

Additional File 2: A stringent “hit/no-hit” model of 10-mer coverage was applied to a dataset of 2285 Gag sequences using a tetra-valent ACDE natural sequence vaccine formulation as the query set. As shown below a near-Gaussian distribution of the percent coverage of all isolates in the dataset was obtained. Near identical mean coverage results were obtained using the Epicover algorithm (as implemented in the LANL HIV Sequence Database) and a larger dataset of 3585 Gag sequences. While a broad distribution of “hits per isolate” was obtained (range = 13%-75%), the near identical mean coverage results for the Epicover algorithm and the “hit/no-hit” model resulted in no difference between the two methods for the calculation of epitope coverage requirements and the multiple epitope hit requirement model.

