ARTICLE

ACS chemical biology

Antiviral Compounds Discovered by Virtual Screening of Small−Molecule Libraries against Dengue Virus E Protein

Zhigang Zhou[†,§], Mansoora Khaliq[‡], Jae-Eun Suk[†], Chinmay Patkar[‡,¶], Long Li[‡], Richard J. Kuhn[‡], and Carol Beth Post[†,‡,*]

[†]Department of Medicinal Chemistry and Molecular Pharmacology and [‡]Department of Biological Sciences, Markey Center for Structural Biology and Purdue Cancer Center, Purdue University, West Lafayette, Indiana 47907. [§]Present address: Computational Biology Branch, NCBI NIH, Bethesda, MD 20894-6075. [¶]Present address: Center for Infectious Disease and Vaccine Research, Univeristy of Massachusetts Medical School, Worcester, MA 01655.

ABSTRACT Infection by the mosquito-borne dengue virus causes dengue fever and the sometimes fatal dengue hemorrhagic fever. The increasing number of dengue infections per year suggests that the virus is becoming more virulent and its transmission is expanding. Nevertheless, no effective treatment for dengue infection currently exists. In a search for antiviral agents effective against dengue virus, we investigated the potential of targeting a structural protein site rather than an enzymatic one. Using this approach, we now report the discovery of a small molecule ligand that inhibits viral growth. Our results also provide the first evidence that the binding site, a pocket located at the hinge between domains 1 and 2 of the envelope protein (E protein) on the virus surface, is a valid target for antiviral therapy. Ligand candidates were identified from libraries of ~142,000 compounds using a computational high-throughput screening protocol targeting this pocket of the E protein. Cell-based assays were conducted on 23 top-ranked compounds. Among four with good antiviral activity profiles, the compound P02 was found to inhibit viral reproduction at micromolar concentrations. Using saturation transfer difference NMR spectroscopy, we also show that the compound binds virus and competes for binding E protein with the known ligand *N*-octyl-β-D-glucoside. Together, the results are consistent with an inhibition mechanism against maturation or host-cell entry mediated by ligand binding to the E-protein pocket. P02 is a promising lead compound for future development of an effective treatment against dengue virus and related flaviviruses.

Mosquito-borne flaviviruses are human pathogens and a major burden in many regions of the world by causing diseases that include yellow fever and the sometimes fatal dengue hemorrhagic fever and the dengue shock syndrome (*1*). In 2007 the four serotypes of dengue viruses were estimated to cause 50−100 million annual human infections worldwide and 22,000 deaths. The number of dengue fever cases per year is increasing steadily, including in the United States where dengue virus has spread to 36 states since 1985, and the risk of an outbreak is recognized (*2*). Nevertheless, there are no known antiviral compounds and no therapeutic treatment against dengue virus. Safe vaccines against the yellow fever flavivirus exist, and progress toward a dengue virus vaccine is being made, but the availability of vaccines is often limited, and in the United States people are not likely to be vaccinated. Therefore, a better defense against outbreaks of dengue infection and exposure to the disease is offered by antiviral agents as a treatment strategy.

Dengue virus is an enveloped virus with a ssRNA genome that is translated as a polyprotein. Cleavage of the polyprotein produces three structural proteins (capsid, C; premembrane, prM; and the glycosylated envelope, E, protein) that form the virus particle and seven nonstructural proteins that form the replicase complex for genome replication. The immature dengue virus at neutral pH is an icosahedral particle with a diameter of ~600 Å and an external coat of trimers of prM:E heterodimers, as shown by cryo-electron microscopy reconstruction (*3*). Notable protrusions occur at the position of the 6-fold symmetry axis with significant open regions

*Corresponding author, cbp@purdue.edu.

Received for review May 9, 2008 and accepted October 6, 2008.

Published online November 24, 2008

10.1021/cb800176t CCC: $40.75

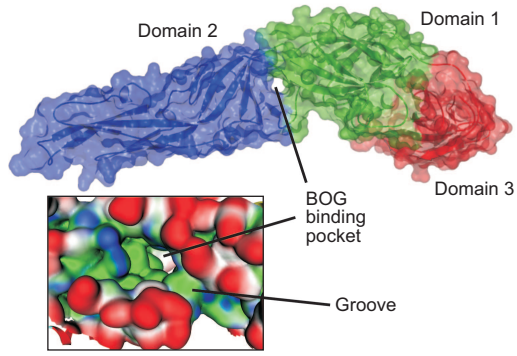


Figure 1. Monomeric dengue E protein showing the targeted BOG site. Upper: Surface representation of E protein domains 1 (green), 2 (blue), and 3 (red). Lower: Expanded view of the BOG site and groove region with solvent-accessible surface with 1.4 Å radius solvent probe (red) and buried surface (green for hydrophobic, blue for hydrophilic).

in the surface exposing the lipid. By contrast, the mature virus is ~500 Å in diameter and 90 E:E homodimers coat the viral surface (*4*) by forming a close-packed smooth, outer shell of uniform thickness. E protein comprises three domains and is elongated in shape (Figure 1).

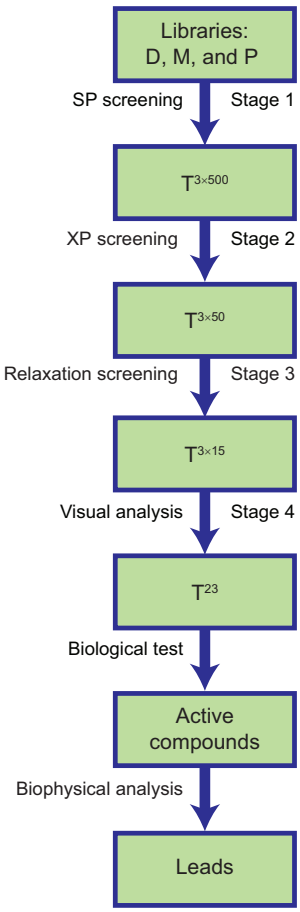
Structural comparisons of E protein in the immature and mature virus support an approach for inhibitor design distinct from pursuing the more common target of viral enzymes. Fitting into cryo-EM density for these two viral forms requires a change in the relative orientation of domains 1 and 2 (*5*). Further variation in the domain orientation was observed in the crystallographic structure of a postfusion form of E protein (*6*). The overall differences in the angular orientation between domains 1 and 2 span ~35° with the largest variation occurring between the mature and postfusion forms. In addition, the interface between domains 1 and 2 has been identified as a site for ligand binding; *N*-octyl-β-D-glucoside (BOG) was observed in one crystallographic structure (*7*) to lie in a buried pocket formed between domains 1 and 2 (Figure 1). This pocket appears to result from an induced-fit binding because in other structures of the E protein from dengue (*5, 8, 9*) or tick-borne encephalitis virus (*10*) a loop connecting the two domains adopts an alternative conformation that closes over and eliminates the pocket. The large-scale structural changes between immature and mature virus, including the differences in domain orientations of E protein, suggest ligand binding to this BOG pocket could inhibit a step in the virus lifecycle at which the virus E protein undergoes a structural transition (*7*). A similar strategy of targeting a buried pocket in a viral capsid protein was exploited with human rhinovirus (*11*), leading to the development of numerous antiviral compounds (*12–14*).

Here we report the use of a four-stage computational high-throughput screening (HTS) (*15, 16*) of three NCI compound libraries to identify small-molecule ligands with potential to bind the BOG pocket in dengue E protein. The screening of >142,000 compounds to select 23 for experimental validation is described. The results from in-cell assays and from saturation transfer difference (STD) NMR spectroscopy, a direct measure of binding and competition with BOG, are then presented. We propose that a compound discovered to have antiviral activity and shown to bind E protein is a strong lead candidate for the development of new therapeutics against dengue virus and related flaviviruses.

RESULTS AND DISCUSSIONS

Screening NCI Libraries against the BOG Pocket. A four-stage protocol (Scheme 1) was used to computationally screen ligands to bind the BOG pocket of E protein and enhance the potential for finding antiflavivirus activity in a biological assay conducted on a smaller scale. The compounds in the diversity, D, (1990 compounds) and mechanistic, M, (879 compounds) libraries have diverse or mechanistically desirable properties, respectively, and are largely subsets of the significantly larger plated, P, library (~140,000 molecules). When this study was initiated, the compounds were available from NCI for biological evaluation. The reliability of the docking method applied to the targeted site of dengue E protein was demonstrated by docking BOG and comparing the results with the known structure of the crystallographic complex (see Supporting Information).

Stages 1 and 2: SP and XP Docking. The ranked Gscore and Emodel scores for the three libraries following standard precision (SP) docking (stage 1) are shown in Figure 2, where lower ranked compounds are plotted by increasing value along the abscissa. Supplementary Figure S2 shows analogous plots for stage 2 extra-precision (XP) docking. The trends in Figure 2 are similar to a Gaussian profile with a rapid initial drop over a small number of the most favorable compounds, followed by more slowly decreasing scores for a substan-



Scheme 1. Virtual screening flow chart for the Diversity (D, 1990 compounds), Mechanistic (M, 879 compounds), and Plated (P, ~140,000 compounds) NCI libraries

tially larger number of compounds and a sharp fall off as the scores approach zero in value. The percentages of ligand complexes with negative scores are comparable in all cases: 1342 (67.4%), 578 (65.8%), and 90,337 (64.5%) of the molecules from the D, M and P libraries, respectively, for Gscore profiles. Rankings with Gscore and Emodel differ, although there is a 15−20% overlap in the stage 1 top-ranked compounds. The arrows in Figure 2 delineate the cutoff values needed for either Gscore or Emodel to select the top-ranked 250 compounds to compose the $T^{3\times500}$ set for stage 2. The best P library scores (Figure 2, panels e and f) are superior to those for D and M libraries (Figure 2, panels a−d). Recalling that D and M libraries are a subset of P, many more than 3000 ligands for Gscore and Emodel would have been required to carry forward to stage 1 in order to include the D and M compounds selected for $T^{3\times500}$.

Stage 3: Thermal Relaxation. Stage 3 allowed internal flexibility of both protein and ligand atoms by energy minimization. The calculated energies for the $T^{3\times50}$ set are provided in Supplementary Table S2.

Compounds were ranked by the interaction energy (Figure 3, panel a), which includes a solvation contribution. Overall, the van der Waals energy contributed most to the interaction energy, but the electrostatic energy showed the greatest variation and was therefore the major factor for the ranking of molecules in stage 3. Energy minimization of the combined set of 2×25 compounds selected on the basis of either Gscore or Emodel from stage 2 resulted in a continuous energy profile, which suggests both scoring functions favor compounds similarly suited for the BOG binding site. In addition, there was no overlap in these top-ranked 50 ligand molecules among the three libraries even though the D and M libraries are largely subsets of the P library. Independent screening of the three overlapping libraries therefore selected three exclusive sets of potential compounds for binding E protein.

The Gscore value is plotted in Figure 3, panel b against the same compound rankings as Figure 3, panel a. It is clear from the many maxima and minima in Figure 3, panel b that GScore does not directly paral-

TABLE 1. Biological activity measured for compounds selected from computational HTS

		IC_{50} (μM)	
Compound	CC_{50} (μM)[a]	YFV-IRES-Luc virus[b]	YF-RLuc2A-RP[c]
D01	95 ± 6	28 ± 5	31 ± 7
D02	343 ± 80	98 ± 39	181 ± 30
D03	61 ± 3	31 ± 0.3	12 ± 5
D04	350 ± 2	70 ± 5	230 ± 40
D05	880 ± 85	500 ± 10	1300 ± 400
M01	309 ± 5	179 ± 21	27 ± 8
M02	70 ± 12	51 ± 7	N/A
P01	563 ± 33	376 ± 177	195 ± 6
P02	371 ± 47	13 ± 3	17 ± 3

[a]Host-cell cytotoxicity. [b]Inhibition in viral growth. [c]Inhibition of replication of the YF-R.luc2A-RP replicon.

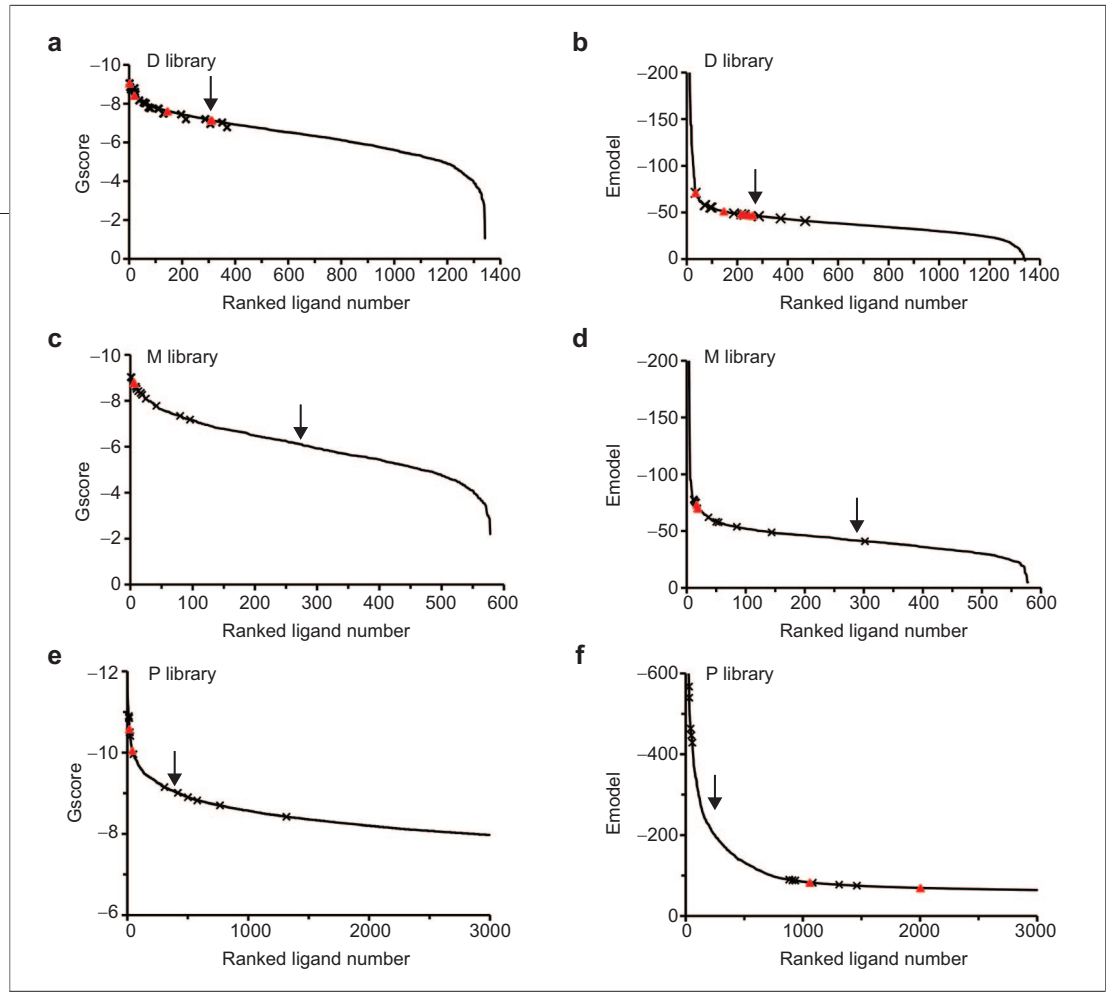


Figure 2. Stage 1, ranking all → $T^{3\times500}$ set. GScore (panels a, c, and e) and Emodel (panels b, d, and f) scores ranked for libraries D (panels a and b), M (panels c and d), and P (panels e and f) from SP screening. The arrow indicates the cutoff points for selection of the $T^{3\times500}$ set, ×'s indicate the compounds selected for the $T^{3\times15}$ set. Red triangles indicate the compounds active in the YFV-IRES-Luc virus growth assay (Table 1). Symbols on the right of the arrow indicate active compounds that were selected by computational screening based on the alternative scoring function.

lel the interaction energy, and therefore both energy and GScore were used to select the candidates for the final graphical analysis. This $T^{3\times15}$ set was composed of 10 molecules defined by the most favorable interaction energy list and 5 defined by the most favorable Gscore for each library.

Chemical properties of the $T^{3\times50}$ and $T^{3\times15}$ molecules from stage 2 and stage 3, respectively, were analyzed to identify the features selected by docking. The presence of hydrogen bonding and cationic and anionic groups was evaluated, as well as the amount of hydrophobic and aromatic contact surface area (Supplementary Table S3). The largest change in these features to evolve over the stage of the screening is the increase in ligand hydrogen donor groups. As expected on the basis of the hydrophobic nature of the buried BOG binding site, few of the selected compounds have cationic (4% and 2%) or anionic groups (2% for both sets).

Stage 4: Visual Analysis. The final selection for the T^{23} set to submit to experimental testing considered the chemical nature of the ligand and the structural quality of the fit in the binding site. The T^{23} set comprised 11, 7, and 5 molecules from the D, M, and P libraries, respectively.

Inhibition of YFV-IRES-Luc Virus Growth. To test the ability of the T^{23} compounds to inhibit viral growth, confluent BHK cells were infected with YFV-IRES-Luc at a low multiplicity of infection (MOI) and inhibition of virus growth was determined from the reduction in luciferase activity (see Methods). Of the 23 compounds tested, nine compounds (Figure 4) showed inhibitory activity against YFV-IRES-Luc growth (Table 1) at concentrations less than CC_{50}. IC_{50} values ranged from ~13 μM for P02 to 500 μM. The antiviral activity of P02 is well-substantiated by the more than 25-fold lower IC_{50} value compared with CC_{50}, although the viral reduction activities of the other compounds shown in Table 1 appear to be toxicity-mediated.

The antiviral activity of compound P02 was further characterized by determining its effect on virus production measured from a virus titer. The levels of virus release in the presence of 10 μM and 100 μM P02 were

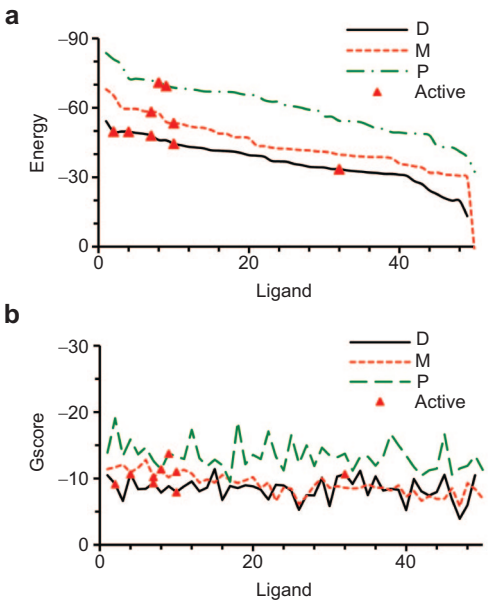


Figure 3. Stage 3, ranking $T^{3\times50} \rightarrow T^{3\times15}$. a) Binding energy and b) GScore ranked by energy minimization of the $T^{3\times50}$ set from the D, M, and P libraries. The red triangles indicate the active compounds. The $T^{3\times15}$ set comprises the top 10 compounds based on energy and the top 5 based on Gscore. In panel b, compounds are ranked the same as in panel a.

determined by plaque assay (data provided in Supporting Information). As expected, the virus titer is reduced ~10 orders of magnitude at 100 μM P02.

Inhibition of YF-R.luc2A-RP Replication. To investigate the possibility that the observed inhibition of YFV-IRES-Luc virus resulted from interference with viral protein production or genome RNA replication, the compounds were tested for inhibition of replication of the YF-R.luc2A-RP replicon (Table 1). The YF-R.luc2A-RP replicon lacks the viral structural proteins, including E protein. Thus, activity in this assay reflects the effect of compounds on genome replication or protein synthesis, as opposed to assembly and maturation of virus particles or host-cell entry. Only the nine compounds that were active against YFV-IRES-Luc were tested in this assay, in which luciferase activity indicates the level of genome RNA replication. IC_{50} values are listed in Table 1 (IC_{50} YF-RLuc2A-RP). For the compounds D02, D04, and D05, IC_{50} values obtained for inhibition of YF-R.luc2A-RP replication are higher than those for inhibiting YFV-IRES-Luc virus growth, while P02 has comparable activities

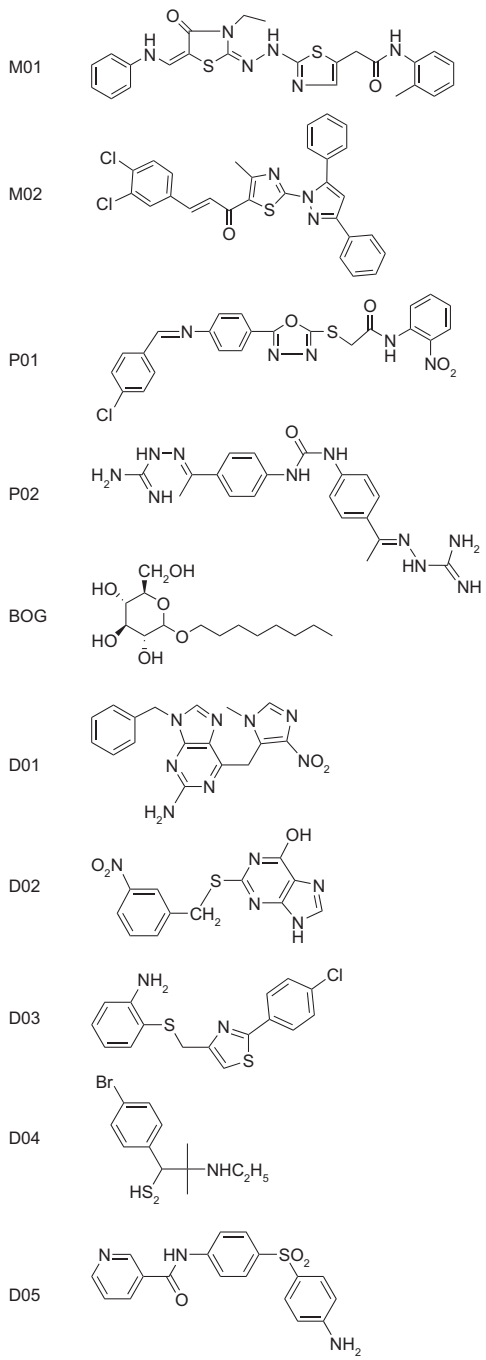


Figure 4. Molecular structures of the compounds listed in Table 1 and BOG.

in viral growth and replicon assays. These results suggest that compounds D02, D04, and D05 inhibit the vi-

rus life cycle at steps other than replication, consistent with inhibition of maturation or virus entry into cells as a result of binding E protein, and P02 activity may involve multiple targets.

Binding Measured by NMR. NMR, including saturation transfer difference (STD) NMR, was used to directly test the binding of a compound with antiviral growth activity to either whole virus or isolated dengue E protein. One compound only has been tested thus far due to the substantial amount of whole virus or purified E protein required for NMR. Ligand resonances appear in an STD spectrum as a result of the transfer of magnetization from the large-molecular-weight species to a bound small-molecule ligand; efficient transfer is possible only when the ligand and macromolecular protons are spatially close.

STD measurements were first conducted in the presence of yellow fever virus for the compound P02, which has the lowest IC_{50} (Table 1, 13 μM) for inhibiting YFV-IRES-Luc virus growth. The results (Supplementary Figure S3) indicate P02 binds the virus. P02 likely associates with E protein given that E protein covers the entire external surface of the virus and the lipid layer is largely inaccessible (*4*).

To further test for specific binding of the compound P02 to E protein and explore binding to the BOG pocket targeted in the computational screening, NMR experiments were conducted on purified dengue E protein with BOG only and BOG in the presence of P02 to determine if P02 competes with BOG (drawn in Figure 4 along with the nine compounds from Table 1). The reference 1D ^{1}H spectrum of a mixture of P02 (0.45 mM) and BOG (0.45 mM) in the absence of E protein is shown in Figure 5, panel a. Resonances from BOG and P02 are labeled. Upon addition of dengue E protein at a concentration (20 μM) too low to detect, the 1D spectral intensities of the P02 resonances are significantly reduced (Figure 5, panel b). The loss of intensity indicates that P02 associates with E protein, and the unbound state observed by the resonances in Figure 5 is in slow exchange with the E protein-bound state. On the other hand, changes are imperceptible in Figure 5, panel b for resonances from BOG, the known E protein ligand (*7*), so that under the conditions of the experiment, BOG is in fast exchange between the free and bound states. The differences in the NMR exchange behavior of P02 and BOG indicate that the binding affinity of P02 is greater than that of BOG. For the sample conditions used here and assuming typical chemical shift differences between free and bound ligands, the slow exchange behavior of P02 is consistent with a dissociation constant of 1 μM or smaller.

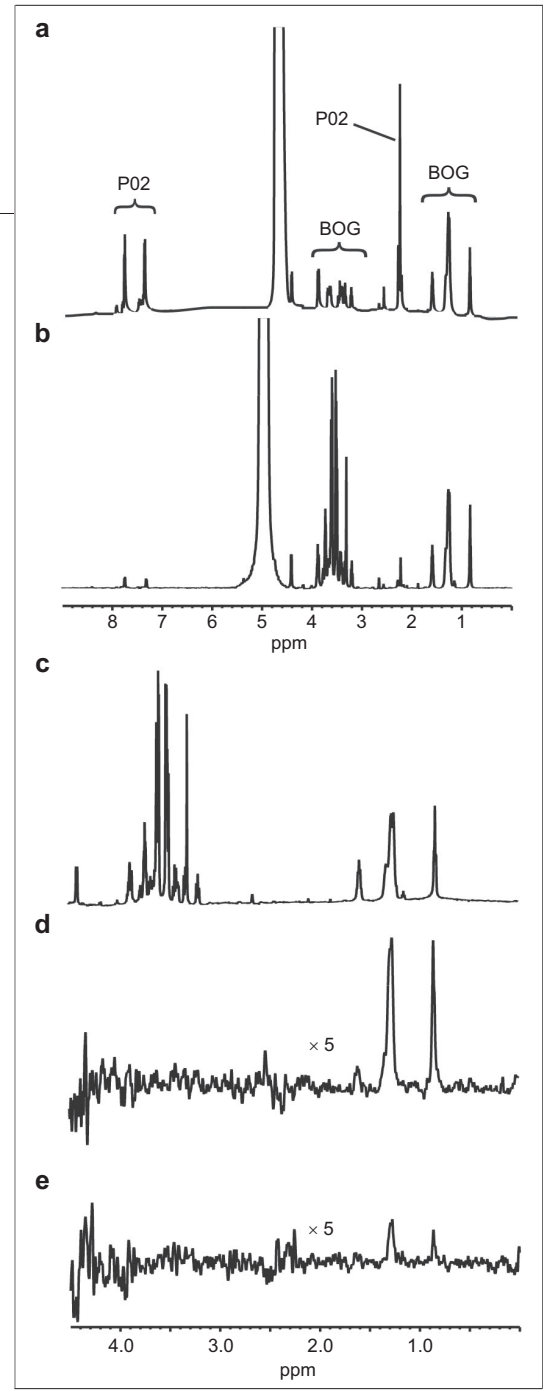


Figure 5. NMR spectra (600 MHz, 277 K) to assess P02 binding to dengue E protein. a) 1D ^{1}H reference spectrum of P02 and BOG with the resonances labeled. The large resonance at 4.8 ppm is from water. b) Same as panel a after addition of dengue E protein. c) ^{1}H reference spectrum of BOG plus dengue E protein with no irradiation. d) ^{1}H STD spectrum for sample shown in panel c with on-resonance irradiation at −2 ppm. e) same as panel d after addition of compound P02.

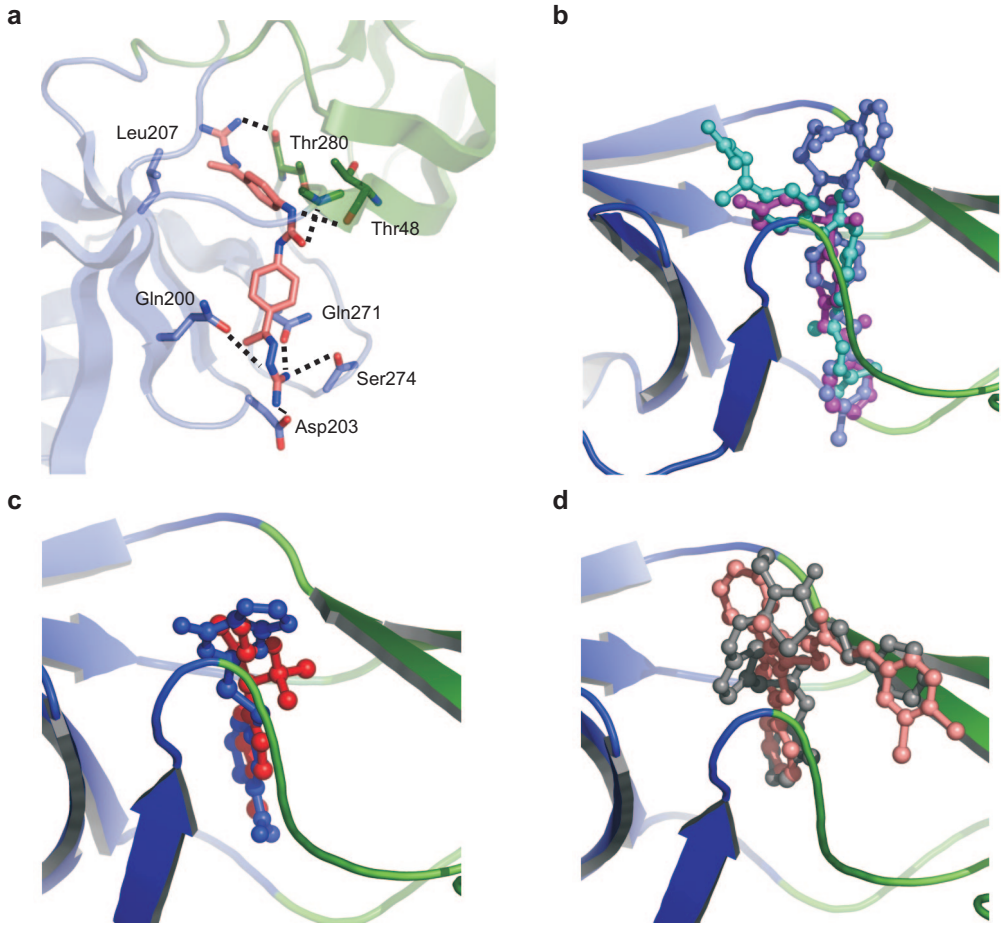


Figure 6. Binding features of the active compounds (Table 1) docked in the BOG channel of E protein. Ribbon drawing of E protein colored light green for domain 1 and light blue for domain 2. The molecule is rotated ~90° about a horizontal axis from Figure 1. Ligands are shown by stick (panel a) or ball-and-stick (panels b–d) models. a) P02 contacts with selected E protein residues drawn as sticks. b) Longer ligands P01, P02, and D05. c) Shorter ligands D03 and D04. d) Ligands M01 and M02 extend into the groove (upper right).

Competition experiments were conducted to further explore the nature of P02 binding. STD experiments were measured by selective irradiation of E protein at low concentration with either BOG only or BOG in the presence of P02. The ^{1}H spectrum of BOG plus dengue E protein, without saturation of the E protein resonances, is provided in Figure 5, panel c for reference. The STD spectrum of this sample calculated from spectra measured with and without saturation (Figure 5, panel d) has strong signals corresponding to the BOG resonances at 0.8 and 1.3 ppm. Thus there is efficient magnetization transfer from E protein to BOG upon association. Addition of P02 to this sample of BOG plus dengue E protein results in nearly complete loss of intensity from the BOG resonances in an STD spectrum (Figure 5, panel e). The loss of these signals indicates that association of BOG with E protein is greatly reduced and that P02 competes with BOG for binding to E protein. (Resonances from P02 are not observed in the STD spectrum because P02 is in slow exchange and magnetization is transferred to the bound-state P02 resonances, which are too broad and low intensity to detect in the STD spectrum.) The data in Figure 5 provide unambiguous evidence that P02 associates with E pro-

tein and competes with BOG for binding. This competition supports the proposal that P02 binds in the BOG pocket.

Assessment of Computational Screening. How the scores of the nine compounds with antiviral activity (Table 1) ranked during the course of the computational screening was examined to assess the utility of screening three libraries, where D and M are relatively small subsets of the P library, and the use of both Gscore and Emodel rankings.

The nine compounds with antiviral activity, marked by red triangles in Figures 2 and 3, are observed to come from all three libraries; however, the compounds identified from the D and M libraries scored well below the cutoff value used for the P library in stage 1 (see Supporting Information) and did not pass the selection filter for screening the large P library. Therefore the objective of finding a set enriched with active compounds was met, although the method may have failed to extract all potentially active compounds from the large number of molecules in the structurally redundant P library.

Gscore ranking at stage 1 screening of the large P library yielded active compounds, but Emodel ranking did not. This result is shown by the occurrence of red triangles to the left of the arrow in Figure 2, panel e, whereas none occur in Figure 2, panel f (also shown in Supplementary Figure 2). Because most compounds with antiviral activity were identified by the Gscore and Emodel rankings were less efficient, we conclude that the predictive power of Gscore is superior.

Binding Site Interactions of Docked Compounds. The BOG binding pocket is a channel with open access at both ends so that linear molecules of different length can be accommodated.

Details of the interactions predicted for P02, the compound demonstrated here by NMR to bind E protein (see above), are shown in Figure 6, panel a. As docked, P02 has high chemical complementary with the channel of E protein. Numerous energetically favorable interactions are observed, including many from the E protein polar residues noted above. P02 heteroatoms are in close contact with side chains of Gln200, Asp203, Gln271, Ser274, and Thr48 and main chain peptide groups of Thr280 and Ala50. Favorable hydrophobic contacts, illustrated with Leu207 in Figure 6, panel a, are also observed.

The docked structures for the nine compounds listed in Table 1 cluster into three types of complexes (Figure 6, panels b–d). P02 is oriented similarly to P01 and D05 (Figure 6, panel b), which are also relatively longer molecules and span the full length of the binding channel. These ligands also make contact with the polar groups from E protein located near the end of the channel (Glu49, Gln200, Asp203, and Gln271). Exploitation of these interactions in future design efforts could enhance binding affinity and confer specificity, as previously recognized from simulations (*17, 18*). The docked orientations of compounds D02 and D04 (Figure 6, panel c) and (M01 and M02 in Figure 6, panel d) are highly overlapped with the phenyl rings nearly superimposed and located in a hydrophobic area of the barrel. M01 and M02 have an additional group positioned to the right in Figure 6, panel d and lies in a shallow groove in the E protein surface (Figure 1). This additional contact area has the potential to enhance binding affinity by burial of hydrophobic surface of the protein and ligand.

Conclusion. A hierarchical four-stage computational HTS was used to identify small-molecule compounds that bind to the BOG pocket of the E protein of dengue virus with the goal of developing antiviral agents against flaviviruses. The screening reduced ~142,000 molecules from three NCI libraries to only 23 top-ranked compounds, which were tested for antiviral activity in biological assays. P02 was demonstrated both to bind E protein and to have antiviral activity. Moreover, P02 was shown by STD NMR to compete with BOG for binding to E protein, which unambiguously demonstrates association of P02 with E protein and supports a binding site that overlaps the BOG pocket targeted by computational screening. P02 binding in the BOG pocket is reasonable in view of the numerous favorable intermolecular interactions between P02 and BOG-site residues (Figure 6, panel a) in the modeled complex. The evidence for binding E protein and the antiviral activity, together, support P02 as a strong lead compound for future design efforts and provide a foundation for exploring an antiviral mechanism based on disruption of the essential viral process of maturation or cell entry by stabilizing selected conformational states of flaviviral structural proteins in a manner similar to that known for rhinovirus (*14, 19–22*).

Three additional compounds (D02, D04, and D05) were found to have activity against virus growth in the

μM concentration range, while exhibiting reduced activity in a viral replication assay (Table 1). Although the cytotoxicity of these compounds infers a toxicity-mediated viral reduction, this activity profile, which is consistent with inhibition of the virus life cycle by blocking virus entry into cells or disruption of virus maturation, nonetheless renders these compounds worthy of consideration for future development given the need and early stage of identifying inhibitors against dengue virus.

In accordance with other recent studies (*15, 16*), the results reported here are strong testimony to the value of the computational screening approach. The objective to identify ligands of dengue virus E protein relied on computational docking with Glide. Ligand flexibility was taken into account throughout the computation, starting with the first stage of docking by use of course-grained docking of explicit rotamers generated for all compounds. Careful visual inspection of the complexes after energy relaxation yielded the final set of the number of compounds targeted for biological assays. There was no overlap in the top 50 ligand molecules ranked out of the three libraries even though the D and M libraries are subsets of the P library, and thus independent screening of these three overlapping libraries selected a unique set of potential E-protein ligands from each library.

METHODS

Screening Strategy. A hierarchical four-stage protocol (Scheme 1) was implemented for HTS against dengue E protein using three NCI small-molecule libraries: diversity (D), mechanistic (M), and plated (P). The objective was to balance the simulation time and docking accuracy in order to select ~50 compounds, a small number amenable to visual inspection and biological assays. Details on the setup and scoring are in Supporting Information.

The initial screening stage used the SP docking procedure in Glide (*FirstDiscovery* v2.7, Schrödinger, Inc.) (*23, 24*), designed for handling a large number of compounds and minimizing the number of false negatives by using a "soft" potential that permits minor steric clashes. Multiple conformations were generated for each molecule in the library by exhaustive enumeration of the energy minima for the rotatable torsion angles and prescreened for docking by eliminating high-energy conformers. Generated poses of a specific ligand conformer positioned in the active site were scored and ranked by Gscore and Emodel functions, and the top 500 compounds from each of the three libraries were selected to carry forward to the second stage screening. Of the 500 compounds, 250 were selected on the basis of Gscore ranking and 250 on the basis of Emodel ranking. In the case of some number (N) of compounds ranked at the top with both Gscore and Emodel, one-half of these ($N/2$) were retained for each scoring function and the cutoff increased until 250 compounds were selected. These 1500 compounds defined the $T^{3\times500}$ set.

The second stage utilized the XP docking procedure of Glide (*23–25*), which incorporates a more accurate, finer-grained docking algorithm designed to eliminate false positives that survive the SP stage 1. As in SP docking, the XP protocol includes ligand flexibility by docking multiple conformers in a rigid receptor, and the resulting complexes were ranked by both Gscore and Emodel. The top 50 molecules from each of the three libraries (25 from each ranking by Gscore or Emodel) generated the $T^{3\times50}$ set of compounds.

Stage 3 involved thermal relaxation of the $T^{3\times50}$ set of molecules. The receptor−ligand complexes were refined by energy minimization executed with the Liaison module (*FirstDiscovery* v2.7). The energy was minimized by the truncated Newton method with optimization of the atomic coordinates of all ligand atoms and protein atoms within 15 Å of any ligand atom while the atomic coordinates of all other protein atoms were fixed. The energy function was the OPLS-AA force field (*26*), including van der Waals and electrostatic terms. Solvation was modeled implicitly using the surface generalized Born model, which includes the solvent reaction-field energy, U_{rxnf}, and the cavity energy, U_{cav}, proportional to the change in exposed surface area of the ligand. A 10 Å cutoff was used for long-range interactions. Energy minimization was carried out for 500 steps, or until the system reached a root-mean-square gradient less than or equal to 0.01 kcal mol^{-1} $Å^{-1}$. This final stage generated the set $T^{3\times15}$, or 15 compounds from each of the P, D and M libraries, selected on the basis of either the energy function used for minimization (10 compounds) or Gscore (5 compounds) criteria.

Stage 4 was visual examination of the selected complexes and assessment of the compound physico/chemical features. $T^{3\times15}$ molecules were evaluated for a maximal number of functional groups, minimal solvent exposure and all ligand atoms having significant contact within the binding site, maximal number of intermolecular hydrogen bonds, and balance between the hydrophobic and hydrophilic groups of the ligand to maintain a proper partition coefficient.

YFV-IRES-Luc Virus Growth Assay. BHK cells were plated in a 96-well plate and grown at 37 °C. At confluency, cells were infected with YFV-IRES-Luc virus at a MOI of 0.1. Details on cell type and generation of YFV-IRES-Luc virus are in Supporting Information. A low MOI was utilized to ensure that fewer cells were infected so that spread of released virus could be monitored. After 1 h of incubation with the virus, cells were overlaid with culture media containing serial dilutions of compounds at concentrations below the CC_{50} values (see below). Compound stock solutions were in DMSO solvent. Controls included uninfected cells, infected cells, and DMSO-treated infected cells. Cells were incubated at 37 °C, 5% CO_2 for ~36 h and lysed using 50 μL of Cell Culture lysis buffer (Promega Inc.), and 10 μL samples of cell extracts were placed into a 96-well opaque plate. Luciferase activity was determined from the luminescence generated with fire-fly luciferase substrate (Promega Inc.). Luminescense was measured in a 96-well-plate luminometer, LMax II (Molecular Devices). A reduction in luciferase activity indicates inhibition of YFV-IRES-Luc virus growth. The luciferase luminescence as a function of compound concentration was analyzed by nonlinear regression analysis using GraphPad Prizm to estimate the IC_{50} of each compound. The IC_{50} was defined as the concentration of the compound to cause 50% reduction of luciferase activity in infected cells as compared to the DMSO-treated cells.

YF-R.luc2A-RP Replication Assay. BHK cells were plated in a 96-well plate and grown at 37 °C. At confluency, cells were infected with YFV pseudoinfectious particles (PIPs) (see Supporting Information) containing YF-R.Luc2A-RP. Cells were then overlaid with culture media containing serial dilutions of compounds at concentrations below the CC_{50} values. Controls included uninfected cells, infected cells, and DMSO-treated infected cells. Cells were incubated at 37 °C, 5% CO_2 for ~36 h and lysed using 50 μL of *Renilla* lysis buffer (Promega Inc.), and 10 μL samples of cell extracts were placed into a 96-well opaque plate. Luciferase activity was determined as described above.

Cell Viability Assay. In accordance with the manufacturer's protocol for determining cell viability as a measurement of cytotoxicity, BHK cells were plated in a 96-well plate at a confluency of $\sim 4 \times 10^4$ cells/well. The cells were incubated at 37 °C for 24 h to a confluency of $\sim 5 \times 10^5$ cells/well, at which time the cells were overlaid with culture media containing serial dilutions of compounds. Compound stocks were generated by dissolving compounds in DMSO and added to the media to give a final DMSO concentration of 1%. Untreated and 1% DMSO-only treated cells served as positive controls. Cells were then incubated at 37 °C, 5% CO_2 for ~36 h. At ~36 h post-treatment, the media on the cells was replaced with fresh media to remove the compounds. Next, 10 μL of XTT-substrate from the Quick Cell Proliferation Kit was added to each well. Cells were incubated at 37 °C for a further 2 h. Plates were then removed, and the optical density, OD_{450}, was measured using a 96-well plate reader (Molecular Devices) to determine the amount of formazan dye generated by cellular mitochondrial dehydrogenases, as an indicator of cell viability. CC_{50} is the concentration of compound that reduces the OD_{450} value by 50% compared to that for DMSO-only treated cells. OD_{450} values for untreated cells were nearly equal to those for DMSO-only treated cells.

STD NMR. STD NMR experiments (*27*) were performed on a Varian 600 MHz spectrometer at 277 K, without sample spinning. E protein purification is described in Supporting Information. Samples were in 12 mM *d*-Tris, 100 mM NaCl, 95% D_2O, and 5% *d*-DMSO solution. 1D STD spectra were obtained *via* phase cycling by subtracting an on-resonance FID, with selective saturation of protein resonances at −2 ppm, from an off-resonance FID, with saturation at 30 ppm, using a series of 50 ms Gaussian-shaped saturation pulses (2 s total saturation time) and a 2-s recycle delay. For high quality data, 1D ^{1}H STD spectra were recorded with 10 K data points covering 8000 Hz, and up to 16 K difference scans were accumulated. During the signal accumulation, FIDs over different time periods (320 scans, ~1280 scans, and ~10 K scans) were stored separately to compare against one another.

Acknowledgment: This work was sponsored by the NIH/NIAID Regional Center of Excellence for Biodefense and Emerging Infectious Diseases Research (RCE) Program. The authors wish to acknowledge membership within and support from the Region V "Great Lakes" RCE (NIH award 1-U54-AI-057153). The authors thank M. Cushman, L. Ze, and R. Perera for valuable discussions and S. Lok for assistance in the preparation of E protein.

Supporting Information Available: This material is free of charge *via* the Internet.

REFERENCES

1. Lindenbach, B. D., and Rice, C. M. (2001) Flaviviridae: The viruses and their replication, in *Fields Virology* (Knipe, D., and Howley, P., Eds.), pp 991−1041, Lippincott Williams & Wilkins, Philadelphia.
2. Morens, D. M., and Fauci, A. S. (2008) Dengue and hemorrhagic fever: A potential threat to public health in the United States, *JAMA, J. Am. Med. Assoc. 299*, 214−216.
3. Zhang, Y., Corver, J., Chipman, P. R., Zhang, W., Pletnev, S. V., Sedlak, D., Baker, T. S., Strauss, J. H., Kuhn, R. J., and Rossmann, M. G. (2003) Structures of immature flavivirus particles, *EMBO J. 22*, 2604−2613.
4. Kuhn, R. J., Zhang, W., Rossmann, M. G., Pletnev, S. V., Corver, J., Lenches, E., Jones, C. T., Mukhopadhyay, S., Chipman, P. R., Strauss, E. G., Baker, T. S., and Strauss, J. H. (2002) Structure of dengue virus: Implications for flavivirus organization, maturation, and fusion, *Cell 108*, 717−725.
5. Zhang, Y., Zhang, W., Ogata, S., Clements, D., Strauss, J. H., Baker, T. S., Kuhn, R. J., and Rossmann, M. G. (2004) Conformational changes of the flavivirus E glycoprotein, *Structure 12*, 1607−1618.
6. Modis, Y., Ogata, S., Clements, D., and Harrison, S. C. (2004) Structure of the dengue virus envelope protein after membrane fusion, *Nature (London, U.K.) 427*, 313−319.
7. Modis, Y., Ogata, S., Clements, D., and Harrison, S. C. (2003) A ligand-binding pocket in the dengue virus envelope glycoprotein, *Proc. Natl. Acad. Sci. U.S.A. 100*, 6986−6991.
8. Zhang, W., Chipman, P. R., Corver, J., Johnson, P. R., Zhang, Y., Mukhopadhyay, S., Baker, T. S., Strauss, J. H., Rossmann, M. G., and Kuhn, R. J. (2003) Visualization of membrane protein domains by cryo-electron microscopy of dengue virus, *Nat. Struct. Biol. 10*, 907−912.
9. Modis, Y., Ogata, S., Clements, D., and Harrison, S. C. (2005) Variable surface epitopes in the crystal structure of dengue virus type 3 envelope glycoprotein, *J. Virol. 79*, 1223−1231.
10. Rey, F. A., Xeinz, F. X., Mandl, C., Kunz, C., and Harrison, S. C. (1995) The envelope glycoprotein from tick-borne encephalitis virus at 2 Å resolution, *Nature (London, U.K.) 375*, 291−298.
11. Diana, G. D., McKinlay, M. A., Otto, M. J., Akullian, V., and Oglesby, C. (1985) (4,5-Dihydro-2-oxazolyl)phenoxy alkyl isoxazoles. Inhibitors of picornavirus uncoating, *J. Med. Chem. 28*, 1906−1910.
12. Goncalves, R. B., Mendes, Y. S., Soares, M. R., Katpally, U., Smith, T. J., Silva, J. L., and Oliveira, A. C. (2007) VP4 protein from human rhinovirus 14 is released by pressure and locked in the capsid by the antiviral compound WIN, *J. Mol. Biol. 366*, 295−306.
13. Hadfield, A. T., Oliveira, M. A., Kim, K. H., Minor, I., Kremer, M., Heinz, B. A., Shepard, D., Pevear, D. C., Rueckert, R. R., and Rossmann, M. G. (1995) Structural studies on human rhinovirus 14 drug-resistant compensation mutants, *J. Mol. Biol. 253*, 61−73.
14. Zhang, Y., Simpson, A. A., Ledford, R. M., Bator, C. M., Chakravarty, S., Skochko, G. A., Demenczuk, T. M., Watanyar, A., Pevear, D. C., and Rossmann, M. G. (2004) Structural and virological studies of the stages of virus replication that are affected by antirhinovirus compounds, *J. Virol. 78*, 11061−11069.
15. Szymkowski, D. E. (2005) Creating the next generation of protein therapeutics through rational drug design, *Curr. Opin. Drug Discovery Dev. 8*, 590−600.
16. Vajda, S., and Guarnieri, F. (2006) Characterization of protein-ligand interaction sites using experimental and computational methods, *Curr. Opin. Drug Discovery Dev. 9*, 354−362.
17. Zhou, Z., Madrid, M., Evanseck, J. D., and Madura, J. D. (2005) Effect of a bound non-nucleoside RT inhibitor on the dynamics of wild-type and mutant HIV-1 reverse transcriptase, *J. Am. Chem. Soc. 127*, 17253−17260.
18. Zhou, Z., and Madura, J. D. (2004) Relative free energy of binding and binding mode calculations of HIV-1 RT inhibitors based on dock-MM-PB/GS, *Proteins: Struct., Funct., Bioinf. 57*, 493−503.
19. Li, Y., Zhou, Z., and Post, C. B. (2005) Dissociation of an antiviral compound from the internal pocket of human rhinovirus 14 capsid, *Proc. Natl. Acad. Sci. U.S.A. 102*, 7529−7534.
20. Phelps, D. K., and Post, C. B. (1999) Molecular dynamics investigation of the effect of an antiviral compound on human rhinovirus, *Protein Sci. 8*, 2281−2289.
21. Speelman, B., Brooks, B. R., and Post, C. B. (2001) Molecular dynamics simulations of human rhinovirus and an antiviral compound, *Biophys. J. 80*, 121−129.

22. Grant, R. A., Hiremath, C. N., Filman, D. J., Syed, R., Andries, K., and Hogle, J. M. (1994) Structures of poliovirus complexes with antiviral drugs: implications for viral stability and drug design, *Curr. Biol. 4*, 784–797.
23. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., and Shenkin, P. S. (2004) Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy, *J. Med. Chem. 47*, 1739–1749.
24. Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., and Banks, J. L. (2004) Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening, *J. Med. Chem. 47*, 1750–1759.
25. Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C., and Mainz, D. T. (2006) Extra precision Glide: Docking and scoring incorporating a model of hydrophobic enclosure for protein–ligand complexes, *J. Med. Chem. 49*, 6177–6196.
26. Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J. (1996) Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, *J. Am. Chem. Soc. 118*, 11225–11236.
27. Mayer, M., and Meyer, B. (2001) Group epitope mapping by saturation transfer difference NMR to identify segments of a ligand in direct contact with a protein receptor, *J. Am. Chem. Soc. 123*, 6108–6117.

Supporting Information

Antiviral compounds discovered by virtual screening of small-molecule libraries against dengue virus E protein

Zhigang Zhou[1,%], Mansoora Khaliq[2], Jae-Eun Suk[1], Chinmay Patkar[2,#], Long Li[2], Richard J. Kuhn[2], Carol Beth Post[1,2,]*

[1]Department of Medicinal Chemistry and Molecular Pharmacology, [2]Department of Biological Sciences, Markey Center for Structural Biology and Purdue Cancer Center,

Purdue University, West Lafayette, Indiana 47907.

TITLE RUNNING HEAD: Antiviral screening of Dengue E protein.

To whom correspondence should be addressed,

e-mail: cbp@purdue.edu,

Fax: 765-496-1189,

Phone: 765-494-5980.

Supplementary Methods

Small-Molecule Library and Protein Preparation: The plated, P, diversity, D and mechanistic, M NCI libraries (http://dtp.nci.nih.gov/docs/3d_database/Structural_information/structural_data.html) were screened for binding to the BOG pocket of E protein. The P library is a collection of approximately 140,000 compounds from the NCI repository of synthetic and natural products that have been evaluated as potential anticancer and anti-HIV agents. The D library comprises 1990 compounds selected from the P library to have structural diversity. The M library comprises 879 compounds chosen from 37,836 compounds that have been tested in the NCI human tumor 60-cell-line screen. The M library is based on the GI50 activity in the 60-cell-line screen and was designed to represent a broad range of growth inhibition patterns.

Heavy-atom coordinates of the molecules in the three libraries were taken from the NCI database. Hydrogen atoms were added to each molecule either according to geometry, in the case of a fixed position, or in the lowest energy conformation determined by a single side chain torsion-angle search. Parameters, including partial charges, were assigned and the energy of the isolated compounds was minimized using the OPLS-AA force field (*1*) implemented by the MOE program package.

Protein coordinates were from the crystallographic solution of dengue E protein and N-octyl-β-D-glucoside (BOG) complex (PDB entry: 1OKE).(*2*) The protein coordinates were prepared following the Glide protocol of adding hydrogen atoms and assigning partial charges using the OPLS-AA force field. A coarse energy minimization with a small number of steps was performed on the complex to relax side chains and the newly added hydrogen atoms. The BOG ligand was used to determine the location of a docking grid box and was removed prior to grid generation in next step.

Docking calculations: Docking by SP and XP protocols was restricted to the binding site region of the BOG ligand in the crystal structure of the E protein complex using a grid with dimensions (8 x 8 x 14) Å and centered at the geometric center of BOG. Ligands were positioned by placing the ligand molecular center at points of a cubic grid. The ligand atoms were required to lie within the volume of a (23 x 23 x 29) Å box, and the binding space was explicitly sampled by rotation of the ligand molecule. Poses with ligand atoms falling outside the larger box were eliminated.

Compounds with less than 150 atoms and less than 35 rotatable bonds were considered for this study. Conformational flexibility of the ligand was included by generation of rotamers from explicit rotation of single-bond torsion angles. Multiple conformations were generated for each molecule in the library by exhaustive enumeration of the energy minima for the rotatable torsion angles, and prescreened for docking by eliminating high-energy conformers(*3-5*). All rotamers for a given ligand were docked and oriented in the site, and grid-refinement executed for 5000 poses. The top 400 refined poses per ligand were subjected to energy minimization using 100 steps of conjugate gradient minimization. Poses of a given compound were not saved unless the poses differed from previous docked poses by more than 0.2 Å rms deviation in heavy atom coordinates, or involve atomic displacements greater than 0.5 Å.

Scoring docked poses: Two empirical scoring functions, GlideScore 2.5 (GScore) and Emodel(*4, 6*), were used to rank and select compounds in stages 1 and 2. The standard-precision (SP) form of GScore has 'softer' potentials to give lower penalties for minor steric clashes that occur as a result of limited sampling. The extra-precision (XP) form of GScore is more physically accurate. The GScore extends the fast scoring function ChemScore(*7*), but with different parameterization of the interaction terms and further division of the types of interactions.

GScore = a*vdW + b**Coul* + *Lipo* +*Hbond* + *Metal* + *BuryP* + *RotB* + *Site*

where the included terms are from van der Waals (vdW) and Coulombic (Coul) energies, a lipophilic contact term (*Lipo*), explicit hydrogen bonding (*Hbond*), a metal-binding term (*Metal*), a penalty for buried charge (*BuryP*), a penalty for freezing rotatable bonds (*RotB*), and a term for desolvation of polar surface in the binding site (Site). The coefficients of vdW and *Coul* were: a = 0.065, b = 0.025 for SP, and a = 0.060, b = 0.020 for XP. A distance-dependent dielectric constant was used.

Emodel is a specific combination of GScore with an internal strain energy incorporated to allow for flexible ligand docking and a ligand-receptor molecular mechanics energy based on non-bonded

contact(*4, 6*). Applications have shown that Emodel score correlates with experimental activity better than GScore in some cases(*8*).

Computational time: The Glide SP docking calculation of each molecule required 0.5-3 minutes of CPU time depending on the number of rotatable bonds. The XP mode for docking utilized 3-10 minutes of CPU time for each molecule. Thermal relaxation required an average of 15 minutes of CPU time for each complex. Calculations were carried out on a 16-node cluster of 1.4 GHz dual-CPU AMD processors running Linux, and workstations configured with Intel 3.0 GHz Xeon dual-processors running Linux.

BHK cells: Clone 15 from BHK-21 cells (baby hamster kidney cells) obtained from the American Type Culture Collection were maintained in minimal essential medium (MEM) (Invitrogen) containing 10% fetal bovine serum (FBS). Cells were grown in incubators at 37ºC in the presence of 5% CO_2.

Generation of YFV-IRES-Luc virus: A fire-fly luciferase reporter gene was inserted into pYF23, a derivative of pACNR which is the full-length cDNA clone of yellow fever virus YFV 17D, to construct YFV-IRES-Luc, a luciferase-reporting full-length virus. To facilitate this construction, an *Nsi*I restriction site was introduced at the beginning of the 3' NTR immediately following the UGA termination codon of NS5 in pYF23 using standard overlapping PCR mutagenesis. To construct YFV-IRES-Luc, an IRES-FF.Luc (EMCV IRES-fire fly luciferase) cassette was amplified by PCR from YFRP-IRES-Luc, an YFV replicon(*9*), and inserted into the *Nsi*I restriction site.

In vitro transcribed YFV-IRES-Luc RNA was transfected into BHK cells using Lipofectamine (Invitrogen). At 4 days post-transfection, the resulting YFV-IRES-Luc virus was harvested and the titer of the virus determined by a standard plaque assay. The infectivity of the virus could be assayed directly as a measure of the luciferase amounts produced in infected cells over a period of time.

Generation of YFV pseudo-infectious particles (PIPs): Generation of YFV pseudo-infectious particles (PIPs) and construction of YF-R.luc2A-RP, a YFV replicon with the *Renilla* luciferase (R.luc) gene, are described in Jones, et al(*9*). Briefly, YF-R.luc2A-RP transcripts were first electroporated into BHK cells, followed by a second electroporation with the Sin-CprME replicon. This replicon is derived from Sindbis virus, and used to express the YFV structural proteins. There was an interval of 24 hr between the two electroporations to allow for sufficient amplification of the initially transfected YF-R.luc2A-RP RNA. PIPs were harvested from the culture media following the second electroporation.

Virus release assay: BHK cells ($5x10^5$) were infected with wild type YFV at an MOI of 1.0 and overlaid with media containing the compound P02 dissolved in DMSO to give a final concentration of 10 or 100μM P02 and 1% DMSO. Cells grown in media with 1% DMSO served as the reference control. Cells were incubated at 37°C and aliquots of culture supernatants were harvested at 12, 24, 36 and 48 hours post infection, and the virus titer was determined by plaque assay. For plaque assays, 10–fold serial dilutions of harvested culture supernatants were made using phosphate–buffered saline (PBS) containing 1% FBS. BHK cells at ~90% confluency were inoculated with the culture dilutions for 1 hour at room temperature and then overlaid with MEM containing 5% FBS and 1% agarose. Cells were incubated at 37°C for 3 days, stained with neutral red and the plaque number determined.

Expression and purification of dengue 2 E protein: The gene of the soluble E protein (1-394 aa) (dengue virus serotype 2, strain 16681) was cloned into the vector pMT-BiP-V5/His A (Invitrogen). *Drosophila* S2 cells were transfected with 20μg plasmid and 1μg pCoHygro vector by the standard calcium phosphate precipitation method. To generate the stable cell line, the transfected S2 cells were selected in the presence of 300μg/ml Hygromycin-B for a month. The stably transfected cells were scaled up and cultured in the serum free medium (Invitrogen) at 27ºC. The expression was induced by addition of 500μM $CuSO_4$ when the cells reached a density of $4\text{-}6\times10^6$ cell/ml. The medium was harvested at days 4 to 6 after induction. The conditioned medium was then concentrated and loaded on an immunoaffinity column (NHS-activated Sepharose, GE Healthcare) cross-linked with the E protein specific 4G2 antibody (*10*). The E protein was eluted by 100mM glycine-HCl buffer (pH 2.8). The protein was further purified on a Mono Q column and a superdex 200 column (GE Healthcare). The purified protein was concentrated to 2-3mg/ml in 10mM HEPES buffer (pH7.0).

Supplementary Results and Discussion

The large-scale structural changes between immature and mature virus, particularly the differences in domain orientations of E protein, suggest ligand binding to this BOG pocket could inhibit a step in the virus lifecycle at which the virus E protein undergoes a structural transition(*2*). A similar strategy of targeting a buried pocket in a viral capsid protein was exploited with human rhinovirus (*11, 12*), leading to the development of numerous antiviral compounds(*13-20*). Receptor-based computational screening has been applied to generate leads and to facilitate drug development in previous studies(*21-30*). A three-stage computational screening was used to rank >142,000 compounds for binding the BOG pocket of dengue E protein, and a manual 4th stage selected 23 of the computationally top-ranked 45 compounds for experimental validation using biological assays of antiviral activity. One compound was tested directly for association with purified E protein using saturation transfer difference NMR spectroscopy (STD).

BOG docking test case: The reliability of the XP docking method applied to the targeted site of dengue E protein was demonstrated by docking BOG and comparing the results with the known structure of the crystallographic complex. The results for 20 docked poses (defined by ligand conformation and docked position) in the best cluster scored using Gscore and Emodel functions are listed in **Table S1**. Gscores ranged from -3.82 to -3.04 and Emodel from -46.10 to -40.46. Emodel ranked some poses with smaller RMSD values higher than the Gscore ranking. The internal structure of BOG differs by less than 0.3 Å for the RMSD calculated after least-squares superposition of BOG alone. The root-mean-square deviations (RMSDs) of BOG heavy atoms between the docked poses and the crystallographic coordinates after superposition of E protein range from 0.5 Å to 1.2 Å. Structural differences between the Glide docked structures and the crystallographic structure are visualized in **Figure S1**. Shown are an overlay of the top 10 poses, as ranked in Table S2, and pose #1 and #10 superimposed individually on the crystal structure. The results are evidence that Glide accurately reproduces the known structure of the E protein-BOG complex.

The rankings of BOG poses based on Gscore and Emodel functions differ. Gscore is more commonly used for Glide docking, however, Emodel more accurately ranked certain poses of BOG than Gscore, and the literature(*8*) reports advantages with Emodel. Accordingly, we elected to use both scoring functions to attempt to enhance the potential for identifying 'hits.'

Screening NCI libraries against the BOG pocket of dengue E protein.

The Diversity (D) library comprises 1990 compounds; The Mechanistic (M) library comprises 879 compounds, and the Plated (P) library is a collection of approximate 140,000 compounds from the NCI repository of synthetic and natural products that have been evaluated as potential anticancer and anti-HIV agents. The hierarchical, four-stage protocol reduces the number of compounds by approximately 100-fold (Stage 1), 10-fold (Stage 2), 3-fold (Stage 3), and 2-fold (visual analysis).

Stage II, XP docking. The ranked scores of the T^{3x500} compounds subjected to the XP docking protocol are depicted in **Figure S2**. The relative scores for Gscore are P<D<M, and Emodel are P<D,M. As was the case for SP docking, the greatest discrimination, based on slope, is observed for a small number of highest-ranked compounds, while scores of most compounds are slowly varying. The highest ranked 25 compounds based on Gscore (**Figure S2A**) and Emodel (**Figure S2B**) for each library composed the T^{3x50} set carried forward to the final stage of screening.

The red triangles in **Figure S2** are the compounds listed in main text **Table 1**, and their rankings at stage 2 of the computational HTS where the T^{3x500} set was reduced to the T^{3x50} set by retaining the top scoring 25 compounds from Gscore and Emodel. The active compounds are ranked among the first 60 for Gscore but as low as 200 for Emodel, indicating the better predictive power of Gscore.

Screening the BOG pocket, Stage 3, thermal relaxation. The final stage III utilized the OPLS molecular mechanics energy function and internal flexibility of both protein and ligand atoms by energy minimization. The calculated energies for the T^{3x50} set are listed in **Table S2**. Compounds were ranked

by the interaction energy, equal to the sum of vdW, electrostatic energy, and the difference in SGB solvation energy between free molecules and the complexes.

The benefit of a final structure relaxation step by energy minimization using a physically based, molecular mechanics force field has been demonstrated(*21, 31*). Here, screening the best 20% of the low-energy scores lead to 8 active compounds (8 red triangles in main text **Figure 4** fall in the top-ranked set of 10 for each library carried forward) and 1 active compound was obtained from the Gscore-based ranking. That 8/9 active compounds from 23 tested were obtained from the lowest energy values is consistent with the notion that structure relaxation enhances the enrichment factor, although additional biological testing of the lower ranked compounds in Figure 4 is necessary to confirm this conclusion.

Gscore appeared to have the best predictive power when ranking compounds. The third stage utilized a physically based molecular mechanics force field for energy minimization to optimize both protein and ligand coordinates of 150 complexes. The results indicated that energy relaxation, which is a straightforward albeit limited approach to account for protein flexibility, can influence the ranking of the top compounds and avoid some shortcomings of rigid-receptor docking protocols.

Chemical properties of the T^{3x50} and T^{3x15} molecules from stage 2 and stage 3. The presence of electrostatic interactions in terms of hydrogen bonding, cationic and anionic groups, and the amount of hydrophobic and aromatic contact area were evaluated to identify the chemical features selected by docking. (**Table S3**). Most of the molecules include hydrogen bond acceptor groups (95% and 98% for T^{3x50} and T^{3x15} subsets, respectively), plus hydrophobic (91% and 94%) and aromatic area (97% and 96%). Only 79% molecules in T^{3x50} have hydrogen bond donors and this percentage increases to 91% in T^{3x15}.

Pharmacophore elements of functional ligands: Pharmacophore elements of ligands are identified using the Polarity-Charge-Hydrophobicity (PCH) scheme in MOE (Chemical Computing Group). Six pharmacophore elements are identified by the scheme: hydrogen bond donor (H_{don}), hydrogen bond acceptor (H_{acc}), cation, anion, hydrophobic area, and aromatic area. The molecular fragments of a molecule are identified using a set of drug-like index rules, which is partially based on Xu's rules(*32*). All fragments are grouped based on one or multiple attachment points.

All active compounds have an appropriate logP(o/w) value for biological efficacy, and at least one hydrophobic center and one hydrophilic center separated by a distance of 4 to 8 Å. This feature appears to be important for ligand/receptor recognition of E protein system to complement the chemical composition of the BOG binding pocket, which is largely hydrophobic internally, and hydrophilic from the polar residues that occur at the top and bottom of the channel near the solvent exposed regions. A ligand that matches this feature should have stronger interactions with the receptor.

Assessment of computational screening: The ranking of the nine compounds with antiviral activity (main text **Table 1**) during the course of the screening gives some insight on the HTS procedure. These rankings are marked by red triangles in main text **Figures 2-3** and **Figure S2**. Compounds surviving until the final T^{3x15} set considered for experimental verification are plotted with ×'s for reference.

The T^{23} set contained compounds from all three NCI libraries, even though the most favorable scores for the D and the M library compounds are substantially inferior to scores for the top-ranked 500 P-library compounds of stage 1. The nine compounds with antiviral activity also included candidates from the three libraries: five compounds from D, and two each from M and P. The active compounds identified from D and M libraries, when ranked within the large P library in stage 1, were scored 1359 at best and most were below 3000, so that identification of these compounds by screening only the P library would have required more than twice the number of compounds carried forward to stage 2. Therefore the objective of finding a set enriched with active compounds was met, although the method may have failed to extract all potentially active compounds from the large number of molecules in the structurally redundant P library. The diversity and mechanistic basis utilized to design these two libraries appears effective in the case of screening against the flavivirus E protein.

Both Gscore and Emodel were used for scoring functions. The arrows in Figure 2 (main text) delineate cutoff values for selection based on either the Gscore or Emodel function, so that symbols

falling to the right of the arrow indicate active compounds that were selected by computational screening based on the alternative scoring function. Although both successfully identified compounds from the smaller D and M libraries, only Gscore correctly predicted active compounds out of stage 1 screening of the larger P library as shown by the red triangles left of the arrow in main text **Figure 2E**. Emodel ranking of the P library yielded none of the nine active compounds, as shown by the lack of red triangles to the left of the arrow in main text **Figure 2F**. A similar result was observed at stage 2 (**Figure S2**). Because most compounds with antiviral activity were identified by the Gscore while Emodel rankings did not provide additional active compounds, we conclude that the predictive power of Gscore is sufficient. These results suggest that a more efficient screening protocol than the one used here would be to include twice the number of compounds ranked solely by Gscore.

Binding Assessed by Saturation Transfer Difference NMR: STD measurements were conducted in the presence of yellow fever virus for the compound P02. The results are shown in **Figure S3**. The bottom panel is the reference spectrum with no irradiation for a sample of P02 (0.45 mM) in the presence of yellow fever virus (0.8 μM). The top panel is the STD spectrum for the same sample. The presence of P02 peaks in the STD spectrum indicate P02 binds virus, and suggest P02 associates with E protein given that E protein covers the entire external surface of the virus and the lipid layer is largely inaccessible.

Inhibition of Virus Release: The antiviral activity of the compound P02 was characterized by determining its effect on virus production measured from a virus titer to further investigate the inhibitory activity observed in YFV–IRES–Luc virus growth assays. The levels of virus release in the presence of 10 μM and 100 μM P02 is compared to the DMSO control in **Figure S4**. No plaques were detected at 12 hours post–infection. The values in terms of percent reduction in the presence of 10 μM P02 are 29% (24hrs), 58% (35hrs) and 63% (48hrs), and in the presence of 100 μM P02 the values are 1.7% (24hrs), 0.4% (35hrs) and 1.0% (48hrs). As expected, the virus titer is reduced approximately 10 orders of magnitude at 100 μM P02.

References

1. Jorgensen, W. L., Maxwell, D. S., and Tirado-Rives, J. (1996) Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids, *J. Am. Chem. Soc. 118*, 11225-11236.
2. Modis, Y., Ogata, S., Clements, D., and Harrison, S. C. (2003) A ligand-binding pocket in the dengue virus envelope glycoprotein, *Proceedings of the National Academy of Sciences of the United States of America 100*, 6986-6991.
3. Halgren, T. A., Murphy, R. B., Friesner, R. A., Beard, H. S., Frye, L. L., Pollard, W. T., and Banks, J. L. (2004) Glide: A new approach for rapid, accurate docking and scoring. 2. Enrichment factors in database screening, *Journal of Medicinal Chemistry 47*, 1750-1759.
4. Friesner, R. A., Banks, J. L., Murphy, R. B., Halgren, T. A., Klicic, J. J., Mainz, D. T., Repasky, M. P., Knoll, E. H., Shelley, M., Perry, J. K., Shaw, D. E., Francis, P., and Shenkin, P. S. (2004) Glide: A new approach for rapid, accurate docking and scoring. 1. method and assessment of docking accuracy, *Journal of Medicinal Chemistry 47*, 1739-1749.
5. Schrödinger. (2005) FirstDiscovery 3.5 - Technical Notes, pp 5-7, Schrödinger Press, New York.
6. Friesner, R. A., Murphy, R. B., Repasky, M. P., Frye, L. L., Greenwood, J. R., Halgren, T. A., Sanschagrin, P. C., and Mainz, D. T. (2006) Extra Precision Glide: Docking and Scoring Incorporating a Model of Hydrophobic Enclosure for Protein-Ligand Complexes, *Journal of Medicinal Chemistry 49*, 6177-6196.
7. Eldridge, M. D., Murray, C. W., Auton., T. R., Paolini, G. V., Mee, R. P., and , . (1997) Empirical scoring functions: I. The development of a fast empirical scoring function to esti-mate the binding affinity of ligands in receptor complexes, *J. Comput.-Aided Mol Des*, 425–445.

8. Bytheway, I., and Cochran, S. (2004) Validation of Molecular Docking Calculations Involving FGF-1 and FGF-2, *Journal of Medicinal Chemistry 47*, 1683-1693.
9. Jones, C. T., Patkar, C. G., and Kuhn, R. J. (2005) Construction and applications of yellow fever virus replicons, *Virology 331*, 247-259.
10. Henchal, E. A., Gentry, M. K., McCown, J. M., and Brandt, W. E. (1982) Dengue virus-specific and flavivirus group determinants identified with monoclonal antibodies by indirect immunofluorescence, *Am J Trop Med Hyg 31*, 830-836.
11. Diana, G. D., McKinlay, M. A., Otto, M. J., Akullian, V., and Oglesby, C. (1985) (4,5-Dihydro-2-Oxazolyl)Phenoxy Alkyl Isoxazoles - Inhibitors of Picornavirus Uncoating, *Journal of Medicinal Chemistry 28*, 1906-1910.
12. Diana, G. D., McKinlay, M. A., and Treasurywala, A. (1997) The use of structural information in the design of picornavirus capsid binding agents, *Struct. Biol. Viruses*, 432-450.
13. Goncalves, R. B., Mendes, Y. S., Soares, M. R., Katpally, U., Smith, T. J., Silva, J. L., and Oliveira, A. C. (2007) VP4 Protein from Human Rhinovirus 14 Is Released by Pressure and Locked in the Capsid by the Antiviral Compound WIN, *Journal of Molecular Biology 366*, 295-306.
14. Steindl, T. M., Crump, C. E., Hayden, F. G., and Langer, T. (2005) Pharmacophore Modeling, Docking, and Principal Component Analysis Based Clustering: Combined Computer-Assisted Approaches To Identify New Inhibitors of the Human Rhinovirus Coat Protein, *Journal of Medicinal Chemistry 48*, 6250-6260.
15. Lewis, J. K., Bothner, B., Smith, T. J., and Siuzdak, G. (1998) Antiviral agent blocks breathing of the common cold virus, *Proceedings of the National Academy of Sciences of the United States of America 95*, 6774-6778.
16. Wang, W., Lee, W.-M., Mosser, A. G., and Rueckert, R. R. (1998) WIN 52035-dependent human rhinovirus 16: assembly deficiency caused by mutations near the canyon surface, *Journal of Virology 72*, 1210-1218.
17. Hadfield, A. T., Oliveira, M. A., Kim, K. H., Minor, I., Kremer, M., Heinz, B. A., Shepard, D., Pevear, D. C., Rueckert, R. R., and Rossmann, M. G. (1995) Structural studies on human rhinovirus 14 drug-resistant compensation mutants, *Journal of Molecular Biology 253*, 61-73.
18. Li, Y., Zhou, Z., and Post, C. B. (2005) Dissociation of an antiviral compound from the internal pocket of human rhinovirus 14 capsid, *Proceedings of the National Academy of Sciences of the United States of America 102*, 7529-7534.
19. Zhang, Y., Simpson, A. A., Ledford, R. M., Bator, C. M., Chakravarty, S., Skochko, G. A., Demenczuk, T. M., Watanyar, A., Pevear, D. C., and Rossmann, M. G. (2004) Structural and virological studies of the stages of virus replication that are affected by antirhinovirus compounds, *Journal of Virology 78*, 11061-11069.
20. Muckelbauer, J. K., Kremer, M., Minor, I., Diana, G., Dutko, F. J., Groarke, J., Pevear, D. C., and Rossmann, M. G. (1995) he structure of coxsackievirus B3 at 3.5 A resolution, *Structure 3*, 653-667.
21. Song, H., Wang, R., Wang, S., and Lin, J. (2005) A low-molecular-weight compound discovered through virtual database screening inhibits stat3 function in breast cancer cells, *Proceedings of the National Academy of Sciences of the United States of America 102*, 4700-4705.
22. Conway, K. A., Rachet, J.-C., Bieganski, R. M., and Lansbury, P. T., Jr. (2001) Kinetic stabilization of the a-synuclein protofibril by a dopamine-a-synuclein adduct, *Science 294*, 1346-1349.
23. Ray, S. S., Nowak, R. J., Brown, R. H., Jr., and Lansbury, P. T., Jr. (2005) Small-molecule-mediated stabilization of familial amyotrophic lateral sclerosis-linked superoxide dismutase mutants against unfolding and aggregation, *Proceedings of the National Academy of Sciences of the United States of America 102*, 3639-3644.
24. Koyama, M., Tsuboi, H., Endou, A., Takaba, H., Kubo, M., Del Carpio, C. A., and Miyamoto, A. (2007) Combinatorial computational chemistry approach for materials design: applications in

deNOx catalysis, Fischer-Tropsch synthesis, lanthanoid complex, and lithium ion secondary battery, *Combinatorial Chemistry & High Throughput Screening 10*, 99-110.
25. Mirkovic, N., Li, Z., Parnassa, A., and Murray, D. (2007) Strategies for high-throughput comparative modeling: applications to leverage analysis in structural genomics and protein family organization, *Proteins: Structure, Function, and Bioinformatics 66*, 766-777.
26. Vajda, S., and Guarnieri, F. (2006) Characterization of protein-ligand interaction sites using experimental and computational methods, *Current Opinion in Drug Discovery & Development 9*, 354-362.
27. Szymkowski, D. E. (2005) Creating the next generation of protein therapeutics through rational drug design, *Current Opinion in Drug Discovery & Development 8*, 590-600.
28. Joseph-McCarthy, D. (2005) Structure-based lead optimization, *Annual Reports in Computational Chemistry 1*, 169-183.
29. Mizutani, M. Y., and Itai, A. (2004) Efficient Method for High-Throughput Virtual Screening Based on Flexible Docking: Discovery of Novel Acetylcholinesterase Inhibitors, *Journal of Medicinal Chemistry 47*, 4818-4828.
30. Alvarez, J. C. (2004) High-throughput docking as a source of novel drug leads, *Current Opinion in Chemical Biology 8*, 365-370.
31. Zhou, Z., Bates, M., and Madura, J. D. (2006) Structure modeling, ligand binding, and binding affinity calculation (LR-MM-PBSA) of human heparanase for inhibition and drug design, *Proteins: Structure, Function, and Bioinformatics 65*, 580-592.
32. Xu, J., and Stevenson, J. (2000) Drug-like Index: A New Approach To Measure Drug-like Compounds and Their Diversity, *Journal of Chemical Information and Computer Sciences 40*, 1177-1187.

Table S1. The energies and scoring properties of top20 poses from docking simulation of **BOG** to E protein of Dengue virus. The poses differ primarily in van der Waals, electrostatic and ligand internal energies, while other factors (e.g. Ulipo, Usite), which would distinguish between ligands of different chemical structure, are relatively constant.

Pose #	GScore	Emodel	U_{vdW}	U_{coul}	U_{Lipo}	U_{hbond}	U_{Bury}	U_{Site}	U_{Intl}[1]	RMSD[2]
1	-3.82	-40.46	-22.81	-8.54	-3.21	-1.5	0	-0.02	8.85	0.65
2	-3.76	-40.78	-20.38	-10.69	-3.23	-1.5	0	-0.03	4.65	0.52
3	-3.59	-46.10	-28.86	-7.38	-3.81	-1.5	1.5	-0.34	4.03	1.13
4	-3.58	-45.62	-28.80	-7.19	-3.82	-1.5	1.5	-0.33	3.95	1.17
5	-3.57	-46.05	-28.52	-7.40	-3.82	-1.5	1.5	-0.34	3.72	1.21
6	-3.57	-45.59	-28.50	-7.20	-3.82	-1.5	1.5	-0.34	3.78	1.03
7	-3.43	-43.96	-26.24	-9.45	-3.82	-1.5	1.5	-0.29	9.98	0.98
8	-3.37	-42.04	-25.99	-9.45	-3.79	-1.5	1.5	-0.28	12.09	0.87
9	-3.33	-40.63	-24.57	-8.51	-3.84	-1.5	1.5	-0.29	11.42	0.72
10	-3.33	-43.09	-24.65	-9.84	-3.81	-1.5	1.5	-0.29	8.06	0.77
11	-3.31	-40.98	-24.39	-8.81	-3.82	-1.5	1.5	-0.29	11.23	0.93
12	-3.30	-41.09	-24.13	-8.97	-3.82	-1.5	1.5	-0.29	11.38	0.69
13	-3.27	-40.84	-23.84	-9.49	-3.80	-1.5	1.5	-0.29	11.53	0.60
14	-3.21	-43.02	-25.35	-9.59	-3.64	-1.5	1.5	-0.29	10.44	1.06
15	-3.13	-41.96	-22.81	-10.62	-3.70	-1.5	1.5	-0.29	10.42	0.56
16	-3.09	-45.68	-29.51	-6.11	-3.80	-1.5	2	-0.33	3.76	1.26
17	-3.08	-44.87	-28.40	-6.54	-3.84	-1.5	2	-0.35	3.62	0.84
18	-3.07	-45.18	-29.42	-6.14	-3.79	-1.5	2	-0.33	4.69	1.09
19	-3.07	-45.35	-29.15	-6.11	-3.81	-1.5	2	-0.34	3.71	0.88
20	-3.04	-43.79	-28.44	-6.16	-3.81	-1.5	2	-0.34	5.10	0.70

1 E_{intl}: Internal energy of ligand.

2 RMSD: Root Mean Square Deviation of BOG heavy atoms between docked pose and crystal complex.

Table S2. The energies and scoring properties of T^{3x50} set of complexes. The variances of the binding energy for the three sets are 41.03, 38.22, and 51, respectively. The ratios of the variance/largest value (absolute value) for three sets are 0.75, 0.56, and 0.61, respectively.

	U_{vdw}		U_{ele}		U_{Rxnf}		U_{Cav}		Gscore		$E_{binding}$	
	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min	Max	Min
D library	-60.88	-31.53	-53.65	9.81	2.22	30.78	-3.20	-1.79	-11.39	-3.92	-54.25	-13.22
M library	-67.76	-36.48	-63.76	-1.43	3.20	23.93	-3.46	-1.96	-12.79	-5.77	-68.64	-29.82
P library	-97.15	-42.75	-84.44	1.21	6.42	50.96	-4.21	-2.32	-19.06	-9.48	-83.65	-32.37

Table S3. The search results for functional groups of two subsets of selected compounds. The hydrogen bonding has apparent change when the selected compounds were narrowed from T^{3x50} set to T^{3x15}.

Compound set	H_{don}*	H_{acc}	Cation	Anion	Hydrophobic	Aromatic
T^{3x50}	119/79%	143/95%	6/4%	3/2%	141/94%	146/97%
T^{3x15}	41/91%	44/98%	1/2%	1/2%	41/91%	43/96%

* H_{don} – hydrogen bond donor functional group; H_{acc} – hydrogen bond acceptor functional group; Cation – cation group; Anion – anion group; hydrophobic – hydrophobic area; Aromatic - aromatic group.

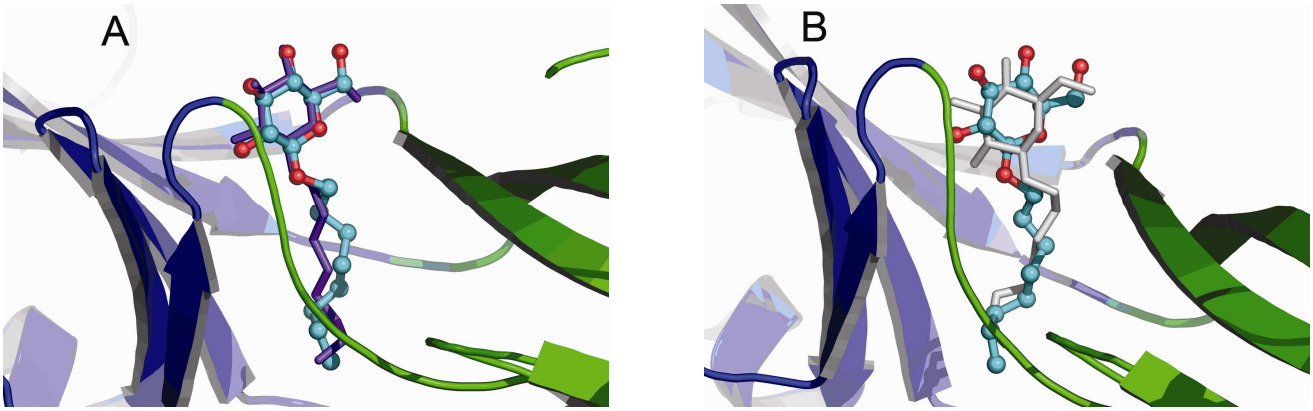


Figure S1: The superposition of Glide docked poses (stick) with the crystal structure of BOG (stick-ball, cyan and red) bound to E protein (ribbon) in the pocket between domain 1 (green) and domain 2 (blue).

A: Pose #2 (purple) superposed with crystal structure.

B: Pose #8 (grey)) superposed with crystal structure (stick-ball mode, carbon colored in cyan) of BOG.

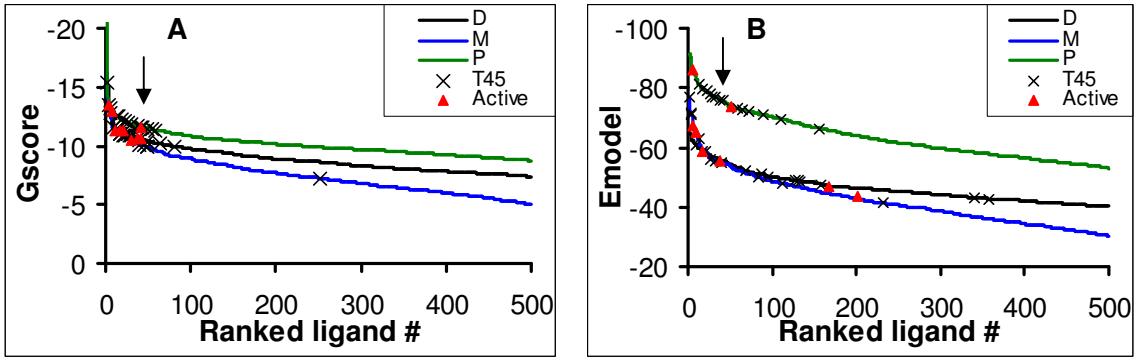


Figure S2. Stage 2: GScore and Emodel plots of T^{3x500} set of in the GLIDE XP docking screening ($T^{3x500} \rightarrow T^{3x50}$). The symbol, ×, indicates compounds selected for T^{3x15} from three libraries and red triangles indicate the compounds active in the YFV-IRES-Luc virus growth assay. The arrow indicates the cutoff points for selection of T^{3x500} set.

A. Gscore profiles of **D, M, and P** libraries; **B.** Emodel score profiles of **D, M, and P** libraries.

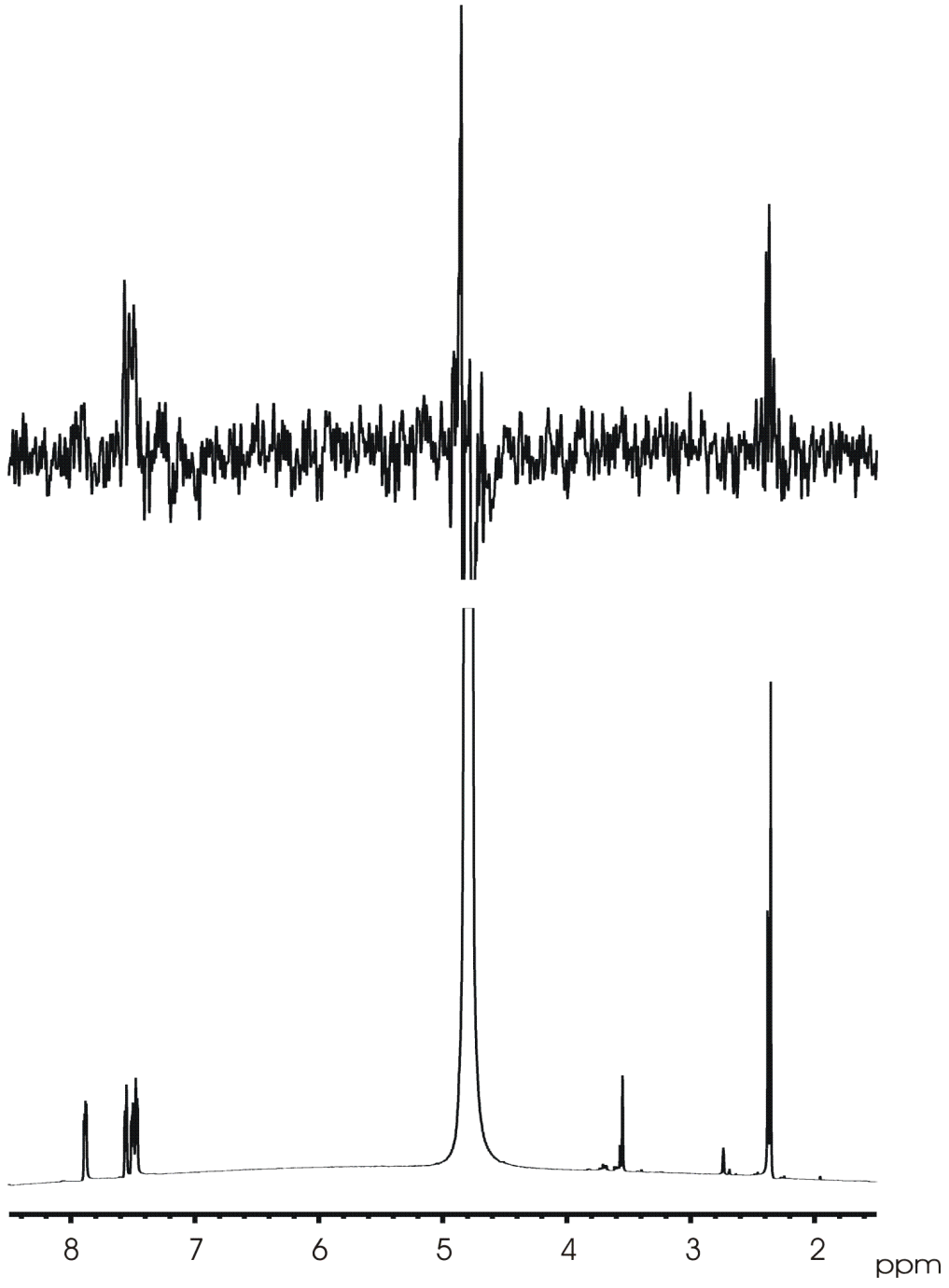


Figure S3: NMR spectra (600 MHz) to assess P02 (0.45 mM) binding to yellow fever virus (0.8 µM). (bottom) 1D 1H spectrum of P02 and YFV. (top) ^{1}H STD spectrum with on-resonance irradiation at -2 ppm for the sample shown on bottom. Total acquisition time was 28 hrs.

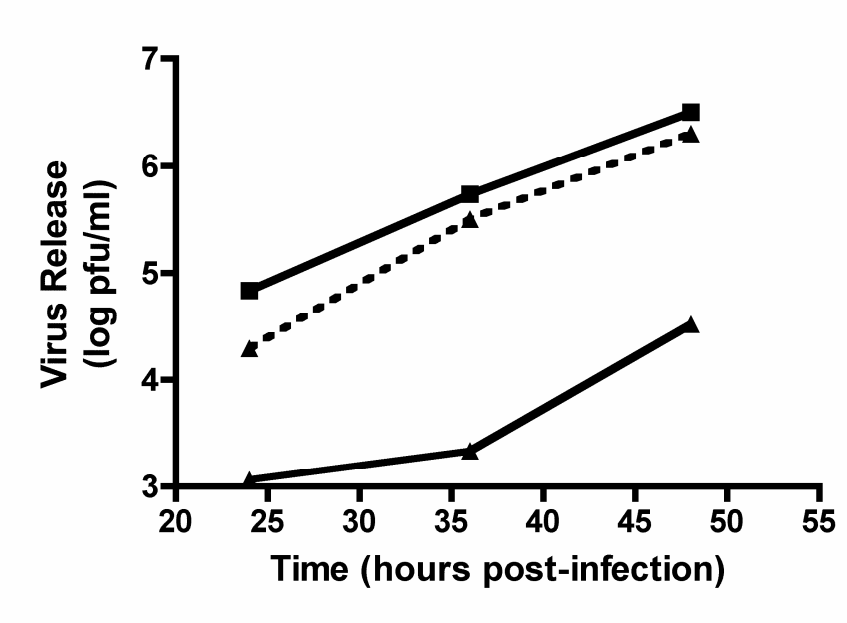


Figure S4. Release of YFV from BHK cells in the presence of compound P02. The pfu/ml is shown as a function of time for the cells overlaid with media containing 1% DMSO and 0 μM P02 (■, solid line), 10 μM P02 (▲, dashed line), or 100 μM P02 (▲, solid line). The data represent an average of two independent experiments.