

К 40-летию Института физиологически активных веществ РАН**ЭКСПЕРИМЕНТАЛЬНЫЕ ИССЛЕДОВАНИЯ****БИНАРНАЯ КЛАССИФИКАЦИЯ СОЕДИНЕНИЙ, ПРОХОДЯЩИХ И НЕ ПРОХОДЯЩИХ ЧЕРЕЗ ГЕМАТОЭНЦЕФАЛИЧЕСКИЙ БАРЬЕР, СОЗДАННАЯ МЕТОДОМ ЛОГИСТИЧЕСКОЙ РЕГРЕССИИ****О.А. Раевский*, Д.Е. Полианчик, О.Е. Раевская**Институт физиологически активных веществ Российской академии наук,
142432, Черноголовка Московской обл., Северный проезд, 1; *эл. почта: raevsky@ipac.ac.ru

Созданы устойчивые классификационные предсказательные модели 83 лекарственных средств, преодолевающих и не преодолевающих гематоэнцефалический барьер, построенные методом логистической регрессии с использованием физико-химических дескрипторов, характеризующих стерические и электростатические взаимодействия и энергию водородной связи. Модели сбалансированы, их уровень предсказания составил 75-80%.

Ключевые слова: ЦНС; ГЭБ; бинарная классификация; дескрипторы; взаимосвязи структура-активность**DOI:** 10.18097/BMCRM00065**ВВЕДЕНИЕ**

В настоящее время проблема лекарств, воздействующих на центральную нервную систему (ЦНС) приобретает для международного сообщества исключительно важное значение. Увеличение продолжительности жизни населения влечёт за собой увеличение прослойки пожилых людей и, соответственно, рост различных заболеваний ЦНС. Лечение таких болезней является чрезвычайно дорогостоящим и подчас неэффективным [1]. Достаточно упомянуть, что сейчас на фармацевтическом рынке есть шесть малоэффективных лекарств для терапии болезни Альцгеймера. Высокая стоимость открытия новых лекарств и их предклинических и клинических испытаний побуждает фармацевтическую индустрию к развитию и использованию современных компьютерных технологий для отбора перспективных кандидатов.

Одним из важнейших защитных механизмов ЦНС является гематоэнцефалический барьер (ГЭБ). Под этим термином условно объединяют совокупность анатомических и функциональных систем, которые поддерживают постоянство и регулируют состав внутренней среды головного и спинного мозга [2]. ГЭБ регулирует распределение различных веществ между током крови и мозгом и обладает высокой избирательностью. Барьерная функция состоит в предотвращении проникновения вредных веществ (лекарства также рассматриваются ГЭБ как вредные) из крови в мозг и их активном выведении в тех случаях, если они каким-то образом попали в ЦНС. Существует множество механизмов проникновения веществ из кровеносного русла в мозг. Большинство веществ проходит через ГЭБ посредством диффузии через клеточные мембраны по градиенту концентрации из тока крови в мозг без энергетических затрат. Кроме этого, различные транспортные системы могут осуществлять как перенос веществ из кровотока к мозгу (influx), так и обратный перенос из ткани мозга в кровоток (efflux).

Созданию моделей *in silico* для предсказания транспорта через ГЭБ, являющегося сложнейшим биохимическим процессом, уделяется огромное внимание, поскольку это может заменить тотальный биологический скрининг разумными избирательными процедурами без неоправданного дорогостоящего синтеза. Здесь всё возрастающую роль играют методы установления взаимосвязей “структура-активность” (QSAR).

Обширные исследования в области компьютерного моделирования транспорта через ГЭБ ведутся с 90-х годов прошлого века. Здесь нужно упомянуть о “наивных” эмпирических правилах медицинских химиков [3-7], знаменитом “правиле-5” Липинского [8], Мульти-Оптимизационном подходе (MOP) [9-11]. Первые наши работы в этой области были выполнены вместе со швейцарскими коллегами в середине 90-х годов [12, 13]. Результаты недавних наших работ опубликованы в [14-16]. Однако, несмотря на интенсивные исследования, проблема создания стабильных предсказательных моделей транспорта через ГЭБ требует привлечения новых баз данных, дескрипторов и методов [17].

МАТЕРИАЛЫ И МЕТОДЫ

Настоящая публикация кратко описывает результаты работы по созданию модели бинарной классификации соединений хорошо и плохо проходящих ГЭБ (BV^+ и BV^-). Порогом такого разделения служит величина $\log BV=0$. В работе использованы тщательно отобранные данные $\log BV$ в равновесном состоянии. Объектом тестирования были крысы [18].

Для BV^+/BV^- классификации была использована логистическая регрессия (LR). LR во многом похожа на линейную регрессию. Однако, в отличие от последней, LR используется для моделирования вероятности какого-либо события (явления) как линейной функции ряда переменных (дескрипторов) [19].



Делается предположение о том, что вероятность (P) наступления события BBB^+ (вещество проходит ГЭБ; $Y=1$):

$$P\{Y = 1|X\} = f(z) \quad (1).$$

Логистическая функция $f(z)$, именуемая также “логит” или сигмоидная функция, записывается как:

$$f(z) = \frac{1}{1 + e^{-z}} \quad (2),$$

где $z = a_0 + a_1X_1 + a_2X_2 + \dots + a_iX_i$, a_0 – свободный член модели, X_1, \dots, X_i – дескрипторы с их соответствующими регрессионными коэффициентами – a_1, \dots, a_i .

Вероятность наступления события BBB^- (вещество не проходит ГЭБ; $Y=0$):

$$P\{Y = 0|X\} = 1 - f(z) \quad (3).$$

При построении QSAR модели с использованием логистической регрессии коэффициенты при дескрипторах рассчитываются из данных по соединениям обучающей выборки, а затем рассчитываются вероятности события для соединений тестовой выборки при фиксированных значениях коэффициентов, установленных для соединений обучающей выборки. При этом выбор события между BBB^+ и BBB^- ($Y = 1$ или 0) определяется по рассчитанному значению функции $f(z)$. Если рассчитанное значение $f(z) \geq 0.5$, то соединение относится к категории BBB^+ ($Y=1$), если $f(z) < 0.5$ – к категории BBB^- ($Y=0$).

Подобно множественной линейной регрессии коэффициенты логистической регрессии могут описывать влияние молекулярных дескрипторов на результат предсказания. Когда коэффициент имеет большое значение, он показывает, что молекулярный дескриптор сильно влияет на вероятность события, в то время как нулевое значение коэффициента свидетельствует, что этот дескриптор не оказывает

влияния на конечный результат. Аналогично, положительные знаки коэффициентов увеличивают вероятность события, а отрицательные знаки уменьшают эту вероятность. Примеры успешного применения логистической регрессии в QSAR моделировании представлены в публикациях [20-24]. В данной работе для проведения расчётов LR использован программный комплекс SSPS [25]. Дескрипторы рассчитаны программами DRAGON [26] и HYBOT [27].

РЕЗУЛЬТАТЫ

В таблице 1 приведены результаты расчётов бинарной классификации BBB^+/BBB^- для 83 соединений. Перед расчётами соединения были ранжированы по величине $\log BB$ и каждое пятое соединение отбирали для тестовой выборки. В работе использована k -fold=4 кросс-валидация, с числом итераций равным 100.

Если результат классификации положительный, и истинное значение тоже положительное, то речь идёт об истинно-положительном значении (true-positive, TP); если результат классификации положительный, но истинное значение отрицательное, ложно-положительное значение (false-positive, FP); если результат классификации отрицательный, и истинное значение тоже отрицательное, истинно-отрицательное значение (true-negative, TN); если результат классификации отрицательный, но истинное значение положительно, ложно-отрицательное значение (false-negative, FN). Чувствительность определяется как $Sens = TP / (TP + FN)$, специфичность как $Spec = TN / (TN + FP)$, точность как $Acc = (TP + TN) / (TP + FN + TN + FP)$ и коэффициент корреляции Мэттьюса (MCC) как:

$$MCC = \frac{TP \times TN - FP \times FN}{\sqrt{(TP + FP)(TP + FN)(TN + FP)(TN + FN)}} \quad (4).$$

Таблица 1. Статистические параметры бинарной классификации BBB^+/BBB^- лекарств методом логистической регрессии

| Sets | N | Descriptors | TP | FN | TN | FP | Sens | Spec | Acc | MCC |
|----------|----------|------------------------------|----|----|----|----|-------|-------|-------|-------|
| training | 67 16 | AlogP | 32 | 8 | 14 | 13 | 0.800 | 0.519 | 0.687 | 0.333 |
| CV | | | 32 | 8 | 14 | 13 | 0.800 | 0.519 | 0.687 | 0.333 |
| test | | | 6 | 4 | 5 | 1 | 0.600 | 0.833 | 0.688 | 0.423 |
| training | 67 16 | AlogP, HBD | 34 | 6 | 16 | 11 | 0.850 | 0.593 | 0.748 | 0.462 |
| CV | | | 34 | 6 | 12 | 12 | 0.850 | 0.556 | 0.731 | 0.429 |
| test | | | 5 | 5 | 6 | 0 | 0.500 | 1.000 | 0.688 | 0.522 |
| training | 67 16 | AlogP, HBD, MW | 31 | 9 | 17 | 10 | 0.775 | 0.630 | 0.716 | 0.407 |
| CV | | | 31 | 9 | 15 | 12 | 0.775 | 0.556 | 0.687 | 0.338 |
| test | | | 5 | 5 | 6 | 0 | 0.500 | 1.000 | 0.688 | 0.522 |
| training | 67 16 | AlogP, HBD, TPSA | 34 | 6 | 20 | 7 | 0.850 | 0.741 | 0.806 | 0.595 |
| CV | | | 34 | 6 | 20 | 7 | 0.850 | 0.741 | 0.806 | 0.595 |
| test | | | 6 | 4 | 6 | 0 | 0.600 | 1.000 | 0.750 | 0.600 |
| training | 67 16 | AlogP, HBD, MW, TPSA | 33 | 7 | 20 | 7 | 0.825 | 0.741 | 0.791 | 0.566 |
| CV | | | 32 | 8 | 19 | 8 | 0.800 | 0.704 | 0.761 | 0.504 |
| test | | | 6 | 4 | 6 | 0 | 0.600 | 1.000 | 0.750 | 0.600 |
| training | 67 16 | $\alpha, \sum C_a, \sum C_d$ | 35 | 5 | 18 | 9 | 0.875 | 0.667 | 0.791 | 0.560 |
| CV | | | 32 | 8 | 17 | 10 | 0.800 | 0.630 | 0.731 | 0.436 |
| test | | | 7 | 3 | 6 | 0 | 0.700 | 1.000 | 0.812 | 0.683 |

Прежде всего, было испытано влияние гидрофобности, выраженного как AlogP. Эта модель оказалась неудовлетворительной вследствие плохого распознавания ВВВ-соединений. Последующее поочередное добавление дескрипторов MW, HBD, TPSA существенно улучшило этот параметр. В результате четырёх-параметровая модель с указанными дескрипторами программы DRAGON обеспечила правильное распознавание 82.5% ВВВ⁺ и 74.1% ВВВ- при обучении и 60.0% ВВВ⁺ и 100% ВВВ⁻ в тестовой выборке. Однако лучшей QSAR моделью для данного ряда соединений оказалась модель из трёх дескрипторов программы НУВОТ (α , ΣC_a , ΣC_d). Эта модель дала такую же точность для обучающей выборки, но существенно лучшие результаты для тестовой выборки (81.2%). Как показывает данная работа, метод LR является простым и удобным для медицинских химиков бинарным классификатором свойства (активности). При его использовании особенно просто оценить вклад каждого дескриптора в модель.

БЛАГОДАРНОСТИ

Работа выполнена в рамках государственного задания на 2018 год (тема № 0090-2017-0020).

ЛИТЕРАТУРА

1. Wager, T.T., Chandrasekaran, R.Y., Hou, X., Troutman, M.D., Verhoest, P.R., Villalobos, A., Will, Y. (2010) Defining desirable central nervous system drug space through the alignment of molecular properties, in vitro ADME, and safety attributes. *ACS Chemical Neuroscience*, 1(6), 420-434. DOI: 10.1021/cn100007x
2. Bradbury, M.W.B. (1979) The concept of a blood-brain barrier. John Wiley & Sons.
3. Young, R.C., Mitchell, R.C., Brown, T.H., Ganellin, C.R., Griffiths, R., Jones, M., Rana, K.K., Saunders, D., Smith, I.R., Sore, N.E., Wilks, T.J. (1988) Development of a new physicochemical model for brain penetration and its application to the design of centrally acting H2 receptor histamine antagonists. *Journal of Medicinal Chemistry*, 31(3), 656-671. DOI: 10.1021/jm00398a028
4. van de Waterbeemd, H.D., Kansy, M. (1992) Hydrogen-bonding capacity and brain penetration. *Chimia*, 46(7-8), 299-303.
5. Kelder, J., Grootenhuis, P.D.J., Bayada, D.M., Delbressine, L.P., Ploemen, J.P. (1999) Polar molecular surface as a dominating determinant for oral absorption and brain penetration of drugs. *Pharmaceutical Research*, 16(10), 1514-1519. DOI: 10.1023/A:1015040217741
6. Gleeson, M.P. (2008) Generation of a set of simple, interpretable ADMET rules of thumb. *Journal of Medicinal Chemistry*, 51(4), 817-834. DOI: 10.1021/jm701122q
7. Waring, M.J. (2009) Defining optimum lipophilicity and molecular weight ranges for drug candidates-Molecular weight dependent lower logD limits based on permeability. *Bioorganic & Medicinal Chemistry Letters*, 19(10), 2844-2851. DOI: 10.1016/j.bmcl.2009.03.109
8. Lipinski, C.A., Lombardo, F., Dominy, B.W., Feeney, P.J. (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Review*, 23(1-3), 3-26. DOI: 10.1016/S0169-409X(96)00423-1
9. Wager, T.T., Hou, X., Verhoest, P.R., Villalobos, A. (2010) Moving beyond rules: the development of a central nervous system multiparameter optimization (CNS MPO) approach to enable alignment of druglike properties. *ACS Chemical Neuroscience*, 1(6), 435-439. DOI: 10.1021/cn100008c
10. Rankovic, Z. (2015) CNS drug design: balancing physicochemical properties for optimal brain exposure. *Journal of Medicinal Chemistry*, 58(6), 2584-2608. DOI: 10.1021/jm501535r
11. Rankovic, Z. (2017) CNS physicochemical property space shaped by a diverse set of molecules with experimentally determined exposure in the mouse brain. *Journal of Medicinal Chemistry*, 60(14), 5943-5954. DOI: 10.1021/acs.jmedchem.6b01469
12. van de Waterbeemd, H.D., Camenisch, G., Folkers, G., Raevsky, O.A. (1996) Estimation of CACO-2 cell permeability using calculated molecular descriptors. *Quantitative Structure-Activity Relationships*, 15(6), 480-490. DOI: 10.1002/qsar.19960150604
13. van de Waterbeemd, H.D., Camenisch, G., Folkers, G., Chretien, J.R., Raevsky, O.A. (1998) Estimation of blood-brain barrier crossing of drugs using molecular size and shape, and H-bonding descriptors. *Journal of Drug Targeting*, 6(2), 151-165. DOI: 10.3109/10611869808997889
14. Raevsky, O.A., Grigorev, V.Y., Polianczyk, D.E., Sandakov, G.I., Solodova, S.L., Yarkov, A.V., Bachurin, S.O., Dearden, J.C. (2016) Physicochemical property profile for brain permeability: comparative study by different approaches. *Journal of Drug Targeting*, 24(7), 655-662. DOI: 10.3109/1061186X.2015.1132224
15. Raevsky, O.A. (2016) CNS multiparameter optimization approach: is it in accordance with Occam's razor principle? *Molecular Informatics*, 35(3-4), 94-98. DOI: 10.1002/minf.201500109
16. Raevsky, O.A., Polianczyk, D.E., Mukhametov, A., Grigorev, V.Y. (2016) Assessment of the classification abilities of the CNS multi-parametric optimization approach by the method of logistic regression. *SAR and QSAR in Environmental Research*, 27(8), 629-635. DOI: 10.1080/1062936X.2016.1212922
17. Raevsky, O.A. (2018) Hydrogen Bond Contribution to Drug Bioavailability: cheminformatics approach. *Biomedical Chemistry: Research and Methods*, 1(3), e00060. DOI: 10.18097/BMCRM00060
18. Raevsky, O.A., Solodova, S.L., Lagunin, A.A., Poroikov, V.V. (2014) Computer Modeling of Blood-Brain Barrier Permeability for Physiologically Active Compounds. *Biochemistry (Moscow) Supplement Series B: Biomedical Chemistry*, 60(2), 161-181. DOI: 10.18097/PBMC20146002161
19. Ooms, F., Weber, P., Carrupt, P.-A., Testa, B. (2002) A simple model to predict blood-brain barrier permeation from 3D molecular fields. *Biochimica et Biophysica Acta*, 1587(2-3), 118-125. DOI: 10.1016/S0925-4439(02)00074-1
20. Raevskij, O.A. (2015) Modelirovanie sootnoshenij "struktura-svoystva", Dobrosvet, M.
21. Singh, N., Chaudhury, S., Liu, R., AbdulHameed, M.D.M., Tawa, G., Wallqvist, A. (2012) QSAR Classification Model for Antibacterial Compounds and Its Use in Virtual Screening. *Journal of Chemical Information and Modeling*, 52(10), 2559-2569. DOI: 10.1021/ci300336v
22. Riniker, S., Wang, Y., Jenkins, J.L., Landrum, G.A. (2014) Using Information from Historical High-Throughput Screens

- to Predict Active Compounds. Journal of Chemical Information and Modeling, 54(7), 1880-1891. DOI: 10.1021/ci500190p
23. Iwata, H., Sawada, R., Mizutani, S., Yamanishi, Y. (2015) Systematic Drug Repositioning for a Wide Range of Diseases with Integrative Analyses of Phenotypic and Molecular Data. Journal of Chemical Information and Modeling, 55(2), 446-459. DOI: 10.1021/ci500670q
24. Yee, L.C., Wei, Y.C. (2012) Statistical Modelling of Molecular Descriptors in QSAR/QSPR, Vol. 2 (Eds: Dehmer, M., Varmuza, K., Bonchev, D.), Wiley-VCH, Verlag GmbH & Co. KGaA., 1-31.
25. SPSS Inc. Released 2008. SPSS Statistics for Windows, Version 17.0. Chicago: SPSS Inc.
26. DRAGON, version 5.5, Talete srl: Milano, Italy (2011).
27. Raevsky, O.A., Grigor'ev, V.Y., Trepalin, S.V. HYBOT program, registration by Russian State Patent Agency No. 990090 of 26.02.99

Поступила: 21. 08. 2018.
Принята к публикации: 10. 09. 2018.

BINARY CLASSIFICATION OF BLOOD-BRAIN BARRIER PENETRATION BY THE LOGISTIC REGRESSION METHOD

O.A. Raevsky, D.E. Polianczyk, O.E. Raevskaja*

Institute of Physiologically Active Compounds of the Russian Academy of Sciences,
1 Severny proezd, Moscow region, Chernogolovka, 142432 Russia; *e-mail: raevsky@ipac.ac.ru

Stable classification predictive models of 83 drugs with different blood-brain barrier penetration capacity have been constructed by the logistic regression method using physicochemical descriptors characterizing steric, electrostatic interactions and hydrogen bond energy. The models are balanced, with the prediction level of 75-80%.

Key words: QSAR; CNS; blood-brain barrier; binary classification; descriptors

ACKNOWLEDGMENTS

The work was performed within the framework of the state task for 2018 (topic number 0090-2017-0020).