Supporting Information

FragRep: A Web Server for structure-based drug design by fragment replacement

Jinwen Shan\textsuperscript{1,2}, Xiaolin Pan\textsuperscript{1,2}, Xingyu Wang\textsuperscript{2}, Xudong Xiao\textsuperscript{1,2}, Changge Ji\textsuperscript{1,2,*}

\textsuperscript{1}Shanghai Engineering Research Center for Molecular Therapeutics and New Drug Development, School of Chemistry and Molecular Engineering, East China Normal University, Shanghai 200062 China

\textsuperscript{2}NYU-ECNU Center for Computational Chemistry at NYU Shanghai, Shanghai, 200062 China

*Email: Chicago.ji@gmail.com
Descriptors for the fragment.

To search suitable fragments efficiently from the database, the following three types of descriptors were calculated for fast database queries.

a. The Geometric relationship between linking bonds (also cutting bonds in the query ligand).
   When there are more than two linking points on the fragment, we use the strategy developed by Lauri and Bartlett (J. Comp. Aided Mol. Des., 8, 51) to describe the geometric relationship between any two linking points. Each linking point is represented by one bond vector. Firstly, we extract bond vectors of the two linking points. As shown in figure a, two linking points were defined by bond vectors \((b_1, t_1)\) and \((b_2, t_2)\) respectively. Secondly, we calculate the Euclidean distance of two points in space \(b_1, b_2\), the angle of three points in space \(t_1, b_1, b_2\), the angle of three points in space \(t_2, b_2, b_1\), and the dihedral angle of four points of space \(t_1, b_1, b_2,\) and \(t_2\) (as shown in figure S1b). The four attribute values serve as the geometric relationship between the two linking points.

![Figure S1](image1.png)

Figure S1. (a) Bond vectors act as linking points on a 3D fragment structure, (b) Distances, angles, and dihedral angles between two linking points.

b. Size of the fragment measured from each linking atom.
   We defined atoms \(b\) and \(t\) and numbered all the remaining heavy atoms in Figure S2, then we calculate the distance between the \(b\) atom and all the remaining atoms in Figure S2. After that, we select the maximum distance as the fingerprint representing the fragment size.

![Figure S2](image2.png)

Figure S2. The maximum distance between linking points and other atoms on the fragment.
c. Polar atom information.
Firstly, we identify polar atoms on the fragment, including hydrogen bond donor, hydrogen bond acceptor, positively charged centers, negatively charged centers. Secondly, we examine geometric relationships between polar atoms and the linking points. As shown in figure 3S, the polar atom is labeled as ‘FA’ and the linking point is represented by the bond vector (b, t). We calculate the distance between ‘FA’ and ‘b’ and the angle between vector (b, t) and the vector (b, FA) to represent relative position between the linking point and the polar atom.

Figure S3. a). Bond vector (b, t) represents the linking point and FA represents the polar atom. b) Distance and angle to represent relative position between the linking point and the polar atom.

Schema of the Fragment Database.
**Fragment_info**

Frag_ID: The id of the fragment

Frag_name: The name of the fragment

Frag_N_atoms: The number of atoms on the fragment

Frag_Wt: The molecular mass of the fragment

Frag_Smiles: The SMILES of the fragment

**Conformation_info**

Conformation_ID: The id of the conformation

Frag_ID: The id of the fragment

Conformation_Block: 3D coordinate information of the conformation

**Polar_atom_info**

Polar_info_ID: The id of the polar atom information

Conformation_ID: The id of the conformation

Polar_info_FP: The index of the polar info type.

Base_Atom_ID: The id of the base atom

Tip_Atom_ID: The id of the tip atom

Polar_Atom_ID: The id of the polar atom

**Polar detail**

Polar_info_ID: The id of the polar atom information

Polar_atom_type: Atom type of the polar atom

Dist: The distance between the base atom and the polar atom

Angle: The angle between the vector (b,t) and the vector (b,FA)

**Size_info**

Size_info_ID: The id of size info

Conformation_ID: The id of the conformation

Base_Atom_ID: The id of the base atom

Tip_Atom_ID: The id of the tip atom

Size: Size of the fragment measured from the base atom

Size_FP: The index of size.

**Geometry_Info**
Geometry_ID: The id of the geometry
Conformation_ID: The id of the conformation
Geometry_info_FP: The index the geometry type.
Base_Atom_ID_A: The id of the A base atom
Tip_Atom_ID_A: The id of the A tip atom
Base_Atom_ID_B: The id of the B base atom
Tip_Atom_ID_B: The id of the B tip atom

Geometry_detail
Geometry_ID: The id of the geometry
Dist: The distance between two base atoms.
Angle1: The angle of three points in space t1,b1,b2.
Angle2: The angle of three points in space t2,b2,b1
Dihedral: the dihedral angle of four points of space t1, b1, b2, and t2

Protein-ligand binding scoring function.

We trained an empirical scoring function to predict the binding affinity of protein-ligand interaction. The scoring function is composed of five individual energy items, including van der Waals interaction, hydrogen bond interaction, hydrophobic interaction, metal-ligand interaction, and the number of rotatable bonds of the ligand. Protein residues within 5.5Å of the ligand were used as binding pocket to identify various interactions.

Details of each interaction type and model training and testing are described as follows.

\[ \Delta G_{\text{binding}} = w_1 \sum_a \sum_b VDW_{ab} + w_2 \sum_a \sum_b H B_{ab} + w_3 \sum_a \sum_b H C_{ab} + w_4 \sum_a \sum_b M_{ab} + w_5 \Delta G_{\text{rot}} + w_6 \]  

(1.)

Van der Waals interaction. Van der Waals interaction is an important energy item in predicting the binding affinity of protein-ligand interaction. Lenard-Jones potential is
used to describe Van der Waals interactions.

\[ VD_{Wab} = \left( \frac{d_{ab}}{d_0} \right)^8 - 2 \times \left( \frac{d_{ab}}{d_0} \right)^4 \]  

(2.)

where \( VD_{Wab} \) denotes the van der Waals interaction energy between protein pocket atom \( a \) and the ligand atom \( b \), \( d_{ab} \) is the distance between the atom \( a \) of protein binding pocket and the atom \( b \) of ligand; \( d_0 \) is the sum of atomic radii of atom \( a \) and \( b \) (all the radii parameters were derived from Wang R, Lai L, Wang S. 2002, J Comput Aided Mol Des, 16(1): 11-26). The Van der Waals interaction is calculated by summarizing all the atom pairs between the protein pocket atoms and the ligand atoms.

**Hydrogen bond interaction.** Two criteria are used to identify whether hydrogen bond exists between donor and acceptor: a) whether the distance between the hydrogen bond donor and the hydrogen bond acceptor is less than 3.5 Å; b) whether the angle between the donor, the hydrogen and the acceptor is greater than 120°. The following equation is used to calculate energetic contribution from a hydrogen bond,

\[ HB_{ab} = \frac{1}{\left( 1 + \left( \frac{d_{ab}}{2.6} \right)^{12} \right)^{0.58}} \]  

(3.)

where \( HB_{ab} \) denotes hydrogen bond energy between binding pocket atom \( a \) and the ligand atom \( b \); \( d_{ab} \) is the distance between binding pocket atom \( a \) and ligand atom \( b \), where \( a \) and \( b \) are the heavy atoms of hydrogen bond donor or acceptor.

**Hydrophobic contact.** We adopted the method from ChemScore to calculate hydrophobic energy, as follows:

\[ HC_{ab} = \begin{cases} 
1.0 & d_{ab} \leq d_0 + 0.5 \\
\left( \frac{1}{1.5 \times (d_0 + 2.0 - d_{ab})} \right) & d_0 + 0.5 < d_{ab} \leq d_0 + 2.0 \\
0.0 & d_{ab} > d_0 + 2.0 
\end{cases} \]  

(4.)

where \( HC_{ab} \) represents the hydrophobic interaction energy between binding pocket
atom \( a \) and the ligand atom \( b \), which is calculated by considering all the hydrophobic contacts between the ligand and the binding pocket; \( d_0 \) is the sum of the atomic radii of atom \( a \) and \( b \). \( d_{ab} \) is the distance between the binding pocket atom \( a \) and ligand atom \( b \).

**Metal-ligand interaction.** The metal-ligand interaction is calculated by the metal-ligand atom pairs formed between the ligand and metal atom in the binding pocket. It’s calculated by the following equation:

\[
M_{ab} = \begin{cases} 
1.0 & d_{ab} < 2.0 \\
3.0 - d_{ab} & 2.0 < d_{ab} < 3.0 \\
0.0 & d_{ab} \geq 3.0 
\end{cases} \tag{5.}
\]

where \( M_{ab} \) represents the metal-ligand interaction between metal atom \( a \) and the ligand atom \( b \), \( d_{ab} \) is the distance between metal atom \( a \) and the ligand atom \( b \).

**Conformational Entropy.** We use the number of rotatable bonds on the ligand to describe the contribution of conformational entropy to the binding affinity.

**Atomic Parameters.** Atomic radii for both protein and ligand atoms are set as follows: carbon, 1.9 Å; oxygen, 1.7 Å; nitrogen, 1.8 Å; phosphorus, 2.1 Å; sulfur, 2.0 Å; fluorine, 1.5 Å; chlorine, 1.8 Å; bromine, 2.0 Å; iodine, 2.2 Å; metals, 1.2 Å; hydrogen, 0 Å. Radii of other atoms not included in the above are set at 1.8 Å.

**Model Training and testing.** We used a standard benchmark set (PDBbind2018 set) to train and test the performance of scoring function. The PDBbind2018 set contains three parts, refined set, general set and core set. The refined set of PDBbind2018 (contains 4197 complexes) was used to train the model. The core set of PDBbind2018 (contains 285 complexes) was used to test the performance of our scoring function. A simple linear regression model was used for model construction. There is no overlap
between the training and test data. We compared the performance our scoring function (rbScore) with two other widely used scoring functions, the autodock vina scoring function and the XScore. The performance is shown in Figure S4.

Figure S4. Performances of the empirical score function on the test set (PDBBind-v2018 Core Set).

**Tutorial for job submission and running on the FragRep server.**

1. Please prepare the input protein file and the ligand file beforehand. You can prepare the input files by splitting an X-ray cocrystal structure of a protein-ligand complex. Or you can prepare the input files from molecular docking. On the xundrug.cn/fragrep website, please choose a protein and a ligand file and then click the upload button.

Figure S5. Input page of the FragRep server.
2. The server analyzes the binding pocket of the protein and all replaceable substructures on the input ligand. Then the server presents all replaceable substructures on the input ligand. Please select the substructure you want to optimize and click the 'Submit' button and wait for two minutes.

![Figure S6. Replaceable substructures on the input ligand.](image)

3. The system finds suitable fragments that fit well with the local protein environment and assemble new molecules. The result page shows the information of newly generated analogues. The user can download the new ligand by clicking “sdf”, explore 3D protein-ligand interactions by clicking ‘3DView’ or explore 3D view of the new ligand overlapped with the original ligand by clicking ‘OverlayView’. The user can also download all the generated ligand as a .xlsx file.

![Figure S7. The result page of FragRep.](image)
**Table S1.** Known active inhibitors of p38 MAP kinase rediscovered by a FragRep search.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Reference DOI</th>
<th>Test Assay</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Structure1" /></td>
<td>10.1021/jm0610970</td>
<td>Inhibition of p38-alpha MAPK by non-radioactive immunosorbent assay</td>
<td>IC50 = 1800.0 nM</td>
</tr>
<tr>
<td><img src="image2.png" alt="Structure2" /></td>
<td>10.1016/j.bmc.2014.05.045</td>
<td>Inhibition of full length human p38alpha MAP kinase by Off-chip mobility shift assay</td>
<td>IC50 = 1310.0 nM</td>
</tr>
<tr>
<td><img src="image3.png" alt="Structure3" /></td>
<td>10.1021/jm0301787</td>
<td>In vitro inhibitory concentration against Mitogen-activated protein kinase p38 alpha</td>
<td>IC50 = 3300.0 nM</td>
</tr>
<tr>
<td><img src="image4.png" alt="Structure4" /></td>
<td>10.1021/jm0301787</td>
<td>In vitro inhibitory concentration against Mitogen-activated protein kinase p38 alpha</td>
<td>IC50 = 24.0 nM</td>
</tr>
<tr>
<td><img src="image5.png" alt="Structure5" /></td>
<td>10.1021/jm901297e</td>
<td>Displacement of N,N'-((2,2'-(3,3'-disulfanediylbis(2,5-dioxopyrrolidine-3,1-diyl))bis(ethane-2,1-diyl))bis(2-(3-(3-tert-butyl-5-(3-naphthalen-1-ylureido)-1H-pyrazol-1-yl)phenylamino)acetamido) from inactive form of p38alpha expressed in</td>
<td>IC50 = 820.0 nM</td>
</tr>
<tr>
<td>IC50 Value</td>
<td>Reaction Type</td>
<td>Substrate</td>
<td>Inhibition Value</td>
</tr>
<tr>
<td>------------</td>
<td>---------------</td>
<td>-----------</td>
<td>-----------------</td>
</tr>
<tr>
<td>86.0 nM</td>
<td>In vitro inhibitory concentration against Mitogen-activated protein kinase p38 alpha</td>
<td>Escherichia coli BL21(DE3) cells by enzyme fragment complementation assay</td>
<td>IC50 = 86.0 nM</td>
</tr>
<tr>
<td>600.0 nM</td>
<td>Inhibition of p38a MAPK (unknown origin) by radiometric assay</td>
<td></td>
<td>IC50 = 600.0 nM</td>
</tr>
<tr>
<td>50.0 %</td>
<td>Inhibition of p38alpha assessed as residual activity at 10 uM by Cherenkov counting relative to control Activity</td>
<td></td>
<td>= 50.0 %</td>
</tr>
<tr>
<td>20.0 %</td>
<td>Inhibition of p38alpha assessed as residual activity at 10 uM by Cherenkov counting relative to control Activity</td>
<td></td>
<td>= 20.0 %</td>
</tr>
<tr>
<td>230.0 nM</td>
<td>Inhibition of p38-alpha MAPK</td>
<td></td>
<td>IC50 = 230.0 nM</td>
</tr>
<tr>
<td>26.50 %</td>
<td>Inhibition of p38-alpha MAPK at 10 uM</td>
<td></td>
<td>= 26.50 %</td>
</tr>
<tr>
<td>Structure</td>
<td>Reference</td>
<td>Description</td>
<td>IC50</td>
</tr>
<tr>
<td>-----------</td>
<td>-----------</td>
<td>-------------</td>
<td>------</td>
</tr>
<tr>
<td><img src="image1.png" alt="Structure 1" /></td>
<td>10.1016/j.bmcl.2008.09.040</td>
<td>Inhibition of GST-fused p38alpha expressed in Escherichia coli</td>
<td>IC50 = 120.0 nM</td>
</tr>
<tr>
<td><img src="image2.png" alt="Structure 2" /></td>
<td>10.1016/j.bmcl.2008.09.040</td>
<td>Inhibition of GST-fused p38alpha expressed in Escherichia coli</td>
<td>IC50 = 160.0 nM</td>
</tr>
<tr>
<td><img src="image3.png" alt="Structure 3" /></td>
<td>10.1016/j.bmcl.2008.09.040</td>
<td>Inhibition of GST-fused p38alpha expressed in Escherichia coli</td>
<td>IC50 = 150.0 nM</td>
</tr>
<tr>
<td><img src="image4.png" alt="Structure 4" /></td>
<td>10.1016/j.bmcl.2016.10.001</td>
<td>Inhibition of human recombinant full length GST-tagged MAPK14 phosphorylation expressed in Escherichia coli in presence of 500 uM of ATP by direct Z-LITE assay</td>
<td>IC50 = 1350.0 nM</td>
</tr>
<tr>
<td><img src="image5.png" alt="Structure 5" /></td>
<td>10.1016/j.bmcl.2008.09.040</td>
<td>Inhibition of GST-fused p38alpha expressed in Escherichia coli</td>
<td>IC50 = 150.0 nM</td>
</tr>
<tr>
<td><img src="image6.png" alt="Structure 6" /></td>
<td>10.1021/jm801467h</td>
<td>Inhibition of p38alpha MAP kinase</td>
<td>IC50 = 79.0 nM</td>
</tr>
<tr>
<td><img src="image7.png" alt="Structure 7" /></td>
<td>10.1021/jm801467h</td>
<td>Inhibition of p38alpha MAP kinase</td>
<td>IC50 = 740.0 nM</td>
</tr>
</tbody>
</table>
Inhibition of p38α-mediated ATF2 phosphorylation in presence of ATP

IC50 = 3190.0 nM

Inhibition of p38α-mediated ATF2 phosphorylation in presence of ATP

IC50 = 3150.0 nM

Inhibition of p38α-mediated ATF2 phosphorylation in presence of ATP

IC50 = 3700.0 nM

Table S2. Known active inhibitors of BCL-XL rediscovered by a FragRep search.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Reference DOI</th>
<th>Test Assay</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Structure" /></td>
<td>10.1021/jm400556w</td>
<td>Binding affinity to human GST-tagged Bcl-Xl assessed as dissociation rate by surface plasmon resonance assay</td>
<td>Kd = 79.0 nM</td>
</tr>
<tr>
<td><img src="image2.png" alt="Structure" /></td>
<td>10.1021/jm400556w</td>
<td>Binding affinity to human GST-tagged Bcl-Xl by luminescence proximity assay</td>
<td>IC50 = 14.0 nM</td>
</tr>
<tr>
<td><img src="image3.png" alt="Structure" /></td>
<td>10.1021/ml500030p</td>
<td>Displacement of 26-mer BIMBH3 peptide from GST-tagged Bcl-Xl (unknown origin) by AlphaScreen assay</td>
<td>IC50 = 720.0 nM</td>
</tr>
</tbody>
</table>
Table S3. Known active inhibitors of Glycogen phosphorylase rediscovered by a FragRep search.

<table>
<thead>
<tr>
<th>Structure</th>
<th>Reference</th>
<th>Test Assay</th>
<th>Activity</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image1.png" alt="Structure 1" /></td>
<td>10.1016/j.bmc.2009.12.043</td>
<td>Inhibition of rabbit skeletal muscle glycogen phosphorylase b by Lineweaver-Burke plot analysis</td>
<td>IC50 = 600000.0 nM</td>
</tr>
<tr>
<td><img src="image2.png" alt="Structure 2" /></td>
<td>10.1016/j.bmc.2014.06.058</td>
<td>Inhibition of rabbit skeletal muscle glycogen phosphorylase b assessed as inorganic phosphate release</td>
<td>Ki = 8500.0 nM</td>
</tr>
<tr>
<td><img src="image3.png" alt="Structure 3" /></td>
<td>10.1016/j.bmc.2014.06.058</td>
<td>Inhibition of rabbit skeletal muscle glycogen phosphorylase b assessed as inorganic phosphate release</td>
<td>Ki = 1800.0 nM</td>
</tr>
</tbody>
</table>