Journal of Medicinal Chemistry Article

J. Med. Chem. **2010,** *53,* 1483–1495 **1483**
DOI: 10.1021/jm900776m

Novel Inhibitors of Dengue Virus Methyltransferase: Discovery by in Vitro-Driven Virtual Screening on a Desktop Computer Grid

Michael Podvinec,[†] Siew Pheng Lim,[‡] Tobias Schmidt,[†] Marco Scarsi,[†] Daying Wen,[‡] Louis-Sebastian Sonntag,[‡] Paul Sanschagrin,[§] Peter S. Shenkin,[§] and Torsten Schwede*[,†]

[†]*Swiss Institute of Bioinformatics and Biozentrum, University of Basel, Klingelbergstrasse 50, CH-4056 Basel, Switzerland,* [‡]*Novartis Institute for Tropical Diseases, 10 Biopolis Road, Chromos #05-01, 138670 Singapore, and* [§]*Schrödinger LLC, 120 West 45th Street, 29th Floor, New York, New York 10036-4041*

Received June 1, 2009

Dengue fever is a viral disease that affects 50−100 million people annually and is one of the most important emerging infectious diseases in many areas of the world. Currently, neither specific drugs nor vaccines are available. Here, we report on the discovery of new inhibitors of the viral NS5 RNA methyltransferase, a promising flavivirus drug target. We have used a multistage molecular docking approach to screen a library of more than 5 million commercially available compounds against the two binding sites of this enzyme. In 263 compounds chosen for experimental verification, we found 10 inhibitors with IC_{50} values of <100 μM, of which four exhibited IC_{50} values of <10 μM in in vitro assays. The initial hit list also contained 25 nonspecific aggregators. We discuss why this likely occurred for this particular target. We also describe our attempts to use aggregation prediction to further guide the study, following this finding.

Introduction

Dengue fever is a viral disease that is transmitted between human hosts by *Aedes* mosquitoes, particularly *Aedes aegyptii*. In 1997, 20 million cases of dengue fever were estimated to occur annually.[1,2] Partially because of increased urbanization and failure to effectively control the spread of the insect vector, more recent estimates suggest this number has risen to 50−100 million, and dengue fever is now seen as one of the most important emerging infectious diseases in many areas of the world.[3−5] Mild cases of dengue fever result in severe flulike symptoms, including fever, headache, and myalgia, but more severe cases can progress into a hemorrhagic fever and shock syndrome with considerable lethality.[6] Current treatment practice is nonspecific and symptomatic with a regimen of analgesics and fluid replacement, as neither specific drugs nor vaccines are available.[1]

Dengue virus is a plus-strand RNA virus belonging to the *Flavivirus* genus of the Flaviviridae family. Four serotypes have been isolated (DENV1−DENV4), and exposure to each of the serotypes conveys only partial immunity. Moreover, the presence of heterologous antibodies against a serotype other than the present infection may precipitate the more severe forms of dengue fever in patients.[7] In the absence of efficient and cost-effective vaccines, the development of inhibitors of viral or cellular enzyme targets as antiviral therapeutic agents is of particular interest.

The dengue genome, a single RNA strand 10.7 kb in length, is translated into a single polyprotein and later cleaved by viral and cellular proteases into 10 mature proteins. Three of the proteins have a structural role (C, prM, and E). In addition, seven nonstructural proteins (NS1, NS2A, NS2B, NS3, NS4A, NS4B, and NS5) are formed.[8]

Of the latter, NS3 and NS5 are the best understood to date, and both enzymes exhibit multiple domains and functions.[9,10] NS5 is the largest (900 amino acids) and most conserved protein in the dengue genome (67% sequence identity among serotypes 1−4).[8] It contains the RNA methyltransferase (MTase)[a] domain, as well as the RNA-dependent RNA polymerase necessary for virus replication. In this study, we focus on the discovery of compounds inhibiting the NS5 MTase, which has been proposed as a promising drug target against flaviviruses by us and others.[11−13]

The 5′ end of the dengue genome contains a type 1 cap structure, followed by the nucleotides AG, which are conserved in all flaviviruses.[14] Appropriate capping of cellular and viral RNA is known to increase translation efficiency as well as RNA half-life.[15,16] Host RNA is transcribed in the nucleus and processed by the cellular capping machinery. Dengue virus replication, however, occurs at the membrane of the endoplasmic reticulum; hence, a viral MTase is required for capping of the nascent viral RNA. Of the four steps necessary in *Flavivirus* cap formation, the final two methylation reactions are catalyzed

*To whom correspondence should be addressed: Swiss Institute of Bioinformatics and Biozentrum, University of Basel, Klingelbergstrasse 50, CH-4056 Basel, Switzerland. Telephone: +41 61 267 15 81. Fax: +41 61 267 15 84. E-mail: torsten.schwede@unibas.ch.

[a] Abbreviations: ATA, aurintricarboxylic acid; CF-I, cell-based flavivirus immunodetection; Cpd, compound; DENV, dengue virus; FN, false negative; FP, false positive; MTase, methyltransferase; NCI DTP, National Cancer Institute Developmental Therapeutics Program; MM-GBSA, molecular mechanics-generalized Born surface area; NS, nonstructural protein; RF, random forest; rmsd, root-mean-square distance; RTP, ribavirin triphosphate; SAH, *S*-adenosyl-L-homocysteine; SAM, *S*-adenosyl-L-methionine; SPA, scintillation proximity assay; TN, true negative; TP, true positive.

Published on Web 01/28/2010

by NS5 MTase with *S*-adenosyl-L-methionine (SAM) as the methyl donor, generating *S*-adenosyl-L-homocysteine (SAH) as a byproduct.[17,18] The cap guanine is methylated at the N7 position, resulting in a type 0 cap structure. Subsequently, the first RNA base, adenosine, is methylated at the 2′-OH group of the ribose, resulting in the formation of a type 1 cap structure.

The three-dimensional (3D) structure of the dengue NS5 MTase domain was the first *Flavivirus* MTase structure to be determined by X-ray crystallography.[18] Structures of the MTase complexed with *S*-adenosyl-L-homocysteine (SAH), the nonhydrolyzable GTP analogue GDPMP, ribavirin triphosphate (RTP), and a variety of RNA cap analogues (GpppA, GpppG, ^{7Me}GpppA, ^{7Me}GpppG, and ^{7Me}GpppG$_{2'OMe}$) have been published.[17−20]

Dengue MTase has an overall globular fold and shares a common fold with many SAM-dependent MTases, consisting of a seven-stranded β-sheet enclosed by four α-helices (subdomain 2).[21] This domain is surrounded by subdomain 1, an N-terminal extension of a helix−turn−helix motif, followed by a β-strand, an α-helix, and subdomain 3, a C-terminal extension consisting of an α-helix and two β-strands, spatially located between subdomain 2 and the N-terminal extension.[18] The enzyme has two specific binding sites where ligands have been cocrystallized (cf. Figure 1). The position of SAH indicates the binding site of the methyl donor, SAM. RNA cap analogues bind to a shallow second pocket formed between subdomains 1 and 2 (cf. Figure 1A). The two binding sites are connected by a common Y-shaped cleft, which suggests the placement of capped RNA along the cleft, positioning the first RNA nucleotide close to SAM, compatible with 2′-O-methylation. These positions are in accordance with observed positions of the RNA and cofactor in a complex structure of vaccinia virus VP39 MTase.[22]

Here, we present the results of our efforts to find novel classes of compounds inhibiting dengue MTase, potentially blocking viral replication. We have used a combination of large-scale structure-based computational analysis and enzyme inhibition assays. On the basis of structural analysis of dengue MTase, separate binding sites for RTP and SAM were targeted. For both sites, competitive inhibitors are known: SAH, sinefungin, and dehydrosinefungin have been characterized as efficient submicromolar competitive inhibitors of this MTase, and structural similarity to SAM strongly suggests their interaction with the SAM pocket.[23] Furthermore, two inhibitors of dengue MTase were published concomitant to this work. An inhibitor ($IC_{50} = 60.5\ \mu M$) has been found by Luzhkov et al. based on structural similarity to SAM,[13] and a docking study by Milani et al. has found aurintricarboxylic acid (ATA) to be a low-micromolar inhibitor of dengue MTase ($IC_{50} = 2.3\ \mu M$).[24] On the basis of the specific structural interactions of RTP ($IC_{50} = 101\ \mu M$)[19] and nucleotide or cap analogues with the RNA cap binding site, we consider this site a valid second target for inhibitors.

Our virtual screening approach was based on initial high-throughput docking calculations performed on a library of more than 5 million commercially available compounds. Using a personal computer (PC) grid to harness the idle computing power of our university's PCs, we were able to perform these calculations without prior focusing of the compound library. After the compounds had been docked, compound poses were refined, and promising candidates were assayed in vitro. Insights from these assays combined with pharmacophoric searches based on the predicted binding mode of actives were then used to select further compounds for follow-up testing. In the following, we will discuss our combined screening study, as well as the results obtained computationally and in vitro.

Materials and Methods

Chemical Compounds. All compounds in the docking database were associated with purchasing information, and compounds selected for inhibition assays were obtained from a variety of vendors. Compounds **1**−**9** (Table 1) were obtained from the NCI DTP Open Chemical Repository (http://dtp.nci.nih.gov) with the following compound codes: NSC12451, NSC15765, NSC26899, NSC49419, NSC54771, NSC84407, NSC91788, NSC14778, and NSC140047, respectively. Compounds **10**−**12**, **14**, **15**, **17**, **18**, **20**, **21**, **27**, and **33** were obtained from ChemBridge Corp. (San Diego, CA) (codes 5654575, 6490771, 7018889, 7936171, 7208655, 7746191, 7778100, 5219400, 7364286, 5255882, and 5917902, respectively). Compounds **13**, **24**, **26**, and **35** were from Enamine Ltd. (Kiev, Ukraine) (codes T0520-2463, T0511-8111, T5237786, and T5285909, respectively). Compounds **16**, **19**, **22**, and **23** were from InterBioScreen (Moscow, Russia) (codes STOCK1N-55803, STOCK2S-36613, STOCK3S-13122, and STOCK5S-06910, respectively). Compounds **25** and **28** were from InterChim (Montluçon, France) (codes STOCK1N-17364 and UZI/9041345, respectively). Compounds **29** and **30** were from Aurora Fine Chemicals (San Diego, CA) (codes Kenb-0135169 and Kina-0056391, respectively). Compound **31** was from Ambinter SARL (Paris, France) (code PHAR058572). Compound **32** was from TimTec LLC (Newark, DE) (code ST057026), and compound **35** was from Life Chemicals (Burlington, ON) (code F0777-1485).

Molecular Modeling. (i) Analysis of Dengue Methyltransferase Mutations and Structural Variability. For structural studies and for docking, an X-ray crystallographic structure of DENV2 MTase with bound SAH and RTP was used [Protein Data Bank (PDB) entry 1R6A]. To assess the conservation of protein residues, we extracted dengue MTase sequences from a database of all dengue sequences in UniProtKB release 14.0 using a blastp search with the sequence of PDB entry 1R6A as a query.[25,26] From the retrieved set of sequences, redundant sequences were removed, and 127 unique sequences were aligned using ClustalW with standard parameter settings.[27] Next, identity histogram values (I_p) were calculated at each position, where $I_p = (M - 1)/(N - 1)$, with p being the position in the alignment, M the number of prevalent residues in row p, and N the total number of residues in row p. Finally, residues were colored by identity histogram values in the Chimera software package.[28]

To study the structural variability seen in dengue MTase crystal structures, we obtained all available X-ray structures from the PDB[20] and optimally superposed their backbone atoms to the reference structure (PDB entry 1R6A). The average per-residue root-mean-square distances (rmsd) between the 1R6A structure and all other structures were calculated using VMD version 1.8.6[29] and colored accordingly.

(ii) Library of Purchasable Chemical Compounds. The compound library for screening was collected as follows. The all-purchasable subset of the ZINC V5 database, comprising ~2.7 million molecules from a variety of vendors, was obtained from http://zinc5.docking.org/.[30] To this collection were added 2.4 million nonredundant compounds from the Schrödinger in-house CACDB database of commercially available compounds. Ligands were prepared for docking using the *LigPrep* process (Schrödinger Suite 2007, Schrödinger LLC, NY). Briefly, the procedure was as follows. Ligands were desalted, neutralized, and parametrized using the OPLS 2005 force field. Next, tautomers and ionization states expected to occur in the pH range of 5.0−9.0 were generated using *ionizer* (Schrödinger Suite 2007). Wherever the stereochemistry of chiral centers

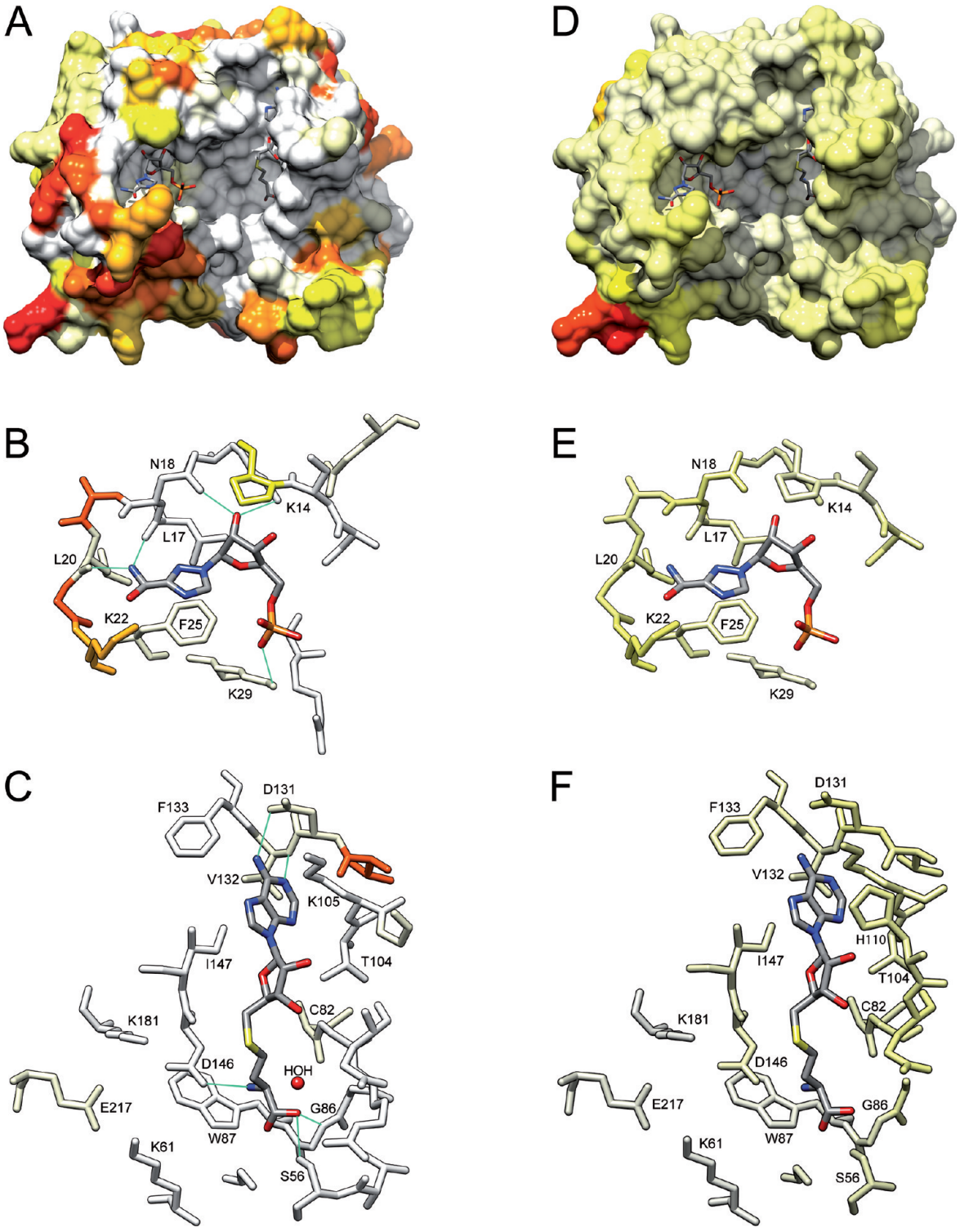

Figure 1. Sequence and structural conservation of DENV2 MTase. (A−C) Sequence conservation of dengue MTase. Sequence conservation is expressed as the identity histogram (I) of an alignment of 127 nonredundant dengue MTase sequences retrieved from UniProtKB. (A) Overall structure of DENV2 MTase in complex with RTP (left) and SAH (right). Ligands are displayed as element-colored licorice sticks. Surface gradient: from light gray (I = 1) to yellow (I = 0.947, i.e., 95% identical residues) to red (I = 0.323, i.e., 33% identical residues). (B) RNA cap binding site. Residues surrounding the inhibitor RTP are shown as licorice sticks, colored by degree of conservation as in panel A. RTP is shown in element-colored sticks (only one of three phosphate groups is shown). Residues undergoing key interactions with the ligand are labeled, and hydrogen bonds are depicted as cyan lines. (C) SAM binding site. Residues surrounding the reaction byproduct SAH are shown as licorice sticks, colored by degree of conservation as in panel A. SAH is shown in element-colored sticks. Residues undergoing key interactions with the ligand are labeled, and hydrogen bonds are depicted as cyan lines. (D−F) Structural variation calculated as a per-residue root-mean-square distance (rmsd) between the displayed structure and all other available DENV2 MTase crystal structures with and without bound ligands. (D) Overall MTase structure. The average rmsd is expressed as a color gradient: from light gray (rmsd = 0.0 Å) to yellow (rmsd = 1.0 Å) to red (rmsd ≥ 2.0 Å). (E) RNA cap binding site with residues close to the inhibitor RTP, colored as in panel D. (F) SAM binding site with residues close to the reaction byproduct SAH, colored according to the rmsd as in panel D. Surfaces were calculated using the MSMS package.[57]

Table 1. Predicted Binding Pocket and Measured Inhibition of Docked Compounds

Cpd	ID[a]	binding pocket	IC_{50} (μM) (Hill coefficient)	IC_{50} (μM) with 0.1% TX100 (μM) (Hill coefficient)	activity retained	EC_{50} (μM)	CC_{50} (μM)
1	NSC12451	SAM	29.9 (n.d.)	> 100 (n.d.)			
2	NSC15765	SAM	14.3 (1.9)	43.4 (2.3)	yes	> 100	> 100
3	NSC26899	SAM	25.3 (3.1)	> 100 (n.d.)			
4	NSC49419	SAM	27.59 (2.5)	> 100 (n.d.)			
5	NSC54771	SAM	27.53 (2.9)	> 100 (n.d.)			
6	NSC84407	SAM	31.43 (2.3)	> 100 (n.d.)			
7	NSC91788	SAM	29.03 (1.4)	> 100 (n.d.)			
8	NSC14778	SAM	1.52 (3.1)	9.46 (2.5)	yes	> 100	> 100
9	NSC140047	SAM	8.78 (1.9)	4.47 (2.2)	yes	> 100	> 100
10	ZINC 02911543	RNA cap	7.56 (1.5)	7.14 (1.4)	yes	> 100	> 100
11	ZINC 01174529	RNA cap	6.83 (2.9)	> 100 (n.d.)			
12	ZINC 03461039	RNA cap	7.11 (2)	> 100 (n.d.)			
13	ZINC 03287966	RNA cap	8.81 (2.3)	> 100 (n.d.)			
14	ZINC 01078518	RNA cap	9.28 (2.4)	64.2 (4.4)	yes	12	22.7
15	ZINC 01138375	RNA cap	11.35 (3.2)	> 100 (n.d.)			
16	ZINC 02129857	RNA cap	11.92 (1.9)	> 100 (n.d.)			
17	ZINC 01112283	RNA cap	13.16 (2.5)	> 100 (n.d.)			
18	ZINC 02849675	RNA cap	17.64 (2.8)	> 100 (n.d.)			
19	ZINC 00632055	RNA cap	20.32 (1.3)	> 100 (n.d.)			
20	ZINC 01467812	RNA cap	37.46 (1.5)	> 100 (n.d.)			
21	ZINC 02826899	SAM	2.91 (2.7)	> 25 (n.d.)			
22	ZINC 01878835	SAM	4.29 (4.1)	> 25 (n.d.)			
23	ZINC 01758620	SAM	9.62 (2.1)	> 100 (n.d.)			
24	ZINC 00633950	SAM	12.84 (1.7)	> 100 (n.d.)			
25	CACDB 1751080	SAM	16.87 (2.2)	79.8 (0.9)	yes	10.9	30.7
26	ZINC 02642996	SAM	16.09 (1.6)	> 100 (n.d.)			
27	ZINC 01226983	SAM	21.11 (2)	> 100 (n.d.)			
28	ZINC 02750651	RNA cap	2.81 (1.6)	19.55 (1.3)	yes	> 100	> 100
29	CACDB964942	RNA cap	13.50 (1.7)	87.1 (2.3)	yes	50.0	75.1
30	CACDB1563494	SAM	9.84 (2.1)	> 100 (n.d.)			
31	ZINC 01832826	RNA cap	4.42 (1.8)	44.5 (4.1)	yes	> 100	> 100
32	ZINC 01078734	RNA cap	12.39 (1.8)	> 100 (n.d.)			
33	ZINC01196449	RNA cap	7.99 (1.3)	> 100 (n.d.)			
34	ZINC02379945	RNA cap	14.50 (1.4)	> 100 (n.d.)			
35	ZINC03369470	RNA cap	4.80 (2.1)	4.91 (1.6)	yes	> 100	> 100

[a] Compound structures are depicted in Table S4 of the Supporting Information. n.d. = not determined.

was not specified, a maximum of two chiral centers was expanded into stereoisomers. Up to two low-energy conformations were produced for ligands with flexible ring systems. Ligand structures were minimized in implicit solvent using *bmin* (Schrödinger Suite 2007). The final library consisted of 5428096 structures.

(iii) Protein Preparation. In preparation for virtual screening, the enzyme structure from PDB entry 1R6A was modified as follows. All sulfate ions and water molecules found in the crystal structure were removed with the exception of HOH11, a structural water molecule found close to SAH. Moreover, the pose of the flexible Lys22 residue was replaced with an alternate rotamer, opening up the front of the RNA cap site to potentially accommodate larger ligands.

(iv) Ligand Docking and Compound Selection Procedures. Virtual screening and docking were performed using Glide version 4.5 (Schrödinger Suite 2007) using default docking parameter settings. A set of docking grids was generated independently for the RNA cap site and the SAM binding site using the default parameters. For the SAM site, the ligand's ability to form a hydrogen bond to the backbone N of Val132 (as observed with SAH) was required as a docking constraint. Next, a "funnel" strategy was employed for virtual screening. Initially, all compounds were docked using Glide in HTVS (High-throughput Virtual Screening) mode. After this rapid screening, the following compounds were selected for the next round. (1) All compounds ranked in the top 10% by GlideScore were picked. (2) All isomers (enantiomers, tautomers, and ring conformers) or alternate protonation states of compounds selected under 1 were chosen. (3) All docked poses forming a hydrogen bond to the Val132 backbone nitrogen were selected using a relaxed distance criterion of 3 Å. In the next round, these compounds were docked into the respective binding sites, using the Glide SP (Standard Precision) protocol. From this stage, compounds were selected as follows. (1) The top 10% of the compounds for each binding site by GlideScore were chosen. (2) Isomers of compounds selected in step 1 were included if found in the top 20% of compounds. These compounds were finally docked using the Glide XP (Extended Precision) procedure, and the 4000 top-ranked molecules from each binding site were selected for further refinement. Details on the number of compounds selected in each step are given in Results.

Following docking, selected compounds were passed through further refinement steps. (1) Additional input conformations for each selected compound were generated by reconstructing the geometry of each of the hit compounds and minimizing in implicit solvent or vacuum using the OPLS-AA or MMFF94 force fields[31,32] using MacroModel (Schrödinger Suite 2007). Alternate conformations were docked using Glide XP, and only the best-scoring pose was retained. The rationale for this enhanced sampling procedure was to ensure that found poses and scores are not influenced by subtle biases in the starting conformations of compounds induced by the force field. (2) We next applied a correction term to the docking score to account for internal ligand strain. The ligand strain correction term was calculated by optimizing the docked pose of the free ligand resulting from step 1 with torsion angle restraints and then

without such restraints. A portion of the difference in minimized energies (25% of the strain energy in excess of 4 kcal/mol) was used to calculate the correction term, which is applied to the original Glide docking score. (3) We then applied the Prime MM-GBSA rescoring protocol (Schrödinger Suite 2007). This procedure estimates the ligand binding free energy by performing minimization of the receptor−ligand complex, or the receptor and ligand alone. After refinement, a sorted list of the top compounds for each binding site was generated using the best-scoring Glide XP score of the multiple conformations, the strain-corrected XP score, and the MM-GBSA binding free energy, and three ranks were assigned to each compound, along with a consensus rank, calculated as the mean of the individual ranks. In summary, the refinement procedure first resampled ligand conformations and then provided the Glide XP score and two additional scores (strain-corrected Glide XP scores and MM-GBSA binding free energy estimates), as well as a consensus score for the selection of ligands for experimental verification.

For retrospective analysis, initial (nondocked) conformations of all 263 compounds tested in assays (regardless of activity) were collected and docked into both binding sites using Glide XP. Next, refinement and rescoring procedures as described above were applied to all resulting poses. To allow comparisons between the binding sites, all compound scores were converted into Z scores. The better-scoring pose of the two binding sites was then used in the generation of enrichment plots. With two exceptions, all the different scoring schemes employed in this process selected the same binding pocket as predicted in our previous docking attempts. We predicted compound **8** to dock into the SAM site, but as it does not form the H-bond to Val132, its pose in the RNA cap site was chosen. This bond was not required in the initial screening of the NCI library. For compound **31**, only the Prime MM-GBSA approach favored a pose in the SAM site over the RNA cap site, which was predicted by all other scores along with our previous predictions.

(v) Pharmacophore Searches. Pharmacophore generation and database searching were performed using Phase version 2.5 (Schrödinger LLC). Three different 3D pharmacophore hypotheses were generated on the basis of (A) a cluster consisting of five compounds [**10**, **13**, **17**, **20**, and **32** (Table 1)], including the confirmed hit compound **10**, (B) compound **10** alone, and (C) the cocrystallized MTase ligands ribavirin monophosphate and guanosine monophosphate. Pharmacophoric features were chosen so that they resembled the interactions between the ligands and the protein predicted by docking (A and B) or present in the cocrystallized structures[33] (C).

To allow flexible pharmacophore matching, a conformational search was performed on all compounds of the library of purchasable chemical compounds described above. For subsequent filtering of the full compound library, Phase default parameters were applied and all features defined in the pharmacophore hypotheses were required to match.

All compounds returned by the pharmacophore searches were subsequently docked to the RNA cap binding site and scored using Glide XP version 4.5 (Schrödinger LLC). The best-scoring pose for each compound was saved for further evaluation. As pharmacophore (A) resulted in a large number of hits, results were clustered by similarity, using MACCS structural keys fingerprints, a Tanimoto metric, and a degree of similarity of 60% using MOE 2007.09 (Chemical Computing Group, Montreal, QC). A diverse result set was obtained by picking the representative with the best GlideScore from each cluster.

Aggregation Prediction. (i) Test Sets. Three test sets were assembled to evaluate different predictors of compound aggregation. Briefly, the DenV test set consists of the molecules tested in this study, and the Med and AmpC test sets are taken from previously published studies on aggregation behavior.[34−36] See the Supporting Information for details.

(ii) Decision Tree Aggregation Prediction. We assembled a decision tree similar to that which Seidler et al. derived using recursive partitioning.[37] See the Supporting Information for details.

(iii) Random Forest Modeling of Aggregation Behavior. We applied the random forest method[38] to model aggregation in a manner similar to the approach published by Feng et al.[35] See the Supporting Information for details.

Computational Infrastructure. While visual analysis of protein and small-molecule structures, as well as analysis of physicochemical properties, was performed on standard Linux workstations, the preparation and filtering of the library for docking were performed on a Beowulf-type Linux cluster. Still, these resources were not sufficient to allow us to execute the planned large-scale molecular docking campaigns against the MTase. In the early stages of the project, we therefore built up a grid computing infrastructure at that time consisting of approximately 300 desktop PCs running the Windows 2000 or Windows XP operating system located in our institution's laboratories and classrooms, as well as some laboratory computers from another academic institution. These computers were tasked with docking or pharmacophore search jobs using Univa UD GridMP version 5.3. Schrödinger Suite 2007 supported this resource management system natively. As these types of computations are embarrassingly parallel, this resource's ready availability allowed us to screen large libraries for suitable compounds within reasonable amounts of time.

In Vitro Assays. (i) Methyltransferase Activity Assay. Unless otherwise stated, all compounds were first tested at a single maximum concentration of 25 or 100 μM followed by IC_{50} determinations with 2-fold serial dilutions starting from 25 or 100 μM following a previously described protocol.[23] In brief, inhibitors were assayed in a 96-well white opaque plates (Corning Costar, Lowell, MA) in 50 mM Tris-HCl (pH 7.0), 10 mM KCl, 2 mM $MgCl_2$, 2 mM $MnCl_2$, 0.05% (v/v) CHAPS, 2 mM DTT, and 5 units of RNasin inhibitor (Promega, Madison, WI). Typically, 25 nM DENV2 MTase enzyme and 40 nM biotinylated RNA substrate were preincubated with the test compounds at room temperature for 20 min, and the reaction was initiated by addition of 0.56 μM [*methyl*-^{3}H]AdoMet (72 Ci/mmol) (Amersham Biosciences, Piscataway, NJ). Under these conditions, the inhibitory effect of the reaction end product is negligible, as the amount of SAH produced during the reaction time (5−10 nM SAH produced in 20 min) is too small to have a significant impact on enzyme activity, as shown in Figures 3 and 4 of Lim et al.[23] To detect aggregators, the assay was varied as follows. Detergent sensitivity experiments were performed via addition of 0.01 or 0.1% Triton X-100 to the reaction mix.[39,40] For spin-down experiments, compound solutions were centrifuged for 15 min at 14000 rpm and room temperature before addition. For IC_{50} shift assays, 8 or 80 nM DENV2 MTase was used to reach a 10-fold difference in enzyme concentration. All other conditions were kept the same.

Reactions were stopped with buffer containing 100 mM Tris-HCl (pH 7), 50 mM EDTA, 300 mM NaCl, 8 mg/mL streptavidin SPA beads (Amersham Biosciences), and 125 μM cold *S*-adenosyl-L-methionine. Plates were read in a Trilux microbeta counter (Perkin-Elmer, Boston, MA) with a counting time of 1 min/well. All data points were measured in duplicate wells. IC_{50} curves were plotted with average counts per minute against the log of compound concentration. The standard deviation was calculated by the nonbiased $n - 1$ method, where standard deviation = $\{[n\sum x^2 - (\sum x)^2]/[n(n-1)]\}^{1/2}$. Nonlinear regression (curve fit) and the equation for the sigmoidal dose response (variable slope) from GraphPad Prism version 3.02 (GraphPad Prism, Inc., San Diego, CA) were used to interpolate values for IC_{50}. The equation is as follows:

$$Y = \text{bottom} + (\text{top} - \text{bottom})/[1 + 10^{(\log IC_{50} - X) \times \text{Hill slope}}]$$

where X is the logarithm of concentration and Y is the response. Y starts at bottom and goes to top with a sigmoid shape. This is identical to the "four-parameter logistic equation".[41,42]

(ii) Cell-Based *Flavivirus* Immunodetection (CF-I) Assays and Cytotoxicity Assay. The ability of compounds to inhibit dengue replication in a cell-based system (EC_{50}) as well as the cytotoxicity of test compounds (CC_{50}) was assayed as previously described.[43]

Results

Structural Analysis of DENV2 Methyltransferase. Viruses are known to benefit from short generation times and fast evolution rates in escaping host defense reactions as well as in developing resistance against therapeutic molecules.[44,45] Inhibitors must remain efficient against different serotypes and common mutations to qualify for further development. Figure 1A shows the conservation of individual MTase residues among peptide sequences from clinical isolates deposited in the UniProt database,[25] mapped onto the surface of DENV2 MTase in complex with RTP and SAH. For each residue, surface color is determined by the identity histogram value of an alignment of 127 unique DENV MTase sequences, shown in a spectrum from light gray (100% conservation) through yellow (95% conservation) to red (33% conservation). The Y-shaped central cavity is clearly visible, with the RNA cap located at the left-hand pocket (occupied by RTP in the structure shown) and SAM located to the right in place of SAH. As shown in Figure 1B, the RNA cap binding site is rather shallow and open. The aromatic ring of Phe25 undergoes π-stacking interaction with aromatic ring systems of the ligand. Hydrogen bonds can be formed from the ribose moiety to backbone oxygens of Asn18 and Lys14. The backbone oxygens of Leu20 and Leu17 likewise accept H-bonds, stabilizing the "front end" of ribavirin or the cap guanine. Moreover, electrostatic interactions between the phosphate groups of RTP and Lys29, Ser150, and Ser214 further stabilize ligand binding. The binding pocket of SAM (Figure 1C) is considerably more closed than the cap binding site. Important interactions are hydrogen bonds at both ends of the elongated SAM molecule, with Asp131 and Val132 fixing the adenine moiety and Gly86, Ser56, and Asp146 fixing the amino acid moiety at the opposite end. The elongated binding pocket is lined with predominantly apolar residues. Figure 1C also shows the catalytic tetrad comprised of Lys61, Asp146, Lys181, and Glu217 essential for RNA 2′-O-methylation.[17,46]

Most current docking algorithms treat the protein as a rigid structure. It is therefore of great importance to investigate and account for possible structural rearrangements between the unliganded and liganded states. Figure 1D shows the variations present in the available X-ray structures of the dengue MTase with the average per-residue root-mean-square distance (rmsd) mapped onto the surface of dengue MTase in complex with RTP and SAH. The variability of each residue is encoded by a color gradient from light gray (0 Å rmsd) to yellow (1 Å rmsd) to red (≥2 Å rmsd). Overall, the structural variability is small and mainly located on the outer, solvent-exposed surface of the protein. The RNA cap binding site (Figure 1E) and the SAM binding site (Figure 1F) show very little structural variability.

The small number of mutations observed in the MTase active sites and the relative rigidity of the enzyme's active sites, as witnessed by structural comparisons, make the MTase binding sites evolutionary and structurally stable targets for rigid protein molecular docking approaches.

Validation of the Docking and Assay Pipeline. In light of the small number of known inhibitors of the MTase, we favored structure-based over ligand-based computational approaches for the identification of promising inhibitory compounds. As a first step in validating our approach, we therefore redocked the cocrystallized ligands RTP and SAH into their respective binding pockets. These calculations were performed using Glide 4.5 in XP mode. The best-scored resulting poses were found to generally reproduce the binding mode of the crystallized ligands with heavy atom rmsd values of 0.74 and 1.33 for RTP and SAH, respectively (Figure 2).

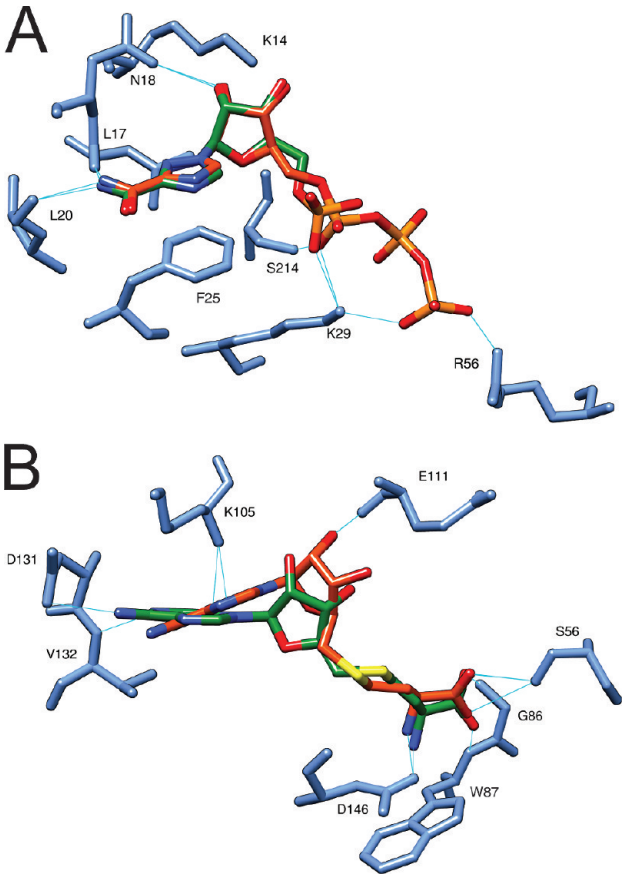

Figure 2. Redocking of RTP and SAH. (A) Crystallized pose of RTP shown with green carbons. The best redocked pose of RTP is shown with orange carbons. Surrounding residues undergoing important interactions are colored light blue. (B) Crystallized pose of SAH shown with green carbons. The best redocked pose of SAH is shown with orange carbons. Main interacting residues are colored light blue.

Despite the fact that the RNA cap site is rather shallow, the redocked pose of RTP closely resembles that of the crystallized compound. Notably, the terminal nitrogen of Lys14 was designated as protonated in our protein preparation procedure, inducing a slight difference in the position of the ribose 3′-hydroxyl group with regard to the experimental structure. As the original data contain coordinates for only ribavirin monophosphate modeled into the $F_o - F_c$ and $2F_o - F_c$ electron density maps,[19] the positions of the RTP β- and γ-phosphates were predicted, with the γ-phosphate stabilized by hydrogen bonds to Arg57 and Lys29. Redocking of SAH resulted in a pose similar to that of the SAH modeled into the electron density maps with a few differences. As a structural water molecule is absent, the sugar 2′-oxygen forms a direct H-bond to the backbone nitrogen of Glu111 rather than a water-mediated interaction as in the

experimental structure. This leads to a tilt in the plane of the adenine and a loss of the hydrogen bonds to Val132 and Asp131. Finally, the amino acid moiety of the SAH molecule is well anchored to Asp146, Trp87, Gly86, and Ser56 in a less strained conformation than that observed in the published structure. Comparable poses were also obtained for further known ligands of dengue MTase, GDPMP, SAM, and sinefungin (data not shown).

Our first effort was to screen a small set of 127000 compounds from the National Cancer Institute Developmental Therapeutics Program (NCI DTP).[47] Separate docking calculations were set up targeting either the RNA cap or SAM binding site. This study was conducted before the development of the protocol described in Materials and Methods and used a somewhat different procedure. Glide 4.5 was used in HTVS mode as a rapid first pass, mainly to eliminate compounds unlikely to fit to the binding pockets. Resulting poses were ranked by GlideScore, and the top 40% of hits against each binding site (79888 poses for the RNA cap site and 79946 poses for the SAM site) were extracted and redocked using Glide in SP mode. From the SP results, 100 compounds per site were chosen, on the basis of their ranked GlideScore and visual inspection for plausibility. Of this compound list, 40 were randomly selected for in vitro testing for dengue MTase inhibition. Of 36 compounds obtained from the NCI, nine compounds (Table 1, compounds **1**−**9**) were found to be inhibitors of the MTase in vitro and will be further discussed below. With the workflow from computation to inhibition data in place, we now were ready to screen the large library of commercially available compounds previously compiled.

High-Throughput Docking. Both binding sites were next targeted in molecular docking campaigns using the whole compound library described above. During the whole procedure, the two binding sites were treated separately. For each site, compounds were docked using the funnel strategy described in Materials and Methods. After the first screening step, 1.01×10^6 compounds for the RNA cap site and 6.72×10^5 compounds for the SAM site were docked using the SP protocol; 1.09×10^5 compounds (RNA cap site) and 7.7×10^4 compounds (SAM site) were finally subjected to the XP procedure, and the 4000 top-ranking molecules from each binding site were selected for further refinement.

To prevent the lowest-energy pose finally retained from being influenced by artifacts of minimization on a 3D grid, each hit molecule was reconstructed and minimized under different conditions (solvent and force field). These additional input conformations were docked, and only the best-scored pose was retained for each compound (3392 compounds for the RNA cap site and 3365 compounds for the SAM site), discarding isomers along with suboptimal poses.

Candidate Selection and Inhibition Assays. Next, we calculated additional scores for the docked compounds. First, a correction term for ligand strain was applied to the GlideScore. Second, ligand binding energies were estimated using the MM-GBSA protocol in Prime version 1.6. The three scores were subsequently combined into a rank-based consensus score. Final short lists for each binding site were as follows: the 200 top compounds by consensus and the top 100 compounds by each of the individual scores. These overlapping criteria resulted in approximately 350 compounds per binding site, which were visually inspected and prioritized by a jury panel drawn from our institutions, involving four or five independent jurors. In this step, different criteria were considered: diversity of chemical moieties covered by selection, credibility of pose based on experience in protein X-ray crystallography, similarity to known inhibitor poses, and presence of one or more key and additional intermolecular contacts with penalization of very close distance contacts. The consensus opinion of the jurors was used to produce a short list of 100 prioritized compounds for each binding site.

In total, 183 compounds could be obtained from vendors and were assayed for their ability to inhibit the transfer of a ^{3}H-labeled methyl group from SAM to a short synthetic GTP-capped RNA oligonucleotide using a scintillation proximity assay.[23] Initial testing for MTase inhibition was conducted at a single concentration (25 or 100 μM), and IC_{50} concentrations were then determined for compounds showing substantial inhibition ($>40\%$) in these experiments. Of the compounds tested, 23 were found to be inhibitors of DENV2 MTase with a spectrum of IC_{50} values ranging from 2.62 to 37.46 μM (Table 1, compounds **10**−**32**).

Pharmacophore Screening. To retrieve further active compounds from the compound database, we next built a five-feature pharmacophore hypothesis from the predicted binding modes of compounds **10**, **13**, **17**, **20**, and **32** in the RNA cap site, reasoning that factors important for ligand binding may be inferred from the predicted binding poses of hit compounds. From a 3D superposition of these five compounds, a pharmacophore hypothesis was built using Phase version 2.5, consisting of pharmacophore features common to all compounds, resembling the protein−ligand interactions as predicted by docking. The hypothesis consists of five pharmacophoric features: two aromatic rings corresponding to (1) the diphenylamino ring stacking with Phe25, (2) the benzenesulfonate moiety, as well as hydrogen bond (3) donor and (4) acceptor features representing the diamino/amide moiety interacting with Ser150 and Ser214, and (5) a negatively charged group on the sulfonate moiety, interacting with Lys29 and Arg212 (Figure 3).

Searching through the compound database resulted in 4200 hits to the pharmacophore, which were subsequently docked into the RNA cap binding site and scored using Glide XP. To reduce the number of compounds but retain structural diversity, the 308 most diverse structures with the best GlideScore were selected on the basis of structural similarity. Resulting poses were visually evaluated by a jury as described above. Eighteen compounds were selected and subsequently tested in vitro as described above. Dose-dependent inhibition of compounds **33**−**35** was found to exhibit IC_{50} values between 4.8 and 14.5 μM (Table 1). Additional pharmacophore searches of the compound database were performed using six-feature models derived from either compound **10** only or the experimental structures of GMP and RTP in the binding pocket, retrieving 149 and 193 compounds, respectively. Of these, five and four compounds were selected for in vitro assays as described above, respectively, but no further inhibitors of MTase were found.

Testing Hit Compounds for Unspecific Inhibition. Of the compounds tested for inhibition, a large fraction exhibited a Hill coefficient substantially larger than unity. A large Hill coefficient signifies that a small increase in inhibitor concentration leads to an anomalously large change in inhibition, which can stem from positive ligand cooperativity, enzymes with equivalent binding sites, but also from non-ideal, nonspecific behavior that leads to abrupt enzyme inhibition above a critical concentration. One such mechanism

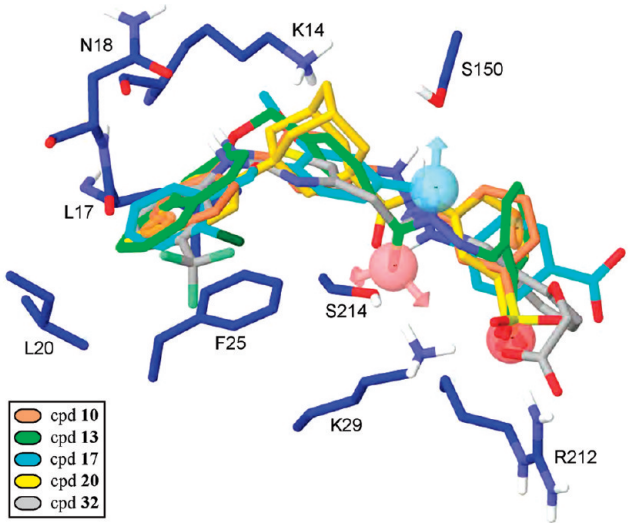

Figure 3. Pharmacophore query. Predicted binding modes of five compounds, obtained by docking to the RNA cap binding site and pharmacophore hypothesis created from the predicted binding modes. Ligands are shown as licorice sticks with colored carbons. Compound numbers in the legend refer to Table 1. Surrounding residues are labeled and shown with blue carbons. The pharmacophore hypothesis consists of five features: two aromatic rings (orange rings), one hydrogen bond acceptor and one donor (red and blue spheres with arrows, respectively), and one negatively charged group (red sphere).

is aggregation,[48] which has been recognized as a major cause of false positives in high-throughput screening.[37,39,49,50] The addition of the nonionic surfactant Triton X-100 (0.1%) to the assay solution is often used to prevent the formation of compound aggregates without influencing enzymatic activity. While the addition of 0.1% Triton X-100 to our assay did not abolish inhibition by sinefungin, it had a marked effect on the majority of the previously chosen compounds.

Of the 35 compounds, 10 retained inhibitory activity (IC_{50} $< 100\ \mu M$) in this assay (Table 1, compounds **2**, **8**−**10**, **14**, **25**, **28**, **29**, **31**, and **35**) (Figure 4), and 25 were rejected, exhibiting essentially flat dose−response curves in the presence of surfactant. The IC_{50} values of four of the actives changed ≤3-fold compared to measurements without surfactant: 14.3 μM for **2**, 4.47 μM for **9**, 7.14 μM for **10**, and 4.91 μM for **35**. Compound **8** exhibited a 6.18-fold change with the addition of Triton X-100 but retained an IC_{50} value of $< 10\ \mu M$. Compounds **14**, **25**, **28**, **29**, and **31** have IC_{50} values shifted more than 3-fold from the values obtained without Triton X-100 and do not saturate the MTase enzyme at the highest inhibitor concentrations (50 and 100 μM) used in the assays (Figure 4). In Figure 5, the two-dimensional structures and predicted binding modes are shown for hits with IC_{50} values of $< 10\ \mu M$. As these low-micromolar inhibitors are of particular interest, we performed two additional experiments to further rule out aggregation as the cause of inhibition.

First, compounds were assayed after centrifugation at 14000 rpm for 15 min (spin-down assay) to deplete the solution of any colloid particles. IC_{50} values obtained under these conditions are comparable to those obtained in previous assays with Triton X-100: 6.22 μM vs 9.46 μM (Cpd **8**), 10.52 μM vs 4.47 μM (Cpd **9**), 10.68 μM vs 7.14 μM (Cpd **10**), and 4.34 μM vs 4.91 μM (Cpd **35**). Second, IC_{50} values of the compounds were assayed in the presence of 8 or 80 nM MTase. Nonaggregating compounds should be insensitive to this shift in enzyme concentration. Compounds **8** and **10** exhibit relatively small changes in IC_{50}, which may be attributed to experimental variation: 2.3-fold for Cpd **8** [from 3.16 μM (8 nM) to 7.14 μM (80 nM)] and 1.3-fold for Cpd **10** [from 7.62 μM (8 nM) to 10.24 μM (80 nM)]. The two remaining compounds, **9** and **35**, exhibit larger shifts: 6.4-fold for Cpd **9** [from 3.27 μM (8 nM) to 20.82 μM (80 nM)] and 8.0-fold for Cpd **35** [from 1.32 μM (8 nM) to 10.6 μM (80 nM)].

All 10 active compounds were further assessed by the CF-I assay to examine their activities against dengue virus replication. The assay is based on quantitative immunodetection of dengue virus E protein production in target cells.[43] EC_{50} and CC_{50} values are reported in Table 1. Of the tested compounds, only compounds **14** ($EC_{50} = 12\ \mu M$), **25** ($EC_{50} = 10.9\ \mu M$), and **29** ($EC_{50} = 50\ \mu M$) elicited a response in the cellular assays. Unfortunately, these compounds also exhibit cytotoxicity in the midmicromolar range (CC_{50} values of 22.7, 30.7, and 75.1 μM, respectively).

Modeling of Aggregation Behavior. The unexpected elimination of many compounds that initially tested positive due to their action as aggregators prompted us to further investigate this nonspecific effect. Different computational approaches that are trained on experimental data to predict aggregation behavior of chemical compounds have been published.[35,37] For our study, we were interested in how well similar predictors can be applied to different biological targets and assay conditions, as training and validation sets are usually measured under identical conditions. To assess the transferability of the proposed methods and evaluate them for our biological target, we have assembled three data sets of compounds with known aggregation properties in their respective assays: one based on the data from this work and two based on previously published large-scale aggregator detection assays.

We adapted the decision tree proposed by Seidler et al.[37] to molecular descriptors available to us and applied this predictor to the three test sets (Table S2 of the Supporting Information). Of the 182 compounds assayed from the large docking study, 23 initially appeared to be active in vitro. Of these, one compound was reliably not aggregating ($IC_{50} <$ 10 μM in the presence of detergent). In this data set, our decision tree classifier underestimated the overall aggregation tendency: 15 (65%) of these compounds were predicted to aggregate. Furthermore, compound **10**, the highest-affinity nonaggregator, was wrongly classified. Overall, the aggregation tendency was predicted correctly 61% of the time. Prior to in vitro testing of the 27 compounds selected from the pharmacophoric search, we ran the same aggregation predictor against this data set. Six (22%) of the compounds were predicted to aggregate. Despite the imperfections of the aggregation predictor, this value was sufficiently low, compared to the 65% prediction for the first-round actives, that we were confident that the second-round compounds would exhibit a significantly weaker aggregation tendency.

We also evaluated the use of a random forest classifier as described by Feng et al.[35] to predict aggregators in the three test sets, on the basis of calculated physicochemical properties. Validation within test sets yielded acceptable false positive (FP) and false negative (FN) rates (see the diagonal elements in Table S3 of the Supporting Information), comparable to results reported elsewhere.[35] When random forest

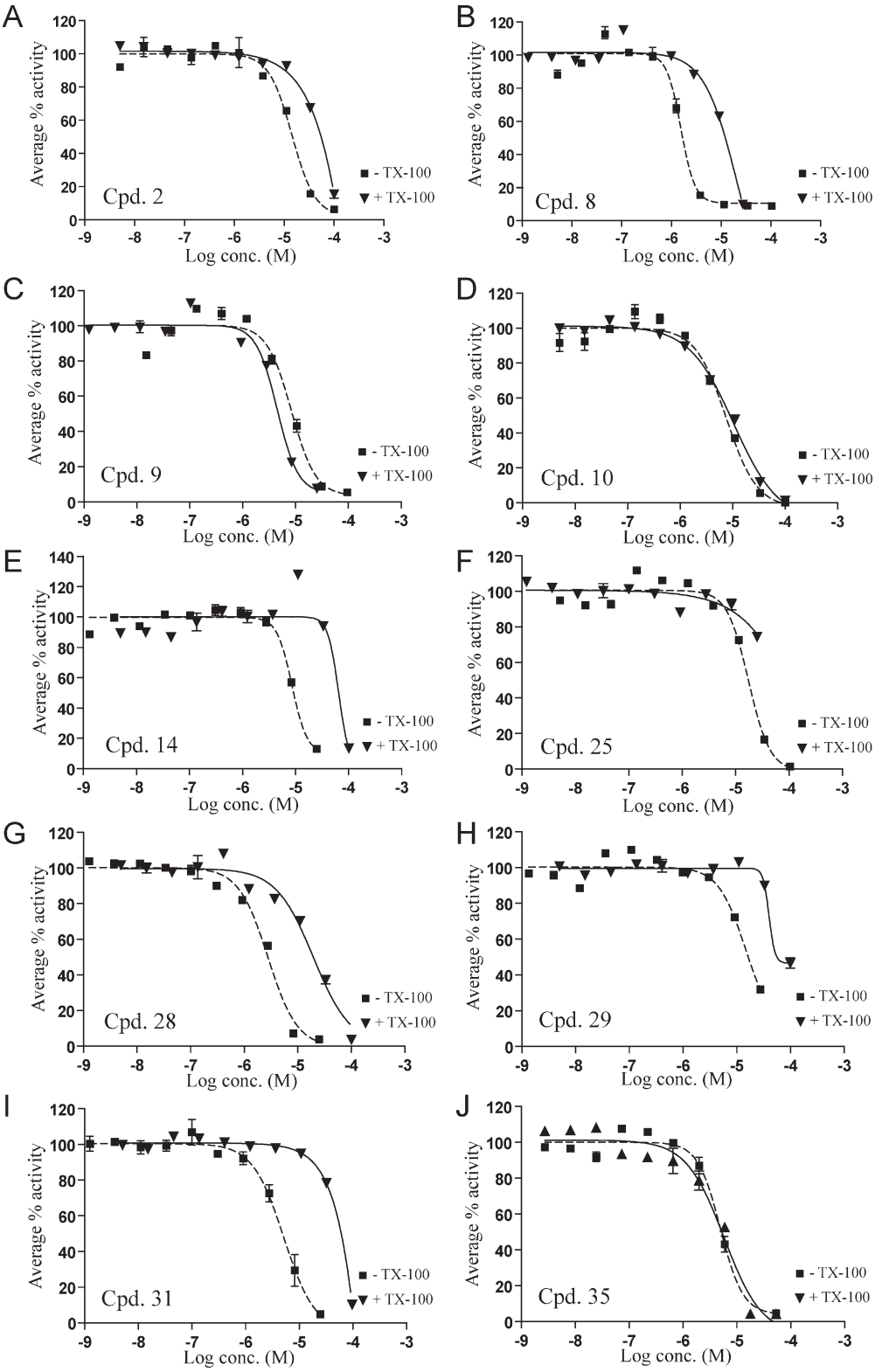

Figure 4. Dose-dependent inhibition of MTase activity. Recombinant NS5 MTase was preincubated with RNA cap analogue and varying concentrations of inhibitor. After addition of radiolabeled SAM, the transfer of the labeled methyl group to the RNA substrate was quantified using a scintillation proximity assay. Boxes and dashed lines show data for inhibition measured without addition of Triton X-100, and triangles and solid lines represent measurements in the presence of 0.1% Triton X-100. Measured counts per minute were normalized by dividing by the top curve value. (A−J) Inhibition of MTase by compounds **2**, **8**−**10**, **14**, **25**, **28**, **29**, **31**, and **35** (see Table 1).

models trained on the full data set for one assay condition were applied to data obtained with other biological targets or assay conditions (Table S3 of the Supporting Information), significantly larger false positive and false negative rates resulted. In summary, we found both predictors to suffer from a weak ability to discriminate nonaggregating compounds (specificity), particularly when applied to conditions other than those on which they were trained.

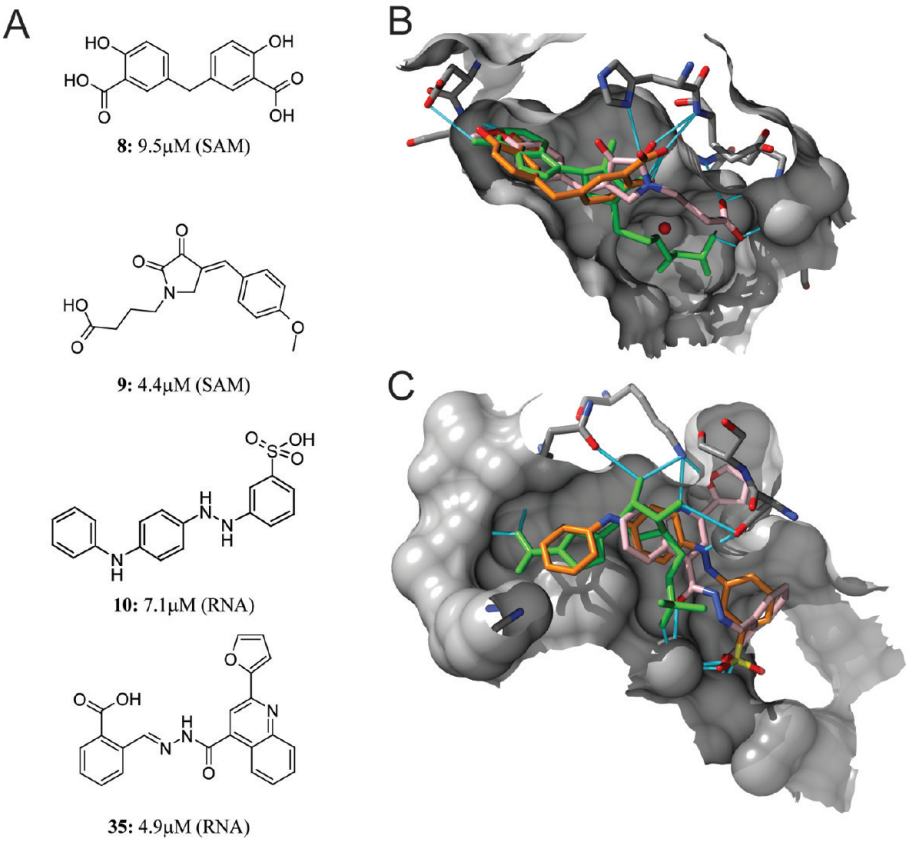

Figure 5. Structure and binding mode of DENV2 MTase inhibitors. (A) Structure of DENV2 MTase inhibitors with IC_{50} values of $< 10\,\mu M$. Compound numbers refer to Table 1, followed by the IC_{50} value. The predicted target site is given in parentheses (SAM, SAM site; RNA, RNA cap site). (B) Crystallized pose of SAH shown as green licorice sticks. Compound **8** is shown with orange carbons, and compound **9** is shown with pink carbons. (C) Crystallized pose of RTP shown as green licorice sticks. Compounds **10** and **35** are shown with orange and pink carbons, respectively.

Of the 27 compounds selected for laboratory screening from our pharmacophoric searches, 23 came from models that included compound **10**, the best nonaggregating hit from the large-scale docking study. It is tempting to speculate that the reason this data set exhibited the lower fraction of aggregators among the apparent hits might be that the nonaggregator found in the earlier search was used to derive these models. It is not possible to draw this conclusion, given the limited number of compounds tested. However, the limited results we have obtained do suggest as a subject for future investigation that similarity to nonaggregating hits found in an early screen could be a useful criterion to incorporate into the development of a second-generation data set for a later screen.

Retrospective Analysis of Refinement and Rescoring Procedures. As the exact procedures for compound screening, refining, and rescoring evolved over the course of this work, we next performed a coherent retrospective analysis of the total set of 263 compounds tested for MTase inhibition in this work. In this analysis, the enrichment of compounds active in the inhibition assay (**2**, **8**−**10**, **14**, **28**, **29**, **31**, and **35**) was calculated at different steps of the procedure. All compounds were first passed through the final screening pipeline: (1) docking to the two binding sites using Glide XP, (2) refinement using resampled ligand conformations, and (3) generation of two additional scores (strain-corrected Glide XP scores and Prime MM-GBSA binding free energy estimates), as well as a consensus score. We next plotted enrichment of actives before and after the refinement procedure (Figure 6A). Following the procedure applied in docking the large compound library, we next compared enrichment of actives achieved with each of the scores (Glide XP score of the refined pose, Glide XP score with internal strain correction, and Prime MM-GBSA binding free energy) and a consensus score, the rank-based average of individual scores (Figure 6B). As the number of actives in this data set is rather small, results from these analyses necessarily remain indicative, rather than conclusive. Nevertheless, our data suggest that the choice of rescoring scheme has an impact on the resulting hit list, whereas the impact of resampling ligand conformations is at least in our system less apparent.

Discussion

In this work, we describe the outcome of a combined computational and experimental study searching for novel inhibitors of dengue NS5 MTase among commercially available compounds. The viral methyltransferase possesses two binding sites that can be targeted in principle. However, the RNA cap site is rather shallow and solvent-exposed, so that molecules interacting firmly with this site are challenging to find. The second site binds the ubiquitous cofactor SAM, which invokes problems of specificity and off-target activity. For instance, the submicromolar inhibitor of dengue MTase sinefungin showed promise as an antibiotic, antiviral, and antiparasitic agent but was not further pursued because of its severe nephrotoxicity and lack of specificity.[51−53]

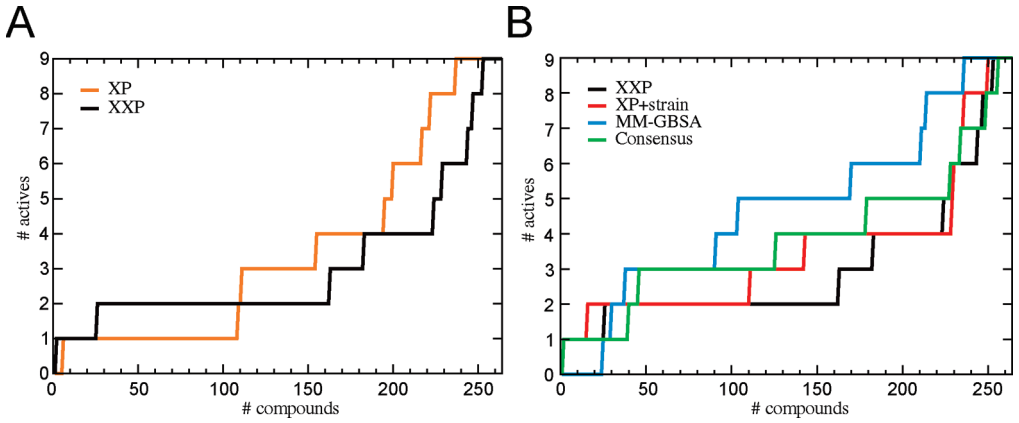

Figure 6. Enrichment of active vs inactive inhibitors. The number of actives recovered is plotted on the *Y*-axis against the size of the database screened on the *X*-axis. (A) Enrichment obtained with standard Glide XP docking (XP, orange) and with the best-scored Glide XP pose after increased sampling of starting conformations (XXP, black). (B) Enrichment obtained with individual scoring schemes and consensus score: Glide XP Score (XXP, black), Glide XP Score with internal strain correction (XP+strain, red), Prime MM-GBSA (MM-GBSA, blue), and rank-based consensus score (Consensus, green).

After an initial investigation of the target for suitability, we launched two virtual screening campaigns. Since little information was published about inhibitors of dengue MTase and their binding mode at the outset of the project, we pursued a broad-sweeping approach, using molecular docking to screen a library of several million compounds, rather than a small focused library. This approach was leveraged by the ready availability of computing resources through a computing grid that put more than 300 idle desktop PCs at our disposal.

The hits found must be considered starting points. The lack of an effect in cell-based assays observed with seven of the compounds may indicate cell permeation or stability problems, and the compounds (**14** and **29**) eliciting a response in these assays show an onset of cytotoxicity close to the half-maximal efficient dose. Nevertheless, these compounds inhibit the enzyme in in vitro assays, represent diverse chemotypes and predicted binding modes due to the study design, and therefore provide a number of inroads toward more focused approaches, which are currently under investigation.

Having at our disposal a set of 263 compounds characterized in inhibition assays, we retrospectively investigated the effect of decisive steps in our compound selection procedure on the resulting hit list. Notably, this set is a mixture of actives and very difficult decoys. All compounds initially docked well to one of the binding sites, were scored high, and were selected by a human panel as promising. Thus, absolute enrichment is not as interesting as are the differences between enrichments obtained by the various methods. In our system, increased sampling of starting conformations had no positive effect on the outcome, indicating that found poses and scores were not influenced by starting conformation biases. This is a positive finding. It indicates that the docking approach used here was not hampered by artifacts introduced by grid-based energy minimization. The comparison of different scoring procedures proves to be more interesting. In this case, where we try to discriminate between compounds scored closely together, Prime MM-GBSA binding free energies seem to provide the best enrichment of true actives for this target, closely followed by the consensus score obtained by the rank average of all three scoring methods used. Future virtual screening efforts against dengue or closely related MTases may benefit from this finding.

Why did our virtual screening result in the discovery of many aggregators? Since, in a virtual screen, there is no opportunity for multiple copies of the ligand to interact with each other, there must be some other explanation for this observation. Most likely, the explanation lies in the fact that the binding sites studied here lend themselves to binding by long, flat molecules which, due to their shape and aromatic nature, have a strong tendency to aggregate. If this is correct, we would expect high-ranking virtual screening candidates obtained against binding sites more polar or more compact than those studied here to exhibit a much weaker tendency to aggregate. To put matters into perspective, the overall percentage of aggregators picked up in our study (25 of a total of 263 compounds tested, or 9.5%) is lower than that found in another study, where 19% of randomly chosen druglike compounds were acting as aggregators.[35] While docking does not select against aggregators, it does not select for them, either.

We tested two computational methods to predict compound aggregation to see if nonspecific binders can be identified before the in vitro stage. When a trained classifier is applied to a set of compounds assayed under different conditions, the level of misclassification is significantly increased in a manner independent of the training set used, making these predictions of limited practical use. As the tested models take compound properties but not the conditions of the assay into account, this is not surprising. Interestingly, the decision tree model proved to be more transferable than the random forest models.

As described earlier, we were confident, on the basis of the results of aggregation prediction, that the second-round compounds would exhibit a lower fraction of aggregates than the first-round actives. Laboratory results were in accord with this expectation. Of 27 compounds assayed, only two (7.4%) turned out to be aggregators. Of the three apparent hits, one (33%) turned out to be a nonaggregating hit, compared to one of 23 compounds (4.3%) from the large-scale docking screen. Apparent hits in the phamacophoric screen were considerably more enriched in nonaggregators than the hits from the large-scale docking screen.

As the majority of compounds selected from the pharmacophoric searches came from models that included compound **10**, the best nonaggregating hit, it is tempting to speculate that similarity to nonaggregating hits found in an early screen could be a useful criterion to incorporate into the development

of a second-generation data set for a later screen, which would be a worthwhile topic for a future study.

In summary, our results, as well as the observations of others, do not suggest the broad use of computational methods in filtering out aggregators for two reasons. First, available methods lack accuracy and transferability between different assay conditions. Second, a number of active ligands and marketed drugs were found to be bona fide aggregators in inhibition assays[37] but pharmacologically active at concentrations lower than that at the onset of aggregation. This strongly suggests that the right way to control aggregation is by preventing it from interfering with the inhibition assay wherever possible. Aggregation prediction has a role, however, in the smart selection of limited follow-up compound sets from primary hits.

The cost of a screening campaign is a major concern, especially when neglected diseases are being targeted. We explored the extent to which in vitro experiments can be replaced by comparably inexpensive computational analysis. Roughly 5.5 million compounds were initially considered, and 263 molecules were assayed in the laboratory. Of these, 10 were initially characterized as inhibitors. When only compounds exhibiting IC_{50} values of $< 10\ \mu M$ and a Hill coefficient of ≤ 2.5 were selected, four of the 10 compounds remain (**8**−**10** and **35**) as potentially interesting inhibitors. To further rule out aggregation as their mode of action, we applied two additional assays to these candidates. While the spin-down assay validated all four compounds, only two of the compounds (**8** and **10**) did not exhibit marked shifts in IC_{50} when the enzyme concentration was changed 10-fold, strengthening the case for these two compounds in particular.

Compounds showing inhibition in the higher micromolar range (in particular, compounds **2** and **28**) are worth following up, as well, but may reveal themselves to be insufficiently specific in further tests. The future use of compounds **8**−**10** and **35** (Table 1 and Figure 6) should be considered in light of their properties. The object of this investigation was to find tool compounds and starting points for further optimization, using a set of commercially available compounds. While the compounds identified are not druglike in their properties and may not readily enter cells, they can serve as tools for cocrystallization or as starting points for scaffold hopping and bioisostere replacements. Interestingly, compound **8** exhibits a striking similarity to ATA, a low-micromolar inhibitor of DENV2 MTase found in a docking study published while this manuscript was being prepared.[24] Compound **9** is currently under further investigation in a program directed at identifying SAM competitive inhibitors.

In conclusion, the outcome of our study demonstrates that iterative combination of virtual screening and validation in the laboratory is a viable approach for the discovery of new hits against drug targets and can serve as a model for similar endeavors against other diseases. Computing power is becoming increasingly inexpensive or even free, as volunteers openly welcome requests for support of projects with a charitable aspect. Indeed, a number of grid-based drug discovery efforts have recently been launched.[54−56] However, we strongly believe that the key to success is not access to virtually unlimited computing capacity, but rather establishing a tight interaction cycle between the computational and experimental parts of the project even before the first calculation takes place.

Acknowledgment. We acknowledge Jeremy R. Greenwood (Schrödinger LLC), Jürgen Kopp (formerly of University of Basel), Eric Vangrevelinghe (Novartis Institutes for Biomedical Research), and Shahul Nilar (Novartis Institute of Tropical Diseases) for helpful discussions concerning methodology as well as valuable input in compound selection. We thank Hao Ying Xu and Boping Liu (Novartis Institute of Tropical Diseases) for their assistance in testing compounds in CF-I assays. Furthermore, we thank Alex Matter (formerly of Novartis Institute of Tropical Diseases), Manuel Peitsch (formerly of Novartis Institutes for Biomedical Research), and Jörg Weiser (Schrödinger LLC), who were instrumental in the origination of the collaboration.

Supporting Information Available: Implementation of decision tree and random forest aggregation predictors, and structures of tested compounds (Tables S1−S4). This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) Dengue and Dengue Hemorrhagic Fever: Information for Health Care Practitioners. http://www.cdc.gov/NCIDOD/dvbid/dengue/dengue-hcp.htm (accessed September 29, 2008).
(2) *Dengue haemorrhagic fever: Diagnosis, treatment, prevention and control*, 2nd ed., World Health Organization: Geneva, 1997.
(3) Morens, D. M.; Fauci, A. S. Dengue and hemorrhagic fever: A potential threat to public health in the United States. *JAMA, J. Am. Med. Assoc.* **2008**, *299* (2), 214–216.
(4) Dengue and dengue haemorrhagic fever. http://www.who.int/mediacentre/factsheets/fs117/en/ (accessed September 29, 2008).
(5) Guzman, M. G.; Kouri, G. Dengue and dengue hemorrhagic fever in the Americas: Lessons and challenges. *J. Clin. Virol.* **2003**, *27* (1), 1–13.
(6) Gould, E. A.; Solomon, T. Pathogenic flaviviruses. *Lancet* **2008**, *371* (9611), 500–509.
(7) Whitehead, S. S.; Blaney, J. E.; Durbin, A. P.; Murphy, B. R. Prospects for a dengue virus vaccine. *Nat. Rev. Microbiol.* **2007**, *5* (7), 518–528.
(8) Perera, R.; Kuhn, R. J. Structural proteomics of dengue virus. *Curr. Opin. Microbiol.* **2008**, *11* (4), 369–377.
(9) Padmanabhan, R.; Mueller, N.; Reichert, E.; Yon, C.; Teramoto, T.; Kono, Y.; Takhampunya, R.; Ubol, S.; Pattabiraman, N.; Falgout, B.; Ganesh, V. K.; Murthy, K. Multiple enzyme activities of flavivirus proteins. *Novartis Found. Symp.* **2006**, *277*, 74–84, 251−253 .
(10) Mukhopadhyay, S.; Kuhn, R. J.; Rossmann, M. G. A structural perspective of the flavivirus life cycle. *Nat. Rev. Microbiol.* **2005**, *3* (1), 13–22.
(11) Podvinec, M.; Schwede, T.; Peitsch, M. C. Docking for neglected diseases as community efforts. In *Computational Structural Biology: Methods and Applications*; Schwede, T., Peitsch, M. C., Eds.; World Scientific Publishing: Singapore, 2008; pp 683−704.
(12) Dong, H.; Zhang, B.; Shi, P. Y. Flavivirus methyltransferase: A novel antiviral target. *Antiviral Res.* **2008**, *80* (1), 1–10.
(13) Luzhkov, V. B.; Selisko, B.; Nordqvist, A.; Peyrane, F.; Decroly, E.; Alvarez, K.; Karlen, A.; Canard, B.; Qvist, J. Virtual screening and bioassay study of novel inhibitors for dengue virus mRNA cap (nucleoside-2′O)-methyltransferase. *Bioorg. Med. Chem.* **2007**, *15* (24), 7795–7802.
(14) Cleaves, G. R.; Dubin, D. T. Methylation status of intracellular dengue type 2 40 S RNA. *Virology* **1979**, *96* (1), 159–165.
(15) Furuichi, Y.; Shatkin, A. J. Viral and cellular mRNA capping: Past and prospects. *Adv. Virus Res.* **2000**, *55*, 135–184.
(16) Ray, D.; Shah, A.; Tilgner, M.; Guo, Y.; Zhao, Y.; Dong, H.; Deas, T. S.; Zhou, Y.; Li, H.; Shi, P. Y. West Nile virus 5′-cap structure is formed by sequential guanine N-7 and ribose 2′-O methylations by nonstructural protein 5. *J. Virol.* **2006**, *80* (17), 8362–8370.
(17) Egloff, M. P.; Decroly, E.; Malet, H.; Selisko, B.; Benarroch, D.; Ferron, F.; Canard, B. Structural and functional analysis of methylation and 5′-RNA sequence requirements of short capped RNAs by the methyltransferase domain of dengue virus NS5. *J. Mol. Biol.* **2007**, *372* (3), 723–736.
(18) Egloff, M. P.; Benarroch, D.; Selisko, B.; Romette, J. L.; Canard, B. An RNA cap (nucleoside-2′-O-)-methyltransferase in the flavivirus RNA polymerase NS5: Crystal structure and functional characterization. *EMBO J.* **2002**, *21* (11), 2757–2768.
(19) Benarroch, D.; Egloff, M. P.; Mulard, L.; Guerreiro, C.; Romette, J. L.; Canard, B. A structural basis for the inhibition of the NS5

dengue virus mRNA 2′-O-methyltransferase domain by ribavirin 5′-triphosphate. *J. Biol. Chem.* **2004**, *279* (34), 35638–35643.
(20) PDB entries 1L9K, 2P1D, 1R6A, 2P3L, 2P3Q, 2P3O, 2P40, and 2P41.
(21) Faumann, E.; Blumenthal, R.; Cheng, X. Structure and evolution of AdoMet-dependent methyltransferases. In *S-Adenosylmethionine-Dependent Methyltransferases: Structure and Functions*; Cheng, X., Blumenthal, R., Eds.; World Scientific Publishing: Singapore, 1999; pp 1−38.
(22) Hodel, A. E.; Gershon, P. D.; Quiocho, F. A. Structural basis for sequence-nonspecific recognition of 5′-capped mRNA by a cap-modifying enzyme. *Mol. Cell* **1998**, *1* (3), 443–447.
(23) Lim, S. P.; Wen, D.; Yap, T. L.; Yan, C. K.; Lescar, J.; Vasudevan, S. G. A scintillation proximity assay for dengue virus NS5 2′-O-methyltransferase: Kinetic and inhibition analyses. *Antiviral Res.* **2008**, *80*, 360–369.
(24) Milani, M.; Mastrangelo, E.; Bollati, M.; Selisko, B.; Decroly, E.; Bouvet, M.; Canard, B.; Bolognesi, M. Flaviviral methyltransferase/RNA interaction: Structural basis for enzyme inhibition. *Antiviral Res.* **2009**, *83* (1), 28–34.
(25) The universal protein resource (UniProt). *Nucleic Acids Res.* **2008**, *36* (Database issue), D190−D195.
(26) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **1990**, *215* (3), 403–410.
(27) Thompson, J. D.; Gibson, T. J.; Higgins, D. G. Multiple sequence alignment using ClustalW and ClustalX. *Current Protocols in Bioinformatics*; Wiley: New York, 2002; Chapter 2, Unit 2.3.
(28) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. UCSF Chimera: A visualization system for exploratory research and analysis. *J. Comput. Chem.* **2004**, *25* (13), 1605–1612.
(29) Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. *J. Mol. Graphics* **1996**, *14* (1), 33.
(30) Irwin, J. J.; Shoichet, B. K. ZINC: A free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2005**, *45* (1), 177–182.
(31) Jorgensen, W. L.; Maxwell, D. S.; TiradoRives, J. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. *J. Am. Chem. Soc.* **1996**, *118* (45), 11225–11236.
(32) Halgren, T. A. Merck molecular force field. 1. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* **1996**, *17* (5−6), 490–519.
(33) PDB entries 1R6A and 2P1D.
(34) Babaoglu, K.; Simeonov, A.; Irwin, J. J.; Nelson, M. E.; Feng, B.; Thomas, C. J.; Cancian, L.; Costi, M. P.; Maltby, D. A.; Jadhav, A.; Inglese, J.; Austin, C. P.; Shoichet, B. K. Comprehensive mechanistic analysis of hits from high-throughput and docking screens against β-lactamase. *J. Med. Chem.* **2008**, *51* (8), 2502–2511.
(35) Feng, B. Y.; Shelat, A.; Doman, T. N.; Guy, R. K.; Shoichet, B. K. High-throughput assays for promiscuous inhibitors. *Nat. Chem. Biol.* **2005**, *1* (3), 146–148.
(36) Feng, B. Y.; Simeonov, A.; Jadhav, A.; Babaoglu, K.; Inglese, J.; Shoichet, B. K.; Austin, C. P. A high-throughput screen for aggregation-based inhibition in a large compound library. *J. Med. Chem.* **2007**, *50* (10), 2385–2390.
(37) Seidler, J.; McGovern, S. L.; Doman, T. N.; Shoichet, B. K. Identification and prediction of promiscuous aggregating inhibitors among known drugs. *J. Med. Chem.* **2003**, *46* (21), 4477–4486.
(38) Breiman, L. Random forests. *Mach. Learn.* **2001**, *45* (1), 5–32.
(39) Feng, B. Y.; Shoichet, B. K. A detergent-based assay for the detection of promiscuous inhibitors. *Nat. Protoc.* **2006**, *1* (2), 550–553.
(40) Ryan, A. J.; Gray, N. M.; Lowe, P. N.; Chung, C. W. Effect of detergent on "promiscuous" inhibitors. *J. Med. Chem.* **2003**, *46* (16), 3448–3451.
(41) Finney, D. J. Radioligand assay. *Biometrics* **1976**, *32* (4), 721–740.
(42) Rodbard, D.; Hutt, D. M. *Statistical analysis of radioimmunoassays and immunoradiometric (labelled antibody) assays. A generalized weighted, iterative, least-squares method for logistic curve fitting*; 1974; pp 165−192.
(43) Wang, Q. Y.; Patel, S. J.; Vangrevelinghe, E.; Xu, H. Y.; Rao, R.; Jaber, D.; Schul, W.; Gu, F.; Heudi, O.; Ma, N. L.; Poh, M. K.; Phong, W. Y.; Keller, T. H.; Jacoby, E.; Vasudevan, S. G. A small-molecule dengue virus entry inhibitor. *Antimicrob. Agents Chemother.* **2009**, *53* (5), 1823–1831.
(44) Dunham, E. J.; Holmes, E. C. Inferring the timescale of dengue virus evolution under realistic models of DNA substitution. *J. Mol. Evol.* **2007**, *64* (6), 656–661.
(45) Twiddy, S. S.; Holmes, E. C.; Rambaut, A. Inferring the rate and time-scale of dengue virus evolution. *Mol. Biol. Evol.* **2003**, *20* (1), 122–129.
(46) Zhou, Y.; Ray, D.; Zhao, Y.; Dong, H.; Ren, S.; Li, Z.; Guo, Y.; Bernard, K. A.; Shi, P. Y.; Li, H. Structure and function of flavivirus NS5 methyltransferase. *J. Virol.* **2007**, *81* (8), 3891–3903.
(47) DTP: 2D and 3D Structural Information. http://dtp.nci.nih.gov/docs/3d_database/Structural_information/structural_data.html (accessed October 5, 2008).
(48) Copeland, R. A. *Evaluation of Enzyme Inhibitors in Drug Discovery: A Guide for Medicinal Chemists and Pharmacologists*; John Wiley & Sons: New York, 2005.
(49) Shoichet, B. K. Screening in a spirit haunted world. *Drug Discovery Today* **2006**, *11* (13−14), 607–615.
(50) Shoichet, B. K. Interpreting steep dose-response curves in early inhibitor discovery. *J. Med. Chem.* **2006**, *49* (25), 7274–7277.
(51) Zweygarth, E.; Schillinger, D.; Kaufmann, W.; Rottcher, D. Evaluation of sinefungin for the treatment of *Trypanosoma* (*Nannomonas*) *congolense* infections in goats. *Trop. Med. Parasitol.* **1986**, *37* (3), 255–257.
(52) Vedel, M.; Lawrence, F.; Robert-Gero, M.; Lederer, E. The antifungal antibiotic sinefungin as a very active inhibitor of methyltransferases and of the transformation of chick embryo fibroblasts by Rous sarcoma virus. *Biochem. Biophys. Res. Commun.* **1978**, *85* (1), 371–376.
(53) Yebra, M. J.; Sanchez, J.; Martin, C. G.; Hardisson, C.; Barbes, C. The effect of sinefungin and synthetic analogues on RNA and DNA methyltransferases from *Streptomyces*. *J. Antibiot.* **1991**, *44* (10), 1141–1147.
(54) Chang, M. W.; Lindstrom, W.; Olson, A. J.; Belew, R. K. Analysis of HIV wild-type and mutant structures via in silico docking against diverse ligand libraries. *J. Chem. Inf. Model.* **2007**, *47* (3), 1258–1262.
(55) Kasam, V.; Zimmermann, M.; Maass, A.; Schwichtenberg, H.; Wolf, A.; Jacq, N.; Breton, V.; Hofmann-Apitius, M. Design of new plasmepsin inhibitors: A virtual high throughput screening approach on the EGEE grid. *J. Chem. Inf. Model.* **2007**, *47* (5), 1818–1828.
(56) Zhang, W.; Du, X.; Ma, F.; Zhang, J.; Shen, J. DDGrid: Harness the Full Power of Supercomputing Systems. Fifth International Conference on Grid and Cooperative Computing Workshops (GCCW '06), **2006**.
(57) Sanner, M. F.; Olson, A. J.; Spehner, J. C. Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers* **1996**, *38* (3), 305–320.