

Метод атом–атомных корреляционных функций в приближении RISM для изучения гидратации макромолекул.

Д.А. Тихонов¹ Р.В. Полозов,¹ А.В. Горелов,² К.А. Доусон²
Э.Г. Тимошенко,² Ю.А. Кузнецов,² Л.А. Панченко,³

Аннотация

Предложен численный алгоритм решения интегральных уравнений теории жидкости в приближении RISM для бесконечно разбавленных растворов молекул с большим числом атомов. Алгоритм основан на применении нестационарных итерационных методов решения линейных систем. С использованием этого алгоритма вычислены атом–атомные корреляционные функции фрагментов дуплексов ДНК различной длины в воде.

1 Введение.

Для изучения сольватации многоатомных систем наряду с методами Монте–Карло и молекулярной динамики используют метод интегральных уравнений статистической теории жидкости. Вариантом метода интегральных уравнений, разработанным для изучения сольватации многоатомных молекул, является приближение RISM (Reference Interaction Site Model) [1]. Достаточно полный обзор метода RISM, и результатов, полученных на его основе приведен в работе [2]. Основным достоинством метода RISM по сравнению с методами имитационного моделирования являются значительно меньшие вычислительные затраты. Примером применения RISM для расчета гидратации биомолекул, является расчет аланинового дипептида, проведенного в работе [3]. В исследовании более сложных макромолекул методом RISM часто используются грубые модели структуры молекул, поскольку учет корреляционных функций всех атомов в молекуле приводит к необходимости поиска численного решения системы интегральных уравнений порядка числа атомов в макромолекуле. В работах [4, 5, 6] выполнены исследования сольватации модельных структур ДНК, в которых нуклеотидные пары аппроксимировались еденичными силовыми центрами, в качестве растворителя выбиралась вода и бинарная смесь заряженных частиц (примитивная модель электролита). При этом, вопрос о том, к каким погрешностям приводит огрубление молекулярной структуры остается открытым.

Грубые модели гидратации представляют в основном чисто теоретический интерес, на их основе можно рассмотреть неспецифические, не связанные с детальной молекулярной структурой, но общие данному классу биомолекул, эффекты гидратации. В работе [7] изучалась гидратация линейной неполярной цепочки в водном растворителе с целью выявления общих закономерностей гидратации протяженных молекулярных структур. Были получены некоторые сведения о локальном поведении термодинамических и структурных величин вдоль молекулярной цепи, в частности, было изучено влияние концевых атомов цепочек на локальные характеристики гидратации. Однако модели молекулярной цепочки, состоящей из одинаковых атомов, также как и используемая в [5] модель той - же цепочки, закрученной в спираль, достаточно далеки от реальной молекулярной структуры ДНК. Поэтому возникает вопрос об учете всех атомов, составляющих фрагмент ДНК, поскольку заранее неизвестно вкладом каких атомов в гидратацию можно пренебречь.

В такой постановке задачи мы сталкиваемся с чисто техническими трудностями при решении системы интегральных уравнений, поскольку полный учет всех корреляций атомов фрагмента дуплекса ДНК с атомами воды приводит к системе порядка нескольких сотен нелинейных интегральных уравнений.

В последнее время получили развитие нестационарные итерационные методы решения больших систем линейных уравнений. Обзор этих методов приведен в работе [8]. Применение итерационных методов в комбинации с методом Ньютона–Рафсона позволяет решать достаточно большие систе-

мы нелинейных алгебраических уравнений, к которым сводится система интегральных уравнений RISM.

Мы развили новый эффективный численный алгоритм решения интегральных уравнений теории жидкости в приближении RISM, который позволяет рассчитывать атом–атомные корреляционные функции больших фрагментов двойной спирали ДНК любой последовательности нуклеотидов с водой, в том числе он применим и к функционально важным участкам ДНК. Атом–атомные корреляционные функции являются исходными для вычисления термодинамических и структурных величин сольватации.

Целью настоящей работы является описание алгоритма, позволяющего рассчитывать атом–атомные корреляционные функции макромолекул в водном растворе. Работа организована следующим образом: приведены общие уравнения RISM для атом–атомных корреляционных функций, обсуждены численные схемы решения этих уравнений и описан вычислительный алгоритм. Также приведена статистика расчетов молекулярных систем с различным числом атомов, выполненных по предложенному алгоритму.

2 Основные уравнения.

Основой современной статистической теории жидкостей являются интегральные уравнения для корреляционных функций, при помощи которых можно вычислить термодинамические величины различных физических систем, от простых одноатомных жидкостей до сложных многоатомных многокомпонентных растворов. Все интегральные уравнения теории жидкостей являются по сути приближенными, поскольку их вывод основан на формализме бесконечной цепочки интегродифференциальных уравнений Боголюбова–Борна–Грина–Кирквуда–Ивона (ББГКИ). Собственно способ обрыва и замыкания бесконечной цепочки ББГКИ порождает те или иные приближенные уравнения. Каждое интегральное уравнение имеет свою область применимости, адекватность каждого уравнения может быть проверена путем сравнения рассчитанных корреляционных функций с экспериментально полученными в опытах по рассеянию рентгеновских лучей на атомах жидкости.

Наибольшего прогресса теория жидкости достигла при исследовании простых одноатомных жидкостей [9, 10]. Основным уравнением здесь является уравнение Орнштейна–Цернике, которое связывает прямую и полную корреляционные функции в фурье пространстве

$$(1 - \rho \tilde{C}(k))(1 + \rho \tilde{h}(k)) = 1, \quad (1)$$

где ρ — плотность жидкости, функции $\tilde{C}(k)$ и $\tilde{h}(k)$ — прямая и полная корреляционные функции в фурье-представлении, соответственно. Здесь и далее знак "тильда" будет указывать на фурье-представление. Выпишем формулы преобразования Фурье, осуществляющие переход между координатным и k -пространствами для произвольной функции x :

$$\begin{aligned} \tilde{x}(k) = \mathcal{F}^f x(r) &= \frac{4\pi}{k} \int_0^\infty x(r) \sin(kr) r dr, \\ x(r) = \mathcal{F}^b \tilde{x}(k) &= \frac{1}{2\pi^2 r} \int_0^\infty \tilde{x}(k) \sin(kr) k dk. \end{aligned}$$

Полная корреляционная функция дает информацию о локальной микроскопической структуре жидкости, а именно, величина $h(r) + 1$ является условной плотностью вероятности обнаружения произвольной частицы на конце радиус–вектора длины r при условии, что в начале координат находится другая частица. Корреляции между частицами жидкости обусловлены двумя причинами. Во первых наличием потенциала взаимодействия между частицами. Корреляции, обусловленные потенциалом взаимодействия частиц $u(r)$, называют прямыми корреляциями, такие корреляции отражены в прямой корреляционной функции $C(r)$. Во вторых, поскольку суммарный исключенный объем жидкости сравним со свободным объемом, то на корреляцию пары частиц влияют соседние частицы жидкости. Вклад соседних частиц в корреляцию пары отражается в непрямой корреляционной функции $\gamma(r) = h(r) - C(r)$. По сути уравнение (1) является формально строгим соотношением между прямой и полной корреляционной функциями, однако оно не может быть решено без знания дополнительного уравнения, связывающего прямую и полную корреляционные функции. Такое уравнение называется уравнением замыкания.

Общая форма уравнения замыкания имеет вид:

$$C(r) = \exp(-u(r)/k_B T + \gamma(r) + \mathcal{B}[\gamma(r)]) - \gamma(r) - 1, \quad (2)$$

где $\mathcal{B}[\gamma(r)]$ так называемый мостиковый функционал, k_B — постоянная Больцмана, T — абсолютная температура. До настоящего времени не найдено точного выражения для мостикового функционала, однако существует множество приближенных выражений, применимость которых ограничена определенным классом потенциалов взаимодействия $u(r)$. Мы приведем два наиболее известных выражения:

$$\mathcal{B}[\gamma(r)] = 0, \quad (3)$$

$$\mathcal{B}[\gamma(r)] = \ln(1 + \gamma(r)) - \gamma(r). \quad (4)$$

Выражение (3) соответствует так называемому гиперцепному приближению, выражение (4) — приближению Перкуса–Йевики. Известно, что (3) хорошо работает для дальнедействующих потенциалов, а (4) — для короткодействующих.

К настоящему времени проведено большое число расчетов термодинамических и структурных величин простых жидкостей на основе уравнения Орнштейна–Цернике, причем результаты расчетов находятся в хорошем согласии с имеющимися экспериментальными данными.

Обобщение теории на случай одноатомных многокомпонентных систем проводится естественным образом. Для этого в уравнении (1) достаточно считать, что полная и прямая корреляционные функции являются матрицами \hat{h} , \hat{C} размерности равной числу различных компонент смеси, элементами этих матриц являются корреляционные функции соответствующих одноатомных компонент. Плотность ρ при этом становится диагональной матрицей, $\hat{\rho}$, элементами которой являются плотности соответствующих компонент. Отметим, что уравнение для многокомпонентных одноатомных систем также является формально строгим уравнением.

Идея метода RISM основана на введении внутримолекулярных корреляционных функций в уравнение Орнштейна–Цернике для многокомпонентных одноатомных систем. При этом многокомпонентная смесь разбивается на группы жестко связанных между собой атомов. Корреляционные функции между атомами i и j внутри группы атомов считаются заданными, и равными $v_{ij}(r) = \delta(r - r_{ij})/4\pi r^2$, где r_{ij} — межатомные расстояния. В фурье-пространстве уравнения RISM записываются в матричном виде:

$$\left(\hat{v}(k) + \hat{\rho}\hat{h}(k) \right) \left(\hat{v}(k)^{-1} - \hat{\rho}\hat{c}(k) \right) = \hat{I}, \quad (5)$$

здесь \hat{v} — блочно диагональная матрица с числом блоков равным числу групп атомов (молекул). Элементами каждого блока являются внутримолекулярные корреляционные функции $v_{ij}(k) = \sin(kr_{ij})/kr_{ij}$. Диагональная матрица $\hat{\rho}$ содержит парциальные плотности атомов смеси, она имеет такую же блочную структуру как и \hat{v} , каждый блок матрицы $\hat{\rho}$ может быть записан в виде произведения скалярной величины плотности молекулярной компоненты на еденичную матрицу размерности равной числу атомов в молекуле. Атом–атомные корреляционные функции $\tilde{c}_{ij}(k)$, $\tilde{h}_{ij}(k)$ являются элементами заполненных матриц \hat{c} и \hat{h} , матрица \hat{I} — еденичная матрица, размерность которой равна полному числу атомов молекулярной смеси.

Рассмотрим двухкомпонентный молекулярный раствор. Плотность одной компоненты (растворителя) положим ρ , плотность другой компоненты (раствора) пусть будет ρ_u . В пределе бесконечного разбавления : $\rho_u/\rho \rightarrow 0$ из уравнения (5) можно выделить два матричных уравнения :

$$\hat{\gamma}^v = \hat{v}^v \hat{c}^v \hat{v}^v \left(\hat{I} - \rho \hat{c}^v \hat{v}^v \right)^{-1} - \hat{c}^v(k) \quad (6)$$

$$\hat{\gamma}^{uv} = \hat{v}^u \hat{c}^{uv} \left(\hat{v}^v + \rho \hat{h}^v \right) - \hat{c}^{uv} \quad (7)$$

Индекс v относится к компонентам растворителя, индекс u — к компонентам растворенного вещества. Систему уравнений (6),(7) можно решить последовательно. На первом этапе решается первое уравнение по отношению к γ^v . Далее решения γ^v используются как входные в уравнение (7), которое фактически является линейным относительно γ^{uv} .

Как уже отмечалось, уравнения RISM являются приближенными. Основное допущение состоит в том, что описание молекулярных корреляций сводится к набору атом–атомных сферически

симметричных корреляционных функций. Хорошо известны недостатки RISM приближения [2, 11], однако вместе с соответствующими уравнениями замыкания она является едва ли не единственным инструментом теоретического исследования сложных молекулярных систем в растворах.

Нетрудно видеть, что уравнение (6) для корреляционных функций чистого растворителя совпадает с уравнением (5). Число атомов молекулы растворителя обычно невелико, поэтому решить уравнение (6) не составляет труда. Основное внимание будет сосредоточено на решении уравнения (7).

Мы будем изучать гидратацию большой молекулярной системы и намерены дать практически пригодный алгоритм решения уравнений RISM для атом–атомных корреляционных функций молекулярной системы, состоящей из нескольких сотен атомов в водном растворителе. С этой целью, чтобы не загромождать изложение несущественными обозначениями, перепишем уравнение (7) для атом–атомных корреляционных функций всех атомов N -атомной молекулярной системы с атомами О и Н молекулы воды в следующем виде:

$$G = (\hat{V} \otimes \hat{W} - \hat{I}_{2N}) C, \quad (8)$$

здесь G, C — вектор–столбцы размерности $2N$, элементами которых являются атом–атомные корреляционные функции:

$$G = \begin{bmatrix} \tilde{\gamma}_{1H}(k) \\ \vdots \\ \tilde{\gamma}_{NH}(k) \\ \tilde{\gamma}_{1O}(k) \\ \vdots \\ \tilde{\gamma}_{NO}(k) \end{bmatrix}, \quad C = \begin{bmatrix} \tilde{c}_{1H}(k) \\ \vdots \\ \tilde{c}_{NH}(k) \\ \tilde{c}_{1O}(k) \\ \vdots \\ \tilde{c}_{NO}(k) \end{bmatrix}.$$

Первый индекс у корреляционной функции является порядковым номером атома в растворенной молекуле, второй индекс обозначает тип атома (О или Н) в молекуле воды. Матрица \hat{W} размерности $N \times N$ — внутримолекулярная структурная матрица растворенной молекулы, элементы которой являются функциями и выражаются через межатомные расстояния r_{ij} следующим образом:

$$W_{ij} = \frac{\sin(kr_{ij})}{kr_{ij}}$$

Матрица \hat{V} размерности 2×2 , отражает геометрическую и корреляционную структуры водного растворителя.

$$\hat{V} = \begin{bmatrix} v_1 + v_2 & v_3 \\ 2v_3 & v_4 \end{bmatrix},$$

где

$$\begin{aligned} v_1 &= 1 + \rho \tilde{h}_{HH}(k), & v_2 &= \sin(kr_{HH})/kr_{HH} + \rho \tilde{h}_{HH}(k), \\ v_3 &= \sin(kr_{OH})/kr_{OH} + \rho \tilde{h}_{OH}(k), & v_4 &= 1 + \rho \tilde{h}_{OO}(k). \end{aligned}$$

Здесь r_{HH}, r_{OH} — межатомные расстояния в молекуле воды, ρ — плотность воды, $\tilde{h}_{HH}, \tilde{h}_{OH}, \tilde{h}_{OO}$ — атом–атомные полные корреляционные функции водного растворителя. Символом " \otimes " в уравнении (8) обозначается прямое произведение матриц, матрица \hat{I}_{2N} является еденичной матрицей размерности $2N$.

Вторым уравнением, выражающим c_{iX} через γ_{iX} , является уравнение замыкания, в которое входит потенциал атом–атомного взаимодействия. В том случае, когда потенциал взаимодействия включает кулоновскую часть наиболее употребительным является гиперцепное замыкание (см. (2), (3)). С использованием перенормировки кулоновского потенциала предложенной Энгом [12], уравнение замыкания имеет вид:

$$c_{iX}^s(r) = \exp(-U_{iX}(r)/k_B T + f_{iX}(r) + \gamma_{iX}^s(r)) - \gamma_{iX}^s(r) - 1, \quad (9)$$

где $U_{iX}(r)$ — некулоновская часть атом–атомного потенциала; индекс i — порядковый номер атома в молекуле раствора; индекс X обозначает атомы О или Н в молекуле воды. Функция f_{iX} является перенормированным кулоновским потенциалом, который выражается через заряды q_i и q_X :

$$f_{iX}(r) = -q_i q_X (1 - \text{erf}(r)) / r k_B T.$$

Индекс s обозначает короткодействующую часть корреляционной функции. Корреляционные функции, входящие в (8), выражаются через короткодействующие функции и перенормированный кулоновский потенциал следующим образом:

$$\tilde{\gamma}_{iX}(k) = \tilde{\gamma}_{iX}^s(k) - \tilde{f}_{iX}(k), \quad \tilde{c}_{iX}(k) = \tilde{c}_{iX}^s(k) + \tilde{f}_{iX}(k).$$

Для полноты приведем аналитическое выражение перенормированного кулоновского потенциала в k -пространстве:

$$\tilde{f}_{iX}(k) = -4\pi q_i q_X \exp(-k^2/4) / k k_B T.$$

За недостатком места, мы не обсуждаем здесь технику перенормировок кулоновского потенциала, отсылая читателя к работе [2]. Достаточно отметить, что введение короткодействующих корреляционных функций, позволяет легко переходить от координатного к k -представлению при помощи численного преобразования Фурье.

3 Численная схема.

В этом разделе мы предлагаем способ численного решения системы $2N$ интегральных уравнений, образованной уравнениями (8) и (9), для набора неизвестных функций $\tilde{\gamma}$, для случая, когда N достаточно велико. При этом мы полагаем заданными плотность растворителя ρ , температуру $k_B T$ а также матрицы \hat{W} , \hat{V} и потенциалы взаимодействия молекулы с растворителем, которые входят в уравнение замыкания.

Основная идея, лежащая в основе численного решения системы нелинейных интегральных уравнений, состоит в сведении, путем выбора дискретной сетки, системы интегральных уравнений к системе нелинейных алгебраических уравнений. Действительно, выберем L узлов, равномерно распределенных в r - и k -пространствах:

$$\begin{aligned} \Delta r \Delta k &= \frac{\pi}{L} \\ r_i &= i \Delta r \quad k_j = j \Delta k \\ i, j &= 1 \dots L. \end{aligned} \tag{10}$$

Пологаем неизвестным $2NL$ -компонентный вектор столбец \vec{X} , аппроксимирующий набор $\tilde{\gamma}_{iX}^s(k)$ функций в $j\Delta k$ узлах:

$$\vec{X} = \begin{bmatrix} \tilde{\gamma}_{1H}^s(\Delta k) \\ \vdots \\ \tilde{\gamma}_{1H}^s(L\Delta k) \\ \vdots \\ \tilde{\gamma}_{NO}^s(L\Delta k) \end{bmatrix}.$$

Вводя конечномерные аппроксимации Фурье–операторов \mathcal{F}^f , \mathcal{F}^b :

$$\hat{F}^f = \frac{4\pi(\Delta r)^2}{\Delta k} \hat{S}, \quad \hat{F}^b = \frac{(\Delta k)^2}{2\pi^2 \Delta r} \hat{S},$$

получаем систему из $2NL$ нелинейных алгебраических уравнений:

$$\vec{X} - \vec{Z}[\vec{X}] \equiv \vec{X} - \hat{A}(\hat{I}_{2N} \otimes \hat{F}^f) \vec{C}[(\hat{I}_{2N} \otimes \hat{F}^b) \vec{X}] - \vec{Q} = 0, \tag{11}$$

где \hat{A} — матрица размерности $2NL \times 2NL$, состоящая из диагональных блоков размерности $L \times L$, являющихся дискретным представлением функциональных элементов матрицы $\hat{V} \otimes \hat{W} - \hat{I}_{2N}$, \hat{S} — матрица размерности $L \times L$ с компонентами:

$$S_{ij} = \frac{j}{i} \sin\left(\frac{\pi}{L} ij\right), \quad i, j = 1, \dots, L.$$

Вектор \vec{Q} аналогичный по структуре вектору \vec{X} , получается при дискретизации $2N$ -компонентного функционального вектора Q , даваемого выражением:

$$Q = (\hat{V} \otimes \hat{W}) \begin{bmatrix} \tilde{f}_{1H}(k) \\ \vdots \\ \tilde{f}_{NO}(k) \end{bmatrix}.$$

Вектор-функция векторного аргумента $\vec{C}[\vec{Y}]$, которая возникает при дискретизации уравнения замыкания (9), дается в соответствии со следующим определением:

$$C_l \equiv C_l[Y_l] = M_l \exp(Y_l) - Y_l - 1, \quad l = 1, \dots, 2NL, \quad (12)$$

где M_l — компоненты $2NL$ -мерного вектора \vec{M} , который получается при дискретизации в $i\Delta r$ -узлах функционального вектора:

$$\begin{bmatrix} \exp(f_{1H}(r) - U_{1H}(r)) \\ \vdots \\ \exp(f_{NO}(r) - U_{NO}(r)) \end{bmatrix}.$$

Отметим, что вектор-функция $\vec{C}[\vec{Y}]$ является диагональной, в том смысле, что ее l -компонента зависит только от такой же компоненты вектора-аргумента, правда нелинейным образом. Этим она отличается от вектор-функции $\vec{Z}[\vec{X}]$, которая является недиагональной и нелинейной функцией векторного аргумента.

Простейшим способом решения нелинейной системы (11) является метод прямых итераций или метод Пиккарда. Выбирая некоторое нулевое приближение $\vec{X}^{(0)}$, дальнейшее уточнение решения находим по рекуррентной формуле:

$$\vec{X}^{(n+1)} = \vec{Z}[\vec{X}^{(n)}], \quad n = 0, 1, \dots \quad (13)$$

Если при этом, начиная с некоторого n , норма разности $\vec{X}^{(n+1)} - \vec{X}^{(n)}$ оказывается меньше некоторой наперед заданной величины, то мы полагаем, что решение найдено. Часто применяют некоторые модификации метода прямых итераций, используя несколько предыдущих итерационных шагов для обеспечения более быстрой сходимости. Удачным примером такой модификации, основанной на векторной экстраполяции, служит работа [19]. Для решения уравнений RISM прямые итерации использованы в работе [13], где для улучшения сходимости применялся принцип монотонности невязки. Однако на практике, к сожалению, сходимость прямых итераций для произвольных потенциалов и зарядов оказывается скорее исключением, чем правилом.

Значительно более эффективным способом решения системы (11) является метод Ньютона-Рафсона, в котором итерации проводятся по следующей схеме:

$$\vec{X}^{(n+1)} = \vec{X}^{(n)} + \Delta \vec{X}^{(n)}, \quad (14)$$

$$\Delta \vec{X}^{(n)} = \hat{J}^{-1} (\vec{Z}[\vec{X}^n] - \vec{X}^n), \quad (15)$$

где \hat{J} — матрица Якоби, которая получается дифференцированием уравнения (11):

$$\hat{J} \equiv \hat{I}_{2NL} - \frac{d\vec{Z}}{d\vec{X}} \equiv \hat{I}_{2NL} - (2/L) \hat{A} (\hat{I}_{2N} \otimes \hat{S}) \hat{D} (\hat{I}_{2N} \otimes \hat{S}). \quad (16)$$

Компоненты диагональной матрицы \hat{D} , даются следующим выражением:

$$D_{ll} = M_l \exp(Y_l) - 1, \quad l = 1, \dots, 2NL.$$

Узким местом в методе Ньютона–Рафсона является то, что на каждом итерационном шаге необходимо решать систему линейных уравнений (15) порядка $2NL$. Порядок этой системы чрезвычайно большой, поскольку даже в случае небольшого числа атомов N , число узлов L , необходимое для хорошей дискретизации функций должно быть порядка 1000. Выход из этой ситуации находят в комбинировании метода Пиккарда с методом Ньютона–Рафсона таким образом, чтобы понизить порядок линейной системы. Более точно, для решения уравнения (1) использовалась "грубая" аппроксимация системы нелинейных алгебраических уравнений системой нелинейных алгебраических уравнений меньшего порядка с использованием предположения о гладкости решения, которая может быть решена методом Ньютона–Рафсона с последующей "очисткой" грубого решения пиккардовскими итерациями [16]. Далее было замечено, что достаточно решать методом Ньютона–Рафсона лишь первые m уравнений, решение остальных уравнений достигается применением прямых итераций [17]. В работе [15], для решения (1) используется метод Ньютона–Рафсона без округления, при этом для решения системы линейных уравнений большого порядка использовался итерационный метод сопряженных градиентов.

При конечномерной дискретизации система интегральных уравнений (8) отличается от уравнения (1) только порядком системы получающихся нелинейных алгебраических уравнений. Поэтому были разработаны численные методы решения (8), на основе методов решения (1) [2, 18]. Однако эти алгоритмы становятся практически непригодными при использовании метода исключения Гаусса решения системы линейных уравнений для больших N , поскольку с ростом числа атомов вычислительные затраты растут как N^3 .

Мы предлагаем использовать для решения линейной системы нестационарные итерационные методы. Эти методы решения линейных систем уравнений в основе своей являются разновидностью метода сопряженных градиентов. Главной особенностью этих методов является лучшая сходимость для более широкого класса матриц. Основным достоинством итерационных методов по сравнению с методом исключения Гаусса является то, что они не требуют размещения матрицы системы в памяти, для нахождения решения достаточно обеспечить процедуру матрично–векторного умножения для произвольного вектора \vec{x} :

$$\vec{y} = \hat{J}\vec{x}. \quad (17)$$

Если мы выберем число узлов L равным целой степени двух, то для вычисления (17) можно воспользоваться алгоритмом быстрого преобразования Фурье, и несложно видеть, что такая процедура потребует LN^2 операций с плавающей точкой при больших N . Дополнительные затраты при осуществлении одной итерации нестационарного метода пропорциональны NL . Так, например, в методе $Bi - CGSTAB$ на одну итерацию требуется два вычисления (17) и дополнительно $48NL$ операций с плавающей точкой. При этом требования к памяти ограничиваются $20NL$ ячейками.

Несмотря на то, что сходимость итераций Ньютона–Рафсона чрезвычайно быстрая, на практике затраты на решение линейных уравнений с матрицей Якоби размерности $2NL$, даже при помощи итерационной процедуры все же неоправданно высоки. Мы предлагаем способ понижения порядка решаемой линейной системы в m раз без существенного ухудшения общей сходимости. Идея состоит в использовании итераций Пиккарда и Ньютона–Рафсона одновременно на каждом итерационном шаге алгоритма.

Выберем некоторое подмножество узлов сетки в координатном и k -пространстве:

$$r'_i = i m \Delta r, \quad k'_j = j \Delta k; \quad i, j = 1 \dots \frac{L}{m}. \quad (18)$$

Будем обозначать матрицу Якоби, рассчитанную на данном подмножестве, как \hat{J}_m . Размерность этой матрицы равна $(2NL/m) \times (2NL/m)$. Обозначим через k'' — дополнительное подмножество компонент. Выпишем формулы для итераций:

$$\vec{R}^{(n)} = \vec{X}^{(n)} - \vec{Z}[\vec{X}^{(n)}], \quad (19)$$

$$\vec{X}^{(n+1)}(k') = \vec{X}^{(n)}(k') - \hat{J}_m^{-1} \vec{R}^{(n)}(k'), \quad (20)$$

$$\vec{X}^{(n+1)}(k'') = \vec{X}^{(n)}(k'') - \vec{R}^{(n)}(k''). \quad (21)$$

Итерации проводятся до тех пор, пока норма вектора $\vec{R}^{(n)}$ не будет меньше некоторого наперед заданного значения.

Особенностью предложенной итерационной схемы является то, что она объединяет две итерационные схемы. В случае $m = L$ итерационная схема будет чисто пиккардовской, а в случае $m = 1$ итерационная схема совпадает со схемой Ньютона–Рафсона. Величину m следует выбирать как целую степень двух, она может меняться в процессе итераций в сторону уменьшения, если сходимость замедляется, или увеличиваться, если сходимость достаточно быстрая. В Приложении приводится схема алгоритма.

4 Результаты расчетов.

В качестве исследуемой системы мы выбрали фрагмент двухспиральной молекулы ДНК, образованный одинаковыми нуклеотидными парами GC. Мы рассчитали фрагменты различной длины (от одной до пяти пар), при этом число атомов N варьировалось от 63 до 315. Геометрия фрагментов выбиралась стандартной [14]. Потенциалы взаимодействия всех пар в данном фрагменте с водой полагались одинаковыми. Некулоновская часть потенциала взаимодействия атомов ДНК с атомами воды, согласно [20] выбиралась в виде:

$$U_{iX}(r) = A_{iX}/r^{12} - B_{iX}/r^6. \quad (22)$$

В том случае, когда атом участвует в образовании водородной связи, выбиралось следующее выражение для потенциала:

$$U_{iX}(r) = A'_{iX}/r^{12} - B'_{iX}/r^{10}. \quad (23)$$

Параметры потенциала взяты из работы [20], заряды на атомах — из работы [21].

В качестве растворителя выбиралась вода при нормальной плотности 1 г/см^3 и температуре 25°C . Корреляционные функции чистой воды рассчитывались по уравнению (6) в гиперцепном замыкании для потенциальной модели TIPS. Параметры модели приведены в работе [22]. Заряд на атоме кислорода воды q_O равен -0.86 , на атомах водорода: $q_H = -q_O/2$.

Межатомные расстояния фрагмента дуплекса ДНК округлялись с точностью до сотой доли ангстрема для понижения числа различных элементов в матрице W , что позволило сильно сократить требования к оперативной памяти. Такой прием, в случае, например, пяти нуклеотидных пар позволяет сократить объем требуемой оперативной памяти до 4% от объема, который потребовался бы для размещения матрицы в случае без округления. При этом, проведенные нами исследования показывают, что корреляционные функции, рассчитанные с округлением и без него, меняются незначительно. Число узлов L выбиралось равным 1024, шаг сетки $\Delta r = 0.025 \text{ \AA}$. Параметр редукции m был выбран равным 8. Порядок линейной системы при этом оказывается равным 80640. Для решения редуцированной линейной системы использовался *TFQMR*-метод [8]. Параметры выхода ϵ , δ были равны 10^{-5} и 10^{-6} , соответственно.

Расчеты проводились по следующей схеме. Вначале была рассчитана одна нуклеотидная пара, затем полученные корреляционные функции использовались в качестве начального приближения при расчете молекулы, состоящей из двух нуклеотидных пар. И так далее, каждая предыдущая молекула использовалась для построения нулевого приближения для следующей. Такой прием позволил сократить число итераций при расчетах молекул с числом нуклеотидных пар отличным от 1. В таблице приведена статистика расчетов фрагментов ДНК. Мы приводим для каждого рассчитанного фрагмента: число атомов N во фрагментах дуплекса, число различных элементов структурной матрицы N^W (для выбранного округления межатомных расстояний), среднее число итераций N^{TFQMR} , потребовавшееся для решения линейной системы, и общее число итераций N^{ITER} , потребовавшихся для расчета данного фрагмента.

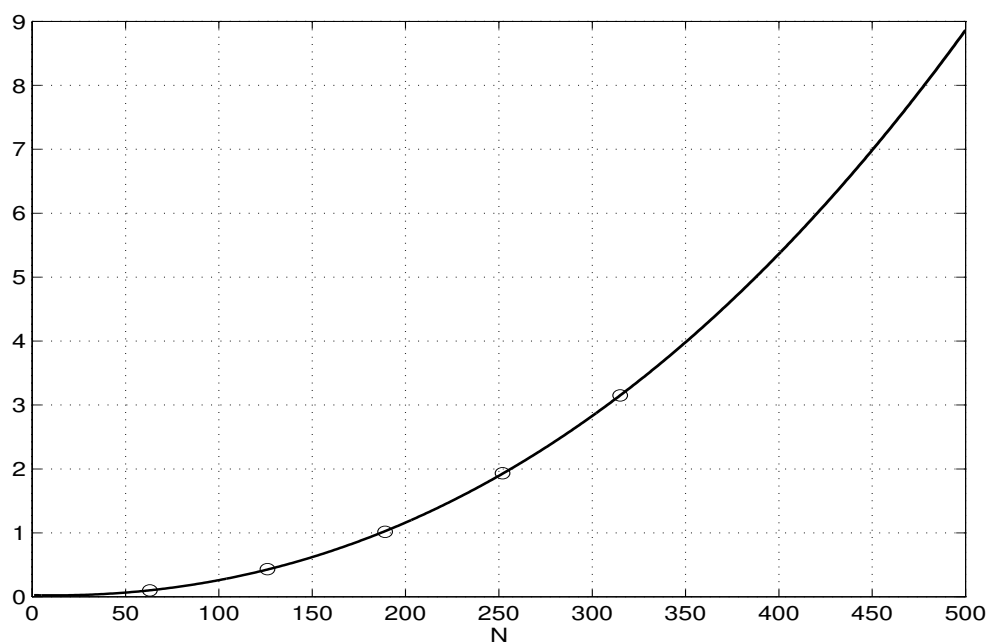
Из данных таблицы, видно, что вычислительные затраты не растут катастрофически с увеличением размера системы. Число итераций остается приемлемо низким для фрагментов различной длины. Время расчета в основном определяется временем необходимым для вычисления матрично-векторного умножения, которое растет приблизительно как квадрат размерности решаемой редуцированной линейной системы, равной $2NL/m$.

Общие затраты времени для осуществления одной итерации метода приведены на рисунке.

Таб. 1: Статистика расчетов гидратации ДНК методом RISM.

N^{GC}	N	N^W	N^{TFQMR}	N^{ITER}
1	63	1038	124	7
2	126	1574	142	8
3	189	1715	149	8
4	252	1798	140	7
5	315	2020	148	7

Рис. 1 Зависимость затрат времени (в часах) на одну итерацию от числа атомов в молекуле N для компьютера Pentium-PRO с тактовой частотой 200 МГц и 64 Мб ОЗУ. Кружками показаны расчеты от одной до пяти нуклеотидных G:C пар, сплошная линия — аппроксимация полиномом третьей степени.



5 Заключение.

Мы построили эффективный алгоритм решения интегральных уравнений теории жидкости для изучения гидратации макромолекул. Предложенный подход позволяет исследовать гидратацию без привлечения упрощающих допущений о молекулярной структуре макромолекул. Алгоритм позволяет исследовать гидратацию длинных (несколько десятков пар нуклеотидов) функционально значимых фрагментов ДНК, например, исследовать гидратацию промоторов и терминаторов генома.

Точный учет детальной структуры макромолекулы позволяет обнаруживать эффекты, которые не проявляются в более грубых моделях. Однако, поскольку мы использовали жесткую модель ДНК, мы не учитывали того, что в водном растворе конформация молекулы может меняться, достигая некоторой оптимальной конформации с минимумом свободной энергии. Несмотря на то, что приближение RISM не допускает учета гибкости молекулы, однако на основе рассчитанных корреляционных функций атомов ДНК - вода, можно вычислить эффективные потенциалы взаимодействия участков ДНК друг с другом в водной среде, так называемые потенциалы средней силы. На основе полученных потенциалов средней силы для каждой конформации молекулы можно решить задачу об оптимальной конформации ДНК с привлечением метода Монте-Карло. Достоинством такого подхода, применявшегося в работе [13] для расчета оптимальной конформации неполярных полимерных молекул в водной среде, является то, что не нужно явным образом учитывать водное окружение, что существенно сокращает вычислительные затраты.

Выражаем благодарность С.С. Колесникову и Ю.В. Бобкову за техническую помощь при проведении расчетов.

Данная работа выполнена при поддержке гранта INTAS-93-2084.

6 Приложение.

Здесь мы приводим схему алгоритма решения уравнения (11). Предполагается, что известна геометрия N -атомной молекулы, некулоновская часть атом-атомных потенциалов взаимодействия с атомами воды U_{iX} и заряды на атомах q_i . Предполагается также, что предварительно рассчитаны атом-атомные корреляционные функции $h_{\text{HH}}, h_{\text{OH}}, h_{\text{OO}}$ чистой воды при плотности ρ и температуре $k_{\text{B}}T$, исходя из заданной потенциальной модели воды. Геометрическая структура молекулы воды $r_{\text{OH}}, r_{\text{HH}}$, и заряды $q_{\text{H}}, q_{\text{O}}$ предполагаются известными.

Алгоритм для нахождения неизвестного вектора \vec{X} описывается следующими шагами:

1. Выбираем число узлов L , шаг сетки Δr , параметр редукции m . Задаем параметры выхода: ϵ, δ .
2. Вычисляем матрицы \hat{W}, \hat{V} и вектора \vec{M}, \vec{Q} .
3. Задаем начальное приближение для вектора \vec{X} .
4. Вычисляем вектор невязки \vec{R} , в том случае если норма этого вектора удовлетворяет условию $|\vec{R}| < \epsilon$, то выходим из программы, иначе переходим к следующему шагу.
5. При помощи нестационарного итерационного метода решения линейных систем находим вектор $\Delta\vec{X}(k')$, удовлетворяющий условию:

$$|\hat{J}_m \Delta\vec{X}(k') + \vec{R}(k')| < \delta |\vec{R}(k')|$$

6. Вычисляем оставшиеся компоненты вектора $\Delta\vec{X}$: $\Delta\vec{X}(k'') = -\vec{R}(k'')$.
7. Вычисляем новое приближение: $\vec{X} = \vec{X} + \Delta\vec{X}$, затем переходим к шагу 4.

Остается сделать несколько замечаний относительно выбора алгоритма нестационарного итерационного метода. На наш взгляд этот выбор ограничивается двумя методами: *Bi-CGSTAB* и *TFQMR*, поскольку остальные методы требуют значительно больше памяти или обладают менее устойчивой сходимостью. Оба метода вполне конкурентны, они требуют вычисления двух матрично-векторных произведений на итерацию.

Список литературы

- [1] Chandler D., Andersen H.C. // J. Chem. Phys. 1972. V. 57. n 5. P. 1930.
- [2] Monson P.A., Morriss G.P. // Adv. Chem. Phys. / Eds. by I. Prigogine and S. Rice. 1990 V. 77 P. 451.
- [3] Pettitt B.M., Karplus M., Rossky P.J. // J. Phys. Chem. 1986. V. 90. N. 23. P. 6335.
- [4] Hirata F., Rossky P.J., Pettitt B.M. // J. Chem. Phys. 1983. V. 78 P. 4133.
- [5] Hirata F., Levy R.M. // Chem. Phys. Lett. 1987. V. 136 P. 267.
- [6] Hirata F., Levy R.M. // J. Phys. Chem. 1989. V. 93. P. 479.
- [7] Тихонов Д.А., Саркисов Г.Н. // Журн. физ. химии. 1997. Т. 71. С. 480.
- [8] Kelley. C. T. "Iterative Methods for Linear and Nonlinear Equations SIAM Frontiers 16, 1995.
- [9] Hansen J.-P., McDonald I. R. "Theory of Simple Liquids Academic Press, London, 1986.
- [10] Martynov G.A. "Fundamental Theory of Liquids Hilger, London, 1992.
- [11] Cammings P.T., Stell G. // Mol. Phys. 1982. V. 46. P. 383.

-
- [12] Ng Kin-Chue // J. Chem. Phys. 1974. V. 61. P. 2680.
- [13] Талицких С.К., Халатур П.К. // Журн. физ. химии. 1995. Т. 69. С. 5.
- [14] Landoldt-Bornstein Numerical Data and Functional Relationships in Science and Technology, New Series, V||/1b, W. Saenger, ed., Springer-Verlag, Berlin, 1989.
- [15] Zerah G. // J. Comput. Phys. 1985. V. 61, P. 280.
- [16] Gillan. M.J. // Mol. Phys. 1979. V. 38, N.6. P. 1781.
- [17] Labik S. Malievsky A., Vonka. P. // Mol. Phys. 1985. V. 56. N. 3, P. 709.
- [18] Ichiye T., Haymet A.O.J. // J. Chem. Phys. 1988. N. 7. P. 4315.
- [19] Homeier H.H.H., Rast S., Krienke H. // Comp. Phys. Commun. 1995. V. 92. P. 188.
- [20] Poltev V.I., Teplukhin A.V., Malenkov G.G. // Int. J. Quant. Chem. 1992. V. 42 P. 1499.
- [21] Журкин В.Б., Полтев В.И., Флорентьев В.Л. // Молекуляр. биология. 1980. Т. 14. С. 1116.
- [22] Jorgensen W.L. // J. Amer. Chem. Soc. 1981. V. 103. P. 341.

METHOD OF ATOM-ATOM CORRELATION FUNCTIONS IN RISM APPROXIMATION FOR
STUDYING OF HYDRATION OF MACROMOLECULES.

D.A. Tikhonov, R.V. Polozov
Institute of Theoretical and Experimental Biophysics, Russian Academy of Sciences
A.V. Gorelov, E.G. Timoshenko, Yu.A. Kuznetsov
Department of Chemistry, University College Dublin
L.A. Panchenko
Moscow M.V. Lomonosov State University
K.A. Dawson
Department of Chemistry, University College Dublin

An efficient numerical algorithm for solving integral equations of the theory of liquids in the RISM approximation for infinitely diluted solution of macromolecules with a large number of atoms is proposed. The algorithm is based on applying the non-stationary iterative methods for solving systems of linear algebraic equations. Using this technique we have calculated the solvent-solute atom-atom correlation functions for short fragments of DNA duplexes of varying length in aqueous solution.