


Перевод на английский язык <https://vavilov.elpub.ru/jour>

Применение метода взвешенных гистограмм для расчета термодинамических параметров формирования комплексов олигодезоксирибонуклеотидов

И.И. Юшин^{1,2}, В.М. Голышев^{1,2}, Д.В. Пышный¹, А.А. Ломзов^{1,2} 

¹ Институт химической биологии и фундаментальной медицины Сибирского отделения Российской академии наук, Новосибирск, Россия

² Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия


 lomzov@niboch.nsc.ru

Аннотация. На сегодняшний день разработан широкий спектр производных и аналогов нуклеиновых кислот. Некоторые из них нашли применение при решении научно-исследовательских задач и задач биомедицины. Детальная информация о свойствах таких соединений является основой их эффективного использования. Одну из наиболее значимых физико-химических характеристик олигонуклеотидов – термодинамическую стабильность их дуплексов с ДНК и РНК – можно рассчитывать лишь для некоторых производных нуклеиновых кислот: LNA, мостиковых олигонуклеотидов и PNA. Существующие подходы основаны на анализе экспериментальных данных и построении прогностических моделей. Проводятся пилотные исследования, направленные на разработку методов прогнозирования свойств нуклеиновых кислот с использованием методов компьютерного моделирования, основанные только на знании структуры олигомеров. В данной работе исследована применимость метода взвешенных гистограмм (WHAM) при анализе зонтичной выборки для расчета термодинамических параметров формирования ДНК-дуплексов: изменения энтальпии ΔH° , энтропии ΔS° и свободной энергии Гиббса ΔG_{37}° . Отработана процедура расчета гибридационных свойств олигодезоксирибонуклеотидов с использованием метода взвешенных гистограмм. Подобраны оптимальные параметры проведения моделирования и расчета термодинамических параметров. На примере представительной выборки из 21 олигонуклеотида длиной от 4 до 16 нт и долей G/C пар от 14 до 100 % показана возможность расчета ΔH° , ΔS° и ΔG_{37}° . Ошибки расчета термодинамических параметров составляют 11.4, 12.9 и 11.8 % соответственно, а температура плавления прогнозируется со средней ошибкой 5.5 °C. Такая высокая точность расчетов сопоставима с экспериментальной и с другими прогностическими методами расчета энергии комплексообразования. В настоящей работе впервые систематически исследовано применение метода WHAM для расчета энергии формирования ДНК-дуплексов. Полученные результаты показывают потенциальную возможность достоверного расчета гибридационных свойств новых, в том числе еще не синтезированных производных нуклеиновых кислот. Это открывает новые горизонты для рационального дизайна конструкций на основе нуклеиновых кислот для решения задач биомедицины и биотехнологии.

Ключевые слова: ДНК; гибридизация; термодинамические параметры; свободная энергия Гиббса; метод взвешенных гистограмм; WHAM; молекулярная динамика.


Для цитирования: Юшин И.И., Голышев В.М., Пышный Д.В., Ломзов А.А. Применение метода взвешенных гистограмм для расчета термодинамических параметров формирования комплексов олигодезоксирибонуклеотидов. *Вавиловский журнал генетики и селекции*. 2023;27(7):807-814. DOI 10.18699/VJGB-23-93

Application of the weighted histogram method for calculating the thermodynamic parameters of the formation of oligodeoxyribonucleotide duplexes

I.I. Yushin^{1,2}, V.M. Golyshev^{1,2}, D.V. Pyshnyi¹, A.A. Lomzov^{1,2} 

¹ Institute of Chemical Biology and Fundamental Medicine of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia

² Novosibirsk State University, Novosibirsk, Russia

 lomzov@niboch.nsc.ru

Abstract. To date, many derivatives and analogs of nucleic acids (NAs) have been developed. Some of them have found uses in scientific research and biomedical applications. Their effective use is based on the data about their properties. Some of the most important physicochemical properties of oligonucleotides are thermodynamic parameters of the formation of their duplexes with DNA and RNA. These parameters can be calculated only for a few NA derivatives: locked NAs, bridged oligonucleotides, and peptide NAs. Existing predictive approaches are based on an analysis of experimental data and the consequent construction of predictive models. The ongoing pilot studies aimed at devising methods for predicting the properties of NAs by computational modeling techniques are based only on knowledge about the structure of oligonucleotides. In this work, we studied the applicability of the weighted

histogram analysis method (WHAM) in combination with umbrella sampling to the calculation of thermodynamic parameters of DNA duplex formation (changes in enthalpy ΔH° , entropy ΔS° , and Gibbs free energy ΔG_{37}°). A procedure was designed involving WHAM for calculating the hybridization properties of oligodeoxyribonucleotides. Optimal parameters for modeling and calculation of thermodynamic parameters were determined. The feasibility of calculation of ΔH° , ΔS° , and ΔG_{37}° was demonstrated using a representative sample of 21 oligonucleotides 4–16 nucleotides long with a GC content of 14–100 %. Error of the calculation of the thermodynamic parameters was 11.4, 12.9, and 11.8 % for ΔH° , ΔS° , and ΔG_{37}° , respectively, and the melting temperature was predicted with an average error of 5.5 °C. Such high accuracy of computations is comparable with the accuracy of the experimental approach and of other methods for calculating the energy of NA duplex formation. In this paper, the use of WHAM for computation of the energy of DNA duplex formation was systematically investigated for the first time. Our results show that a reliable calculation of the hybridization parameters of new NA derivatives is possible, including derivatives not yet synthesized. This work opens up new horizons for a rational design of constructs based on NAs for solving problems in biomedicine and biotechnology.

Key words: DNA; hybridization; thermodynamic parameters; Gibbs free energy; Weighted Histogram Analysis Method; WHAM; molecular dynamics.

For citation: Yushin I.I., Golyshev V.M., Pyshnyi D.V., Lomzov A.A. Application of the weighted histogram method for calculating the thermodynamic parameters of the formation of oligodeoxyribonucleotide duplexes. *Vavilovskii Zhurnal Genetiki i Seleksii = Vavilov Journal of Genetics and Breeding*. 2023;27(7):807-814. DOI 10.18699/VJGB-23-93

Введение

На сегодняшний день разработан широкий спектр производных и аналогов нуклеиновых кислот. Некоторые из них нашли применение при решении научно-исследовательских задач и задач биомедицины (Wang et al., 2022). Возможность эффективного использования нуклеиновых кислот (НК) обусловлена наличием детальной информации об их физико-химических, молекулярно-биологических и биологических свойствах. Данная информация доступна лишь для ограниченного числа производных НК, таких как замкнутые НК (LNA) (McTigue et al., 2004), пептидные НК (PNA) (Griffin, Smith, 1998), фосфогриотатные производные (PS) (Eckstein, 2014), фосфорамидатные морфолиновые олигомеры (PMO) (Summerton, Weller, 1997), мозаичные олигонуклеотиды (BO) (Lomzov et al., 2006). Разработка подходов прогнозирования свойств нуклеиновых кислот, их аналогов и производных совершенно необходима для рационального дизайна олигонуклеотидных конструкций при решении всех вышеописанных задач. Наличие таких инструментов в существенной мере упрощает процесс научных исследований, в которые вовлечены такие соединения, а также создания коммерческих продуктов, например систем молекулярной диагностики или терапевтических нуклеиновых кислот.

Одним из основных физико-химических свойств производных НК является их способность и эффективность формирования комплексов с комплементарными последовательностями ДНК и РНК. Разработаны модели прогностического расчета термодинамических характеристик формирования дуплексных структур ДНК (SantaLucia, Hicks, 2004) и РНК (Xia et al., 1998), гибридных ДНК/РНК (Sugimoto et al., 1995; Banerjee et al., 2020) и некоторых производных нуклеиновых кислот: LNA (McTigue et al., 2004), BO (Lomzov et al., 2006) и PNA (Griffin, Smith, 1998). Они основаны на анализе экспериментальных данных о гибридационных свойствах этих олигомеров и построении с их использованием прогностических аналитических моделей. Кроме того, проводятся пилотные исследования, направленные на разработку методов достоверной оценки энергии формирования комплексов нуклеиновых кислот с использованием методов компьютерного моделирования.

Последние перспективны с точки зрения создания подходов для априорного предсказания свойств производных нуклеиновых кислот, которые еще не были синтезированы. В недавней работе D. Dowerah с коллегами предложен ряд новых аналогов LNA с различными линкерами между O2' и C4' атомами (Dowerah et al., 2023). Это указывает на высокий потенциал и востребованность методов предсказания свойств модифицированных нуклеиновых кислот на основании лишь их химической структуры.

Давно зарекомендовавшим себя подходом для расчета свободной энергии Гиббса является метод анализа взвешенных гистограмм (weighted histogram analysis method, WHAM) при анализе зонтичной выборки (umbrella sampling) (Kumar et al., 1992). Общий принцип расчета заключается в проведении молекулярного моделирования методом зонтичной выборки и анализа полученных траекторий методом WHAM (рис. 1). При молекулярном моделировании методом зонтичной выборки на систему накладывается дополнительный (обычно гармонический) потенциал вдоль координаты реакции (ξ), который удерживает систему в положении ξ_i ($i = 1 \dots i_{\max}$) с определенной силой. Для каждого окна зонтичной выборки (i) получают гистограмму, которая представляет собой распределение вероятности по координате реакции, смещенной удерживающим потенциалом. Широко распространенным методом расчета потенциала средней силы по гистограммам является WHAM. В рамках данного подхода оценивают статистическую неопределенность несмещенного распределения вероятностей с учетом зонтичных гистограмм, а затем рассчитывают потенциал средней силы (PMF), который соответствует наименьшей неопределенности (Kumar et al., 1992). Это позволяет вычислить свободную энергию и другие наблюдаемые величины (Grossfield, 2018).

В настоящей работе исследована возможность расчета энергии образования совершенных ДНК-дуплексов различной длины и нуклеотидного состава с использованием WHAM при анализе зонтичной выборки. Проведение расчета свободной энергии Гиббса формирования дуплексов при различных температурах должно позволить рассчитывать энтальпийный (ΔH°) и энтропийный (ΔS°) вклады. С помощью величин ΔH° и ΔS° можно рассчитать наи-

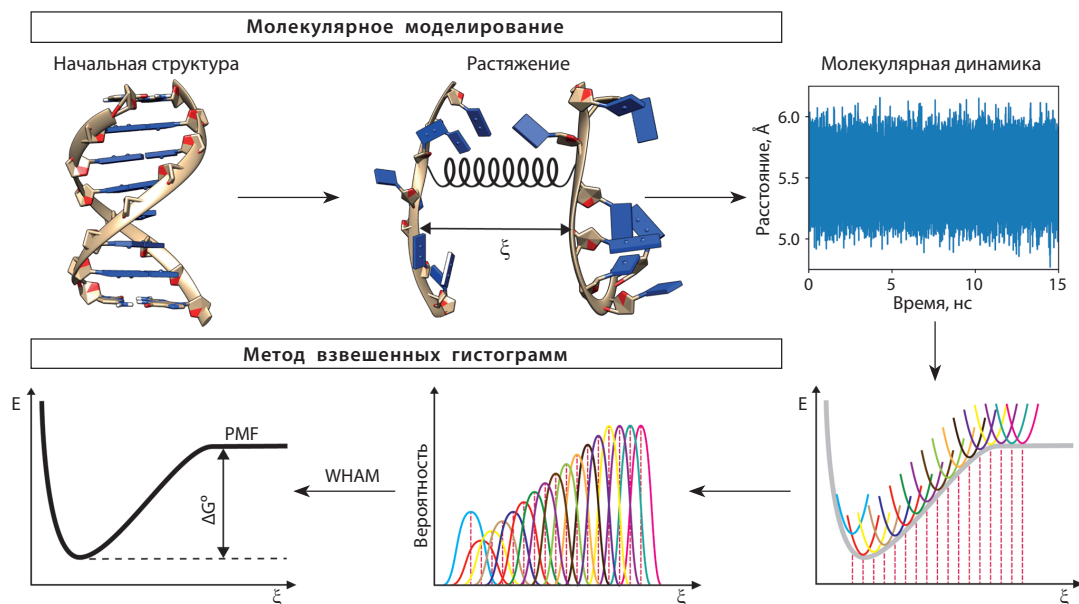


Рис. 1. Протокол проведения расчета свободной энергии Гиббса образования двойной спирали НК с использованием метода взвешенных гистограмм.

более наглядную и широко используемую характеристику для описания термической стабильности комплексов нуклеиновых кислот – температуру плавления ($T_{пл}$).

Методы

Структура ДНК-дуплексов создана в программе NAB пакета программ AmberTools18 (Case et al., 2018). Стартовые структуры имели В-форму двойной спирали.

Моделирование методом молекулярной динамики (МД) выполнено в пакете программ AMBER18 (Case et al., 2018) с параллельными вычислениями на центральных процессорах и графических ускорителях. Моделирование ДНК осуществляли в силовом поле ff99bsc0 (Pérez et al., 2007) в неявной водной оболочке (Tsui, Case, 2000) при фиксированной температуре в диапазоне от 273 до 333 К с шагом 10 градусов. Регулирование температуры проводили с помощью термостата Берендсена и масштабирования скоростей с периодом привязки 10 пс (Omelyan, Kovalenko, 2013). Для возможности использования шага интегрирования уравнений движения 2 фс применяли алгоритм SHAKE.

Процедура моделирования включала восемь шагов:

- 1) создание структуры ДНК-дуплекса и сохранение в формате PDB (с использованием программы NAB пакета AmberTools18). Сохранение структуры в формате файлов amber (tleap);
- 2) минимизация структуры в течение 10 000 шагов (pmemd.cuda);
- 3) ступенчатый нагрев системы от 0 до 100 К за 50 пс и от 100 К до заданной температуры (от 273 до 333 К с шагом 10 К) за 150 пс (pmemd.cuda). Шаг интегрирования 0.5 фс;
- 4) растяжение двух цепей от 0 до 45 Å в течение 10 нс путем наложения потенциала 10 ккал/моль на расстояние между центрами масс выбранных атомов цепей (pmemd.cuda);

5) извлечение из траектории растяжения двух цепей ДНК структур, для которых расстояние между центрами масс составляет от 0 до 45 Å (или 60) с шагом 0.5 Å (pmemd.cuda);

6) моделирование извлеченных структур методом молекулярной динамики в течение 15 нс с наложением гармонического потенциала 10 ккал/моль на расстояние между центрами масс выбранных атомов цепей (pmemd.cuda);

7) расчет энергии взаимодействия цепей методом взвешенных диаграмм в программе WHAM (Grossfield, 2018). Число точек вдоль координаты реакции для дискретизации профиля свободной энергии было выбрано равным 150 (см. ниже), критерий сходимости WHAM 10^{-6} ;

8) расчет энергий взаимодействия цепей методом покомпонентного расчета изменения свободной энергии на основе симуляции методом молекулярной динамики по обобщенной модели Борна (Molecular Mechanics/Generalized Born Surface Area, MMGBSA) выполнен с использованием модуля MMPBSA.py пакета AmberTools18.

Структуры молекул визуализировали в программе UCSF Chimera (Pettersen et al., 2004).

Результаты и обсуждение

Для отработки протокола моделирования был выбран набор ДНК олигомеров различной длины (от 4 до 16 п. о.) и GC-состава (от 14 до 100 %). Нуклеотидные последовательности приведены в таблице. Общий протокол моделирования и анализа представлен на рис. 1. Мы выбрали метод проведения исследований, в котором в качестве координаты реакции использовали расстояние между двумя цепями ДНК. То есть осуществляли поэтапное разнесение двух цепей в пространстве и рассчитывали потенциал средней силы (PMF) в зависимости от расстояния между ними. Этот подход с привлечением метода взвешен-

Термодинамические параметры, рассчитанные методами WHAM, MMGBSA и определенные экспериментально

Последовательность олигонуклеотидов от 5'-к 3'-концу	WHAM					MMGBSA					Эксперимент				
	ΔH°	ΔS°	ΔG_{37}°	$T_{пл}$	R^2	ΔH°	ΔH°	ΔS°	ΔG_{37}°	$T_{пл}$					
AATTGGAC	-43.7	-115.6	-7.8	36.1	0.918	-77.4 ± 5.1	-56.9	-161	-6.9	31.7					
ACGACCTC	-64.0	-169.1	-11.5	55.3	0.959	-85.1 ± 7.3	-59.8	-165	-8.6	40.5					
AGAGCTCT	-64.3	-166.1	-12.8	62.5	0.958	-78.4 ± 8.1	-49.8	-134	-8.2	38.8					
AGCATTAGACGGACCT	-166.0	-434.1	-31.3	87.8	0.960	-162.8 ± 7.9	-123.9	-335	-19.9	70.4					
AGCCG	-39.0	-103.1	-7.0	29.8	0.923	-58.0 ± 5.8	-39.0	-108	-5.5	18.7					
AGTTGC	-31.2	-82.0	-5.8	16.8	0.857	-65.4 ± 8.4	-37.0	-101	-5.7	19.0					
ATATGGAC	-46.4	-130.6	-5.9	23.9	0.907	-77.6 ± 7.5	-53.8	-153	-6.5	28.0					
CAAATAAAG	-67.9	-208.1	-3.4	17.5	0.963	-76.7 ± 8.6	-58.6	-168	-6.5	29.5					
CACAG	-26.6	-71.2	-4.6	2.0	0.979	-56.6 ± 6.5	-33.7	-97	-3.6	1.7					
CCGCGG	-60.8	-158.3	-11.7	57.2	0.932	-83.9 ± 7.6	-41.4	-106	-8.4	41.4					
CGCG	-27.9	-68.3	-6.7	23.7	0.892		-36.3	-103	-4.5	9.1					
CGCGCG	-45.5	-113.8	-10.2	52.9	0.905	-79.6 ± 6.7	-46.4	-121	-8.7	43.3					
GCACCGAC	-92.3	-249.5	-14.9	62.2	0.986	-87.9 ± 7.6	-71.0	-196	-10.2	47.2					
GCATGC	-58.8	-160.4	-9.1	43.1	0.948	-69.5 ± 7.9	-42.2	-117	-6.0	22.7					
GCCCGGAC	-69.1	-182.8	-12.4	58.4	0.949	-94.9 ± 6.4	-61.4	-165	-10.3	48.9					
GCCTGC	-48.3	-126.6	-9.0	44.0	0.915	-73.6 ± 8.6	-37.5	-100	-6.5	25.3					
TACTGGAC	-62.7	-168.9	-10.3	48.9	0.934	-81.0 ± 7.7	-58.5	-165	-7.2	33.7					
TCTATGCA	-44.3	-109.8	-10.3	54.4	0.813	-79.5 ± 6.5	-51.7	-145	-6.6	29.8					
TGCGCA	-61.2	-162.3	-10.8	52.3	0.980	-76.9 ± 6.9	-42.5	-114	-7.3	31.2					
TGTTGC	-41.1	-112.8	-6.1	23.6	0.979	-65.8 ± 7.9	-37.2	-101	-5.8	20.6					
ACATTATTATACA	-148.7	-442.7	-11.4	44.3	0.947	-125.4 ± 13.6	-89.9	-254	-11.1	48.3					

Примечание. Размерность величин термодинамических параметров ΔH° и ΔG_{37}° – ккал/моль, ΔS° – кал/моль/К, $T_{пл}$ – °C. R^2 – коэффициент корреляции Пирсона для линейной зависимости $\Delta G^\circ(T)$. Ошибки экспериментальных значений ΔH° , ΔS° и ΔG_{37}° и $T_{пл}$ составляют 10, 10 и 8 % и 0.5 °C соответственно.

ных гистограмм (WHAM) позволяет определить энергию Гиббса взаимодействия двух цепей непосредственно из компьютерного эксперимента. Проводя такой *in silico* эксперимент при разной температуре, можно вычислить изменение энтальпии и энтропии комплексообразования из линейной температурной зависимости свободной энергии Гиббса. На первом этапе необходимо было определить параметры, оптимальные для проведения расчетов.

В качестве координаты реакции (ξ) мы выбрали расстояние между центрами масс (r) C4' атомов всех нуклеотидов каждой из двух цепей. Начальное расстояние было выбрано равным 0 Å, чтобы рассмотреть возможность «сжатия» двойной спирали и при анализе определить наличие минимума зависимости $\Delta G_T^\circ(r)$, а энергию комплексообразования рассчитывать как разницу между минимумом и максимумом данной зависимости (см. рис. 1). Анализ разнесения центров масс цепей показал, что максимального расстояния в 45 Å достаточно для полной диссоциации комплексов олигонуклеотидов длиной от 4 до 9 п.о., а для комплексов размером 14 и 16 п.о. оно должно быть увеличено до 60 Å. При выбранной таким образом координате реакции диссоциация двух цепей для большинства комплексов олигонуклеотидов происходит в соответствии

с моделью «застежка-молния» (Cantor, Schimmel, 1980; Volkov, Solov'ov, 2009), с расплетанием двойной спирали с одного из концов, либо в смешанном режиме «сдвиг/застежка-молния» (Mosayebi et al., 2015; Kurus, Dultsev, 2018). Пример изменения конформаций олигомеров вдоль координаты реакции приведен в Приложении 1¹. Механизм диссоциации комплексов в проводимом исследовании не является критически важным ввиду того, что рассматриваются только два предельных состояния: релаксированная структура дуплекса и два одноцепочечных не взаимодействующих олигонуклеотида. Соответствие механизма перехода спираль–клубок наблюдаемым экспериментальными методами подтверждает адекватность выбранного подхода для описания диссоциации двойной спирали ДНК.

Согласно общепринятым требованиям для проведения WHAM анализа, необходимо, чтобы перекрытие между гистограммами было не менее 20 %. Анализ показал, что это достигается при ~0.7 Å между соседними окнами моделирования. Пример зависимости гистограмм распределения от расстояния между цепями для дуплекса

¹ Приложения 1–8 см. по адресу:
<https://vavilovj-icg.ru/download/pict-2023-27/appx26.pdf>

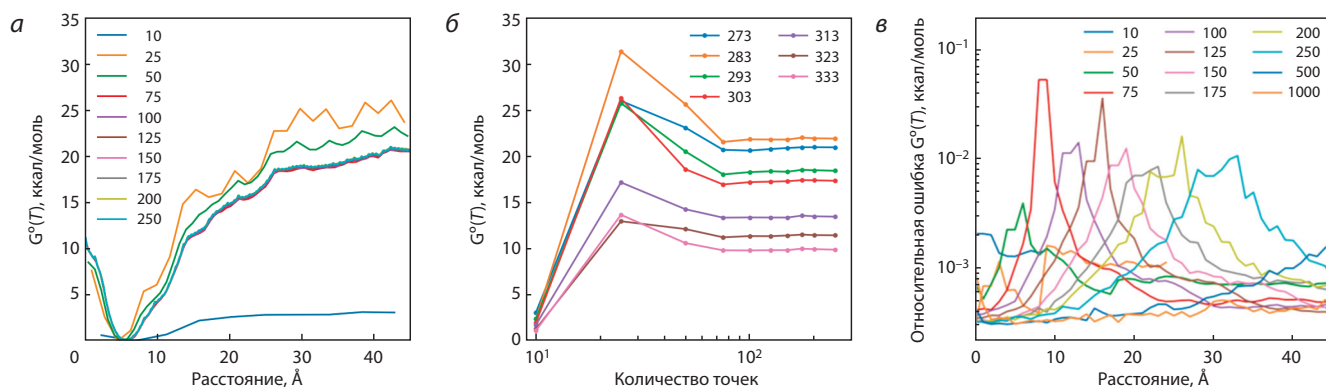


Рис. 2. Поиск параметров моделирования и анализа MD траекторий.

a – зависимость профиля свободной энергии Гиббса от расстояния между центрами масс C4' атомов углерода двух цепей ДНК при различном числе точек вдоль координаты реакции, выбранных для дискретизации профиля свободной энергии при 273 К; *б* – зависимость энергии Гиббса формирования комплекса от числа точек вдоль координаты реакции, выбранных для дискретизации профиля свободной энергии; *в* – зависимость относительной ошибки расчета свободной энергии Гиббса от расстояния между центрами масс C4' атомов углерода двух цепей ДНК при различном числе точек вдоль координаты реакции, выбранных для дискретизации профиля свободной энергии.

5'-GCACCGAC-3'/5'-GTCGGTGC-3' приведен в Приложении 2. Мы выбрали шаг координаты реакции равным 0.5 Å, чтобы заведомо соответствовать данному критерию.

При расчете энергии методом WHAM важным параметром является число точек (bins) вдоль координаты реакции, выбранных для дискретизации профиля свободной энергии. При числе точек 100 и более (до 1000 разбиений) достигается плато как для формы и положения профилей свободной энергии Гиббса (рис. 2, *a*, Приложение 3), так и для величин изменения свободной энергии Гиббса при различных температурах (см. рис. 2, *б*). При этом величины относительных значений ошибок, рассчитанных методом бутстреп (Grossfield, 2018), не превышают 6 % (см. рис. 2, *в*).

Свободную энергию Гиббса при определенной температуре рассчитывали как разницу между минимумом и максимумом на профиле потенциала средней силы: $\Delta G^\circ(T) = \text{PMF}_{\min} - \text{PMF}_{\max}$. Для получения температурных зависимостей свободной энергии Гиббса был выбран диапазон от 273 до 333 К с шагом 10 К. Нижнее значение выбрано в соответствии с температурой заморозки воды, а верхнее – чтобы ограничить денатурацию цепей НК при моделировании в течение заданного диапазона времени для коротких олигомеров. Такой диапазон является достаточно широким, чтобы можно было построить зависимость $\Delta G^\circ(T)$ для достоверного определения ΔH° и ΔS° методом линейной регрессии с использованием уравнения $\Delta G^\circ(T) = \Delta H^\circ - T\Delta S^\circ$.

Длина траектории при моделировании методом молекулярной динамики при каждом фиксированном расстоянии между выбранными центрами масс и заданной температуре была выбрана равной 15 нс, чтобы получить минимально достоверную траекторию в невязной водной оболочке (Lomzov et al., 2015). Таким образом, для каждого дуплекса были получены траектории длиной 15 нс при 90 (или 120) различных расстояниях между центрами масс при семи значениях температуры. Длина траектории для каждого дуплекса варьировала от 9.45 до 12.6 мкс. Полная продолжительность траекторий для всех комплексов составила более 200 мкс.

Проведен расчет методом WHAM зависимости энергии Гиббса взаимодействия двух олигонуклеотидов от расстояния между центрами масс C4' атомов каждой из цепей (*r*) при семи исследованных температурах для 21 исследованного комплекса (см. таблицу). Типичная зависимость энергий Гиббса комплексообразования от *r* при температурах от 273 до 333 К для комплекса 5'-GCACCGAC-3'/5'-GTCGGTGC-3' приведена на рис. 3, *a*. Зависимость свободной энергии Гиббса имеет четко выраженный минимум в районе 6 Å и возрастает при сжатии и растяжении двойной спирали. При растяжении график зависимости проходит через максимум и незначительно снижается. Точка максимума соответствует расстоянию, на котором прекращается взаимодействие между цепями.

Для оценки адекватности моделирования мы провели сравнение геометрии двойной спирали релаксированной формы ДНК-дуплекса 5'-GCACCGAC-3'/5'-GTCGGTGC-3' с литературными данными (Приложение 4). Все структурные параметры хорошо согласуются с данными для структуры додекамера Дрю-Диккерсона (DDD, 5'-CGCGAATTCGCG-3'), определенной экспериментально методами ЯМР спектроскопии (PDB ID: 1NAJ) и рентгеновской кристаллографии (PDB ID: 1BNA).

Проведен расчет свободной энергии Гиббса комплексообразования при различных температурах ($\Delta G^\circ(T)$). Установлено, что данная зависимость имеет линейный характер с высоким коэффициентом корреляции R^2 (более 0.83); среднее значение для всех исследованных комплексов составило 0.93 (см. рис. 3, *б* и таблицу). На основании полученных зависимостей (Приложение 5) рассчитаны величины изменения энтальпии и энтропии комплексообразования (см. таблицу). Сопоставление рассчитанных методом WHAM величин термодинамических параметров с величинами, определенными экспериментально (по данным (Lomzov et al., 2015)), показывает линейную связь между ними с высокими коэффициентами корреляции R^2 : 0.87, 0.82, 0.88 и 0.75 для ΔH° , ΔS° , ΔG_{37}° и $T_{пл}$ соответственно (Приложение 6). В качестве температуры плавления выбирали такое значение, при котором половина олигонуклеотидов находится в двуцепочечном,

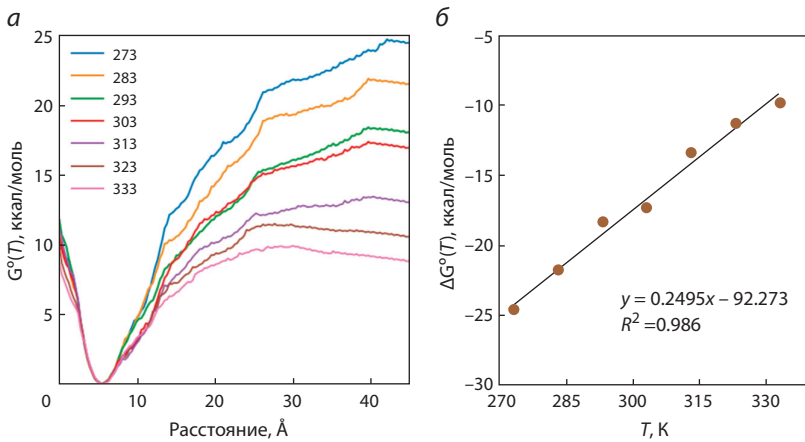


Рис. 3. Зависимость свободной энергии Гиббса: *a* – от расстояния между молекулами при различных температурах (273, 283, 293, 303, 313, 323, 333 K); *б* – от температуры модельного дуплекса 5'-GCACCGAC-3'/5'-GTCGGTGC-3'.

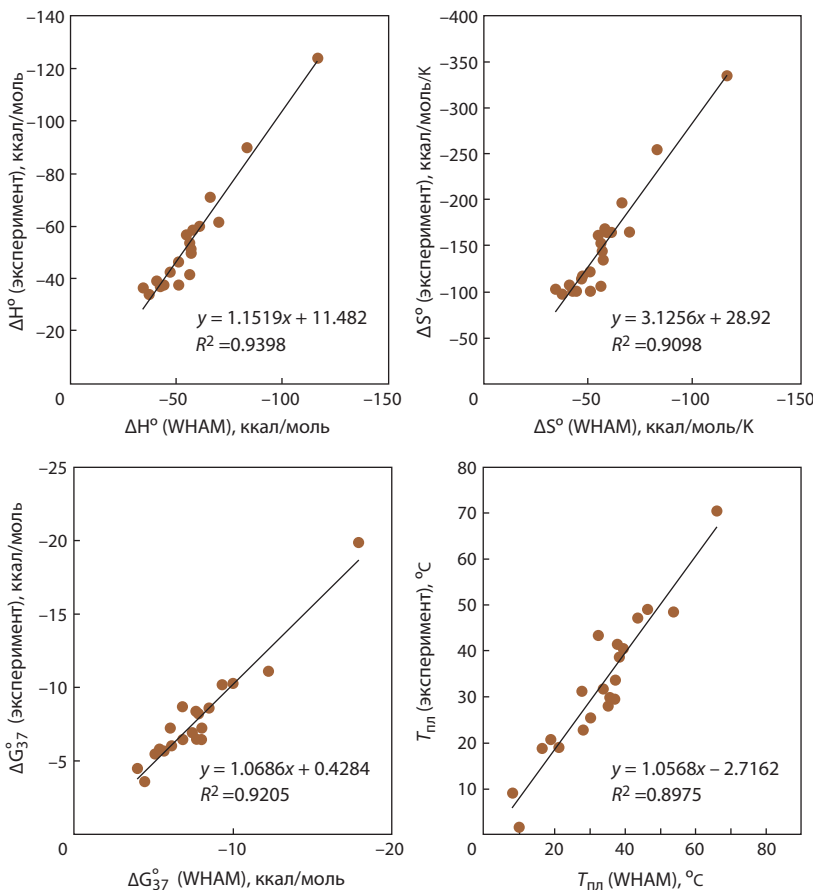


Рис. 4. Корреляция термодинамических параметров ΔH° , ΔS° и ΔG_{37}° и температуры плавления комплексов, рассчитанных методом WHAM с учетом линейных поправок, с величинами, определенными экспериментально (по данным (Lomzov et al., 2015)).

а вторая – в одноцепочечном состоянии. Величина $T_{пл}$ была рассчитана с использованием термодинамических параметров (Lomzov, Pyshnyi, 2012):

$$T_{пл} = \Delta H^\circ / (\Delta S^\circ + R \ln \left[\frac{Ct}{4} \right]),$$

где R – универсальная газовая постоянная, Ct – полная концентрация олигонуклеотидов в системе. Значение Ct было взято равным 10 мкМ в

соответствии с типичными экспериментальными значениями.

Наклон линейной зависимости термодинамических параметров близок к 0.5, а величины свободных членов значительны по сравнению с анализируемыми величинами (см. Приложение 6). Поэтому, как предложено в наших предыдущих работах (Lomzov et al., 2015; Golyshev et al., 2021), можно ввести линейные поправки на значения рассчитанных термодинамических параметров ΔH° и ΔS° . После применения такой коррекции коэффициенты корреляции для свободной энергии Гиббса и температур плавления существенно улучшаются – до 0.94 и 0.86 соответственно (рис. 4). При этом средние абсолютные значения ошибок расчета термодинамических параметров становятся равными 11.4, 12.9 и 11.8 % и 5.5 °C для ΔH° , ΔS° и ΔG_{37}° и $T_{пл}$ соответственно. Для данной выборки аналогичные значения ошибок величин термодинамических характеристик, полученных методом MMGBSA в работах (Lomzov et al., 2015; Golyshev et al., 2021) с учетом линейных поправок, несколько ниже и составляют 7.6, 11.4 и 10.6 % и 4.3 °C. Точность расчета термодинамических параметров в настоящей работе сопоставима с экспериментальной и с точностью при использовании метода ближайших соседей – наиболее распространенного метода расчета эффективности комплексообразования олигонуклеотидов, которая составляет около 10 % для энтальпии и энтропии и около 8 % для свободной энергии Гиббса комплексообразования (SantaLucia, Hicks, 2004; Lomzov et al., 2006).

Для дополнительной проверки качества результатов траектории проанализировали методом MMGBSA и сопоставили рассчитанные величины с данными, рассчитанными методом WHAM, и с ранее полученными методом MMGBSA величинами (Lomzov et al., 2015). Типичный вид зависимости энергии MMGBSA от расстояния между центрами масс C4' атомов цепей схож с зависимостью свободной энергии Гиббса от расстояния, приведенной на рис. 3, *a* (Приложение 7). При расстояниях, близких к максимальному, значение энергии формирования двойной спирали ДНК выходит на плато, равное нулю, что свидетельствует об отсутствии взаимодействия между цепями при данном способе анализа траекторий. Наблюдается дно потенциальной ямы в области 2–7 Å, что соответствует релаксированной форме двойной спирали ДНК, а ее глобаль-

ный минимум в районе 7 Å близок к минимуму, наблюдающемуся в зависимости энергии Гиббса, определенной в рамках WHAM (см. рис. 3, *a* и Приложение 7). Присутствует слабая зависимость энергии комплексообразования, рассчитанной методом MMGBSA, от температуры, что соответствует малому значению изменения теплоемкости ΔC_p . Достоверно определить величину изменения теплоемкости из компьютерных экспериментов не представляется возможным ввиду больших ошибок расчета.

Величины энтальпии комплексообразования, рассчитанные в настоящей работе и определенные ранее, хорошо коррелируют ($R^2 = 0.97$) с наклоном, близким к единице (0.95), и свободным членом линейной зависимости, близким к нулю (4 ккал/моль) (Приложение 8, *a*). Кроме того, аналогичная линейная корреляция наблюдается для величин энтальпии комплексообразования, рассчитанных методами MMGBSA и WHAM в данной работе. Таким образом, полученные нами МД траектории релевантны.

Одним из важных аспектов в исследованном ранее подходе расчета энергии методом MMGBSA является неопределенность, связанная со структурой одноцепочечного состояния олигонуклеотидов. Ее извлекали из траектории двойной спирали. Тем не менее это позволило вычислить энтальпию комплексообразования с достаточно хорошей точностью. В данной работе при анализе методом взвешенных гистограмм одноцепочечное состояние олигонуклеотидов неплохо представлено в МД траекториях (насколько это возможно сделать в рамках приближения неявной водной оболочки и выбранного силового поля). Такой подход дает хорошие результаты при расчете энергии формирования двойной спирали. Вместе с тем основным преимуществом использования метода WHAM является расчет непосредственно изменения свободной энергии Гиббса формирования двойной спирали ДНК. Тот факт, что эта характеристика оказалась в наших расчетах линейной в широком диапазоне исследованных температур, указывает на то, что выбранные параметры моделирования, а также заложенные в моделирование и анализ модели достаточно хорошо описывают физику как дву-, так и одноцепочечной ДНК. Для последней в пользу этого утверждения свидетельствует конформация олигонуклеотидов при моделировании цепей с большим расстоянием между их центрами масс (см. Приложение 1). Олигомеры не остаются линейными, как в составе дуплекса, что использовали при анализе методом MMGBSA, и не становятся полностью разупорядоченными цепочками, а сохраняют несколько гетероциклических оснований подряд в стэкинге. Это коррелирует с персистентной длиной одноцепочечных олигонуклеотидов, которая составляет несколько нуклеотидов (в зависимости от нуклеотидной композиции и ионной силы раствора) (Chen et al., 2012). Кроме того, наблюдающаяся линейная зависимость $\Delta G^\circ(T)$ в исследованном нами методе позволяет напрямую вычислять энтропию комплексообразования.

Вместе с тем разработанный подход далек от совершенства. В частности, для более достоверного моделирования структуры и динамики ДНК необходимо применять наиболее современные силовые поля и модель явной водной

оболочки. Анализ параметров силового поля для такого моделирования является отдельной сложной задачей. Кроме того, сложность и продолжительность расчетов в явной водной оболочке кратно увеличиваются. Так, основные вычислительные затраты происходят на этапе молекулярно-динамического расчета. Для девятизвенного ДНК-дуплекса, детально проанализированного в данной работе, скорость расчета в неявной водной оболочке при использовании современной видеокарты NVIDIA GTX3080 составляет около 800 нс/день. Таким образом, время расчета одного модельного дуплекса составляет 12 суток. В случае явной водной оболочки моделируемая периодическая ячейка будет содержать около 15–20 тыс. молекул ввиду максимального расстояния между цепями 45 Å. Это снизит производительность до ~100 нс/день или составит около 3 месяцев. В явной водной оболочке существенно снизится конформационная подвижность ДНК, что потребует увеличения длины траектории для каждого окна моделирования и, соответственно, увеличения вычислительных затрат. Тем не менее такое усложнение представляется необходимым для увеличения достоверности расчетов.

Еще одним перспективным направлением развития исследованного подхода является его апробация на примерах известных модифицированных нуклеиновых кислот. Это должно ответить на вопрос о его применимости для рационального дизайна химической структуры новых, еще не синтезированных химически производных нуклеиновых кислот, что может быть использовано для решения конкретных задач биомедицины и биотехнологии. Проведенный анализ показывает высокий потенциал и реалистичность применения метода WHAM для расчета энергии формирования комплексов нуклеиновых кислот, их аналогов и производных.

Заключение

Отработана процедура метода взвешенных гистограмм для расчета гибридационных свойств олигодезоксирибонуклеотидов. Подобраны оптимальные параметры для проведения моделирования и расчета термодинамических параметров формирования ДНК-дуплексов. На примере представительной выборки, содержащей 21 олигонуклеотид длиной от 4 до 16 нт и долей G/C пар от 14 до 100 %, показана возможность расчета энтальпии, энтропии и свободной энергии Гиббса формирования комплексов олигонуклеотидов методом взвешенных гистограмм (WHAM) при анализе МД траекторий с использованием в качестве координаты реакции расстояния между центрами масс C4' атомов углерода каждой из цепей. Установлена линейная зависимость свободной энергии Гиббса от температуры, при которой проводится моделирование. Это позволяет проводить расчет энтальпии и энтропии комплексообразования из анализа результатов, полученных методом взвешенных гистограмм. Рассчитываемые термодинамические параметры линейно коррелируют с экспериментально определенными величинами с высоким коэффициентом корреляции R^2 (более 0.83). С учетом линейной поправки данной зависимости величины ошибок расчета значений термодинамических параметров сопоставимы с экспериментальными и составляют 11.4, 12.9

и 11.8 % для ΔH° , ΔS° и ΔG_{37}° , а температура плавления прогнозируется со средней ошибкой 5.5 °C. Таким образом, впервые систематически исследовано применение метода WHAM для расчета энергии формирования ДНК-дуплексов. Установлена высокая точность таких расчетов, сопоставимая с экспериментальной и другими методами расчета энергии комплексообразования.

Список литературы / References

- Banerjee D., Tateishi-Karimata H., Ohya T., Ghosh S., Endoh T., Takahashi S., Sugimoto N. Improved nearest-neighbor parameters for the stability of RNA/DNA hybrids under a physiological condition. *Nucleic Acids Res.* 2020;48(21):12042-12054. DOI 10.1093/nar/gkaa572
- Cantor C.R., Schimmel P.R. *Biophysical Chemistry. Part I: The Conformation of Biological Macromolecules.* New York: W.H. Freeman & Company, 1980
- Case D.A., Walker R.C., Cheatham T.E., Simmerling C., Roitberg A., Merz K.M., Luo R., Darden T. *Amber 18. Reference Manual.* San Francisco: Univ. of California, 2018
- Chen H., Meisburger S.P., Pabit S.A., Sutton J.L., Webb W.W., Pollack L. Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. *Proc. Natl. Acad. Sci. USA.* 2012;109(3):799-804. DOI 10.1073/pnas.1119057109
- Dowerah D., Uppuladinne M.V.N., Sarma P.J., Biswakarma N., Sonavane U.B., Joshi R.R., Ray S.K., Namsa N.D., Dekka R.C. Design of LNA analogues using a combined density functional theory and molecular dynamics approach for RNA therapeutics. *ACS Omega.* 2023;8(25):22382-22405. DOI 10.1021/acsomega.2c07860
- Eckstein F. Phosphorothioates, essential components of therapeutic oligonucleotides. *Nucleic Acid Ther.* 2014;24(6):374-387. DOI 10.1089/nat.2014.0506
- Golyshev V.M., Pyshnyi D.V., Lomzov A.A. Calculation of energy for RNA/RNA and DNA/RNA duplex formation by molecular dynamics simulation. *Mol. Biol.* 2021;55(6):927-940. DOI 10.1134/S002689332105006X
- Griffin T.J., Smith L.M. An approach to predicting the stabilities of peptide nucleic acid:DNA duplexes. *Anal. Biochem.* 1998;260(1):56-63. DOI 10.1006/abio.1998.2686
- Grossfield A. WHAM: the weighted histogram analysis method. 2018.
- Kumar S., Rosenberg J.M., Bouzida D., Swendsen R.H., Kollman P.A. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *J. Comput. Chem.* 1992;13(8):1011-1021. DOI 10.1002/jcc.540130812
- Kurus N.N., Dultsev F.N. Determination of the thermodynamic parameters of DNA double helix unwinding with the help of mechanical methods. *ACS Omega.* 2018;3(3):2793-2797. DOI 10.1021/acsomega.7b01815
- Lomzov A.A., Pyshnyi D.V. Considering the oligonucleotide secondary structures in thermodynamic and kinetic analysis of DNA duplex formation. *Biophysics (Oxf).* 2012;57(1):19-34. DOI 10.1134/S0006350912010137
- Lomzov A.A., Pyshnaya I.A., Ivanova E.M., Pyshnyi D.V. Thermodynamic parameters for calculating the stability of complexes of bridged oligonucleotides. *Dokl. Biochem. Biophys.* 2006;409(1):211-215. DOI 10.1134/S1607672906040053
- Lomzov A.A., Vorobjev Y.N., Pyshnyi D.V. Evaluation of the Gibbs free energy changes and melting temperatures of DNA/DNA duplexes using hybridization enthalpy calculated by molecular dynamics simulation. *J. Phys. Chem. B.* 2015;119(49):15221-15234. DOI 10.1021/acs.jpcc.5b09645
- McTigue P.M., Peterson R.J., Kahn J.D. Sequence-dependent thermodynamic parameters for locked nucleic acid (LNA)-DNA duplex formation. *Biochemistry.* 2004;43(18):5388-5405. DOI 10.1021/bi035976d
- Mosayebi M., Louis A.A., Doye J.P.K., Ouldrige T.E. Force-induced rupture of a DNA duplex: from fundamentals to force sensors. *ACS Nano.* 2015;9(12):11993-12003. DOI 10.1021/acsnano.5b04726
- Omelyan I., Kovalenko A. Generalised canonical-isokinetic ensemble: speeding up multiscale molecular dynamics and coupling with 3D molecular theory of solvation. *Mol. Simul.* 2013;39(1):25-48. DOI 10.1080/08927022.2012.700486
- Pérez A., Marchán I., Svozil D., Sponer J., Cheatham T.E., Laughton C.A., Orozco M. Refinement of the AMBER force field for nucleic acids: Improving the description of α/γ conformers. *Biophys. J.* 2007;92(11):3817-3829. DOI 10.1529/biophysj.106.097782
- Pettersen E.F., Goddard T.D., Huang C.C., Couch G.S., Greenblatt D.M., Meng E.C., Ferrin T.E. UCSF Chimera – a visualization system for exploratory research and analysis. *J. Comput. Chem.* 2004;25(13):1605-1612. DOI 10.1002/jcc.20084
- SantaLucia J., Hicks D. The thermodynamics of DNA structural motifs. *Annu. Rev. Biophys. Biomol. Struct.* 2004;33(1):415-440. DOI 10.1146/annurev.biophys.32.110601.141800
- Sugimoto N., Nakano S., Katoh M., Matsumura A., Nakamuta H., Ohmichi T., Yoneyama M., Sasaki M. Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry.* 1995;34(35):11211-11216. DOI 10.1021/bi00035a029
- Summerton J., Weller D. Morpholino antisense oligomers: design, preparation, and properties. *Antisense Nucleic Acid Drug Dev.* 1997;7(3):187-195. DOI 10.1089/oli.1.1997.7.187
- Tsui V., Case D.A. Theory and applications of the generalized born solvation model in macromolecular simulations. *Biopolymers.* 2000;56(4):275-291. DOI 10.1002/1097-0282(2000)56:4<275::AID-BIP10024>3.0.CO;2-E
- Volkov S.N., Solov'yov A.V. The mechanism of DNA mechanical unzipping. *Eur. Phys. J. D.* 2009;54(3):657-666. DOI 10.1140/epjd/e2009-00194-5
- Wang F., Li P., Chu H.C., Lo P.K. Nucleic acids and their analogues for biomedical applications. *Biosensors.* 2022;12(2):93. DOI 10.3390/bios12020093
- Xia T., SantaLucia J., Burkard M.E., Kierzek R., Schroeder S.J., Jiao X., Cox C., Turner D.H. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry.* 1998;37(42):14719-14735. DOI 10.1021/bi9809425

ORCID ID

I.I. Yushin orcid.org/0000-0001-5954-641X
V.M. Golyshev orcid.org/0000-0002-0521-6228

D.V. Pyshnyi orcid.org/0000-0002-2587-3719
A.A. Lomzov orcid.org/0000-0003-3889-9464

Благодарности. Исследование поддержано в рамках государственного задания ИХБФМ СО РАН № 121031300042-1.

Конфликт интересов. Авторы заявляют об отсутствии конфликта интересов.

Поступила в редакцию 15.07.2023. После доработки 20.09.2023. Принята к публикации 21.09.2023.