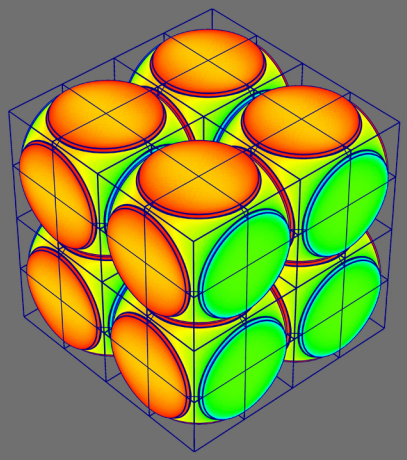
ISSN 2309-3684



Математическое моделирование и численные методы

Плохотников К.Э. Численное моделирование матрицы формы молекулы ДНК. Математическое моделирование и численные методы, 2026, № 1, с. 99–118.

Источник: https://mmcm.bmstu.ru/articles/401/

Параметры загрузки:
IP: 216.73.217.38
16.07.2026 00:23:52

УДК 519.6 + 539.182 DOI: 10.18698/2309-3684-2026-1-99118

Численное моделирование матрицы формы молекулы ДНК

© К.Э. Плохотников[1,2]

[1]МГУ им. М.В. Ломоносова, 119991, Москва, Россия
[2]ФГБОУ ВО Финансовый университет при Правительстве РФ, 125167, Москва, Россия

Понятие матрицы формы или морфы для описания произвольной молекулярной системы (кластера) введено автором ранее в предыдущей работе. Матрица формы всегда может быть построена по имеющейся конфигурации атомов кластера. В общем же случае морфа это симметричная матрица с неотрицательными элементами размером $N \times N$, где N — число атомов в кластере. Вариация матрицы формы кластера позволила регулировать число локально равновесных конфигураций аффилированной функции потенциальной энергии и, тем самым, в известных пределах решить проблему выбора равновесной конфигурации кластера из набора возможных. Рассмотрен пример построения первичной структуры небольшого фрагмента молекулы ДНК, именуемой еще олигонуклеотидом, в качестве локально равновесной конфигурации подходящей функции потенциальной энергии. Матрица формы олигонуклеотида строится последовательно путем построения матриц формы всех входящих компонентов, включая четыре азотистых основания ДНК: аденина, гуанина, цитозина и тимина. Определены матрицы формы углеводной компоненты нуклеиновой кислоты — дезоксирибозы, а также комбинации углеводной компоненты и фосфатной группы нуклеиновой кислоты — дезоксирибозы-фосфат. Наконец, построены матрицы формы нуклеозидмонофосфатов (нуклеотидов): дезоксиаденозина, дезоксигуанозина, дезоксицитидина, тимидина. Искомая матрица формы олигонуклеотида (молекулы ДНК) определяется в виде блочно-диагональной матрицы, в которой в качестве блоков выступают матрицы формы отдельных нуклеотидов. Построенная матрица формы ДНК имеет перспективу обобщения данного формата описания на случай вторичной и последующих структур молекулы ДНК.

Ключевые слова: *локально равновесная конфигурация, градиентный спуск, вычислительный эксперимент, проблема выбора решений, мультиатомный кластер, бинарный потенциал, матрица формы, морфа, нуклеотид, дезоксирибоза, олигонуклеотид, молекула ДНК*

Введение. В работах автора [1,2] изучалась проблема выбора решений при классическом формате описания мультиатомного кластера. Под мультиатомным кластером понимается такая молекулярная система, которая состоит из набора, вообще говоря, различных атомов. Специальным образом строилась функция потенциальной энергии для описания мультиатомного кластера. Изучались локально равновесные стационарные точки функции потенциальной энергии. Как оказалось число локально равновесных конфигураций функции

потенциальной энергии может быть значительно уменьшено вплоть до одной-единственной конфигурации с помощью понятия матрицы формы или морфы. Более того, можно построить функцию потенциальной энергии, имеющую одну-единственную локально-равновесную конфигурацию, выступающую в качестве глобального минимума.

Математическая модель. Пусть мультиатомный кластер состоит из N атомов, позиции которых описываются радиус-векторами $\mathbf{r}_i = (x_i, y_i, z_i)$, $i = 1,...,N$. Определим функцию потенциальной энергии U мультиатомного кластера в следующем виде:

$$U = \varepsilon \sum\nolimits_{1=i<j=N} \phi(\tfrac{r_{i,j}}{p_{i,j}}), \tag{1}$$

где

$$r_{i,j} = |\mathbf{r}_i - \mathbf{r}_j| = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2 + (z_i - z_j)^2},$$

ε — неотрицательный параметр. Функция $\phi = \phi(a)$ одной переменной a аимеет один-единственный минимум. Для определенности функцию ϕ выберем в форме потенциала Ми (Леннарда-Джонса), т.е.

$$\phi(a) = \tfrac{m}{n-m} a^{-n} - \tfrac{n}{n-m} a^{-m},$$

где n, m ($n>m>0$) — некоторые константы.

В (1) присутствует матрица формы $\{p_{i,j}\}$, $i, j = 1,...,N$, которая по определению является симметричной матрицей с неотрицательными элементами, т.е.

$$p_{i,j} = p_{j,i} \geq 0, \; i, j = 1,...,N.$$

Рассмотрим специальный случай, когда локально равновесная геометрия мультиатомного кластера известна и характеризуется набором радиус-векторов $\mathbf{p}_i = (u_i, v_i, w_i)$, $i = 1,...,N$. В этом случае матрицу формы по определению подсчитаем согласно формуле:

$$p_{i,j} = |\mathbf{p}_i - \mathbf{p}_j| = \sqrt{(u_i - u_j)^2 + (v_i - v_j)^2 + (w_i - w_j)^2}, \tag{2}$$

где $i \neq j$, $i, j = 1,...,N$.

Определим процедуру нахождения локального минимума функции (1) методом градиентного спуска, рассматривая в качестве переменных спуска $3N$ переменных $\mathbf{r}_i$, $i = 1,...,N$. В этом случае запишем:

$$\begin{aligned}
\dot{x}_k &= -\tfrac{\partial U}{\partial x_k} = \tfrac{nm}{n-m}\varepsilon \sum\nolimits_{j=1(j\neq k)}^{N} \mu(\tfrac{p_{k,j}}{r_{k,j}})\tfrac{x_{k,j}}{r_{k,j}^2},\\
\dot{y}_k &= -\tfrac{\partial U}{\partial y_k} = \tfrac{nm}{n-m}\varepsilon \sum\nolimits_{j=1(j\neq k)}^{N} \mu(\tfrac{p_{k,j}}{r_{k,j}})\tfrac{y_{k,j}}{r_{k,j}^2},\\
\dot{z}_k &= -\tfrac{\partial U}{\partial z_k} = \tfrac{nm}{n-m}\varepsilon \sum\nolimits_{j=1(j\neq k)}^{N} \mu(\tfrac{p_{k,j}}{r_{k,j}})\tfrac{z_{k,j}}{r_{k,j}^2}.
\end{aligned} \tag{3}$$

где точка над величинами обозначает производную по переменной, по которой осуществляется градиентный спуск, а

$$\mu(a) = a^n - a^m,\ x_{k,j} = x_k - x_j,\ y_{k,j} = y_k - y_j,$$
$$z_{k,j} = z_k - z_j,\ k, j = 1,...,N.$$

Система уравнений (3) решалась численно, при этом начальные положения $3N$ переменных x_k, y_k, z_k, $k = 1,...,N$ выбирались равномерно случайными числами из отрезка $[-L, L]$, где L некоторый положительный параметр.

С учетом предыдущих работ автора, где введено понятие матрицы формы, следует ожидать, что локально равновесная конфигурация молекулярной системы $\mathbf{R}_i$, $i = 1,...,N$ уравнений (3) должна быть такой, что

$$R_{i,j} = |\mathbf{R}_i - \mathbf{R}_j| = p_{i,j},\ 1 = i < j = N.$$

Кроме того, полученная конфигурация должна выступать в качестве глобального минимума функции потенциальной энергии (1). При этом значение функции потенциальной энергии на равновесной конфигурации

$$U|_{\mathbf{r}_i=\mathbf{R}_i,\, p_{i,j}=R_{i,j},\, 1=i<j=N} = U_{eq}$$

должно совпасть с числом бинарных связей кластера, т.е.

$$U_{eq} = -\tfrac{1}{2}N(N-1)\varepsilon.$$

Алгоритм численного решения. Существует множество подходов к моделированию динамики отдельных молекул ДНК [3–9], а также молекул ДНК, упакованных в клетках эукариотов [10–13]. Сравним понятие матрицы формы с аппроксимационным, мультипараметрическим потенциалом взаимодействия атомов кластера ряда компьютерных программ молекулярно динамического моделирования [14–16]. Для определенности выберем пакет молекулярно динамического моделирования сложных биологических молекул AMBER [17,18]. Его мультипараметрический потенциал, описывающий силовые поля в методе молекулярной динамики представлен на ресурсе [19]. Потенциал AMBER, а также другие программные средства в ча-

сти спектра настраиваемых параметров заметно избыточен, тогда как матрица формы специфична для описания данной молекулярной системы, она достаточна для (практически) однозначного воспроизведения в пространстве позиций атомов кластера.

Результаты моделирования азотистых оснований молекулы ДНК. В начале построим в качестве равновесных конфигураций четыре азотистых основания нуклеиновых кислот ДНК. Для этого потребуются определить соответствующие матрицы формы. В сети можно получить полную информацию о пространственном строении азотистых оснований молекул аденина [20], гуанина [21], цитозина [22] и тимина [23].

Для каждого из четырех нуклеотидов решалась задача градиентного спуска (3) по поиску равновесной конфигурации $\mathbf{r}_i = (x_i, y_i, z_i)$, $i = 1, ..., N$ при фиксированной матрице формы, найденной с помощью (2), где набор векторов

$$\mathbf{p}_i = (u_i, v_i, w_i),\ i = 1, ..., N,$$

считался известным, он выбирался последовательно среди радиусов-векторов атомов аденина, гуанина, цитозина и тимина. На рис. 1 приведены все четыре конфигурации, найденный в качестве локально равновесных, при этом каждая из конфигураций соответствовала глобальному минимуму функции потенциальной энергии (1).

На рис.1 прямые линии проведены для тех пар атомов в молекуле, которые не превышают двух значений минимального парного расстояния. Расчеты локально равновесных конфигураций каждой из молекул проводились в количестве $M = 150$. В качестве начальных распределений атомов выбирались равномерно случайные позиции в кубе со стороной, равной максимальному парному расстоянию. В каждом из расчетов воспроизводилась одна и та же каноническая конфигурация, задаваемая соответствующей матрицей формы. Во всех расчетах считалось, что $n = 12$, $m = 6$.

Моделирование дезоксирибозы и дезоксирибозы-фосфат молекулы ДНК. Построим матрицу формы углеводной компоненты нуклеиновой кислоты — дезоксирибозы, точнее β-2-дезокси-D-рибозы (β-D-2-Deoxyribose). На ресурсе [24] представлены 6 конформаций дезоксирибозы. Там же можно получить доступ к пространственному строению всех атомов дезоксирибозы каждой из конформаций. В дезоксирибозу входят 4, 5 и 10 атомов кислорода, углерода и водорода соответственно, т.е. всего атомов $N = 19$.

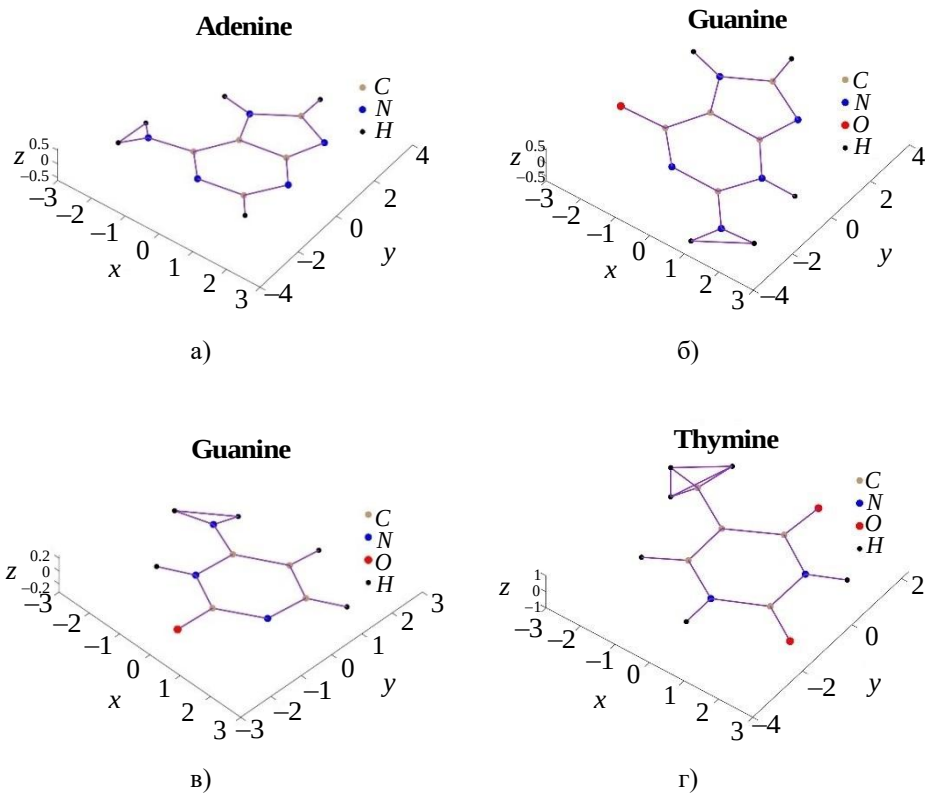


Рис. 1. Равновесная конфигурация молекулы:
а) — аденина; б) — гуанина; в) — цитозина; г) — тимина

Матрицу формы дезоксирибозы вычислим с помощью нахождения гармонической средней по всем имеющимся конформациям. Пусть

$$\{p_{\alpha,i,j}\},\ \alpha = 1,\ldots,\mathrm{A} = 6,\ i,j = 1,\ldots,N = 19$$

— матрицы формы каждой из A конформаций дезоксирибозы. Матрицу формы каждой из конформаций найдем согласно формуле (2), т.е. положим, что

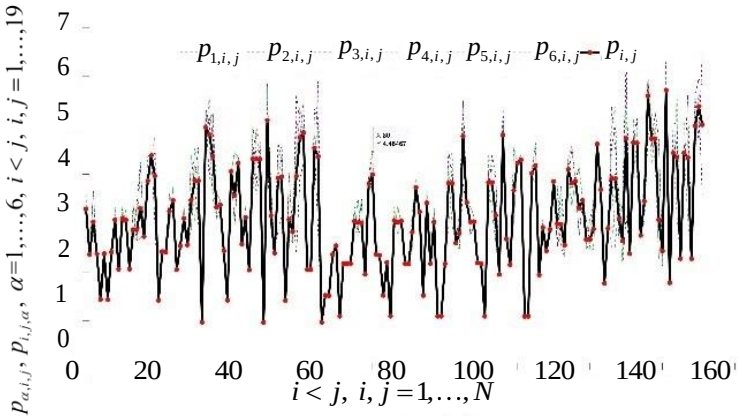
$$p_{\alpha,i,j} = |\mathbf{p}_{\alpha,i} - \mathbf{p}_{\alpha,j}|,\ \alpha = 1,\ldots,\mathrm{A} = 6,\ i,j = 1,\ldots,N = 19,$$

где $\mathbf{p}_{\alpha,i}$ — радиус-вектор i-го атома a-й конформации дезоксирибозы. Найдем усредненную матрицу формы дезоксирибозы $\{p_{i,j}\}$ без диагональных элементов по формуле:

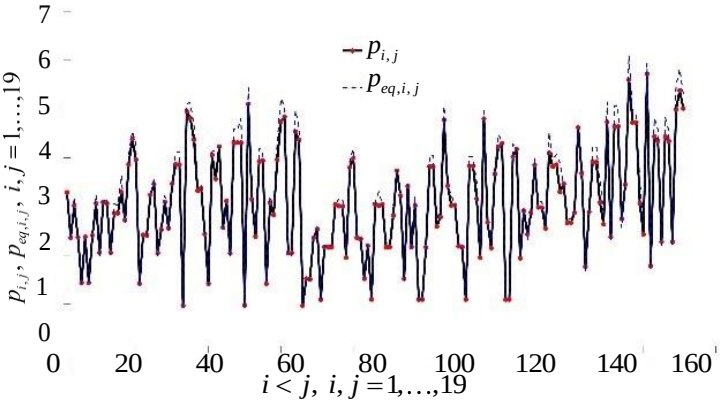
$$\frac{1}{p_{i,j}} = \frac{1}{\mathrm{A}}\sum_{\alpha=1}^{\mathrm{A}} \frac{1}{p_{\alpha,i,j}},\ i,j(i \neq j) = 1,\ldots,N\ . \qquad (4)$$

Гармоническая средняя (4) выбрана с учетом того, что недиагональные элементы матрицы формы по определению положительны. Отметим, что выбор формулы усреднения (средние арифметическая, геометрическая и пр.) незначительно влияют на конечный результат.

На рис.2,а приведены семь распределений элементов матриц $\{p_{1,i,j}\}$, …, $\{p_{6,i,j}\}$, $\{p_{i,j}\}$ в последовательности $i<j$, $i,j=1,...,19$, т.е. всего $\frac{19\cdot 18}{2}=171$ элементов. Черная линия с красными маркерами в виде точек обозначает распределение элементов усредненной по формуле (4) матрицы формы $\{p_{i,j}\}$.



а)



б)

Рис. 2. Сравнение элементов:
а) — матриц форм шести конформаций дезоксирибозы с усредненной матрицей формы;
б) — усредненных матриц форм: исходной и полученной после усреднения равновесных конфигураций дезоксирибозы

Выберем в качестве потенциала взаимодействия атомов дезоксирибозы выражение (1) с матрицей формы (4). Решая задачу градиентного спуска (3), найдем локально равновесные конфигурации каждой из шести конформаций дезоксирибозы. В качестве начальных распределений атомов при численном решении системы уравнений (3) выберем имеющиеся конфигурации матрицы формы, элементы которых приведены на рис. 2,а.

Обозначим через $\mathbf{p}_{eq,\alpha,i}$ — радиус-вектор i-го атома, α-ой конформации равновесной конфигурации дезоксирибозы. С помощью формулы (2) найдем матрицу формы α-ой конформации, $\{p_{eq,\alpha,i,j}\}$. По формуле (4) построим среднюю гармоническую матрицу формы $\{p_{eq,i,j}\}$, исходя из матриц формы равновесных конфигураций $\{p_{eq,\alpha,i,j}\}$, $\alpha = 1,...,6$. На рис. 2,б приведено сравнение элементов двух средних матриц форм $\{p_{i,j}\}$ и $\{p_{eq,i,j}\}$ в виде двух кривых. Первая обозначена черной линией с красными маркерами в виде точек, вторая — синей штриховой линией. Видно неплохое соответствие элементов обеих матриц.

Рассмотрим совместное позиционирование имеющихся конформаций дезоксирибозы с локально равновесными, которые получены из первых путем градиентного спуска при решении задачи (3), когда была использована найденная выше матрица формы $\{p_{i,j}\}$, $i, j = 1,...,19$.

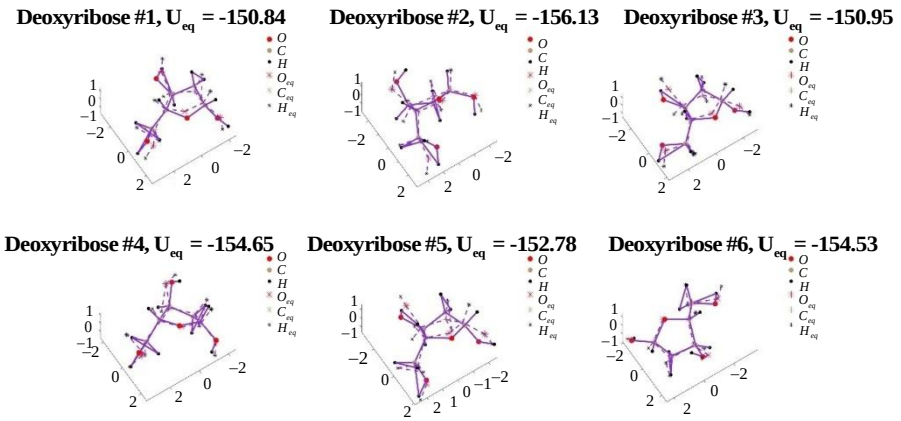


Рис. 3. Совместное позиционирование имеющихся конформаций дезоксирибозы с локально равновесными конфигурациями

На рис.3 приведены шесть имеющихся конформаций дезоксирибозы (сплошные линии), а также соответствующие локально равновесные конфигурации (штриховые линии). Атомы для первого и вто-

рого типа конфигураций помечены маркерами различных размеров в виде точек и звезд соответственно. Субиндекс “eq” обозначает отношение к локально равновесным конфигурациям. Видно, что локально-равновесные конфигурации заметно отличаются от исходных, взятых из имеющейся базы данных, при этом качественно структура дезоксирибозы не нарушается. Линии, соединяющие пару атомов дезоксирибозы на рис.3, проведены в случаях, когда длины линий меньше двух значений минимального бинарного расстояния.

Отметим, что средние матрицы формы $\{p_{i,j}\}$, $\{p_{eq,i,j}\}$, $i, j = 1,...,19$ не удовлетворяют условию (2), т.е. одна-единственная конфигурация дезоксирибозы не существует. По этой причине следует ожидать, что с любой из двух выбранных матриц формы может быть связано множество конформаций, которые выступают в виде набора локально равновесных конфигураций. При этом среди всех этих конформаций глобальный минимум потенциальной энергии, равный значению

$$-\tfrac{\varepsilon}{2} N(N-1)\,|_{\varepsilon=1,N=19} = -171$$

не достижим.

В дальнейших расчетах в качестве матрицы формы дезоксирибозы будем иметь в виду матрицу $\{p_{eq,i,j}\}$, $i, j = 1,...,19$, полученную после усреднения матриц форм отдельных равновесных конфигураций. Для простоты субиндекс «eq» в обозначении матрицы формы опустим. Учтем, что потенциальная энергия равновесных конфигураций, представленных на рис. 3, не превышает значения $U_{\max} = -150.84$. Путем многократного решения системы уравнений (3) в количестве $M = 10^5$ подсчитаем число локально равновесных конфигураций, N_{eq}, энергия которых не превышает значения $U_{\max}$. В рамках указанной совокупности решений системы уравнений (3) с параметрами $n = 12$, $m = 6$, $L = 3.0478$ оказалось, что $N_{eq} = 3'172$. В каждом расчете начальные позиции атомов дезоксирибозы выбирались равномерно случайно в кубе $[-L, L]^3$, при этом ребро куба $2L$ выбиралось равным максимальному бинарного расстоянию между атомами.

Среди найденного количества локально равновесных конфигураций могут быть совпадающие. Для учета возможных совпадений введем вектор бинарных расстояний кластера с учетом матрицы формы по формуле:

$$R = \mathrm{sort}\left(\frac{r_{1,2}}{p_{1,2}}, ..., \frac{r_{N-1,N}}{p_{N-1,N}} \right), \tag{5}$$

где $sort$ — операция сортировки по возрастанию.

В силу определения, вектор (5) имеет $\frac{1}{2}N(N-1)$ компонент. Введем меру различия $d(c_1, c_2)$ пары локально равновесных конфигураций c_1, c_2 с помощью соответствующих векторов бинарных расстояний $R^{(1)}$, $R^{(2)}$ кластеров по формуле:

$$d(c_1, c_2) = \max_{1 \le i \le \frac{1}{2}N(N-1)} |R_i^{(1)} - R_i^{(2)}| > \delta_c . \qquad (6)$$

Согласно критерию (6) две локально равновесные конфигурации кластера c_1, c_2 считаются различными, если расстояние $d(c_1, c_2)$ между ними превышает некоторое пороговое значение δ_c. В расчете, упомянутом выше, при $\delta_c = 10^{-3}$ осталось 291 различных локально равновесных конфигурации дезоксирибозы, т.е., как оказалось, $N_{eq} = 291$, при этом дальнейшее уменьшение значения параметра δ_c не влияло на число различных локально равновесных конфигураций. На рис. 4 приведены шесть локально равновесных конфигураций дезоксирибозы, потенциальная энергия которых не превышает значения $U_{\max}$. Найденное число $N_{eq} = 291$ локально равновесных конфигураций следует рассматривать в качестве нижней границы. Все найденные равновесные конфигурации субординированы по возрастанию потенциальной энергии, при этом первые три и вторые три конфигурации имеют наименьшие и наибольшие значения потенциальных энергий соответственно.

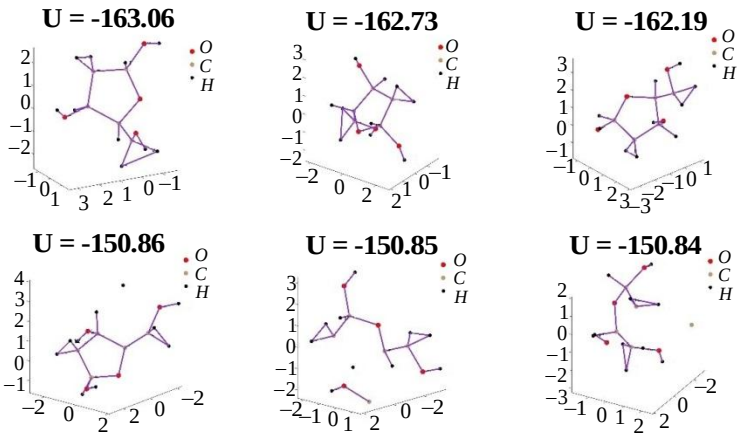


Рис. 4. Образцы локально равновесных конфигураций дезоксирибозы, потенциальная энергия которых не превышает значения $U_{\max} = -150.84$

Построим матрицу формы комбинации углеводной компоненты и фосфатной группы нуклеиновой кислоты — дезоксирибозы-фосфат

(2-Deoxyribose-5-phosphate). На ресурсе [25] приводится 10 конформаций данной компоненты. Применим методику построения матрицы формы, апробированную на предыдущем примере построения матрицы формы дезоксирибозы.

В начале по формуле (4) была вычислена средняя гармоническая матрица формы $\{p_{i,j}\}$ всех десяти конформаций дезоксирибозы-фосфат,

$$\{p_{\alpha,i,j}\},\ \alpha = 1,...,\mathrm{A} = 10,\ i, j = 1,..., N = 24$$

Найденная матрица формы была использована при поиске локально равновесных конфигураций кластера путем решения системы дифференциальных уравнений (3), (4). В качестве начальных позиций атомов кластера выбирались имеющиеся на указанном ресурсе позиции атомов каждой из десяти конформаций. В итоге были найдены десять локально равновесных конфигураций, $\{p_{eq,\alpha,i,j}\}$, $\alpha = 1,...,10$, $i, j = 1,..., 24$, среди которых, кстати, нашлись четыре пары совпадающих с номерами (2,3), (2,7), (3,7), (5,10). Потенциальные энергии всех десяти конформаций вошли в интервал $[-244.55, -225.68]$.

В дальнейших расчетах в качестве матрицы формы $\{p_{i,j}\}$, $i, j = 1,..., 24$ выбиралось среднее гармоническое набора локально равновесных матриц формы $\{p_{eq,\alpha,i,j}\}$, $\alpha = 1,...,10$, $i, j = 1,..., 24$. Отметим, что глобальный минимум для дезоксирибозы-фосфат, состоящей из $N = 24$ атомов, составляет значение $-\frac{\varepsilon}{2} N(N-1)|_{\varepsilon=1, N=24} = -276$.

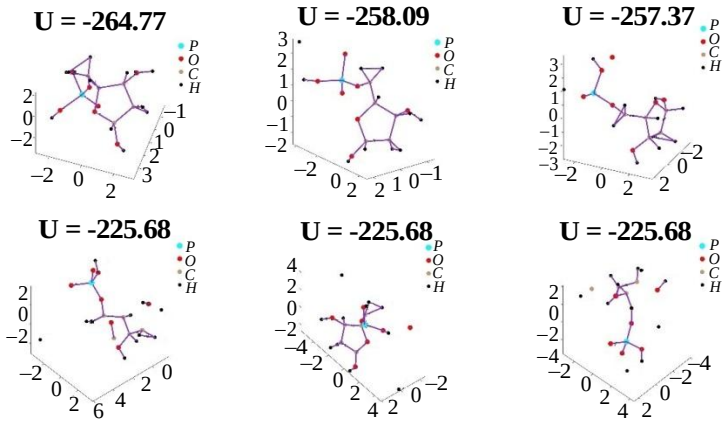


Рис. 5. Образцы локально равновесных конформаций дезоксирибозы-фосфат

Оценим потенциальное количество конформаций N_{eq} кластера дезоксирибозы-фосфат. Для этого решим задачу (3), (4) $M = 10^5$ раз, выбирая в каждом случае начальные позиции атомов кластера равномерно случайно в кубе $[-L, L]^3$, $L = 4.1184$. Отберем лишь те локально равновесные конфигурации, потенциальная энергия которых не превышает значения $U_{\max} = -225.68$, их количество оказалось равным 9′880, из них согласно критерию (5), (6) отличными друг от друга оказалось $N_{eq} = 6'956$. Остальные значения параметров расчета совпадали с теми, которые были использованы при построении матрицы формы дезоксирибозы.

На рис. 5 приведены шесть образцов локально равновесных конформаций дезоксирибозы-фосфат. Все конфигурации субординированы по возрастанию потенциальной энергии, причем первые три имеют наименьшие значения из выбранного диапазона, а вторые три — наибольшие значения. Видно, что с ростом потенциальной энергии не только отдельные атомы водорода, но целые фрагменты кластера отдаляются от других фрагментов на расстояние, большее двух значений минимального бинарного расстояния.

Результаты моделирования матрицы формы нуклеотидов молекулы ДНК. Построим матрицы формы нуклеозидмонофосфатов (нуклеотидов) — следующих по сложности компонент дезоксирибонуклеиновых кислот. К ним относятся дезоксиаденозин-5′-фосфат (2′-Deoxyadenosine-5′-monophosphate [26]), дезоксигуанозин-5′-фосфат (2′-Deoxyguanosine-5′-phosphate [27]), дезоксицитидин-5′-фосфат (Deoxycytidine-5′-monophosphate [28]), тимидин-5′-фосфат (Thymidine-5′-phosphate [29]). Воспользуемся стандартными сокращенными обозначениями dAMP, dGMP, dCMP, dTMP для каждого из перечисленных выше нуклеотидов. На указанных выше ресурсах для каждого из четырех нуклеотидов приведено по десять пространственных конформаций.

Таблица 1

Число локально равновесных конфигураций, N_{eq}, потенциальная энергия которых не превышает значения $U_{\max}$

	N	$U_{\min,gl} = -\frac{1}{2}N(N-1)$	$U_{\max}$	N_{eq}
dAMP	36	– 630	– 537.55	6′954
dGMP	37	– 666	– 559.12	2′871
dCMP	34	– 561	– 475.01	4′941
dTMP	36	– 630	– 521.69	4′073

Как и в случае дезоксирибозы-фосфат были найдены все четыре средние гармонические матрицы формы нуклеотидов, подсчитанные по соответствующему набору равновесных конфигураций. Для каждого из нуклеотидов решалась задача (3) в количестве $M = 10^5$ раз. Начальные позиции атомов нуклеотидов при решении задачи (3) выбирались равномерно случайно в кубе $[-L, L]^3$, где длина ребра куба $2L$ равнялась максимальному бинарному расстоянию в кластере. Подсчитывалось число, N_{eq} тех локально равновесные конфигурации, потенциальные энергии которых не превышали соответствующие пороговые значения $U_{\max}$. Кроме того, с помощью критерия (5), (6) при $\delta_c = 10^{-3}$ проверялось различие учтенных локально равновесных конфигураций. В таблице №1 сосредоточены результаты подсчета числа локально равновесных конфигураций (конформаций) каждого из нуклеотидов.

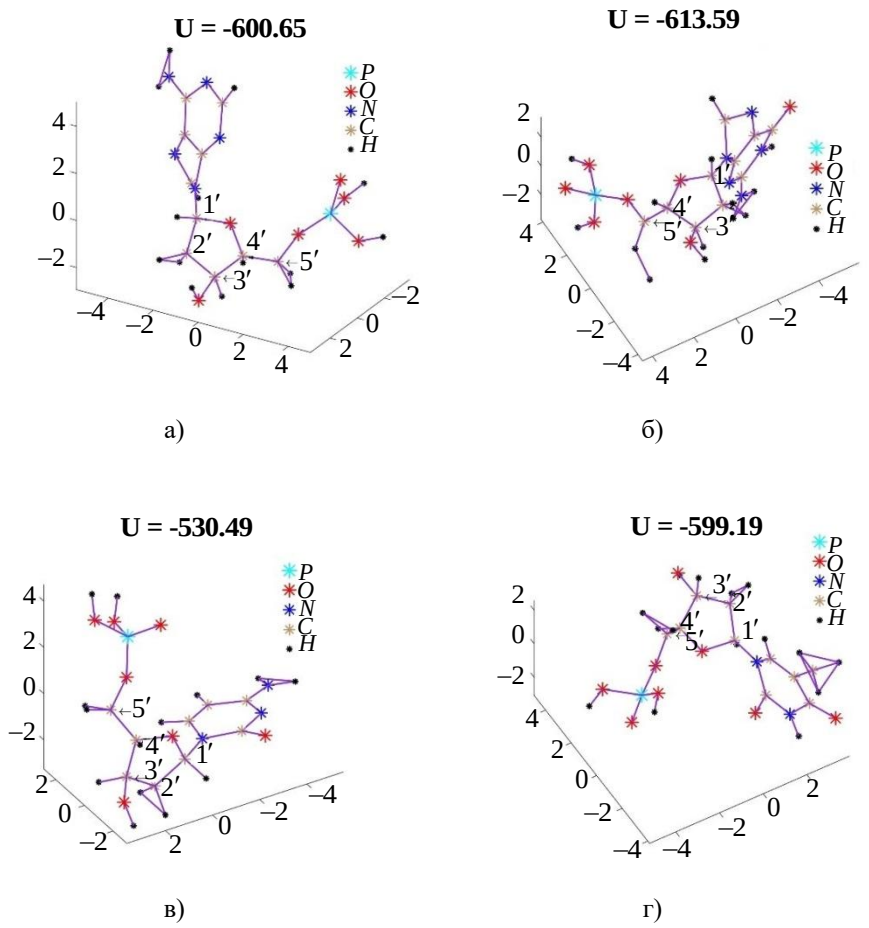


Рис. 6. Дезоксиаденозин-$5'$-фосфат:
а) — dAMP; б) — dGMP; в) — dCMP; г) — dTMP

В таблице 1 приведено число атомов, N, а также глобальные минимумы, $U_{\min,gl}$ нуклеотидов, которые недостижимы в силу того, что элементы соответствующих матриц формы вычислялись как средние гармонические величины каждой из десяти имеющихся конформаций. Число равновесных конфигураций N_{eq} следует рассматривать в качестве нижних границ.

На рис.6 приведены геометрические образцы всех четырех нуклеотидов, найденных в качестве локально равновесных конфигураций. На всех четырех рисунках приведена стандартная нумерация пяти атомов углерода (1′, 2′, 3′, 4′, 5′), входящих в дезоксирибозу каждого из нуклеотидов. Длина линий, соединяющих атомы, не превышала двух значений минимального бинарного расстояния каждого из нуклеотидов.

Имея в виду построенные выше матрицы формы четырех нуклеотидов, построим матрицу формы небольшой последовательности нуклеотидов, т.е. построим первичную структуру фрагмента ДНК, которая еще называется олигонуклеотидом. Как известно, каждый нуклеотид связан с другим фосфодиэфирными мостиками, причем 3′-конец одного нуклеотида соединяется с 5′-концом соседнего нуклеотида. Таким образом, любой полинуклеотид, помимо набора (быть может и пустого) внутренних нуклеотидов, содержит два концевых нуклеотида, которые еще называются 3′-концевой (со свободной 3′-OH-группой) и 5′-концевой (5′-OH-группой).

Моделирование матрицы формы молекулы ДНК. Под матрицей формы олигонуклеотида, P будем понимать блочно-диагональную матрицу, в которой в качестве блоков выступают матрицы формы отдельных нуклеотидов, p_{dAMP}, p_{dGMP}, p_{dCMP}, p_{dTMP}. В дальнейшем, упрощая обозначения, матрицы формы отдельных нуклеотидов обозначим буквами A, G, C, T. В этом случае блочные матрицы формы олигонуклеотидов можно описать выражениями вида: $P = \mathrm{A} \div \mathrm{G} \div \mathrm{G} \div \mathrm{C} \div \mathrm{T} \div \mathrm{T} \div \mathrm{A} \div \ldots$, где символ $\div$ — обозначает блочно-диагональную композицию пары матриц. Опуская символ $\div$, определим еще более компактную запись $P = \mathrm{AGGCTTA}\ldots$

При нашем формате описания полинуклеотида фосфодиэфирный мостик между парой нуклеотидов допускает два способа композиции. И в первом, и во втором способах гидроксильные группы: 3′-OH и две группы 5′-OH преобразуются (см. рис.6). При этом в расчете локально равновесной конфигурации олигонуклеотида равномерно случайно для каждого из четырех нуклеотидов выбирается одна из двух групп 5′-OH.

Согласно первому способу (схема 3′←5′), атом кислорода 3′ выступает в качестве 5′ атома кислорода последующего нуклеотида.

При этом выбранная гидроксильная группа 5′-OH отбрасывается, и позиция атома кислорода отброшенной группы замещается позицией кислорода гидроксильной группы 3′-OH предыдущего нуклеотида. Атом водорода последней гидроксильной группы отбрасывается.

Согласно второму способу композиции (схема 3′→5′), наоборот, атом кислорода 5′ выступает в качестве 3′ атома кислорода предыдущего нуклеотида. Гидроксильная группа 3′-OH отбрасывается, и позиция атома кислорода отброшенной группы предыдущего нуклеотида помещается в месте атома кислорода выбранной ранее гидроксильной группы 5′-OH последующего нуклеотида. Атом водорода последней гидроксильной группы отбрасывается.

На рис.7 приведены примеры обеих композиций матриц форм полинуклеотидов, состоящих из пяти нуклеотидов. На рис.7 точками отмечены ненулевые элементы матриц форм. Построение матриц форм осуществляется по главной диагонали с левого верхнего угла до правого нижнего, при этом первый и последний нуклеотид не претерпели изменений в гидроксильных группах 5′-OH и 3′-OH соответственно. Последовательность нуклеотидов на рис. 7 выбиралась равномерно случайно. Матрицы формы отдельных нуклеотидов помечены стандартной аббревиатурой. Скопление точек в виде прямых линий на рис. 7 обозначает проявление схем композиции 3′↔5′ последующего нуклеотида с предыдущим.

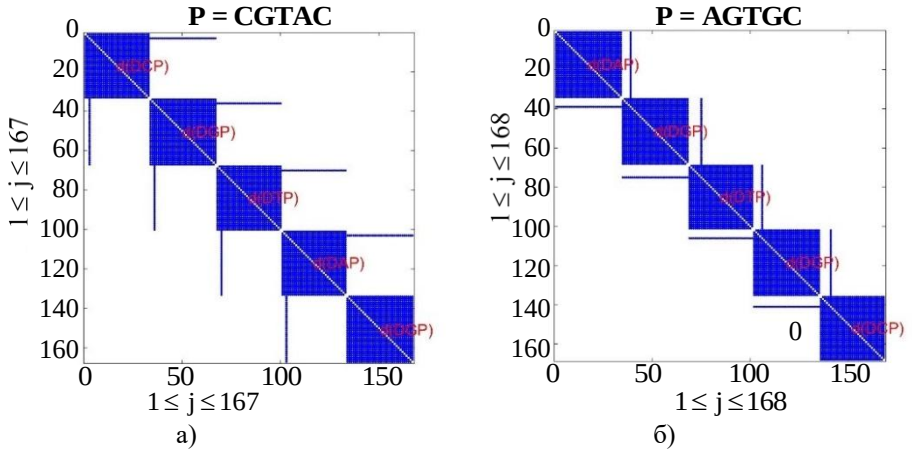


Рис. 7. Формат композиции матрицы формы:
а) — первый; б) — второй

С помощью вычислительного эксперимента построим олигонуклеотид в качестве локально равновесной конфигурации. Для этого выберем набор нуклеотидов и один из двух способов композиции соответствующих матриц формы в единую матрицу формы всего оли-

гонуклеотида. Решим численно систему уравнений градиентного спуска (3). Для приготовления начального распределения атомов олигонуклеотида необходимо сделать ряд замечаний.

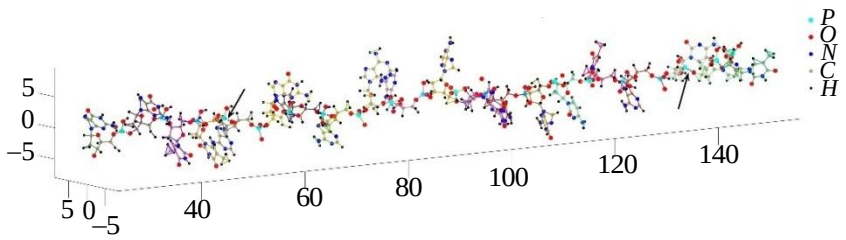
Во-первых, начальные позиции атомов каждого из нуклеотидов выбираем равномерно случайно в кубе со стороной, равной максимальному бинарному расстоянию между атомами данного нуклеотида. Во-вторых, кубы с начальными данными распределяются, примыкая друг к другу с право на лево по оси абсцисс с учетом последовательности нуклеотидов в матрице формы P. В-третьих, начальные данные перед использованием в системе уравнений (3) преобразуются с помощью минимизации методом градиентного спуска функционала вида:

$$F = \tfrac{1}{4}\sum\nolimits_{1\le i<j\le N,\, P_{i,j}>0} (r_{i,j}^2 - P_{i,j}^2)^2 \,, \qquad (7)$$

где $r_{i,j}$ — расстояние между i-м и j-м атомами олигонуклеотида, $\{P_{i,j}\}$, $i, j = 1,\ldots,N$ — матрица формы олигонуклеотида. Суммирование в (7) распространяется лишь на те пары атомов, для которых соответствующие элементы матрицы формы отличны от нуля.

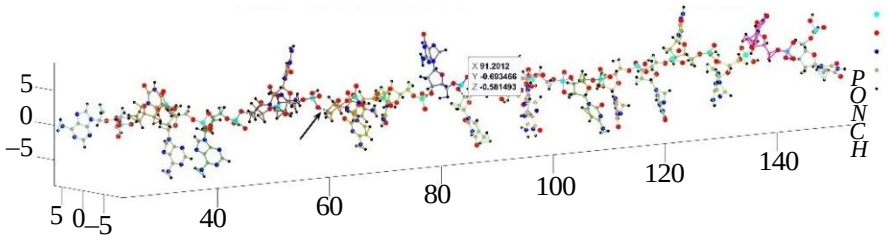
Отметим, что минимизация функционала (7) значительно облегчает поиск минимума потенциальной энергии (1) при решении системы уравнений (3). Конструкция функционала (7) обусловлена характерным представлением матрицы формы (2), когда матрица формы определяется набором бинарных расстояний между N точками в трехмерном пространстве.

P = TCCATCGCGAAAAAGATTGG, U = -10472.10



а)

P = TTACGGCTGTGCCCCCAATA, U = -10299.31



б)

Рис. 8. Локально равновесный фрагмент молекулы ДНК, построенный:
а) — согласно первой схеме композиции отдельных нуклеотидов;

В итоговых расчетах олигонуклеотида как локально равновесной конфигурации обращает внимание то обстоятельство, что локально равновесных конфигураций оказалось небольшое число в пределах десяти.

Заключение. В работе рассмотрен метод моделирования молекулярных систем с точки зрения подсчета энергии кластера и определения геометрии распределения атомов кластера в пространстве в терминах бинарного потенциала и аффилированной с ней матрицы формы.

Для иллюстрации понятия матрицы формы (морфы) рассмотрена задача описания первичной структуры молекулы ДНК в качестве локально равновесной конфигурации подходящей функции потенциальной энергии. На данном примере иллюстрируется применение универсальной методики композиции матрицы формы для произвольной биологической молекулы.

Предлагаемая методика позволила шаг за шагом в последовательности: азотистые основания (аденин, гуанин, цитозин, тимин) → дезоксирибоза → дезоксирибоза-фосфат → нуклеотиды → олигонуклеотид (небольшой фрагмент ДНК) построить искомую матрицу формы. Последняя позволила воспроизвести в качестве равновесной выбранный фрагмент молекулы ДНК с произвольной последовательностью азотистых оснований.

ЛИТЕРАТУРА

[1] Плохотников К.Э. Определение матрицы формы мультиатомного кластера. *Математическое моделирование*, 2024, т. 36, № 6, с.153–169.

[2] Плохотников К.Э. О множестве локально равновесных конфигураций потенциальной энергии мультиатомного кластера. *Ученые записки физического факультета Московского университета*, 2024, №5, 2450101, с. 1–19.

[3] Lu X. J., Olson W.K. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nature Protocols*, 2008, vol. 3, no.7, p. 1213—1227; DOI:

[4] Schlick T. *The mathematics of DNA structure, mechanics, and dynamics. molecular modeling and simulation: an interdisciplinary guide*. New York, Springer, 2010, pp. 95 p.

[5] Pérez A., Luque F. J., Orozco M. Frontiers in molecular dynamics simulations of DNA. *Accounts of Chemical Research*, 2012, vol. 45, no. 2, pp. 196–205.

[6] Cheatham T. E., Case D. A. Twenty-five years of nucleic acid simulations. *Biopolymers*, 2013, vol. 99, no. 12, pp. 969–977.

[7] Mu Z.-C., Tan Y.-L., Liu J., Zhang B.-G., Shi Y.-Z. Computational modeling of DNA 3D structures: from dynamics and mechanics to folding. *Molecules*, 2023, vol. 28, no. 12, art. 4833.

[8] B. Knappeová, V. Mlýnský, M. Pykal et al. Comprehensive Assessment of Force-Field Performance in Molecular Dynamics Simulations of DNA/RNA Hybrid Duplexes. *Journal of Chemical Theory and Computation*, 2024, vol. 20, no. 15, pp. 6917–6929.

[9] De Silva N., Perez A. Predicting rare DNA conformations via dynamical graphical models: a case study of the B→A transition. *Nucleic Acids Research*, 2025, vol. 53, no. 13. DOI: 10.1093/nar/gkaf601.

[10] Eidelman, Y. A., Slanina, S. V., Gusev, O. A., Andreev, S. G. High-performance computational modeling of chromosome conformations. *Supercomputing Frontiers and Innovations*, 2018, vol. 5, no. 3, pp. 38–41. DOI: https://doi.org/10.14529/jsfi180305.

[11] M. Ohno, T. Ando, D. G. Priest, et al. Sub-nucleosomal genome structure reveals distinct nucleosome folding motifs. *Cell*, 2019, vol. 176, no. 3, pp. 520–534. DOI: 10.1016/j.cell.2018.12.014.

[12] Liang, J., A. Perez-Rathke. Minimalistic 3D chromatin models: Sparse interactions in single cells drive the chromatin fold and form many-body units. *Current Opinion in Structural Biology*, 2021, vol. 71, pp. 200–214. DOI: 10.1016/j.sbi.2021.06.017.

[13] Vercellone F. A., Chiariello A. M., Esposito A. [et al.] Multiscale perspective on chromatin architecture through polymer physics. *Physiology*, 2025, vol. 40, no. 3. DOI: 10.1152/physiol.00050.2024

[14] Christen M., Hünenberger P. H., Bakowies D. [et al.] The GROMOS software for biomolecular simulation: GROMOS05. *Journal of Computational Chemistry*, 2005, vol. 26, no. 16, pp. 1719–1751. DOI: 10.1002/jcc.20303.

[15] Harder E., Damm W., Maple J. [et al.] OPLS3: A force field providing broad coverage of drug-like small molecules and proteins. *Journal of Chemical Theory and Computation*, 2016, vol. 12, no. 1, pp. 281–296. DOI: 10.1021/acs.jctc.5b00864.

[16] Ahmad K., Javed A. T., Lanphere C. [et al.] Structure and dynamics of an archetypal DNA nanoarchitecture revealed via cryo-EM and molecular dynamics simulations. *Nature Communications*, 2023, vol. 14, no. 1, art. 3630. DOI: 10.1038/s41467-023-38681-5.

[17] Case D. A., Aktulga H. M., Belfon K. [et al.] *Amber 2024 Reference Manual*. San Francisco, University of California, 2024, 1043 p.

[18] Case D. A., Cerutti D. S., Cruzeiro V. W. D. [et al.]. Recent developments in amber biomolecular simulations. *Journal of Chemical Information and Modeling*, 2025, vol. 65, no. 15, pp. 7835–7843. DOI: 10.1021/acs.jcim.5c01063.

[19] Maier J. A., Martinez C., Kasavajhala K. [et al.] ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of*

Chemical Theory and Computation, 2015, vol. 11, no. 8, pp. 3696–3713. DOI: 10.1021/acs.jctc.5b00255
[20] URL: https://bmrb.io/metabolomics/mol_summary/?molName=adenine [Электронный ресурс].
[21] URL: https://bmrb.io/metabolomics/mol_summary/?molName=guanine [Электронный ресурс].
[22] URL: https://bmrb.io/metabolomics/mol_summary/?molName=cytosine [Электронный ресурс].
[23] URL: https://www.chem.purdue.edu/jmol/molecules/thymine.html# [Электронный ресурс].
[24] URL: https://pubchem.ncbi.nlm.nih.gov/compound/439576#section=2D-Structure [Электронный ресурс].
[25] URL: https://pubchem.ncbi.nlm.nih.gov/compound/2-Deoxyribose-5-phosphate [Электронный ресурс].
[26] URL: https://pubchem.ncbi.nlm.nih.gov/compound/12599 [Электронный ресурс].
[27] URL: https://pubchem.ncbi.nlm.nih.gov/compound/135398597 [Электронный ресурс].
[28] URL: https://pubchem.ncbi.nlm.nih.gov/compound/13945 [Электронный ресурс].
[29] URL: https://pubchem.ncbi.nlm.nih.gov/compound/9700 [Электронный ресурс].

Статья поступила в редакцию 21.05.2025

Ссылку на эту статью просим оформлять следующим образом:
Плохотников К.Э. Численное моделирование матрицы формы молекулы ДНК. *Математическое моделирование и численные методы*, 2026, № 1, с. 99–118.

Плохотников Константин Эдуардович — д-р физ.-мат. наук, вед. научн. сотр., кафедра математического моделирования и информатики физического факультета МГУ им. М.В. Ломоносова. e-mail: psygma@yandex.ru

Numerical modeling of the shape matrix of a DNA molecule

© A.A. Valishin[1], A.V. Zaprivoda[2], S.S. Tsukhlo[1]

[1]M.V.Lomonosov Moscow State University, Moscow 119991 Russia
[2]Financial University under the Government of the Russian Federation, 125993, Moscow, Russia

The concept of a shape matrix or morph to describe an arbitrary molecular system (cluster) was introduced by the author earlier in a previous paper. The shape matrix can always be constructed according to the available configuration of cluster atoms. In general, a morph is a symmetric matrix with $N \times N$ *non—negative elements, where* N *is the number of atoms in the cluster. The variation of the cluster shape matrix made it possible to regulate the number of locally equilibrium configurations of the affiliated potential energy function and, thus, within known limits, solve the problem of choosing an equilibrium cluster configuration from a set of possible ones. An example of constructing the primary structure of a small fragment of a DNA molecule, also called an oligonucleotide, as a locally equilibrium configuration of a suitable potential energy function is consid-*

ered. The oligonucleotide shape matrix is constructed sequentially by constructing the shape matrices of all the constituent components, including the four nitrogenous bases of DNA: adenine, guanine, cytosine and thymine. The matrices of the form of the carbohydrate component of nucleic acid — deoxyribose, as well as combinations of the carbohydrate component and the phosphate group of nucleic acid — deoxyribose-phosphate were determined. Finally, matrices of the form of nucleoside monophosphates (nucleotides) were constructed: deoxyadenosine, deoxyguanosine, deoxycytidine, thymidine. The desired matrix of the oligonucleotide shape (DNA molecule) is defined as a block-diagonal matrix, in which the blocks are the matrices of the shapes of individual nucleotides. The constructed DNA shape matrix has the prospect of generalizing this description format to the case of secondary and subsequent structures of the DNA molecule.

Keywords*: locally equilibrium configuration, gradient descent, computational experiment, decision problem, multiatomic cluster, binary potential, shape matrix, morph, nucleotide, deoxyribose, oligonucleotide, DNA molecule*

REFERENCES

[1] Plokhotnikov K.E. Definition of the shape matrix multiatomic cluster. *Mathematical Models and Computer Simulations*, 2024, vol. 36, no. 6, pp. 153–169.

[2] Plokhotnikov K.E. O mnozhestve lokal'no ravnovesnyh konfiguracij potencial'noj energii mul'tiatomnogo klastera [On the set of locally equilibrium potential energy configurations of a multiatomic cluster]. *Uchenye zapiski fizicheskogo fakul'teta Moskovskogo universiteta* [Scientific notes of the Faculty of Physics of Moscow University], 2024, no. 5, art. 2450101, pp. 1–19.

[3] Lu X. J., Olson W.K. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nature Protocols*, 2008, vol. 3, no.7, p. 1213—1227; DOI:

[4] Schlick T. *The mathematics of DNA structure, mechanics, and dynamics. molecular modeling and simulation: an interdisciplinary guide*. New York, Springer, 2010, pp. 95 p.

[5] Pérez A., Luque F. J., Orozco M. Frontiers in molecular dynamics simulations of DNA. *Accounts of Chemical Research*, 2012, vol. 45, no. 2, pp. 196–205.

[6] Cheatham T. E., Case D. A. Twenty-five years of nucleic acid simulations. *Biopolymers*, 2013, vol. 99, no. 12, pp. 969–977.

[7] Mu Z.-C., Tan Y.-L., Liu J., Zhang B.-G., Shi Y.-Z. Computational modeling of DNA 3D structures: from dynamics and mechanics to folding. *Molecules*, 2023, vol. 28, no. 12, art. 4833.

[8] B. Knappeová, V. Mlýnský, M. Pykal et al. Comprehensive Assessment of Force-Field Performance in Molecular Dynamics Simulations of DNA/RNA Hybrid Duplexes. *Journal of Chemical Theory and Computation*, 2024, vol. 20, no. 15, pp. 6917–6929.

[9] De Silva N., Perez A. Predicting rare DNA conformations via dynamical graphical models: a case study of the B→A transition. *Nucleic Acids Research*, 2025, vol. 53, no. 13. DOI: 10.1093/nar/gkaf601.

[10] Eidelman, Y. A., Slanina, S. V., Gusev, O. A., Andreev, S. G. High-performance computational modeling of chromosome conformations. *Supercomputing Frontiers and Innovations*, 2018, vol. 5, no. 3, pp. 38–41. DOI: https://doi.org/10.14529/jsfi180305.

[11] M. Ohno, T. Ando, D. G. Priest, et al. Sub-nucleosomal genome structure reveals distinct nucleosome folding motifs. *Cell*, 2019, vol. 176, no. 3, pp. 520–534. DOI: 10.1016/j.cell.2018.12.014.

[12] Liang, J., A. Perez-Rathke. Minimalistic 3D chromatin models: Sparse interactions in single cells drive the chromatin fold and form many-body units. *Current Opinion in Structural Biology*, 2021, vol. 71, pp. 200–214. DOI: 10.1016/j.sbi.2021.06.017.

[13] Vercellone F. A., Chiariello A. M., Esposito A. [et al.] Multiscale perspective on chromatin architecture through polymer physics. *Physiology*, 2025, vol. 40, no. 3. DOI: 10.1152/physiol.00050.2024

[14] Christen M., Hünenberger P. H., Bakowies D. [et al.] The GROMOS software for biomolecular simulation: GROMOS05. *Journal of Computational Chemistry*, 2005, vol. 26, no. 16, pp. 1719–1751. DOI: 10.1002/jcc.20303.

[15] Harder E., Damm W., Maple J. [et al.] OPLS3: A force field providing broad coverage of drug-like small molecules and proteins. *Journal of Chemical Theory and Computation*, 2016, vol. 12, no. 1, pp. 281–296. DOI: 10.1021/acs.jctc.5b00864.

[16] Ahmad K., Javed A. T., Lanphere C. [et al.] Structure and dynamics of an archetypal DNA nanoarchitecture revealed via cryo-EM and molecular dynamics simulations. *Nature Communications*, 2023, vol. 14, no. 1, art. 3630. DOI: 10.1038/s41467-023-38681-5.

[17] Case D. A., Aktulga H. M., Belfon K. [et al.] *Amber 2024 Reference Manual*. San Francisco, University of California, 2024, 1043 p.

[18] Case D. A., Cerutti D. S., Cruzeiro V. W. D. [et al.]. Recent developments in amber biomolecular simulations. *Journal of Chemical Information and Modeling*, 2025, vol. 65, no. 15, pp. 7835–7843. DOI: 10.1021/acs.jcim.5c01063.

[19] Maier J. A., Martinez C., Kasavajhala K. [et al.] ff14SB: Improving the Accuracy of Protein Side Chain and Backbone Parameters from ff99SB. *Journal of Chemical Theory and Computation*, 2015, vol. 11, no. 8, pp. 3696–3713. DOI: 10.1021/acs.jctc.5b00255

[20] URL: https://bmrb.io/metabolomics/mol_summary/?molName=adenine [Electronic resource].

[21] URL: https://bmrb.io/metabolomics/mol_summary/?molName=guanine [Electronic resource].

[22] URL: https://bmrb.io/metabolomics/mol_summary/?molName=cytosine [Electronic resource].

[23] URL: https://www.chem.purdue.edu/jmol/molecules/thymine.html# [Electronic resource].

[24] URL: https://pubchem.ncbi.nlm.nih.gov/compound/439576#section=2D-Structure [Electronic resource].

[25] URL: https://pubchem.ncbi.nlm.nih.gov/compound/2-Deoxyribose-5-phosphate [Electronic resource].

[26] URL: https://pubchem.ncbi.nlm.nih.gov/compound/12599 [Electronic resource].

[27] URL: https://pubchem.ncbi.nlm.nih.gov/compound/135398597 [Electronic resource].

[28] URL: https://pubchem.ncbi.nlm.nih.gov/compound/13945 [Electronic resource].

[29] URL: https://pubchem.ncbi.nlm.nih.gov/compound/9700 [Electronic resource].

Plokhotnikov K.E., Dr. Sc. (Phys. — Math.), Leading scientist of Department of Mathematical Modeling and Computer Science, Faculty of Physics, Lomonosov Moscow State University. e-mail: psygma@yandex.ru