Вавиловский журнал генетики и селекции. 2023;27(7):807-814
DOI 10.18699/VJGB-23-93

Original Russian text https://vavilovj-icg.ru/

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. Pyshnyi[1], 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 Selektsii = Vavilov Journal of Genetics and Breeding*. 2023;27(7):807-814. DOI 10.18699/VJGB-23-93

Применение метода взвешенных гистограмм для расчета термодинамических параметров формирования комплексов олигодезоксирибонуклеотидов

И.И. Юшин[1, 2], В.М. Голышев[1, 2], Д.В. Пышный[1], 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 гибридизационных 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 гибридизационных 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.

Introduction

To date, a wide range of derivatives and analogs of nucleic acids (NAs) have been developed, many of which have found applications in solving research problems and problems of biomedicine (e. g., (Wang et al., 2022)). Their effective use is possible due to the availability of detailed information about their physicochemical, molecular-biological, and biological properties. This information exists only for a limited number of derivatives of NAs such as locked nucleic acids (LNAs) (McTigue et al., 2004), peptide NAs (Griffin, Smith, 1998), phosphorothioate derivatives (Eckstein, 2014), phosphoramidate morpholino oligomers (Summerton, Weller, 1997), and bridged oligonucleotides (Lomzov et al., 2006). The development of approaches to the prediction of the properties of NAs, their analogs, and derivatives is absolutely necessary for rational design of oligonucleotide constructs in all the abovementioned applications. The availability of such approaches will greatly simplify both scientific research involving such compounds and the creation of commercial products, for example, molecular diagnostic systems or therapeutic NAs.

One of the key physicochemical properties of NA derivatives is their ability to form (and the efficiency of formation of) complexes with complementary sequences of DNA and RNA. Models have been devised to predictively calculate thermodynamic characteristics of the formation of duplex DNA structures (SantaLucia, Hicks, 2004), of duplex RNA structures (Xia et al., 1998), of hybrid DNA/RNA duplexes (Sugimoto et al., 1995; Banerjee et al., 2020), and of some NA derivatives: LNAs (McTigue et al., 2004), bridged oligonucleotides (Lomzov et al., 2006), and peptide NAs (Griffin, Smith, 1998). Such studies are based on analysis of experimental data about hybridization properties of these oligomers with consequent construction of predictive analytical models. In addition, pilot studies are being conducted that are aimed at designing techniques for reliable estimation of formation energy of NA complexes by computer modeling methods. The latter are promising from the standpoint of development of approaches to *a priori* prediction of properties of NA derivatives that have not yet been synthesized. In a recent paper, D. Dowerah and coworkers proposed a series of new analogs of LNAs with different linkers between O2′ and C4′ atoms (Dowerah et al., 2023). This work indicates high potential and demand for methods predicting the properties of modified NAs by means of only their chemical structure.

One well-established approach to the computation of Gibbs free energy is the weighted histogram analysis method (WHAM) combined with an analysis of umbrella sampling (e. g., (Kumar et al., 1992)). The general principle behind this calculation is to carry out molecular modeling by the umbrella sampling procedure and to analyze the resulting trajectories by the WHAM (Fig. 1). In molecular modeling by the umbrella sampling procedure, an additional (usually harmonic) potential is imposed on the system along the reaction coordinate (ξ), and this potential holds the system at position ξ_i ($i = 1 \ldots i_{max}$) with a certain force. For each umbrella sampling window (i), a histogram is obtained that represents a probability distribution along the reaction coordinate skewed by the holding potential. One of the most common techniques for calculating the potential of mean force (PMF) from histograms is the WHAM. Within this approach, researchers estimate statistical uncertainty of an unbiased (unshifted) probability distribution taking into account umbrella histograms and then compute the PMF that corresponds to the lowest uncertainty (Kumar et al., 1992). This approach allows to calculate free energy and other observable parameters (Grossfield, 2018).

In this work, we investigated the feasibility of calculating the formation energy of perfect DNA duplexes having various lengths and GC contents by the WHAM coupled with an analysis of an umbrella sample. The computation of the Gibbs free energy of duplex formation at different temperatures should enable us to calculate enthalpy (ΔH°) and entropy (ΔS°) contributions. By means of ΔH° and ΔS° values, it is possible to calculate the most illustrative and widely used characteristic for describing thermal stability of NA complexes: melting temperature (T_m).

Methods

The structure of DNA duplexes was created using the NAB program from the AmberTools18 software suite (Case et al., 2018). Starting structures had a B-form of the double helix.

A molecular dynamics (MD) simulation was performed in the AMBER18 software (Case et al., 2018) *via* parallel com-

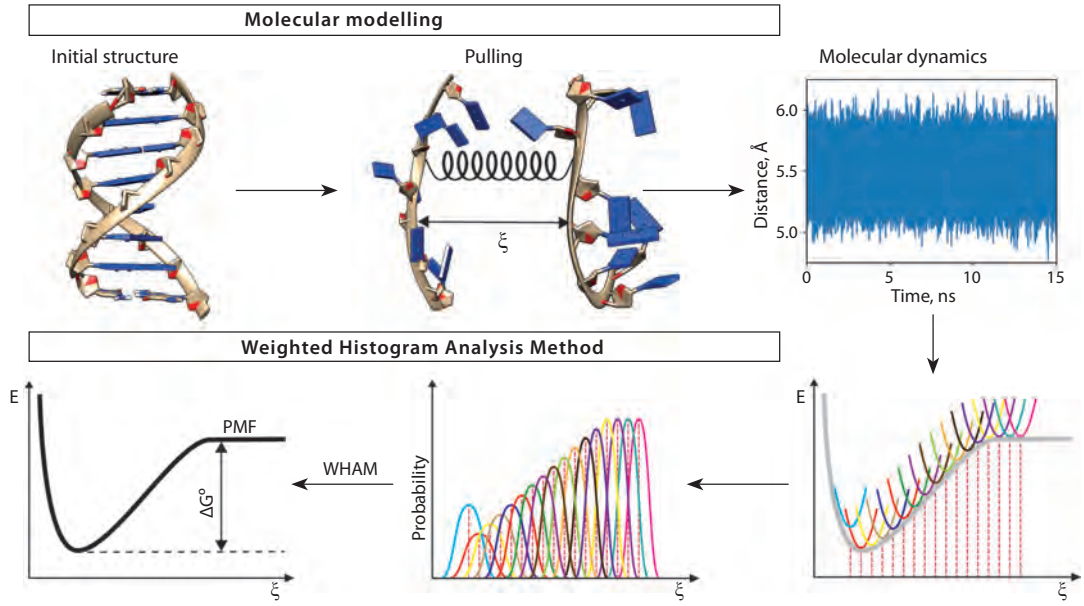

Fig. 1. The protocol for calculating the Gibbs free energy of formation of an NA double helix by the WHAM.

puting on central processing units and graphics accelerators. The ff99bsc0 force field was chosen to model DNA (Pérez et al., 2007). The MD simulation was carried out in an implicit water shell (Tsui, Case, 2000) at a fixed temperature in the range of 273 to 333 K with a step of 10 degrees by means of a Berendsen thermostat with a time constant of 10 ps (Omelyan, Kovalenko, 2013). To enable the step of integration of 2 fs motion equations, we employed the SHAKE algorithm.

The modeling procedure included eight stages:

1. Creating the structure of a DNA duplex and saving it in PDB format (with the help of the NAB program from the AmberTools18 software suite). Saving the structure in the amber file format (tleap).
2. Structure minimization for 10,000 steps (pmemd.cuda).
3. Stepwise heating of the system: from 0 to 100 K for 50 ps and from 100 K to a desired temperature (273 to 333 K in steps of 10 K) for 150 ps (pmemd.cuda). An integration time step of 0.5 fs was used.
4. Separation of two strands from 0 to 45 Å for 10 ns by applying 10 kcal/mol potential to the distance between the centers of mass of selected atoms of the strands (pmemd.cuda).
5. From the separation trajectory of the two DNA strands, extraction of structures for which the distance between the centers of mass was 0 to 45 (or 60) Å with a step of 0.5 Å (pmemd.cuda).
6. MD simulation of the extracted structures for 15 ns with the imposition of 10 kcal/mol harmonic potential on the distance between the centers of mass of the strands' selected atoms (pmemd.cuda).
7. Computation of interaction energy of the strands by the WHAM in the WHAM software (Grossfield, 2018). The number of points along the reaction coordinate for sampling of the free-energy profile was set to 150 (see below), and the convergence criterion of the WHAM was 10^{-6}.
8. Calculation of strand interaction energies *via* component-wise computation of free-energy changes based on MD simulation according to the generalized Born model (molecular mechanics/generalized Born surface area, MMGBSA) was performed using the MMPBSA.py module from the AmberTools18 software suite.

Molecular structures were visualized in the UCSF Chimera software (Pettersen et al., 2004).

Results and discussion

To refine the modeling protocol, a set of DNA oligomers having various lengths (4 to 16 bp) and GC contents (14 to 100 %) was chosen. Nucleotide sequences are given in the Table. The general protocol of the modeling and analysis is presented in Figure 1. We selected an approach where the distance between two DNA strands served as the reaction coordinate. That is, we carried out step-by-step separation of two strands in space and calculated the PMF depending on the distance between them. This approach combined with the WHAM makes it possible to determine the Gibbs energy of interaction between two strands directly in a computational experiment. If such an *in silico* experiment is conducted at different temperatures, it is possible to calculate a change in the enthalpy and entropy of complexation from a linear dependence of Gibbs free energy on temperature. At the first stage, it is necessary to determine the optimal parameters for performing such calculations.

As the reaction coordinate (ξ), we chose the distance (r) between centers of mass of C4′ atoms of all nucleotides from both strands. The initial distance was set to 0 Å in order to (i) examine the possibility of "compression" of the double helix, (ii) determine in the analysis the existence of a minimum of dependence $\Delta G^{\circ}_{T}(r)$, and (iii) compute the energy of complexation as the difference between the minimum and maximum of this dependence (see Fig. 1). Analyzing the separation of the strands' centers of mass showed that a maximum distance of 45 Å is sufficient for complete dissociation of the duplexes of oligonucleotides with a length of 4 to 9 bp, and for complexes

Duplex formation thermodynamic parameters calculated by the WHAM and the MMGBSA method and determined experimentally

Oligonucleotide sequence from the 5′ to the 3′end	WHAM					MMGBSA	Experimental values			
	ΔH°	ΔS°	ΔG°_{37}	T_m	R^2	ΔH°	ΔH°	ΔS°	ΔG°_{37}	T_m
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
ACATTATTATTACA	–148.7	–442.7	–11.4	44.3	0.947	–125.4 ± 13.6	–89.9	–254	–11.1	48.3

Note. Sequence only one strand of each duplex is shown. Units of thermodynamic parameters ΔH° and ΔG°_{37}: kcal/mol, ΔS°: cal/(mol · K), and T_m: °C. R^2: Pearson's correlation coefficient for linear dependence ΔG°(*T*). Error of experimental ΔH°, ΔS°, ΔG°_{37} and T_m values is 10, 10, 8 %, and 0.5 °C, respectively.

of 14 and 16 bp in size, the distance should be increased to 60 Å. With the reaction coordinate chosen in this manner, the dissociation of the two strands for most oligonucleotide duplexes proceeded in accordance with the unzipping model (Cantor, Schimmel, 1980; Volkov, Solov'yov, 2009), which involves the unwinding of the double helix from one of the ends, or in accordance with a mixed shearing/unzipping mode (Mosayebi et al., 2015; Kurus, Dultsev, 2018). An example of alterations of oligonucleotide conformations along the reaction coordinate is given in Supplementary Material 1[1]. The mechanism of dissociation of duplexes in the current paper is not critically important because only two limiting states are being examined: a relaxed duplex structure and two noninteracting single-stranded oligonucleotides. The match between the helix–coil transition mechanism and the mechanisms observed by experimental methods confirmed the adequacy of the chosen approach for describing the dissociation of a DNA double helix.

According to generally accepted requirements for using the WHAM, it is necessary that the overlap between histograms be at least 20 %. Our analysis indicated that this is achieved at ~0.7 Å between adjacent simulation windows. An example of the dependence of distribution histograms on the distance between strands for duplex 5′-GCACCGAC-3′/5′-GTCGGTGC-3′ is given in Supplementary Material 2. We chose the reaction coordinate step of 0.5 Å to reliably meet this criterion.

When energy is calculated by the WHAM, an important parameter is the number of points ("bins") along the reaction coordinate that are chosen for sampling of the free energy profile. When the number of points was 100 or more (up to 1,000 partitions), a plateau was reached for the shape and position of the Gibbs free energy profiles (Fig. 2, *a* and Supplementary Material 3) and for the magnitude of the change in Gibbs free energy at different temperatures (see Fig. 2, *b*). At the same time, relative error values calculated by the bootstrap method (Grossfield, 2018) did not exceed 6 % (see Fig. 2, *c*).

Gibbs free energy at a certain temperature was calculated as the difference between a minimum and a maximum in the PMF profile: $\Delta G^\circ(T) = PMF_{min} - PMF_{max}$. To determine dependence of Gibbs free energy on temperature, the range from 273 to 333 K with a step of 10 K was chosen. The lower value was selected in accordance with the freezing temperature of water,

[1] Supplementary Materials 1–8 are available at:
https://vavilov.elpub.ru/jour/manager/files/Suppl_Yushin_Engl_27_7.pdf

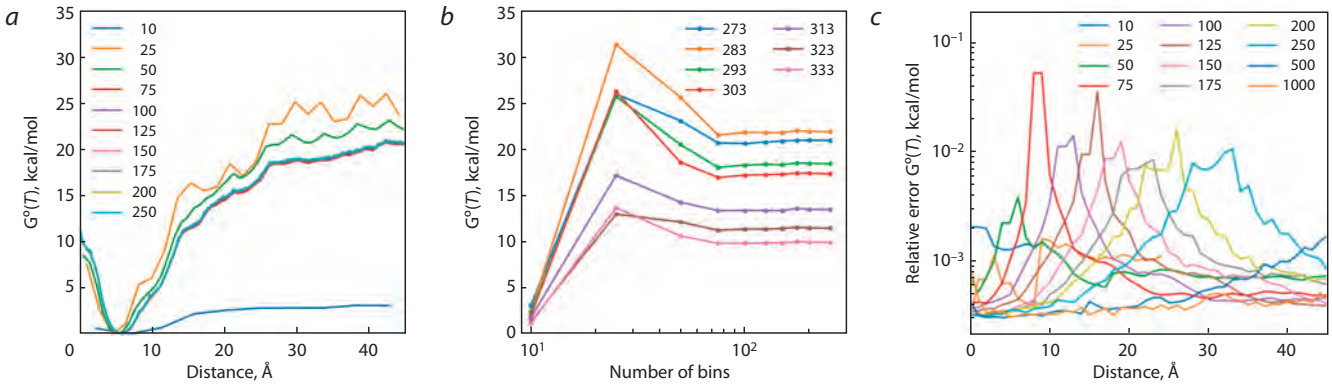

Fig. 2. Determination of parameters for the modeling and analysis of MD trajectories.
a, The dependence of the Gibbs free energy profile on the distance between the centers of mass of C4′ carbon atoms in two DNA strands for different numbers of points along the reaction coordinate that were chosen to sample the free-energy profile at 273 K. *b*, The dependence of the Gibbs energy of complexation on the number points along the reaction coordinate that were chosen to sample the free-energy profile. *c*, The dependence of relative error of Gibbs free energy computation on the distance between the centers of mass of C4′ carbon atoms from the two DNA strands for different numbers of points along the reaction coordinate that were chosen for sampling the free-energy profile.

and the upper value was chosen to limit the denaturation of an NA duplex in the modeling during the selected time range for short oligonucleotides. This range is wide enough for constructing dependence $\Delta G°(T)$ for reliable determination of $\Delta H°$ and $\Delta S°$ by linear regression analysis via the equation $\Delta G°(T) = \Delta H° - T\Delta S°$.

The trajectory length in the MD simulation at each fixed distance between the selected centers of mass and at a given temperature was set to 15 ns in order to obtain a minimally sufficient trajectory in an implicit water shell for the calculation of thermodynamic parameters (Lomzov et al., 2015). Thus, for each duplex, trajectories 15 ns long were obtained at 90 (or 120) different distances between the centers of mass at seven temperatures. Accordingly, trajectory length for each duplex ranged from 9.45 to 12.6 μs. The total length of trajectories for each complex was more than 200 μs.

By the WHAM, the dependence of the Gibbs energy of interaction of two oligonucleotides on the distance (r) between the centers of mass of C4′ atoms of each strand was calculated at seven tested temperatures for 21 studied duplexes (see the Table). A typical dependence of Gibbs energies of complex formation on r at temperatures of 273 to 333 K for the 5′-GCACCGAC-3′/5′-GTCGGTGC-3′ complex is depicted in Figure 3, *a*. The dependence of Gibbs free energy has a clear-cut minimum near 6 Å and increases with either an approach or dissociation of the double helix strands. During the dissociation, the dependence passes through a maximum and diminishes slightly. The maximum corresponds to the distance at which the interaction between the strands disappears.

To assess the adequacy of the modeling, we compared the geometry of the DNA double helix of the 5′-GCACCGAC-3′/5′-GTCGGTGC-3′ duplex in a relaxed form with literature data (Supplementary Material 4). All structural parameters are in good agreement with the data on Drew-Dickerson dodecamer (DDD, 5′-CGCGAATTCGCG-3′) structure determined experimentally by NMR spectroscopy (Protein Data Bank [PDB] ID: 1NAJ) and by X-ray crystallography (PDB ID: 1BNA).

Gibbs free energy of complexation was computed at various temperatures. It was established that $\Delta G°(T)$ is linear, with a high correlation coefficient R^2 of more than 0.83 and an average for all the analyzed complexes of 0.93 (see Fig. 3, *b* and the Table). Based on the obtained dependences (Supplementary Material 5), changes in the enthalpy and entropy of complexation were calculated next (see the Table). A comparison of thermodynamic parameters calculated by the WHAM with those determined experimentally (data from (Lomzov et al., 2015)) indicated a linear relation between them with high correlation coefficients R^2: 0.87, 0.82, 0.88, and 0.75 for $\Delta H°$, $\Delta S°$, $\Delta G°_{37}$ and T_m, respectively (Supplementary Material 6). As melting temperature, the temperature at which half of oligonucleotides are in a double-stranded state, and the remaining oligonucleotides are in a single-stranded state was chosen. T_m was computed from the thermodynamic parameters (Lomzov, Pyshnyi, 2012) as follows:

$$T_m = \Delta H°/(\Delta S° + R \ln\left(\frac{Ct}{4}\right)),$$

where R is the universal gas constant, and Ct is the total concentration of oligonucleotides in the system. Ct was set to 10 μM in accordance with typical experimental values.

The slope of the linear dependence of the thermodynamic parameters was found to be close to 0.5, and values of free terms of the linear dependence are substantial as compared to the analyzed values (see Supplementary Material 6). Therefore, as suggested in our previous papers (Lomzov et al., 2015; Golyshev et al., 2021), it is possible to apply linear corrections to calculated thermodynamic parameters $\Delta H°$ and $\Delta S°$. After this correction was applied, correlation coefficients for Gibbs free energy and melting temperatures improved considerably to 0.94 and 0.86, respectively (Fig. 4). In this context, the average absolute error of calculation of thermodynamic parameters became 11.4, 12.9, 11.8 %, and 5.5 °C for $\Delta H°$, $\Delta S°$, $\Delta G°_{37}$, and T_m, respectively. For our set of oligonucleotides, such error values for thermodynamic characteristics that have been obtained by the MMGBSA method in some studies (Lomzov et al., 2015; Golyshev et al., 2021) taking into account linear corrections are slightly lower: 7.6, 11.4, 10.6 %, and 4.3 °C. The accuracy of the computation of thermodynamic parameters in the present work is comparable to the accuracy of the experimental approach and to that of

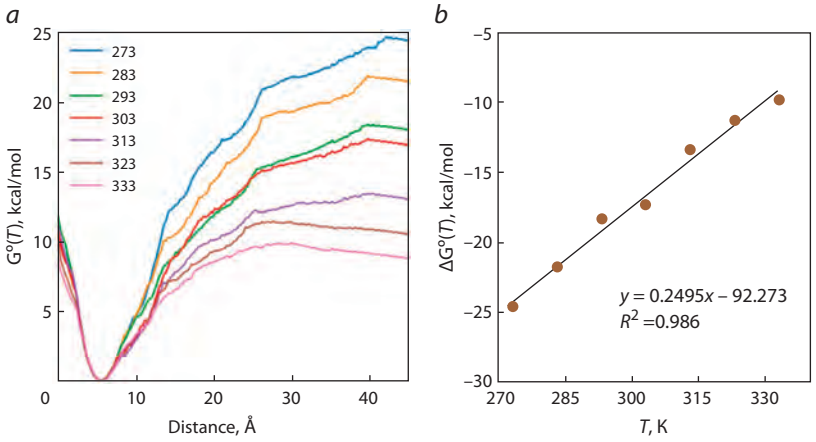

Fig. 3. The dependence of Gibbs free energy: *a*, on the distance between molecules at different temperatures (273, 283, 293, 303, 313, 323, and 333 K); *b*, on the temperature of model duplex 5′-GCACCGAC-3′/5′-GTCGGTGC-3′.

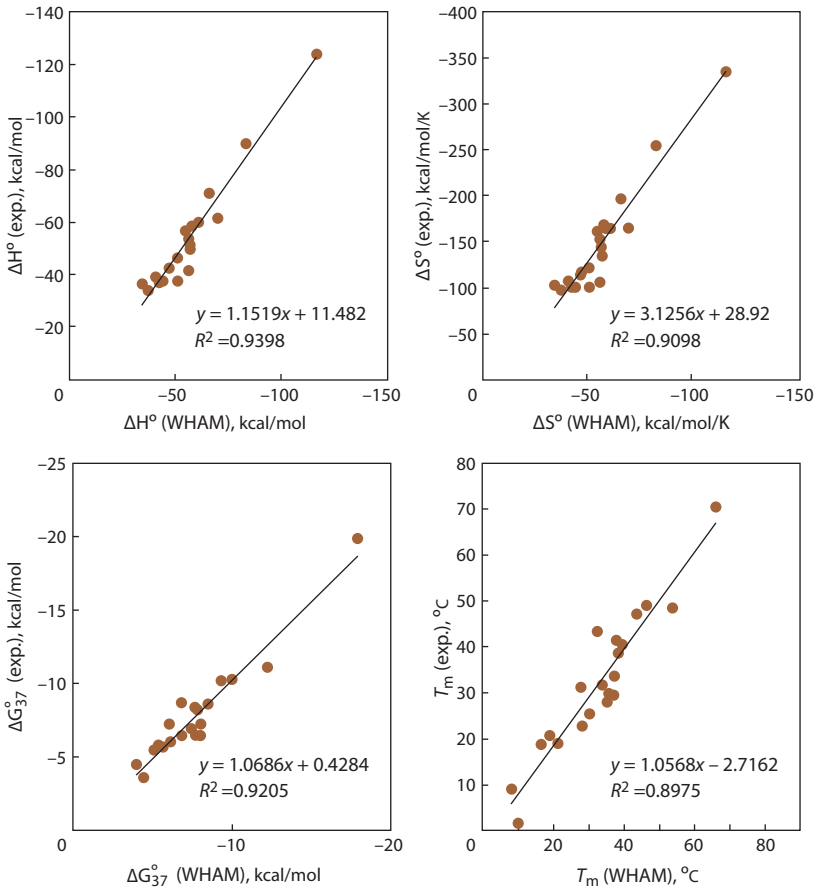

Fig. 4. Correlation of thermodynamic parameters ΔH°, ΔS°, ΔG°_{37}, and melting temperature of complexes – calculated by the WHAM taking into account linear corrections – with experimentally determined parameters (data from (Lomzov et al., 2015)).

the most common procedure for calculating the efficiency of oligonucleotides hybridization (the nearest-neighbor method): ~10 % for enthalpy and entropy and approximately 8 % for the Gibbs free energy of complexation (SantaLucia, Hicks, 2004; Lomzov et al., 2006).

To further check the quality of the results, the obtained trajectories were analyzed by the MMGBSA method, and the computed values were compared with the data of the WHAM and with values obtained by the MMGBSA method previously (Lomzov et al., 2015). The typical shape of the dependence of MMGBSA-calculated energy on the distance between the centers of mass of the strands' C4′ atoms proved to be similar to the dependence of Gibbs free energy on the distance depicted in Figure 3, *a* (Supplementary Material 7). At distances close to the maximum, the energy of formation of a DNA double helix reaches a plateau of zero, indicating the absence of interaction between the strands when this method of trajectory analysis is employed. The bottom of the potential well was observed in the region of 2–7 Å, which matches the relaxed form of the DNA double helix, and its global minimum near 7 Å is close to the minimum observed in Gibbs energy's dependence determined by the WHAM (see Fig. 3, *a* and Supplementary Material 7). There is a weak dependence of the complexation energy calculated by the MMGBSA method on temperature, implying a small change in heat capacity, ΔCp. It seems impossible to reliably determine the change in heat capacity by computational experiments owing to large values of calculation error.

The complexation enthalpy values computed in this work and those determined previously correlate well (R^2 = 0.97), with a slope close to unity (0.95) and the free term of the linear dependence close to zero (4 kcal/mol) (Supplementary Material 8, *a*). Additionally, a similar linear correlation was observed between the complexation enthalpy values calculated by the MMGBSA method and those determined by the WHAM in this work. Thus, the MD trajectories obtained in our paper are realistic.

One of the key aspects of the previously researched energy calculation by the MMGBSA method is the uncertainty associated with the structure of the single-stranded state of oligonucleotides. This state was extracted from the MD trajectory of a double helix. Nevertheless, this approach allowed to calculate the enthalpy of complexation with sufficient accuracy. In this work, during analysis by the WHAM, the single-stranded state of oligonucleotides was fairly well discernible in MD trajectories (as far as this can be done within the framework of the implicit water shell approximation and the force field in question). This approach yielded good results when the energy of double-helix formation was computed. Meanwhile, the main advantage of the WHAM is direct calculation of the change in the Gibbs free energy of DNA double-helix formation. This parameter

turned out to be linear in our calculations across a wide range of tested temperatures. This finding suggests that the modeling parameters selected by us and those included in the simulation and model analysis describe the physics of both double- and single-stranded DNA rather well. For the latter, this statement is supported by the conformation of oligonucleotides seen during the modeling of strands with a large distance between their centers of mass (see Supplementary Material 1). Oligonucleotides did not remain linear (in contrast to the duplex), and this observation was utilized in the MMGBSA analysis; they did not become completely disordered strands either but retained several heterocyclic bases in a row in stacking interactions. This finding is consistent with the persistent length of single-stranded regions of oligonucleotides, which is several nucleotides (depending on the GC content and on ionic strength of the solution) (e. g., (Chen et al., 2012)). Furthermore, the linear dependence $\Delta G°(T)$ observed in the procedure evaluated in the present paper makes it possible to directly calculate complexation entropy.

Nonetheless, the newly developed approach is far from perfect. In particular, for more accurate modeling of the structure and dynamics of DNA, it is necessary to employ the most modern force fields and an explicit water shell model. The analysis of force field parameters for such modeling is a separate, rather complicated task. Besides, the use of an explicit water shell greatly increases the complexity and duration of calculations. For instance, the main computational costs are incurred at the stage of MD calculations. For the 9 bp DNA duplex analyzed in detail in the current work, with the implicit water shell, the calculation speed for a modern video card (NVidia GTX 3080) is ~800 ns/day. Therefore, the computation time for one model duplex is 12 days. With an explicit water shell, the periodic cell being modeled will contain approximately 15–20 thousand molecules owing to the maximum distance between strands of 45 Å. This situation will reduce productivity to ~100 ns/day or approximately 3 months in total. In addition, in an explicit water shell, conformational mobility of DNA will significantly decrease, which will require extending trajectory length for each simulation window, thereby leading to higher computational costs. Nevertheless, such a complication seems necessary to improve the reliability/accuracy of the calculations.

Another promising avenue for the development of the proposed approach is its testing on known modified NAs as examples. This testing should answer the question of the applicability of our approach to the rational design of the chemical structure of new NA derivatives not yet chemically synthesized. The answer can help to solve specific problems of biomedicine and biotechnology. Our present analysis shows high potential and feasibility of the WHAM for calculating the formation energy of duplexes of NAs, their analogs, and derivatives.

Conclusion

A WHAM procedure for computing hybridization properties of oligodeoxyribonucleotides was refined here. Optimal parameters were selected for modeling and calculating thermodynamic parameters of the formation of DNA duplexes. By means of a representative sample of 21 oligonucleotides 4 to 16 nucleotides long with a GC content of 14 to 100 %, we demonstrated that calculating the enthalpy, entropy, and Gibbs free energy of the formation of oligonucleotide complexes by the WHAM is possible when MD trajectories are analyzed using the following reaction coordinate: the distance between the centers of mass of C4′ carbons of the two strands. A linear dependence of Gibbs free energy on the temperature at which the simulation is performed was documented. This finding enables researchers to compute the enthalpy and entropy of complexation *via* an analysis of WHAM results. The calculated thermodynamic parameters linearly correlate with experimentally determined values, with a high correlation coefficient R^2 (greater than 0.83). With a linear correction of this dependence, the error of calculation of thermodynamic parameters is comparable with the experimental one and amounts to 11.4, 12.9, and 11.8 % for $\Delta H°$, $\Delta S°$, and $\Delta G°_{37}$, while melting temperature is predicted with an average error of 5.5 °C. Thus, the use of the WHAM for calculating the formation energy of DNA duplexes was systematically investigated for the first time. High accuracy of such calculations was demonstrated, which is comparable with the accuracy of experimental and other techniques for computing the energy of complexation.

References

Banerjee D., Tateishi-Karimata H., Ohyama 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., Deka 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.jpcb.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., Ouldridge 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

Acknowledgements. This work was supported by the Russian state-funded project for ICBFM SB RAS (grant number 121031300042-1).

Conflict of interest. The authors declare no conflict of interest.

Received July 15, 2023. Revised September 20, 2023. Accepted September 21, 2023.