



Fig S8. Comparison of the entropy per residue of the amino acid MSA and that of the binarized MSA for (A) HIV Gag, (B) HIV Nef, (C) HCV NS3-4A, and (D) HCV NS4B. A high positive Pearson correlation r (close to 1) with very high statistical significance (very small P) is obtained between the entropy per residue of the amino acid MSA (H_i) and that of the binarized MSA (H_i^{bin}) in all cases; for details on the entropy computation, see [1]. This demonstrates that the binary MSA is a good approximation of the amino acid MSA for the considered viral proteins.

References

1. Quadeer AA, Louie RHY, Shekhar K, Chakraborty AK, Hsing IM, McKay MR. Statistical linkage analysis of substitutions in patient-derived sequences of genotype 1a hepatitis C virus nonstructural protein 3 exposes targets for immunogen design. *J Virol.* 2014;88(13):7628–44. doi:10.1128/JVI.03812-13.