Биомедицинская химия, 2015 том 61, вып. 6, с. 712-716.

УДК 004.094, 539.6, 577.3
©Коллектив авторов

СУПЕРКОМПЬЮТЕРНОЕ ИССЛЕДОВАНИЕ СПЕКТРА НИЗКОЭНЕРГЕТИЧЕСКИХ МИНИМУМОВ ЭНЕРГИИ СИСТЕМЫ БЕЛОК-ЛИГАНД

И.В. Оферкин[1], А.В. Сулимов[1,2], Е.В. Каткова[1,2]*, Д.К. Кутов[1,2], Ф.В. Григорьев[1,2], О.А. Кондакова[1,2], В.Б.Сулимов[1,2]

[1]ООО "Димонта", 117186, Москва, ул. Нагорная, д. 15, корп. 8
[2]Научно-исследовательский вычислительный центр,
Московский государственный университет имени М.В.Ломоносова,
119991, Москва, ул. Ленинские горы, д. 1, стр. 4; тел.: +7 (495) 939-40-04;
эл. почта: katkova@dimonta.com

Адекватный выбор целевой функции при проведении докинга может значительно повлиять как на точность позиционирования лиганда, так и на точность расчетов энергии связывания белка и лиганда. В данной работе проведена оценка пяти различных целевых функций, применяемых в докинге, включая функции, базирующиеся на силовом поле MMFF94, а также на квантовохимическом методе PM7, как с учетом, так и без учета растворителя (в рамках моделей растворителя PCM, COSMO, SGB). Для проведения подобной оценки мы сравнили положения лигандов в минимумах функций с экспериментально известными положениями лиганда в активном центре белка (так называемыми кристаллическими положениями лиганда), а также сравнили получившиеся в результате вычислений свободные энергии связывания с энергиями, полученными из экспериментальных данных. Каждая из функций была исследована на тестовом наборе, состоящем из 16 комплексов белок-лиганд. Для нахождения минимумов была разработана новая параллельная программа докинга FLM, использующая алгоритм Монте-Карло и осуществляющая исчерпывающий поиск низкоэнергетических минимумов, а также вычисление свободной энергии связывания. Полученные данные показывают, что целевая функция докинга на основе силового поля MMFF94 в вакууме может быть использована для обнаружения кристаллического или близких к нему позиций лиганда. Улучшение как точности позиционирования лиганда, так и корреляции значений энергий связывания со значениями, полученными в эксперименте, было показано при замене силового поля MMFF94 на новый метод PM7 с учётом воды в неявной модели COSMO.

Ключевые слова: молекулярное моделирование, докинг, силовые поля, квантовая химия, мультигармоническое приближение, суперкомпьютерные вычисления

DOI: 10.18097/PBMC20156106712

ВВЕДЕНИЕ

Компьютерное моделирование может существенно ускорить разработку новых лекарственных средств – ингибиторов заданных белков-мишеней [1]. Среди прочих для этих целей применяются программы докинга, осуществляющие позиционирование молекул-кандидатов (лигандов) в заданных белках-мишенях и оценивающие изменение свободной энергии при образовании комплекса белок-лиганд. Данная величина связана широко известным в термодинамике уравнением с константой равновесия; чем она больше по абсолютной величине, тем сильнее связывается лиганд, и тем эффективнее в большинстве случаев он работает как лекарственное соединение. К настоящему времени достигнуты определённые успехи в применении программ докинга для поиска и разработки ингибиторов [2, 3]. Однако, использование методов докинга не всегда может гарантировать положительные результаты. Это связано с тем, что ни одна программа докинга не обладает достаточно высокой точностью расчёта энергии связывания белок-лиганд, которая необходима, чтобы отличать слабые ингибиторы от средних, а средние – от сильных ингибиторов. Для этого точность расчётов энергии связывания должна быть не хуже 1 ккал/моль.

* - адресат для переписки

Парадигма, в рамках которой создаётся и существует большинство современных программ докинга, состоит в том, что положение лиганда в белке находится вблизи глобального минимума энергии системы белок-лиганд. В соответствии с этим, цель работы программы докинга – поиск глобального (или близкого к нему локального) минимума энергии комплекса белок-лиганд. В ходе поиска вычисляется свободная энергия связывания или хотя бы дается её грубая оценка (так называемая оценивающая или "скоринг" функция). Ситуация часто усугубляется тем, что целевые функции во многих программах докинга кроме энергетических вкладов (зависящих от выбранного силового поля) содержат дополнительные настраиваемые параметры. Эти параметры появляются перед каждым составляющим силового поля (электростатическое взаимодействие, Ван-дер-Ваальсово взаимодействие, взаимодействие с растворителем и т.д), а их значения определяются таким образом, чтобы для заданного множества комплексов значения целевых функций наилучшим образом согласовались с экспериментом. Эти подгоночные параметры, как правило, не имеют физического смысла, и их использование может затруднить априорную оценку точности позиционирования лигандов и точности вычисления энергии связывания белок-лиганд для заданного силового поля.

Чтобы выявить причины недостаточно высокой точности таких расчётов, мы провели большой объём вычислений на суперкомпьютере МГУ Ломоносов [4] с помощью нашей новой программы докинга FLM. В данной работе представлены результаты исследования точности докинга по позиционированию лиганда для разных целевых функций и сравнения экспериментальных энергий связывания белок-лиганд со значениями, рассчитанными на основе вычисления статистической суммы с учетом не одного глобального минимума, а целого спектра низкоэнергетических локальных минимумов и учёта в них колебаний атомов. Дана оценка правомерности использования докинга на основе поиска глобального минимума энергии системы белок-лиганд.

МЕТОДИКА

Программа докинга FLM

Программа докинга FLM (Find Local Minima, авторы: Оферкин И.В., Сулимов В.Б., Сулимов А.В.) [5] предназначена для поиска спектра низкоэнергетических локальных минимумов и расчёта энергии связывания системы белок-лиганд. Для поиска локальных минимумов энергии используется алгоритм Монте-Карло, формирующий пространство случайных начальных положений лиганда, и выполняется многократная локальная оптимизация структуры комплексов по положениям атомов лиганда.

Для вычисления изменения свободной энергии Гиббса молекулярной системы при образовании комплекса используется мультигармоническое приближение, в рамках которого потенциальная энергия молекулярной системы представляется совокупностью параболических ям, то есть совокупностью многомерных гармонических осцилляторов, каждый из которых соответствует одному из минимумов, найденных программой FLM. Это приближение похоже на метод "mining minima" Гилсона [6], но отличается от него главным образом тем, что конфигурационное пространство исследуется с помощью метода Монте-Карло более равномерно, чем в методе Гилсона, где поиск минимумов проводится вдоль направлений мод низкочастотных колебаний.

Анализ зависимости числа обновлений набора низкоэнергетических минимумов от числа проведённых локальных оптимизаций показал, что максимальная величина в 100000 оптимизаций достаточна практически для всех комплексов, включая лиганды, имеющие порядка 70 атомов (найденное значение энергии глобального минимума не обновляется после 100000 итераций).

Программа FLM не может конкурировать с существующими лучшими программами докинга по быстродействию, но её задача – найти весь спектр низкоэнергетических локальных минимумов в заданной области фазового пространства комплекса белок-лиганд.

Тестовый набор комплексов белок-лиганд

В качестве тестового использовался набор из 16 комплексов, структуры которых есть в базе PDB [7] с разрешением 2 Å и лучше, для которых известны константы ингибирования:

- 4 комплекса белка CHK1 (Checkpoint kinase 1) (4FT0, 4FT9, 4FSW, 4FTA);
- 2 комплекса белка ERK2 (Extracellular signal-regulated kinase 2) (4FV5, 4FV6);
- 2 комплекса белка тромбина (1DWC, 1TOM);
- 6 комплексов белка урокиназы (1C5Y, 1F5L, 1O3P, 1SQO, 1VJ9, 1VJA);
- 2 комплекса белка фактора Xa (2P94, 3CEN).

В этих комплексах лиганды представлены набором разнообразных молекул от маленьких (4FSW лиганд содержит 26 атомов) до больших и более конформационно подвижных (1VJ9 лиганд содержит 74 атомов). На основании этих комплексов был сформирован тестовый набор лигандов. Далее мы будем называть положение лиганда, взятое из базы данных PDB, кристаллическим положением, не анализируя, совпадает ли это экспериментально известное по данным рентгеноструктурного анализа положение лиганда с его положением в живой системе. Кристаллические положения лигандов в комплексах были локально оптимизированы, и в тестовый набор вошли только те комплексы, в которых RMSD (среднеквадратичное отклонение) между кристаллическим положением лиганда и локально оптимизированным кристаллическим положением лиганда было меньше 1,5 Å, так что

оптимизированное кристаллическое положение лиганда может представлять саму позу, полученную из PDB.

РЕЗУЛЬТАТЫ И ОБСУЖДЕНИЕ

Позиционирование лигандов – исследование разных целевых функций

В данной работе мы исследовали целевые функции на основе силового поля MMFF94 [8] и квантовохимического метода PM7 [9]. Вычисления проводили "в вакууме", учитывая растворитель в неявных моделях: PCM (модель проводящего континуума), COSMO (модель экранировки проводникового типа) и S-GB (поверхностная обобщенная модель Борна) [10, 11]. Все функции были использованы без подгоночных параметров на одном и том же наборе комплексов белок-лиганд.

Обозначим как "{1}MMFF94" набор минимумов, найденный программой FLM при выполнении локальных оптимизаций лиганда в комплексе (белок конформационно неподвижен) и при отборе низкоэнергетических минимумов в силовом поле MMFF94 в вакууме. Далее, энергии данных отобранных низкоэнергетических минимумов были пересчитаны с учётом взаимодействия с растворителем в модели PCM [11], обозначим данный набор как "{1}MMFF94+PCM". Следующий набор минимумов "{1}PM7" был получен при локальной оптимизации энергии комплексов из набора "{1}MMFF94" с помощью программы MOPAC [12] в рамках метода PM7 [9] в вакууме. Затем энергии этих минимумов были пересчитаны с учётом растворителя в модели COSMO с помощью MOPAC, и этот набор обозначен как "{1}PM7+COSMO".

Отбор минимумов был проведён и с учётом растворителя: FLM проводила локальную оптимизацию в вакууме в рамках MMFF94, но отбор в набор низкоэнергетических минимумов осуществлялся только после пересчёта энергии комплекса с учётом растворителя в модели PCM [11]. Такой набор мы обозначили как "{2}MMFF94+PCM". Энергии этих минимумов были пересчитаны и в вакууме – набор "{2}MMFF94", и с учётом растворителя в обобщённой модели Борна [10] – набор "{2}MMFF94+SGB".

Для анализа получившихся результатов ранжируем набор минимумов по энергии и припишем каждому минимуму номер, начиная с самого низкого по энергии, которому дадим номер 1. Обозначим как INN индекс самого низкоэнергетического минимума, среди тех, которые находятся в пространстве близко (RMSD < 2 Å) к положению лиганда в кристалле. Добавим к набору минимумов еще один, получающийся при локальной оптимизации лиганда из его начального положения в кристалле, и ранжируем по энергии все минимумы, начиная с самого низкого по энергии с номером 1. Обозначаем как IN индекс оптимизированного кристаллического положения в этом списке. Значения IN=1 и INN=1 говорят о том, что для данного комплекса выбранная целевая функция соответствует парадигме докинга, то есть её глобальный минимум находится вблизи положения лиганда в кристалле (INN=1), а полученное в результате оптимизации координат из кристалла положение имеет самую низкую энергию.

Результаты расчётов показали (таблица), что энергия в рамках силового поля MMFF94 в вакууме может быть использована в качестве целевой функции для докинга только для трёх комплексов: 1C5Y, 1F5L и 1SQO. Только для этих комплексов IN=1 и INN=1. Для остальных комплексов энергия оптимизированного кристаллического состояния выше энергии найденного глобального минимума на значительную (по сравнению с kT) величину: от 4,7 ккал/моль для комплекса 1VJA до 91,2 ккал/моль для комплекса 4FTA. Тем не менее, в этом наборе низкоэнергетических минимумов для большинства комплексов имеется близкий к кристаллическому положению минимум (INN=1, 2, …, 28, 131). Только для 4 комплексов (4FTA, 4FV6, 1DWC, 1TOM) можно с уверенностью сказать, что силовое поле MMFF94 в вакууме не подходит для докинга: для них энергия оптимизированного кристаллического положения выше энергий всех (1024) найденных низкоэнергетических минимумов (IN=*inf*), и среди них нет минимумов близких к кристаллическому положению (INN=*inf* за исключением комплекса 1DWC).

Из таблицы также видно, что использование квантовохимического метода PM7 в целом несколько улучшает позиционирование. Например, для комплексов 4FTA, 4FV6 и 1DWC индекс IN изменяется от *"inf*" (то есть все низкоэнергетические минимумы из соответствующего набора лежат ниже оптимизированного кристаллического положения лиганда) до конечного значения по сравнению с полем MMFF94. Однако наиболее сильное улучшение результатов (снижение значений индексов IN и INN) достигается при учёте растворителя, в том числе при отборе минимумов в низкоэнергетический набор. Оказалось, что более точная модель PCM работает хуже, чем обобщенная модель Борна. Возможно, это связано с итерационным решением уравнений и слишком грубыми аппроксимациями для матриц в модели PCM [11], в том время как метод S-GB более гладкий (расчёт энергии осуществляется по явной формуле), то есть дает близкие значения энергии взаимодействия с растворителем для близких конфигураций комплекса.

Вычисление свободной энергии связывания, сравнение с экспериментом

Проведенные расчёты энергии связывания с учётом всех найденных низкоэнергетических минимумов для комплекса и свободного лиганда дали следующие результаты. Во-первых, рассчитанные в мультигармоническом приближении с учётом энтропийных вкладов (учитывались потери степеней свободы на трансляцию и вращение лиганда как целого) свободные энергии связывания ΔG_{bind} для всех 16 комплексов значительно отличаются

Таблица. Значения индексов IN / INN для всех 16 тестированных комплексов и всех 7 полученных наборов минимумов. Символ "inf" для IN означает что все (1024) минимума из соответствующего набора имеют энергии ниже энергии оптимизированного кристаллического положения лиганда. Символ "inf" для INN означает, что все низкоэнергетические минимумы из соответствующего набора лежат далеко (RMSD > 2 Å) от кристаллического положения лиганда.

PDB ID	{1}MMFF94	{1}MMFF94+PCM	{1}PM7	{1}PM7 +COSMO	{2}MMFF94	{2}MMFF94+PCM	{2}MMFF94+SGB
4FT0	36 / 20	8 / 7	37 / 12	1 / 1	180 / 99	164 / 159	8 / 6
4FT9	45 / 28	1 / 1	25 / 6	1 / 1	194 / 125	3 / 1	1 / 1
4FSW	5 / 5	6 / 6	40 / 40	12 / 13	110 / 102	134 / 140	21 / 3
4FTA	*inf* / *inf*	4 / *inf*	379 / *inf*	1 / *inf*	*inf* / *inf*	186 / 187	97 / 97
4FV5	204 / 131	3 / 3	253 / 194	2 / 1	186 / 134	6 / 3	5 / 5
4FV6	*inf* / *inf*	1 / *inf*	49 / *inf*	1 / *inf*	86 / 289	3 / 68	1 / 24
1DWC	*inf* / 670	245 / 25	689 / 661	158 / 141	245 / 114	250 / 35	107 / 8
1TOM	*inf* / *inf*	13 / *inf*	*inf* / *inf*	1 / *inf*	*inf* / *inf*	13 / 4	7 / 1
1C5Y	1 / 1	2 / 1	7 / 1	2 / 1	1 / 1	2 / 1	1 / 1
1F5L	1 / 1	1 / 1	43 / 16	69 / 30	1 / 1	10 / 1	1 / 1
1O3P	20 / 18	21 / 1	5 / 1	3 / 1	69 / 62	274 / 1	130 / 2
1SQO	1 / 1	2 / 1	1 / 1	1 / 1	1 / 1	54 / 1	5 / 1
1VJ9	46 / 1	86 / 51	32 / 1	26 / 8	6 / 1	11 / 18	10 / 14
1VJA	42 / 3	7 / 1	7 / 4	6 / 4	4 / 49	1 / 2	1 / 1
2P94	36 / 2	19 / 1	23 / 6	7 / 1	22 / 1	35 / 1	21 / 1
3CEN	96 / 1	18 / 1	13 / 1	3 / 1	90 / 1	35 / 1	13 / 1

от экспериментальных значений этой величины ΔG_{exp}, полученных пересчётом экспериментально измеренных констант ингибирования, взятых из литературы. Например, при использовании минимумов из набора "{1}MMFF94" наиболее близкие к эксперименту вычисленные значения ΔG_{bind} были равны -16,3 (-6,0) и -4,6 (-10,6) ккал/моль для комплексов 1C5Y и 1SQO соответственно (в скобках значения ΔG_{exp}). Для некоторых же комплексов расхождение расчёта и эксперимента было очень значительно: +82,2 (-10,1) и -224,1 (-11,8) для 4FT0 и 1TOM соответственно. При этом вклад от всех минимумов кроме глобального составил для всех комплексов величины не более 1,5 ккал/моль, так что ΔG_{bind} определялась только вкладом от глобальных минимумов энергии комплекса и свободного лиганда. Вклад от колебаний в глобальных минимумах составил величины от десятых долей до нескольких единиц ккал/моль, а энтропийный вклад (положительный) от потери лигандом вращательных и трансляционных степеней свободы слабо менялся от комплекса к комплексу и составлял величины близкие к 20 ккал/моль, так что главный специфический вклад в свободную энергию связывания определялся в первую очередь значениями потенциальных энергий в глобальных минимумах комплекса и свободного лиганда. Коэффициент корреляции Пирсона между ΔG_{bind} и ΔG_{exp} был на уровне 0,5 и существенно не изменялся как при учёте растворителя, так и при расчётах энтальпии связывания с помощью квантовохимического метода PM7 вместо силового поля MMFF94. Тем не менее, и учёт растворителя, и переход к квантовохимическому описанию системы значительно уменьшил разброс рассчитанных энергий связывания: 311 (для {1}MMFF94), 87 (для {1}MMFF94+PCM), 226 (для {1}PM7) и 36 ккал/моль (для {1}PM7+COSMO), при том, что значения экспериментальных энергий связывания ΔG_{exp} для этих же 16 комплексов находились всего лишь в диапазоне -13,0 – -6,0 ккал/моль. Кроме того, для варианта "{1}PM7+COSMO" все значения ΔG_{bind} получались отрицательными, а их коэффициент корреляции Пирсона с ΔG_{exp} составил 0,6.

Отметим, что при выбранном нами здесь способе вычисления энергии связывания энергия внутренних напряжений лиганда (другими словами энергия деформации лиганда при связывании) учитывалась автоматически – это разность энергии, рассчитанной для свободного лиганда в его глобальном минимуме, и для конформации лиганда, полученной в глобальном минимуме комплекса. Для нашего тестового набора комплексов эта энергия составила значительные величины – от нескольких единиц до нескольких десятков ккал/моль.

ЗАКЛЮЧЕНИЕ И ВЫВОДЫ

Таким образом показано, что для силового поля MMFF94 в вакууме парадигма докинга выполняется только для трёх из 16 комплексов: оптимизированное кристаллическое положение представляет собой глобальный минимум энергии по сравнению со всеми найденными минимумами, и найденный при поиске глобальный минимум находится вблизи закристаллизованного положения лиганда. Заметное улучшение позиционирования наблюдается при учёте взаимодействия с водой в неявных

моделях, и имеется некоторое улучшение позиционирования при переходе от силового поля MMFF94 к квантовохимическому описанию молекул (PM7). Показано, что исследование спектра низкоэнергетических минимумов с помощью программы FLM представляет собой инструмент апробирования адекватности того или иного способа расчёта энергии комплекса белок-лиганд в качестве целевой функции докинга. Кроме того, полученные наборы минимумов можно использовать для тестирования алгоритмов докинга.

Рассчитанные в мультигармоническом приближении свободные энергии связывания сильно отличаются от "экспериментальных" значений и плохо с ними коррелируют для всех исследованных энергетических функций. Ближе всего к эксперименту значения энергии связывания, рассчитанные с помощью метода PM7 с учётом воды в неявной модели COSMO. Основной вклад в свободную энергию связывания даёт её энтальпийная составляющая, которая в программе FLM определяется разностью потенциальных энергий комплекса, с одной стороны, и свободного белка и лиганда – с другой, в их глобальных минимумах с учётом взаимодействия с водой. Проведенные исследования показали, что для повышения точности докинга необходимо учитывать взаимодействие молекул с растворителем, а также искать вместо MMFF94 другое силовое поле и/или реализовать докинг с помощью методов квантовой химии.

Работа выполнена при поддержке гранта РНФ № 15-11-00025.

ЛИТЕРАТУРА

1. *Садовничий В.А., Сулимов В.Б.* (2009) в: Суперкомпьютерные технологии в науке, образовании и промышленности (под редакцией Садовничего В.А. , Савина Г.И. и Воеводина Вл.В.), Издательство Московского университета, Москва, сс. 16-23.
2. *Sinauridze E.I., Romanov A.N., Gribkova I.V., Kondakova O.A., Surov S.S., Gorbatenko A.S., Butylin A.A., Monakov M.Yu., Bogolyubov A.A., Kuznetsov Yu.V., Sulimov V.B., Ataullakhanov F.I.* (2011) PLoS ONE, **6**(5), e19969.
3. *Bustanji Y., Al-Masri I.M., Qasem A., Al-Bakri A.G, Taha M.O.* (2009) Chem. Biol. Drug Des., **74**, 258-265.
4. Moscow University Supercomputing Center, Moscow, Russia, http://hpc.msu.ru/.
5. Dimonta, Ltd, Moscow, Russia, http://dimonta2.1gb.ru/ru/node/56.
6. *Chen W., Gilson M.K., Webb S.P., Potter M.J.* (2010) J. Chem. Theory Comput., **6**, 3540-3557.
7. *Berman H.M., Westbrook J., Feng Z., Gilliland G., Bhat T.N., Weissig H., Shindyalov I.N., Bourne P.E.* (2000) Nucl. Acids Res., **28**, 235-242. http://www.pdb.org/.
8. *Halgren T.A.* (1996) J. Comput. Chem., **17**, 490-519.
9. *Stewart J.J.P.* (2013) J. Mol. Model., **19**, 1-32.
10. *Купервассер О.Ю., Жабин С.Н., Мартынов Я.Б., Федулов К.М., Офёркин И.В., Сулимов А.В., Сулимов В.Б.* (2011) Вычислительные методы и программирование, **12**, 246-261.
11. *Михалев А.Ю., Офёркин И.В., Оселедец И.В., Сулимов А.В., Тыртышников Е.Е., Сулимов В.Б.* (2014) Вычислительные методы и программирование, **15**, 9-21.
12. MOPAC2012, James J.P. Stewart, Stewart Computational Chemistry, Colorado Springs, CO, USA, http://OpenMOPAC.net

Поступила: 29. 07. 2015.

SUPERCOMPUTER INVESTIGATION OF THE PROTEIN-LIGAND SYSTEM LOW-ENERGY MINIMA

I.V. Oferkin[1], A.V. Sulimov[1,2], E.V. Katkova[1,2], D.K. Kutov[1,2], F.V. Grigoriev[1,2], O.A. Kondakova[1,2], V.B. Sulimov[1,2]

[1]Dimonta, Ltd., Moscow, Russia
[2]Research Computer Center, Moscow State University,
Leninskie Gory 1, bld 4, Moscow, 119992 Russia; tel.: +7 (495) 939-40-04; e-mail: katkova@dimonta.com

The accuracy of the protein-ligand binding energy calculations and ligand positioning is strongly influenced by the choice of the docking target function. This work demonstrates the evaluation of the five different target functions used in docking: functions based on MMFF94 force field and functions based on PM7 quantum-chemical method accounting or without accounting the implicit solvent model (PCM, COSMO or SGB). For these purposes the ligand positions corresponding to the minima of the target function and the experimentally known ligand positions in the protein active site (crystal ligand positions) were compared. Each function was examined on the same test-set of 16 protein-ligand complexes. The new parallelized docking program FLM based on Monte Carlo search algorithm was developed to perform the comprehensive low-energy minima search and to calculate the protein-ligand binding energy. This study demonstrates that the docking target function based on the MMFF94 force field can be used to detect the crystal or near crystal positions of the ligand by the finding the low-energy local minima spectrum of the target function. The importance of solvent accounting in the docking process for the accurate ligand positioning is also shown. The accuracy of the ligand positioning as well as the correlation between the calculated and experimentally determined protein-ligand binding energies are improved when the MMFF94 force field is substituted by the new PM7 method with implicit solvent accounting.

Key words: molecular modeling, docking, force field, quantum chemistry, multiwell approximation, high-performance computing