

# Мультипроцессорный расчет констант ассоциации низкомолекулярных лигандов в водном растворе методом Монте-Карло

А.В. Теплухин

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 различных концентраций и выявлены основные типы конфигураций молекулярных агрегатов.

## 1. Введение

В современных технологиях разработки лекарств широко применяются процедуры предварительного отбора, основанные на оценках степени сродства различных низкомолекулярных лигандов к активному центру заданного фермента, выполняемых с помощью компьютерного моделирования. Тем не менее, в большинстве случаев отобранные «кандидаты в лекарство» оказываются непригодными для практического применения. Одним из возможных путей повышения результативности отбора является использование более сложных моделей взаимодействия компонентов изучаемой системы.

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 взаимодействиям между молекулами лекарства (а также сайт-сайт и иным, в том числе и многочастичным корреляциям). В свою очередь, зная равновесные концентрации самого комплекса (пары сайт-лиганд) и его «не связавшихся» компонентов, можно вычислить константу равновесия соответствующей реакции димеризации, а затем использовать её как количественную характеристику степени сродства.

В качестве объекта исследования мы выбрали достаточно хорошо изученную экспериментальными методами систему ДНК+кофеин+вода [1,2]. Далее в статье будут представлены результаты изучения взаимодействия кофеина с ДНК в водном растворе при комнатной температуре, полученные для трех концентраций (разбавленный, концентрированный и перенасыщенный) методом компьютерного моделирования.

## 2. Методика моделирования

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 на 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 [3]. 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 вычисляемые данные обычно получают путем усреднения по наборам «мгновенных» конфигураций, генерируемых либо детерминист-

скими (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 рассчитать требуемые данные путем усреднения по соответствующим наборам мест связывания.

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-спектроскопии [1] 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.

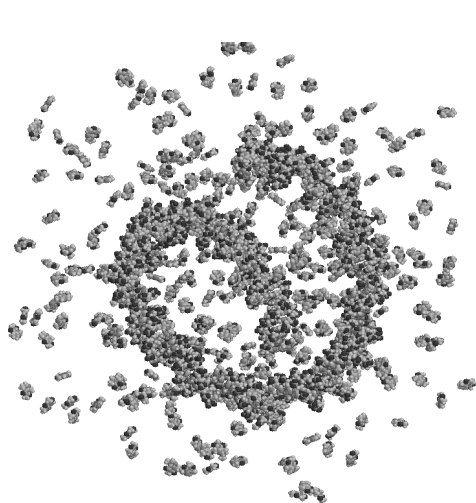
Характеристики систем, состоящих из нескольких сотен молекул воды, легко рассчитываются на обычных персональных компьютерах. Однако при моделировании водного окружения биополимеров приходится использовать элементарные ячейки очень большого объема. Количество молекул воды в таких моделях исчисляется сотнями 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 [4], позволяющий выполнять классическую процедуру Метрополиса [5] одновременно для нескольких (определяется числом процессоров, задействованных в решении данной задачи) молекул. Для этого моделируемая элементарная ячейка кубической формы представляется в виде совокупности кубов (далее Р-куб) меньшего объема, образующих регулярную пространственную решетку. Каждый из процессоров занят расчетами движения молекул «своего» Р-куба, периодически обмениваясь необходимыми данными с «соседними» (в пространстве моделируемой ячейки) процессорами о перемещениях молекул как внутри Р-кубов, так и между ними. Программное обеспечение, необходимое для моделирования водного раствора кофеина, было разработано на алгоритмическом языке Fortran77 с применением стандартных коммуникационных подпрограмм библиотеки MPI 1.2 [6].

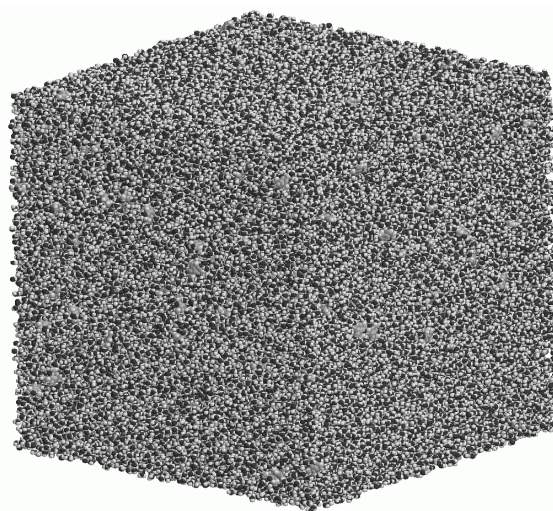
Мы использовали жесткую трехцентровую модель молекулы воды из работы [7]. Молекулы кофеина также рассматривались как твердые тела с молекулярной геометрией, описанной в работе [8]. Неподвижный двуспиральный фрагмент ДНК (GAGAAAAAGA)<sub>14</sub>:(TCTTTTCTC)<sub>14</sub>, состоящий из 140 нуклеотидных пар (14 витков двойной спирали в В-форме) и образующий петлю суперспирали (диаметр 13.4 нм, шаг 7.6 нм, если измерить 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 оси двойной спирали ДНК. Оригинальные координаты атомов декамера GAGAAAAAGA:TCTTTTCTC с параметрами для их преобразования были предоставлены В.Б. Журкиным. Важно отметить, что размер гликозидного желоба 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 -CH<sub>2</sub>-PO<sub>3</sub> группы были заменены метилами, а два 3'-концевых кислорода стали гидроксилами. Таким образом, исследуемый образец ДНК состоит из 8914 атомов.

Параметры атом-атомных потенциальных функций для взаимодействий вода-вода, вода-кофеин и вода-ДНК те же, что в работах [7,9], а для взаимодействий кофеин-кофеин и кофеин-ДНК – из работы [10]. Величины парциальных атомных зарядов рассчитывались с помощью простых методов MO-LCAO и Хюккеля.

Было проведено три компьютерных «эксперимента» по изучению взаимодействия кофеин-ДНК в разбавленном, насыщенном и перенасыщенном водных растворах (концентрации кофеина, соответственно, 0.05М, 0.1М и 0.2М) при температуре 300К. Для этого при проведении расчетов использовались три набора по 216 процессоров суперкомпьютера МВС100К в МСЦ РАН ([www.jssc.ru](http://www.jssc.ru)). Стартовые конфигурации, подготовленные независимо для каждой из трех систем, содержали, соответственно: 175 и 192869, или 350 и 187224, или 700 кофеиновых и 181730 водных молекул, 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. 1.



**Рис. 1а.** Расположение фрагмента ДНК и 350 молекул кофеина в элементарной ячейке. Одна из мгновенных конфигураций, молекулы воды не показаны



**Рис. 1б.** Общий вид моделируемой системы. Показаны все атомы

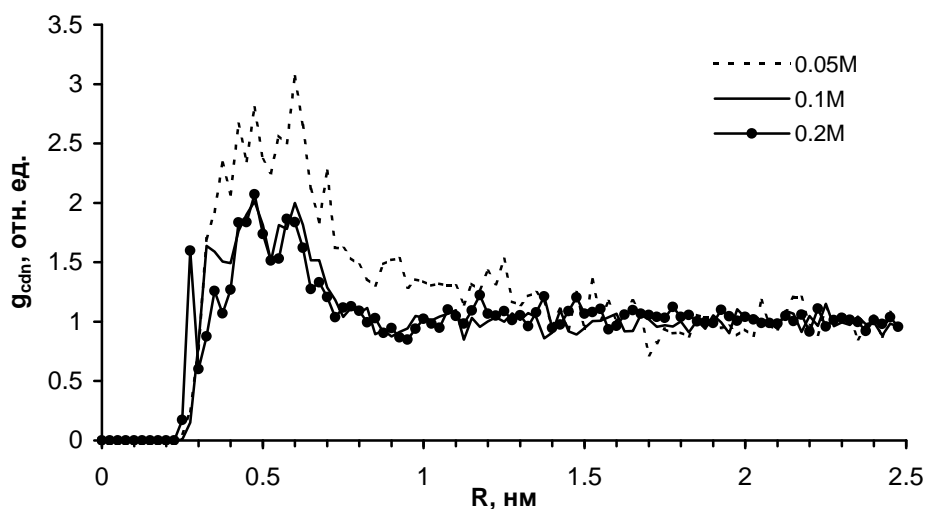
Чтобы устранить влияние произвольности выбора стартовой конфигурации на результаты, первоначальный участок марковской цепи, соответствующий 5 миллионам испытаний в расчете на каждую молекулу, исключался из рассмотрения. Расчет и анализ структурных и энергетических характеристик выполнялся далее по ансамблям конфигураций, соответствующим не менее 30 миллионам испытаний, выполненным с каждой молекулой.

### 3. Результаты расчетов

При изучении локального упорядочения в простых жидкостях часто используют аппарат радиальных функций распределения (РФР). В нашем случае (для исследования пространственных корреляций в молекулярной 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).

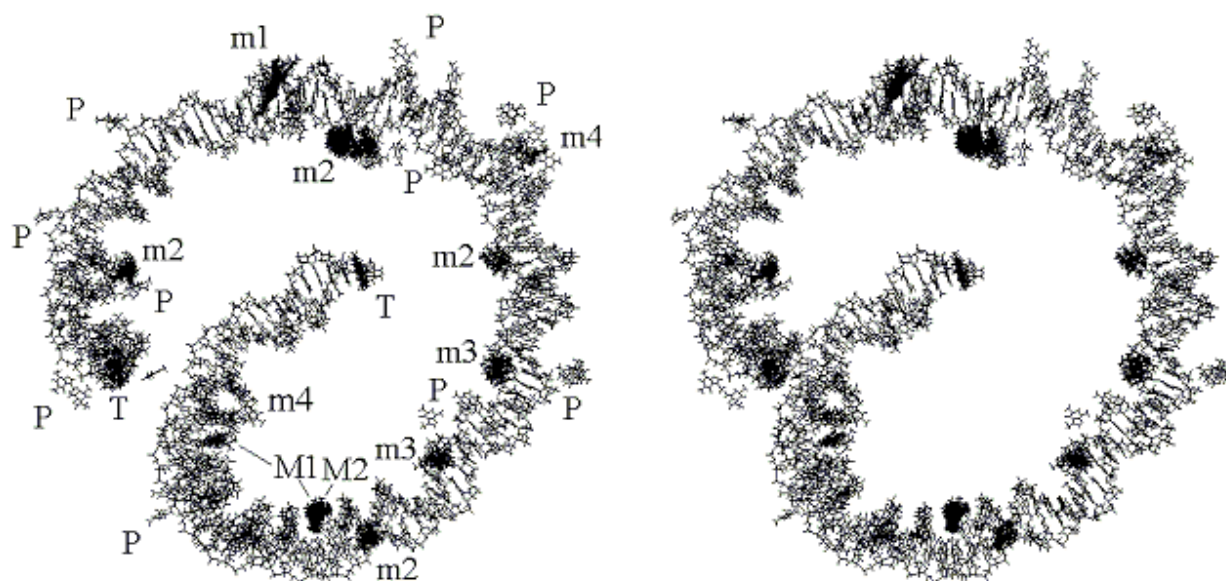
На Рис. 2 представлены РФР кофеина вокруг ДНК, рассчитанные вышеописанным способом по статистически значимой выборке «мгновенных» молекулярных конфигураций, генерируемых в ходе процедуры Метрополиса [5]. Прежде всего, на этом рисунке выделяется сильно заселенная (по сравнению с равномерным распределением по объему) область вокруг ДНК толщиной около 0.4 нм, образующуюся 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, об-

разовавшие 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.



**Рис. 2.**  $g_{cdn}$  – относительная плотность радиального распределения молекул кофеина вокруг ДНК. Данные для трех концентраций раствора

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. 2, только молекулы кофеинов, 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 100000-го из 10 миллионов циклов процедуры Метрополиса по всем молекулам и дальнейшей «фильтрацией» с помощью геометрического критерия близости. Каждый такой набор файлов представляет собой выборку, содержащую не менее 100 вариантов «кофеиновых оболочек» ДНК. На Рис. 3 представлен общий вид выборки для раствора 0.1M.



**Рис. 3.** Конфигурационные ансамбли молекулярных комплексов кофеин-ДНК в водном растворе (стереопара, молекулы окружающего раствора не показаны)

В ходе анализа полученных выборок удалось не только выявить основные типы комплексов кофеин-ДНК, образующихся в водной среде, и установить 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. 4). Такие конфигурации реализуются в зауженных участках гликозидного желоба (5'-AAAGA) как с образованием водородной связи между аминогруппой гуанина и гидрофильным атомом кофеина – тип m1, так и без неё – тип m2. Причем в очень концентрированном растворе преобладает первый тип, а в более разбавленных растворах чаще встречается конфигурация без Н-связи (см. таблицу 1).

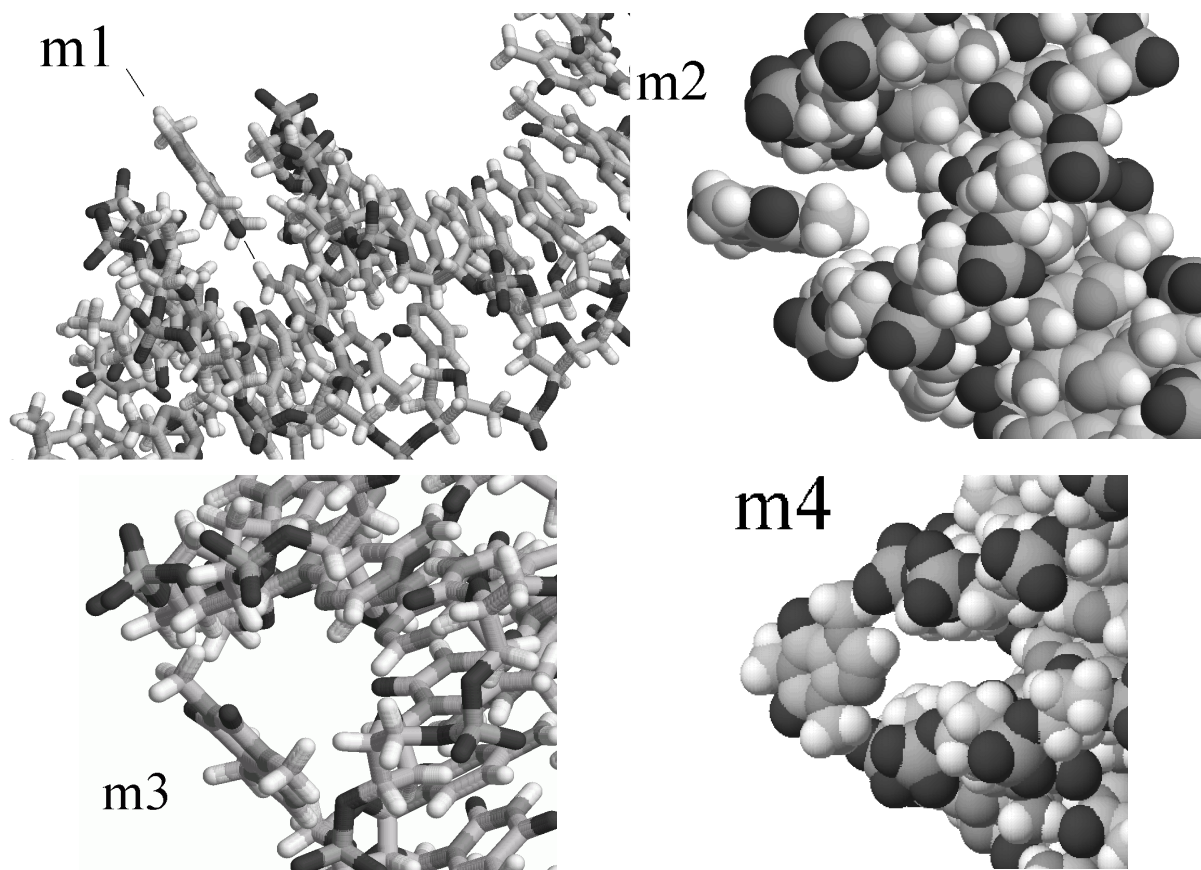


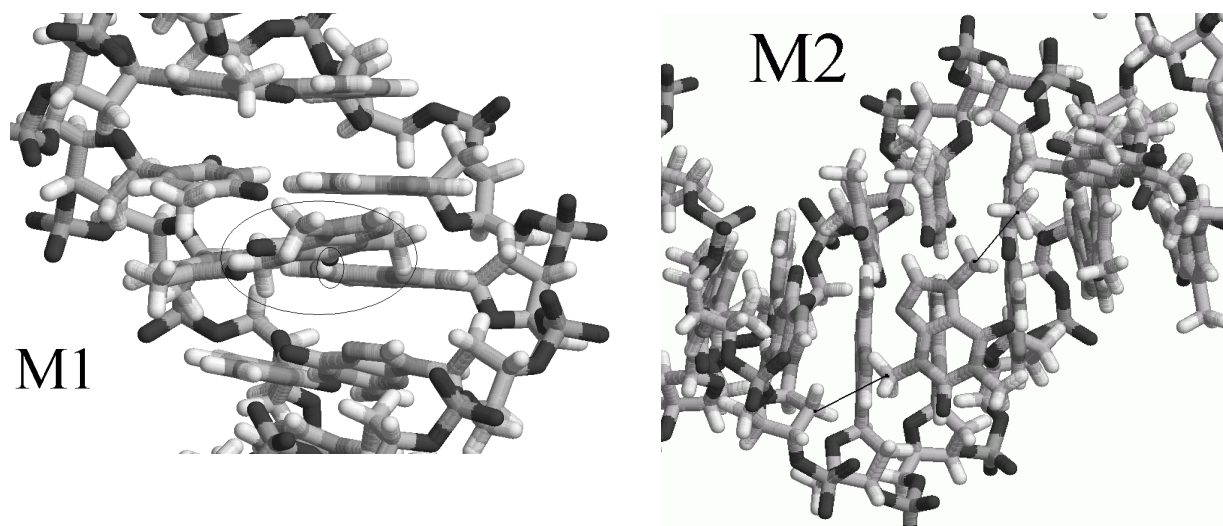
Рис. 4. Конфигурации основных типов комплексов кофеин-ДНК в гликозидном желобе

В следующий класс следует объединить комплексы типов m3, m4 и P (Рис. 3 и 4), 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 полинуклеотидных цепей. 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. таблицу 1). «Ареалом» данного класса являются участки 5'-GAAA гликозидного желоба, имеющие относительно большую ширину.

Таблица 1. Вероятность образования комплексов кофеин-ДНК в расчете на один сегмент GAGAAAAGA:TCTTTTCTC

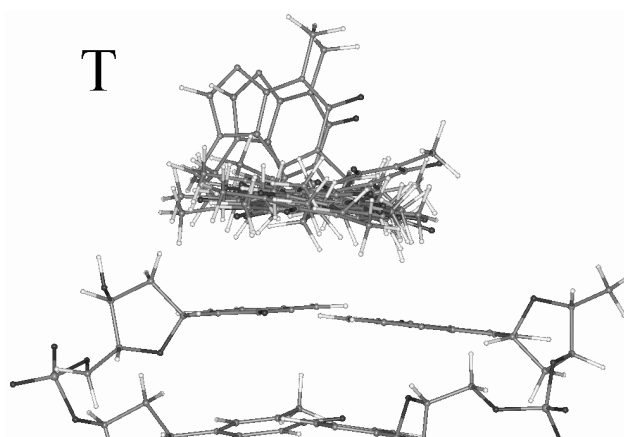
Концентрация раствора	Типы комплексов кофеин-ДНК							
	m1	m2	m3	m4	M1	M2	T	P
0.05M	0.01	0.02	0.01	0.08	<0.01	0.01	<0.01	0.02
0.1 M	0.01	0.08	0.01	0.04	0.01	0.03	0.13	0.08
0.2 M	0.31	0.03	0.01	0.03	0.04	0.05	0.23	0.21

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 Н-связи между протон-акцепторным атомом кофеина и NH<sub>2</sub>-группой аденина (Рис. 5, овалы) или цитозина. Структуры комплексов M2 довольно разнообразны и могут осуществляться либо путем установления CH<sub>3</sub>-группами кофеина гидрофобных контактов с CH<sub>3</sub>-группой тимина и/или C2'-H<sub>2</sub>-группой дезоксирибозы, либо через формирование своеобразного π-комплекса, когда химическая связь атома C5 тимина с его CH<sub>3</sub>-группой направлена в центр молекулы кофеина перпендикулярно плоскости ароматических колец.



**Рис. 5.** Конфигурации основных типов комплексов кофеин-ДНК в негликозидном желобе

Конфигурации, возникающие при взаимодействии молекул кофеина с концевыми участками двойной спирали, составляют особый класс. Чаще всего формируются комплексы стопочной или Т-образной структуры (см. выборку «Т» на Рис. 6), а также «переходные» между ними конфигурации. При увеличении концентрации растворов вероятность появления таких комплексов быстро возрастает (см. таблицу 1). Сравнительно редко (с вероятностью 0.02-0.05) могут образовываться Н-связи между гидрофильными атомами кофеина и протонами 3'-концевых гидроксидов.

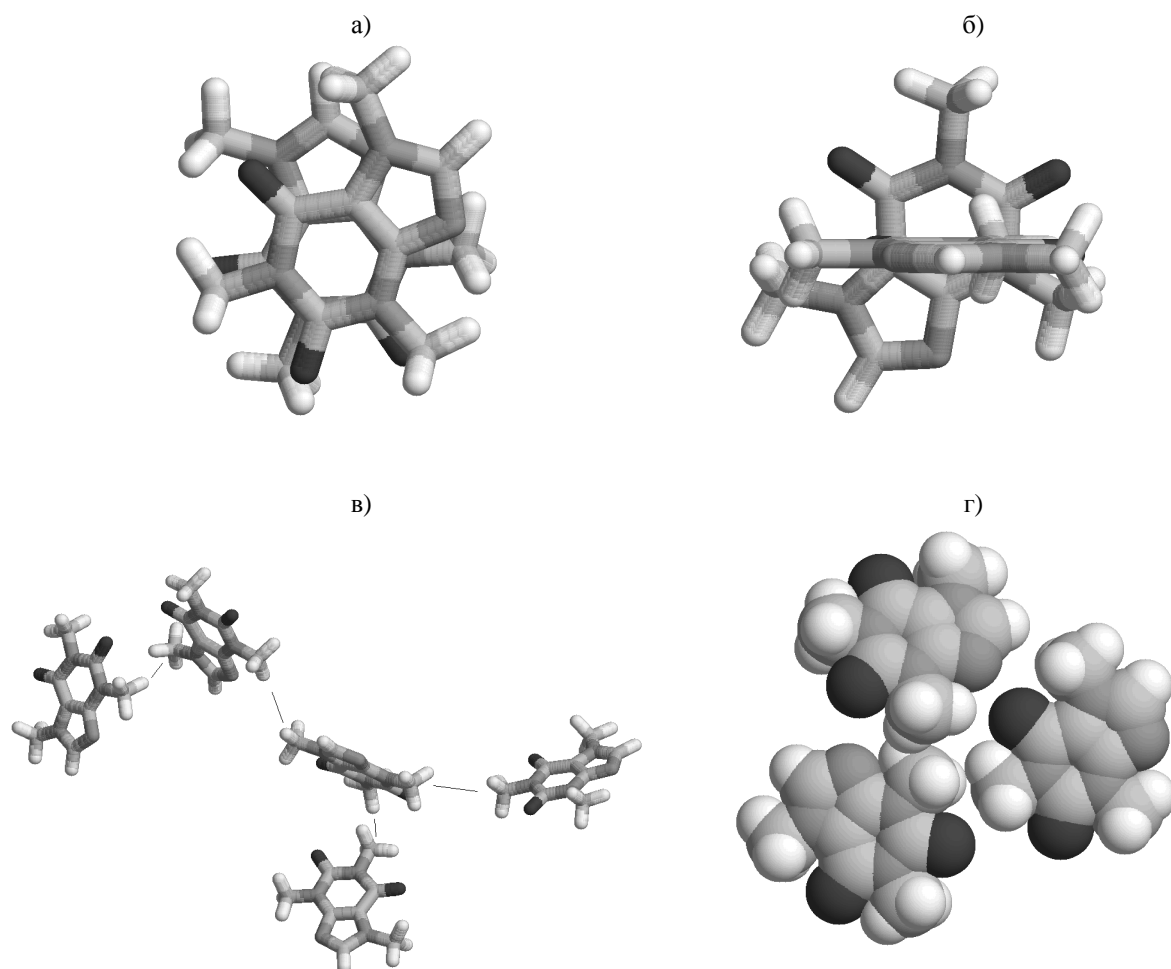


**Рис. 6.** Выборка «Т»-конфигураций молекулы кофеина у концевого участка ДНК

Экспериментальные данные, имеющиеся в научной литературе, дают слишком скудную, но очень важную структурную информацию, подтверждающую наши выводы о взаимодействии молекул кофеина с ДНК. В частности отмечается [1], что кофеин взаимодействует с сахаро-фосфатным остовом и NH<sub>2</sub>-группами азотистых оснований.

Аналогичный подход (РФР, выборки) можно применить и для изучения самоассоциации кофеина. Взяв в качестве критерия расстояние 0.49 нм (первый минимум РФР) мы получили, что «ассоциированными» оказывается лишь 6%, 10% или 18% (для 0.05М, 0.1М и 0.2М растворов) растворенных молекул кофеина. Причем «стопочные» (Рис. 7а) и «ортогональные» ассоциаты (Рис. 7б) составляют, соответственно, 15% и 20% от общего количества кофеиновых димеров. В остальных 65% случаев молекулы кофеина располагаются бок о бок, осуществляя контакт через СН<sub>3</sub>-группы (Рис. 7в и 7г).

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 [11]. 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 и второй пик РФР. Новые значения  $K_p$  (2.8, 3.5 и 6.1 л/моль для концентраций 0.05М, 0.1М и 0.2М) – 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.

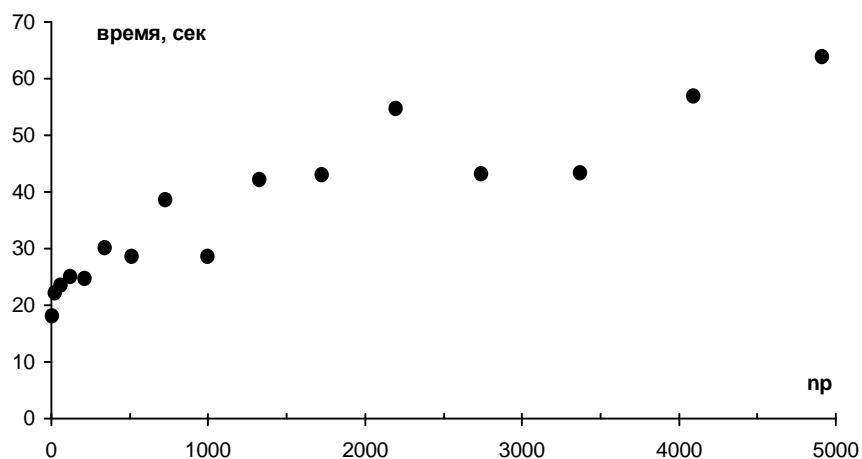


**Рис. 7.** Наиболее типичные конфигурации ассоциатов кофеина в водном растворе (молекулы окружающего раствора не показаны)

## 4. Заключение

Представленный подход к моделированию молекулярных систем не имеет формальных ограничений ни на размер элементарной ячейки, ни на количество атомов в исследуемом объекте. Для иллюстрации эффективности нашего подхода была выполнена (МВС100К МСЦ РАН) серия расчетов систем различного объема (кубической формы), содержащих только молекулы воды. В этом случае  $k$ -я система может быть представлена в виде пространственной решетки из  $(k+1)^3$  Р-кубов одинакового размера ( $k$  – натуральные числа от 1 до 16). Каждый Р-куб содержит, в среднем, 230 молекул воды. Для расчета  $k$ -й системы используется столько же процессоров (ядер) сколько в ней Р-кубов. Так, система минимального размера (2x2x2 Р-кубов, 5514 атомов) рассчитывалась на 8 процессорных ядрах, а самая большая (17x17x17 Р-кубов, 3386673 атома) – на  $17^3$  (т.е. 4913) ядрах. На Рис. 8 приведено время выполнения 1000 «всемолекулярных» циклов процедуры Метрополиса для ряда «водных кубов» различного объема.

Моделирование молекулярных систем размерами в сотни нанометров становится вполне осуществимым, благодаря распределенным параллельным вычислениям на современных многопроцессорных суперкомпьютерах. Все это в свою очередь позволяет вывести технологии компьютерной разработки лекарств на новый уровень.



**Рис. 8.** Время, требуемое для расчета «водных кубов», расположенных в порядке возрастания их размера. Здесь **np** – одновременно и количество Р-кубов, составляющих конкретный водный куб, и число процессорных ядер, задействованных при расчете.

## Литература

1. Шестопалова А.В. Гидратация кофеина и его взаимодействие с системой ДНК-вода // Дис....канд. физ.-мат. наук. – Тбилиси: Ин-т физики АН ГССР, 1987.
2. Davies D.B., Veselkov D.A., Djimant L.N., Veselkov A.N. // Eur. Biophys. J. –2001. –Vol. 30. – P. 354-366.
3. Falk M., Gil M., Iza N. // Can. J. Chem. –1990. –Vol. 68. –P. 1293-1299.
4. Теплухин А.В. // Матем. моделирование. –2004. –Том 16, № 11. –С. 15-24.
5. Metropolis N.A., Rosenbluth A.W., Rosenbluth M.N., Teller A.H., Teller E. // J. Chem. Phys. – 1953. –Vol. 21. –P. 1087-1092.
6. Snir M., Otto S., Huss-Lederman S., Walker D., Dongarra J. // MPI: The Complete Reference. – Boston: MIT Press, 1996.
7. Poltev V.I., Grokhлина T.I., Malenkov G.G. // J.Biomol.Struct.Dyn. –1984. –Vol. 2. –P. 413-429.
8. Sutor D.J. // Acta Cryst. –1958. –Vol. 11. –P. 453-458.
9. Teplukhin A.V., Malenkov G.G., Poltev V.I. // J.Biomol.Struct.Dyn. –1998. –Vol. 16. –P. 289-300.
10. Poltev V.I., Shulyupina N.V. // J.Biomol.Struct.Dyn. –1986. –Vol.3. –P. 739-765.
11. Kan L.-S., Borer P.N., Cheng D.M., Ts'o P.O.P. // Biopolymers. –1980. –Vol. 19. –P. 1641-1658.