

УДК 577.32

# Роль стэкинг-взаимодействий в комплексах белков с аденин- и гуанин-содержащими лигандами

Т. В. Пырков<sup>1, 2#</sup>, Д. В. Пыркова<sup>1</sup>, Е. Д. Балицкая<sup>1, 3</sup>, Р. Г. Ефремов<sup>1</sup><sup>1</sup>Учреждение Российской академии наук Институт биоорганической химии им. академиков М. М. Шемякина и Ю. А. Овчинникова<sup>2</sup>Московский физико-технический институт (государственный университет)<sup>3</sup>Московский государственный университет им. М. В. Ломоносова,

#e-mail: pyrkov@nmr.ru

## ВВЕДЕНИЕ

Биологическая функция белков часто определяется взаимодействием с низкомолекулярными соединениями – их лигандами и субстратами. При этом белки выступают, соответственно, в роли рецепторов и ферментов. Знание молекулярных основ взаимодействия рецепторов с лигандами, в частности – пространственной структуры комплекса белок-лиганд, является важнейшей предпосылкой к пониманию функции белка и его роли в биохимических процессах, регулирующих жизнедеятельность клетки. Кроме того, знание такой структуры дает возможность осуществлять на практике рациональное конструирование прототипов лекарственных соединений, которые, как правило, являются лигандами для определенных белков-мишеней в организме.

Инструментальные методы определения пространственной структуры белков и их комплексов с лигандами, такие как рентгеноструктурный анализ и спектроскопия ЯМР,

имеют ряд существенных ограничений. Часто получение структуры комплекса белка с лигандом оказывается затруднительным даже в случае, когда структура самого белка уже известна. Особенно остро проблемы, связанные с экспрессией и кристаллизацией, встают при исследовании трансмембранных белков, которые включают важнейший класс G-белок сопряженных рецепторов. Некоторый оптимизм внушают последние успехи на данном направлении – расшифровка структуры бета-адренергического и аденозинового рецепторов [1].

Технические трудности, сдерживающие применение экспериментальных подходов, привели к развитию методов компьютерного молекулярного моделирования. Одним из таких методов является «докинг» (от англ. docking – стыковка) – алгоритм, позволяющий вписать лиганд в трехмерную структуру белка-рецептора и оценить достоверность предсказанной таким образом структуры комплекса. На сегодняшний день метод до-

кинга активно применяют для исследования структурно-функциональной роли аминокислотных остатков белков, для поиска новых лекарственных соединений, селективно действующих на определенные белки-мишени [2, 3].

Развитие метода докинга происходит за счет применения новых алгоритмов конформационного поиска и новых методов оценки энергии межмолекулярных взаимодействий (оценочных функций, ОФ). ОФ могут иметь различный вид и включать либо термы межмолекулярных взаимодействий, заимствованные из силовых полей [2], либо эмпирические оценки (например, выявление водородных связей по геометрическим параметрам) [4]. В данной работе мы подробно остановимся на стэкинг-взаимодействиях, которые в широко используемых ОФ часто не рассматривают в явном виде.

### ПАРАМЕТРЫ СТЭКИНГ-ВЗАИМОДЕЙСТВИЙ

Среди различных типов контактов в биомолекулярных комплексах (водородные связи, электростатические взаимодействия и др.) стэкинг ароматических фрагментов заслуживает особого внимания. Большинство лекарственных препаратов содержат ароматические циклы, и стэкинг часто играет важную роль в молекулярном узнавании рецептор-лиганд. Ранее мы показали [5], что явный учет стэкинг-взаимодействий существенно повышает эффективность докинга АТФ. Стэкинг-контакты описывали функцией, зависящей от геометрических параметров взаимного расположения двух ароматических фрагментов – высоты  $h$  и угла  $\alpha$  между ними и сдвига  $d$  одного из колец относительно другого (рис. 1).

Диапазон этих параметров, определяющий наличие или отсутствие стэкинга, до сих пор остается не выясненным и в оценочных критериях выбирается достаточно произвольно [6, 7]. Его уточнение могло бы повысить эффективность оценки качества и достоверности структур белок-лиганд, предсказываемых методами молекулярного моделирования. С этой целью мы провели анализ экспериментально установленных пространственных структур атомного разрешения для комплексов различных белков с лигандами, содержащими наиболее распространенные пуриновые основания – аденин и гуанин.

Известный пример стэкинг-взаимодействий – параллельная упаковка азотистых оснований нуклеотидов в ДНК [8, 9]. Но некоторые ароматические соединения стремятся расположиться не только параллельно, но еще и перпендикулярно друг другу, как это показано для аминокислот в белках [7, 10] и в модельных системах, состоящих из простых углеводов – бензола, нафталина [11-14]. Кроме того, такие соединения имеют тенденцию участвовать в  $\pi$ -катионном взаимодействии, при котором образуется контакт между положительно заряженными группами и  $\pi$ -электронным облаком [15-17].

Поэтому мы исследовали распределение параметров  $h$  и  $d$  в зависимости от угла  $\alpha$  относительно азотистого основания лиганда для ароматических боковых цепей остатков Phe, Tyr, Trp и His, а также для положительно заряженных гуанидиновой группы Arg и аминокислотной группы Lys. На рис. 2 приведены результаты для лигандов, содержащих гуанин.

Показано, что для остатка Phe характерны два альтернативных положения над плоскостью гуанинового коль-

Рис. 1. Геометрические параметры, с помощью которых описывали стэкинг-взаимодействия между двумя ароматическими кольцами: смещение ( $d$ ) и высота ( $h$ ) центра одного из циклов относительно другого цикла; угол ( $\alpha$ ) между нормальными к плоскостям циклов

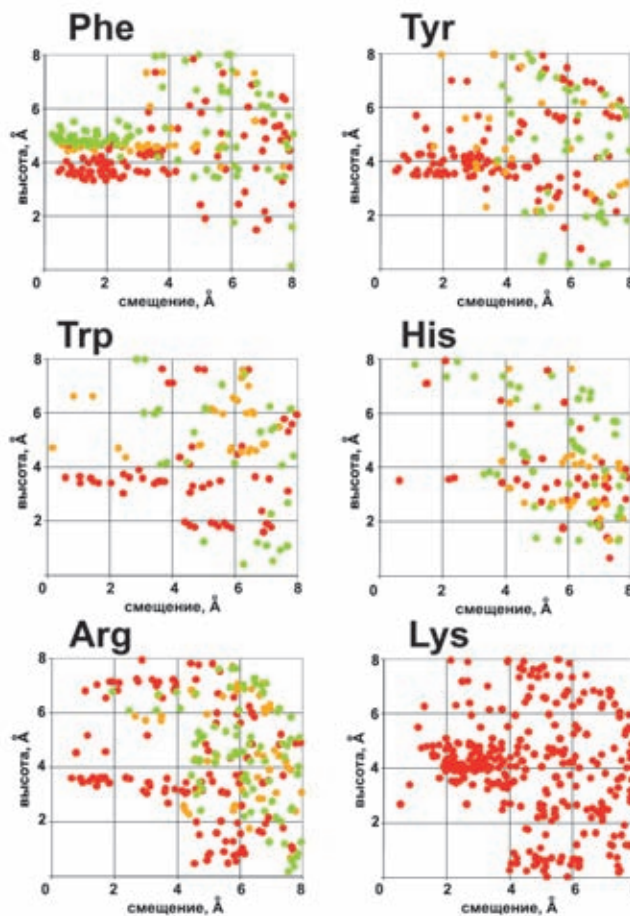
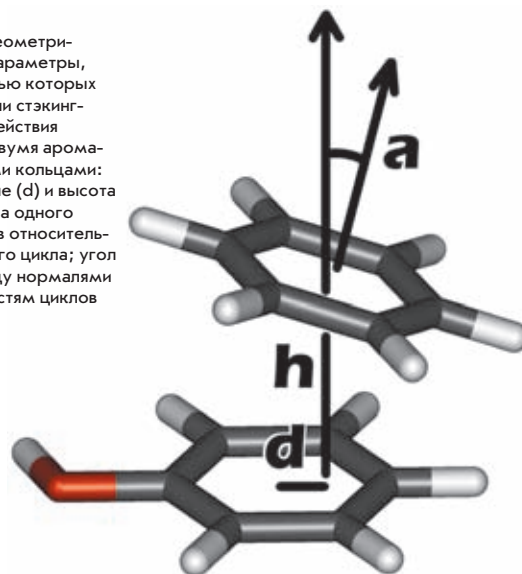


Рис. 2. Распределение центров ароматических колец и положительно заряженных боковых групп аминокислотных остатков белка-рецептора относительно центра гуанинового основания в комплексах белков с гуанин-содержащими лигандами. Красный цвет соответствует  $\cos^2\alpha = 0.0-0.4$  (параллельное расположение), зеленый – промежуточным значениям, где  $\alpha$  – угол между плоскостями ароматических циклов. Для Lys угол  $\alpha$  не определен

ца – параллельное и перпендикулярное (на рис. 2 показано красным и зеленым цветом соответственно). Для обоих типов контактов смещение  $d$  может меняться в достаточно большом диапазоне – до 3 Å. При этом два типа контактов хорошо различаются по высоте  $h$ , которая не превышает 4.5 Å для параллельной и 5.5 Å для перпендикулярной ориентации колец. Аналогичные распределения были получены и для остатков Tyr, Trp и His, хотя здесь тенденции проследить труднее ввиду меньшей распространенности данных остатков. Кроме того, видно, что для ароматических фрагментов остатков Tyr, Trp и His T-образная геометрия не так характерна, как для Phe.

Интересно, что распределение для гуанидиновой группировки Arg очень похоже на Tyr, Trp и His с преобладанием параллельной ориентации относительно азотистого основания. Для аминокетонной боковой цепи Lys характерно как положение над плоскостью гуанина ( $\pi$ -катионное взаимодействие), так и в его плоскости. Последнее объясняется образованием водородных связей с гетероатомами гуанинового основания.

Для аденина распределения в точности повторяют полученные для гуанина (данные не приведены). Полученные данные по геометрическим параметрам, описывающим стэкинг-взаимодействия, могут в дальнейшем быть использованы при разработке критериев оценки результатов докинга или включены в ОФ, используемые в процессе докинга.

### ГУАНИН-СПЕЦИФИЧНАЯ ОЦЕНОЧНАЯ ФУНКЦИЯ

Эффективность явного учета стэкинг-взаимодействий в ОФ продемонстрировали на примере оценки результатов докинга ГТФ для 14 комплексов с различными белками-рецепторами. Все структуры комплексов, сгенерированные с помощью докинга, были разделены на корректные и неправильные по значению среднеквадратичного отклонения (СКО) по координатам атомов гуанина (см. Методы). Были сконструированы несколько оценочных функций в виде линейной комбинации заранее заданных термов межмолекулярных взаимодействий. При этом все комплексы были разделены на две равные группы – обучающий и тестовый наборы. Весовые коэффициенты при термах подбирали методом линейной регрессии на значения дискретной

функции (1 – корректное решение, 0 – в противном случае) по результатам докинга для 7 комплексов обучающего набора. Для проверки устойчивости полученных весовых коэффициентов использовали метод перекрестной проверки путем изъятия из обучающего набора по одному комплексу. Относительная ошибка для всех коэффициентов при термах составила <30 %, что свидетельствует о надежности полученных результатов.

Новые ОФ были использованы для сортировки структур, сгенерированных с помощью докинга для каждого комплекса по отдельности. Видно, что эффективность ОФ *goldscore* [18], используемой в алгоритме докинга, составляет около 50 % – из возможных 7 корректные структуры выявлены только для 4 и 3 комплексов обучающего и тестового наборов соответственно (табл. 1). Эффективность сортировки по значению термина  $T_{stack}$  описывающего стэкинг-контакты, выше, чем по ОФ *goldscore*. Комбинирование ОФ *goldscore* и  $T_{stack}$  (ОФ1) также дает лучшие результаты, чем ОФ *goldscore*. Такой же эффект наблюдали при включении термина  $T_{stack}$  в ОФ, учитывающую липофильные контакты и водородные связи гуанин-белок (ОФ2 и ОФ3, табл. 1).

Наиболее эффективной из полученных ОФ является функция ОФ3. Число комплексов, где верное решение ранжировано на первом месте по ОФ3, заметно превышает те же показатели для стандартной функции *goldscore*. При этом также значительно улучшается и средний лучший ранг верного решения, т.е. качество ранжирования повышается в целом равномерно для всех комплексов. Это видно и из анализа результатов, полученных для каждого комплекса (рис. 3).

### ЗАКЛЮЧЕНИЕ

Анализ структурных данных по комплексам белков с гуанин- и аденинсодержащими соединениями позволил уточнить геометрические параметры стэкинг-взаимодействий лигандов с ароматическими боковыми цепями аминокислотных остатков рецептора. На примере докинга ГТФ было продемонстрировано, что явный учет стэкинг-взаимодействий в оценочных критериях позволяет повысить эффективность поиска достоверных (близких к нативной) структур комплексов белок-лиганд, предсказываемых с помощью методов компьютерного моделиро-

Табл. 1. Ранжирование решений докинга ГТФ

Метод ранжирования	Обучающий набор, 7 комплексов		Тестовый набор, 7 комплексов	
	Число комплексов, где верное решение ранжировано на 1 месте	Средний лучший ранг верного решения	Число комплексов, где верное решение ранжировано на 1 месте	Средний лучший ранг верного решения
<i>goldscore</i>	4	4.7	3	12.1
$T_{stack}$	5	2.3	3	7.0
$ОФ1 = -1.3 + 0.21 \times T_{stack} + 0.016 \times goldscore$	5	1.7	4	6.9
$ОФ2 = 0.06 + 0.007 \times T_{lipophilic} + 0.43 \times T_{h-bond}$	5	2.3	4	6.9
$ОФ3 = 0.05 + 0.004 \times T_{lipophilic} + 0.39 \times T_{h-bond} + 0.22 \times T_{stack}$	6	2.0	5	6.4

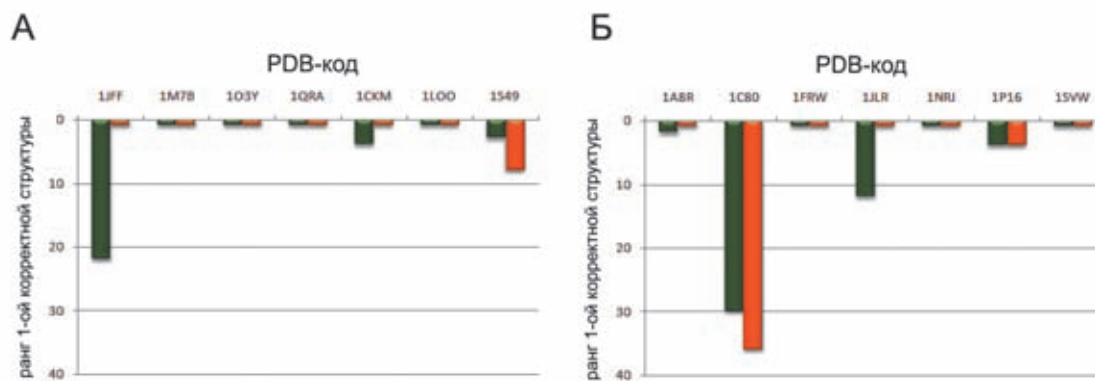


Рис. 3. Эффективность ранжирования результатов докинга ГТФ по ОФ goldscore (зеленый) и ОФ3 (красный) для обучающего (А) и тестового (Б) наборов. Показано лучшее по значению ОФ положение корректной структуры комплекса, сгенерированной с помощью процедуры молекулярного докинга

вания. В дальнейшем полученные результаты могут быть использованы для повышения точности докинга более широкого класса нуклеотид-содержащих соединений.

### МЕТОДЫ

Структуры комплексов гуанин- и аденинсодержащих лигандов с белками были взяты из базы данных PDB (Brookhaven Protein Data Bank) [19]. Веб-сервер PDBlig [20] использовали для поиска записей комплексов, лиганды в которых содержат аденин и гуанин как подструктуру. Лиганды с модифицированными азотистыми основаниями, а также структуры, содержащие цепочки нуклеиновых кислот, не рассматривали. Для выбора представительного набора рецепторов использовали кластеризацию по аминокислотной последовательности с помощью программы Clustalw [21] и выбирали структуру с наилучшим разрешением.

Докинг ГТФ проводили с помощью программы GOLD [18] с использованием функции ранжирования gold-

score. Использовали параметры докинга, установленные по умолчанию. Для каждой структуры белка-мишени было получено по 60 решений докинга. Предсказание докинга считали корректным, если СКО от экспериментальной структуры по координатам атомов гуанина составляло менее 2.5 Å.

Для расчета площади гидрофобного контакта ГТФ-белок (как меры гидрофобного взаимодействия,  $T_{\text{lipophilic}}$ ) использовали веб-сервер PLATINUM [22], смещение шкалы атомных констант гидрофобности взяли равным +0.2 по поверхности лиганда, что позволило более реалистично описать распределение гидрофобных/гидрофильных свойств ГТФ.

Терм  $T_{\text{h-bond}}$  представляет собой дискретную функцию, принимающую значение 1 в случае контакта гуанина с определенным мотивом водородных связей и значение 0 в противном случае. Такими мотивами считали водородные связи атомов N1, N2, O6 гуанина с остатками  $i$ ,  $i$ ,  $i$  либо  $i$ ,  $i$ ,  $i-2$ . ●

### Список литературы

- Hanson M.A., Stevens R.C. Structure. 2009. 17. 8-14.
- Kitchen D.B., Decornez H., Furr J.R., Bajorath J. Nat. Rev. Drug. Discov. 2004. 3. 935-949.
- Moitessier N., Englebienne P., Lee D., Lawandi J., Corbeil C.R. Br. J. Pharmacol. 2008. 153. S7-S26.
- Eldridge M.D., Murray C.W., Auton T.R., Paolini G.V., Mee R.P. J. Comput. Aided. Mol. Des. 1997. 11. 425-445.
- Pyrkov T.V., Kosinsky Y.A., Arseniev A.S., Priestle J.P., Jacoby E., Efremov R.G. PROTEINS 2007. 66. 388-398.
- Deng Z., Chuaqui C., Singh J. J. Med. Chem. 2004. 47. 337-344.
- Chakrabarti P., Bhattacharyya R. Prog. Biophys. Mol. Biol. 2007. 95. 83-137.
- Gotoh O. Adv. Biophys. 1983. 16. 1-52.
- Sponer J., Leszczynski J., Hobza P. J. Biomol. Struct. Dynam. 1996. 14. 117-135.
- Chelli R., Gervasio F.L., Procacci P., Schettino V. J. Am. Chem. Soc. 2002. 124. 6133-6143.
- Tsuzuki S., Honda K., Uchimaru T., Mikami M., Tanabe K. J. Am. Chem. Soc. 2001. 124. 104-112.
- Small D., Zaitsev V., Jung Y., Rosokha S.V., Head-Gordon M., Kochi J.K. J. Am. Chem. Soc. 2004. 126. 13850-13858.
- Sato T., Tsuneda T., Hirao K. J. Chem. Phys. 2005. 123. 104307.
- Malathy Sony S.M., Ponnuswamy M.N. Crystal Growth & Design 2006. 6. 736-742.
- Waters M.L. Curr. Opin. Chem. Biol. 2002. 6. 736-741.
- Meyer E.A., Castellano R.K., Diederich F. Angew. Chem. Int. Ed. 2003. 42. 1210-1250.
- Tewari A.K., Dubey R. Bioorg. Med. Chem. 2008. 16. 126-143.
- Jones G., Willett P., Glen R.C., Leach A.R., Taylor R.D. J. Mol. Biol. 1997. 267. 727-748.
- Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E. Nucleic. Acid. Res. 2000. 28. 235-242.
- Chalk A.J., Worth C.L., Overington J.P., Chan A.W. J. Med. Chem. 2004. 47. 3807-3816.
- Larkin M.A., Blackshields G., Brown N.P., Chenna R., McGettigan P.A., McWilliam H., Valentin F., Wallace I.M., Wilm A., Lopez R., Thompson J.D., Gibson T.J., Higgins D.G. Bioinformatics. 2007. 23. 2947-2948.
- Pyrkov T.V., Chugunov A.O., Krylov N.A., Nolde D.E., Efremov R.G. Bioinformatics. 2009. DOI: 10.1093/bioinformatics/btp111