Electronic Supplementary Material (ESI) for Physical Chemistry Chemical Physics.

Supporting Information

Reinforcement Learning to Boost Molecular Docking upon Protein Conformational Ensemble

Bin Chong[1], Yingguang Yang[2], Zi-Le Wang[3], Han Xing[1], and Zhirong Liu[1,*]

[1] College of Chemistry and Molecular Engineering, and Beijing National Laboratory for Molecular Sciences (BNLMS)
Peking University
Beijing 100871 (China)
[2] School of Cyberscience
University of Science and Technology of China
Hefei 230026 (China)
[3] State Key Laboratory of Low Dimensional Quantum Physics, Department of Physics
Tsinghua University
Beijing 100084 (China)

Correspondence and requests for materials should be addressed to Z.L.(e-mail: LiuZhiRong@pku.edu.cn).

Calculation Procedures

Binding site prediction. We used the code CAVITY (Y. X. Yuan, J. F. Pei, L. H. Lai, *Curr. Pharm. Design* **2013**, *19*, 2326-2333) to determine the potential binding sites of each conformation for molecular docking. In case of multiple sites for one conformation, the pocket with the maximal potential binding affinity was chosen for docking.

Molecular docking. We used the open-source program Vina (O. Trott, A. J. Olson, *J. Comput. Chem.* **2010**, *31*, 455-461) for molecular docking. Default parameters were chosen, e.g., the grid size is 22×22×22, while the energy_range, exhaustiveness and num_modes are set to be 3, 8 and 8, respectively. The grid center is set to be the geometric center of the conformational cavities obtained from CAVITY.The docking score given by Vina is the binding free energy $\Delta G_{i,j}$ (in a unit of kcal/mol) between the conformation i and the ligand j. Open Babel (N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, G. R. Hutchison, *J. Cheminformatics* **2011**, *3*, 33.) was used to preprocess the file format of proteins and ligands.

Conformational ensemble of IDPs. The conformational ensemble of the oncoprotein c-Myc was borrowed from a previous study of large-scale MD simulations (F. Jin, C. Yu, L. H. Lai, Z. R. Liu, *PLoS Comput. Biol.* **2013**, *9*, e1003249), giving 16716 conformations. More specifically, Hammoudeh *et al.* (D. I. Hammoudeh, A. V. Follis, E. V. Prochownik, S. J. Metallo, *J. Am. Chem. Soc.* **2009**, *131*, 7390-7401.) measured the chemical shifts and several NOE signal of c-$Myc_{370\text{-}409}$ with and without small molecules by NMR, and predicted the average dihedral angles of the main chain; Jin *et al.* utilized these angles to construct and optimized the single apo and holo structures which were further used as initial conformation in MD simulations; Jin *et al.* then conducted large-scale (with a total simulation time of 34.5 μs) replica-exchange MD using AMBER software and AMBER99SB force field, where the ionic strength of the system is set to 0.2 M.

Ligand library. The positive and negative control ligands (10074-A4 and AJ292) as well as 6 ligand molecules (YC-1101, YC-1201~YC-1205) obtained from virtual screening by Yu *et al.* (C. Yu, X. Niu, F. Jin, Z. Liu, C. Jin, L. Lai, *Sci. Rep.* **2016**, *6*, 22298) were selected in priority. Additional 275 ligands were selected from the library of SPECS, a worldwide provider of compound management services besides being a main supplier of screening compounds in drug discovery. Specifically, small compound set with amount of more than 50 mg, which contains ~140,000 compounds, were selected from the SPECS libraries. Based on the calculated fingerprint of each compound, Leader-Follower clustering and K-Mean clustering were conducted in sequence. Finally, the 275 target compounds were obtained. In total, 283 ligands were used in our study.

Real dataset-I. Each of the 16716 conformations from the ensemble of c-Myc was docked with each of the 283 ligands above, and $K_{a,j}^{(app)}$ for each ligand is calculated with Eq.(2) of the main text. The distribution properties were analysed detailedly before (B. Chong, Y. G. Yang, C. G. Zhou, Q. J. Huang, and Z. R. Liu, *J. Chem Inf. Model.,* 2020, **60**, 4967).

Synthetic datasets. The ensemble distribution of $\Delta G_{i,j}$ for each ligand j is approximately described by a Gaussian distribution with a mean μ_j and a variance σ_j^2, i.e., $p(\Delta G_{i,j}) = N(\Delta G_{i,j}|\mu_j,\sigma_j^2)$. The distributions of μ_j and σ_j for ligands of the synthetic datasets are also assumed to be Gaussian, whose parameters were inferred from the real dataset-I with 283 ligands to give $p(\mu_j) = N(\mu_j| -5.1,0.65^2)$ and $p(\sigma_j) = N(\sigma_j|0.44,0.08^2)$ (in units of kcal/mol). The apparent association constant is calculated by

$lnK_{a,j}^{(app)} = -\frac{1}{RT}\left(\mu_j - \frac{\sigma_j^2}{2RT}\right)$ according to the reference (B. Chong, Y. G. Yang, C. G. Zhou, Z. R. Liu, *J. Chem Inf. Model.*, 2020, **60**, 4967).

Real dataset-II. To provide a validation set, we selected additional 828 ligands randomly (which are different from those in real dataset-I) from the library of SPECS and docked them with all 16716 conformations from the ensemble of c-Myc. The rusults were used as the dataset to validate the effects of the rUCB algorithm.

Large ligand library. A large library with 28479 ligands (including dataset-I and dataset-II) from SPECS was constructed. The rUCB algorithm was used to screen this large library to target c-Myc.

Details of the rUCB algorithm

The aim of the rUCB algorithm is to predict (pick out) top m ligands with the largest $K_{a,j}^{(app)}$ from a library with N ligands, by scheduling docking process with reinforcement learning approach, within a limited number of docking times T. Unless otherwise stated, $T = 2N$ is adopted in this study. The rUCB algorithm is composed of the following steps:

Initialization: Dock each ligand once, with IDP conformation randomly selected from the ensemble. The total number of docking times in this stage is N.

Loop ($t = N + 1, T$): Choose one ligand to dock once (with IDP conformation randomly selected from the ensemble) by

— Pre-choose m ligands that maximize an indexed function as

$$\arg\max_j \left[logK_{a,j}^{(app)}(t) + c\sigma_{logK,j}(t)\sqrt{\frac{1}{n_j(t)}} \right], \tag{S1}$$

We adopt $Q_j = logK_{a,j}^{(app)}$ with a logarithm operation to avoid the improper influence of the fact that $K_{a,j}^{(app)}$ spans a few orders of magnitude. $n_j(t)$ is the number of times that the ligand-j has been docked prior to t. $logK_{a,j}^{(app)}(t)$ is the estimate of $logK_{a,j}^{(app)}$ based on the prior docking results, while $\sigma_{logK,j}(t)$ is the estimated standard deviation of $logK_{a,i,j}$. For $n_j(t) = 1$, a straightforward determination of $\sigma_{logK,j}(t)$ is not available, then a default value of 0.35 inferred from the real docking dataset with 283 ligand is used. For in t a range of $[N + 1{:}T]$ with $T = 2N$ and large N, the range of variation of the lnt term in usual UCB algorithm is very small, so we discard the lnt term. An advantage of this simplification is that we need not update the index function for all ligands at each step t, i.e., we need only update the index value of the ligand that was docked in the latest round and its ranking.

— Among m pre-chosen ligands, choose the one that minimize another indexed function as

$$\arg\min_j \left[logK_{a,j}^{(app)}(t) - c\sigma_{logK,j}(t)\sqrt{\frac{1}{n_j(t)}} \right], \tag{S2}$$

and dock it once (with IDP conformation randomly selected from the ensemble).

Final prediction: Pick m ligands that maximize $logK_{a,j}^{(app)}(T)$ as the predicted top ligands from virtual screening.

Example codes are provided as the attached iDockLearn.rar.

Results and Discussion

An Oversimplified Analysis on the docking times

By ignoring the difference of $\sigma_{logK,j}$ for ligands, the indexed function in selecting ligands in Eq. (S1) becomes

$$logK_{a,j}^{(app)}(t) + c' \sqrt{\frac{1}{n_j(t)}}. \tag{S3}$$

When the total docking number is large, the indexed function reaches a constant (denoted as $logK_a^{(0)}$), so it yields

$$logK_{a,j}^{(app)} + c' \sqrt{\frac{1}{n_j}} = logK_a^{(0)}. \tag{S4}$$

So

$$n_j = \frac{c'^2}{\left(logK_{a,j}^{(app)} - logK_a^{(0)}\right)^2}. \tag{S5}$$

It diverges when $logK_{a,j}^{(app)} = logK_a^{(0)}$. In rUCB algorithm, $logK_{a,j}^{(app)}$ is estimated from the prior docking results, which introducing fluctuation. So we add an extra term to reflect such an effect:

$$n_j = \frac{c'^2}{\left(logK_{a,j}^{(app)} - logK_a^{(0)}\right)^2 + \gamma^2}, \tag{S6}$$

which is an Cauchy-Lorentz function.

Analyses on Eq. (S2) gives similar result.

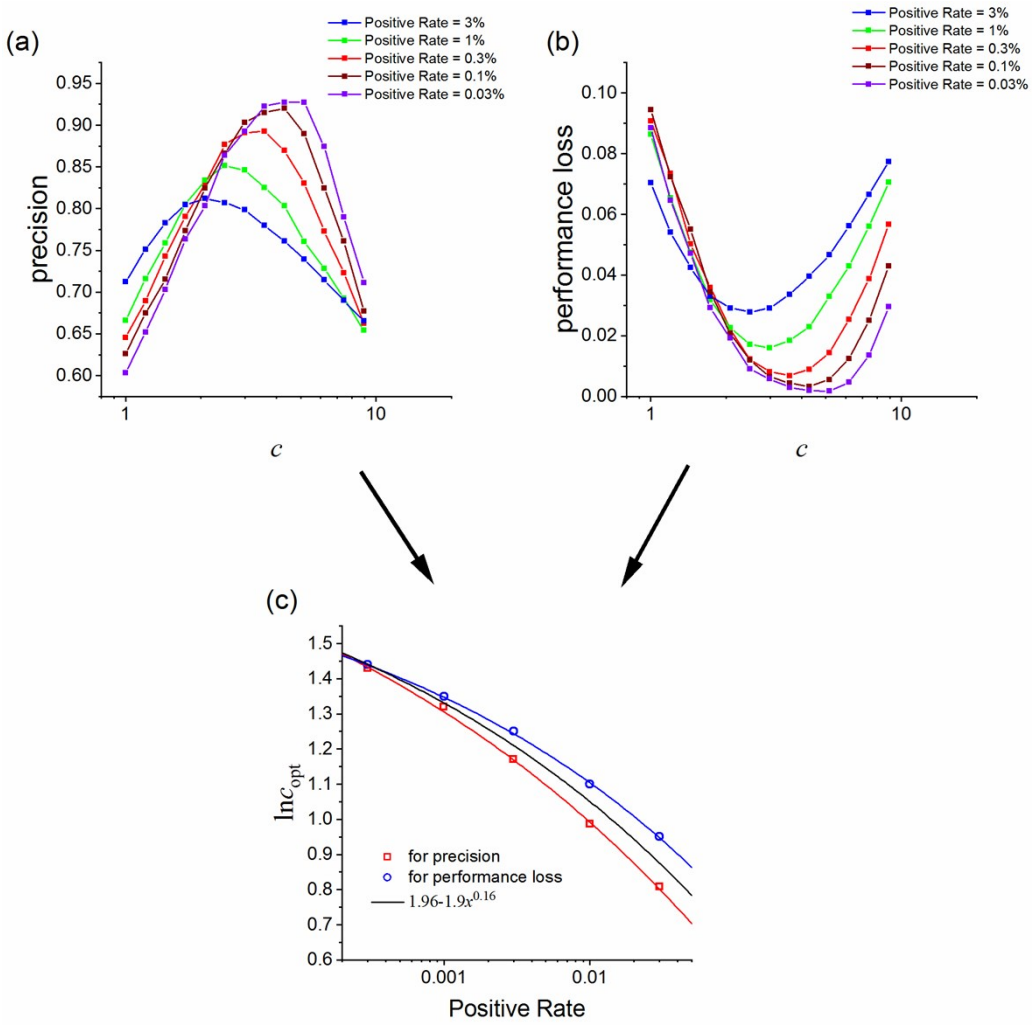

Figure S1. Influence of the learning parameter c for the reversible UCB algorithm. Synthetic dataset with $N = 10^4$ is used. (a) The precision as a function of c. (b) The performance loss as a function of c. (c) The optimal c as a function of the positive rate, where red squares were derived from (a) to achieve the maximal precision, and green circles were derived from (b) to achieve the minimal performance loss, while the black lines represents a compromise as $c = e^{1.96 - 1.9x^{0.16}}$, where x is the positive rate defined as m/N. The total docking number in each run is two times the ligand number, i.e., $T = 2N$. Each data point in (a) and (b) was averaged from 200 runs of simulation with different random seeds.

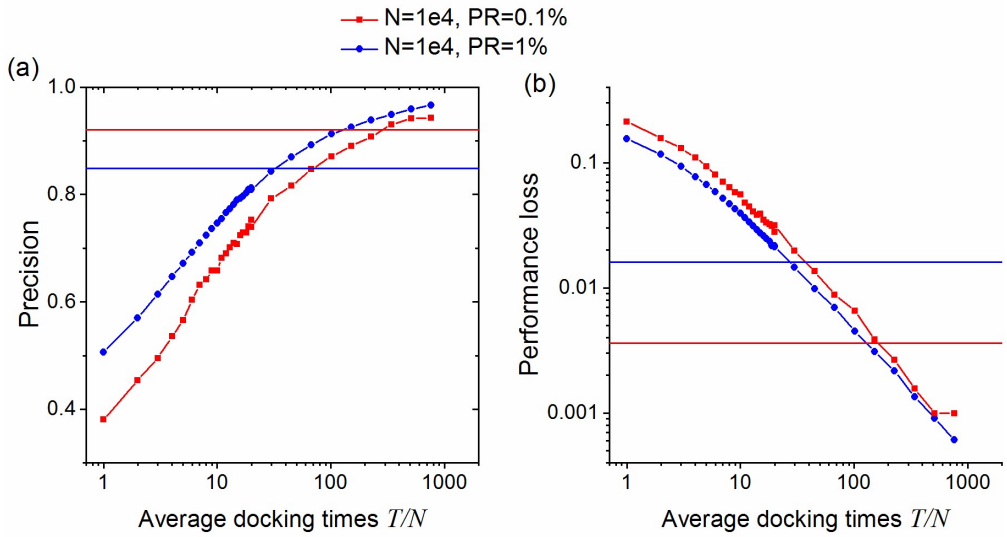

Figure S2. Effects of the uniform algorithm (curves with symbols) as functions of the average docking times per ligand, T/N. Synthetic dataset with $N = 10^4$ is used. The positive rate is 0.1% (curves with squares) or 1% (curves with circles). The rUCB results with $T/N = 2$ were shown as horizontal lines for comparison. Each data point was averaged from 200 runs of simulation.

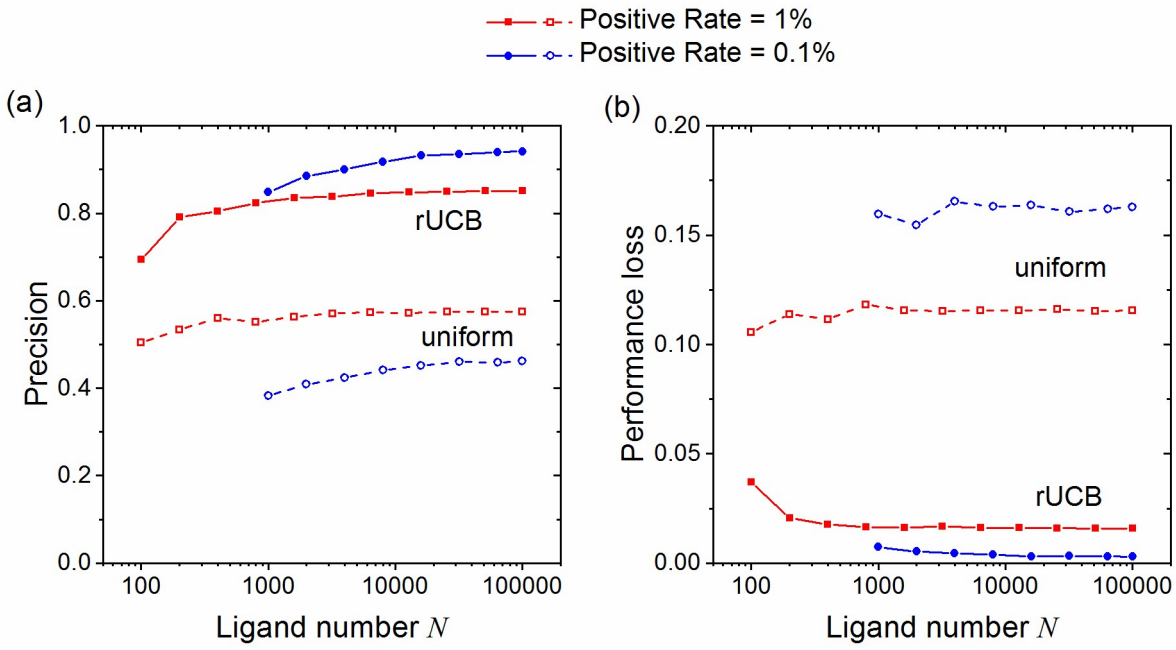

Figure S3. Effects of the rUCB algorithm (filled symbols) in comparison with those of the uniform algorithm (opened symbols) as functions of the ligand number N. Synthetic datasets with a positive rate of 1% (squares) and 0.1% (circles) were used. The average docking times per ligand is 2, i.e., $T/N = 2$. Each data point was averaged from 500 runs of simulation.

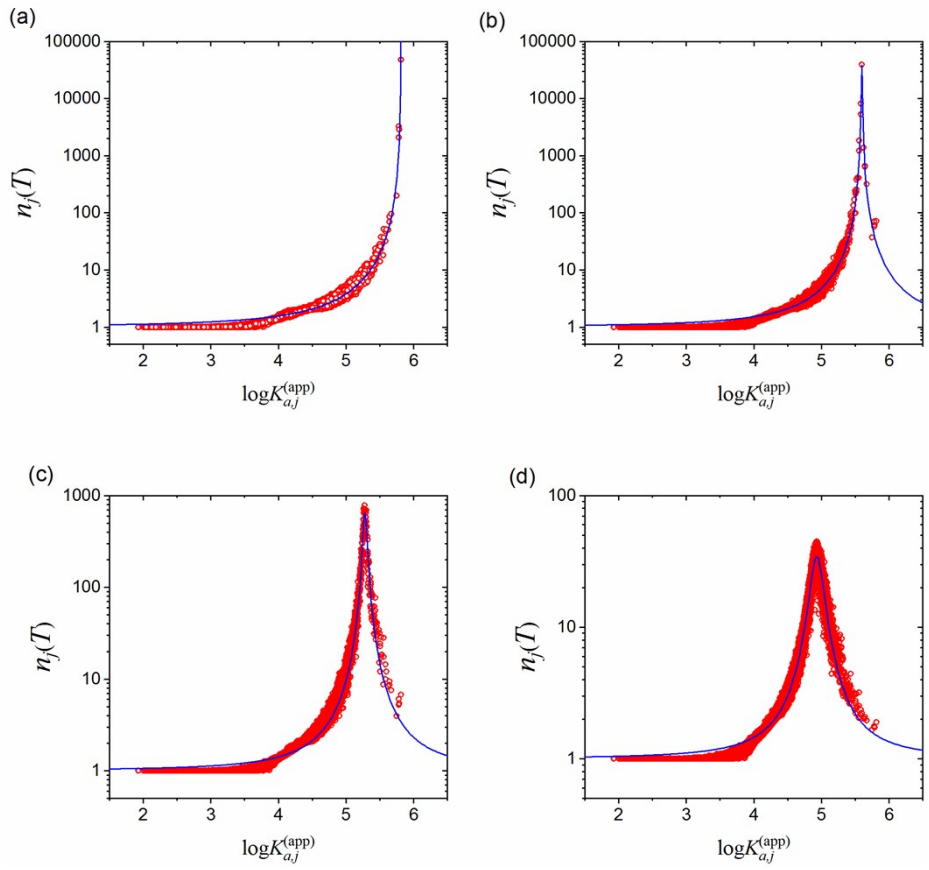

Figure S4. Distribution of docking times under rUCB for synthetic dataset with $N = 10^5$ and (a-d) $m = 1, 10, 100, 1000$, respectively, to give a positive rate of 0.001%, 0.01%, 0.1%, 1%. $T/N = 2$. Results from 400 simulations were averaged for each panel, where the blue line is the fitted Cauchy-Lorentz function.

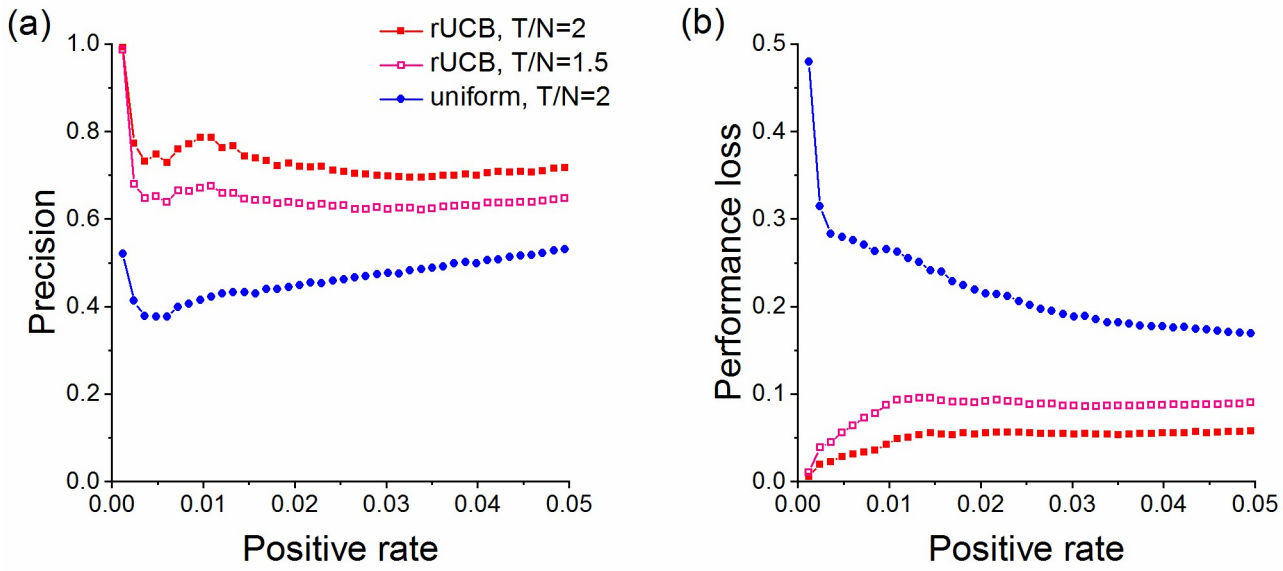

Figure S5. Validation of rUCB on the real dataset-II with 828 ligands. For each data point, 2000 runs of simulation were conducted to get average.

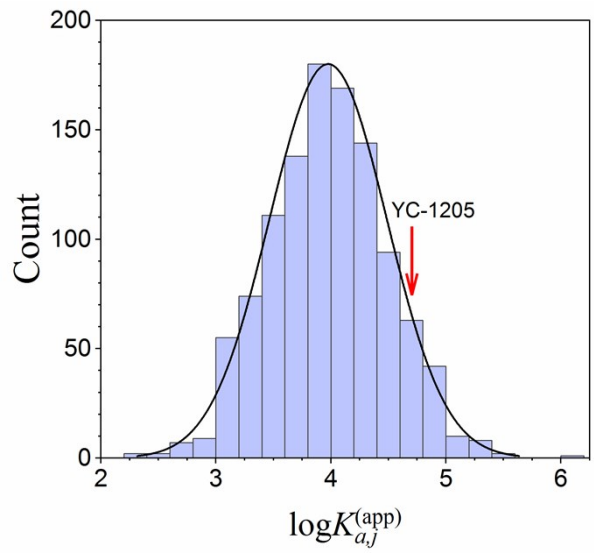

Figure S6. Distribution of $logK_{a,j}^{(app)}$ for 1111 ligands as a combination of Real dataset-I and Real dataset-II. The location of YC-1205 with $logK_{a,j}^{(app)} = 4.70$ was indicated by an arrow.