

УДК 004.051; 539.6; 544.18; 615.01

ПРИМЕНЕНИЕ КВАНТОВО-ХИМИЧЕСКОГО ПОЛУЭМПИРИЧЕСКОГО МЕТОДА РМ7 ДЛЯ РАЗРАБОТКИ НОВЫХ ИНГИБИТОРОВ УРОКИНАЗЫ

Е. В. Каткова¹, И. В. Офёркин², В. Б. Сулимов³

Представлены результаты использования полуэмпирических квантово-химических методов с помощью программного комплекса МОРАС применительно к задаче поиска новых ингибиторов урокиназы. Проведено исследование оптимальных условий работы в рамках комплекса МОРАС с методом параметризации РМ7, позволяющим учитывать различные виды нековалентных межмолекулярных взаимодействий, таких как дисперсионные взаимодействия, галогенные и водородные связи. Показано, что различные способы расстановки водородов на первоначальные рентгено-кристаллические структуры значительно влияют на энергии взаимодействия белка и лиганда и, как следствие, немаловажную роль играет подвижность водородов белка в процессе локальной оптимизации комплекса. На 22 комплексах, включающих в себя уже не только урокиназные белки, была продемонстрирована важность учета растворителя, а также было показано, что расчет энергий взаимодействия методом квантовой химии дает лучшую корреляцию с экспериментом по сравнению с расчетами в силовом поле MMFF94. Для нативных комплексов урокиназа–ингибитор рассчитаны энтальпии связывания лиганд–белок и найдены коэффициенты корреляции с экспериментом для результатов расчетов в комплексе МОРАС. Приведены результаты использования комплекса МОРАС в разработке новых ингибиторов урокиназы.

Ключевые слова: молекулярное моделирование, квантовая химия, полуэмпирические методы, поиск новых ингибиторов, эффективность, оптимальные параметры работы, сеточный докинг, прямой докинг.

1. Введение. Компьютерный (*in silico*) дизайн новых ингибиторов для заданных белков-мишеней в настоящее время становится важнейшей составляющей процесса рациональной разработки новых лекарственных препаратов [1] и новым инструментом исследований в молекулярной биологии и фармацевтике. Использование компьютерного молекулярного моделирования значительно сокращает число соединений, которые необходимо синтезировать и экспериментально тестировать в процессе разработки нового лекарства. В последнее время стали появляться публикации, в которых молекулярное моделирование сыграло существенную роль при разработке новых ингибиторов для заданных белков-мишеней [2, 3].

Ключевым звеном в разработке нового лекарства является поиск вещества, способного связаться с выбранным белком-мишенью и заблокировать патологические функции данного белка. Однако несмотря на то что было разработано множество компьютерных подходов для оценки свободной энергии связывания белок–лиганд [4–11], на данный момент нет ни одного эффективного и надежного метода, позволяющего в общем случае достаточно аккуратно (с точностью лучше 1 ккал/моль или 0.04 эВ) рассчитывать измеряемые в экспериментах свободные энергии связывания лигандов с белками-мишенями, а также иметь возможность проводить скрининг виртуальных библиотек химических соединений.

В отличие от чисто эмпирических или статистических методов, более строгие молекулярные структурно-ориентированные подходы сильно зависят от реалистичного и точного описания взаимодействия белок–лиганд. К ним относятся “free-energy perturbation” (возмущение свободной энергии [12, 13]), “linear-response approximation” (приближение линейного отклика [14]) и комбинирование расчетов молекулярной механики с континуальными моделями растворителя (Molecular Mechanics/Poisson–Boltzmann surface

¹ ООО “Димонта”, ул. Нагорная, д. 15, корп. 8, 117186, Москва; мл. науч. сотр., e-mail: katkova@dimonta.com

² Научно-исследовательский вычислительный центр, Московский государственный университет им. М.В. Ломоносова, Ленинские горы, д. 1, стр. 4, 119992, Москва; вед. программист, e-mail: io@dimonta.com

³ Научно-исследовательский вычислительный центр, Московский государственный университет им. М.В. Ломоносова, Ленинские горы, д. 1, стр. 4, 119992, Москва; зав. лабораторией, e-mail: vladimir.sulimov@gmail.com

area [15, 16]). Данные методы, как и методы докинга (позиционирования лиганда в активном центре белка), основаны на использовании силовых полей (силовое поле — это набор потенциалов, описывающих взаимодействия между атомами как классическими частицами) и, таким образом, включают в себя как их преимущества, так и недостатки. Возрастающие мощности современных компьютеров и разработка новых компьютерных методологий делают уже сейчас возможным также и применение более точных квантово-химических вычислений для нескольких тысяч атомов, что *a priori* может давать более строгое и полное описание молекулярной системы белок–лиганд, чем описание с помощью силовых полей. В настоящей статье представлены результаты одного из первых применений полуэмпирического квантово-химического метода PM7, реализованного в программном комплексе MOPAC (Molecular Orbital PACkage) [17, 18], для расчета энергий связывания белка и лиганда и поиска новых ингибиторов урокиназы. Исследовано, насколько сильно такие факторы, как подвижность атомов белка и учет растворителя, влияют на результаты расчета энергии связывания. Показано, что расчеты с помощью примененного нами квантово-химического метода дают лучшую корреляцию с экспериментом, чем силовое поле MMFF94 [19]. Проведенные исследования позволили успешно применить пакет квантово-химических программ MOPAC в комбинации с докингом для разработки новых ингибиторов урокиназы.

2. Методы.

2.1. MOPAC: новое дыхание полуэмпирических методов.

2.1.1. Линейная скалируемость с ростом размера системы. Применение неэмпирических (*ab initio*) квантово-химических методов имеет ограничение по размеру исследуемой системы в несколько десятков атомов (для методов DFT (Density Functional Theory) — несколько сотен атомов), поэтому для задачи вычисления энергии взаимодействия между белком (тысячи атомов) и лигандом (десятки атомов) могут быть реально применены пока только полуэмпирические квантово-химические методы. Эти методы занимают промежуточное положение между точными *ab initio* методами квантовой химии и методами молекулярной механики. Они гораздо более универсальны, чем методы молекулярной механики, и в значительной степени опережают неэмпирические методы по скорости вычислений благодаря использованию приближенных расчетов в тех случаях, где требуются значительные вычислительные ресурсы, и благодаря эмпирическим параметрам, определенным таким образом, чтобы предсказанные результаты наилучшим образом соответствовали экспериментальным данным. Однако и для них вопрос скалируемости (т.е. зависимости времени счета от размера системы) остается актуальным. Программный комплекс MOPAC [17, 18] позволяет проводить расчеты молекул, содержащих несколько тысяч атомов (например, белки), благодаря встроенному модулю MOZYME [20]. Причина, по которой использовать полуэмпирические методы для больших систем представляется нецелесообразным, заключается в способах решения уравнений самосогласованного поля. Для полуэмпирических методов используется приближение LCAO (Linear Combination of Atomic Orbitals), в котором молекулярные орбитали ψ_i представляют собой линейные комбинации атомных орбиталей ϕ_λ : $\psi_i = \sum_{\lambda=1}^M c_{\lambda i} \phi_\lambda$, где суммирование ведется по числу используемых атомных орбиталей M .

В результате того, что молекулярные орбитали зависят от матрицы Фока, а матрица Фока зависит от молекулярных орбиталей, уравнения самосогласованного поля должны решаться самосогласованно методом итераций. В этом приближении матрица плотности P , составленная из функций молекулярных орбиталей, имеет вид

$$\hat{P} = 2 \sum_{i=1}^N c_{\lambda i} c_{\sigma i}.$$

Для системы из N атомов число атомных орбиталей и число занятых молекулярных орбиталей возрастают пропорционально N . Таким образом, поскольку число элементов в матрице плотности возрастает пропорционально N^2 , а число операций, требуемых для вычислений каждого матричного элемента, пропорционально N , то полное число операций, требуемых для вычислений, возрастает как N^3 . Вследствие этого в приближении LCAO даже при использовании сравнительно быстрых полуэмпирических методов расчеты систем, содержащих более 1000 атомов, практически невозможны.

Модуль MOZYME [20], появившийся в программном комплексе сравнительно недавно (в версии MOPAC2009), заменяет обычное приближение LCAO на метод локализованных молекулярных орбиталей LMO (Localized Molecular Orbital). Это дает ряд преимуществ: например, все взаимодействия, включающие в себя локализованные молекулярные орбитали, разделенные большим расстоянием, автоматически обнуляются, причем чем больше система, тем большую роль играет этот фактор.

Таким образом, благодаря использованию преимущества локализованных молекулярных орбиталей

уравнения самосогласованного поля могут быть решены за времена, пропорциональные размеру системы. В результате модуль MOZYME позволяет в рамках полуэмпирических методов, реализованных в пакете MORAS, оптимизировать геометрию молекулярных систем, имеющих до 15 000 атомов, и проводить вычисления в одной точке для систем, содержащих до 18 000 атомов.

2.1.2. Включение поправок на межмолекулярные взаимодействия. Одним из недостатков полуэмпирических квантово-химических методов долгое время оставалось плохое описание дисперсионных взаимодействий и водородных связей даже в наиболее популярных и надежных методах на основе приближения NDDO (Neglect of Diatomic Differential Overlap — пренебрежение двухатомным дифференциальным перекрытием): AM1 [21], PM3 [22], PM6 [23] и RM1 [24]. Эти методы базируются на получаемом из уравнения Хартри–Фока уравнении Рутаана, которое в матричной форме записывается следующим образом: $\hat{F}\hat{C} = \hat{S}\hat{C}\hat{\epsilon}$, где \hat{F} — матрица Фока, \hat{S} — матрица перекрытия, $\hat{\epsilon}$ — диагональная матрица орбитальных энергий, c_i — матрица коэффициентов разложения орбиталей по базису.

Матрица Фока может быть представлена в виде

$$F_{\mu\nu} = H_{\mu\nu}^{\text{core}} + \sum_{\lambda\sigma} P_{\lambda\sigma} \left[(\mu\nu|\sigma\lambda) - \frac{1}{2}(\mu\lambda|\sigma\nu) \right],$$

где $(\mu\nu|\lambda\sigma) = \int d\mathbf{r}_1 d\mathbf{r}_2 \phi_\mu^*(\mathbf{r}_1)\phi_\nu(\mathbf{r}_1) \frac{1}{r_{12}} \phi_\lambda^*(\mathbf{r}_2)\phi_\sigma(\mathbf{r}_2)$, $H_{\mu\nu}^{\text{core}}$ — матрица одноэлектронного гамильтониана, а $P_{\lambda\sigma}$ — матрица плотности.

В методе NDDO матрица перекрытия заменяется единичной матрицей, что позволяет существенно сократить количество требуемых к вычислению интегралов, поскольку в этом случае нулю приравняются интегралы, описывающие перекрытие атомных орбиталей, центрированные на разных атомах.

Полная энергия системы выражается формулой

$$E_{\text{tot}} = E_0 + \sum_A \sum_{B>A} \frac{Z_A Z_B}{R_{AB}},$$

где суммирование ведется по всем атомам A и всем атомам $B > A$, Z_A и Z_B — заряды ядер, а E_0 — электронная энергия молекулы на каждом шаге итераций:

$$E_0 = \frac{1}{2} \sum_{\mu} \sum_{\nu} P_{\nu\mu} \left(H_{\mu\nu}^{\text{core}} + F_{\mu\nu}(\hat{P}) \right).$$

Для вычисления оставшихся интегралов в полуэмпирических методах используются формулы, содержащие эмпирические параметры, и параметризация проводится таким образом, чтобы воспроизводить экспериментальные данные. В результате исследования метода PM6 [23] на модельных системах, включающих в себя биологические макромолекулы и кристаллические решетки, были выявлены некоторые ошибки. Метод PM7 [25] параметризовался так, чтобы эти ошибки не возникали. Кроме того, этот метод позволил ввести более аккуратное согласование дальних и ближних взаимодействий.

Метод PM7 на данный момент является последним по времени создания в этом семействе методов NDDO. В нем значительно улучшено описание дисперсионных взаимодействий, а также образование водородных и галогенных связей за счет введения еще на стадии параметризации соответствующих поправок, опирающихся на аналогичные поправки, успешно применяемые в последние годы для улучшения описания межмолекулярных взаимодействий в методе функционала плотности DFT [26, 27]. В первую очередь идея применения таких поправок на дисперсионные взаимодействия была успешно апробирована на уже существующем методе PM6 [23]. Позднее были также успешно апробированы и поправки к PM6 для описания водородных и галогенных связей [28–31]. Несмотря на многие преимущества, без этих поправок метод PM6 терпел неудачи при описании нековалентных межмолекулярных взаимодействий, но включение дополнительных эмпирических поправок для всех групп элементов и переходных металлов на дисперсионные взаимодействия, галогенные и водородные связи позволили значительно улучшить ситуацию (PM6-DH+ [31], PM6-DH2 [29, 30], PM6-DH2X [28]). Дополнительный дисперсионный член для парного межатомного взаимодействия выглядит следующим образом:

$$\sum E_{\text{disp}} = - \sum f_{\text{damp}}(r_{ij}, R_{ij}^0) C_{6ij} r_{ij}^{-6},$$

где r^{-6} — быстро убывающий член, позволяющий избежать перекрытия с квантово-механическим потенциалом, который описывает должным образом отталкивание на коротких расстояниях; R_{ij}^0 — атомные

ван-дер-ваальсовы радиусы; C_6 — коэффициенты, также являющиеся атомными параметрами и не зависящие от квантово-химического метода; параметры демпфирующей функции f_{damp} оптимизируются таким образом, чтобы воспроизводить энергии взаимодействия комплексов, в которых присутствуют дисперсионные связи.

Поправки на галогенные взаимодействия определяются по формуле

$$E_{\text{corr}} = ae^{-br},$$

где r — расстояние между галогеном и акцептором, а a и b — параметры.

Поправка, описывающая водородные связи, записывается в форме

$$E_{\text{H-bond}} = \left[a \times \frac{q_A \times q_H}{r^b} + c \times d^r \right] \times \cos \theta \times \cos \phi \times \cos \psi,$$

где q_A и q_H — заряды атомов акцептора A и водорода H ; r — расстояние между этими атомами; θ — угол $A \cdots H - D$ (D — атом донора); ϕ и ψ — отклонения торсионных углов $R_2 - A \cdots H$ и $R_1 R_2 - A \cdots H$ от идеальных оптимальных значений для водородной связи; a , b , c и d — параметры.

Эти поправки вошли в метод PM7 самосогласованным образом еще на стадии его параметризации, для которой были использованы экспериментальные и высокоуровневые *ab initio* расчетные данные для более чем 9000 соединений. Добавление функции дисперсии [27] и дополнительных членов, ответственных за водородные связи, значительно улучшило точность предсказаний межмолекулярных взаимодействий в PM7. На нескольких тестах для нековалентных взаимодействий было показано, что с учетом данных поправок достигается точность вычислений, характерная для современных DFT-D (DFT с дисперсионной коррекцией) подходов, примененных для этого типа задач. Кроме того, сравнение с *ab initio* расчетами показывает, что использование таких поправок позволяет предсказать энергию взаимодействия галогенных связей с ошибкой менее чем 10%.

2.1.3. Вычисление энтальпии связывания белок–лиганд с помощью МОРАС. Подбор оптимальных параметров работы. Для того чтобы рассчитать энергию взаимодействия белка и лиганда в МОРАС, структура комплекса белок–лиганд оптимизировалась методом PM7 при использовании инструмента MOZYME (программный комплекс МОРАС2012) в декартовых координатах. При этом при поиске оптимальных параметров этой оптимизации использовалась нативная структура комплекса, т.е. структура белка с лигандом из базы данных PDB (Protein Data Bank), а при поиске новых ингибиторов мы использовали положение лиганда в белке, полученное при докинге. Геометрия свободных лигандов для дальнейшей оптимизации в МОРАС находилась с помощью программы поиска глобального минимума FLM (см. ниже).

После оптимизации комплекса мы разделяли его на белок и лиганд и снова проводили оптимизацию уже отдельных молекул. Энтальпия взаимодействия вычислялась следующим образом:

$$\Delta H_{\text{bind}} = H_{\text{complex}} - H_{\text{protein}} - H_{\text{ligand}},$$

где H_{complex} — энтальпия образования, полученная при оптимизации комплекса в МОРАС; H_{protein} — энтальпия образования, полученная при оптимизации белка в МОРАС; H_{ligand} — энтальпия образования, полученная при оптимизации лиганда в МОРАС из положения, найденного в программе FLM. При проведении дальнейшего анализа мы сравнивали эти значения ΔH_{bind} со значением $R \times T \times \ln(K_i)$ — логарифмом констант ингибирования белков соответствующими лигандами, где константы ингибирования берутся из экспериментальных данных.

Оптимизация в МОРАС возможна в различных условиях. Одним из важных аспектов расчета энергии взаимодействия белок–лиганд является взаимодействие с растворителем. МОРАС предоставляет возможность включить в расчеты неявную модель растворителя в рамках приближения COSMO (COnductor-like Screening Model), когда весь растворитель заменяется непрерывной средой с бесконечной диэлектрической проницаемостью — металлом; например, для воды это оправдано, так как ее проницаемость при комнатной температуре составляет 78.5. Однако подобное включение в процессе оптимизации, во-первых, существенно увеличивает время счета, а во-вторых, может внести дополнительные погрешности, связанные с погрешностями модели растворителя. Еще одним вариантом является оптимизация системы белок–лиганд в вакууме и добавление растворителя на последнем шаге. Кроме того, программный комплекс МОРАС позволяет фиксировать заданные пользователем атомы, положение которых в процессе оптимизации будет оставаться неизменным. Оптимизация всего комплекса белок–лиганд требует значительных вычислительных ресурсов, однако может оказаться непрактичной, поскольку движение атомов белка,

находящихся вдалеке от центра связывания, может быть интересно с точки зрения задачи фолдинга (сворачивания в пространственную структуру) белка, однако не принесет пользы для задачи расчета энергии взаимодействия белка с лигандом. Поэтому в процессе оптимизации имеет смысл оставлять свободными атомы лиганда и фиксировать атомы белка, находящиеся от лиганда на расстоянии больше заданного. Отдельный интерес может представлять также свободное движение атомов водорода белка.

2.2. Программа прямого докинга FLM. Нами использована программа прямого бессеточного докинга FLM (Finder of Local Minima) [32], которая отыскивает различные локальные минимумы системы белок–лиганд в вакууме методом Монте-Карло: начальное состояние системы задается при помощи случайных торсионных деформаций лиганда и его вращений-трансляций как целого, затем проводится локальная оптимизация энергии лиганда как функции его декартовых координат. Взаимодействие лиганда с белком описывается непосредственно в рамках силового поля MMFF94 без каких-либо упрощений и заранее рассчитанных сеток потенциалов. В результате работы программы находятся не более 1024 низкоэнергетических локальных минимумов лиганда в белке или в свободном состоянии. По найденным локальным минимумам рассчитывается свободная энергия молекулярной системы в мультигармоническом приближении. Найденные программой FLM локальные минимумы использовались в дальнейшем в качестве начальных конфигураций при локальной оптимизации с помощью MORAC методом RM7.

2.3. Структура белка и лигандов. Подготовка. Урокиназная система активно вовлечена в процессы ремоделирования и регенерации сосудов и тканей, однако в случае патологических процессов эти механизмы оказываются задействованными также и в опухолевом росте. Урокиназная система начинает также играть роль в развитии новообразований, включая регуляцию опухолевого ангиогенеза, деление, адгезию и миграцию злокачественных клеток, метастазирование опухолей [33–35]. Кроме того, урокиназная система является мощным независимым прогностическим фактором в случае многих патологий, например в случае рака молочной железы. Это означает, что ее компоненты также могут быть подходящими мишенями для противоопухолевой терапии [36]. В настоящее время одним из эффективных терапевтических подходов представляется ингибирование протеолитической функции урокиназы. Таким образом, задача разработки новых ингибиторов протеолитического центра урокиназы представляется в настоящее время актуальной.

Все вычисления проводились на комплексах урокиназа–лиганд. Урокиназа представляет собой многофункциональный мультидоменный 53 kDa белок, состоящий из 411 остатков. В структуре урокиназы выделяют три домена: N-концевой домен, подобный эпидермальному фактору роста (EGF) и отвечающий за связывание с uPAR (рецептором урокиназы), крингл-домен и C-концевой каталитический домен, осуществляющий превращение плазминогена в плазмин. Трехмерные структуры белков урокиназы были взяты из базы данных PDB [37]. Из найденных 60 структур урокиназы были закристаллизованы в комплексе с прямыми обратимыми, нековалентными ингибиторами протеолитического центра (рис. 1). Десять комплексов урокиназа–лиганд, содержащих разные по размеру и активным группам соединения, мы выбрали для нахождения оптимальных параметров работы программного комплекса MORAC применительно к нашей задаче. Все комплексы были подготовлены к работе следующим образом: из .pdb-файла были исключены все молекулы растворителя, структуры белка и лиганда были разделены в отдельные файлы, а для белка и лиганда были добавлены водороды программой Aplite [38] в соответствии с заданными зарядовыми состояниями. После этого из них формировались .dat-файлы для работы с MORAC. Кроме того, нами использовались ненативные лиганды. Для этих лигандов также было добавлено водороды и получены трехмерные структуры.

3. Результаты и обсуждение.

3.1. Результаты учета растворителя. Для 10 нативных комплексов урокиназа–лиганд мы провели расчеты как с учетом растворителя, так и без него, а также и с учетом растворителя только на последнем шаге оптимизации от нативного положения, которое было начальным при оптимизации. Сама оптимизация проводилась в декартовых координатах только по атомам лиганда, а атомы белка оставались фиксированными. Оценка результатов проводилась с помощью вычисления коэффициента корреляции между рассчитанными с помощью MORAC энтальпиями связывания и экспериментальными энергиями

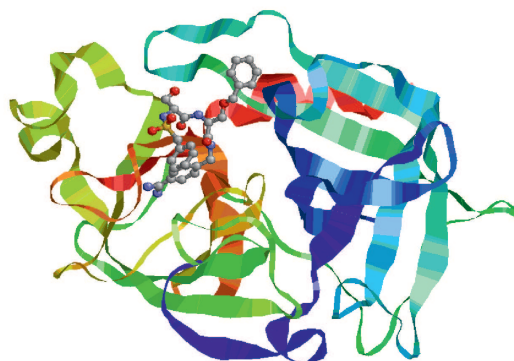


Рис. 1. Структура комплекса 1vj9 — белок урокиназы и закристаллизованный с ней лиганд (редактор RasMol)

связывания, полученными из констант ингибирования. Кроме того, мы сравнивали среднеквадратичное отклонение оптимизированного лиганда от нативного положения. Во всех случаях локальная оптимизация (причем как в растворителе, так и без него) не уведила далеко геометрию лиганда, и RMSD (Root Mean Square Deviation) оказывалось меньше 1 Å, причем RMSD относительно нативного положения почти во всех случаях, хотя и не намного (не более чем на десятые доли ангстрема), но уменьшалось при оптимизации в растворителе.

Если не задавать дополнительных параметров оптимизации, то расчеты с растворителем в MORAC часто останавливаются на больших значениях градиента энергии, чем расчеты без растворителя. Поэтому для того чтобы точности вычислений энергии совпадали для разных случаев, мы в каждом случае задавали градиент энергии, при котором оптимизация заканчивается, равным 1 ккал/моль/Å. При расчетах в растворителе нужно обратить также внимание на тот факт, что при использовании меньших значений градиента энергии могут возникнуть проблемы со сходимостью.

Время, которое потребовалось для оптимизации структуры с растворителем, во всех случаях оказывалось больше времени, которое потребовалось для оптимизации без растворителя. Для небольших лигандов эта разница составила около часа (для комплекса 1f5k: 4.5 часа в отсутствие растворителя и 5.5 часов в присутствии растворителя), для больших до 14 часов (для комплекса 2viv: 13 часов в отсутствие растворителя и 27 часов в присутствии растворителя). Значения энтальпий образования комплекса, белка и лиганда по отдельности при добавлении растворителя становятся отрицательнее.

Далее, мы сравнивали коэффициенты корреляции между энтальпиями связывания белок–лиганд в MORAC для всех случаев с энергиями, рассчитанными из экспериментальных данных. В случае оптимизации без растворителя этот коэффициент составил 0.53, в случае оптимизации с растворителем — 0.55, в случае добавления растворителя на последнем шаге оптимизации — 0.53, как и в случае без растворителя. Отсюда видно, что использование COSMO в процессе оптимизации не намного улучшает корреляцию с экспериментом, что, вероятно, может быть связано с неточностями самой континуальной модели растворителя.

Тем не менее нужно заметить, что величина диапазона изменения энергий связывания при учете растворителя (от -100 до -50 ккал/моль) оказывается меньше величины диапазона изменений энергий связывания при расчетах без растворителя (от -230 до -100 ккал/моль), что приближает значения энергий взаимодействия к экспериментальным результатам и указывает на то, что растворитель в значительной степени экранирует энергию кулоновских взаимодействий.

3.2. Результаты учета подвижности атомов белка. Мы также провели расчеты энтальпии связывания не только для фиксированных атомов белка, но и для ситуаций, когда в процессе оптимизации некоторые атомы белка имеют подвижность. Оптимизация здесь проводилась в декартовых координатах. Один из таких случаев — это оптимизация по всем атомам лиганда и по всем легким атомам белка (водородам). Рассматривались также случаи, когда свободными оставались атомы белка, расположенные не дальше определенного расстояния от крайних атомов лиганда. В настоящей работе были взяты расстояния (радиусы): 2 Å, 3 Å, 4 Å, 5 Å и 10 Å. В каждом случае оптимизация проводилась до того момента, пока градиент по энергии не становился меньше 1 ккал/моль/Å. Представляется закономерным тот факт, что чем большее количество атомов белка было отпущено, тем соответственно ниже оказывались теплоты образования комплекса и белка. Однако, как уже было упомянуто выше, слишком большое количество отпущенных атомов может привести к тому, что начнется перестройка участков белка, далеких от активного центра, что с точки зрения поиска энергии взаимодействия с лигандом не интересно.

Что касается времен счета, то, например, для комплекса 1f92 (16 торсионных степеней свободы лиганда, т.е. 16 связей, вокруг которых возможно вращение, изменяющее конформацию лиганда) оптимизация только лиганда в декартовых координатах занимала 16.5 часов, оптимизация лиганда и атомов белка на расстоянии 2 Å от лиганда — уже более 25 часов, оптимизация лиганда и атомов белка на расстоянии 3 Å от лиганда — около 31 часа, оптимизация лиганда и атомов белка на расстоянии 4 Å от лиганда — около 36 часов, а оптимизация лиганда и всех водородов белка — 41.5 часов. Для лигандов меньшего размера расчеты занимали меньше времени, однако время все равно возрастало при увеличении количества свободных атомов. Чтобы уменьшить время расчетов в последнем случае при оптимизации водородов белка, возможно проводить оптимизацию не по всем водородам, а только по тем, которые могут сыграть роль при взаимодействии с лигандом, например по водородам белка, находящимся на расстоянии не более $2 \div 3$ Å от крайних атомов лиганда.

Выгоду от увеличения количества свободных атомов белка мы рассматривали на основании полу-

Таблица 1

Коэффициенты корреляции между рассчитанными в МОРАС
энтальпиями связывания и экспериментальными энергиями

Радиус оптимизации в белке	Коэффициент корреляции
Только лиганд (с COSMO / без COSMO)	0.53 / 0.53
Все водороды белка (с COSMO / без COSMO)	0.59 / 0.50
2Å (с COSMO / без COSMO)	0.59 / 0.54
3Å (с COSMO / без COSMO)	0.43 / 0.20
4Å (с COSMO / без COSMO)	0.55 / 0.50
5Å (с COSMO / без COSMO)	0.59 / 0.49
10Å (с COSMO / без COSMO)	0.50 / 0.24

чившихся коэффициентов корреляции между рассчитанными энтальпиями связывания белка и лиганда и энергиями, полученными из экспериментальных данных, для различных вариантов оптимизации атомов белка. Эти коэффициенты приведены в табл. 1. Растворитель в модели COSMO учитывался на последнем шаге оптимизации или не учитывался совсем (данные для этих двух режимов разделены в таблице слешем “/”). Из таблицы видно, что оптимизация водородов или некоторого количества ближайших атомов позволяет улучшить корреляцию расчетов с экспериментом, хотя и незначительно. Здесь следует принять во внимание, что оптимизированный лиганд — нативный; если не рассматривать возможные неточности кристаллической структуры комплекса, то именно это положение его в белке в идеале является положением с минимальной энергией, которая соответствует экспериментальным данным. Дополнительная оптимизация атомов белка, скорее всего, позволяет исправить неточности рентгеноструктурного анализа. Кроме того, дополнительная оптимизация белка может найти новые минимумы энергии, однако может и увести конформацию комплекса (и как следствие — энергию взаимодействия белка и лиганда) от нативной конформации. Тем не менее при расчете энтальпий связывания для нативных лигандов в отсутствие оптимизации коэффициент корреляции с экспериментальными данными получился меньше (для неоптимизированного комплекса в присутствии растворителя он был равен 0.1), что подтверждает необходимость локальной оптимизации.

Таблица 2

Коэффициенты корреляции между рассчитанными в МОРАС
энтальпиями связывания и экспериментальными энергиями

Радиус оптимизации в белке	Коэффициент корреляции
Только лиганд (с COSMO / без COSMO)	0.72 / 0.54
Все водороды белка (с COSMO / без COSMO)	0.74 / 0.51
2Å (с COSMO / без COSMO)	0.81 / 0.55
3Å (с COSMO / без COSMO)	0.61 / 0.20
4Å (с COSMO / без COSMO)	0.73 / 0.51
5Å (с COSMO / без COSMO)	0.79 / 0.50
10Å (с COSMO / без COSMO)	0.66 / 0.26

Следует также заметить, что исключение одного из двух комплексов из выборки (1c5w или 1ejn) довольно сильно улучшает согласованность с экспериментом. В табл. 2 приведены коэффициенты корреляции рассчитанных энтальпий связывания белка и лиганда с энергиями, полученными из экспериментальных данных, без учета комплекса 1ejn. Растворитель в модели COSMO учитывался на последнем шаге опти-

мизации или не учитывался совсем (данные для этих двух режимов разделены в таблице слешем “/”). В этом случае оптимизация небольшого количества атомов белка вокруг лиганда так же остается выгодной с точки зрения улучшения согласованности с экспериментом, а также более заметным становится то, что учет растворителя приводит к ухудшению результатов.

Таким образом, из табл. 1 и 2 видно, что включение в оптимизацию атомов белка заметно улучшает корреляцию расчетов с экспериментом, причем разумно учитывать подвижность атомов белка только в ближайшей окрестности атомов лиганда — не далее 2 \AA хотя бы от одного атома лиганда.

3.3. Влияние положения водородов. В наших исследованиях мы обнаружили, что положения водородов и их начальная расстановка играют значительную роль при расчете энергии связывания. Это было показано на примере одного из комплексов (1sqo), в котором подобная расстановка была проведена двумя разными способами за счет возможности по-разному протонировать гистидины. В одном случае протонирование проводилось программой Aplite [38] и все гистидины были нейтральными, в другом — 5 гистидинов белка (гистидины с номерами 22, 106, 107, 184 и 257) получали по дополнительному водороду. Такая разница привела к тому, что после запуска программы поиска глобального минимума FLM в вакууме в силовом поле MMFF94 при всех фиксированных атомах белка, включая все его водороды, найденные глобальные минимумы оказались близки между собой и близки к нативному положению, однако после локальной оптимизации атомов лиганда в MORAC из найденного глобального минимума рассчитанные в программе MORAC значения энтальпий взаимодействия различались более чем в два раза: -149.97 ккал/моль для первого случая и -58.77 ккал/моль для второго. Кроме того, была подготовлена еще одна структура, отличающаяся от второй положением одного водорода в активном центре белка: одна из связей водород-кислород (водород гидроксильной группы серина 206) была развернута на 180 градусов, причем в непосредственной близости от водорода гидроксильной группы серина 206 располагалась гуанидиновая группа лиганда; таким образом, изменение положения водорода давало возможность для образования водородной связи (расстояние между водородом гуанидиновой группы и кислородом серина составляло 1.97 \AA). В этом случае после процедуры поиска глобального минимума программой FLM найденное положение лиганда также оказалось близко к найденным минимумам для первых двух структур. Однако при оптимизации атомов лиганда в MORAC из найденного глобального минимума для этой структуры энтальпия взаимодействия белка и лиганда составила -64.97 ккал/моль. Таким образом, изменение положения одного водорода уже заметным образом сказывалось на энергии взаимодействия. При этом добавление растворителя во всех случаях сглаживало эту разницу: энтальпии взаимодействия белок-лиганд, рассчитанные в MORAC (после локальной оптимизации найденного в FLM положения лиганда в вакууме с учетом растворителя на последнем шаге оптимизации), составили соответственно -84.24 ккал/моль (первая структура), -71.67 ккал/моль (вторая структура) и -71.32 ккал/моль (третья структура). Здесь снова становится понятной важность правильного учета растворителя, как и то, что при квантово-химических расчетах (как, в общем, и при расчетах в силовом поле) различные положения водородов могут заметно повлиять на энергию взаимодействия и требуется их оптимизация в рамках той системы, в которой будет проводиться вычисление энергии.

3.4. Соединение методов прямого докинга и квантовой химии. Как и в случае с сеточным докингом, мы провели вычисление энтальпии в программном комплексе MORAC, отталкиваясь от предполагаемого глобального минимума, найденного с помощью программы FLM. Поскольку программа FLM проводит поиск глобального минимума и оптимизацию комплекса в вакууме, мы рассмотрели два варианта вычисления энтальпии в MORAC — без учета растворителя и с учетом растворителя на последнем шаге оптимизации. В первом случае коэффициент корреляции рассчитанных в MORAC энтальпий со значениями энергий, полученными из экспериментальных данных, составил 0.53 , а во втором — 0.25 . Подобная разница, вероятно, является следствием отсутствия учета растворителя при поиске глобального минимума в программе FLM. Чтобы проверить это предположение и провести более адекватное сравнение с экспериментом расчетов в силовом поле MMFF94, методом PM7 в вакууме и методом PM7 в вакууме с учетом растворителя в модели COSMO мы провели следующие расчеты.

Во-первых, мы расширили тестовый набор комплексов, включив в него не только комплексы урокиназы, но и других белков — всего 22 комплекса с нативными лигандами, для которых известны константы ингибирования. Это были 4 комплекса СНК1 (4FT0, 4FT9, 4FSW, 4FTA), 2 комплекса ERK2 (4FV5, 4FV6), 3 комплекса PIN1 (3IKD, 3IKG, 3JYJ), 3 комплекса рибонуклеазы А (3D6O, 3D6P, 3D8Z), 2 комплекса тромбина (1DWC, 1TOM), 6 комплексов урокиназы (1C5Y, 1F5L, 1O3P, 1SQO, 1VJ9, 1VJA) и 2 комплекса фактора Ха (2P94, 3CEN). Для этих комплексов, а также для соответствующих свободных лигандов в силовом поле MMFF94 были найдены низкоэнергетические локальные минимумы (в количестве $\sim 10^2$ – 10^3 для каждого комплекса и свободного лиганда) при помощи программы FLM. К этому набору было

добавлено также и оптимизированное нативное положение. На основании глобальных минимумов потенциальной энергии MMFF94 для комплекса и свободного лиганда были рассчитаны энергии связывания лиганда с белком:

$$\Delta E_{\min} = E_{\min}^{\text{complex}} - E_{\min}^{\text{free ligand}} - E^{\text{free protein}},$$

где ΔE_{\min} — энергия связывания лиганда с белком, $E_{\min}^{\text{complex}}$ — глобальный минимум энергии комплекса, $E_{\min}^{\text{free ligand}}$ — глобальный минимум энергии свободного лиганда, $E^{\text{free protein}}$ — энергия свободного белка, который полагался жестким.

Во-вторых, не имея пока возможности реализовать непосредственно процедуру FLM поиска низкоэнергетических локальных минимумов с помощью квантово-химического метода PM7, мы провели локальную оптимизацию лиганда как в комплексе, так и в свободном состоянии, выбрав в качестве начальных все локальные минимумы, найденные с помощью FLM при использовании силового поля MMFF94, и добавив сюда оптимизированное в PM7 исходное нативное положение. При этом были определены глобальные минимумы (на самом деле это были условно глобальные минимумы, поскольку пространство поиска было существенно ограничено) энтальпии образования PM7 и получены соответствующие энергии связывания лиганда с белком в рамках метода PM7. Далее, энергии эти локальных минимумов были пересчитаны в MORAC с PM7 и COSMO без дополнительной оптимизации и были получены соответствующие энергии связывания для PM7 с учетом растворителя.

Эти значения ΔE_{\min} энергии связывания лиганда с белком для трех вариантов функции энергии сравнивались с $R \times T \times \ln(K_i)$, т.е. с логарифмом констант ингибирования белков соответствующими лигандами. Коэффициенты корреляции рассчитанной энергии связывания с логарифмом экспериментально измеренной константы ингибирования соответственно равны: 0.54 — при использовании MMFF94; 0.66 — при использовании PM7; 0.59 — при использовании PM7 с COSMO.

Зависимость рассчитанной энергии связывания (ΔE_{\min}) от логарифма экспериментально измеренной константы ингибирования ($R \times T \times \ln(K_i)$) для всех трех вариантов функции энергии связывания (MMFF94, PM7, PM7 COSMO) показана на рис. 2. При построении графика использовались 22 комплекса белок–лиганд с низкоэнергетическими локальными минимумами комплекса и свободного лиганда, найденными программой FLM. Белок полагался жестким. Энергия связывания рассчитывалась на основе глобальных минимумов соответствующей функции энергии для комплекса и свободного лиганда.

Как видим, для данного набора комплексов с экспериментом лучше всего коррелируют энергии связывания, рассчитанные при помощи PM7 в вакууме, а хуже всего — при помощи MMFF94, хотя разница между коэффициентами корреляции сравнительно невелика. Хотя корреляция с экспериментом в случае PM7 с COSMO хуже, чем в случае PM7 в вакууме, оптимизированное нативное положение лиганда оказывается ниже по энергии, чем у всех найденных минимумов для 6 комплексов из 22, в случае PM7 без COSMO — для одного комплекса из 22, а для силового поля ни для одного комплекса оптимизированное нативное положение не является глобальным минимумом. Напомним, однако, что поиск минимумов энергий при помощи PM7 и PM7 с COSMO был ограничен изначально найденным в поле MMFF94 набором низкоэнергетических минимумов. Интересно отметить, что по порядку величины диапазон изменения энергий связывания PM7 с COSMO (около 60 ккал/моль) заметно меньше диапазона изменений энергий связывания MMFF94 (около 400 ккал/моль) или энергий связывания PM7 (около 300 ккал/моль), что ближе к экспериментальным данным. Это опять указывает на важность учета растворителя, который существенно экранирует энергию кулоновских взаимодействий в рассматриваемых молекулярных системах. Мы показали также, что энергии найденных локальных минимумов комплексов и свободных лигандов существенно перераспределяются при использовании квантово-химического метода PM7 по сравнению с энергиями в силовом поле MMFF94. Таким образом, для улучшения совпадения рассчитанных энергий связывания с экспериментом докинг надо проводить сразу в рамках полуэмпирического метода PM7.

3.5. Поиск новых ингибиторов урокиназы методами докинга и квантовой химии. Отра-

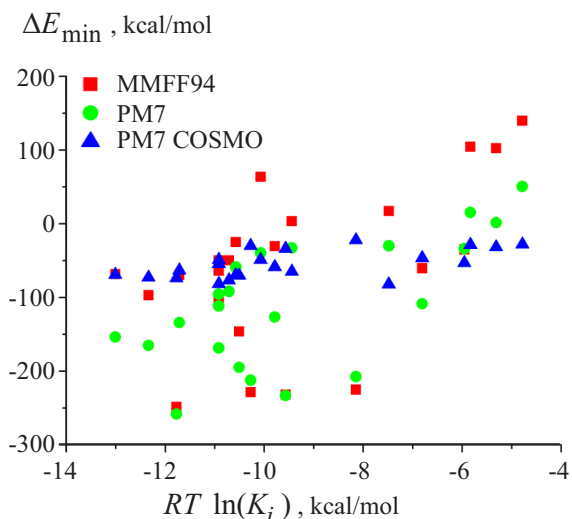


Рис. 2. Зависимость рассчитанной энергии связывания от ее экспериментального значения

ботав методику использования квантово-химических методов на нативных урокиназных комплексах, мы применили МОРАС к задаче поиска новых ингибиторов урокиназы. Для этого мы проанализировали значения энтальпий связывания десяти нативных лигандов, упомянутых выше, а также значения энтальпий связывания для некоторых ненативных лигандов, полученных в результате поиска новых ингибиторов урокиназы и оказавшихся неактивными (при расчете энтальпии оптимизация в МОРАС проводилась только по атомам лиганда, а растворитель учитывался на последнем шаге оптимизации). В качестве порогового значения энтальпии связывания между плохими и хорошими ингибиторами мы выбрали значение -50 ккал/моль. Таким образом, если энтальпия связывания белка с лигандом, рассчитанная в программе МОРАС, меньше -50 ккал/моль, то мы полагаем, что лиганд будет связываться с белком, и соответственно наоборот. Следует отметить, что эти значения энергий оказываются почти на порядок выше, нежели свободные энергии Гиббса, рассчитанные из констант ингибирования и равные примерно $-10 \div -5$ ккал/моль. Это обусловлено, во-первых, учетом только энтальпии и неучетом энтропийной составляющей свободной энергии Гиббса, которая повысит энергию связывания на $20-30$ ккал/моль. Во-вторых, определенную роль может сыграть как использование более точной модели растворителя, так и включение в расчеты и оптимизацию отдельных молекул воды, что может привести к появлению дополнительных водородных связей. И наконец, несмотря на довольно аккуратный учет межмолекулярных взаимодействий для РМ7, метод параметризации также может оказать влияние на величину энергии связывания. В качестве начального положения ненативного лиганда — кандидата в ингибиторы (в отличие от нативных лигандов оно неизвестно) в комплексе бралось положение лиганда, найденное программой докинга SOL [38–40], которая хорошо зарекомендовала себя в предыдущих исследованиях [2, 38]. Программа SOL находит глобальный минимум энергии взаимодействия системы белок–лиганд с помощью генетического алгоритма в рамках силового поля MMFF94 [19] для жесткого белка и подвижного лиганда; при этом белок моделируется заранее рассчитанной сеткой потенциалов взаимодействия пробных атомов лиганда с белком. Далее, из найденного SOL глобального минимума в силовом поле проводилась локальная оптимизация лиганда в МОРАС.

Таблица 3

Синтезированные и проверенные в эксперименте лиганды, их скоринг-функции в SOL, энтальпии в МОРАС и экспериментальные активности

Номер лиганда	Скоринг-функция SOL, ккал/моль	Энтальпия МОРАС, ккал/моль	Результаты эксперимента, мкМ
U002R	-4.73	-42.55	n/b
U001R	-4.62	-48.77	n/b
U007R	-4.86	-43.50	n/b
U003P	-4.59	-49.50	n/b
U005r	-5.04	-59.63	500
U006R	-4.73	-46.95	n/b
U009T	-7.14	-43.82	n/b
U012T	-6.91	-55.57	n/b
U013T	-7.62	-57.21	n/b
U014R	-4.15	-55.08	200
U014aR	-4.57	-47.95	n/b
U015R	-4.13	-67.14	n/b
U017R	-4.15	-58.98	200
U019R	-4.21	-51.08	500
U020R	-4.02	-49.56	500
U021R	-4.74	-60.75	500
U024R	-5.1	-60.04	33
U026R	-5.56	-61.31	10

Было рассчитано 18 соединений, предложенных к синтезу, и получены как скоринг-функции программы SOL, так и энтальпии в программе MOPAC (табл. 3). Все эти соединения были проверены в эксперименте и получены их экспериментальные активности (в табл. 3: n/b — non-binder — неактивный лиганд); лидерное соединение ((5-бромбензизотиазол-3-ил)гуанидин) показало активность в 10 мкМ. Из этой таблицы видно, что все соединения, показавшие в эксперименте хотя бы небольшую ингибирующую активность (кроме U020R — ложноотрицательный результат), действительно имеют значение энтальпии меньше -50 ккал/моль. Видно также, что есть два ложноположительных результата. Для программы докинга SOL (если выбрать пороговое значение скоринг-функции, разделяющее ингибиторы и неактивные соединения, равным -4.5 ккал/моль [38], что соответствует активности в 500 мкМ) количество ложноотрицательных и ложноположительных результатов оказалось значительно больше: 4 и 9 соответственно. Кроме того, мы рассчитали коэффициенты корреляции с экспериментальными данными для скоринг-функции SOL (-0.2) и для энтальпии MOPAC (0.52). В данном случае квантово-химические вычисления показали лучшие результаты, тогда как программа докинга на основе силового поля MMFF94 не смогла выделить активные и неактивные лиганды.

4. Заключение. В настоящей статье мы описали одно из первых применений полуэмпирического квантово-химического метода PM7 для оценки энергий связывания лигандов и белков на примере работы с комплексами урокиназа–лиганд и поиска новых ингибиторов урокиназы. Расчет подобных больших систем с помощью квантовой химии стал возможен не только благодаря возрастающим компьютерным мощностям, но и новым алгоритмам и методам. Повышение точности расчетов для данного полуэмпирического метода достигалось за счет добавления поправок (на водородные и галогенные связи и на дисперсионные взаимодействия), учтенных в методе PM7. В результате показано, что использование квантово-химических полуэмпирических методов позволяет улучшить корреляцию рассчитанных энергий связывания белок–лиганд с энергиями связывания, полученными из эксперимента по сравнению с методом докинга с использованием силового поля, и, таким образом, делает перспективным использование указанных методов в разработке новых лекарственных средств.

Одним из основных выводов работы является демонстрация того факта, что положения водородов белка, которые обычно отсутствуют в PDB-структурах и добавляются при помощи различных программ молекулярного моделирования, значительным образом влияют на энергию связывания белок–лиганд. Изменение положения одного водорода может приводить к образованию или исчезновению водородной связи, энергетический вклад которой составляет от 6 до 10 ккал/моль, что и было показано на примере комплекса 1SQO. По этой же причине важным аспектом при оптимизации комплекса белок–лиганд является подвижность атомов белка. Показано также, что включение атомов водородов белка в оптимизацию комплекса в пакете программ MOPAC улучшает корреляцию рассчитанных энергий с экспериментом.

Учет подвижности белка при оптимизации нативных структур приводил к общему понижению энергии комплекса и белка и до определенного радиуса (расстояния от атомов лиганда, на котором атомы белка включались в оптимизацию) улучшал корреляцию рассчитанных энтальпий с экспериментальными данными, хотя время расчетов при этом соответственно увеличивалось. Не нужно также исключать и того, что проведение подобной оптимизации, возможно, позволяет обнаружить новые локальные минимумы, но при этом уводит геометрию комплекса белок–лиганд от нативной конформации, соответствующей определенной константе связывания. Тем не менее учет подвижности как тяжелых атомов белка, так и водородов дает значительные преимущества при вычислении энергии связывания; эти факторы следует учитывать, например, при проведении прямого докинга, в котором, в отличие от сеточного, возможно реализовать свободное движение атомов белка. Немаловажную роль также играет учет растворителя, который реализован в программном комплексе MOPAC методом COSMO. При рассмотрении 9 комплексов из 10 учет растворителя приводил к значительному улучшению корреляции с экспериментом. Очевидно, что более аккуратный его подсчет может значительным образом повлиять на значения энергий, однако расчет энтальпии связывания белок–лиганд в присутствии растворителя в рамках MOPAC представляется пока проблематичным из-за слишком медленной и не очень точной оптимизации геометрии с учетом растворителя, реализованной в MOPAC. Поиск глобального минимума в рамках силового поля MMFF94 в присутствии растворителя в рамках программы прямого докинга FLM также пока не реализован из-за слишком медленных расчетов полярной части энергии взаимодействия молекул с растворителем в рамках существующей программы DISOLV [41]. В обоих случаях решение проблемы учета растворителя возможно с помощью недавно разработанного мультитарядового метода работы с большими плотными матрицами [42]. Учет растворителя сказывается не только на коэффициентах корреляции с экспериментом, но и на значениях энергии связывания, поскольку он существенно экранирует энергию кулоновских взаимодействий, а значит, при проведении расчетов следует также внимательно следить за зарядовыми состояниями молекул исследуемой системы.

Стремясь выделить основные факторы, влияющие на улучшение точности расчетов энергии связывания белок–лиганд, нужно также уточнить, что для более аккуратного расчета свободной энергии Гиббса связывания белок–лиганд, которая напрямую связана с измеряемой в экспериментах константой ингибирования, следует учитывать также и энтропийную составляющую, которая по величине может быть близка к энтальпии связывания, но противоположна ей по знаку. Следует также отметить, что полуэмпирические квантово-химические методы все же имеют некоторые ограничения как по времени счета, обусловленного довольно большими размерами исследуемых систем (комплексы белок–лиганд), так и по точности вычислений, поскольку более аккуратный расчет энергии будет требовать учета многих локальных минимумов для корректного расчета энтропии, а также более аккуратного учета растворителя и, возможно, отдельных молекул воды.

Однако несмотря на все вышеперечисленные ограничения, присутствующие на данный момент в квантово-химических расчетах с помощью программного комплекса МОРАС, их применение для поиска новых ингибиторов урокиназы в качестве постпроцессинга (т.е. после нахождения программой докинга глобального минимума) представляется актуальным. В этом случае вначале проводится поиск глобального минимума энергии связывания с помощью прямого или сеточного докинга, а затем — локальная оптимизация и оценка энтальпии связывания полуэмпирическим квантово-химическим методом РМ7. Используя данную процедуру, мы смогли выделить 18 новых лигандов — кандидатов в ингибиторы урокиназы, которые были синтезированы и проверены в эксперименте. Ингибирующей активностью в отношении урокиназы обладали 8 лигандов, из которых 2 соединения были отмечены как лидеры и стали отправной точкой для последующей разработки. В отличие от постпроцессинга с помощью силового поля MMFF94, квантово-химические расчеты показали лучшие результаты, чем при докинге; для синтезированных лигандов корреляция энергий, вычисленных в МОРАС и рассчитанных исходя из экспериментальных данных, составила 0.52. Кроме того, для 22 комплексов белок–лиганд, для которых энергии взаимодействия были рассчитаны как в силовом поле MMFF94, так и с помощью полуэмпирического метода РМ7, коэффициенты корреляции с экспериментом оказались выше в случае квантово-химических расчетов. Это дает основания полагать, что включение квантовой химии в процесс при расчетах свободной энергии связывания может улучшить точность вычислений без использования дополнительных подгоночных коэффициентов, которые присутствуют в формулах для скоринг-функций многих программ (в том числе программы SOL), а значит, позволит не только приблизительно разделять лиганды на “хорошие” и “плохие”, но также отделять “хорошие” ингибиторы от “удовлетворительных” и “очень хороших”. Однако поскольку времена, требуемые на расчет одного комплекса, варьируются от одного до десяти часов (а то и более — если брать во внимание оптимизацию белка) в зависимости от размера лиганда, расчет большого количества молекул (например, скрининг баз данных, содержащих десятки и сотни тысяч соединений) с помощью квантовой химии представляется пока нереализуемым, если только не использовать соответствующее количество вычислительных ядер на суперкомпьютере, распараллелив вычисления по каждому лиганду: одно вычислительное ядро — один лиганд.

Работа выполнена при финансовой поддержке РФФИ (код проекта 14–07–31126), а также в рамках проекта МГУ имени М. В. Ломоносова “Математическое и молекулярное моделирование для медицины и фармакологии”.

СПИСОК ЛИТЕРАТУРЫ

1. Садовничий В.А., Сулимов В.Б. Суперкомпьютерные технологии в медицине // Суперкомпьютерные технологии в науке, образовании и промышленности / Под ред. В.А. Садовничего, Г.И. Савина, Вл.В. Воеводина. М: Изд-во Моск. ун-та, 2009. 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., Ataulakhanov F.I. New synthetic thrombin inhibitors: molecular design and experimental verification // PLoS ONE. 2011. 6, N 5. Article No. e19969.
3. Chen W., Gilson M.K., Webb S.P., Potter M.J. Modeling protein–ligand binding by mining minima // J. Chem. Theory Comput. 2010. 6, N 11. 3540–3557.
4. Wang R., Lu Y., Wang S. Comparative evaluation of 11 scoring functions for molecular docking // J. Med. Chem. 2003. 46, N 12. 2287–2303.
5. Grzybowski B.A., Ishchenko A.V., Shimada J., Shakhnovich E.I. From knowledge-based potentials to combinatorial lead design in silico // Acc. Chem. Res. 2002. 35, N 5. 261–269.
6. Halperin I., Ma B., Wolfson H., Nussinov R. Principles of docking: an overview of search algorithms and a guide to scoring functions // Proteins. 2002. 47, N 4. 409–443.
7. Taylor R.D., Jewsbury P.J., Essex J.W. A review of protein–small molecule docking methods // J. Comput.-Aided Mol. Des. 2002. 16, N 3. 151–166.

8. Gohlke H., Klebe G. Statistical potentials and scoring functions applied to protein–ligand binding // *Curr. Opin. Struct. Biol.* 2001. **11**, N 2. 231–235.
9. Pérez C., Ortiz A.R. Evaluation of docking functions for protein–ligand docking // *J. Med. Chem.* 2001. **44**, N 23. 3768–3785.
10. Muegge I., Rarey M. Small molecule docking and scoring // *Reviews in Computational Chemistry*. Vol. 17. New York: Wiley, 2001. 1–60.
11. Ehrlich L.P., Wade R.C. Protein–protein docking // *Reviews in Computational Chemistry*. Vol. 17. New York: Wiley, 2001. 61–97.
12. Jorgensen W.L. Free energy calculations: a breakthrough for modeling organic chemistry in solution // *Acc. Chem. Res.* 1989. **22**, N 5. 184–189.
13. Kollman P. Free energy calculations: applications to chemical and biochemical phenomena // *Chem. Rev.* 1993. **93**, N 7. 2395–2417.
14. Aqvist J., Medina C., Samuelsson J.-E. A new method for predicting binding affinity in computer-aided drug design // *Protein Eng.* 1994. **7**, N 3. 385–391.
15. Srinivasan J., Cheatham T.E., Cieplak P., Kollman P.A., Case D.A. Continuum solvent studies of the stability of DNA, RNA and phosphoramidate–DNA helices // *J. Am. Chem. Soc.* 1998. **120**, N 37. 9401–9409.
16. Kollman P.A., Massova I., Reyes C., Kuhn B., Huo S., Chong L., Lee M., Lee T., Duan Y., Wang W., Donini O., Cieplak P., Srinivasan J., Case D.A., Cheatham T.E. Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models // *Acc. Chem. Res.* 2000. **33**, N 12. 889–897.
17. Stewart J.J.P. MOPAC: a semiempirical molecular orbital program // *J. Comput.-Aided Mol. Des.* 1990. **4**, N 1. 1–103.
18. Stewart J.J.P. MOPAC2012 (<http://OpenMOPAC.net>).
19. Halgren T.A. Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94 // *J. Comput. Chem.* 1996. **17**, N 5/6. 490–519.
20. Stewart J.J.P. Application of localized molecular orbitals to the solution of semiempirical self-consistent field equations // *Int. J. Quant. Chem.* 1996. **58**, N 2. 133–146.
21. Dewar M.J.S., Zoebisch E.G., Healy E.F., Stewart J.J.P. Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical model // *J. Am. Chem. Soc.* 1985. **107**, N 13. 3902–3909.
22. Stewart J.J.P. Optimization of parameters for semiempirical methods I. Method // *J. Comp. Chem.* 1989. **10**, N 2. 209–220.
23. Stewart J.J.P. Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements // *J. Mol. Modeling.* 2007. **13**, N 12. 1173–1213.
24. Rocha G.B., Freire R.O., Simas A.M., Stewart J.J.P. RM1: a reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I // *J. Comp. Chem.* 2006. **27**, N 10. 1101–1111.
25. Stewart J.J.P. Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters // *J. Mol. Modeling.* 2013. **19**, N 1. 1–32.
26. Grimme S. Accurate description of van der Waals complexes by density functional theory including empirical corrections // *J. Comput. Chem.* 2004. **25**, N 12. 1463–1473.
27. Jurečka P., Černý J., Hobza P., Salahub D.R. Density functional theory augmented with an empirical dispersion term: interaction energies and geometries of 80 noncovalent complexes compared with *ab initio* quantum mechanics calculations // *J. Comput. Chem.* 2007. **28**, N 2. 555–569.
28. Řezáč J., Hobza P. A halogen-bonding correction for the semiempirical PM6 method // *Chem. Phys. Lett.* 2011. **506**, N 4–6. 286–289.
29. Řezáč J., Fanfrlík J., Salahub D., Hobza P. Semiempirical quantum chemical PM6 method augmented by dispersion and H-bonding correction terms reliably describes various types of noncovalent complexes // *J. Chem. Theory Comput.* 2009. **5**, N 7. 1749–1760.
30. Korth M., Pitoňák M., Řezáč J., Hobza P. A transferable H-bonding correction for semiempirical quantum-chemical methods // *J. Chem. Theory Comput.* 2010. **6**, N 1. 344–352.
31. Korth M. Third-generation hydrogen-bonding corrections for semiempirical QM methods and force fields // *J. Chem. Theory Comput.* 2010. **6**, N 12. 3808–3816.
32. Oferkin I.V., Katkova E.V., Sulimov V.B. Direct generalized docking program FLM: protein–ligand binding free energy calculations by multiwell approximation // *Journal of Chemical Information and Modeling*. 2014 (submitted).
33. Duffy M.J. The urokinase plasminogen activator system: role in malignancy // *Curr. Pharm. Des.* 2004. **10**, N 1. 39–49.
34. Choong P.F.M., Nadesapillai A.P.W. Urokinase plasminogen activator system: a multifunctional role in tumor progression and metastasis // *Clin. Orthop. Relat. Res.* 2003. **415**. S46–S58.
35. Blasi F., Carmeliet P. uPAR: a versatile signalling orchestrator // *Nat. Rev. Mol. Cell Biol.* 2002. **3**, N 12. 932–943.
36. Ulisse S., Baldini E., Sorrenti S., D’Armiento M. The urokinase plasminogen activator system: a target for anti-cancer therapy // *Current Cancer Drug Targets*. 2009. **9**, N 1. 32–71.
37. The Protein Data Bank (<http://www.rcsb.org/>).

38. *Sulimov A.V., Kutov D.C., Oferkin I.V., Katkova E.V., Sulimov V.B.* Application of the docking program SOL for CSAR Benchmark // *J. Chem. Inf. Model.* 2013. **53**, N 8. 1946–1956.
39. *Романов А.Н., Кондакова О.А., Григорьев Ф.В., Сулимов А.В., Луцкекина С.В., Мартынов Я.Б., Сулимов В.Б.* Компьютерный дизайн лекарственных средств: программа докинга SOL // *Вычислительные методы и программирование.* 2008. **9**. 213–233.
40. *Офёркин И.В., Сулимов А.В., Кондакова О.А., Сулимов В.Б.* Реализация поддержки параллельных вычислений в программах докинга SOLGRID и SOL // *Вычислительные методы и программирование.* 2011. **12**. 9–23.
41. *Купервассер О.Ю., Жабин С.Н., Мартынов Я.Б., Федулов К.М., Офёркин И.В., Сулимов А.В., Сулимов В.Б.* Континуальная модель растворителя: программа DISOLV — алгоритмы, реализация и валидация // *Вычислительные методы и программирование.* 2011. **12**. 247–261.
42. *Михалев А.Ю., Офёркин И.В., Оселедец И.В., Сулимов А.В., Тьртмышников Е.Е., Сулимов В.Б.* Применение мультязарядового приближения больших плотных матриц в рамках модели поляризуемого континуума для растворителя // *Вычислительные методы и программирование.* 2014. **15**. 9–21.

Поступила в редакцию
25.03.2014

Application of the PM7 Quantum Chemical Semi-Empirical Method to the Development of New Urokinase Inhibitors

E. V. Katkova¹, I. V. Oferkin², and V. B. Sulimov³

¹ *Dimonta, Ltd; ulitsa Nagornaya 15, stroenie 8, Moscow, 117186, Russia; Junior Scientist, e-mail: katkova@dimonta.com*

² *Research Computing Center, Lomonosov Moscow State University; Leninskie Gory, Moscow, 119992, Russia; Leading Programmer, e-mail: io@dimonta.com*

³ *Research Computing Center, Lomonosov Moscow State University; Leninskie Gory, Moscow, 119992, Russia; Ph.D., Head of Laboratory, e-mail: vladimir.sulimov@gmail.com*

Received March 25, 2014

Abstract: The application of quantum-chemistry semi-empirical methods using the MOPAC programs to the search for new inhibitors is considered. The best operating conditions for the PM7 parameterization method in the framework of MOPAC are studied. This method allows taking into account various noncovalent intermolecular interactions, such as dispersion interactions and halogen and hydrogen bonds. It is shown that different methods of protein protonation have a great influence on the protein–ligand binding enthalpies. As a consequence, the hydrogen mobility of the protein plays an important role during the local optimization of the complex. Using the set of 22 different protein–ligand complexes, the importance of the solvent is demonstrated; it is also shown that the correlation coefficients between computed and experimental energies are better for quantum chemical calculations than for the MMFF94 force field calculations for this set. The protein–inhibitor binding enthalpies for native urokinase complexes from the PDB database are calculated, and the correlation coefficients between MOPAC calculations and experiments are found. Some results of MOPAC usage for new urokinase inhibitors design are discussed.

Keywords: molecular modeling, quantum chemistry, semi-empirical methods, new inhibitors design, efficacy, optimal operating parameters, grid docking, direct docking.

References

1. V. A. Sadovnichii and V. B. Sulimov, “Supercomputing Technologies in Medicine,” in *Supercomputing Technologies in Science, Education, and Industry*, Ed. by V. A. Sadovnichii, G. I. Savin, and Vl. V. Voevodin (Mosk. Gos. Univ., Moscow, 2009), pp. 16–23.
2. E. I. Sinauridze, A. N. Romanov, I. V. Gribkova, et al., “New Synthetic Thrombin Inhibitors: Molecular Design and Experimental Verification,” *PLoS ONE* **6** (5), Article No. e19969 (2011).
3. W. Chen, M. K. Gilson, S. P. Webb, and M. J. Potter, “Modeling Protein–Ligand Binding by Mining Minima,” *J. Chem. Theory Comput.* **6** (11), 3540–3557 (2010).

4. R. Wang, Y. Lu, and S. Wang, "Comparative Evaluation of 11 Scoring Functions for Molecular Docking," *J. Med. Chem.* **46** (12), 2287–2303 (2003).
5. B. A. Grzybowski, A. V. Ishchenko, J. Shimada, and E. I. Shakhnovich, "From Knowledge-Based Potentials to Combinatorial Lead Design in Silico," *Acc. Chem. Res.* **35** (5), 261–269 (2002).
6. I. Halperin, B. Ma, H. Wolfson, and R. Nussinov, "Principles of Docking: An Overview of Search Algorithms and a Guide to Scoring Functions," *Proteins* **47** (4), 409–443 (2002).
7. R. D. Taylor, P. J. Jewsbury, and J. W. Essex, "A Review of Protein-Small Molecule Docking Methods," *J. Comput.-Aided Mol. Des.* **16** (3), 151–166 (2002).
8. H. Gohlke and G. Klebe, "Statistical Potentials and Scoring Functions Applied to Protein-Ligand Binding," *Curr. Opin. Struct. Biol.* **11** (2), 231–235 (2001).
9. C. Pérez and A. R. Ortiz, "Evaluation of Docking Functions for Protein-Ligand Docking," *J. Med. Chem.* **44** (23), 3768–3785 (2001).
10. I. Muegge and M. Rarey, "Small Molecule Docking and Scoring," in *Reviews in Computational Chemistry* (Wiley, New York, 2001), Vol. 17, pp. 1–60.
11. L. P. Ehrlich and R. C. Wade, "Protein-Protein Docking," in *Reviews in Computational Chemistry* (Wiley, New York, 2001), Vol. 17, pp. 61–97.
12. W. L. Jorgensen, "Free Energy Calculations: A Breakthrough for Modeling Organic Chemistry in Solution," *Acc. Chem. Res.* **22** (5), 184–189 (1989).
13. P. Kollman, "Free Energy Calculations: Applications to Chemical and Biochemical Phenomena," *Chem. Rev.* **93** (7), 2395–2417 (1993).
14. J. Aqvist, C. Medina, and J.-E. Samuelsson, "A New Method for Predicting Binding Affinity in Computer-Aided Drug Design," *Protein Eng.* **7** (3), 385–391 (1994).
15. J. Srinivasan, T. E. Cheatham, P. Cieplak, et al., "Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate-DNA Helices," *J. Am. Chem. Soc.* **120** (37), 9401–9409 (1998).
16. P. A. Kollman, I. Massova, C. Reyes, et al., "Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models," *Acc. Chem. Res.* **33** (12), 889–897 (2000).
17. J. J. P. Stewart, "MOPAC: A Semiempirical Molecular Orbital Program," *J. Comput.-Aided Mol. Des.* **4** (1), 1–103 (1990).
18. J. J. P. Stewart, "MOPAC2012," <http://OpenMOPAC.net>.
19. T. A. Halgren, "Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94," *J. Comput. Chem.* **17** (5/6), 490–519 (1996).
20. J. J. P. Stewart, "Application of Localized Molecular Orbitals to the Solution of Semiempirical Self-Consistent Field Equations," *Int. J. Quant. Chem.* **58** (2), 133–146 (1996).
21. M. J. S. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. P. Stewart, "Development and Use of Quantum Mechanical Molecular Models. 76. AM1: A New General Purpose Quantum Mechanical Model," *J. Am. Chem. Soc.* **107** (13), 3902–3909 (1985).
22. J. J. P. Stewart, "Optimization of Parameters for Semiempirical Methods I. Method," *J. Comp. Chem.* **10** (2), 209–220 (1989).
23. J. J. P. Stewart, "Optimization of Parameters for Semiempirical Methods V: Modification of NDDO Approximations and Application to 70 Elements," *J. Mol. Model.* **13** (12), 1173–1213 (2007).
24. G. B. Rocha, R. O. Freire, A. M. Simas, and J. J. P. Stewart, "RM1: A Reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I," *J. Comp. Chem.* **27** (10), 1101–1111 (2006).
25. J. J. P. Stewart, "Optimization of Parameters for Semiempirical Methods VI: More Modifications to the NDDO Approximations and Re-Optimization of Parameters," *J. Mol. Model.* **19** (1), 1–32 (2013).
26. S. Grimme, "Accurate Description of van der Waals Complexes by Density Functional Theory Including Empirical Corrections," *J. Comput. Chem.* **25** (12), 1463–1473 (2004).
27. P. Jurečka, J. Černý, P. Hobza, and D. R. Salahub, "Density Functional Theory Augmented with an Empirical Dispersion Term: Interaction Energies and Geometries of 80 Noncovalent Complexes Compared with *ab initio* Quantum Mechanics Calculations," *J. Comput. Chem.* **28** (2), 555–569 (2007).
28. J. Řezáč and P. Hobza, "A Halogen-Bonding Correction for the Semiempirical PM6 Method," *Chem. Phys. Lett.* **506** (4–6), 286–289 (2011).
29. J. Řezáč, J. Fanfrlík, D. Salahub, and P. Hobza, "Semiempirical Quantum Chemical PM6 Method Augmented by Dispersion and H-Bonding Correction Terms Reliably Describes Various Types of Noncovalent Complexes," *J. Chem. Theory Comput.* **5** (7), 1749–1760 (2009).
30. M. Korth, M. Pitoňák, J. Řezáč, and P. Hobza, "A Transferable H-Bonding Correction for Semiempirical Quantum-Chemical Methods," *J. Chem. Theory Comput.* **6** (1), 344–352 (2010).

31. M. Korth, "Third-Generation Hydrogen-Bonding Corrections for Semiempirical QM Methods and Force Fields," *J. Chem. Theory Comput.* **6** (12), 3808–3816 (2010).
32. I. V. Oferkin, E. V. Katkova, and V. B. Sulimov, "Direct Generalized Docking Program FLM: Protein–Ligand Binding Free Energy Calculations by Multiwell Approximation," *J. Chem. Inf. Model.* (submitted).
33. M. J. Duffy, "The Urokinase Plasminogen Activator System: Role in Malignancy," *Curr. Pharm. Des.* **10** (1), 39–49 (2004).
34. P. F. M. Choong and A. P. W. Nadesapillai, "Urokinase Plasminogen Activator System: A Multifunctional Role in Tumor Progression and Metastasis," *Clin. Orthop. Relat. Res.* **415**, S46–S58 (2003).
35. F. Blasi and P. Carmeliet, "uPAR: A Versatile Signalling Orchestrator," *Nat. Rev. Mol. Cell Biol.* **3** (12), 932–943 (2002).
36. S. Ulisse, E. Baldini, S. Sorrenti, and M. D'Armiento, "The Urokinase Plasminogen Activator System: A Target for Anti-Cancer Therapy," *Curr. Cancer Drug Targets* **9** (1), 32–71 (2009).
37. The Protein Data Bank. <http://www.rcsb.org/>.
38. A. V. Sulimov, D. C. Kutov, I. V. Oferkin, et al., "Application of the Docking Program SOL for CSAR Benchmark," *J. Chem. Inf. Model.* **53** (8), 1946–1956 (2013).
39. A. N. Romanov, O. A. Kondakova, F. V. Grigor'ev, et al., "The SOL Docking Package for Computer-Aided Drug Design," *Vychisl. Metody Programm.* **9**, 213–233 (2008).
40. I. V. Oferkin, A. V. Sulimov, O. A. Kondakova, and V. B. Sulimov, "Implementation of Parallel Computing in SOLGRID and SOL Docking Programs," *Vychisl. Metody Programm.* **12**, 9–23 (2011).
41. O. Yu. Kupervasser, S. N. Zhabin, Ya. B. Martynov, et al., "A Continuum Solvent Model: The DISOLV Program (Algorithms, Implementation, and Validation)," *Vychisl. Metody Programm.* **12**, 247–261 (2011).
42. A. Yu. Mikhalev, I. V. Oferkin, I. V. Oseledets, et al., "Application of the Multicharge Approximation for Large Dense Matrices in the Framework of the Polarized Continuum Solvent Model," *Vychisl. Metody Programm.* **15**, 9–21 (2014).