

Московский Государственный Университет им. М. В. Ломоносова
Механико-математический факультет
Кафедра дифференциальной геометрии и приложений



Дипломная работа

студентки 507 группы

Киреевой Альбины Александровны.

Компьютерное моделирование топологии модельного ансамбля
молекул ДНК.

Научный руководитель:

д. ф.-м. н., профессор Голо Войслав Любомирович

Москва, 2010

Введение.

Дезоксирибонуклеиновая кислота (ДНК) — один из двух типов нуклеиновых кислот, обеспечивающих хранение, передачу из поколения в поколение и реализацию генетической программы развития и функционирования живых организмов. Основная роль ДНК в клетках — долговременное хранение информации о структуре РНК и белков. В клетках эукариот (например, животных или растений) ДНК находится в ядре клетки в составе хромосом, а также в некоторых клеточных органоидах (митохондриях и пластидах). С химической точки зрения, ДНК — это длинная полимерная молекула, состоящая из повторяющихся блоков, нуклеотидов. Каждый нуклеотид состоит из азотистого основания, сахара (дезоксирибозы) и фосфатной группы (см. рис.1). Связи между нуклеотидами в цепи образуются за счёт дезоксирибозы и фосфатной группы. В подавляющем большинстве случаев (кроме некоторых вирусов, содержащих одноцепочечную ДНК) макромолекула ДНК состоит из двух цепей, ориентированных азотистыми основаниями друг к другу. Структура молекулы ДНК получила название «двойной спирали». В ДНК встречается четыре вида азотистых оснований (аденин, гуанин, тимин и цитозин). Азотистые основания одной из цепей соединены с азотистыми основаниями другой цепи водородными связями согласно принципу комплементарности: аденин соединяется только с тимином, гуанин — только с цитозином. Полимер ДНК обладает довольно сложной структурой. Нуклеотиды соединены между собой ковалентно в длинные полинуклеотидные цепи. Эти цепи в подавляющем большинстве случаев (кроме некоторых вирусов, обладающих одноцепочечными ДНК-геномами) попарно объединяются при помощи водородных связей в структуру, получившую название двойной спирали. Остов каждой из цепей состоит из чередующихся фосфатов и сахаров. [1].

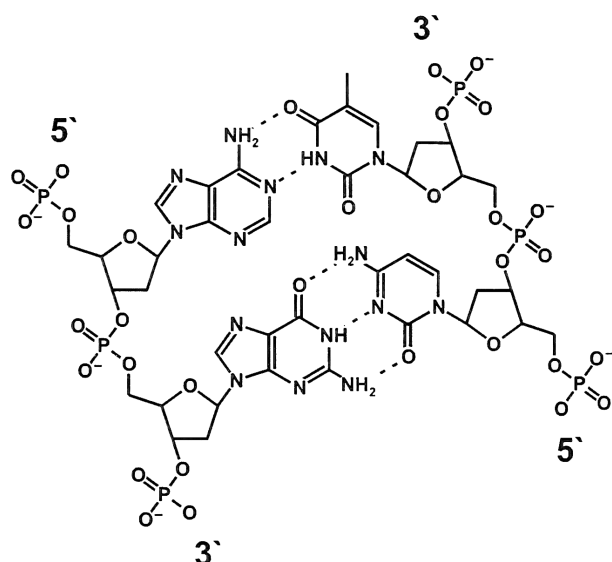


Рис. 1: Фрагмент молекулы ДНК.

Как уже было сказано выше, у подавляющего большинства живых организмов ДНК состоит не из одной, а из двух полинуклеотидных цепей. Эти две длинные цепи закручены одна вокруг другой в виде двойной спирали, стабилизированной водородными связями, образующимися между обращёнными друг к другу азотистыми основаниями входящих в неё цепей. В природе эта спираль, чаще всего, правозакрученная. Ширина двойной спирали составляет от 22 до 24 Å, или 2,2 – 2,4 нм, расстояние между соседними парами оснований 3,4 Å. ДНК обладает длиной персистентности. Экспериментально выяснено, что отрезки ДНК длиной 100–150 пар оснований (500 Å) образуют жидкокристаллические фазы при определенных параметрах плотности и температуры [2][3].

Жидкий кристалл - состояние вещества, сочетающее в себе свойства кристалла (пространственная упорядоченность в одном или двух направлениях) и жидкости. Выделяют следующие виды жидких кристаллов в зависимости от ориентационного порядка: нематические жидкие кристаллы (наличествует дальний ориентационный порядок, все молекулы расположены примерно вдоль одной оси), холестерические жидкие кристаллы (от слоя к слою происходит поворот оси, вдоль которой ориентируются молекулы) и смектики (молекулы расположены слоями). [5][6].

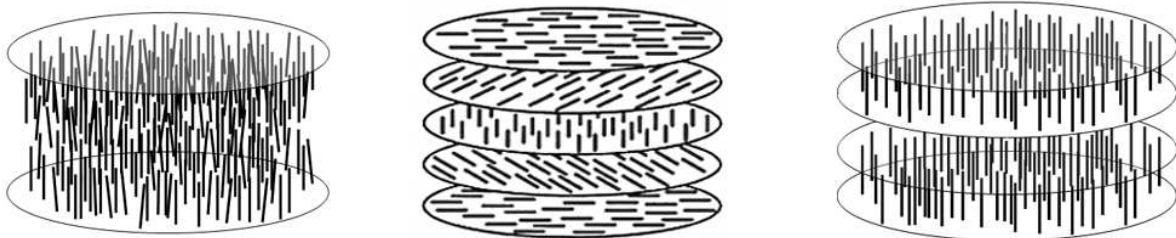


Рис. 2: Нематик, холестерик, смектик

В экспериментах с отрезками ДНК, описанных в работах [2][3], наблюдаются различные фазы, зависящие от концентрации молекул в растворе и температуры. Для расстояний между осями молекул от 28 до 35 Å, отрезки молекул ДНК в растворе образуют твердую гексагональную фазу. Для расстояний более 45 Å появляется холестерическая структура с шагом холестерической спирали, варьирующемся в пределах $(1.5 \div 3.5)10^3 \text{ Å}$. Принято считать [10], что электростатическое взаимодействие должно играть существенную роль в образовании холестерической фазы, поскольку молекулы ДНК сильно заряжены.

Поскольку система очень сложна и невозможно, даже применяя численные методы и большие компьютерные мощности, точно учесть все взаимодействия, необходим модельный подход. Для разных целей возможны разные модели ДНК. ДНК - сильно заряженная молекула, поэтому предполагается, что природа взаимодействия определяется электростатическими силами. Реализовывать эту гипотезу в модели можно различными способами. Распространенный подход основан на представлении молекулы ДНК как цилиндра с непрерывным винтообразным распределением зарядов [10]. Однако, интересно также рассмотреть случай дискретного распределения зарядов. В настоящей работе рассматривается подход, основанный на представлении о дискретном распределении зарядов с учетом геликоидальной симметрии. Задачей данной работы является моделирование ансамбля из отрезков молекул ДНК, что является существенным продвижением относительно сделанных ранее работ [8][9], в которых исследовался исключительно парный потенциал.

Модель.

В настоящей работе отрезок молекулы ДНК моделируется одномерной решеткой, с дискретно расположенными зарядами, соответствующими фосфатным группам, и диполями, соответствующими парам оснований.



Рис. 3: Схематическое изображение модели одной молекулы ДНК: крупными точками обозначены заряды, стрелки обозначают дипольные моменты пар оснований. Диполи перпендикулярны оси молекулы, относительный угол поворота между диполями - $\pi/5$

Расстояние между соседними диполями берется равным 3.4\AA , что копирует пространственную конформацию диполей пар оснований. Диполи располагаются перпендикулярно оси стержня, угол поворота следующего диполя относительно предыдущего равен $\pi/5$, чтобы учесть спираль. Заряды считаются лежащими точно посередине между соседними диполями. Все величины для упрощения счета переводятся в систему единиц ДНК:

- $M = 10^{-22} \text{ gr}$ единица массы, порядок массы пары оснований;
- $L = 3 \times 10^{-8} \text{ cm}$ единица длины, расстояние между соседними парами оснований;
- $T = 10^{-13} \text{ sec}$ единица времени, согласно верхней границе звуковых колебаний в ДНК.

Исходя из такой перенормировки, величины заряда электрона, дипольного момента в 1 Debye и энергии в 1 erg примут следующие значения:

- $1 e = 1$;
- $1 \text{ Debye} = 0.06$;
- $1 \text{ erg} = 10^{11}$.

Согласно данным работы [4], электрические дипольные моменты пар оснований измеряются в пределах $3 \div 9 \text{ Debye}$, что в системе единиц ДНК соответствует $0.2 \div 0.6$.

Энергия взаимодействия двух молекул считается как сумма элементарных энергий взаимодействия:

$$U = u_{dd} + u_{cd} + u_{dc} + u_{cc},$$

Где u_{dd} – энергия взаимодействия диполя первой молекулы с диполем второй, u_{cd} – энергия взаимодействия заряда первой молекулы с диполем второй, u_{dc} – энергия взаимодействия диполя первой с зарядом второй, u_{cc} – энергия взаимодействия зарядов первой и второй. Взаимодействия вычисляются по следующим формулам [7]:

$$u_{dd} = \left(\frac{1}{\rho^3} (\vec{p}, \vec{p}') - 3 \frac{(\vec{p}, (\vec{r} - \vec{r}')) (\vec{p}', (\vec{r} - \vec{r}'))}{\rho^5} \right)$$

$$u_{dc} = Q' \frac{(\vec{p}, (\vec{r} - \vec{r}'))}{\rho^3}$$

$$u_{cd} = Q \frac{(\vec{p}', (\vec{r} - \vec{r}'))}{\rho^3}$$

$$u_{cc} = \frac{QQ'}{\rho}$$

$$\rho = | \vec{r} - \vec{r}' |$$

Где \vec{r}, \vec{r}' - вектора из начала координат в соответствующие узлы решетки молекулы, Q, Q' - заряды, заданные соответствующими фосфатными группами.

Важным вопросом является экранировка. В настоящее время пока не выяснено, насколько применима теория Дебая-Хюккеля для молекул ДНК в растворе. Поскольку модель, рассматриваемая в данной работе, является очень грубой, экранировка учитывается лишь косвенно, за счет изменения величины зарядов.

В данной работе так же не учитывается диэлектрическая проницаемость среды.

В работах [2][3] сообщается об экспериментах с жидкокристаллическими фазами ДНК для случая отрезков длины порядка 500\AA (150 – 152 пары оснований). Но при моделировании ансамбля молекул такой длины возникают вычислительные трудности - компьютерный счет занимает длительное время. Поэтому исследуются конформации специального вида для ансамбля более коротких молекул - длины 32 пары оснований. Можно предположить, что проведенные вычисления позволяют качественно интерпретировать результаты [2][3].

Энергия взаимодействия ансамбля молекул считается как сумма энергий попарных взаимодействий.

Численное моделирование

Основной задачей работы является исследование ансамбля молекул ДНК при влиянии температуры. Во всей полноте численное моделирование раствора молекул ДНК в настоящее время едва ли возможно по причине недостаточной мощности компьютеров и отсутствия соответствующих алгоритмических методов. Одной из существенных проблем является также проверка результатов. Ввиду этого мы ограничиваемся частным случаем, который позволяет получить качественную информацию.

Для моделирования выбрана следующая система конфигураций: молекулы лежат одна над другой, их центры фиксированы и лежат на одной прямой. Все молекулы лежат в плоскостях, перпендикулярных этой прямой. Расстояние между плоскостями соседних молекул одинаково и фиксировано (в данной работе это расстояние считалось равным 4 в безразмерных единицах, в пересчете 12\AA) и является параметром, отражающим концентрацию молекул в объеме. На рисунке 4 изображена идеальная холестерическая конфигурация. Если электростатические силы играют основную роль при формировании холестерических фаз, то выбранная таким образом система предположительно будет это отражать, и можно будет сделать качественные выводы об экспериментах [2][3].

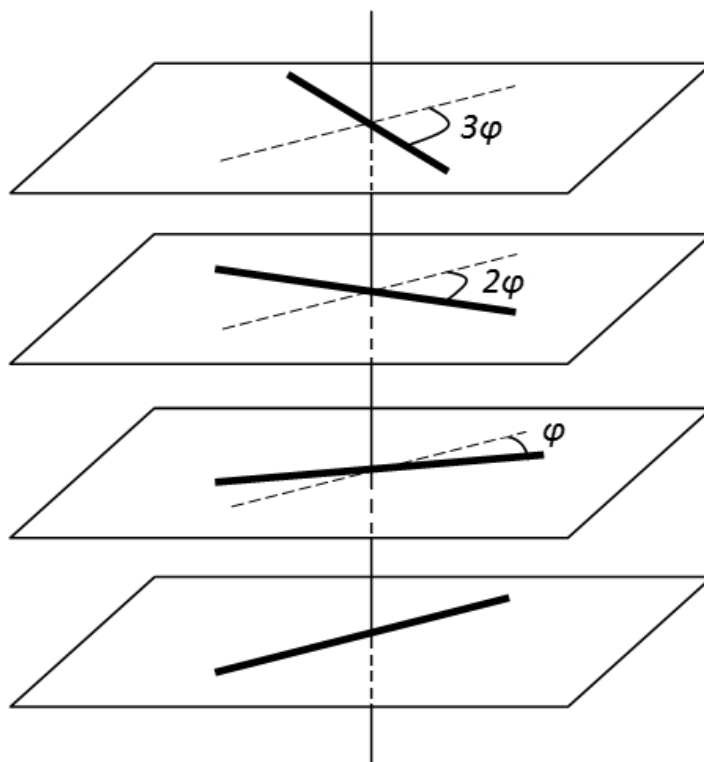


Рис. 4: На рисунке схематично показаны молекулы, находящиеся в идеальной холестерической конфигурации. Молекулы находятся в плоскостях, перпендикулярных общей прямой, расстояние между плоскостями соседних молекул фиксировано и одинаково. Обозначен угол отклонения относительно первой молекулы (пунктирная линия показывает параллельно пренесенное положение первой молекулы). В изображенной конфигурации относительные углы между соседними молекулами одинаковы.

Выбор базовой конфигурации.

Для дальнейшего исследования нужно выбрать начальную конфигурацию, которая потом будет подвергаться влиянию температуры в термостате Гиббса (при фиксированном объеме, фиксированной температуре и числе молекул). Кроме того, нужно определить какой эффективный заряд в модели может привести к результатам, качественно похожим на эксперименты с жидкокристаллическими фазами ДНК. С этой целью были сконструированы различные положения системы и подсчитаны энергии для получившихся различных положений.

При разных значениях заряда построены профили энергии для систем, в которых молекулы расположены по "винту": угол поворота следующей относительно предыдущей одинаков (см. рис. 4). На рисунках 5, 6, 7, 8 для различных значений зарядов изображены профили энергии сконструированных "винтовых" конфигураций в зависимости от относительного угла поворота соседних молекул.

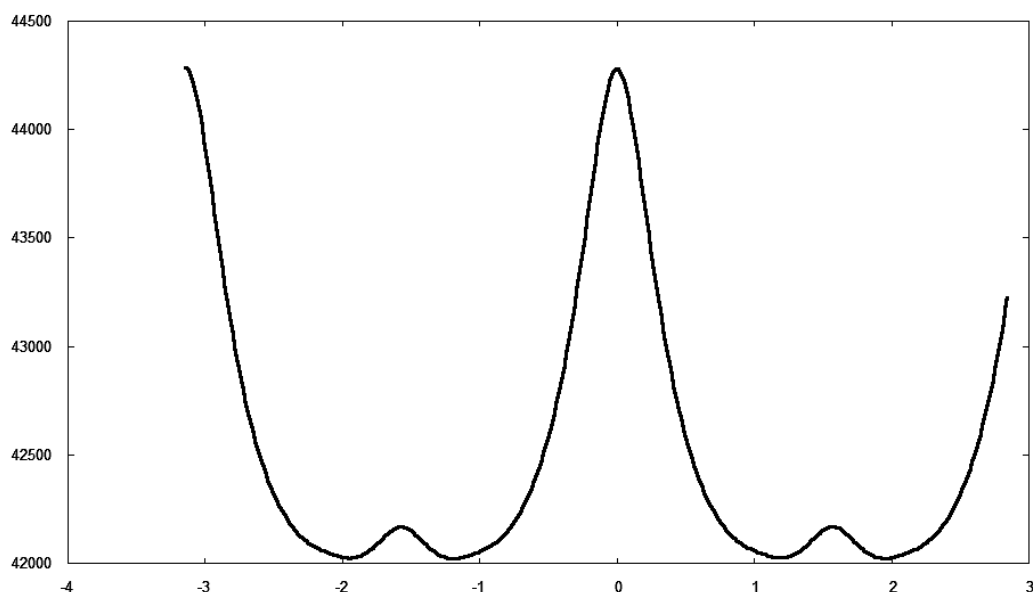


Рис. 5: Профиль энергии сконструированных систем в зависимости от угла поворота между соседними молекулами. По оси абсцисс - угол поворота между соседними молекулами. По оси ординат - значение энергии системы. **Заряды 1**

На графике (рис. 4) видны глобальные минимумы в областях углов $2\pi/5$ и $3\pi/5$, $-2\pi/5$ и $-3\pi/5$. Из графика следует, что при значениях заряда 1 при обсчете этой системы в термостате маловероятно сохранение конфигурации с малым относительным углом поворота. Поскольку нас интересуют именно малые относительные углы, такие заряды пока подробно не рассматривались.

С уменьшением заряда происходит сдвиг минимума в область углов 0.1 (см. рис.6 заряды 0.08 а также рис.7 заряды 0.05).

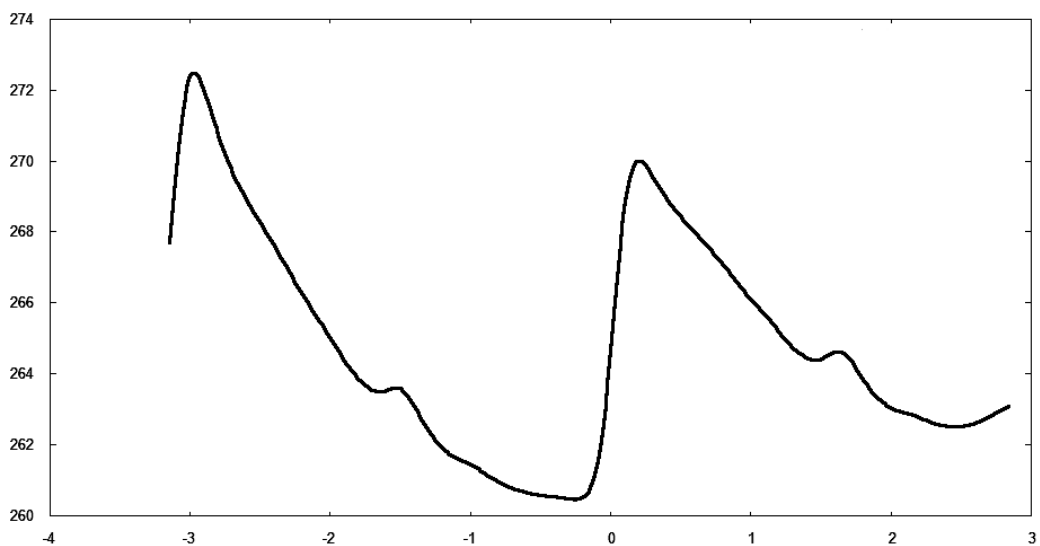


Рис. 6: Профиль энергии сконструированных систем в зависимости от угла поворота между соседними молекулами. По оси абсцисс - угол поворота между соседними молекулами. По оси ординат - значение энергии системы. **Заряды 0,08**

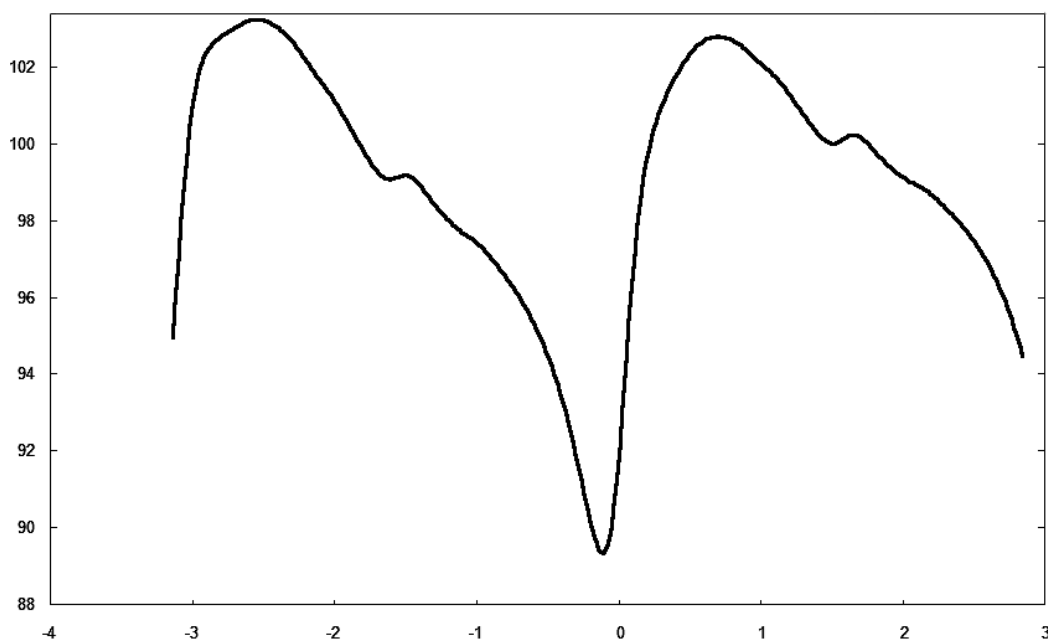


Рис. 7. Профиль энергии сконструированных систем в зависимости от угла поворота между соседними молекулами. По оси абсцисс - угол поворота между соседними молекулами. По оси ординат - значение энергии системы. **Заряды 0,05**

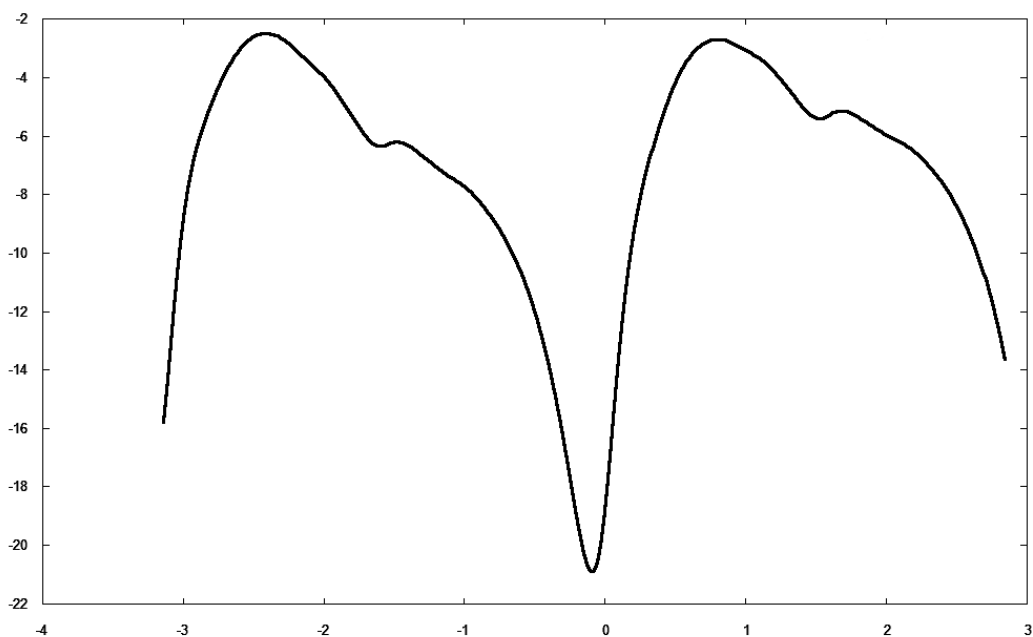


Рис. 8. Профиль энергии системы в зависимости от угла поворота между соседними молекулами без проведения метода Монте-Карло. По оси абсцисс - угол поворота между соседними молекулами. По оси ординат - значение энергии системы. Заряды 0

Все графики выше были построены без применения Монте-Карло для выяснения того, какие эффективные заряды могут после применения метода Монте-Карло дать холестерик. Из графиков видно, что с уменьшением эффективных зарядов в модели происходит углубление минимума в области малых углов. В частности, если рассматривать модель без зарядов, то минимум в этой области становится как на рис. 8.

Исходя из построенных графиков для сконструированных состояний, для дальнейшего моделирования выбраны заряды 0.05. Начальное положение системы таково: идеальный холестерик с относительным углом закрутки 0.1.

Для моделирования термостата используется метод Монте-Карло и алгоритм Метрополиса. Описание метода находится в приложении.

Важную роль при построении марковской цепи играет генератор следующего состояния. В этой задаче использовались два типа генераторов:

- Каждая молекула поворачивается на случайный малый угол вокруг общей оси и на случайный малый угол вокруг своей оси;
- Случайно выбирается молекула и поворачивается на случайный малый угол вокруг общей оси и на случайный малый угол вокруг своей оси.

Работа с первым генератором приводила к длительному счету, подбор подходящего состояния осуществлялся в течение долгого времени. Поэтому в ходе работы предпочтение было отдано второму генератору марковской цепи. Благодаря такому изменению стало возможным за то же время производить большее число шагов метода Монте-Карло.

К работе прилагается код программы с комментариями (см. приложение 3).

После проведения моделирования термостата методом Монте-Карло при различных температурах получены следующие результаты.

На рисунке 9 показан абсолютный угол отклонения направления молекулы от начального положения. Из графика следует, что угол отклонения в среднем растет, но наблюдаются отклонения.

На рисунке 10 показан относительный поворот $\Delta\varphi_n - \Delta\varphi_{n+1}$, характеризующий отклонения от холестерической структуры. Колебания относительного поворота порядка 0.02.

Таким образом, выяснено, что изначально заданная структура холестерика под воздействием температуры нарушается, но не разрушается полностью. Энергия в полученном состоянии отличается от начальной энергии на $1,7 \cdot 10^{-2}\%$

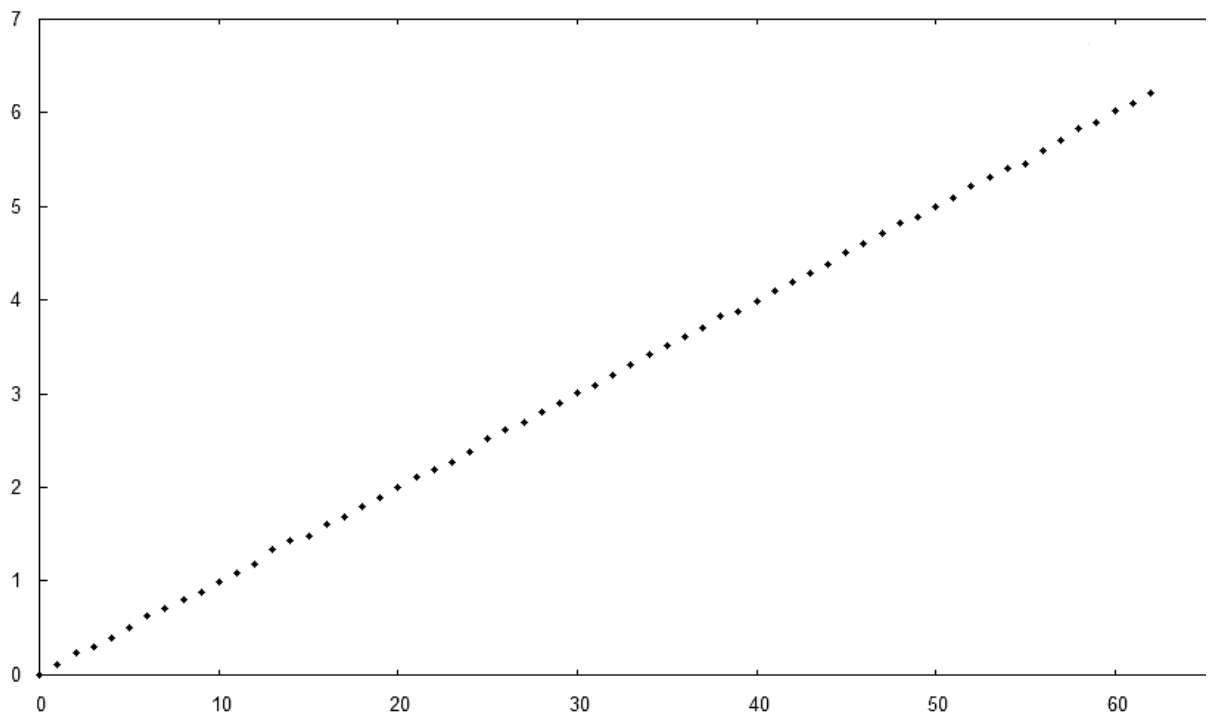


Рис. 9. Абсолютный угол отклонения молекулы от направления $(1, 0, 0)$. По оси абсцисс - последовательные номера молекул. По оси ординат - значение угла в радианах. Температура 300 К.

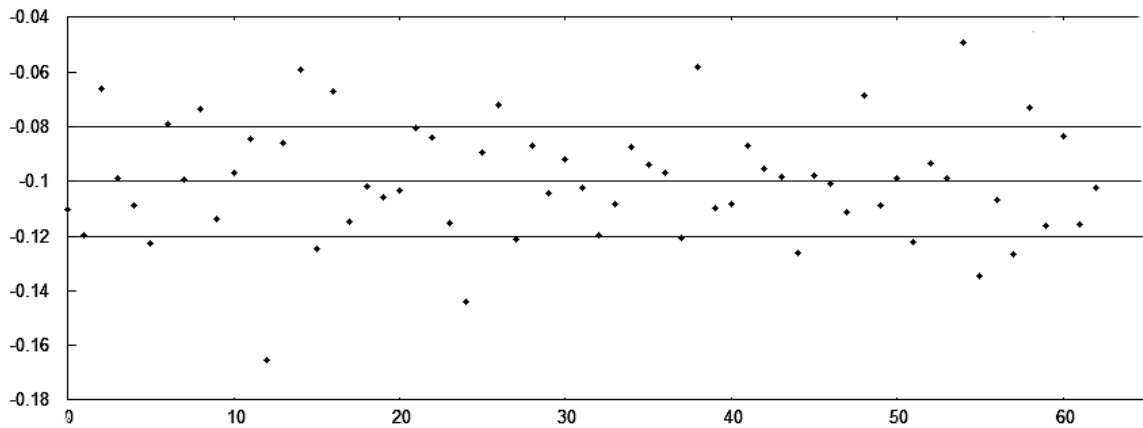


Рис. 10. Относительный угол отклонения молекул. По оси абсцисс - последовательные номера пар молекул. По оси ординат - значение угла в радианах. Температура 300 К.

Выводы.

1. Рассматриваемая модель дает неплохое качественное согласие с экспериментом при полученных величинах зарядов, что может при дальнейшем исследовании в какой-то степени прояснить вопрос с экранировкой для растворов ДНК.

2. Под влиянием температуры для рассмотренных параметров системы качественно сохраняется идеальная холестерическая структура, но появляются дефекты в виде отклонений к геликоидальной симметрии.

Список литературы

- [1] Бреслер С.Е. "Введение в молекулярную биологию" М. Наука, 1966.
- [2] F. Livolant, A. Leforestier "Condensed Phases Of DNA: Structures and Phase Transitiones" Prog. Polym. Sci., Vol. 21, 1115-1164, 1996.
- [3] Ю.М.Евдокимов, Жидкие кристаллы 3, 10 (2003).
- [4] J. Sponer, J. Leszczynski and P. Hobza, Biopolymers 61, 3 (2002).
- [5] А. Н. Захлевных "Основы статистической физики жидких кристаллов" Пермь 2006.
- [6] П. де Жен "Физика жидких кристаллов" издательство "Мир" 1977.
- [7] Л. Д. Ландау, И. М. Лифшиц "Теоретическая физика" т. 2 "Теория поля" М.: Наука 1988
- [8] В. Л. Голо, Е. И. Кац и Ю. С. Волков, Письма в ЖЭТФ 86, 311 (2007)
- [9] Ю. С. Волков, В. Л. Голо, Е. И. Кац, С. А. Кузнецова "Жидкокристаллические фазы, образованные дуплексами ДНК, содержащими пирофосфатные группы" 2008.
- [10] А. А. Kornyshev, S. Leikin, J. Chem. Phys. 107, 3656 (1997).

Приложение 1

Алгоритм(метод Монте - Карло).

Метод основан на предположении эргодичности, то есть совпадении гиббсовских средних и временных. В этой задаче мы предполагаем эргодичность.

Строим марковский процесс $p_k^\alpha \rightarrow p_{k+1}^\beta$ где p_i^ξ - вероятность на i -м шаге находиться в состоянии ξ , w_α^β - вероятность совершить переход $p_k^\alpha \rightarrow p_{k+1}^\beta$.

$$\sum_{\alpha} p_k^\alpha = 1$$
$$\sum_{\beta} w_\alpha^\beta = 1$$

Запишем p_{k+1}^β в виде:

$$p_{k+1}^\beta = p_k^\beta + \sum_{\alpha} w_\alpha^\beta p_k^\alpha - p_k^\beta \cdot 1 = p_k^\beta + \sum_{\alpha} (w_\alpha^\beta p_k^\alpha - w_\beta^\alpha p_k^\beta).$$

Из этого получим равенство

$$p_{k+1}^\beta - p_k^\beta = \sum_{\alpha} (w_\alpha^\beta p_k^\alpha - w_\beta^\alpha p_k^\beta).$$

Достаточное условие стационарности процесса:

$$w_\alpha^\beta p^\alpha - w_\beta^\alpha p^\beta = 0.$$

$$\text{Подставим } p^\beta = \frac{1}{Z} e^{-\beta E_\beta}, p^\alpha = \frac{1}{Z} e^{-\beta E_\alpha}.$$

Где β - обратная температура, E_α - энергия в состоянии α . Получим $\frac{w_\alpha^\beta}{w_\beta^\alpha} = e^{-\beta(E_\beta - E_\alpha)}$.

Для программирования используется алгоритм Метрополиса. Считаем, что

$$w_\alpha^\beta = \begin{cases} 1, & E - E < 0; \\ e^{-\beta(E_\beta - E_\alpha)}, & E - E > 0, 0 < q < 1, q - \text{случайно.} \end{cases}$$

Если $0 < q < e^{-\beta(E_\beta - E_\alpha)}$, тогда переходим к следующему шагу Монте-Карло. Если нет, то выбираем другое положение системы, а этот шаг не учитываем.

Алгоритм Метрополиса позволяет учесть зависимость от температуры. Если система пришла в равновесие, то среднее значение можно найти усреднением по последним состояниям цепи.

Приложение 2

Описание программы.

Вычисление парного потенциала взаимодействия - процесс, требующий большого количества вычислительных операций. Для ускорения работы программы была использована технология параллельного программирования с использованием графических процессоров. Эта технология позволяет получить существенное ускорение по сравнению с линейным счетом на процессоре, причем с увеличением системы ускорение существенно возрастает.

Архитектура графических процессоров позволяет вычислять большое количество однотипных независимых друг от друга действий одновременно. Поэтому общий потенциал допускает распараллеливание следующим образом: все данные молекул (диполи, заряды и их координаты) записываются в один массив, один графический процессор считает энергию взаимодействия элементарной ячейки одной молекулы с элементарной ячейкой другой молекулы (см. рис. 11). Потом все эти элементарные энергии складываются. Общая энергия при этом получается удвоенной.

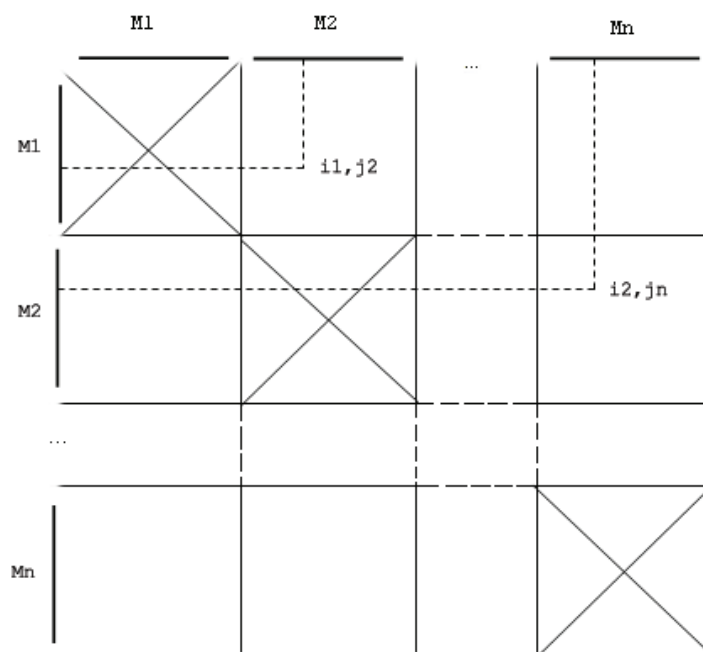


Рис. 11. Способ распараллеливания. Каждый процессор считает элементарную энергию взаимодействия ячейки одной молекулы с ячейкой другой.

Приложение 3

Код программы.