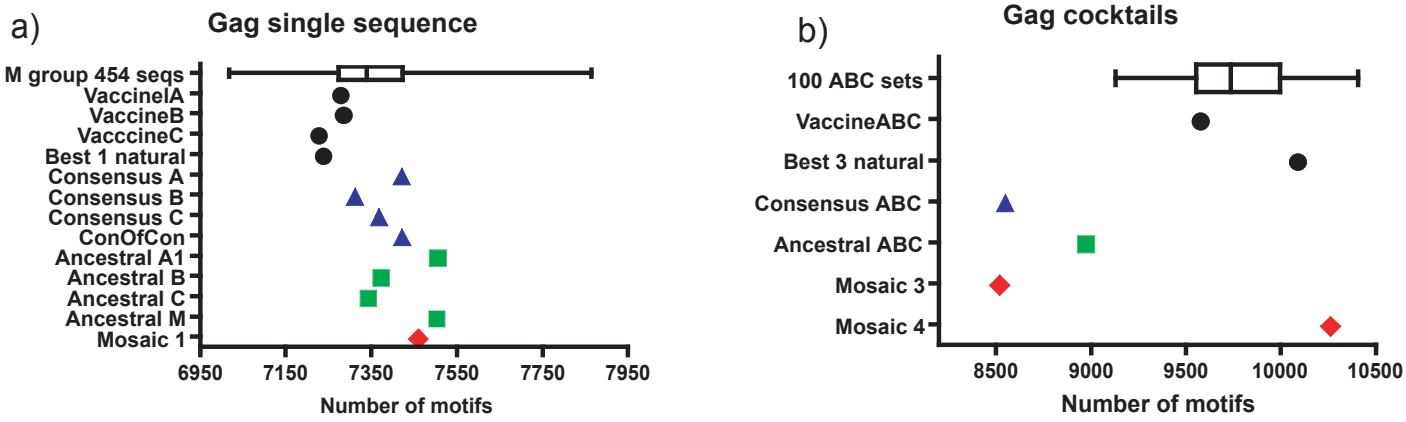
HLA binding motifs



Unfavorable amino acids

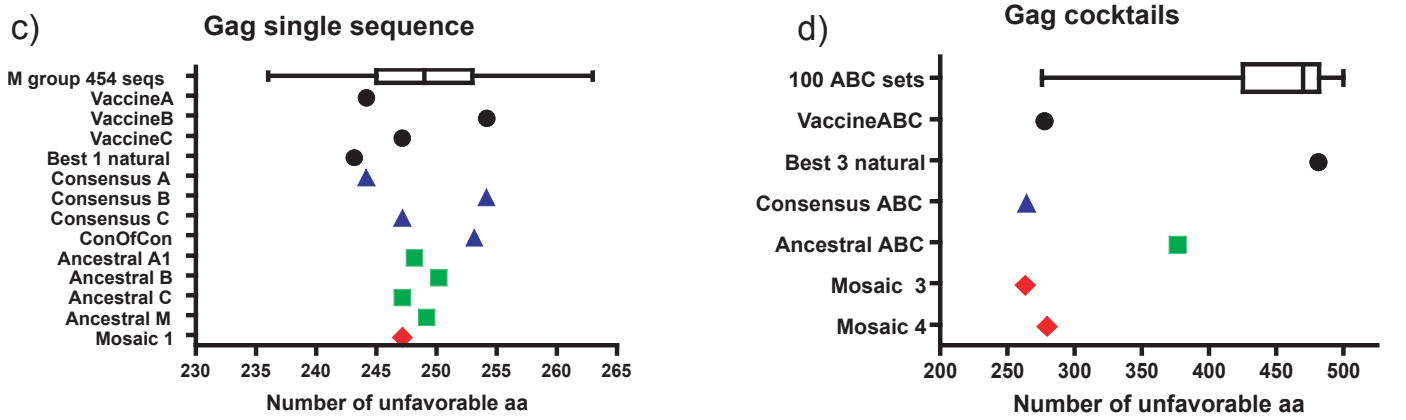


Figure S3. HLA binding potential of vaccine candidates. a,b) HLA binding motif counts. c,d) number of unfavorable amino acids. In all graphs: natural sequences are marked with black circles (●);consensus sequences with blue triangles (▲); inferred ancestral sequences with green squares (■); and mosaic sequences with red diamonds (♦). Left panel (graphs a and c) shows HLA-binding-motif counts (a) and counts of unfavorable amino acids (c) calculated for individual sequences; Right panel (graphs b) and d) shows HLA binding motifs counts (b) and counts of unfavorable amino acids (d) calculated for sequence cocktails. The top portion of each graph (box-and-whiskers graph) shows the distribution of respective counts (motif counts or counts of unfavorable amino acids) based either on alignment of M group sequences (for individual sequences, graphs a) and c) or on 100 randomly composed cocktails of three sequences, one from each A, B and C subtypes (for sequence cocktails, graphs b) and d). The alignment was downloaded from the Los Alamos HIV database. The box extends from the 25 percentile to the 75 percentile, with the line at the median. The whiskers extending outside the box show the highest and lowest values. Amino acids that are very rarely found as C-terminal anchor residues are G, S, T, P, N, Q, D, E, and H, and tend to be small, polar, or negatively charged[1]. Results are shown for Gag, but the same qualitative results hold for Nef core and complete Nef. The same procedure was done for supertype motifs with results qualitatively similar to the results for HLA binding motifs (data not shown).

1. Yusim, K. et al. Clustering patterns of cytotoxic T-lymphocyte epitopes in human immunodeficiency virus type 1 (HIV-1) proteins reveal imprints of immune evasion on HIV-1 global variation. *J Virol* **76**, 8757-68 (2002).