In silico design of novel proton-pump inhibitors with reduced adverse effects

Xiaoyi Li[1], Hong Kang[2], Wensheng Liu[3], Sarita Singhal[3], Na Jiao[1], Yong Wang[4], Lixin Zhu (✉)[3,5], Ruixin Zhu (✉)[1]

[1]Department of Gastroenterology, Shanghai East Hospital, School of Life Sciences and Technology, Tongji University, Shanghai 200092, China; [2]School of Biomedical Informatics, The University of Texas Health Science Center at Houston, 7000 Fannin St, Houston, TX 77030, USA; [3]Digestive Diseases and Nutrition Center, Department of Pediatrics, The State University of New York at Buffalo, Buffalo, NY 14260, USA; [4]Basic Medical College, Beijing University of Chinese Medicine, Beijing 100029, China; [5]Genome, Environment and Microbiome Community of Excellence, The State University of New York at Buffalo, Buffalo, NY 14214, USA

Abstract The development of new proton-pump inhibitors (PPIs) with less adverse effects by lowering the pKa values of nitrogen atoms in pyrimidine rings has been previously suggested by our group. In this work, we proposed that new PPIs should have the following features: (1) number of ring II = number of ring I + 1; (2) preferably five, six, or seven-membered heteroatomic ring for stability; and (3) 1 < pKa1 < 4. Six molecular scaffolds based on the aforementioned criteria were constructed, and R groups were extracted from compounds in extensive data sources. A virtual molecule dataset was established, and the pKa values of specific atoms on the molecules in the dataset were calculated to select the molecules with required pKa values. Drug-likeness screening was further conducted to obtain the candidates that significantly reduced the adverse effects of long-term PPI use. This study provided insights and tools for designing targeted molecules *in silico* that are suitable for practical applications.

Keywords proton-pump inhibitor; adverse effect; pharmacological mechanism; toxicological mechanism; pKa calculation

Introduction

Proton-pump inhibitors (PPIs) are benzimidazoles or their analogs that directly target H^+K^+-ATPase proton pump for gastric acid secretion. PPIs are widely used in the treatment of peptic ulcer disease, *Helicobacter pylori* infection, gastroesophageal reflux, nonsteroidal anti-inflammatory drug-induced gastrointestinal lesions, and Zollinger–Ellison syndrome [1]. Since the introduction of the first PPI omeprazole in 1988 [2], several other PPIs, such as lansoprazole, pantoprazole, esomeprazole, rabeprazole, tenatoprazole, and leminoprazole, have been successively developed. PPIs are widely used and relatively safe drugs for the treatment of gastric acid related diseases due to their significant curative effects attributed to potent inhibition of gastric acid secretion and few adverse reactions. However, reports on the adverse effects of long-term PPI use have increased recently. This condition has led to a new round of research on the mechanisms underlying those adverse effects. Recent studies have postulated two possible mechanisms that were primarily based on clinical trials or on experimental biology that focused on a certain target or pathway. The first mechanism suggested that the long-term use of PPIs led to the increase in gastric pH and the risk of *H. pylori* infection that aggravated adverse effects, such as inflammation [3]. The second mechanism suggested that PPIs were activated in the environment of lysosomes (pH 5) and affected the activities of lysosomal enzymes, which weakened the immune system and induced adverse effects, such as cancer or infectious diseases [4]. The two studies have only focused on few diseases, such as cancer and infections. Many adverse effects, such as abdominal pain, diarrhea, constipation, emotional disturbance, distractibility, anemia, and angioedema, were reported that were still not explained with the abovementioned mechanisms [1,5–7]. To analyze the side effects of PPIs, we investigated the gene expression profiles of rat liver tissue under PPI perturbation and we established the relationships among PPIs, their adverse

effects, and the involved genes [8]. We determined that the PPIs enriched in the acidic organelles of cells, became activated, and inhibited V-ATPase and hydrolytic enzymes in those acid organelles, which blocked the antigen presenting process and the synthesis and excretion of cytokines, complements, and blood coagulation factors. On this basis, we presented *in silico* design for novel PPI molecules with reduced adverse effects.

Materials and methods

Extraction of R groups

The extraction of fragments (R groups) from small molecules in various databases was performed with the method of the breaking of retrosynthetically interesting chemical substructures (rdkit.Chem.BRICS package in RDkit package) [9] by breaking the synthetically accessible bonds to fragment these small molecules. The generated fragments were merged, and the frequency of these R groups was calculated after the redundancy was removed.

Molecule enumeration

The enumeration was conducted to derive all the possible combinations within the molecular scaffolds of benzimidazoles and abovementioned R groups. All the attachment points of molecular scaffolds were bonded with one random R group to generate a new set of candidate molecules.

Calculation of pKa value

The pKa values were calculated by pKaPlugin in the ChemAxon Marvin toolkit [10]. The candidates were screened by $1 < pKa1 < 4$ of the nitrogen atom on its pyridine ring.

Drug-likeness screening

A drug-likeness screening was performed based on the "Rule of Five" as follows: (1) less than 5 hydrogen bond donors; (2) less than 10 hydrogen bond receptors; (3) less than 500 Dalton molecular mass; (4) greater than 5 lipid-water partition coefficients [11]. RDKit was applied in the calculation. The candidate molecules that complied with these rules were used for screening.

All the scripts of core methods in this study are available at Supplementary Codes of Methods.

Results and discussion

We first generated a virtual data set of benzimidazole molecules based on the knowledge of PPI pharmacology and toxicology. These molecules were screened by the pKa value of their nitrogen atoms on pyridine ring. Subsequently, the drug-likeness evaluation was conducted, and those passed the evaluation were considered as the candidates for new PPIs.

Generation of virtual data set of benzimidazole molecules

The design of benzimidazole molecular scaffolds was enlightened by PPI pharmacology. PPIs are prodrugs that require activation in an acidic environment. Activated PPIs can bind to the alpha region of the H^+K^+-ATPases on secretory capillaries and can form disulfide bonds with cysteine residues that inhibit the activities of those acid secretion pumps (Fig. 1) [12]. Stable Ring I and Ring II successively formed after PPI activation under acidic environment, which were critical for the drug efficacy. Considering the one more atom constantly existed in Ring II than in Ring I, PPIs were preferred to be 5 or 6 or 7-membered heteroatomic ring for their stability.

In addition, with regard to PPI toxicology, pH of the tube cavity on gastric wall is approximately 1, pH around acid organelles, such as lysosome, is approximately 5, and pKa1 (Fig. 2; pKa of nitrogen atom on pyridine ring) of current PPIs is close to 4. However, there are some factors that make PPIs enriched in weak acid environment before playing its effective role, which can lead to series of adverse effects. These factors include the following: (1) PPIs should go through the blood circulation before reaching gastric parietal cells and getting close to their target proton pump; (2) the activation of nitrogen atom on its pyridine ring is a typical reversible reaction; (3) PPIs form irreversible covalent bond with their target proton pump, which enable PPIs to reach an "irreversible" chemical equilibrium in other weak acid environments before reaching their target cell.

On this basis, we proposed a clear target to design new PPIs by modifying the structure of the current PPIs to decrease its pKa1, which increases the acidity of the nitrogen atom on its pyridine ring. New PPI molecules should have the following characteristics:

(1) Number Ring II = Number Ring I + 1

(2) Prefer to be 5 or 6 or 7-membered heteroatomic ring for its stability

(3) $1 < pKa1 < 4$

Under the same conditions (pH, temperature, etc.), the pKa value of a certain atom can be affected by the substituents on or near the atom. However, no substituent can be added to the nitrogen atom on its pyrimidine ring in PPI, which indicates that the modification should focus on the nearing sites. Considering the structure of several existing benzimidazoles, such as tenatoprazole, we designed six molecular scaffolds, as illustrated in Fig. 3.
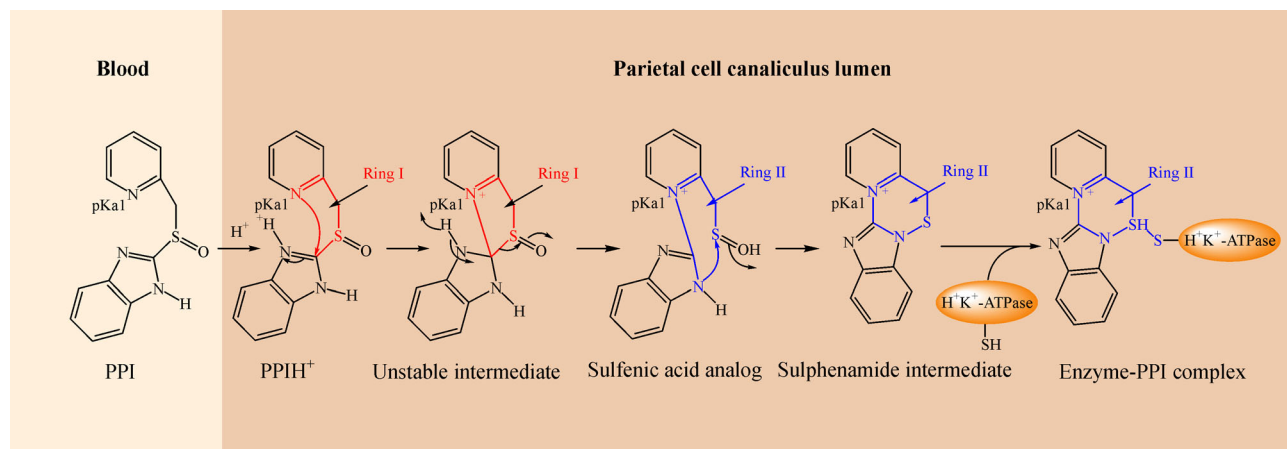
Fig. 1 Key steps of PPI pharmacology. PPIs are made as a prodrug. PPIs are not reactive at neutral pH. PPIs are hydrophobic and freely move across membranes. PPIs become protonated, carry a positive charge, and cannot go across plasma membranes once they reach an acidic environment. Protonation triggers a 3-step transformation that results in the sulfinamide form of PPI, which is the active form of the drug. Sulfinamide forms a di-sulfide bond with another chemical with a free thio group. The covalent bond formation between PPI and H, K inactivates H,K-ATPase and inhibits acid secretion.
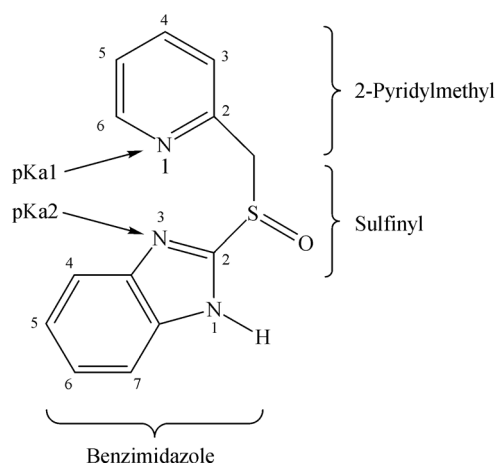

Fig. 2 Structure of PPI. PPI can be split into two parts. The first part is the substituted pyridine, and the second part is the substituted benzimidazoles (or substituted imidazopyridines in several cases).

Extraction of R groups

To generate potential candidate molecules in comprehensive aspects, we collected and organized massive compound structure data from multiple databases for the extraction of R groups, which included DrugBank [13], TCM Database@Taiwan [14], TTD [15], GDB [16,17], HMDB [18], KEGG COMPOUND [19], BitterDB [20], CancerResource [21], T3DB [22,23], and FooDB [24]. After redundancy removal, we calculated the frequency for the remaining R groups. Next, the attachment point of each R group was marked and was bonded to six molecular scaffolds. Through direct downloading or manual collection, the molecular structures from the 10 abovementioned databases were derived and R groups were split. The representative results are illustrated in Fig. 4, and the complete results can be found at the corresponding authors' website. Please contact us if you are interested.
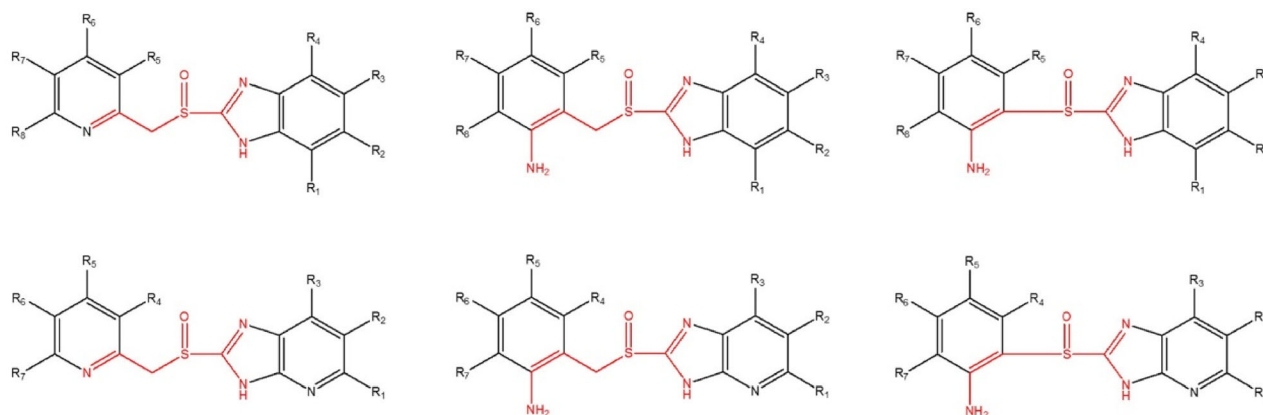

Fig. 3 Designed molecular scaffolds. Six different molecular scaffolds of PPIs are designed based on the analysis on PPIs pharmacology and toxicology.
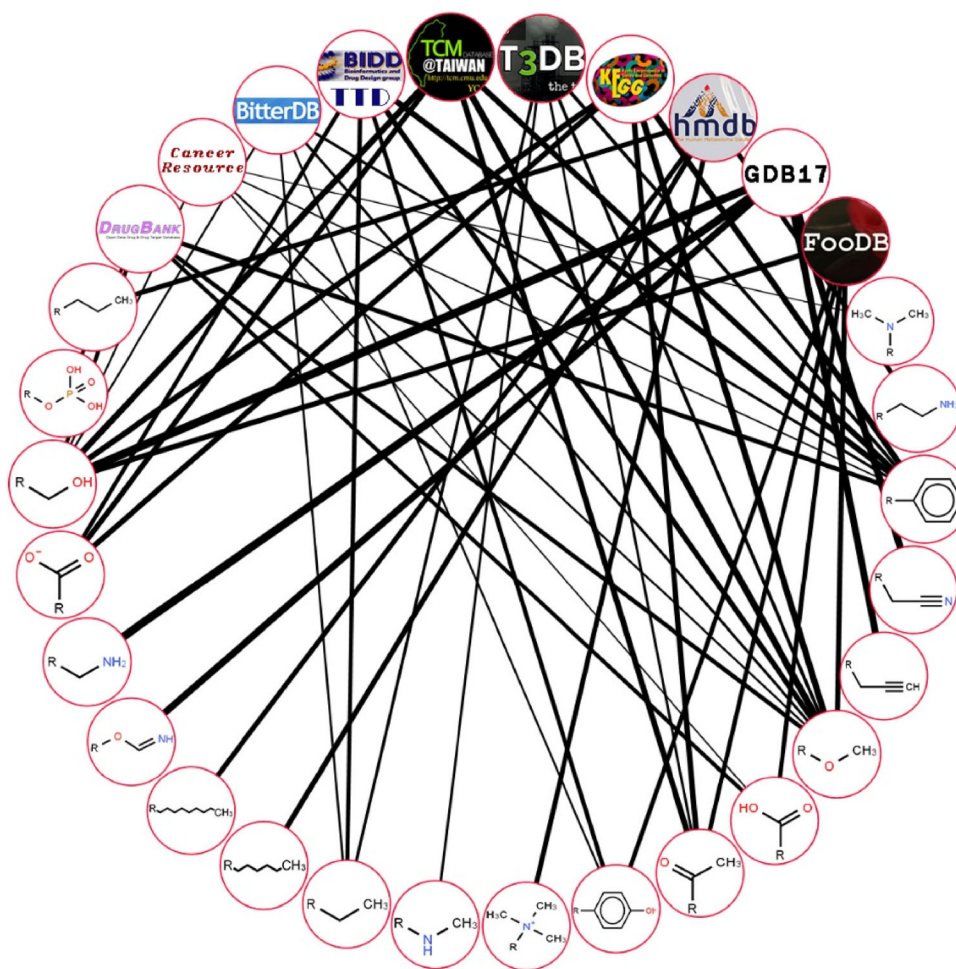
Fig. 4 Relationship network of the first five R groups in the databases. The lines indicate from which database a certain molecule is generated, and the thickness of the line reflects the log value of the number of certain molecule in the corresponding database.

Molecule enumeration

Full enumeration was performed on all possible combinations within the six molecular scaffolds, and different sets of R groups were derived from 10 different databases. Ten virtual data sets of benzimidazole molecules were generated and can be found at the corresponding authors' website. Please contact us if you are interested.

pKa prediction and molecule screening

On the basis of knowledge of PPI pharmacology and toxicology, the pKa value of nitrogen atom on pyridine ring is the key factor for PPIs to activate in acid organelles and induce adverse effects. Therefore, the R groups on PPI molecular scaffolds can be replaced to decrease those adverse effects caused by long-term use of PPIs and to decrease the pKa value of the nitrogen atom on its pyridine ring. Therefore, the pKa value prediction was applied to all virtual molecules.

Currently available pKa prediction tools include ACD/ pKa DB [10], Epik [25], Jaguar [26], Pallas pKalc [27], Pipeline Pilot [28], and Marvin [10]. They are all empirical methods that involve linear free-energy relationships [25], quantitative structure-property relationships (QSPR) [29,30], and database lookup. Marvin belongs to QSPR and can calculate the pKa value of a certain atom on molecules, which is easy to expand and customize through its API. In this study, Marvin was adopted for predicting the pKa1 values and for screening all virtual benzimidazole molecules.

After prediction, all molecules were screened by $1 < pKa1 < 4$. The remaining molecules were collected for subsequent study.

Fig. 5 illustrates the pKa1 distributions of virtual molecules. The detailed information for virtual molecules derived from the first three R groups of different databases can be found at the corresponding authors' website. Please contact us if you are interested. The results significantly vary between different data sources, which indicate the necessity to be as comprehensive and complete as possible to collect molecules from different databases.
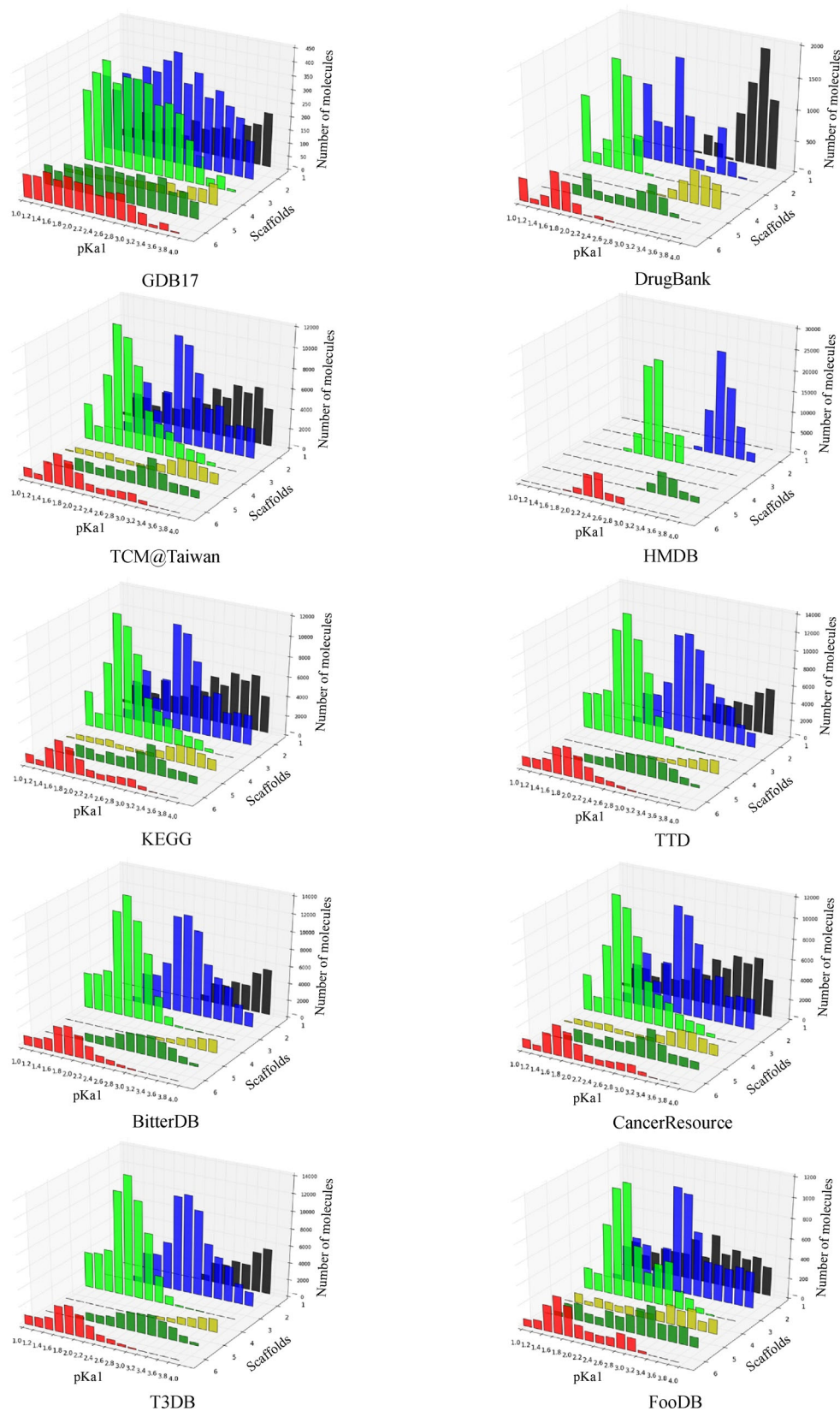
Fig. 5 pKa1 distributions of virtual molecules generated from R groups of 10 different databases. The pKa1 distribution significantly varies in different R group databases and scaffolds.

Drug-likeness screening

Since Lipinski *et al.* [11] established the first and renowned "Rule of Five" drug-likeness rule in 1997, a series of similar rules, known as "rules of thumb," has been successively set. Typical rules include rapid elimination of swill (REOS) by Walters *et al.* [31] in 1998 and quantitative estimate of drug-likeness index by Hopkins *et al.* [32] in 2012. To improve the accuracy of prediction, many machine learning approaches have also been employed to fit the complex relationships within various properties and drugs in recent years [33].

We adopted Lipinski's "Rule of Five" for drug-likeness screening in our work by considering the following facts: (1) this rule is generally recognized and widely used; (2) the disadvantage of relatively few molecule properties can be compensated by later ADMET theoretical evaluation, which incorporates many properties; (3) The results can be repeatable by other researchers because many commercial and open-source software are embedded in this rule. The new candidate PPI molecule data set that is provided to effectively decrease the adverse effect induced by long-term therapy is listed in Supplementary Table S1. The scripts of "Rule of Three (Lead-like drugs)" and REOS are also provided as an extension to ROF application for researchers at Supplementary Codes of Methods.

The pKa value distributions of virtual molecules with different scaffolds (Fig. 6A) demonstrate that the differences in scaffolds have an effect on the pKa value of molecules to some extent, in which the pKa value for scaffolds 3, 6 generated molecules are relatively low, and for 1, 4, 5 are relatively high. Scaffolds 2, 3 generated many molecules that meet the requirements of

$1 < pKa1 < 4$ and "Rule of Five." Fig. 6B demonstrates that scaffolds 1, 2, 3, 4, 5, and 6 generated molecules that have similar molecular structures that are mainly reflected in the PC2 axis. In addition, all molecules can be divided into two major parts in the PC1 axis, which is caused by difference of the R groups in the molecules.

Conclusions

We designed new PPIs based on their pharmacological and toxicological mechanisms. We first constructed six molecular scaffolds that complied with three abovementioned criteria. Next, R groups were extracted from compound molecules in 10 different databases, such as DrugBank, TCM Database@Taiwan, and GDB. On this basis, we employed the "virtual structure generation" technique to establish a virtual molecule data set. Subsequently, the pKa values of specific atoms on the generated molecules were calculated to select those molecules with required pKa values. Finally, drug-likeness screening was conducted, in which we obtained the remaining molecules that can be the new PPI candidates to significantly reduce the adverse effects. Although these new PPI candidates with reduced adverse effects required experimental validation, we showed that they can be achieved theoretically by lowering the pKa values of nitrogen atoms on their pyrimidine ring. The study also provided insights and tools for designing the targeted molecules *in silico* that were suitable in practical applications. Considering that the annual sales of PPIs have exceeded $20 billion on the global market [34], this work may have significant effect on global economy. All the scripts and core functions of these methods in this study are available at Supplementary Codes of Methods.
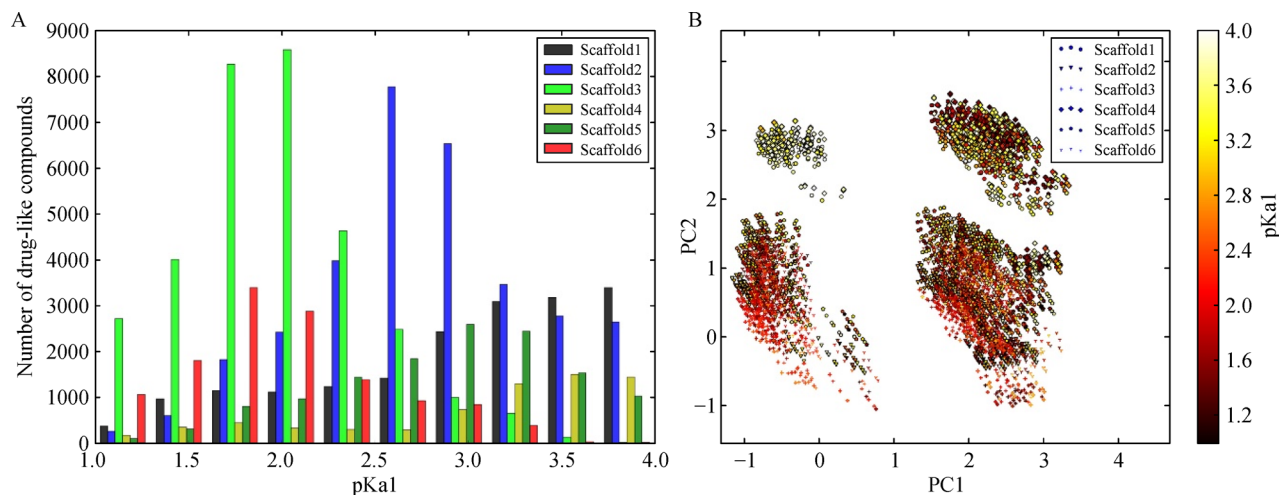

Fig. 6 Distributions of pKa values and structural similarity of the generated virtual molecules from different scaffolds. (A) The distributions of pKa values of nitrogen atom on pyrimidine ring. (B) The PCA analysis on the FingerPrints of 115 511 virtual molecules. FingerPrints are described by MACCS from RDkit, and different colors indicate the pKa value.

Acknowledgements

Compliance with ethics guidelines

Xiaoyi Li, Hong Kang, Wensheng Liu, Sarita Singhal, Na Jiao, Yong Wang, Lixin Zhu, and Ruixin Zhu declare that they have no conflict of interest. This manuscript does not contain any studies with human or animal subjects.

Electronic Supplementary Material Supplementary material is available in the online version of this article at https://doi.org/ 10.1007/s11684-018-0630-3 and is accessible for authorized users.

References

1. Martín de Argila C. Safety of potent gastric acid inhibition. Drugs 2005; 65(Suppl 1): 97–104

2. Hetzel DJ, Dent J, Reed WD, Narielvala FM, Mackinnon M, McCarthy JH, Mitchell B, Beveridge BR, Laurence BH, Gibson GG, Grant AK, Shearman DJC, Whitehead R, Buckle PJ. Healing and relapse of severe peptic esophagitis after treatment with omeprazole. Gastroenterology 1988; 95(4): 903–912

3. Hagiwara T, Mukaisho K, Nakayama T, Sugihara H, Hattori T. Long-term proton pump inhibitor administration worsens atrophic corpus gastritis and promotes adenocarcinoma development in Mongolian gerbils infected with *Helicobacter pylori*. Gut 2011; 60 (5): 624–630

4. Liu W, Baker SS, Trinidad J, Burlingame AL, Baker RD, Forte JG, Virtuoso LP, Egilmez NK, Zhu L. Inhibition of lysosomal enzyme activities by proton pump inhibitors. J Gastroenterol 2013; 48(12): 1343–1352

5. Polimeni G, Cutroneo P, Gallo A, Gallo S, Spina E, Caputi AP. Rabeprazole and psychiatric symptoms. Ann Pharmacother 2007; 41(7): 1315–1317

6. Sarzynski E, Puttarajappa C, Xie Y, Grover M, Laird-Fick H. Association between proton pump inhibitor use and anemia: a retrospective cohort study. Dig Dis Sci 2011; 56(8): 2349–2353

7. Bowlby HA, Dickens GR. Angioedema and urticaria associated with omeprazole confirmed by drug rechallenge. Pharmacotherapy 1994; 14(1): 119–122

8. Wu D, Qiu T, Zhang Q, Kang H, Yuan S, Zhu L, Zhu R. Systematic toxicity mechanism analysis of proton pump inhibitors: an *in silico* study. Chem Res Toxicol 2015; 28(3): 419–430

9. Landrum G. Rdkit: a software suite for cheminformatics, computational chemistry, and predictive modeling. 2013

10. Settimo L, Bellman K, Knegtel RM. Comparison of the accuracy of experimental and predicted pKa values of basic and acidic compounds. Pharm Res 2014; 31(4): 1082–1095

11. Lipinski CA, Lombardo F, Dominy BW, Feeney PJ. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. Adv Drug Deliv Rev 2001; 46(1-3): 3–26

12. Olbe L, Carlsson E, Lindberg P. A proton-pump inhibitor expedition: the case histories of omeprazole and esomeprazole. Nat Rev Drug Discov 2003; 2(2): 132–139

13. Law V, Knox C, Djoumbou Y, Jewison T, Guo AC, Liu Y, Maciejewski A, Arndt D, Wilson M, Neveu V, Tang A, Gabriel G, Ly C, Adamjee S, Dame ZT, Han B, Zhou Y, Wishart DS. DrugBank 4.0: shedding new light on drug metabolism. Nucleic Acids Res 2014; 42(Database issue D1): D1091–D1097

14. Chen CY. TCM Database@Taiwan: the world's largest traditional Chinese medicine database for drug screening *in silico*. PLoS One 2011; 6(1): e15939

15. Zhu F, Shi Z, Qin C, Tao L, Liu X, Xu F, Zhang L, Song Y, Liu X, Zhang J, Han B, Zhang P, Chen Y. Therapeutic target database update 2012: a resource for facilitating target-oriented drug discovery. Nucleic Acids Res 2012; 40(Database issue): D1128–D1136

16. Reymond JL, Ruddigkeit L, Blum L, van Deursen R. The enumeration of chemical space. Wiley Interdiscip Rev Comput Mol Sci 2012; 2(5): 717–733

17. Ruddigkeit L, van Deursen R, Blum LC, Reymond JL. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17. J Chem Inf Model 2012; 52(11): 2864–2875

18. Wishart DS, Jewison T, Guo AC, Wilson M, Knox C, Liu Y, Djoumbou Y, Mandal R, Aziat F, Dong E, Bouatra S, Sinelnikov I, Arndt D, Xia J, Liu P, Yallou F, Bjorndahl T, Perez-Pineiro R, Eisner R, Allen F, Neveu V, Greiner R, Scalbert A. HMDB 3.0 — The Human Metabolome Database in 2013. Nucleic Acids Res 2013; 41(Database issue): D801–D807

19. Kanehisa M, Goto S, Sato Y, Furumichi M, Tanabe M. KEGG for integration and interpretation of large-scale molecular data sets. Nucleic Acids Res 2012; 40(Database issue D1): D109–D114

20. Wiener A, Shudler M, Levit A, Niv MY. BitterDB: a database of bitter compounds. Nucleic Acids Res 2012; 40(Database issue): D413–D419

21. Ahmed J, Meinel T, Dunkel M, Murgueitio MS, Adams R, Blasse C, Eckert A, Preissner S, Preissner R. CancerResource: a comprehensive database of cancer-relevant proteins and compound interactions supported by experimental knowledge. Nucleic Acids Res 2011; 39 (Database issue suppl_1): D960–D967

22. Wishart D, Arndt D, Pon A, Sajed T, Guo AC, Djoumbou Y, Knox C, Wilson M, Liang Y, Grant J, Liu Y, Goldansaz SA, Rappaport SM. T3DB: the toxic exposome database. Nucleic Acids Res 2015; 43(Database issue): D928–D934

23. Lim E, Pon A, Djoumbou Y, Knox C, Shrivastava S, Guo AC, Neveu V, Wishart DS. T3DB: a comprehensively annotated

database of common toxins and their targets. Nucleic Acids Res 2010; 38(Database issue suppl_1): D781–D786

24. Song H, Chu Q, Yan F, Yang Y, Han W, Zheng X. Red pitaya betacyanins protects from diet-induced obesity, liver steatosis and insulin resistance in association with modulation of gut microbiota in mice. J Gastroenterol Hepatol 2016; 31(8): 1462–1469

25. Shelley JC, Cholleti A, Frye LL, Greenwood JR, Timlin MR, Uchimaya M. Epik: a software program for pK(a) prediction and protonation state generation for drug-like molecules. J Comput Aided Mol Des 2007; 21(12): 681–691

26. Klicić JJ, Friesner RA, Liu SY, Guida WC. Accurate prediction of acidity constants in aqueous solution via density functional theory and self-consistent reaction field methods. J Phys Chem A 2002; 106(7): 1327–1335

27. Balogh GT, Gyarmati B, Nagy B, Molnar L, Keseru GM. Comparative evaluation of in silico pK(a) prediction tools on the Gold Standard Dataset. QSAR Comb Sci 2009; 28(10): 1148–1155

28. Vellay S, Miller Latimer N, Paillard G. Interactive text mining with Pipeline Pilot: a bibliographic web-based tool for PubMed. Infect Disord Drug Targets 2009; 9(3): 366–374

29. Rupp M, Körner R, V. Tetko I. Predicting the pKa of small molecule. Comb Chem High Throughput Screen 2011; 14(5): 307–327

30. Grüber C, Buss V. Quantum-mechanically calculated properties for the development of quantitative structure-activity relationships (QSAR'S). pKa-values of phenols and aromatic and aliphatic carboxylic acids. Chemosphere 1989; 19(10-11): 1595–1609

31. Walters WP, Stahl MT, Murcko MA. Virtual screening — an overview. Drug Discov Today 1998; 3(4): 160–178

32. Bickerton GR, Paolini GV, Besnard J, Muresan S, Hopkins AL. Quantifying the chemical beauty of drugs. Nat Chem 2012; 4(2): 90–98

33. Tian S, Wang J, Li Y, Li D, Xu L, Hou T. The application of in silico drug-likeness predictions in pharmaceutical research. Adv Drug Deliv Rev 2015; 86: 2–10

34. Durand C, Willett KC, Desilets AR. Proton pump inhibitor use in hospitalized patients: is overutilization becoming a problem? Clin Med Insights Gastroenterol 2012; 5: CGast. S9588