

УДК 519.6

ПРИМЕНЕНИЕ МУЛЬТИЗАРЯДОВОГО ПРИБЛИЖЕНИЯ БОЛЬШИХ ПЛОТНЫХ МАТРИЦ В РАМКАХ МОДЕЛИ ПОЛЯРИЗУЕМОГО КОНТИНУУМА ДЛЯ РАСТВОРИТЕЛЯ

А. Ю. Михалев¹, И. В. Офёркин², И. В. Оселедец³, А. В. Сулимов⁴,
Е. Е. Тыртышников⁵, В. Б. Сулимов⁶

Рассмотрены расчеты полярной части энергии взаимодействия молекул с водой в рамках континуальной модели растворителя РСМ (Polarized Continuum Model). Предложен новый алгоритм решения уравнения для поляризационных зарядов, позволяющий получить ускорение в сотни раз по сравнению с обычными итерационными методами без существенной потери точности. Этот новый алгоритм основан на использовании мультизарядового приближения больших плотных матриц, возникающих при триангуляции поверхности, разделяющей молекулу и растворитель, и дискретизации непрерывной плотности поляризационных зарядов, наводимых на поверхности диэлектрика-растворителя атомными зарядами молекулы. Новый алгоритм реализован в программе MCBHSOLV, написанной на языке Python, а сравнение результатов расчетов и их быстроедействие проводилось с программой DISOLV. Сравнение проведено на тестовом наборе комплексов белок–лиганд, содержащем различные белки и лиганды. Особенностью этого сравнения является то, что в обеих программах использовались одни и те же молекулы, одни и те же молекулярные поверхности и их триангуляционные сетки и одни и те же заряды соответствующих атомов рассмотренных молекул, а сравнение проводилось при расчетах полярной составляющей энергии сольватации белков, лигандов и их комплексов, а также при расчетах энергии десольватации, играющей важную роль при связывании лигандов (молекул-кандидатов в ингибиторы) с белками. Для тестового набора комплексов исследовано влияние на результат шага триангуляционной сетки на поверхности раздела молекула–растворитель. Показано, что для расчета энергии десольватации при связывании лиганда с белком с помощью программы MCBHSOLV с точностью лучше 1 ккал/моль достаточно проводить расчеты с шагом сетки 0.2 Å, затрачивая при этом всего несколько минут на комплекс.

Ключевые слова: континуальная модель растворителя, поляризационные заряды, триангуляционная сетка, молекулярная поверхность, поверхность исключенного из растворителя объема, энергия сольватации, энергия десольватации, энергия взаимодействия белок–лиганд, растворитель, мультизарядовый метод.

1. Введение. В настоящее время рациональная разработка лекарств становится не только все более актуальной, но и возможной благодаря интенсивному развитию молекулярной биологии, а также методов рентгеноструктурного анализа белков. Для многих патологий сравнительно быстро выявляются белки-мишени, блокирование работы которых позволяет остановить развитие заболевания, определяются их трехмерные атомно-молекулярные структуры, а в них — так называемые активные центры, ответственные за функционирование этих белков. Блокировка работы этих белков-мишеней осуществляется путем избирательного связывания с их активными центрами сравнительно небольших молекул-ингибиторов,

¹ Московский государственный университет им. М.В. Ломоносова, факультет вычислительной математики и кибернетики, Ленинские горы, д. 1, стр. 52, 119991, Москва; аспирант, e-mail: mikhailovsk@gmail.com

² ООО “Димонта”, ул. Нагорная, д. 15, корп. 8, 117186, Москва; программист, e-mail: io@dimonta.com

³ Институт вычислительной математики РАН, ул. Губкина, д. 8, 119333, Москва; ст. науч. сотр.; e-mail: ivan.oseledets@gmail.com

⁴ ООО “Димонта”, ул. Нагорная, д. 15, корп. 8, 117186, Москва; зам. директора по информационным технологиям, e-mail: sulimovv@mail.ru

⁵ Институт вычислительной математики РАН, ул. Губкина, д. 8, 119333, Москва; директор, e-mail: eugene.tyrtyshnikov@gmail.com

⁶ Научно-исследовательский вычислительный центр, Московский государственный университет им. М.В. Ломоносова, Ленинские горы, д. 1, корп. 4, 119992, Москва; зав. лабораторией, e-mail: vladimir.sulimov@gmail.com

которые и становятся основой новых лекарств. Поиск или разработка таких ингибиторов представляет собой и ключевой, и наиболее сложный этап разработки новых лекарств. Существенную помощь в решении этой непростой задачи может дать применение методов компьютерного молекулярного моделирования [1]. Способность ингибитора блокировать работу данного белка-мишени определяет энергия его связывания с белком — чем она больше, тем выше эффективность лекарства на основе этого ингибитора.

В компьютерной разработке новых ингибиторов первое место занимает докинг лигандов [2–4], т.е. молекул-кандидатов в ингибиторы, в активные центры белков-мишеней, в результате которого происходит позиционирование лигандов в этих центрах и вычисление свободной энергии их связывания с белком-мишенью. Точность вычисления этой величины — ключевой параметр, определяющий надежность предсказания ингибирующей способности лигандов и напрямую влияющий на эффективность применения методов молекулярного моделирования для разработки новых лекарств. Достаточно высокая практическая предсказуемость достигается при погрешности расчета энергии связывания белок–лиганд, не превосходящей 1 ккал/моль. Эта точность определяется многими факторами, среди которых важную роль играет описание взаимодействия молекул с растворителем — водой. В настоящее время при молекулярном моделировании используются два метода описания взаимодействия молекул с растворителем. Первый — явная модель растворителя, в которой окружающий молекулу растворимого вещества растворитель представляется в виде большого массива более или менее равномерно распределенных молекул воды. Из-за больших вычислительных затрат такие методы никогда не используются в задачах докинга, как, впрочем, и при квантово-химическом описании молекулярных систем. Второй метод описания растворителя — это неявная, континуальная модель растворителя [5–7], в которой весь растворитель, окружающий данную молекулу, представляется в виде, как правило, однородного диэлектрика с диэлектрической проницаемостью растворителя (для воды 78.5 при $T = 300$ К). Такие модели позволяют сравнительно точно проводить расчеты кулоновских взаимодействий в системе белок–лиганд–растворитель, где растворитель (вода) вносит существенный вклад в экранировку кулоновских взаимодействий белок–лиганд. Неявные модели растворителя имеют в настоящее время сравнительно неплохие параметризацию [8, 9] и быстрое действие, однако последнего далеко не достаточно, чтобы применять эти модели в задачах докинга, в которых при позиционировании лиганда в белке приходится рассматривать миллионы молекулярных конфигураций. Поэтому на повестке дня рациональной разработки новых лекарств весьма остро стоит задача о существенном повышении скорости решения уравнений неявных моделей растворителя без потери точности.

В настоящей статье предложен новый алгоритм решения матричных уравнений РСМ (Polarized Continuum Model) [5, 6, 9], позволяющий получить ускорение в сотни раз по сравнению с обычными итерационными методами [9, 10] без существенной потери точности. Этот новый алгоритм основан на использовании мультизарядового приближения больших матриц [11]. Большие матрицы в уравнении РСМ возникают при триангуляции поверхности, разделяющей молекулу и растворитель, и при дискретизации непрерывной плотности поляризованных зарядов, наводимых на поверхности диэлектрика атомными зарядами молекулы. Для получения высокой точности расчетов кулоновского взаимодействия белков с водой нужен достаточно малый шаг триангуляционной сетки (≤ 0.1 Å), что приводит к сотням тысяч поверхностных элементов и, соответственно, к большим размерам матриц, причем чем меньше шаг сетки и выше точность расчетов, тем больше получаем ускорение при применении мультизарядового разложения больших матриц для решения уравнений РСМ. Приведено интегральное уравнение РСМ для вычисления поляризованных зарядов, его дискретизация, кратко описан способ его решения обычным методом [10], реализованным в программе DISOLV [9], описан алгоритм и способ решения матричных уравнений РСМ с помощью мультизарядового приближения для больших матриц [11], реализованный в оригинальной программе MCBHSOLV, и проведено сравнение результатов расчетов обоими методами электростатической части энергии взаимодействия белков, лигандов и их комплексов с водой (далее для краткости — энергии сольватации), а также энергии десольватации при связывании этих белков и лигандов в комплексы, которая вычисляется как разность энергии сольватации комплекса и суммы энергий сольватации белка и лиганда, для набора комплексов белок–лиганд, среди которых имеются различные белки и лиганды разных размеров. Энергия десольватации непосредственно входит в свободную энергию связывания белок–лиганд, и точность ее вычисления напрямую влияет на точность вычисления последней, а значит и на надежность предсказания констант ингибирования молекул-кандидатов в ингибиторы. Особенностью этого сравнения является то, что в обеих программах (DISOLV и MCBHSOLV) используются одни и те же молекулы, одни и те же молекулярные поверхности и их триангуляционные сетки и одни и те же заряды атомов соответствующих молекул.

2. Уравнение РСМ и его дискретизация на молекулярной поверхности. Модель РСМ

(Polarized Continuum Model) [5, 6, 9, 12] является моделью для рассмотрения электростатических взаимодействий точечных зарядов с диэлектриком, причем заряды находятся в полости диэлектрика, а диэлектрик вне полости является однородным и бесконечным. Подобная модель может быть применена к описанию электростатического взаимодействия растворенной молекулы с растворителем, если считать растворитель диэлектриком, полость в котором образована атомами молекулы, каждый атом которой представлен в виде сферы определенного радиуса с точечным зарядом в центре. Свободная энергия электростатического взаимодействия растворенной молекулы с растворителем ΔG_{pol} определяется кулоновским взаимодействием атомных и поляризационных зарядов, причем поверхностная плотность последних $\sigma(\mathbf{r})$ определяется из решения интегрального уравнения

$$\sigma(\mathbf{r}) = \frac{1 - \varepsilon}{2\pi(1 + \varepsilon)} \left(\sum_i \frac{Q_i ((\mathbf{r} - \mathbf{R}_i) \cdot \mathbf{n})}{|\mathbf{r} - \mathbf{R}_i|^3} + \int_{\Omega} \frac{\sigma(\mathbf{r}') ((\mathbf{r} - \mathbf{r}') \cdot \mathbf{n})}{|\mathbf{r} - \mathbf{r}'|^3} dS' \right), \quad (1)$$

где Q_i — заряд i -го атома молекулы; \mathbf{R}_i — радиус-вектор i -го атома молекулы; Ω — поверхность, формирующая полость (полости) в растворителе; σ — поверхностная плотность поляризационного заряда; \mathbf{r} — радиус-вектор элемента поверхности; \mathbf{n} — нормаль к поверхности полости, направленная в растворитель; ε — относительная диэлектрическая проницаемость растворителя. В уравнении (1) полагается, что относительная диэлектрическая проницаемость внутри полости равна 1.

Для решения уравнения (1) поверхность полости (полостей) триангулируется и, таким образом, аппроксимируется связным набором дискретных элементов, где i -й дискретный элемент характеризуется своим центром \mathbf{r}_i , нормалью \mathbf{n}_i , площадью S_i и поляризационным зарядом q_i . После триангуляции интегральное уравнение (1) преобразуется к матричному уравнению

$$A\mathbf{q} = B\mathbf{Q}. \quad (2)$$

В уравнении (2) введены следующие обозначения: \mathbf{q} — вектор-столбец поляризационных зарядов дискретных элементов поверхности и \mathbf{Q} — вектор-столбец зарядов атомов молекулы. Матрица A имеет размеры $N \times N$, где N — количество дискретных элементов поверхности; матрица B имеет размеры $N \times M$, где M — количество атомов в молекуле. Элементы матриц A и B вычисляются по формулам

$$A_{ij} = \begin{cases} \frac{(\varepsilon - 1)}{4\pi(1 + \varepsilon)} \frac{((\mathbf{r}_i - \mathbf{r}_j) \cdot \mathbf{n}_i) S_i}{|\mathbf{r}_i - \mathbf{r}_j|^3}, & i \neq j, \\ \frac{\varepsilon}{1 + \varepsilon} - \sum_{k \neq j} A_{kj}, & i = j, \end{cases} \quad (3)$$

$$B_{ij} = \frac{1 - \varepsilon}{4\pi(1 + \varepsilon)} \frac{((\mathbf{r}_i - \mathbf{R}_j) \cdot \mathbf{n}_i) S_i}{|\mathbf{r}_i - \mathbf{R}_j|^3}. \quad (4)$$

Полярная часть энергии сольватации вычисляется по формуле

$$\Delta G_{\text{pol}} = \frac{1}{2} \mathbf{Q}^T D \mathbf{q}, \quad (5)$$

где элементы матрицы D вычисляются по формуле

$$D_{ij} = \frac{1}{|\mathbf{R}_i - \mathbf{r}_j|}. \quad (6)$$

Итак, чтобы решить электростатическую задачу определения поляризационных зарядов на поверхности диэлектрика, имитирующего растворитель и окружающего молекулу, необходимо построить эту поверхность, которая в неявных моделях растворителя называется поверхностью исключенного из растворителя объема, или SES — Solvent Excluded Surface. В настоящей работе для построения этой поверхности используется программа TAGSS [9, 13–15], которая строит и триангулирует поверхность SES следующим образом. Поверхность строится за две основные стадии: первичная и вторичная обкатка [16, 17], а затем генерирование сетки триангуляции с использованием параметров, полученных на первой стадии. Молекула представляется совокупностью жестких сфер, центры которых расположены на ядрах составляющих молекулу атомов; радиусы этих сфер различны для атомов разных типов и являются параметрами континуальной модели растворителя [9]. Первичная обкатка — это образование молекулярной поверхности

путем перемещения по молекуле пробной сферы, имитирующей молекулу растворителя, так что в любой момент времени этого процесса должна быть как минимум одна точка касания пробной сферы и какой-либо сферы, соответствующей некоторому атому молекулы. Все возможные точки касания сферы первичной обкатки и сфер атомов определяют точки молекулярной поверхности SES. Такая модель построения хорошо воспроизводит форму молекулы с точки зрения растворителя, а детали поверхности, фактически, ограничиваются размером молекулы растворителя.

Процедура первичной обкатки иногда, особенно в случае больших молекул, может приводить к нежелательным самопересечениям поверхности и изломам [9]; для их сглаживания применяется процедура вторичной обкатки [17]. Суть метода вторичной обкатки состоит в замене фрагментов поверхности вблизи изломов и самопересечений на другие гладкие фрагменты, которые определяются путем обкатки небольшой сферой (называемой сферой вторичной обкатки) поверхности SES вблизи изломов. Диаметр пробной сферы вторичной обкатки выбирается так, чтобы он был меньше диаметра пробной сферы первичной обкатки, но больше определенного критического расстояния, зависящего от средней величины дискретного элемента поверхности. Обе стадии обкатки (первичная и вторичная) позволяют получить описание поверхности молекулы в виде наборов координат положения и ориентации сферических и тороидальных фрагментов, а также геометрической связанности этих фрагментов друг с другом. Исходными параметрами, на основе которых производятся все последующие вычисления, являются данные о координатах атомов, их радиусах (в данной работе они взяты из [9]), а также значения радиусов сфер первичной (1.4 Å) и вторичной обкаток (0.4 Å), а также критического расстояния (0.15 Å) — последние три параметра использовались в [9] при определении атомных радиусов. Критическое расстояние определяет некоторое минимальное расстояние, допустимое между двумя близкими, но поверхностно-удаленными точками поверхности. Близкие точки считаются поверхностно-удаленными, если минимальное расстояние между ними, вычисленное вдоль поверхности, велико по сравнению с прямым расстоянием.

3. Решение уравнения РСМ в программе DISOLV. Для решения линейной системы (2) в программе DISOLV используется одношаговый итерационный метод [18] со специально подобранным начальным решением и набором параметров τ_{n+1} : $D_A \frac{q^{(n+1)} - q^{(n)}}{\tau_{n+1}} + Aq^{(n)} = BQ$, где матрица D_A — диагональная часть матрицы A . Начальное приближение для поляризационных зарядов рассчитывается по следующей формуле:

$$q^{(0)} = \frac{1 + \varepsilon}{\varepsilon} BQ. \quad (7)$$

Эта формула (7) является следствием уравнения (1) в предположении отсутствия взаимного влияния поляризационных зарядов друг на друга за исключением того, что коэффициент перед суммой изменен, так как суммарный поляризационный заряд должен удовлетворять соотношению

$$\sum_i q_i = \frac{1 - \varepsilon}{\varepsilon} \sum_k Q_k. \quad (8)$$

Для выбора τ_{n+1} используются следующие числовые значения, предложенные в работе [10]: если $\sum_i q_i^{(n)}$ и $\sum_i q_i^{(n+0.5)}$ находятся по разные стороны от равновесного значения суммарного поляризационного заряда $\frac{1 - \varepsilon}{\varepsilon} \sum_k Q_k$, то $\tau_{n+1} = \left(\sum_i q_i^{(n)} - \frac{1 - \varepsilon}{\varepsilon} \sum_k Q_k \right) \left(\sum_i q_i^{(n)} - \sum_i q_i^{(n+0.5)} \right)^{-1}$, иначе

$$\tau_{n+1} = 0.65. \quad (9)$$

Кроме того, после шестой итерации τ_{n+1} всегда вычисляется по формуле (9).

Итерации останавливаются тогда, когда изменение свободной энергии сольватации становится меньше заданного порога на протяжении трех итераций подряд. Свободная энергия сольватации рассчитывается на каждой итерации по формуле (5).

4. Мультизарядовый метод. Для нахождения неизвестных q необходимо решить систему линейных уравнений $Aq = BQ$ с плотной матрицей A . Так как матрица A является плотной, использование прямых методов решения потребует $O(N^3)$ операций, поэтому для решения системы необходимо использовать итерационные методы. Основной операцией для итерационных методов является матрично-векторное произведение. Чем быстрее вычисляется это произведение, тем меньше времени потребуется для получения решения. Умножение матрицы A на вектор с помощью стандартного алгоритма имеет сложность $O(N^2)$,

где N — размер матрицы. Поэтому при больших N , т.е. $N \sim 10^5$, время решения становится очень большим. Таким образом, появляется необходимость в новом подходе, имеющем меньшую вычислительную сложность.

Значения элементов матрицы A вычисляются по формуле (3). Важно, что их можно представить в виде $A_{ij} = f(X_i, Y_j)$, где X_i и Y_j — дискретные элементы поверхности. Таким образом, элемент матрицы A_{ij} задает *элементарное взаимодействие* двух дискретных элементов. Действительно, вычисление одного элемента вектора Aq для всех i можно представить в виде $\sum_j A_{ij}q_j = \sum_j f(X_i, Y_j)q_j$, что имеет простую “физическую” интерпретацию: источники с зарядами q_j , находящиеся в Y_j , создают “потенциал” в приемнике X_i , а элементарное взаимодействие задается матричным элементом A_{ij} . В нашем случае в качестве источников и приемников используются дискретные элементы поверхности SES.

Для эффективной работы мультизарядового метода требуется наличие \mathcal{H}^2 -структуры [19, 20] у матрицы A . Существование такой структуры доказано для определенного класса задач. Можно показать, что описанное ранее взаимодействие принадлежит этому классу задач. Таким образом, матрицы, порожденные элементарными взаимодействиями дискретных элементов поверхности в соответствии с формулой (3), хорошо приближаются \mathcal{H}^2 -матрицами. Такая \mathcal{H}^2 -структура подразумевает наличие мозаичного разбиения [21] матрицы A . Мозаичным разбиением называется представление всей матрицы в виде набора подматриц так, что каждый элемент матрицы принадлежит ровно одной подматрице из этого набора. Это разбиение строится стандартным образом на основе критерия геометрической разделенности источников и приемников. Подматрицы, соответствующие геометрически разделенным источникам и приемникам, называются *дальними*. Построение малоранговых аппроксимаций для каждой отдельной дальней подматрицы составляет основу мозаично-скелетного метода [35]. Такое представление также называют \mathcal{H} -матрицами [36, 37]. Основные элементы \mathcal{H}^2 -структуры (блочные строки и блочные столбцы) являются специальными объединениями дальних подматриц и имеют иерархические зависимости. Фактически, \mathcal{H}^2 -структура является алгебраическим аналогом быстрого мультиполюсного метода [22].

Мультизарядовый метод, предложенный в работе [11], позволяет строить приближение \mathcal{H}^2 -матрицы по значениям ее элементов с высокой надежностью и не требует дополнительной информации. Другие возможные подходы основаны либо на \mathcal{H} -матрицах, либо требуют априорной информации о структуре или свойствах ядра взаимодействия (мультиполюсный метод [22], матричный мультиполюсный метод [23], nested cross approximation [24]).

Основой мультизарядового метода является скелетное разложение [25] $A \approx C\hat{A}^{-1}R$, где C состоит из базисных столбцов матрицы A , R — из базисных строк матрицы A , а \hat{A} — пересечение базисных строк и столбцов матрицы A . Несложным преобразованием можно получить:

$$A \approx (C\hat{A}^{-1})R = \hat{C}R, \quad A \approx (C\hat{A}^{-1})\hat{A}(\hat{A}^{-1}R) = \hat{C}\hat{A}\hat{R}.$$

Обозначим через $A(t, s)$ подматрицу на строках t и столбцах s матрицы A . Если подматрица на пересечении набора строк t и набора столбцов s принадлежит мозаичному разбиению и имеет малый вычислительный ранг, то $A(t, s) \approx \hat{C}_t R_s$,

$$A(t, s) \approx \hat{C}_t \hat{A}(t, s) \hat{R}_s. \tag{10}$$

Пусть набор строк t при мозаичном разбиении делится на наборы строк t_1 и t_2 , а набор столбцов s выбран так, что подматрица на пересечении строк t и столбца s обладает малым вычислительным рангом. Тогда верно следующее: $A(t, s) \approx \hat{C}_t R_0$, $A(t_1, s) \approx \hat{C}_{t_1} R_1$, $A(t_2, s) \approx \hat{C}_{t_2} R_2$, где R_0 состоит из базисных строк подматрицы $A(t, s)$, R_1 состоит из базисных строк подматрицы $A(t_1, s)$, R_2 состоит из базисных строк подматрицы $A(t_2, s)$. Из построения t получаем

$$A(t, s) = \begin{bmatrix} A(t_1, s) \\ A(t_2, s) \end{bmatrix} \approx \begin{bmatrix} \hat{C}_{t_1} & 0 \\ 0 & \hat{C}_{t_2} \end{bmatrix} \begin{bmatrix} R_1 \\ R_2 \end{bmatrix} \approx \begin{bmatrix} \hat{C}_{t_1} & 0 \\ 0 & \hat{C}_{t_2} \end{bmatrix} M_t R_0.$$

Отсюда заключаем, что матрицы \hat{C} обладают рекурсивной зависимостью $\hat{C}_t = \begin{bmatrix} \hat{C}_{t_1} & 0 \\ 0 & \hat{C}_{t_2} \end{bmatrix} M_t$, а матрицы

перехода M_t можно вычислять по базисным строкам: $\begin{bmatrix} R_1 \\ R_2 \end{bmatrix} \approx M_t R_0$. Однако для вычисления M_t знать все элементы матриц R_1 и R_2 не обязательно, достаточно знать те столбцы, которые содержат подматрицу максимального объема [26]. Зачастую “хорошие” столбцы выбирают по геометрическим признакам.

В работе [11] предложен итерационный метод выбора “хороших” столбцов и строк, составляющий основу мультизарядового метода. Мультизарядовый метод является алгебраическим и требует геометрическую информацию только для построения мозаичного разбиения. Сложность мультизарядового метода составляет $O(N)$ операций как для аппроксимации самой матрицы A , так и для умножения матрицы в данном представлении на вектор.

Мультизарядовый метод реализован в качестве модуля на языке программирования Python. Вычислительно узкие места (такие как вычисление базисных строк, столбцов и матриц перехода) реализованы при помощи Cython [27] и достигают скорости работы программы, написанной на языке Си. Кроме того, библиотека Intel Math Kernel Library [28] может быть использована для ускорения вычислений (в численных экспериментах используется последовательная версия этой библиотеки). Весь программный модуль и информацию по его использованию можно получить по ссылке <https://bitbucket.org/muxas/h2tools/>.

5. Решение уравнения РСМ в программе MCBHSOLV. Входными параметрами для мультизарядового метода являются дискретизация поверхности молекулы и соответствующая этой дискретизации матрица A (см. (3); передается программе в виде функции от набора строк и столбцов, которая возвращает подматрицу на их пересечении). Программа MCBHSOLV сама строит иерархическое кластерное дерево для элементов дискретизации (по центрам этих элементов) и использует это дерево для построения \mathcal{H}^2 -разложения плотной матрицы A с требуемой точностью.

Построение \mathcal{H}^2 -разложения, как и матрично-векторное произведение в \mathcal{H}^2 -формате, требует $O(N)$ операций, где N — порядок квадратной матрицы A . Именно быстрое построение разложения и еще более быстрые матрично-векторные операции позволяют успешно применять мультизарядовый метод для значительного ускорения вычисления энергий сольватации и десольватации различных молекул. В отличие от программы DISOLV, для решения уравнения (2) программа MCBHSOLV использует не одношаговый итерационный метод, а итерационный метод GMRES [29]. Достоинством метода GMRES является то, что решение ищется на крыловских подпространствах [30], что значительно сокращает количество итераций, необходимых для решения уравнения с требуемой точностью.

Итоговый алгоритм работы MCBHSOLV представлен ниже. Для численных экспериментов использовалась последовательная версия программы.

Алгоритм работы программы MCBHSOLV

Дано: файл с данными о дискретных элементах поверхности SES (r_i , n_i , s_i для каждого элемента поверхности) и атомах (координаты R и численные значения зарядов Q).

Вычислить: полярная составляющая энергии сольватации молекулы.

1. Считывание данных о дискретных элементах поверхности SES и атомах из файла при помощи пакета h2tools.
 2. Построение иерархического кластерного дерева для дискретных элементов поверхности SES (по r_i , встроено в пакет h2tools).
 3. Вычисление очереди обработки кластерного дерева (встроено в пакет h2tools).
 4. Вычисление правой части уравнения (2) умножением соответствующей матрицы B (4) на вектор зарядов атомов Q : $f = BQ$.
 5. Вычисление выражения $Q^T D$, необходимого для вычисления полярной составляющей энергии сольватации (5): $d = Q^T D$.
 6. Построение мультизарядового разложения матрицы A (3) с 1 итерацией и параметром точности $\tau = 10^{-4}$ (по данному иерархическому кластерному дереву и очереди обработки кластерного дерева, встроено в пакет h2tools).
 7. Решение уравнения (2) с вычисленной ранее правой частью при помощи метода GMRES $Aq = f$.
 8. Вычисление полярной составляющей энергии сольватации (5): $\Delta G_{\text{pol}} = \frac{1}{2} dq$.
-

6. Численные эксперименты. В данном разделе приведены результаты расчетов энергий сольватации и десольватации (только их электростатические составляющие) для 22 комплексов белок–лиганд, вычисленные тремя различными методами.

1. Решение интегрального уравнения (1) в рамках континуальной модели PCM. Система линейных уравнений (2) решалась одношаговым итерационным методом, описанным в разделе 2, с помощью программы DISOLV, тщательная валидация которой представлена в работе [9].

2. Решение интегрального уравнения (1) в рамках континуальной модели PCM. Матрица A из системы (2) аппроксимировалась мультизарядовым методом с относительной точностью 10^{-4} , при этом для хранения аппроксимации требовалась дополнительная оперативная память (до 4.5 гигабайт). Решение строилось при помощи итерационного метода GMRES [29]. В настоящем разделе статьи коротко обозначается как MCBH.

3. Вычисление электростатической составляющей взаимодействия молекул с растворителем с помощью эвристического метода Surface Generalized Born [9, 31], реализованного в программе DISOLV. В настоящем разделе коротко обозначается как SGB. Высокая скорость вычислений (для каждого атома молекулы вычисляются четыре поверхностных интеграла) при сравнительно высокой точности сделала подобные методы весьма популярными, особенно применительно к макромолекулам [7].

Для каждой молекулы построение поверхности и ее триангуляция проводились с помощью программы TAGSS (подпрограмма DISOLV), а заряды атомов вычислялись в соответствии с силовым полем MMFF94 [32]. Трехмерная структура комплексов бралась из базы данных Protein Data Base [33], идентификаторы комплексов приведены по горизонтальным осям соответствующих рисунков (см. ниже). Структуры белков были получены при помощи удаления из PDB-комплексов всех небелковых включений, водороды на белках были расставлены при помощи программы APLITE [4]. Лиганды также были взяты из PDB-комплексов, расстановка водородов на них осуществлялась с помощью программы Avogadro [34]. В комплексах и в свободном состоянии положения лигандов получались локальной оптимизацией в вакууме в рамках силового поля MMFF94 (типизация атомов в этом поле проводилась программой APLITE) по всем атомам лигандов (в комплексах — при жестких белках) из начального нативного положения, т.е. из положения лиганда в закристаллизованном комплексе, взятом непосредственно из базы PDB.

Для каждого белка и комплекса белок-лиганд поверхностные приближения были построены с тремя различными шагами сетки (0.3, 0.2 и 0.1 Å), а для каждого лиганда — с шагами 0.3, 0.2, 0.1 и 0.05 Å. Расчеты, соответствующие шагам сетки 0.3, 0.2 и 0.1 Å, отмечены на графиках (см. далее) метками 0.30, 0.20 и 0.10 соответственно.

Оказалось, что различие в энергии сольватации между SGB и PCM почти не зависит от шага сетки. Это было учтено при построении графиков: для сравнения точности приводится только результат SGB на сетке с шагом 0.10 Å.

Для рассматриваемых молекул рассчитанные энергии сольватации находятся в диапазонах: для комплексов и белков $\approx (-4000)$ – (-990) ккал/моль, для лигандов $\approx (-160)$ – (-10) ккал/моль, а время счета CPU (без учета времени построения сеток) при использовании метода PCM программы DISOLV составляет при шаге сетки 0.1 Å: для комплексов и белков $\approx (27000)$ – (200000) с, для лигандов $\approx (17)$ – (150) с в зависимости от размера молекулы. Время счета программы DISOLV обратно пропорционально шагу сетки в четвертой степени. Так, переход от шага сетки 0.1 к 0.2 Å уменьшает время счета примерно в 15 раз, а переход от шага сетки 0.2 к 0.3 Å уменьшает время счета примерно в 5 раз.

На рис. 1 для каждого комплекса приведены значения модуля разности между энергиями сольватации комплекса, полученными с помощью метода PCM из программы DISOLV и с помощью метода PCM из программы MCBHSOLV (или с помощью метода SGB из программы DISOLV). Сравнение проводится для каждого из трех шагов сеток дискретизации поверхности молекулы (0.1, 0.2 и 0.3 Å). Для разных методов расчета каждой молекулы использовались одни и те же триангуляционные сетки для данного шага. На горизонтальной оси отложены идентификаторы комплексов в базе данных Protein Data Base. Как и во всей статье, имеется в виду разность электростатических составляющих энергии сольватации.

Как можно видеть, результаты расчетов энергии сольватации быстрым (см. ниже) методом MCBH отличаются от результатов расчетов PCM на величину, меньшую 1 ккал/моль для триангуляционных сеток с шагами 0.1 и 0.2 Å (параметр относительной точности расчета MCBH был задан равным 10^{-4}). Результаты расчетов быстрым эвристическим методом SGB отличаются от результатов PCM значительно больше — на 100–200 ккал/моль. Для энергий сольватации белков получаем аналогичную картину. Для лигандов расхождение значений энергии сольватации для PCM и MCBH (или SGB) значительно меньше: между результатами, полученными с помощью PCM и MCBH для шага сетки 0.05 Å, модуль разности энергий сольватации для лигандов разных комплексов составляет величину 10^{-4} – 10^{-2} ккал/моль и даже

для шага 0.3 \AA не превосходит 1 ккал/моль. Результаты расчетов лигандов методом SGB отличаются от результатов расчетов методом PCM не более, чем на 5 ккал/моль.

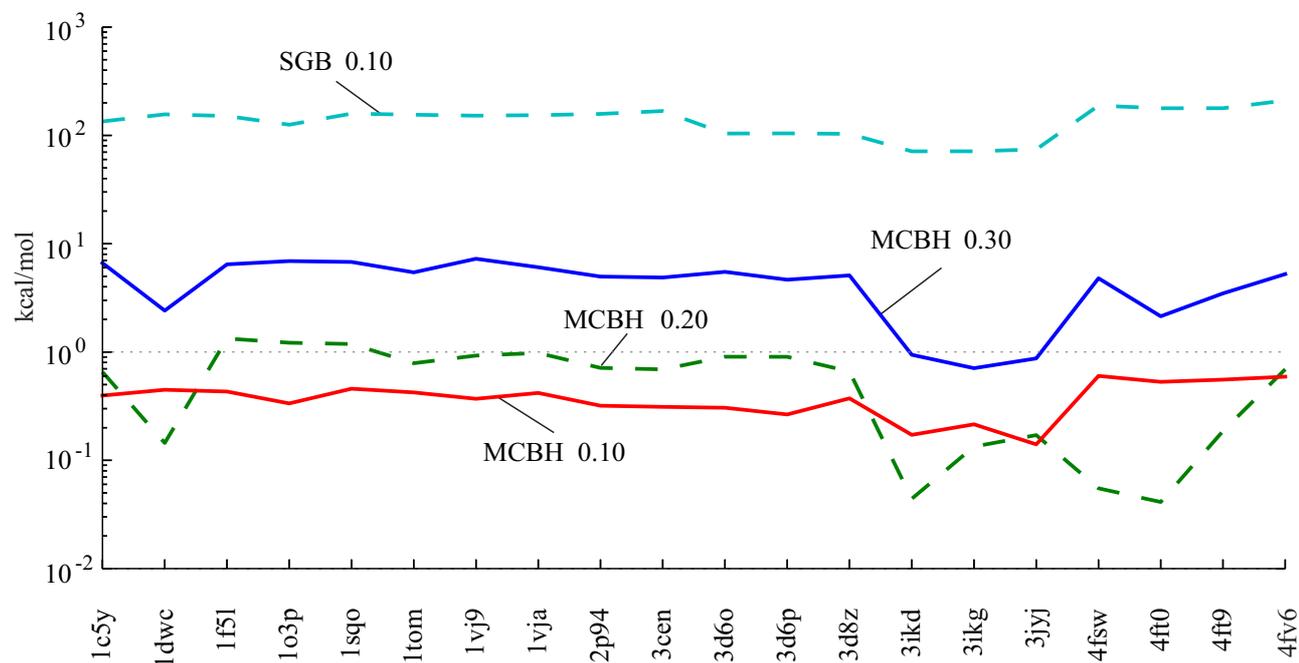


Рис. 1. Модуль разности энергий сольватации комплексов белок–лиганд, вычисленных методами PCM, MCBH и SGB

Энергии десольватации различных комплексов могут различаться как по знаку, так и по величине от единиц до сотен ккал/моль; при ее вычислении происходит заметная компенсация различий значений энергий сольватации белка, лиганда и их комплекса, полученных разными методами. Результаты расчетов методом MCBH отличаются от соответствующих результатов PCM менее, чем на 1 ккал/моль для всех рассмотренных шагов триангуляционных сеток, а для шага 0.1 \AA различие составляет всего 10^{-2} – 10^{-1} ккал/моль (параметр относительной точности расчета MCBH был задан равным 10^{-4}). Результаты расчетов энергии десольватации методом SGB отличаются от результатов PCM не более, чем на 10 ккал/моль.

На рис. 2 показано ускорение вычисления энергии десольватации при помощи метода MCBH (или SGB) относительно метода PCM на соответствующих сетках (не учитывается время построения поверхности SES вокруг молекул).

Из этого рисунка видно, что в методе MCBH получается существенное ускорение по сравнению с методом PCM на тех же сетках, причем это ускорение тем больше, чем меньше шаг сетки, а значит и существенно больше количество дискретных элементов поверхности и размер матриц A и B в уравнении для вычисления поляризованных зарядов. Это ускорение слабо меняется от комплекса к комплексу и при шаге сетки 0.1 \AA для отдельных комплексов достигает 300 раз! Столь большое ускорение обусловлено двумя причинами:

(i) переходом от одношагового итерационного метода [18] решения матричного уравнения (2) в программе DISOLV к гораздо более эффективному итерационному методу GMRES [29], реализованному в программе MCBHSOLV,

(ii) использованием быстрого построения \mathcal{H}^2 -разложения плотной матрицы A и еще более быстрых матрично-векторных произведений в \mathcal{H}^2 -формате.

Оперативная память, необходимая для хранения \mathcal{H}^2 -разложения матрицы A из уравнения (2) для сетки с шагом 0.1 \AA , составляла от 1.5 до 4.5 гигабайт в зависимости от размера и формы комплекса. В связи с тем, что матрицы $\hat{A}(t, s)$ из уравнения (10) являются подматрицами исходной матрицы A , их можно не хранить, а вычислять по мере необходимости, что позволяет сократить максимальное требование по оперативной памяти с 4.5 гигабайт до 0.9 гигабайт, но увеличивает время, необходимое для каждого матрично-векторного произведения в \mathcal{H}^2 -формате (в худшем случае это увеличивает полное время работы программы MCBHSOLV в два раза). Однако для численных экспериментов использовалась та версия

программы, которая хранит все матрицы $\hat{A}(t, s)$.

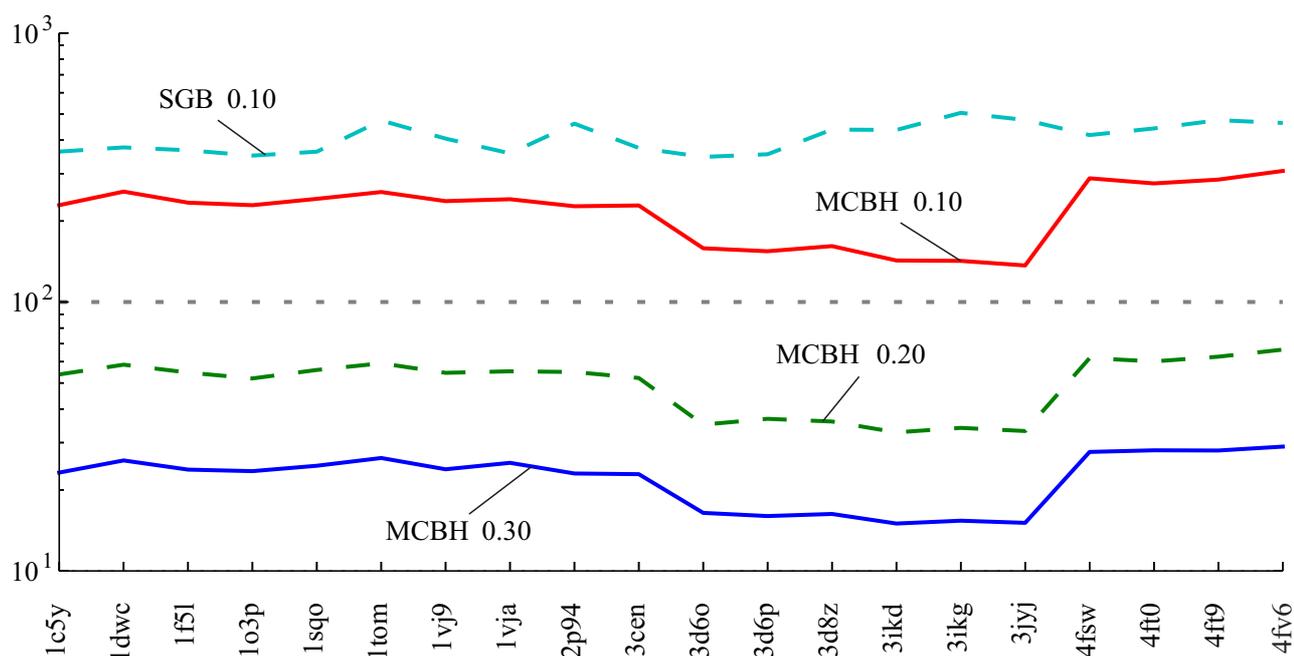


Рис. 2. Ускорение расчетов при помощи метода MCBH относительно метода PCM на соответствующих сетках

При увеличении требуемой точности возрастет и требуемая память, которую можно сократить за счет вышеописанного действия (что повлечет за собой увеличение времени расчетов). В результате полученного ускорения для наиболее аккуратных вычислений (шаг сетки 0.1 Å) вместо 4.7 суток (для комплекса 4fv6), которые необходимы для расчета энергии десольватации с помощью программы DISOLV методом PCM, программе MCBHSOLV требуется всего около 20 минут для получения практически того же значения энергии десольватации. Ускорение расчета энергии десольватации методом SGB по сравнению с PCM программы DISOLV для шага сетки 0.1 Å близко к ускорению MCBH для шага сетки 0.1 Å.

7. Заключение. В настоящей работе рассматриваются расчеты полярной части энергии взаимодействия молекул с водой в континуальной модели растворителя PCM — электростатическая составляющая энергии сольватации больших и малых молекул с водой.

Рассмотрен способ существенного ускорения решения уравнения для поляризационных зарядов на поверхности SES раздела молекула–растворитель в рамках континуальной модели растворителя.

Разработаны новые алгоритмы, их программная реализация MCBHSOLV и проведено сравнение результатов и времен счета с соответствующими величинами, полученными с помощью программы DISOLV [9]. Сравнение проведено на тестовом наборе комплексов белок–лиганд, содержащем различные белки и лиганды. Особенностью этого сравнения является то, что в обеих программах используются одни и те же молекулы, одни и те же молекулярные поверхности и их триангуляционные сетки и одни и те же заряды соответствующих атомов рассмотренных молекул. Сравнение показало, что программа MCBHSOLV работает быстрее программы DISOLV в 100 и более раз практически без потери точности вычислений, причем чем меньше шаг триангуляции сольватационной поверхности, тем больше ускорение.

Ускорение получено применением двух методов. Во-первых, с помощью недавно разработанного мультитарядового метода аппроксимации матриц больших размеров [11], которые хорошо приближаются \mathcal{H}^2 -матрицами [19, 20]. Быстрое построение \mathcal{H}^2 -разложения и еще более быстрые матрично-векторные операции в \mathcal{H}^2 -формате позволяют успешно применить мультитарядовый метод для существенного ускорения решения матричного уравнения PCM для поляризационных зарядов. Во-вторых, путем перехода от одношагового итерационного метода решения уравнения для поляризационных зарядов в континуальной модели растворителя, реализованного в программе DISOLV [9], к итерационному методу GMRES [29], основным достоинством которого является то, что решение ищется на крыловских подпространствах [30], а это приводит к значительному сокращению количества итераций, необходимых для решения уравнения для поляризационных зарядов с требуемой точностью.

Отличия в результатах по энергии десольватации в модели РСМ на идентичных сетках между программами MCBHSOLV и DISOLV не превосходят ≈ 1.0 , ≈ 0.6 и ≈ 0.2 ккал/моль для шагов сетки 0.3, 0.2 и 0.1 Å соответственно. Таким образом, для вычисления энергии десольватации с приемлемой точностью можно использовать программу MCBHSOLV с шагом сетки 0.2 Å и тратить при этом всего несколько минут на комплекс.

Работа выполнена при поддержке Министерства образования и науки Российской Федерации (соглашение № 14.132.21.1346), при поддержке РФФИ (коды проектов 12-01-00565-а и 12-01-33013-мол-а-вед), а также при поддержке ООО “Димонта” (Москва, РФ).

СПИСОК ЛИТЕРАТУРЫ

1. Садовничий В.А., Сулимов В.Б. Суперкомпьютерные технологии в медицине // Суперкомпьютерные технологии в науке, образовании и промышленности / Под ред. Вл.В. Воеводина. В.А. Садовничего, Г.И. Савина. М.: Изд-во Моск. гос. ун-та, 2009. 16–23.
2. 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.
3. 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.
4. 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.
5. Tomasi J., Persico M. Molecular interactions in solution: an overview of method based on continuous distributions of the solvent // Chem. Rev. 1994. **94**, N 7. 2027–2094.
6. Cramer C., Truhlar D. Implicit solvation models: equilibria, structure, spectra, and dynamics // Chem. Rev. 1999. **99**, N 8. 2161–2200.
7. Onufriev A. Continuum electrostatics solvent modeling with the generalized Born model // Modeling Solvent Environments: Applications to Simulations of Biomolecules. Weinheim: Wiley, 2010. 127–166.
8. Bordner A.J., Cavasotto C.N., Abagyan R.A. Accurate transferable model for water, n-Octanol, and n-Hexadecane solvation free energies // J. Phys. Chem. B. 2002. **106**, N 42. 11009–11015.
9. 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**. 246–261.
10. Pomelli C.S., Tomasi J. A new formulation of the PCM solvation method: PCM-QINTn // Theor. Chem. Acc. 1997. **96**, N 1. 39–43.
11. Mikhalev A. Yu., Oseledets I.V. Adaptive nested cross approximation of non-local operators (arXiv preprint: 1309.1773). 2013 (URL: <http://arxiv.org/abs/1309.1773>).
12. Totrov M., Abagyan R. Rapid boundary element solvation electrostatics calculations in folding simulations: successful folding of a 23-residue peptide // Peptide Science. 2001. **60**, N 2. 124–133.
13. Жабин С.Н., Сулимов В.Б. Свидетельство № 2006613753 о государственной регистрации программ для ЭВМ. Зарегистрировано в реестре программ для ЭВМ Федеральной службы по интеллектуальной собственности, патентам и товарным знакам 27 октября 2006.
14. Жабин С.Н., Сулимов В.Б. Программа построения доступной растворителю поверхности для произвольных органических молекул и интерактивный просмотр положений лигандов в активных центрах белков // Сб. материалов XIII Российского национального конгресса “Человек и лекарство”. 3–7 апреля 2006. М., 2006. 15.
15. Жабин С.Н., Сулимов В.Б. Построение гладких молекулярных поверхностей с адаптивной триангуляцией: программа TAGSS // Научная визуализация. 2011. **3**, № 2. 27–53.
16. Connolly M.L. Solvent-accessible surfaces of proteins and nucleic acids // Science. 1983. **221**, N 4612. 709–713.
17. Vorobjev Y.N., Hermans J. SIMS: computation of a smooth invariant molecular surface // Biophys. J. 1997. **73**. 722–732.
18. Самарский А.А. Введение в численные методы. СПб.: Лань, 2005.
19. Hackbusch W., Khoromskij B.N., Sauter S.A. On \mathcal{H}^2 -matrices // Lectures on Applied Mathematics. Heidelberg: Springer, 2000. 2–29.
20. Börm S. Efficient numerical methods for non-local operators: \mathcal{H}^2 -matrix compression, algorithms and analysis. Zürich: Eur. Math. Soc., 2010.
21. Voevodin V.V. On a method of reducing the matrix order while solving integral equations // Numerical Analysis on FORTRAN. Moscow: Moscow Univ. Press, 1979. 21–26.
22. Greengard L., Rokhlin V. A fast algorithm for particle simulations // J. Comput. Phys. 1987. **73**, N 2. 325–348.
23. Ying L., Biros G., Zorin D. A kernel-independent adaptive fast multipole algorithm in two and three dimensions // J. Comput. Phys. 2004. **196**, N 2. 591–626.

24. *Bebendorf M., Venn R.* Constructing nested bases approximations from the entries of non-local operators // Numer. Math. 2012. **121**, N 4. 609–635.
25. *Goreinov S.A., Tyrtysnikov E.E., Zamarashkin N.L.* A theory of pseudo-skeleton approximations // Linear Algebra and Its Applications. 1997. **261**. 1–21.
26. *Goreinov S.A., Oseledets I.V., Savostyanov D.V., Tyrtysnikov E.E., Zamarashkin N.L.* How to find a good submatrix. ICM Research Report 08-10. Hong Kong: Kowloon Tong, 2008.
27. *Behnel S., Bradshaw R., Citro C., Dalcin L., Seljebotn D.S., Smith K.* Cython: the best of both worlds // Comput. Sci. Eng. 2011. **13**, N 2. 31–39.
28. Intel Math Kernel Library (<http://software.intel.com/en-us/articles/intel-mkl/>).
29. *Saad Y., Schultz M.H.* GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems // SIAM J. Sci. and Stat. Comput. 1986. **7**, N 3. 856–869.
30. *Крылов А.Н.* О численном решении уравнения, которым в технических вопросах определяются частоты малых колебаний материальных систем // Изв. АН СССР. VII серия. 1931. № 4. 491–539.
31. *Romanov A.N., Jabin S.N., Martynov Ya.B., Sulimov A.V., Grigoriev F.V., Sulimov V.B.* Surface generalized Born method: a simple, fast, and precise implicit solvent model beyond the Coulomb approximation // J. Phys. Chem. A. 2004. **108**, N 43. 9323–9327.
32. *Halgren T.A.* Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94 // J. Comput. Chem. 1996. **17**. 490–519.
33. Protein Data Bank (<http://www.pdb.org/pdb/home/home.do>).
34. Avogadro — free cross-platform molecule editor (http://avogadro.openmolecules.net/wiki/Main_Page).
35. *Tyrtysnikov E.E.* Mosaic-skeleton approximations // Calcolo. 1996. **33**, N 1. 47–57.
36. *Hackbusch W.* A sparse matrix arithmetic based on \mathcal{H} -matrices. Part I: introduction to \mathcal{H} -matrices // Computing. 1999. **62**, N 2. 89–108.
37. *Hackbusch W., Khoromskij B.N.* A sparse \mathcal{H} -matrix arithmetic. Part II: application to multi-dimensional problems // Computing. 2000. **64**, N 1. 21–47.

Поступила в редакцию
09.12.2013

Application of the Multicharge Approximation for Large Dense Matrices in the Framework of the Polarized Continuum Solvent Model

A. Yu. Mikhalev¹, I. V. Oferkin², I. V. Oseledets³, A. V. Sulimov⁴,
E. E. Tyrtysnikov⁵, and V. B. Sulimov⁶

¹ *Lomonosov Moscow State University, Faculty of Computational Mathematics and Cybernetics; Leninskie Gory, Moscow, 119992, Russia; Graduate Student, e-mail: muxasizhevsk@gmail.com*

² *Dimonta, Ltd, Nagornaya ulitsa 15, stroenie 8, Moscow, 117186, Russia; Programmer, e-mail: io@dimonta.com*

³ *Institute of Numerical Mathematics, Russian Academy of Sciences; ulitsa Gubkina 8, Moscow, 119333, Russia; Ph.D., Senior Scientist, e-mail: ivan.oseledets@gmail.com*

⁴ *Dimonta, Ltd, Nagornaya ulitsa 15, stroenie 8, Moscow, 117186, Russia; Deputy Director on Information Technologies, e-mail: sulimov@mail.ru*

⁵ *Institute of Numerical Mathematics, Russian Academy of Sciences; ulitsa Gubkina 8, Moscow, 119333, Russia; Professor, Director, e-mail: eugene.tyrtysnikov@gmail.com*

⁶ *Research Computing Center, Lomonosov Moscow State University; Leninskie Gory, Moscow, 119992, Russia; Ph.D., Head of Laboratory, e-mail: vladimir.sulimov@gmail.com*

Received December 9, 2013

Abstract: The polar part of the molecular interaction energy with a water solvent and its computation are considered in the frame of the PCM continuum implicit solvent model (Polarized Continuum Model). A new algorithm of solving the polarized charge equation is proposed. This algorithm is more than two orders of magnitude faster than the regular iterative methods without any critical loss of accuracy. The algorithm is based on the multicharge method for approximating large dense matrices generated by discretizing the molecular Solvent Excluded Surface (SES) and the polarized charges on it. The algorithm is implemented in the MCBHSOLV program written on the Python programming language. Numerical results obtained with the

MCBHSOLV program and its performance are compared with those of the DISOLV program. The comparison is based on computing the polar part of the solvation energy of different proteins, ligands, and protein-ligand complexes as well as on computing the desolvation energy being an important part of the protein-ligand interaction energy. Both the programs used the same test data: molecules, surfaces, discretizations, and atomic charges. The influence of the discretization step size on the molecule solvation and desolvation energies is examined. It is shown that the computations of the desolvation energy with an error less than 1 kcal/mol can be performed with a discretization step of 0.2 Å, which requires only several minutes per one set of protein, ligand, and protein-ligand complexes.

Keywords: implicit solvent model, polarized charges, triangulation network, molecular surface, Solvent Excluded Surface (SES), solvation energy, desolvation energy, protein–ligand binding energy, solvent, multicharge method.

References

1. V. A. Sadovnichii and V. B. Sulimov, “Supercomputing Technologies in Medicine,” in *Supercomputing Technologies in Science, Education, and Industry*, Ed. by Vl. V. Voevodin, V. A. Sadovnichii, and G. I. Savin (Mosk. Gos. Univ., Moscow, 2009), pp. 16–23.
2. 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).
3. 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).
4. 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).
5. J. Tomasi and M. Persico, “Molecular Interactions in Solution: An Overview of Method Based on Continuous Distributions of the Solvent,” *Chem. Rev.* **94** (7), 2027–2094 (1994).
6. C. Cramer and D. Truhlar, “Implicit Solvation Models: Equilibria, Structure, Spectra, and Dynamics,” *Chem. Rev.* **99** (8), 2161–2200 (1999).
7. A. Onufriev, “Continuum Electrostatics Solvent Modeling with the Generalized Born Model,” in *Modeling Solvent Environments: Applications to Simulations of Biomolecules* (Wiley, Weinheim, 2010), pp. 127–166.
8. A. J. Bordner, C. N. Cavasotto, and R. A. Abagyan, “Accurate Transferable Model for Water, *n*-Octanol, and *n*-Hexadecane Solvation Free Energies,” *J. Phys. Chem. B* **106** (42), 11009–11015 (2002).
9. 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).
10. C. S. Pomelli and J. Tomasi, “A New Formulation of the PCM Solvation Method: PCM-QINTn,” *Theor. Chem. Acc.* **96** (1), 39–43 (1997).
11. A. Yu. Mikhalev and I. V. Oseledets, *Adaptive Nested Cross Approximation of Non-Local Operators*, arXiv preprint: 1309.1773v1 [math.NA] (Cornell Univ. Library, Ithaca, 2013), available at <http://arxiv.org/abs/1309.1773/>.
12. M. Totrov and R. Abagyan, “Rapid Boundary Element Solvation Electrostatics Calculations in Folding Simulations: Successful Folding of a 23-Residue Peptide,” *Peptide Sci.* **60** (2), 124–133 (2001).
13. S. N. Zhabin and V. B. Sulimov, RF Patent No. 2 006 613 753 (27 October 2006).
14. S. N. Zhabin and V. B. Sulimov, “Software for Constructing the Surface-Accessible Surfaces for Arbitrary Organic Molecules and for Positioning the Ligands in the Active Centers of Proteins,” in *Proc. 13th Rus. Intl. Conf. on Humans and Medicine, Moscow, April 3–7, 2006* (Ross. Akad. Gos. Sluzhby, Moscow, 2006), p. 15.
15. S. N. Zhabin and V. B. Sulimov, “Construction of Smooth Molecular Surfaces with Adaptive Triangulation: the TAGSS Program,” *Nauch. Vizualiz.* **3** (2), 27–53 (2011).
16. M. L. Connoly, “Solvent-Accessible Surfaces of Proteins and Nucleic Acids,” *Science* **221** (4612), 709–713 (1983).
17. Y. N. Vorobjev and J. Hermans, “SIMS: Computation of a Smooth Invariant Molecular Surface,” *Biophys. J.* **73**, 722–732 (1997).
18. A. A. Samarskii, *Introduction to Numerical Methods* (Lan’, St. Petersburg, 2005) [in Russian].
19. W. Hackbusch, B. N. Khoromskij, and S. A. Sauter, “On \mathcal{H}^2 -Matrices,” in *Lectures on Applied Mathematics* (Springer, Heidelberg, 2000), pp. 2–29.
20. S. Börm, *Efficient Numerical Methods for Non-Local Operators: \mathcal{H}^2 -Matrix Compression, Algorithms, and Analysis* (Eur. Math. Soc., Zürich, 2010).
21. V. V. Voevodin, “On a Method of Reducing the Matrix Order while Solving Integral Equations,” in *Numerical Analysis on Fortran* (Mosk. Gos. Univ., Moscow, 1979), pp. 21–26.

22. L. Greengard and V. Rokhlin, "A Fast Algorithm for Particle Simulations," *J. Comput. Phys.* **73** (2), 325–348 (1987).
23. L. Ying, G. Biros, and D. Zorin, "A Kernel-Independent Adaptive Fast Multipole Algorithm in Two and Three Dimensions," *J. Comput. Phys.* **196** (2), 591–626 (2004).
24. M. Bebendorf and R. Venn, "Constructing Nested Bases Approximations from the Entries of Non-Local Operators," *Numer. Math.* **121** (4), 609–635 (2012).
25. S. A. Goreinov, E. E. Tyrtyshnikov, and N. L. Zamarashkin, "A Theory of Pseudo-Skeleton Approximations," *Linear Algebra and Its Appl.* **261**, 1–21 (1997).
26. S. A. Goreinov, I. V. Oseledets, D. V. Savostyanov, E. E. Tyrtyshnikov, and N. L. Zamarashkin, *How to Find a Good Submatrix*, Research Report 08-10 (Kowloon Tong, Hong Kong, 2008).
27. S. Behnel, R. Bradshaw, C. Citro, et al., "Cython: the Best of Both Worlds," *Comput. Sci. Eng.* **13** (2), 31–39 (2011).
28. Intel Math Kernel Library, available at <http://software.intel.com/en-us/articles/intel-mkl/>.
29. Y. Saad and M. H. Schultz, "GMRES: a Generalized Minimal Residual Algorithm for Solving Nonsymmetric Linear Systems," *SIAM J. Sci. Stat. Comput.* **7** (3), 856–869 (1986).
30. A. N. Krylov, "Numerical Solution of the Equation to Determine the Frequency of Small Oscillations in Material Systems," *Izv. Akad. Nauk SSSR, Mat. Ser.*, No 4, 491–539 (1931).
31. A. N. Romanov, S. N. Zhabin, Ya. B. Martynov, et al., "Surface Generalized Born Method: a Simple, Fast, and Precise Implicit Solvent Model Beyond the Coulomb Approximation," *J. Phys. Chem. A* **108** (43), 9323–9327 (2004).
32. T. A. Halgren, "Merck Molecular Force Field. I. Basis, Form, Scope, Parameterization, and Performance of MMFF94," *J. Comput. Chem.* **17**, 490–519 (1996).
33. Protein Data Bank, available at <http://www.pdb.org/pdb/home/home.do/>.
34. Avogadro : Free Cross-Platform Molecule Editor, available at http://avogadro.openmolecules.net/wiki/Main_Page/.
35. E. E. Tyrtyshnikov, "Mosaic-Skeleton Approximations," *Calcolo* **33** (1), 47–57 (1996).
36. W. Hackbusch, "A Sparse Matrix Arithmetic Based on \mathcal{H} -Matrices. Part I: Introduction to \mathcal{H} -Matrices," *Computing* **62** (2), 89–108 (1999).
37. W. Hackbusch and B. N. Khoromskij, "A Sparse \mathcal{H} -Matrix Arithmetic. Part II: Application to Multi-Dimensional Problems," *Computing* **64** (1), 21–47 (2000).