

## Electronic Supplementary Information (ESI†)

### Virtual Coformer Screening by a Combined Machine Learning and Physics-Based Approach

Jiuchuang Yuan<sup>†a</sup>, Xuetao Liu<sup>†b,a</sup>, Simin Wang<sup>a</sup>, Chao Chang<sup>a</sup>, Qiao Zeng<sup>a</sup>, Zhengtian Song<sup>a</sup>, Yingdi Jin<sup>a</sup>, Qun Zeng<sup>a,\*</sup>, GuangXu Sun<sup>a</sup>, Shigang Ruan<sup>a</sup>, Chandler Greenwell<sup>c</sup>, Yuriy A. Abramov<sup>c,d,\*</sup>

<sup>a</sup> XtalPi Inc., Shenzhen Jingtai Technology Co., Ltd. Floor 3, Sf Industrial Plant, No. 2 hongliu Road, Fubao Community, Fubao Street, Futian District, Shenzhen, 518100 China

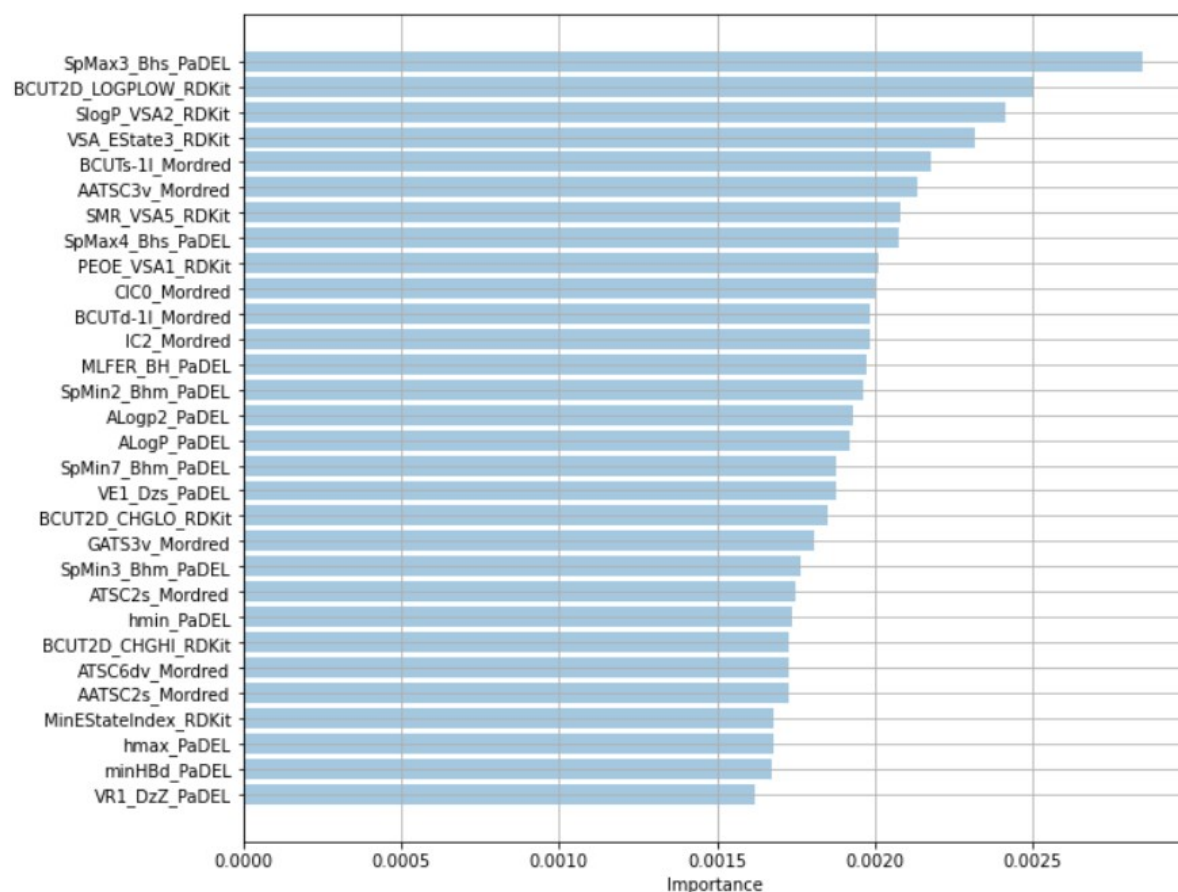
<sup>b</sup> Lab of Computational Chemistry and Drug Design, State Key Laboratory of Chemical Oncogeomics, Peking University Shenzhen Graduate School, Shenzhen, 518055 China

<sup>c</sup> XtalPi Inc, Cambridge, Massachusetts 02142, United States

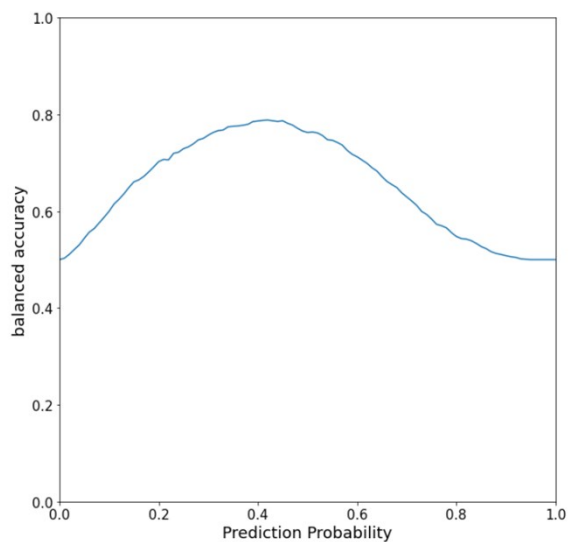
<sup>d</sup> Eshelman School of Pharmacy, University of North Carolina, Chapel Hill, North Carolina 27599, United States

\*Corresponding author.

†These authors contributed equally to this work



**Figure S1.** Thirty most important features of the ML RF model.



**Figure S2.** The balanced accuracy distribution of the cross-validation results of the RF model.

**Table S1.** Published ibuprofen coformer screening results.<sup>1-4</sup>

NO.	Name	Exp.
1	Succinic Acid	0
2	Citric Acid	1
3	L-Glutamic Acid	0
4	L-Phenylalanine	0
5	L-Histidine	0
6	L-Threonine	0
7	Lysine	0
8	L(+)-Arginine	1
9	Nicotinamide	1
10	Acetamide	0
11	Saccharin	1
12	Salicylamide	0
13	Oxalic acid	1
14	L-Proline	1
15	L-Alanine	1
16	Glycine	0

**Table S2.** A proposed ranking scheme for the ML+COSMO-RS model.

Order	AD	ML prob	$H_{ex}$ , kJ/mol	Ranked by
1	Inside	>0.62	any	ML prob
2	Inside	>0.42	<-4.4	ML prob
3	Inside	>0.42	$\geq$ -4.4	ML prob
4	Inside	$\leq$ 0.42	<-4.4	ML prob
5	Outside	any	<-4.4	$H_{ex}$
6	Inside	$\leq$ 0.42	$\geq$ -4.4	ML prob
7	Outside	any	$\geq$ -4.4	$H_{ex}$

## References

1. S. A. Hashib, N. Anuar, N. Jamburi, N.F. Ahmad, S.A. Rahim, Applied Mechanics and Materials, 2015, 754,1002.
2. M.F. Othman, N. Anuar, S.A. Ad, N.A.A. Taifuddin, IOP Conf. Ser.: Mat. Sci. Eng., 2018, 358, 012065.
3. M. F. Othman, N. Jamburi, N. Anuar, S.A. Rahim, N. H. Rohalim, MATEC Web Conf., 2016, 69, 03002.
4. Shen Z M, A Study on the Preparation of Ibuprofen Cocrystals (Masteral dissertation In Chinese), 2013, Tianjin University