Biophysical Journal, Volume 99

Supporting Material

Title: Discovery of Entry Inhibitors for HIV-1 via a New De Novo Protein

Authors: Meghan L Bellows, Martin S Taylor, Philip A Cole, Lin Shen, Robert F Siliciano, Hoki Fung, and Christodoulos A Floudas

Ensemble Generation

Structure Prediction

A 3-dimensional (3D) structure of each sequence is needed. This is done using RosettaAbinitio, part of the Rosetta++ software package. The strategy behind the RosettaAbinitio algorithm is based upon experimental observation that the local structure of the protein is influenced, but not uniquely determined, by the local sequence of the protein. The final overall protein structure is determined when the modification of local structures come together to give a compact structure, accounting for non-local interactions such as buried hydrophobic residues, paired β strands, and specific side chain interactions.

RosettaAbinitio uses Monte Carlo to determine the final folded structure of the sequences. Since these structures represent local minima in free energy, not necessarily the global minimum, 1000 possible structures, or decoys, are generated per sequence.

Clustering

The 1000 decoys from the structure prediction step are then clustered based upon their φ and ψ angles using OREO in order to elucidate representative backbone structures of the whole set. The average structure from the ten largest clusters is chosen to dock to the target protein. In addition, the overall lowest-energy structure is selected. This gives 11 backbone structures for each peptide sequence, incorporating backbone flexibility into the ensemble generation.

Clustering ensures that 11 unique peptide backbone conformations are selected for each sequence. Typically, 15-30 clusters are formed, with a Cα RMSD among the structures in a cluster around 0.5-2 Å. This indicates that the average structure selected is very representative of the cluster. The Cα RMSD of the average structures selected from each cluster differ from each other by 3 - 5 Å.

Docking Prediction

Docking prediction is done using RosettaDock. For each sequence, each of the 11 peptide backbone structures is docked against the target protein. In this case, since the binding site is known, the peptides are placed near the binding site and allowed to translate 3 Å normal to the binding site, 8 Å parallel to the binding site, and rotate 8°. RosettaDock also uses Monte Carlo, so 1000 decoys for each docking run are generated. The ten lowest energy decoys in each of the 11 runs are used as starting structures in the final rotamerically-based conformation ensemble generation (110 starting structures per sequence).

The conformations that are selected from the docking prediction introduce further flexibility into the ensemble. These conformations differ by 0.5 – 3 Å (Cα RMSD) and this spread is consistent regardless of sequence or peptide backbone structure. Selecting the lowest-energy docked structures ensures that conformations that will contribute highly to the partition function (i.e. lowest energy) are included in the ensemble.

Final Ensemble Generation

RosettaDesign is used to generate the final rotamerically-based conformation ensemble because it can be used to generate a number of decoys by only adjusting the rotamers on the side-chains. It therefore leaves the backbone and sequence intact, which is exactly what is needed to apply the K* approach. RosettaDesign is given a number of starting structures, and for each structure, a residue is randomly chosen and the rotamer changed. The probability of accepting a move (P_{accept}) is based upon the Metropolis Criterion (Eq. 1), where ΔE is the change in energy of a move and T is the temperature.

$$P_{accept} = \min(1, e^{-\Delta E/T}) \qquad (1)$$

This is repeated until thousands of rotamer substitutions are attempted and gives a final low-energy conformation that will contribute highly to the partition function. The full atom RMSD among the rotamer conformations are in the range of 0.5 - 1 Å and is consistent among all sequences.

To generate the peptide ensemble, the ten lowest-energy peptide structures

from each of the ten largest clusters plus the ten overall lowest-energy peptide structures are used as starting structures for RosettaDesign (110 total starting structures). For each starting structure, 200 rotamer conformers are generated, giving a final ensemble of 22,000 structures. The ensemble incorporates both backbone flexibility (by using 110 different backbone starting structures) and rotamer flexibility (by generated 200 rotamer conformers per starting structure).

The complex ensemble is generated similarly by taking the 110 starting structures from the docking prediction step and generating 200 rotamer conformers per starting structure. The final ensemble size is 22,000 structures. Flexibility is taken into account by the various peptide backbone structures used (11 different backbones total), the various docked conformations (10 per peptide backbone), and the rotamer conformers for each starting structure.

The protein ensemble is generated by running RosettaDesign on just the target protein structure (typically a crystal structure from the PDB). In this case, 2000 rotamer conformations are generated for the single starting structure, so the final ensemble size is 2000 structures. Backbone flexibility is not introduced in this ensemble because the target protein often has hundreds of residues, making it computationally infeasible to predict backbone structures.

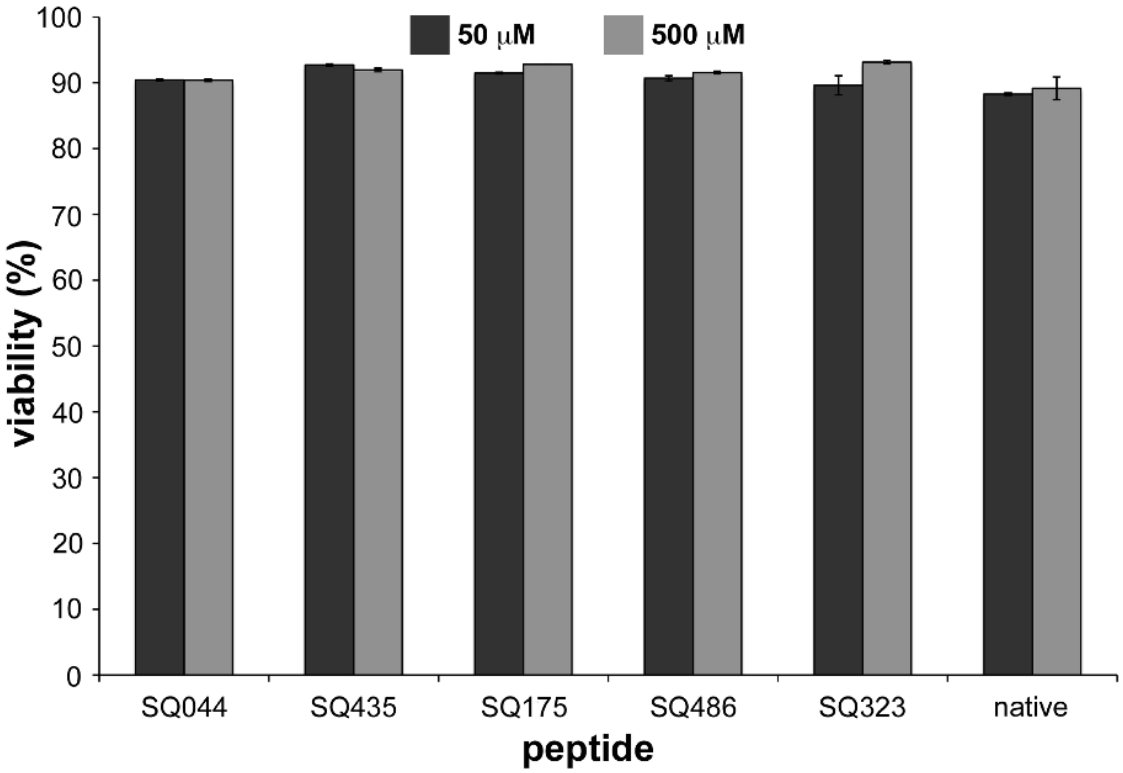

Figure S1: Viability results for the top five mutant sequences and the native sequence. Experiments were done in duplicate from one cell donor at two peptide concentrations, 50μM and 500μM. 100% viability indicates no cell deaths due to the peptide. The error bars indicate the span of values obtained.

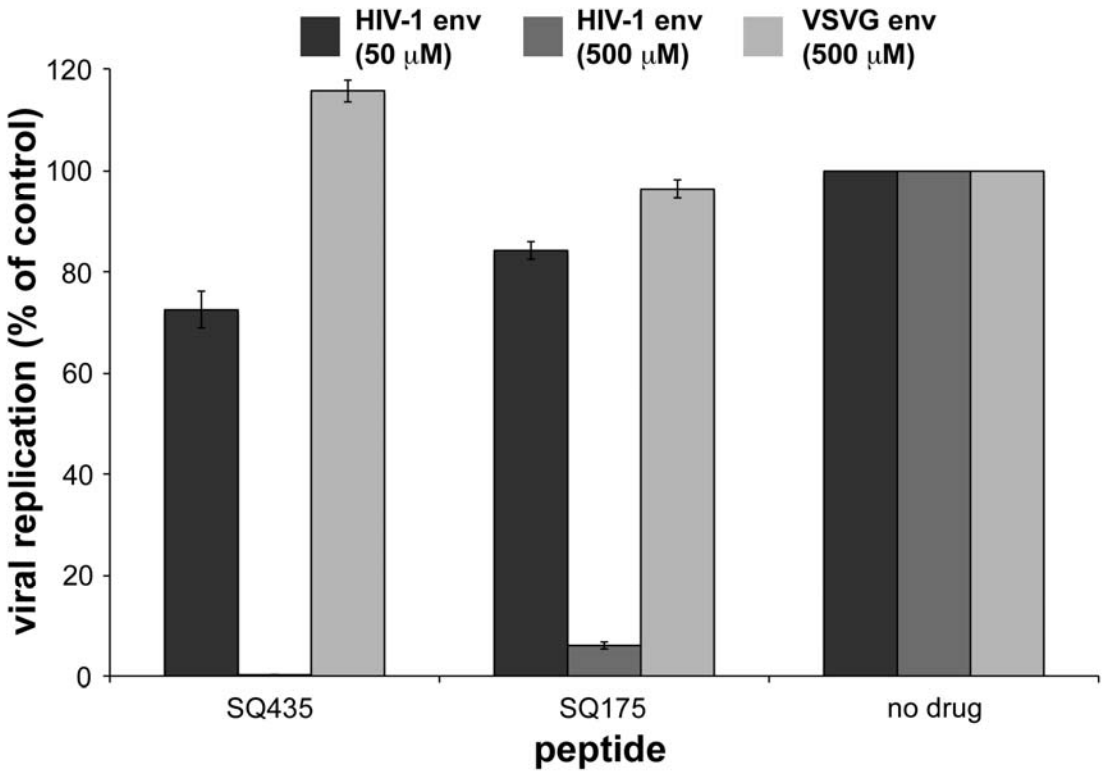

Figure S2: Inhibition of entry of wild type HIV and HIV enclosed in a VSVG capsid into primary CD4+ T cells by SQ435 and SQ175. Experiments were done in duplicate with cells from one. Replication (% of control) was determined by single-round infectivity assay with GFP-encoding recombinant HIV. Experiments were carried out at two peptide concentrations for the HIV-1 envelope, 50μM and 500μM, and at one peptide concentration for the VSVG envelope, 500μM. The error bars indicate the span of values obtained.

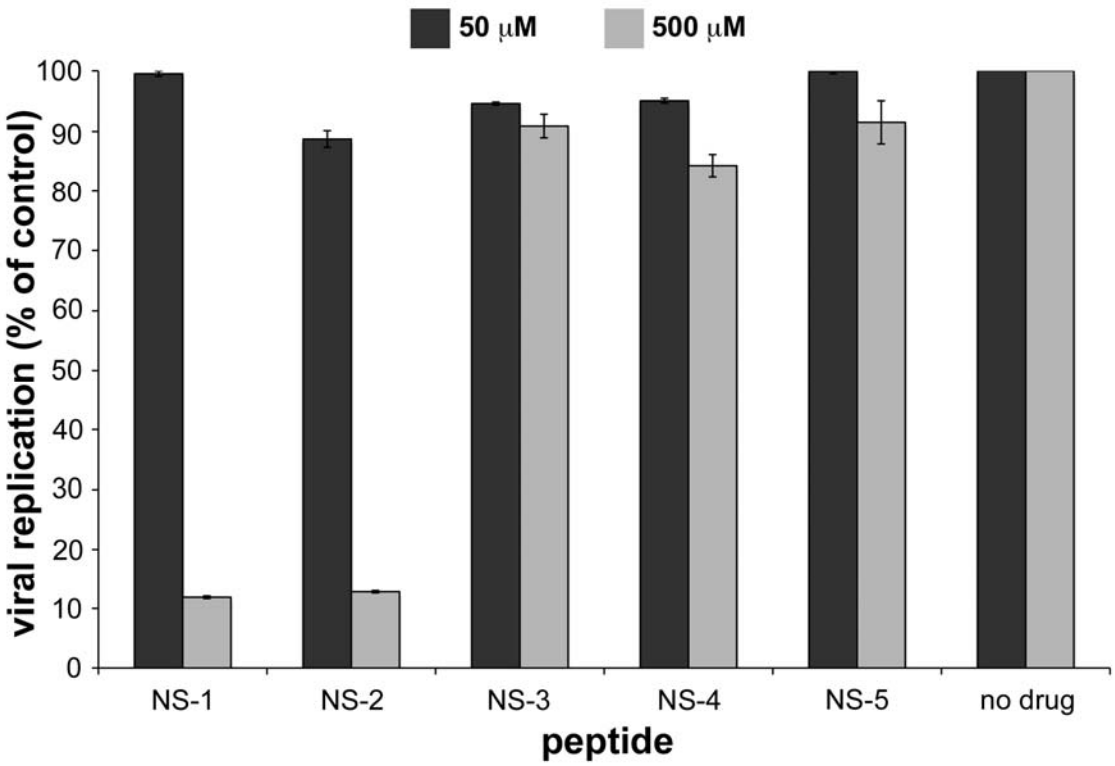

Figure S3: Inhibition of entry of wild type HIV into primary CD4+ T cells by five new sequences. Inhibition results for the five new sequences. Experiments were done in duplicate with cells from two donors and at two peptide concentrations, 50μM and 500μM. Replication (% of control) was determined by single-round infectivity assay with GFP-encoding recombinant HIV. The error bars indicate the span of values obtained.