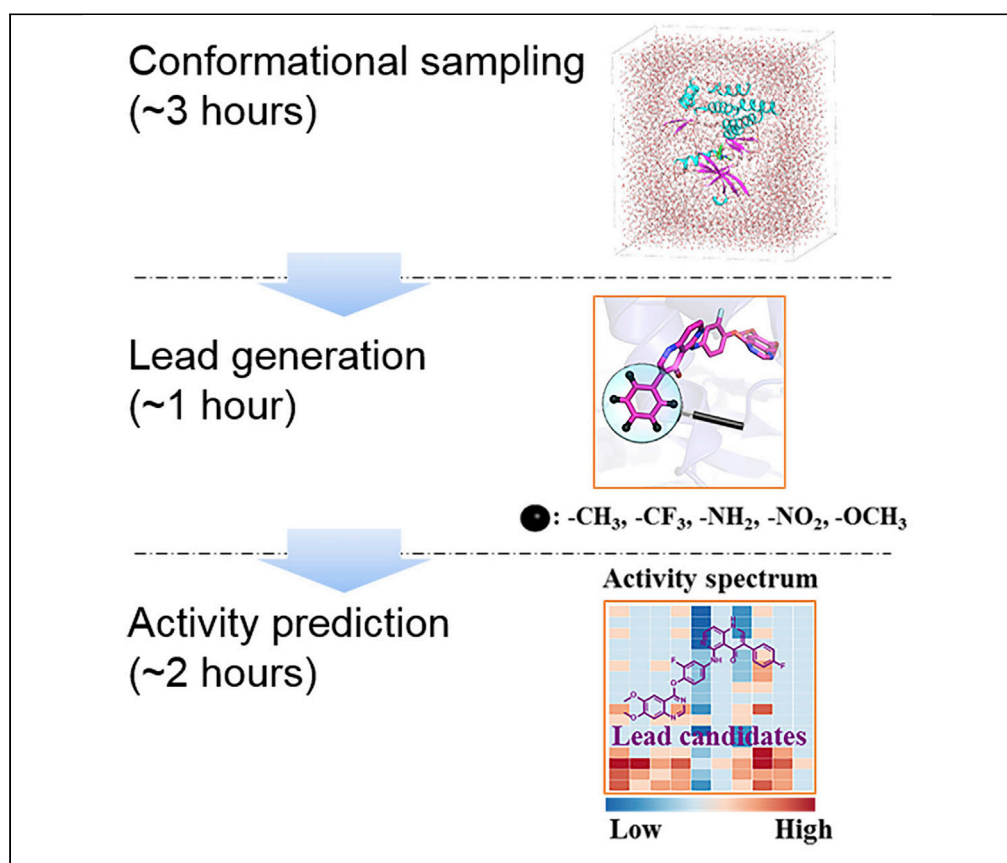


Protocol

Protocol for hit-to-lead optimization of compounds by auto *in silico* ligand directing evolution (AILDE) approach



Longcan Mei,
Fengxu Wu, Gefei
Hao, Guangfu Yang
gefei_hao@foxmail.com

HIGHLIGHTS

Detailed instructions are given for ligand modification by fragment growing

Conformational sampling is performed by molecular dynamics simulation

Binding free energy of a protein-ligand complex is computationally evaluated

Hit-to-lead (H2L) optimization is crucial for drug design, which has become an increasing concern in medicinal chemistry. A virtual screening strategy of auto *in silico* ligand directing evolution (AILDE) has been developed to yield promising lead compounds rapidly and efficiently. The protocol includes instructions for fragment compound library construction, conformational sampling by molecular dynamics simulation, ligand modification by fragment growing, as well as the binding free energy prediction.

Mei et al., STAR Protocols 2,
100312
March 19, 2021 © 2021 The
Author(s).
[https://doi.org/10.1016/
j.xpro.2021.100312](https://doi.org/10.1016/j.xpro.2021.100312)



Protocol

Protocol for hit-to-lead optimization of compounds by auto *in silico* ligand directing evolution (AILDE) approachLongcan Mei,^{1,2,5} Fengxu Wu,^{1,2,5} Gefei Hao,^{1,2,3,6,7,*} and Guangfu Yang^{1,2,4}¹Key Laboratory of Pesticide & Chemical Biology, Ministry of Education, College of Chemistry, Central China Normal University, Wuhan 430079, China²International Joint Research Center for Intelligent Biosensor Technology and Health, Central China Normal University, Wuhan 430079, China³State Key Laboratory Breeding Base of Green Pesticide and Agricultural Bioengineering, Key Laboratory of Green Pesticide and Agricultural Bioengineering, Ministry of Education, Research and Development Center for Fine Chemicals, Guizhou University, Guiyang 550025, China⁴Collaborative Innovation Center of Chemical Science and Engineering, Tianjin 300072, China⁵These authors contributed equally⁶Technical contact⁷Lead contact*Correspondence: gefei_hao@foxmail.com
<https://doi.org/10.1016/j.xpro.2021.100312>

SUMMARY

Hit-to-lead (H2L) optimization is crucial for drug design, which has become an increasing concern in medicinal chemistry. A virtual screening strategy of auto *in silico* ligand directing evolution (AILDE) has been developed to yield promising lead compounds rapidly and efficiently. The protocol includes instructions for fragment compound library construction, conformational sampling by molecular dynamics simulation, ligand modification by fragment growing, as well as the binding free energy prediction.

For complete details on the use and execution of this protocol, please refer to Wu et al. (2020).

BEFORE YOU BEGIN

Auto *in silico* ligand directing evolution (AILDE) is an effective approach to rapidly explore the structure-activity relationship for H2L optimization (Wu et al., 2020). AILDE is utilized to optimize a hit compound by minor chemical modifications on its scaffold. Minimal increases or losses for the binding efficiency of ligand could be obtained by these modifications. Hence, this method can help us for the exploration of the chemical space around a hit compound and promote the optimization of the hit compound into more drug-like lead compounds. Figure 1 shows the workflow of AILDE. AILDE is based on molecular dynamics (MD) simulation complemented with one-step free energy perturbation (FEP) (Hao et al., 2015; Wu et al., 2018). MD simulation is firstly performed on the protein-hit complex. The equilibrated conformational ensemble is then collected. The hit ligand in each conformation is modified to new ligand with substitution using the fragment library. The binding free energy changes between protein-hit and hit analogs are calculated. New leads are finally determined based on the rank of the binding affinity changes.

We should provide a productive computational environment for the usage of AILDE. We need a high performance computer with a Unix-based operating system to perform this protocol. A graphics processing unit (GPU) equipment is advised to be used to accelerate the process of molecular dynamics simulation. The prerequisite software (in the materials and equipment section) is freely available on the corresponding websites. The detail introduction about the installation is provided in their user manuals. These software are utilized in different procedures. We download the source



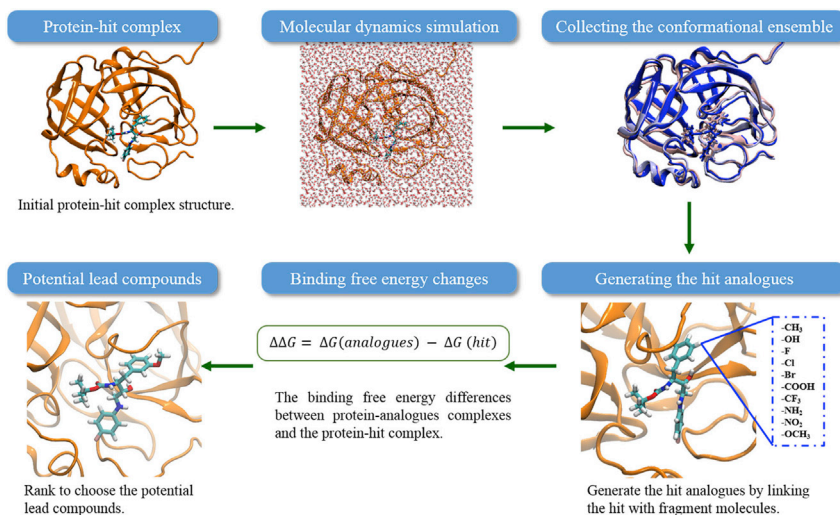


Figure 1. Workflow of the AILDE method

Once an initial protein-hit complex structure is provided, AILDE will conduct the hit-to-lead optimization according to the following steps: (1) molecular dynamics simulation: this allow to simulate the structural flexibility of the protein-hit complex; (2) collecting the conformational ensemble: adequately collect the representative conformations of a protein-hit complex; (3) generating the hit analogues: a fragment growing based method is utilized to convert a hit to analogue compounds. The conformations of protein-analogues complex are obtained from the representative conformations of the protein-hit complex; (4) binding free energy change: comparing the binding free energy of protein-hit and protein-analogues complexes. Identifying the potential lead compounds according to the rank of the binding free energy changes.

code of AILDE at <https://github.com/fwangccnu/AILDE>. After configuration of the computational environment, we prepare the input data according to the following steps.

Preparing the input structure file of the protein complex with a hit ligand

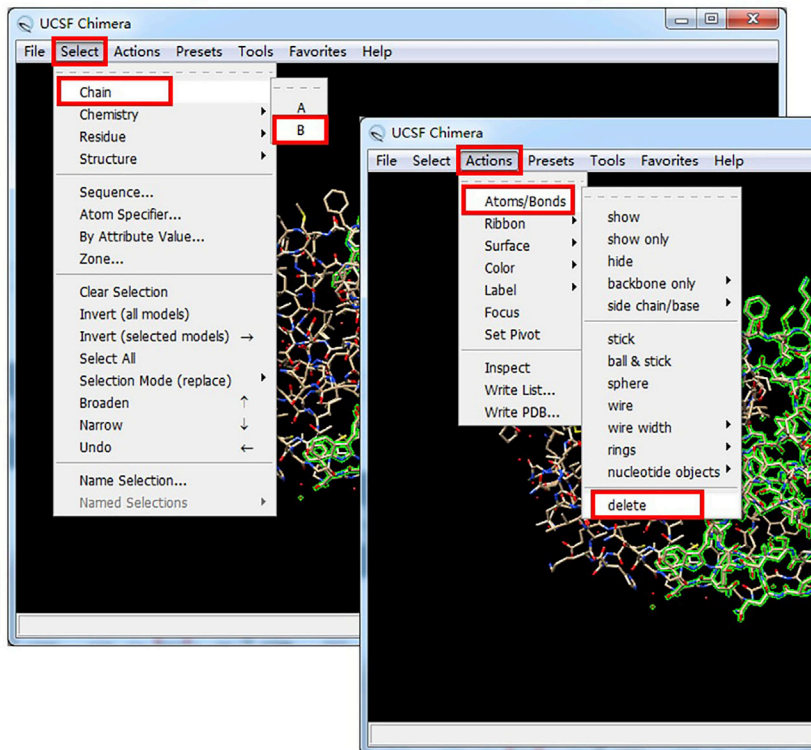
⌚ Timing: ~15 min

1. Download the structure file of our interest protein complex with a hit ligand from RCSB Protein Data Bank (PDB) (<https://www.rcsb.org/>). For example, search the PDB ID of 2P3B in PDB database, and obtain the result of the crystal structure of the subtype B wild type HIV protease complexes with TL-3 inhibitor. And save the PDB format structure file.
2. Open the structure file by Chimera software, keep the hit ligand and its direct interacting protein chain and delete all the irrelative molecules, and save as a processed PDB format structure file. [Figure 2A](#) shows the example of deleting the irrelative atoms by Chimera software. A detail document about the usage of Chimera is available on the website (<https://www.cgl.ucsf.edu/chimera/docindex.html>).
3. Text edit the processed structure file. Rename MSE residues as MET, and rename HIS as HID, HIE, or HIP based on the protonated states ([Figure 2B](#)). And add a "TER" record to indicate that the chains are not physically connected to each other. If there are alternative conformations of the residues, keep only the highest occupancy. These operations can be conducted by Chimera software ([Figure 2C](#)).

Note: Alternatively, we can obtain the complex structure of a protein with a hit ligand by molecular docking method ([Meng et al., 2011](#)).

△ CRITICAL: It requires the complex structure file to strictly abide by a standard PDB format, and a description about PDB format file is accessed at <http://www.wwpdb.org/documentation/file-format-content/format33/v3.3.html>.

A



B

ATOM	555	CA	GLY	A	68	35.725	-14.017	11.431	1.00
ATOM	556	C	GLY	A	68	36.063	-12.846	12.348	1.00
ATOM	557	O	GLY	A	68				
ATOM	558	N	HIS	A	69				
ATOM	559	CA	HIS	A	69				
ATOM	560	C	HIS	A	69				
ATOM	561	O	HIS	A	69				
ATOM	562	CB	HIS	A	69				
ATOM	563	CG	HIS	A	69				
ATOM	564	ND1	HIS	A	69				
ATOM	565	CD2	HIS	A	69				
ATOM	566	CE1	HIS	A	69				
ATOM	567	NE2	HIS	A	69				
ATOM	568	N	LYS	A	70				
ATOM	569	CA	LYS	A	70				

ATOM	279	OE1	GLU	A	35	10.134	-11.361	11.200	
ATOM	280	OE2	GLU	A	35	11.072	-11.405	13.224	
ATOM	281	N	MSE	A	36	14.461	-14.360	12.937	
ATOM	282	CA	MSE	A	36				ATOM 279 OE1 GLU A 35
ATOM	283	C	MSE	A	36				ATOM 280 OE2 GLU A 35
ATOM	284	O	MSE	A	36				ATOM 281 N MET A 36
ATOM	285	CB	MSE	A	36				ATOM 282 CA MET A 36
ATOM	286	CG	MSE	A	36				ATOM 283 C MET A 36
ATOM	287	SD	MSE	A	36				ATOM 284 O MET A 36
ATOM	288	CE	MSE	A	36				ATOM 285 CB MET A 36
ATOM	289	N	SER	A	37				ATOM 286 CG MET A 36
ATOM	290	CA	SER	A	37				ATOM 287 SD MET A 36
ATOM	291	C	SER	A	37				ATOM 288 CE MET A 36
ATOM	289	N	SER	A	37				ATOM 289 N SER A 37
ATOM	290	CA	SER	A	37				ATOM 290 CA SER A 37
ATOM	291	C	SER	A	37				ATOM 291 C SER A 37

ATOM	793	CZ	PHE	A	99	35.218	8.296	-4.457	
ATOM	794	OXT	PHE	A	99	39.959	5.701	-6.345	
TER									
ATOM	796	N	PRO	B	1	39.913	8.589	-2.684	
ATOM	797	CA	PRO	B	1	39.697	8.722	-1.247	
ATOM	798	C	PRO	B	1	39.481	7.384	-0.548	

NC(C)C(=O)O

 HID

NC(C)C(=O)O

 HIE

NC(C)C(=O)O

 HIP

C

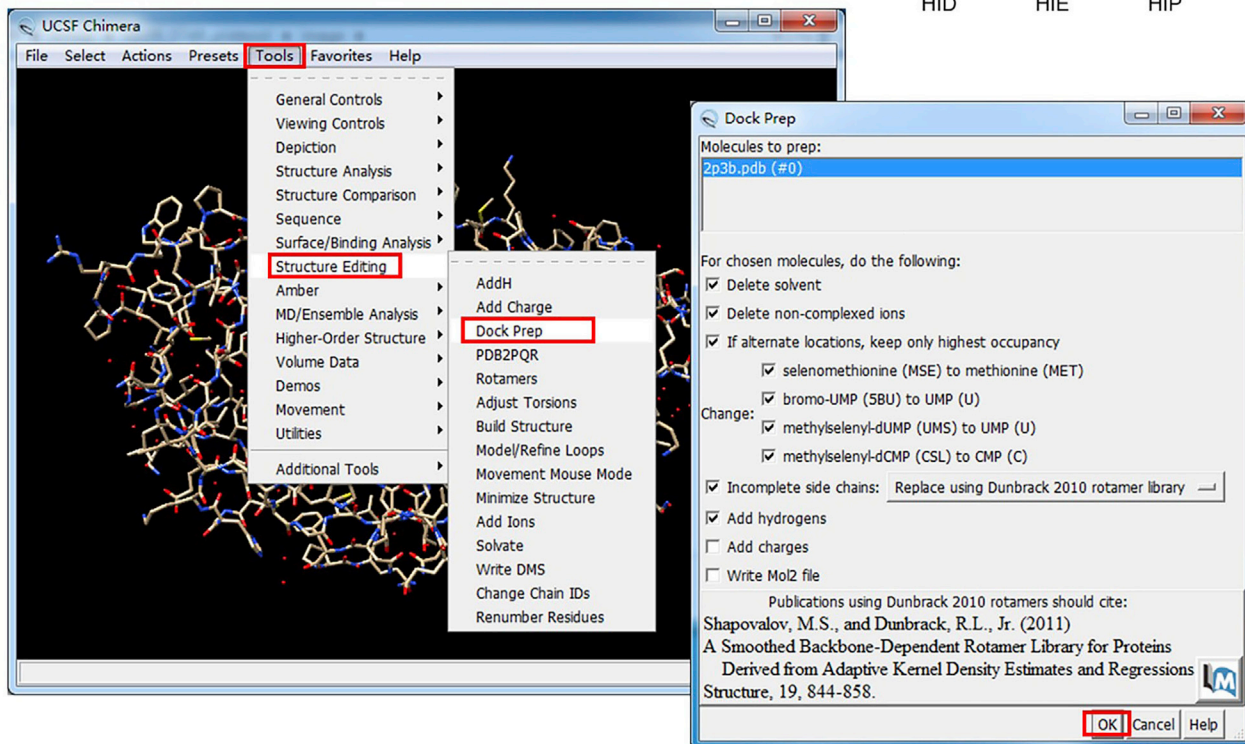


Figure 2. The complex structure file preparation

(A) Illustration of selecting and deleting the irrelative atoms from the complex by Chimera software.
 (B) Example of renaming amino acids according to their protonated states and adding a “TER” record. The positions of proton in different protonated states are highlighted in red: HID represents histidine with hydrogen on the delta nitrogen, HIE on the epsilon nitrogen, and HIP on both nitrogens.
 (C) Illustration of the structure preparation by the Dock Prep module in Chimera. This module can perform some complex operations, such as deleting solvent and isolated ions, keeping only the conformation with highest occupancy, repairing missing atoms, and adding hydrogen atoms.

KEY RESOURCES TABLE

REAGENT or RESOURCE	IDENTIFIER
Deposited data	
Three-dimensional structure	RCSB Protein Data Bank (PDB) PDB: 2P3B
ZINC	(Irwin and Shoichet, 2005) http://zinc12.docking.org/browse/subsets/
FDB-17	(Visini et al., 2017) https://gdb.unibe.ch/tools/
PADFrag	(Yang et al., 2018) http://chemyang.ccnuc.edu.cn/ccdb/database/PADFrag/index.php/help/download
Software and algorithms	
AILDE source code	This study https://github.com/fwangccnu/AILDE

MATERIALS AND EQUIPMENT

Equipment

- Computer: a computational workstation installed with proper software listed below is required.
 List of hardware of the computational workstation:
 CPU: Intel Core i5-6400 3.3GHz 4-core processor
 Memory: Kingston KVR16N11/8 8GB DDR3-1600
 GPU: GeForce GTX 1060 3GB
 Storage: Western Digital WD10EZEX 1TB

Software

- Chimera (version 1.13 or later, <https://www.cgl.ucsf.edu/chimera/>)
- OpenBabel (version 2.3.2 or later, http://openbabel.org/wiki/Main_Page)
- Python (version 2.7, <https://www.python.org/>)
- Amber and AmberTools (version 16, <https://ambermd.org/>)
- Autogrow (<http://autogrow.ucsd.edu/>)
- R package (version 3.0.0 or later, <https://www.r-project.org/>)
- A molecular visualizing software for viewing the complex structures, such as VMD (<https://www.ks.uiuc.edu/Research/vmd/>) or PyMOL (<https://pymol.org/2/>).

STEP-BY-STEP METHOD DETAILS

A fragment library construction

⌚ Timing: ~3 h

1. Create the three-dimensional structures of all fragments by Chimera software (Figure 3). The fragments are used to link with the hit compound to generate the hit analogs (Hoffer et al., 2018). All the hydrogen atoms should be deleted in the fragment molecules, except the hydrogen atoms located at the linking point.

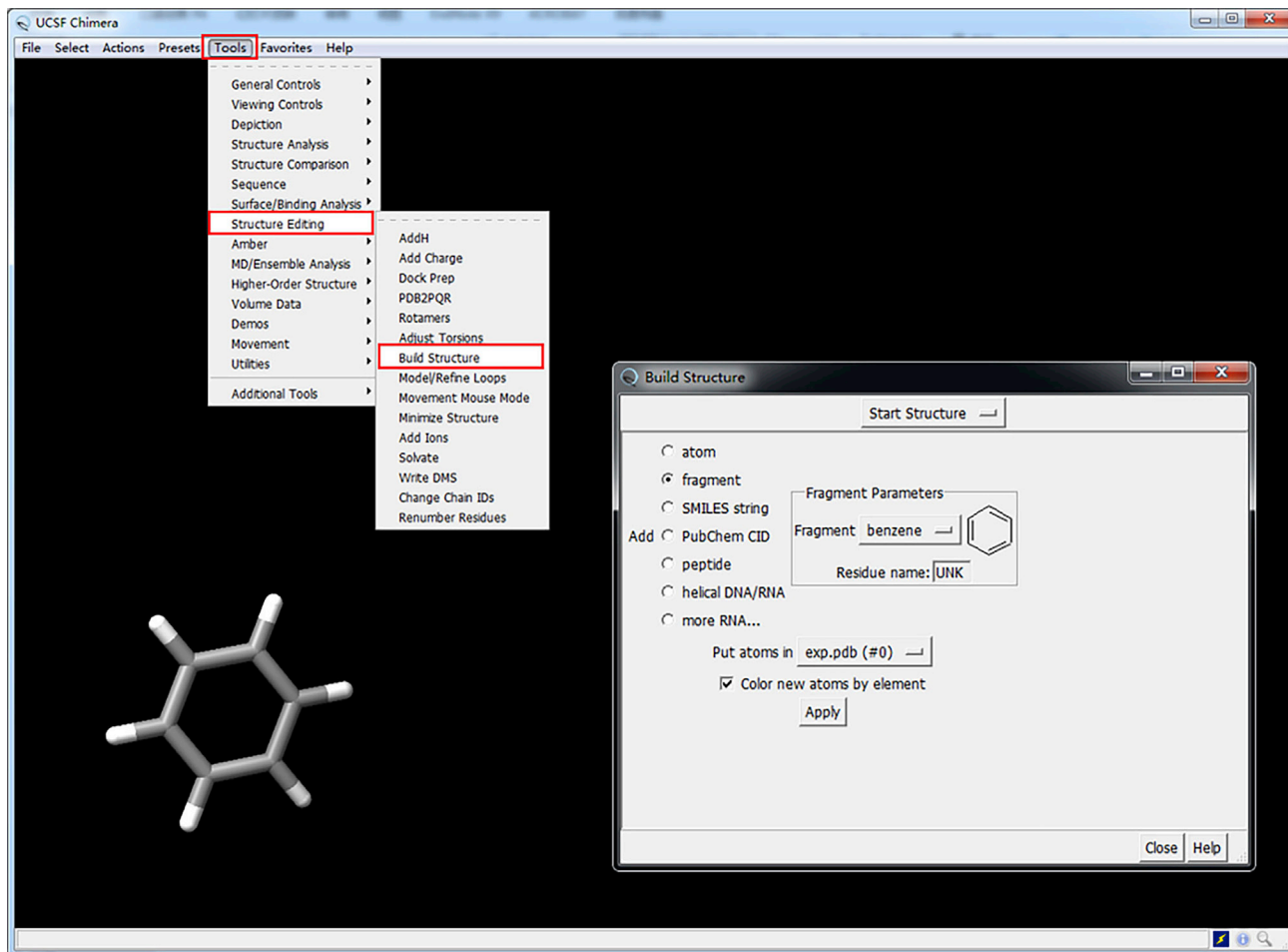


Figure 3. Construction of stereoscopic conformation of a fragment molecule

The Build Structure module in Chimera provides several ways to manipulate a molecular structure. This snapshot shows the procedure for building a benzene molecule.

Due to the fixed format of a PDB file, it is easy to do these by the scripts in batch mode. We can delete all the hydrogen atoms in a PDB file except the linking point of a hydrogen atom (serial number: *id*) by a line of bash command:

```
awk '{if ($NF == "H" && $2 != id) {} else {print}}' fragment_H.pdb > fragment_noH.pdb
```

where *id* is the serial number of a hydrogen atom as the linking point, *fragment_H.pdb* is the pdb file of a fragment molecule with all hydrogen atoms, *fragment_noH.pdb* is the fragment molecule with only a hydrogen atom.

For a large number of fragments, we just need to use the loop function to deal with each individual repeatedly:

```
for i in $(ls *.pdb)
do
  filename=$(basename ${i} .pdb)
  for j in $( awk '{if ($NF == "H") {print $2}}' ${i})
  do
```

```
awk '{if ($NF == "H" && $2 != ""$j)} else {print}}' ${i} > ${fragname}_${j}.pdb
done
done
```

We can save the above bash codes in a text file (for example: a file named *del_H.sh*) in the same directory as the fragment PDB files. And run the scripts by type the following in the command line:

```
bash del_H.sh
```

The processed PDB files of fragments are generated in the directory.

2. Perform the conformational optimization on the fragment molecules by Chimera software.
3. Save the structure files of those processed fragment molecules in PDB format.

Note: The fragment molecules can be download from public fragment databases ([Irwin and Shoichet, 2005](#); [Visini et al., 2017](#); [Yang et al., 2018](#)). The related website links are available in the [key resources table](#).

Preparing input parameter data

⌚ Timing: ~5 min

Prepare an input parameter file, in which all the parameters for AILDE calculation are contained.

4. Create a parameter file which contains all the necessary parameters for AILDE protocol ([Figure 4](#)). Mandatory parameters are "step," "complex," "ligand_name," "ligand_charge," "parallel," "MDstart," "MDend," "mm_pbsa," "md_request," and "entropy." Each parameter has a default value.
5. An example parameter file is provided in the AILDE source code ([Figure 4](#)). A standard format must be followed to define the parameters. Among those parameters, the following three parameters should be adjusted according to actual situation:
 - a. complex: define the structure file name of the protein complex with hit ligand.
 - b. ligand_name: define the name of the hit ligand in the complex structure file.
 - c. md_request: define whether structural relaxation is performed on the newly obtained hit analogs-protein complex by molecular dynamics (MD) simulation.
6. Other parameters can be maintained as default value and the detail descriptions about these parameters are recorded in the template parameter file ([Troubleshooting 1](#)).

⚠ CRITICAL: The parameters should be accordant with the actual situations and abide by the format in the example parameter file.

Molecular dynamics simulation of the protein-hit complex

⌚ Timing: ~3 h

In this stage, conformational sampling on the protein complex with hit ligand is conducted by molecular dynamics simulation ([Maximova et al., 2016](#)).

7. The topology and coordinate file for the solvated protein-hit complex is generated by *tleap* module in Amber program. Protein was parameterized by Amber ff14SB force field and the hit compound molecule by the general Amber force field (GAFF). AM1-BCC charge was assigned for the hit compound by *antechamber* module in Amber. The complex system was then solvated in the cube box of the TIP3P water model with a distance of 10 Å between the

A

fwangccnu Update README.md		1500f1e on 22 Apr	🕒 2 commits
📁 packages	first commit		7 months ago
📄 AILDE_RUN	first commit		7 months ago
📄 AILDE_run_all.py	first commit		7 months ago
📄 README	first commit		7 months ago
📄 README.md	Update README.md		7 months ago
📄 parameter.txt	first commit		7 months ago

B

```
#####this file is used to specify the parameters#####
#####1.specify the complex file#####

step          01-18          #default: total 15 steps, use '-' or ',' to specify the st
                  #e.g. 01,03,04,05,06   or 01-07,10-15
                  #space is not allowed when link the number

#####2.specify the complex file#####
complex        com.pdb          #default: com.pdb

#####3.specify the ligand name#####
ligand_name    LIG            #default: LIG

#####4.specify the ligand name charge#####
ligand_charge  gas           #default: gas

#####5.whether to use parallel#####
parallel       YES          #default: YES and 4 process

#####6.MD job settings#####
MDstart        3            #default: 3  should be more than 3 and smaller than MDend
MDend          15          #default: 15

#####7.mm_pbsa.in locations#####
mm_pbsa        ./packages/files/mmpbsa/mm_pbsa.in  #must be specified

#####8.md refine request#####
md_request     NO           #default: NO

#####9.entropy calculation file locations#####
entropy        ./packages/files/entropy/      #must be specified
```

Figure 4. Preparing the input parameter file

(A) The snapshot of webpage for downloading the example parameter file at <https://github.com/fwangccnu/AILDE>.

(B) The content in the example parameter file.

solute and box. The counter-ions (Cl^- or Na^+) were then added to neutralize the solvated protein-hit system.

8. Three cycles of energy minimization are performed on the system by *sander* module in Amber program. Each cycle includes 5,000 steps of steepest descent, followed by 5,000 steps of conjugated gradient minimization with a final convergence criterion of $0.2 \text{ kcal}\cdot\text{mol}^{-1}$. The following shows a brief description about three cycles of energy minimization:
 - a. Cycle 1: the heavy atoms in protein and hit compound are restrained with an elastic constant of $50 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$. The water molecules, ions, and hydrogen atoms are allowed to move for relaxation.
 - b. Cycle 2: protein side-chains are also allowed to relax in this step. Only the heavy atoms in the backbone of the protein are restrained with an elastic constant of $50 \text{ kcal}\cdot\text{mol}^{-1}\cdot\text{\AA}^{-2}$.
 - c. Cycle 3: the whole system is minimized.
9. Molecular dynamics simulation is then performed on the protein-hit complex with the periodic boundary condition in the NPT ensemble. The detail steps are described as the following:
 - a. The solvated system is heated gradually from 0 to 298 K in the NVT ensemble during a period of 500 ps.
 - b. The heated system is equilibrated in the NPT ensemble during a period of 1 ns.
 - c. Production simulation at the temperature of 298 K and the pressure of 1 atm. The SHAKE algorithm is used to constrain all covalent bonds involving hydrogens. The particle mesh Ewald (PME) algorithm is used to calculate long-range electrostatic interactions and the short-range van der Waals (vdW) interactions is handled with a cutoff distance of 10 \AA . The integration of the equations of motion is conducted at a time step of 2 fs. Finally, 1 ns production simulation is performed and the structure snapshots are collected every 1 ps ([Troubleshooting 2](#)).
10. The fluctuations of the protein and hit compound are measured by root-mean-square deviations (RMSD) against their initial structures. The stability of the MD simulation is reflected by their RMSD values, which reveal the conformational changes during the time period of MD simulation. A plateau RMSD value means that the protein-hit complex is in the equilibration state, and that the simulation moves toward the stability status.

Collecting the conformational ensemble of protein-hit complex

⌚ Timing: ~10 min

After the molecular dynamics simulation, we collected the representative conformations from the equilibrated simulation trajectories. Consideration of structural flexibility has an effect on the accuracy of binding free energy calculation. Predicting free energy changes using structural ensembles has better performance than that using static structures ([Benedix et al., 2009](#)).

11. 50 snapshots are extracted from the last 1 ns simulation trajectory by the *cpptraj* module in Amber program.

Fragment growing to generate the hit analogs

⌚ Timing: ~1 h

In this stage, AILDE generates hit analogs by linking the fragment molecules with the hit ligand in the initial protein complex ([Figure 5](#)).

12. All fragment molecules in our fragment library are linked with the hit ligand by AutoGrow software.
13. The generated compounds with the protein are saved as protein-analogs complexes. The complex structures are relaxed by energy minimization via one of the following two strategies:

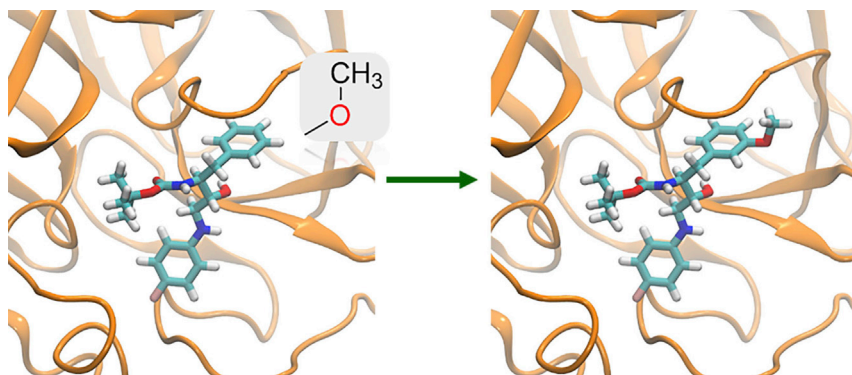


Figure 5. Lead compounds generation by fragment growing strategy

This illustrates the process of converting a hit compound to a lead compound by fragment linking with the former.

- a. When the parameter “md_request” is set to “NO,” structural relaxation without molecular dynamics simulation will be performed. First, all the backbone atoms of protein and the ligand are restrained with an elastic constant of $50 \text{ kcal} \cdot \text{mol}^{-1} \cdot \text{\AA}^{-2}$. Then the whole system is minimized. Each of the two minimization steps includes 5,000 cycles of steepest descent and 5,000 cycles of conjugated gradient optimization with a convergence criterion of $0.2 \text{ kcal} \cdot \text{mol}^{-1}$.
 - b. When the parameter “md_request” is set to “YES,” structural relaxation with molecular dynamics simulation will be performed. The above two cycles of energy minimizations is performed. Then a short period of MD simulation is performed to relax the side-chain atoms.
14. All the structure snapshots of protein-hit complex are relaxed by the same strategy as mentioned above ([Troubleshooting 3](#)).

Note: In our previous study ([Wu et al., 2020](#)), structural relaxation with molecular dynamics simulation may provide more accurate binding free energy results, but at the cost of more computing resources.

Predicting free energy changes for binding affinities

⌚ Timing: ~2 h

In this stage, we calculate the binding affinity (ΔG_{bind}) of protein-hit and protein-analogs complexes, then compare the differences of binding free energy change ($\Delta\Delta G$) between protein-hit and protein-analogs complexes.

15. The binding affinity ΔG_{bind} can be decomposed into three main energy terms, including the molecular mechanical gas-phase interaction energy (ΔE_{MM}), desolvation energy (ΔG_{solv}) and the change of conformational entropy ($-T\Delta S$). ΔE_{MM} and ΔG_{solv} are calculated by the MMPBSA module in Amber program. $-T\Delta S$ is calculated using our in-house scripts. The entropy contribution is temperature-dependent and can be divided to two contributions: solvation entropy (ΔS_{solv}) and conformational entropy (ΔS_{conf}). The solvation entropy consists of the apolar and polar solvation entropy, which is typically given by an empirical formula that is proportional to the apolar and polar solvent accessible surface area differences. We use the *molsurf* and *surf* modules in Amber program to calculate the solvent accessible surface area. The conformational entropy is related to the change of the number of rotatable bonds during the binding process. The change of the conformational entropy is proportional to the number of the lost rotatable bonds during the binding process. The *conf* module in Amber is used to calculate the number of the rotatable bonds. More details can be found in the previous work ([Pan et al., 2008](#)). The binding affinity ΔG_{bind} for a complex is the average value of 50 structural snapshots.

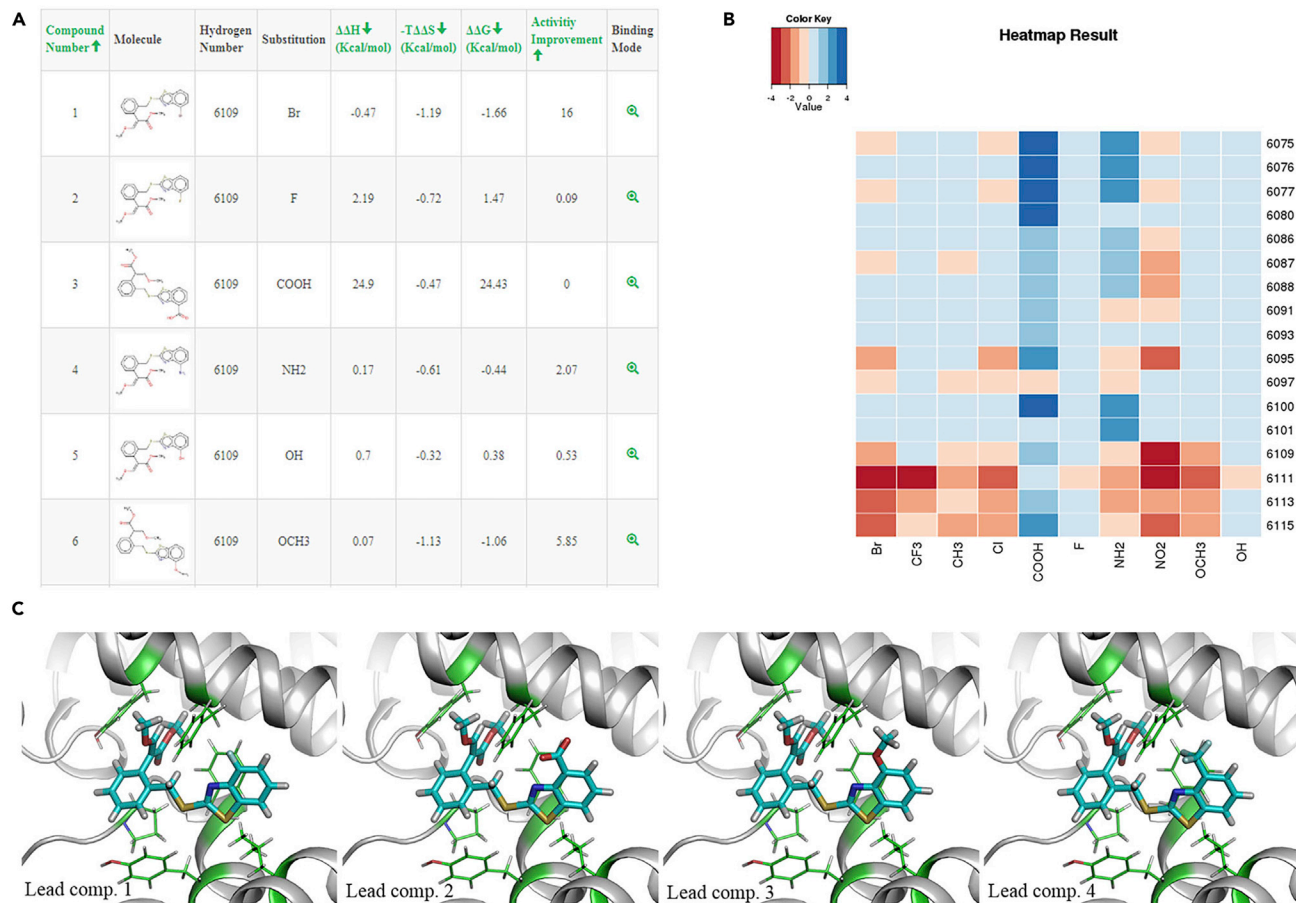


Figure 6. Examples of the expected outcomes of AILDE method

(A) The calculated binding free energy of the lead compounds with a protein receptor.

(B) Heatmap showing the activities of the lead compounds obtained by fragment linking. The horizontal axis represents the fragment molecules, and the vertical axis represents the atomic number of the atoms connecting the fragments. The darker the red, the more active the compound is. The darker the blue, the less active the compound is.

(C) The interaction modes of the lead compounds with protein receptor.

16. The binding affinity differences $\Delta\Delta G$ between the protein-hit and protein-analogs complexes are then obtained as $\Delta\Delta G = \Delta G_{\text{bind}}(\text{Analogs}) - \Delta G_{\text{bind}}(\text{Hit})$.

EXPECTED OUTCOMES

Figure 6 illustrates the examples of the expected outcomes of this protocol.

We obtain a series of hit analogs from the initial hit molecule. The binding affinity of each analog compound interacting with the protein has been evaluated. Then the binding affinity changes between the protein-analogs and protein-hit complexes are calculated. A negative change of binding free energy indicates an increase in the binding affinity, while a positive change of binding free energy indicates a decrease in the binding affinity. Potential lead compounds are identified based on the rank of the changes of binding free energy. Figure 6A shows the results of the binding affinity changes, which contain the information about the analog compound structures, fragment molecules, substituent positions, and binding affinity changes. In addition, a heatmap is plotted to graphically display the relationship between the generated analog compounds and their binding affinity changes (Figure 6B). The horizontal axis represents fragment molecules, and the vertical axis represents hydrogen atoms where the fragment molecules are linked. The darker the red, the more active

the compound is. The darker the blue, the less active the compound is. AILDE not only generates the analog compounds, also predict their binding poses in the protein. The binding modes of the compounds interacting with the protein are explored by analyze the complex structures (Figure 6C).

LIMITATIONS

AILDE is based on an assumption that there are no substantial changes in the binding modes between protein-hit and protein-analogs complexes. AILDE could not achieve a good performance in a complicated situation of activity cliffs. This is caused by the compounds with similar structures, but have large differences in binding affinity (Stumpfe and Bajorath, 2012). In this protocol, a small fragment library is applied to generate the potential lead compounds. It is suggested to utilize larger chemical space to search and predict more possible lead compounds. The molecular weight of a fragment molecule is recommend to be less than 90 g/mol for a good performance.

TROUBLESHOOTING

Problem 1

AILDE reminders that there are any files not found.

Potential solution

All the input files for AILDE are identified in the parameter file. We should prepare a suitable parameter file according to the example file provided in AILDE source code. We recommend changing the paths of the necessary data file in the example file.

Problem 2

The molecular dynamics simulation terminates abnormally.

Potential solution

We recommend using the protein complex with only one ligand compound as the initial structure to perform the molecular dynamics simulation. In preparing the input structure file of the protein complex with a hit ligand, we could remove the unconcerned chemical compounds in the complex structure.

Problem 3

The structures of the hit analogs generated by fragment growing are weird.

Potential solution

We recommend that the conformational optimization should be performed for the fragment molecules in the step of a fragment library construction. Chimera software could be used to conduct the optimization.

RESOURCE AVAILABILITY

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Gefei Hao (gefei_hao@foxmail.com).

Materials availability

This study did not generate new unique reagents.

Data and code availability

The online web service of AILDE can be accessed at <http://chemyang.cnu.edu.cn/ccb/server/AILDE>. The code of AILDE can be downloaded at <https://github.com/fwangcnu/AILDE>.

ACKNOWLEDGMENTS

This work was supported by the National Natural Science Foundation of China (21772059, 91853127, and 31960548), Program of Introducing Talents of Discipline to Universities of China (111 Program, D20023), and Frontiers Science Center for Asymmetric Synthesis and Medicinal Molecules, Department of Education, Guizhou Province (Qianjiaohe KY number (2020)004).

AUTHOR CONTRIBUTIONS

Conceptualization, G.H.; programming, F.W.; writing – original draft, L.M. and F.W.; writing – review & editing, G.H. and G.Y.

DECLARATION OF INTERESTS

The authors declare no competing interests.

REFERENCES

- Benedix, A., Becker, C.M., de Groot, B.L., Caflich, A., and Bockmann, R.A. (2009). Predicting free energy changes using structural ensembles. *Nat. Methods* 6, 3–4.
- Hao, G.-F., Yang, S.-G., Huang, W., Wang, L., Shen, Y.-Q., Tu, W.-L., Li, H., Huang, L.-S., Wu, J.-W., Berry, E.A., et al. (2015). Rational design of highly potent and slow-binding cytochrome bc1 inhibitor as fungicide by computational substitution optimization. *Sci. Rep.* 5, 13471.
- Hoffer, L., Voitovich, Y.V., Raux, B., Carrasco, K., Muller, C., Fedorov, A.Y., Derviaux, C., Amouric, A., Betzi, S., Horvath, D., et al. (2018). Integrated strategy for lead optimization based on fragment growing: the diversity-oriented-target-focused-synthesis approach. *J. Med. Chem.* 61, 5719–5732.
- Irwin, J.J., and Shoichet, B.K. (2005). ZINC—a free database of commercially available compounds for virtual screening. *J. Chem. Inf. Model.* 45, 177–182.
- Maximova, T., Moffatt, R., Ma, B., Nussinov, R., and Shehu, A. (2016). Principles and overview of sampling methods for modeling macromolecular structure and dynamics. *PLoS Comput. Biol.* 12, e1004619.
- Meng, X.Y., Zhang, H.X., Mezei, M., and Cui, M. (2011). Molecular docking: a powerful approach for structure-based drug discovery. *Curr. Comput. Aid. Drug Des.* 7, 146–157.
- Pan, Y., Gao, D., and Zhan, C.G. (2008). Modeling the catalysis of anti-cocaine catalytic antibody: competing reaction pathways and free energy barriers. *J. Am. Chem. Soc.* 130, 5140–5149.
- Stumpfe, D., and Bajorath, J. (2012). Exploring activity cliffs in medicinal chemistry. *J. Med. Chem.* 55, 2932–2942.
- Visini, R., Awale, M., and Reymond, J.L. (2017). Fragment database FDB-17. *J. Chem. Inf. Model.* 57, 700–709.
- Wu, F.X., Wang, F., Yang, J.F., Jiang, W., Wang, M.Y., Jia, C.Y., Hao, G.F., and Yang, G.F. (2018). AIMMS suite: a web server dedicated for prediction of drug resistance on protein mutation. *Brief. Bioinform.* 21, 318–328.
- Wu, F.X., Zhuo, L.S., Wang, F., Huang, W., Hao, G.F., and Yang, G.F. (2020). Auto in silico ligand directing evolution to facilitate the rapid and efficient discovery of drug lead. *iScience* 23, 101179.
- Yang, J.F., Wang, F., Jiang, W., Zhou, G.Y., Li, C.Z., Zhu, X.L., Hao, G.F., and Yang, G.F. (2018). PADFrag: a database built for the exploration of bioactive fragment space for drug discovery. *J. Chem. Inf. Model.* 58, 1725–1730.