JCTC Journal of Chemical Theory and Computation

Article

pubs.acs.org/JCTC

Designing Short Peptides with High Affinity for Organic Molecules: A Combined Docking, Molecular Dynamics, And Monte Carlo Approach

Rolando P. Hong Enriquez,*[,†,‡] Silvia Pavan,[‡] Fabio Benedetti,[‡] Alessandro Tossi,[§] Adriano Savoini,[∥] Federico Berti,[‡] and Alessandro Laio[†]

[†]SISSA/ISAS - International School for Advanced Studies, Trieste 34100, Italy

[‡]Dipartimento di Scienze Chimiche e Farmaceutiche and [§]Dipartimento di Scienze della Vita, Università di Trieste, via Giorgieri 1, 34127 Trieste, Italy

[∥]T&B associati srl, Area Science Park, Padriciano 91, Trieste, Italy

Supporting Information

ABSTRACT: We present a method for designing artificial receptors capable of binding with high affinity to a chosen target organic molecule. The primary sequence of the peptide is optimized to maximize its binding affinity. Our algorithm builds on a combination of molecular dynamics, semiflexible docking, and replica exchange Monte Carlo and performs simultaneous sampling in sequence and conformational spaces carefully selecting the degree of flexibility in the mutated peptides. The approach is used to design a decapeptide able to bind efavirenz. The calculated binding energy of the designed peptide (approximately −12 kcal/mol) was confirmed experimentally by fluorescence measurements. NMR spectroscopy confirmed the interactions between the peptide and the efavirenz molecule predicted by the algorithm.

1. INTRODUCTION

The capability of proteins of recognizing each other and other molecules is possibly the most important working principle of biochemistry. The mechanism at the basis of this process has been optimized during billions of years of evolution and relies on the presence of binding sites able to establish multiple specific interactions between the partner molecules. The ability to design and synthesize artificial molecules with properties similar to those of an antibody or of an enzyme would offer invaluable applications in the technological and medical areas. Due to its utmost importance, this topic is currently at the center of intense investigation, and important progress has been made. Protein designers[1,2] are currently able to modify natural proteins with rationally chosen mutations, with the scope of improving their binding affinity toward a target or tuning their catalytic activity. Computational procedures combining optimization and evolutionary strategies have also been successfully applied to this problem.[1,3] However, embedding a recognition site in a protein scaffold retaining its structural stability to mutations is far from trivial. Additionally, proteins with a stable scaffold are normally large and difficult to synthesize or express in large quantities.

The design of short high-affinity peptides with 15 amino acids or less has recently emerged as a possible alternative to large proteins scaffolds.[4] This is due to at least three important reasons: (i) The number of different molecules (and functionalities) that can be obtained by combining the 20 natural amino acids is still overwhelmingly large even if the peptide is short and can be further extended using non-natural amino acids; (ii) molecular biology and chemistry offer several powerful techniques for the rapid screening of peptide libraries; and (iii) the automatic synthesis of peptides of this size even in large quantities is easy and inexpensive. From the computational point of view, however, designing a binding site on a small peptide rather than on a larger and stable scaffold poses severe problems: A small peptide does not show, normally, a stable structure and can change conformation upon binding to its target. Moreover, its preferred conformation can easily change due to a single or few mutations. As a consequence, several authors prefer to take advantage of available structural information of protein−protein and protein−peptide complexes to select candidate scaffolds for the binding site(see refs 5 and 6). On the other hand, algorithms that do not exploit structural information have to cope with the combinatorial explosion of conformational space and are normally used to design rather small drug-size peptides (∼3−4 residues long).[7,8]

Short peptides are also able to recognize and bind small ligands. This was demonstrated for example by Wu and Lo[9] who prepared peptides derived from dog olfactory receptors which bind trimethylamine with some selectivity. In related studies, small peptides binding porphyrin[10] or monosaccharides[11] were also isolated from experimentally generated peptide libraries. More recently, oligopeptides have also been shown to catalyze the aldol addition, proving that short peptides are capable of highly specific molecular recognition.[12] These examples demonstrate that designing short peptides capable of binding strongly and selectively to small ligands is possible and technologically promising. But again, from the computational point of view, designing such a peptide is even more challenging than designing drug-like peptides capable of binding to a protein. In fact, if the target is a small ligand, a correct strategy should not exploit available structural information; any viable optimization technique should instead take into account the high

Received: December 5, 2011
Published: February 13, 2012

flexibility of the peptide while trying to find the best binding pose.

In this work we introduce a novel approach aimed at designing peptides of 5−15 residues with high affinity toward small organic molecules, such as, for example, a drug or a metabolite. The algorithm is based on a combination of molecular dynamics (MD), semiflexible docking, and replica exchange Monte Carlo (MC) that allows performing importance sampling simultaneously in sequence and structure space. This allows taking fully into account the flexibility of the peptide during the sequence optimization. The algorithm has been implemented in a bash script embedding functionalities from Modeler[13] and Autodock Vina.[14] We apply this approach to design a peptide capable of binding efavirenz (EFV hereafter, see Figure 1), a potent

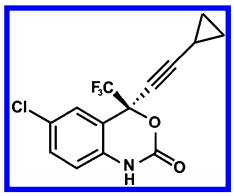

Figure 1. Structure of efavirenz.

non-nucleoside reverse transcriptase inhibitor widely used in HIV therapy.[15,16] A nanomolar binding affinity between a designed peptide and EFV was confirmed by fluorescence quenching and NMR experiments.

The choice of EFV as a target stems from our interest to develop rapid and efficient methods for the detection of anti-HIV drugs,[17] as designed peptides could provide an alternative to specific antibodies generally used for detection. Thus, beside being a testing system for the proposed algorithm, the high binding affinity for EFV has potential practical fallouts. Indeed, drug monitoring systems during HIV therapy should help to provide a personalized attention to the patients and reduce the appearance of resistance related phenomena.[18]

2. RESULTS

We implemented a computational procedure that, starting from a poly alanine, is able to optimize simultaneously the sequence and conformation of small peptides in order to reach a high binding affinity to a target organic molecule. The computational search is performed within the MC optimization scheme that, in this particular case, consists in a random walk in sequence and structure space. In this random walk each state is defined by the primary sequence SEQ and atomic coordinates R of a peptide. The energy of the state is the docking binding energy of the peptide with structure R and sequence SEQ versus a target molecule (see Methods Section). Starting from a state (R, SEQ), a new state (R_{new}, SEQ_{new}) is found using the following procedure:

(1) First the peptide is mutated at a random position, replacing the amino acid with a randomly selected new one. The new sequence will then be SEQ_{new}. The length of the peptide is not affected.

(2) The newly mutated structure is then relaxed by MD to avoid close contacts. The exact protocol of this step is described in Methods Section.

(3) Starting from the final structure of the relaxation, semiflexible docking between the peptide and the target is performed. In this docking step only the side chain and Ramachandran dihedral angles of the mutated amino acid are considered as flexible (see Methods Section for details). We will show below that restricting the flexibility to only a few dihedrals is essential to achieve a good compromise between computational speed and efficiency.

(4) Finally, the peptide docking pose with the lowest binding energy is selected; this pose defines the new coordinates R_{new}, and the docking score function defines the new energy E_{new}.

The new state is accepted or rejected according to the Metropolis criterion, namely with a probability equal to the minimum between 1 and $\exp((E_{new} - E)/T$. Here T is a fictitious "temperature" parameter that controls the probability of acceptance for the new configuration. If the new state is accepted, (R, SEQ, E) are set to (R_{new}, SEQ_{new}, E_{new}), otherwise R and SEQ are left unchanged, and a new mutation is attempted. In Figure 2 we plot the value of the energy E as a

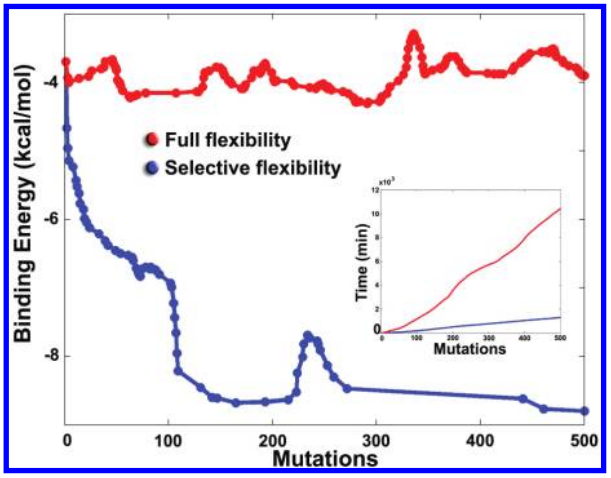

Figure 2. Importance of limited flexibility for the combined docking MC algorithm. Two runs of the algorithm with T = 0.2 performed leaving all the dihedrals free to move (red curve) or allowing only dihedral relaxations of the mutated residue (blue curve) are shown. To help visualization, the fluctuations in energy have been smoothed by taking a running average over a window of 10 mutations. Inset: computational time per mutation during the two runs. The algorithm with selective flexibility (blue) is considerably faster than its full flexible counterpart.

function of the number of mutations during a run that started with a deca-alanine. The temperature was set to T = 0.2. After only 150 mutations the estimated binding energy drops from an initial value of −5 to −9 kcal/mol. Even with this simple one temperature optimization, the algorithm was able to find a peptide with high binding energy for EFV (−8.8 kcal/mol) with sequence FPFPGWPPK.

2.1. Docking with Localized Flexibility Essential for Convergence. In this algorithm, limiting the flexibility only to the mutated amino acid and its adjacent Ramachandran angles (see scheme in Figure 7) is of paramount importance. In Figure 2, together with the energy per mutation obtained using selective flexibility, we show the results of a run performed with full flexibility of all the dihedrals in the peptides. Strikingly, in the latter case, the algorithm was not able to optimize the primary sequence, and the energy did not reach values below −5 kcal/mol. This is possibly due to the fact that most docking programs (like Autodock Vina[14] used here) are not designed to handle the large flexibility of small peptides. An additional advantage of limiting the flexibility is the gain in computational efficiency. In the inset of Figure 2, the time employed by the

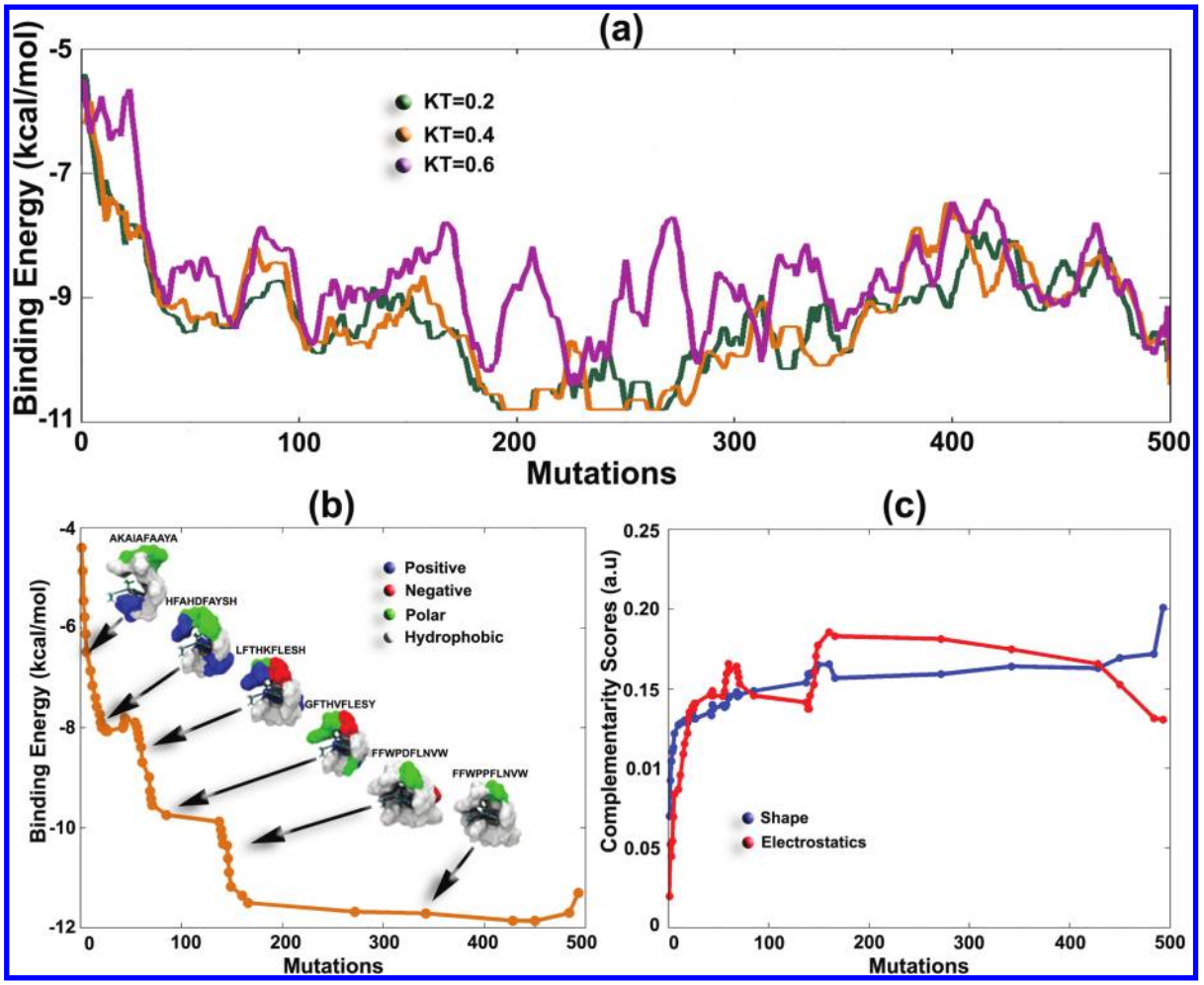

Figure 3. Optimization performed with the combined docking−MD−REMC algorithm. (a) Binding energy as a function of the number of mutations for a run performed at three temperatures (T = 0.2, 0.4, 0.6). (b) Binding energy as a function of the number of mutations at the lowest temperature, T = 0.2, for a statistically independent run of the algorithm. Six peptide sequences and the structures of the corresponding EFV complexes are also shown. Note the evolution of the peptides toward sequences and conformations that saturate the available surface contacts of EFV. The last structure, with sequence FFWPPFLNVW and binding energy of −12.1 kcal/mol, was selected for experimental validation. (c) Shape and complementarity scores of all the accepted peptides in the run in panel (b).

algorithm as a function of the number of performed mutations is presented for both cases; clearly, from the computational point of view, considering the flexibility of all the dihedrals at each docking step is at least eight times more expensive than performing docking with selected flexibility.

2.2. Replica Exchange Allows Avoiding Traps during Minimization. In Figure 2 (blue), after ∼200 mutations the energy slightly increases, as can happen in stochastic optimizations; the system then remains trapped in a fixed configuration for the last 200 mutations. To avoid these "traps" as much as possible and to improve the exploratory capabilities of the procedure, in our implementation we run N MC optimizations simultaneously at different fictitious temperatures (T_1,T_2, ..., T_N). At given times, one attempts swapping the configurations (R, SEQ) of two replicas according to a replica exchange Monte Carlo (REMC) scheme.[19] A swap between two replicas of temperatures T_i and T_{i+1} is accepted with probability equal to the minimum between 1 and $\exp[(E_i - E_{i+1})(1/T_i - 1/T_{i+1})]$, where E_i is the energy of the configuration in replica i. If the move is accepted, the sequence and the coordinates of the two replicas are swapped.

In Figure 3a, we plot the value of the energy E for a simulation performed with three replicas (N = 3) at temperatures T = 0.2, 0.4, and 0.6. The setup is otherwise identical to that of Figure 2. The exchanges between replicas allowed a better exploration of the mutational/conformational spaces; the solutions of the algorithm were improved by ∼3 kcal/mol with respect to the simple MC procedure. Figure 3b shows the binding energies for the T = 0.2 replica of a statistically independent replica exchange run. The best binding energy also in this case is approximately equal to −12 kcal/mol. In many other attempts, we could never obtain better solutions; this indicates that this value is likely to represent a lower limit for the binding affinity between a decapeptide and EFV. Six structures of peptide−EFV complexes obtained in this optimization, with a color-coded surface distribution of amino acid types, provide a simple pictorial representation of the 'evolution' of the algorithm.

In Figure 3c we plot the shape and the electrostatic complementarity scores (see Methods Section) of the accepted mutations. These scores offer a quantitative measure of the fitness of the designed peptides. The shape complementarity quantifies the extent of the contact surface between the peptide and EFV. The electrostatic complementarity quantifies the consistency between the electrostatic fields generated by the

two molecules at the contact points. Remarkably, both scores have the tendency to increase during the optimization, even if they were not explicitly taken into account in the MC acceptance condition. This indicates that our optimization algorithm, using the Vina scoring function, produces solutions that are also correct under other quality criteria.

2.3. Optimizing the Length of the Peptide. The algorithm described above allows the sequence and conformational optimization of short peptides maximizing their binding energies versus a target molecule. However, before using this procedure, an appropriate length for the initial peptide should be estimated according to the target's structure. Qualitatively, a larger number of contacts between the target and the peptide is expected to lead to stronger binding energies, and larger peptides are expected to form more contacts. In Figure 4a we present

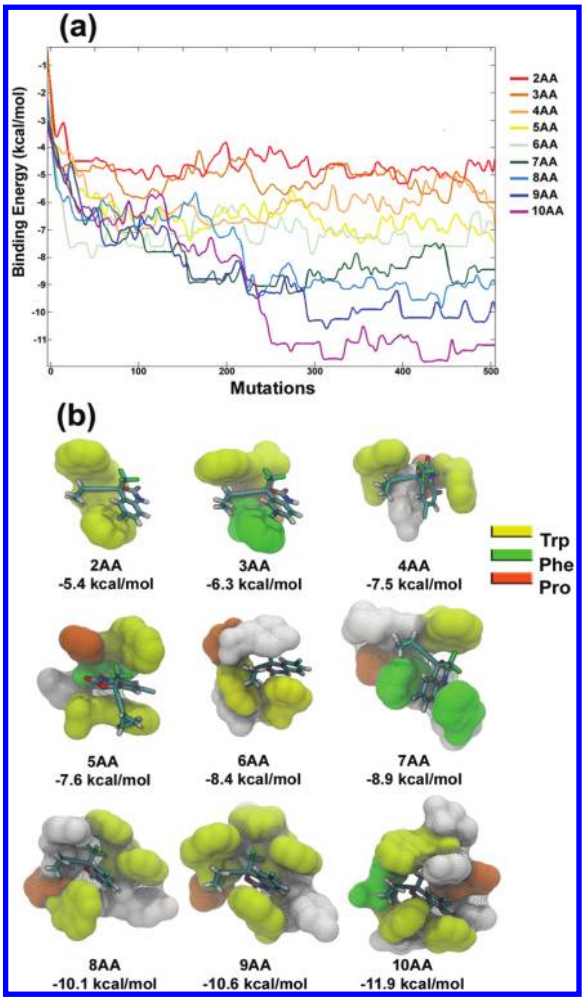

Figure 4. Optimization with different peptide lengths. (a) Binding energy as a function of the number of mutations for eight runs of peptides with length between 2 (red) and 10 AA (magenta). (b) Best docked complexes for each of the runs in (a).

the results of several optimizations versus EFV performed on peptides with sizes between 2 and 10. As can be seen, the binding energy improves as the size of the initial peptide increases. For dipeptides (red in Figure 4a) we rarely observe structures with binding energies better than −5 kcal/mol. Binding energies slowly improve and get better than −11 kcal/mol for decapeptides (magenta in Figure 4a). At this size, the peptide in its optimal conformation saturates most of the available contacts with EFV, as shown in Figure 4b. Therefore, it is unlikely that longer peptides would increase the number of contacts or improve the binding energies. Furthermore, larger peptides may adopt more stable secondary structures, decreasing the potential number of contacts with the target. Thus, for EFV, we decided to experimentally validate the results of the calculations selecting a REMC accepted peptide with size 10 AA.

2.4. Experimental Validation of the Algorithm. Using the REMC algorithm described above, we optimize the binding affinity of a decapeptide versus EFV(Figure 3b. A peptide (hereafter called pepHA) with sequence FFWPPFLNVW and theoretical binding energy of −12.1 kcal/mol was selected for experimental validation and synthesized using an automated microwave synthesizer.

2.4.1. Measurement of EFV−pepHA Affinity. The presence of fluorescent tryptophan residues in the sequence of pepHA allows exploiting fluorescence spectroscopy to assess the formation of a complex between the two molecules, obtain information about its stoichiometry, and estimate the dissociation equilibrium constant (K_d) at 298 K.

The peptide was synthesized by solid-phase techniques, and after LC purification, preliminary fluorescence experiments were performed at high peptide concentration in DMSO. The emission spectrum of pepHA was recorded as a synchronous scan[20] at Δ = 60 nm to selectively obtain the tryptophan fluorescence (Figure 5a); in the spectrum there is an overall maximum at 284 nm (i.e., 344 nm excitation), corresponding to

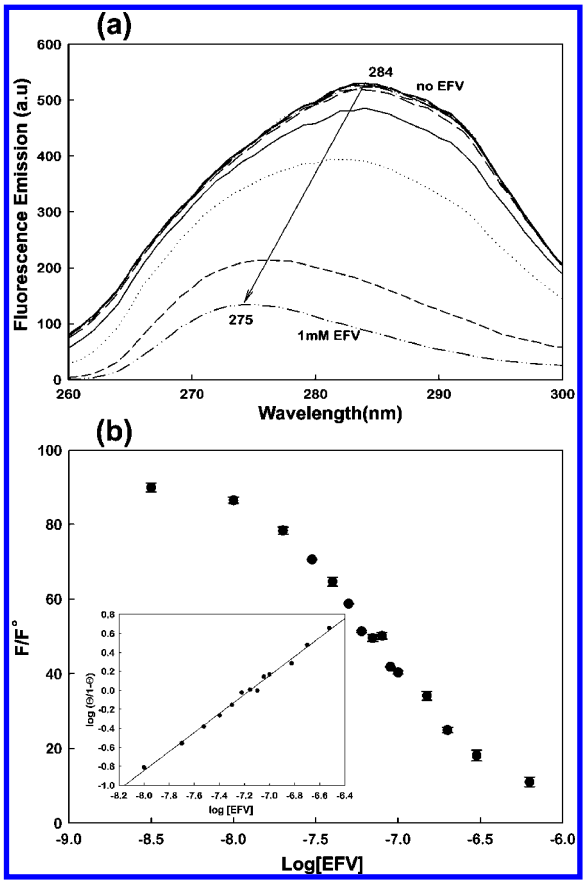

Figure 5. (a) Synchronous fluorescence emission spectrum (Δ = 60 nm) of 100 nM pepHA in DMSO and fluorescence quenching upon addition of EFV (50 nM to 1 mM). (b) Fluorescence quenching of 10 nM pepHA in 50 mM phosphate, 20% acetonitrile by EFV. Excitation 285 nm. Insert: Hill plot for the quenching data.

dx.doi.org/10.1021/ct200873y | J. Chem. Theory Comput. 2012, 8, 1121−1128

the typical tryptophan emission, and the quantum yield of this emission is comparable to that observed with similar peptides in similar conditions on the same instrument. Quenching is observed upon addition of EFV, with a blueshift from 284 to 275 nm. This indicates that the environment surrounding the Trp residues is becoming less polar, most likely resulting from replacement of solvent molecules by EFV. After this preliminary evidence of interaction, the solvent was changed to water, as the peptide has been designed in this solvent, and thanks to the relatively high emission, the concentration of peptide was lowered in order to allow K_d measurement. Quenching was observed again in 50 mM phosphate buffer, pH 7.4, containing 10% acetonitrile (Figure 5b). A Stern−Volmer analysis of the quenching data was carried out first to determine if the observed quenching is caused by binding or by collisional phenomena; a resulting apparent Stern−Volmer quenching constant of 9.09×10^6 L mol^{-1} was obtained. Assuming a decay time for tryptophan fluorescence of 5 ns, the observed Stern−Volmer constant gives an apparent bimolecular quenching constant as high as 1.8×10^{15} L mol^{-1}s^{-1}. This value is orders of magnitude higher than the limit for diffusion-controlled collisional quenching and consistent with static quenching originating from the association of the fluorophore and the quencher in a bimolecular complex.[20] The dissociation constant (K_d) for the 1:1 complex between EFV and pepHA was therefore evaluated at 298 K from a Hill analysis of the fluorescence data. The experimental value found was K_d = 64 nM, which corresponds to a binding energy E = −9.7 kcal/mol.

2.4.2. Interactions between pepHA and EFV Probed by NMR. Since fluorescence spectroscopy confirms that the designed peptide interacts significantly with EFV, we performed NMR experiments to verify if the structure of the complex predicted by the algorithm is correct. The partial ^{1}H NMR spectra of free EFV and of EFV/pepHA mixtures are shown in Figure 6a (see Supporting Information for EFV and peptide NMR assignments). Even at 0.5:1 EFV/pepHA ratio, a significant upfield drift in the EFV NH proton chemical shift is noticeable which could be related to a direct involvement of this proton in a hydrogen bond. This interaction is consistent with the model structure obtained with our algorithm (Figure 3b) where the EFV NH proton is predicted to be at a hydrogen-bond distance with the carbonyl oxygen of Pro5. Close contacts of EFV with pepHA are also supported by the observation of a set of 2D nuclear Overhauser enhancement spectroscopy (NOESY) cross-peaks between peptide and ligand signals at a 3:1 EFV/pepHA molar ratio (Figure 6b,c). The 2D NOESY provides detailed information about intermolecular contacts since correlations are observed only for pairs of protons separated by less than 5 Å, and the cross peak intensities depend on the inverse sixth power of the internuclear separations. The NH proton of EFV shows cross-peaks with the NH amide proton of Leu7 and with an aromatic proton of Trp3 (Figure 6c), in agreement with the structure predicted by the algorithm. Moreover, the cyclopropyl methyne proton of EFV displays cross-peaks with aromatic protons of Trp10 and with the indole NH of Trp3 (Figure 6b); these interactions were also predicted by the theoretical model. Therefore the binding between the peptide and EFV is likely to involve: (i) a hydrogen bond between the NH group of EFV and acceptors at the N-terminal region of pepHA; (ii) $\pi-\pi$ interactions between the benzoxazin-2-one ring of EFV and Trp residues; and (iii) other hydrophobic interactions at the cyclopropyl ring level. Both the N- and C-terminals of the peptide are thus involved in interactions with the ligand.

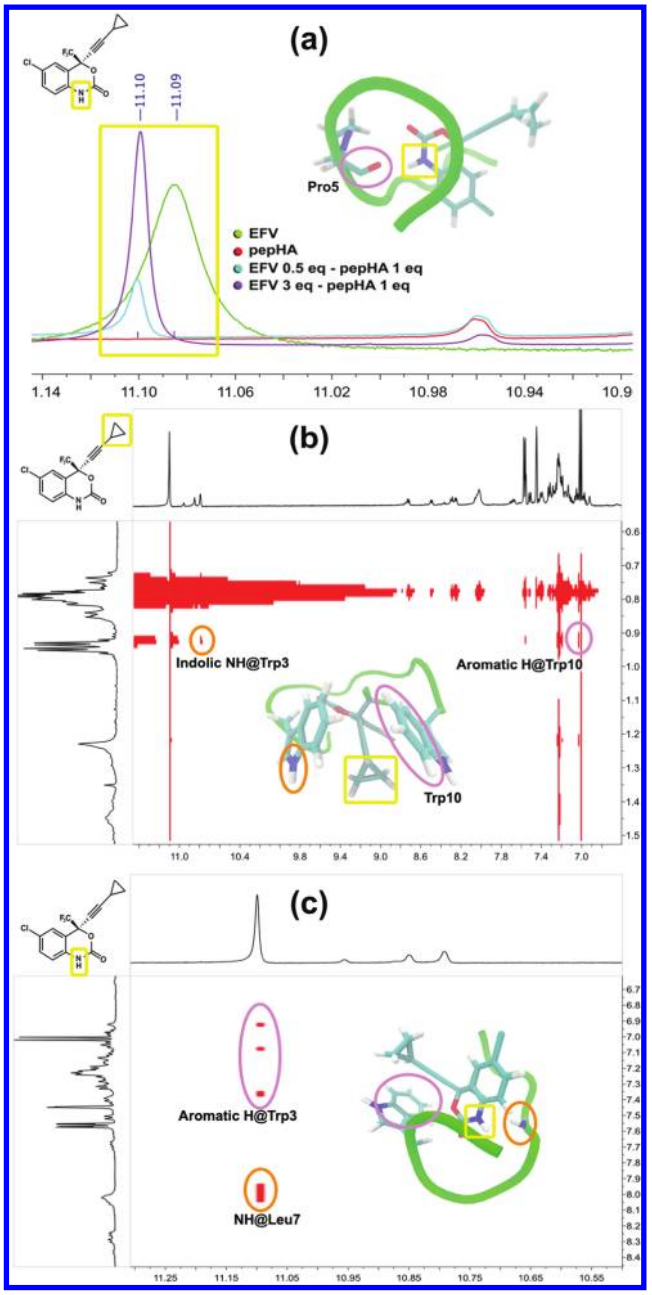

Figure 6. Comparison between EFV−pepHA interactions observed by NMR experiments in DMSO and the theoretical structure of the complex (see Figure 3). (a) ^{1}H NMR 500 MHz spectra for 5 mM EFV (green), 5 mM pepHA (red), and mixtures of EFV/pepHA (light blue 1:2; purple 3:1). The NMR reported hydrogen bond is predicted by the model between NH@EFV and O@Pro5. (b) Details of the 2D NOESY NMR spectrum of a 3:1 mixture of EFV and pepHA. Several EFV protons (yellow frames) are at NOE distance with peptide protons (colored frames). With the exception of the contact between H_ε@Trp3 showed in (c), all the NMR interactions are predicted by the model.

3. CONCLUSIONS

We introduced an algorithm able to find small peptides with high affinity for a predetermined small organic ligand. The algorithm explores simultaneously the sequence and conformational spaces of the peptide using a combination of MD and semiflexible docking within the framework of REMC. This allows taking fully into account the flexibility of the peptide during the sequence

dx.doi.org/10.1021/ct200873y | J. Chem. Theory Comput. 2012, 8, 1121−1128

optimization for chain lengths of 10 amino acids or more. The key ingredients for the success of the algorithm are at least three: (i) the flexibility of the peptide is taken into account by generating at each mutation all the rotamers of the side chain and Ramachandran dihedrals of the mutated residue. This choice, in the framework of importance sampling, still allows visiting all the possible conformations of the peptide without explicitly considering all the rotamers at each step, with a significant gain in efficiency. (ii) The use of MD after every mutation, that allows relaxing the structures of each conformer/rotamer before the docking step. (iii) The use of a REMC scheme, in which the high-temperature replicas provide an ensemble of conformations and sequences highly uncorrelated to the low-temperature replicas.

The algorithm is able to find solutions in which the peptide "wraps" the target molecule, forming with it a large number of favorable contacts. This is demonstrated by monitoring the evolution of a shape and electrostatic complementarity score that tend to grow systematically during the optimization. For a specific target, an EFV molecule, we verified that statistically independent runs lead to similar estimated binding energies. The primary sequences of the peptides optimized in these runs are different, as one would expect in a stochastic optimization procedure in a space of enormous extension (10^{20} for a peptide of length 10). However, the best solutions always contain similar residues, indicating a certain degree of reproducibility in the type of optimized interactions.

The high binding energy toward EFV of a selected peptide optimized with our procedure was confirmed experimentally by fluorescence measurements. Most of the contacts between the peptide and EFV predicted by the algorithm were confirmed by NMR experiments, indicating that also the theoretically predicted structure of the complex is essentially correct. Clearly, the reliability of the algorithm should be further tested on other peptides designed for EFV and/or other targets. We hope that this work will trigger an important validation effort to expose hidden deficiencies of the procedure that can lead us, ultimately, to improve our methodology. In the current form, the main known limitations of the algorithm are the limited accuracy of the scoring function used to estimate the binding affinity (~2.75 kcal/mol)[14] and the need for more effective ways to control the specificity of the candidate peptides. Satisfactory solutions to these limitations are, however, far from trivial.

In conclusion, the present work introduces an algorithm to find small peptides with high binding affinity to a preselected target. This method can still be improved to cope with challenges like the selectivity of the designed peptides; a better estimator of the binding affinity would also be valuable. Still, the method successfully addresses the problem of sequence/conformational sampling, allowing the design of relatively large peptides without using any structural template. Additionally, the method is general, and as such, it can be used with any molecular target, opening a wide range of applications in sensor, diagnostic, or clinical areas.

4. METHODS

4.1. Molecular Dynamics. Mutations, minimizations, and MD functionalities were performed using MODELER v 9.7.[13] For the mutations, the coordinates and parameters of the mutated side chains were extracted from an internal Charmm22 force field database.[21] To avoid close contacts caused by the mutation, subsequent steps including (i) conjugate gradient (CG) energy minimization, (ii) MD with simulated annealing, and (iii) a final CG energy minimization were performed locally in the mutated side chain and nonbonded neighboring atoms. In more details, we used the following protocol:[22] The first CG minimization phase actually consists of five successive minimizations of up to 200 steps each, increasing the scaling factors for the nonbonded restraints from 0, 0.01, 0.1, 0.5, to 1.0, respectively. The MD phase used a time step of 4 fs and consists of a fast heating with 200 steps of MD at 150, 250, 400, 700, and 1000 K and a slower cooling using 600 steps of MD at 1000, 800, 600, 500, 400, and 300 K. A final phase of 1000 steps of CG minimization completes the protocol.

4.2. Docking. After a properly relaxed mutated structure of the peptide is obtained, the next step in the algorithm is a semiflexible docking of the peptide and the target molecule, i.e., EFV. In the docking setup we used the center of mass of EFV as the center of a cubic docking grid with side dimensions approximately equal to the length of the extended initial peptide (~25 Å for deca-alanine). Most of the dihedrals in the peptide were kept rigid with the exception of those in the side chain of the mutated amino acid and the Ramachandran dihedrals connecting the mutated amino acid with its neighbors (Figure 7). Binding energies were obtained using the scoring function of

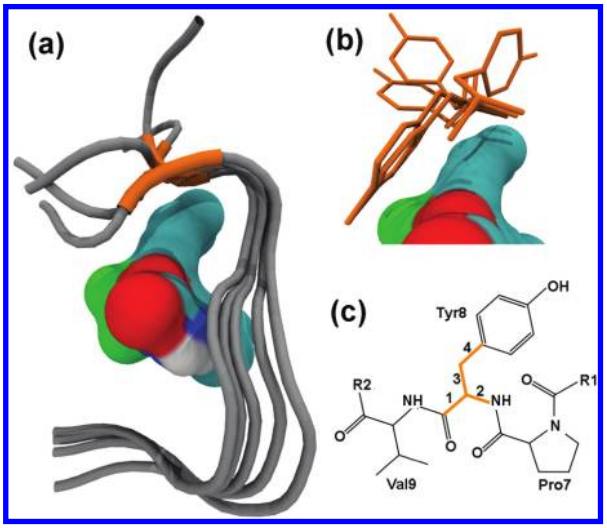

Figure 7. Scheme of the semiflexible docking setup used in each step of the algorithm. (a) Tube representation of five poses of a peptide with sequence AYWPGYPYVY docked versus EFV (Connolly surface representation). The last mutated amino acid is Tyr8, (orange). The five poses differ by rotations in the two Ramachandran angles of Tyr8. (b) Poses of the side chains of Tyr8 probed in the same docking step. The two dihedral angles around the $C_\alpha-C_\beta$ and the $C_\beta-C_\gamma$ bonds are allowed to rotate. (c) 2D representation of the rotatable dihedrals. Dihedrals 1 and 2 are the Ramachandran angles of Tyr8. Dihedrals 3 and 4 are the dihedral angles around the $C_\alpha-C_\beta$ and the $C_\beta-C_\gamma$ bonds of Tyr8.

Autodock Vina.[14] Autodock 4 structure files (.pdbqt) were used to define the dihedrals that are not held fixed, the two Ramachandran angles of the residue that are mutated, and all the dihedral angles of its side chain. To compare with experiments, it was assumed that the error of a Vina binding energy docking calculation was 2.75 kcal/mol.[14] Molecular docking is the computationally most time-consuming step in our implementation. Using semiflexible docking allows increasing the speed of each step at least by a factor of 8 (with respect to fully flexible docking). A single docking calculation with our setup requires ~1 min on 8 AMD 2200 MHz processors.

4.3. Shape and Electrostatic Complementarity. In order to quantify the extent of the contact surface between the interacting molecules (i.e., the peptide and EFV) we used a shape complementarity score. Independent surface triangulations of the docked molecules using a rolling probe of a diameter of 1.4 Å were calculated using the program MSMS.[23] The complementarity score SC is then defined as

$$\mathrm{SC} = \sum_{i\in \mathrm{A}, j\in \mathrm{B}} g(r_{ij})$$

where the sum runs over the vertices i and j belonging to the two surfaces A and B, r_{ij} is the distance between vertex i and j, and $g(r) = 1$, if $r < 0.5$ Å and 0 otherwise. In the spirit of ref 24, we also defined an electrostatic complementarity measure between the two molecules in the following manner: (i) We assigned partial charges to the molecules using the program PDB2PQR;[25,26] (ii) we solved the Poisson−Boltzmann equation for the interacting molecules using the APBS program;[27] (iii) we computed the value of the electrostatic potential U_i on the vertices of the two molecular surfaces. The electrostatic complementarity score EC is then defined as

$$\mathrm{EC} = \sum_{i\in \mathrm{A}, j\in \mathrm{B}} g(r_{ij})|U_i - U_j|$$

Since one molecule (EFV) is modeled as rigid, we calculate its electrostatic potential only once. An increase in the score can therefore be only caused by mutations that increase the number of contacts and/or induce a larger difference $|U_i - U_j|$ in contacting vertices and then a stronger electrostatic interaction.

4.4. Chemistry: Materials. FMOC-protected amino acids and FMOC-Trp(Boc)-NovaSyn-TGA resin were purchased from Novabiochem. DMSO-d_6 was acquired from Cambridge Isotope Laboratories, Inc. All the other reagents and solvents were obtained at the highest purity available from Sigma-Aldrich and used without further purification. The phosphate buffer used for the binding analysis was 50 mM of $Na_2HPO_4$and 100 mM of NaCl (pH 7.4). EFV was synthesized from 4-chloroaniline via enantioselective addition of Li-cyclopropyl acetylide to *p*-methoxybenzyl-protected ketoaniline, in the presence (1*R*,2*S*)-*N*-pyrrolidinylnorephedrine lithium alkoxide.[28]

4.5. Peptide Synthesis and Purification. The peptide was synthesized using an automatic microwave peptide synthesizer (Liberty CEM) at 100 μmol scale. Standard solid-phase peptide synthesis protocols for FMOC chemistry were used throughout. The peptide was assembled at the C-terminal carboxyl using FMOC-Trp(Boc)-NovaSyn-TGA resin 0.23 mmol/g substituted. Coupling was accomplished with 5 equiv of PyBop using 5 equiv of amino acids and 10 equiv of DIPEA. The FMOC group was removed with 0.1 M HOBt in 20% piperidine in DMF. Finally, the side-chain-protecting groups and the resin support were removed with TFA/DODT/thioanisole/phenol/TIPS/H_2O (83:8:3:2:2:2) for 2 h. After cleavage, the resin was removed by filtration and the filtrate concentrated to approximately 1 mL under a stream of nitrogen. The deprotected peptide was precipitated by addition of cold *tert*-butyl methyl ether, centrifuged, and washed three times with cold diethyl ether. The peptide (pepHA) was purified by reversed-phase HPLC (Amersham Pharmacia Biotech) on a preparative Waters column (Delta-Pak, C18, 15 μm, 300 Å, 25 × 100 mm) using appropriate water/acetonitrile linear gradients in the presence of 0.1% trifluoroacetic acid. Identity was confirmed by electrospray mass spectra analysis (Esquire 4000 Bruker, Daltonics).

4.6. NMR Spectroscopy. The complex formed by pepHA and EFV was also studied by NMR spectroscopy. All NMR experiments were recorded at 500 MHz on a Varian 500 spectrometer. All spectra were run at 298 K. DMSO-d_6 was chosen for the NMR experiments because the peptide is readily soluble in such solvent, while it is hardly soluble in aqueous buffer. Advantageously, in DMSO-d_6 solution, the risk of self-aggregation decreases, and also peptide NH proton resonances are clearly observed allowing unambiguous resonance assignments. The peak of residual nondeuterated solvent (δ = 2.50 ppm in proton spectra, δ = 39.7 ppm in ^{13}C experiments) was used as reference for the chemical shift calibration in all the experiments. The sample of the pure peptide was prepared at 6.25 mM concentration in DMSO-d_6. Complete backbone and partial side-chains ^{1}H assignments for pepHA were obtained by analysis of correlation spectroscopy (COSY), total correlation spectroscopy (TOCSY), and NOESY experiments and ^{13}C assignments from gradient heteronuclear single quantum coherence adiabatic experiments (gHSQCAD). The EFV−pepHA complex formation was studied by addition of 0.5, 1, 2, and 3 equiv of EFV to the DMSO-d_6 peptide solution, in a final total volume of 840 μL for each sample.

4.7. Fluorescence Spectroscopy. The binding affinity of pepHA and EFV was studied by intrinsic tryptophan fluorescence quenching. In DMSO, 1 ×10^{-3} mol L^{-1} pepHA and EFV stock solutions were prepared. Steady-state fluorescence spectra were recorded at 298 K on a CARY Eclipse (Varian) spectrofluorimeter equipped with a 1.0 cm quartz cuvette. Synchronous fluorescence spectra were measured by setting the excitation wavelength in the 295−400 nm range, with increments of 60 nm, and the emission was recorded in the 250−450 nm range. The width of the excitation and emission slits was set to 5.0 nm for all measurements. In a first set of experiments, the spectra were recorded in DMSO, at 100 nM of pepHA. The concentration of the peptide was then lowered to at 10 nM in 50 mM phosphate buffer containing 20% acetonitrile, and the concentration of EFV was increased from 0 to 3 × 10^{-7}mol L^{-1}. After each addition and equilibration of the ligand, the intensity of the fluorescence at the maximum emission wavelength and the drift of such maximum were measured.

■ ASSOCIATED CONTENT

Ⓢ Supporting Information

A pseudocode of the algorithm presented here. A collection of NMR spectral data for EFV and pepHA. Coordinates of the complexes portrayed in Figure 3. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*E-mail: rolando.hong@gmail.com.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We are grateful to Commissariato del Governo nella Regione Friuli-Venezia Giulia (Fondo Trieste grant 597/08 "Piccoli peptidi per lo sviluppo di biosensori") for having funded our research project and for grants to R.H. and S.P. We warmly acknowledge Pilar Cossio, Francesco Colizzi, and Giacinto Scoles for useful discussions.

REFERENCES

(1) Leisola, M.; Turunen, O. Protein engineering: opportunities and challenges. *Appl. Microbiol. Biotechnol.* **2007**, *75*, 1225−1232.

(2) Jackel, C.; Kast, P.; Hilvert, D. Protein design by directed evolution. *Annu. Rev. Biophys.* **2008**, *37*, 153−173.

(3) Lippow, S.; Tidor, B. Progress in computational protein design. *Curr. Opin. Biotechnol.* **2007**, *18*, 305−311.

(4) Vanhee, P.; van der Sloot, A.; Verschueren, E.; Serrano, L.; Rousseau, F.; Schymkowitz, J. Computational design of peptide ligands. *Trends Biotechnol.* **2011**, *29*, 231−239.

(5) Fernandez-Ballester, G.; Beltrao, P.; Gonzalez, J.; Song, Y.; Wilmanns, M.; Valencia, A.; Serrano, L. Structure-Based Prediction of the Saccharomyces cerevisiae SH3-Ligand Interactions. *J. Mol. Biol.* **2009**, *338*, 902−916.

(6) Raveh, B.; London, N.; Schueler-Furman, O. Sub-angstrom modeling of complexes between flexible peptides and globular proteins. *Proteins* **2010**, *78*, 2029−2040.

(7) Abe, K.; Kobayashi, N.; Sode, K.; Ikebukuro, K. Peptide ligand screening of a-synuclein aggregation modulators by in silico panning. *BMC Bioinformatics* **2007**, *8*, 451.

(8) Unal, E.; Gursoy, A.; Erman, B. VitAL: Viterbi Algorithm for de novo Peptide Design. *PLoS One* **2010**, *5*, e10926.

(9) Wu, T.; Lo, Y. Synthetic peptide mimicking of binding sites on olfactory receptor protein for use in 'electronic nose'. *J. Biotechnol.* **2000**, *80*, 63−73.

(10) Nakamura, C.; Inuyama, Y.; Shirai, K.; Sugimoto, N.; Miyake, J. Detection of porphyrin using a short peptide immobilized on a surface plasmon resonance sensor chip. *Biosens. Bioelectron.* **2001**, *16*, 1095−1100.

(11) Sugimoto, N.; Miyoshi, D.; Zou, J. Development of small peptides recognizing a monosaccharide by combinatorial chemistry. *Chem. Commun.* **2001**, 2295−2296.

(12) Jiang, L.; Althoff, E.; Clemente, F.; Doyle, D.; Rothlisberger, D.; Zanghellini, A.; Gallaher, J.; Betker, J.; Tanaka, F.; Barbas, C. III; Hilvert, D.; Houk, K.; Stoddard, B.; Baker, D. De Novo Computational Design of Retro-Aldol Enzymes. *Science* **2008**, *319*, 1387−1391.

(13) Eswar, N.; Marti-Renom, M.; Webb, B.; Madhusudhan, M.; Eramian, D.; Shen, M.; Pieper, U.; Sali, A. Development of small peptides recognizing a monosaccharide by combinatorial chemistry. *Chem. Commun.* **2001**, 2295−2296.

(14) Trott, O.; Olson, A. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization and multithreading. *J. Comput. Chem.* **2010**, *31*, 455−461.

(15) Hammer, S.; Eron, J. Jr; Reiss, P.; Schooley, R.; Thompson, M.; Walmsley, S.; Cahn, P.; Fischl, M.; Gatell, J.; Hirsch, M.; Jacobsen, D.; Montaner, J.; Richman, D.; Yeni, P.; Volberding, P. Antiretroviral treatment of adult HIV infection: 2008 recommendations of the International AIDS Society-USA panel. *JAMA, J. Am. Med. Assoc.* **2008**, *300*, 555−570.

(16) De Clerck, E. The history of antiretrovirals: key discoveries over the past 25 years. *Rev. Med. Virol.* **2009**, *19*, 287−299.

(17) Bastiani, E.; Benedetti, F.; Berti, F.; Campaner, P.; Donadel, E.; Montagna, M.; Regazzi, M.; Rinaldi, S.; Savoini, A.; Venturini, R. Development and evaluation of an immunoassay for the monitoring of the anti-HIV drug amprenavir. *J. Immunol. Methods* **2007**, *325*, 35−41.

(18) Liu, X.; Ma, Q.; Zhang, F. Therapeutic drug monitoring in highly active antiretroviral therapy. *Exp. Opin. Drug Safety* **2010**, *9*, 743−758.

(19) Iba, Y. Entended ensemble monte carlo. *Int. J. Mod. Phys. C* **2001**, *12*, 623−656.

(20) Lackowicz, J. *Principles of Fluorescence Spectroscopy*; Springer Science + Business Media LCC: New York, 2006.

(21) MacKerell, A. Jr; Bashford, D.; Bellott, M.; Dunbrack, R. Jr; Evanseck, J.; Field, M.; Fischer, S.; Gao, J.; Guo, H.; Ha, S.; Joseph-McCarthy, D.; Kuchnir, L.; Kuczera, K.; Lau, F.; Mattos, C.; Michnick, S.; Ngo, T.; Nguyen, D.; Prodhom, B.; Reiher, W. III; Roux, B.; Schlenkrich, M.; Smith, J.; Stote, R.; Straub, J.; Watanabe, M.; Wiorkiewicz-Kuczera, J.; Yin, D.; Karplus, M. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* **1998**, *102*, 3586−3616.

(22) Feyfant, E.; Sali, A.; Fiser, A. Modeling mutations in protein structures. *Protein Sci.* **2009**, *16*, 2030−2041.

(23) Sanner, M.; Olson, A.; Spehner, J. Fast and Robust Computation of Molecular Surfaces. *Proc. 11th ACM Symp. Comput. Geom.* **1995**, C6−C7.

(24) McCoy, A.; Chandana, E.; Colman, P. Electrostatic complementarity at protein-protein interfaces. *J. Mol. Biol.* **1997**, *298*, 570−584.

(25) Dolinsky, T.; Nielsen, J.; McCammon, J.; Baker, N. PDB2PQR: an automated pipeline for the setup, execution, and analysis of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res.* **2004**, *32*, W665−W667.

(26) Dolinsky, T.; Czodrowski, P.; Li, H.; Nielsen, J.; Jensen, J.; Klebe, G.; Baker, N. PDB2PQR: Expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res.* **2007**, *35*, W522−W525.

(27) Baker, N.; Sept, D.; Joseph, S.; Holst, M.; McCammon, J. Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc. Natl. Acad. Sci. U.S.A.* **2001**, *98*, 10037−10041.

(28) Pierce, M.; Parsonos, R.; Radesca, L.; Silverman, S.; Moore, J.; Islam, Q.; Choudhury, A.; Fortunak, J.; Nguyen, D.; Luo, C.; Morgen, S.; Davis, W.; Gonfalone, P.; Chen, C.; Tillyer, R.; Fey, L.; Tan, L.; Xu, F.; Zhao, D.; Thompson, A.; Corley, E.; Grabowski, E.; Reamer, R.; Reider, P. Practical asymmetric synthesis of efavirenz (DMP 266), an HIV-1 reverse transcriptase inhibitor. *J. Org. Chem.* **1998**, *63*, 8536−8543.