**Supplementary data:**

**Enrichments of Ensemble Docking Strategy Based on the Bayesian Model**

Yu Leia,b, Sheng Guoa,b, Yi Liua,b\*, Zhili Zuoa,b\*

a State Key Laboratory of Phytochemistry and Plant Resources in West China, Kunming Institute of Botany, Chinese Academy of Sciences, Kunming, Yunnan, 650201, China

b School of Chemical Engineering, Sichuan University of Science & Engineering, 180 Xueyuan Street, Huixing Road, Zigong, Sichuan, 643000, China

\* Corresponding author: Zhili Zuo, E-mail addresses: [zuozhili@mail.kib.ac.cn](mailto:zuozhili@mail.kib.ac.cn) Yi Liu, E-mail addresses: yiliu@suse.edu.cn

1. Supplementary Methods

**Docking Simulations**. GOLD 2020.1 was employed to conduct the docking calculations. Before molecular docking, the preparation of the representative protein structures for each target was carried out using Discovery Studio 4.0 by removing water, adding hydrogen atoms, loading AMBER99SB force field. The structure was optimized by using SYBYL-X software (version 2.0). The binding site was defined by the location of co-crystallized ligand in the structure and all atoms are within 6 Å. The genetic algorithm (GA) runs was set to 25 times. The rest parameters were set at default values.

**MD Simulations.** The MD simulations were performed with GROMACS 5.0.2 software(2015; Jonathan., et al., 2012)(2015; Jonathan., et al., 2012)(2015; Jonathan., et al., 2012). The missing residues and its side chains were added. Then, the topology file of the small molecule was generated using AMBER program with the general AMBER force filed (GAFF). The protein was loaded with AMBER99SB force field, the topology file was obtained likewise. The topology file of complex was created through a combination of *gro* format files of small molecule and protein. A dodecahedron periodic box was defined and each system was solvated with a simple point charge (spc216) water model. Initially, each system was minimized using the steepest descent algorithm. The particle mesh Ewald (PME) was used for calculating long-range electrostatic. Then, in equilibrated stage, each system was gradually heated from 0 to 300 K over 100 ps under NVT conditions and then equilibrated under NPT conditions for 1000 ps at 300 K. Finally, each system was simulated for 40 ns and the MD time step was 2 fs. Coordinate trajectories were saved every 10 ps for the whole MD runs. For each system, there are 4000 snapshots in each trajectory and named as PDB ID\_n (n=1, 2, 3, …, 4000).

**Bayesian Model.** The Bayesian categorization model was built to distinguish active ligands from baseline ligands with numeric, fingerprint or textual properties employing as the evaluation criteria. With this method, multiple numeric properties such as docking scores used in this study could be integrated with efficacy to discriminate the actives from decoys. A five-fold cross validation is performed to optimize the Bayesian Model, furthermore, the developed Bayesian Model could be validated by an external test set and these reduplicative validations make it a more credible model.

As a simple probabilistic classification model, Bayesian categorization model is based on Bayes' theorem:

where

h is the hypothesis or model

d is the observed data

p(h) is the prior belief (probability of hypothesis h before observing any data)

p(d) is the data evidence (marginal probability of the data)

p(d|h) is the likelihood (probability of data d if hypothesis h is true)

p(h|d) is the posterior probability (probability of hypothesis h being true given the observed data d)

Bayesian statistics not only considers the likelihood of a model, it also takes into consideration the complexity of the model. As a result, it automatically picks the simplest model that can explain the observed data. In the implementation of the Bayesian modeling in this application, the learned models are created with a learn-by-example paradigm: the user marks the sample data that is of interest (good), and then the system learns to distinguish them from background data. No tuning parameters are required beyond the selection of the input descriptors from which to learn. The learning process generates a large set of Boolean features from the input descriptors. It then collects the frequency of occurrence of each feature in the good subset and in all data samples. To apply the model to a particular sample, the features of the sample are generated, and a weight is calculated for each feature using a Laplacian-adjusted probability estimate. The weights are summed to provide a probability estimate, which is a relative predictor of the likelihood of that sample being from the good subset.

The Bayesian Model of each ensemble was established by using the *Create Bayesian Model* module in DS 4.0 by integrating the docking scores of each representative independent docking runs. Firstly, the *Create Bayesian Model* module is setting up and running by employing *Create QSAR Model* tools and click *Create Bayesian Model* to open the Create Bayesian Model dialog. The training ligands used for the calculation and test ligands used to test the performance of the model in the external validation step were input. The simulated docking scores of the input ligands were set as User Properties and Calculable Properties were set as null. Further, we have checked the Cross Validation to perform cross validation on the model and the number of groups was set as five. All other parameters were kept as default values.

1. Supplementary Figures and Tables
   1. Structural Clustering

Table S1. Structural clustering type, PDB ID and resolution for the crystal structures of the BTK

|  |  |  |
| --- | --- | --- |
| **Cluster** | **PDB ID** | **Resolution(Å)** |
| **Type Ⅰ** | 3pj2 | 1.75 |
| 3pj1 | 2 |
| 3pj3 | 1.85 |
| 3piz | 2.21 |
| 3piy | 2.55 |
| **6o8i** | **1.42** |
| 3pix | 1.85 |
| **Type Ⅱ** | 6s90 | 1.82 |
| 5zz4 | 2.90 |
| 5t18 | 1.5 |
| 5jrs | 1.97 |
| 4nwm | 2.03 |
| 6bkw | 1.5 |
| 6bkh | 1.79 |
| 6bik | 1.9 |
| 6aua | 1.66 |
| **4rx5** | **1.36** |
| 5vgo | 1.62 |
| 5kup | 1.39 |
| 4otf | 1.95 |
| 6aub | 1.65 |
| 3ocs | 1.8 |
| **Type Ⅲ** | 5fbn | 1.8 |
| 5p9g | 1.75 |
| 5vfi | 1.59 |
| 5p9f | 1.71 |
| 4ot6 | 2.05 |
| 4rg0 | 2.5 |
| 4otr | 1.95 |
| 4rfy | 1.7 |
| 4otq | 1.55 |
| **4rfz** | **1.17** |
| 4ot5 | 1.55 |
| **Type Ⅳ** | 6nzm | 1.72 |
| 3gen | 1.6 |
| 5bq0 | 1.57 |
| **5p9k** | **1.28** |
| 4zly | 1.65 |
| 4zlz | 2 |
| 4z3v | 1.6 |

The PDB ID of the bold font is selected as the representative crystal structures.

Table S2. Structural clustering type, PDB ID and resolution for the crystal structures of the JAK

|  |  |  |
| --- | --- | --- |
| **Cluster** | **PDB ID** | **Resolution(Å)** |
| **Type Ⅰ** | 6n79 | 2.27 |
| 6dbn | 2.48 |
| **6bbu** | **2.08** |
| **Type Ⅱ** | 6n78 | 1.83 |
| 6n7d | 1.78 |
| 6n7b | 1.81 |
| 6n77 | 1.64 |
| 6n7c | 1.69 |
| **6n7a** | **1.33** |
| 6ggh | 1.7 |
| **Type Ⅲ** | 4k6z | 2.73 |
| 6smb | 2.04 |
| **6aah** | **1.83** |
| 6sm8 | 1.85 |
| 5e1e | 2.3 |
| 5wo4 | 1.84 |
| 4i5c | 2.1 |
| 4k77 | 2.4 |
| 4ivd | 1.93 |
| 4ei4 | 2.22 |
| **Type Ⅳ** | 4fk6 | 2.2 |
| **4ehz** | **2.17** |

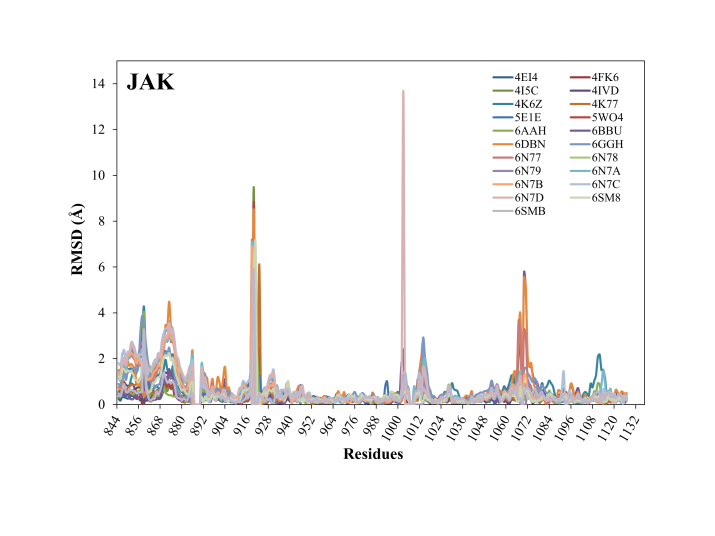
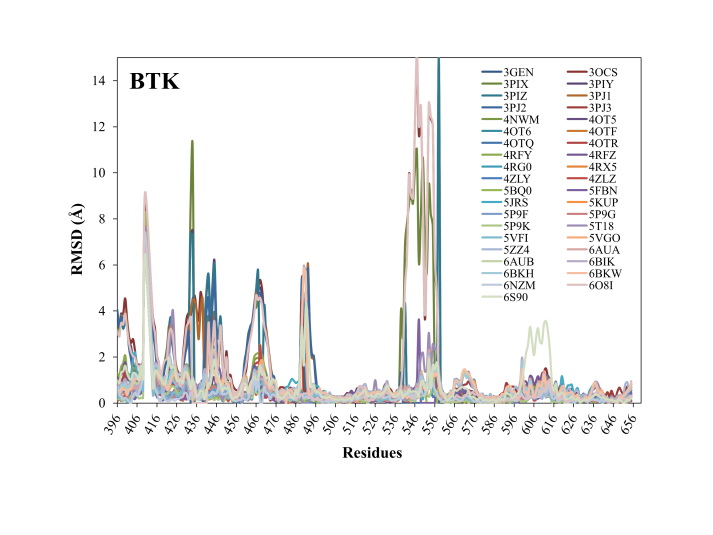
The PDB ID of the bold font is selected as the representative crystal structures.

Table S3. Structural clustering type, PDB ID and resolution for the crystal structures of the PARP

|  |  |  |
| --- | --- | --- |
| **Cluster** | **PDB ID** | **Resolution(Å)** |
| **Type Ⅰ** | 5ha9 | 4.01 |
| **4zzz** | **1.9** |
| 1uk1 | 3 |
| 1uk0 | 3 |
| **Type Ⅱ** | 5wtc | 2.2 |
| 5ws0 | 2.6 |
| **5ws1** | **1.9** |
| 5wrz | 2.2 |
| 5wry | 2.3 |
| 4l6s | 2.2 |
| **Type Ⅲ** | 4uxb | 3.22 |
| 6ghk | 2.28 |
| 4r5w | 2.84 |
| **6i8t** | **2.1** |
| 6i8m | 2.1 |
| 4r6e | 2.2 |
| 5xsu | 2.4 |
| 4hhz | 2.71 |
| 5a00 | 2.75 |
| 4rv6 | 3.19 |
| 4gv7 | 2.5 |
| 3l3m | 2.5 |
| 3l3l | 2.5 |
| 3gn7 | 2.5 |
| 3gjw | 2.3 |
| 2rcw | 2.8 |
| 1wok | 3 |
| **Type Ⅳ** | 4und | 2.2 |
| 5xst | 2.3 |
| 5xsr | 2.3 |
| 4pjt | 2.35 |
| 4hhy | 2.36 |
| 6nri | 2.2 |
| 6nrj | 1.65 |
| **6nrh** | **1.5** |
| 6nrg | 1.7 |
| 6nrf | 2 |
| 6bhv | 2.3 |

The PDB ID of the bold font is selected as the representative crystal structures.

The scatterplots of RMSD values for crystal structures versus residues were shown in Figure S1. The RMSD values versus residues for BTK fluctuate greatly between amino acids 536-556 (1Å-15Å), and there are several great deviations between 406 and 496 as well (around 6Å), but the RMSD values of key residues are generally below 1Å which indicate that the protein structures of each cluster may have well-matched performance on docking calculations. For JAK, the residues around 856, 874, 920, 1004 and 1072 have shown quite inferior performance in RMSD with the lowest RMSD value (4Å). While the residues (Glu957, Leu959, Asn1008) making key interactions with inhibitors and its vicinity exhibit extremely low RMSD values below 1Å which also means similar performance on docking calculations. For PARP, the residues’ RMSD values are around 1Å; only a few positions deviated fiercely. This uneven deviation suggests a significantly different performance on docking calculations.



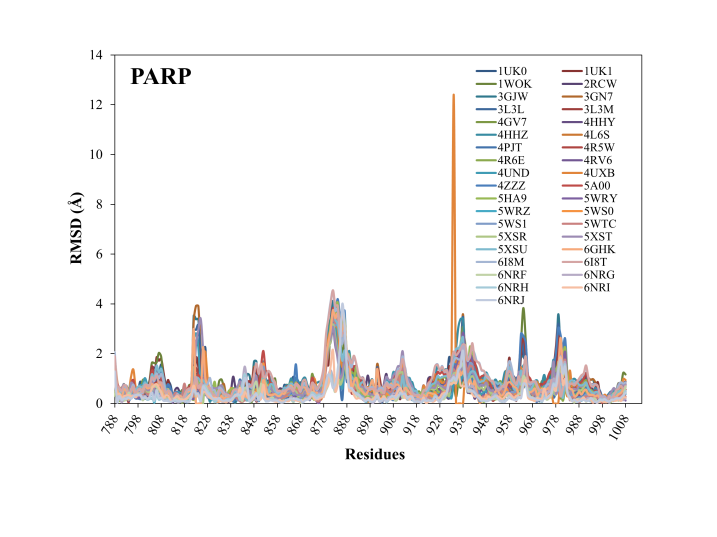


Figure S1. RMSD per residue of the crystal structures for three targets. 4Z3V, 4EHZ and 6BHV serve as the reference structures of BTK, JAK and PARP, respectively.

* 1. Docking Calculations

Table S4. The identification of scoring function in docking program (GOLD) applied to docking simulations

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **BTK** | | | | | |
| **Scoring functions** | | **PDB ID** | | | |
| 6O8I | 4RX5 | 4RFZ | 5P9K |
| CHEMPLP | Score | 73.32 | 77.21 | 105.63 | 85.71 |
| RMSD(Å) | 1.6808 | 0.8877 | 0.8824 | 1.6393 |
| GoldScore | Score | 62.10 | 78.89 | 103.02 | 64.09 |
| RMSD(Å) | 1.6561 | 1.0211 | 0.6289 | 5.4334 |
| ChemScore | Score | 33.53 | 35.68 | 34.73 | 29.79 |
| RMSD(Å) | 1.6681 | 8.0292 | 4.1673 | 3.5949 |
| ASP | Score | 29.82 | 42.25 | 33.30 | 40.82 |
| RMSD(Å) | 1.0889 | 0.8749 | 6.1873 | 3.7713 |
| **JAK** | | | | | |
| **Scoring functions** | | **PDB ID** | | | |
| 6BBU | 6N7A | 6AAH | 4EHZ |
| CHEMPLP | Score | 70.34 | 76.40 | 73.91 | 63.20 |
| RMSD(Å) | 0.7786 | 0.3008 | 0.5392 | 0.5708 |
| GoldScore | Score | 61.07 | 66.89 | 65.52 | 57.37 |
| RMSD(Å) | 0.9370 | 0.53467 | 0.4248 | 0.6064 |
| ChemScore | Score | 23.94 | 26.76 | 28.97 | 28.12 |
| RMSD(Å) | 1.4776 | 1.9800 | 4.7367 | 0.8783 |
| ASP | Score | 31.02 | 31.44 | 31.10 | 28.29 |
| RMSD(Å) | 1.2172 | 6.5891 | 0.7981 | 0.6296 |
| **PARP** | | | | | |
| **Scoring functions** | | **PDB ID** | | | |
| 4ZZZ | 5WS1 | 6I8T | 6NRH |
| CHEMPLP | Score | 73.94 | 70.56 | 80.55 | 89.41 |
| RMSD(Å) | 2.0444 | 2.3612 | 0.9136 | 2.3873 |
| GoldScore | Score | 55.26 | 54.94 | 63.23 | 73.78 |
| RMSD(Å) | 2.4966 | 2.6435 | 0.6697 | 0.5923 |
| ChemScore | Score | 27.61 | 32.64 | 35.89 | 40.57 |
| RMSD(Å) | 2.0446 | 2.2064 | 0.6956 | 3.9138 |
| ASP | Score | 42.34 | 45.65 | 45.61 | 55.88 |
| RMSD(Å) | 1.6429 | 1.7098 | 0.6877 | 4.0107 |

* 1. The Correlation Analysis of Docking Performance Based on Crystal Structures and Structures from MD Simulations

It is observed from Table S5 that the Pearson correlation coefficients for BTK are generally higher than JAK and PARP with all correlation coefficients greater than 0.8 and one pair of structures above 0.9 (0.923 of 5P9K\_2526 and 5P9K\_2629). The protein structure with inferior AUC value is discarded. The Pearson correlation coefficients for JAK are high between 0.7-0.9, likewise with one pair of structures (6N7A\_3398 and 6N7A\_2681) which Pearson correlation coefficient is 0.901. For PARP, as shown in Table 3, the correlation coefficients are distributing widespread from 0.5 to 0.9. Four pairs of structures: 6I8T\_3041 and 6I8T\_3906, 6I8T\_3041 and 6I8T\_2072, 6I8T\_3041 and 6I8T\_3274, 6I8T\_3906 and 6I8T\_3274, give the quite high r values (＞0.9). The least two r values belong to the structure 6I8T with 4ZZZ and 6I8T with 5WS1.

The Spearman ranking correlation coefficients (ρ) were illustrated in Table S6. Similar to the Pearson correlation coefficients, high ranking correlation is observed in BTK with almost all correlation coefficients greater than 0.8 means the strong similarity of rankings of the docking scores for each pair structures in BTK. Alike, the ρ values in JAK show the good ranking correlation. Compared with BTK and JAK, the Spearman ranking correlation in PARP accommodates two degrees of relevance including moderate correlation and high correlation.

Table S5. Pearson Correlation Coefficients (r) of the Docking Scores of the Known Inhibitors and Non-inhibitors Based on the Representative Crystal Structures and MD Simulations of Three Targets.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **BTK** | | | | | | | | |
| Pearson/r | 4RFZ | 6O8I | 4RX5 | 5P9K | 5P9K\_3768 | 5P9K\_2526 | 5P9K\_2992 | 5P9K\_2629 |
| 4RFZ | 1.000 | 0.859\*\* | 0.867\*\* | 0.855\*\* | 0.831\*\* | 0.835\*\* | 0.850\*\* | 0.845\*\* |
| 6O8I | 0.859\*\* | 1.000 | 0.884\*\* | 0.894\*\* | 0.876\*\* | 0.863\*\* | 0.885\*\* | 0.875\*\* |
| 4RX5 | 0.867\*\* | 0.884\*\* | 1.000 | 0.853\*\* | 0.860\*\* | 0.863\*\* | 0.874\*\* | 0.873\*\* |
| 5P9K | 0.855\*\* | 0.894\*\* | 0.853\*\* | 1.000 | 0.837\*\* | 0.852\*\* | 0.857\*\* | 0.861\*\* |
| 5P9K\_3768 | 0.831\*\* | 0.876\*\* | 0.860\*\* | 0.837\*\* | 1.000 | 0.862\*\* | 0.878\*\* | 0.870\*\* |
| 5P9K\_2526 | 0.835\*\* | 0.863\*\* | 0.863\*\* | 0.852\*\* | 0.862\*\* | 1.000 | 0.879\*\* | 0.923\*\* |
| 5P9K\_2992 | 0.850\*\* | 0.885\*\* | 0.874\*\* | 0.857\*\* | 0.878\*\* | 0.879\*\* | 1.000 | 0.884\*\* |
| 5P9K\_2629 | 0.845\*\* | 0.875\*\* | 0.873\*\* | 0.861\*\* | 0.870\*\* | 0.923\*\* | 0.884\*\* | 1.000 |
| **JAK** | | | | | | | | |
| Pearson/r | 6BBU | 6N7A | 6AAH | 4EHZ | 6N7A\_3398 | 6N7A\_2014 | 6N7A\_2681 | 6N7A\_2010 |
| 6BBU | 1.000 | 0.837\*\* | 0.768\*\* | 0.885\*\* | 0.805\*\* | 0.802\*\* | 0.779\*\* | 0.777\*\* |
| 6N7A | 0.837\*\* | 1.000 | 0.764\*\* | 0.868\*\* | 0.860\*\* | 0.851\*\* | 0.838\*\* | 0.816\*\* |
| 6AAH | 0.768\*\* | 0.764\*\* | 1.000 | 0.810\*\* | 0.747\*\* | 0.747\*\* | 0.716\*\* | 0.740\*\* |
| 4EHZ | 0.885\*\* | 0.868\*\* | 0.810\*\* | 1.000 | 0.828\*\* | 0.822\*\* | 0.798\*\* | 0.802\*\* |
| 6N7A\_3398 | 0.805\*\* | 0.860\*\* | 0.747\*\* | 0.828\*\* | 1.000 | 0.896\*\* | 0.901\*\* | 0.862\*\* |
| 6N7A\_2014 | 0.802\*\* | 0.851\*\* | 0.747\*\* | 0.822\*\* | 0.896\*\* | 1.000 | 0.870\*\* | 0.860\*\* |
| 6N7A\_2681 | 0.779\*\* | 0.838\*\* | 0.716\*\* | 0.798\*\* | 0.901\*\* | 0.870\*\* | 1.000 | 0.822\*\* |
| 6N7A\_2010 | 0.777\*\* | 0.816\*\* | 0.740\*\* | 0.802\*\* | 0.862\*\* | 0.860\*\* | 0.822\*\* | 1.000 |
| **PARP** | | | | | | | | |
| Pearson/r | 4ZZZ | 5WS1 | 6I8T | 6NRH | 6I8T\_3041 | 6I8T\_3906 | 6I8T\_2072 | 6I8T\_3274 |
| 4ZZZ | 1.000 | 0.844\*\* | 0.511\*\* | 0.745\*\* | 0.798\*\* | 0.809\*\* | 0.792\*\* | 0.798\*\* |
| 5WS1 | 0.844\*\* | 1.000 | 0.533\*\* | 0.788\*\* | 0.821\*\* | 0.816\*\* | 0.819\*\* | 0.818\*\* |
| 6I8T | 0.511\*\* | 0.533\*\* | 1.000 | 0.543\*\* | 0.568\*\* | 0.580\*\* | 0.564\*\* | 0.568\*\* |
| 6NRH | 0.745\*\* | 0.788\*\* | 0.543\*\* | 1.000 | 0.820\*\* | 0.838\*\* | 0.828\*\* | 0.833\*\* |
| 6I8T\_3041 | 0.798\*\* | 0.821\*\* | 0.568\*\* | 0.820\*\* | 1.000 | 0.914\*\* | 0.910\*\* | 0.917\*\* |
| 6I8T\_3906 | 0.809\*\* | 0.816\*\* | 0.580\*\* | 0.838\*\* | 0.914\*\* | 1.000 | 0.899\*\* | 0.921\*\* |
| 6I8T\_2072 | 0.792\*\* | 0.819\*\* | 0.564\*\* | 0.828\*\* | 0.910\*\* | 0.899\*\* | 1.000 | 0.887\*\* |
| 6I8T\_3274 | 0.798\*\* | 0.818\*\* | 0.568\*\* | 0.833\*\* | 0.917\*\* | 0.921\*\* | 0.887\*\* | 1.000 |

\*\*. Correlation is significant at the 0.01 level (2-tailed).

Table S6. Spearman Ranking Correlation Coefficients (ρ) of the Docking Scores of the Known Inhibitors and Non-inhibitors Based on the Representative Crystal Structures and MD Simulations of Three Targets.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **BTK** | | | | | | | | |
| Spearman/ρ | 4RFZ | 6O8I | 4RX5 | 5P9K | 5P9K\_3768 | 5P9K\_2526 | 5P9K\_2992 | 5P9K\_2629 |
| 4RFZ | 1.000 | 0.827\*\* | 0.831\*\* | 0.829\*\* | 0.798\*\* | 0.804\*\* | 0.810\*\* | 0.810\*\* |
| 6O8I | 0.827\*\* | 1.000 | 0.847\*\* | 0.850\*\* | 0.853\*\* | 0.829\*\* | 0.848\*\* | 0.842\*\* |
| 4RX5 | 0.831\*\* | 0.847\*\* | 1.000 | 0.802\*\* | 0.837\*\* | 0.830\*\* | 0.821\*\* | 0.837\*\* |
| 5P9K | 0.829\*\* | 0.850\*\* | 0.802\*\* | 1.000 | 0.798\*\* | 0.812\*\* | 0.804\*\* | 0.816\*\* |
| 5P9K\_3768 | 0.798\*\* | 0.853\*\* | 0.837\*\* | 0.798\*\* | 1.000 | 0.826\*\* | 0.844\*\* | 0.835\*\* |
| 5P9K\_2526 | 0.804\*\* | 0.829\*\* | 0.830\*\* | 0.812\*\* | 0.826\*\* | 1.000 | 0.837\*\* | 0.891\*\* |
| 5P9K\_2992 | 0.810\*\* | 0.848\*\* | 0.821\*\* | 0.804\*\* | 0.844\*\* | 0.837\*\* | 1.000 | 0.840\*\* |
| 5P9K\_2629 | 0.810\*\* | 0.842\*\* | 0.837\*\* | 0.816\*\* | 0.835\*\* | 0.891\*\* | 0.840\*\* | 1.000 |
| **JAK** | | | | | | | | |
| Spearman/ρ | 6BBU | 6N7A | 6AAH | 4EHZ | 6N7A\_3398 | 6N7A\_2014 | 6N7A\_2681 | 6N7A\_2010 |
| 6BBU | 1.000 | 0.842\*\* | 0.772\*\* | 0.887\*\* | 0.817\*\* | 0.809\*\* | 0.785\*\* | 0.781\*\* |
| 6N7A | 0.842\*\* | 1.000 | 0.769\*\* | 0.867\*\* | 0.847\*\* | 0.839\*\* | 0.813\*\* | 0.805\*\* |
| 6AAH | 0.772\*\* | 0.769\*\* | 1.000 | 0.807\*\* | 0.749\*\* | 0.753\*\* | 0.719\*\* | 0.747\*\* |
| 4EHZ | 0.887\*\* | 0.867\*\* | 0.807\*\* | 1.000 | 0.833\*\* | 0.829\*\* | 0.800\*\* | 0.804\*\* |
| 6N7A\_3398 | 0.817\*\* | 0.847\*\* | 0.749\*\* | 0.833\*\* | 1.000 | 0.894\*\* | 0.892\*\* | 0.858\*\* |
| 6N7A\_2014 | 0.809\*\* | 0.839\*\* | 0.753\*\* | 0.829\*\* | 0.894\*\* | 1.000 | 0.862\*\* | 0.853\*\* |
| 6N7A\_2681 | 0.785\*\* | 0.813\*\* | 0.719\*\* | 0.800\*\* | 0.892\*\* | 0.862\*\* | 1.000 | 0.816\*\* |
| 6N7A\_2010 | 0.781\*\* | 0.805\*\* | 0.747\*\* | 0.804\*\* | 0.858\*\* | 0.853\*\* | 0.816\*\* | 1.000 |
| **PARP** | | | | | | | | |
| Spearman/ρ | 4ZZZ | 5WS1 | 6I8T | 6NRH | 6I8T\_3041 | 6I8T\_3906 | 6I8T\_2072 | 6I8T\_3274 |
| 4ZZZ | 1.000 | 0.824\*\* | 0.518\*\* | 0.725\*\* | 0.772\*\* | 0.788\*\* | 0.765\*\* | 0.775\*\* |
| 5WS1 | 0.824\*\* | 1.000 | 0.547\*\* | 0.774\*\* | 0.805\*\* | 0.801\*\* | 0.804\*\* | 0.805\*\* |
| 6I8T | 0.518\*\* | 0.547\*\* | 1.000 | 0.555\*\* | 0.591\*\* | 0.598\*\* | 0.588\*\* | 0.588\*\* |
| 6NRH | 0.725\*\* | 0.774\*\* | 0.555\*\* | 1.000 | 0.811\*\* | 0.824\*\* | 0.822\*\* | 0.824\*\* |
| 6I8T\_3041 | 0.772\*\* | 0.805\*\* | 0.591\*\* | 0.811\*\* | 1.000 | 0.908\*\* | 0.906\*\* | 0.910\*\* |
| 6I8T\_3906 | 0.788\*\* | 0.801\*\* | 0.598\*\* | 0.824\*\* | 0.908\*\* | 1.000 | 0.895\*\* | 0.915\*\* |
| 6I8T\_2072 | 0.765\*\* | 0.804\*\* | 0.588\*\* | 0.822\*\* | 0.906\*\* | 0.895\*\* | 1.000 | 0.883\*\* |
| 6I8T\_3274 | 0.775\*\* | 0.805\*\* | 0.588\*\* | 0.824\*\* | 0.910\*\* | 0.915\*\* | 0.883\*\* | 1.000 |

\*\*. Correlation is significant at the 0.01 level (2-tailed).

* 1. Performance of Bayesian Models Based on Each Single Structure and Ensemble of Protein Structures

Table S7. The Ensemble Member Details to Each Ensemble Number for Three Targets a

|  |  |  |  |
| --- | --- | --- | --- |
| **Ensemble No.** | **Ensemble Member** | | |
| **BTK** | **JAK** | **PARP** |
| Ensemble 1 | 4RFZ\_5P9K | 6N7A\_4EHZ | 6I8T\_6NRH |
| Ensemble 2 | 2526\_2992 | 2014\_2010 | 3041\_3274 |
| Ensemble 3 | 5P9K\_2992 | 6N7A\_2010 | 6I8T\_3041 |
| Ensemble 4 | 4RFZ\_4RX5\_5P9K | 6BBU\_6N7A\_4EHZ | 5WS1\_6I8T\_6NRH |
| Ensemble 5 | 3768\_2526\_2992 | 3398\_2014\_2010 | 3041\_3906\_3274 |
| Ensemble 6 | 4RFZ\_5P9K\_2992 | 6BBU\_6N7A\_2010 | 6I8T\_6NRH\_3041 |
| Ensemble 7 | 4RFZ\_6O8I \_4RX5\_5P9K | 6BBU\_6N7A\_6AAH\_4EHZ | 4ZZZ\_5WS1\_6I8T\_6NRH |
| Ensemble 8 | 3768\_2526\_2992\_2629 | 3398\_2014\_2681\_2010 | 3041\_3906\_2072\_3274 |
| Ensemble 9 | 4RFZ\_5P9K\_3768\_2526 | 6BBU\_6N7A\_4EHZ\_2010 | 6I8T\_3041\_3906\_3274 |
| Ensemble 10 | 4RFZ\_6O8I \_4RX5\_5P9K \_2992 | 6BBU\_6N7A\_6AAH\_4EHZ\_2010 | 6I8T\_6NRH\_3041\_3906\_3274 |
| Ensemble 11 | 4RFZ\_5P9K\_3768\_2526\_2992 | 6BBU\_6N7A\_6AAH\_4EHZ\_2014 | 4ZZZ\_6I8T\_6NRH\_3041\_3274 |
| Ensemble 12 | 4RFZ\_6O8I \_4RX5\_5P9K \_2992\_2629 | 6BBU\_6N7A\_6AAH\_4EHZ\_2014\_2010 | 6I8T\_6NRH\_3041\_3906\_2072\_3274 |
| Ensemble 13 | 4RFZ\_6O8I \_4RX5\_5P9K \_3768\_2992\_2629 | 6BBU\_6N7A\_6AAH\_4EHZ\_3398\_2014\_2010 | 4ZZZ\_6I8T\_6NRH\_3041\_3906\_2072\_3274 |
| Ensemble 14 | 4RFZ\_6O8I\_4RX5\_5P9K \_3768\_2526\_2992\_2629 | 6BBU\_6N7A\_6AAH\_4EHZ\_3398\_2014\_2681\_2010 | 4ZZZ\_5WS1\_6I8T\_6NRH\_3041\_3906\_2072\_3274 |

a 3768, 2526, 2992 and 2629 in the BTK represent the structure 5P9K\_3768, 5P9K\_2526, 5P9K\_2992 and 5P9K\_2629, respectively. 3398, 2014, 2681 and 2010 in the JAK represent the structure 6N7A\_3398, 6N7A\_2014, 6N7A\_2681 and 6N7A\_2010, respectively. 3041, 3906, 2072 and 3274 in the PARP represent the structure 6I8T\_3041, 6I8T\_3906, 6I8T\_2072 and 6I8T\_3274, respectively.

Table S8. Performance of the Bayesian Models Based on the Docking Scores of Each Single Representative Complex and Ensemble for BTK, JAK and PARP in the Training Seta

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 5-Fold Cross-Validation Result Using Training Set | | | | | | | | | | |
| **BTK** | | | | | | | | | | |
| Panel | Ensemble | ROC Score | ROC Rating | TP | FN | FP | TN | SE | SP | C |
| Crystal  Structure | 4RFZ | 0.957 | Excellent | 76 | 8 | 192 | 1475 | 0.905 | 0.885 | 0.886 |
| 6O8I | 0.96 | Excellent | 74 | 10 | 134 | 1533 | 0.881 | 0.92 | 0.918 |
| 4RX5 | 0.958 | Excellent | 74 | 10 | 155 | 1512 | 0.881 | 0.907 | 0.906 |
| 5P9K | 0.961 | Excellent | 76 | 8 | 157 | 1510 | 0.905 | 0.906 | 0.906 |
| MD Simulation | 3768 | 0.951 | Excellent | 80 | 4 | 201 | 1466 | 0.952 | 0.879 | 0.883 |
| 2526 | 0.963 | Excellent | 76 | 8 | 165 | 1502 | 0.905 | 0.901 | 0.901 |
| 2992 | 0.968 | Excellent | 78 | 6 | 166 | 1501 | 0.929 | 0.9 | 0.902 |
| 2629 | 0.967 | Excellent | 71 | 13 | 77 | 1590 | 0.845 | 0.954 | 0.949 |
| Two-size Ensemble | Ensemble 1 | 0.975 | Excellent | 74 | 10 | 54 | 1613 | 0.881 | 0.968 | 0.963 |
| Ensemble 2 | 0.974 | Excellent | 77 | 7 | 107 | 1560 | 0.917 | 0.936 | 0.935 |
| Ensemble 3 | 0.978 | Excellent | 80 | 4 | 185 | 1482 | 0.952 | 0.889 | 0.892 |
| Three-size Ensemble | Ensemble 4 | 0.979 | Excellent | 82 | 2 | 194 | 1473 | 0.976 | 0.884 | 0.888 |
| Ensemble 5 | 0.976 | Excellent | 76 | 8 | 80 | 1587 | 0.905 | 0.952 | 0.95 |
| Ensemble 6 | 0.981 | Excellent | 77 | 7 | 98 | 1569 | 0.917 | 0.941 | 0.94 |
| Four-size Ensemble | Ensemble 7 | 0.981 | Excellent | 80 | 4 | 134 | 1533 | 0.952 | 0.92 | 0.921 |
| Ensemble 8 | 0.977 | Excellent | 77 | 7 | 107 | 1560 | 0.917 | 0.936 | 0.935 |
| Ensemble 9 | 0.981 | Excellent | 75 | 9 | 83 | 1584 | 0.893 | 0.95 | 0.947 |
| Five-size Ensemble | Ensemble 10 | 0.983 | Excellent | 82 | 2 | 194 | 1473 | 0.976 | 0.884 | 0.888 |
| Ensemble 11 | 0.982 | Excellent | 81 | 3 | 196 | 1471 | 0.964 | 0.882 | 0.886 |
| Six-size Ensemble | Ensemble 12 | 0.983 | Excellent | 80 | 4 | 184 | 1483 | 0.952 | 0.89 | 0.893 |
| Seven-size Ensemble | Ensemble 13 | 0.983 | Excellent | 82 | 2 | 201 | 1466 | 0.976 | 0.879 | 0.884 |
| Eight-size Ensemble | Ensemble 14 | 0.982 | Excellent | 83 | 1 | 213 | 1454 | 0.988 | 0.872 | 0.878 |
| **JAK** | | | | | | | | | | |
| Panel | Ensemble | ROC Score | ROC Rating | TP | FN | FP | TN | SE | SP | C |
| Crystal  Structure | 6BBU | 0.862 | Good | 89 | 13 | 594 | 1483 | 0.873 | 0.714 | 0.721 |
| 6N7A | 0.906 | Excellent | 88 | 14 | 368 | 1709 | 0.863 | 0.823 | 0.825 |
| 6AAH | 0.87 | Good | 82 | 20 | 397 | 1680 | 0.804 | 0.809 | 0.809 |
| 4EHZ | 0.89 | Good | 95 | 7 | 574 | 1503 | 0.931 | 0.724 | 0.733 |  |
| MD Simulation | 3398 | 0.828 | Good | 86 | 16 | 606 | 1471 | 0.843 | 0.708 | 0.715 |
| 2014 | 0.864 | Good | 82 | 20 | 427 | 1650 | 0.804 | 0.794 | 0.795 |
| 2681 | 0.796 | Fair | 78 | 24 | 619 | 1458 | 0.765 | 0.702 | 0.705 |
| 2010 | 0.855 | Good | 87 | 15 | 540 | 1537 | 0.853 | 0.74 | 0.745 |
| Two-size Ensemble | Ensemble 1 | 0.913 | Excellent | 94 | 8 | 450 | 1627 | 0.922 | 0.783 | 0.79 |
| Ensemble 2 | 0.884 | Good | 93 | 9 | 566 | 1511 | 0.912 | 0.727 | 0.736 |
| Ensemble 3 | 0.912 | Excellent | 86 | 16 | 340 | 1737 | 0.843 | 0.836 | 0.837 |
| Three-size Ensemble | Ensemble 4 | 0.916 | Excellent | 94 | 8 | 472 | 1605 | 0.922 | 0.773 | 0.78 |
| Ensemble 5 | 0.879 | Good | 85 | 17 | 455 | 1622 | 0.833 | 0.781 | 0.783 |
| Ensemble 6 | 0.917 | Excellent | 96 | 6 | 505 | 1572 | 0.941 | 0.757 | 0.765 |
| Four-size Ensemble | Ensemble 7 | 0.926 | Excellent | 88 | 14 | 293 | 1784 | 0.863 | 0.859 | 0.859 |
| Ensemble 8 | 0.874 | Good | 86 | 16 | 485 | 1592 | 0.843 | 0.766 | 0.77 |
| Ensemble 9 | 0.919 | Excellent | 98 | 4 | 517 | 1560 | 0.961 | 0.751 | 0.761 |
| Five-size Ensemble | Ensemble 10 | 0.926 | Excellent | 95 | 7 | 366 | 1711 | 0.931 | 0.824 | 0.829 |
| Ensemble 11 | 0.924 | Excellent | 97 | 5 | 462 | 1615 | 0.951 | 0.778 | 0.786 |
| Six-size Ensemble | Ensemble 12 | 0.924 | Excellent | 98 | 4 | 498 | 1579 | 0.961 | 0.76 | 0.77 |
| Seven-size Ensemble | Ensemble 13 | 0.918 | Excellent | 89 | 13 | 376 | 1701 | 0.873 | 0.819 | 0.821 |
| Eight-size Ensemble | Ensemble 14 | 0.914 | Excellent | 89 | 13 | 399 | 1678 | 0.873 | 0.808 | 0.811 |
| **PARP** | | | | | | | | | | |
| Panel | Ensemble | ROC Score | ROC Rating | TP | FN | FP | TN | SE | SP | C |
| Crystal  Structure | 4ZZZ | 0.779 | Fair | 64 | 26 | 315 | 1472 | 0.711 | 0.824 | 0.818 |
| 5WS1 | 0.775 | Fair | 75 | 15 | 695 | 1092 | 0.833 | 0.611 | 0.622 |
| 6I8T | 0.878 | Good | 87 | 3 | 469 | 1318 | 0.967 | 0.738 | 0.749 |
| 6NRH | 0.843 | Good | 60 | 30 | 237 | 1550 | 0.667 | 0.867 | 0.858 |
| MD Simulation | 3041 | 0.879 | Good | 80 | 10 | 418 | 1369 | 0.889 | 0.766 | 0.772 |
| 3906 | 0.853 | Good | 77 | 13 | 457 | 1330 | 0.856 | 0.744 | 0.75 |
| 2072 | 0.841 | Good | 84 | 6 | 595 | 1192 | 0.933 | 0.667 | 0.68 |
| 3274 | 0.867 | Good | 83 | 7 | 493 | 1294 | 0.922 | 0.724 | 0.734 |
| Two-size Ensemble | Ensemble 1 | 0.905 | Excellent | 87 | 3 | 463 | 1324 | 0.967 | 0.741 | 0.752 |
| Ensemble 2 | 0.892 | Good | 87 | 3 | 504 | 1283 | 0.967 | 0.718 | 0.73 |
| Ensemble 3 | 0.912 | Excellent | 79 | 11 | 259 | 1528 | 0.878 | 0.855 | 0.856 |
| Three-size Ensemble | Ensemble 4 | 0.897 | Good | 87 | 3 | 495 | 1292 | 0.967 | 0.723 | 0.735 |
| Ensemble 5 | 0.895 | Good | 80 | 10 | 426 | 1361 | 0.889 | 0.762 | 0.768 |
| Ensemble 6 | 0.918 | Excellent | 82 | 8 | 301 | 1486 | 0.911 | 0.832 | 0.835 |
| Four-size Ensemble | Ensemble 7 | 0.893 | Good | 88 | 2 | 520 | 1267 | 0.978 | 0.709 | 0.722 |
| Ensemble 8 | 0.895 | Good | 79 | 11 | 411 | 1376 | 0.878 | 0.77 | 0.775 |
| Ensemble 9 | 0.919 | Excellent | 87 | 3 | 414 | 1373 | 0.967 | 0.768 | 0.778 |
| Five-size Ensemble | Ensemble 10 | 0.916 | Excellent | 76 | 14 | 264 | 1523 | 0.844 | 0.852 | 0.852 |
| Ensemble 11 | 0.914 | Excellent | 84 | 6 | 343 | 1444 | 0.933 | 0.808 | 0.814 |
| Six-size Ensemble | Ensemble 12 | 0.914 | Excellent | 80 | 10 | 353 | 1434 | 0.889 | 0.802 | 0.807 |
| Seven-size Ensemble | Ensemble 13 | 0.91 | Excellent | 82 | 8 | 378 | 1409 | 0.911 | 0.788 | 0.794 |
| Eight-size Ensemble | Ensemble 14 | 0.906 | Excellent | 81 | 9 | 353 | 1434 | 0.9 | 0.802 | 0.807 |

a TP, true positive; FN, false negative; FP, false positive; TN, true negative; SE, sensitivity; SP, specificity; C, concordance rate; ROC Score, Bayesian score; ROC Rating, model quality. 3768, 2526, 2992 and 2629 in the BTK represent the structure 5P9K\_3768, 5P9K\_2526, 5P9K\_2992 and 5P9K\_2629, respectively. 3398, 2014, 2681 and 2010 in the JAK represent the structure 6N7A\_3398, 6N7A\_2014, 6N7A\_2681 and 6N7A\_2010, respectively. 3041, 3906, 2072 and 3274 in the PARP represent the structure 6I8T\_3041, 6I8T\_3906, 6I8T\_2072 and 6I8T\_3274, respectively.

Table S9. Performance of the Bayesian Models Based on the Docking Scores of Each Single Representative Complex and Ensemble for BTK, JAK and PARP in the Test Seta

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Validation Result Using External Test Set | | | | | | | | | | |
| **BTK** | | | | | | | | | | |
| Panel | Ensemble | ROC Score | ROC Rating | TP | FN | FP | TN | SE | SP | C |
| Crystal  Structure | 4RFZ | 0.95 | Excellent | 32 | 9 | 26 | 807 | 0.78 | 0.969 | 0.96 |
| 6O8I | 0.948 | Excellent | 37 | 4 | 65 | 768 | 0.902 | 0.922 | 0.921 |
| 4RX5 | 0.942 | Excellent | 36 | 5 | 69 | 764 | 0.878 | 0.917 | 0.915 |
| 5P9K | 0.97 | Excellent | 38 | 3 | 21 | 812 | 0.927 | 0.975 | 0.973 |
| MD Simulation | 3768 | 0.948 | Excellent | 35 | 6 | 72 | 761 | 0.854 | 0.914 | 0.911 |
| 2526 | 0.941 | Excellent | 37 | 4 | 61 | 772 | 0.902 | 0.927 | 0.926 |
| 2992 | 0.983 | Excellent | 41 | 0 | 63 | 770 | 1 | 0.924 | 0.928 |
| 2629 | 0.952 | Excellent | 32 | 9 | 28 | 805 | 0.78 | 0.966 | 0.958 |
| Two-size Ensemble | Ensemble 1 | 0.973 | Excellent | 40 | 1 | 29 | 804 | 0.976 | 0.965 | 0.966 |
| Ensemble 2 | 0.986 | Excellent | 36 | 5 | 17 | 816 | 0.878 | 0.98 | 0.975 |
| Ensemble 3 | 0.992 | Excellent | 41 | 0 | 46 | 787 | 1 | 0.945 | 0.947 |
| Three-size Ensemble | Ensemble 4 | 0.982 | Excellent | 36 | 5 | 29 | 804 | 0.878 | 0.965 | 0.961 |
| Ensemble 5 | 0.985 | Excellent | 37 | 4 | 26 | 807 | 0.902 | 0.969 | 0.966 |
| Ensemble 6 | 0.991 | Excellent | 40 | 1 | 49 | 784 | 0.976 | 0.941 | 0.943 |
| Four-size Ensemble | Ensemble 7 | 0.98 | Excellent | 36 | 5 | 30 | 803 | 0.878 | 0.964 | 0.96 |
| Ensemble 8 | 0.982 | Excellent | 37 | 4 | 31 | 802 | 0.902 | 0.963 | 0.96 |
| Ensemble 9 | 0.981 | Excellent | 34 | 7 | 16 | 817 | 0.829 | 0.981 | 0.974 |
| Five-size Ensemble | Ensemble 10 | 0.988 | Excellent | 38 | 3 | 35 | 798 | 0.927 | 0.958 | 0.957 |
| Ensemble 11 | 0.989 | Excellent | 37 | 4 | 35 | 798 | 0.902 | 0.958 | 0.955 |
| Six-size  Ensemble | Ensemble 12 | 0.987 | Excellent | 38 | 3 | 52 | 781 | 0.927 | 0.938 | 0.937 |
| Seven-size Ensemble | Ensemble 13 | 0.986 | Excellent | 38 | 3 | 40 | 793 | 0.927 | 0.952 | 0.951 |
| Eight-size Ensemble | Ensemble 14 | 0.985 | Excellent | 38 | 3 | 51 | 782 | 0.927 | 0.939 | 0.938 |
| **JAK** | | | | | | | | | | |
| Panel | Ensemble | ROC Score | ROC Rating | TP | FN | FP | TN | SE | SP | C |
| Crystal  Structure | 6BBU | 0.856 | Good | 42 | 8 | 229 | 734 | 0.84 | 0.762 | 0.766 |
| 6N7A | 0.883 | Good | 46 | 4 | 221 | 742 | 0.92 | 0.771 | 0.778 |
| 6AAH | 0.854 | Good | 44 | 6 | 291 | 672 | 0.88 | 0.698 | 0.707 |
| 4EHZ | 0.87 | Good | 47 | 3 | 321 | 642 | 0.94 | 0.667 | 0.68 |
| MD Simulation | 3398 | 0.783 | Fair | 40 | 10 | 356 | 607 | 0.8 | 0.63 | 0.639 |
| 2014 | 0.796 | Fair | 38 | 12 | 333 | 630 | 0.76 | 0.654 | 0.659 |
| 2681 | 0.768 | Fair | 42 | 8 | 434 | 529 | 0.84 | 0.549 | 0.564 |
| 2010 | 0.834 | Good | 47 | 3 | 435 | 528 | 0.94 | 0.548 | 0.568 |
| Two-size Ensemble | Ensemble 1 | 0.894 | Good | 48 | 2 | 270 | 693 | 0.96 | 0.72 | 0.731 |
| Ensemble 2 | 0.836 | Good | 40 | 10 | 290 | 673 | 0.8 | 0.699 | 0.704 |
| Ensemble 3 | 0.887 | Good | 46 | 4 | 212 | 751 | 0.92 | 0.78 | 0.787 |
| Three-size Ensemble | Ensemble 4 | 0.896 | Good | 47 | 3 | 238 | 725 | 0.94 | 0.753 | 0.762 |
| Ensemble 5 | 0.835 | Good | 39 | 11 | 284 | 679 | 0.78 | 0.705 | 0.709 |
| Ensemble 6 | 0.885 | Good | 43 | 7 | 182 | 781 | 0.86 | 0.811 | 0.813 |
| Four-size Ensemble | Ensemble 7 | 0.911 | Excellent | 47 | 3 | 227 | 736 | 0.94 | 0.764 | 0.773 |
| Ensemble 8 | 0.829 | Good | 38 | 12 | 281 | 682 | 0.76 | 0.708 | 0.711 |
| Ensemble 9 | 0.899 | Good | 46 | 4 | 240 | 723 | 0.92 | 0.751 | 0.759 |
| Five-size Ensemble | Ensemble 10 | 0.912 | Excellent | 46 | 4 | 214 | 749 | 0.92 | 0.778 | 0.785 |
| Ensemble 11 | 0.9 | Excellent | 47 | 3 | 241 | 722 | 0.94 | 0.75 | 0.759 |
| Six-size  Ensemble | Ensemble 12 | 0.901 | Excellent | 46 | 4 | 238 | 725 | 0.92 | 0.753 | 0.761 |
| Seven-size Ensemble | Ensemble 13 | 0.894 | Good | 46 | 4 | 278 | 685 | 0.92 | 0.711 | 0.722 |
| Eight-size Ensemble | Ensemble 14 | 0.888 | Good | 46 | 4 | 281 | 682 | 0.92 | 0.708 | 0.719 |
| **PARP** | | | | | | | | | | |
| Panel | Ensemble | ROC Score | ROC Rating | TP | FN | FP | TN | SE | SP | C |
| Crystal  Structure | 4ZZZ | 0.791 | Fair | 24 | 20 | 132 | 761 | 0.545 | 0.852 | 0.838 |
| 5WS1 | 0.828 | Good | 35 | 9 | 227 | 666 | 0.795 | 0.746 | 0.748 |
| 6I8T | 0.908 | Excellent | 41 | 3 | 206 | 687 | 0.932 | 0.769 | 0.777 |
| 6NRH | 0.865 | Good | 33 | 11 | 128 | 765 | 0.75 | 0.857 | 0.852 |
| MD Simulation | 3041 | 0.906 | Excellent | 40 | 4 | 205 | 688 | 0.909 | 0.77 | 0.777 |
| 3906 | 0.92 | Excellent | 41 | 3 | 196 | 697 | 0.932 | 0.781 | 0.788 |
| 2072 | 0.87 | Good | 36 | 8 | 201 | 692 | 0.818 | 0.775 | 0.777 |
| 3274 | 0.934 | Excellent | 41 | 3 | 208 | 685 | 0.932 | 0.767 | 0.775 |
| Two-size Ensemble | Ensemble 1 | 0.925 | Excellent | 41 | 3 | 168 | 725 | 0.932 | 0.812 | 0.818 |
| Ensemble 2 | 0.934 | Excellent | 43 | 1 | 205 | 688 | 0.977 | 0.77 | 0.78 |
| Ensemble 3 | 0.938 | Excellent | 37 | 7 | 105 | 788 | 0.841 | 0.882 | 0.88 |
| Three-size Ensemble | Ensemble 4 | 0.923 | Excellent | 33 | 11 | 112 | 781 | 0.75 | 0.875 | 0.869 |
| Ensemble 5 | 0.938 | Excellent | 43 | 1 | 213 | 680 | 0.977 | 0.761 | 0.772 |
| Ensemble 6 | 0.939 | Excellent | 41 | 3 | 151 | 742 | 0.932 | 0.831 | 0.836 |
| Four-size Ensemble | Ensemble 7 | 0.92 | Excellent | 40 | 4 | 142 | 751 | 0.909 | 0.841 | 0.844 |
| Ensemble 8 | 0.933 | Excellent | 44 | 0 | 199 | 694 | 1 | 0.777 | 0.788 |
| Ensemble 9 | 0.948 | Excellent | 42 | 2 | 143 | 750 | 0.955 | 0.84 | 0.845 |
| Five-size Ensemble | Ensemble 10 | 0.949 | Excellent | 43 | 1 | 167 | 726 | 0.977 | 0.813 | 0.821 |
| Ensemble 11 | 0.942 | Excellent | 42 | 2 | 136 | 757 | 0.955 | 0.848 | 0.853 |
| Six-size  Ensemble | Ensemble 12 | 0.946 | Excellent | 44 | 0 | 172 | 721 | 1 | 0.807 | 0.816 |
| Seven-size Ensemble | Ensemble 13 | 0.942 | Excellent | 43 | 1 | 188 | 705 | 0.977 | 0.789 | 0.798 |
| Eight-size Ensemble | Ensemble 14 | 0.939 | Excellent | 42 | 2 | 149 | 744 | 0.955 | 0.833 | 0.839 |

a TP, true positive; FN, false negative; FP, false positive; TN, true negative; SE, sensitivity; SP, specificity; C, concordance rate; ROC Score, Bayesian score; ROC Rating, model quality. 3768, 2526, 2992 and 2629 in the BTK represent the structure 5P9K\_3768, 5P9K\_2526, 5P9K\_2992 and 5P9K\_2629, respectively. 3398, 2014, 2681 and 2010 in the JAK represent the structure 6N7A\_3398, 6N7A\_2014, 6N7A\_2681 and 6N7A\_2010, respectively. 3041, 3906, 2072 and 3274 in the PARP represent the structure 6I8T\_3041, 6I8T\_3906, 6I8T\_2072 and 6I8T\_3274, respectively.

GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. 2015.

Jonathan., L.*, et al.* Olaparib Maintenance Therapy in Platinum-Sensitive Relapsed Ovarian Cancer. *N Engl J Med* 2012;366(15):1382-1392.