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

Supporting Information

AromTool: Predict Aromatic Stacking Energy Using an Atomic Neural Network Model

Wengan He, [‡a] Danhong Liang, [‡ a] Kai Wang, [a,b] Nan Lv, [a,c] Hongjuan Diao, [a] and Ruibo Wu, [*a]

[a] *Guangdong Provincial Key Laboratory of New Drug Design and Evaluation, School of Pharmaceutical Sciences, Sun Yat-sen University, Guangzhou 510006, People's Republic of China*

[b] *School of agriculture and biology, Zhongkai University of Agriculture and Engineering, Guangzhou 510000, P. R. China*

[c] *State Key Laboratory of Physical Chemistry of Solid Surfaces and Fujian Provincial Key Laboratory of Theoretical and Computational Chemistry, College of Chemistry and Chemical Engineering, Xiamen University, Xiamen 361005, China*

[#] W. He and D. Liang contributed equally to this work.

*** E-mail**: wurb3@mail.sysu.edu.cn

Table of Contents

1. The stacking energies calculation for PDBbind

Protein-ligand complexes in PDBbind[1](PDBbind v2019 Protein-ligand complexes: The refined set) were downloaded for one of application cases, the dataset containing 4852 protein-drug systems. Both DFT(XYGJOS/6-311++G*) and AromTool were applied to calculate the stacking energies for the ligand-benzene containing aromatic stacking in this database. For the DFT calculations, they were done with Q-CHEM (ver. 4.0) software[2] for the aromatic stacking conformations based on three approximation – cutoff, reference ring surrogate and gas phase modelling. The cutoff and surrogate method was proposed because most of the PDB crystal structures' resolutions are low, which makes some conformations unreasonable. Therefore, it is difficult and even unable to calculate the stacking energies by DFT methods due to convergences problem. To keep consistency among all the samples, all the aromatic rings extracted from crystal structures were cut off and substituted with corresponding reference rings. Gas phase assumption supposes that the energies distribution and change trend in solution approximate to those in gas phase. Wheeler's work[3] found that the overall trends of the stacking interaction energies in pas phase remain unchanged compared with those in protein-like environments, so here in our study, we also calculate the stacking interaction energies in gas phase. As for AromTool calculation, 4852 protein-drug compounds were input to automatically search for the contact pair as well as calculate the stacking energies. Finally, AromTool gave a report, which includes the contact type, contact distance, contact angel and contact energy of each contact pair.

2. The stacking energies calculation for MD trajectory of OSC

Prepare for the protein-ligand complex

The X-ray crystal structure of human oxidosqualene-lanosterol cyclase (OSC) in complex with Ro-48-8071 was solved in 2004[4], obtained from the Protein Data Bank (PDB ID :1W6J). The crystal structure was prepared using Molecular Operating Environment (MOE) 2014 software[5], the detergents and waters molecules were removed. The protonation states of charged residues were determined by the H++ program[6] and carefully examining their individual local hydrogen bond networks. The resp charge of the ligand was calculated at the HF/6-31G* level using the Gaussian 09[7]. The simulation model was neutralized by adding Na+ ions by employing the AmberTool and solvated into a about 102*104*92 Å rectangular box of water molecules. The TIP3P model[8] and Amber99SB force field[9] were employed for water molecules and the protein,

respectively. And the force field parameters of ligand was generated from General Amber Force Field (GAFF)[10].

Classical MD simulation

The MD simulation was performed by using AMBER16[11] molecular dynamics package. The simulation system was optimized by three multistep minimizations. Then, the system was heated from 0 to 310K gradually under the *NVT* ensemble for 100ps. Afterwards, another 100ps MD simulations were employed under the NPT ensemble to relax density to the target pressure of 1.0 atm. In the NPT ensemble, the Berendsen thermostat method[12] was used to control the system temperature. Finally, a long timescale 50ns MD simulations under the NVT ensemble were performed with the periodic boundary condition. During the MD simulations, the SHAKE algorithm[13] was applied to constrain all hydrogen-containing bonds. The Langevin dynamics method was used to control the system temperature. And a cutoff of 12 Å was set for both van der Waals and electrostatic interactions. After the simulation, 4474 different conformations of the complex were extracted in chronological order from the trajectory file. Then, the Ben–Phe521 stacking conformation formed by the ligand benzene ring and the protein residue Phe521 side chain were extracted through AromTool, and then the aromatic stacking energy was calculated via the built-in AromNN model. Finally, we obtained the corresponding geometry and energy.

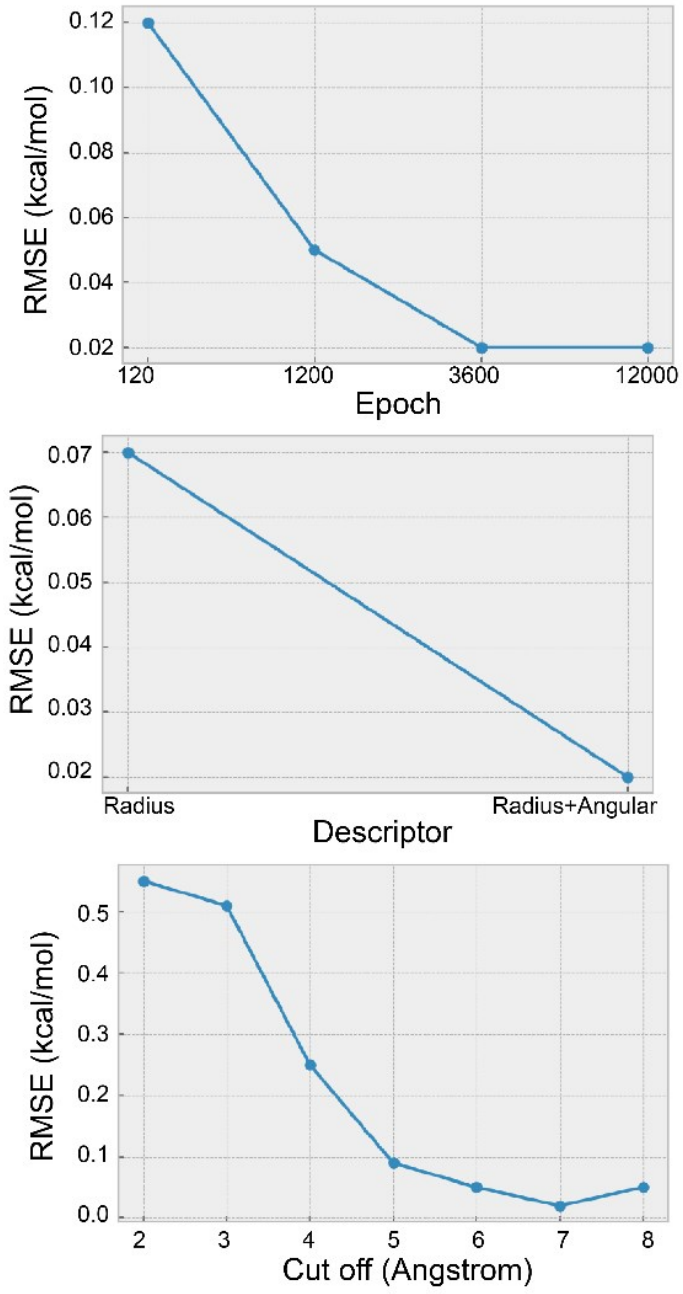

Figure S1. Hyperparameter optimization profiles of epoch, descriptor and cut off. The hyperparameter tuning was done on the 20% part of the training samples. See details in the corresponding part in the method section.

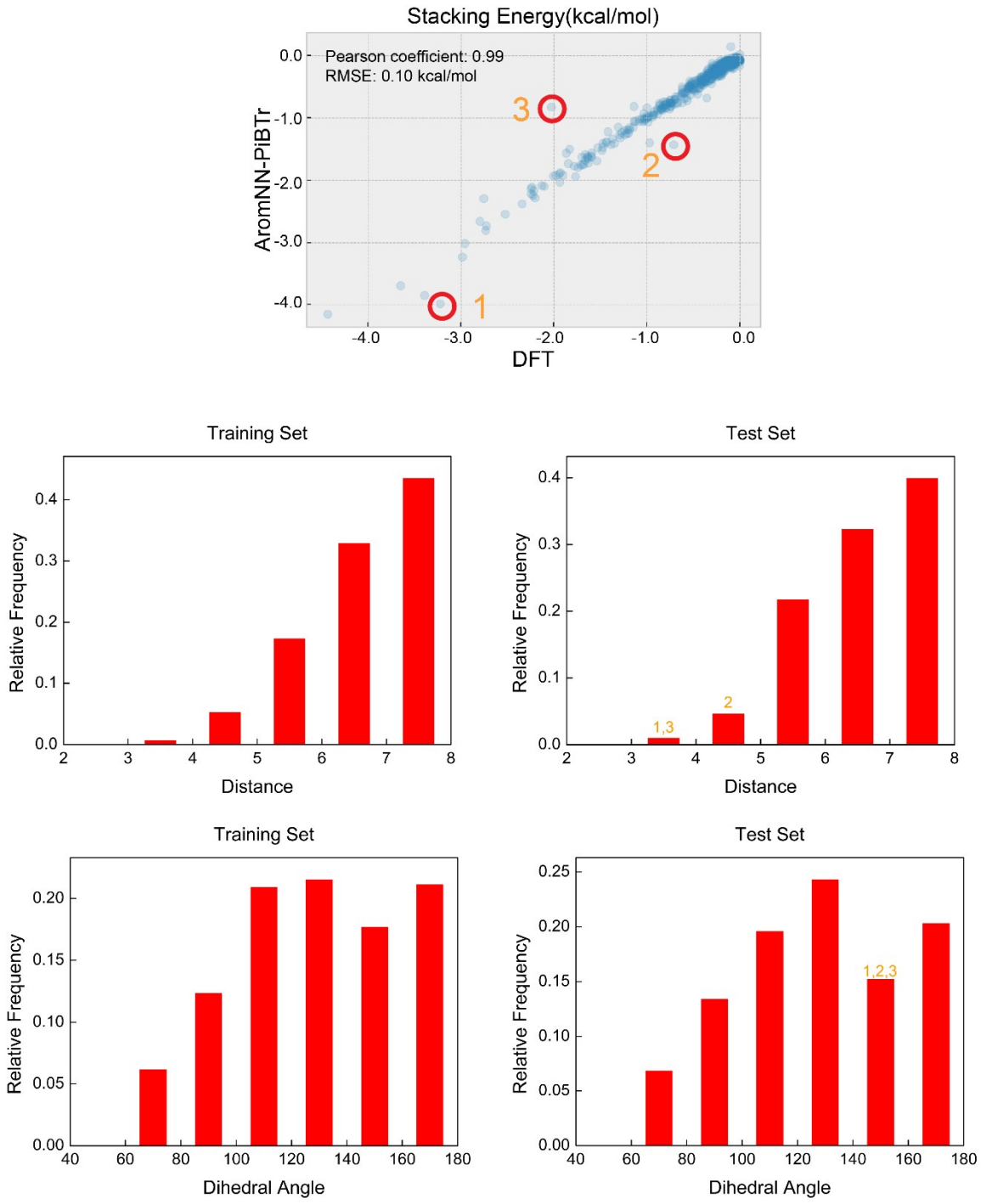

Figure S2. The three representative outliers in Figure5(C) for PiBTr's model and its corresponding geometry distribution in training and test set.

Table S1. The performance on the benchmark dataset for five DFT methods (B3LYP-D3, XYGJOS, LXYGJOS, M062X, M062X-D3) under the same basis set of 6-311++G**.

Table S2. The detailed results with energies, angle and distance analysis for PDBbind refine set.

Please see Table S1 and Table S2 in the excel files.

Reference

1. Z. Liu, Y. Li, L. Han, J. Li, J. Liu, Z. Zhao, W. Nie, Y. Liu and R. Wang, *Bioinformatics*, 2015, **31**, 405-412.
2. Y. Shao, Z. Gan, E. Epifanovsky, A. T. B. Gilbert, M. Wormit, J. Kussmann, A. W. Lange, A. Behn, J. Deng, X. Feng, D. Ghosh, M. Goldey P. R. Horn, L. D. Jacobson, I. Kaliman, R. Z. Khaliullin, T. Kús, A. Landau, J. Liu, E. I. Proynov, Y. M. Rhee, R. M. Richard, M. A. Rohrdanz, R. P. Steele, E. J. Sundstrom, H. L. Woodcock III, P. M. Zimmerman, D. Zuev, B. Albrecht, E. Alguire, B. Austin, G. J. O. Beran, Y. A. Bernard, E. Berquist, K. Brandhorst, K. B. Bravaya, S. T. Brown, D. Casanova, C.-M. Chang, Y. Chen, S. H. Chien, K. D. Closser, D. L. Crittenden, M. Diedenhofen, R. A. DiStasio Jr., H. Dop, A. D. Dutoi, R. G. Edgar, S. Fatehi, L. Fusti-Molnar, A. Ghysels, A. Golubeva-Zadorozhnaya, J. Gomes, M. W. D. Hanson-Heine, P. H. P. Harbach, A. W. Hauser, E. G. Hohenstein, Z. C. Holden, T.-C. Jagau, H. Ji, B. Kaduk, K. Khistyaev, J. Kim, J. Kim, R. A. King, P. Klunzinger, D. Kosenkov, T. Kowalczyk, C. M. Krauter, K. U. Lao, A. Laurent, K. V. Lawler, S. V. Levchenko, C. Y. Lin, F. Liu, E. Livshits, R. C. Lochan, A. Luenser, P. Manohar, S. F. Manzer, S.-P. Mao, N. Mardirossian, A. V. Marenich, S. A. Maurer, N. J. Mayhall, C. M. Oana, R. Olivares-Amaya, D. P. O'Neill, J. A. Parkhill, T. M. Perrine, R. Peverati, P. A. Pieniazek, A. Prociuk, D. R. Rehn, E. Rosta, N. J. Russ, N. Sergueev, S. M. Sharada, S. Sharmaa, D. W. Small, A. Sodt, T. Stein, D. Stück, Y.-C. Su, A. J. W. Thom, T. Tsuchimochi, L. Vogt, O. Vydrov, T. Wang, M. A. Watson, J. Wenzel, A. White, C. F. Williams, V. Vanovschi, S. Yeganeh, S. R. Yost, Z.-Q. You, I. Y. Zhang, X. Zhang, Y. Zhou, B. R. Brooks, G. K. L. Chan, D. M. Chipman, C. J. Cramer, W. A. Goddard III, M. S. Gordon, W. J. Hehre, A. Klamt, H. F. Schaefer III, M. W. Schmidt, C. D. Sherrill, D. G. Truhlar, A. Warshel, X. Xua, A. Aspuru-Guzik, R. Baer, A. T. Bell, N. A. Besley, J.-D. Chai, A. Dreuw, B. D. Dunietz, T. R. Furlani, S. R. Gwaltney, C.-P. Hsu, Y. Jung, J. Kong, D. S. Lambrecht, W. Liang, C. Ochsenfeld, V. A. Rassolov, L. V. Slipchenko, J. E. Subotnik, T. Van Voorhis, J. M. Herbert, A. I. Krylov, P. M. W. Gill, and M. Head-Gordon. Advances in molecular quantum chemistry contained in the Q-Chem 4 program package.
3. A. N. Bootsma, A. C. Doney and S. E. Wheeler, *J Am Chem Soc*, 2019, **141**, 11027-11035.
4. R. Thoma, T. Schulz-Gasch, B. D'Arcy, J. Benz, J. Aebi, H. Dehmlow, M. Hennig, M. Stihle and A. Ruf, *NATURE*, 2004, **432**, 118-122.
5. V. Santiago, C. Giorgio and M. Stefano, *Current Topics in Medicinal Chemistry*, 2008, **8**, 1555-1572.
6. J. C. Gordon, J. B. Myers, T. Folta, V. Shoja, L. S. Heath and A. Onufriev, *Nucleic acids research*, 2005, **33**, W368-W371.
7. M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria,M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone,B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato,X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng,J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda,J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao,H. Nakai, T. Vreven, J. A. Montgomery Jr, J. E. Peralta,F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers,K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand,K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar,J. Tomasi, M. Cossi, N. Rega, J. M. Millam, M. Klene,J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo,R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin,R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin,K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador,J. J. Dannenberg, S. Dapprich, A. D. Daniels,

O¨. Farkas,J. B. Foresman, J. V. Ortiz, J. Cioslowski and D. J. Fox,Gaussian 09, Gaussian, Inc., Wallingford Ct, 2009.
8. W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey and M. L. Klein, *The Journal of Chemical Physics*, 1983, **79**, 926-935.
9. V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins: Structure, Function, and Bioinformatics*, 2006, **65**, 712-725.
10. J. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *Journal of Computational Chemistry*, 2004, **25**, 1157-1174.
11. D. A. Case, R. M. Betz, D. S. Cerutti, T. E. Cheatham, Iii,T. A. Darden, R. E. Duke, T. J. Giese, H. Gohlke, A. W. Goetz,N. Homeyer, S. Izadi, P. Janowski, J. Kaus, A. Kovalenko,T. S. Lee, S. Legrand, P. Li, C. Lin, T. Luchko, R. Luo,B. Madej, D. Mermelstein, K. M. Merz, G. Monard,H. Nguyen, H. T. Nguyen, I. Omelyan, A. Onufriev,D. R. Roe, A. Roitberg, C. Sagui, C. L. Simmerling,W. M. Botello-Smith, J. Swails, R. C. Walker, J. Wang,R. M. Wolf, X. Wu, L. Xiao and P. A. Kollman, Amber2016, University of California, San Francisco, 2016.
12. H. J. C. Berendsen, J. P. M. Postma, W. F. v. Gunsteren, A. DiNola and J. R. Haak, *The Journal of Chemical Physics*, 1984, **81**, 3684-3690.
13. J.-P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *Journal of Computational Physics*, 1977, **23**, 327-341.