ACS Medicinal Chemistry Letters

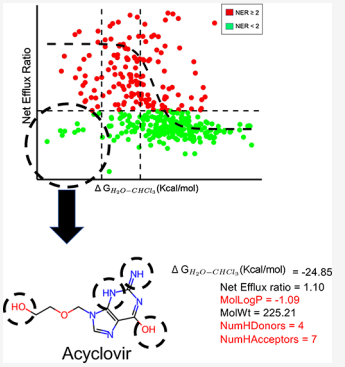
pubs.acs.org/acsmedchemlett
Letter

Impact of Hydrogen Bonding on P-Glycoprotein Efflux Transport as Revealed by Evaluation of a De Novo Prediction Model

Yulong Gou, Suyong Re, Kenji Mizuguchi, and Chioko Nagao*

Cite This: ACS Med. Chem. Lett. 2024, 15, 54−59
Read Online

ABSTRACT: Drug efflux transport by the permeability glycoprotein (P-gp) frequently diminishes drug efficacy, making the prediction of the P-gp efflux ratio critically significant. Most contemporary computational methods rely on binary predictions through machine learning. However, the accuracy and applicability of these predictions are heavily influenced by the training data sets. In contrast, we evaluated the validity of a de novo prediction model, which employs solvation free energies both within and outside the P-gp-binding pocket. Utilizing our in-house data set of 397 compounds, we discovered that this model struggles to predict a certain class of compounds accurately and that such outliers often exhibit favorable solvation free energy, with an increased number of hydrogen bond donors compared to other compounds investigated. Further examination of the functional groups in these compounds highlighted the significance of their specific interactions with P-gp. Considering these specific drug-P-gp interactions could enhance the accuracy of mechanism-based de novo prediction methods.



KEYWORDS: *P-gp, net efflux ratio, hydrogen bond donors, fragments*

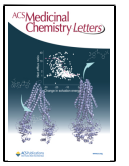
Permeability glycoprotein (P-gp), part of the ATP-binding cassette (ABC) superfamily and produced by the ABCB1 gene, is expressed as a principal component in various cells and organs.[1] P-gp actively expels a wide range of substances from cells, playing a significant role in pharmacokinetics and physical barriers.[2] Within the blood-brain barrier (BBB), P-gp is essential in efflux transport, serving to prevent the buildup of harmful and toxic compounds in the central nervous system.[3] In the context of Alzheimer's disease, a deficiency in P-gp contributes to the accumulation of amyloid β-peptide in the central nervous system (CNS),[4] leading to a deterioration in thinking and memory. Conversely, the overexpression of P-gp in cancer cells correlates with multidrug resistance to anticancer agents.[5] Thus, during drug discovery, it is vital to evaluate candidate compounds for their efflux transport to ascertain their permeability.

The net efflux ratio (NER) is frequently employed to gauge whether a compound is a P-gp substrate. The efflux ratio (ER) is expressed as the ratio of apparent permeability across cell monolayers in two directions: basolateral to apical (B → A) divided by apical to basolateral (A → B). An ER greater than 1 signifies higher efflux than passive transport. NER is determined as the ratio of ER in P-gp-overexpressing cells to ER in wild-type cells, offering a solid measure of P-gp's contribution. Compounds are deemed P-gp substrates for a NER ≥ 2 and non-P-gp substrates for a NER < 2.[6]

Several computational models for predicting P-gp substrates or inhibitors have been crafted using pharmacophore[7] and machine-learning-based quantitative structure−activity (property) relationships (QSAR/QSPR).[8] Most of these are binary classifiers designed to differentiate between substrates and nonsubstrates, or inhibitors and noninhibitors. However, predictive accuracy is often compromised by the utilization of small data sets or those derived from heterogeneous sources. Our group assembled an in-house data set of 397 compounds and developed regression models to predict P-gp-mediated efflux ratios, as well as three-category classification models using SVM and random forests.[9] Nonetheless, the reliance on pharmacophore-based or machine learning-based methods confines the models' applicability, prompting the need for more versatile approaches.

P-gp-mediated efflux occurs in a three-step process: initially, neutral compounds bind to the intracellular-facing state of P-gp; next, a conformational change to the extracellular-facing state occurs upon ATP binding;[10] and finally, the compounds are released into the extracellular domain. Gunaydin et al. hypothesized that the migration of compounds from the P-gp binding pocket of the hydrophobic environment to the extracellular domain of the hydrophilic environment primarily determines NER. They also presented a model based on the correlation between the calculated solvation free energy differences and NER.[11] They calculated the solvation free

Received: August 25, 2023
Revised: November 29, 2023
Accepted: November 29, 2023
Published: December 21, 2023



https://doi.org/10.1021/acsmedchemlett.3c00376
ACS Med. Chem. Lett. 2024, 15, 54−59

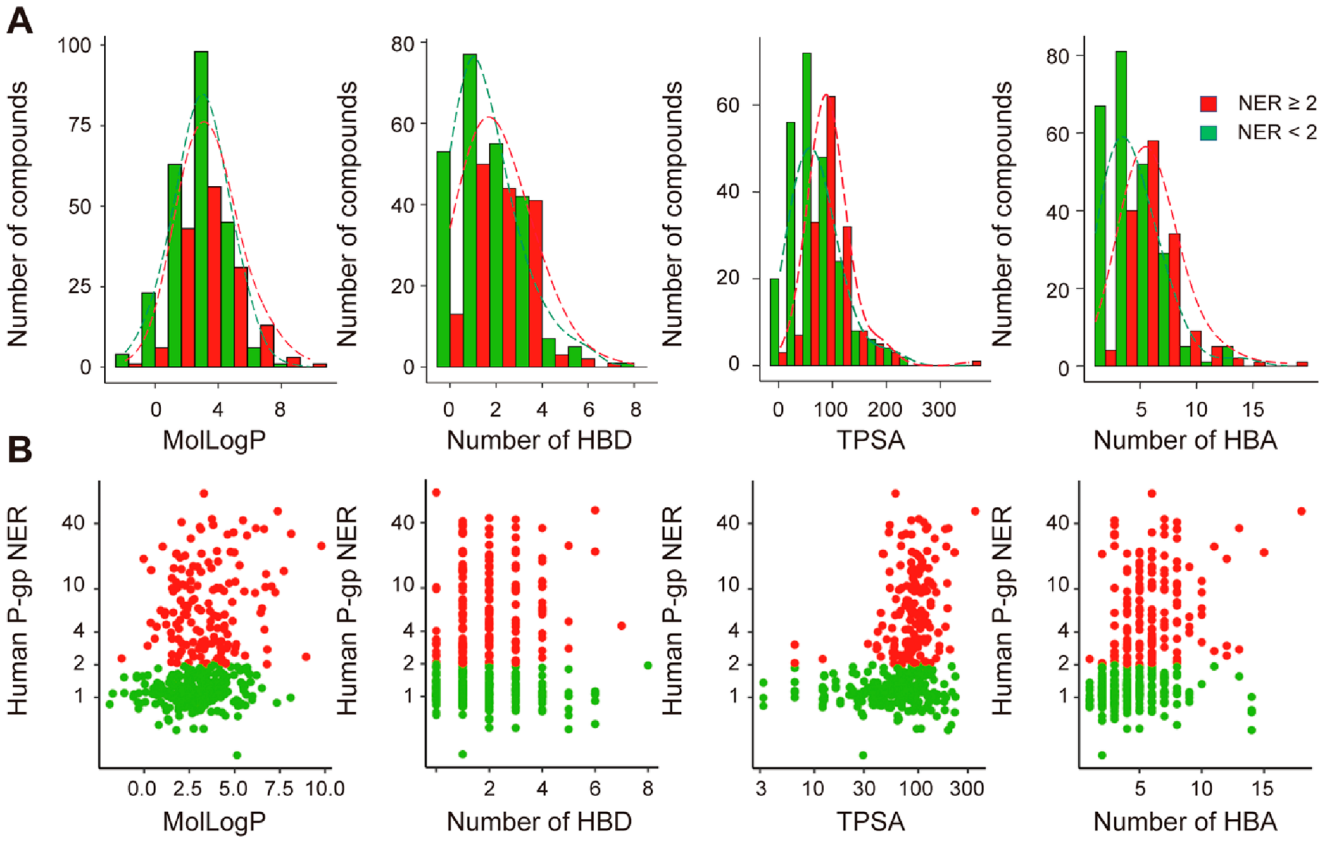


Figure 1. (A) Graph illustrating the distribution of the physicochemical properties: MolLogP, Number of HBDs, Number of HBAs, and TPSA. (B) Correlation between the human P-gp net efflux ratio and physicochemical properties. P-gp substrates are denoted by red bars, while non-P-gp substrates are shown in green.

energies of 282 compounds (not publicly available) in both water and chloroform to estimate the free energy change in the transfer of compounds from the P-gp-binding pocket to the extracellular domain. The resultant free energy differences displayed a sigmoidal relationship with the NER, facilitating binary-level classification of compounds. However, the applicability of this simplistic model remains uncertain. In this study, we assessed Gunaydin's model utilizing our data set of 397 commercially available compounds from Watanabe et al.[12] (Table S1).

Figure 1 illustrates the distributions of LogP, the number of hydrogen bond (H-bond) donors (HBDs), the number of H-bond acceptors (HBAs), and the topological polar surface area (TPSA) in the compounds and their associations with NER. The NER values, obtained through an in vitro P-gp transport assay utilizing the LLC-PK1 cell line, ranged from 0.29 to 40. Compared to Gunaydin's data set, depicted in Figure 4 of their article,[11] the physicochemical properties in our data set are distributed over a broader spectrum. For example, the maximum number of HBDs in the Gunaydin data set was three, while our data set included compounds with over seven HBDs, which violates one of the Lipinski Rule of Five (Ro5) criteria. Our data set contained 73 compounds that did not meet at least one of the four criteria of the Ro5, of which an additional nine compounds were the beyond Ro5 (bRo5) compounds (Table S2).

For comparison, we employed the same computational model as Gunaydin but with some technical updates. Specifically, we used RDKit with the ETKDG method[13] to create 3D structures of 397 molecules from SMILES, and all compounds were minimized using the MMFF94 force field[14] for later quantum chemical calculations. We applied the polarizable continuum model using the integral equation formalism variant (IEFPCM)[15] to compute the solvation free energies in water (ΔG_{water}) and chloroform (ΔG_{CHCl3}), using Gaussian 16 software at the M06-2X/6-31+G* level[16] (Figure S1). For metal atoms unsupported by the 6-31+G* basis set, we employed the Lanl2TZ basis sets.[17] The difference in free energies between the two media, $\Delta G_{water\text{-}CHCl3}$, symbolized the free energy shift from the P-gp binding pocket to the hydrophilic environment of the extracellular domain. We further corrected the solvation free energy for compounds with ionizable groups by calculating the pK_a values [$\Delta G = -1.4$ (pH − pK_a)]. MolGpka,[18] an open-source program employing a graph-convolutional neural network, was utilized for pK_a computation. While the updated methods shift ΔG values, our results were consistent with the predicted pK_a and corrected ΔG values reported by Gunaydin et al. for six compounds (refer to Figure S2 and Table S3).

Initially, we investigated the correlation between experimentally determined P-gp NERs and calculated solvation free energy changes (Figure 2A). Our data set's $\Delta G_{water\text{-}CHCl3}$ values covered a slightly more expansive range (−30 to 0 kcal/mol) than the data set of Gunaydin et al. (−25 to −5 kcal/mol). Remarkably, our data set was widely spread and did not perfectly fit the sigmoidal relationship that Gunaydin et al. proposed. Figure 2B shows the box plots of the $\Delta G_{water\text{-}CHCl3}$ with NER thresholds of two and ten. Both analyses show significant differences between two groups (p-values of 0 and 8e-07, respectively), indicating that the sigmoidal function

https://doi.org/10.1021/acsmedchemlett.3c00376
ACS Med. Chem. Lett. 2024, 15, 54−59

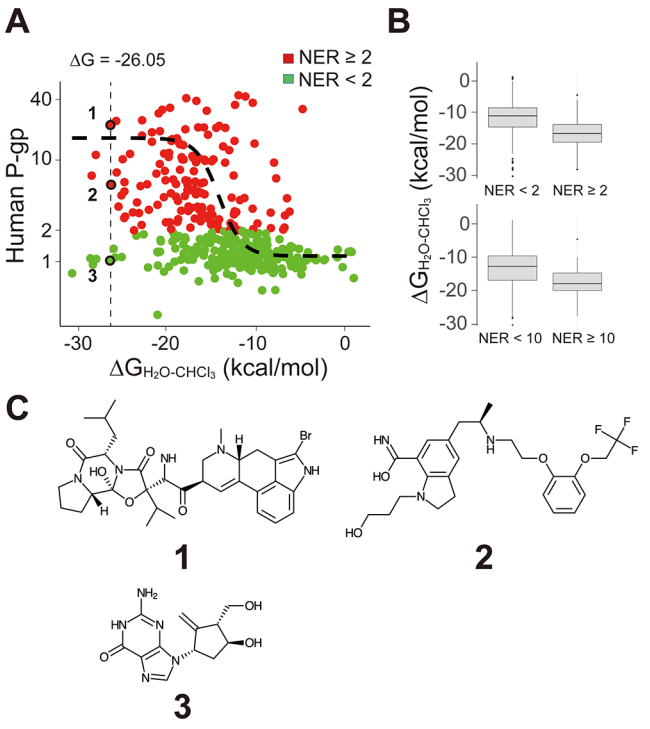


Figure 2. Scatter plot (A) and box plots (B) illustrating the relationship between the net efflux ratio and $\Delta G_{water-CHCl3}$, and the arbitrary chosen compounds having similar NERs (C). The boxes indicate the interquartile range, and the line in the box indicates the median. The whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range from the box.

fitting is somewhat valid. However, in the boxplot with a NER threshold of ten, there is an overlap of the two groups, and there are non-negligible outliers regardless of the threshold value. For instance, three compounds with nearly identical $\Delta G_{water-CHCl3}$ values (~ −26 kcal/mol) demonstrated varied NER values from 1 to 22 (Figure 2 and Table 1). While

Table 1. Comparison of Three Compounds That Exhibit Similar ΔG Values but Have Differing NERs

	compound		
	1	2	3
ΔG (kcal/mol)	−26.11	−26.04	−26.16
NER	22.34	5.71	1.01
number of HBDs	3	4	5
number of HBAs	6	6	7
molecular weight	654.61	495.54	277.28
MolLogP	4.03	3.86	−0.58

compounds 1 and 2 are near the minimum value of $\Delta G_{water-CHCl3}$, compound 3 is an outlier. On analyzing these compounds' physicochemical properties, a weak connection was observed between NER and the number of HBDs, along with the previously recognized relationship between NER, MolLogP, and molecular weight (MW).[19,20] We also analyzed the relationships between physicochemical properties and $\Delta G_{water-CHCl3}$ values (Figure S3 and Table S4). All physicochemical properties were weakly correlated with $\Delta G_{water-CHCl3}$ values, whereas, as shown in Figure S3, even compounds with the same physicochemical value have various $\Delta G_{water-CHCl3}$ values, indicating that a single physicochemical property is not the primary explanatory variable.

To scrutinize compounds that markedly deviated from the sigmoidal curve, we segmented all compounds into six regions, labeled A−F, by fitting the $\Delta G_{water-CHCl3}$ and NER distributions with Gaussian functions (Figure 3). The demarcation thresholds for each region are as follows: $\Delta G \leq -22$ kcal/mol, NER ≥ 2 for region A; −22 kcal/mol < $\Delta G \leq -16$ kcal/mol, NER ≥ 2 for region B; $\Delta G > -16$ kcal/mol, NER ≥ 2 for region C; $\Delta G \leq -22$ kcal/mol, NER < 2 for region D; −22 kcal/mol < $\Delta G \leq -16$ kcal/mol, NER < 2 for region E; $\Delta G > -16$ kcal/mol, NER < 2 for region F.

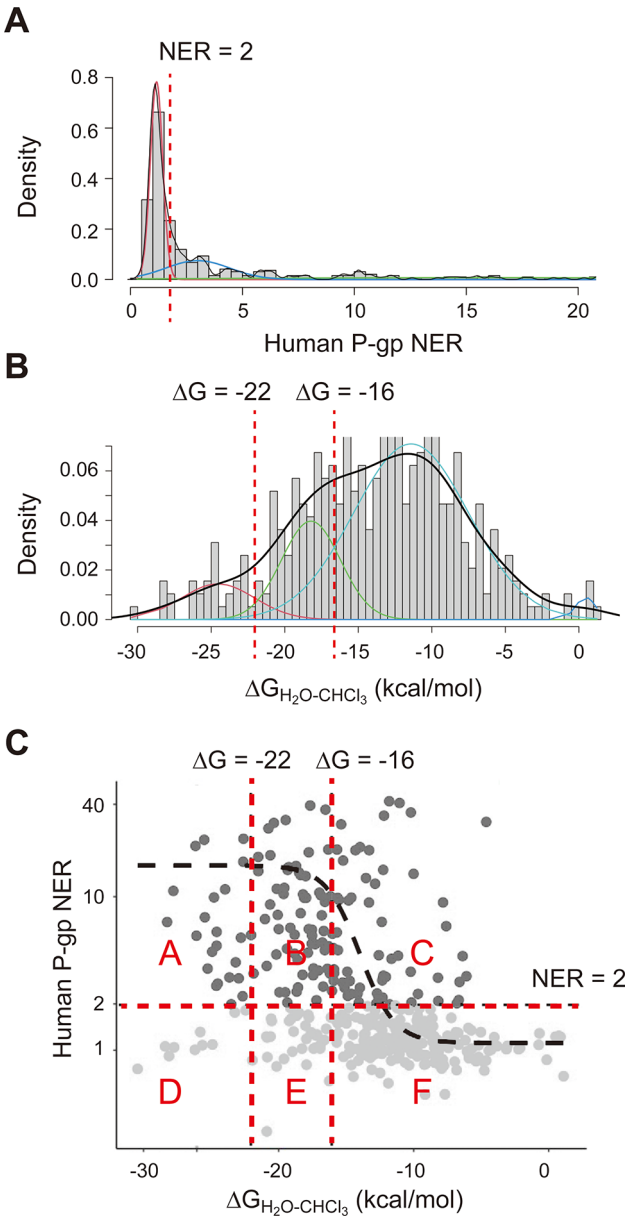


Figure 3. (A) Density plot for NER with an intersection at a value of 2. (B) Density plot for ΔG with intersections at values of −22 and −16. (C) Scatter plot segmented by density distributions into six parts: A−F.

The compounds within each region were analyzed following Ro5, considering parameters such as the number of HBDs, the number of HBAs, MW, and MolLogP. In addition, TPSA

https://doi.org/10.1021/acsmedchemlett.3c00376
ACS Med. Chem. Lett. 2024, 15, 54−59

parameters of the compounds were analyzed. The average values of these physicochemical traits for the six regions are detailed in Table 2. Figures 4, S4 and Table S5 show box plots

Table 2. Classification and Physicochemical Properties of Regions A−F

	thresholds						
	ΔG	NER	no. of compds	avg HBDs	avg HBAs	regions	avg MolLogP
A	≤−22	≥2	22	3.18	7.54	508.77	3.18
B	−22 < and ≤ −16	≥2	65	1.92	6.42	437.68	3.03
C	−16 <	≥2	64	1.41	4.64	390.66	3.82
D	≤−22	<2	10	4.5	6.7	351.49	0.61
E	−22 < and ≤ −16	<2	35	2.31	5.06	345.38	2.09
F	−16 <	<2	190	1.17	3.61	307.56	3.22

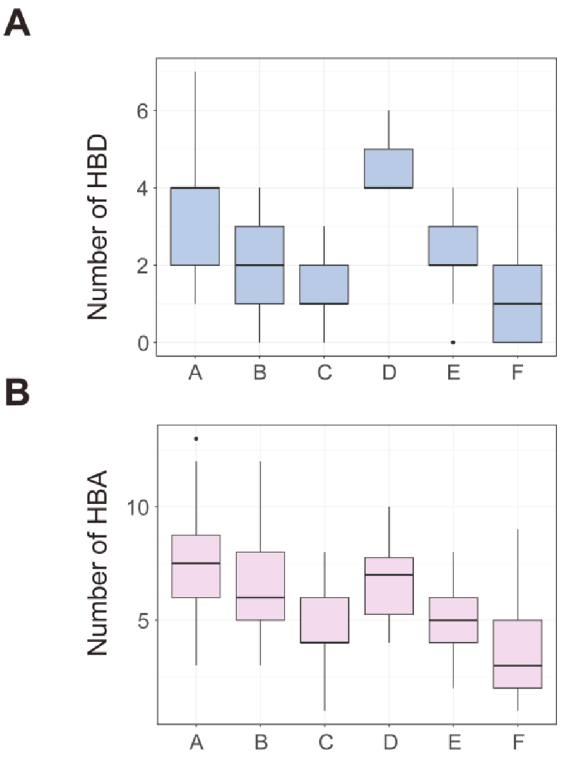


Figure 4. Box plots of the physicochemical properties of regions A−F. (A) Distribution of number of HBDs and (B) number of HBAs. The boxes indicate the interquartile range, and the line in the box indicates the median. The whiskers extend to the most extreme data point which is no more than 1.5 times the interquartile range from the box.

and the results from Tukey's test. Notably, the average MolLogP value was considerably higher (3.82) in region C and markedly lower (0.61) in region D. This signifies that compounds in region C with small $\Delta G_{water\text{-}CHCl3}$ values lean toward a hydrophobic setting, while those in region D with large $\Delta G_{water\text{-}CHCl3}$ values favor a hydrophilic environment. Intriguingly, region D has an average HBD number of 4.5, which is considerably greater than that of any other region and a substantial average number of HBAs (6.7, with region A having the maximum at 7.6).

To delve into the chemical characteristics of these compounds, we examined the occurrence of functional groups within each region. Utilizing SMARTS, we calculated the main molecular fragments, including commonly found groups like esters, carboxylic acids, phenols, azoles, and arene carbons. Region D exhibited a significant prevalence of ester, carbonyl acid, phenol, azole, and primary ketimine fragments, consistent with its low MolLogP value (0.61) and high number of HBDs (4.5) (Figure 5). Counter to expectations, compounds with a

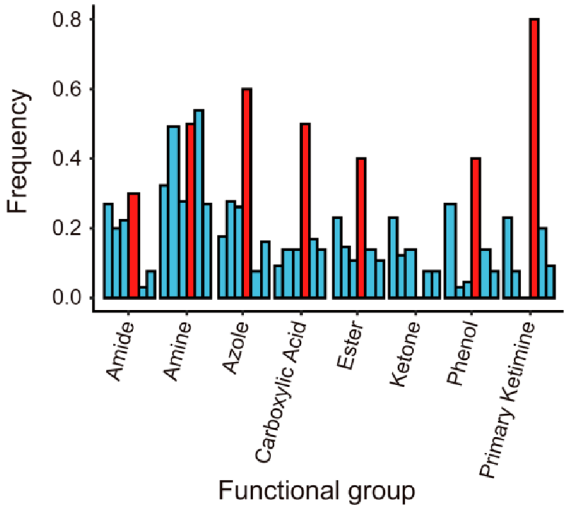


Figure 5. Histograms showcasing the presence of amide, amine, azole, carboxylic acid, ester, ketone, phenol, and primary ketimines. Histograms are organized from left to right corresponding to regions A−F. The histograms associated with the D region are highlighted in red.

greater number of HBDs and favorable solvation in bulk water have low NER values. No definitive distribution of compound fragments was detected in region C.

These findings imply that a predictive model solely based on solvation free energy cannot encompass the NER for a diverse set of compounds. The approach by Gunaydin et al. calculates solvation free energy using a continuum solvation model, handling the interactions of compounds with proteins and solvents implicitly. In this context, P-gp NER is determined merely by the compounds' physicochemical attributes, neglecting specific interactions with the surrounding environment (protein-binding site or solvent water). Our investigation underscores that in addition to a holistic view of solvation free energy and physicochemical properties, specific substrate recognition by P-gp plays an essential role in P-gp NER. Based on the analysis of a series of compounds, the number and strength of the HBA/HBD ratio affect P-gp NER,[21] which is consistent with our results. Intramolecular hydrogen bonds are also said to affect the interaction with P-gp and membrane permeability.[21] Our data set includes the compounds that may have intramolecular hydrogen bonds; thus, further analysis is required.

Generally, P-gp transports hydrophobic and weakly cationic substrates. A broad substrate specificity of P-gp has been brought by a large cavity of ∼6000 Å. Through X-ray crystallography and cryo-electron microscopy studies, complex structures of P-gp have been revealed, displaying multiple binding sites and modes.[22] The electrostatic potential of the intracellular-facing state of P-gp's binding pocket is portrayed in Figure S5. This inner surface, adjacent to the substrate-binding pocket, is hydrophobic, with some negatively charged acidic areas. Compounds in region D, which prefer a

https://doi.org/10.1021/acsmedchemlett.3c00376
ACS Med. Chem. Lett. 2024, 15, 54−59

hydrophilic environment, are less likely to be substrates, whereas hydrogen bonds may form between the HBDs and charged acidic residues. Compounds with elevated NER often contain both HBAs and HBDs, hinting that the P-gp NER is affected by specific substrate recognition through hydrogen bonds. Docking studies have shed light on potential compound binding sites and crucial residues, considering their various roles as substrates, inhibitors, or modulators.[23] Current structural information is confined to the intracellular-facing state of P-gp, leaving the dynamics of the interaction between the compound and P-gp during the ATP-induced conformational shift to the extracellular-facing state unclear. This understanding would be crucial for transporting the compound out of the cell. Thus, our results advocate that specific interactions between compounds and P-gp binding site residues should also be considered in the development of universally applicable mechanism-based de novo prediction methods.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acsmedchemlett.3c00376.

Comparison of our calculated solvation free energy and pK_a with those reported by Gunaydin et al.,[11] relationships between physicochemical properties and ΔG or NER, and example of Gaussian input files (PDF)

All compound information (XLSX)

■ AUTHOR INFORMATION

Corresponding Author

Chioko Nagao − *Institute for Protein Research, Osaka University, Suita-shi, Osaka 565-0871, Japan; Artificial Intelligence Center for Health and Biomedical Research, National Institutes of Biomedical Innovation, Health and Nutrition, Settu-City, Osaka 566-0002, Japan;* orcid.org/0000-0002-7721-0642; Email: c_nagao@protein.osaka-u.ac.jp

Authors

Yulong Gou − *Institute for Protein Research, Osaka University, Suita-shi, Osaka 565-0871, Japan;* orcid.org/0009-0005-2718-5812

Suyong Re − *Artificial Intelligence Center for Health and Biomedical Research, National Institutes of Biomedical Innovation, Health and Nutrition, Settu-City, Osaka 566-0002, Japan;* orcid.org/0000-0002-3752-6554

Kenji Mizuguchi − *Institute for Protein Research, Osaka University, Suita-shi, Osaka 565-0871, Japan; Artificial Intelligence Center for Health and Biomedical Research, National Institutes of Biomedical Innovation, Health and Nutrition, Settu-City, Osaka 566-0002, Japan;* orcid.org/0000-0003-3021-7078

Complete contact information is available at: https://pubs.acs.org/10.1021/acsmedchemlett.3c00376

Author Contributions

C.N. and K.M. conceptualized the study. Y.G., S. R. and C.N., carried out the calculations and data analysis, and substantially contributed to the manuscript drafting. All authors critically reviewed and revised the manuscript draft and approved the final version for submission.

Notes

The authors declare no competing financial interest.

■ ACKNOWLEDGMENTS

This work was supported by JSPS KAKENHI, Grant Number JP 21K12125, the Promotion Program for Frontier Protein Research from the Institute for Protein Research, Osaka University, and the Collaborative Research Program of the Institute for Protein Research, Osaka University (CRa-23-04). Special thanks to Dr. Reiko Watanabe for engaging in meaningful discussions that contributed to the work.

■ ABBREVIATIONS

P-gp, Permeability glycoprotein; BBB, Blood-brain barrier; ABC, ATP-binding cassette; NER, Net efflux ratio; HBD, Hydrogen bond donor; HBA, Hydrogen bond acceptor; TPSA, Topological polar surface area; ΔG_{water}, Solvation free energy in water; ΔG_{CHCl3}, Solvation free energy in chloroform; $\Delta G_{water\text{-}CHCl3}$, Solvation free energy difference between water and chloroform

■ REFERENCES

(1) Doige, C. A.; Ames, G. F.-L. ATP-Dependent Transport Systems in Bacteria and Humans: Relevance to Cystic Fibrosis and Multidrug Resistance. *Annu. Rev. Microbiol.* **1993**, *47* (1), 291−319.

(2) Cordon-Cardo, C.; O'Brien, J. P.; Boccia, J.; Casals, D.; Bertino, J. R.; Melamed, M. R. Expression of the Multidrug Resistance Gene Product (P-Glycoprotein) in Human Normal and Tumor Tissues. *J. Histochem. Cytochem.* **1990**, *38* (9), 1277−1287.

(3) Bendayan, R.; Lee, G.; Bendayan, M. Functional Expression and Localization of P-Glycoprotein at the Blood Brain Barrier. *Microsc. Res. Technol.* **2002**, *57* (5), 365−380.

(4) Hardy, J. A.; Higgins, G. A. Alzheimer's Disease: the Amyloid Cascade Hypothesis. *Science* **1992**, *256* (5054), 184−185.

(5) Endicott, J. A.; Ling, V. The Biochemistry of P-glycoprotein-Mediated Multidrug Resistance. *Annu. Rev. Biochem.* **1989**, *58* (1), 137−171.

(6) U.S. FDA. *In Vitro Drug Interaction Studies—Cytochrome P450 Enzyme- and Transporter-Mediated Drug Interactions Guidance for Industry*; Food and Drug Administration (FDA), Center for Drug Evaluation and Research (CDER): Silver Spring, MD, 2020.

(7) Li, W. X.; Li, L.; Eksterowicz, J.; Ling, X. B.; Cardozo, M. Significance Analysis and Multiple Pharmacophore Models for Differentiating P-Glycoprotein Substrates. *J. Chem. Inf. Model.* **2007**, *47* (6), 2429−2438.

(8) Kovatcheva, A.; Golbraikh, A.; Oloff, S.; Xiao, Y. D.; Zheng, W.; Wolschann, P.; Buchbauer, G.; Tropsha, A. Combinatorial QSAR of Ambergris Fragrance Compounds. *J. Chem. Inf. Comput. Sci.* **2004**, *44* (2), 582−595.

(9) Huang, J.; Ma, G.; Muhammad, I.; Cheng, Y. Identifying P-Glycoprotein Substrates Using a Support Vector Machine Optimized by a Particle Swarm. *J. Chem. Inf. Model.* **2007**, *47* (4), 1638−1647.

(10) Prajapati, R.; Sangamwar, A. T. Translocation Mechanism of P-Glycoprotein and Conformational Changes Occurring at Drug-Binding Site: Insights From Multi-targeted Molecular Dynamics. *Biochim. Biophys. Acta* **2014**, *1838* (11), 2882−2898.

(11) Gunaydin, H.; Weiss, M. M.; Sun, Y. De Novo Prediction of p-Glycoprotein-Mediated Efflux Liability for Druglike Compounds. *A.C.S. Med. Chem. Lett.* **2013**, *4* (1), 108−112.

(12) Watanabe, R.; Esaki, T.; Ohashi, R.; Kuroda, M.; Kawashima, H.; Komura, H.; Natsume-Kitatani, Y.; Mizuguchi, K. Development of an In Silico Prediction Model for p-Glycoprotein Efflux Potential in Brain Capillary Endothelial Cells Toward the Prediction of Brain Penetration. *J. Med. Chem.* **2021**, *64* (5), 2725−2738.

(13) Riniker, S.; Landrum, G. A. Better Informed Distance Geometry: Using What We Know to Improve Conformation Generation. *J. Chem. Inf. Model.* **2015**, *55* (12), 2562−2574.
(14) Halgren, T. A. Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94. *J. Comput. Chem.* **1996**, *17* (5−6), 490−519.
(15) Takano, Y.; Houk, K. N. Benchmarking the Conductor-Like Polarizable Continuum Model (CPCM) for Aqueous Solvation Free Energies of Neutral and Ionic Organic Molecules. *J. Chem. Theory Comput.* **2005**, *1* (1), 70−77.
(16) Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Petersson, G. A.; Nakatsuji, H. et al. *Gaussian 16*, revision C. 01; Gaussian, Inc.: Wallingford, CT, 2016.
(17) Roy, L. E.; Hay, P. J.; Martin, R. L. Revised Basis Sets for the LANL Effective Core Potentials. *J. Chem. Theory Comput.* **2008**, *4* (7), 1029−1031.
(18) Pan, X.; Wang, H.; Li, C.; Zhang, J. Z. H.; Ji, C. MolGpka: A Web Server for Small Molecule pKa Prediction Using a Graph-Convolutional Neural Network. *J. Chem. Inf. Model.* **2021**, *61* (7), 3159−3165.
(19) Löscher, W.; Potschka, H. Role of Drug Efflux Transporters in the Brain for Drug Disposition and Treatment of Brain Diseases. *Prog. Neurobiol.* **2005**, *76* (1), 22−76.
(20) Wildman, S. A.; Crippen, G. M. Prediction of Physicochemical Parameters by Atomic Contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39* (5), 868−873.
(21) Desai, P. V.; Raub, T. J.; Blanco, M. J. How Hydrogen Bonds Impact P-Glycoprotein Transport and Permeability. *Bioorg. Med. Chem. Lett.* **2012**, *22* (21), 6540−6548.
(22) Nosol, K.; Romane, K.; Irobalieva, R. N.; Alam, A.; Kowal, J.; Fujita, N.; Locher, K. P. Cryo-EM Structures Reveal Distinct Mechanisms of Inhibition of the Human Multidrug Transporter ABCB1. *Proc. Natl. Acad. Sci. U. S. A.* **2020**, *117* (42), 26245−26253.
(23) Vilar, S.; Sobarzo-Sánchez, E.; Uriarte, E. In Silico Prediction of P-Glycoprotein Binding: Insights From Molecular Docking Studies. *Curr. Med. Chem.* **2019**, *26* (10), 1746−1760.

https://doi.org/10.1021/acsmedchemlett.3c00376
ACS Med. Chem. Lett. 2024, 15, 54−59