



Article

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

Takatsugu Kosugi and Masahito Ohue *

Department of Computer Science, School of Computing, Tokyo Institute of Technology, G3-56-4259 Nagatsutacho, Midori-ku, Yokohama 226-8501, Kanagawa, Japan; kosugi@li.c.titech.ac.jp

* Correspondence: ohue@c.titech.ac.jp; Tel.: +81-45-924-5522

Abstract: Drug-likeness quantification is useful for screening drug candidates. Quantitative estimates of drug-likeness (QED) are commonly used to assess quantitative drug efficacy but are not suitable for screening compounds targeting protein-protein interactions (PPIs), which have recently gained attention. Therefore, we developed a quantitative estimate index for compounds targeting PPIs (QEPPi), specifically for early-stage screening of PPI-targeting compounds. QEPPi is an extension of the QED method for PPI-targeting drugs that models physicochemical properties based on the information available for drugs/compounds, specifically those reported to act on PPIs. FDA-approved drugs and compounds in iPPI-DB, which comprise PPI inhibitors and stabilizers, were evaluated using QEPPi. The results showed that QEPPi is more suitable than QED for early screening of PPI-targeting compounds. QEPPi was also considered an extended concept of the “Rule-of-Four” (RO4), a PPI inhibitor index. We evaluated the discriminatory performance of QEPPi and RO4 for datasets of PPI-target compounds and FDA-approved drugs using F-score and other indices. The F-scores of RO4 and QEPPi were 0.451 and 0.501, respectively. QEPPi showed better performance and enabled quantification of drug-likeness for early-stage PPI drug discovery. Hence, it can be used as an initial filter to efficiently screen PPI-targeting compounds.

Keywords: drug discovery; protein-protein interaction (PPI); PPI-targeting drug; virtual screening; QED; QEPPi



Citation: Kosugi, T.; Ohue, M. Quantitative Estimate Index for Early-Stage Screening of Compounds Targeting Protein-Protein Interactions. *Int. J. Mol. Sci.* **2021**, *22*, 10925. <https://doi.org/10.3390/ijms222010925>

Academic Editors: Koji Sode, Takenori Satomura and Wataru Yoshida

Received: 13 September 2021
Accepted: 7 October 2021
Published: 10 October 2021

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Protein-protein interactions (PPIs) have attracted attention as drug targets since the early 2000s [1–5]. However, it is difficult to design drugs for PPIs based on conventional rules, such as Lipinski’s rule of five (RO5) [6,7], because their physicochemical characteristics are very different from those of conventional drug targets [8,9]. In fact, only a few PPI inhibitors have been approved to date, and few PPI-targeting drug candidates have advanced in clinical trials to subsequent phases [10]. Therefore, an index that can be used to computationally select compounds that are likely to target PPIs is needed.

The quantitative estimate of drug-likeness (QED), which was proposed in 2012 [11], is an index of drug-likeness modeled using information available on marketed drugs and is widely used in current small-molecule drug discovery for computational methods [12,13] and to evaluate drug-like properties [14]. The QED index models these properties using data available from 771 orally administered drugs already approved by the U.S. Food and Drug Administration (FDA). However, this index is not an appropriate measure for PPI-targeting compounds, which require a relatively large surface area of the protein with which to interact. Therefore, new measures should be developed for PPI-targeting drugs [15].

QEX [16] and QEPT [17] are examples of QED remodeling methods that are based on the concept of QED and involve modeling physicochemical properties. In the case of QEX, the target compounds act on each target protein. In the case of QEPT, the target compounds are organic chemicals obtained from plant roots. These compounds represent quintessential

successful models for evaluating physicochemical properties. The idea is based on the remodeling of a PPI-targeting drug based on an already approved PPI-targeting drug. However, many molecular optimizations must be performed before approval. Even PPI-targeting compounds have been optimized to exhibit the general characteristics of drugs, such as RO5 (low-molecular-weight and water solubility). However, indices such as QED are mainly used in the early stages of drug discovery, that is, seed compound discovery. The metrics modeled from PPI-targeting drugs that are already available on the market are idealistic and unsuitable for the early stage.

In recent years, small-molecule de novo design methods have increasingly been used in combination with generative modeling and reinforcement learning [12,13,18,19]. However, it is difficult to design PPI-targeting compounds using these methods, and there is currently no continuous index of the likelihood of PPI-targeting compounds.

In this study, we developed a method named QEPPi (Quantitative Estimate Index for Compounds Targeting Protein-Protein Interactions), which is useful for early-stage PPI-targeting drug discovery based on data from compounds that have undergone extensive PPI inhibition or stabilization experiments, rather than on data from marketed PPI-targeting drugs. The code and easy-to-use environment of QEPPi are available on Google Colaboratory at <https://github.com/ohuelab/QEPPi>, accessed on 30 September 2021.

We published a preliminary version of this work in conference proceedings [20], including only limited experiments and discussion. In the current study, we have expanded the introduction and a discussion based on the results of physicochemical characterization of PPI modulators by Truong et al. [21]. In addition, we developed an implementation code to make QEPPi publicly available.

2. Results

2.1. Model Building for QEPPi

QEPPi is an index in the early-stages of PPI drug discovery, and the prerequisites for using the dataset to model QEPPi are as follows:

- Not limited to those in the approval phase or marketed after approval, as various optimizations will be performed during the approval phase.
- Not limited to PPI structures or complexes of protein and PPI-targeting compounds with known structures.

The reason for these requirements is that a drug undergoes many molecular optimizations before it is approved. Therefore, if only approved compounds are used as the dataset to create the model, too many ideal compounds will be identified as unsuitable for this purpose. In addition, data on various candidate compounds are necessary for the initial stage of the search. According to a recent review by Shin et al. [15], if only compounds with known structures are selected for X-ray crystallography of proteins and ligands, the amount of data that can be handled will involve tens to hundreds of compounds, although more than 720,000 human PPIs are known BioGrid [22] Current Build Statistics (4.3.196)—April 2021). Considering the number of known PPIs, structural information on PPIs likely remains insufficient. We hypothesized that information on the target proteins and their ligand compounds would be sufficient without requiring three-dimensional structures.

Therefore, we used iPPI-DB [23], which was manually curated from the literature. In total, 2361 PPI-targeting compounds are registered in this database (as of 21 April 2021), which are primarily derived from PPI inhibition or stabilization experiments. The number of compounds registered in DrugBank was 43, which is approximately 1.8% of the total. The quality and quantity of the data meet the requirements of the dataset for modeling QEPPi.

We built the QEPPi model using data selected after clustering for non-redundancy (see Methods for details). The histograms of the distributions of seven molecular physicochemical properties, MW, ALogP, HBD, HBA, TPSA, ROTB, and AROM, are shown in Figure 1. The distribution peaks and optimized weights w_i of each physicochemical property are shown in Table 1.

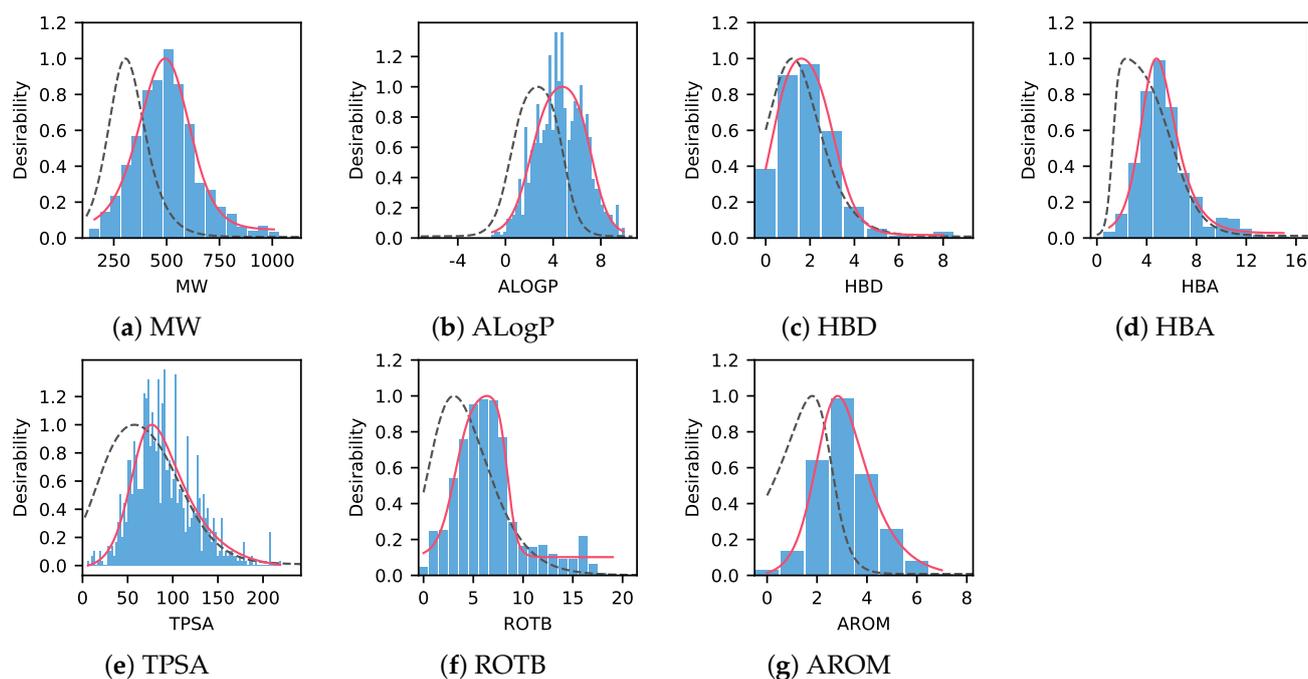


Figure 1. Histograms of seven molecular physicochemical properties for a set of non-redundant compounds of iPPI-DB. Molecular weight (MW) (a), LogP value estimated by Ghose-Crippen method (ALogP) (b), number of hydrogen bond donors (HBD) (c), number of hydrogen bond acceptors (HBA) (d), topological molecular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

Table 1. 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 |
| | QEPI | 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 |
| | QEPI | 0.47 | 0.10 | 0.82 | 0.81 | 0.37 | 0.53 | 0.89 |

* QED was modeled as a function that includes ALERTS; the peak value of ALERTS in QED was -24.6 , and its weight w_{ALERTS} was 0.95 .

Figure 1 and Table 1 show that oral drugs and PPI-targeting compounds have very different properties. Table 1 shows that the peak values of all properties were higher for QEPI than for QED. Particularly, the major difference between QEPI and QED is the peak value of ALogP (QEPI: 4.78, QED: 2.70), suggesting that low lipophilicity and high hydrophilicity are important for oral drugs in terms of oral absorption. This suggests that QEPI can capture PPI-targeting drug-like properties compared to QED and has a different role in the seed compound discovery process, which is the early-stage of drug discovery.

2.2. Evaluation of QEPI

To evaluate whether QEPI, which was developed in this study, is a more useful index for early-stage PPI drug discovery compared to QED, we obtained data on 321 PPI-targeting compounds from the iPPI-DB that were not used for model building (iPPI-DB dataset). In addition, we obtained data on 1596 FDA-approved drugs, excluding duplicates and approved drugs targeting PPI (FDA dataset). The QED score was calculated using these data; the distribution of these values is shown in Figure 2a. Similarly, the QEPI score was calculated, and the distribution of the values is shown in Figure 2b.

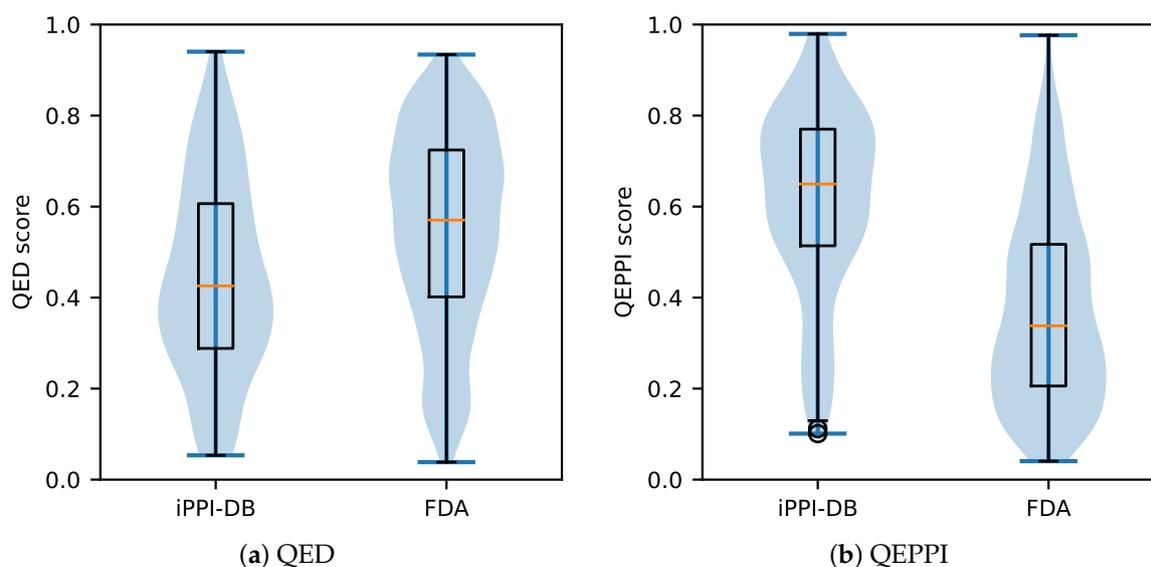


Figure 2. Distribution of QED and QEPPI in the PPI-targeting compounds dataset and FDA-approved drug dataset. Each filled area extends to represent the entire data range, with optional lines at the median. The QED score was calculated for both datasets (a). The QEPPI score was calculated for both datasets (b).

Figure 2a shows that PPI-targeting compounds exhibit a lower distribution of QED scores compared to conventional drugs, suggesting that QED is not an appropriate measure for PPI-targeting compounds, as it typically represents oral drug-like properties rather than drug-likeness. Figure 2b shows that PPI-targeting compounds have a higher distribution of QEPPI scores compared to conventional drugs, and a QEPPI threshold of 0.5 is sufficient to identify approximately 75% of PPI-targeting compounds. Furthermore, PPI-target drugs have been removed from the FDA dataset based on the literature [21]; as there are few PPI-targeting compounds in the FDA dataset, the smaller QEPPI scores in the FDA dataset compared to those in the iPPI-DB dataset are consistent. However, the results in Figure 2 show that for each dataset, QED and QEPPI had almost opposite trends, and thus, $1 - \text{QED}$ may be a similar index to QEPPI. Therefore, we calculated the ROC curve and area under the curve (AUC) to evaluate the quantitative performance of QEPPI and $1 - \text{QED}$ in identifying PPI-targeting compounds. The true-positive rate and false-positive rate were calculated to plot the ROC curve. Figure 3 shows the ROC curves obtained from QEPPI, QED, and the value of $1 - \text{QED}$. For QED, the AUC was less than 0.5 (0.362), which was worse than that obtained using randomly selected compounds. This is consistent with the results shown in Figure 2a, which shows that the AUC of QEPPI (0.789) was higher than that of $1 - \text{QED}$ (0.638), clearly indicating that QEPPI performs better than $1 - \text{QED}$ in identifying whether a compound is likely to be a PPI-targeting compound. The FDA dataset removed 13 compounds that overlapped with PPI-targeting drugs in the “Truong approved dataset”. We also performed calculations using the dataset of 1609 compounds without removing the 13 PPI-targeting compounds, which gave an AUC of QEPPI, QED, and $1 - \text{QED}$ values of 0.789, 0.365, and 0.635, respectively (see Supplementary Figure S1).

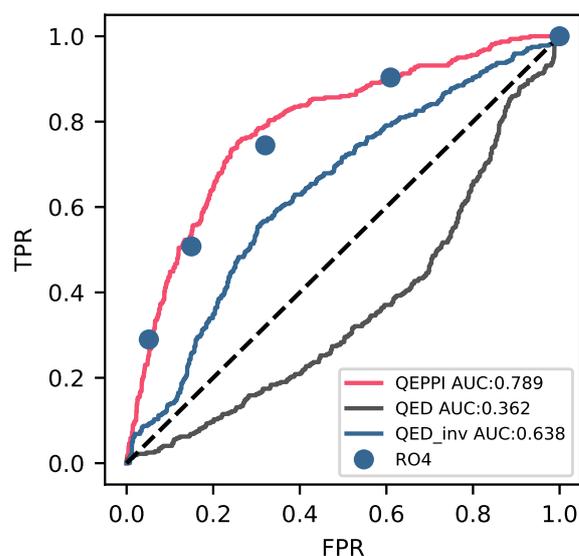


Figure 3. Comparison of QEPPi with other measures of drug-like properties in receiver operating characteristic (ROC) curves. All ROC curves show that the true-positive rate against the false-positive rate describes the differences in performance for classifying compounds as PPI-targeting compounds. The red, black, and blue lines represent the ROC curves for QEPPi, QED, and $1 - \text{QED}$ (QED_inv), respectively. The five blue dots are plotted as points that allowed 0 to 4 violations of RO4. The dashed black line represents a random prediction of the dataset.

2.3. QEPPi Extends the Rule-of-Four

Morelli et al. proposed the “rule-of-four” (RO4) to evaluate PPI inhibitors [8]. This proposal was based on a statistical analysis of 39 PPI inhibitors in 2P2Idb [24] (currently, 2P2Idb is not accessible. Thus, we cannot use data from 2P2Idb). They calculated the general characteristics of the chemical space in which PPI inhibitors differed from FDA-approved drugs. As a result, the RO4 consists of the following four criteria for physicochemical properties:

- MW must be higher than 400;
- ALogP must be higher than 4;
- HBA must be higher than 4;
- The number of rings (RING) must be higher than 4.

Figure 3 shows that we could convert a discrete value index into a continuous value index, as the ROC curve of QEPPi and each point of RO4 were very close to each other. The result suggests that QEPPi is a general extension of the RO4 concept.

The threshold value of QEPPi can be adjusted. We calculated the threshold value of QEPPi (QEPPi scores with a threshold value of 0.5196) such that the F-score was maximized. We then used the iPPI-DB data as a positive sample and the FDA data as a negative sample to compare the discriminative performance of RO4, allowing one violation in the QEPPi. The confusion matrix and F-score results for RO4 and QEPPi are described in Tables 2 and 3.

Table 3 shows that the F-score of QEPPi and RO4 are 0.501 and 0.451, respectively, indicating that QEPPi performs better than RO4. We performed the calculation again using the dataset of 1609 compounds without removing the 13 PPI-targeting compounds, obtaining the F-score of 0.499 and 0.446 for QEPPi and RO4, respectively (see Supplementary Tables S1–S3).

Table 2. Confusion matrix based on RO4 with one violation.

| | Passed | Failed |
|----------|--------|--------|
| positive | 163 | 158 |
| negative | 239 | 1357 |

Table 3. Precision, Recall, and F-score values for one violation of RO4 and QEPPi scores with a threshold value of 0.5196.

| | Precision | Recall | F-Score |
|-------|-----------|--------|---------|
| RO4 | 0.405 | 0.508 | 0.451 |
| QEPPi | 0.379 | 0.735 | 0.501 |

Finally, to compare the classification performance of two different metrics, namely, RO4 (rule-based) and QEPPi (threshold-based), we compared the value of Recall between the same value of Precision and value of Precision between the same value of Recall. The Precision-Recall curve is shown in Figure 4. Because RO4 is rule-based, we plotted the curves for all violations from one to four. As a result, each point of RO4, although not all RO4 points, was plotted on the lower side of the Precision-Recall curve of QEPPi. We also performed the calculation using the dataset of 1609 compounds without removing the 13 PPI-targeting compounds, resulting in AUC values for QEPPi, QED, and $1 - QED$ values of 0.422, 0.134, and 0.238, respectively (see Supplementary Figure S2).

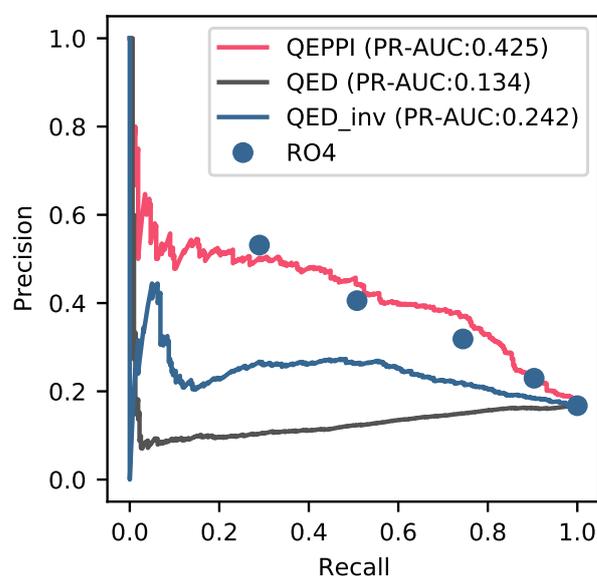


Figure 4. Comparison of QEPPi with RO4 in a Precision-Recall curve. The Precision-Recall curve shows Precision against the Recall value, which describes differences in performance for classifying compounds as PPI-target compounds. The red, black, and blue lines represent the Precision-Recall curves for QEPPi, QED, and $1 - QED$ (QED_inv), respectively. The five blue dots represent the points that allowed 0 to 4 violations of RO4.

3. Discussion

3.1. Advantage of QEPPi

Theoretically, we represent the ideal values for each physicochemical property characteristic of that dataset. This is because the frequency of compounds with that property was highest in that dataset. Therefore, these properties are expected to reflect the nature of the target proteins. Furthermore, because QED is modeled using FDA-approved oral drugs, it is expected to reflect absorption, distribution, metabolism, excretion, and toxicity. In contrast, the dataset used for QEPPi involves many PPI-targeting compounds and does not involve any optimization. Hence, the peak values for all physicochemical properties were higher for QEPPi than those for QED.

The advantage of QEPPi is that it allows model building using only target data. It does not require appropriate negative samples. The performance of machine learning classifiers is poor in problem settings where positive and negative samples are imbalanced [25].

Therefore, QEPPI may be more effective than machine learning models under conditions in which appropriate negative samples are difficult to obtain from public databases.

RO4 is rule-based; therefore, it is nearly impossible to adjust certain threshold values. However, the threshold values of QEPPI developed in this study can be adjusted such that the desired sensitivity and specificity are achieved.

QEPPI indices are primarily intended to be used in the early-stage of PPI drug discovery, which is the seed compound discovery stage. Hence, better discrimination performance is desirable. Figure 4 shows that QEPPI has higher Precision at the same Recall and higher Recall at the same Precision than those for RO4 with one, two, and four violations of RO4. The Precision-Recall AUC values for QEPPI, QED (a measure of oral drug-like properties), and QED_inv (a measure of $1 - \text{QED}$) were 0.425, 0.134, and 0.242, respectively, indicating that among these measures, QEPPI most accurately identified PPI-targeting compounds. The rules of RO4 are based on only 39 PPI inhibitors and, as with RO5, the strict cutoff for each physicochemical property is controversial. For example, a molecular weight of 401 is a pass, whereas 399 is a violation. In fact, Table 1 shows that the peak value for MW is approximately 500 and peak values for ALogP and HBA are slightly higher than 4. This means that many compounds violate the RO4 criteria. For the 1007 PPI-targeting compounds used in the QEPPI model, the results of calculating whether each physicochemical property used in RO4 violates the four criteria are shown in Table 4. Table 4 shows that the violation percentages of WM, ALOP, and HBA were 24.1%, 37.5%, and 34.5%, respectively. For RING, which is a physicochemical property used only for RO4, more than 50% of the compounds violated this property.

Table 4. RO4 violations in the dataset used for QEPPI modeling.

| | MW | ALOGP | HBA | RING |
|----------------|-------|-------|-------|-------|
| violation | 243 | 378 | 347 | 532 |
| no violation | 764 | 629 | 660 | 475 |
| violation rate | 0.241 | 0.375 | 0.345 | 0.528 |

The results above suggest that QEPPI is more useful and suitable compared to the conventional drug discovery indices QED and RO4. Hence our proposal is a useful index of PPI-targeting compounds in designing for early detection of PPI drugs.

Additionally, in further studies, QEPPI can be used as a reward in sequence-based molecular generation models using reinforcement learning such as REINVENT [18,19], and as a condition for sequence-based molecular generation models using conditional Wasserstein generative adversarial networks (WGANs) and Variational Autoencoders (VAEs), such as gcWGAN [26] and CVAE [27], which will enable molecular design with high PPI-targeting compound properties.

3.2. Application of QEPPI to PPI-Targeting Compounds That Are Approved or in Clinical Trials, and Other Small Compounds

In 2020, Shin et al. reported a review of PPI-targeting drug designs. We applied QEPPI to one dataset in this review [15]. The dataset is described as the non-PPI dataset in the review (Soga dataset) [28]. In 2021, Truong et al. explored which physicochemical parameters are necessary for a PPI modulator to become a clinical drug by analyzing the physicochemical properties of small-molecule PPI modulators that are either on the market, in clinical trials, or have been published. They found that PPI modulators currently on the market have a wide range of values for most physicochemical parameters, whereas PPI modulators in clinical trials conform much more closely to standard drug-like parameters, and therefore, a new PPI-specific screening library could be designed. This suggests that when designing new PPI-specific screening libraries, it is necessary to remain within parameters similar to those of standard drugs to obtain clinical candidates [21]. As suggested by the authors, PPI modulators undergoing clinical trials tend to have physicochemical

properties more similar to those of standard drugs than those of PPI modulators currently on the market.

We also applied QEPPi to the above datasets. The distribution of the QEPPi is shown in Figure 5. 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, Figure 5. Although the physicochemical properties of the PPI-targeting compounds registered in iPPI-DB and FDA-approved drugs are different, as shown in Figure 1 and Table 1, the QEPPi modeled from iPPI-DB shows potential to be adapted to more recent PPI modulators.

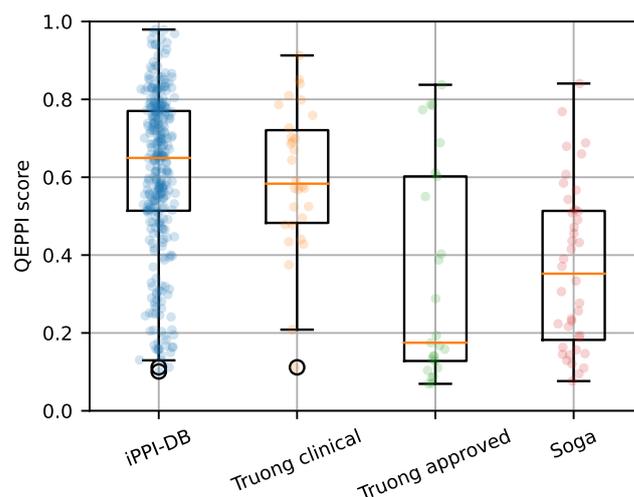


Figure 5. Distribution of QEPPi with respect to compounds in the clinical phase or approved PPI-targeting compounds dataset. The Truong clinical and Truong approved datasets represent clinical and FDA-approved PPI-targeting compound data, respectively. The iPPI-DB and Soga datasets represent positive and negative controls, respectively. The jitter overlaid on the boxplots shows the QEPPi scores for all samples in each data set.

In addition, we also looked at when the PPI-targeted compounds included in the Truong approved data were marketed and when the PPI-targeted compounds included in the Truong clinical data were used in clinical trials. Figure 6 shows the QEPPi of PPI-targeting marketed drugs and compounds in clinical trials within the last 30 years (in detail Supplementary Table S4). Figure 6a shows the PPI-targeting drugs on the market, year the drug was first marketed (as identified in DrugBank), QEPPi value, and target PPI for each drug. PPI-targeting drugs launched in the 1990s showed lower QEPPi scores, whereas drugs marketed more recently tended to have higher QEPPi scores. Figure 6b shows the PPI-targeting compounds in clinical trials, year of the first clinical trial (identified in [ClinicalTrials.gov](https://clinicaltrials.gov), accessed on 15 September 2021), EU Clinical Trials Register (or NIPH Clinical Trial Search in Japan), QEPPi value, and target PPI for each compound. Regardless of the year, the QEPPi scores showed a high transition. This is consistent with the fact that the QEPPi scores of the marketed drugs in Figure 6a exhibited a recent trend toward higher values.

In this study, some PPI-targeting compounds with low QEPPi scores showed small molecular weights compared to those at the peak. A previous study showed that the size and complexity of the binding interface of PPIs varies depending on the target. If the interface is relatively less complex and small, some PPI-targeting compounds with relatively small molecular weights can sufficiently block the binding interface. When the binding interface is more complex, the binding interface tends to be wide, and only a PPI-targeting compound with a large molecular weight can sufficiently block the binding interface [29].

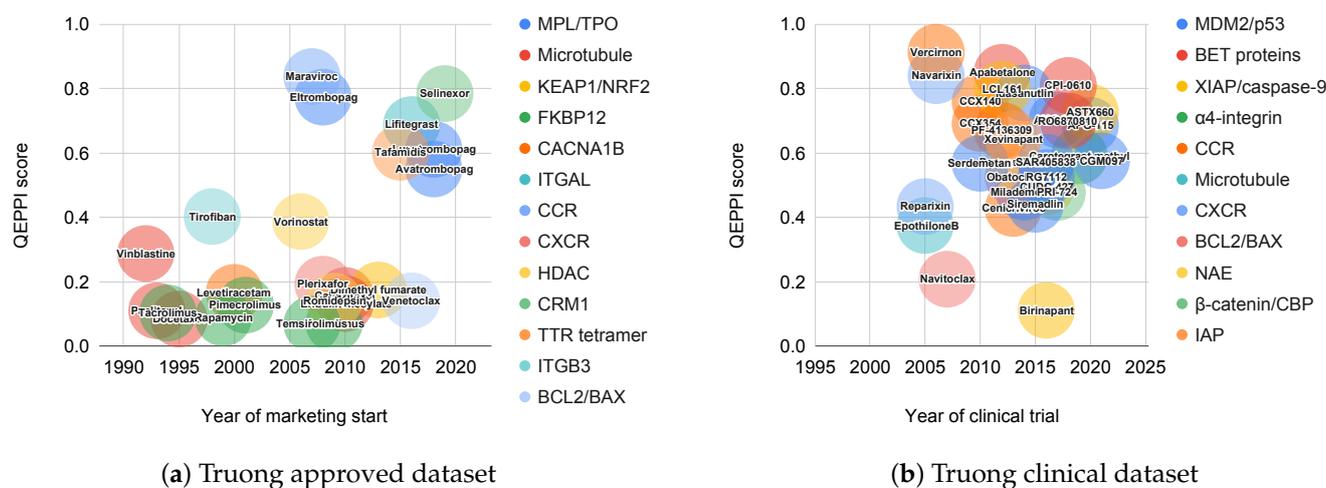


Figure 6. QEPPi scores for PPI-targeting marketed drugs within the past 30 years in Truong approved dataset and compounds in clinical trials within the past 30 years in Truong clinical dataset. QEPPi scores of PPI-targeting marketed drugs in Truong approved dataset (a), and QEPPi scores of PPI-targeting compounds in clinical trials in Truong clinical dataset (b). The color of the circle in each figure indicates the target for that drug or compound (in detail Supplementary Table S4).

Figure 7 shows that there is a difference in the distribution of QEPPi for each PPI family in the iPPI-DB dataset. This result shows that the QEPPi scores of compounds targeting Bromodomain/Histone [29], XIAP/Smac [29], LFA/ICAM [29], and CD4/gp120 [30], which have primary epitopes (such as linear peptide), tend to be higher than those of compounds targeting Bcl2/Bax [29], p53/MDM2 [29], and CD80/CD28 [31], which have secondary epitopes (such as the helix structure). As the LEDGF/IN interface area (400 Å) and transthyretin (TTR) dimer-dimer interface area are much smaller than the interface area of other PPIs [15,29,32], the QEPPi scores of these PPI-targeting compounds tended to be low. Thus, the difference in the complexity of the PPI interface may affect the physico-chemical properties of PPI-targeting compounds such as molecular weight; furthermore, the complexity of the PPI interface is related to the QEPPi score.

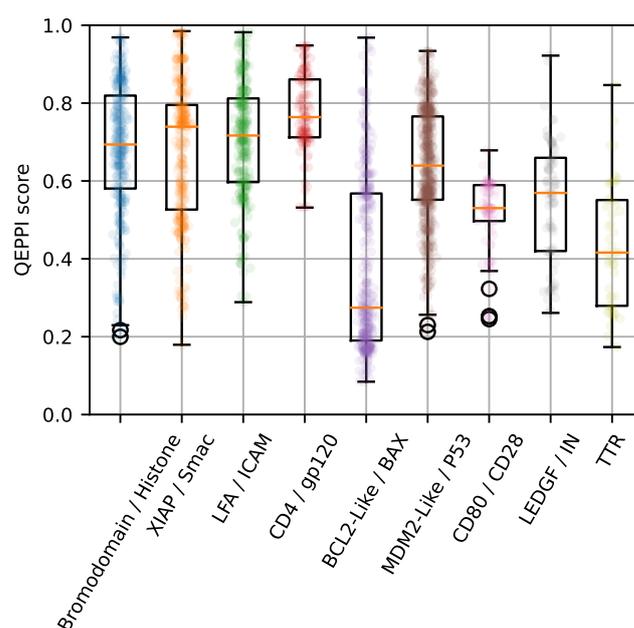


Figure 7. Distribution of QEPPi scores for 9 PPI families with more than 50 compounds targeting each PPI family in iPPI-DB. The jitter overlaid on the boxplots shows the QEPPi scores for all samples in each dataset. Statistics related to this figure are shown in Supplementary Table S5.

Therefore, evaluating average PPI-targeting compounds using the iPPI-DB as a dataset of various types of PPI-targeting compounds would be advantageous. Our further studies will focus on designing indices that are more specific to PPI-targeting compounds, such as the size of the binding interface or PPI family. This is similar to the proposal of QEX. The approach will become feasible as more data are deposited in the database.

Various types of PPI-targeting compounds have been reviewed, and QEPPi scores were calculated for several types of PPI target compounds mentioned in the review published by Mabonga et al. [33]. For example, nine compounds that target chemokine receptors were included in the Truong approved and Truong clinical datasets used in this study, with an average QEPPi score of 0.641 (see Supplementary Table S6). In addition, compounds targeting MDM2/p53, a cancer-related PPI, were also included in the Truong clinical dataset with an average QEPPi score of 0.593 (see Supplementary Table S6). The average QEPPi scores of 79-6 (PubChem CID5721353) targeting BCL6/SMRT and FMP-API-1 targeting AKAP18 δ /PKA were 0.468 and 0.410, respectively (see Supplementary Table S7). In addition, LFA/ICAM, a PPI related to T-cell activation, was included in the data downloaded from iPPI-DB, and the average QEPPi score was 0.706 (see Supplementary Table S5). These results suggest that QEPPi is effective for PPI modulators that have been developed to date. However, the QEPPi scores of approved drugs targeting FKBP12 (e.g., pimecrolimus, tacrolimus, everolimus, rapamycin, temsirolimus) and approved drugs targeting microtubules with a molecular weight greater than 800 (cabazitaxel, docetaxel, eribulin mesylate, paclitaxel, vinblastine) have lower QEPPi scores (see Supplementary Table S6). This is because the iPPI-DB does not include macrocyclic compounds that target FKBP12, and only approximately 4% of the total iPPI-DB of compounds have molecular weights exceeding 800; the QEPPi score of such compounds that deviate from the average is considered low because of the nature of the method. Further studies are needed to expand the compound space that is not covered by iPPI-DB.

To date, COVID-19 has claimed the lives of more than 4.7 million people and infected another 230 million, making it a global pandemic. In response to this critical situation, the development of drugs targeting the etiological agent, SARS-CoV-2, is ongoing. Some researchers are focusing on compounds that target PPIs. One of the most promising PPI targets is the interaction between the SARS-CoV-2 S protein and human angiotensin-converting enzyme 2 receptor [34–36]. Therefore, we also calculated QEPPi for small-molecules targeting PPIs against SARS-CoV/SARS-CoV-2 (see Supplementary Table S8). The median QEPPi score for these compounds was 0.511. Although this is only an example, QEPPi may be effective for host-pathogen PPIs and other PPIs.

3.3. Limitations and Challenges

In this study, the physicochemical properties of the structural alerts were not used in the model building process. As mentioned in the methods, because of the nature of the database, the ratio of compounds in iPPI-DB that do not contain structural alerts is higher than the ratio of compounds in oral FDA-approved drugs which were used in QED that do not contain structural alerts. Therefore, addition of physicochemical properties of ALERTS may bias the model to give undue preference to compounds with low ALERTS values. In the early stages of the actual screening process, it is necessary to filter out structural alerts such as highly reactive functional groups. For example, it is possible to filter out compounds with an “unwanted group” structure described in the supplementary material of the paper published by Brenk et al. [37], which was also used in QED. It is also necessary to filter out other artifacts that do not depend on specific drug-like interactions between molecules and proteins, known as PAINS (pan assay interfering compounds) [38]. During actual operation, these filters should be used in conjunction with the QEPPi score in screening.

In the present QEPPi model, we focused on the physicochemical properties and did not consider the stereo-coordination of the compounds targeting PPI. However, it has been suggested that 3D conformation properties are important for PPI-targeted compounds [29].

It is likely that many PPI-targeted compounds are large and have a high degree of freedom in their bond angles, which may also be characteristic of their stereo-coordination. Therefore, the investigation of methodologies that consider stereo-coordination is one of the important issues to be addressed in the future.

4. Materials and Methods

4.1. Calculation of QEPPi

QEPPi was calculated using essentially the same procedure as used for the original QED, except that it was modeled using compounds curated in the iPPI-DB. We did not use 'ALERTS' among the physicochemical properties, because approximately 58.7% of compounds in the iPPI-DB do not contain a substructure on the list of structural alerts and 22.6% of compounds had only one substructure on the list of structural alerts. Because of the nature of the database, the percentage of compounds without structural alerts is large and may create biases if they are added to the physicochemical properties of the modeling, and thus, these compounds were excluded from the physicochemical properties during modeling. The algorithms used are described below. In the first modeling step, RDKit (2020.09.1) was used to calculate seven molecular physicochemical properties: the molecular weight (MW), LogP value estimated by the Wildman-Crippen method [39] (ALogP), number of hydrogen bond donors (HBD), number of hydrogen bond acceptors (HBA), topological molecular polar surface area (TPSA), number of rotatable bonds (ROTB), and number of aromatic rings (AROM). Table 5 lists the RDKit functions used to calculate these properties.

Table 5. RDKit functions used to calculate the molecular properties used in quantitative estimate of protein-protein interaction targeting drug-likeness (QEPPi) and rule-of-four (RO4).

| Property | RDKit Function |
|----------|---|
| MW | Chem.rdMolDescriptors.CalcExactMolWt |
| ALogP | Chem.Crippen.MolLogP |
| HBD | Chem.rdMolDescriptors.CalcNumHBD |
| HBA | Chem.rdMolDescriptors.CalcNumHBA |
| TPSA | Chem.rdMolDescriptors.CalcTPSA |
| ROTB | Chem.rdMolDescriptors.CalcNumRotatableBonds |
| AROM | Chem.rdMolDescriptors.CalcNumAromaticRings |
| RING | Chem.rdMolDescriptors.CalcNumRings |

A histogram of each property was created and fitted to the asymmetric double sigmoid function $Q(x)$ shown in Equation (1) by implementing the Levenberg-Marquardt algorithm in SciPy (version 1.6.1).

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

All fitting functions ($Q_{MW}(x)$, $Q_{ALogP}(x)$, $Q_{HBD}(x)$, $Q_{HBA}(x)$, $Q_{TPSA}(x)$, $Q_{ROTB}(x)$, and $Q_{AROM}(x)$) were divided by the maximum value and normalized to a maximum value of 1. The normalized function $\tilde{Q}_i(x)$ ($i \in \{MW, ALogP, HBD, HBA, TPSA, ROTB, AROM\}$) was used as the desirability function. Finally, the QEPPi score of compound k was assigned as the weighted geometric mean of all desirability functions [40], as shown in Equation (2).

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

The seven weights were tested from 0 to 1 in increments of 0.25, and the average of the 1000 combinations of weights resulting in the highest Shannon entropy was adopted.

The Shannon entropy of the model was calculated as shown in Equation (3), where n represents the number of compounds used for modeling.

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

4.2. Calculation of QED and RO4

To evaluate the filtering performance of QEPPI, QED and RO4 were calculated and used for comparison. The QED score was calculated using the Chem.QED.qed method in RDKit. RO4 is calculated from four properties: the MW, ALogP, HBA, and number of ring structures (RING). MW, ALogP, and HBA were calculated using the same methods as those used for QEPPI. RING was calculated using the Chem.rdMolDescriptors.CalcNumRings method in RDKit (Table 5).

4.3. Dataset

To create a non-redundant dataset for the QEPPI model, we downloaded 2361 SMILES and other data of compounds registered in iPPI-DB, and 1007 compounds were selected from all clusters individually with the best activities determined by clustering with Bemis-Murcko atomic frameworks [41].

- As a dataset for evaluating QEPPI, 321 compounds were selected from all clusters individually with the best activities from all clusters of compounds that were not used for model building (iPPI-DB dataset).
- As a dataset for small-molecule compounds, we obtained SMILES and other data of compounds known as “DrugBank FDA only” compounds from the catalog of ZINC [42] and removed duplicates by InChI, resulting in 1609 compounds. In addition, 13 compounds that overlapped with the “Truong approved dataset” PPI-targeting drugs (as explained below) were removed, resulting in 1596 compounds (FDA dataset).
- As a dataset for PPI-targeting compounds that are either in the clinical stage or have been approved, 30 PPI-targeting compounds in clinical trial stages and 25 FDA-approved PPI-targeting compounds were obtained from [21] (Truong clinical and Truong approved datasets, respectively).
- As a dataset of non-PPI ligands, which involved known non-redundant protein-ligand complexes evaluated using X-ray crystallography, we obtained 40 PDB IDs of single-molecule ligands obtained from [28] (Soga dataset).

For datasets for which only PDB IDs were available, the IDs were converted to SMILES using PDB’s GraphQL-based API [43].

4.4. Performance Measures

Using the “iPPI-DB dataset” as the positive dataset and “FDA dataset” as the negative dataset, samples scoring above a certain threshold by QEPPI or QED were predicted as positive, and samples scoring below the threshold were predicted as negative. The performance measures used are shown in Equations (4) and (5):

$$\text{TPR} = \text{Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \quad (4)$$

$$\text{FPR} = \frac{\text{FP}}{\text{FP} + \text{TN}} \quad (5)$$

where TP, FP, FN, TN, TPR, and FPR are the numbers of true-positives, false-positives, false-negatives, true-negatives, and true-positive and false-positive ratios, respectively.

Furthermore, the F-score shown in Equation (6) was used to evaluate the discrimination performance, and the precision shown in Equation (7) was used for the Precision-Recall curve (Recall is already shown in Equation (4)).

$$\text{F-score} = \frac{2 \cdot \text{TP}}{2 \cdot \text{TP} + \text{FP} + \text{FN}} \quad (6)$$

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} \quad (7)$$

To evaluate the receiver operating characteristic (ROC) and Precision-Recall curves, the QEPPI threshold was calculated based on all QEPPI scores in the data set, TP, FP, FN, and TN. In Table 6, the threshold was calculated, in which the F-score was maximized.

Table 6. Confusion matrix based on QEPPI scores with a threshold value of 0.5196.

| | Passed | Failed |
|----------|--------|--------|
| positive | 236 | 85 |
| negative | 386 | 1210 |

5. Conclusions

QEPPI is based on the concept of QED, which models the physicochemical properties of a target compound and can quantify the PPI-targeting drug-likeness of interest compounds using PPI inhibitors and stabilizers as the target compound. The metric is useful in the early detection of PPI drugs.

RO4 was originally proposed as a rule-based approach with respect to a statistical analysis of the physicochemical characteristics of actual PPI inhibitors. QED is also based on the distribution data of the physicochemical properties of oral drugs and has gained attention in early-stage drug discovery. However, this method is not suitable for early-stage screening of PPI-targeting compounds because the physicochemical properties of PPI-targeting compounds differ significantly from those of oral drugs. In addition, compared to the rule-based approach of RO4, QEPPI is based on the basic distribution data of physicochemical properties of more PPI-targeting compounds. Unlike rule-based indices, when many parameters of physicochemical properties are ideal, certain unfavorable parameters of properties may still be acceptable, making it an extremely useful index, specifically for early-stage screening of compounds targeting PPIs.

QEPPI may lead to the development of PPI-based drugs along with consequent improvements in the accuracy of QEPPI as more PPI-targeting compounds are registered in the database.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/ijms222010925/s1>.

Author Contributions: Conceptualization, M.O. and T.K.; methodology, M.O. and T.K.; software, T.K.; validation, T.K.; formal analysis, T.K.; investigation, T.K.; resources, M.O. and T.K.; data curation and cleaning, T.K.; writing—original draft preparation, T.K.; writing—review and editing, M.O.; visualization, T.K.; supervision, M.O.; project administration, M.O.; funding acquisition, M.O. Both authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by KAKENHI (20H04280) from the Japan Society for the Promotion of Science (JSPS), ACT-X (JPMJAX20A3) from the Japan Science and Technology Agency (JST), and the Mizuho Foundation for the Promotion of Sciences.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The compound data sets used in this study, implementation code, and easy-to-use environment on Google Colaboratory are available under the MIT license at <https://github.com/ohuelab/QEPPI>, accessed on 30 September 2021.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

| | |
|--------|--|
| QED | Quantitative estimate of drug-likeness |
| PPI | Protein-protein interaction |
| QEPI | Quantitative estimate index for compounds targeting protein-protein interactions |
| RO4 | Rule of four |
| RO5 | Rule of five |
| PPI | Protein-protein interaction |
| FDA | The U.S. Food and Drug Administrations |
| MW | Molecular weight |
| ALogP | LogP value estimated by the Wildman-Crippen method |
| HBD | The number of hydrogen bond donors |
| HBA | The number of hydrogen bond acceptors |
| TPSA | Topological molecular polar surface area |
| ROBT | The number of rotatable bonds |
| AROM | The number of aromatic rings |
| RING | The number of ring structures |
| ALERTS | The number of structural alerts |
| ROC | Receiver operating characteristic |
| AUC | Area under the curve |
| WGAN | Wasserstein generative adversarial network |
| VAE | Variational Autoencoder |
| TTR | Transthyretin |
| PAINS | Pan assay interfering compounds |

References

1. Toogood, P.L. Inhibition of Protein-Protein Association by Small Molecules: Approaches and Progress. *J. Med. Chem.* **2002**, *45*, 1543–1558. [[CrossRef](#)]
2. Arkin, M.R.; Wells, J.A. Small-molecule inhibitors of protein-protein interactions: Progressing towards the dream. *Nat. Rev. Drug Discov.* **2004**, *3*, 301–317. [[CrossRef](#)] [[PubMed](#)]
3. Dev, K.K. Making protein interactions druggable: Targeting PDZ domains. *Nat. Rev. Drug Discov.* **2004**, *3*, 1047–1056. [[CrossRef](#)]
4. Jin, L.; Wang, W.; Fang, G. Targeting Protein-Protein Interaction by Small Molecules. *Annu. Rev. Pharmacol. Toxicol.* **2014**, *54*, 435–456. [[CrossRef](#)]
5. Ivanov, A.A.; Khuri, F.R.; Fu, H. Targeting protein-protein interactions as an anticancer strategy. *Trends Pharmacol. Sci.* **2013**, *34*, 393–400. [[CrossRef](#)] [[PubMed](#)]
6. Lipinski, C.A.; Lombardo, F.; Dominy, B.W.; Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv. Drug Deliv. Rev.* **1997**, *23*, 3–25. [[CrossRef](#)]
7. Lipinski, C.A. Lead- and drug-like compounds: The rule-of-five revolution. *Drug Discov. Today Technol.* **2004**, *1*, 337–341. [[CrossRef](#)] [[PubMed](#)]
8. Morelli, X.; Bourgeas, R.; Roche, P. Chemical and structural lessons from recent successes in protein-protein interaction inhibition (2P2I). *Curr. Opin. Chem. Biol.* **2011**, *15*, 475–481. [[CrossRef](#)] [[PubMed](#)]
9. Sperandio, O.; Reynès, C.H.; Camproux, A.-C.; Villoutreix, B.O. Rationalizing the chemical space of protein-protein interaction inhibitors. *Drug Discov. Today* **2010**, *15*, 220–229. [[CrossRef](#)]
10. Wishart, D.S.; Feunang, Y.D.; Guo, A.C.; Lo, E.J.; Marcu, A.; Grant, J.R.; Sajed, T.; Johnson, D.; Li, C.; Sayeeda, Z.; et al. DrugBank 5.0: A major update to the DrugBank database for 2018. *Nucleic Acids Res.* **2018**, *46*, D1074–D1082. [[CrossRef](#)]
11. Bickerton, G.R.; Paolini, G. V.; Besnard, J.; Muresan, S.; Hopkins, A.L. Quantifying the chemical beauty of drugs. *Nat. Chem.* **2012**, *4*, 90–98. [[CrossRef](#)]
12. De Cao, N.; Kipf, T. MolGAN: An implicit generative model for small molecular graphs. In Proceedings of the ICML'18 Workshop on Theoretical Foundations and Applications of Deep Generative Models, Stockholm, Sweden, 10–15 July 2018; pp. 1–11.
13. Putin, E.; Asadulaev, A.; Ivanenkov, Y.; Aladinskiy, V.; Sanchez-Lengeling, B.; Aspuru-Guzik, A.; Zhavoronkov, A. Reinforced Adversarial Neural Computer for de Novo Molecular Design. *J. Chem. Inf. Model.* **2018**, *58*, 1194–1204. [[CrossRef](#)] [[PubMed](#)]
14. Warner, K.D.; Hajdin, C.E.; Weeks, K.M. Principles for targeting RNA with drug-like small molecules. *Nat. Rev. Drug Discov.* **2018**, *17*, 547–558. [[CrossRef](#)] [[PubMed](#)]
15. Shin, W.-H.; Kumazawa, K.; Imai, K.; Hirokawa, T.; Kihara, D. Current challenges and opportunities in designing protein-protein interaction targeted drugs. *Adv. Appl. Bioinform. Chem.* **2020**, *13*, 11–25. [[CrossRef](#)] [[PubMed](#)]
16. Mochizuki, M.; Suzuki, S.D.; Yanagisawa, K.; Ohue, M.; Akiyama, Y. QEX: Target-specific druglikeness filter enhances ligand-based virtual screening. *Mol. Divers.* **2019**, *23*, 11–18. [[CrossRef](#)]

17. Limmer, M.A.; Burken, J.G. Plant Translocation of Organic Compounds: Molecular and Physicochemical Predictors. *Environ. Sci. Technol. Lett.* **2014**, *1*, 156–161. [[CrossRef](#)]
18. Olivecrona, M.; Blaschke, T.; Engkvist, O.; Chen, H. Molecular de-novo design through deep reinforcement learning. *J. Cheminform.* **2017**, *9*, 48. [[CrossRef](#)]
19. Blaschke, T.; Arús-Pous, J.; Chen, H.; Margreitter, C.; Tyrchan, C.; Engkvist, O.; Papadopoulos, K.; Patronov, A. REINVENT 2.0: An AI Tool for de Novo Drug Design. *J. Chem. Inf. Model.* **2020**, *60*, 5918–5922. [[CrossRef](#)]
20. 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), Melbourne, Australia, 13–15 October 2021. [[CrossRef](#)]
21. Truong, J.; George, A.; Holien, J.K. Analysis of physicochemical properties of protein-protein interaction modulators suggests stronger alignment with the “Rule-of-Five”. *RSC Med. Chem.* **2021**. [[CrossRef](#)]
22. Oughtred, R.; Rust, J.; Chang, C.; Breitkreutz, B.-J.; Stark, C.; Willems, A.; Boucher, L.; Leung, G.; Kolas, N.; Zhang, F.; et al. The BioGRID database: A comprehensive biomedical resource of curated protein, genetic, and chemical interactions. *Protein Sci.* **2021**, *30*, 187–200. [[CrossRef](#)]
23. Torchet, R.; Druart, K.; Ruano, L.C.; Moine-Franel, A.; Borges, H.; Doppelt-Azeroual, O.; Brancotte, B.; Mareuil, F.; Nilges, M.; Ménager, H.; et al. The iPPI-DB initiative: A community-centered database of protein–protein interaction modulators. *Bioinformatics* **2021**, *37*, 89–96. [[CrossRef](#)]
24. Basse, M.-J.; Betzi, S.; Morelli, X.; Roche, P. 2P2Idb v2: Update of a structural database dedicated to orthosteric modulation of protein–protein interactions. *Database* **2016**, *2016*, baw007. [[CrossRef](#)]
25. Xu, C.; Jackson, S.A. Machine learning and complex biological data. *Genome Biol.* **2019**, *20*, 76. [[CrossRef](#)] [[PubMed](#)]
26. Karimi, M.; Zhu, S.; Cao, Y.; Shen, Y. De Novo Protein Design for Novel Folds Using Guided Conditional Wasserstein Generative Adversarial Networks. *J. Chem. Inf. Model.* **2020**, *60*, 5667–5681. [[CrossRef](#)] [[PubMed](#)]
27. Lim, J.; Ryu, S.; Kim, J.W.; Kim, W.Y. Molecular generative model based on conditional variational autoencoder for de novo molecular design. *J. Cheminform.* **2018**, *10*, 31. [[CrossRef](#)] [[PubMed](#)]
28. Soga, S.; Shirai, H.; Kobori, M.; Hirayama, N. Use of Amino Acid Composition to Predict Ligand-Binding Sites. *J. Chem. Inf. Model.* **2007**, *47*, 400–406. [[CrossRef](#)] [[PubMed](#)]
29. Arkin, M.R.; Tang, Y.; Wells, J.A. Small-molecule inhibitors of protein-protein interactions: Progressing toward the reality. *Chem. Biol.* **2014**, *21*, 1102–1114. [[CrossRef](#)] [[PubMed](#)]
30. Biorn, A.C.; Cocklin, S.; Madani, N.; Si, Z.; Ivanovic, T.; Samanen, J.; Van Ryk, D.I.; Pantophlet, R.; Burton, D.R.; Freire, E.; et al. Mode of Action for Linear Peptide Inhibitors of HIV-1 gp120 Interactions. *Biochemistry* **2004**, *43*, 1928–1938. [[CrossRef](#)]
31. Srinivasan, M.; Lu, D.; Eri, R.; Brand, D.D.; Haque, A.; Blum, J.S. CD80 binding polyproline helical peptide inhibits T cell activation. *J. Biol. Chem.* **2005**, *280*, 10149–10155. [[CrossRef](#)] [[PubMed](#)]
32. Bulawa, C.E.; Connelly, S.; DeVit, M.; Wang, L.; Weigel, C.; Fleming, J.A.; Packman, J.; Powers, E.T.; Wiseman, R.L.; Foss, T.R.; et al. Tafamidis, a potent and selective transthyretin kinetic stabilizer that inhibits the amyloid cascade. *Proc. Natl. Acad. Sci. USA* **2012**, *109*, 9629–9634. [[CrossRef](#)]
33. Mabonga, L.; Kappo, A.P. Protein-protein interaction modulators: Advances, successes and remaining challenges. *Biophys. Rev.* **2019**, *11*, 559–581. [[CrossRef](#)] [[PubMed](#)]
34. Chang, C.K.; Lin, S.M.; Satange, R.; Lin, S.C.; Sun, S.C.; Wu, H.Y.; Kehn-Hall, K.; Hou, M.H. Targeting protein-protein interaction interfaces in COVID-19 drug discovery. *Comput. Struct. Biotechnol. J.* **2021**, *19*, 2246–2255. [[CrossRef](#)] [[PubMed](#)]
35. Kao, R.Y.; Tsui, W.H.W.; Lee, T.S.W.; Tanner, J.A.; Watt, R.M.; Huang, J.-D.; Hu, L.; Chen, G.; Chen, Z.; Zhang, L.; et al. Identification of Novel Small-Molecule Inhibitors of Severe Acute Respiratory Syndrome-Associated Coronavirus by Chemical Genetics. *Chem. Biol.* **2004**, *11*, 1293–1299. [[CrossRef](#)] [[PubMed](#)]
36. Lin, S.M.; Lin, S.C.; Hsu, J.N.; Chang, C.K.; Chien, C.M.; Wang, Y.S.; Wu, H.Y.; Jeng, U.S.; Kehn-Hall, K.; Hou, M.H. Structure-Based Stabilization of Non-native Protein-Protein Interactions of Coronavirus Nucleocapsid Proteins in Antiviral Drug Design. *J. Med. Chem.* **2020**, *63*, 3131–3141. [[CrossRef](#)] [[PubMed](#)]
37. Brenk, R.; Schipani, A.; James, D.; Krasowski, A.; Gilbert, I.H.; Frearson, J.; Wyatt, P.G. Lessons Learnt from Assembling Screening Libraries for Drug Discovery for Neglected Diseases. *ChemMedChem* **2008**, *3*, 435–444. [[CrossRef](#)] [[PubMed](#)]
38. Baell, J.; Walters, M.A. Chemistry: Chemical con artists foil drug discovery. *Nature* **2014**, *513*, 481–483. [[CrossRef](#)]
39. Wildman, S.A.; Crippen, G.M. Prediction of Physicochemical Parameters by Atomic Contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 868–873. [[CrossRef](#)]
40. Harrington, E.C. The desirability function. *Ind. Qual. Control* **1965**, *21*, 494–498.
41. Bemis, G.W.; Murcko, M.A. The Properties of Known Drugs. 1. Molecular Frameworks. *J. Med. Chem.* **1996**, *39*, 2887–2893. [[CrossRef](#)]
42. Irwin, J.J.; Tang, K.G.; Young, J.; Dandarchuluun, C.; Wong, B.R.; Khurelbaatar, M.; Moroz, Y.S.; Mayfield, J.; Sayle, R.A. ZINC20—A Free Ultralarge-Scale Chemical Database for Ligand Discovery. *J. Chem. Inf. Model.* **2020**, *60*, 6065–6073. [[CrossRef](#)]
43. Burley, S.K.; Bhikadiya, C.; Bi, C.; Bittrich, S.; Chen, L.; Crichlow, G.V.; Christie, C.H.; Dalenberg, K.; Di Costanzo, L.; Duarte, J.M.; et al. RCSB Protein Data Bank: Powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res.* **2021**, *49*, D437–D451. [[CrossRef](#)] [[PubMed](#)]