

# LigMerge: A Fast Algorithm to Generate Models of Novel Potential Ligands from Sets of Known Binders

Steffen Lindert<sup>1,\*†</sup>, Jacob D. Durrant<sup>2,†</sup>  
and J. Andrew McCammon<sup>1,3,4</sup>

<sup>1</sup>Department of Pharmacology, University of California San Diego, La Jolla, CA 92093, USA

<sup>2</sup>Department of Chemistry and Biochemistry, University of California San Diego, La Jolla, CA 92093, USA

<sup>3</sup>Department of Chemistry and Biochemistry, NSF Center for Theoretical Biological Physics, National Biomedical Computation Resource, University of California San Diego, La Jolla, CA 92093, USA

<sup>4</sup>Howard Hughes Medical Institute, University of California San Diego, La Jolla, CA 92093, USA

\*Corresponding author: Steffen Lindert, slindert@ucsd.edu

†These authors contributed equally to this work.

**One common practice in drug discovery is to optimize known or suspected ligands in order to improve binding affinity. In performing these optimizations, it is useful to look at as many known inhibitors as possible for guidance. Medicinal chemists often seek to improve potency by altering certain chemical moieties of known/endogenous ligands while retaining those critical for binding. To our knowledge, no automated, ligand-based algorithm exists for systematically 'swapping' the chemical moieties of known ligands to generate novel ligands with potentially improved potency. To address this need, we have created a novel algorithm called 'LigMerge'. LigMerge identifies the maximum (largest) common substructure of two three-dimensional ligand models, superimposes these two substructures, and then systematically mixes and matches the distinct fragments attached to the common substructure at each common atom, thereby generating multiple compound models related to the known inhibitors that can be evaluated using computer docking prior to synthesis and experimental testing. To demonstrate the utility of LigMerge, we identify compounds predicted to inhibit peroxisome proliferator-activated receptor gamma, HIV reverse transcriptase, and dihydrofolate reductase with affinities higher than those of known ligands. We hope that LigMerge will be a helpful tool for the drug design community.**

**Key words:** biophysical chemistry, drug design, structure-based drug design

Received 23 February 2012, revised 13 April 2012 and accepted for publication 8 May 2012

Given the exponential growth of computer speed and power, the role computers play in modern drug discovery, already important, is likely to increase in prominence in coming years. Virtual screening is one application of computer-aided drug design that is already commonplace. Rather than testing millions of compounds in high-throughput screens, experiments that are costly in both time and treasure, many researchers first use docking programs to predict small-molecule binding *in silico*. Virtual screening approaches enrich a pool of candidate ligands for true binders; only a limited number of the best-scoring compounds are then tested experimentally, leading to greater hit rates and decreased cost (1–3). These methodologies have been used successfully to identify many experimentally validated ligands, including inhibitors of *Trypanosoma brucei*-RNA editing ligase 1 (4,5), *T. brucei* UDP-galactose 4'-epimerase (6), *T. brucei* farnesyl diphosphate synthase (7), *M. tuberculosis* dTDP-6-deoxy-L-lyxo-4-hexulose reductase (8), and *H. sapiens* stromelysin-1 (9).

Critical to any virtual screening project is the selection of a good database of small-molecule models whose real-world counterparts are readily available for experimental validation. These databases generally consist of compounds carefully designed to represent diverse scaffolds (i.e., diversity sets), compounds derived from common reactions (combinatorial libraries), compounds with known pharmacological properties (e.g., the set of all approved drugs), or analogs of known ligands.

In part because of the advent of high-throughput screening, many protein receptors are associated with a plethora of experimentally validated ligands (10). In designing novel small-molecule databases for virtual screening, it makes sense to consider the pharmacophoric features of known ligands. New ligands that combine the observed features of validated binders are more likely to be potent binders themselves.

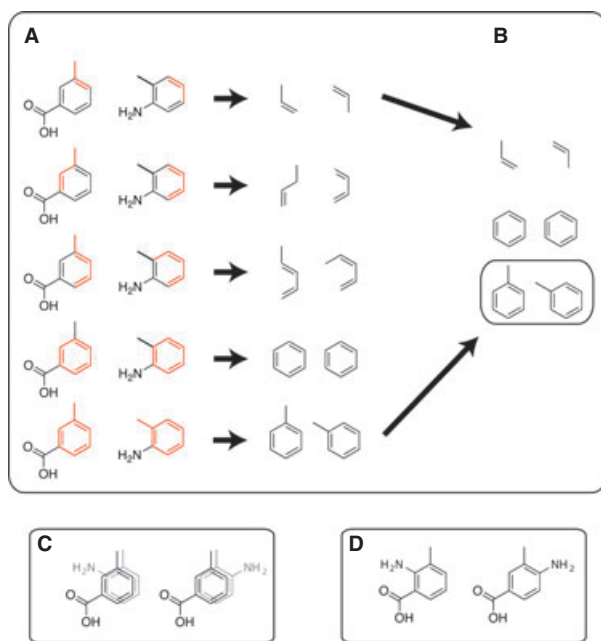
BREED (11), an algorithm developed by Vertex pharmaceuticals, overlays known receptor–ligand complexes to generate novel ligands that bind with improved affinity. BREED is a receptor-based algorithm that relies on the presence of high-resolution crystal or NMR structures to overlay known ligands. To our knowledge, there is no stand-alone, ligand-based tool for recombining the three-dimensional structures of known ligands into novel potential binders.

Here, we present a program called LigMerge that provides a fast and easy way to generate molecular models derived from known inhibitors without the need for information about the receptor. We expect the program will be useful for those designing custom virtual screening, small-molecule databases when many ligands, potent or otherwise, have been identified experimentally or theoretically *via* virtual screening. LigMerge is implemented in Python and so is easily editable, customizable, and platform independent. A copy can be downloaded free of charge from <http://www.nbcrc.net/ligmerge/>.

## Materials and Methods

### The LigMerge algorithm

As input, LigMerge accepts two three-dimensional, PDB-formatted compound models. PDB files are the only supported input format. SDF or MOL files must be converted to the PDB format before using LigMerge. These models are processed in three steps. First, the maximum (largest) common substructure of the two models is identified (Figure 1A,B). Second, the two models are translated and rotated, so that these two substructures are superimposed (Figure 1C). Third, the two models are merged by mixing and matching



**Figure 1:** A schematic representing the LigMerge algorithm. (A) Stretches of connected atoms consisting of identical elements in sequence are identified from two distinct compounds. (B) Those stretches of connected atoms that have identical geometries are identified as common substructures. The maximum (largest) common substructure is subsequently identified (highlighted in a separate box). (C) The two distinct compounds are aligned so that their greatest common substructures are superimposed. All possible superimpositions are considered. (D) Novel compounds are generated by mixing and matching the moieties connected to each of the superimposed atoms of the maximum common substructure.

the distinct fragments of each model attached at each common, superimposed atom (Figure 1D).

### Finding the maximum common substructure (MCS)

Exhaustive lists of atom indices/element types for all heavy atoms in the two structures are first generated (Figure 1A). Hydrogen atoms are not included in this analysis. Stretches of connected atoms comprised of the same sequence of elements occurring in both structures are identified and stored, regardless of geometry. As no structural information beyond connectivity is encoded in these lists, the criterion for consideration is necessary but not sufficient for identifying a common substructure. Many of the identified common fragments will eventually be rejected for having distinct geometries, but all true common substructures are nevertheless among those enumerated. The shortest stretches considered are three-atom fragments, as shorter fragments (i.e., single atoms or mere pairs of bonded atoms) cannot reasonably be considered distinctive common substructures. Consecutively, larger fragments are likewise stored. While ideally MCSs of at least ten atoms are preferable to ensure as unique an overlay as possible, we judge three to be sufficient in extreme cases because, in addition to connectivity, the algorithm will eventually also account for the three-dimensional structures of these models. While three is set as the program default, the minimum number of common atoms can also be specified explicitly by the user.

Having identified candidate common substructures, the next step is to test for identical geometries (Figure 1B). To facilitate geometric comparison, a sorted pairwise distance matrix (i.e., a distance 'fingerprint') describing the distance between all atom pairs is calculated for each fragment. Two fragments are considered geometrically identical if all pairwise distances are identical within a specified tolerance. Comparisons between fragments begin with the largest candidate substructures; subsequently, smaller candidates are considered if larger candidates are found to be geometrically dissimilar. Setting the `-output_mcs` command-line parameter to true causes the program to output the maximum common substructure in addition to merged compound models.

Without further consideration, the above protocol ignores questions of symmetry. For example, consider two models whose greatest common substructure is a toluene. Two symmetry-related superimpositions exist (i.e., two rotations about the axis defined by the methyl-phenyl bond). The `-all_symmetry_relations` command-line parameter can be used to specify whether the algorithm should consider all symmetry assignments when generating merged compound models or whether it should randomly choose a single assignment from those identified (Figure 1C). The `-all_symmetry_relations` command-line parameter only creates multiple ligands if the overlay of the determined MCS is ambiguous.

It is important to note that LigMerge ignores ligand flexibility when performing geometric comparisons. It is therefore prudent to use ligand models of compounds in docked or crystallographic poses or to choose ligand models with inflexible segments (e.g., benzene rings). As the number of publicly available crystal structures is ever

increasing and inflexible segments are common in bioactive molecules, we expect that these two limitations will not be too problematic.

### Superimposition and fragment merging

All possible substructure assignments of maximum length and identical geometry are subsequently considered. For each, a transformation matrix is identified that minimizes the RMSD between substructures. Although only the common substructures are considered in generating this matrix, entire molecules are subjected to the transformation, essentially positioning the two models so that their maximum common substructures are superimposed (Figure 1C). When identifying the MCS, LigMerge is sensitive to the ligand conformation. If the user wishes to consider multiple ligand conformations, they need only to provide multiple pdb files representing each conformation.

Moieties from each model, comprised of fragments with atoms bound to those of the common substructure, are next identified. The common-substructure atoms to which these moieties are bound are designated 'handle atoms'. If the command-line parameter `-all_substituent_combinations` is set to false, a random fragment is selected for each handle atom, and a single merged compound is generated by combining the common substructure and the selected fragments. Special consideration is given to 'multiple-handle fragments', that is, fragments that externally connect to two or more handle atoms. If fragments containing more than one handle atom are selected, these fragments essentially determine the selection at multiple handle-atom locations. If the command-line parameter `-all_substituent_combinations` is set to true, multiple merged structures with all possible combinations of fragments are generated and saved to separate PDB files (Figure 1D). If a specific fragment combination will create a molecule with steric clashes between the fragments, the merged molecule will not be generated and the fragment combination will be skipped.

### Docking of LigMerge-generated compounds

To demonstrate the utility of the LigMerge algorithm, compounds generated by applying LigMerge to a variety of known binders were docked into three receptors: peroxisome proliferator-activated receptor (PPAR) gamma, HIV reverse transcriptase (RT), and dihydrofolate reductase (DHFR).

### Receptor preparation

Crystal structures of peroxisome proliferator-activated receptor (PPAR) gamma in complex with ligand 570 (PDB ID: 1FM9 (12)), HIV reverse transcriptase in complex with inhibitor 14 (PDB ID: 3C6T (13)), and DHFR in complex with methotrexate [PDB ID: 3DFR (14)] were used for the virtual screening studies. All crystallographic water molecules as well as the ligand molecules themselves were removed from the PDB files. Hydrogen atoms were added using PDB2PQR (15). In the case of DHFR, the hydrogen atoms associated with the NDP cofactor were derived from those present in the DUD database (10). All PQR files were then converted to the PDBQT format using MGLTools (16).

### Ligand preparation

The BindingDB (17) was used to identify PPAR, RT, and DHFR ligands. SMILES strings of the thirty unique PPAR, RT, and DHFR ligands with the lowest  $IC_{50}$  values, respectively, were obtained from PubChem (18). The LigPrep module of Schrodinger's Maestro computer program was used to build the molecular models in three dimensions, to add missing hydrogen atoms, and to generate all possible protonation states in a pH range of 5.0–9.0. For PPAR, LigPrep generated 30 unique molecular models from the top 15 known binders and 64 models from the top 30 binders. For RT, LigPrep generated 37 unique molecular models from the top ten known binders and 105 models from the top 30 binders. For DHFR, LigPrep generated 66 unique molecular models from the top 30 binders.

### LigMerge compound merging

The 30 models derived from the top fifteen known PPAR inhibitors, the 37 models derived from the top ten known RT inhibitors, and the 66 models derived from the top thirty DHFR inhibitors were processed using LigMerge with the `-all_symmetry_relations` and `-all_substituent_combinations` flags set to true. The `-ligands_dir` flag was used to automatically run LigMerge on all possible pairs of ligands in the specified directory. Following a second LigPrep run undertaken to minimize the structures, 896, 3959, and 3974 unique potential inhibitors were identified for PPAR, RT, and DHFR, respectively.

### Docking protocol

For PPAR, AutoDock Vina (19) was used to dock both the 896 LigMerge-generated models and the 64 models of known inhibitors into the 1FM9 binding site using a box size of  $40.9 \text{ \AA} \times 44.3 \text{ \AA} \times 46.8 \text{ \AA}$ . For RT, the 3959 LigMerge-generated models and the 105 models of known inhibitors were docked into the 3C6T binding site using a box size of  $18.0 \text{ \AA} \times 18.0 \text{ \AA} \times 18.0 \text{ \AA}$ . For DHFR, the 3974 LigMerge-generated compounds as well as the 66 models of known inhibitors were likewise docked into a crystallographic binding pocket (3DFR), using a box size of  $42.9 \text{ \AA} \times 44.8 \text{ \AA} \times 44.0 \text{ \AA}$ .

### Custom decoy libraries

For each of the LigMerge-generated ligand sets corresponding to the three receptors, the molecular weight (MW), logP, and polar surface area (PSA) were calculated using obprop (20). The set of 896 LigMerge compounds generated from known PPAR inhibitors had an average molecular weight of  $611 \pm 166 \text{ Da}$ , an average logP of  $7.1 \pm 0.7$ , and an average PSA of  $106 \pm 38 \text{ \AA}^2$ . The set of 3959 LigMerge compounds generated from known RT inhibitors had an average molecular weight of  $493 \pm 139 \text{ Da}$ , an average logP of  $4.6 \pm 1.2$ , and an average PSA of  $107 \pm 40 \text{ \AA}^2$ . Finally, the set of 3974 LigMerge compounds generated from known DHFR inhibitors had an average molecular weight of  $478 \pm 140 \text{ Da}$ , an average logP of  $3.6 \pm 1.3$ , and an average PSA of  $178 \pm 73 \text{ \AA}^2$ .

For each of the LigMerge-generated compound sets, an in-house script was used to generate a decoy set equal in size and chemical properties. MW, logP, and PSA statistics were calculated for each

of the approximately 11 000 000 compounds of the ZINC 'All Clean' data set (21) using obprop (20). Subsets of the 'All Clean' database were then identified with chemical properties similar to those of each of the LigMerge-generated sets in terms of both averages and standard deviations. In this way, decoy libraries were generated for PPAR (896 compounds), RT (3959 compounds), and DHFR (3974 compounds) that had average MW, logP, and PSA values within 1% of the values derived for the corresponding reference LigMerge data sets of same size. PPAR was the only exception; the average molecular weight of the PPAR decoy library was within 13% (MW = 532 Da) of the reference LigMerge-generated set because there was an insufficient number of high-molecular-weight compounds in the ZINC 'All Clean' data set. Additionally, it was ensured that standard deviations for these quantities did not exceed values in the corresponding LigMerge-generated sets. The decoy libraries were docked into their respective receptors with the same parameters used to dock the LigMerge-generated compound sets.

## Results and Discussion

LigMerge is an open-source, easy-to-use tool for generating novel compounds with structural features similar to those of known ligands. Compounds derived from known ligands are more likely to be true binders themselves. Once generated, LigMerge-derived compounds can be docked into receptor structures to identify likely inhibitors for subsequent synthesis and experimental validation.

To demonstrate the utility of the LigMerge algorithm, three protein drug targets with many known inhibitors were chosen as test systems: peroxisome proliferator-activated receptor (PPAR) gamma, HIV reverse transcriptase (RT), and dihydrofolate reductase (DHFR). For each of these systems, novel compounds were generated using LigMerge by combining features of known inhibitors. Predicted binding affinities were then assessed by computer docking. To demonstrate that LigMerge can generate a set of compounds enriched for high-affinity binders above and beyond screens of chemically similar molecules chosen at random, we also dock appropriate decoy databases into each of the three receptors studied to facilitate comparison.

### Peroxisome proliferator-activated Receptor

A total of 896 LigMerge-generated compounds were derived from the top fifteen experimentally verified PPAR binders listed in the Binding-DB database (17) as of October 2011. These compounds, together with the top thirty experimentally known inhibitors, were docked into an PPAR crystal structure using AutoDock Vina (19). The best-scoring LigMerge molecule (compound **1**, Figure 2A) and known inhibitor (Figure 3A) had estimated binding affinities of  $-13.2$  and  $-11.1$  kcal/mol, respectively. In fact, there were 109 LigMerge-generated models that scored better than the best-known inhibitor.

The best-scoring binding pose of the top-ranked LigMerge molecule (compound **1**) is shown in Figure 3A, together with the crystallographic pose of a known inhibitor (in yellow, taken from PDB ID: 2HFP (22)). The similarities between the poses of these two compounds are noteworthy. The predicted pose of compound **1** positions benzene and anisole substructures coincident with those of the known inhibitor. Additionally, the trifluoromethyl group of com-

ound **1** is predicted to be proximal to a sulfonamide group of the known compound, and the trifluoromethyl group of the known inhibitor is positioned proximal to the predicted location of a compound-**1** carboxylate group. Others have suggested that fluorinated methyl groups might be bioisosteres of the carboxylate group (compare PDB structures 3AEB and 3AE6) and the sulfonamide group (compare PDB structures 2XBV and 2XBX (23)). The similarities of these binding modes are not likely the result of mere chance; they give credence to the hypothesis that the LigMerge-generated compounds have improved docking scores specifically because they are based on known inhibitors and therefore build on pharmacophores known to be relevant to the receptor of interest.

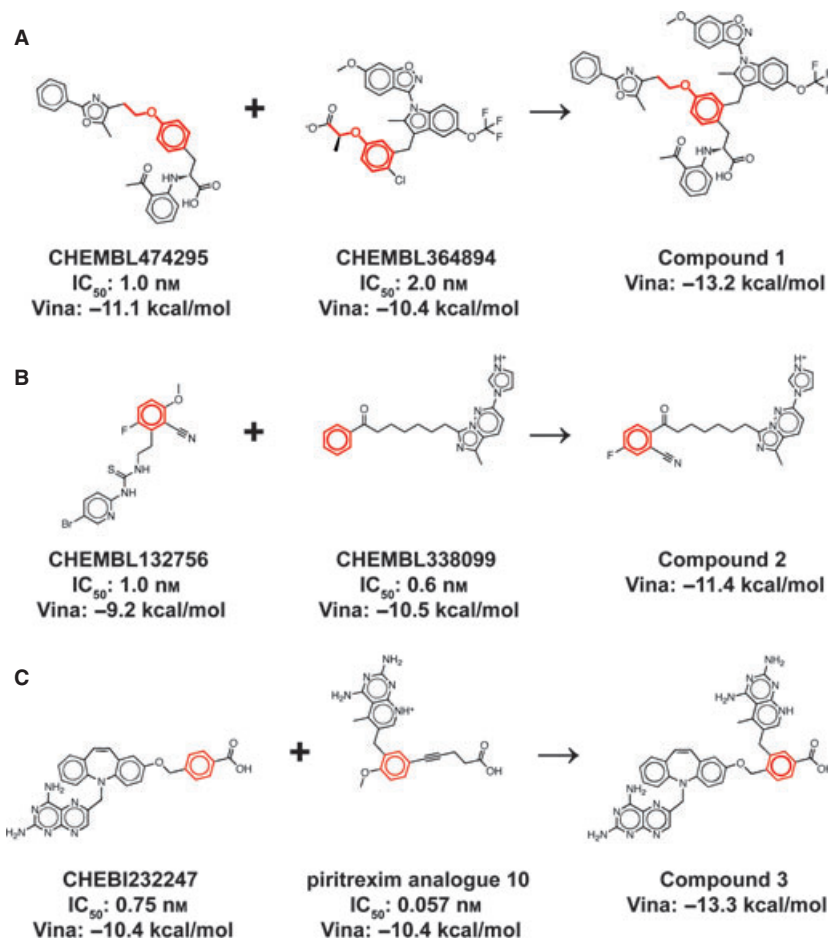
Figure 4A shows a normalized histogram of the Vina scores associated with the LigMerge-generated and top 30/known inhibitor compounds. The docking score distribution of the LigMerge-generated compounds is markedly broader than that of the known inhibitors. As expected, LigMerge generated a number of compounds that scored worse than the known inhibitors; as compound models are generated through an exhaustive combinatorial process, it is unsurprising that some LigMerge compounds had reduced predicted binding affinities. However, the docking score distribution of the LigMerge-generated compounds also extends further toward high affinities than that of the known inhibitors. About 5% of the known inhibitors had docking scores better than  $-11.0$  kcal/mol, suggesting tight binding. In contrast, more than 15% of the docked LigMerge compounds scored in that range, suggesting a genuine enrichment for strong binders.

### Reverse transcriptase

A total of 3959 LigMerge-generated compounds were derived from the top ten experimentally verified RT binders listed in the Binding-DB database (17) as of October 2011. These compounds, together with the top thirty experimentally known inhibitors, were docked into an RT crystal structure using AutoDock Vina (19). The best-scoring LigMerge molecule (compound **2**, Figure 2B) and known inhibitor (Figure 3B) had estimated binding affinities of  $-11.4$  and  $-10.5$  kcal/mol, respectively. In fact, there were 132 LigMerge-generated models that scored better than the best-known inhibitor.

The best-scoring binding pose of the top-ranked LigMerge molecule (compound **2**) is shown in Figure 3B, together with the crystallographic pose of a known inhibitor (in yellow, taken from PDB ID: 3C6T (13)). The similarities between the poses of these two compounds are noteworthy. Aside from the fact that they are predicted to occupy the same general space in the binding pocket, the 3-fluorobenzonitrile moiety of compound **2** docked at the same location as the analogous 3-chlorobenzonitrile moiety of the co-crystallized inhibitor. Additionally, the aromatic imidazo[1,5-b]pyridazine moiety of compound **2** docks at the same location, and in the same plane, as a crystallographic benzene moiety of the known inhibitor. Again, the similarities of these binding modes are not likely the result of mere chance; LigMerge-generated compounds likely have improved docking scores because they are based on known inhibitors rather than chosen at random.

Figure 4B shows a normalized histogram of the Vina scores associated with the models of both the top 30 known inhibitors and the



**Figure 2:** The top LigMerge-generated compounds and the known inhibitors from which they are derived. The maximum common substructures are highlighted in red. (A) Peroxisome proliferator-activated receptor gamma. (B) Reverse transcriptase. (C) Dihydrofolate reductase.

LigMerge-generated compounds. While some LigMerge-generated compounds again performed a good deal worse than known inhibitors, as expected, LigMerge did generate a number of compounds with higher-scoring predicted affinities; over 4% of the LigMerge compounds scored better than -10.0 kcal/mol, compared to fewer than 2% of the known inhibitors.

### Dihydrofolate reductase

A total of 3974 LigMerge-generated compounds were derived from the top thirty experimentally verified DHFR binders listed in the BindingDB database (17) as of October 2011. These compounds, together with the thirty experimentally known inhibitors themselves, were docked into a DHFR crystal structure. The best-scoring LigMerge molecule and known inhibitor (Figure 3C) had estimated binding affinities of -13.3 and -10.4 kcal/mol, respectively. LigMerge generated 608 models that scored better than the best-known inhibitor.

An analysis of the predicted binding pose of the top-scoring LigMerge-generated compound, compound **3** (Figure 2C), suggests that, as before, the enhanced predicted affinity over known inhibitors has not arisen by chance alone. The top predicted ligand was derived from the two best-scoring known inhibitors (Figure 2C):

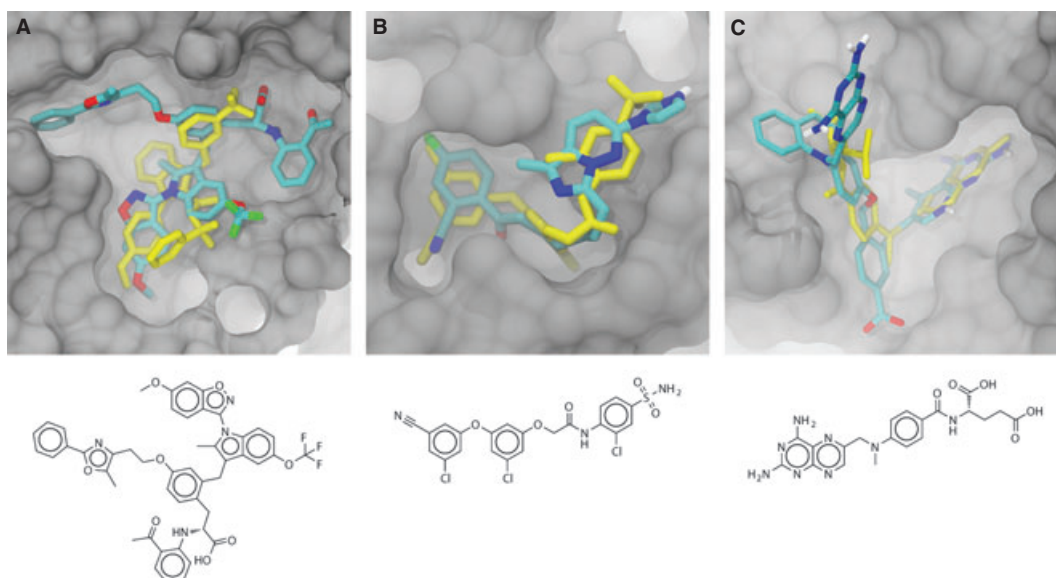
CHEBI232247 (24) and pirtrexim analog 10 (25), with IC<sub>50</sub> values of 0.75 and 0.057 nM, respectively. Additionally, the top Vina pose of compound **3** places the moiety analogous to the pteridine-2,4-diamine of methotrexate, a known inhibitor (in yellow, taken from PDB ID: 3DFR (14)), deep within the same folate-binding pocket. The novel ligand binds in ways that are similar to known inhibitors, as expected given that pharmacophoric information from known ligands has essentially been leveraged in the design of these novel compounds.

Histograms showing the Vina score distributions of the 30 known DHFR inhibitors, as well as the LigMerge-derived compound models, are shown in Figure 4C. The distribution of the Vina scores associated with the LigMerge-generated compounds was again generally wider than that of the known inhibitors. As before, some of the LigMerge compounds were certainly incompatible with potent binding, but 15.5% of the compound models were predicted to bind more potently than any known inhibitor.

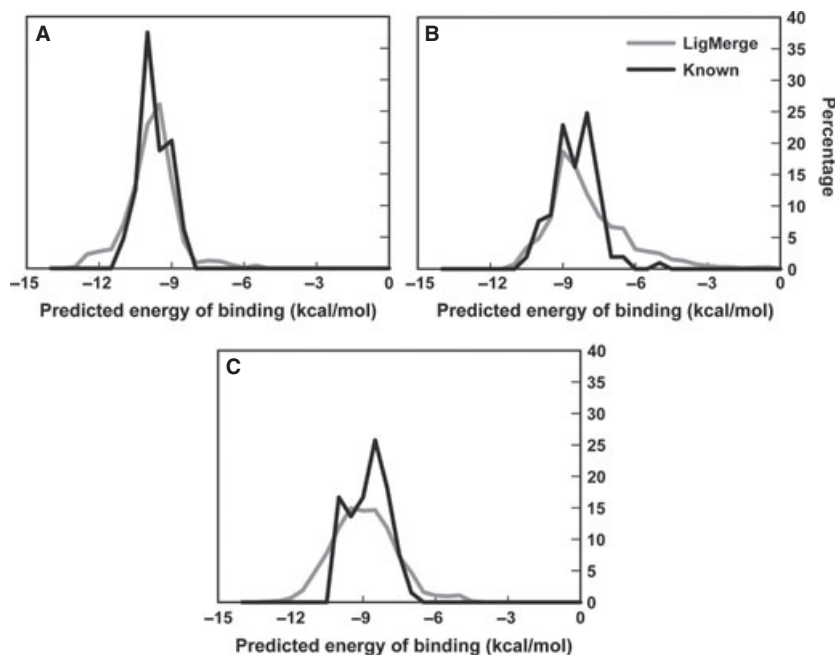
### Decoy library docking

We propose that the LigMerge-generated compound set included compounds with improved predicted affinities over those of known inhibitors because LigMerge generates novel compounds from





**Figure 3:** The predicted binding poses of the top LigMerge-generated compounds docked into their respective receptors. In all panels, some portions of the protein have been removed to facilitate visualization. Docked LigMerge-generated compounds are colored by element, and known co-crystallized compounds are colored yellow. The following are the standard representations of the co-crystallized ligands: (A) compound **1**, docked into PPAR. The crystallographic ligand (in yellow) is compound 2a, a known binder. (B) Compound **2**, docked into HIV reverse transcriptase. The crystallographic ligand (in yellow) is inhibitor 14, a known binder. (C) Compound **3**, docked into dihydrofolate reductase. The crystallographic ligand (in yellow) is methotrexate, a known binder.

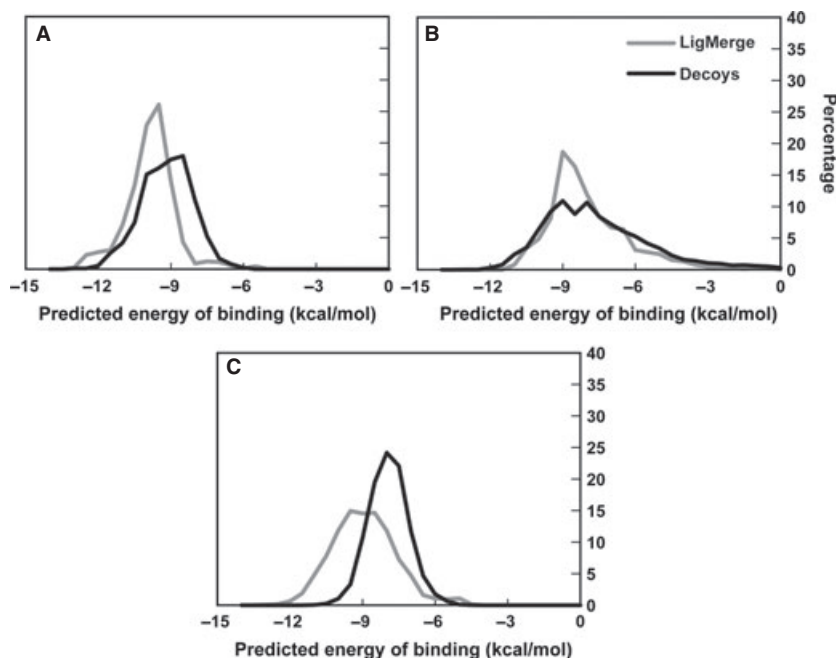


**Figure 4:** Vina score Histograms: LigMerge-generated Compounds versus Known Inhibitors. The LigMerge-generated compounds are shown in gray, and the known inhibitors are shown in black. (A) The histograms for PPAR gamma. (B) The histograms for HIV RT. (C) The histograms for DHFR.

known inhibitors in an intelligent and systematic way. However, it could be that these improved compounds were identified simply because the set of LigMerge-generated compounds was much larger than the set of known inhibitors, making it statistically more likely that a high-affinity predicted ligand would be found. To rule out this possibility, we compared the docking performance of the

three LigMerge-generated compound sets to that of decoy libraries similar in size and average chemical properties.

Figure 5 shows normalized histograms of the Vina score distributions for LigMerge and decoy docking into PPAR (Figure 5A), RT (Figure 5B), and DHFR (Figure 5C). For PPAR and DHFR, the LigMerge distributions



**Figure 5:** Vina score Histograms: LigMerge-generated Compounds versus Random Decoys. The LigMerge-generated compounds are shown in gray, and the decoy compounds are shown in black. (A) The histograms for PPAR gamma. (B) The histograms for HIV RT. (C) The histograms for DHFR.

are clearly shifted toward higher binding affinities, suggesting legitimate enhancement beyond what would be expected by docking compounds chosen at random. In contrast, the LigMerge score distribution associated with RT is similar to that of the decoy library. This may well be a consequence of Vina's inability to discriminate between native-like and non-native-like ligands for the HIV RT test system. These results demonstrate that for two of three systems, LigMerge provided a useful enrichment for high-affinity predicted binders.

## Conclusion

We here present an algorithm called LigMerge that considers two three-dimensional models of known or suspected small-molecule inhibitors and forms derivative models with similar chemical features. In the process of merging models, LigMerge first identifies the maximum common substructure (MCS). The MCS, which can be saved for later examination, may itself be a valuable tool for ligand evaluation. Next, the program aligns the two molecules by their mutual MCS, so that they are partially superimposed. Finally, the chemical moieties attached to each superimposed atom of the maximum common substructure are recombined, producing composite molecules similar to known or suspected inhibitors, but with potentially higher affinities. LigMerge is freely available through the National Biomedical Computation Resource (NBCR) and can be downloaded at <http://www.nbcrc.net/ligmerge/>.

## References

- Kitchen D.B., Decornez H., Furr J.R., Bajorath J. (2004) Docking and scoring in virtual screening for drug discovery: methods and applications. *Nat Rev Drug Discov*;3:935–949.
- Cross J.B., Thompson D.C., Rai B.K., Baber J.C., Fan K.Y., Hu Y.B., Humblet C. (2009) Comparison of several molecular docking programs: pose prediction and virtual screening accuracy. *J Chem Inf Model*;49:1455–1474.
- Kruger D.M., Evers A. (2010) Comparison of structure- and ligand-based virtual screening protocols considering hit list complementarity and enrichment factors. *ChemMedChem*;5:148–158.
- Amaro R.E., Schnauffer A., Interthal H., Hol W., Stuart K.D., McCammon J.A. (2008) Discovery of drug-like inhibitors of an essential RNA-editing ligase in *Trypanosoma brucei*. *Proc Natl Acad Sci*;105:17278–17283.
- Durrant J.D., Hall L., Swift R.V., Landon M., Schnauffer A., Amaro R.E. (2010) Novel naphthalene-based inhibitors of *Trypanosoma brucei* RNA editing ligase 1. *PLoS Negl Trop Dis*;4:e803.
- Durrant J.D., Urbaniak M.D., Ferguson M.A., McCammon J.A. (2010) Computer-aided identification of *Trypanosoma brucei* uridine diphosphate galactose 4'-Epimerase inhibitors: toward the development of novel therapies for African Sleeping Sickness. *J Med Chem*;53:5025–5032.
- Durrant J.D., Cao R., Gorfe A.A., Zhu W., Li J., Sankovsky A. et al. (2011) Non-bisphosphonate inhibitors of isoprenoid biosynthesis identified via computer-aided drug design. *Chem Biol Drug Des*;78:323–332.
- Wang Y., Hess T.N., Jones V., Zhou J.Z., McNeil M.R., McCammon J.A. (2011) Novel inhibitors of *Mycobacterium tuberculosis* dTDP-6-deoxy-L-lyxo-4-hexulose Reductase (RmlD) identified by virtual screening. *Bioorg Med Chem Lett*;21:7064–7067.
- Puerta D.T., Mongan J., Tran B.L., McCammon J.A., Cohen S.M. (2005) Potent, selective pyrone-based inhibitors of stromelysin-1. *J Am Chem Soc*;127:14148–14149.
- Huang N., Shoichet B.K., Irwin J.J. (2006) Benchmarking sets for molecular docking. *J Med Chem*;49:6789–6801.

11. Pierce A.C., Rao G., Bemis G.W. (2004) BREED: generating novel inhibitors through hybridization of known ligands. Application to CDK2, p38, and HIV protease. *J Med Chem*;47:2768–2775.
12. Gampe R.T., Montana V.G., Lambert M.H., Miller A.B., Bledsoe R.K., Milburn M.V. *et al.* (2000) Asymmetry in the PPAR gamma/RXR alpha crystal structure reveals the molecular basis of heterodimerization among nuclear receptors. *Mol Cell*;5:545–555.
13. Tucker T.J., Saggar S., Sisko J.T., Tynebor R.M., Williams T.M., Felock P.J. *et al.* (2008) The design and synthesis of diaryl ether second generation HIV-1 non-nucleoside reverse transcriptase inhibitors (NNRTIs) with enhanced potency versus key clinical mutations. *Bioorg Med Chem Lett*;18:2959–2966.
14. Bolin J.T., Filman D.J., Matthews D.A., Hamlin R.C., Kraut J. (1982) Crystal-structures of *Escherichia-Coli* and *Lactobacillus-Casei* dihydrofolate-reductase refined at 1.7 Å Resolution. 1. General features and binding of methotrexate. *J Biol Chem*;257:3650–3662.
15. Dolinsky T.J., Czodrowski P., Li H., Nielsen J.E., Jensen J.H., Klebe G. *et al.* (2007) PDB2PQR: expanding and upgrading automated preparation of biomolecular structures for molecular simulations. *Nucleic Acids Res*;35:W522–W525.
16. Sanner M.F. (1999) Python: a programming language for software integration and development. *J Mol Graph Model*;17:57–61.
17. Liu T., Lin Y., Wen X., Jorissen R.N., Gilson M.K. (2007) Binding-DB: a web-accessible database of experimentally determined protein-ligand binding affinities. *Nucleic Acids Res*;35:D198–D201.
18. Wang Y., Xiao J., Suzek T.O., Zhang J., Wang J., Bryant S.H. (2009) PubChem: a public information system for analyzing bioactivities of small molecules. *Nucleic Acids Res*;37:W623–W633.
19. Trott O., Olson A.J. (2009) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*;31:455–461.
20. O'Boyle N.M., Banck M., James C.A., Morley C., Vandermeersch T., Hutchison G.R. (2011) Open Babel: an open chemical toolbox. *J Cheminform*;3:33.
21. Irwin J.J., Shoichet B.K. (2005) ZINC – a free database of commercially available compounds for virtual screening. *J Chem Inf Model*;45:177–182.
22. Hopkins C.R., O'Neil S.V., Laufersweiler M.C., Wang Y.L., Pokross M., Meke M. *et al.* (2006) Design and synthesis of novel N-sulfonyl-2-indole carboxamides as potent PPAR-gamma binding agents with potential application to the treatment of osteoporosis. *Bioorg Med Chem Lett*;16:5659–5663.
23. Anselm L., Banner D.W., Benz J., Zbinden K.G., Himber J., Hilpert H. *et al.* (2010) Discovery of a factor Xa inhibitor (3R,4R)-1-(2,2-difluoro-ethyl)-pyrrolidine-3,4-dicarboxylic acid 3-[(5-chloropyridin-2-yl)-amide] 4-[[2-fluoro-4-(2-oxo-2H-pyridin-1-yl)-phenyl]-amide} as a clinical candidate. *Bioorg Med Chem Lett*;20:5313–5319.
24. Rosowsky A., Fu H.N., Chan D.C.M., Queener S.F. (2004) Synthesis of 2,4-diamino-6-[2'-O-(omega-carboxyalkyl)oxydibenz[b,f]azepin-5-yl]-methylpteridines as potent and selective inhibitors of *Pneumocystis carinii*, *Toxoplasma gondii*, and *Mycobacterium avium* dihydrofolate reductase. *J Med Chem*;47:2475–2485.
25. Chan D.C.M., Fu H.N., Forsch R.A., Queener S.F., Rosowsky A. (2005) Design, synthesis, and antifolate activity of new analogues of piritrexim and other diaminopyrimidine dihydrofolate reductase inhibitors with omega-carboxyalkoxy or omega-carboxy-1-alkynyl substitution in the side chain. *J Med Chem*;48:4420–4431.