

УДК 532.74

**ИЗУЧЕНИЕ ВЗАИМОДЕЙСТВИЯ КОФЕИНА С ДНК В ВОДНОМ РАСТВОРЕ  
МЕТОДОМ МОНТЕ-КАРЛО С ИСПОЛЬЗОВАНИЕМ ПАРАЛЛЕЛЬНЫХ  
ВЫЧИСЛЕНИЙ**© 2009 М.Д. Калугин<sup>1</sup>, А.В. Теплухин<sup>2\*</sup><sup>1</sup>Институт системного программирования РАН, Москва<sup>2</sup>Институт математических проблем биологии РАН, Пущино*Статья поступила 13 ноября 2008 г.*

Методом Монте-Карло с использованием параллельных вычислений проведено моделирование водных растворов кофеина, содержащих сверхспиральный фрагмент В-ДНК. Установлены места связывания молекул кофеина с ДНК и наиболее вероятные структуры образующихся комплексов. Выполнена оценка степени ассоциированности молекул кофеина в водных растворах различной концентрации и выявлены основные типы конфигураций молекулярных агрегатов.

**Ключевые слова:** вода, структура, кофеин, ДНК, компьютерное моделирование, Монте-Карло, параллельные вычисления.

**ВВЕДЕНИЕ**

Взаимодействие молекул кофеина с ДНК в водном растворе интенсивно изучается как теоретическими, так и экспериментальными методами. Главная цель этих исследований — установить на молекулярном уровне различные механизмы активности кофеина, когда он может либо сам выступать в роли лекарства, либо действовать в качестве "перехватчика лекарств", либо становиться мутагеном [1—4]. Основным источником экспериментальных данных в этой области служат работы, выполненные на основе методов ЯМР или ИК/УФ спектроскопии, а собственно информация о строении молекулярных комплексов получается в процессе интерпретации спектральных кривых концентрационной или/и температурной зависимости. С математической точки зрения данная проблема является типичным представителем широкого класса так называемых обратных задач, причем единственным средством для ее решения зачастую остается метод подбора (чем больше вариантов решения рассмотрено, тем выше надежность выводов). В этой ситуации компьютерное моделирование становится незаменимым инструментом, не только стимулируя интуицию исследователя, но и предоставляя ему широкий набор готовых вариантов для анализа. Таким образом, к задачам данной работы можно отнести как изучение строения возможных молекулярных агрегатов, так и разработку высокоэффективного метода моделирования водных растворов ДНК, содержащих кофеин.

Связывание низкомолекулярных лигандов с ДНК в водном растворе обычно осуществляется двумя способами: в зависимости от своей химической природы молекула лиганда действует либо как интеркалятор, встраиваясь между плоскостями соседних нуклеотидных пар ДНК, либо вступает в тесный контакт с атомами на боковой поверхности двойной спирали ДНК. В первом случае структура ДНК претерпевает существенные изменения. Во втором случае связывание лиганда сопровождается лишь незначительными возмущениями геометрии двойной спирали (например, легкими изменениями ширины желобов ДНК). По данным ИК спектроско-

---

\* E-mail: tepl@impb.psn.ru

пии [ 5 ] молекулы кофеина в водном растворе достаточно сильно взаимодействуют лишь с фосфатами ДНК либо с боковыми атомными группами азотистых оснований, что говорит о низкой вероятности процессов интеркаляции. В свою очередь степень консерватизма структуры двойной спирали при связывании по второму способу дает нам достаточно аргументов для "замораживания" внутренних степеней свободы исследуемого образца ДНК в ходе компьютерного моделирования, позволяя существенно ускорить расчет.

В отличие от большинства ферментов двойная спираль ДНК не предоставляет хорошо выраженных компактных "карманов" связывающимся с ней молекулам. Поэтому степень сродства (иными словами сила связи) в паре ДНК—лиганд не только зависит от наличия специфической комбинации гидрофильных и/или гидрофобных атомных групп в месте связывания, но и может существенно "модулироваться" локальными вариациями геометрии желобов ДНК. Кроме того, хорошо известно, что молекулы кофеина в водном растворе имеют тенденцию образовывать димеры и ассоциаты более высокого порядка [ 6 ]. Все эти факты говорят о необходимости использования крупных элементарных ячеек при моделировании.

Особое внимание должно быть уделено и требованию статистической надежности результатов компьютерного моделирования, поскольку любые вычислимые данные обычно получают путем усреднения по наборам "мгновенных" конфигураций, генерируемых либо детерминистскими (уравнения молекулярной динамики), либо стохастическими (Монте-Карло) методами. С этой целью можно было бы выполнить несколько независимых расчетов, каждый раз используя неповторяющиеся (случайные) наборы начальных данных. Оптимальней, тем не менее, использовать в исследовании большой образец ДНК, состоящий из нескольких идентичных (и по нуклеотидной последовательности и по геометрическим параметрам) фрагментов, а затем рассчитывать требуемые данные путем усреднения по соответствующим наборам мест связывания.

Характеристики систем, состоящих из нескольких сотен молекул воды, легко рассчитываются на обычных персональных компьютерах. Однако при моделировании водного окружения биополимеров приходится использовать элементарные ячейки очень большого объема. Количество молекул воды в таких моделях исчисляется сотнями тысяч. Использование последовательных алгоритмов для решения этих задач становится крайне неэффективным не только из-за роста объемов вычислений при выполнении расчетов, но и в связи с необходимостью размещения очень больших массивов данных в оперативной памяти компьютера. Обычно расчеты таких молекулярных систем выполняют на многопроцессорных комплексах с использованием параллельных вычислений.

В данной работе применен новый подход к моделированию очень больших (до  $10^9$  атомов) молекулярных агрегатов в рамках метода Монте-Карло. В его основу положен оригинальный алгоритм [ 7 ], позволяющий выполнять классическую процедуру Метрополиса [ 8 ] одновременно для нескольких (определяется числом процессоров, задействованных в решении данной задачи) молекул. Программное обеспечение, необходимое для моделирования водного раствора кофеина, было разработано на алгоритмическом языке Fortran77 с применением стандартных коммуникационных подпрограмм библиотеки MPI 1.2 [ 9 ]. Доработку алгоритмов и оптимизацию параллельного кода выполняли с использованием специализированного пакета ParJava [ 10 ].

Благодаря распределенным параллельным вычислениям изучение молекулярных систем размерами в сотни нанометров становится вполне выполнимым на современных многопроцессорных суперкомпьютерах. Это позволяет, в частности, вывести технологии компьютерной разработки лекарств на новый уровень. Например, вероятность и силу связывания молекулы лекарства с активным центром можно оценивать непосредственно, как долю пар сайт—лиганд, образующихся в моделируемой ячейке, содержащей не только большое количество молекул растворителя, но и значительное число молекул лекарства и мест связывания (в соответствии с их реальными физиологическими концентрациями). При этом не только становится излишней процедура расчета приращений свободной энергии (обязательная при использовании стандартных подходов и требующая больших вычислительных ресурсов), но и появляется возможность корректного учета очень важной компоненты свободной энергии раствора, отвечающей взаимо-

действиям между молекулами самого лекарства (а также сайт—сайт и иным, в том числе и многочастичным корреляциям).

### МЕТОДИКА РАСЧЕТОВ

**Стандартный ("последовательный") алгоритм моделирования воды методом Монте-Карло.** Для расчета структурных и термодинамических характеристик воды или водных растворов методом Монте-Карло обычно применяют процедуру Метрополиса [ 8 ], реализованную по классической схеме [ 11 ]. В качестве модельной системы чаще всего рассматривают элементарную ячейку кубической формы с периодическими граничными условиями, содержащую десятки или сотни молекул воды. Для расчета энергии межмолекулярных взаимодействий используют атом-атомные потенциальные функции (в виде кулоновских и ван-дер-ваальсовых вкладов) ограниченного радиуса действия (сферическое обрезание) и метод ближайшего образа [ 11 ]. Размер ячейки, температура и число частиц в ходе расчетов не изменяются (*NVT*-ансамбль). Начальное положение молекул выбирается случайным образом в пределах ячейки.

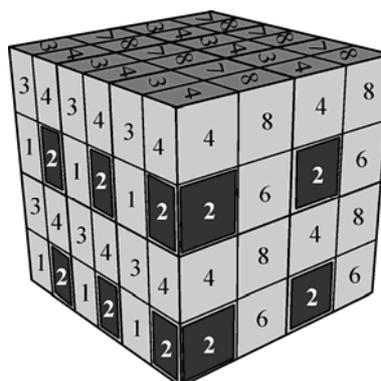
В нескольких фразах схема стандартной ("последовательной") процедуры Метрополиса для *NVT*-ансамбля может быть изложена следующим образом: конкретная молекула поворачивается на случайный угол вокруг случайно выбранной оси, проходящей через ее центр масс, и смещается на случайный вектор (с учетом периодических граничных условий); рассчитывается изменение энергии системы. Если энергия снижается, новое положение молекулы принимается. В противном случае новое положение может быть принято с вероятностью, определяемой больцмановским фактором этого приращения энергии. Если перемещение, ведущее к росту энергии, отвергнуто, молекула остается в исходном положении, а молекулярная конфигурация, не изменившаяся в ходе этого испытания, повторно принимается в качестве очередного состояния марковской цепи. Затем эта последовательность действий применяется к следующей молекуле (циклически, по номерам молекул) и т.д. В нормальных условиях максимальное смещение за одно испытание составляет 0,02 нм, а максимальный угол поворота — 0,15 рад. Более подробное изложение можно найти в книге [ 11 ].

**"Параллельный" алгоритм метода Монте-Карло для моделирования воды.** Компьютерное моделирование растворов биополимеров требует рассмотрения элементарных ячеек очень большого объема, чтобы разместить в них многие миллионы атомов. Изучать такие системы на обычных настольных компьютерах, используя вычислительные процедуры последовательного типа, крайне неэффективно. В то же время параллельные вычисления на многопроцессорных суперкомпьютерах имеют очень высокий потенциал для решения задач такого рода. Ниже мы кратко изложим основные идеи параллельного алгоритма Монте-Карло [ 7 ], реализованного в данной работе для моделирования взаимодействия кофеина с ДНК в водном растворе.

Прежде всего следует распределить пространство элементарной ячейки (*C*-куб) среди доступных для решения задачи процессоров (пусть их будет *NP* штук) так, чтобы каждый из них "управлял" молекулами своей части (далее *P*-куб) моделируемой системы. В этом случае многопроцессорная вычислительная среда дает возможность одновременного выполнения *NP* испытаний по алгоритму Метрополиса, что существенно ускоряет расчеты. Корректность вычислений по такой параллельной схеме возможна при условии статистической независимости *NP* молекул, одновременно изменяющих свое положение в пространстве на каждом шаге расчетов методом Монте-Карло. Кроме того, в памяти каждого задействованного процессора должна храниться и своевременно обновляться информация, необходимая для расчета энергии взаимодействия молекул соответствующего *P*-куба с молекулами из окружающих *P*-кубов.

Выполнение этих требований можно обеспечить следующим образом. Разобьем каждый *P*-куб еще на 8 одинаковых кубов (далее *L*-кубы) взаимно перпендикулярными плоскостями, проходящими через центры *P*-кубов и параллельными соответствующим граням *C*-куба (рис. 1). При этом длина ребра *L*-куба не должна быть менее  $R_c + 2\delta$ . Здесь  $R_c$  — максимальный радиус обрезания атом-атомных потенциальных функций;  $\delta$  — максимальная величина смещения молекулы воды вдоль какой-либо координатной оси в результате одного испытания методом Монте-Карло при заданной температуре.

Рис. 1. Фрагмент моделируемой системы из 12  $P$ -кубов (жирные линии). Показано разбиение на  $L$ -кубы (тонкие линии) и порядок их выбора при моделировании (цифры)



Каждый процессор выполняет процедуру Метрополиса (одно испытание) для молекулы, случайно выбранной в одном из  $L$ -кубов "своего"  $P$ -куба.  $L$ -кубы, содержащие испытываемые молекулы, назначаются по единой для всех  $NP$  вычислительных процессов схеме: циклический обход с первого до восьмого (см. порядок на рис. 1). Иными словами,  $L$ -кубы, с которыми на данной итерации "работают" процессоры, располагаются в одних и тех же местах соответствующих  $P$ -кубов. Например,  $i$ -е испытание процессоры выполняют каждый со своей молекулой, случайно выбранной в правом  $L$ -кубе из ближнего к нам ряда в нижнем слое соответствующего  $P$ -куба (см. рис. 1, темно-серые клетки с номером 2). В результате такой организации вычислений любые одновременно испытываемые молекулы всегда находятся на расстояниях, превышающих величину  $R_c$ , а значит, не могут оказывать влияния на движение друг друга в рамках нашей модели.

Для того чтобы корректно обрабатывать перемещения молекул между  $P$ -кубами, а также рассчитывать "межпроцессорные" вклады в энергию системы, в памяти каждого вычислительного процесса необходимо иметь данные не только о молекулах воды из "своих" восьми  $L$ -кубов, но также и о молекулах из  $L$ -кубов, принадлежащих соседним  $P$ -кубам. Таким образом, в оперативной памяти каждого процессора должна храниться информация о молекулах 64  $L$ -кубов, включая 8 своих и 56 соседних. После очередного испытания, благодаря использованию обрезания потенциала, каждой паре "соседних" процессоров ( $P$ -кубов) достаточно обменяться информацией только об  $L$ -кубах, имеющих общую грань, ребро или вершину. Таким образом, объем оперативной памяти, требуемый от каждого вычислительного узла для проведения расчетов, не зависит от размера моделируемой системы ( $C$ -куба) и определяется постоянной величиной  $R_c$ . То же самое верно и для максимальной длины межпроцессорного сообщения, и для количества вычислительной работы, приходящейся на каждый процессор в одном цикле процедуры Метрополиса.

Количество обменов сообщениями, в расчете на один процессор, также не зависит от размера модели (и числа задействованных процессоров). В общем случае (начиная с модели для расчетов на 27 ( $3 \times 3 \times 3$ ) процессорах), каждому процессору необходимо обменяться сообщениями не более чем с 26 другими, сделав 26 вызовов стандартной функции *SendRecv* из библиотеки MPI [ 9 ]. Количество сообщений можно уменьшить до 6 (только 3 пары обменов в направлениях, перпендикулярных граням  $P$ -кубов [ 12 ]), накапливая обмениваемую информацию. Однако в этом случае размер сообщения после первых двух пар обменов может вырасти девятикратно. Тем не менее схема с шестью обменами не только легко программируется с использованием MPI (посредством организации пространственной "декартовой" топологии межпроцессорных связей), но и обеспечивает высокую вычислительную эффективность. Следует заметить, что при расчетах на восьми процессорах (модель  $2 \times 2 \times 2$   $P$ -кубов), благодаря периодическим граничным условиям, у каждого  $P$ -куба имеется только 7 фактических соседей. В такой модели можно обойтись тремя обменами на каждом испытании.

Межпроцессорные коммуникации даже в современных суперкомпьютерах сильно снижают эффективность параллельных вычислений, поэтому совершать обмен данными после каждого испытания (см. выше, порядок назначения  $L$ -кубов) нерационально (процессоры простаивают, ожидая данных). Скорость вычислений значительно возрастет, если каждый процессор сначала выполнит процедуру Метрополиса со всеми молекулами назначенного  $L$ -куба, а затем начнет обмен данными. Более того, перед пересылкой полезно, используя технику списков соседних молекул *à la Verlet* [ 11 ], выполнить подряд несколько циклов испытаний с молекулами одного и того же  $L$ -куба, но злоупотреблять этим не стоит, так как могут возникнуть проблемы с эргодичностью.

Применение для межпроцессорного обмена данными библиотечных подпрограмм *Send-Recv* блокирующего типа дополнительно выполняет очень важную функцию по общей синхронизации вычислений по всему коллективу процессоров, используемых в решении данной задачи. Некоторая рассогласованность в работе процессоров, возникающая, несмотря на принудительную синхронизацию в ходе пересылок, из-за неодинаковости данных или по другим причинам, положительно влияет на производительность, поскольку снижает вероятность пиковых перегрузок линий межпроцессорных коммуникаций.

Начальные данные готовятся специальной программой, равномерно (с помощью генератора случайных чисел) заполняющей элементарную ячейку молекулами до достижения заданной плотности, не допуская при этом слишком коротких межатомных контактов. Эта программа выполняется параллельно на совокупности процессоров, соответствующей размеру заданного объекта исследований. Каждый процессор генерирует атомные координаты только для молекул внутри своего  $P$ -куба, а затем топологически соседние процессоры обмениваются данными, необходимыми для инициализации перекрывающихся частей их доменов.

Следует отметить, что представленный подход к моделированию молекулярных систем не имеет формальных ограничений ни на размер элементарной ячейки, ни на количество атомов в исследуемом объекте.

**Использование параллельного алгоритма для моделирования более сложных объектов.** При моделировании водной оболочки биомолекул возникает специфическая проблема, связанная с расчетом энергии их взаимодействия с молекулами окружающей воды. Здесь уместно напомнить, что интересующие нас биологические молекулы являются полимерами, т.е. представляют собой цепочки из небольших групп атомов, соединенных химическими связями в определенном порядке (аминокислоты в молекулах белков и нуклеотиды в ДНК и РНК). Таким образом, любую биомолекулу можно представить в виде совокупности компактных и, по возможности, электронейтральных фрагментов, содержащих небольшое число атомов. Так, образец ДНК разбивается на нуклеотиды, а в каждом из них выделяется фосфатная группа, остаток дезоксирибозы и азотистое основание, определяющее тип соответствующего нуклеотида. При таком способе разбиения получают фрагменты из 5—15 атомов, расположенные в пределах сферы радиуса  $R_b = 2—3 \text{ \AA}$ . Для каждого фрагмента определяются координаты его геометрического центра, и, в соответствии с их значениями, программа генерации начальных данных распределяет фрагменты по  $P$ -кубам. Весовыми коэффициентами при расчете координат центра служат абсолютные величины парциальных атомных зарядов, выраженные в единицах модуля заряда электрона.

Согласно принятой в данной работе "жесткой" модели ДНК все ее фрагменты остаются неподвижными в ходе моделирования, а расчет энергии взаимодействия вода—ДНК сводится к суммированию энергий взаимодействия отдельных молекул воды с каждым из фрагментов ДНК, расстояние до геометрического центра которого не превышает  $R_{cb} = R_c + R_b$ . Параметр  $R_c$  в настоящей работе равен 0,9 нм, а величина  $R_b$  для молекулы воды приравнена к нулю. В общем случае (включая молекулы кофеина)  $R_{cb}(i,j) = R_c + R_b(i) + R_b(j)$ , где  $i$  и  $j$  — типы фрагментов. Длина ребра  $L$ -куба должна превосходить наибольшее из этих  $R_{cb}(i,j)$  хотя бы на  $2\delta$ . Начальная конфигурация молекул воды получается из предварительно подготовленного "водного куба" требуемых размеров путем исключения молекул воды, расположенных неестественно близко к атомам ДНК или молекул кофеина.

**Молекулярные модели, параметры потенциальных функций и композиция элементарных ячеек.** Мы использовали жесткую трехцентровую модель молекулы воды из работы [13]. Молекулы кофеина (рис. 2) также рассматривались как твердые тела с молекулярной геометрией, описанной в работе [14]. Неподвижный двуспиральный фрагмент ДНК  $(GAGAAAAGA)_{14}:(TCTTTTCTC)_{14}$ , состоящий из 140 нуклеотидных пар (14 витков двойной спирали в  $B$ -форме) и образующий петлю суперспирали (диаметр 13,4 нм, шаг 7,6 нм между соответствующими фосфатами), был помещен (центр в центр) в элементарную ячейку кубической формы с ребром 18 нм. Координаты атомов фрагмента ДНК рассчитаны с помощью процедуры репликации, использующей криволинейные периодические граничные условия вдоль

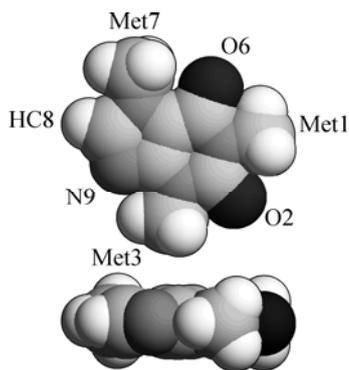


Рис. 2 (слева). Молекула кофеина (две проекции)

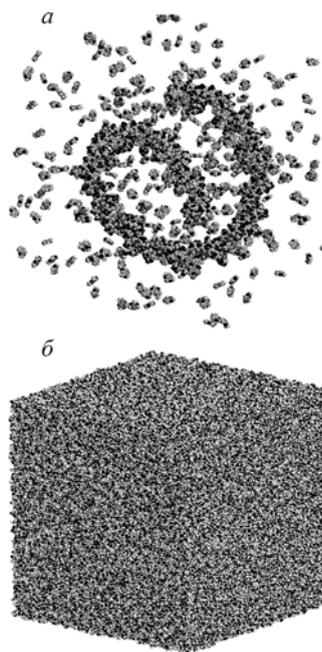


Рис. 3 (справа). Расположение фрагмента ДНК и 350 молекул кофеина в элементарной ячейке (одна из мгновенных конфигураций, молекулы воды не показаны) — а, общий вид моделируемой системы (показаны все атомы) — б

оси двойной спирали ДНК. Оригинальные координаты атомов декамера GAGAAAAGA: TCTTTTCTC, а также параметры для их преобразования были любезно предоставлены В.Б. Журкиным. Важно отметить, что размер гликозидного желоба [ 15 ] этого 140-мера двуспиральной ДНК периодически (14 раз) варьирует от своего минимального значения в конце А-тракта до максимальной величины в его начале. Кроме того, обе 5'-концевые группы  $-\text{CH}_2-\text{PO}_3$  были заменены метильными, а два 3'-концевых атома кислорода стали гидроксильными. Таким образом, исследуемый образец ДНК состоит из 8914 атомов.

Параметры потенциальных функций для взаимодействий вода—вода, вода—кофеин и вода—ДНК те же, что и использованные в наших предыдущих работах [ 13, 15, 16 ], а для взаимодействий кофеин—кофеин и кофеин—ДНК взяты из работы [ 17 ]. Величины парциальных атомных зарядов рассчитывались с помощью простых методов MO-LCAO и Хюккеля с использованием набора параметров работы [ 18 ]. Соответствующие значения (в единицах модуля заряда электрона) для всех типов ингредиентов, формирующих изучаемую систему, представлены в табл. 1.

Было проведено три компьютерных "эксперимента" по изучению взаимодействия кофеин—ДНК в разбавленном, насыщенном и перенасыщенном водных растворах (концентрации кофеина 0,05, 0,1 и 0,2 М соответственно) при температуре 300 К. Для этого при проведении расчетов использовались три набора по 216 процессоров суперкомпьютера МВС100К в МСЦ РАН (<http://www.jssc.ru>). Стартовые конфигурации, подготовленные независимо для каждой из трех систем, содержали 175 и 192869, или 350 и 187224, или 700 кофеиновых и 181730 водных молекул соответственно, а также один и тот же образец ДНК. Таким образом, каждая из моделей состоит из 591721, 578986 и 570904 атомов соответственно. Общий вид одной из элементарных ячеек представлен на рис. 3. Чтобы устранить влияние произвольности выбора стартовой конфигурации на результаты, первоначальный участок марковской цепи, соответствующий 5 миллионам испытаний в расчете на каждую молекулу, исключали из рассмотрения. Расчет и анализ структурных и энергетических характеристик выполняли далее по ансамблям конфигураций, соответствующим не менее 30 миллионам испытаний, выполненных с каждой молекулой.

#### РЕЗУЛЬТАТЫ И ИХ ОБСУЖДЕНИЕ

**Распределение молекул кофеина в водном растворе вокруг ДНК.** При изучении локального упорядочения в простых жидкостях часто используют аппарат радиальных функций распределения (РФР). В нашем случае (для исследования пространственных корреляций в мо-

Т а б л и ц а 1

Парциальные заряды  $q$  (в единицах заряда позитрона) компонентов ДНК и атомов кофеина

| Ade  |        | Thy  |        | Gua  |        | Cyt  |        | Sugar |        | Phos |        | Кофеин     |        |
|------|--------|------|--------|------|--------|------|--------|-------|--------|------|--------|------------|--------|
| Атом | $q$    | Атом | $q$    | Атом | $q$    | Атом | $q$    | Атом  | $q$    | Атом | $q$    | Атом       | $q$    |
| N1   | -0,518 | N1   | -0,111 | N1   | -0,254 | N1   | -0,065 | C1'   | 0,139  | O1P  | -0,730 | N1         | -0,157 |
| C2   | 0,313  | C2   | 0,443  | C2   | 0,545  | C2   | 0,437  | C2'   | -0,045 | O2P  | -0,730 | C1M        | -0,088 |
| N3   | -0,510 | O2   | -0,353 | N2   | -0,417 | O2   | -0,392 | C3'   | 0,082  | O5'  | -0,313 | C2         | 0,428  |
| C4   | 0,311  | N3   | -0,325 | N3   | -0,628 | N3   | -0,667 | O3'   | -0,313 | P    | 0,802  | O2         | -0,354 |
| C5   | 0,112  | C4   | 0,334  | C4   | 0,285  | C4   | 0,398  | C4'   | 0,097  | H51' | 0,053  | N3         | -0,137 |
| C6   | 0,340  | O4   | -0,350 | C5   | 0,027  | N4   | -0,431 | O4'   | -0,280 | H52' | 0,053  | C3M        | -0,088 |
| N6   | -0,439 | C5   | -0,097 | C6   | 0,342  | C5   | -0,160 | C5'   | 0,014  |      |        | C4         | 0,256  |
| N7   | -0,563 | C5M  | -0,151 | O6   | -0,371 | C6   | 0,234  | H1'   | 0,055  |      |        | C5         | 0,000  |
| C8   | 0,291  | C6   | 0,125  | N7   | -0,545 | H41  | 0,229  | H21'  | 0,045  |      |        | C6         | 0,324  |
| N9   | -0,013 | H3   | 0,195  | C8   | 0,213  | H42  | 0,229  | H22'  | 0,045  |      |        | O6         | -0,350 |
| H2   | 0,070  | H6   | 0,059  | N9   | 0,004  | H5   | 0,055  | H3'   | 0,054  |      |        | N7         | -0,015 |
| H61  | 0,230  | HM1  | 0,054  | H1   | 0,196  | H6   | 0,061  | H4'   | 0,051  |      |        | C7M        | -0,090 |
| H62  | 0,230  | HM2  | 0,054  | H21  | 0,231  |      |        |       |        |      |        | C8         | 0,276  |
| H8   | 0,068  | HM3  | 0,054  | H22  | 0,231  |      |        |       |        |      |        | N9         | -0,554 |
|      |        |      |        | H8   | 0,068  |      |        |       |        |      |        | HM11,12,13 | 0,052  |
|      |        |      |        |      |        |      |        |       |        |      |        | HM31,32,33 | 0,052  |
|      |        |      |        |      |        |      |        |       |        |      |        | HM71,72,73 | 0,057  |
|      |        |      |        |      |        |      |        |       |        |      |        | H8         | 0,067  |

Примечание. В дезоксирибозах (Sugar) с 5'-концов полинуклеотидных цепей атом C5' должен быть заменен CH<sub>3</sub>-группой с парциальными зарядами -0,089 и 0,053 для атома углерода и каждого из трех ее атомов водорода соответственно. В дезоксирибозах с 3'-концов атом O3' должен быть заменен OH-группой с зарядами -0,483 и 0,226 для атомов кислорода и водорода соответственно. Водородные атомы H51' и H52', будучи химически связаны с атомом C5' дезоксирибозы, включены формально в состав фосфатного фрагмента и, наоборот, атом O3' включен во фрагмент "Sugar". Парциальные заряды на атомах молекулы воды [13]: -0,68 для кислорода и 0,34 для каждого водородного атома.

лекулярной подсистеме кофеин—ДНК) можно ввести в рассмотрение некоторый аналог таких функций. Иными словами, чтобы охарактеризовать положение конкретной молекулы кофеина по отношению к ДНК, при расчете РФР предлагается использовать расстояние между ближайшей парой их атомов (первый атом пары принадлежит молекуле кофеина, а второй — ДНК, атомы водорода исключены из рассмотрения).

Следует отметить, что на расстояниях, достаточно малых по сравнению с величиной радиуса кривизны суперспирали, эта зависимость обладает цилиндрической симметрией, причем для обеспечения правильной нормировки распределения (выход на единицу при больших удалениях) область внутри цилиндра радиусом 0,9 нм должна быть исключена из рассмотрения. Таким образом, мы получим зависимость относительной плотности вероятности встретить ближайший к ДНК атом молекулы кофеина (не водород) на заданном удалении от поверхности криволинейной нанотрубы, имитирующей усредненную боковую поверхность ДНК. Фактически наличие желобов дает возможность молекулам кофеина появляться глубоко внутри габаритного цилиндра В-ДНК радиусом 1,1 нм, но, как показал дальнейший анализ радиальных распределений, величина 0,9 нм более точно учитывает среднее расстояние от оси двойной спирали до ближайшего атома молекулы кофеина. Некоторое систематическое занижение на малых удалениях значений модифицированных таким способом РФР не является в данном случае принципиальным ввиду своей относительной малости.

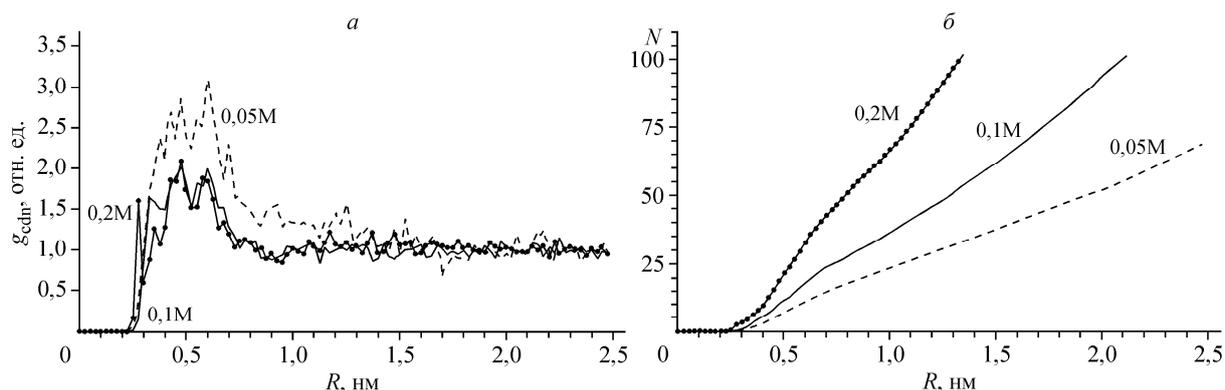


Рис. 4.  $g_{\text{cdn}}$  — относительная плотность радиального распределения молекул кофеина вокруг ДНК (а);  $N$  — среднее значение количества молекул кофеина, окружающих фрагмент ДНК слоем толщиной  $R$ , нм (см. пояснения в тексте) (б). Данные для трех концентраций раствора

На рис. 4, а представлены РФР кофеина вокруг ДНК, рассчитанные вышеописанным способом по статистически значимой выборке "мгновенных" молекулярных конфигураций, генерируемых в ходе процедуры Метрополиса. Прежде всего, на этом рисунке выделяется сильно заселенная (по сравнению с равномерным распределением по объему) область вокруг ДНК толщиной около 0,4 нм, образующаяся вне зависимости от концентрации раствора. Далее следует отметить ярко выраженные пики на расстояниях 0,35 нм (контакты через метильные группы), 0,47 нм ("водные мостики"; здесь ближайшую пару составляют гидрофильные атомы, образовавшие водородные связи с одной и той же молекулой воды) и 0,6 нм (диффузные скопления молекул кофеина вдоль сахаро-фосфатного остова со стороны негликозидного желоба ДНК). Резкий максимум на 0,275 нм говорит об образовании водородных связей между молекулами кофеина и ДНК. Среднее значение количества молекул кофеина, присутствующих в оболочке заданной толщины, охватывающей 140-мер ДНК, может быть оценено по интегральным кривым, изображенным на рис. 4, б. Например, в среднем только 0,1, 0,2 или 2,4 молекулы кофеина могут образовывать водородную связь с протоно-донорными группами ДНК в 0,05, 0,1 или 0,2 М растворах соответственно.

Для того чтобы выявить основные места связывания и установить конфигурацию возможных молекулярных комплексов, необходимо проанализировать представительную выборку "мгновенных" молекулярных конфигураций моделируемой ячейки. Для этого мы должны в каждой такой конфигурации (представляемой набором координат всех атомов) отобрать, используя геометрический критерий близости, основанный на данных рис. 4, а, только молекулы кофеина, контактирующие с ДНК (т.е., когда расстояние между ближайшими атомами пары кофеин—ДНК не превосходит 0,365 нм). На заключительном этапе моделирования было подготовлено три набора файлов (для каждой концентрации свой). Это делалось путем отбора последней конфигурации каждого стотысячного из десяти миллионов циклов процедуры Метрополиса по всем молекулам и дальнейшей "фильтрацией" с помощью геометрического критерия близости. Каждый такой набор файлов и представляет собой выборку, содержащую не менее 100 вариантов "кофеиновых оболочек" ДНК. На рис. 5 представлен общий вид (стереопара) такой выборки для раствора 0,1 М.

В ходе анализа полученных выборок удалось не только выявить основные типы комплексов кофеин—ДНК, наиболее часто образующихся в водной среде, и установить предпочтительные места связывания молекул кофеина на поверхности двойной спирали, но и обнаружить их корреляции с нуклеотидной последовательностью. Наиболее важный класс составляют комплексы, характеризующиеся структурами типа m1 или m2 (рис. 6). Такие конфигурации реализуются в зауженных участках гликозидного желоба (5'-AAAGA) как с образованием водородной связи между аминокгруппой гуанина и гидрофильным атомом кофеина (как правило, карбонильным

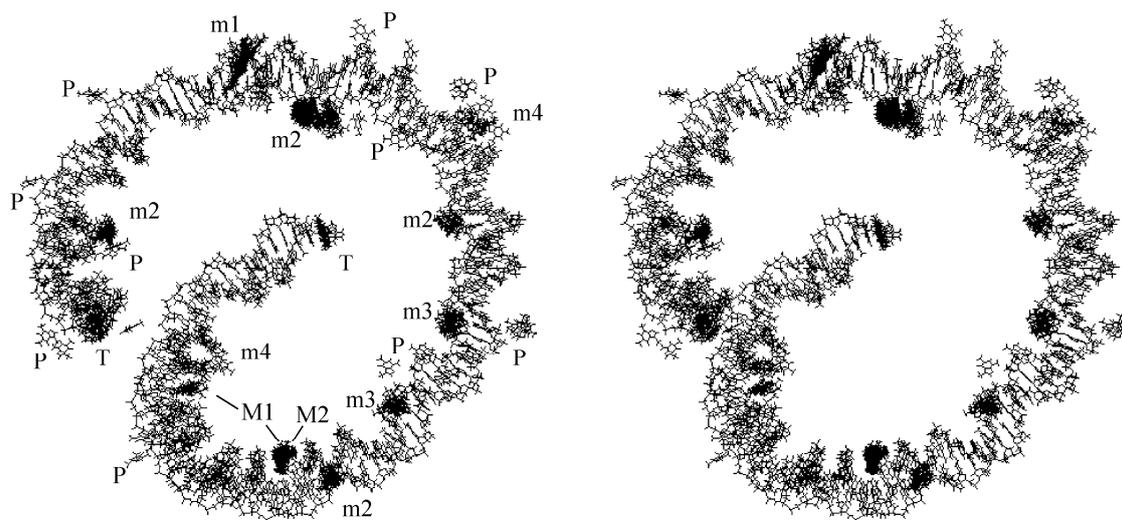


Рис. 5. Конфигурационные ансамбли молекулярных комплексов кофеин—ДНК в водном растворе (стереопара, молекулы окружающего раствора не показаны)

атомом кислорода) — тип m1, так и без нее — тип m2. Причем если в очень концентрированном растворе преобладает первый тип, то в более разбавленных растворах чаще встречается конфигурация без водородной связи (табл. 2).

В следующий класс следует объединить комплексы типов m3, m4 и P (см. рис. 5 и 6), характеризующиеся внешним (т.е. без внедрения в гликозидный желоб) связыванием с сахарофосфатным остовом ДНК. Другая их общая черта — тесные контакты метильных групп кофеина с алифатическими участками (дезоксирибозы между фосфатами) полинуклеотидных цепей. Следует отметить, что граница между типами m3 и m4 весьма условна, поскольку достаточно часто плоскость ароматических колец молекулы кофеина находится в промежуточном положении. Вероятность образования бидентатных комплексов (m3 и m4) несколько снижается при увеличении концентрации раствора кофеина, в то время как доля агрегатов типа P при этом резко увеличивается (см. табл. 2). "Ареалом" данного класса являются участки 5'-GAAA гликозидного желоба, имеющие относительно большую ширину.

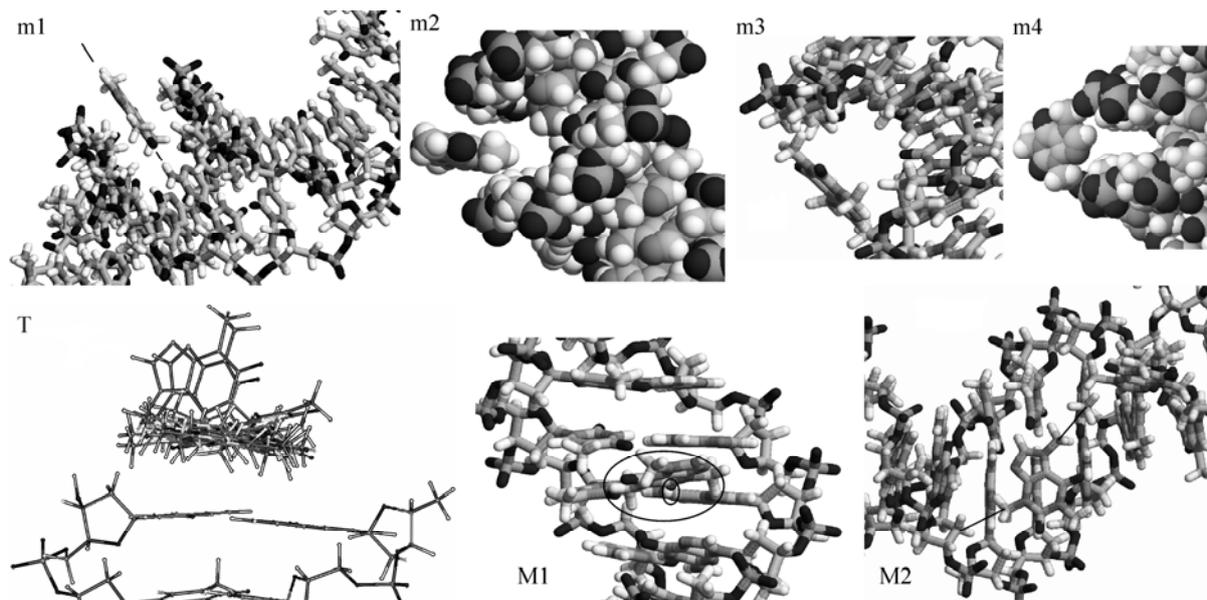


Рис. 6. Наиболее устойчивые конфигурации основных типов комплексов кофеин—ДНК

Т а б л и ц а 2

Вероятность образования комплексов кофеин—ДНК  
(в расчете на один сегмент GAGAAAAAGA:TCTTTTCTC)

| Тип комплексов кофеин—ДНК (рис. 8) | Концентрация водного раствора кофеина, М |      |      | Тип комплексов кофеин—ДНК (рис. 8) | Концентрация водного раствора кофеина, М |      |      |
|------------------------------------|--|------|------|------------------------------------|--|------|------|
|                                    | 0,05                                     | 0,1  | 0,2  |                                    | 0,05                                     | 0,1  | 0,2  |
| m1                                 | 0,01                                     | 0,01 | 0,31 | M1                                 | < 0,01                                   | 0,01 | 0,04 |
| m2                                 | 0,02                                     | 0,08 | 0,03 | M2                                 | 0,01                                     | 0,03 | 0,05 |
| m3                                 | 0,01                                     | 0,01 | 0,01 | T <sub>AT</sub>                    | < 0,01                                   | 0,12 | 0,27 |
| m4                                 | 0,08                                     | 0,04 | 0,03 | T <sub>GC</sub>                    | < 0,01                                   | 0,14 | 0,19 |
|                                    |  |      |      | P                                  | 0,02                                     | 0,08 | 0,21 |

Комплексы типов M1 и M2, чаще всего встречающиеся в наиболее узкой части (на участках 5'-GAAA) негликозидного желоба двойной спирали, образуют третий структурный класс. Характерной особенностью конфигурации типа M1 является наличие водородной связи между протоно-акцепторным атомом кофеина (чаще всего это один из двух карбонильных атомов кислорода) и аминогруппой аденина (см. рис. 6, овалы) или цитозина. Структуры комплексов типа M2 довольно разнообразны и могут осуществляться либо путем установления метильными группами кофеина гидрофобных контактов с группой CH<sub>3</sub> тимина и/или метиленовой (при атоме C2') группой дезоксирибозы (см. рис. 6, линии на изображении M2), либо через формирование своеобразного  $\pi$ -комплекса, когда химическая связь атома C5 тимина с его группой CH<sub>3</sub> направлена в центр молекулы кофеина перпендикулярно плоскости ароматических колец.

Конфигурации, возникающие при взаимодействии молекул кофеина с концевыми участками нерасплетенной двойной спирали, составляют особый класс "ограниченного пользования", поскольку экспериментальных данных о наличии именно такого пространственного строения этих областей ДНК в водном растворе недостаточно. Тем не менее полученные нами данные адекватно описывают взаимодействия молекул кофеина с азотистыми основаниями и в рассматриваемой ситуации. Чаще всего формируются комплексы стопочной или T-образной структуры (см. выборку "T" на рис. 6), а также "переходные" между ними конфигурации, когда плоскости ароматических колец взаимодействующих молекул не параллельны и не ортогональны. При увеличении концентрации растворов вероятность появления таких комплексов быстро возрастает (см. табл. 2). Сравнительно редко (с вероятностью 0,02—0,05) могут образовываться водородные связи между гидрофильными атомами кофеина и протонами 3'-концевых гидроксиллов.

Экспериментальные данные, имеющиеся в научной литературе, дают слишком скудную, но очень важную структурную информацию, подтверждающую наши выводы о взаимодействии молекул кофеина с двойной спиралью ДНК. В частности, отмечается [5], что кофеин взаимодействует с сахаро-фосфатным остовом и аминогруппами азотистых оснований в желобах ДНК.

**Взаимное расположение молекул кофеина в водном растворе.** Для изучения пространственных корреляций между положениями молекул кофеина были рассчитаны РФР двух типов (рис. 7):  $g_{aa}$  — функция расстояния между ближайшей парой атомов двух молекул (атомы водорода в этой процедуре не участвуют) и  $g_{cc}$  — функция расстояния между геометрическими центрами пары молекул. Для обеспечения правильной нормировки распределения  $g_{aa}$  область внутри сферы радиусом 0,35 нм должна быть исключена из рассмотрения. Таким образом, мы получим зависимость относительной плотности вероятности встретить один из атомов (не водород) кофеина на заданном удалении от атомов на боковой поверхности другой молекулы кофеина. Сопоставляя данные этих графиков и принимая во внимание дискообразную (средний диаметр — 0,95 нм, толщина — 0,34 нм, см. рис. 2) форму ван-дер-ваальсовой поверхности молекулы кофеина, можно выявить ряд важных закономерностей взаимного расположения этих молекул в водном растворе.

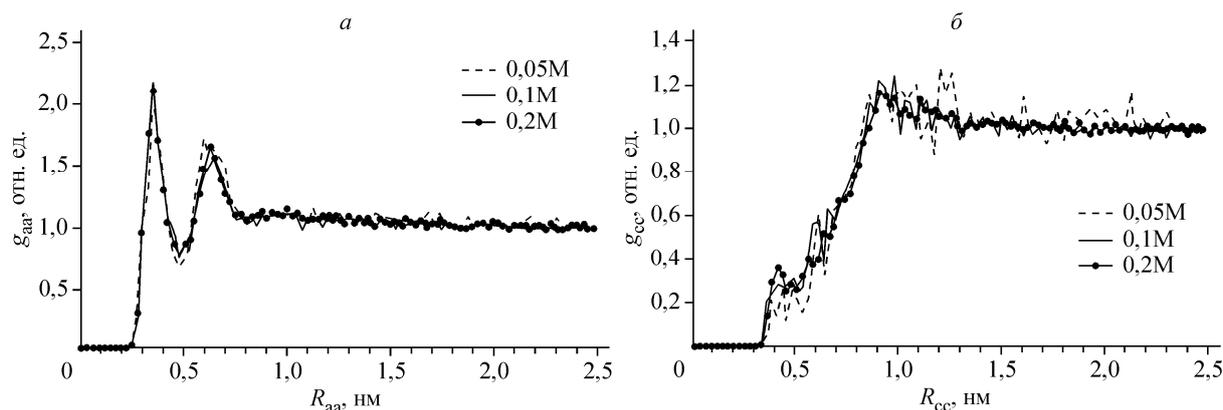


Рис. 7. РФР кофеина в водном растворе для трех концентраций (см. пояснения в тексте)

Графики распределений  $g_{aa}$  (см. рис. 7, *a*) для трех концентраций раствора очень похожи между собой. Первый пик ( $R_{aa} = 0,35$  нм) соответствует гидрофобным контактам молекул кофеина с участием метильных групп (рис. 8, *a*). Вторым ( $R_{aa} = 0,63$  нм) и слабо выраженный третий ( $R_{aa} = 0,95$  нм) пики отражают молекулярную "дискретность" растворителя и соответствуют конфигурациям, когда две молекулы кофеина разделены одной или более молекулами воды.

Четкий пик  $g_{cc}$  на 0,95 нм (см. рис. 7, *б*) говорит о наличии конфигураций, когда диски кофеиновых молекул стремятся расположиться в одной плоскости бок о бок, "соприкасаясь" метильными группами (см. рис. 8, *a* и *б*). При больших концентрациях такие молекулы формируют протяженные агрегаты в виде ветвящихся цепочек (см. рис. 8, *a*). Ненулевые значения функции  $g_{cc}$  слева от этого пика отражают существование непрерывного континуума конфигураций с постепенно уменьшающимся углом между плоскостями кофеиновых дисков вплоть до их складывания стопкой ( $R_{cc} = 0,34$  нм). Некоторые из этих расположений обладают повышенной устойчивостью (пики на 0,6 и 0,38 нм) и соответствуют Т-образной структуре (см. рис. 8, *в*) или стопочной укладке (см. рис. 8, *г*).

Используя методику анализа представительных выборок с помощью геометрического критерия комплексообразования (см. раздел о взаимодействии кофеин—ДНК), мы смогли не только оценить степень ассоциированности молекул кофеина в водном растворе, но и выявить наиболее вероятные конфигурации молекулярных агрегатов. Взяв в качестве критерия расстояние 0,49 нм (первый минимум функции  $g_{aa}$ ), мы получили, что "ассоциированными" оказывается лишь 6, 10 или 18 % (для 0,05, 0,1 и 0,2 М растворов соответственно) растворенных молекул кофеина. Причем стопочные (см. рис. 8, *г*) и Т-ассоциаты (см. рис. 8, *в*) составляют лишь 15 и 20 % соответственно от общего количества кофеиновых димеров. В остальных 65 % случаев молекулы кофеина располагаются бок о бок, осуществляя ван-дер-ваальсов контакт через метильные группы (см. рис. 8, *a* и *б*).

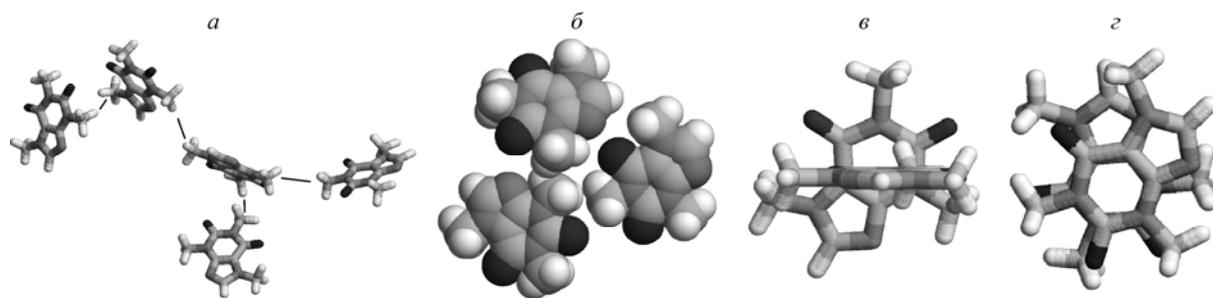


Рис. 8. Наиболее типичные конфигурации ассоциатов кофеина в водном растворе (молекулы окружающего раствора не показаны)

Выполненное нами компьютерное моделирование говорит о значительном преобладании "нестопочных" конфигураций кофеиновых ассоциатов в водном растворе. Такое положение может оказаться вполне соответствующим реальности. Ведь молекула кофеина содержит не только гидрофобные участки, но и три очень "водолюбивых" атома (два кислорода и азот N9). Расположение бок о бок позволяет минимизировать общую гидрофобную поверхность и сохранить, в отличие от стопочной конфигурации, доступность гидрофильных атомов для воды. Следует отметить также, что конфигурационные ансамбли для "стопочных", Т-образных и "бок о бок" ассоциатов в значительной мере перекрываются за счет большого количества переходных вариантов. Поэтому при изучении растворов кофеина методами спектроскопии и ЯМР необходимо рассматривать все конфигурационные возможности при интерпретации концентрационных и температурных зависимостей спектров.

Расчет константы равновесия  $K_p$  для реакции димеризации кофеина по приведенным выше оценкам степени их ассоциированности дает значения 0,3—0,25 л/моль. В то же время среди экспериментальных работ в научных журналах можно встретить значения для этой "константы", располагающиеся в диапазоне от 5 до 13 л/моль [19]. Причина столь значительного расхождения кроется, по-видимому, в расплывчатости самого понятия "ассоциат". Мы провели дополнительные расчеты, увеличив предельное значение межатомного расстояния в ассоциате до 0,775 нм, захватив, таким образом, и второй пик функции  $g_{aa}$ . Новые значения  $K_p$  (2,8, 3,5 и 6,1 л/моль для концентраций 0,05, 0,1 и 0,2 М соответственно) оказались в значительно лучшем согласии с экспериментальными данными. В связи с этим можно предположить, что большую долю ассоциатов, регистрируемых экспериментально, представляют собой относительно устойчивые агрегаты, в которых соседние молекулы кофеина разделены водной прослойкой.

#### СПИСОК ЛИТЕРАТУРЫ

1. Larsen R.W., Jasuja R., Hetzler R. et al. // *Biophys. J.* – 1996. – **70**. – P. 443.
2. Traganos F., Kapuscinski J., Darzynkiewicz Z. // *Cancer Res.* – 1991. – **51**. – P. 3682.
3. Davies D.B., Veselkov D.A., Djimant L.N., Veselkov A.N. // *Eur. Biophys. J.* – 2001. – **30**. – P. 354.
4. Lyles M.B., Cameron I.L. // *Cell Biol. Int.* – 2002. – **26**. – P. 145.
5. Шестопалова А.В. Гидратация кофеина и его взаимодействие с системой ДНК—вода: Дис...канд. физ.-мат. наук. – Тбилиси: Ин-т физики АН ГССР, 1987.
6. Falk M., Gil M., Iza N. // *Canad. J. Chem.* – 1990. – **68**. – P. 1293.
7. Теплухин А.В. // *Матем. моделир.* – 2004. – **16**, № 11. – С. 15.
8. Metropolis N.A., Rosenbluth A.W., Rosenbluth M.N. et al. // *J. Chem. Phys.* – 1953. – **21**. – P. 1087.
9. Snir M., Otto S., Huss-Lederman S. et al. *MPI: The Complete Reference.* – Boston: MIT Press, 1996.
10. Иванников В.П., Гайсарян С.С., Аветисян А.И., Падарян В.А. // *Программирование.* – 2006. – **32**, № 4. – С. 203.
11. Allen M.P., Tildesley D.J. *Computer simulation of liquids.* – N. Y.: Oxford University Press, 1987.
12. Heffelfinger G.S. // *Comput. Phys. Commun.* – 2000. – **128**. – P. 219.
13. Poltev V.I., Grokhlina T.I., Malenkov G.G. // *J. Biomol. Struct. Dyn.* – 1984. – **2**. – P. 413.
14. Sutor D.J. // *Acta Crystallogr.* – 1958. – **11**, N 7. – P. 453.
15. Теплухин А.В., Журкин В.Б., Джерниган Р., Полтев В.И. // *Молекуляр. биология.* – 1996. – **30**, № 1. – С. 121. (Engl. transl. *Mol. Biol. (Moscow)*. – 1996. – **30**, N 1, pt. 2. – P. 75–84.)
16. Teplukhin A.V., Malenkov G.G., Poltev V.I. // *J. Biomol. Struct. Dyn.* – 1998. – **16**. – P. 289.
17. Poltev V.I., Shulyupina N.V. // *Ibid.* – 1986. – **3**. – P. 739.
18. Berthod H., Pullman A. // *J. Chim. Phys.* – 1965. – **62**. – P. 942.
19. Kan L.-S., Borer P.N., Cheng D.M., Ts'o P.O.P. // *Biopolymers.* – 1980. – **19**. – P. 1641.