# Structure-Based Fragment Hopping for Lead Optimization Using Predocked Fragment Database

Fang-Yu Lin[†] and Yufeng J. Tseng[†,‡,*]

[†]Graduate Institute of Biomedical Electronics and Bioinformatics and [‡]Department of Computer Science and Information Engineering, National Taiwan University, No. 1, Sec. 4, Roosevelt Road, Taipei, Taiwan 106

**ABSTRACT:** In this work, we describe a structure-based de novo optimization process, called "LeadOp" (short for *lead optimization*), that decomposes a compound into fragments of different molecular components either by chemical or user-defined rules. Each fragment is evaluated through a predocked fragment database that ranks fragments according to specific fragment—receptor binding interactions, replacing fragments that contribution the least to binding and finally reassembling the fragments to form a new ligand. The fundamental idea is to replace "bad" fragments of a ligand with "good" fragments while leaving the core of the ligand intact, thus improving the compound's activity. The molecular fragments were selected from a collection of 27 417 conformers that are the fragments of compounds in the DrugBank database. The collection of molecular fragments are docked to the target's binding site and evaluated using group efficiency (calculated binding affinity divided by the number of heavy atoms), and the "strongest" binder is selected. The LeadOp method was tested with two biomolecular systems: mutant B-Raf kinase and human 5-lipoxygenase. The LeadOp methodology was able to optimize the query molecules and systematically developed improved analogs for each of our example systems.
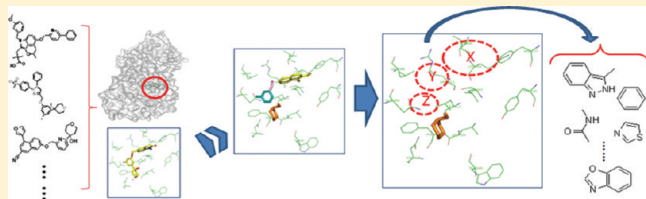
## INTRODUCTION

Lead optimization typically involves substituent replacement paired with a QSAR (quantitative structure—activity relationship) model to refine and evaluate new compounds related to a specific biological end point or druglike properties. The use of QSAR optimization relies on the availability of confirmed chemical and biological data for a series of molecules to build the QSAR model that is able to predict the bioactivity (or end point) for new compounds in the hope of designing either better compounds or finding a novel series of compounds.[1] Scaffold hopping aims to substitute the existing chemical core structure with a novel chemical structure while maintaining—or improving—the biological activity of the original molecule and uses one of two approaches: (i) virtual screening of the entire molecule, not a specific scaffold, to find novel chemical structures in molecular databases of available or virtual compounds[2] or (ii) replacing the core structure with a different chemical motif that preserves similar ligand—receptor interactions via crucial ligand terminal groups.[3,4]

The QSAR approach in the search for new scaffolds depends mostly on the molecular similarity of the initial compound of interest and the compounds in the database. The molecular similarity search techniques include shape,[5] pharmacophore,[6] and fingerprint-based[2] methods or a combination of these strategies[7] to identify similar molecules based on molecular features and potential similar bioactivities. The type of structural features and the molecular similarity cutoff value affects which molecules are selected.[8,9] To overcome the molecular similarity bias that is commonly seen in ligand-based methods,[10] fragment-based

approaches have become widely used.[11] Fragment libraries of possible molecular replacements (substituent) can be constructed by searching for bioisosteres,[12] locating similar ring systems,[13] replacing a central atom of the scaffold, using simple chemical rules (SMART matches, an extension of SMILES strings[14] used to locate molecular substructures to condense the current compound databases[15]), or defining fragmentation schemes of known ligands.[16]

Prior knowledge of the ligand—receptor interactions by means of a cocrystal structure allows the incorporation of these molecular interactions in the search for compounds with different core structures while preserving similar biological activity.[17] Bergmann et al.[18] combined the GRID[19]-based interaction profile of the target protein with the geometrical description of a ligand scaffold to obtain new scaffolds with discrete structural features. Favorable regions for potential ligand—receptor interactions are identified through the creation (calculation) of isocontours. The molecular probes used to calculate the molecular interaction field isocontours include a water molecule, a methyl group, an amine nitrogen, a carboxyl oxygen, and a hydroxyl group. Each probe visits each grid point of a uniformly constructed grid that contains the receptor or a user-defined region of the receptor such as the binding site. Another methodology, GANDI,[20] is fragment-based and generates new molecules by connecting predocked—to the receptor's binding site—fragments and linkers within the binding site. Successive force-field-based (molecular mechanics) energy
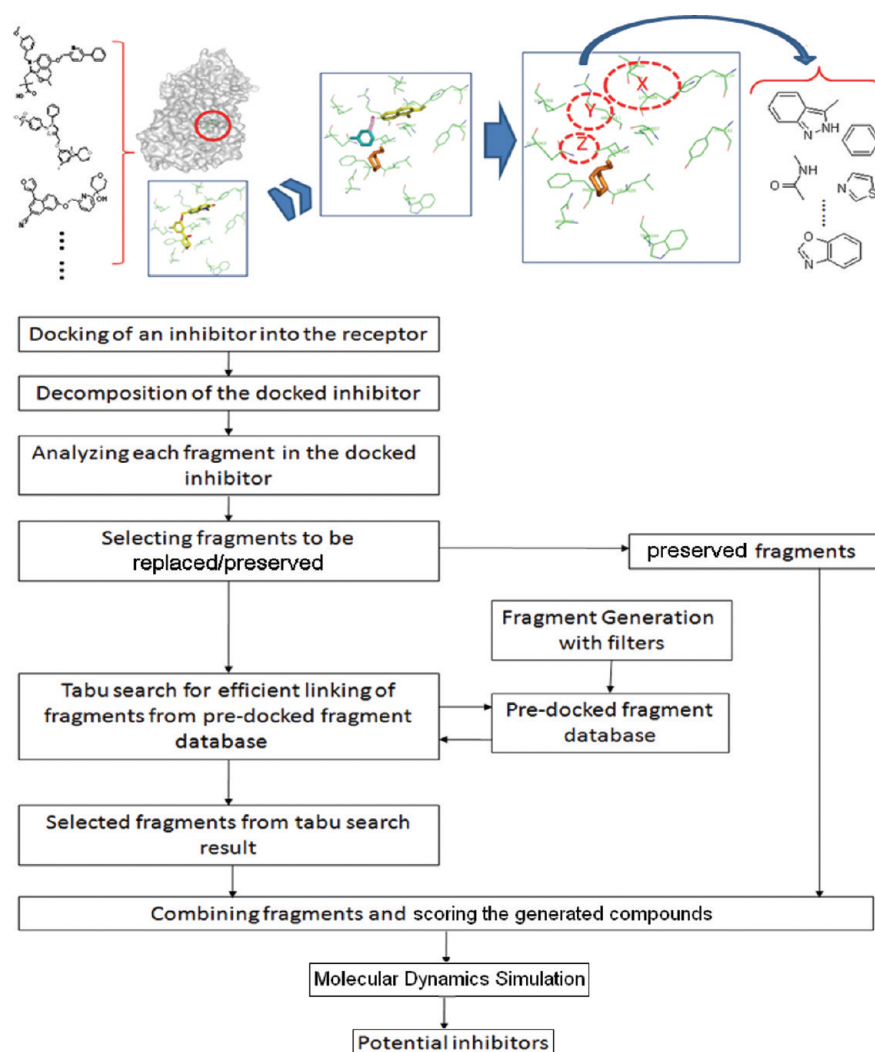
**Figure 1.** Illustration of the LeadOp optimization steps. Starting with a query molecule in its binding pose at the active site, it is decomposed into fragments. The molecular fragments are evaluated and those with the least amount of contribution to binding, based on group efficiency, are replaced with a fragment database through an evaluation process, while the remaining parts were preserved. New compounds are generated by linking the fragments, and the newly proposed compounds are ranked on the basis of a calculated binding free energy.

minimization of the new complex is carried-out to remove steric clashes and optimize the ligand–receptor interactions to mirror the 2D-similarity and 3D-overlap of the original compound's known binding mode(s) by way of a genetic algorithm. The GANDI protocol was assessed using the cyclin-dependent kinase 2 (CDK2) biomolecular system. New bioactive compounds for CDK2 were suggested that contained unique scaffolds and transformed substituents, which preserved the main binding motifs, along with corresponding to known CDK2 inhibitors.

In this study, we discuss the development and implementation of the lead optimization algorithm, "LeadOp" for short, that (i) decomposes a lead drug compound (structure) into fragments either by chemical rules or user-defined rules, (ii) evaluates each fragment's (fragments from the LeadOp predocked fragment database that includes those of the initial ligand and DrugBank[21]) efficiency within the binding site and ranks the fragments according to specific fragment–receptor binding interactions, (iii) replaces ligand fragments that contribute the least to binding with more favorable binders, (iv) reassemble the fragments from each part of the binding site to construct a new ligand, and finally

(v) calculate the new binding affinity of the compound(s). The substitution of the original ligand's substituents (fragments) from a database of ranked fragments that are prioritized as a result of docking each fragment to the receptor facilitates the generation of new structures that are possibly stronger binders. Also, by evaluating the contribution of each fragment toward binding with group efficiency instead of relying on similarity of fragments or overall structure, LeadOp selects fragments that bind stronger yet possess less heavy atoms.[22,23] The assembly of the fragments into a molecule is based on the bond distance and angles of each fragment, with respect to each other. Only the combinations of fragments that form reasonable chemical bonds are selected to generate the new compound that is added to the initial list of proposed compounds. The final list of suggested compounds are reduced using Lipinski's rules-of-five[24] and a set of in-house filters based on our observations of compounds within DrugBank.[21] The compounds that passed the molecular property filters comprised the final list of proposed compounds. The compounds were then energy-minimized and ranked on the basis of the overall ligand–receptor binding energy. To investigate the

1704

dx.doi.org/10.1021/ci200136j |*J. Chem. Inf. Model.* 2011, 51, 1703–1715

interactions between the newly assembled molecules and their receptor, molecular dynamics simulations were performed to investigate the poses and their interactions with the solved crystal structure of the receptor.

To demonstrate the LeadOp algorithm, we selected the mutant B-Raf kinase enzyme[25] and human 5-lipoxygenase (5-LOX)[26] with their associated inhibitors as model systems. B-Raf is a Ras-activated serine/threonine protein kinase that belongs to the Ras protein family (A, B, C)[27] and forms part of a conserved signal transduction pathway that regulates cellular responses to extracellular signals. B-Raf is critical for cell survival, growth, and proliferation and is most frequently shown as a mutated protein kinase (a valine amino acid residue is mutated to a glutamic acid at residue position 600, B-Raf$^{V600E}$) in the development of cancer.[28] Mutagenesis studies revealed that 40−60% of melanomas and 7−8% of all cancers carry an activating mutation in the gene encoding B-Raf and involve the mitogen-activated protein (MAP) kinase pathway (Ras/B-Raf/MEK/ERK) in a large panel of common cancers.[29] The activity of mutant B-Raf$^{V600E}$ kinase is elevated 500-fold (due to the mutation), providing cancer cells with both proliferation and survival signals, allowing them to grow as tumors in model systems.[30] With the kinase domain of mutant B-Raf being reported (solved X-ray structure),[25] this has allowed the optimization of the compounds for preferential binding to mutant B-Raf over wild-type B-Raf. Thus, the development of inhibitors that target mutant B-Raf is of particular interest for the design of cancer[31] therapeutics through target-based approaches. Lipoxygenases are a family of iron-containing enzymes found in a large variety of organisms, such as bacteria and animals. It catalyzes the dioxygenation of polyunsaturated fatty acids containing a *cis*-1,4-pentadiene structure—the first committed structure in a casade of metabolic pathway—and is involved in the initiation of signaling molecule synthesis and inducing structural or metabolic changes.[32] In animal, four major isozymes of lipoxygenases have been identified,[33] including 5-, 8-, 12-, and 15-LOX, which are key enzymes in the metabolism of prostaglandins and leukotrienes. In particular, leukotrienes are produced through the 5-LOX pathway, and the increased activity of the 5-LOX pathway, which includes another protein termed FLAP (5-lipoxygenase—activating protein),[34] is strongly associated with atherosclerosis. As these biological pathways and byproducts lead to inflammation, discovery of a 5-lipoxygenase inhibitor is important in the fields of inflammatory and allergic diseases.[35]

## ■ MATERIALS AND METHODS

**Overall Procedure.** The overall protocol for LeadOp is illustrated in Figure 1 and the details of each step are described in the following sections. The molecule to be modified is docked to the receptor's known drug binding site and then decomposed into molecular fragments. Each fragment of the query ligand was evaluated with the degree of interaction based on group efficiency or user-defined/scientific knowledge to determine which fragments were to be replaced. Molecular fragments of the ligand that possess an unfavorable interaction with the receptor, based on the initial evaluation, are marked for replacement while those with more favorable interactions are retained. Before the substitution of ligand fragments, a fragment library (consisting of fragments from the initial ligand and the DrugBank database[21]) was constructed and predocked into the receptor's binding site. All predocked fragments are sorted (ranked) by their group

efficiency—and ligand attachment point—creating a predocked fragment database used to draw potential ligand−fragment replacements for the noted ligand fragments possessing unfavorable interactions with the receptor. Tabu searching[36] was implemented to search for the "superior" substituent from the predocked database. Once an optimal set of fragments for substitution was found, fragments are linked with the remaining portion of the initial molecule to generate a new compound. Finally, all the compounds generated with this strategy were ranked, providing a series of new de novo compounds.

**Example Systems.** B-Raf kinase (PDB ID: 3idp), a ras-activated proto-oncogene serine/theronione protein kinase,[25] and human 5-LOX enzyme (obtained from the homology model by Caroline et al.,[37] a key enzyme in leukotriene biosynthesis, were selected as our model systems to examine the LeadOp approach. One B-Raf kinase inhibitor, compound **16** (aminoisoquinoline series) in ref 25 (denoted as compund **A** in this study), and a human 5-LOX inhibitor, compound **7** (substituted coumarins) in ref 26 (denoted as compound **F** in this study), were selected as LeadOp examples.

**Generation of Fragments.** The library of potential substitution fragments was generated by decomposing 4855 molecules from the "small molecule structures" property descriptions of the "drug structure" section in the DrugBank database.[21] The DrugBank database contains chemical, pharmacological, and pharmaceutical drug data along with sequence, structure, and pathway information for various drug targets. The DrugBank compounds were energy-minimized and subsequently decomposed with DAIM[38] to generate the fragments; duplicate fragments were removed, resulting in 1688 fragments being added to the LeadOp fragment library from DrugBank. LeadOp fragment library also included 1311 amine building blocks from SciFinder (heterocycles such as quinolines, imidazoles, biaryls, pyrrolizines, thiopyranol[2,3,4-*c*,*d*]indoles, naphthalenic lignan lactones, phenoxymethylpyrazoles, methoxytetrahydropyrans) and substituted coumarins from a previous studies.[25,39,40] Fragments were removed if (i) the number of oxygen, nitrogen, sulfur, phosphates, and halogens in a fragment was greater than two, (ii) there was more than one double and/or triple bond, and (iii) there was more than two hydrogen-bonding donors or acceptors.

**Predocked Fragment Database Construction.** Each fragment of the LeadOp fragment library, generated in the previous step, was docked into the B-Raf and 5-LOX binding site via SEED,[41] which explicitly calculated the desolvation energy of the fragment while exploring the fragment's possible binding modes. Each docked fragment resulted in multiple poses and associated binding energies. A representative fragment pose was selected using a cutoff energy of 5 kcal/mol; this yielded 236 585 conformations for 1688 docked fragments. All fragments were ranked according to group efficiency, calculated by dividing the fragment's docked binding energy with the number of heavy atoms within the fragment. The resulting prioritized, predocked fragments database contained 27 417 conformers for 1688 fragments.

**Preparation for Optimization.** Compounds to be docked were geometry optimized with the MM+ force field in HyperChem 7.0[42] and docked into the target protein binding sites with AutoDock Vina[43] using the default settings.

**Selection of Fragments To Be Replaced.** The ability to indicate how the docked inhibitors are decomposed along with which fragments are retained are user specifications within the LeadOp protocol. The decomposition retains the docked orientation and position of each fragment, and the group efficiency of

1705

dx.doi.org/10.1021/ci200136j |*J. Chem. Inf. Model.* 2011, 51, 1703–1715

each fragment is calculated. The top 20% of the original fragments (from the original query molecule), on the basis of group efficiency, are automatically retained while the remainder of the original fragments undergo replacement.

**Tabu Search for Better Replacement and Compounds Assembly.** To efficiently search and determine reasonable replacement fragments, a look-up table consisting of the bond distances and angles between the fragments and the original compound's attachment points (location of substituents to be exchanged) is constructed. Acceptable bond distance(s) and angle(s) between the fragment and the potential attachment point are a key indicator to determine if the docked fragment should be a possible replacement. The new compounds are generated by connecting all the possible combinations of fragments to the remaining initial ligand based on appropriate bond lengths and angles.

**Trimming the Potential Compound Library.** After the assembling the compounds and removing those that violate Lipinski's rules-of-five, the following filters are applied to reduce the total number of new compounds. Compounds with (i) four or more double bonds (excluding aromatic bonds) or triple bonds with no more than three of each type or (ii) 11 or more triple bond are removed from the potential set of compounds. After reducing the compounds that violate the above rules, each compound is energy minimized and prioritized (ranked) using the overall binding energy.

**Molecular Dynamics Simulations.** The bound pose of the newly constructed compound, as determined with AutoDock Vina,[43] is refined from the lowest binding free energy and the number of favorable ligand–receptor interactions within the binding site. The unfavorable contacts between the docked pose of the energy-minimized "constructed" compound (fragments connected to the remaining initial compound) and the residues within the binding site are removed using molecular dynamics simulations, thus allowing the complex to explore local energy minima. The best complex pose was selected and molecular dynamics was performed using GROMACS version 4.03[44] and the GROMOS 53A6 force field.[45] The complexes are placed in a simple cubic periodic box of SPC216-type water molecules,[46] and the distance between protein and each edge of the box was set as 0.9 nm. To maintain overall electrostatic neutrality and isotonic conditions, $Na^+$ and $Cl^-$ ions were randomly positioned within this solvation box. To maintain the proper structure and remove unfavorable van der Waals contacts, a 1000-step energy minimization using the steepest descent algorithm was employed with an energy minimization convergence criteria of a between-step difference smaller than $1000 \text{ kJ mol}^{-1} \text{ nm}^{-1}$. After the energy minimization, the system was subjected to a 1200 ps molecular dynamics simulation at constant temperature (300 K), pressure (1 atm), and a time step of 0.002 ps (2 fs) with the coordinates of the systems recorded every 1 ps.

## ■ RESULT AND DISCUSSION

**Structure-Based Fragment Hopping of B-Raf Inhibitors.** For the B-Raf inhibitors example, a mutant B-Raf and a ras-activated proto-oncogene serine/theronione protein kinase were selected. An aminoisoquinolines series of mutant B-Raf pathway inhibitors was investigated in the literature,[25] and a cocrystal structure of inhibitor L1E with B-Raf shows the interactions in the B-Raf active site (PDB ID: 3idp). In this cocrystal structure, the purine group of L1E forms several stabilizing interactions
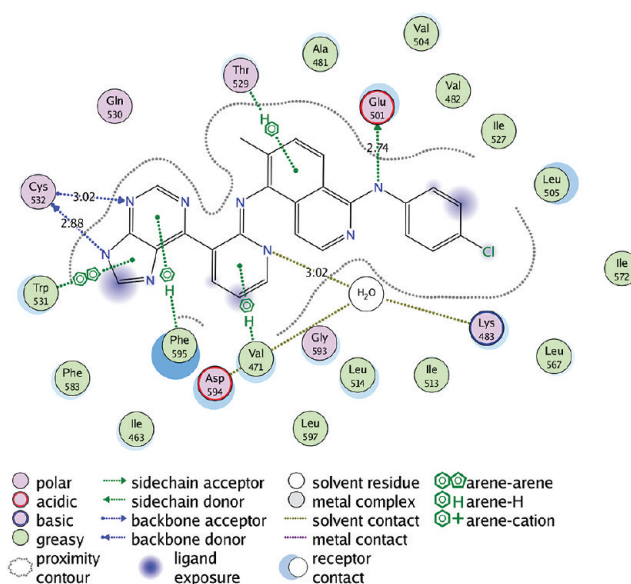


**Figure 2.** Illustration of the ligand–protein interaction of mutent B-Raf and LIE from cocrystal stsructure (PDB ID: 3idp) Chemical characteristic of each residue and interaction within the complex are colored and described in the following.

with the receptor: (i) two hydrogen bonds with Cys532 of B-Raf (one with the backbone amine and the other with the backbone carboxyl group), (ii) π-stacking with the side chain of Trp531, and (iii) a π–hydrogen atom interaction with Phe595. Figure 2 illustrates the ligand–receptor interaction of this cocrystal structure.

A pose similar to the solved crystal structure of L1E bound to B-Raf was determined through our docking study. Therefore, the same AutoDock Vina parameters were used to dock compound **A**, from the same series, into the binding pocket; Figure 3a illustrates the docked pose. Compound **A** was selected for optimization by the LeadOp algorithm in this example.

The aminoisoquinoline core was preserved during the fragment hopping due to its kinase selectivity and favorable pharmacokinetic properties.[25] Compound **A**—docked to B-Raf—was decomposed (fragmented) into six fragments (Frag-0 to Frag-5 in Table 1, indicated using different colors in Figure 3a), and the group efficiency scores were calculated. More positive group efficiency values infer a weaker binding interaction than fragments with lower values. Thus, the original ligand fragments with the most positive group efficiency scores were selected for replacement (Frag-0, Frag-1, and Frag-5 in Table 1) under the user-defined selection mode. The new compounds were constructed after replacement of the weakly performing (binding) fragments with fragments considered to have "better" interactions with the receptor. The last step of LeadOp is the ranking of the new compounds based on their calculated binding energy. For this example, 5576 new B-Raf inhibitors were generated, evaluated, and ranked.

To evaluate our algorithm, we compared all of the LeadOp-generated compounds to the proposed aminoisoquinoline analogs from the original literature and found that six of the LeadOp compounds (Figure 3b) are among the 12 proposed aminoisoquinoline analogs that have been synthesized and measured.[25] The inclusive replacement of fragments (substituents) combined with systematically examining the proposed fragment's interactions with the receptor while retaining the core generated six compounds that have more potent $IC_{50}$ values than the original compound
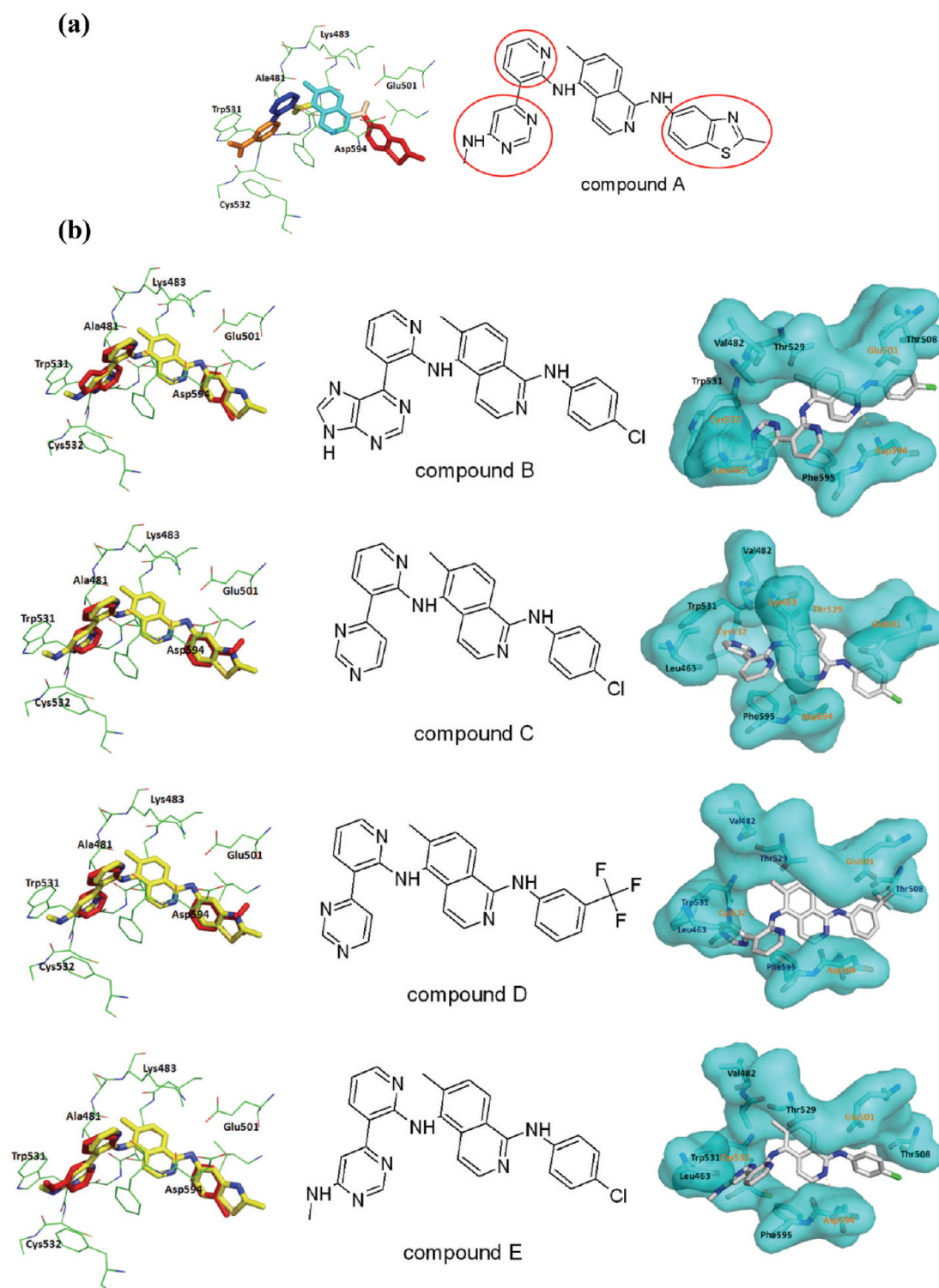
**Figure 3.** LeadOp result in B-Raf model system: (a) Each fragment of compound **A** is colored differently (left). The red ovals indicated the fragments selected to be replaced (right). (b) The carbon atoms of the original compound **A** are colored yellow and the new fragments' carbon atoms, of the generated compound (middle), are colored red (left). Amino acid residues that participate in hydrogen-bonding interactions with the proposed compound, at the binding site (right), are depicted with cyan molecular surfaces.
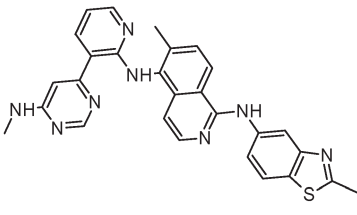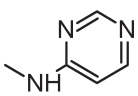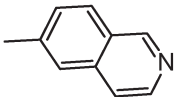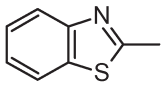
(compound **A**). Four (compounds **B**–**E**) of the six generated compounds were selected for further investigation of their ligand–receptor interactions to represent diverse IC$_{50}$ values.

The poses, ligand–receptor interactions, and the replaced fragments (in red) of these four compounds are shown in Figure 3b. It is

interesting to note that even though Frag-0, Frag-1, and Frag-5 were possible replacement locations, these three fragments are retained in their original location for several of the final structures.

Compound **B** (the most active compound among the four proposed with an IC$_{50}$ = 1.6 nM) preserved Frag-1 in one of the

**Table 1. Evaluation of the Six Fragments, Frag-0 to Frag-5, from Compound A for the B-Raf Biological System with Binding Free Energy ($\Delta G$) and Group Efficiency (GrpEff)**[b]

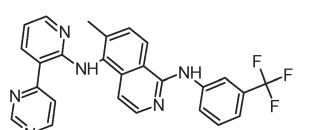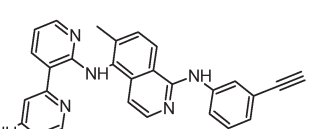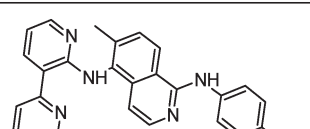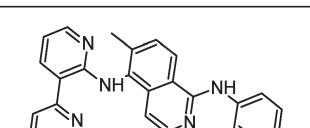| | Structure | $\Delta G$ (kcal·mol$^{-1}$) | $\Delta$GrpEff (kcal·mol$^{-1}$ HA$^{-1}$)[a] | T/F |
|---|---|---|---|---|
| Compound A |  | -10.23 | -0.28 | |
| Frag-0 |  | -3.60 | -0.43 | T |
| Frag-1 |  | -2.57 | -0.42 | T |
| Frag-2 |  | -0.42 | -0.42 | F |
| Frag-3 |  | -6.10 | -0.55 | F |
| Frag-4 |  | -3.12 | -0.56 | F |
| Frag-5 |  | -4.59 | -0.45 | T |
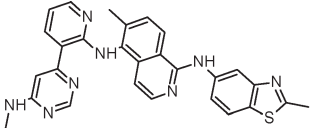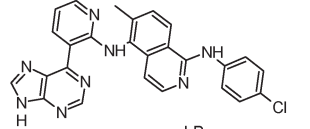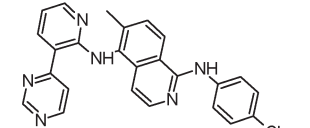
[a] HA is the number of non-hydrogen atoms in the fragment. [b] The fragments selected to be replaced are marked as T and those preserved are marked as F.

final proposed compounds, while Frag-0 and Frag-5 were replaced with a purine and a phenylchloro group, respectively. It is interesting that compound **B**, generated with the LeadOp algorithm, is the same structure as the original ligand (inhibitor L1E) of the cocrystal structure (PDB ID: 3idp). Compound **C** kept Frag-1 in its final state while Frag-0 and Frag-5 were replaced with pyrimidine and phenylchloro groups, respectively. Compound **D** retained Frag-1 in the final compound, and Frag-0 and Frag-5 were replaced with pyrimidine and trifluoromethylphenyl groups, respectively. Compound **E** combined Frag-0 and Frag-1, resulting in Frag-0, yet

Frag-5 was replaced with the phenylchloro group. The detailed rankings from our algorithm for the compounds **B**–**E**, **X**, and **Y** on the basis of biologically measured IC$_{50}$, depicted structure, and the predicted binding energy are reported in Table 2.

Molecular dynamics simulation studies were performed to further investigate the resulting ligand–receptor interactions as suggested by our algorithm (LeadOp) and to explore the possible interactions within the cocrystal complex of B-Raf and compound L1E.[25] The generated compounds **B**–**E** were energically optimized and docked into the receptor's binding site as

**Table 2. Ranking of the New Compounds Generated by the LeadOp Algorithm and Their Biologically Determined Inhibition Potency ($IC_{50}$) of B-Raf from the Literature[42] [a]**

| Rank | Compound | $IC_{50}$ (nM) | Predicted binding energy ($kcal \cdot mol^{-1}$) | Original rank |
|---|---|---|---|---|
| Query | compound A | 110 | -10.23 | |
| 1 | compound B | 1.6 | -12.64 | 21 |
| 2 | compound C | 3.4 | -12.53 | 65 |
| 3 | compound D | 17 | -11.86 | 584 |
| 4 | compound X | 18 | -11.37 | 1035 |
| 5 | compound Y | 39 | -11.11 | 1371 |
| 6 | compound E | 56 | -10.83 | 2056 |

[a] All new compounds have a higher potency than the query compound, and the suggested priority of the new compounds with the predicted binding energy as well as their original rankings (out of 5576) from the algorithm have a similar trend as the $IC_{50}$ potency values from the literature.

described previously in the Materials and Methods. Molecular dynamics simulation studies were performed with the final poses of the compounds **B**−**E** with respect to B-Raf, and the unique low-energy conformations of the complexes, from the last 50 ps of the MDS (50 configurations), are shown in Figure 3b.

The available cocrystal of the B-Raf-L1E complex shows hydrogen-bonding interactions between Cys532 of B-Raf and the purine group, hydrogen bond interactions between Glu501 and a nitrogen atom connecting two aromatic groups, a hydrogen bond between an aromatic nitrogen of L1E and a bound water
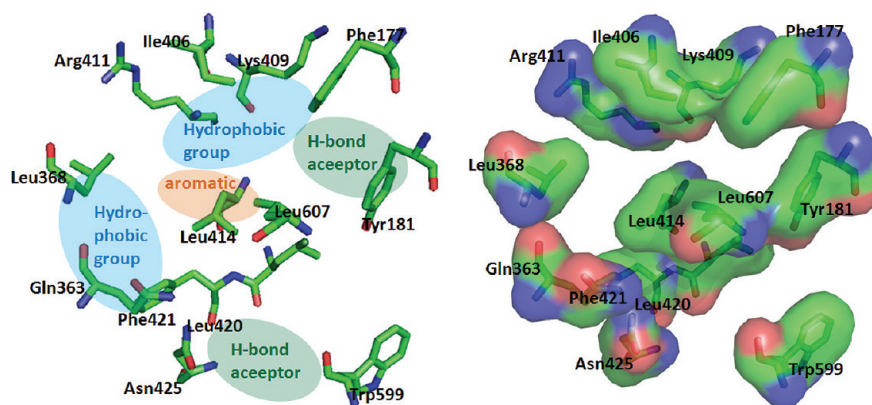
**Figure 4.** Schematic representation of the human 5-LOX active site (left) and the binding pocket (right). The perceived pharmacophores of the binding site of 5-LOX involve two hydrophobic groups (blue ovals), two hydrogen-bond acceptors (green ovals), and an aromatic ring (orange oval) for ligand binding at the binding cavity.

that is hydrogen bonded to Asp594 and Lys483 of B-Raf, and a potential favorable $\pi$-stacking interaction with the side chain of Trp531.[25] We observe similar hydrogen-bonding interactions between the aminoisoquinoline group in compound **B** with binding site residues Asp594 and Glu507 and between the purine group of compound **B** and residues Leu463 and Cys532 of the receptor.

Compound **C** has a similar set of hydrogen bond interactions—as compared to B-Raf—L1E complex—between itself and Asp594 and Cys532 along with two additional hydrogen-bond interactions with residues Lys483 and Thr529. Compounds **D** and **E** also display key hydrogen-bond interactions that are similar to those between L1E's three nitrogen groups and the surrounding binding site residues (the nitrogen atom bridging two aromatic ring groups and Glu501, a nitrogen atom in an aromatic ring and Asp594 via a bound water molecule, and two nitrogen atoms of an aromatic ring group and the backbone hydrogen-bond acceptor and donor of Cys532).

**Structure-Based Fragment Hopping of Human 5-Lipoxygenase Inhibitors.** The human 5-lipoxygenase (5-LOX) enzyme with the well-known 5-LOX inhibitors was selected as the second LeadOp test case. To design better 5-LOX inhibitors, structural insight of the 5-LOX active site and its associated interactions with ligands would be helpful; unfortunately, the crystal structure of this enzyme has yet to be elucidated. We selected a theoretical model (comparative/homology protein structure/model) of 5-LOX[37] that has good agreement with mutagenesis studies.[47,48] The proposed active site of 5-LOX forms a deep and bent cleft that extends from Phe177 and Tyr181 on the top of the cleft to the Trp599 and Leu420 at the bottom (shown in Figure 4). Most of the residues lining the cleft are hydrophobic with several polar residues (Gln363, Asn425, Gln557, Ser608, and Arg411) distributed along the channel that have the ability to interact with ligands during the binding process. A small side pocket off of the main channel is composed of hydrophobic residues (Phe421, Gln363, and Lue368), and it is postulated that lipophilic interactions with the ligand may enhance activity. The purported major pharmacophore interactions needed for ligand binding to 5-LOX include (i) two hydrophobic groups, (ii) a hydrogen-bond acceptor, (iii) an aromatic ring, and (iv) two secondary interactions. These two secondary interactions are between the ligand and an acidic moiety and a hydrogen-bond acceptor within the binding pocket

of the receptor. The hydrogen-bond acceptor probably interacts with the key anchoring points (Tyr181, Asn425, and Arg411) to form the hydrogen bond, while Leu414 and Phe421 form a hydrophobic interaction between the ligand and the binding cavity.[37]

The 5-LOX inhibitor compound **F** (compound **6** in the literature[26]) was selected as our initial molecule for lead optimization and has a biologically determined $IC_{50}$ value of 145 nM. Compound **F** was docked into the theoretical 5-LOX binding site and the lowest energy conformation was submitted to LeadOp. This selected conformation possesses similar interactions that have been previously reported[37] and discussed above within at the 5-LOX active site (Figure 4). The oxochromen group favorably interacts with the hydrophobic residue Leu414 ($CH\cdots\pi$ interaction) in the middle of the cavity, while the fluorophenyl group extends to the hydrogen-bond-acceptor region in the lower cleft of the active site. The docked conformation was selected as the query molecule and was decomposed into the five fragments shown in Figure 5a.

The group efficiency was evaluated for each of the decomposed fragments to determine if it is eligible for replacement. The oxochromen and fluorophenyl groups (Frag-0 and Frag-1 in Table 3, respectively) were considered the largest contributing features for ligand binding to 5-LOX according to the literature[37] and our observations from the docking simulation, decomposition, and group efficiency calculation. On the basis of these circumstances, the oxochromen and fluorophenyl groups were therefore preserved during the replacement portion of LeadOp. As in the B-Raf example, LeadOp can identify analogs (compounds **G**−**I** in Figure 5b) that were previously proposed, synthesized, and had their biological end points measured while also discovering compound **F** in the literature.[26]

In the final set of proposed compounds, compound **G** (the strongest inhibitor among those that were previously proposed; $IC_{50}$ = 10 nM) and compound **I** ($IC_{50}$ = 130 nM) were the most potent; compound **G** was generated by replacing Frag-2, Frag-3, and Frag-4 of compound **F** with a secondary amine, an oxadiazole ring, and a $-C(CH_2CH_3)(CF_3)OH$, respectively, and compound **I** was created by replacing Frag-4 of compound **F** with $-C(CH_2CH_3)_2OH$. Compound **H** ($IC_{50}$ = 64 nM) preserved Frag-3 and Frag-4 of compound **F**, while Frag-2 was replaced with an alkyl group. The three compounds suggested by LeadOp, based on the query molecule compound **F**, were ranked with
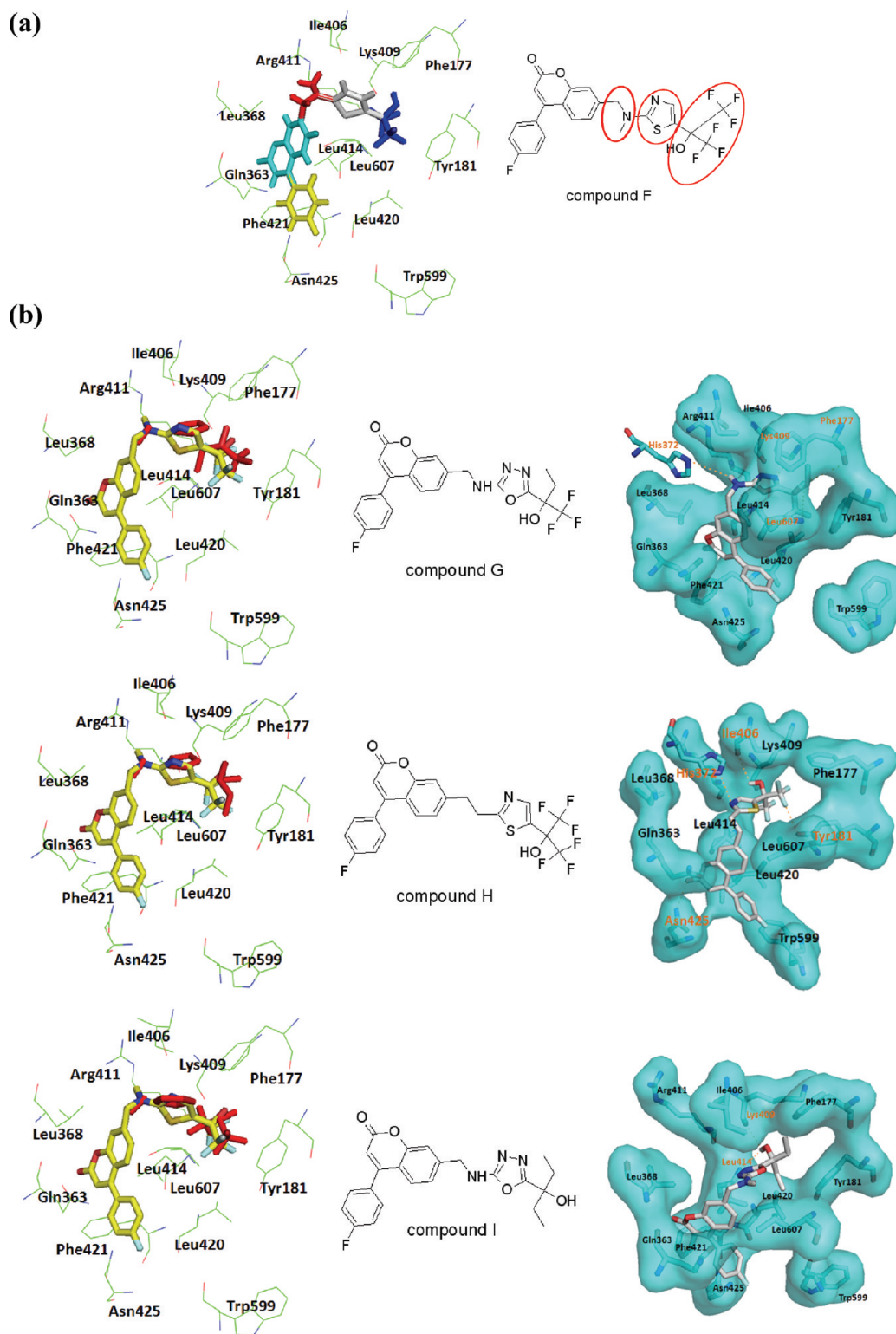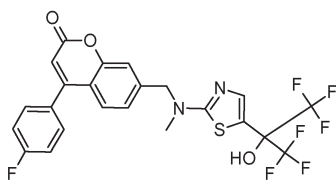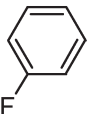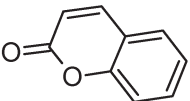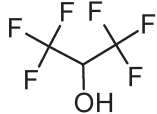
**Figure 5.** LeadOp result in 5-LOX model system: (a) Each fragment of compound **F** is colored differently (left). The red ovals indicated the fragments selected to be replaced (right). (b) The original compound **F** carbon atoms are colored yellow and the new fragments' carbon atoms, of the generated compound (middle), are colored red (left). Amino acid residues that participate in hydrogen-bonding interactions with the proposed compound, at the binding site, (right) are depicted with cyan molecular surfaces.

respect to their predicted binding energy. Depicted representations of compounds **F**−**I**, as well as the corresponding inhibition data from the biological experiments and their predicted binding energy, are listed in Table 4.

The three LeadOp proposed compounds were submitted to molecular dynamics simulations (MDSs) to analyze the ligand−receptor interactions within the 5-LOX active site. Figure 5b displays the last conformation from the MDS along

**Table 3. Evaluation of the Five Fragments, Frag-0 to Frag-4, from Compound F, a Human 5-LOX Inhibitor with Binding Free Energy ($\Delta G$) and Group Efficiency (GrpEff)[b]**

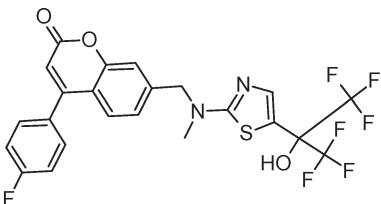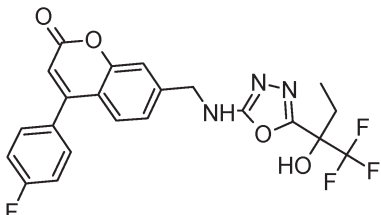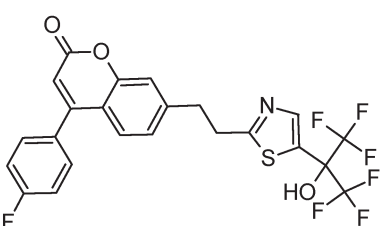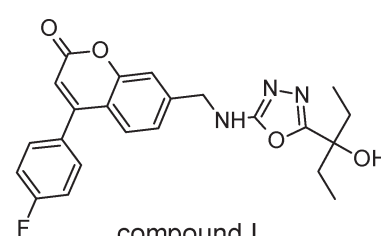| | Structure | $\Delta G$ (kcal·mol$^{-1}$) | $\Delta GE$ (kcal·mol$^{-1}$ HA$^{-1a}$) | T/F |
|---|---|---|---|---|
| Compound F | | -7.00 | -0.19 | |
| Frag-0 | | -3.04 | -0.44 | F |
| Frag-1 | | -5.32 | -0.48 | F |
| Frag-2 | | -0.87 | -0.29 | T |
| Frag-3 | | -1.23 | -0.25 | T |
| Frag-4 | | -3.41 | -0.34 | T |

[a] HA is the number of non-hydrogen atoms in the fragment. [b] The fragments selected to be replaced are marked as T and those preserved are marked as F.

with the interaction between each ligand and the 5-LOX binding site. The interactions of compounds **G–I** all contain the hydrogen-bonding interactions between the oxygen or nitrogen atoms of the thiazol group at the Frag-2 or Frag-3 position. In compounds **G** and **H**, the fluoro group at the Frag-4 position extends to the hydrogen-bond acceptor in the upper domain of the active site and interacts with Lys409 through hydrogen bonding. In addition, the oxochromen ring of Frag-1 is in close proximity to Leu414 and is potentially an important CH$\cdots\pi$ contact, as indicated in the literature.[37] Also, Frag-3 of compound **G** interacts with 5-LOX hydrophobic residues Leu420 and Leu607, which have been suggested to improve binding in the 5-LOX system via complementary hydrophobic interaction between the ligand and receptor, which probably explains

compound **G**'s better inhibition compared to compounds **F**, **H**, and **I**. These optimized results indicate that hydrogen-bonding and hydrophobic interactions are important for ligands binding to and inhibition of 5-LOX, as previously reported.[37]

The diversity of the fragment database is a critical factor when searching for substituent fragments. The number of different poses determined by docking fragments to each binding location is always important. The more substructural classes and docked conformations in the fragment database, for the system of interest, results in a greater number of possible combinations that are available to generate new compounds. As LeadOp is an optimization algorithm that starts with a query molecule, better lead optimization occurs when starting with a strong inhibitor.

**Table 4. Ranking of the New Compounds Generated by the LeadOp Algorithm and the Inhibition Potency (IC$_{50}$) of Human 5-LOX from the Literature[43] [a]**

| Rank | Compound | IC$_{50}$(nM) | Predicted binding energy ( kcal·mol$^{-1}$ ) | Original rank |
|------|----------|---------------|----------------------------------------------|---------------|
| Query |  compound F | 145±(15) | -7.00 | |
| 1 |  compound G | 10±(3) | -10.56 | 81 |
| 2 |  compound H | 64±(3) | -9.92 | 204 |
| 3 |  compound I | 130±(25) | -7.05 | 1339 |

[a] All new compounds have a higher potency than the query compound, and the suggested priority of the new compounds with the predicted binding energy as well as their original rankings (out of 1637) from the algorithm have a similar trend as the IC$_{50}$ potency values from the literature.

## ■ CONCLUSION

In this study, we have implemented a structure-based ligand optimization algorithm called "LeadOp". The mutant B-Raf and human 5-LOX series of inhibitors was selected to demonstrate how structured-based fragment hopping can improve the ligand—receptor interactions from a query molecule by optimizing the query molecule through decomposition and selection of alternative substructures from thousands of fragments in our in-house library. In the case of mutant B-Raf and its inhibitors, LeadOp generates a set of potential compounds that exhibit better inhibition. The docking and molecular dynamic simulation analysis demonstrates that the generated structures preserve the important ligand—receptor interactions as seen in the crystal structures. In the case of human of 5-LOX, without a cocrystal structure, lead optimization was performed with a theoretical 5-LOX receptor model (comparative or homology protein model) and a known inhibitor. The optimization process was able to improve the known inhibitor through the decomposition and replacement of fragments with substituents that possess better efficiency score(s).

1713

dx.doi.org/10.1021/ci200136j |*J. Chem. Inf. Model.* 2011, 51, 1703–1715

For the proposed compounds with biological inhibition values ($IC_{50}$) from the literature, LeadOp calculated inhibition values correspond to the literature values when ranking the generated structures according to binding energy. The interactions between the inhibitors and proteins, as noted in the literature, were observed in the molecular dynamic simulation. Moreover, we observed fragments with more ligand—protein interactions than the original fragment.

In short, LeadOp is an algorithm that can automatically optimize a query molecule by searching and replacing fragments from a predocked fragment database in the active site to generate structures with better binding without prior knowledge of better fragments. Also, users can specify parts of structures to be optimized on the basis of known interactions or preference.

# ■ AUTHOR INFORMATION

**Corresponding Author**

* Phone: +886.2.3366.4888 #529; fax:+886.2.23628167; e-mail: yjtseng@csie.ntu.edu.tw.

# ■ ACKNOWLEDGMENT

# ■ REFERENCES

(1) Brown, N.; Lewis, R. A. Exploiting QSAR Methods in Lead Optimization. *Curr. Opin. Drug Discovery Dev.* **2006**, *9*, 419–424.

(2) Zhang, Q.; Muegge, I. Scaffold Hopping through Virtual Screening Using 2D and 3D Similarity Descriptors: Ranking, Voting, and Consensus Scoring. *J. Med. Chem.* **2006**, *49*, 1536–1548.

(3) Senger, S. Using Tversky Similarity Searches for Core Hopping: Finding the Needles in the Haystack. *J. Chem. Inf. Model.* **2009**, *49*, 1514–1524.

(4) Böhm, H.-J.; Flohr, A.; Stahl, M. Scaffold Hopping. *Drug Discovery Today: Technol.* **2004**, *1*, 217–224.

(5) Nilakantan, R.; Bauman, N.; Venkataraghavan, R. New Method for Rapid Characterization of Molecular Shapes: Applications in Drug Design. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 79–85.

(6) Langer, T.; Krovat, E. M. Chemical Feature-Based Pharmacophores and Virtual Library Screening for Discovery of New Leads. *Curr. Opin. Drug Discovery Dev.* **2003**, *6*, 370–376.

(7) Sykes, M. J.; Sorich, M. J.; Miners, J. O. Molecular Modeling Approaches for the Prediction of the Nonspecific Binding of Drugs to Hepatic Microsomes. *J. Chem. Inf. Model.* **2006**, *46*, 2661–2673.

(8) Vogt, M.; Stumpfe, D.; Geppert, H.; Bajorath, J. Scaffold Hopping Using Two-Dimensional Fingerprints: True Potential, Black Magic, or a Hopeless Endeavor? Guidelines for Virtual Screening. *J. Med. Chem.* **2010**, *53*, 5707–5715.

(9) Kirchmair, J.; Distinto, S.; Markt, P.; Schuster, D.; Spitzer, G. M.; Liedl, K. R.; Wolber, G. How To Optimize Shape-Based Virtual Screening: Choosing the Right Query and Including Chemical Information. *J. Chem. Inf. Model.* **2009**, *49*, 678–692.

(10) Fitzgerald, S. H.; Sabat, M.; Geysen, H. M. Diversity Space and Its Application to Library Selection and Design. *J. Chem. Inf. Model.* **2006**, *46*, 1588–1597.

(11) Hajduk, P. J.; Greer, J. A Decade of Fragment-Based Drug Design: Strategic Advances and Lessons Learned. *Nat. Rev. Drug Discovery* **2007**, *6*, 211–219.

(12) Ertl, P. Cheminformatics Analysis of Organic Substituents: Identification of the Most Common Substituents, Calculation of Substituent Properties, and Automatic Identification of Drug-like Bioisosteric Groups. *J. Chem. Inf. Comput. Sci.* **2002**, *43*, 374–380.

(13) Lewell, X. Q.; Jones, A. C.; Bruce, C. L.; Harper, G.; Jones, M. M.; McLay, I. M.; Bradshaw, J. Drug Rings Database with Web Interface. A Tool for Identifying Alternative Chemical Rings in Lead Discovery Programs. *J. Med. Chem.* **2003**, *46*, 3257–3274.

(14) Weininger, D. SMILES, A Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. *J. Chem. Inf. Comput. Sci.* **1988**, *28*, 31–36.

(15) 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.

(16) Fechner, U.; Schneider, G. Flux (2): Comparison of Molecular Mutation and Crossover Operators for Ligand-Based de Novo Design. *J. Chem. Inf. Model.* **2007**, *47*, 656–667.

(17) Grant, M. A. Protein Structure Prediction in Structure-Based Ligand Design and Virtual Screening. *Comb. Chem. High Throghput Screening* **2009**, *12*, 940–960.

(18) Bergmann, R.; Linusson, A.; Zamora, I. SHOP: Scaffold HOPping by GRID-Based Similarity Searches. *J. Med. Chem.* **2007**, *50*, 2708–2717.

(19) Goodford, P. J. A Computational Procedure for Determining Energetically Favorable Binding Sites on Biologically Important Macromolecules. *J. Med. Chem.* **1985**, *28*, 849–857.

(20) Dey, F.; Caflisch, A. Fragment-Based de Novo Ligand Design by Multiobjective Evolutionary Optimization. *J. Chem. Inf. Model.* **2008**, *48*, 679–690.

(21) Wishart, D. S.; Knox, C.; Guo, A. C.; Cheng, D.; Shrivastava, S.; Tzur, D.; Gautam, B.; Hassanali, M. DrugBank: A Knowledgebase for Drugs, Drug Actions and Drug Targets. *Nucleic Acids Res.* **2008**, *36*, D901–D906.

(22) Orita, M.; Ohno, K.; Niimi, T. Two 'Golden Ratio' Indices in Fragment-Based Drug Discovery. *Drug Discovery Today* **2009**, *14*, 321–328.

(23) Congreve, M.; Chessari, G.; Tisi, D.; Woodhead, A. J. Recent Developments in Fragment-Based Drug Discovery. *J. Med. Chem.* **2008**, *51*, 3661–3680.

(24) 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.

(25) Smith, A. L.; DeMorin, F. F.; Paras, N. A.; Huang, Q.; Petkus, J. K.; Doherty, E. M.; Nixey, T.; Kim, J. L.; Whittington, D. A.; Epstein, L. F.; Lee, M. R.; Rose, M. J.; Babij, C.; Fernando, M.; Hess, K.; Le, Q.; Beltran, P.; Carnahan, J. Selective Inhibitors of the Mutant B-Raf Pathway: Discovery of a Potent and Orally Bioavailable Aminoisoquinoline. *J. Med. Chem.* **2009**, *52*, 6189–6192.

(26) Ducharme, Y.; Blouin, M.; Brideau, C.; Chateauneuf, A.; Gareau, Y.; Grimm, E. L.; Juteau, H.; Laliberte, S.; MacKay, B.; Masse, F.; Ouellet, M.; Salem, M.; Styhler, A.; Friesen, R. W. The Discovery of Setileuton, a Potent and Selective 5-Lipoxygenase Inhibitor. *ACS Med. Chem. Lett.* **2010**, *1*, 170–174.

(27) Pipaon, C.; Gutierrez, P.; Montero, J. C.; Lorenzo, M.; Eguiraun, A.; De la Fuente, J. A.; Pandiella, A.; Leon, J.; Ortiz, J. M. Mitogen-Activated Protein Kinase Routes as Targets in the Action of Diaza-Anthracene Compounds with a Potent Growth-Inhibitory Effect on Cancer Cells. *Mol. Cancer Ther.* **2002**, *1*, 811–819.

(28) Wellbrock, C.; Karasarides, M.; Marais, R. The RAF Proteins Take Centre Stage. *Nat. Rev. Mol. Cell Biol.* **2004**, *5*, 875–885.

(29) Flaherty, K. T.; Puzanov, I.; Kim, K. B.; Ribas, A.; McArthur, G. A.; Sosman, J. A.; O'Dwyer, P. J.; Lee, R. J.; Grippo, J. F.; Nolop, K.;

1714

dx.doi.org/10.1021/ci200136j |*J. Chem. Inf. Model.* 2011, 51, 1703–1715

Chapman, P. B. Inhibition of Mutated, Activated BRAF in Metastatic Melanoma. *N. Engl. J. Med.* **2010**, *363*, 809–819.

(30)  Garnett, M. J.; Marais, R. Guilty as Charged: B-RAF Is a Human Oncogene. *Cancer Cell* **2004**, *6*, 313–319.

(31)  Tsai, J.; Lee, J. T.; Wang, W.; Zhang, J.; Cho, H.; Mamo, S.; Bremer, R.; Gillette, S.; Kong, J.; Haass, N. K.; Sproesser, K.; Li, L.; Smalley, K. S. M.; Fong, D.; Zhu, Y. L.; Marimuthu, A.; Nguyen, H.; Lam, B.; Liu, J.; Cheung, I.; Rice, J.; Suzuki, Y.; Luu, C.; Settachatgul, C.; Shellooe, R.; Cantwell, J.; Kim, S. H.; Schlessinger, J.; Zhang, K. Y. J.; West, B. L.; Powell, B.; Habets, G.; Zhang, C.; Ibrahim, P. N.; Hirth, P.; Artis, D. R.; Herlyn, M.; Bollag, G. Discovery of a Selective Inhibitor of Oncogenic B-Raf Kinase with Potent Antimelanoma Activity. *Proc. Natl. Acad. Sci. U. S. A.* **2008**, *105*, 3041–3046.

(32)  Steele, V. E.; Holmes, C. A.; Hawk, E. T.; Kopelovich, L.; Lubet, R. A.; Crowell, J. A.; Sigman, C. C.; Kelloff, G. J. Lipoxygenase Inhibitors as Potential Cancer Chemopreventives. *Cancer Epidemiol., Biomarkers Prev.* **1999**, *8*, 467–483.

(33)  Ivanov, I.; Heydeck, D.; Hofheinz, K.; Roffeis, J.; O'Donnell, V. B.; Kuhn, H.; Walther, M. Molecular Enzymology of Lipoxygenases. *Arch. Biochem. Biophys.* **2010**, *503*, 161–174.

(34)  Woods, J. W.; Evans, J. F.; Ethier, D.; Scott, S.; Vickers, P. J.; Hearn, L.; Heibein, J. A.; Charleson, S.; Singer, I. I. 5-Lipoxygenase and 5-Lipoxygenase-Activating Protein Are Localized in the Nuclear Envelope of Activated Human Leukocytes. *J. Exp. Med.* **1993**, *178*, 1935–1946.

(35)  Shaffer, D. N.; Mansmann, P. T. Leukotriene Inhibition and Advances in the Treatment of Asthma: A Pharmacological Review. *Pediatr. Asthma Allergy* **1997**, *11*, 171–179.

(36)  Glover, F. Future Paths for Integer Programming and Links to Artificial Intelligence. *Comput. Oper. Res.* **1986**, *13*, 533–549.

(37)  Charlier, C.; Hénichart, J.-P.; Durant, F.; Wouters, J. Structural Insights into Human 5-Lipoxygenase Inhibition: Combined Ligand-Based and Target-Based Approach. *J. Med. Chem.* **2006**, *49*, 186–195.

(38)  Kolb, P.; Caflisch, A. Automatic and Efficient Decomposition of Two-Dimensional Structures of Small Molecules for Fragment-Based High-Throughput Docking. *J. Med. Chem.* **2006**, *49*, 7384–7392.

(39)  Pommery, N.; Taverne, T.; Telliez, A.; Goossens, L.; Charlier, C.; Pommery, J.; Goossens, J.-F.; Houssin, R.; Durant, F.; Hénichart, J.-P. New COX-2/5-LOX Inhibitors: Apoptosis-Inducing Agents Potentially Useful in Prostate Cancer Chemotherapy. *J. Med. Chem.* **2004**, *47*, 6195–6206.

(40)  Barbey, S.; Goossens, L.; Taverne, T.; Cornet, J.; Choesmel, V.; Rouaud, C.; Gimeno, G.; Yannic-Arnoult, S.; Michaux, C.; Charlier, C.; Houssin, R.; Henichart, J. P. Synthesis and Activity of a New Methoxytetrahydropyran Derivative as Dual Cyclooxygenase-2/5-Lipoxygenase Inhibitor. *Bioorg. Med. Chem. Lett.* **2002**, *12*, 779–782.

(41)  Majeux, N.; Scarsi, M.; Apostolakis, J.; Ehrhardt, C.; Caflisch, A. Exhaustive Docking of Molecular Fragments with Electrostatic Solvation. *Proteins: Struct. Funct. Genet.* **1999**, *37*, 88–105.

(42)  HyperChem, Version 7.0; Hypercube, Inc.: Gainesville, FL, 2007.

(43)  Trott, O.; Olson, A. J. Software News and Update AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010**, *31*, 455–461.

(44)  Hess, B.; Kutzner, C.; van der Spoel, D.; Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem. Theory Comput.* **2008**, *4*, 435–447.

(45)  Oostenbrink, C.; Soares, T. A.; van der Vegt, N. F. A.; van Gunsteren, W. F. Validation of the 53A6 GROMOS Force Field. *Eur. Biophys. J.* **2005**, *34*, 273–284.

(46)  Berendsen, H. P., JPM; van Gunsteren, W. F.; Hermans, J. Interaction Models for Water in Relation to Protein Hydration. In *Intermolecular Forces*; Pullman, B., Ed.; Reidel: Dordrecht, The Netherlands. 1981; pp 331−342.

(47)  Hammarberg, T.; Zhang, Y. Y.; Lind, B.; Radmark, O.; Samuelsson, B. Mutations at the C-Terminal Isoleucine and Other Potential Iron Ligands of 5-Lipoxygenase. *Eur. J. Biochem.* **1995**, *230*, 401–407.

(48)  Schwarz, K.; Walther, M.; Anton, M.; Gerth, C.; Feussner, I.; Kuhn, H. Structural Basis for Lipoxygenase Specificity. *J. Biol. Chem.* **2001**, *276*, 773–779.

1715

dx.doi.org/10.1021/ci200136j |*J. Chem. Inf. Model.* 2011, 51, 1703–1715