Docking Ligands into Flexible and Solvated Macromolecules. 2. Development and Application of FITTED 1.5 to the Virtual Screening of Potential HCV Polymerase Inhibitors

Christopher R. Corbeil,[†] Pablo Englebienne,[†] Constantin G. Yannopoulos,[‡] Laval Chan,[‡] Sanjoy K. Das,[‡] Darius Bilimoria,[‡] Lucille L'Heureux,[‡] and Nicolas Moitessier*[,†]

Department of Chemistry, McGill University, 801 Sherbrooke St W, Montreal, QC, Canada H3A 2K6, and ViroChem Pharma Inc., 275 Armand-Frappier Blvd., Laval, QC, Canada H7V 4A7

Received November 1, 2007

HCV NS5B polymerase is a validated target for the treatment of hepatitis C, known to be one of the most challenging enzymes for docking programs. In order to improve the low accuracy of existing docking methods observed with this challenging enzyme, we have significantly modified and updated FITTED 1.0, a recently reported docking program, into FITTED 1.5. This enhanced version is now applicable to the virtual screening of compound libraries and includes new features such as filters and pharmacophore- or interaction-site-oriented docking. As a first validation, FITTED 1.5 was applied to the testing set previously developed for FITTED 1.0 and extended to include hepatitis C virus (HCV) polymerase inhibitors. This first validation showed an increased accuracy as well as an increase in speed. It also shows that the accuracy toward HCV polymerase is better than previously observed with other programs. Next, application of FITTED 1.5 to the virtual screening of the Maybridge library seeded with known HCV polymerase inhibitors revealed its ability to recover most of these actives in the top 5% of the hit list. As a third validation, further biological assays uncovered HCV polymerase inhibition for selected Maybridge compounds ranked in the top of the hit list.

INTRODUCTION

Docking-Based Virtual Screening. Various approaches to the design or identification of new drugs have recently been developed and successfully applied, including both experimental (e.g., SAR by NMR[1]) and computational approaches (e.g., docking[2,3]). In modern drug design, docking-based virtual screening (VS) methods provide a quick and cost-effective alternative to high-throughput screening (HTS).[2,3] Many recent VS applications have been reported and demonstrate an increasing level of accuracy for the currently available methods.[4,5] However, to date, only a few docking programs (e.g., FlexX-Ensemble[6] and AutoDock 4.0[7]) can take into account conformational changes that occur as a result of binding to a ligand. FITTED 1.0 (*Flexibility Induced Through Targeted Evolutionary Description*) is a docking program that was recently developed and validated against a set of cocrystallized protein/ligand complexes.[8,9] This program not only accurately predicts the ligand binding mode, but it also predicts the optimal protein conformation and the presence or absence and location of water molecules with a high level of accuracy. Recently, hepatitis C virus (HCV) polymerase has been found to be a very challenging protein for docking programs. Our long-term goal is to develop a docking-based virtual screening tool that can be applied to as large a number of proteins as possible. However, the initial version of FITTED (v. 1.0) was found to be too slow and not applicable to large VS studies. Efforts to increase the speed without affecting the accuracy were necessary. In addition, its accuracy for HCV polymerase inhibitors was low. We report herein an enhanced version of FITTED (v. 1.5), its validation, and its application to the discovery of HCV polymerase inhibitors.

Hepatitis C. HCV is the major causative agent responsible for non-A and non-B hepatitis, affecting over 170 million people worldwide. Chronic HCV infection often results in liver fibrosis, liver cirrhosis, hepatocellular carcinoma, and other forms of liver dysfunction.[10,11] Given the widespread impact of this disease, there is a substantial medical need for the discovery of new and effective anti-HCV agents to complement current therapies.[12] The impetus for the identification of agents that will be part of a potent and effective combination regimen is growing in view of the inevitability of the development of drug-resistant mutations. Extensive efforts have been devoted toward the study of the NS5B RNA-dependent RNA polymerase due to its critical function in the replication cycle of the virus. Positive results from several clinical trials have indeed validated the HCV NS5B polymerase as a target for the therapy of HCV infections. For example, nucleoside analogs (e.g., valopicitabine (NM283), **1**,[13,14] and R1626, **2**[15]) and non-nucleoside or allosteric (e.g., HCV-796, **3**[16]) inhibitors of HCV NS5B have been shown to be effective either alone or in combination with interferon (Figure 1). Others such as VCH-759[17,18] and GSK625433[19] are currently being evaluated in clinical trials.

HCV Polymerase, a Flexible Protein. We have recently shown that at least two major conformations can be adopted by the HCV polymerase upon the binding of inhibitors to an allosteric site located in the thumb region.[20,21] The main

* To whom correspondence should be addressed. Phone: 514-398-8543. Fax: 514-398-3797. E-mail: nicolas.moitessier@mcgill.ca.
[†] McGill University.
[‡] ViroChem Pharma Inc.

10.1021/ci700398h CCC: $40.75
Published on Web 03/15/2008

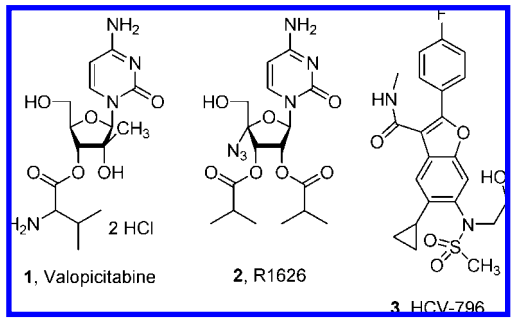

Figure 1. Selected HCV polymerase inhibitors.

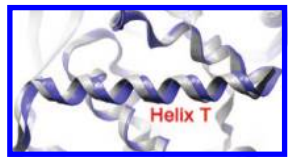

Figure 2. Helix T perturbation upon inhibitor binding. Blue and grey ribbon representations are from two different X-ray complexes.[20,21]

difference appeared to be a significant shift of the α helix T located in the binding site (Figure 2). It is therefore critical to account for this HCV polymerase thumb binding site plasticity in both hit identification and inhibitor design stages. In a recent comparative study by Warren et al., it was shown that none of the assessed docking programs predicted the experimentally observed binding modes of HCV polymerase inhibitors with high accuracy.[22] In fact, the flexibility of this protein can in part explain this poor accuracy. This challenge was the starting point of the development of FITTED 1.5.

THEORY AND IMPLEMENTATION

FITTED 1.0 and 1.5. As discussed above, two major aspects have to be considered for implementation into FITTED 1.5. First, the newer version should be applicable to VS studies. Second, FITTED 1.5 should be accurate enough with HCV polymerase inhibitors.

Flexible ligand/flexible protein docking programs have seldom been applied to VS.[23] In practice, taking flexibility into account significantly enlarges the search space, thus reducing throughput and drastically impeding implementation in VS campaigns. We hypothesized that evaluating FITTED in this context would assess the role of flexibility in VS studies against HCV polymerase. FITTED 1.0 is a suite of programs for docking that considers ligand and protein flexibility by means of a genetic algorithm[24] while water molecule displacement is accounted for by means of a specific potential energy function.[8,25] During the docking process, the protein side chains and backbone conformations, the water molecule positions, and the ligand torsion angles are coded as genes and optimized through a combined Lamarckian/Darwinian evolution. This early version of the program was developed to dock single compounds in proof-of-concept studies with no consideration for CPU time requirements.[8] This, obviously, is a serious limitation in the context of a large VS study. In order to optimize the software for efficiency and speed, a stepwise approach to identify and remove inappropriate candidates (poses) early in the process was implemented in FITTED 1.5. In addition, preliminary studies have shown that the accuracy of FITTED 1.0 with HCV

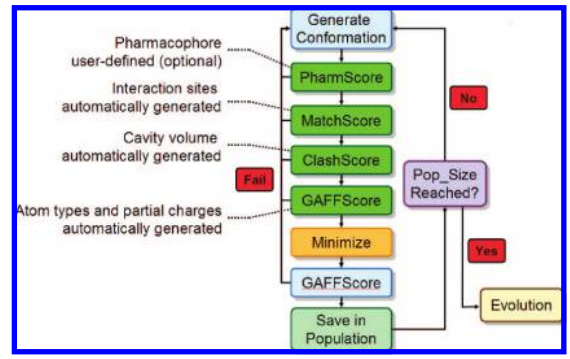

Figure 3. Consensus docking. Application to the generation of the initial population.

polymerase inhibitors should be improved. The various modifications and implementations, which required major rewriting of the program, are listed and described in the following sections.

SMART. SMART (*Small Molecule Atom typing and Rotatable Torsion Assignment*) is a module of FITTED used to prepare the ligands to be docked. In contrast to the original program developed with FITTED 1.0, the current version now describes the compound features with a bit string added to the compound's mol 2 file. The bit string includes the following descriptors: molecular weight, number of rotatable bonds, net charge, and the presence of functional groups such as known toxicophores or reactive groups (e.g, nitro groups, aldehydes, and Michael acceptors) or labile imines. The descriptions are then used by FITTED to filter out compounds not fulfilling Lipinski's rules[26] or having undesired (user-defined) functionalities. These descriptors can also restrict the search to compounds with needed functionalities (e.g., aldehydes and nitriles for reversible covalent inhibitors). Although this simple approach is not expected to accurately discriminate between druglike molecules and non-druglike molecules, it will orient our study toward a "cleaner" compound set.

Interaction Site Filter. FITTED 1.0 included a functionality that filtered out poses that did not fulfill constraints imposed by the user (e.g, binding to metals). It also included a function that ensured that poses were within the binding site (ClashScore) prior to any further optimization or more complex scoring. ClashScore, which is a binary score, uses a series of spheres representing the accessible cavity space. Each pose is then compared to this set of spheres, and a score ("in" or "out") is computed. This crude score is used to discard poses that are not located within the binding site. After modifications, FITTED 1.5 can preselect the poses that are the most apt to be successfully docked with a number of predefined interaction sites (Figure 3). A set of interaction sites is similar to a pharmacophore but automatically generated by PROCESS from the protein structure alone. FITTED also allows for the use of a manually created pharmacophore which may exploit user expertise, such as the one shown in Figure 4, or for the use of both automatically generated interaction sites and user-defined pharmacophores (Figure 3). The inclusion of a pharmacophore component in virtual screening has been shown to enhance efficiency and accuracy of docking methods in previous studies,[27,28] including HCV polymerase.[22]

The interaction sites (and pharmacophores) are represented by a series of spheres. A sphere diameter defines the allowed

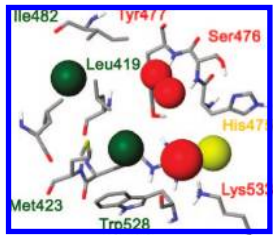

Figure 4. Binding site pharmacophore for HCV polymerase: red, hydrogen-bond acceptor; green, hydrophobic/aromatic; yellow, either hydrogen-bond acceptor or hydrophobic/aromatic.

volume of the constraint and a weight (w in eq 1) other than 1 can be assigned to each sphere.

$$W_{\text{sphere}} = 100 \times w / \sum w \quad (1)$$

$$\text{MatchScore}_{\text{pose}} = \sum W_{\text{matchedspheres}} \quad (2)$$

Each generated pose is compared to the interaction site (and/or pharmacophore), and a MatchScore (and/or PharmScore) (eq 2) ranging from 0 to 100% is computed. If the atom types of the ligand atoms lying within the volume of the sphere match the interaction site/pharmacophore sphere's pharmacophoric properties, the weight of the sphere is added to the MatchScore (or PharmScore) for that pose. FITTED 1.5 then discards poses with a low MatchScore (and/or PharmScore), thereby reducing the required CPU time by directing the docking toward strongly interacting poses. It is well-known that higher success rates are obtained when rescoring of the poses is performed using other scoring functions (consensus scoring). In the present version of FITTED, up to four scores are computed while docking (ClashScore mentioned above, MatchScore, PharmScore, and GAFFScore derived from the computed General AMBER Force Field[29] (GAFF) potential energy) and can be combined to discriminate active from inactive compounds (Figure 3). These four scores are used in decreasing order of speed and allow FITTED to eliminate poses exhibiting bad scoring with one function before proceeding with the next one. This filtering of poses is carried out both during the generation of the initial population, as in Figure 3, and also during the evolution, and it can be viewed as consensus docking.[30] This feature significantly reduces the time required to dock a single compound and increases its accuracy in three ways. First, the MatchScore and PharmScore are quicker to compute than the GAFF potential energy. Second, "bad" poses are not considered for energy minimization, a time-consuming step in the docking process. Third, poses with reasonable GAFFScores but poor chemical complementarity with the protein were found using FITTED 1.0. Conversely, FITTED 1.5 assigns a low MatchScore to these poses, thus reducing the number of false positives.

PROCESS. PROCESS (*Protein Conformation Ensemble System Setup*) is the second module of FITTED used to prepare the protein file. As described in our previous report,[8] PROCESS also prepares the set of spheres representing the cavity space used to compute ClashScore (see above). The current version of PROCESS can now derive a set of interaction sites such as ideal locations for hydrogen-bond donors and acceptors and hydrophobic and aromatic groups.

Quick Docking. When docking a potential ligand, FITTED generates an initial population and then simulates its evolution. Although this is appropriate for "good" binders, it may be inappropriate for compounds which are, for instance, too large or too hydrophobic and therefore should be excluded prior to this time-consuming conformational search. For this purpose, additional filters were implemented to prevent undesirable compounds from being docked. First, compounds lacking the required pharmacophoric groups are excluded. Then, FITTED generates a maximum number of random poses to produce the initial population. If 100 000 possible binding modes are generated without accepting one into the initial population (low MatchScore and/or not in the cavity), the program aborts and docks the next compound on the list. This stage, based on simple shape discrimination, does not require any CPU-intensive energy or score computation and can be done within a few seconds per compound.

Refined Docking. A close look at the evolution of the score of the top pose as the docking proceeds revealed that the scores computed after a few generations are typically within 1.0−1.5 kcal/mol of the score of the final pose. It also indicated that, at this stage of the evolution, poses close to the native pose are not always identified. These two observations demonstrate (1) the high quality of the initial population and therefore the identification of poses with good scores early in the evolution and (2) the need for a multigeneration evolution process to produce the correct (i.e., experimentally observed) pose. Thus, if after a few generations (e.g., five) the score is not satisfactory, the docking can be aborted. We have therefore implemented new functions (and keywords) into FITTED to account for this intermediate evaluation. In practice, more than one of these intermediate evaluations can be used to further reduce the number of generations carried out with a potentially inactive compound.

Scoring. The force field used in FITTED 1.0 (AMBER84) was not appropriate for most of the small druglike molecules like the ones found in virtual libraries. Instead, we have used the GAFF[29] for the description of the small molecules. This required a series of modifications to the force field itself, as a specific format has to be used to be readable by FITTED. SMART was also modified to assign these new atom types. Finally, a simple automated parameter estimator was developed. Although GAFF parameters span a large variety of functional groups, some are missing but could be guessed on-the-fly by FITTED 1.5. These parameters are simply derived from the input structures; bond lengths and angles of the input structures are used as equilibrium values and reported in a specific log file. These listed missing parameters can later be further optimized and added to the force field for the next study or to perform a second run on these specific molecules. In fact, our own version of GAFF is regularly updated to include more functional groups and heteroaromatic rings.

Whereas ClashScore, PharmScore, MatchScore, and GAFFScore are used upon docking to identify the correct pose, the RankScore scoring function reported previously[24] is used to assign the final poses a score describing their binding affinities.

VALIDATION OF FITTED 1.5

FITTED 1.0 versus 1.5. As all of these changes may affect the accuracy of FITTED, we used the testing set previously

Table 1. Comparison of FITTED 1.0 with FITTED 1.5

	% success[a]			
	version 1.0		version 1.5	
mode	<1.0 Å	<2.0 Å	<1.0 Å	<2.0 Å
Rigid (self-docking)	33	79	84	93
Rigid (cross-docking)	21	47	51	74
Semiflexible	36	73	78	84
Flexible	57	73	78	88

[a] Two criteria of success are shown. A docking run is considered successful if the rmsd between the modeled and experimental binding modes is within 1.0 or 2.0 Å, respectively.

prepared for FITTED 1.0 to evaluate the accuracy of the current version. This set consists of ligands complexed with HIV-1 protease, thymidine kinase, factor Xa, trypsin, and MMP-3. Table 1 summarizes the accuracy obtained for the self-docking of these 33 inhibitors ("Rigid" mode) as well as their docking to flexible proteins. The "Semiflexible" mode, as defined in our previous report,[8] corresponds to the docking of ligands to conformational ensembles of protein structures, while the "Flexible" mode corresponds to a fully flexible protein structure.[8] The detailed results for each of the 33 systems are given as Supporting Information. The computed root-mean-square deviations (RMSDs) compare the modeled "docked" binding mode to the observed one (from crystal structures). For this study, the interactions sites were generated by PROCESS.

Overall, accuracy is significantly increased from version 1.0 to 1.5. More specifically, there is an enhanced accuracy when examining systems for which the RMSDs are below 1.0 Å. This observation is most likely due to the use of interaction sites to guide the docking. With this implementation, FITTED 1.5 generates and considers only poses that already passed the MatchScore and/or PharmScore filter. Thus, the quality of the initial population as well as that of the children offspring produced during the evolution of the population are of higher quality than those with FITTED 1.0. Along with the increased accuracy, a 3-fold increase in speed was observed. In addition, as observed previously with FITTED 1.0, the docking to flexible proteins is nearly as accurate as self-docking and much more accurate than cross-docking.

With these encouraging results in hand, we focused our attention to HCV polymerase inhibitors. For this purpose, two sets of protein/inhibitor complexes were initially built and very recently extended as novel crystal structures were reported and made available. The first set includes 16 inhibitors bound to the allosteric site described above, while the second set includes seven inhibitors bound to the catalytic site. A second allosteric site has been reported but is not used herein.[31] We first used the interaction sites as for the testing set above. The results summarized in Table 2 show that the accuracy obtained when docking to the allosteric site of the HCV polymerase was not as high as for the testing set used above, but they were nevertheless considered reasonable. In order to increase the accuracy of this program and to eventually increase the enrichment factor of the VS study, we manually defined a pharmacophore used in place of the interaction sites by FITTED (Figure 3). In this case, the accuracy was slightly increased, while the required CPU time was not affected. In this study, we used the empty space found within a sphere of 40 Å designated as large cavity in Table 2. A more focused binding site (25 Å) led to a significant increase in accuracy when self-docking was considered but only a slight increase in accuracy when a flexible protein was used. Interestingly, the use of flexible protein was found to be significantly more accurate than cross-docking, indicating that its implementation should increase the accuracy of FITTED in VS studies against HCV polymerase.

We then turned our attention to the catalytic site; however, docking of the seven reported inhibitors was initially unsuccessful (Table 3). Considering the size of this very large binding site, this disappointing result is not surprising. While this work was ongoing, Warren et al. reported a large comparative study including 13 HCV polymerase inhibitors.[22] In their study, only two programs docked 1 out of the 13 inhibitors to the catalytic site with a RMSD below 2.0 Å, while the other eight programs failed with all of the inhibitors. Although their set (not given) and ours may be different, there is at least one HCV polymerase inhibitor common to both sets.

Warren et al. also mentioned that "no docking program was able to generate docked poses within 2 Å for ≥40% of the compounds" when only the NTP site is considered. Unfortunately, no details were provided. In order to orient the docking toward this binding site, we used PROCESS to automatically generate interaction sites and spheres representing the binding site cavity centered on this site. Much to our delight, FITTED was found to dock five out of the seven inhibitors with RMSDs below 1.2 Å in self-docking experiments and the same five with RMSDs below 1.5 Å when the semiflexible mode was selected. These results therefore position our program among the top of the list of assessed programs for this HCV polymerase site. We believe that the good accuracy observed with FITTED is due to the consensus docking approach implemented in FITTED 1.5, which is expected to accurately filter out unreasonable poses. We also found that introducing the protein flexibility led to a significant increase in accuracy relatively to cross-docking.

APPLICATION TO THE SCREENING OF A LIBRARY AGAINST THE HCV POLYMERASE

The previous validation demonstrated that the current version of FITTED docked inhibitors with reasonable accuracy and also demonstrated the key role of the protein flexibility accounted for by FITTED in the HCV polymerase context. However, this validation did not provide any indication about its ability to identify active compounds within a large set (i.e., to rank known inhibitors at the top of the hit list). For this purpose, we selected first the well-studied thiophene site from which we have collected much data from both SAR and X-ray crystallography. Another two sites (catalytic and allosteric) are also validated targets but were not considered here. The Maybridge library of druglike molecules, which was obtained from the ZINC Web site,[32,33] was seeded with known actives ranging from nanomolar to micromolar activities.

To account for the site flexibility, we used two inhibitor-bound in-house crystal structures in a "semiflexible" docking run, an option implemented in FITTED that allows the simultaneous docking of a flexible ligand to more than one

Table 2. Docking of HCV Polymerase Inhibitors to the Allosteric Site with FITTED 1.5

	% success[a]					
mode	interaction sites, large cavity[b]		pharmacophore, large cavity[b]		pharmacophore, focused cavity[c]	
accuracy	<1.0 Å	<2.0 Å	<1.0 Å	<2.0 Å	<1.0 Å	<2.0 Å
Rigid (self-docking)	56	75	63	75	81	81
Rigid (cross-docking)	0	0	13	31	6	38
Semiflexible	38	50	56	69	63	69
Flexible	25	31	47	59	38	63

[a] Two criteria of success are shown. A docking run is considered successful if the rmsd between modeled and experimental binding modes is within 1.0 or 2.0 Å, respectively. [b] 40 Å diameter cavity. [c] 25 Å diameter cavity.

Table 3. Docking of HCV Polymerase Inhibitors to the Catalytic Site

	% success[a]			
	whole cavity		NTP site	
mode	<1.0 Å	<2.0 Å	<1.0 Å	<2.0 Å
Rigid (self-docking)	0	0	57	71
Rigid (cross-docking)	0	0	0	14
SemiFlexible	0	0	43	57
Flexible	0	0	29	43
Flo, Gold, Glide, DockIt, MVP, LigFit, Dock4, FlexX, Fred, MOE		0−8[b]		0−60[b]

[a] Two criteria of success are shown. A docking run is considered successful if the rmsd between modeled and experimental binding mode is within 1.0 or 2.0 Å respectively. [b] Data for self-docking (corresponding to Rigid mode) from Warren et al.[22]

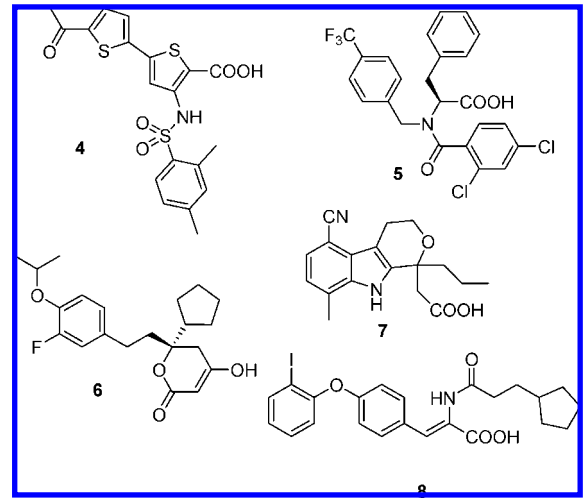

Figure 5. Selected known actives.[35–40]

protein structure. The pharmacophore shown in Figure 4 includes six spheres identifying two hydrophobic pockets (shown in green), three sites for hydrogen-bond acceptors (HBA, shown in red), and a mixed hydrophobic/HBA site (shown in yellow). At this last location, a phenyl ring may interact with His475 via aromatic ring stacking and/or π-cation interaction with Lys533. Experimental data showed that hydrogen bonds with Tyr477 and Ser476 are key interactions and that two out of the three defined hydrophobic pockets are often targeted.[20] A compound that binds without filling the deep hydrophobic pocket delineated by Met423, Trp528, and Leu419 would trap water molecules, a phenomenon that is highly disfavored. Similarly, desolvation of the other hydrophobic pocket defined by Leu419, Ile482, and Leu497 is favored upon binding. To account for these specific situations, the spheres representing these hydrophobic pockets are given a larger priority ($w = 2$) than the other four ($w = 1$), and poses with a MatchScore lower than 60% are discarded.

Stepwise Screening. As described above, the fully automated FITTED 1.5 screening protocol can be broken down into five distinct steps (Figure 5): (1) filtering out nondruglike compounds, (2) filtering out compounds that cannot match the binding site pharmacophore and/or the binding site cavity, (3) quick docking and discarding compounds with RankScore values higher than −5.25, (4) refined docking of the best candidates, and (5) selection of the best scoring compounds.

This stepwise approach was applied to the Maybridge set seeded with 23 known active inhibitors including very weak inhibitors. A large variety of known ligands were selected, and some are illustrated in Figure 5. As an additional test for FITTED, one of these HCV polymerase inhibitors (compound **7**) has been reported to have a high anti-HCV activity, with the (*R*) enantiomer being the most active.[34,35] We therefore spiked our set with the two enantiomeric forms of **7** in order to assess their binding to the allosteric site and their relative predicted binding affinity. In contrast to closely related analogues which bind to the catalytic site, compound **8** is believed to bind to the allosteric site and was added to this set.[36]

The protocol is shown in Figure 6. The entire library was processed using SMART, and a first filtration step was carried out using FITTED 1.5. Compounds with net charges between −1 and +1, a number of hydrogen-bond acceptors lower or equal to 10, a number of hydrogen-bond donors lower or equal to five, a maximum of six rotatable bonds, and molecular weights below or equal to 550 and containing no potentially toxic, reactive, or hydrolyzable groups were retained. This set was comprised of nearly 32 500 compounds containing 19 known active inhibitors (0.058% of the library).

Prior to the actual docking, compounds not featuring the necessary pharmacophoric groups, as well as compounds that could not fit the binding site cavity and/or had a MatchScore lower than 60%, were discarded. None of the known active compounds failed this test, while a further 6% of the filtered library was eliminated. When appropriate poses were found, FITTED started the genetic algorithm optimization and produced the initial population. A quick evolution (five generations) was then applied to this population, and the top three poses (best GAFFScore) were scored using RankScore. Compounds with a RankScore of −5.25 or lower were allowed to progress to the next step. An additional 52% of the filtered library was removed at this stage, while all seed compounds were retained. This observation provides a clear

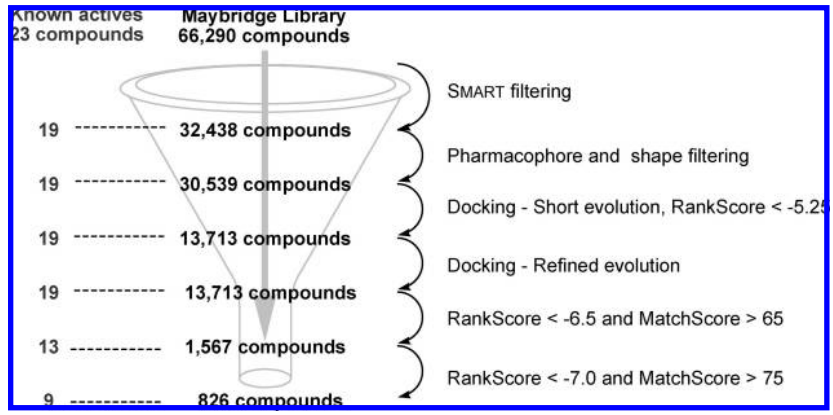

Figure 6. Funnel approach implemented in FITTED.

Table 4. Focused Libraries Based on MatchScore >75 and RankScore as Indicated

RankScore cutoff	hits[a]	known actives	enrichment factor[b]
<−7.0	835	9	18.4
<−7.5	401	8	34.1
<−8.0	147	6	69.7
<−8.5	48	6	214
<−9.0	14	3	366
filtered library	32457	19	1.0

[a] Including known actives. [b] Based on the filtered library.

indication of the usefulness of this intermediate selection. The remaining 13 713 compounds were further optimized (another 95 generations of evolution) and scored using RankScore. Finally, focused libraries of different sizes were compiled on the basis of different score cutoffs. Table 4 summarizes the size, number of recovered known actives, and enrichment factors for these small size libraries.

Enrichment Factors. Performing docking-based virtual screening, tools should prioritize active compounds from a library of druglike molecules. It is common practice to seed a library of druglike molecules with known actives and use the enrichment factor obtained to evaluate the accuracy of the docking and scoring functions of the software. An accurate program should be able to recover the seed compounds at the top of the score-ranked hit list. Comparative studies evaluating the accuracy of docking programs to extract active compounds from large libraries showed that Surflex, GOLD, Glide, and FlexX are among the best programs.[3] For instance, the best performer in Rognan's study, Surflex, was able to rank 10 known thymidine kinase inhibitors from a library of 1000 druglike molecules in the top 10% of the library, with five in the top 3%.[41] In another study, slightly less than 50% of the seed inhibitors were ranked in the top 10% by DOCK, GOLD, and Glide, with 38% in the top 2% when GOLD was used.[42] Overall, a state-of-the-art VS tool rarely extracts 100% of the actives in the top 10% and even more rarely in the top 5%; 50–60% in the top 5% is more commonly observed with the best programs. In contrast, from the focused sets selected by FITTED 1.5 (Table 1), large enrichment factors were computed. Considering that in the initial library and in the filtered library only 0.035% and 0.058%, respectively, were known actives, enrichment factors of over 300 for the top 14 compounds were achieved for this target with FITTED 1.5.

Data analysis illustrated in Figure 7 indicated that a third of the actives were recovered in the top 0.1% and that half

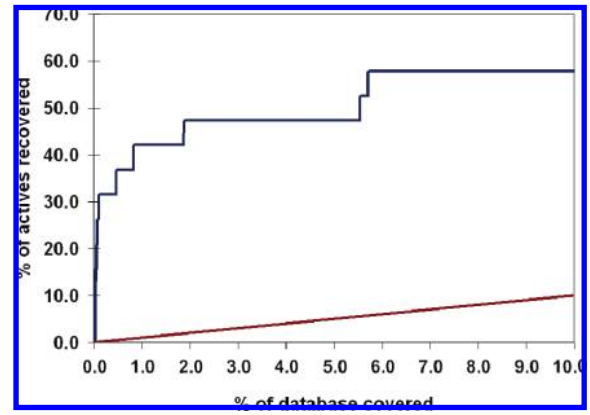

Figure 7. Active compounds recovered. Blue curve, FITTED VS study; orange, random selection.

of the actives were found in the top 2% of the hit list. An average of 12 min of CPU time per compound was needed to dock each of the 32 500 filtered compounds using the semiflexible HCV polymerase structure on desktop Linux PCs (AMD Opteron), while less than 1 s per compound was needed to filter out the bad candidates.

Interestingly, (*R*)-**7**, the most active enantiomer of **7**, was predicted to bind well to the allosteric site, while (*S*)-**7** was given a score much worse than the other 22 actives. This result correlates well with the experimental data and indicates that **7** may bind to the allosteric site. Compound **8** was also assigned a high score and is predicted to bind tightly to the allosteric site, an observation that has only been postulated.[36]

Biological Evaluation. Encouraged by the large enrichment factors obtained for this target, we assessed FITTED's ability to identify new HCV polymerase inhibitors from the Maybridge library by screening the high-ranking compounds in biochemical assays. The top-scoring compounds with a RankScore below −7.0 and a MatchScore higher than 75 were considered for biological evaluation (826 compounds; 1.25% of the Maybridge library) using a scintillation proximity assay (SPA) described in the Experimental Section. Unfortunately, some high-scoring compounds were not available for purchase at the time, and only 659, representing 1% of the total Maybridge library, were acquired. All of these 659 available compounds were screened against the HCV polymerase using a single point concentration and resulted in 220 compounds showing greater than 50% inhibition at 10 μg/mL and 12 compounds which had greater than 90% inhibition at 10 μg/mL. The set of 12 actives were retested

in an 11-point dose response SPA assay, and two druglike compounds were identified to inhibit HCV polymerase with IC_{50} values of 7 and 12 μM.

With these newly discovered actives, a new enrichment factor of 20.4 was computed for the top 835 hits.

CONCLUSION

HCV has been shown to be a challenging enzyme for docking methods and prompted us to assess FITTED in this context. Hence, FITTED 1.0 has been modified to incorporate features for its application to docking and virtual screening, such as ligand- and pharmacophore-based prefiltering. This current version, namely, FITTED 1.5, showed significantly enhanced accuracy and speed relative to the previous version. Validation experiments carried out on two binding sites on HCV polymerase (allosteric and catalytic site) further confirmed its accuracy. We next looked at its ability to identify active HCV polymerase inhibitors from a set of druglike molecules. A virtual screening run on the Maybridge library seeded with known actives gave enrichment factors which were superior to the ones often observed with other available docking programs. Top-scoring compounds representing around 1% of the Maybridge library were purchased and screened in HCV polymerase assays, resulting in the identification of two compounds with IC_{50}'s of 7 and 12 μM. The screening of larger libraries is now ongoing.

FITTED 1.5 and the subsequent versions are now available to the scientific community.[43]

EXPERIMENTAL SECTION

Running FITTED 1.0 Testing Set with FITTED 1.5. The preparation of the testing set has been previously reported and will not be described herein.[8] All protein, interaction site and cavity files were then prepared using PROCESS 1.5 and all ligand files with SMART 1.5. The HCV polymerase/inhibitor complexes were prepared following the same protocol. In order to proceed in the semiflexible and flexible modes, FITTED requires identical sequences (and number of atoms) for the protein structures used as input. However, a large number of differences in the sequence of the various crystal structures have been found. As they were far enough from the binding sites, they are not expected to affect the docking accuracy. Thus, manual mutations were carried out to correct these discrepancies.

Preparation of the Cavity and Pharmacophore Files for VS. Two crystal structures (1NHV and 2GIR) representative of the set of 16 were used for the VS study. Preparation of the protein files was carried out as previously described. PROCESS was then used to prepare the structures for the VS. The center of the active site was defined by the centroid of the ligands present in the crystal structures. A sphere radius of 25 Å was used to generate the binding site cavity file. The pharmacophore was generated manually by examining the known binding modes and the interaction sites identified by PROCESS and extrapolating the six key interactions shown in Figure 3.

Preparation of the Library. The Maybridge library was downloaded from the ZINC database[32] in a mol2 format. Each compound of the library was then prepared by SMART, which added the rotatable bonds and atom types and completed the bit string for each compound.

Docking a Library with FITTED. Each compound of the library was docked individually using FITTED in Semiflexible mode. Compounds containing the following groups where filtered out and were not docked: aldehydes, esters, imines, nitro, acyclic Michael acceptors, azides, isocyanates, and acyl chlorides. As an additional constraint, all compounds were required to have at least one aromatic ring. The screening was carried out on the 872-node Dell PowerEdge cluster of a 3.2 GHz Intel Pentium 4 located at the Réseau Québecois de Calcul de Haute Performance (RQCHP) at the Université de Sherbrooke.

Biological Evaluation of the Selected Compounds. Briefly, 250 ng of a 5′-biotinylated DNA oligonucleotide (oligo dT15) primer, annealed to 10 pmol of a homopolymeric poly rA RNA template, was captured on the surface of streptavidin-coated beads (GE Healthcare, Uppsala, Sweden). The polymerization activity of a 50 nM HCV NS5B enzyme (genotype 1b, BK strain) was quantified by measuring the incorporation of a radiolabeled [^{3}H]-UTP substrate onto the 3′ end of the growing primer at 22 °C for 140 min. Detection was performed by counting the signal using a liquid scintillation counter (Wallac MicroBeta Trilux, Perkin-Elmer, MA). Compounds were initially tested using a single point concentration, and the actives were reconfirmed by 11 point dose responses. Curves were fitted to data points using nonlinear regression analysis, and IC_{50}'s were interpolated from the resulting curves using GraphPad Prism software, version 2.0 (Graphpad Software Inc., San Diego, CA).

ACKNOWLEDGMENT

We thank the Canadian Foundation for Innovation for financial support through the New Opportunities Fund program. C.R.C. holds a CIHR-funded Chemical Biology Scholarship, and P.E. holds a McGill Majors Fellowship (J. W. McConnell Memorial). We also thank CIHR (operating grant), FQRNT (Nouveaux chercheurs), and NSERC for funding and RQCHP for generous allocation of computer resources.

Supporting Information Available: Detailed data on the docking to the validation set. This information is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Shuker, S. B.; Hajduk, P. J.; Meadows, R. P.; Fesik, S. W. Discovering High-Affinity Ligands for Proteins: SAR by NMR. *Science* **1996**, *274*, 1531–1534.

(2) Cavasotto, C. N.; Orry, A. J. W. Ligand Docking and Structure-Based Virtual Screening in Drug Discovery. *Curr. Top. Med. Chem.* **2007**, *7*, 1006–1014.

(3) Moitessier, N.; Englebienne, P.; Lee, D.; Lawandi, J.; Corbeil, C. R. Towards the Development of Universal, Fast and Highly Accurate Docking/Scoring Methods: A Long Way to Go. *Br. J. Pharmacol.* **2007**, in press.

(4) Cozza, G.; Bonvini, P.; Zorzi, E.; Poletto, G.; Pagano, M. A.; Sarno, S.; Donella-Deana, A.; Zagotto, G.; Rosolen, A.; Pinna, L. A.; Meggio, F.; Moro, S. Identification of Ellagic Acid as Potent Inhibitor of Protein Kinase CK2: A Successful Example of a Virtual Screening Application. *J. Med. Chem.* **2006**, *49*, 2363–2366.

(5) De Graaf, C.; Oostenbrink, C.; Keizers, P. H. J.; Van Der Wijst, T.; Jongejan, A.; Vermeulen, N. P. E. Catalytic Site Prediction and Virtual Screening of Cytochrome P450 2D6 Substrates by Consideration of Water and Rescoring in Automated Docking. *J. Med. Chem.* **2006**, *49*, 2417–2430.

(6) Claussen, H.; Buning, C.; Rarey, M.; Lengauer, T. FLEXE: Efficient Molecular Docking Considering Protein Structure Variations. *J. Mol. Biol.* **2001**, *308*, 377–395.
(7) *AutoDock*, 4.0; The Scripps Research Institute: La Jolla, CA, 2006.
(8) Corbeil, C. R.; Englebienne, P.; Moitessier, N. Docking Ligands into Flexible and Solvated Macromolecules. 1. Development and Validation of FITTED 1.0. *J. Chem. Inf. Model.* **2007**, *47*, 435–449.
(9) Englebienne, P.; Fiaux, H.; Kuntz, D. A.; Corbeil, C. R.; Gerber-Lemaire, S.; Rose, D. R.; Moitessier, N. Evaluation of Docking Programs for Predicting Binding of Golgi alpha-Mannosidase II Inhibitors: A Comparison with Crystallography. *Proteins: Struct., Funct., Bioinf.* **2007**, *69*, 160–176.
(10) Bacon, B. R.; Di Bisceglie, A. M.; Korb, J. R.; Tillmann, H. L.; Herold, K. C.; Himelhoch, S.; De Knegt, R. J.; Van Den Berg, A. P.; Bell, B. P.; Walker, B. D.; Lauer, G. M. Hepatitis C Virus Infection [3] (multiple letters). *N. Engl. J. Med.* **2001**, *345*, 1425–1428.
(11) Lauer, G. M.; Walker, B. D. Hepatitis C Virus Infection. *N. Engl. J. Med.* **2001**, *345*, 41–52.
(12) Strader, D. B.; Wright, T.; Thomas, D. L.; Seeff, L. B. Diagnosis, Management, and Treatment of Hepatitis C. *Hepatology* **2004**, *39*, 1147–1171.
(13) Toniutto, P.; Fabris, C.; Bitetto, D.; Fornasiere, E.; Rapetti, R.; Pirisi, M. Valopicitabine Dihydrochloride: A Specific Polymerase Inhibitor of Hepatitis C Virus. *Curr. Opin. Invest. Drugs* **2007**, *8*, 150–158.
(14) Pierra, C.; Amador, A.; Benzaria, S.; Cretton-Scott, E.; D'Amours, M.; Mao, J.; Mathieu, S.; Moussa, A.; Bridges, E. G.; Standring, D. N.; Sommadossi, J. P.; Storer, R.; Gosselin, G. Synthesis and Pharmacokinetics of Valopicitabine (NM283), an Efficient Prodrug of the Potent Anti-HCV Agent 2′-*C*-Methylcytidine. *J. Med. Chem.* **2006**, *49*, 6614–6620.
(15) Smith, D. B.; Martin, J. A.; Swallow, S.; Smith, M. K. A.; Yee, C.; Crowell, M.; Kim, W.; Sarma, K.; Najera, I.; Jiang, W.-R.; Le Pogam, S.; Rajyaguru, S.; Klumpp, K.; Leveque, V.; Ma, H.; Tu, Y.; Chan, R.; Brandl, M.; Alfredson, T.; Wu, X.; Birudaraj, R.; Tran, T.; Cammack, N. From R1479 to R1626: Optimization of a Nucleoside Inhibitor of NS5B for the Treatment of Hepatitis C. In Abstracts of Papers, 232nd ACS National Meeting, San Francisco, CA, Sept. 10–14, 2007.
(16) ViroPharma Incorporated. Therapeutic focus: HCV 796. http://www.viropharma.com/therapeutic/hcv796.asp (accessed Jan 03, 2008).
(17) Virochem Pharma, Corporate Information. http://www.virochempharma.com/ourFocus.html (accessed Jan 03, 2008).
(18) Virochem Pharma, Corporate Information. http://clinicaltrials.gov/ct/gui/show/NCT00389298?order=6 (accessed Jan 03, 2008).
(19) Haigh, D.; Amphlett, E. M.; Bravi, G. S.; Bright, H.; Chung, V.; Chambers, C. L.; Cheasty, A. G.; Convery, M. A.; Ellis, M. R.; Fenwick, R.; Gray, D. F.; Hartley, C. D.; Howes, P. D.; Jarvest, R. L.; Medhurst, K. J.; Mehbob, A.; Mesogiti, D.; Mirzai, F.; Nerozzi, F.; Parry, N. R.; Roughley, N.; Skarzynski, T.; Slater, M. J.; Smith, S. A.; Stocker, R.; Theobald, C. J.; Thomas, P. J.; Thommes, P. A.; Thorpe, J. H.; Wilkinson, C. S.; Williams, E. Identification of GSK625433: A Novel Clinical Candidate for the Treatment of Hepatitis C. In Abstracts of Papers, 233rd ACS National Meeting, Chicago, IL, March 25–29, 2007.
(20) Biswal, B. K.; Cherney, M. M.; Wang, M.; Chan, L.; Yannopoulos, C. G.; Bilimoria, D.; Nicolas, O.; Bedard, J.; James, M. N. G. Crystal Structures of the RNA-dependent RNA Polymerase Genotype 2a of Hepatitis C Virus Reveal two Conformations and Suggest Mechanisms of Inhibition by Non-nucleoside Inhibitors. *J. Biol. Chem.* **2005**, *280*, 18202–18210.
(21) Biswal, B. K.; Wang, M.; Cherney, M. M.; Chan, L.; Yannopoulos, C. G.; Bilimoria, D.; Bedard, J.; James, M. N. G. Non-nucleoside Inhibitors Binding to Hepatitis C Virus NS5B Polymerase Reveal a Novel Mechanism of Inhibition. *J. Mol. Biol.* **2006**, *361*, 33–45.
(22) Warren, G. L.; Andrews, C. W.; Capelli, A. M.; Clarke, B.; LaLonde, J.; Lambert, M. H.; Lindvall, M.; Nevins, N.; Semus, S. F.; Senger, S.; Tedesco, G.; Wall, I. D.; Woolven, J. M.; Peishoff, C. E.; Head, M. S. A Critical Assessment of Docking Programs and Scoring Functions. *J. Med. Chem.* **2006**, *49*, 5912–5931.
(23) Kim, J.; Park, J. G.; Chong, Y. FlexE Ensemble Docking Approach to Virtual Screening for CDK2 Inhibitors. *Mol. Simul.* **2007**, *33*, 667–676.
(24) Moitessier, N.; Therrien, E.; Hanessian, S. A method for Induced-fit Docking, Scoring, and Ranking of Flexible Ligands. Application to Peptidic and Pseudopeptidic β-secretase (BACE 1) Inhibitors. *J. Med. Chem.* **2006**, *49*, 5885–5894.
(25) Moitessier, N.; Westhof, E.; Hanessian, S. Docking of Aminoglycosides to Hydrated and Flexible RNA. *J. Med. Chem.* **2006**, *49*, 1023–1033.
(26) Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Delivery Rev.* **1997**, *23*, 3–25.
(27) Moitessier, N.; Henry, C.; Maigret, B.; Chapleur, Y. Combining Pharmacophore Search, Automated Docking, and Molecular Dynamics Simulations as a Novel Strategy for Flexible Docking. Proof of Concept: Docking of Arginine-glycine-aspartic Acid-like Compounds into the $\alpha_v\beta_3$ Binding Site. *J. Med. Chem.* **2004**, *47*, 4178–4187.
(28) Goto, J.; Kataoka, R.; Hirayama, N. Ph4Dock: Pharmacophore-Based Protein-Ligand Docking. *J. Med. Chem.* **2004**, *47*, 6804–6811.
(29) Wang, J.; Wolf, R. M.; Caldwell, J. W.; Kollman, P. A.; Case, D. A. Development and Testing of a General Amber Force Field. *J. Comput. Chem.* **2004**, *25*, 1157–1174.
(30) Paul, N.; Rognan, D. ConsDock: A New Program for the Consensus Analysis of Protein-Ligand Interactions. *Proteins: Struct., Funct., Genet.* **2002**, *47*, 521–533.
(31) Di Marco, S.; Volpari, C.; Tomei, L.; Altamura, S.; Harper, S.; Narjes, F.; Koch, U.; Rowley, M.; De Francesco, R.; Migliaccio, G.; Carfi, A. Interdomain Communication in Hepatitis C Virus Polymerase Abolished by Small Molecule Inhibitors Bound to a Novel Allosteric Site. *J. Biol. Chem.* **2005**, *280*, 29765–29770.
(32) Irwin, J. J.; Shoichet, B. K. ZINC - A Free Database of Commercially Available Compounds for Virtual Screening. *J. Chem. Inf. Model.* **2005**, *45*, 177–182.
(33) A Free Database for Virtual Screening ZINC. http://blaster.docking.org/zinc/ (accessed Jan 03, 2008).
(34) Gopalsamy, A.; Lim, K.; Ciszewski, G.; Park, K.; Ellingboe, J. W.; Bloom, J.; Insaf, S.; Upeslacis, J.; Mansour, T. S.; Krishnamurthy, G.; Damarla, M.; Pyatski, Y.; Ho, D.; Howe, A. Y. M.; Orlowski, M.; Feld, B.; O'Connell, J. Discovery of Pyrano[3,4-b]indoles as Potent and Selective HCV NS5B Polymerase Inhibitors. *J. Med. Chem.* **2004**, *47*, 6603–6608.
(35) Gopalsamy, A.; Aplasca, A.; Ciszewski, G.; Park, K.; Ellingboe, J. W.; Orlowski, M.; Feld, B.; Howe, A. Y. M. Design and synthesis of 3,4-dihydro-1H-[1]-benzothieno[2,3-c]pyran and 3,4-dihydro-1H-pyrano[3,4-b]benzofuran derivatives as non-nucleoside inhibitors of HCV NS5B RNA dependent RNA polymerase. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 457–460.
(36) Pfefferkorn, J. A.; Nugent, R.; Gross, R. J.; Greene, M.; Mitchell, M. A.; Reding, M. T.; Funk, L. A.; Anderson, R.; Wells, P. A.; Shelly, J. A.; Anstadt, R.; Finzel, B. C.; Harris, M. S.; Kilkuskie, R. E.; Kopta, L. A.; Schwende, F. J. Inhibitors of HCV NS5B polymerase. Part 2: Evaluation of the northern region of (2Z)-2-benzoylamino-3-(4-phenoxy-phenyl)-acrylic acid. *Bioorg. Med. Chem. Lett.* **2005**, *15*, 2481–2486.
(37) Wang, M.; Ng, K. K.-S.; Cherney, M. M.; Chan, L.; Yannopoulos, C. G.; Bedard, J.; Morin, N.; Nguyen-Ba, N.; Alaoui-Ismaili, M. H.; Bethell, R. C.; James, M. N. G. Non-nucleoside Analogue Inhibitors Bind to an Allosteric Site on HCV NS5B Polymerase. *J. Biol. Chem.* **2003**, *278*, 9489–9495.
(38) Biswal, B. K.; Wang, M.; Cherney, M. M.; Chan, L.; Yannopoulos, C. G.; Bilimoria, D.; Bedard, J.; James, M. N. G. Non-nucleoside Inhibitors Binding to Hepatitis C Virus NS5B Polymerase Reveal a Novel Mechanism of Inhibition. *J. Mol. Biol.* **2006**, *361*, 33–45.
(39) Li, H.; Tatlock, J.; Linton, A.; Gonzalez, J.; Borchardt, A.; Dragovich, P.; Jewell, T.; Prins, T.; Zhou, R.; Blazel, J.; Parge, H.; Love, R.; Hickey, M.; Doan, C.; Shi, S.; Duggal, R.; Lewis, C.; Fuhrman, S. Identification and structure-based optimization of novel dihydropyrones as potent HCV RNA polymerase inhibitors. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 4834–4838.
(40) Chan, L.; Pereira, O.; Reddy, T. J.; Das, S. K.; Poisson, C.; Courchesne, M.; Proulx, M.; Siddiqui, A.; Yannopoulos, C. G.; Nguyen-Ba, N.; Roy, C.; Nasturica, D.; Moinet, C.; Bethell, R.; Hamel, M.; L'Heureux, L.; David, M.; Nicolas, O.; Courtemanche-Asselin, P.; Brunette, S.; Bilimoria, D.; Bedard, J. Discovery of thiophene-2-carboxylic acids as potent inhibitors of HCV NS5B polymerase and HCV subgenomic RNA replication. Part 2: Tertiary amides. *Bioorg. Med. Chem. Lett.* **2004**, *14*, 797–800.
(41) Kellenberger, E.; Rodrigo, J.; Muller, P.; Rognan, D. Comparative Evaluation of Eight Docking Tools for Docking and Virtual Screening Accuracy. *Proteins: Struct., Funct., Genet.* **2004**, *57*, 225–242.
(42) Kontoyianni, M.; Sokol, G. S.; McClellan, L. M. Evaluation of Library Ranking Efficacy in Virtual Screening. *J. Comput. Chem.* **2005**, *26*, 11–22.
(43) Moitessier, N.; Corbeil, C. R.; Englebienne, P.; Schwartzentruber, J. FITTED 2.2 can be obtained on request from the developers (nicolas.moitessier@mcgill.ca).

CI700398H