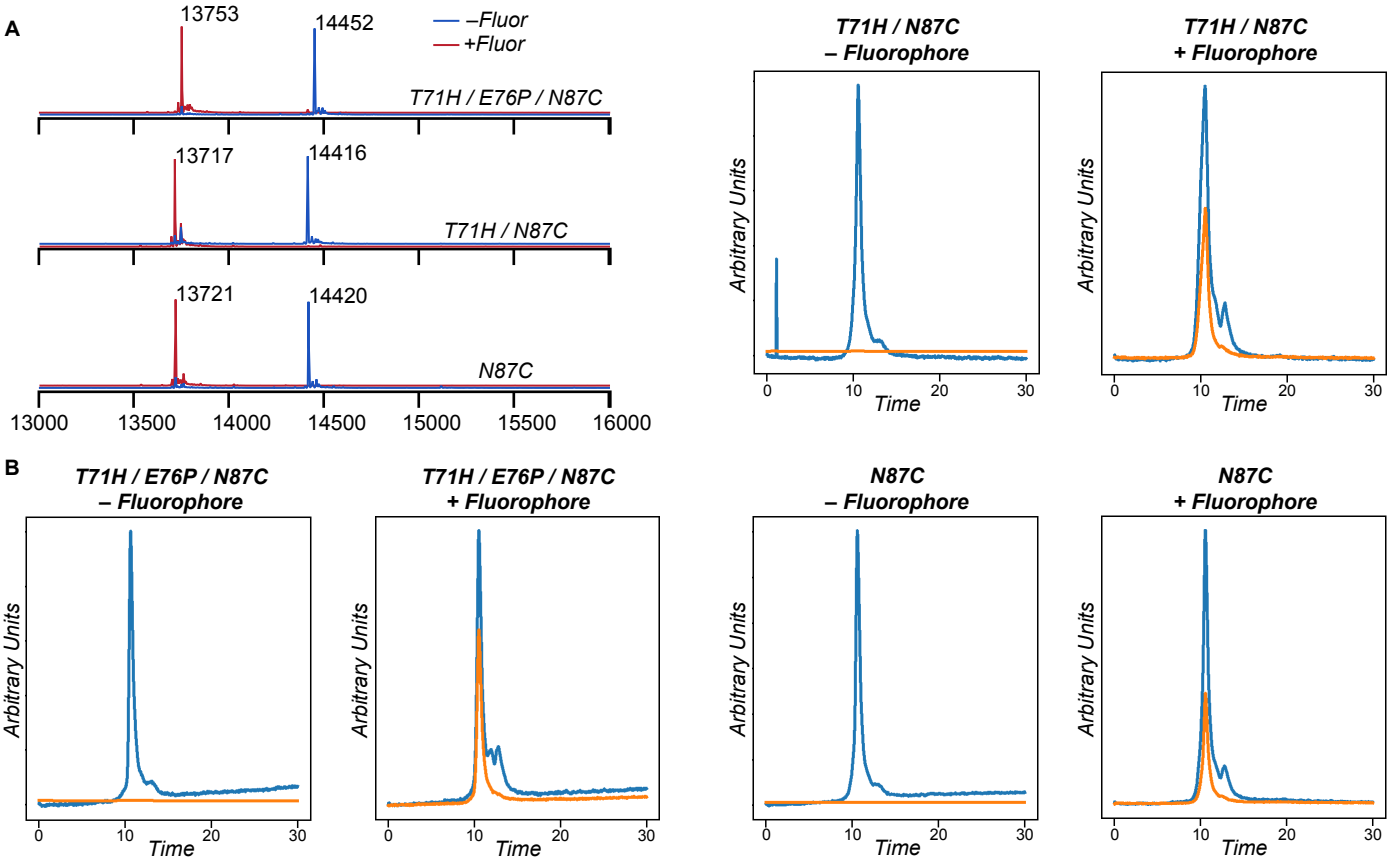

Figure S5. Chemical conjugation of N87C. A) Mass spectrometry indicates near complete modification of the interior cysteine in all cases. HPLC SEC traces of CP[T71H/E76P/N87C] (B,C), CP[T71H/N87C] (D,E), and CP[N87C] (F,G) with and without AlexaFluor488 maleimide, respective. Orange indicates absorbance at 488 nm, while blue indicates absorbance at 280 nm.

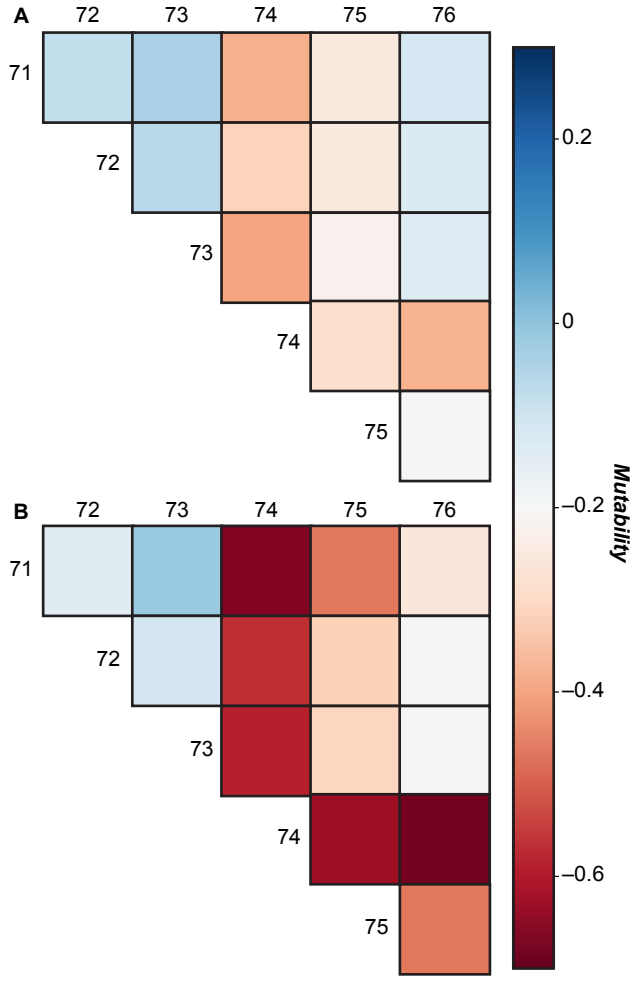

Figure S6. Shannon Entropy is used to calculate the mutability of each pair of residues following the assembly selection (A) and heat selection (B).

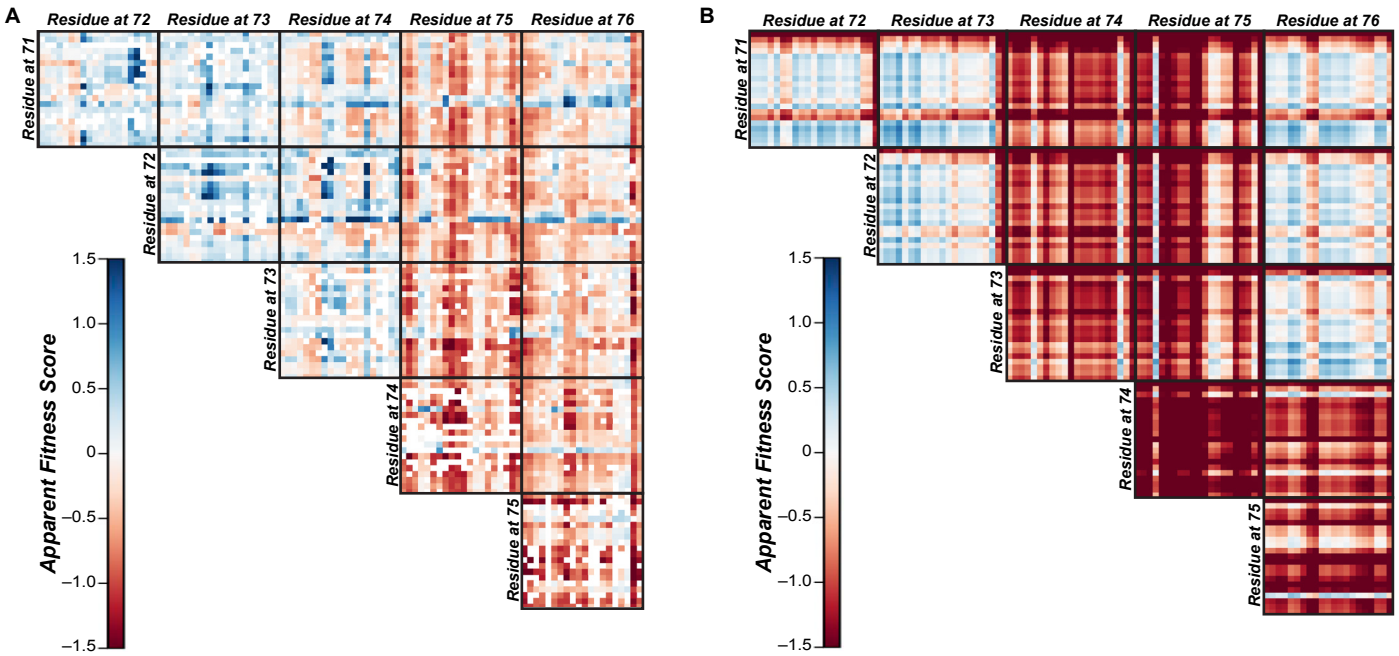

Figure S7. Predicted 2D-AFLs generated from 1D-AFL data using a convolutional neural network (A) and a simple additive method (B). The values indicated in B appear as the green bars in Figure 6 of the main text.

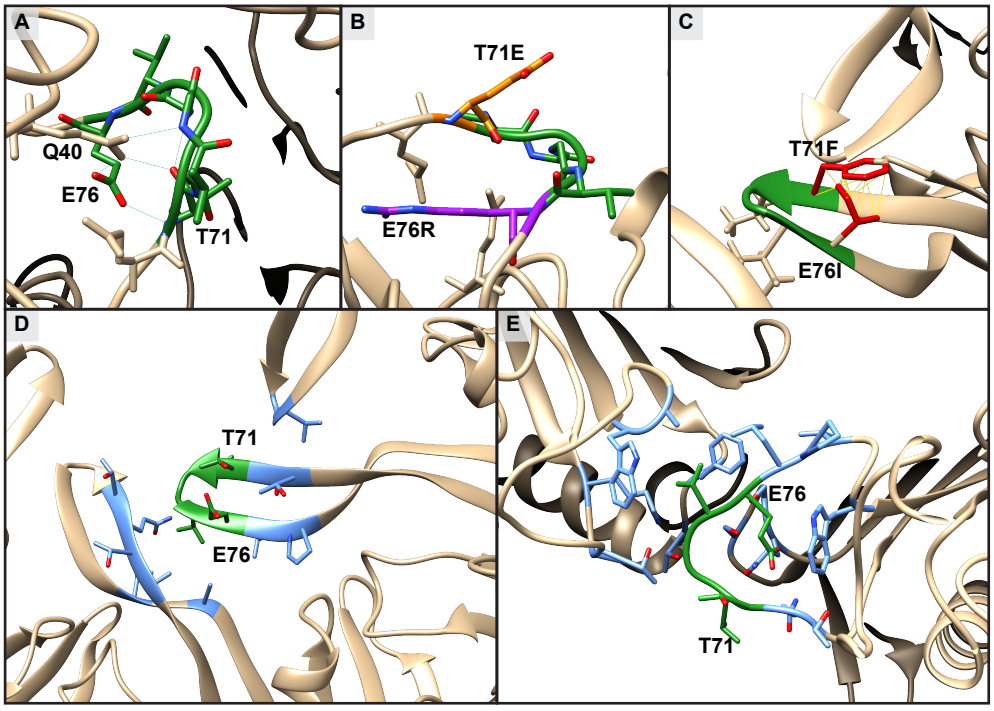

Figure S8. Chimera analysis of mutations with a large effect on VLP formation. In each structure, T71 and E76 are indicated with labels, and the FG loop is shown in green. A) Hydrogen bond networks are shown in the B form of the FG loop. B) CP[T71E / E76R] is visualized at the B form. T71E is shown in orange, while E76R is shown in purple. C) CP[T71F / E76I] is shown at the C form, where clashes are indicated with yellow lines. D,E) The local environment of the FG loop is shown for the A/C form (D) and B form (E). In each case, residues within 5 Å are indicated in light blue.

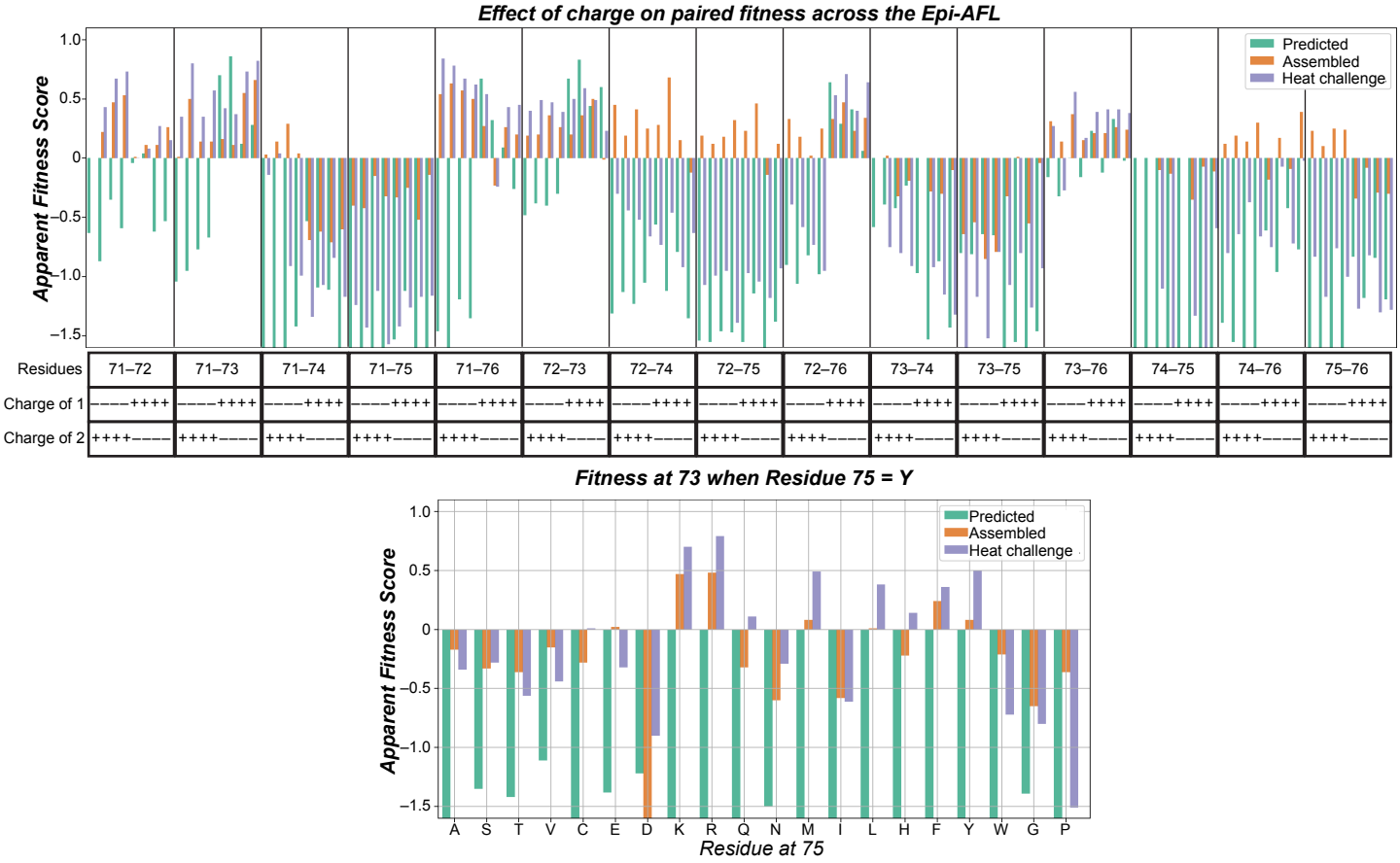

Figure S9. Instances of positive and negative sign epistasis. A) The effect of two mutations to charged residues is compared across the FG loop. B) Combinations of mutations at position G73 and V75 are evaluated for epistasis. In these graphs, AFS values predicted from the 1D AFL data are shown in green. AFS values are also shown for the measured assembly-selected (orange) and heat-selected (purple) datasets. Differences between the predicted and measured scores indicate epistatic interactions.