УДК 544.165

КЛАССИФИКАЦИОННЫЕ МОДЕЛИ ПРОНИЦАЕМОСТИ ХИМИЧЕСКИХ СОЕДИНЕНИЙ В ЦЕНТРАЛЬУЮ НЕРВНУЮ СИСТЕМУ (CNS) МЕТОДОМ СТРУКТУРНОГО СХОДСТВА

Ярков А.В., Трепалин С.В., Григорьев В.Ю., Раевский О.А.

ФГБУН «Институт физиологически активных веществ РАН», Черноголовка, e-mail: yarkov@ipac.ac.ru

Проницаемость в центральную нервную систему (CNS) является ключевым фактором для создания новых лекарственных препаратов для лечения неврологических заболеваний. В настоящей работе проведена классификация проницаемости химических соединений в центральную нервную систему (CNS) с использованием методов структурного сходства (Structural Similarity) и среднеарифметического свойства (Arithmetic Mean Property). В основе метода структурного сходства лежит тот факт, что свойство исследуемого соединения определяется из известных свойств наиболее родственных структурных соединений. На основе этого построена классификационная модель проницаемости химических соединений в нервную систему, результаты классификации которой значительно лучше, чем в широко используемых эмпирических правилах «rule-5» и MPO. Полученная модель довольно проста и может быть использована на ранней стадии поиска новых лекарств.

Ключевые слова: проницаемость в CNS, методы классификации, структурное сходство

CLASSIFICAITION MODELS OF CHEMICAL SUBSTANCES BRAIN PENETRATION BY MEANS OF STRUCTURE SIMILARITY METHOD

Yarkov A.V., Trepalin S.V., Grigorev V.Y., Raevsky O.A.

Institute of Physiologically Active Compounds, Chernogolovka, e-mail: yarkov@ipac.ac.ru

The penetration into central nervous system (CNS) is the key factor for creation of new drugs to treat neurological deceases. In the present work the property of penetration of chemical substances into brain (CNS) was classified using the methods of Structure Similarity and Arithmetic Mean Property. In the basis of the structural similarity lies in the fact that the property of a compound is determined from the known properties of the most structural related compounds. Using above method the classification model was build and the results of classification were significantly better then obtained by empirical methods: «rule-5» and MPO. The resulting model is quite simple and can be used at an early stage of drug discovery.

Keywords: brain penetration CNS, classification methods, structure similarity

Проницаемость в центральную нервную систему (CNS) является ключевым фактором как для создания новых лекарственных препаратов для лечения неврологических заболеваний, так и для защиты нервной системы от вредных веществ. Использование методов QSAR позволяет оценить возможность проникновения в CNS новых веществ на предварительном этапе испытаний, что значительно сокращает затраты и позволяет оптимизировать работы в конструировании лекарств.

В 1997 году Липинским [6] было предложено «правило-5» для оценки того, чтобы вещество могло быть лекарством для CNS. Это правило включает 5 параметров (Молекулярный вес меньше 500, число атомов доноров и акцепторов водорода а также logP меньше 5, пятый параметр заключается в присутствии везде цифры 5). Это правило значительно повлияло на развитие медицинской химии и упоминается в сотнях публикаций. Позднее Вагер и др. [14] предложили более сложный способ оценки для исследования проницаемости в CNS – метод MPO (Multi Parameter Optimization). В методе использованы шесть физико-химических параметров (logP, logD, MW, TPSA, число атомов – доноров H и pKa для наиболее основного центра). Вместо порогового значения в методе MPO используются нормированные (от 0 до 1) значения каждого параметра, и результат определяется суммой этих нормированных параметров (от 0 до 6). В обзоре [10] описаны другие подходы с использованием других дескрипторов для построения QSAR модели проницаемости химических соединений в центральную нервную систему.

В основе метода структурного сходства лежит тот факт, что свойство исследуемого соединения определяется из известных свойств наиболее родственных структурных соединений (подход, аналогичный методу kNN, k-ближайшего соседа). В рамках этого подхода для каждого рассматриваемого соединения устанавливается ряд структурно-родственных соединений, и среднее арифметическое значение их свойства принимается за рассчитанное значение свойства изучаемого соединения. Выбор родственных соединений может быть про-

веден также с использованием и других дескрипторов, используемых при исследованиях QSAR, поэтому метод структурного сходства является частным случаем метода AMP (Arithmetic Mean Properties) [8, 9]. Результаты исследований данным методом показывают, что в большинстве случаев достаточно использование трех ближайших соседей. В качестве структурных дескрипторов могут быть использованы индексы схожести Танимото [12]. Таким образом, для каждой молекулы можно определить n наиболее похожих молекул, а исследуемому свойству приписать среднее свойство его ближайших соседей.

Целью данной работы является построение классификационной модели для проницаемости в центральную нервную систему с помощью структурных дескрипторов методом структурного сходства и сравнение полученных результатов с эмпирическими правилами Липински [6] и MPO [14].

Материалы и методы исследования

Созданная нами база данных содержит информацию о проницаемости в CNS 2294 уникальных соединений и процитирована из публикаций [1, 3, 4, 5, 11, 15]. Среди них 1159 соединений считались проницаемыми через CNS, 735 – нет. С помощью алгоритма, описанного в [12] и реализованного в СУБД CheD [13], все соединения были отсортированы в порядке убывания индекса разнообразия и отобрано по пятьсот соединений CNS+ и CNS-. Это позволило отобрать наиболее непохожие соединения для формирования обучающей и тестовой выборок. Каждое пятое соединение из отобранных пятисот было отобрано для формирования тестовой выборки. Таким образом, были сформированы выборки из 400 обучающих соединений и 100 тестовых. Кроме того, были сформированы также внешние тестовые выборки по 50 соединений – лекарств [2], для которых данные по проницаемости в CNS наиболее достоверны. Для полученных выборок была рассчитана матрица схожести Танимото Ti,j [12].

Программа AMP [8, 9] в качестве входной информации использует матрицу схожести Танимото, данные по активности CNS («+» или «–») и число ближайших соседей для расчета активности. В выходном файле для каждого соединения перечисляются номера ближайших структурных соседей и индексы схожести Танимото (Tc) с каждым соседом. Считается, что соединение имеет свойство CNS+ (или CNS-), если большая часть его соседей имеет активность «+» (или «-»).

Результаты исследования и их обсуждение

В таблице представлены результаты моделирования проницаемости в CNS методом структурного сходства и дано сравнение с эмпирическими правилами Липински (правило «5») [6] и MPO [14].

Приведенные данные демонстрируют значительное улучшение в точности предсказания проницаемости в CNS по сравнению с эмпирическим правилом – «5» и MPO. Немного выпадает точность в классификации CNS-внешней выборки [2], особенно для одного соседа – 0,62. Это может быть связано с тем, что в обучающей выборке недостаточно структурных фрагментов, представленных во внешней тестовой выборке. Остальные результаты близки к экспериментальной точности определения свойств и могут быть использованы при конструировании перспективных лекарств. Полученная модель очень проста, для ее использования не требуется знания сложных машинно-обучаемых методов и программ. Единственным параметром является структура изучаемого вещества. Кроме того, точ-

Результаты классификации проницаемости химических соединений в CNS методом структурного сходства.

k	Выборка	Число CNS+	Число CNS-	Точн. CNS+	Точн. CNS-	Точн. общая	Точн. «5»	Точн. MPO
1	Обуч.(800)	326	312	0,815	0,780	0,798	0,560	0,573
	Тест(200)	76	85	0,760	0,850	0,805	0,630	0,585
	Тест(100, вн.)	40	31	0,800	0,620	0,710	0,550	0,520
3	Обуч.(800)	322	314	0,805	0,785	0,795		
	Тест(200)	77	84	0,770	0,840	0,805		
	Тест(100, вн.)	36	36	0,720	0,720	0,720		
5	Обуч. (800)	315	308	0,788	0,770	0,779		
	Тест(200)	78	79	0,780	0,790	0,785		
	Тест(100, вн.)	35	38	0,700	0,760	0,730		

Примечания. k – число ближайших соседей, число CNS+ и CNS- количество правильно определенных соединений, точность CNS+ CNS- доля правильно классифицированных соединений, точность «5» и MPO – доля правильно классифицированных соединений методами Липински [6] и MPO [14].

ность прогноза легко может быть улучшена за счет расширения обучающей выборки и привлечения в нее более разнообразных соединений.

Список литературы

1. Adenot M., Lahana R. Blood-Brain Barrier Permeation Models: Determinating between Potential CNS and Non-CNS drugs Including P-Glycoprotein Subtrates // J. Chem .Inf. Comput. Sci. – 2004. – Vol. 44. – P. 239–248.

2. Ghose A.K., Herbertz T., Hudkins R.L., Dorsey B.D., Mallamo J.P. Knowledge-based, central nervous system (CNS) lead selection and lead optimization for CNS drug discovery // ACS Chem. Neurosci. – 2012. – Vol. 3. – P. 50–68.

3. Hou T.J., Xu X.J. ADME Evaluation in Drug Discovery. 3. Modeling Blood-Brain Barrier Partitioning Using Simple Molecular Descriptors // J. Chem. Inf. Comput. Sci. – 2003. – Vol. 43. – P. 2137–2152.

4. Kelder J., Grootenhuis P.D.J., Bayada D.M., Delbressine L.P.C., Ploemen J.P. 1999. Polar molecular surface as a dominating determinant for oral absorption and brain penetration of drugs // Pharm. Res. – 1999. – Vol. 16. – P. 1514–1519.

5. Kelly M., Mahar D. Passive Pearmibility and P-Glycoprotein-Mediated Efflux Differentiate Central Nervous System (CNS) and Non-CNS Marketed Drugs // J. Pharmacol. Exp. Therapeut. – 2002. M Vol. 303. № 3. – P. 1029–1037.

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. Del. Rev. – 1997. – Vol. 23. – P. 3–26.

7. Platts J.A., Abraham M.H., Zhao Y.H., Hersey A., Ijaz L., Butina D., Correlation and prediction of a large blood-brain distribution data set – an LFER study // 2001. Eur. J. Med. Chem. Vol. 36. – P. 719–730.

8. Raevsky O.A., Grigor'ev V.Yu., Modina E.A., Worth A. Prediction of Acute Rodent Toxicity on the Basis of Chemical Structure and Physicochemical Similarity // Molecular Informatics. – 2011. – P. 267–275.

9. Raevsky O.A., Grigor'ev V.Yu., Modina E.A., Worth A. Prediction of Acute Toxicity to Mice by the Arithmetic Mean Toxicity (AMT) Modelling Approach // SAR & QSAR in Environmental Research. – 2010. – Vol. 21. – P. 265–275.

10. Raevsky O.A., Solodova S.L., Lagunin A.A., Poroikov V.V. Computer Modeling of Blood Brain Barrier Permeability for Physiologically Active Compounds // Biochemistry (Moscow) Supplement Series B: Biomedical Chemistry. – 2013. – Vol. 7. – P. 95–107.

11. Rose K., L.H. Hall. Modeling Blood-Brain Barrier Partitioning Using the Electrotopological State// J. Chem. Inf. Comput. Sci. – 2002. – Vol. 42. – P. 651–666.

12. Trepalin S.V., Gerasimenko V.A., Kozyukov A.V., Savchuk N.Ph., Ivaschenko A.A. New Diversity Calculations Algorithms Used for Compound Selection// J. Chem. Inf. Comput. Sci. – 2002. – Vol. 42. – P. 249–258.

13. Trepalin S.V., Yarkov A.V., CheD: Chemical Database Compilation Tool, Internet Server, and Client for SQL Servers // J. Chem. Inf. Comput. Sci. – 2001. – Vol. 41. – P. 100–107.

14. Wager T.T., Hou X., Verhoest P.R., Villalobos A. Moving beyond rules: the development of a central nervous system multiparameter optimization (CNS MPO) approach to enable alignment of druglike properties // ACS Chem. Neurosci. – 2010. – Vol. 1. – P. 435–439.

15. 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 antagonist // J. Med. Chem. – 1988. – Vol. 31. – P.656–671.