JOURNAL OF CHEMICAL INFORMATION AND MODELING

Article

pubs.acs.org/jcim

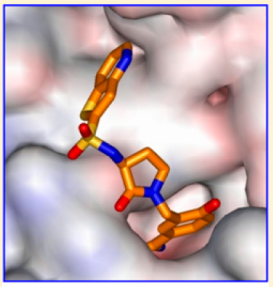
Structure-Based Design Technology Contour and Its Application to the Design of Renin Inhibitors

Alexey Ishchenko, Zhijie Liu, Peter Lindblom, Guosheng Wu, Kam-Chuen Jim, Richard D. Gregg, David A. Claremon, and Suresh B. Singh*

Vitae Pharmaceuticals, 502 West Office Center Drive, Fort Washington, Pennsylvania 19034, United States

Supporting Information

ABSTRACT: It is well-known that the structure-based design approach has had a measurable impact on the drug discovery process in identifying novel and efficacious therapeutic agents for a variety of disease targets. The de novo design approach has inherent potential to generate novel molecules that best fit into a protein binding site when compared to all of the computational methods applied to structure-based design. In its initial attempts, this approach did not achieve much success due to technical hurdles. More recently, the algorithmic advancements in the methodologies and clever strategies developed to design drug-like molecules have improved the success rate. We describe a state-of-the-art structure-based design technology called Contour and provide details of the algorithmic enhancements we have implemented. Contour was designed to create novel drug-like molecules by assembling synthetically viable fragments in the protein binding site using a high-resolution crystal structure of the protein. The technology consists of a sophisticated growth algorithm and a novel scoring function based on a directional model. The growth algorithm generates molecules by dynamically selecting only those fragments from the fragment library that are complementary to the binding site, and assembling them by sampling the conformational space for each attached fragment. The scoring function embodying the essential elements of the binding interactions aids in the rank ordering of grown molecules and helps identify those that have high probability of exhibiting activity against the protein target of interest. The application of Contour to identify inhibitors against human renin enzyme eventually leading to the clinical candidate VTP-27,999 will be discussed here.



1. INTRODUCTION

Current day drug discovery has evolved into a highly sophisticated and complex process. As new experimental and computer-based technologies emerged, they were integrated into the process during the last three decades.[1−3] In addition to the experimental methods, computational and informatics technologies have made a significant impact on the drug discovery process in identifying hits and reducing the time taken to optimize them. The advances in protein X-ray crystallography and the computational methodologies that exploit the information in these structures have had a measurable impact on drug discovery through the structure-based design approach.[3] At present, there are over 71 058 X-ray crystal structures of proteins in the protein data bank (PDB) and about 58 213 ligands contained in them.[4] This number appears to grow annually at a steady pace.[5] The PDB resource has been an invaluable resource for analyzing protein:ligand structures, studying protein−ligand interactions, and developing innovative methods to take advantage of this information.[6,7] The access to high-resolution crystal structures of a large diversity of protein families has influenced the development of novel algorithms and approaches to exploit this information.[8]

A major advancement in the computational approaches for structure-based design was attempted in the early 1990s with the development of de novo design methodologies using novel algorithms.[9−14] These methods generated molecules either by assembling fragments or by evolving molecules in the context of the binding site. The potential impact of de novo design was quite significant, and it correspondingly generated a high level of enthusiasm in the pharmaceutical industry. The ambitious goal adopted by the algorithm designers ran into significant challenges in creating molecules that were drug-like and synthetically tractable. Thus, this methodology diminished and lost its appeal in favor of docking and scoring methodologies. In the past decade, improvements in growth algorithms and scoring functions have led to a resurgence of the de novo design approach.[15−22]

In this Article, we describe our structure-based design technology called Contour and its application to design inhibitors of human renin. Contour consists of a combinatorially efficient growth algorithm that grows molecules de novo, a novel directional model-based scoring function, and a computational infrastructure for high performance structure-based design. The technology borrowed basic concepts from the methodology initially developed at Harvard.[18] Vitae redesigned this methodology by developing an improved growth algorithm with enhanced functionality, flexibility, and the ability to grow drug-like molecules.[19] To guide the growth of the molecules, a novel directional contact model-based

Received: December 16, 2011

   dx.doi.org/10.1021/ci200605k | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

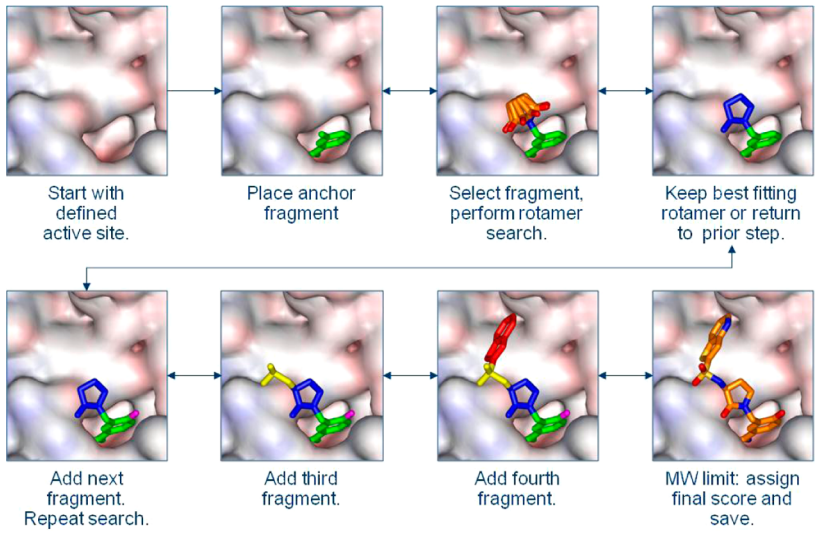


Figure 1. Illustration of the molecule assembly process by the growth algorithm.

scoring function was developed using a diverse set of high-resolution crystal structures of protein:ligand complexes. Training of the scoring function was conducted using a support-vector-based algorithm. In the theory and methodology section, we provide more details underlying Contour.

The renin−angiotensin−aldosterone system (RAAS) plays a key role in the regulation of blood pressure and electrolyte balance.[23] Pharmacological interruption of the RAAS with angiotensin-converting enzyme (ACE) inhibitors or angiotensin receptor blockers (ARBs) has proven effective in blood pressure control. However, renin, which catalyzes the first step in the RAAS cascade, has been the target for blocking its activity to achieve superior cardiovascular benefits and decreased side-effects in comparison to the therapeutic agents used against the targets downstream from it. Renin specifically cleaves angiotensinogen, forming the inactive decapeptide (angiotensin I), which is further converted to the active octapeptide (angiotensin II) by angiotensin converting enzyme (ACE). The conversion of angiotensinogen to Ang I by renin is the rate-limiting step in the formation of Ang II. Direct renin inhibition could be a more effective means of suppressing Ang II production.

Renin is an aspartyl protease. The structure of renin consists of two β-barrel domains, with each domain contributing an aspartic acid residue to the catalytic center.[24] The negatively charged catalytic residues present an ideal environment for cationic moieties to bind strongly. We will describe the use of a cationic functional group that mimics the transition state of the enzyme substrate to design inhibitors of renin with Contour.

2. THEORY AND TECHNOLOGY

Contour consists of a highly sophisticated growth algorithm and a transferable scoring function deployed on a high performance computational infrastructure meant for structure-based design. It is capable of scoring, docking, growing, and optimizing small molecules in protein binding sites.

2.1. Molecular Growth Algorithm. Contour uses a novel algorithm for an efficient search of very large chemical and conformational space for assembling drug-like molecules. Molecules are assembled within a protein binding site starting from an initial fragment. The source of the coordinates for the initial fragment can be from a complexed ligand in a high-resolution crystal structure, from a model, or selected fragments docked to a defined functional interaction site in the binding cavity. Open valencies defined by the atoms containing hydrogen atoms of the starting or initial fragment, referred to as hooks, serve as vectors for attaching additional fragments from the fragment library with single or multiple bonds to grow molecules. The starting fragment can be a single atom, functional group, or larger chemically reasonably fragment, which preferably is derived from a high-resolution crystal structure of a protein:ligand complex. In the absence of a crystal structure of a ligand bound to the target protein, fragments are docked to a functionally relevant site either manually or with Contour. The work described here utilized 1rne structure and manual docking protocols to derive the starting fragments (Figure 6).

There are three possible ways in which the growth algorithm can handle the selection of a fragment from the fragment library at each growth step: (1) a fragment can be selected according to a specified list and/or a sequence, (2) it can be selected stochastically, or (3) the fragment can be selected dynamically guided by the features of the binding site in the neighborhood of the hook to which it needs to be attached.

The attachment of each fragment is followed by a systematic or random search of low energy rotamers and conformational optimization using a novel deterministic algorithm. A small set of chemical rules defined ensures that the structures generated are reasonable and are not chemically unstable. For example, avoid forming a bond between nitrogen and oxygen atoms through a single bond. The optimization algorithm uses the decomposition of the displacement vector to derive movement vectors that optimize hydrogen bond, torsion, and sterics leading to low energy conformations in the binding site. The optimization algorithm performs global rotation, translation, and torsional movements to optimize the interaction energy. This conformational search and optimization protocol guarantees low scoring conformations, but not necessarily low energy conformations identifiable with a force field minimization. A depth-first search of the rotamer tree performed for each configuration enables efficient sampling, which is followed by exclusion of the sterically disallowed conformers. This process of attaching fragments coupled with conformational search and optimization continues until a preset number of

dx.doi.org/10.1021/ci200605k | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

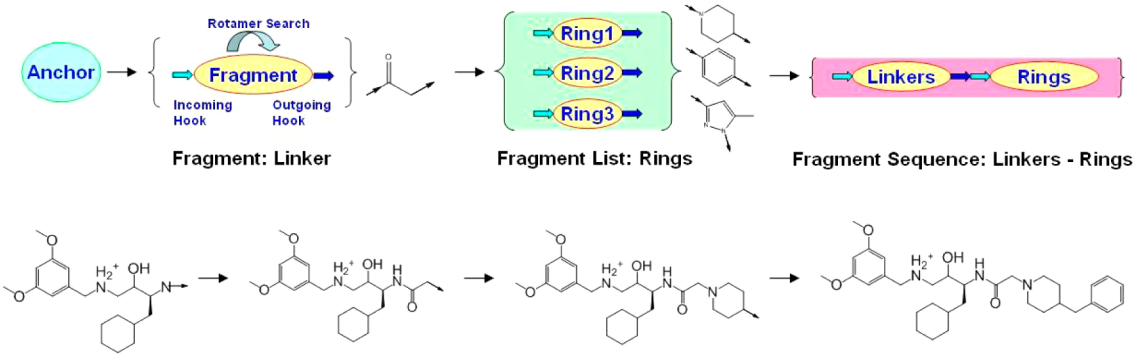


Figure 2. A sample programmable growth instruction set. Anchor is a starting fragment either derived from an X-ray structure or docked to a key site of interaction. Attachment points in the fragments containing hydrogens are labeled as hooks. The hooks in the existing fragments are referred to as outhooks, and the hooks in the fragments that are selected for attachment are referred to as inhooks. Rings can be monocyclic or bicyclic and unsaturated or saturated rings. Linkers in this case are methylketone, methylene, and oxygen atom.

fragments are assembled or until a defined limit is achieved (Figure 1). Grown molecules ideally assume the shape and size of the cavity guided by steric factors to include linear and branched attachments. Each molecule with its unique chain of fragments is assembled in multiple conformations, which are then refined with potential energy minimization using the CHARMm force field and the adopted basis Newton−Raphson minimizer with a command line script.[25] The energy minimized conformations are subsequently optimized with the deterministic optimizer in Contour to fine-tune the intermolecular interactions between the protein and the ligand.

2.1.1. Ligand Chemical Space Sampling. The number of possible combinations of fragments to be assembled in a given growth calculation depends on the number of available fragments in the library. The average number of hooks across the 10 000 fragment library (vide infra) largely containing mono- and bicyclic rings is about 8 (min = 1, max = 26). So each fragment on average has about 8 incoming and 8 outgoing hooks or bond vectors, and 6 possible discrete rotational isomers for each bond (Figure 2). So the total number of possibilities for attaching one fragment to the initial fragment by sampling the extended library containing 10 000 fragments would be $8(8 \times 10\,000) \times 6 = 3.84 \times 10^6$. Therefore, it is not practical to exhaustively sample the available chemical space to identify best-fitting fragments for assembling molecules containing more than two fragments. One way to efficiently sample the chemical space during the assembly of molecules with more than two fragments is to distribute the calculations over a cluster of multiple processors, taking advantage of the hardware architecture. Because exhaustive enumeration is impractical for more than two fragments, algorithmically a random sampling of a subset of the fragment library is another approach. Controlling the sequence and the number of fragments assembled with specified instructions for assembly ensures efficient sampling of the chemical and conformational space accessible within the binding site (e.g., Figure 2). This approach avoids the intractable combinatorial explosion and the need to evaluate a large number of misfits. Random sampling does not guarantee that the best-fitting fragments are selected at each step, but it does provide a practical solution for sampling the sparse chemical and conformational space suitable for a given protein binding site. Thus, to address this sampling issue, Vitae developed a more intelligent solution to enhance the growth process by implementing a dynamic fragment selection algorithm to select the best-fitting fragments.

2.1.2. Dynamic Fragment Selection − Active Site Feature-Based Screening. As was briefly introduced in the previous section, dynamic fragment selection is a novel feature introduced into Contour that uses the physical characteristics of the binding site in selecting complementary fragments during the growth process. This feature significantly reduces the vast chemical and configurational space by selecting only a subset of the fragment library that best matches the shape and features of a given pocket in the binding site.

In the dynamic fragment selection mode, for each hook of a given fragment the protein binding site is characterized with a probe (length $\leq$ 10 Å $\times$ width $\leq$ 5 Å) that calculates steric shape and hydrogen-bonding features in that site. A fragment is then drawn out of a subset of the fragment library either stochastically or in a deterministic manner. The fragment's features are compared against the features in the binding site to calculate its score using the hydrogen bond and solvation energy contributions. The best scoring fragments are then selected at each step to assemble molecules in a piece-wise fashion. The fragment library can be sampled almost exhaustively by distributing the calculation over a dedicated number of processors. For example, for each growth step sampling 100−300 fragments per processor, over 100 processors should efficiently sample 70−95% of the chemical, configurational, and conformational space of the extended library identifying most complementary fragments.

2.1.3. Programmable Molecular Growth. A programmable feature has been implemented to handle one or more instructions for assembling fragments with varying degrees of complexity providing flexibility and sophistication for growth. This allows users to specify growth instructions ranging from very simple ones such as assembling fragments sequentially to complex nested loops consisting of collections of alternate fragments specified by lists and sequential fragments specified by fragment sequences. Thus, users can customize the growth process to their needs. These instructions can be executed in an exhaustive or random fashion to evaluate the possible combinations in assembling fragments. An example of this instruction set is shown in Figure 2.

2.1.4. Protein Conformational Flexibility. The size and scope of sampling of the chemical and conformational space of the ligand gets even more intractable by increasing the number of conformations for the protein to be used during growth. During the initial stages of the exploration of the allowed chemical space for a given binding site, the growth mode keeps

dx.doi.org/10.1021/ci200605k | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

all protein side chains in their starting conformations. In a more focused and detailed exploration, the flexible growth mode allows selected side chains to sample accessible rotameric states. These conformational states are then optimized for hydrogen bonding and steric interactions guided by the scoring function, to allow ligand-induced fit in the protein binding site. A subsequent CHARMm energy minimization process introduces further protein flexibility by allowing the force field to move both side chain and backbone atoms.

2.1.5. Molecular Fragment Libraries. 2.1.5.1. Basic Fragment Library. Molecular fragments in this library consist of frequently occurring core ring fragments without any substitutions derived from the Comprehensive Medicinal Chemistry (CMC) database by chopping up along single bonds.[26] A few selected spirocyclic rings were also added to the library. In addition, the library also contains linkers and simple monofunctional groups. Molecular weights for the fragments in the library range from 16 to 222. Fragments with unsaturated rings are represented by multiple low energy conformations, totaling about 150 fragments.

2.1.5.2. Extended Fragment Library. This library consists of about 23 000 fragments comprised of spirocyclic rings, substructures from commercially available reagents, monocyclic, bicyclic, and tricyclic rings provided to us.[27] A commonly used subset of the fragment library that excludes the tricyclic rings contains about 10 000 fragments with MW range 42−250.

2.1.5.3. ZINC Fragment Library. An independent fragment library was generated from the 2 million drug-like compounds in the ZINC database.[28] The molecules in the ZINC database were processed to generate fragments by chopping up molecules along single bonds excluding large fragments with MW > 250. A unique set of fragments resulting from this process was subjected to conformational sampling and energy minimization. This process yielded 50 000 fragments with MW range 16−250. There is an 80% overlap between the ZINC and the extended libraries.

2.2. Contour Scoring Function. *2.2.1. Functional Form.* Contour adopts an empirical physical model with a sum of discrete linear free energy terms derived from experimental data designed to capture essential features of molecular interactions. The total score for computing protein−ligand binding free energy is given by:

$$F = F_{\text{Interaction}} + F_{\text{Solvation}} \quad (1)$$

where $F_{\text{Interaction}}$ and $F_{\text{Solvation}}$ stand for atom−atom pairwise interaction and solvation scores, respectively (eq 1). The score is expressed in negative log unit of the binding constant ($-\log K_i$). For example, scores of 6.0, 7.0, 8.0, 9.0, and 10.0 correspond to binding constants of 1000, 100, 10, 1.0, and 0.1 nM, respectively.

The first part $F_{\text{Interaction}}$ in the function is a linear combination of contributions from pairwise atom−atom interaction types including hydrogen bonding, electrostatic repulsion, and the nonpolar attraction and repulsion terms (eq 2). The atom types are represented by the AMBER[29] atom types based on the element's hybridization state. The atoms are represented according to their hybridization states with explicit lone pairs and π orbitals. For example, the carbonyl group is represented by an sp^2 carbon attached to an sp^2 oxygen with two lone pairs and p orbitals above and below the plane of the C═O bond. There are no explicit bond orders, but the bond distances and angles are borrowed from the AMBER force field. This physical representation of the atoms forms the basis for our directional contact model. The details of the model along with the description of the scoring function will be published separately.

$$F_{\text{Interaction}} = \sum_i W(t_i, d_i)\cdot f_i \quad (2)$$

where $W(t_i, d_i)$ is a weight derived from the training set for the atom−atom interaction i with the interaction type of t_i at the distance of d_i; and f_i is a form factor of the interaction i, which considers the atom−atom interaction geometry and screening effects. For example, in the hydrogen-bonding term, the form factor f_i is implemented according to the classical definition of the hydrogen-bond geometry (distance and angle) and electronics to reproduce orientation and interactions observed in high-resolution small molecule crystal structures.[30]

The second part $F_{\text{solvation}}$ is a surface area-based solvation term introduced to capture electrostatic and cavity terms in the desolvation of protein and ligand and the solvation of protein−ligand complex. It is computed by the sum of the solvation energies of all atoms involved in protein−ligand binding (eq 3):

$$F_{\text{Solvation}} = \sum_a W(t_a)\cdot f_a \quad (3)$$

where $W(t_a)$ is a weight derived from the training set for the atom a with type t_a, and f_a is a form factor related to the surface area of the atom covered by complex formation.

Steric repulsion is represented by the r^{-12}-based term, which is used as a conformational filter and is not part of the scoring function.

2.3. Training and Validation. *2.3.1. Input Data.* The scoring function weights defined above were derived from a training set of 195 protein−ligand complex X-ray structures spanning 10 families of proteins (Figure 3), and 8 T4 lysozome inhibitor models and 100 renin inhibitors for which crystal structures are not available. For each ligand, at least 25 distinct conformations were generated. These were then optimized with a novel and proprietary support vector bounding algorithm. The input data, training details, and related information will be released along with a separate publication describing the scoring function.

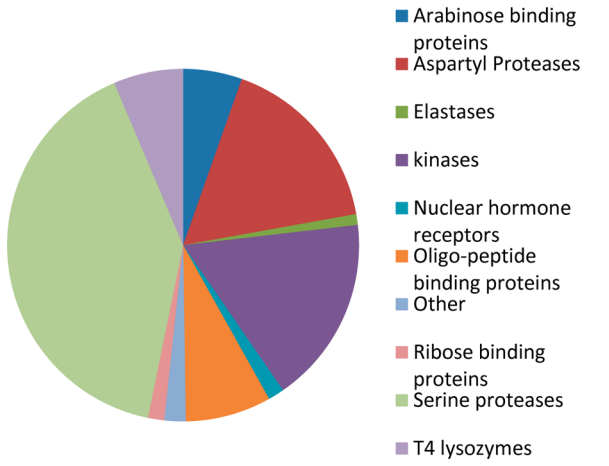


Figure 3. Distribution of protein families used for training the Contour scoring function. Crystal structures of proteins in complex with drug-like molecules from these families were used.

2.3.2. Support Vector Bounding. An algorithm designed to take input data with activity values, discrete bounds, and classification criteria was used to generate the Contour scoring

dx.doi.org/10.1021/ci200605k | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

function. This methodology is a hybrid of support vector machine and support vector regression methods, which combines classification and regression modeling techniques. The algorithm shares several advantages of both of these techniques, such as a sparse solution space, absence of local minima, the ability to control the capacity of the system to prevent overfitting, and the ability to model nonlinear functions using linear operations in a kernel-induced feature space.

Using the support-vector bounding algorithm, all of the input data comprising the scoring function weights were simultaneously fit to the 203 structures of protein−ligand complexes and 100 modeled renin inhibitor complexes with associated binding constants to derive parameters that are consistent with the physical chemical principles of molecular interactions. The iterative process of fitting the input data set and activity prediction of the test set with 20% leave out cross-validation was continued until the scoring function parameters and their weights achieved convergence and displayed physically meaningful signs.

2.4. Computational Infrastructure. A cluster containing 200 nodes of Intel processors (Xeon + Core Duo) with the Red Hat Linux operating system environment comprises the computational server infrastructure. The coordinates and all of the associated properties of the molecules generated are stored in an Oracle database. A Java-based graphical user interface was developed to facilitate 3D graphical visualization of protein−ligand complexes, set up growth calculations, and store 3D coordinates and the associated data in Oracle tables.

The graphical user interface allows building of protocols for stringing the sequence of calculations that form the process of growing and scoring molecules. For instance, a typical Contour calculation has the following sequence of steps: growth, energy minimization, scoring, and storing results in specific output folders. The user interface allows direct navigation of the folders in the Oracle database. Thousands of molecules can be viewed rapidly and seamlessly with a typical physical memory configuration and the Microsoft Windows operating system.

3. RESULTS

3.1. Testing and Validation of Scoring Function. The performance of the scoring function for the set of 203 structures of protein−ligand complexes is shown in Figure 4. The affinities are expressed in negative log scale spanning from millimolar to picomolar range. Greater than 75% of the activities were predicted within 1.5 log units of the experimental data and about 50% within 1.0 log unit. The predicted affinities of 30 renin−inhibitor complexes are shown in triangles and span micromolar to picomolar range. The performance of the scoring function provides a strong basis for separation of the weak millimolar actives from the potent nanomolar compounds and enrichment of the novel computationally generated ideas with high probability of exhibiting predicted activity. The compounds exhibiting scores greater than 5 possess a high probability of exhibiting activity in the enzyme assay, limiting the number of false positives.

An independent validation of Contour scoring function was performed with an external set of six classes of proteins with their ligands. The performance for aurora kinase A inhibitors, one of the six validation sets, is shown in Figure 5. The narrow range of the inhibitory activities of the compounds in Figure 5 provides a more stringent test of the scoring function. The scoring function is able to separate the weak actives from the most potent ones, but lacks a clear rank ordering of the

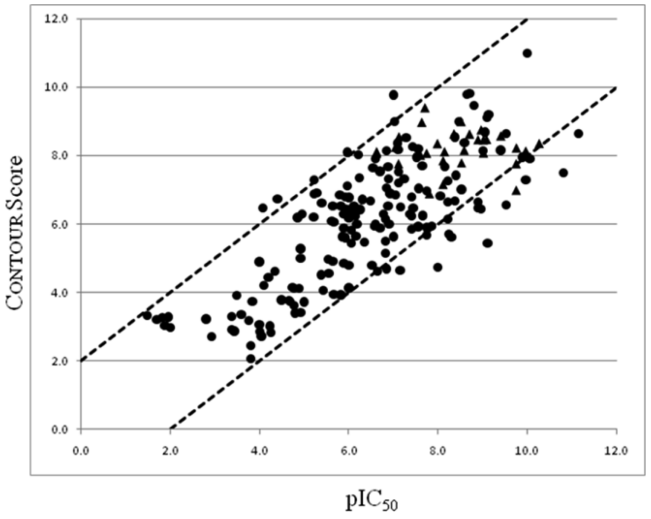


Figure 4. Plot of log(IC_{50}) versus Contour score of the 203 cocomplex structures. The closed circles are the data points for all of the protein−ligand complexes, and the data points shown in triangle symbols are for the renin inhibitor complexes. The Contour scoring function is able to effectively separate the weakly active millimolar compounds from the nanomolar compounds. The predictions shown are from the 30-fold with 20% leave-out cross-validation.

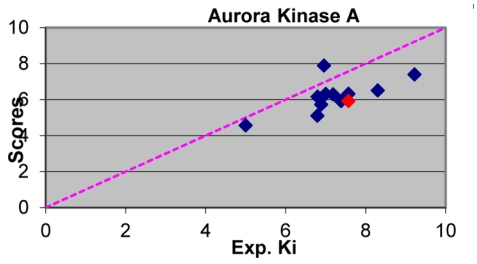


Figure 5. Performance of Contour scoring function on the external validation set of aurora kinase A inhibitiors. The data points shown in blue diamonds are for the predicted models of the inhibitor binding mode using the crystal structure of the complex of the Vitae compound shown with a red diamond.

compounds with activities in the midrange. Nonetheless, consistent with the training set performance, the Contour scoring function provides a means to identify compounds that are likely to exhibit activity against the target when selected above the cutoff score of 5.

3.2. Design of Renin Inhibitors. *Renin Inhibitor Design Strategy.* The Contour design strategy started with a positively charged core fragment interacting with the aspartyl residues of renin's catalytic center, followed by the addition of fragments to this core to generate linear and branched molecules. The active site was explored by limiting the number of fragments assembled at a time to keep the number of combinatorial possibilities tractable to efficiently sample the relevant chemical space in the context of the binding site. Fragments with functional groups complementary to a given subpocket were retained to build further toward the other sites.

Growth Procedure. The design of renin inhibitors was initiated by docking core fragments A and B (Figure 6) starting with the X-ray structure of renin from 1RNE.[31] Orientations with good hydrogen bonds to the catalytic aspartyl residues were chosen. Molecules were grown by stochastically selecting fragments from the fragment library and added to the starting

dx.doi.org/10.1021/ci200605k | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

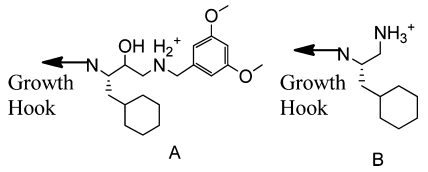


Figure 6. Protonated amines used as starting fragments for designing renin inhibitors. Fragment A derived from the β-secretase inhibitors was manually docked into the renin active site, and the growth calculations were initiated from the amine functionality with the arrow pointing to the left. Fragment B was derived from compound **6** used for optimization for growth calculations.

core with open hooks (points containing hydrogen atoms). After the addition of each fragment, the conformational sampling and optimization was performed, and sterically allowed conformations were retained for the next step. The addition of fragments continued until a preset limit of number of fragments was reached. The assembled molecules were then optimized with the CHARMm force field and scored with the Contour scoring function (eq 1). Ligands with scores above 3 were saved for further detailed graphical visualization and analysis to select those that were acceptable based on viable conformations, binding site complementarity, and calculated molecular physical properties. A 3D graphical visualization with medicinal chemists on a regular basis combined with qualitative assessment of their synthetic feasibility, metabolic stability, and off-target activity led to the selection of compounds for chemical synthesis and biological testing.

4. RESULTS

A high-resolution X-ray crystal structure of recombinant renin in complex with a 2.0 nM peptidomimetic analogue (**1**, 1RNE, Figure 7) served as the starting point for the modeling efforts. The hydroxyethylene moiety in **1** provides functional interaction with catalytic aspartates, and a bulkier cyclohexylmethyl group in the P1 position replaces the leucine residue of the substrate, filling the pocket optimally. One of the design strategies involved borrowing the transition state mimic containing the aminoalcohol headgroup with the prime-side 3−5-dimethoxybenzylamine functional group from inhibitors of the β secretase enzyme (Figure 8). Starting with the docked model of the aminoalcohol headgroup interacting with the catalytic residues and growing the cyclohexylmethyl group from **1** into the S1 site and 3−5-dimethoxybenzylamine moiety into the S1′ pocket with Contour yielded conformations with good fit. The conformation with the best hydrogen-bond interaction between the aminoalcohol group and the catalytic diad was used as the basis for further growth into S2 and S3 pockets. Commercially available 4-phenyl- and 4-benzylpiperidines, 3-phenylpyrrolidines, and other fragments known to exist in active inhibitors of renin were grown. One of the highest scoring molecules was 4-benzylpiperidine derivative **3** shown in Figure 8. At the time of design of compound **3**, the X-ray structure of **2** was not available. So a model of **2** was generated using the 1RNE[9] binding site and superimposed with the model of compound **3**. This superimposed model led to the growth of the repositioned benzyl group to the 3-position of the piperidine ring and attaching a 3-methoxypropoxy chain at the benzylic position. The subsequent synthesis and testing of the modeled compound **4** led to the boost in affinity from 10 μM with **3** to 156 nM. However, the increase in potency came at the cost of increased molecular weight. To reduce the size, compound **4** was truncated on the prime side based on a qualitative analysis that the nonprime side provides significant energetic contributions toward the affinity. This change coupled with the replacement of the alcohol group with a primary amine led to compound **5** with improved binding efficiency relative to **3** as a result of reduced molecular weight. A graphical visualization of the model of compound **5** in the renin active site identified the possibility of an electrostatic repulsion between the ether oxygen of the special pocket chain and the carbonyl oxygen of Gly216. This qualitative analysis led to the conversion of the ether linkage to methylene linker, followed by a Contour growth with functional groups containing donor atom at the tertiary carbon. This growth calculation identified the hydroxyl moiety as the best hydrogen-bond donor for Ser219 in the S4 pocket leading to compound **6**. The predicted binding mode of compound **6** and its graphical visualization suggested the addition of the methyl group to the primary amine would lead to improved electrostatic interactions with the catalytic aspartyl residues as a result of solvent shielding. This led to compound **7**, the most potent inhibitor in the series with 0.5 nM inhibitory activity against the isolated enzyme.[32] The X-ray structure of compound **7** determined subsequently matched the model of the compound bound to renin (Figure 9; 3GW5, rmsd 0.28 Å).

Even though compound **7** exhibited potent activity against renin, it could not be advanced further due to loss of 26-fold activity in the presence of plasma and 1 μM inhibitory activity against major drug metabolizing enzyme CYP3A4. To reduce the shift in renin activity in the presence of plasma and minimize the inhibition of CYP3A4, compound **7** was further optimized with Contour by systematically introducing polarity in the molecule to reduce plasma protein binding.[33,34] The change in polarity of molecules designed was monitored by calculating log $D_{7.3}$ values and total polar surface area with Pipeline Pilot.[35] The goal of the optimization process was to keep the *C* log *D* values in range of 1.0−3.0 and PSA 90−120,

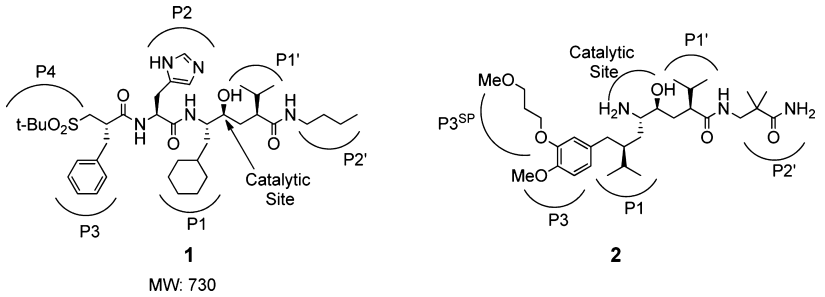


Figure 7. Renin inhibitors: **1**, CGP-38560 (1.0 nM); **2**, Aliskiren (0.5 nM).

dx.doi.org/10.1021/ci200605k | J. Chem. Inf. Model. XXXX, XXX, XXX−XXX

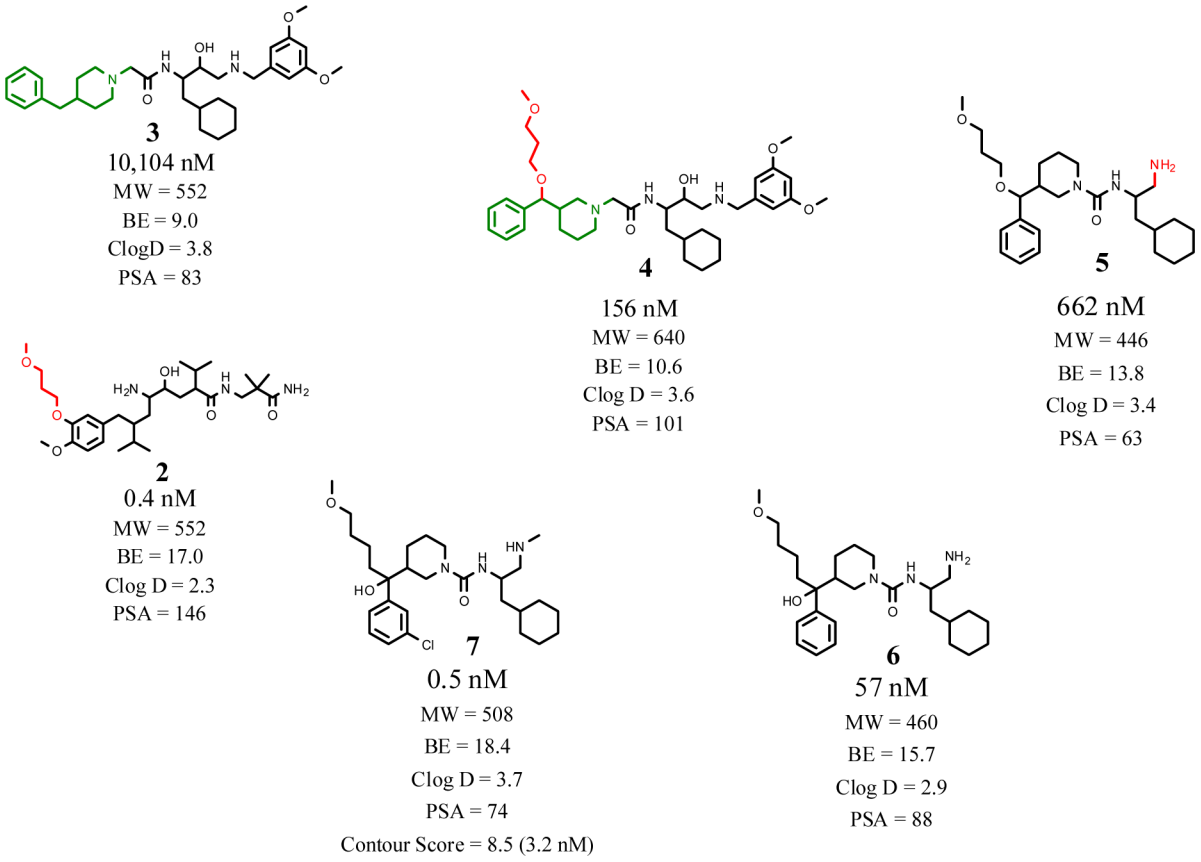


Figure 8. Contour-based design of the lead compound **7**. Fragment A was used to grow compound **3**. Grown model of aliskiren (**2**) superimposed with **3** led to the growth of molecule **4** with benzyl group at the 3-position of piperidine and the methoxypropoxy side chain grown from the benzylic position. Molecule **5** derived from the truncation of the prime side of the catalytic site. Compound **6** optimized using Contour to grow hydroxyl group in S4 and to convert oxygen linker to a carbon linker in the special pocket. Compound **7** optimally fills the p3 pocket with the chlorine atom, making the favorable $\sigma-\pi$ interaction with the phe112 underneath it (Figure 9). BE = binding efficiency defined as the (pIC_{50}/MW) × 1000. pIC_{50} = log $1/IC_{50}$. *C* log *D* values at pH 7.3 and total polar surface area values shown were calculated with Pipeline Pilot 8.5.[35]

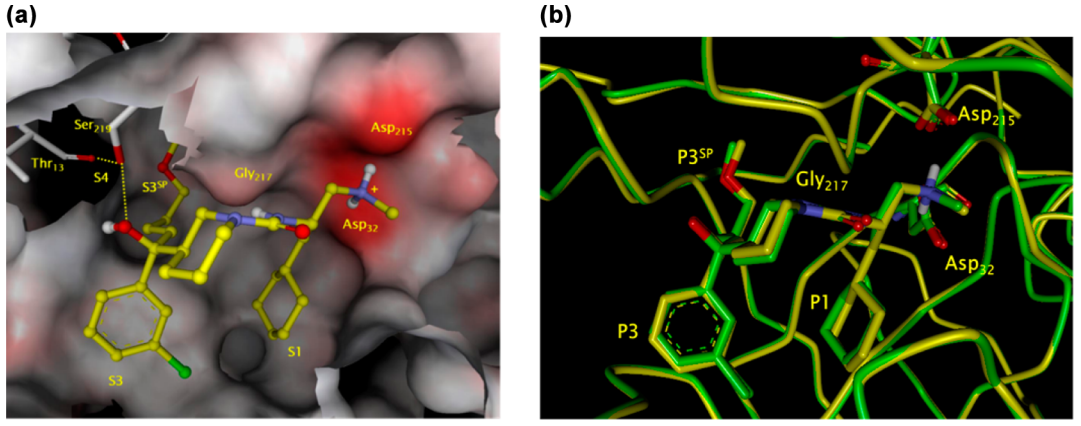
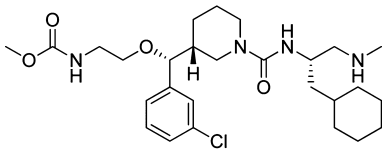


Figure 9. (a) X-ray crystal structure of compound **7** bound to human renin (3GW5). Solvent-accessible surface rendering of the binding site of renin inhibitor with the β-hairpin loop in the front obstructing the view removed. The surface cut-out shows the hydrogen-bond interactions between the p4 hydroxyl group and the s4 subsite residues. (b) Superimposition of the X-ray structure of **7** (yellow) with Contour predicted model (green). Root mean squared deviation of the protein and ligand atoms is 0.28 Å, and that for the ligands is 0.25 Å.

and molecular weight < 550 in the molecules designed by Contour. The hydrogen-bond-donating and -accepting groups of the protein in the S3sp subpocket provided an opportunity to model and test a variety of polar functional groups to replace the terminal methoxy group in compound **7** (*C* log *D* = 3.7, PSA = 74).[33] One of the successful replacements of the 3-methoxypropyl chain occupying the S3sp subpocket with chains incorporating terminal amide and carbamate functional groups afforded more polar compounds such as **8** (*C* log *D* = 3.0, PSA = 101), which exhibited improved activity in the presence of plasma by 16-fold and reduced CYP3A4 activity by 2-fold (Figure 10).[33] The modeled structure of **8** provided a structural basis for the rationale for its most potent activity of the series.

dx.doi.org/10.1021/ci200605k | *J. Chem. Inf. Model.* XXXX, XXX, XXX−XXX

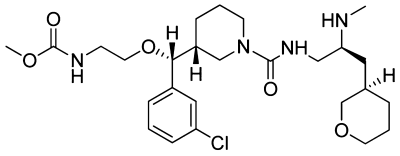


8

IC_{50} = 0.5 nM
MW = 523
Clog D = 3.1
PSA = 92

Figure 10. Optimized renin inhibitor that retains the potency similar to **6** and abolishes the shift to the plasma in the in vitro plasma renin assay (the optimization of **6** leading to compound **8** is presented by Xu et al.[33]).

The CYP3A4 inhibitory activity was further reduced for increasing safety margin in vivo with the exploration of polar functional groups in the P1 pocket in combination with the optimal S3sp chain containing the terminal carbamate group to reduce affinity toward CYP3A4 and plasma proteins. Contour was used to identify a suitable polar ring system that complements the P1 pocket the best, retain the potency, and decrease the activity toward CYP3A4. A variety of oxygen containing hydrogen-bond-accepting ring systems were modeled and tested, of which the 3-tetrahydopyranyl group modeled into the P1 pocket predicted that it would be well tolerated by forming a hydrogen bond with Thr77 on the flap.[34] This change in the p1 moiety with 3-tetrahydopyranyl proved to be the best replacement for the cyclohexyl group in the P1 pocket with *C* log *D* of 1.4 and PSA of 101 for the molecule exhibiting 0.5 nm affinity against the enzyme, unshifted in the presence of plasma, and inhibition > 28 μM against CYP3A4.[32] Among the changes explored for improving bioavailability, rodent pharmacokinetic data for compounds with substitutions α to the amine functionality exhibited improved bioavailability. On the basis of this observation, a model generated by Contour showed that substituting the methyl-3-tetrahydopyranyl group α to the amine functionality could be tolerated, and consequently this change improved its oral bioavailability across three different species (VTP-27,999, Figure 11).[34] The



VTP-27,999

IC_{50} = 0.5 nM
MW = 525
Clog D = 1.4
PSA = 101

Figure 11. Clinical candidate VTP-27,999 (the development of the clinical candidate is presented by Jia et al.[34]).

X-ray structure determined for VTP-27,999 (3Q4B) showed that the oxygen atom of 3-tetrahydropyranyl group is within hydrogen-bonding distance of the Thr77 side chain on the flap and provided tighter interaction with the enzyme. Compound VTP-27,999 was practically unshifted in its inhibitory activity against renin in the presence of plasma and exhibited >30 000 μM affinity toward CYP3A4. Thus, VTP-27,999 exhibiting the best in vitro and in vivo profile was developed as the clinical candidate and recently completed phase I clinical trials successfully.[34]

5. SUMMARY

We described a state-of-the-art structure-based de novo design technology called Contour. The growth algorithm presents novel approaches to efficiently sample the vast chemical space covered by the large number of combinations possible through the fragment libraries described above. The Contour scoring function, developed and validated using a diverse set of protein families and their associated ligands, provides a strong basis for identifying molecules with high likelihood of exhibiting binding activity.

Contour was employed to design a novel structural class of potent, low-molecular-weight, orally bioavailable, efficacious, nonpeptidic, drug-like renin inhibitors using structure-based computer-aided design methodology closely integrated with medicinal chemistry and X-ray crystallography. The models generated for the renin inhibitors allowed diligent optimization of their physical properties by introducing complementary drug-like functional groups assisted by the computational methodology. This approach led to inhibitors exhibiting maximal affinity toward renin, and conferred good oral bioavailability and efficacy in vivo. These efforts resulted in the synthesis of the clinical candidate VTP-27,999 that has a novel structural scaffold with distinct binding mode exhibiting good oral bioavailability (>15%) across four species including humans, and efficacy in animal models.

■ ASSOCIATED CONTENT

Supporting Information

Coordinates of the models for compounds 7 and VTP-27,999 are provided in PDB format. The coordinates are presented in the same frame of reference as PDB entries 3GW5 and 3Q4B, respectively. Also provided is a plot of the experimental pIC_{50}'s versus the Contour score of the renin inhibitors in Figure S1. This material is available free of charge via the Internet at http://pubs.acs.org.

■ AUTHOR INFORMATION

Corresponding Author

*Phone: (215) 461-2048. E-mail: ssingh@vitaerx.com.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

We thank the renin project team for their contributions towards the work described here. We also thank Zhongren Wu and Brian McKeever for the X-ray crystallographic data.

■ REFERENCES

(1) Drews, J. Drug discovery: a historical perspective. *Science* **2000**, *287*, 1960−1964.
(2) Gershell, L. J.; Atkins, J. H. A brief history of novel drug discovery technologies. *Nat. Rev. Drug Discovery* **2003**, *2*, 321−327.
(3) Abola, E.; Calrton, D. D.; Kuhn, P.; Stevens, R. C. Five years of increasing structural biology throughput - A retrospective analysis. In *Structrure Based Discovery*, 1st ed.; Harren, J., Leach, A. R., Eds.; Springer: Dordrecht, The Netherlands, 2007; Vol. *1*, pp 1−24.
(4) RCSB, as of Tuesday Apr 24, 2012 at 2 pm EDT, there are 81 048 entries in the PDB and 71 058 X-ray structures.

(5) Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235−242.
(6) Böhm, H. J. *Prediction of Non-bonded Interactions in Drug Design*; Wiley-VCH Verlag GmbH & Co. KGaA: New York, 2006; pp 3−20.
(7) Williams, M. A.; Ladbury, J. E. *Hydrogen Bonds in Protein-Ligand Complexes*; Wiley-VCH Verlag GmbH & Co. KGaA: New York, 2005; pp 137−161.
(8) Böhm, H.-J.; Schneider, G. *Protein-Ligand Interactions: From Molecular Recognition to Drug Design*; Wiley-VCH Verlag GmbH & Co. KGaA: New York, 2006; p 262.
(9) Danziger, D. J.; Dean, P. M. Automated site-directed drug design: the prediction and observation of ligand point positions at hydrogen-bonding regions on protein surfaces. *Proc. R. Soc. London, Ser. B* **1989**, *236*, 115−124.
(10) Moon, J. B.; Howe, W. J. Computer design of bioactive molecules: A method for receptor-based de novo ligand design. *Proteins: Struct., Funct., Bioinf.* **1991**, *11*, 314−328.
(11) Appelt, K.; Bacquet, R. J.; Bartlett, C. A.; Booth, C. L. J.; Freer, S. T.; Fuhry, M. A. M.; Gehring, M. R.; Herrmann, S. M.; Howland, E. F. Design of enzyme inhibitors using iterative protein crystallographic analysis. *J. Med. Chem.* **1991**, *34*, 1925−1934.
(12) Bohacek, R. S.; McMartin, C. Multiple highly diverse structures complementary to enzyme binding sites: results of extensive application of a de novo design method incorporating combinatorial growth. *J. Am. Chem. Soc.* **1994**, *116*, 5560−5571.
(13) Rotstein, S. H.; Murcko, M. A. GenStar: A method for de novo drug design. *J. Comput.-Aided Mol. Des.* **1993**, *7*, 23−43.
(14) Gillet, V. J.; Newell, W.; Mata, P.; Myatt, G.; Sike, S.; Zsoldos, Z.; Johnson, A. P. SPROUT: Recent developments in the de novo design of molecules. *J. Chem. Inf. Model.* **1994**, *34*, 207−217.
(15) McMartin, C.; Bohacek, R. S. QXP: Powerful, rapid computer algorithms for structure-based drug design. *J. Comput.-Aided Mol. Des.* **1997**, *11*, 333−344.
(16) Ripka, A. S.; Bohacek, R. S.; Rich, D. H. Synthesis of novel cyclic protease inhibitors using Grubbs olefin metathesis. *Bioorg. Med. Chem. Lett.* **1998**, *8*, 357−360.
(17) Wang, R.; Gao, Y.; Lai, L. LigBuilder: a multi-purpose program for structure-based drug design. *J. Mol. Model.* **2000**, *6*, 498−516.
(18) Grzybowski, B. A.; Ishchenko, A. V.; Shimada, J.; Shakhnovich, E. I. From knowledge-based potentials to combinatorial lead design in silico. *Acc. Chem. Res.* **2002**, *35*, 261−269.
(19) Shimada, J. The challenges of making useful protein-ligand free energy predictions for drug discovery. In *Computer Applications in Pharmaceutical Research and Development*; Ekins, S., Ed.; John Wiley & Sons, Inc.: Hoboken, NJ, 2006; pp 321−351.
(20) Jorgensen, W. L.; Ruiz-Caro, J.; Tirado-Rives, J.; Basavapathruni, A.; Anderson, K. S.; Hamilton, A. D. Computer-aided design of non-nucleoside inhibitors of HIV-1 reverse transcriptase. *Bioorg. Med. Chem. Lett.* **2006**, *16*, 663−667.
(21) Kutchukian, P. S.; Shakhnovich, E. I. De novo design: balancing novelty and confined chemical space. *Expert Opin. Drug Discovery* **2010**, *5*, 789−812.
(22) Bajorath, J.; Hartenfeller, M.; Schneider, G. De novo drug design. *Chemoinformatics and Computational Chemical Biology*; Humana Press: Totowa, NJ, 2011; Vol. *672*, pp 299−323.
(23) Atlas, S. A. The renin-angiotensin aldosterone system: pathophysiological role and pharmacologic inhibition. *J. Manag. Care Pharm.* **2007**, *13*, S9−S20.
(24) Tyndall, J. D. A.; Nall, T.; Fairlie, D. P. Proteases universally recognize beta strands in their active sites. *Chem. Rev.* **2005**, *105*, 973−1000.
(25) *CHARMM, Version 27*; Accelrys, Inc.: San Diego, CA, 2007.
(26) *Comprehensive Medicinal Chemistry Database*; Accelrys, Inc.: San Diego, CA, 2003.
(27) Bohacek, R. Boston Denovo Design. Personal communication.
(28) Irwin, J. J.; Shoichet, B. K. ZINC−a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* **2004**, *45*, 177−182.
(29) Cornell, W. D.; Cieplak, P.; Bayly, C. I.; Gould, I. R.; Merz, K. M.; Ferguson, D. M.; Spellmeyer, D. C.; Fox, T.; Caldwell, J. W.; Kollman, P. A. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* **1995**, *117*, 5179−5197.
(30) Allen, F. The Cambridge Structural Database: a quarter of a million crystal structures and rising. *Acta Crystallogr., Sect. B: Struct. Sci.* **2002**, *58*, 380−388.
(31) Rahuel, J.; Priestle, J. P.; Grütter, M. G. The crystal structures of recombinant glycosylated human renin alone and in complex with a transition state analog inhibitor. *J. Struct. Biol.* **1991**, *107*, 227−236.
(32) Tice, C. M.; Xu, Z.; Yuan, J.; Simpson, R. D.; Cacatian, S. T.; Flaherty, P. T.; Zhao, W.; Guo, J.; Ishchenko, A.; Singh, S. B.; Wu, Z.; Scott, B. B.; Bukhtiyarov, Y.; Berbaum, J.; Mason, J.; Panemangalore, R.; Cappiello, M. G.; Müller, D.; Harrison, R. K.; McGeehan, G. M.; Dillard, L. W.; Baldwin, J. J.; Claremon, D. A. Design and optimization of renin inhibitors: Orally bioavailable alkyl amines. *Bioorg. Med. Chem. Lett.* **2009**, *19*, 3541−3545.
(33) Xu, Z.; Cacatian, S.; Yuan, J.; Simpson, R. D.; Jia, L.; Zhao, W.; Tice, C. M.; Flaherty, P. T.; Guo, J.; Ishchenko, A.; Singh, S. B.; Wu, Z.; McKeever, B. M.; Scott, B. B.; Bukhtiyarov, Y.; Berbaum, J.; Mason, J.; Panemangalore, R.; Cappiello, M. G.; Bentley, R.; Doe, C. P.; Harrison, R. K.; McGeehan, G. M.; Dillard, L. W.; Baldwin, J. J.; Claremon, D. A. Optimization of orally bioavailable alkyl amine renin inhibitors. *Bioorg. Med. Chem. Lett.* **2011**, *20*, 694−699.
(34) Jia, L.; Simpson, R. D.; Yuan, J.; Xu, Z.; Zhao, W.; Cacatian, S.; Tice, C. M.; Guo, J.; Ishchenko, A.; Singh, S. B.; Wu, Z.; McKeever, B. M.; Bukhtiyarov, Y.; Johnson, J. A.; Doe, C. P.; Harrison, R. K.; McGeehan, G. M.; Dillard, L. W.; Baldwin, J. J.; Claremon, D. A. Discovery of VTP-27999, an alkyl amine renin inhibitor with potential for clinical utility. *ACS Med. Chem. Lett* **2011**, *2*, 747−751.
(35) *Pipeline Pilot, Version 8.5*; Accelrys Inc.: San Diego, CA, 2012.