

**К 40-летию Института физиологически активных веществ РАН****ПРОТОКОЛЫ ЭКСПЕРИМЕНТОВ, ПОЛЕЗНЫЕ МОДЕЛИ, ПРОГРАММЫ И СЕРВИСЫ****ПРОГРАММА HVBOT 3D MF И ЕЁ ИСПОЛЬЗОВАНИЕ В CoMFA****В.С. Скворцов<sup>1,2\*</sup>, А.В. Микурова<sup>1,2</sup>, О.А. Раевский<sup>1</sup>**<sup>1</sup>Институт физиологически активных веществ Российской академии наук, 142432, Черноголовка Московской обл., Северный проезд, 1; \*эл. почта: vladlen@ibmh.msk.su<sup>2</sup>Научно-исследовательский институт биомедицинской химии имени В.Н. Ореховича, 119121, Москва, Погодинская ул. 10

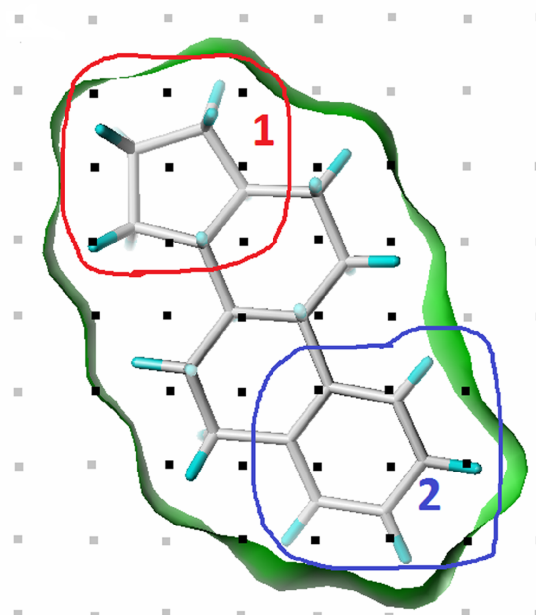
На основе модифицированного с использованием факторов водородного связывания HVBOT потенциала “6-8” разработана программа “HVBOT 3D MF”. Для интеграции программы с модулем “QSAR and Advanced CoMFA” пакета SYBYL-X создан набор макросов, написанных на SPL. Полноценная версия программы доступна по адресу <http://www.ibmc.msk.ru/HVBOT3D>. Показано, что для ряда наборов химических соединений, имеющих выраженные различия в части групп с потенциальной возможностью образовывать водородные связи, использование полей HVBOT 3D MF позволяет улучшить предсказательную силу моделей, созданных по технологии CoMFA.

**Ключевые слова:** QSAR; CoMFA; HVBOT; молекулярные поля**DOI:** 10.18097/BMCRM00073**ВВЕДЕНИЕ**

Использование различных подходов в области QSAR стало неотъемлемой частью современной вычислительной медицинской химии. Среди них особенно интенсивно развивается группа так называемых 3D методов, когда корреляции с активностями (свойствами) строят на основе определённых характеристик химических соединений, распределённых в пространстве. К таким характеристикам можно отнести различные молекулярные поля, такие как электростатическое и стерическое, используемые в одном из самых первых методов 3D QSAR – методе сравнительного анализа молекулярных полей (Comparative Molecular Field Analysis, CoMFA) [1], имеющем 30-летнюю историю. В данном методе для каждой молекулы согласно общепринятой методике (рис. 1) описывается дискретное молекулярное поле с установленными значениями в узлах трёхмерной решётки, которая описывает либо всё пространство вокруг выровненного набора молекул, либо его отдельную часть. В качестве дескрипторов используются суммы значений для выделенных групп узлов (компонент). Метод CoMFA широко используется и развивается и поныне [2]. На базе аналогичных полей построена также программа GRID (Graphic Retrieval and Information Display) [3, 4], в которой есть отличия в определении стерических полей и включены потенциалы водородной связи и гидрофобный потенциал. Потенциалы водородной связи впоследствии появились и в CoMFA, а гидрофобный потенциал можно было добавить как внешнее поле. Метод сравнительного анализа индексов молекулярного подобия (Comparative Molecular Similarity Indices Analysis, CoMSIA) [5, 6] представляет собой дальнейшее развитие CoMFA. В данном методе изменена методология расчёта молекулярного поля, когда в узлах решётки рассчитывают индексы молекулярного подобия, и в стандартной версии [SYBYL] имеется возможность

использовать электростатические, стерические, гидрофобные поля, а также поля водородных связей. Потенциалы в CoMSIA более плавные. Для поиска корреляций и в CoMFA, и в CoMSIA используется один и тот же стандартный метод статистического анализа PLS (Partial Least Squares или метод частичных наименьших квадратов).

Следует отметить, что наряду с методами описания молекул как набора дискретных дескрипторов, существуют и альтернативные способы, использующие наборы непрерывных гладких функций от пространственных координат и методы машинного обучения с использованием статистических ядер (kernels) [7], но рассматриваемые в работе молекулярные поля были адаптированы и испытаны только для дискретных подходов.

**Рисунок 1.** Пример формирования стерического поля для CoMFA и определение переменных (компонент).

В программе GRID для описания водородного связывания используется потенциал “6-8” [8], где изменения энергии при образовании водородной связи описываются следующим набором уравнений:

$$\Delta E = \Delta E_r f(p) f(t) \quad (1);$$

$$E_r = \frac{C}{r^8} - \frac{D}{r^8} \quad (2);$$

$$C = -3 E_m r_m^8 \quad (3);$$

$$D = -4 E_m r_m^6 \quad (4);$$

где  $E_m$  – оптимальная свободная энергия (ккал/моль),  $r$  – расстояние между донором и акцептором водорода ( $r_m$  – оптимальное, Å).

$$f(t) = \cos^2(t) \quad (5);$$

где  $t$  – угол между линиями водород/донор и акцептор/донор ( $^\circ$ ).

$$f(p) = \cos^2(p) \quad (6);$$

где  $p$  – угол между линиями неподелённая пара/акцептор и акцептор/донор ( $^\circ$ ).

При этом устанавливался ограниченный набор пар атомов донор/акцептор для которых были определены оптимальные значения свободной энергии водородной связи и оптимальные расстояния.

В работе [9] нами была предложена модификация потенциала “6-8”, основанная на факторах водородной связи, рассчитанных программой HYBOT [10]. При этом значения оптимальной свободной энергии водородной связи и оптимального расстояния зависели от произведения акцепторного ( $E_a$ ) и донорного ( $E_d$ ) факторов водородного связывания:

$$E_m = 1.33 E_a E_d \quad (7);$$

$$r_m = k_7 \log\left(\frac{k_4 - E_m}{E_m}\right) + k_8 \quad (8).$$

Коэффициенты  $k_4$ ,  $k_7$  и  $k_8$  были подобраны для каждой пары донор/акцептор сгруппированных по типу химических элементов. Например,  $k_4 = -15$ ,  $k_7 = 0.58$ ,  $k_8 = 2.43$  для варианта ‘ОН...О’;  $k_4 = -12$ ,  $k_7 = 0.69$ ,  $k_8 = 2.73$  для ‘ОН...N’;  $k_4 = -10$ ,  $k_7 = 0.94$ ,  $k_8 = 2.80$  для ‘NH...N’ и т.д.

Используя предложенную модификацию потенциала водородного связывания и молекулу воды в качестве “идеального” зонда можно рассчитать 2 типа молекулярных полей, описывающих акцепторную и донорную способности молекулы соответственно, и использовать эти поля в CoMFA. Соответствующее программное обеспечение, названное “HYBOT 3D MF” было создано и доступно по адресу <http://www.ibmc.msk.ru/HYBOT3D>.

## ВОЗМОЖНОСТИ ПРОГРАММЫ “HYBOT 3D MF”

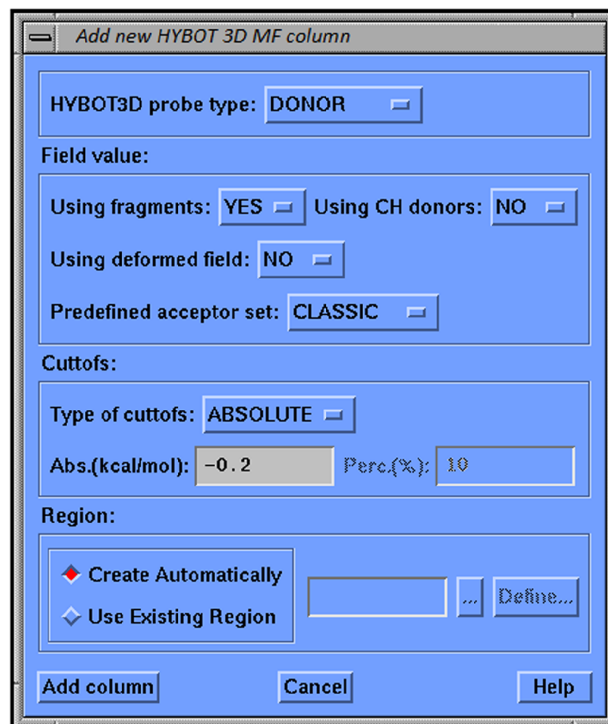
Программа “HYBOT 3D MF” написана на языке C++ в виде приложения, исполняемого из командной строки. Тестовая версия не имеет ограничений и доступна для операционной системы Windows 7 и старше. Для интеграции с модулем “QSAR and

Advanced CoMFA” пакета SYBYL-X [11] имеется набор макросов, написанных на SPL (SYBYL Programming Language). Макросы работоспособны и в более старых версиях программы SYBYL.

После инсталляции модуля “HYBOT 3D MF” в SYBYL-X в списке добавления “Computed column” для таблиц появляется опция “HYBOT 3D MF”, загружающая окно с параметрами для расчёта молекулярного поля (рис. 2).

Пользователь может выбрать тип поля. Если в качестве пробы используется молекула воды как донор водородного связывания, то поле описывает акцепторную способность молекулы, если же проба трактуется как акцептор водородного связывания – то донорную способность. В качестве параметров поля можно задать варианты расчётов факторов водородного связывания:

1. с использованием корректирующих молекулярных фрагментов или без них (только по подобию с молекулами в базе данных);
2. рассчитывать ли факторы для слабых –СН доноров (например, в ароматических системах);
3. определить список атомов – акцепторов водородных связей. CLASSIC – N, O; EXTENDED – N, O, S; FULL – N, O, S, галогены;
4. использовать ли “деформированное” поле, когда в качестве значения в узле решётки берётся не вычисленное в данной точки значение потенциала водородной связи, а минимальное значение, вычисленное в некоторой окрестности (бокс или сфера) на расстоянии в половину шага решётки.



**Рисунок 2.** Внешний вид окна, вызываемого при добавлении к таблице SYBYL-X молекулярного поля, описывающего донорную или акцепторную способность молекулы.

Пользователь может установить уровень, ниже которого значение потенциала водородного связывания обнуляется. Данный уровень устанавливается либо в абсолютных значениях в ккал/моль, либо в процентах от максимально возможного. Также пользователь должен определить регион, для которого рассчитывается поле. Автоматически поле рассчитывается для бокса, включающего все молекулы таблицы плюс 4Å по каждой из координат в обоих направлениях. Можно также подгрузить уже имеющийся файл SYBYL, описывающий нужный регион.

Помимо описанного способа настройку расчёта поля можно скорректировать через специфический набор переменных TAILOR. Вдобавок к уже описанным опциям имеется возможность определить шаг решётки по умолчанию, а также изменить штрафы, налагаемые за отклонение от идеальных углов (формулы 5 и 6). В стандартном варианте программы GRID эти штрафы пропорциональны квадрату косинуса угла ( $\cos^2$ ). В нашей программе допускаются также  $\cos$  и  $\cos^4$ .

После того как колонка с соответствующим полем загружена в таблицу, дальнейший анализ проводится по стандартной для CoMFA процедуре. При загрузке полей имеются некоторые ограничения, а именно: при изменении молекул в ассоциированной с таблицей базе данных автоматический перерасчёт полей HYBOT 3D MF не производится, то есть процедуру следует запускать по новой. Если ячейка не заполняется, то это может быть связано с тем, что в молекуле имеются незаполненные валентности, либо программа не может привести структуру к стандартному описанию. Последнее можно проверить, записав проблемную молекулу в формате MACCS, отключив возможность записи ароматических связей (тип AR). Если молекула не будет преобразована правильно, то, вероятнее всего, поле HYBOT 3D MF также не будет вычислено. Данный вариант был выбран для того, чтобы избежать ложных результатов.

## ПРИМЕРЫ ИСПОЛЬЗОВАНИЯ ПРОГРАММЫ “HYBOT 3D MF”

Следует отметить, что для того чтобы акцепторные и донорные поля помогли улучшить предсказательную силу QSAR моделей, в наборе молекул не только должны присутствовать потенциальные доноры и/или акцепторы водородного связывания, но и данные группы должны иметь при себе различные химические заместители. Если в группе молекул существует область, в которой у части из них имеется группировка C=O, но при этом в ближайшем окружении данной группы нет других различий (то есть акцепторная способность не изменяется от молекулы к молекуле), то выигрыша по сравнению с использованием стандартных донорно-акцепторных полей CoMFA, вероятнее всего, не будет.

В таблице 1 приведён пример использования поля HYBOT 3D MF на выборке из 21 стероида с известной аффинностью к тестостерон-связывающему глобулину (стандартный пример SYBYL [1]). Модель 9, использующая оба варианта полей HYBOT 3D MF, демонстрирует наилучшие показатели. Другой пример (табл. 2) – предсказание ингибиторной активности флавоноидов для альдозоредуктазы (EC 1.1.1.21) [12]. Отличительной особенностью данного исследования является то, что были использованы как CoMFA поля, так и CoMSIA. Поскольку эти два метода различаются только на стадии вычисления молекулярного поля, то использование их комбинации в одной модели вполне оправдано. В данном случае явного преимущества полей HYBOT 3D MF при обучении не наблюдается, но имеется выигрыш в предсказательной силе моделей, демонстрируемой на тестовой выборке. Кроме того, при обучении модели без использования полей HYBOT 3D MF дают наилучшие результаты при большем значении параметра “число компонент модели”, что может быть причиной переобучения модели.

**Таблица 1.** Статистические параметры CoMFA моделей для различных комбинаций переменных (стандартный пример SYBYL, таблица JACS, 21 стероид)

N	q <sup>2</sup>	n	r <sup>2</sup>	SE	F	Относительный вклад переменных				
						HD	HA	STER	ELEC	DA_S
1	0.416	1	0.598	0.764	29.21	1.0				
2	0.646	2	0.893	0.405	75.05		1.0			
3	0.685	3	0.922	0.356	66.56			1.0		
4	0.688	3	0.919	0.363	64.15				1.0	
5	0.687	3	0.940	0.312	88.63					1.0
6	0.688	2	0.915	0.361	96.77	0.314	0.686			
7	0.715	3	0.942	0.308	91.32			0.289	0.711	
8	0.726	3	0.952	0.279	112.42			0.099	0.243	0.658
9	0.763	3	0.962	0.247	144.27	0.325	0.357	0.093	0.225	

Примечание: HD - поле HYBOT 3D MF для донорного зонда; HA - поле HYBOT 3D MF для акцепторного зонда; STER - стерическое поле SYBYL; ELEC - электростатическое поле SYBYL; DA\_S - донорно-акцепторное поле SYBYL.

**Таблица 2.** Набор моделей, использующих CoMFA и CoMSIA для предсказания ингибиторной активности флавоноидов для альдозоредуктазы (EC 1.1.1.21) [12]

№	$N_L$	$N_T$	$N_{comp}$	$Q^2_{cv}$	$SE_{cv}$	$R^2_L$	$SE_L$	$ME_L$	F	$R^2_T$	$SE_T$	$ME_T$	Fields
1	55	14	4	0.65	1.11	0.92	0.55	0.33	134.82	0.78	0.90	0.71	S E
2	55	14	4	0.69	1.04	0.94	0.47	0.32	186.94	0.81	0.96	0.70	S E
3	55	14	7	0.73	1.01	0.98	0.25	0.18	412.64	0.79	1.05	0.59	S E D A
4	55	14	4	0.69	1.05	0.93	0.49	0.32	171.88	0.84	0.90	0.60	S E HD HA
5	55	14	3	0.69	0.74	0.88	0.62	0.50	90.81	0.83	0.76	0.53	S E H D A
6	55	14	5	0.66	1.10	0.94	0.48	0.34	146.66	0.79	0.97	0.59	iS iE
7	55	14	8	0.72	1.04	0.98	0.26	0.19	315.58	0.83	0.89	0.53	iS iE H
8	55	14	3	0.67	1.07	0.87	0.67	0.46	114.23	0.83	0.83	0.55	iS iE H iD iA
9	55	14	3	0.64	1.12	0.88	0.65	0.39	122.16	0.88	0.74	0.47	iS iE HD HA
10	55	14	3	0.72	0.99	0.92	0.52	0.36	202.08	0.86	0.81	0.49	iS iE H HD HA
11	55	14	7	0.72	1.03	0.99	0.20	0.14	624.63	0.83	0.92	0.51	iE H HD HA
12	55	14	3	0.59	1.19	0.82	0.79	0.41	77.32	0.88	0.70	0.52	HD HA

Примечание: № - номер модели;  $N_L$  - количество наблюдений в обучающей выборке;  $N_T$  - количество наблюдений в тестовой выборке;  $N_{comp}$  - число компонент;  $Q^2_{cv}$  -  $Q^2$  для скользящего контроля;  $SE_{cv}$  - стандартная ошибка для скользящего контроля;  $R^2_L$  -  $R^2$  обучения конечной модели;  $SE_L$  - средняя ошибка обучения конечной модели;  $ME_L$  - максимальная ошибка обучения конечной модели; F - критерий Фишера (значимость модели: чем больше, тем лучше);  $R^2_T$  -  $R^2$  предсказания;  $SE_T$  - средняя ошибка предсказания;  $ME_T$  - максимальная ошибка предсказания. CoMFA поля: S - стерическое поле; E - электростатическое; H - гидрофобное; D - донорное; A - акцепторное. CoMSIA поля: iS - стерическое поле; iE - электростатическое; H - гидрофобное; iD - донорное; iA - акцепторное. HYBOT 3D MF поля: HD - донорное; HA - акцепторное.

## ЗАКЛЮЧЕНИЕ

Подводя итог можно отметить, что для ряда наборов химических соединений, имеющих выраженные различия в части групп с потенциальной возможностью образовывать водородные связи, использование полей HYBOT 3D MF позволяет улучшить предсказательную силу моделей. Несмотря на то, что программа "HYBOT 3D MF" тестировалась, в первую очередь, в составе комплекса SYBYL, её можно применять и в других программах QSAR, использующих для анализа различные молекулярные поля, так как по сути это приложение, выполняемое из командной строки, особенности вызова которого описаны в руководстве пользователя. Полноценная версия программы "HYBOT 3D MF" для ОС WINDOWS доступна по адресу <http://www.ibm.msk.ru/HYBOT3D>.

## БЛАГОДАРНОСТИ

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

## ЛИТЕРАТУРА

1. Cramer, R.D., Patterson, D.E., & Bunce, J.D. (1988) Comparative molecular field analysis (CoMFA). 1. Effect of shape on binding of steroids to carrier proteins. *Journal of the American Chemical Society*, 110(18), 5959-5967. DOI: 10.1021/ja00226a005

2. Ansari, S.M., Palmer, D.S. (2018) Comparative Molecular Field Analysis Using Molecular Integral Equation Theory. *Journal of Chemical Information and Modeling*, 58(6), 1253-1265. DOI: 10.1021/acs.jcim.7b00600

3. Goodford, P.J. (1985) A computational procedure for determining energetically favorable binding sites on biologically important macromolecules. *Journal of medicinal chemistry*, 28(7), 849-857. DOI: 10.1021/jm00145a002

4. Cruciani, G., & Watson, K.A. (1994) Comparative molecular field analysis using GRID force-field and GOLPE variable selection methods in a study of inhibitors of glycogen phosphorylase b. *Journal of medicinal chemistry*, 37(16), 2589-2601. DOI: 10.1021/jm00042a012

5. Klebe, G., Abraham, U., & Mietzner, T. (1994) Molecular similarity indices in a comparative analysis (CoMSIA) of drug molecules to correlate and predict their biological activity. *Journal of medicinal chemistry*, 37(24), 4130-4146. DOI: 10.1021/jm00050a010

6. Klebe, G., & Abraham, U. (1999) Comparative molecular similarity index analysis (CoMSIA) to study hydrogen-bonding properties and to score combinatorial libraries. *Journal of computer-aided molecular design*, 13(1), 1-10. DOI: 10.1023/A:1008047919606

7. Zhokhova, N.I., Baskin, I.I., Bakhronov, D.K., Palyulin, V.A., & Zefirov, N.S. (2009, November) Method of continuous molecular fields in the search for quantitative structure-activity relationships. In *Doklady Chemistry* (Vol. 429, No. 1, p. 273). SP MAIK Nauka/Interperiodica.

8. Boobbyer, D.N., Goodford, P.J., McWhinnie, P.M., & Wade, R.C. (1989) New hydrogen-bond potentials for use in determining energetically favorable binding sites on molecules

of known structure. Journal of medicinal chemistry, 32(5), 1083-1094. DOI: 10.1021/jm00125a025

9. Raevsky, O.A., & Skvortsov, V.S. (2002) 3D hydrogen bond thermodynamics (HYBOT) potentials in molecular modelling. Journal of computer-aided molecular design, 16(1), 1-10. DOI: 10.1023/A:1016361910530

10. Raevsky, O.A., Grigor'ev, V.Ju., Trepalin, S.V., HYBOT program package, Registration by Russian State Patent Agency No. 990090 of 26.02.99.

11. SYBYL-X 2.1. Certara, Princeton, NJ, USA

12. Caballero, J. (2010) 3D-QSAR (CoMFA and CoMSIA) and pharmacophore (GALAHAD) studies on the differential inhibition of aldose reductase by flavonoid compounds. Journal of Molecular Graphics and Modelling, 29(3), 363-371. DOI: 10.1016/j.jmgm.2010.08.005

Поступила: 24. 08. 2018.

Принята к публикации: 07. 09. 2018.

## THE HYBOT 3D MF APPLICATION AND IT INTEGRATION WITH COMFA

*V.S. Skvortsov<sup>1,2\*</sup>, A.V. Mikurova<sup>1,2</sup>, O.A. Raevsky<sup>1</sup>*

<sup>1</sup>Institute of Physiologically Active Compounds of the Russian Academy of Sciences,  
1 Severny proezd, Moscow region, Chernogolovka, 142432 Russia; \*e-mail: vladlen@ibmh.msk.su

<sup>2</sup>Institute of Biomedical Chemistry, 10 Pogodinskaya str., Moscow, 119121 Russia

The program 'HYBOT 3D MF' based on the potential "6-8" modified by using HYBOT hydrogen bonding factors has been developed. An integration with "QSAR and Advanced CoMFA" module of the SYBYL-X is performed by set of macros written in SPL. A full working version of the program is available at <http://www.ibmc.msk.ru/HYBOT3D>. For a number of sets of chemical compounds having pronounced differences in groups with the hydrogen bonding potential, the use of HYBOT 3D MF fields makes it possible to improve the predictive power of models created by the CoMFA technology.

**Key words:** QSAR; CoMFA; HYBOT; molecular fields

### ACKNOWLEDGMENTS

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