Supporting Information

Identifying Structure-Property Relationships through SMILES Syntax Analysis with Self-Attention Mechanism

Shuangjia Zheng[1], Xin Yan*[1], Yuedong Yang[2,4] and Jun Xu*[1,3]

[1]Research Center for Drug Discovery, School of Pharmaceutical Sciences, Sun Yat-sen University, 132 East Circle at University City, Guangzhou 510006, China

[2]National Supercomputer Center in Guangzhou, Sun Yat-sen University, Guangzhou 510006, China

[3]School of Computer Science & Technology, Wuyi University, 99 Yingbin Road, Jiangmen 529020, China

[4]Key Laboratory of Machine Intelligence and Advanced Computing (Sun Yat-sen University), Ministry of Education, Guangzhou 510000, China

*To whom correspondence should be addressed.

Xin Yan: yanxin_0736@hotmail.com

Jun Xu：junxu@biochemomes.com or xujun9@mail.sysu.edu.cn

Table S1. Training details. Common to all models are: N_{hidden} *= 128, batch size = 64, embedding size = 64, optimizer = Adam, attention bias = false, regularization = True, dropout = 0.2, clip = 0.3*.

Dataset	N_{epoch}	*LR*[a]	*Loss*	*Row*[b]	*Note*
Solubility	30	0.0007	MSE[c]	10	88% training, 12% internal validation, separate test dataset
Photovoltaics	100	0.001	MSE	15	88% training, 12% internal validation, separate test dataset

Stability	50	0.001	BCE[d]	10	Stratified 10-fold CV
Tox21	50	0.001	BCE	10~18	Stratified 5-fold CV
DUD-E	50	0.001	BCE	15	Stratified 5-fold CV
HIV	100	0.001	BCE	18	Stratified 10-fold CV
Drug efficacy	100	0.0007	MSE	15	88% training, 12% internal validation, separate test dataset

[a]*LR*: Learning rate.
[b]*Row*: Attention row (*r*)
[c]MSE: mean squared error
[d]BCE: Binary cross entropy