

GC and CELPP: Workflows and Insights

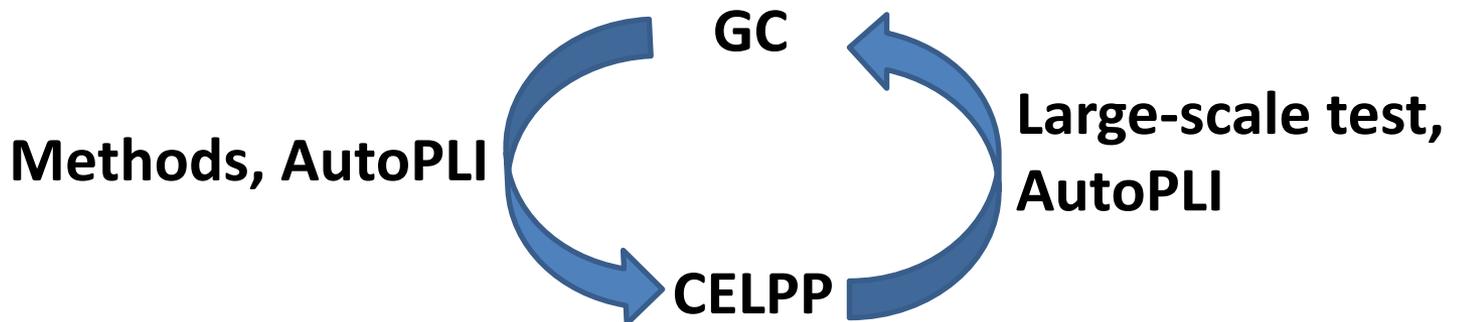
Xianjin Xu, Zhiwei Ma, Rui Duan, Xiaoqin Zou

Dalton Cardiovascular Research Center, Department of Physics and
Astronomy, Department of Biochemistry, & Informatics Institute

University of Missouri - Columbia

Outline

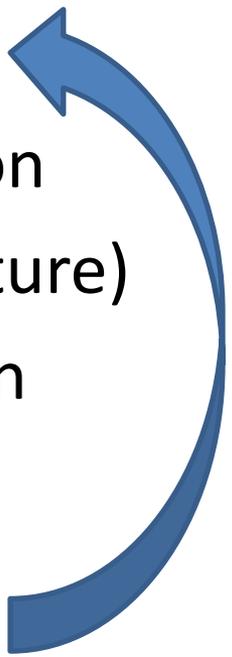
- **Methodology**
- **GC/CELPP results and the lessons we've learned**
- **AutoPLI: A fully automated server for predicting protein-ligand interactions**
- **Conclusion**



Challenges for protein-ligand binding mode and affinity predictions

Binding mode prediction:

- Protein flexibility (prediction of the bound protein structure)
- Scoring function (prediction of the energy minimum)



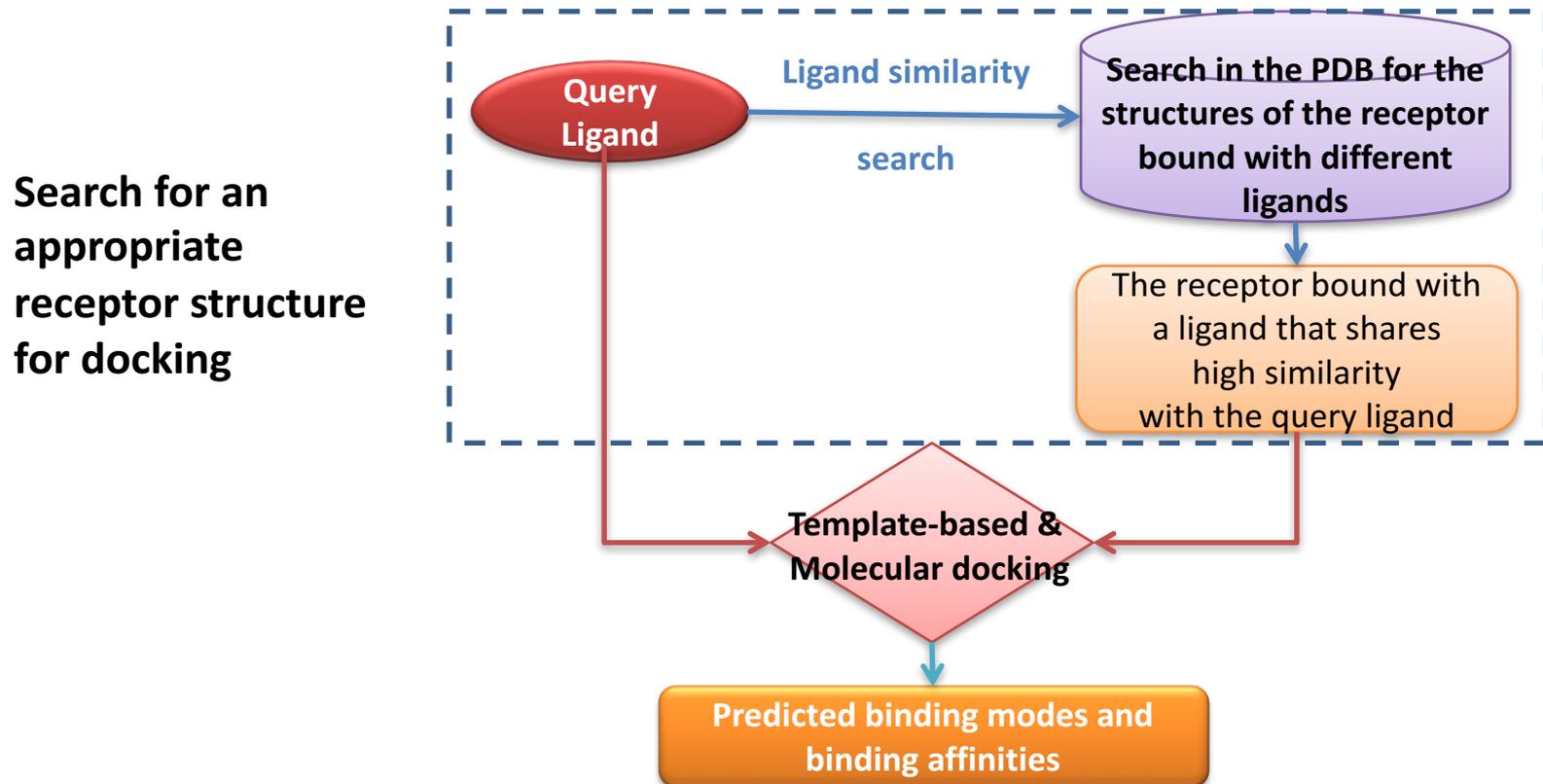
The affinity prediction is dependent of the mode prediction.

Binding affinity/ranking prediction:

- Scoring function (prediction of the whole energy landscape)

Methodology

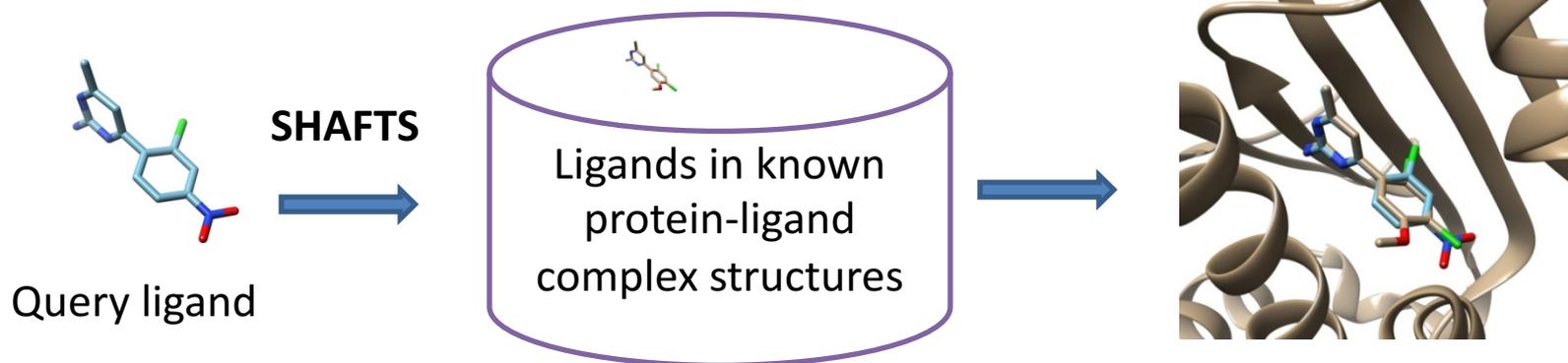
Search for a receptor structure with a bound ligand that shares high similarity with the query ligand for the binding mode/affinity prediction.



If there is no reliable receptor structure for docking, **ensemble docking** (i.e., using multiple protein structures) is performed.

Step 1: Search for an appropriate receptor structure for rigid receptor docking

- **Construct a receptor structural database**, containing all the released protein-ligand complex structures of the receptor in the Protein Data Bank.
- **Ligand similarity calculation: SHAFTS**
The similarity is based on the shape overlay and pharmacophore feature matching.



The receptor structure with a bound ligand that shares the highest similarity with the query ligand will be used for binding mode prediction.

Step 2.1: Template-based method

- 1) Generate up to 500 conformers for the query ligand using **Omega2** (OpenEye Scientific Software)
- 2) Superimpose each conformer on the known co-bound ligands of the receptor, and calculate the ligand similarity scores using **SHAFTS**
- 3) Binding modes are ranked by the similarity scores (or reranked by ITScore), and the top 5 models are refined by MD simulation.

Step 2.2: Molecular docking

Binding mode sampling:

Program: Modified AutoDock Vina 1.0

Receptor: rigid

Ligand: flexible

Exhaustiveness = 30

Output models = Up to 500

Step 3: Scoring and ranking: ITScore

- A statistical potential-based scoring function, ITScore, was used to evaluate the generated models. The scores are also used for binding affinity prediction.
- The scoring function was developed using the iterative method based on the refined set of PDBbind 2012.

Huang and Zou, *J. Comput. Chem.* 2006, 27, 1866-1875.

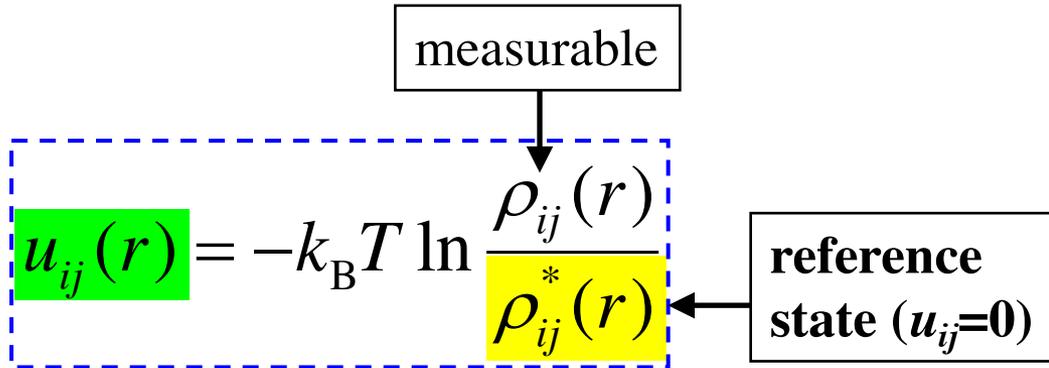
Yan *et al.*, *J. Chem. Inf. Model.* 2016, 56, 1013.

Wang *et al.*, *J. Med. Chem.* 2005, 48, 4111-4119.

Cheng *et al.*, *J. Chem. Inf. Model.* 2009, 49, 1079-1093.

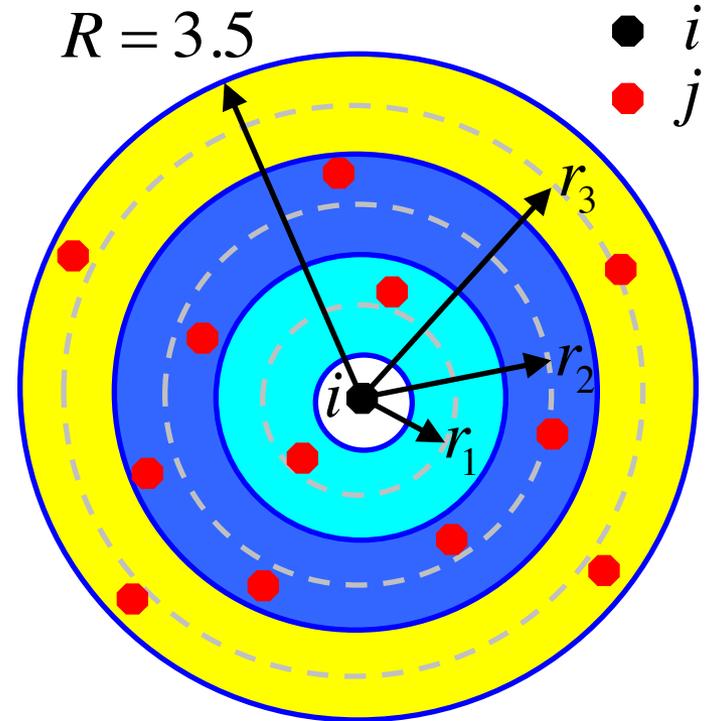
Traditional formalism to derive the statistical pair potentials

$$\rho_{ij}(r) = \rho_{ij}^*(r) \cdot \exp\left(-\frac{u_{ij}(r)}{k_B T}\right)$$



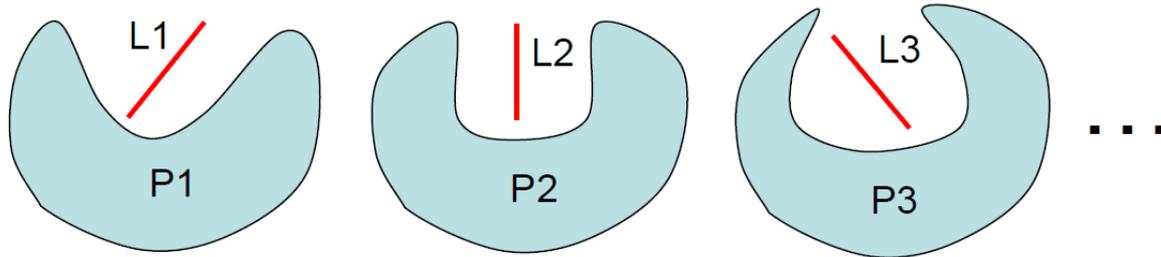
An example: "ideal gas"

$$\Delta G = \sum_{ij} u_{ij}(r)$$



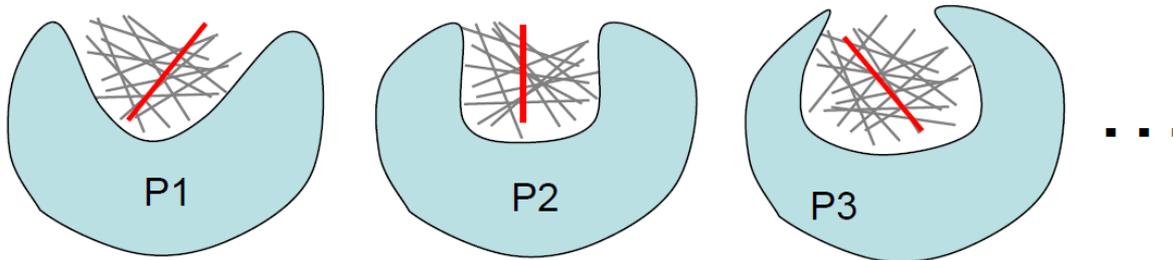
The reference state problem is a big hurdle for this inverse algorithm!

Derivation of the effective pair potentials using statistical mechanical principles



$$g_{ij}^*(r) = \frac{1}{K} \sum_{k=1}^K g_{ij}^{k*}(r)$$

$$g_{ij}^{k*}(r) = \rho_{ij}^{k*}(r) / \rho_{ij,\text{bulk}}^{k*}$$



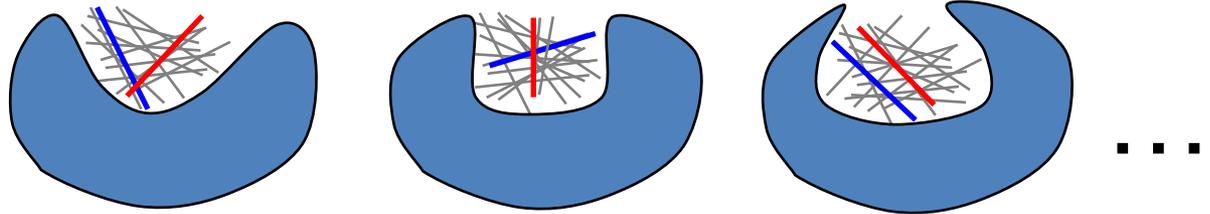
$$g_{ij}^{(n)}(r) = \frac{1}{K} \sum_{k=1}^K \sum_{l=0}^L P_k^l g_{ij}^{kl}(r)$$

$$P_k^l = \frac{e^{-\beta E_k^l}}{Z_k} \quad Z_k = \sum_{l=0}^L e^{-\beta E_k^l}$$

Our physics-based iterative method circumvents the reference state problem

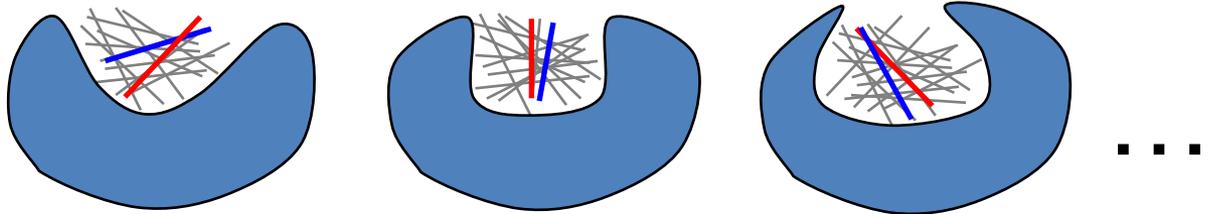
Step 0: $g_{ij}^{obs}(r)$

$$u_{ij}^1(r) = u_{ij}^0(r) + \Delta g_{ij}^0(r)$$



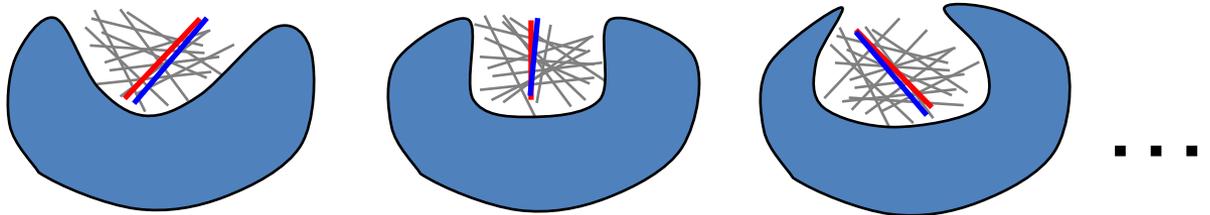
Step n: $g_{ij}^n(r)$

$$u_{ij}^{n+1}(r) = u_{ij}^n(r) + \Delta g_{ij}^n(r)$$



Step N: $g_{ij}^N(r)$

$$u_{ij}^{N+1}(r) = u_{ij}^N(r) + \Delta g_{ij}^N(r)$$



until $g_{ij}^N(r) \longrightarrow g_{ij}^{obs}(r)$

GC2 results and analysis

GC2_2016: FXR

36 compounds (FXR_1 to FXR_36) for binding mode prediction.

A total of 26 known FXR–ligand complex structures were retrieved from the PDB for GC2 prediction.

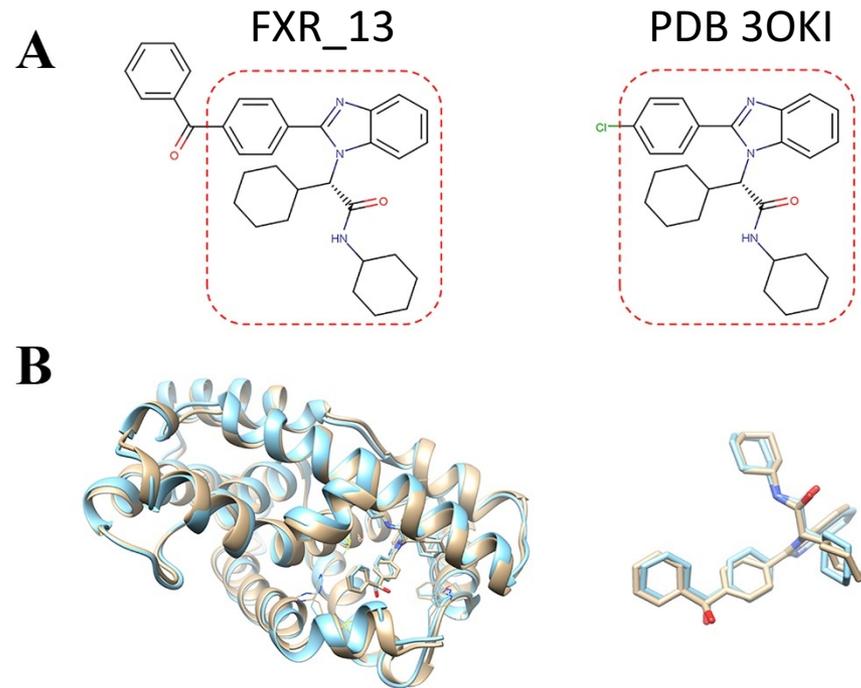


Fig. A. Chemical structures of compound FXR_13 (left) and ligand OKI in PDB entry 3OKI (right). **B.** The top model from the template-based approach (cyan) aligned with the crystal structure (tan) released by D3R.

FXR Results

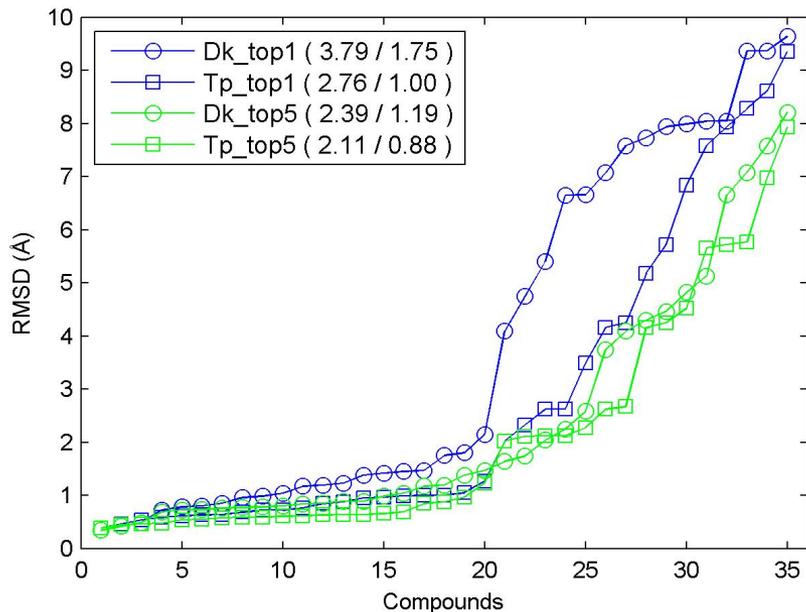


Fig. 1 The results of binding mode prediction for the FXR dataset (i.e., FXR_1–FXR_36) based on the top model (blue) or the best among the top 5 models (green) using the **docking approach (Dk)** or the **template-based approach (Tp)** with the scoring function ITScore.

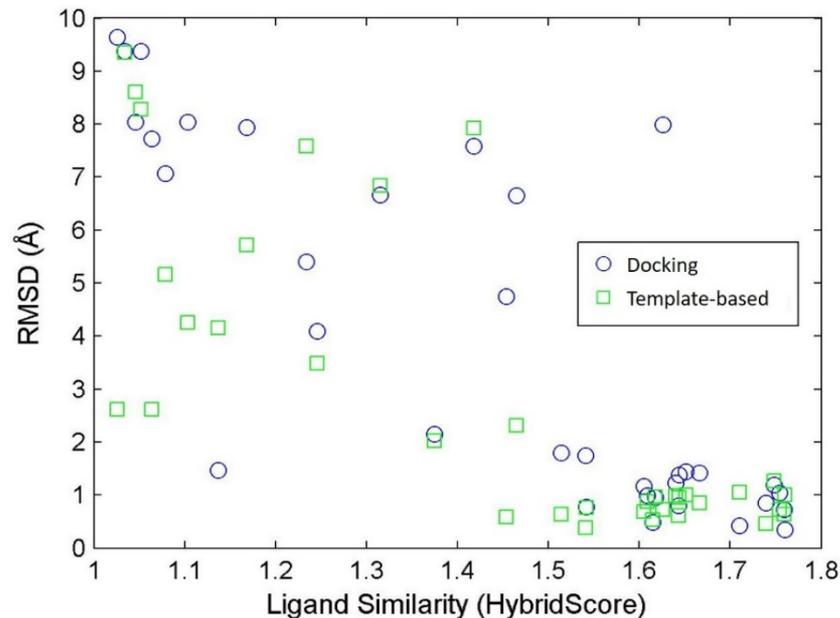


Fig. 2 Ligand RMSD values vs ligand similarities (HybridScore) of the query ligand against the ligand in the template for both the docking approach and the template-based method (based on the top model for each ligand).

For top 5 models, docking-based and template-based methods achieved similar performances.

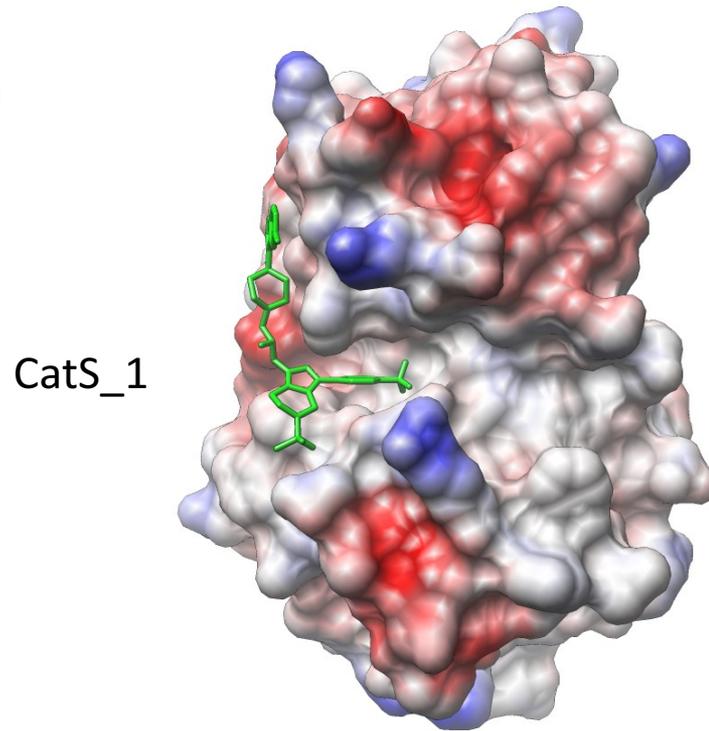
For the top model, better performances were achieved by the template-based approach if similar ligands were found.

GC3 results and analysis

GC3_2017: CatS (Cathepsin S)

24 compounds (CatS_1 to CatS_24) for binding mode prediction.

A total of 27 known CatS–ligand complex structures were retrieved from the PDB for GC3 prediction.



Challenge: Entropy seems to play an important role for binding in this case; Low energy score when using the X-ray complex structure; Possible effect of the neighboring subunit in the crystal

CatS: Binding mode prediction

Submitted Results:

Method	Mean RMSD of Pose 1 (Å)	Median RMSD of Pose 1 (Å)	Top 5: Mean RMSD of lowest-RMSD pose (Å)	Top 5: Median RMSD of lowest-RMSD pose (Å)
Vina	9.9	10.6	8.0	7.9
Vina/ITScore	9.9	8.3	6.5	5.9
Vina_bound	10.8	11.2	8.6	8.6
Vina/ITScore_bound	11.2	10.6	7.3	6.8
Template_based	4.3	3.5	3.4	2.4

The failure of docking methods due to the inaccuracy of current scoring functions, especially for the entropy calculation.

CELPP results and analysis

- Run predictions every week (Targets released on Saturday and submitted on Tuesday)
- **38 weeks** (Joined from week20_2017), and continues
- Predicted over **1400 targets**

For each target, up to 5 protein structures were provided by CELPP:

- **LMCSS:** The **Candidate** protein that contains the ligand with the largest maximum common substructure (MCSS) to the **Target** ligand.
- **SMCSS:** The **Candidate** protein that contains the ligand with the smallest maximum common substructure (MCSS) to the **Target** ligand.
- **hiResHolo:** **Highest resolution** ligand-bound **Candidate** protein.
- **hiResApo:** Highest resolution unbound **Candidate** protein.
- **hiTanimoto:** Similar to LMCSS, a different method for calculating ligand similarities.

CELPP: Binding mode prediction

- Use the **docking method** (Vina sampling and ITCscore reranking); not template-based
- Only **top 1** model was submitted for each protein structure
- Targets with incorrect binding sites were discarded

Protein selection	Number of targets	Mean RMSD of Pose 1 (Å)	Median RMSD of Pose 1 (Å)
LMCSS	1222	4.0	4.6
SMCSS	1222	5.4	6.3
hiResApo	736	6.3	6.7
hiResHolo	1222	5.0	5.5
hiTanimoto	1222	4.2	4.2
Best SHAFTS score	1059	4.0	4.7

Using a proper protein structure for docking significantly improves binding mode prediction. (Suggest to use the protein structure from LMCSS, Best SHAFTS score or hiTanimoto.)

A wish list for CELPP

- 1) Allow top 5 models for each prediction.
- 2) Discard targets with unreasonable ligands, such as tiny ligands (atom numbers < 6), ions, detergents, and cofactors (for which docking alone is incorrect).
- 3) Discard targets with incorrect/unfavorable binding sites.
- 4) Discard trivial targets of which the binding modes are already deposited in the PDB (e.g., using the ligand similarity as a cutoff).

The lessons we've learned from D3R

- 1) The information extracted from known protein-ligand complex structures may significantly improve binding mode prediction.
- 2) If a similar co-bound ligand is found, the template-based method usually achieves better performance than docking methods for top 1 prediction. Performances are similar for top 5.
- 3) Docking with an appropriate receptor structure achieves better performance than docking with multiple receptor structures (ensemble docking).
- 4) If the receptor structure is not accurate, ensemble docking usually achieves better performance than single-receptor docking.

Web Server: AutoPLI

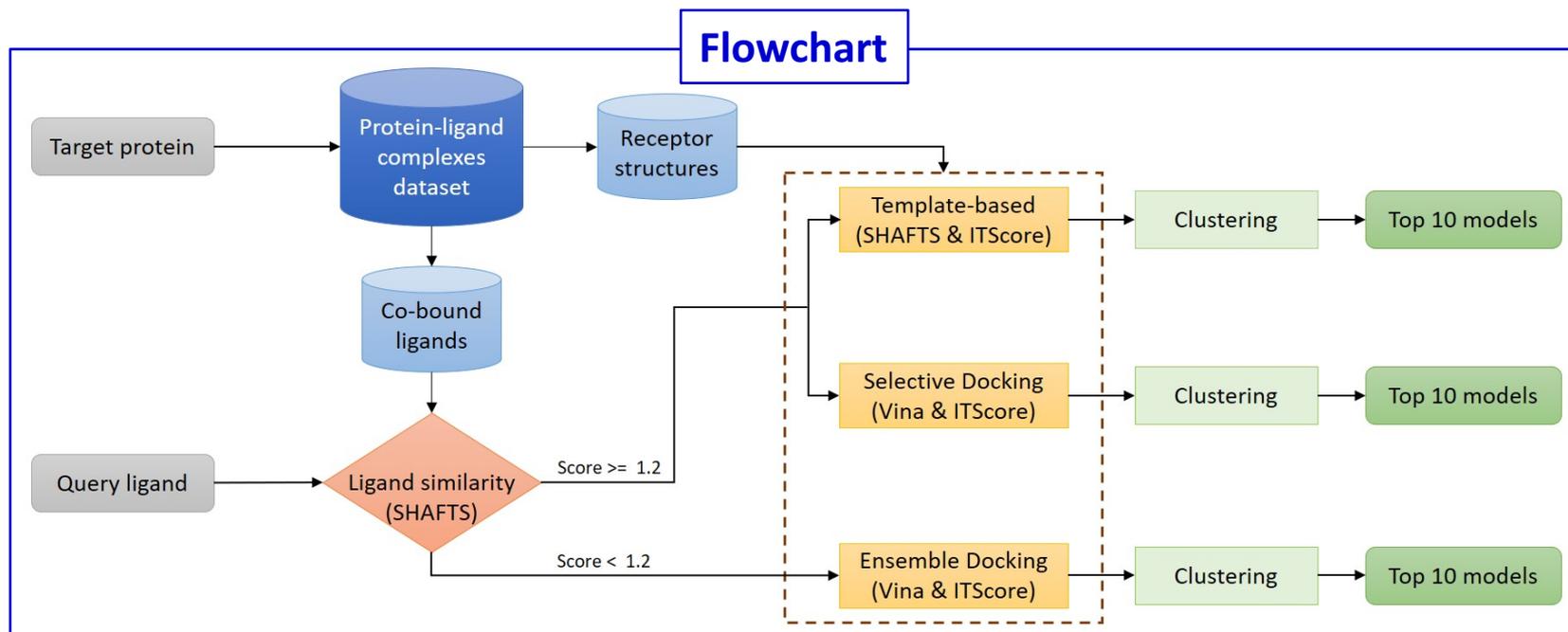
Website: <http://zougrouptoolkit.missouri.edu/autopli>

Inputs: (1) A ligand structure in the MOL2 format;
(2) A target protein with a UniProt ID or a 3D structure in the PDB format.

Output: Top 10 predicted protein-ligand complex structures in the PDB format.

Methods: Template-based; Selective docking; Ensemble docking

AutoPLI is a **fully automated** web server that **uses the information embedded in the available complex structures** for predicting protein-ligand interactions.



Job submission

To submit a job, two inputs are required:

(1) A ligand structure in the MOL2 format;

(2) A target protein with UniProt access number/ID or 3D structure in the PDB format.

Job Submission

→ A ligand structure in MOL2 format:

→ A target protein: ? **OR**

→ Job name (optional):

→ Email address (optional) ? :

→ Do not show my job on the queue page

Upload a PDB structure ▼

upload a protein structure in PDB format:

&& the center of the binding site ? :

/ /

Example: Ligand structure : [lig.mol2](#);

Protein receptor: [UniProt AC Q96RI1](#) (UniProt ID [NR1H4_HUMAN](#)) **or** PDB file: [rec.pdb](#) (Binding site: 43.2 / 14.3 / 2.7)

Queue page

Once the job is successfully submitted, the job status is monitored on the "Queue" page. The user will receive an email (if provided) notification after the job is completed.

Show entries

Search:

Job ID	Job name	Submission time	Email	Status
P00118	test8	2017-12-27/11:09:19	xxx@mail.missouri.edu	Done
P00115	mk_8	2017-12-27/00:09:00	xxx@gmail.com	Done
P00114	mk_7	2017-12-27/00:08:43	xxx@gmail.com	Done
P00113	mk_6	2017-12-27/00:08:27	xxx@gmail.com	Done
P00112	mk_5	2017-12-27/00:08:12	xxx@gmail.com	Done
P00111	mk_4	2017-12-27/00:07:55	xxx@gmail.com	Done
P00110	mk_3	2017-12-27/00:07:39	xxx@gmail.com	Done
P00109	mk_2	2017-12-27/00:07:21	xxx@gmail.com	Done
P00108	mk_1	2017-12-27/00:07:07	xxx@gmail.com	Done
P00107	human_case	2017-12-26/23:49:06	NA	Done

Results

(Template-based)

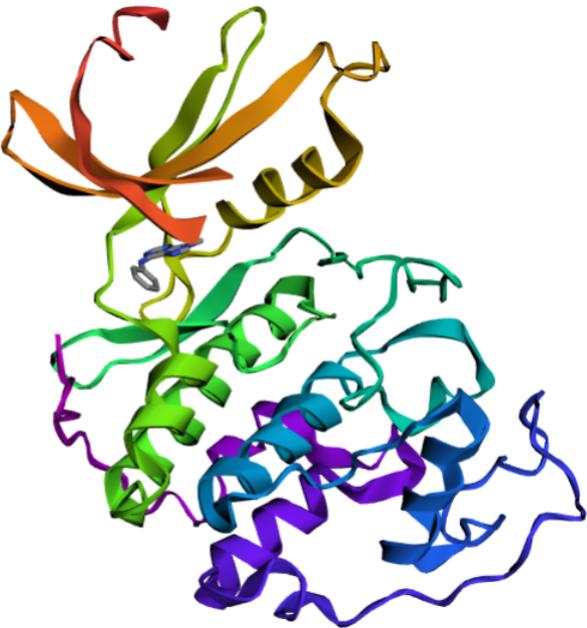
ligand MOL2: [lig.mol2](#)

Protein UniProt AC/ID: [CDK2_HUMAN](#)

Top (up to) 10 predicted binding modes using SHAFTS (template-based)

Click the "View" button on the right panel to display the corresponding binding mode.

Click "Model #" to download the corresponding PDB file.



Models	Active	View	PDB	SHAFTS	ITScore
Model 1	<input checked="" type="checkbox"/>	<input type="button" value="View"/>	2wipA	1.2	-41.8
Model 2	<input type="checkbox"/>	<input type="button" value="View"/>	4ez7A	1.3	-39.5
Model 3	<input type="checkbox"/>	<input type="button" value="View"/>	3qqgA	1.3	-36.5
Model 4	<input type="checkbox"/>	<input type="button" value="View"/>	4fkoA	1.3	-36.1
Model 5	<input type="checkbox"/>	<input type="button" value="View"/>	1e1vA	1.2	-35.4
Model 6	<input type="checkbox"/>	<input type="button" value="View"/>	4fkuA	1.2	-33.9
Model 7	<input type="checkbox"/>	<input type="button" value="View"/>	2c5nA	1.3	-32.7
Model 8	<input type="checkbox"/>	<input type="button" value="View"/>	3rprA	1.2	-32.7
Model 9	<input type="checkbox"/>	<input type="button" value="View"/>	3rprA	1.3	-32.3
Model 10	<input type="checkbox"/>	<input type="button" value="View"/>	5aniA	1.3	-32.0

Download

Click [Here](#) to download all the results.

Results

(selective docking)

Top 10 predicted binding modes (docking-based)

The PDB 1ckpA (SHAFTS score: 1.9) was selected for docking.

Click the "View" button on the right panel to display the corresponding binding mode.



Models	Active	View	ITScore	Download
Model 1	<input checked="" type="checkbox"/>	<input type="button" value="View"/>	-37.7	Download
Model 2	<input type="checkbox"/>	<input type="button" value="View"/>	-37.1	Download
Model 3	<input type="checkbox"/>	<input type="button" value="View"/>	-36.3	Download
Model 4	<input type="checkbox"/>	<input type="button" value="View"/>	-34.9	Download
Model 5	<input type="checkbox"/>	<input type="button" value="View"/>	-34.7	Download
Model 6	<input type="checkbox"/>	<input type="button" value="View"/>	-34.6	Download
Model 7	<input type="checkbox"/>	<input type="button" value="View"/>	-34.3	Download
Model 8	<input type="checkbox"/>	<input type="button" value="View"/>	-33.6	Download
Model 9	<input type="checkbox"/>	<input type="button" value="View"/>	-33.6	Download
Model 10	<input type="checkbox"/>	<input type="button" value="View"/>	-33.5	Download

Download

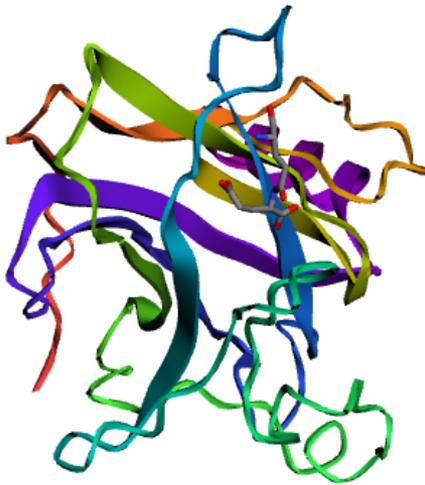
Click [Here](#) to download all the results.

Results

(ensemble docking)

Top 10 predicted binding modes (multiple receptor structures could be used)

Click the "View" button on the right panel to display the corresponding binding mode.



Models	Active	View	PDB	ITScore
Model 1		<input type="button" value="View"/>	3zmeA	-55.8
Model 2		<input type="button" value="View"/>	5aokA	-52.1
Model 3		<input type="button" value="View"/>	5ab9A	-51.6
Model 4		<input type="button" value="View"/>	5aojA	-50.5
Model 5		<input type="button" value="View"/>	5aojA	-49.0
Model 6		<input type="button" value="View"/>	5g4nA	-48.8
Model 7		<input type="button" value="View"/>	3zmeA	-48.8
Model 8		<input type="button" value="View"/>	5abaA	-48.0
Model 9		<input type="button" value="View"/>	5aokA	-47.6
Model 10		<input type="button" value="View"/>	5aokA	-47.5

Download

Click [Here](#) to download all the results.

Conclusion

- We developed an automated strategy using the information from the known protein-ligand complex structures to improve binding mode prediction.
- A ligand similarity calculation method was employed to search for the closest receptor structure with a bound ligand that shares high similarity with the query ligand for binding mode prediction.
- The methods (template-based, selective docking, ensemble docking) have been implemented in a fully automated web server, named as AutoPLI, for the prediction of protein-ligand complex structures.
- Future improvement: Entropy; receptor flexibility

THANKS FOR YOU ATTENTION!

Current Lab Members:

D3R Team

Rui Duan, Ph.D.

Benjamin Merideth

Liming Qiu, Ph.D.

Xianjin Xu, Ph.D.

Zhiwei Ma

Marshal Huang

OpenEye Scientific Software

Supported by:

National Institutes of Health [R01GM109980 & R01HL126774]

NSF CAREER Award [DBI-0953839]

American Heart Association (Midwest Affiliate) [13GRNT16990076]



Zou lab: <http://zoulab.dalton.missouri.edu/>