Development of a Quantitative Estimate Index for Early-Stage Screening of Compounds Targeting Protein-Protein Interactions

Takatsugu Kosugi[1,†] Masahito Ohue[1]

1. Department of Computer Science, School of Computing, Tokyo Institute of Technology, Kanagawa, Japan

† Contact : kosugi@li.c.titech.ac.jp

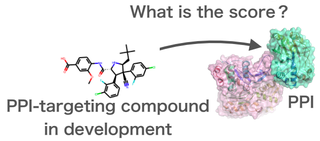
Kosugi, T.; Ohue, M. Quantitative estimate of protein-protein interaction targeting drug-likeness.

In Proceedings of The 18th IEEE International Conference on Computational Intelligence in Bioinformatics and Computational Biology (CIBCB 2021) (in press), ChemRxiv, Preprint. 2021. doi:10.33774/chemrxiv-2021-psqq4-v2.

Introduction

There are indexes to score compounds to develop drugs, Lipinski's rule of five (RO5) and QED are very famous.

Protein-Protein Interaction (PPI) is a recent drug discovery target, but it's very difficult to develop drugs, there are no indexes to score whether to target PPI



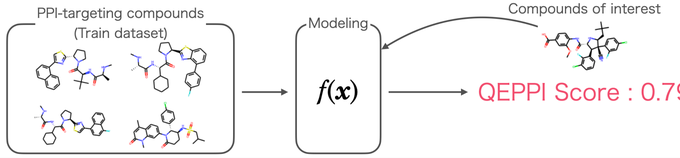
Morelli et al. proposed the "Rule-of-Four" (RO4) to evaluate PPI inhibitors[8].

RO4 is useful for filtering out PPI inhibitors, but it is not well quantitative because it is rule-based and classificated by with or without violations.

Problems

- QED is modeled for oral drugs. The drug-likeness represents is oral drug-like.
- RO4 is useful for filtering PPI inhibitors, but it is not well quantitative (rule-based)

In this study, we solved these problems by developing a index called the QEPPI (Quantitative Estimate Index for Compounds Targeting Protein-Protein Interactions) , specifically for early-stage screening of PPI-targeting compounds.



Methods (Difference from QED)

QEPPI was calculated using essentially the same procedure as that of the original QED, except for a few points.

The differences from the original QED are as follows:

	QED	QEPPI
The aim of the index	Quantitative estimate	Quantitative estimate PPI-targeting drug-likeness
Dataset	771 oral FDA approved drugs	1007 PPI-targeting compounds
The number of Descriptors	8	7 (didn't use "ALERTS")
How to build the model	TableCurve 2D	Levenberg-Marquardt algorithm in SciPy

It is basically the same as the modeling procedure for QED.

QED : Nat. Chem. 2012, 4, 90–98

iPPI-DB : Bioinformatics 2021, 37, 89–96

Methods

Plotting histograms

The same procedure as that of the original QED

Fitting of ADS function

The ADS function $Q(x)$ (Eq. (1)) by implementing the Levenberg-Marquardt algorithm in SciPy

$$Q(x) = a + \frac{b}{1+\exp\left(-\frac{x-c+\frac{d}{2}}{e}\right)}\left[1 - \frac{b}{1+\exp\left(-\frac{x-c-\frac{d}{2}}{f}\right)}\right] \quad (1)$$

Normalization of All fitting functions

All fitting functions ($Q_{\mathrm{MW}}(x)$, $Q_{\mathrm{ALogP}}(x)$, $Q_{\mathrm{HBD}}(x)$, $Q_{\mathrm{HBA}}(x)$, $Q_{\mathrm{TPSA}}(x)$, $Q_{\mathrm{ROTB}}(x)$, and $Q_{\mathrm{AROM}}(x)$) were divided by the maximum value and normalized to give a value 0 to1.

The normalized function $\tilde{Q}_i(x)(i \in \{\mathrm{MW}, \mathrm{ALogP}, \mathrm{HBD}, \mathrm{HBA}, \mathrm{TPSA}, \mathrm{ROTB}, \mathrm{AROM}\})$ was used as the desirability function.

Weighted desirability functions

The QEPPI score of compound *k* was assigned as the weighted geometric mean of all desirability functions (Eq. (2)).

$$\mathrm{QEPPI}_k = \exp\left(\frac{\sum_i w_i \ln\left(\tilde{Q}_i\right)}{\sum_i w_i}\right) \quad (2)$$

Assignment of weights

The seven weights were thoroughly tested from 0 to 1 in increments of 0.25, and the average of the 1,000 combinations of weights that resulted in the highest Shannon entropy was adopted. The Shannon entropy of the model was calculated (Eq. (8)).

$$\mathrm{entropy} = -\sum_{k=1}^{n} \mathrm{QEPPI}_k \log_2 \mathrm{QEPPI}_k \quad (3)$$

Where *n* represents the number of compounds used in the modeling.

Peak and weight of descriptors in QEPPI

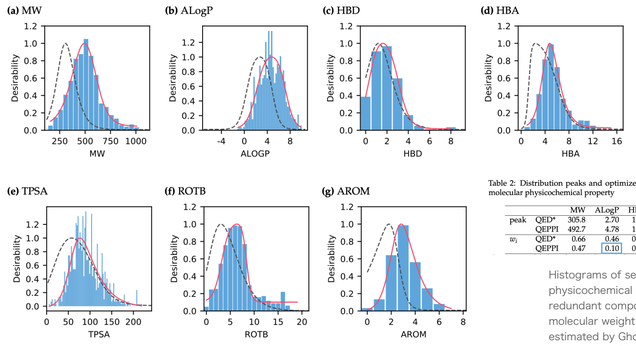


Table 2: Distribution peaks and optimized desirability function weightings of each molecular physicochemical property

		MW	ALogP	HBD	HBA	TPSA	ROTB	AROM
peak	QED*	305.8	2.70	1.20	2.38	57.5	3.04	1.8
	QEPPI	492.7	4.78	1.61	4.79	76.9	6.37	2.8
w_i	QED*	0.66	0.46	0.61	0.05	0.06	0.65	0.48
	QEPPI	0.47	0.10	0.82	0.81	0.37	0.53	0.89

Histograms of seven molecular physicochemical properties for a set of non-redundant compounds of iPPI-DB. (a)–(g), molecular weight (MW) (a), LogP value estimated by Ghose-Crippen method (ALogP) (b), number of hydrogen bond donors (HBD) (c), number of hydrogen bond acceptors (HBA) (d), Topological molecular polar surface area (TPSA) (e), number of rotatable bonds (ROTB) (f), and number of aromatic rings (AROM) (g). The solid red lines describe the ADS function (1) used to model the QEPPI histograms. The black dashed lines describe the ADS function used to model the QED histograms.

These results show that the weight of ALogP, an important descriptor of oral absorption in QED, is mostly ignored in QEPPI. On the other hand, the weights of HBA and TPSA, which are mostly ignored in QED, are given more significance in QEPPI. It suggests that QEPPI can capture PPI-targeting drug-like properties compared to QED and can play a different role in the seed compound discovery process.

Results

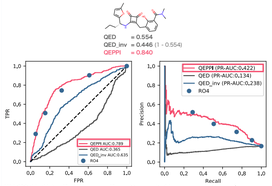
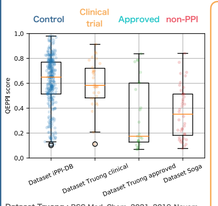


TABLE IV: Precision, Recall, and F-score values for one violation of RO4 and QEPPI score with the threshold value of 0.5196.

	Precision	Recall	F-score
RO4	0.398	0.508	0.446
QEPPI	0.378	0.735	0.499



Dataset Truong : RSC Med. Chem. 2021, 2019-Novem

Dataset Soga : J. Chem. Inf. Model. 2007, 47, 400–406

Clinical trial examples

Vercirnon (QEPPI : 0.91)
(CCR9)
Phase 3

Idasanutlin (QEPPI : 0.79)
(p53/Mdm2)
Phase 3

Apabetalone(QEPPI : 0.85)
(BET)
Phase 3

Approved examples

Romidepsin(QEPPI : 0.14)
(HDAC/PI3K)
Market start in 2009

Paclitaxel (QEPPI : 0.11)
(Microtubule)
Market start in 1993

Temsirolimus (QEPPI : 0.07)
(mTOR)
Market start in 2007

Both ROC-AUC and PR-AUC are higher than 1- QED (QED_inv). In addition, F-score is also higher than RO4. These results suggest that QEPPI performs better than other indexes.

Interestingly, when each value of RO4 was plotted on the ROC and AUC curves of QEPPI, they were very close to each other, suggesting that RO4, an index of discrete value, could be extended to an index of continuous value. The results suggested that QEPPI is a general extension of the RO4 concept.

Our application of QEPPI to the 30 clinical candidates used by Truong *et al.* showed a median value of approximately 0.59, which is higher than that of commercially available PPI modulators. The QEPPI modeled from iPPI-DB has the potential to be adapted to more recent PPI modulators.

Conclusion

- In this study, we solved these problems by developing a quantitative index called the QEPPI (Quantitative Estimate Index for Compounds Targeting Protein-Protein Interactions), specifically for early-stage screening of PPI-targeting compounds.
- QEPPI performs better than other conventional indexes such as RO4 and QED_inv.
- QEPPI was also considered to be an extension of the concept of RO4
- QEPPI has the potential to be more suitable for more recent PPI-targeting compounds.

Acknowledgements

This work was partially supported by;

- KAKENHI (20H04280) from the Japan Society for the Promotion of Science (JSPS)
- ACT-X (JPMJAX20A3) from the Japan Science and Technology Agency (JST)
- Mizuho Foundation for the Promotion of Sciences