JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

pubs.acs.org/jcim

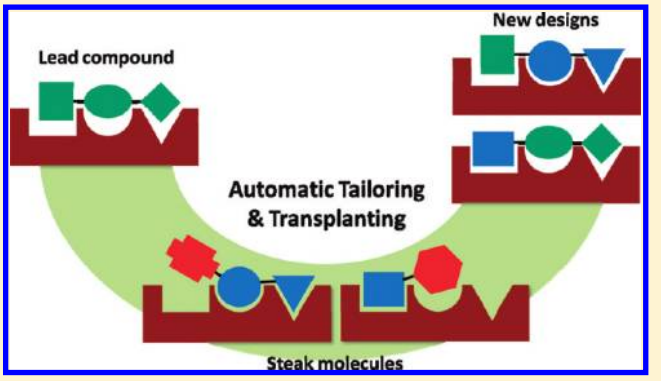
Automatic Tailoring and Transplanting: A Practical Method that Makes Virtual Screening More Useful

Yan Li, Yuan Zhao, Zhihai Liu, and Renxiao Wang*

State Key Laboratory of Bioorganic Chemistry, Shanghai Institute of Organic Chemistry, Chinese Academy of Sciences, 345 Lingling Road, Shanghai 200032, People's Republic of China

Supporting Information

ABSTRACT: Docking-based virtual screening of large compound libraries has been widely applied to lead discovery in structure-based drug design. However, subsequent lead optimizations often rely on other types of computational methods, such as de novo design methods. We have developed an automatic method, namely automatic tailoring and transplanting (AutoT&T), which can effectively utilize the outcomes of virtual screening in lead optimization. This method detects suitable fragments on virtual screening hits and then transplants them onto a lead compound to generate new ligand molecules. Binding affinities, synthetic feasibilities, and drug-likeness properties are considered in the selection of final designs. In this study, our AutoT&T program was tested on three different target proteins, including p38 MAP kinase, PPAR-α, and Mcl-1. In the first two cases, AutoT&T was able to produce molecules identical or similar to known inhibitors with better potency than the given lead compound. In the third case, we demonstrated how to apply AutoT&T to design novel ligand molecules from scratch. Compared to the solutions generated by other two de novo design methods, i.e., LUDI and EA-Inventor, the solutions generated by AutoT&T were structurally more diverse and more promising in terms of binding scores in all three cases. AutoT&T also completed the assigned jobs more efficiently than LUDI and EA-Inventor by several folds. Our AutoT&T method has certain technical advantages over de novo design methods. Importantly, it expands the application of virtual screening from lead discovery to lead optimization and thus may serve as a valuable tool for many researchers.



INTRODUCTION

Identification of promising lead compounds is typically the starting point of a drug discovery project. In many cases, lead compounds are discovered through high-throughput screening of chemical libraries consisting of natural products or synthesized compounds. Computer-aided drug design (CADD)[1] provides new opportunities. Among all CADD techniques, virtual screening based on the three-dimensional structure of a target protein has been widely employed as a practical approach in the discovery of novel lead compounds.[2−4] Virtual screening narrows down the candidate compounds to be experimentally screened from millions to hundreds, leading to an improved success rate of finding active compounds at a much lower cost. As a matter of fact, virtual screening has been applied successfully to more than a hundred molecular targets since 1990s.[5−9]

Nevertheless, the active hits discovered in a virtual screening job often exhibit only low to medium levels of biological activities. These compounds need to be structurally modified to achieve tighter binding to the target protein. Optimization of lead compounds, if necessary, also aims at solving poor bioavailability and pharmacokinetic properties. Traditionally, ideas of structural optimization are proposed by experts with profound knowledge in medicinal chemistry, an approach still widely applied today. Some empirical methods, such as the bioisostere principle[10] and various structure−activity relationship studies,[11] have been developed during this process. But an obvious shortcoming in this approach is that an expert tends to propose the chemical structures that he/she is familiar with. Thus, the success of this approach is largely dependent on the experts themselves.

With the advent of CADD, computer programs have provided an alternative approach for lead optimization. A group of methods called de novo design[12,13] have been developed for this purpose. Currently, there are a variety of available commercial programs for performing de novo design, such as LUDI[14] (as implemented in the Discovery Studio software) and EA-Inventor[15] (as implemented in the Sybyl software) as well as some programs released by academic groups, such as LigBuilder.[16] These methods install various fragments onto the core structure of a given lead compound to generate new ligand molecules fitting better to the target protein. They can also link multiple fragments placed separately inside the binding pocket on the target protein into integrated molecules. Thus, they are also referred to as "build-up" methods. These methods essentially automate the common

Received: January 25, 2011
Published: April 26, 2011

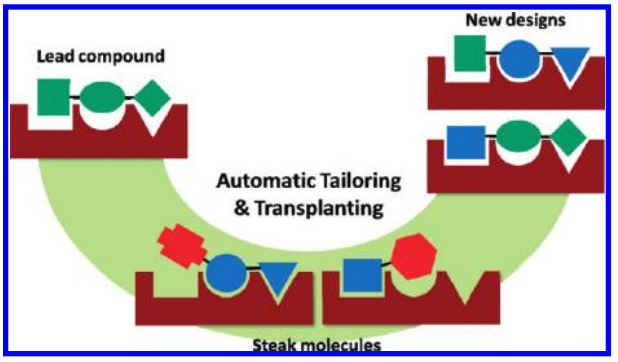


Figure 1. The basic idea of the AutoT&T method: Useful fragments on the "steak molecules" are transplanted onto the lead compound to produce optimized structures.

practice by human experts in a more systematic manner with computer programs. These computer programs can produce more diverse solutions as compared to human designs. The users can make rational choices among the outcomes of these computer programs, or they can be inspired by these outcomes to come up with even better ideas. Many successful applications of de novo design methods have also been reported so far.[12,17]

No matter what lead optimization is performed by human experts or with the aid of computer programs, the outcomes of the virtual screening job conducted at the previous lead discovery stage are not useful any more. Virtual screening often requires high-performance computing facilities since hundreds of thousands of molecules need to be examined through extensive molecular docking to the target protein. In a virtual screening job, only a small group of top-ranked molecules (typically <1%) will be selected for further consideration; whereas others are discarded. In other words, more than 99% of the computational expenses of virtual screening are simply wasted. Although virtual screening only produces virtual results, the costs on purchasing and maintaining the required hardware and software are real, and they are not as cheap as one may think. Considering the wide application of virtual screening, it is certainly desired if virtual screening results can somehow be utilized in lead optimization as well.

Here, we describe a practical method, called automatic tailoring and transplanting (AutoT&T), which is developed primarily for performing lead optimization in structure-based drug design. A set of in-house computer programs have been developed to automate this method. The key idea of this method is illustrated in Figure 1. Basically, AutoT&T needs the information of a lead compound as the starting point. In addition, a docking-driven virtual screening against the given target protein needs to be conducted in prior. AutoT&T identifies suitable fragments on the screened molecules based on their binding modes derived from molecular docking. Then, these selected fragments are transplanted onto the lead compound to form new ligand molecules with higher binding affinities. In this way, the outcomes of virtual screening are effectively "recycled" in lead optimization. AutoT&T was applied to three test cases in this study. It produced structurally more diverse and more promising designs than two standard de novo design methods in a head-to-head comparison. The following sections will provide descriptions on the algorithms of AutoT&T, the results obtained in test cases, and discussion on the technical advantages of AutoT&T over conventional de novo design methods.

METHODS

The overall flowchart of the AutoT&T program is illustrated in Figure 2. The necessary inputs include the three-dimensional structure of a target protein and a given lead compound. The binding pose of the lead compound, which is derived with a computational or experimental method, should also be provided. In addition, the binding poses of a whole compound library need to be supplied. Outcomes of a virtual screening job against the same target protein serve for this purpose perfectly. We will refer to the molecules in this compound library as "steak molecules" throughout this article in a metaphor that they will be truncated to get the useful parts. The protein structure is required to be given in the popular PDB format. All small molecules can be saved in either the SDF or the Tripos Mol2 formats.

The AutoT&T program is composed of three major modules. A structural operation module generates new molecules by swapping fragments between existing molecules. An evaluation module computes the binding affinities of newly generated molecules, which works in couple with the structural operation module. A postprocessing module processes the outputs of the structural operation module and evaluates other aspects of the newly generated molecules. The final outputs of the AutoT&T program are a number of ligand molecules with higher predicted binding affinities than the given lead compound. All source codes are written in the ANSI C++ language. The entire set of programs has been tested on a RedHat Linux platform with Intel Xeon 5140 processors.

Structural Operation Module. This module generates new molecules by tailoring and transplanting appropriate fragments from other molecules onto the lead compound. In order to achieve this goal, a matching algorithm is used to determine if a pair of bonds on the given lead compound and another compound overlaps on the binding poses of the two molecules. This algorithm is similar to the one implemented in the BREED program

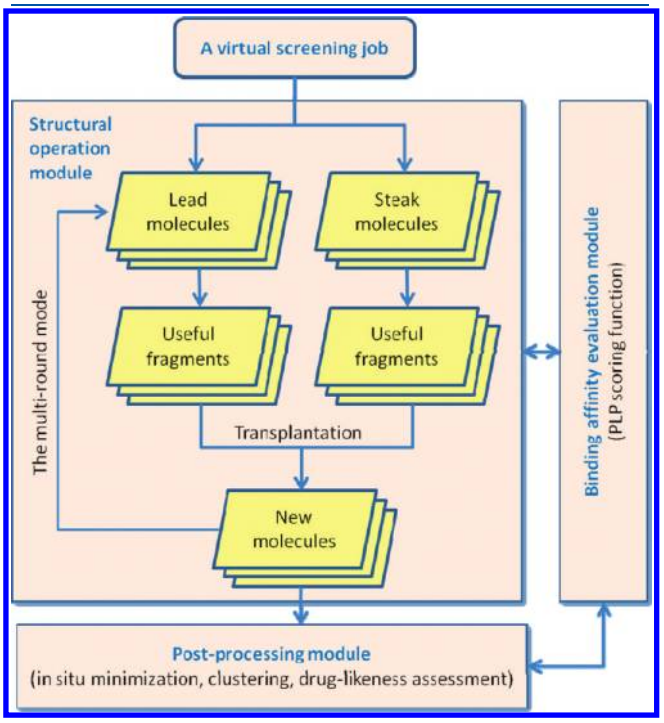


Figure 2. Overall flowchart of the AutoT&T program.

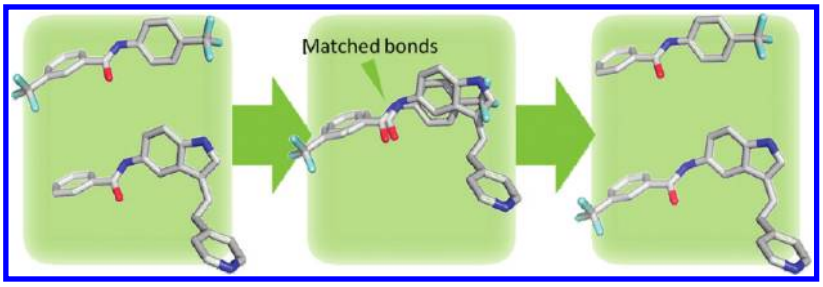


Figure 3. Illustration of how new structures are generated by swapping the fragments at both ends of the matched bonds. Note that if the two fragments at the same end of the matched bonds are identical, this operation will not produce new structures. Duplicate check is thus necessary at this step.

developed by Pierce et al.[18] A pair of matching bonds must meet the following criteria: First, the two bonds must belong to the same category, i.e., single, double, or triple, and must not be contained in a ring. This requirement simplifies subsequent structural operations. Second, both ends of two bonds must be within 1.0 Å from each other, and the centers of two bonds should be within 0.5 Å. Third, the angle between two bonds must be smaller than 15°. These cutoffs can be adjusted by the users. The matching degree (MD) between two bonds is then measured quantitatively as:

$$\mathrm{MD} = \frac{r_{\mathrm{cutoff}} - r_1}{r_{\mathrm{cutoff}}} \times \frac{r_{\mathrm{cutoff}} - r_2}{r_{\mathrm{cutoff}}} \times \frac{\theta_{\mathrm{cutoff}} - \theta}{\theta_{\mathrm{cutoff}}} \qquad \text{(eq. 1)}$$

In this equation, r_1 and r_2 are the distances between both ends of the two matched bonds; r_{cutoff} denotes the cutoff value of these distances; θ is the angle between the two matched bonds; and θ_{cutoff} denotes the cutoff value of this angle. The value of matching degree ranges between 0 and 1. A larger matching degree indicates a better overlapping of the two bonds under consideration.

If a pair of matched bonds can be found between the binding poses of the given lead compound and another molecule (a steak molecule), then the fragments at both sides of this bond on the latter can be used to replace the corresponding part on the lead compound. Note that not every pair of matched bonds is suitable for performing this kind of structural operation. If the fragment to be transplanted is identical to the one to be replaced, then obviously such an operation will not result in any new molecules. Thus, our program identifies such cases as invalid matched bonds. After filtering out such invalid matched bonds, the remaining ones are ranked by their matching degrees. All pairs of matched bonds are considered by their matching degrees in a decreasing order. For each pair of matched bonds, the two relevant molecules are tailored into four fragments at their matched bond. Then, the program checks if any steric bump exists between the fragment to be transplanted and the core structure of the lead compound. The default distance cutoff between a pair of heavy atoms is set as 2.0 Å. If no obvious steric bump is found, then the binding affinity of the fragment to be transplanted and the binding affinity of the fragment to be replaced on the lead compound is respectively computed by the binding affinity prediction module. The subsequent transplantation will be conducted only if the former is higher than the latter. Through transplantation, the selected fragment on the lead compound is removed, and a new chemical bond links the remaining segment of the lead compound and the fragment truncated from the selected steak molecule. This process is illustrated in Figure 3 with an example.

The structural transplantation described above generates a new molecule. It will be saved in a storage pool. Then, the next pair of matched bonds is examined to perform structural transplantation. Whenever a new molecule is generated, it will be compared to the existing molecules in the storage pool to avoid duplicates. In our program, this task is done by comparing the structural fingerprints of two molecules. The modified extended connectivity fingerprints (ECFP) is employed for this purpose.[19] This process is repeated until all pairs of matched bonds found between the given lead compound and the given compound library have been examined. Then, all newly generated molecules in the storage pool will be subjected to the postprocessing module for further analysis. If necessary, the AutoT&T program can be run in a multiround mode. In this mode, all molecules generated at the previous round will be taken in turn as the lead compounds to mate with the molecules in the given compound library. Newly generated molecules will again be added to the storage pool after duplicate check. The multiround mode will produce more diverse molecular structures, but it certainly consumes more computational resources.

Another two features related to the structural operations enabled by the AutoT&T program also need to be described here.

Random and Directional Optimization. By default, the structural operations enabled by our AutoT&T program are conducted in a random manner, i.e., structural transplantations occur at any valid pair of matched bonds to generate new structures. This option enables a more thorough exploration of the possible optimization schemes of the given lead compound. Nevertheless, in many cases of structure-based drug design, a structure−activity relationship has already been derived regarding the lead compound, and thus one has a relatively clear idea of the possible sites for further optimizations. In order to work in such cases, the AutoT&T program also allows the user to specify certain sites on the given lead compound for optimization, and AutoT&T will perform tailoring and transplanting only at these sites. Such directional optimizations produce the outcomes desired by the user more efficiently.

Consideration of Synthetic Feasibility. Our program builds molecular structures by combining chemical fragments. The resulting structures, however, are not always feasible for synthesis. In an attempt to address the problem of synthetic feasibility, the retrosynthetic combinatorial analysis procedure (RECAP) proposed by Lewell et al.[20] is implemented in AutoT&T. According to this scheme, a total of 11 types of chemical bonds are defined to be "breakable", each of which is related to a certain type of real organic reaction (Figure 4). Breaking or recombining chemical structures at these bonds may produce molecules that are more feasible for organic synthesis. In our program, if a pair of matched bonds belong to one of these 11 types, their matching degree will be adjusted to a higher value and thus gain priority in the subsequent structural operations. Note that some de novo design programs choose to address the synthetic feasibility issue

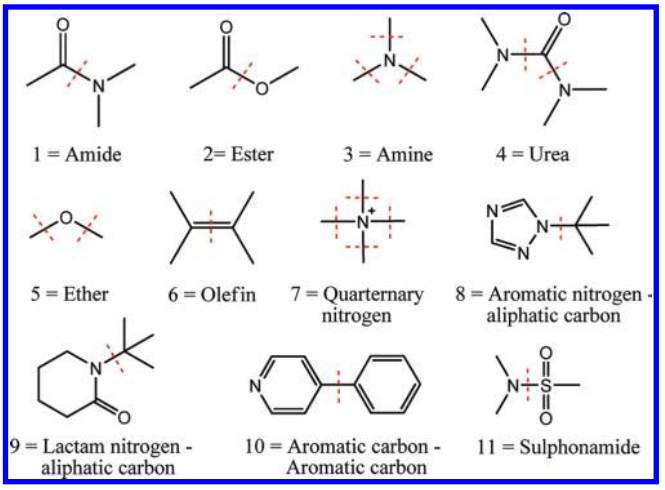


Figure 4. Eleven types of "breakable" chemical bonds defined in the RECAP scheme. This figure is cited from ref 20.

only at the postprocessing step; whereas our method addresses this issue with the RECAP scheme during structural operations. This practice is able to eliminate some unreasonable designs at an early stage.

Binding Affinity Calculation Module. The binding affinity prediction module works with the structural operation module. It is designed to guide lead compound optimization, aiming at achieving improved binding affinities. This module is called twice in each structural operation. Before a certain fragment on a steak molecule is considered for transplantation, its binding affinity to the target protein is computed. Only the fragments with higher binding affinities than that of the fragment to be replaced on the lead compound will be actually selected to be transplanted. After the transplantation is completed, the overall binding affinity of the newly generated molecule will be computed again. Only the molecules with improved binding affinities over the lead compound will be accepted and recorded in the storage pool.

The desired scoring method to be implemented in AutoT&T needs to achieve a compromise between accuracy and speed since binding affinity calculation needs to be conducted hundreds of thousands times in a single job. Among the available methods for binding affinity calculation, only scoring functions[21,22] can fulfill this demand. Scoring functions have been widely applied in structure-based drug design since 1990s. In fact, some studies[23,24] suggest that scoring functions are not necessarily less accurate in binding affinity prediction than other computationally expensive methods. In our work, we chose to implement the piecewise linear potential (PLP) scoring function[25] in the AutoT&T program for calculating protein−ligand binding affinities. PLP is an empirical scoring function, which sums up distance-dependent potentials of the atom pairs between a protein and a ligand. Details of this scoring function can be found in the Supporting Information of this article. Some benchmarks of scoring functions[26,27] suggest that PLP is among the top ones in terms of general performance. Besides, it is convenient to use PLP to compute the binding score of any fragment, since PLP is based on pure pairwise potentials. As explained above, this feature is very important for the function of AutoT&T.

Postprocessing Module. The postprocessing module conducts a series of operations and analyses on the outcomes of the structural operation module, including in situ energy minimization, binding affinity calculation, structural clustering, and drug-likeness assessment.

In situ Energy Minimization. The AutoT&T program generates new molecular structures by combining fragments. Although distance and angular cutoffs are defined to enforce overlapping at the joint point, the chemical bond used to link two fragments is not always in the standard bond length and bond angle. Therefore, it is necessary to adjust the structures of the resulting new molecule through energy minimization. In our program, energy minimizations of the ligand structures are performed within the geometrical constraints of the binding pocket on the target protein. During energy minimization, the protein structure is kept fixed. The AMBER force field[28] and the Tripos force field[29] are implemented in AutoT&T for this purpose. The users may choose among the Simplex, Powell, steepest descent or conjugate gradient algorithms as the minimization method. Since the binding poses of both the given lead compound and the steak molecule are assumed to be derived by a reliable molecular docking program, the new molecular structure only needs minor adjustments. Therefore, it is by default minimized with the Simplex method for 100 cycles with the Tripos force field. After this in situ energy minimization step, the binding affinity calculation module gives the final binding score of each new ligand molecule. The user may specify the number of top-ranked molecules for further consideration.

Structural Clustering. If the AutoT&T program generates a large number of molecules as final outputs, then it will be quite impractical to manually examine them one by one. In such a case, structural clustering will help the user select the representative molecules for visual inspection. It also serves as an objective measurement of the structural diversity among the generated molecules. Two clustering methods are implemented in our program: One is the K-means method[30] and the other is a modified stochastic clustering method.[31] Both clustering methods start from a similarity matrix containing all molecules under consideration. The similarity index between any two molecules is derived based on topological torsion descriptors[32] or atom pair descriptors.[33] According to the K-means method, a number of the most dissimilar molecules are selected as the initial clustering centers. Then, the remaining molecules are allocated into different clusters according to their similarities to the cluster centers. In the modified stochastic cluster analysis method, the pair of molecules with the lowest similarity is chosen as the initial clustering centers. Then, according to a user-specified similarity cutoff, the clustering centers will be extended to those molecules with lower similarities to the existing clustering centers than the cutoff value. Finally, the remaining molecules are allocated into different clusters with the same algorithm as the K-means method. Once the clustering process is completed, the user can easily pick out the most promising one in each cluster according to its binding score or on other criteria.

Drug-Likeness Assessment. To become a potential drug candidate, a compound must have suitable absorption, distribution, metabolism, excretion, and toxicity (ADMET) properties besides a tight binding to the target protein. Some rule-based evaluations of "drug-likeness" are also implemented in the postprocessing module to address this issue. The descriptors considered by us are very similar to those in the popular Lipinski's "rule-of-five",[34] including molecular weight, number of heavy atoms, number of hydrogen-bond donors and acceptors, total number of hydrogen-bond donors and acceptors, number of rotatable bonds, and the octanol−water partition coefficient

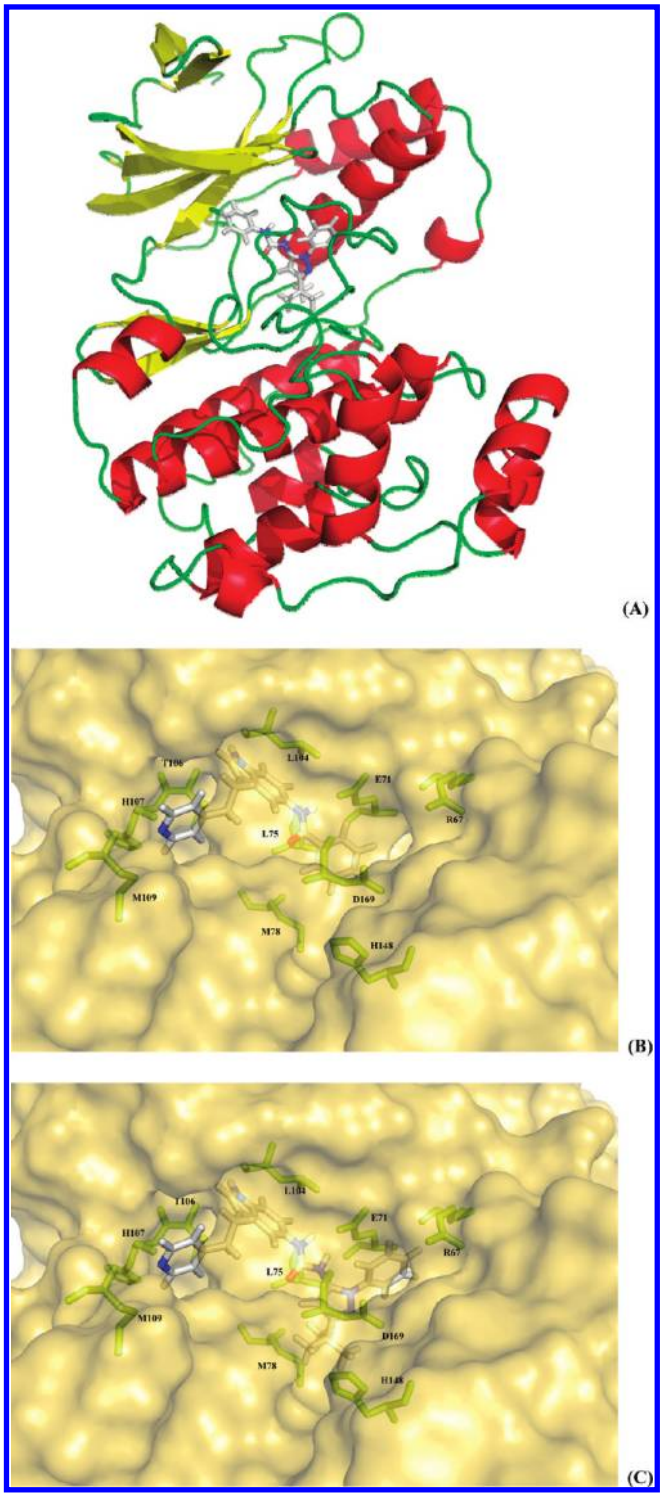


Figure 5. (A) Three-dimensional structure of a p38 MAP kinase complex (PDB entry 1W82). (B) Binding mode of the lead compound derived by molecular docking. (C) Binding mode of an optimized structure generated by AutoT&T (molecule A1-1 in Table 1).

(log *P*). All of these descriptors except the log *P* value can be derived directly from molecular structures. The log *P* value in our program is calculated with the XLOGP3 method.[35] For each descriptor, the user can set a lower cutoff and a higher cutoff. If any of these descriptors of a given molecule falls out of the acceptable range, then this molecule will be labeled as undesired. These filters are employed to eliminate the obviously unreasonable ones among all output molecules.

It should be mentioned that the postprocessing module is an auxiliary module of AutoT&T. Other external methods certainly can be employed to perform the same functions as the postprocessing module, including in situ minimization, rescoring, clustering, and drug-likeness assessment. This is actually a more flexible option for the users.

RESULTS: TEST CASES

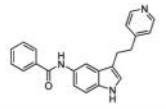
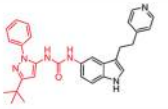
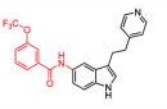
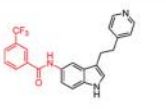
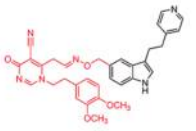
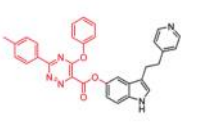
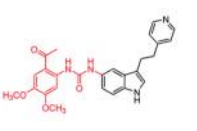
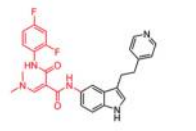
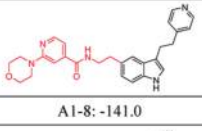
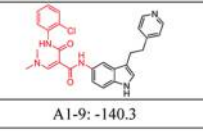
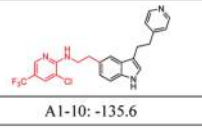
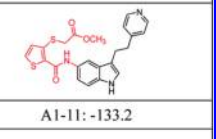
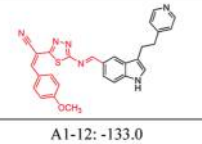
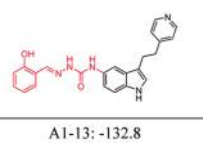
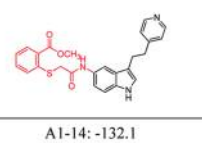
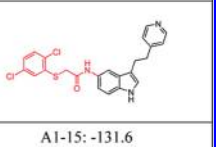
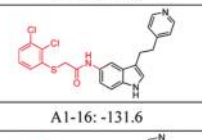
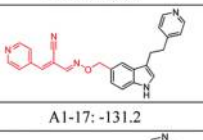
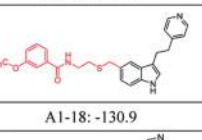
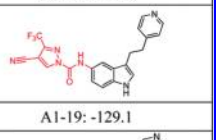
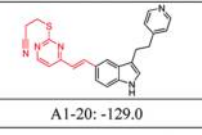
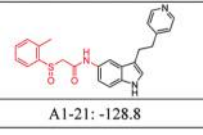
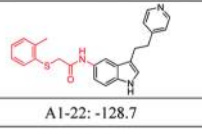
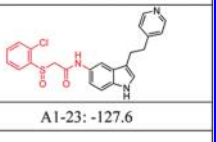
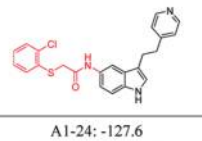
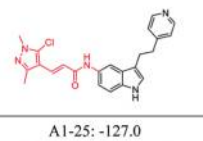
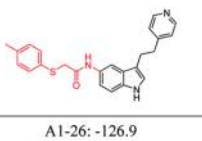
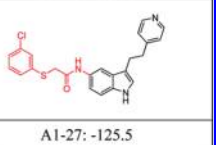
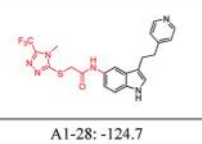
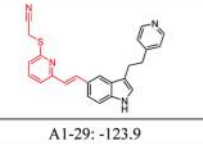
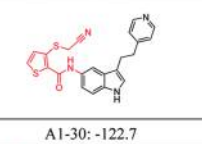
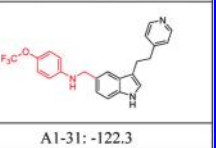
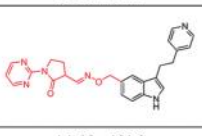
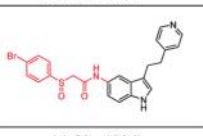
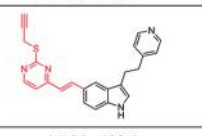
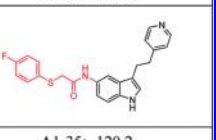
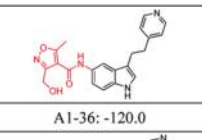
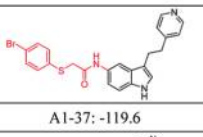
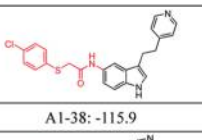
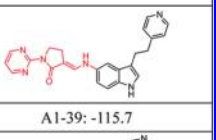
The AutoT&T program was tested in three cases in this study. In the first case, p38 mitogen-activated protein (MAP) kinase was the target protein. New designs were generated by installing suitable fragments onto the core structure of a known p38 MAP kinase inhibitor. In the second case, peroxisome proliferator-activated receptor-α (PPAR-α) was the target protein. New designs of PPAR-α inhibitors were generated by joining two desired fragments on a selected lead compound with some appropriate chemical linkers. Since there are some known inhibitors for both p38 MAP kinase and PPAR-α, these two test cases were chosen mainly for validating the effectiveness of AutoT&T in a retrospective way. In the third case, the Mcl-1 protein, which antagonizes apoptosis process through protein−protein interactions, was chosen as the molecular target, and novel small-molecule binders to Mcl-1 were generated from scratch by using AutoT&T. The last test case demonstrates how AutoT&T can be applied in lead discovery when few lead compounds are known.

Case 1: p38 Mitogen-Activated Protein (MAP) Kinases. The p38 MAP kinases (Figure 5A) are a class of mitogen-activated protein kinases which respond to stress stimuli, such as cytokines, ultraviolet irradiation, heat shock, and osmotic shock. They are involved in cell differentiation and apoptosis and have been well studied as potential drug targets.[36] A variety of p38 MAP kinase inhibitors have been publicly reported in literature.[37−40] Here, we selected one of them (Table 1, IC_{50} = 40 μM) as the lead compound for optimization. This particular compound was selected mainly for two reasons. First, its size is relatively small, and its potency is only at the μM range. Thus, there is still enough space left for further developments. Second, some known p38 MAP kinase inhibitors consist of the same core structure as this lead compound. Thus, if AutoT&T can successfully reproduce them, its effectiveness will be validated in this case.

Since the complex structure formed by p38 MAP kinase and this compound is not available, we used GLIDE,[41] a popular molecular docking program, to derive the binding mode of this compound (Figure 5B). After analyzing the interactions between the lead compound and p38 MAP kinase as well as the shape of the binding pocket, we decided to keep the segment containing the indole moiety and replace the rest part in optimization (Table 1). The lead compound structure was annotated accordingly. Then, a total of 1000 molecules were randomly selected from the MDL Available Chemical Directory (ACD) database as the steak molecules. They were docked onto the same p38 MAP kinase structure with the GLIDE program. AutoT&T finally produced a total of 44 novel structures in the single-round mode. This job took less than one minute on a RedHat Linux desktop workstation with dual Intel Xeon-5140 CPUs at 2.33 GHz.

Some of the top-ranked molecules are listed in Table 1. The structural diversity presented in these structures is quite impressive

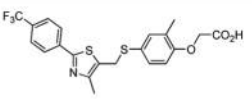
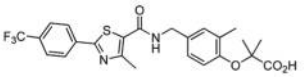
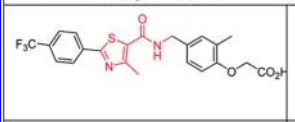
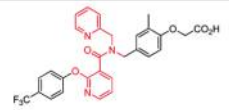
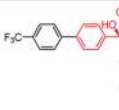
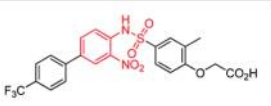
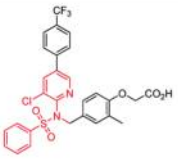
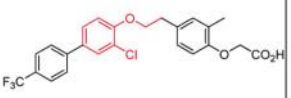
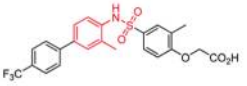
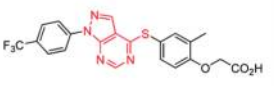
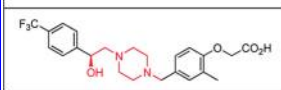
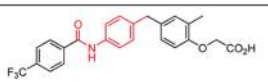
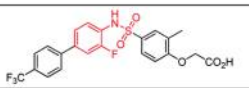
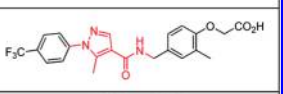
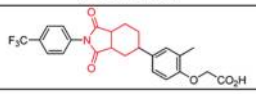
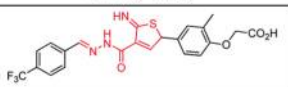
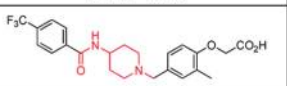
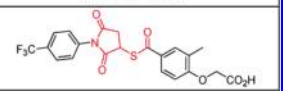
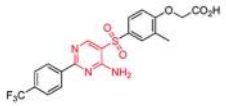
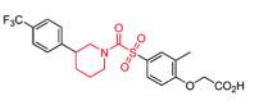
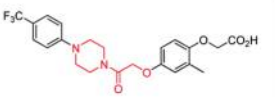
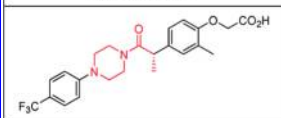
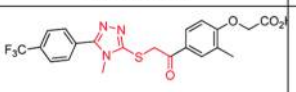
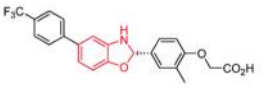
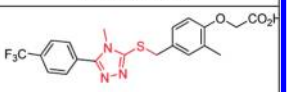
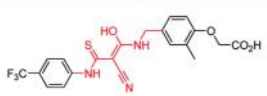
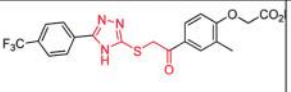
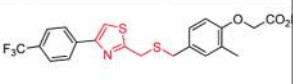
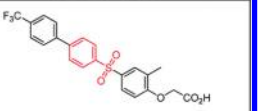
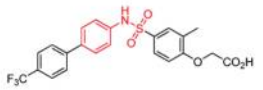
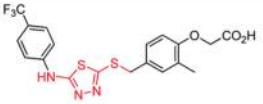
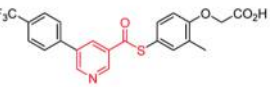
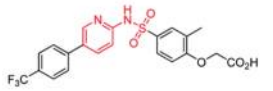
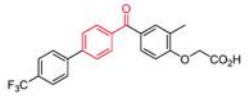
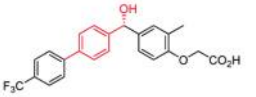
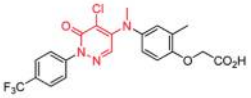
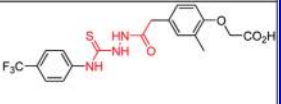
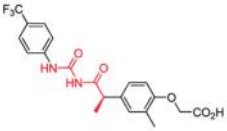
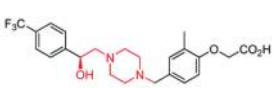
Table 1. Optimization of p38 MAP Kinase Inhibitors: Lead Compound and Optimized Structures Generated by AutoT&T[a]

			
The lead compound (PLP score = -108.7, IC_{50} = 40 μM)	A1-1: -165.2 (IC_{50} = 145 nM)	A1-2: -133.9 (IC_{50} = 2 μM)	A1-3: -129.0 (IC_{50} = 4 μM)
			
A1-4: -156.9	A1-5: -152.4	A1-6: -151.2	A1-7: -144.9
			
A1-8: -141.0	A1-9: -140.3	A1-10: -135.6	A1-11: -133.2
			
A1-12: -133.0	A1-13: -132.8	A1-14: -132.1	A1-15: -131.6
			
A1-16: -131.6	A1-17: -131.2	A1-18: -130.9	A1-19: -129.1
			
A1-20: -129.0	A1-21: -128.8	A1-22: -128.7	A1-23: -127.6
			
A1-24: -127.6	A1-25: -127.0	A1-26: -126.9	A1-27: -125.5
			
A1-28: -124.7	A1-29: -123.9	A1-30: -122.7	A1-31: -122.3
			
A1-32: -121.3	A1-33: -120.7	A1-34: -120.6	A1-35: -120.2
			
A1-36: -120.0	A1-37: -119.6	A1-38: -115.9	A1-39: -115.7
A1-40: -115.6	A1-41: -114.4	A1-42: -114.2	A1-43: -113.6
A1-44: -110.5			

[a] In each structure, the fragments in black are reserved during optimization; whereas those in red are transplanted from steak molecules. The number below each structure is its binding score computed with PLP. Lower PLP scores indicate higher binding affinities.

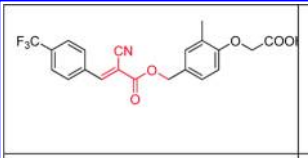
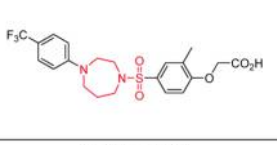
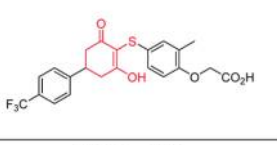
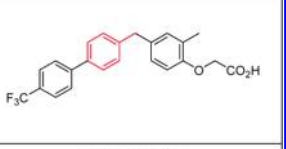
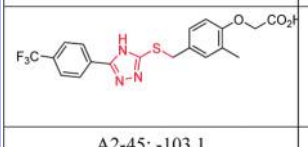
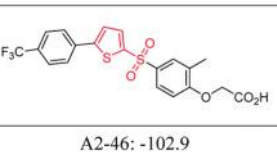
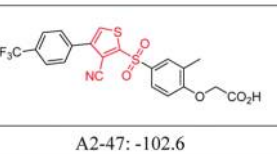
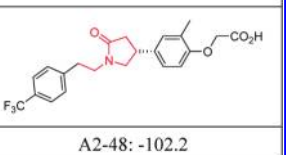
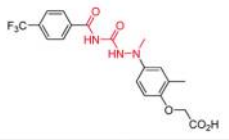
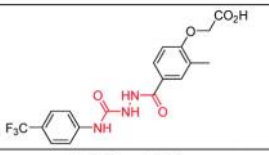
dx.doi.org/10.1021/ci200036m |J. Chem. Inf. Model. 2011, 51, 1474–1491

Table 2. Optimization of PPAR-α Inhibitors: Lead Compound and Optimized Structures Generated by AutoT&T[a]

 The lead compound (GW501516, EC_{50} =1 μM, PLP score = -100.6)		 The target compound (GW590735, EC_{50} = 4 nM, PLP score = -112.6)	
			
A2-1: -113.8	A2-2: -132.6	A2-3: -122.4	A2-4: -119.9
			
A2-5: -117.7	A2-6: -116.6	A2-7: -116.0	A2-8: -115.3
			
A2-9: -113.8	A2-10: -113.3	A2-11: -112.9	A2-12: -112.7
			
A2-13: -112.3	A2-14: -112.1	A2-15: -111.8	A2-16: -111.6
			
A2-17: -110.2	A2-18: -110.1	A2-19: -110.0	A2-20: -110.0
			
A2-21: -109.4	A2-22: -109.1	A2-23: -108.3	A2-24: -108.1
			
A2-25: -107.9	A2-26: -107.7	A2-27: -107.4	A2-28: -107.3
			
A2-29: -107.1	A2-30: -107.0	A2-31: -106.9	A2-32: -106.8
			
A2-33: -106.0	A2-34: -105.2	A2-35: -105.1	A2-36: -105.0
			
A2-37: -104.8	A2-38: -104.7	A2-39: -104.3	A2-40: -104.1

dx.doi.org/10.1021/ci200036m |J. Chem. Inf. Model. 2011, 51, 1474–1491

Table 2. Continued

			
A2-41: -103.7	A2-42: -103.6	A2-43: -103.6	A2-44: -103.5
			
A2-45: -103.1	A2-46: -102.9	A2-47: -102.6	A2-48: -102.2
			
A2-49: -101.5	A2-50: -101.3		

[a] In each structure, the fragments in black are reserved during optimization; whereas those in red are transplanted from steak molecules. The number below each structure is its binding score computed with PLP. Lower PLP scores indicate higher binding affinities.

considering that only 1000 randomly selected molecules were used as the steak molecules in this trial. After examining public literature and filed patents, we found among the outcomes of AutoT&T three known p38 MAP kinase inhibitors (i.e., molecules A1-1, A1-2, and A1-3 in Table 1). All of them indeed have better potency than the lead compound. In addition, it is encouraging to observe that the binding scores computed by the PLP scoring functions are in a good correlation with the experimentally measured biological potency of molecules A1-1, A1-2, and A1-3 and the lead compound. The most potent one among them has an IC_{50} value lower than that of the lead compound by 300 fold. This molecule fits the binding pocket on p38 MAP kinase much better than the lead compound according to the binding mode generated by AutoT&T (Figure 5C).

Case 2: Peroxisome Proliferator-Activated Receptors (PPARs). PPARs are a group of nuclear receptor proteins that function as transcription factors regulating the expression of genes. They play an essential role in the regulation of cellular differentiation development, metabolism, and tumorigenesis of higher organisms.[42] We noticed a study by Sierra et al.[43] who reported a selective antagonist to PPAR-δ (GW501516). This compound inhibits PPAR-δ with an EC_{50} around 1 nM, whereas its EC_{50} against PPAR-α is only about 1 μM. Sierra et al. then developed this compound through a conventional medicinal chemistry approach and finally obtained a potent compound (GW590735) with an EC_{50} of 4 nM against PPAR-α and a 500-fold selectivity versus PPAR-γ and PPAR-δ. This is a good example of lead optimization. We used it in our study to test if AutoT&T is able to realize this process in a virtual manner.

Accordingly, compounds GW501516 and GW590735 were chosen as the lead and the target compounds, respectively (Table 2). Crystal structure of the target compound in complex with PPAR-α has been resolved (Figure 6A).[43] Analysis of the interactions between this compound and PPAR-α indicates that the *p*-trifluoromethyl phenyl and the carboxyl acid terminal moieties on this molecule occupy two separate subpockets at the binding site (Figure 6B). The acidic moiety has interactions with residues Ser280, Tyr314, His440, and Tyr464; whereas the *p*-trifluoromethyl phenyl moiety fills up a hydrophobic subpocket formed by residues Ile241, Leu247, Leu254, Val255, and Val332. Since the major difference between the lead compound and this target compound is the linker moiety in the middle, our optimization plan in this trial was to keep the two important terminal moieties on the lead compound and replace the linker between them. We then examined the outcomes of AutoT&T to see if any designs close to the target compound could be found.

Although the binding mode of the lead compound is unknown, it can be reasonably derived from that of the target compound. In fact, the binding mode of the lead compound predicted by the GLIDE program resembles that of the target compound in the crystal structure very closely (Figure 6B). Binding modes of over 60 000 small-molecule compounds in the MayBridge database were also derived by using the GLIDE program. They were then supplied to the AutoT&T program as inputs. Since the linker in the middle is desired to connect two separate seed structures, two pairs of matched bonds are required simultaneously to perform appropriate structural transplantation. Finally, a total of 50 new structures were generated by linking the two given seed structures with suitable fragments truncated from the steak molecules (Table 2). Again, a remarkable structural diversity was observed among the solutions produced by AutoT&T. In particular, molecule A2-1 in Table 2 is almost identical to the target compound with two missing methyl groups beside the terminal carboxyl group. These two methyl groups are missing simply because they are not on the seed structure selected as input. The binding mode of this molecule derived by AutoT&T is illustrated in Figure 6C, which is very close to the experimentally resolved binding mode of the target compound. This indicates that AutoT&T is able to generate ligand molecules in reasonable bound conformations.

Generating new structures by linking multiple fragments placed at different sites is a more challenging task than the one in the first test case. This test case demonstrates that the

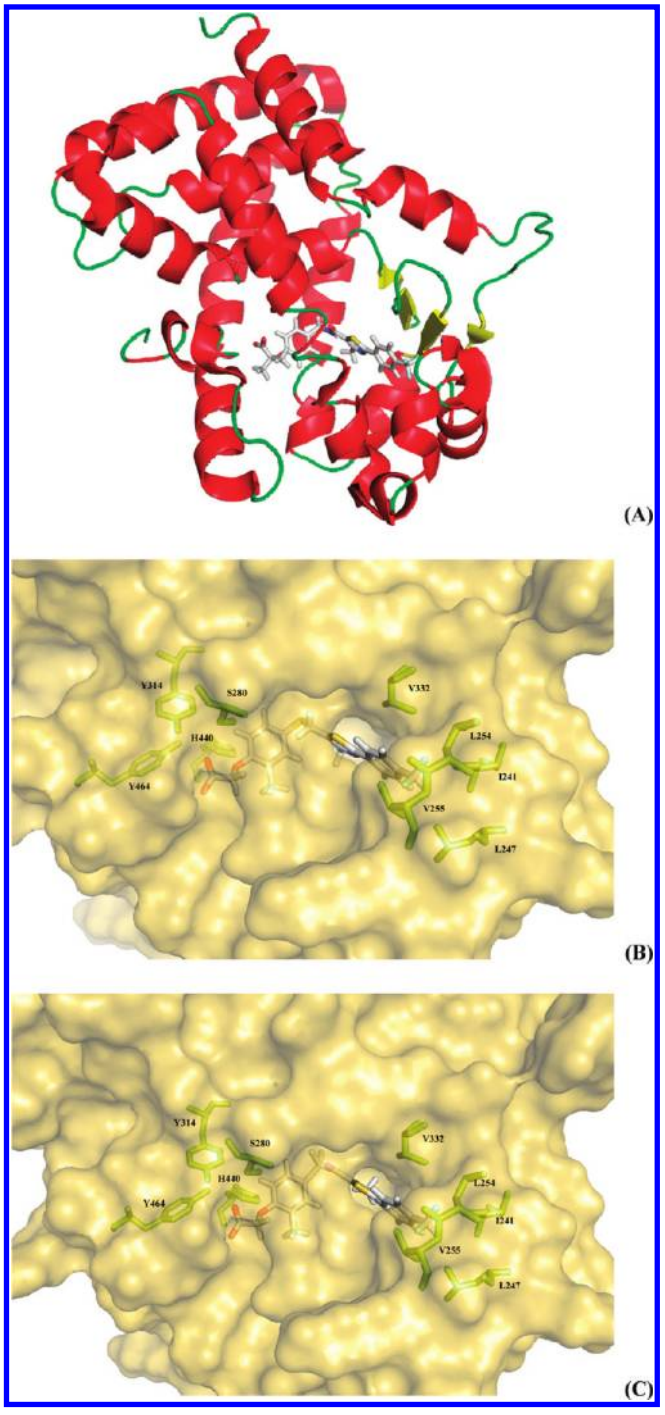


Figure 6. (A) Three-dimensional structure of PPAR-α in complex with a known potent agonist (PDB entry 2P54). (B) Binding mode of the lead compound derived by molecular docking. (C) Binding mode of an optimized structure generated by AutoT&T (molecule A2-1 in Table 2).

AutoT&T program is also very capable for this task. Interestingly, AutoT&T actually produced a number of novel molecules with comparable or even better binding scores than molecule A2-1, which seemed to be reasonable as well. It took the AutoT&T program only about 20 min on the same Linux desktop workstation to complete this job, although a significantly larger number of steak molecules were considered in this case.

Case 3: Design of Small-Molecule Inhibitors of the Mcl-1 Protein. Mcl-1 is a member of the Bcl-2 protein family which plays an important role in regulating the balance between cell proliferation and apoptosis.[44,45] It thus has been received extensive attention in recent years as a potential target for developing new anticancer therapies. Mcl-1 executes its biological function by antagonizing other pro-apoptotic proteins (Figure 7A). Thus, it is relatively difficult to design effective small-molecule inhibitors of Mcl-1 since such compounds are desired to disrupt protein−protein interactions. So far there are only a few known small-molecule inhibitors targeting at Mcl-1.[46] In this trial, we aimed at designing some novel small-molecule Mcl-1 inhibitors without referring to any known Mcl-1 inhibitors.

In a previous work, we performed virtual screening of the MayBridge database (>60 000 compounds) using the GOLD program[47] in an attempt to discover some plausible binders to Mcl-1. The binding site on Mcl-1 consists of two distinctive subpockets: one is formed by residues Met231, Val249, Met250, Val253, Leu267, and Phe270, and the other is formed by residues Val216, Val220, Val265, and Phe319 (Figure 7B). Among the 100 top-ranked hits selected by the GOLD program, we noticed one compound which has a suitable 3-pyridinylurea moiety crossing over the ridge between the two subpockets mentioned above (Table 3). Besides, the urea group on this moiety may form a hydrogen bond with a nearby Thr266 residue (Figure 7B). But the rest parts of this compound do not fit the binding site on Mcl-1 well. Assume that the chemical structure of this compound is represented as A−B−C, in which A is the *p*-*t*-butylphenyl moiety at the left end, B is the 3-pyridinylurea moiety in the middle, and C is the benzothiazole moiety at the right end. In this trial, we chose moiety B as the starting structure and optimized the branches at both ends (A and C) to design more effective binders to Mcl-1.

In this case, the molecule shown in Figure 7B was used as the lead compound; while the >60 000 compounds in the MayBridge database processed in a previous virtual screening task were supplied as the steak molecules. After running the AutoT&T program in the single-round mode, we obtained a number of molecules with either A′−B−C or A−B−C′ structures. In order to obtain structures with new replacements at both ends, i.e., A′−B−C″ or A″−B−C′, the AutoT&T program was further run in the multiround mode in which two rounds of tailoring and transplanting were performed. In fact, the linker in the middle (B) can also be changed to some other fragments, since the structures generated in the first round are all considered as the starting structures in the next round of structural operations. Thus, we finally obtained A′−B″−C″, A″−B″−C′ and other possible combinations. A maximum of 2000 new structures were retained in each round. The final resulting structures were filtered using a set of drug-likeness rules, including (i) molecular weights falling in the range of 250−750; (ii) number of hydrogen-bond donors <5; (iii) number of hydrogen-bond acceptors <10; (iv) number of rotatable bonds <10; (v) number of rings <5; and (vi) computed log *P* values falling in the range of −2.0 to 6.5. After filtering, a total of 514 new structures were output. The entire process took about 1.5 h on the same Linux desktop workstation used in the other two test cases.

Among the final outcomes, 28 molecules have different fragments at both ends as compared to the lead compound (Table 3). All of these new molecules are significantly better than the lead

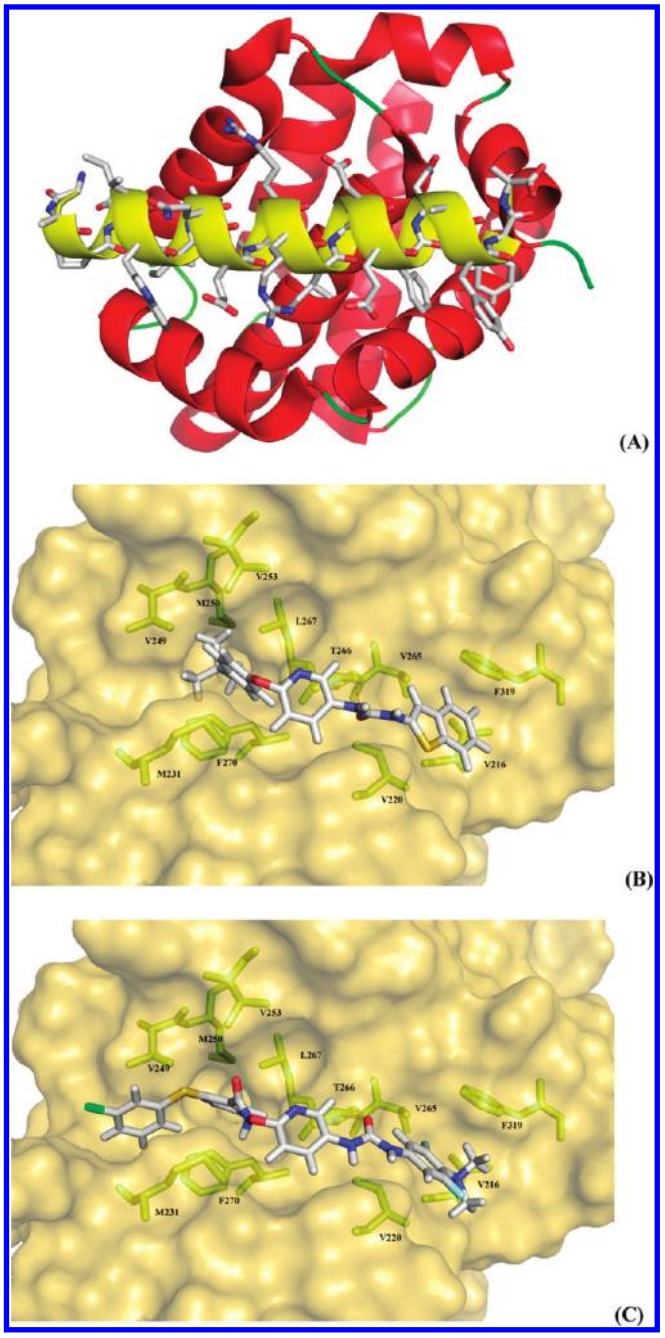


Figure 7. (A) Three-dimensional structure of Mcl-1 in complex with the Bim peptide (PDB entry: 2PQK). (B) Binding mode of an interesting hit selected from virtual screening outcomes. (C) Binding mode of an optimized structure generated by AutoT&T (molecule A3-1 in Table 3).

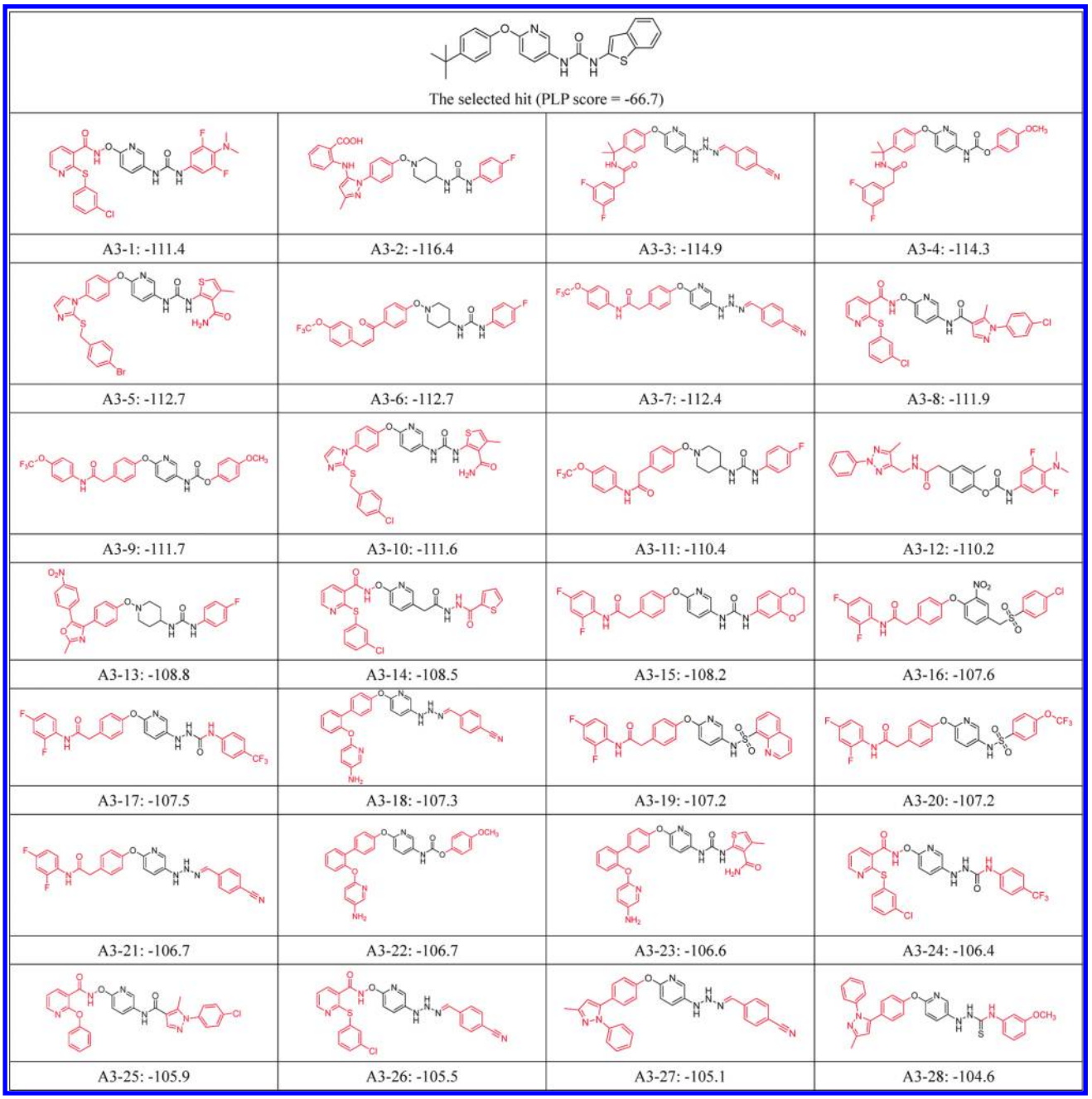
compound in terms of binding scores. Although the biological activities of these molecules need to be eventually verified in experiments, at least they seem to be promising and thus fulfill the goal of molecular design. A representative design (molecule A3-1 in Table 3) is illustrated in Figure 7C. One can see that this molecule fits into the binding site on Mcl-1 nicely according to the binding mode predicted by AutoT&T. Both ends of this molecule are essentially lipophilic, which match the environment of the two noted subpockets. This test case demonstrates how the AutoT&T program can be applied to lead discovery for a target protein with few known lead compounds.

Comparison with Two Other De Novo Design Methods. Conventional de novo design methods are also widely applied to lead optimization. A head-to-head comparison of the performance of AutoT&T with standard de novo design methods will be intriguing. For this purpose, we tested two popular de novo design methods available in commercial software, i.e., LUDI[14] and EA-Inventor,[15] in the same three test cases. In all three cases, the same lead compounds used for running AutoT&T were supplied as the inputs for running LUDI and EA-Inventor. The final solutions generated by both programs were also rescored with the PLP scoring function implemented in the Discovery Studio software (version 2.5, released by Accelrys Inc.) so that they could be compared to the solutions generated by AutoT&T directly. Among the solutions generated by each program, only those with binding scores better than the given lead compound were considered as qualified solutions. The performance of each program was then evaluated in two aspects: the total number of the qualified solutions obtained in each test case, and the average binding scores of the qualified solutions.

The LUDI method considered in our study is implemented in the Discovery Studio software as the "LUDI Link" module. An internal library of 1166 small chemical fragments is used by this implementation. Several main parameters were set as follows in our study: (i) The radius of the site sphere was 10 Å; (ii) the cutoff value of the maximal root-mean-square deviation (RMSD) from maximal protein−ligand interactions was 1.0 Å; and (iii) the bond rotation mode was set as "one at a time". All other parameters were set to their default values. In all three test cases, the starting structures were exactly the same as those used for testing AutoT&T (Tables 1−3). Note that the case of p38 MAP kinase actually tested the "growing" mode of LUDI, whereas the case of PPAR-α tested the "linking" mode of LUDI. The case of Mcl-1 again tested the "growing" mode of LUDI, in which fragment growing was supposed to occur at two sites. Unfortunately, the implementation of LUDI tested in our study does not allow fragment growing to occur simultaneously at two sites. Therefore, we ran LUDI in two consecutive rounds: the first round of fragment growing occurred at one end of the seed structure; whereas the second round of fragment growing occurred at the other end. The outcomes of the first round operation were used as the inputs for the second round. In all three cases, binding scores of the final solutions generated by LUDI were recomputed with the PLP scoring function.

In the case of p38 MAP kinase, LUDI generated a total of 31 solutions, among which 8 of them had better binding scores than the lead compound (Table 4). Apparently, LUDI produced fewer qualified solutions than AutoT&T (8 versus 44) in this case. Note that LUDI failed to reproduce any known p38 MAP kinase inhibitors based on the given seed structure. Furthermore, the binding scores of LUDI solutions were generally less promising than those of AutoT&T solutions. The best binding score among LUDI solutions was −115.8 (molecule L1-1 in Table 4), which was far below the best one among AutoT&T solutions (molecule A1-1 in Table 1). In the second case of PPAR-α, which tested the "linking" mode, LUDI failed to produce any solutions that were able to link the two separate fragments on the given seed structure. In the third case of Mcl-1, LUDI generated a total of 10 qualified solutions (Table 4). In comparison, AutoT&T generated 28 (Table 3). Again, the binding scores of LUDI solutions were generally less promising than those of AutoT&T solutions. Besides binding scores, the structural diversity in the computer-generated designs is also

Table 3. Design of Mcl-1 InhibitorS from Scratch: Starting Structure and Final Designs Generated by AutoT&T[a]

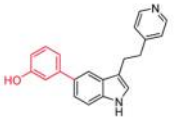
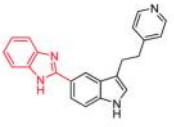
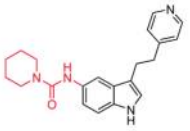
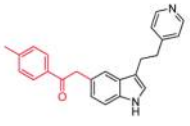
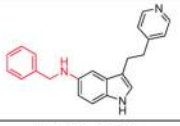
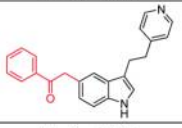
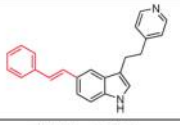
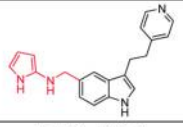
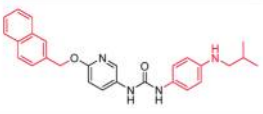
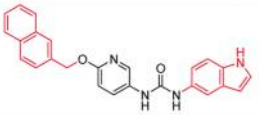
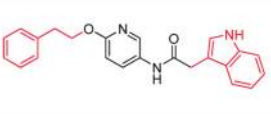
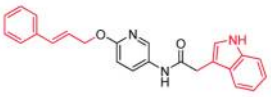
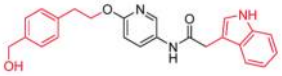
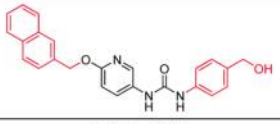
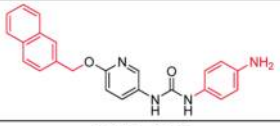
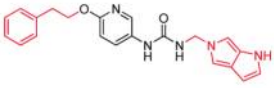
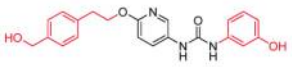
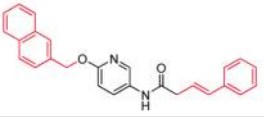


[a] In each structure, the fragments in red fit into the two subpockets in the binding site on Mcl-1; whereas those in back are the linkers crossing over the ridge between the two subpockets. The number below each structure is its binding score computed with PLP. Lower PLP scores indicate higher binding affinities.

important. In this aspect, LUDI solutions are apparently less diverse than AutoT&T solutions in all test cases. The new fragments installed on the seed structure by LUDI are basically some simple cyclic moieties without any substituent groups. This observation indicates the limitation of using a predefined fragment library in structure construction, although the fragment library used by LUDI is fairly large.

The EA-Inventor method considered in our study was implemented in the SYBYL software (version 8.1, released by Tripos Inc.). EA-Inventor employs an evolutionary process to generate desired ligand molecules for a given target protein. Structure generations also start on a given seed structure using an internal building block library containing over 1300 fragments derived from the MDL Drug Data Report (MDDR) database. New structures are passed to Surflex-Dock, which is another module in SYBYL, to be docked to the target protein and compute binding scores. Only the structures with improved binding scores will remain in the final outcomes. Other structures will be considered in the second generation of structure operations for further optimization. This process will be repeated until

Table 4. New Designs Generated by LUDI in the Three Test Cases

p38 MAP kinase inhibitors			
			
L1-1: -115.8	L1-2: -114.6	L1-3: -112.8	L1-4: -111.1
			
L1-5: -110.9	L1-6: -110.4	L1-7: -110.1	L1-8: -109.0
PPAR-α inhibitors: no valid designs			
Mcl-1 inhibitors			
			
L3-1: -76.0	L3-2: -75.0	L3-3: -73.2	L3-4: -71.5
			
L3-5: -70.9	L3-6: -69.7	L3-7: -69.5	L3-8: -68.6
			
L3-9: -68.2	L3-10: -68.2		

[a] In each structure, the fragments in black are reserved during optimization; whereas those in red are newly added fragments. The number below each structure is its binding score computed with PLP. Lower PLP scores indicate higher binding affinities.

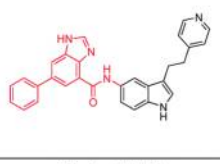
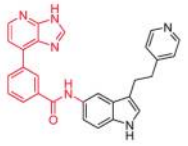
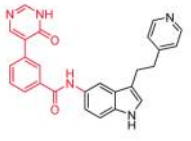
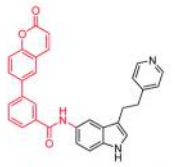
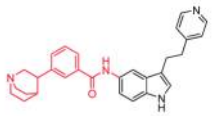
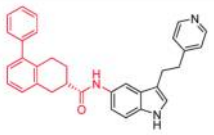
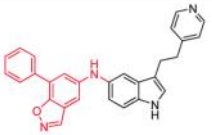
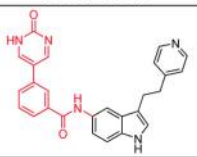
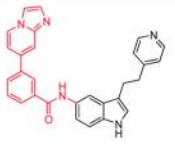
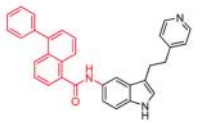
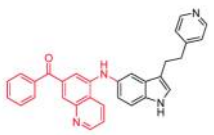
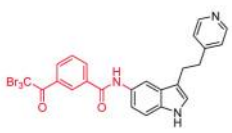
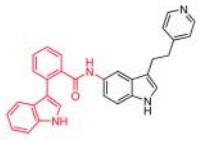
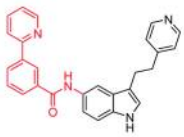
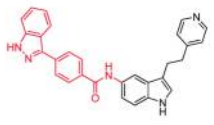
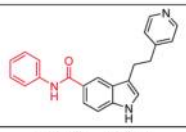
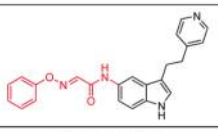
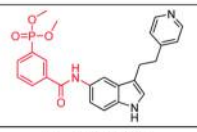
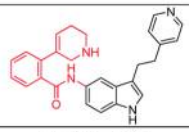
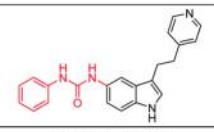
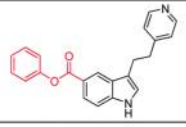
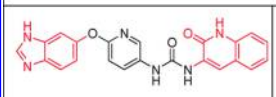
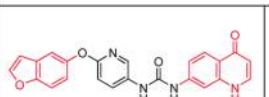
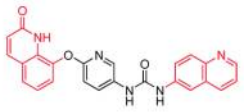
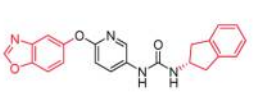
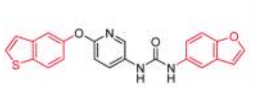
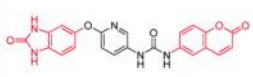
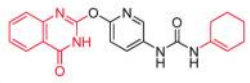
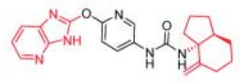
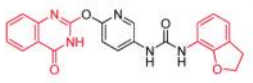
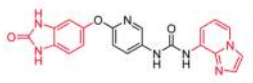
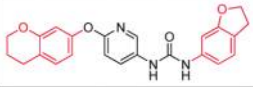
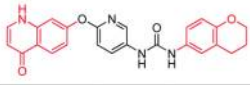
a user-specified generation number has been reached. In our study, desired size of each generation was set as 50. The seed for the pseudorandom number generator was set as 0. Other parameters were assigned the default settings. Note that in the case of Mcl-1, just as LUDI, the optimization process was also divided into two consecutive steps: At the first step, fragment growing occurred at one end of the seed structure, whereas the second round of fragment growing occurred at the other end. In all three cases, binding scores of the final solutions generated by EA-Inventor were also computed with the PLP scoring function implemented in the Discovery Studio software.

In the case of p38 MAP kinase, EA-Inventor generated a total of 21 qualified solutions (Table 5), i.e., those with PLP binding scores better than the lead compound. Apparently, EA-Inventor was more successful than LUDI in this case. It should be attributed to the more sophisticated fragment library used by EA-Inventor. Among the solutions generated by EA-Inventor, one can observe a few that are close to the known p38 MAP kinase inhibitors sharing the same starting structure, such as E1-12 and E1-20 in Table 5. The binding scores of EA-Inventor solutions were generally comparable to those of AutoT&T solutions (Table 1). In the second case of PPAR-α, just like LUDI, EA-Inventor failed to produce any solutions that are able to link the two separate fragments in the given seed structure. In the third case of Mcl-1, EA-Inventor produced a total of 12 qualified solutions (Table 5), much fewer than the solutions produced by AutoT&T (12 versus 28) in the same case. In addition, it seems that these designed molecules do not really fit the binding site on Mcl-1 well. Even the best binding score among EA-Inventor solutions, i.e., −84.7 (E3-1 in Table 5), was lower than the least promising one among AutoT&T solutions, i.e., −104.6 (A3-28 in Table 3). Running EA-Inventor in the multiple-generation mode may produce more promising designs in this case. Unfortunately, it is technically troublesome to set up such a job for EA-Inventor and thus was not attempted in our study. We thus conclude that the performance of EA-Inventor is less robust than AutoT&T in the latter two test cases.

DISCUSSION

Fragment Transplantation: An Old Dog's New Trick. With the rapid progress in structural genomics, structure-based design has become a main-stream approach in drug design and discovery. Continuous innovation of relevant methods is a driving force for pushing this approach forward. Our AutoT&T method optimizes the chemical structure of a given lead compound by replacing certain parts on it with more promising fragments. This has been one of the essential methods for lead optimization. It is also similar in concept to "scaffold hopping".[48,49] The value of our AutoT&T method is that it automates this process with

Table 5. New Designs Generated by EA-Inventor in the Three Test Cases

p38 MAP kinase inhibitors				
				
E1-1: -141.7	E1-2: -141.0	E1-3: -139.7	E1-4: -134.8	E1-5: -130.5
				
E1-6: -130.0	E1-7: -129.4	E1-8: -127.1	E1-9: -124.6	E1-10: -124.2
				
E1-11: -122.3	E1-12: -122.0	E1-13: -121.0	E1-14: -119.3	E1-15: -116.5
				
E1-16: -115.0	E1-17: -114.9	E1-18: -113.6	E1-19: -113.1	E1-20: -110.4
				
E1-21: -109.3				
PPAR-α inhibitors: no valid designs				
Mcl-1 inhibitors				
				
E3-1: -84.7	E3-2: -81.1	E3-3: -78.8	E3-4: -75.4	E3-5: -74.6
				
E3-6: -73.7	E3-7: -73.2	E3-8: -73.0	E3-9: -72.5	E3-10: -71.4
				
E3-11: -68.5	E3-12: -68.1			

[a] In each structure, the fragments in black are reserved during optimization; whereas those in red are newly added fragments. The number below each structure is its binding score computed with PLP. Lower PLP scores indicate higher binding affinities.

computer programs and that it effectively utilizes the outcomes of virtual screening for this purpose. One may think that it is somewhat inconvenient to conduct a virtual screening job first before AutoT&T can be applied. However, virtual screening has become so popular in structure-based drug design. A number of advanced molecular docking programs, such as GLIDE,[41] GOLD,[47] AutoDock,[50] and DOCK[51] and a number of chemical databases, such as the MDL Available Chemical Directory (ACD) database and the ZINC database,[52] can be used for this purpose. Once the target protein is specified, it is almost certain that one will conduct a virtual screening of some chemical databases to identify interesting hits. Thus, the inputs needed by AutoT&T are usually available.

Furthermore, what the AutoT&T method really needs from virtual screening is the binding modes of a large number of molecules. Once the virtual screening of some chemical databases is completed, the outcomes can be saved on hard disk and

reused again and again by the AutoT&T method in the optimization of different lead compounds. The virtual screening job itself does not need to be repeated each time. Note that a virtual screening job could consume a considerable amount of computational resources, especially when the chemical database under screening is large and when the required molecular docking is performed at a high computational level. The ability of recycling virtual screening outcomes in lead optimization is a notable advantage of our AutoT&T method.

It also needs to be emphasized that AutoT&T is actually a general method rather than a particular computer program. In fact, we have applied the basic idea illustrated in Figure 1 to one drug design project[53] before computer programs were developed to automate this process. In that study, we proposed some derivative compounds of tamiflu as dual-site binders of neuraminidase subtype 1 (N1). Over 1000 FDA-approved small-molecule drugs were examined by molecular docking targeting at the open conformation of N1. Suitable chemical moieties on these molecules that can occupy a minor cavity near the main catalytic site on N1[54] were manually selected out and then transplanted onto the core structure of tamiflu. This work demonstrates that one can apply the AutoT&T method without using the AutoT&T program. But such a process has to be automated with computer programs if the number of steak molecules under consideration is large. This was exactly our original motivation for developing the AutoT&T program.

As described in the Methods Section, our AutoT&T method constructs new molecules by identifying matched bonds between a pair of molecules and then swapping the fragments at both ends of the matched bonds. This algorithm was inspired by the BREED method developed by Pierce et al.,[18] which is now implemented in the MOE software (version 2010.10, released by the CCG Inc.). The central idea of the BREED method is to recombine known ligands to a given target protein to generate new ligand molecules. This approach presumably has higher success rates than designing from scratch. However, an obvious shortcoming of the BREED method is that it can only be applied when some ligands are already known in literature. It is not very helpful if the target protein under consideration is novel. Besides, hybridizing the structures of known ligands to the target protein is not much different from what a medicinal chemist could do. In this aspect, the BREED method resembles the conventional practice of scaffold hopping more closely.

Our AutoT&T method can be considered as a substantial expansion to BREED. Of course the users may choose to supply all known ligand molecules of the given target protein as steak molecules, so AutoT&T will be run in a manner similar to BREED. But the steak molecules utilized by AutoT&T can be any organic molecules in principle. If necessary, one can even supply virtual chemical libraries to AutoT&T as inputs. For example, fragment-based drug discovery[55−57] has drawn extensive attention in recent years, since this approach may obtain promising hits quickly by examining a simplified chemical space. In order to realize the idea of fragment-based discovery with AutoT&T, one can construct a special library containing only small chemical fragments, e.g., molecular weights between 150−250, and examine the possible binding locations of these fragments on the target protein with a molecular docking program. Then, these fragments can be utilized by AutoT&T as inputs to optimize a given lead compound or even assemble novel lead compounds.

Comparison with de Novo Design Methods: Pros and Cons. Various types of de novo design methods have been developed in the past 20 years or so.[58] These methods construct molecular structures with building blocks. For this reason, they are also referred to as "build-up" methods. In this aspect, our AutoT&T method is similar to de novo design methods. However, our AutoT&T method also has some obvious difference from them, which needs to be discussed here.

A de novo design program typically has a predefined internal building block library. Compiling such a library needs thoughtful considerations. Ideally, most regions of the entire known chemical space should be reconstructed with these building blocks. To achieve this goal, it is only possible to use a set of essential small-sized fragments, such as methyl, amino, hydroxyl, ketone, and carboxylic groups, phenyl rings, and so on (see ref 16 for an example). It is impractical to use more complex fragments since the size of the building block library will increase exponentially. Only if a de novo design program is intended to produce certain types of molecules, larger fragments can be used as building blocks. For example, the E-Novo method developed by Pearce et al.[59] utilized the R-group scheme. The R-groups defined by them contained only alkyl groups, cyclopentane derivatives, and phosphate groups, and they were used primarily for generating ATP analogs targeting at kinases. Most de novo design programs are designed to be generally applicable. Thus, they still employ small building blocks. Nevertheless, it is not straightforward to construct some complex organic structures with small building blocks. For example, many heteroatom-containing aliphatic and aromatic cyclic moieties are frequently observed among drug-like molecules. Such moieties cannot be decomposed into simpler building blocks easily. Conventional de novo design methods have certain problems in generating such structures. As a matter of fact, the solutions provided by them often look somewhat "dull" (see Tables 4 and 5). Furthermore, if an organic molecule is randomly assembled with some simple building blocks, its synthetic feasibility is often a matter of concern. This is an intrinsic problem in de novo design methods.

Our AutoT&T method can overcome these difficulties at least to some extent. AutoT&T also constructs molecular structures by assembling building blocks, but it does not rely on a predefined building block library. All of the fragments used by AutoT&T in structural operations are truncated in a dynamic manner from steak molecules. In principle, the building blocks used by AutoT&T have no limit in terms of sizes or types. They are limited only by the contents of the compound library supplied as input. As demonstrated in the three test cases, a remarkable level of structural diversity was observed among the AutoT&T outcomes (Tables 1−3), although only a modest number of steak molecules (up to 60 000 molecules) were considered in structural operation. This number is a fraction of what is typically processed in a virtual screening job. In addition, since the fragments used by AutoT&T in structural operation are all truncated from known organic compounds, they are normally associated with established synthetic methods. Thus, the molecules generated by AutoT&T are more synthetically feasible as compared to those produced by regular de novo design programs. Due to the same reason, the molecules generated by AutoT&T may be more drug-like as well.

The second potential problem rooted in conventional de novo design methods is the so-called combinatorial explosion problem. Most de novo design methods build up molecular structures by assembling fragments basically in a sequential manner.

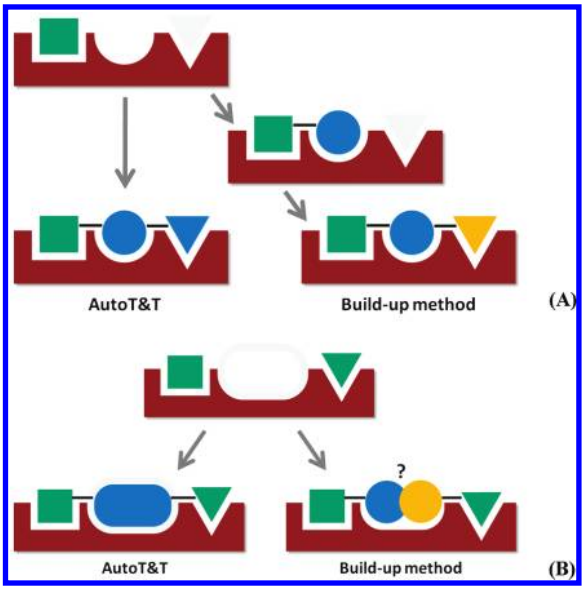


Figure 8. Comparison of AutoT&T and conventional de novo methods in the "growing" (A) and the "linking" (B) modes.

Assume that two fragments need to added onto a seed structure to form an ideal ligand molecule to a given target protein (Figure 8A) and a building block library contains 1000 fragments, then at least $1000 \times 1000 = 1\,000\,000$ possibilities need to be evaluated. In fact, the actual number is much greater than this since there are usually multiple ways to connect two selected fragments. In addition, connecting two fragments often results in a rotatable bond. It is necessary to explore the conformational space regarding this bond to determine the low-energy conformations, which costs an extra amount of computation. Therefore, it is usually not practical for a de novo design program to conduct a systematic search of all possible ligand molecules. A de novo design program normally relies on a sampling algorithm, such as genetic algorithm, to obtain desired results while controlling the computational cost at an acceptable level.

Our AutoT&T method does not have this combinatorial explosion problem since it does not adopt a sequential build-up approach. What AutoT&T does is to compare the lead compound with each given steak molecule and then perform structural transplantation once matched bonds are identified. There are a finite number of possible bond pairs between the lead compound and all steak molecules. Besides, AutoT&T makes rational selections among available fragments for transplantation by considering their binding affinities and synthetic feasibilities. This further reduces unnecessary structural operations. No conformational sampling of the resulting molecules is needed either, since the binding poses of all fragments considered in structural operations, as parts of steak molecules, have been derived by molecular docking in a preparative step. Thus, the overall process of AutoT&T is more efficient than a typical de novo design program. In all of the three test cases considered in this study, AutoT&T indeed completed its jobs very efficiently.

Yet another problem in conventional de novo design methods is that, although they are in theory able to generate new ligand molecules by linking multiple fragments, they often fail in such cases. Indeed, both LUDI and EA-Inventor failed to produce any valid solutions in the second test case shown in our study. This difficulty actually roots in the stepwise building-up process employed by conventional de novo design programs. When a fragment is being installed on the seed structure, the program actually does not have any consideration on what the next fragment should be used. Thus, if the remaining gap between two structures to be connected is too narrow to insert a suitable linker, then the program will fail to complete the final desired molecule (Figure 8B). This problem may be solved by tracing back to replace the fragment installed at the previous step with a more suitable one. However, no de novo design method implements such a recursive process. According to our own experience with conventional de novo design methods, being less effective in the "linking" mode is a serious problem for this type of methods.

In contrast, our AutoT&T method works well in the "linking" mode. As one can see in Table 2, AutoT&T was not only able to provide some solutions in the second test case. The structural diversity among the AutoT&T solutions was also impressive. AutoT&T does not assemble the desired linker segment by segment and thus does not have the same problem as de novo design methods. Instead, it finds a suitable chemical moiety to match both ends at one step. It is like crossing a river through a prefabricated one-piece bridge. This approach is certainly more effective.

Despite of the advantages discussed above, AutoT&T certainly has its own shortcomings. An obvious concern is that whether it is always possible to find desired fragments on steak molecules to transplant onto a given lead compound. The chance for AutoT&T to perform such structural operations successfully is approximately proportional to the total number of steak molecules supplied as inputs. Thus, problems could happen when there are not sufficient steak molecules. However, our test cases suggest that this concern is not critical. In the three test cases described in this study, AutoT&T was run based on 1000 to 60 000 steak molecules, and it produced diverse designs. Some 60 000 molecules are totally practical in reality. Larger chemical libraries can also be considered. Another potential problem is that even if a certain fragment on a steak molecule is suitable to be transplanted onto the lead compound to produce new promising ligand molecules, it has to be placed at the right location inside the binding pocket to be picked up by AutoT&T. Unfortunately, one cannot really control the binding mode of each steak molecule (or any part of it) because it is generated by an external molecular docking program. In theory, increasing the total number of steak molecules can provide a remedy for this problem as well: If a desired fragment on one steak molecule is not placed at the right location, the same fragment on another may be placed at the right location. If this fragment is placed at the right location once among all given steak molecules, then it will not be missed since AutoT&T examines all steak molecules systematically rather than performing a random sampling among them.

Another factor that has a major impact on the quality of the final outcomes of AutoT&T is the scoring method. The primary function of AutoT&T is to generate diverse ligand molecules. These molecules certainly need to be evaluated in order to select the promising ones for further analysis or even experimental tests. A major criterion in selection is the binding affinities between the designed ligand molecules and the target protein. The current version of the AutoT&T program relies on the PLP scoring function to evaluate protein−ligand binding affinities. As mentioned in the Methods Section, this scoring function was chosen since its general performance is among the best scoring functions. It also has certain technical convenience suitable for AutoT&T. No other scoring method is currently implemented in AutoT&T. But the users are encouraged to employ other external scoring

methods, which are not limited to scoring functions, to rescore the outcomes of AutoT&T at the postprocessing step. Other considerations, such as pharmacophore and protein−ligand binding fingerprints, can also be taken into account here. We believe that this is a more flexible strategy for the users.

Different scoring methods may or may not produce consistent results on the same set of molecules. In our study, we have applied the X-Score scoring function[60] and the molecular mechanics-generalized Born/surface area (MM-GB/SA) method[61] to rescore the outcomes produced by AutoT&T, LUDI, and EA-Inventor in all three test cases. The readers are encouraged to examine the Supporting Information of this article for details and corresponding discussion. One can see that these three scoring methods, i.e., PLP, X-Score, and MM-GB/SA, reach a consensus in some cases but not in other cases, which is somewhat unpredictable. Thus, when multiple scoring methods are applied, the results must be interpreted with extreme care.

CONCLUSIONS

Our AutoT&T method is developed primarily for performing lead optimization in structure-based drug design. The most distinctive feature of AutoT&T is that it is able to utilize virtual screening outcomes for this purpose and thus expands the application of virtual screening from lead discovery to lead optimization. AutoT&T identifies suitable chemical fragments on the steak molecules and then transplants them onto appropriate parts on a given lead compound to form new ligand molecules. Unlike conventional de novo design methods, AutoT&T does not reply on a predefined built-in library of building blocks. All fragments used in structural operation are truncated directly from other organic molecules. Thus, AutoT&T has the potential to use relatively complex fragments. Yet, the chemical structures produced by AutoT&T are synthetically more feasible in theory. AutoT&T is also more efficient than conventional de novo design methods since it does not need to perform sampling of the possible combinations of fragments and the possible conformations of the resulting molecules during structural operations.

In our study, the AutoT&T program was tested in different running modes on three target proteins. In the first two cases, i.e., p38 MAP kinase and PPAR-α, AutoT&T was able to produce structures identical or similar to known ligand molecules with better potency than the lead compound. In the third case, we demonstrated how to apply the AutoT&T program to design novel ligand molecules from scratch for the Mcl-1 protein, a relatively challenging target for obtaining small-molecule ligand molecules. In all three test cases, AutoT&T was able to produce a good number of reasonable designs. Compared to the outcomes of LUDI and EA-Inventor on the same target proteins, the outcomes of AutoT&T are structurally more diverse and more promising in terms of binding scores. AutoT&T was also faster than LUDI and EA-Inventor by several fold to complete the assigned jobs. Although AutoT&T has certain technical advantages over conventional de novo design methods, we do not think that it can fully replace the role of de novo design methods. Together they provide a wider range of options for structure-based drug design studies.

ASSOCIATED CONTENT

Supporting Information. Details of the PLP scoring function and the rescoring outcomes produced by AutoT&T, LUDI, and EA-Inventor in all three test cases. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*E-mail: wangrx@mail.sioc.ac.cn. Telephone: 86-21-54925128.

ACKNOWLEDGMENT

The authors are grateful to the financial supports from the Chinese National Natural Science Foundation (grants no. 20772149, 90813006, 21072213, and 21002117) and the Chinese Ministry of Science and Technology (grant no. 2009ZX09501-002).

REFERENCES

(1) Jorgensen, W. L. The many roles of computation in drug discovery. *Science* **2004**, *303*, 1813–1818.

(2) Shoichet, B. K. Virtual screening of chemical libraries. *Nature* **2004**, *432*, 862–865.

(3) Klebe, G. Virtual ligand screening: strategies, perspectives and limitations. *Drug Discovery Today* **2006**, *11*, 580–594.

(4) Muegge, I.; Oloff, S. Advances in virtual screening. *Drug Discovery Today: Technol.* **2006**, *3*, 405–411.

(5) Schapira, M.; Raaka, B. M.; Samuels, H. H.; Abagyan, R. Rational discovery of novel nuclear hormone receptor antagonists. *Proc. Natl. Acad. Sci. U.S.A.* **2000**, *97*, 1008–1013.

(6) Powers, R. A.; Morandi, F.; Shoichet, B. K. Structure-based discovery of a novel, noncovalent inhibitor of AmpC β-lactamase. *Structure* **2002**, *10*, 1013–1023.

(7) Liu, H.; Li, Y.; Song, M.; Tan, X.; Cheng, F.; Zheng, S.; Shen, J.; Luo, X.; Ji, R.; Yue, J.; Hu, G.; Jiang, H.; Chen, K. Structure-based discovery of potassium channel blockers from natural products virtual screening and electrophysiological assay testing. *Chem. Biol.* **2003**, *10*, 1103–1113.

(8) Liu, Z.; Huang, C.; Fan, K.; Wei, P.; Chen, H.; Liu, S.; Pei, J.; Shi, L.; Li, B.; Yang, K.; Liu, Y.; Lai, L. Virtual screening of novel noncovalent inhibitors for SARS-Cov 3C-like proteinase. *J. Chem. Inf. Model.* **2005**, *45*, 10–17.

(9) Tanrikulu, Y.; Schneider, G. Pseudoreceptor models in drug design: bridging ligand- and receptor-based virtual screening. *Nat. Rev. Drug Discovery* **2008**, *7*, 667–677.

(10) Patani, G. A.; LaVoie, E. J.; Bioisosterism, A Rational Approach in Drug Design. *Chem. Rev.* **1996**, *96*, 3147–3176.

(11) Hansch, C.; Selassie, C. Quantitative Structure−Activity Relationship − A Historical Perspective and the Future. In *Comprehensive Medicinal Chemistry II*; Taylor, J. B., Triggle, D. J., Ed.; Elsevier: Burlington, MA, 2006; Vol.4, pp 43−63.

(12) Schneider, G.; Fechner, U. Computer-based de novo design for drug-like molecules. *Nat. Rev. Drug Discovery* **2005**, *4*, 649–663.

(13) Mauser, H.; Guba, W. Recent developments in de novo design and scaffold hopping. *Curr. Opin. Drug Discovery Dev.* **2008**, *11*, 365–374.

(14) Böhm, H. J. The computer program LUDI: a new method for the de novo design of enzyme inhibitors. *J. Comput.-Aided Mol. Des.* **1992**, *6*, 61–78.

(15) Liu, Q.; Masek, B.; Smith, K.; Smith, J. Tagged fragment method for evolutionary structure-based de novo lead generation and optimization. *J. Med. Chem.* **2007**, *50*, 5392–5402.

(16) Wang, R.; Gao, Y.; Lai, L.; LigBuilder, A Multi-Purpose Program for Structure-Based Drug Design. *J. Mol. Model.* **2000**, *6*, 498–516.

(17) Zaliani, A.; Boda, K.; Seidal, T.; Herwig, A.; Schwab, C. H.; Gasteiger, J.; Clauben, H.; Lemmen, C.; Degen, J.; Parn, J.; Rarey, M. Second-generation de novo design: a view from a medicinal chemist perspective. *J. Comput.-Aided. Mol. Res.* **2009**, *23*, 593–602.

(18) Pierce, A. C.; Rao, G.; Bemis, G. W. BREED: Generating novel inhibitors through hybridization of known ligands. Application to CDK2, P38, and HIV Protease. *J. Med. Chem.* **2004**, *47*, 2768–2775.

(19) Rogers, D.; Brown, R. D.; Hahn, M. Using extended-connectivity fingerprints with Laplacian-modified Bayesian analysis in high-throughput screening follow-up. *J. Biomol. Screen.* **2005**, *10*, 682–686.

(20) Lewell, X. Q.; Judd, D. B.; Watson, S. P.; Hann, M. M. RECAP-retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 511–522.

(21) Böhm, H. J.; Stahl, M. The use of scoring functions in drug discovery applications. In *Reviews in Computational Chemistry*; Lipkowitz, K. B. Boyd, D. B., Eds.; Wiley-VCH Inc.: Hoboken, NJ, 2002; Vol. 18, pp 41−88.

(22) Schulz-Gasch, T.; Stahl, M. Scoring functions for protein-ligand interactions: a critical perspective. *Drug Discovery Today* **2004**, *1*, 231–239.

(23) Li, Y.; Liu, Z.; Wang, R. Test MM-PB/SA on true conformational ensembles of protein-ligand complexes. *J. Chem. Inf. Model.* **2010**, *50*, 1682–1692.

(24) Zhang, X.; Li, X.; Wang, R. X. Interpretation of the binding affinities of PTP1B inhibitors with the MM-GB/SA method and the X-Score scoring function. *J. Chem. Inf. Model.* **2009**, *49*, 1033–1048.

(25) Gehlhaar, D. K.; Verkhivker, G. M.; Rejto, P. A.; Sherman, C. J.; Fogel, D. B.; Fogel, L. J.; Freer, S. T. Molecular recognition of the inhibitor AG-1343 by HIV-1 protease: conformationally flexible docking by evolutionary programming. *Chem. Biol.* **1995**, *2*, 317–324.

(26) Cheng, T.; Li, X.; Li, Y.; Liu, Z.; Wang, R. X. Comparative assessment of scoring functions on a diverse test set. *J. Chem. Inf. Model.* **2009**, *49*, 1079–1093.

(27) Wang, R.; Lu, Y.; Fang, X.; Wang, S. An extensive test of 14 scoring functions using the PDBbind refined set of 800 protein-ligand complexes. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2114–2125.

(28) Case, D. A., Darden, T. A., Cheatham, T. E., Simmerling, C. L., Wang, J., Duke, R. E., Luo, R., Merz, K. M., Pearlman, D. A., Crowley, M., Walker, R. C., Zhang, W., Wang, B., Hayik, S., Roitberg, A., Seabra, G., Wong, K. F., Paesani, F., Wu, X., Brozell, S., Tsui, V., Gohlke, H., Yang, L., Tan, C., Mongan, J., Hornak, V., Cui, G., Beroza, P., Mathews, D. H., Schafmeister, C., Ross, W. S., Kollman, P. A. *AMBER 9*, : University of California: San Francisco, 2006.

(29) Clark, M.; Cramer, R. D.; Van Opdenbosch, N. Validation of the general purpose Tripos 5. 2 force field. *J. Comput. Chem.* **1989**, *10*, 982–1012.

(30) Likas, A.; Vlassis, N.; Verbeek, J. J. The global k-means clustering algorithm. *Pattern Recognit.* **2003**, *36*, 451–461.

(31) Reynolds, C. H.; Druker, R.; Pfahler, L. B. Lead discovery using stochastic cluster analysis (SCA): a new method for clustering structurally similar compounds. *J. Chem. Inf. Comput. Sci.* **1998**, *38*, 305–312.

(32) Nilakantan, R.; Bauman, N.; Dixon, J. S.; Venkataraghavan, R. Topological torsion: a new molecular descriptor for SAR applications. Comparison with other descriptors. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 82–85.

(33) Sheridan, R. P.; Miller, M. D.; Underwood, D. J.; Kearsley, S. K. Chemical similarity using geometric atom pair descriptors. *J. Chem. Inf. Comput. Sci.* **1996**, *36*, 128–136.

(34) 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.

(35) Cheng, T.; Zhao, Y.; Li, X.; Lin, F.; Xu, Y.; Zhang, X.; Li, Y.; Wang, R.; Lai, L. Computation of octanol-water partition coefficients by guiding an additive model with knowledge. *J. Chem. Inf. Model.* **2007**, *47*, 2140–2148.

(36) Kumar, S.; Boehm, J.; Lee, J. C. P38 MAP kinases: key signalling molecules as therapeutic targets for inflammatory diseases. *Nat. Rev. Drug Discovery* **2003**, *2*, 717–726.

(37) Wilson, K. P.; McCaffrey, P. G.; Hsiao, K.; Pazhanisamy, S.; Galullo, V.; Bemis, G. W.; Fitzgibbon, M. J.; Caron, P. R.; Murcko, M. A.; Su, M. S. The structural basis for the specificity of pyridinylimidazole inhibitors of p38 MAP kinase. *Chem. Biol.* **1997**, *4*, 423–431.

(38) Stelmach, J. E.; Liu, L.; Patel, S. B.; Pivnichny, J. V.; Scapin, G.; Singh, S.; Hop, C. E.; Wang, Z.; Strauss, J. R.; Cameron, P. M.; Nichols, E. A.; O'Keefe, S. J.; O'Neill, E. A.; Schmatz, D. M.; Schwartz, C. D.; Thompson, C. M.; Zaller, D. M.; Doherty, J. B. Design and synthesis of potent, orally bioavailable dihydroquinazolinone inhibitors of p38 MAP kinase. *Bioorg. Med. Chem. Lett.* **2003**, *13*, 277–280.

(39) Gill, A.; Cleasby, A.; Jhoti, H. The discovery of novel protein kinase inhibitors by using fragment-based high-thoughput X-ray crystallography. *ChemBioChem* **2005**, *6*, 506–512.

(40) de Dios, A.; Shih, C.; Lopez de Uralde, B.; Sanchez, C.; del Prado, M.; Martin Cabrejas, L. M.; Pleite, S.; Blanco-Urgoiti, J.; Lorite, M. J.; Nevill, C. R., Jr.; Bonjouklian, R.; York, J.; Vieth, M.; Wang, Y.; Magnus, N.; Campbell, R. M.; Anderson, B. D.; McCann, D. J.; Giera, D. D.; Lee, P. A.; Schultz, R. M.; Li, L. C.; Johnson, L. M.; Wolos, J. A. Design of potent and selective 2-aminobenzimidazole-based p38α MAP kinase inhibitors with excellent in vivo efficacy. *J. Med. Chem.* **2005**, *48*, 2270–2273.

(41) Friesner, R. A.; Banks, J. L.; Murphy, R. B.; Halgren, T. A.; Klicic, J. J.; Mainz, D. T.; Repasky, M. P.; Knoll, E. H.; Shelley, M.; Perry, J. K.; Shaw, D. E.; Francis, P.; Shenkin, P. S. Glide: A new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J. Med. Chem.* **2004**, *47*, 1739–1749.

(42) Michalik., L.; Auwerx, J.; Berger, J. P.; Chatterjee, V. K.; Glass, C. K.; Gonzalez, F. J.; Grimaldi, P. A.; Kadowaki, T.; Lazar, M. A.; O'Rahilly, S.; Palmer, C. N.; Plutzky, J.; Reddy, J. K.; Spiegelman, B. M.; Staels, B.; Wahli, W. International Union of Pharmacology. LXI. Peroxisome proliferator-activated receptors. *Pharmacol. Rev.* **2006**, *58*, 726–741.

(43) Sierra, M. L.; Beneton, V.; Boullay, A. B.; Boyer, T.; Brewster, A.; Donche, F.; Forest, M. C.; Fouchet, M. H.; Gellibert, F. J.; Grillot, D. A.; Lambert, M. H.; Laroze, A.; Grumelec, C. L.; Linget, J. M.; Montana, V. G.; Nguyen, V. L.; Nicodeme, E.; Patel, V.; Penfornis, A.; Pineau, O.; Pohin, D.; Potvain, F.; Paulain, G.; Ruault, C. B.; Saunders, M.; Toum, J.; Xu, H. E.; Xu, R. X.; Pianetti, P. M. Substituted 2-[(4-aminomethyl)phenoxy]-2-methylpropionic acid PPARα agonists. 1. Discovery of a novel series of potent HDLc rasing agents. *J. Med. Chem.* **2007**, *50*, 685–695.

(44) Day, C. L.; Chen, L.; Richardson, S. J.; Harrison, P. J.; Huang, D. C. S.; Hinds, M. G. Solution structure of prosurvival Mcl-1 and characterization of its binding by proapoptotic BH3-only ligands. *J. Biol. Chem.* **2005**, *280*, 4738–4744.

(45) Meng, X. W.; Lee, S.-H.; Dai, H.; Loegering, D.; Yu, C.; Flatten, K.; Schneider, P.; Dai, N. T.; Kumar, S. K.; Smith, B. D.; Karp, J. E.; Adjei, A. A.; Kaufmann, S. H. Mcl-1 as a buffer for proapoptotic BCL-2 family members during TRAIL-induced apoptosis. *J. Biol. Chem.* **2007**, *282*, 29831–29846.

(46) Lessene, G.; Czabotar, P. E.; Colman, P. M. BCL-2 family antagonists for cancer therapy. *Nat. Rev. Drug Discovery* **2008**, *7*, 989–1000.

(47) Jones, G.; Willett, P.; Glen, R. C. Molecular recognition of receptor sites using an algorithm with a description of desolvation. *J. Mol. Biol.* **1995**, *245*, 43–53.

(48) Tsunoyama, K.; Amini, A.; Sternberg, M.; Muggleton, S. H. Scaffold hopping in drug discovery using inductive logic programming. *J. Chem. Inf. Model.* **2008**, *48*, 949–957.

(49) Wale, N.; Watson, I. A.; Karypis, G. Indirect similarity based methods for effective scaffold-hopping in chemical compounds. *J. Chem. Inf. Model.* **2008**, *48*, 730–741.

(50) Goodsell, D. S.; Olson, A. J. Automated docking of substrates to proteins by simulated annealing. *Proteins* **1990**, *8*, 195–202.

(51) Kuntz, I. D.; Blaney, J. M.; Oatley, S. J.; Langridge, R.; Ferrin, T. E. A geometric approach to macromolecule-ligand interactions. *J. Mol. Biol.* **1982**, *161*, 269–288.

(52) 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.
(53) Li, Y.; Zhou, B.; Wang, R. Rational design of Tamiflu derivatives targeting at the open conformation of neuraminidase subtype 1. *J. Mol. Graphics Modell.* **2009**, *28*, 203–219.
(54) Russell, R. J.; Haire, L. F.; Srevens, D. J.; Collins, P. J.; Lin, Y. P.; Blackburn, G. M.; Hey, A. J.; Gamblin, S. J.; Skehel, J. J. The structure of H5N1 avian influenza neuraminidase suggests new opportunities for drug design. *Nature* **2006**, *443*, 45–49.
(55) Erlanson, D. A. Fragment-based lead discovery: a chemical update. *Curr. Opin. Biotechnol.* **2006**, *17*, 643–652.
(56) Murray, C. W.; Rees, D. C. The rise of fragment-based drug design. *Nat. Chem.* **2009**, *1*, 187–192.
(57) Coyne, A. G.; Scott, D. E.; Abell, C. Drugging challenging targets using fragment-based approaches. *Curr. Opin. Chem. Biol.* **2010**, *14*, 299–307.
(58) Schneider, G.; Fechner, U. Computer-based de novo design of drug-like molecules. *Nat. Rev. Drug Discov.* **2005**, *4*, 649–663.
(59) Pearce, B. C.; Langley, D. R.; Kang, J.; Huang, H.; Kulkarni, A. *E-novo*: A automated workflow for efficient structure-based lead optimization. *J. Chem. Inf. Model.* **2009**, *49*, 1797–1809.
(60) Wang, R.; Lai, L.; Wang, S. Further development and validation of empirical scoring functions for structure-based binding affinity prediction. *J. Comput.-Aided Mol. Res.* **2002**, *16*, 11–26.
(61) Kollman, P. A.; Massova, I.; Reyes, C.; Kuhn, B.; Huo, S.; Chong, L.; Lee, M.; Lee, T.; Duan, Y.; Wang, W.; Donini, O.; Cieplak, P.; Srinivasan, J.; Case, D. A.; Cheatham, T. E. Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models. *Acc. Chem. Res.* **2000**, *33*, 889–897.