Supporting Information

Conservation of the GGA bulge in the HIV-1 frameshift site stem-loop

The GGA bulge sequence in the HIV-1 frameshift site stem-loop is highly conserved (Figure S1). Only two out of the seven residues comprising the three-purine bulge and its flanking base pairs are variable; residue 16 (as in Fig. 1a) can be either a uridine or a cytidine, and residue 43 can be either a guanosine or an adenosine. These variants conserve base-pairing between residue 16 and residue 41, and also conserve the three-purine nature of the bulge. Overall, this observation suggests the presence of similar bulge structures in all HIV-1 sequences analyzed. Additional constraints on the HIV-1 frameshift-inducing RNA sequence such as the need to maintain protein function in two reading frames, as well as the conservation of the frameshift-inducing RNA structure, may necessarily result in a low degree of mutation, as previously suggested[1-5]; thus the virus may not readily evolve resistance to small molecules targeting this site.

Supplementary Figure Legends

Figure S1. HIV-$1_{8\text{-}52}$ sequence conservation. The consensus sequences of 1204 HIV-1/SIVcpz Group M (all subtypes A-K) and 619 HIV-1 Group M subtype B sequences are shown. Sequences were obtained from the Los Alamos National Laboratories HIV Sequence Database (http://www.hiv.lanl.gov/). Sequence conservation is shown according to the figure inset. Residues conserved in fewer than 80% of the sequences also have the second most prevalent nucleotide indicated. The exact percentage of guanosine to adenosine mutations in the conserved GGA sequence is indicated.

Figure S2. Twenty selected compounds. These compounds were selected for further study based on their favorable fluorescence and UV absorbance profiles.

Figure S3. Preliminary fluorescence-monitored compound titrations. Each titration was performed a single time to provide an initial measure of the apparent binding affinity of the compound for the HIV-1_{screen} RNA. Data are shown in the same grid positions as in Figure S3. The fluorescence response has been normalized relative to the maximum level observable by the instrument ($3.8*10^6$) and is plotted along the ordinate axis for each graph in arbitrary units. In some instances, the fluorescence response was multiplied by 10 to allow the graph to fit in the grid space; this is indicated on the graph by a (10x). When possible, the estimated dissociation constant has been indicated. The four compounds selected for further examination are indicated by black stars.

Figure S4. Structurally similar compound classes. Compounds were clustered by atom pair similarity using the Tanimoto coefficient[6, 7] and a similarity cutoff of 0.5. Clustering was performed in R (http://www.R-project.org). The number of compounds populating each class is indicated in the upper-left corner of each grid box. Representative compounds from each class are shown as examples.

Figure S5. Model of doxorubicin interaction with the HIV-1 frameshift site RNA. Doxorubicin was hand-docked to the HIV-1 RNA in an orientation similar to that found in a daunorubicin-DNA structure[8] (PDB: 1D33) using Sybyl 7.3 (Tripos, Inc.). Doxorubicin was oriented with its amino group positioned in a manner where it may hydrogen-bond to the amino of the central

guanosine in the three purine bulge in the HIV-1 RNA. The anthracyclene conjugated ring system of doxorubicin was positioned to stack upon the U-A base-pair at the top of the lower helix in the RNA. Charges were added to the complex by the Gasteiger-Huckel method (Tripos), followed by 500 steps of energy minimization using the Powell algorithm and the Tripos force-field. The stem-loop of the RNA is colored in red, the lower helix in blue, and the GGA bulge with its flanking base-pairs is colored in violet. Doxorubicin is shown as a ball-and-stick model.

Supplementary Figures

Figure S1. HIV-1_{8-52} sequence conservation.

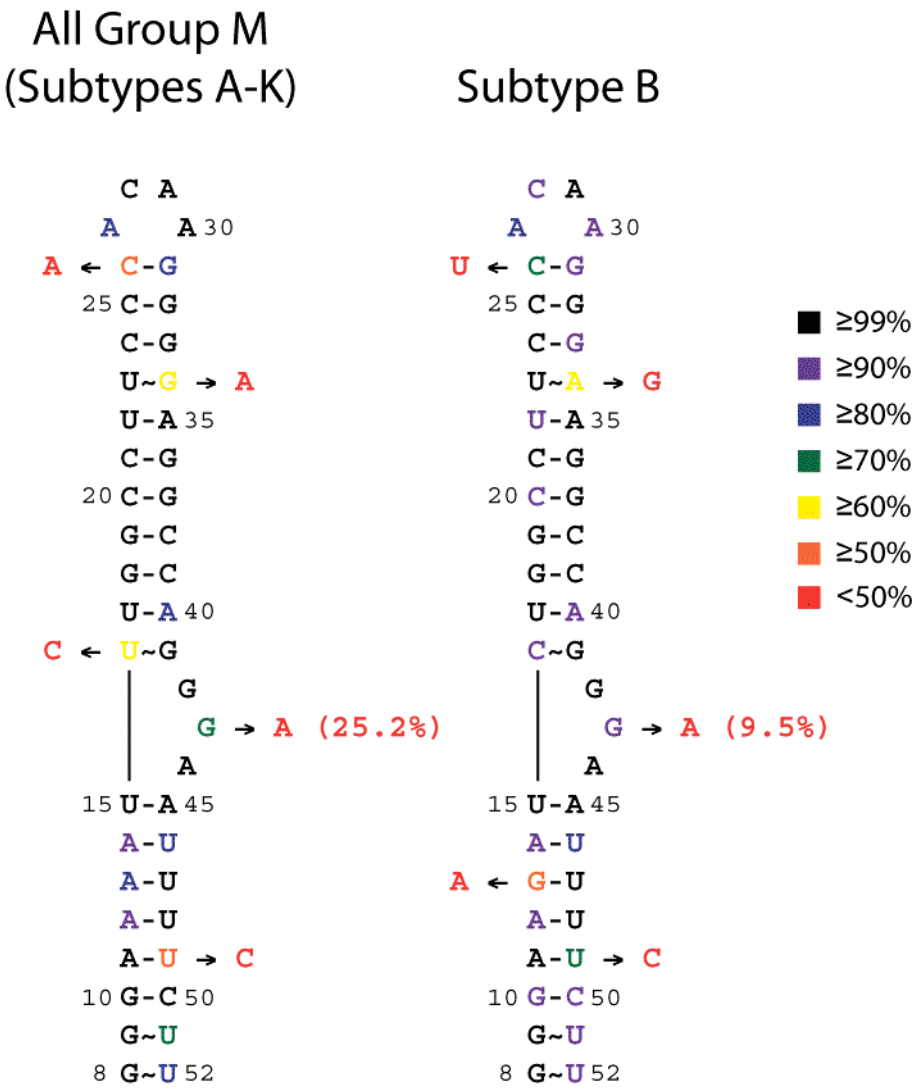


Figure S2. Twenty selected compounds.

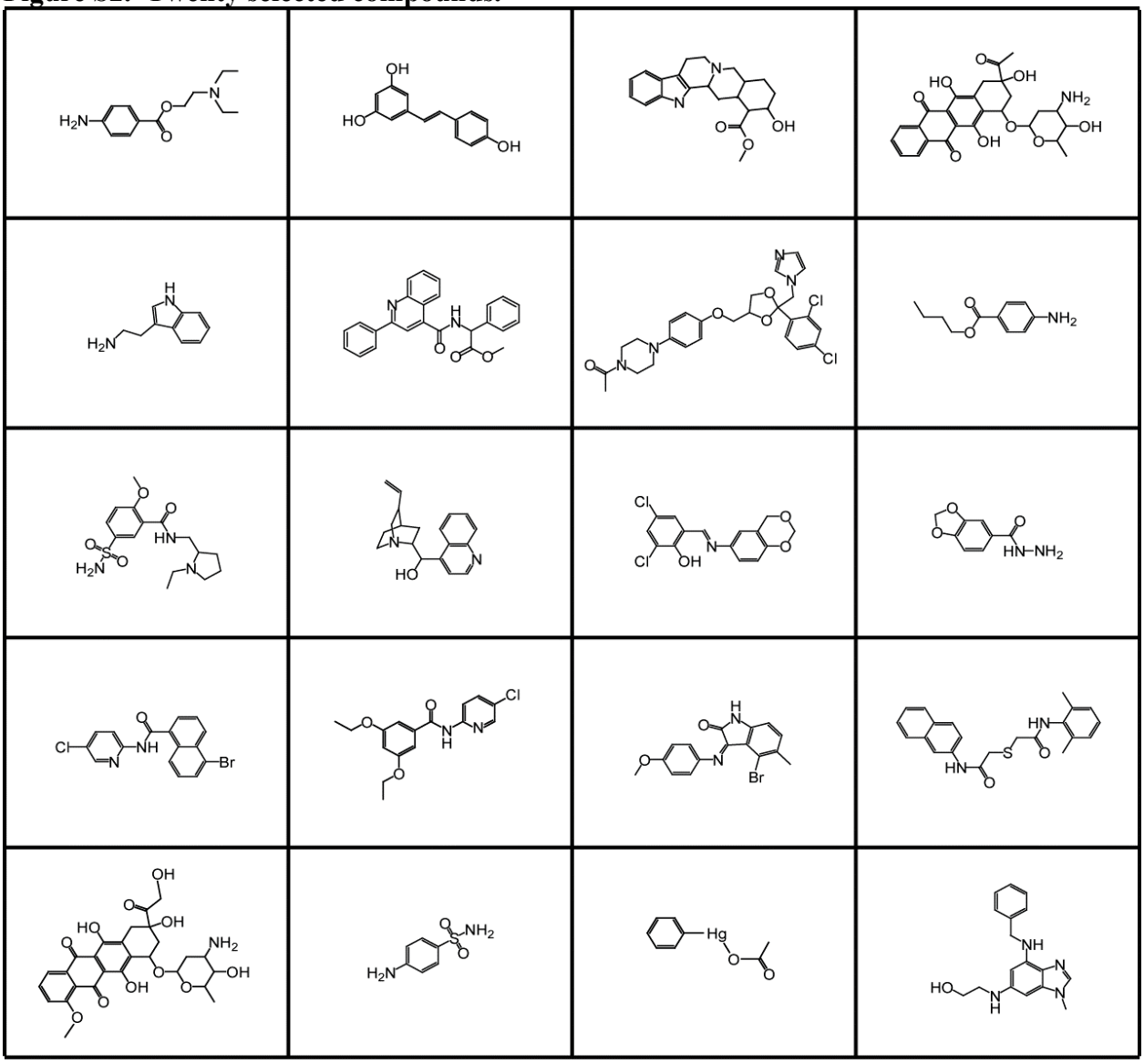


Figure S3. Preliminary fluorescence-monitored compound titrations.

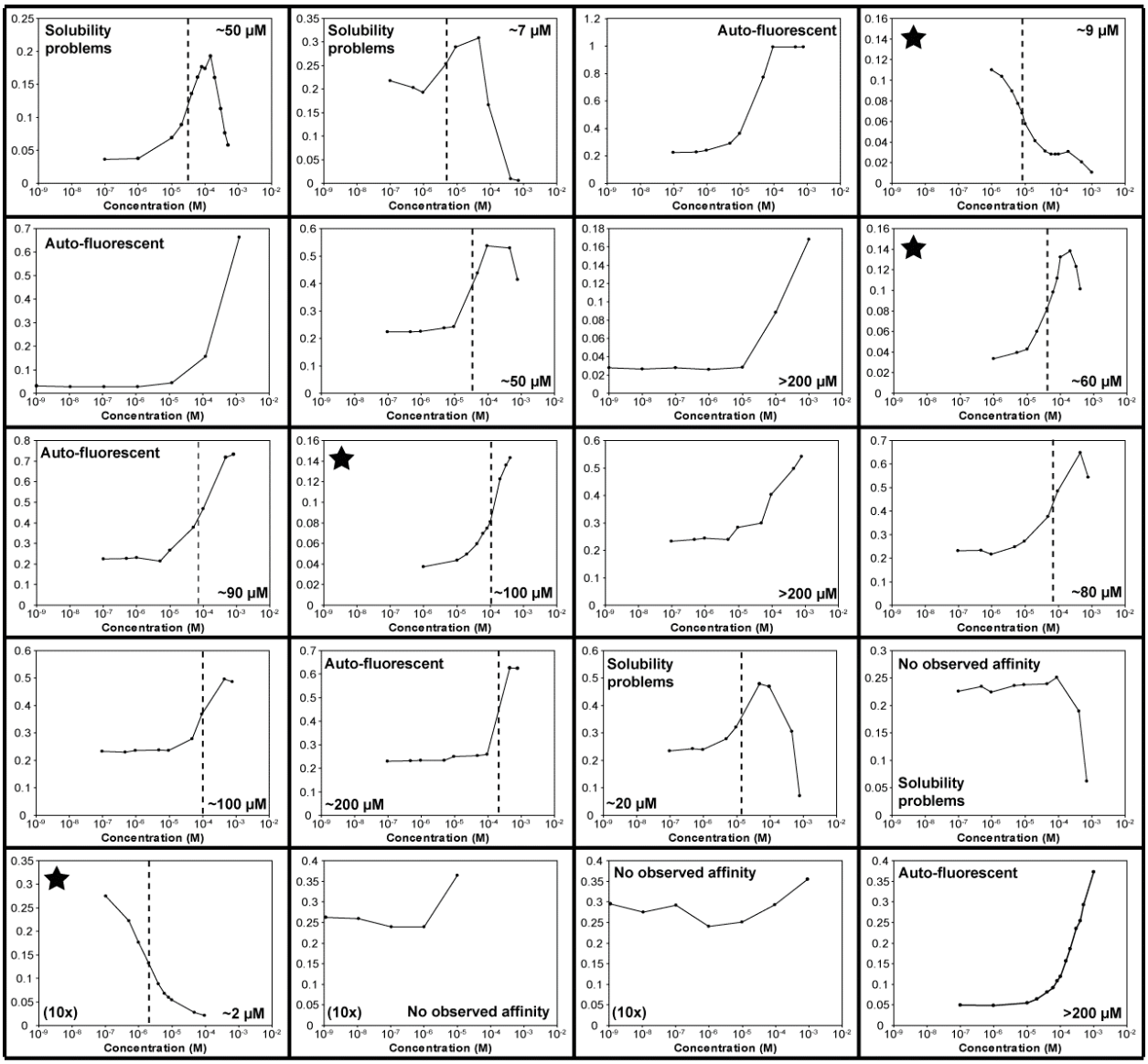


Figure S4. Structurally similar compound classes.

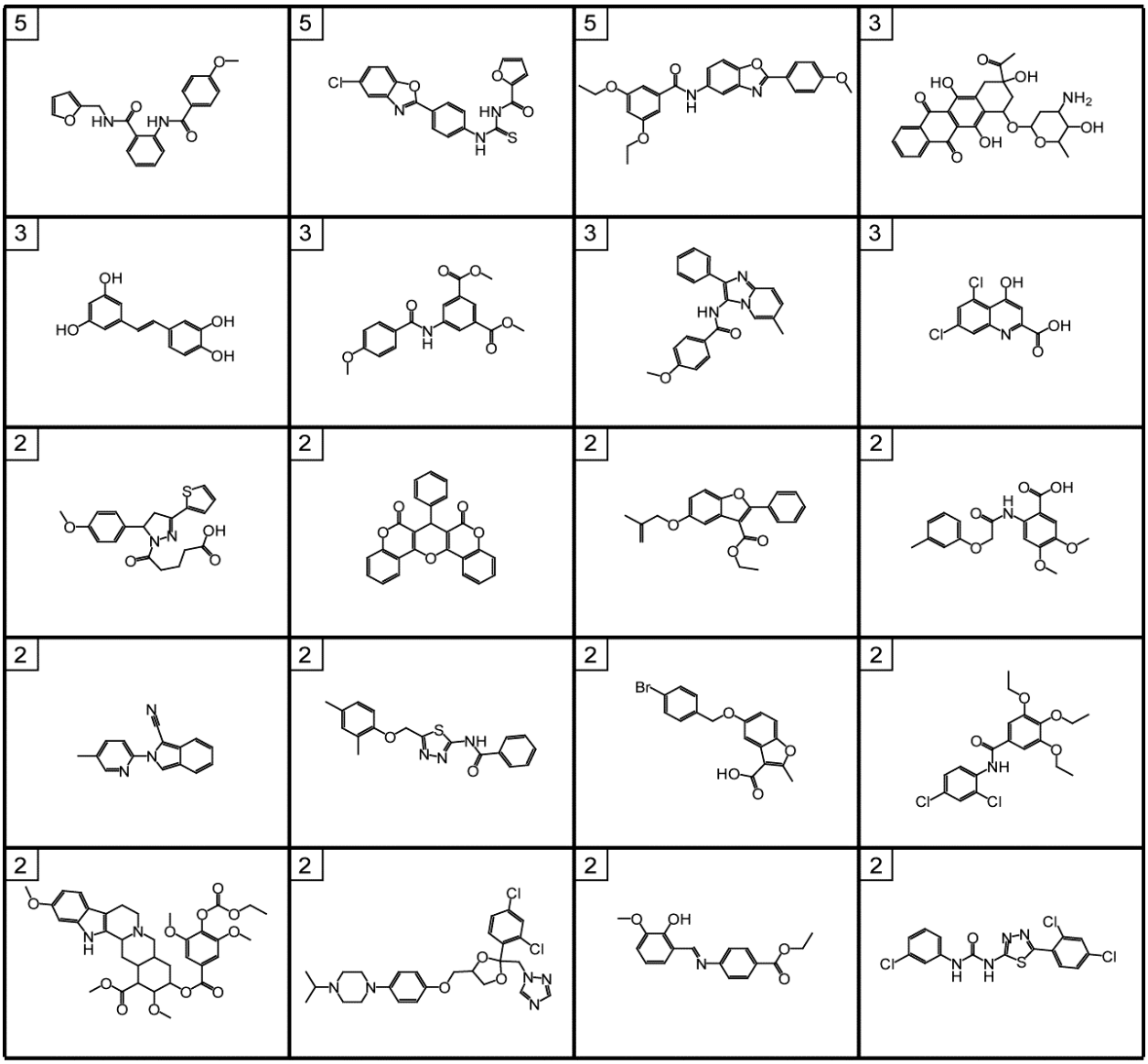
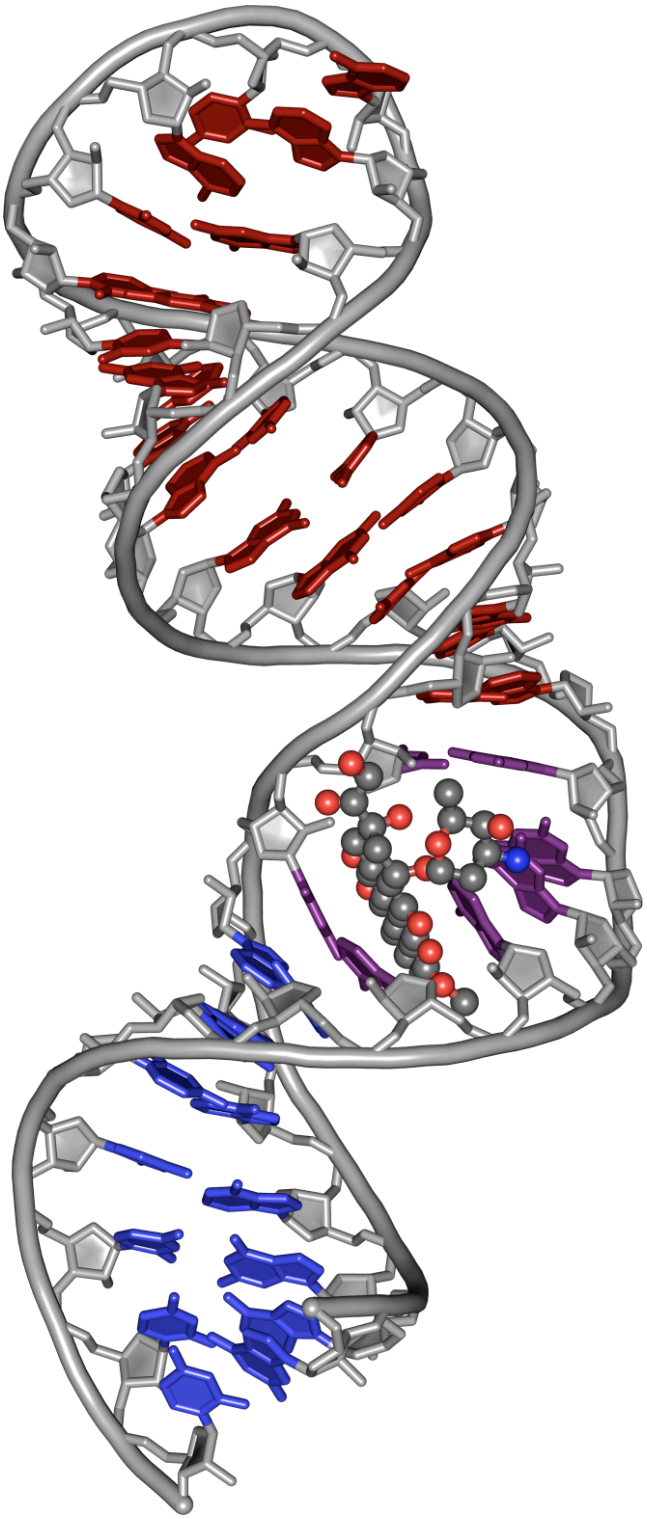


Figure S5. Model of doxorubicin interaction with the HIV-1 frameshift site RNA



Supplementary Methods

Frameshifting assay, continued. Complementary synthetic oligonucleotides (Integrated DNA Technologies, Inc.) representing a naturally occurring HIV-1 group M subtype B sequence (http://www.hiv.lanl.gov, accession number AB078005) with BamH I and Sac I compatible ends were cloned into the p2luc plasmid using the unique BamH I and Sac I sites between the *rluc* and *fluc* reporter genes. Oligonucleotides comprising the sequence (5'-GATCC**T**TT**T**TTAGGGAAGATCTGGCCTTCCCACAAGGGAAGGCCAGGGAATTTTCTT CAGAGCAGACCAGAGCCAACAGCCGCACCGAGCT-3'), and its complements were phosphorylated, annealed, and ligated into the p2luc vector to produce the experimental constructs. This places the *fluc* gene in the -1 reading frame relative to *rluc*; analogous to the orientation of the *gag* and *pol* genes in the HIV-1 genome. Positive control sequences and their complements were also cloned into p2luc and have two thymidine residues (bold) in the slippery sequence (underlined) replaced with cytidines and an additional nucleotide that places the *rluc* and *fluc* genes in-frame inserted immediately before the Sac I complementary sequence (GAGCT). Resultant products were transformed into *E. coli* competent cells (JM109, Promega). Plasmid DNA was purified from cell cultures (Qiagen) and the sequences of all constructs were verified (University of Wisconsin Biotechnology Center). RNA was transcribed as previously described[9].

Luminescence was measured using a Veritas microplate luminometer equipped with dual-injectors (Turner Biosystems), as previously described[9]. For each compound concentration, a minimum of 10 replicates each of the experimental and positive control reactions were assayed using the Dual-Luciferase Reporter Assay System (Promega).

Ratios of firefly/*Renilla* luminescence were calculated for each of the experimental and control reactions at the various compound concentrations. Replicate experiments were normalized to their respective controls and averaged. The ratio of averages (experimental/control) for each compound concentration was then determined, and frameshifting efficiencies were reported. Measurement error was propagated through these calculations to yield a final s.e.m. for each of the determined frameshift efficiencies.

Statistical Analyses. Data were analyzed using Prism 4.3 (GraphPad). The built-in one-way ANOVA function with a Dunnett's Multiple Comparison post-test was used to analyze the data. Data are plotted with associated s.e.m. in all instances.

Supplementary References

1. Dulude, D.; Berchiche, Y. A.; Gendron, K.; Brakier-Gingras, L.; Heveker, N., Decreasing the frameshift efficiency translates into an equivalent reduction of the replication of the human immunodeficiency virus type 1. *Virology* **2006,** 345, (1), 127-36.
2. Baril, M.; Dulude, D.; Gendron, K.; Lemay, G.; Brakier-Gingras, L., Efficiency of a programmed -1 ribosomal frameshift in the different subtypes of the human immunodeficiency virus type 1 group M. *RNA* **2003,** 9, (10), 1246-53.
3. Gareiss, P. C.; Miller, B. L., Ribosomal frameshifting: an emerging drug target for HIV. *Curr Opin Investig Drugs* **2009,** 10, (2), 121-8.
4. Brierley, I.; Dos Ramos, F. J., Programmed ribosomal frameshifting in HIV-1 and the SARS-CoV. *Virus Res* **2006,** 119, (1), 29-42.
5. Biswas, P.; Jiang, X.; Pacchia, A. L.; Dougherty, J. P.; Peltz, S. W., The human immunodeficiency virus type 1 ribosomal frameshifting site is an invariant sequence determinant and an important target for antiviral therapy. *J Virol* **2004,** 78, (4), 2082-7.
6. Fligner, M. A.; Verducci, J. S.; Blower, P. E., A Modification of the Jaccardâ€“Tanimoto Similarity Index for Diverse Selection of Chemical Compounds Using Binary Strings. *Technometrics* **2002,** 44, (2), 110-119.
7. Godden, J. W.; Xue, L.; Bajorath, J., Combinatorial preferences affect molecular similarity/diversity calculations using binary fingerprints and Tanimoto coefficients. *J Chem Inf Comput Sci* **2000,** 40, (1), 163-6.
8. Wang, A. H.; Gao, Y. G.; Liaw, Y. C.; Li, Y. K., Formaldehyde cross-links daunorubicin and DNA efficiently: HPLC and X-ray diffraction studies. *Biochemistry* **1991,** 30, (16), 3812-5.
9. Marcheschi, R. J.; Staple, D. W.; Butcher, S. E., Programmed ribosomal frameshifting in SIV is induced by a highly structured RNA stem-loop. *J Mol Biol* **2007,** 373, (3), 652-63.