STRUCTURAL–FUNCTIONAL ANALYSIS OF BIOPOLYMERS AND THEIR COMPLEXES

UDC 57.013

Calculation of Energy for RNA/RNA and DNA/RNA Duplex Formation by Molecular Dynamics Simulation

V. M. Golyshev^{a, b}, D. V. Pyshnyi^{a, b}, and A. A. Lomzov^{a, b, *}

^a Institute of Chemical Biology and Fundamental Medicine, Siberian Branch of the Russian Academy of Sciences, Novosibirsk, 630090 Russia

^b Novosibirsk State University, Novosibirsk, 630090 Russia

*e-mail: lomzov@niboch.nsc.ru

Received March 7, 2021; revised May 21, 2021; accepted May 21, 2021

Abstract—The development of approaches for predictive calculation of hybridization properties of various nucleic acid (NA) derivatives is the basis for the rational design of the NA-based constructs. Modern advances in computer modeling methods provide the feasibility of these calculations. We have analyzed the possibility of calculating the energy of DNA/RNA and RNA/RNA duplex formation using representative sets of complexes (65 and 75 complexes, respectively). We used the classical molecular dynamics (MD) method, the MMPBSA or MMGBSA approaches to calculate the enthalpy (ΔH°) component, and the quasi-harmonic approximation (Q-Harm) or the normal mode analysis (NMA) methods to calculate the entropy (ΔS°) contribution to the Gibbs energy (ΔG_{37}) of the NA complex formation. We have found that the MMGBSA method in the analysis of the MD trajectory of only the NA duplex and the empirical linear approximation allow calculation of the enthalpy of formation of the DNA, RNA, and hybrid duplexes of various lengths and GC content with an accuracy of 8.6%. Within each type of complex, the combination of rather efficient MMGBSA and Q-Harm approaches being applied to the trajectory of only the bimolecular complex makes it possible to calculate the ΔG_{37} of the duplex formation with an error value of 10%. The high accuracy of predictive calculation for different types of natural complexes (DNA/RNA, DNA/RNA, and RNA/RNA) indicates the possibility of extending the considered approach to analogs and derivatives of nucleic acids, which gives a fundamental opportunity in the future to perform rational design of new types of NA-targeted sequence-specific compounds.

Keywords: thermal stability, hybridization, molecular dynamics, DNA/RNA, RNA/RNA, duplexes, molecular dynamics, MMGBSA, MMPBSA, oligonucleotides

DOI: 10.1134/S002689332105006X

INTRODUCTION

Derivatives and analogs of nucleic acids (NA) are used to solve the problems of targeted therapy of various diseases [1, 2], biosensor technologies, molecular biology [3], and biotechnology [4]. Although there are numerous NA derivatives with improved properties relative to natural analogs [4], they do not fully meet the requirements of researchers, and their use in practice is often limited by patent law. Thus, the creation of new derivatives with predetermined biological and physicochemical properties remains an urgent task in

chemistry and physical chemistry of nucleic acids. One of the most important properties of these analogs is their ability to efficiently interact with complementary NA sequences. To date, there are no methods for evaluating the hybridization properties of NA derivatives before chemical synthesis. For well-studied compounds, several approaches based on the analysis of empirical dependencies have been developed that allow calculation of the melting temperature [13, 14] and the energy of the formation of their complexes with DNA [5–11] and/or RNA [12], which depend on the length and nucleotide composition of the interacting strands. In the case of new compounds, it is necessary to use approaches that do not rely on experimental information on their hybridization properties. Methods based on molecular mechanics and/or molecular dynamics (MD) are suitable for large compounds consisting of more than 100 atoms.

Over the past decade, there has been great progress in the development of various approaches for calculating the thermodynamic parameters of intermolecular

Abbreviations: MD, molecular dynamics; NA, nucleic acid; MMGBSA (Molecular Mechanics/Generalized Born Surface Area), a component-by-component approach for calculating the free energy change based on simulation by the molecular dynamics method using the generalized Born model; MMPBSA (Molecular Mechanics/Poisson Boltzmann Surface Area), a component-by-component method for calculating the free energy change based on a molecular dynamics simulation using the Poisson-Boltzmann model; NMA (Normal Mode Analysis), analysis of normal modes of oscillation; Q-Harm, Quasi-Harmonic approximation.

interactions [15]. The most common methods are the MMPBSA (Molecular Mechanics/Poisson Boltzmann Surface Area) [16], MMGBSA (Molecular Mechanics/Generalized Born Surface Area) [17, 18], and methods for analyzing the potential mean force (PMF), for example the weighted histogram analysis method (WHAM) [19] or the method of the variational free energy profile (vFEP) [20]. Over the past decade, several approaches have also been developed which allow one to effectively cover the analyzed phase space for a more productive calculation of the free energy. Among these methods, one can distinguish methods of alchemical transformations [21], e.g., methods of free energy perturbation (FEP) [22] and thermodynamic integration (TI) [23]. Their disadvantages include large expenditures of computing power and calculation time to obtain reliable modeling results. Therefore, the above-mentioned approaches cannot be applied to calculate the energies of a set of complexes of different lengths and GC composition.

Simple and computationally efficient techniques, such as MMGBSA and MMPBSA are very popular for evaluating the energy of complex formation. In these methods, the energies of the entire complex and its components obtained by MD modeling are calculated for each conformation of molecules, and the solvent (which is presented in the simulation explicitly) is replaced by a continuous medium with a given dielectric permittivity. In these approaches, the binding energy (ΔG_{bind}) consists of a change in the total molecular mechanical energy (ΔE_{MM}), the solvation energy (ΔG_{solv}), and the contribution of conformational entropy ($T\Delta S$) [24]:

$$\Delta G_{\text{bind}} = \Delta E_{\text{MM}} + \Delta G_{\text{solv}} - T\Delta S. \quad (1)$$

In turn, the internal energy (ΔE_{int}) (deformation of bonds, angles, and dihedral angles), electrostatic energy (ΔE_{ele}), and Van der Waals interactions (ΔE_{vdW}) contribute to the ΔE_{MM} value:

$$\Delta E_{\text{MM}} = \Delta E_{\text{int}} + \Delta E_{\text{ele}} + \Delta E_{\text{vdW}}. \quad (2)$$

The solvation energy is the sum of the contributions from the polar ($\Delta G_{\text{PB/GB}}$) and non-polar (ΔG_{SA}) interaction of a studied molecule with solvent:

$$\Delta G_{\text{solv}} = \Delta G_{\text{PB/GB}} + \Delta G_{\text{SA}}. \quad (3)$$

The first member of equation (3), $\Delta G_{\text{PB/GB}}$, is the electrostatic energy, which can be calculated either by the direct solution of the Poisson–Boltzmann equation (PB, the MMPBSA method), or using the generalized Born method (GB, the MMGBSA method). The contribution of nonpolar interactions, ΔG_{SA} , is usually represented as a linear function of the area of the molecule available for interaction with the solvent (SASA) [25]:

$$\Delta G_{\text{SA}} = \gamma \text{SASA} + b, \quad (4)$$

where γ and b are coefficients of the linear equation.

To calculate the Gibbs free energy, the entropy contribution including the conformation component is usually evaluated by normal mode analysis (NMA) [26] or quasi-harmonic approximation (Q-Harm) [27]. The former method is characterized by high computational complexity and allows analysis of the entropy contribution for only a small number of snapshots of the MD trajectory (up to several thousand), while the latter method is significantly faster and more efficient.

These methods of estimating the energy of intermolecular interactions are widely used when studying the interactions of biopolymers with small molecules, e.g., in molecular biological studies [28] or the development of new drugs [29]. The successful application of these approaches has also been demonstrated for calculation of the energy of complex formation with DNA and RNA of both native [30, 31] and chemically modified oligonucleotides [32–36]. However, in the latter case, the research is not systematic; it is rather an illustration of the possibilities of the methods.

The creation of an approach for the predictive calculation of hybridization properties of different classes of NA derivatives could become the basis for the rational design of these compounds and the subsequent directed search for methods of chemical synthesis. At the first stage, it is necessary to verify the applicability of calculations which have been performed for well-studied unmodified NA complexes, to new, including not yet synthesized, NA analogs. We have previously demonstrated the successful application of the MPB(GB)SA approaches for calculating the energy and thermal stability of a set of more than 300 DNA/DNA duplexes [30]. During this study, we analyzed the possibility of calculating the hybridization properties of DNA/RNA and RNA/RNA complexes by the classical MD method.

The goal of the present work was to test the applicability of the MMPBSA and MMGBSA methods for calculating the enthalpy component and the Q-Harm and NMA methods for calculating the entropy component of the Gibbs energy of DNA/RNA and RNA/RNA duplexes. The MD method was used to simulate representative sets of oligonucleotides and their complexes. The resulting trajectories were analyzed using the MMPBSA and MMGBSA approaches, and the entropy was calculated using the NMA and Q-Harm methods. The data were analyzed and compared with the experimental values of the thermodynamic parameters of oligonucleotide complex formation.

EXPERIMENTAL

Database of thermodynamic parameters. The experimental values of the thermodynamic parameters of the formation of the NA complexes are taken from previously published studies. For complementary DNA/RNA duplexes, we used the data of Sugimoto et al. [7]. Duplexes that contained 3'-terminal phos-

phate moieties were excluded from consideration. Thus, the database contained thermodynamic parameters of complex formation for 65 duplexes with a length from 5 to 12 bp (on average 8 bp) and composition of G/C pairs from 0 to 84% (on average 53%). For complementary RNA/RNA duplexes, we used the data of Xia et al. [37]. Duplexes that contained terminal phosphate groups were also excluded from consideration. Thus, the database contained the thermodynamic parameters of complex formation for 75 duplexes with a length of 4 to 11 bp (on average 7 bp) and composition of G/C pairs from 22 to 100% (on average 58%).

Preparation of molecular systems for modeling. Duplex structures were prepared in the NAB program of the AmberTools14 software package [38] using the Chimera program [39]. The starting structures of RNA/DNA and RNA/RNA duplexes had the A form of a double helix. Single-stranded oligonucleotides were obtained from duplexes by removing one of the strands.

Molecular dynamics simulation. MD modeling was performed using the AMBER 14 software package and parallel computing on CPUs and GPUs with CUDA architecture. The ff99bsc0 [40] and ff99bsc0+chiOL3 [41] force fields were used for DNA and RNA, respectively. MD simulations were performed in an explicit water shell at a temperature of 300 K. The electrostatic interactions were calculated by the Particle Mesh Ewald Method with a lattice step of 1 Å. The pressure in the system was 1 bar (Berendsen barostat); the temperature in the system was maintained by scaling the atom's speeds with a period of 10 ps. The SHAKE algorithm was used to increase the step of simulation up to 2 fs. We used a nonbonded cut-off of 10 Å. The coordinates of each atom in the system were recorded every 10 ps of simulation. Duplexes and single-stranded oligonucleotides were placed in a cubic cell with a distance of 12 Å from the simulated molecule to the cell boundaries. The amount of water in the simulated cell was in the range of 4000–8000 molecules. Sodium ions were used to neutralize the negative charge of NA in the periodic cell.

The simulation procedure consisted of the following steps:

(1) Creation of a PDB file that contains data on the coordinates of each atom in the structure of a duplex or single-stranded NA.

(2) Creation of a water environment (TIP3P water model, cubic periodic conditions, 12 Å); addition of sodium ions to neutralize the periodic system.

(3) Minimization of the entire system with a fixed NA (10000 minimization steps) (PMEMD.MPI).

(4) Heating of the system with a fixed NA for 2.5 ns with an integration step of 0.5 fs from 0 to 300 K (PMEMD.CUDA).

(5) Equilibrating the density of the system at a constant pressure of 1 bar for 500 ps (SANDER) with a fixed NA.

(6) Equilibrating the system at a constant pressure of 1 bar and a temperature of 300 K for 5 ns (PMEMD.CUDA).

(7) MD simulation for 100 ns in an NPT ensemble (1 bar, 300 K) (PMEMD.CUDA).

(8) Calculation of the energy of complex formation using the MMGBSA and MMPBSA methods; calculation of entropy in Q-Form and NAME (MMPBSA.MPI). Calculations were performed using only the duplex trajectory (single-trajectory approach) and independently obtained trajectories of duplex and single-stranded oligonucleotides (three-trajectory approach).

To calculate by the MMGBSA method, we used the parameters of the generalized Born method proposed by Tsui and Case [17]. The ionic strength of the solution was 100 mM. Each recorded snapshot of the MD trajectory was used in calculation by the MMGBSA, MMPBSA, and Q-Harm methods, and every tenth saved conformation was used in the case of NMA.

RESULTS

We have previously shown that computer calculations for evaluating the hybridization properties of DNA/DNA duplexes provide accuracy comparable to that of experimental methods [30]. Therefore, we used the MD modeling and calculation of the energy of duplex formation concerning DNA/RNA and RNA/RNA complexes. We used a database of thermodynamic parameters of the complex formation including experimental values of the enthalpy, entropy, and Gibbs free energy of the formation of NA duplexes. The characteristics of the data sets are given in the Experimental section.

At the first stage, molecular structures were designed and calculated by the MD method in an explicit water shell for 100 ns for 65 DNA/RNA duplexes [7], 75 RNA/RNA duplexes [37], and their single-stranded oligonucleotide components. In the case of self-complementary oligonucleotides, calculations were performed using two different initial velocities of atoms. The resulting MD trajectories were analyzed.

Analysis of Complex Structure

The modeling of the RNA/RNA duplexes showed that the structure of the A form of the double helix was preserved along the entire trajectory with a length of 100 ns for each of the 75 analyzed complexes. The deviation of the geometry of the helix from the initial unbalanced structure during modeling did not exceed 2.5 Å, the standard deviation (RMSD) between the atoms. The complexes after equilibrating for 10 ns slightly changed their conformation during the next 90 ns of modeling. The average RMSD value along the MD trajectory for the entire set of complexes was 1.56 Å and did not exceed 2.56 Å for each snapshot of

the trajectory. Unwinding of the end pairs of nucleotides was observed in 11 of the 75 studied duplexes. In some cases, reversible destruction of the second base pair occurred for duplexes with terminal A/U- or A/T-pairs, while the terminal adenine and the pre-terminal nucleotide of the same strand retained a conformation close to that in the double helix.

When modeling DNA/RNA duplexes, deviation from the starting structure did not exceed 4.1 Å. In 17 of the 65 studied structures, unwinding of the terminal pairs was observed as in the bimolecular RNA/RNA complexes. The average RMSD value along the MD trajectory for the entire set of complexes was 2.26 Å and did not exceed 3.86 Å for each snapshot of the trajectory.

The structures of the DNA/RNA and RNA/RNA duplexes were slightly different. Analysis of RNA/RNA complexes using the X3DNA software [42, 43] revealed that these complexes are in the A form of the double helix. The structure of DNA/RNA duplexes can be characterized as intermediate between the A and B forms. This difference in conformations is known and has been shown by experimental methods, i.e., NMR spectroscopy [44], X-ray diffraction analysis [45], and circular dichroism spectroscopy [44, 46, 47]. We did not consider homonucleotide sequences of the DNA/RNA and RNA/RNA duplexes because of the peculiarities of their structures [48, 49].

The MD analysis of single-stranded oligonucleotides revealed significant conformational rearrangements in both DNA and RNA strands. In most cases, we observed local destruction of single-stranded stacking between adjacent bases and, in some cases, the formation of intramolecular hairpin structures, which can be quite stable over a trajectory length of 100 ns. This leads to significantly higher RMSD values: both averaged values along the trajectory (5.3 Å for the RNA strand and 5.7 Å for the DNA strand, on average for all the studied oligonucleotides) and maximum values (8.0 and 9.5 Å for the RNA and DNA strands, respectively) compared to those for the initial structures. This behavior of single-stranded forms is typical for nucleic acids, which follows from the published experimental data [50, 51] and the results of computer modeling [52–54].

Energy of Formation of RNA/RNA Complexes

The energy of the formation of NA complexes was evaluated as the difference in the energies of the double-stranded and single-stranded states. The change in the internal energy of complex formation was calculated by MD modeling using MMPBSA and MMGBSA. The NME and Q-Harm methods were used to calculate the contribution of the configuration entropy. The contribution of single-stranded states was assessed by analyzing the MD trajectories of two separate oligonucleotides (three-trajectory approach, 3tr).

Alternatively, individual oligonucleotides were isolated from the MD trajectory of RNA duplexes, and the resulting set of conformations was used for calculations (single-trajectory approach, 1tr). The Gibbs energy of complex formation was calculated using various combinations of the enthalpy and entropy values of complex formation obtained in single- or three-trajectory approaches.

A comparison of the experimental values of the thermodynamic parameters and the values obtained in single- and three-trajectory approaches for RNA/RNA complexes is shown in Fig. 1 and Table 1. The best correlation between the experimental and calculated values of the enthalpy of complex formation was observed when using the MMGBSA method and the three-trajectory approach. The correlation coefficient R^2 (where R is the Pearson correlation coefficient) between the experimental and calculated values was 0.72; in this case, the slope of the linear dependence was close to unity (0.95), and the shift coefficient was only 4.6 kcal/mol. When using this dependence to calculate the enthalpy of complex formation, the average absolute value of its prediction error was 8.1 kcal/mol or 16.3%. When using the MMPBSA method, the correlation of the experimental and calculated values was much worse than when using the MMGBSA method, i.e., the average absolute error was much higher in the case of a single-trajectory approach and significantly lower in the case of a three-trajectory approach (Table 1).

The introduction of a linear empirical correction to the calculated enthalpy values of complex formation, which is obtained from the correlation of calculated and experimental values ($\Delta H^0(\text{Exp}) = a\Delta H^0(\text{MD}) + b$), decreases the absolute and relative error values to 3.7 kcal/mol and by up to 8.1% when calculated by the MMGBSA (3 tr) method. This decrease is lower than the common experimental accuracy of evaluating the complexation enthalpy (10%) [37].

The value of the entropy of the complex formation was calculated by the Q-Harm and NMA when analyzing one or three independent trajectories. A comparison of the calculated and experimental values shows that the best correlation ($R^2 = 0.67$) is observed when using a quasi-harmonic approximation for the analysis of a single trajectory (Table 1). The average value of the relative error in calculating the entropy, in this case, is 83.4 cal mol⁻¹ K⁻¹, or 54%. This is significantly worse in percentage terms than in the case of the optimal method of calculating the enthalpy by MMPGSA using three-trajectory analysis. The three-trajectory analysis gives 1.2–2.5 times lower error values in calculating the entropy of complex formation, whereas the observed correlations are significantly worse compared to the single trajectory analysis. It should be noted that linear dependencies demonstrate the proportionality coefficient between the calculated and experimental entropy values in the range 1.13–2.40 and the free term of the linear dependence in the

Fig. 1. Correlation analysis of the thermodynamic parameter values evaluated experimentally and calculated by the MD method for RNA/RNA duplexes. The enthalpy values (ΔH^0) are calculated by the MMGBSA methods using the analysis of one (1tr) and three (3tr) trajectories (a). The Gibbs energy values at 37°C (ΔG_{37}^0) are evaluated experimentally and by the analysis of a single MD trajectory using a combination of the MMGBSA and NMA or Q-Harm methods (b). The order of the equations corresponds to the order of the notation in the legend.

range $-34.99 + 16.48 \text{ cal mol}^{-1} \text{ K}^{-1}$. This means that the configuration entropy is largely correlated with but does not make a determining contribution to the entropy of complex formation. The use of a linear correction to approximate the calculated values of the entropy of complex formation to the experimental data significantly decreases the calculation error to 8.7% when using a quasi-harmonic approximation in a single-trajectory approach (Table 1). The less reliable predictive ability of the entropy calculation methods in the three-trajectory approach is most likely caused by incomplete coverage of the configuration space of single-stranded oligomers at a given length of the trajectory in the classical MD method. In addition, the observed significant difference in the entropy also depends on the ionic strength of the solution, which, when modeled, differs from the standard experimental conditions (1 M NaCl, neutral pH values). In one of the classical concepts, the effect of the concentration of monovalent ions on the thermal stability of intermolecular NA complexes can be represented as an entropy correction, which is proportional to the logarithm of the concentration of cations in solution and the number of negatively charged phosphate residues in the NA strand [37, 55, 56]. This dependence corresponds to the observed trend, i.e., the linear dependence of the experimental and calculated values of the entropy of duplex formation. However, it is not possible to take full account of the contribution of NA duplexes to the thermal stability within the classical MD method.

The Gibbs free energy value is most often used in practice when comparing the hybridization properties

of oligonucleotides [55, 56]. This parameter determines the value of the equilibrium constant in the Van't Hoff equation and, consequently, the ratio of the concentrations of single- and double-stranded states of oligonucleotides in solution at different temperatures. We considered the predictive ability of combinations of entropy and enthalpy calculation methods when evaluating the Gibbs energy of duplex formation in the single- and three-trajectory approaches. The results of correlations of the calculated and experimental values are shown in Fig. 1 and Table 1. The best correlation with $R^2 = 0.85$ in evaluating the enthalpy and entropy of complex formation is observed when using single-trajectory analysis by the NMA and Q-Harm methods, respectively. The linear dependence shows a significant slope ($a = 0.29$) and a nonzero shift coefficient ($b = -4.0 \text{ kcal/mol}$) (Fig. 1b). This leads to large values of the absolute error value. As can be seen from the data for the prediction of the enthalpy and entropy of complex formation (Table 1), both of these components contribute to the non-unity correlation slope of the calculated and experimental values of the Gibbs free energy change and promote intersection with the abscissa axis at a non-zero point. Consideration of the linear correlation significantly reduces the error to 0.68 kcal/mol, or 8.63%, which coincides with the average experimental error for NA duplexes ($\sim 7\%$) [37].

The analysis shows sufficiently close experimental and calculated values (without taking into account linear corrections) of the complex formation energy only for the enthalpy change, which is calculated by the MMGBSA method using the three-trajectory

Table 1. Comparison of the experimental thermodynamic parameter values for RNA/RNA duplexes with those evaluated by analyzing MD trajectories^a

<i>N</i> ^b	Calculation method	<i>a</i>	<i>b</i>	<i>R</i> ²	$\langle \Delta E \rangle^c$	$\langle \Delta E / E \rangle, \%$	$\langle \Delta E^* \rangle^c$	$\langle \Delta E^* / E^* \rangle, \%$
ΔH°								
1tr	MMGBSA	0.98	5.20	0.66	7.2	14.5	5.1	9.29
	MMPBSA	0.68	−33.39	0.16	21.9	38.3	7.9	14.61
3tr	MMGBSA	0.95	4.60	0.72	8.1	16.3	3.7	8.09
	MMPBSA	1.06	5.78	0.68	5.0	10.1	4.1	8.81
ΔS°								
1tr	Q-Harm	2.40	13.64	0.67	83.4	54.0	13.5	8.73
	NMA	1.47	16.48	0.60	118.4	77.2	14.6	9.47
3tr	Q-Harm	1.68	−15.12	0.63	70.6	45.9	12.3	9.16
	NMA	1.13	−34.99	0.42	48.2	30.0	14.8	11.93
ΔG_{37}°								
1tr	MMPBSA + Q-Harm	0.14	−6.22	0.18	5.5	79.3	1.40	19.22
	MMGBSA + Q-Harm	0.29	4.00	0.85	31.9	433.8	0.68	8.63
	MMPBSA + NMA	0.07	−8.08	0.07	10.6	145.5	1.54	20.62
	MMGBSA + NMA	0.30	−0.15	0.77	17.7	238.2	0.74	10.92
3tr	MMPBSA + Q-Harm	0.22	−0.48	0.50	22.9	331.2	0.89	18.34
	MMGBSA + Q-Harm	0.22	0.78	0.58	28.1	407.3	0.84	18.05
	MMPBSA + NMA	0.22	−2.26	0.52	16.1	226.9	0.92	18.26
	MMGBSA + NMA	0.26	0.15	0.71	20.9	297.5	0.87	17.85

^a Here and further: the coefficients *a* and *b* are evaluated by the least-squares method from the correlation of the calculated and experimental values of the thermodynamic parameters within each calculation method; e.g., for the enthalpy values: $\Delta H^\circ(\text{Exp}) = a\Delta H^\circ(\text{MD}) + b$; *R*² is the correlation coefficient of this dependence. $\langle|\Delta E|\rangle$ and $\langle|\Delta E|/|E|\rangle$ are the average values of the absolute and relative error of calculating the thermodynamic parameter. The $\langle|\Delta E^*|\rangle$ and $\langle|\Delta E^*|/|E^*|\rangle$ values are calculated after linear correction (coefficients *a* and *b*) of the thermodynamic parameters obtained by computer analysis.

^b The number of analyzed trajectories.

^c The dimension of the average absolute error in calculating the enthalpy and Gibbs free energy (kcal mol^{−1}) and entropy (kcal mol^{−1} K^{−1}) of complex formation.

approach. The nearest neighbor model gives comparable error values [37]. We compared the previously published increments of the formation of the dinucleotide steps of the helix with the increments based on the calculation of both the experimental and computer analysis data (Fig. 2).

The nearest neighbor model equally well describes the experimental data and the results calculated by the MMPBSA method using the three-trajectory approach (Fig. 2a). All increments are significant (*p*-value < 4 × 10^{−6}) except for the increment of the initiation factor for experimental data (*p*-value = 0.14).

The literature values of the formation of dinucleotide pairs are close to those which are based on the experimental and calculated data, and are the same within the limits of errors. The exceptions were increments for GA/UC and GC/CG, which were calculated by the MMPBSA method (Fig. 2b). In this case,

the difference from the experimental values for the studied database was 4.3 and 3.8 kcal/mol, respectively. The most significant difference was observed for the initiation factor (Fig. 2b, increment “Init”), the value of which was −9.9 kcal/mol when analyzing the MD trajectories and +5.9 kcal/mol when analyzing the same database of experimental values. This difference is probably caused by the low significance of the increment based on the experimental data. Similarly, the correction associated with the terminal A/U pair is more negative, i.e., +5.4 vs. +9.4 kcal/mol, respectively. This may be caused by an incorrect account of the interaction of the terminal base pairs and an insufficiently long MD trajectory, which makes it impossible to take into account the effect of opening the terminal pairs with high confidence. At the same time, the data set indicates the adequacy of the proposed approach including MD modeling (force field, modeling method) and energy calculation using the

Fig. 2. Correlation analysis of the thermodynamic parameter values evaluated experimentally and calculated by the MD method for RNA/RNA duplexes. The predictive efficiency of the nearest neighbor method (NN; 5'-NN/NN-3') for the enthalpy of the RNA/RNA duplex formation evaluated experimentally (Exp) and calculated by the MMGBSA method using the three-trajectory analysis (MMGBSA, 3tr) (a). Comparison of the increments of the nearest neighbor model for RNA/RNA complexes, which were published by Xia et al. [37], evaluated experimentally (Exp), and calculated by the (MMGBSA, 3tr) method (b). The order of the equations corresponds to the order of the notation in the legend. Init., the initiation factor of the double helix formation.

MMPBSA method to account for interactions within the duplex structure of RNA.

Energy of Formation of DNA/RNA Complexes

In the case of hybrid DNA/RNA duplexes, we revealed a linear correlation between the calculated and experimental values of the enthalpy of complex formation (Fig. 3, Table 2). The correlation of the changes in the enthalpy was observed only when analyzing one trajectory by the MMGBSA method ($R^2 = 0.71$). The slope of the correlation straight line is close to unity ($a = 0.841$), and the free term is close to zero ($b = -0.982$ kcal/mol). This method of calculating the enthalpy gave 23% relative error and 12.0 kcal/mol of absolute error. In all other cases, the correlation coefficient R^2 did not exceed 0.64. At the same time, the MMPBSA method in the three-trajectory analysis led to the minimum values of the calculation errors (16% and 11.4 kcal/mol) in the absence of correlation ($R^2 = 0.13$) as in the case of the RNA/RNA complexes (Tables 1 and 2).

The introduction of a linear correction reduces the error to 8.2% (5.1 kcal/mol) in the case of the MMGBSA method when analyzing the trajectory of DNA/RNA duplexes only. This error is close to that for the RNA/RNA complexes.

The values of the entropy of the formation of DNA/RNA complexes calculated by the NMA method in the analysis of a single trajectory are best correlated with the experimental values ($R^2 = 0.74$), while the average absolute error is 47.5 cal mol⁻¹ K⁻¹, or 29% (Table 2). When using the quasi-harmonic approx-

imation in a single-trajectory analysis, the correlation coefficient was lower, $R^2 = 0.55$, and the average absolute error was 78.2 cal mol⁻¹ K⁻¹, or 49%. The proportionality coefficient between the calculated and experimental values of the entropy of the DNA/RNA complexes is in the range of 1.3–1.8, and the free term of the linear dependence is in the range of -3.0...+41 cal mol⁻¹ K⁻¹, which indicates significant differences between the experimental values of the entropy of complex formation and calculated values of the configuration entropy. The correlation coefficient and slope are close to those in the case of the RNA/RNA complexes (Tables 1 and 2), which indicates the general nature of the observed patterns for different types of complexes.

The calculation of the entropy of the formation of RNA-containing duplexes in the quasi-harmonic approximation when analyzing a single trajectory gives a good linear correlation with the experimental data. However, the slope of the linear approximation differs significantly from unity, which makes it difficult to use this approach to evaluate the total entropy of complex formation. It is worth noting that the calculation of the entropy in the quasi-harmonic approximation is several orders of magnitude faster than in the analysis of normal mode analysis. This justifies the use of the Q-Harm approximation in evaluating the Gibbs energy for a large set of DNA/RNA and RNA/RNA complexes.

The values of the Gibbs energy of the formation of hybrid complexes calculated by MMGBSA and using the configuration entropy values are highly correlated with the experimental values for DNA/RNA duplexes (Fig. 3). The correlation coefficients, R^2 , are 0.82 and

Fig. 3. Correlation analysis of the thermodynamic parameter values evaluated experimentally and calculated by the MD method for DNA/RNA duplexes. The enthalpy values (ΔH°) are calculated by the MMGBSA methods using the analysis of one (1tr) and three (3tr) trajectories (a). The Gibbs free energy values at 37°C (ΔG_{37}°) evaluated experimentally and by the analysis of a single MD trajectory using a combination of the MMGBSA and NMA or Q-Harm methods (b). The order of the equations corresponds to the order of the notation in the legend.

0.78 when calculating the entropy by Q-Harm and NMA, respectively. In other cases, there is no significant correlation ($R^2 < 0.06$).

The analysis of the applicability of the nearest neighbor model to describe the values of the enthalpy of complex formation revealed a high predictive efficiency of this model when applied to data obtained experimentally and by the MMPBSA method when analyzing a single trajectory (Fig. 4a). The increments of the nearest neighbor model, which describe the initiation factor of complex formation were not statistically significant, i.e., the p -values were 0.11 and 0.69 for experimental and calculated data, respectively. When analyzing the experimental database, the dinucleotide CU/AG and UU/AA pairs are of low significance. For the enthalpy values calculated by the MMGBSA in the single-trajectory approach, all increments, except for the initiation factor, were statistically significant (p -value $< 4 \times 10^{-5}$).

The increment values for experimental data, which we evaluated and published earlier, differed significantly for the initiation factor and dinucleotide pairs that contained cytosine or guanine at the 5' end of the RNA strand (rCX/dX'G, or rGX/dX'C, where X = A, C, U, or G, except for the rGU/dAC dinucleotide pair) (Fig. 4b). The evaluated increments of rCX/dX'G and rGX/dX'C have amplitudes lower and higher, respectively, compared to the literature data. The UG and UU increments have also a lower contribution to the stability of the double helix. The comparison of the

increments of the collected database shows that the values of the increments of rCX/dX'G, rGG/dCC, and rGU/dAC of the experimental enthalpy values of complex formation are significantly lower compared to those calculated by the MMGBSA method in the single-trajectory approach. This may be caused by the inconsistency of the force fields for DNA and RNA when modeling hybrid complexes, which have been optimized for DNA/DNA and RNA/RNA duplexes. In addition, a relatively small sample and a low variety of lengths and nucleotide compositions of model complexes may lead to sufficiently high error values for individual increments in the nearest-neighbor model (Fig. 4b). Expanding the sample set in the future will allow for more reliable results.

The negative values of the initiation factors and end pairs calculated by the MD method using the nearest neighbor model exceed the corresponding experimental data for the DNA/DNA [30], RNA/RNA, and DNA/RNA complexes. This indicates a systematic difference in the MD modeling or MMGBSA calculations in comparison with experimental data. The NMR method shows that the terminal base pairs are in equilibrium between the open and closed states [57] and the opening of the end pairs is not observed for all complexes. These events are not numerous repeated within the considered trajectory for each modeled complex. Therefore, it makes it impossible to reliably account for contributions from terminal base pairs when calculating thermodynamic parameters. Never-

Table 2. Comparison of the experimental thermodynamic parameter values for DNA/RNA duplexes with those evaluated by analyzing MD trajectories^a

<i>N</i>	Calculation method	<i>a</i>	<i>b</i>	<i>R</i> ²	$\langle \Delta E \rangle^a$	$\langle \Delta E / E \rangle, \%$	$\langle \Delta E^* \rangle^a$	$\langle \Delta E^* / E^* \rangle, \%$
ΔH°								
1tr	MMGBSA	0.84	1.0	0.71	12.0	22.9	5.1	8.2
	MMPBSA	0.49	−44.7	0.64	25.1	44.0	8.9	10.8
3tr	MMGBSA	0.56	−22.3	0.32	11.3	21.3	7.9	12.6
	MMPBSA	0.41	−36.9	0.13	11.4	16.0	8.7	13.8
ΔS°								
1tr	Q-Harm	1.81	−15.3	0.55	78.2	49.0	19.0	14.3
	NMA	1.77	41.1	0.74	47.5	29.2	14.3	8.4
3tr	Q-Harm	1.33	−31.8	0.66	60.9	38.3	16.8	9.5
	NMA	1.55	−3.0	0.53	57.6	35.5	20.0	11.4
ΔG_{37}°								
1tr	MMPBSA + Q-Harm	0.05	−6.97	0.04	4.7	81.0	1.94	23.4
	MMGBSA + Q-Harm	0.19	1.60	0.82	35.8	561.1	0.62	9.3
	MMPBSA + NMA	0.02	−7.38	0.00	10.5	171.6	2.01	24.2
	MMGBSA + NMA	0.21	0.24	0.78	26.3	407.0	0.69	9.6
3tr	MMPBSA + Q-Harm	−0.04	−8.39	0.06	18.4	317.2	1.46	24.7
	MMGBSA + Q-Harm	0.00	−7.36	0.00	26.8	451.4	1.43	24.3
	MMPBSA + NMA	−0.01	−7.65	0.01	17.0	287.1	1.45	24.6
	MMGBSA + NMA	0.03	−6.13	0.03	25.5	423.0	1.38	23.5

^a See Table 1 for all designations and dimensions.

theless, the MMGBSA method gives the best correlation between the experimental and the calculated values for all three types of complexes.

Our results demonstrate the possibility of calculating the thermodynamic parameters for the formation of NA complexes by the classical MD method. When analyzing sets of the RNA/RNA, DNA/RNA, and DNA/DNA duplexes [30], we have found that the enthalpy of complex formation calculated by the MMGBSA method when analyzing the MD trajectory of only the complex has the best correlation, on average, between the calculated and experimental values for each type of complex ($R^2 > 0.73$). The error in calculating the enthalpy of complex formation (15–30%) is significantly higher than the typical experimental error (~10%). A high correlation is also observed when calculating the entropy using the quasi-harmonic approximation in the single-trajectory analysis. In this case, however, the slope and the free term of the linear dependence differed significantly from unity and zero, respectively. The analysis of correlations between the experimental and calculated enthalpy values for complex formation of the DNA/DNA [30], RNA/RNA, and DNA/RNA

duplexes shows that the slope angles and free terms of all linear equations are close for these three pairs (Fig. 5). We combined all our results to evaluate a general universal dependence, which will allow one to recalculate the values obtained by the MMGBSA method in a single-trajectory approximation. This linear function has the following form: $\Delta H^\circ(\text{Exp}) = 0.727 \Delta H^\circ(\text{MMGBSA}, 1\text{tr}) 5.32$.

The use of this linear correction makes it possible to significantly reduce the relative error of calculating the enthalpy of complex formation to 9.3% (7.8%), 12.8% (9.9%), and 7.5% (6.3%) for RNA/RNA, DNA/RNA, and DNA/DNA complexes, respectively (the variance for the given values is indicated in parentheses). On average, the calculation error was 8.6% for all types of complex, which is lower than the typical value of the experimental error (10%).

A significantly higher level of correlation is observed between the value of complex formation energy calculated by the MMGBSA method in the single-trajectory approximation and the experimental value of the Gibbs free energy when all types of complex are considered simultaneously ($R^2 = 0.897$). However, a more in-depth analysis revealed that there

Fig. 4. Correlation analysis of the thermodynamic parameter values evaluated experimentally and calculated by the MD method for DNA/RNA duplexes. Predictive efficiency of the nearest neighbor method (NN; 5'-NN/NN-3') for the enthalpy of the DNA/RNA duplex formation evaluated experimentally (Exp.) and calculated by the MMGBSA method using the three-trajectory analysis (MMGBSA, 3tr) (a). Comparison of the increments of the nearest neighbor model for DNA/RNA complexes, which were published by Sugimoto et al. [7], evaluated experimentally (Exp), and calculated by the (MMGBSA, 3tr) method (b). The order of the equations corresponds to the order of the notation in the legend. Init, the initiation factor of the double helix formation.

Fig. 5. Correlation analysis of the enthalpy (a) and Gibbs free energy (b) values evaluated experimentally and calculated by the MMGBSA method using a single trajectory for different types of complexes. Data on DNA/DNA complexes are taken from Lomzov et al. [30]. The order of the equations corresponds to the order of the notation in the legend.

are significantly different linear dependencies for each of the considered types of complex, and the highest R^2 value was found for a sample of 304 DNA/DNA complexes. The differences in the linear dependence coefficients are caused, among other things, by the different contributions of the entropy for different types of complexes, both in terms of the configuration and solvation component, which characterizes the interaction with ions. This hypothesis is confirmed by various

linear dependencies for the experimental Gibbs free energy values and those calculated from the enthalpy and entropy values obtained by the MMGBSA (1tr) method and quasi-harmonic approximation, respectively (Tables 1 and 2). Thus, the linear dependencies (Fig. 5b) can be used to calculate the Gibbs free energy only for a particular type of complex using the values of complex formation energy calculated by the MMGBSA method in a single-trajectory approxima-

tion. Taking into account corrections in the form of linear functions with their coefficients for each type of complex, the error values of these calculations give ΔG_{37}° 9.9 (6.7), 10.3 (9.0), and 8.0% (6.9%) for RNA/RNA, DNA/RNA, and DNA/DNA, respectively (the variance for the values given is indicated in parentheses).

The use of the MMGBSA-calculated values of the internal energy change and the configuration entropy in the quasi-harmonic approximation to calculate the Gibbs energy of duplex formation gives results that are linearly related to the experimental values with a high correlation coefficient. This dependence is observed for sets of the RNA/RNA, DNA/RNA, and DNA/DNA duplexes of different lengths and GC composition. At the same time, the high variance of the calculation error indicates that it is necessary either to use methods of more complete coverage of the conformational space (sampling methods) or analyze a large number of complexes of different lengths and nucleotide composition, which will make it possible to reliably compare the calculated data for different types of complexes.

The use of computer modeling methods with more complete coverage of the conformational space in the analysis of trajectories makes it possible to evaluate the value of the Gibbs free energy. Previous publications have shown the fundamental possibility of achieving experimental accuracy of ~ 1 kcal/mol. Using the umbrella sampling method when using the distance between the ends of the RNA hairpin structure as the coordinate and the weighted histogram analysis method (WHAM), L. Smith et al. [58] obtained a high consistency of experimental and calculated values of free energy. The main disadvantage of this approach is the need to obtain an extremely long MD trajectory, ~ 200 μ s. A combination of alchemical transformation methods and the exchange of replicas may be more economical. In this case, it has been found that the replacement of one base by another of the same class (purine or pyrimidine) in an RNA duplex with a size of 6 bp results in accuracy of 0.55 kcal/mol for calculating the energy [22]; at the same time, the duration of the MD trajectory is small, ~ 1 μ s, which is an order of magnitude greater than in the approach we propose.

MD methods are used to analyze the structure and hybridization properties of not only native oligonucleotides but also their analogs and derivatives to establish their possible spatial structure, the effect of modifications on the thermal stability of complementary complexes, and the reasons for the observed physicochemical and molecular-biological effects. The energy of the formation of modified complexes was calculated for locked NA (LNA) [59], peptidyl NA (PNA) [35], 2'-O-Me RNA [34], glycine-morpholino analogs [32], and phosphoryl guanidine oligonucleotides [33]. Researchers used various approaches including the most common MMPB(GB)SA method, which we

also implemented. In most cases, correlation has been shown between the calculated energy values and the experimental characteristics of thermal stability, i.e., the enthalpy, the Gibbs free energy of complexation, or the melting point. However, there is no quantitative agreement in most approaches between the calculated and experimental energy values. This is further aggravated by the fact that a small number of complexes of similar length and/or GC composition are usually analyzed.

Taken together, the results indicate that it is impossible to quantitatively compare the Gibbs energy of the formation of different types of complexes, for example, oligonucleotides with a modified chemical structure, using classical MD and MMPB(GB)SA methods for calculating the enthalpy and the Q-Harm or NMA methods for calculating the entropy of complex formation. Our studies show the possibility to quantitatively compare the enthalpy of the formation of the RNA/RNA, DNA/RNA, and DNA/DNA complexes. It will be possible to understand whether this statement is true for NA derivatives after analyzing a set of their complexes of different lengths and GC composition using our proposed approach.

DISCUSSION

In this study, we analyzed the possibility of using the MD method to predict the energy value of the formation of NA duplexes. The MMPBSA and MMGBSA methods were used to calculate the enthalpy of complex formation, and the Q-Harm and NMA methods were used to calculate the entropy of complex formation. The analysis was performed using only the MD trajectory for the NA duplex or three trajectories for double-stranded state and single-stranded oligonucleotides. The value of the Gibbs free energy of duplex formation was evaluated from a combination of the entropy and enthalpy values calculated by different methods using the single- and three-trajectory approaches. In some cases, the resulting data were characterized by high correlation coefficients for experimental and calculated values.

The results lead to the conclusion that quantitative calculation of the enthalpy of complex formation can be carried out for the RNA/RNA, DNA/RNA, and DNA/DNA complexes using the MMGBSA method in a single-trajectory approximation using linear empirical correction. The resulting enthalpy value coincides with the experimental value with an accuracy of 8.6%. The values of the entropy and Gibbs free energy can be compared quantitatively only within one type of complex. In this case, the single-trajectory approach and combination of the MMGBSA and Q-Harm methods give the highest accuracy, i.e., 8.2 and 9.3% on average, for RNA/RNA and DNA/RNA duplexes, respectively, when taking into account the linear correction for each type of the complex. The use of these two calculation methods is the most compu-

tationally efficient and gives an accurate quantitative assessment of the thermodynamic parameters of complex formation.

To extend the proposed approach to NA analogs or derivatives, it is necessary to reliably calculate the structure of oligomers and their complexes with complementary nucleic acids. As soon as these structures or ensembles of structures are found, the method can be used to quickly assess the hybridization ability of NA derivatives. We showed earlier with examples of glycine-morpholino oligomers [32] and phosphoryl guanidine oligonucleotides [33] that the enthalpy values of complex formation calculated by the proposed method were close to the experimental values. However, the reliability of the proposed method should be quantified in additional studies of a set of oligomers of different lengths and GC compositions.

For quantitative calculation of the enthalpy of duplex formation, we recommend the MMGBSA method with an analysis of the trajectory of the duplex structure as an effective and fast approach. In the future, the proposed approaches can also be extended to NA analogs and derivatives, which will provide the opportunity for the rational design of new types of compounds.

ACKNOWLEDGMENTS

The authors thank the general-purpose computing center “Siberian Supercomputer Center SB RAS” (Novosibirsk, Russia) for the opportunity to use the computing clusters for the calculations.

FUNDING

The work was supported by the Russian Foundation for Basic Research and the Government of the Novosibirsk region (grant no. 20-43-5430291) and funded budget project of The Program of Fundamental Scientific Research of State Academies of Sciences no. AAAA-A 17-117020210021-7.

COMPLIANCE WITH ETHICAL STANDARDS

Conflict of interests. The authors declare that they have no conflict of interest.

This article does not contain any studies on the use of humans as objects of research.

OPEN ACCESS

This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article’s Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included

in the article’s Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

REFERENCES

1. Bennett C.F. 2019. Therapeutic antisense oligonucleotides are coming of age. *Annu. Rev. Med.* **70**, 307–321.
2. Smith C.I.E., Zain R. 2019. Therapeutic oligonucleotides: State of the art. *Annu. Rev. Pharmacol. Toxicol.* **59**, 605–630.
3. Benizri S., Gissot A., Martin A., Vialet B., Grinstaff M.W., Barthélémy P. 2019. Bioconjugated oligonucleotides: Recent developments and therapeutic applications. *Bioconjug. Chem.* **30**, 366–383.
4. Glazier D.A., Glazier D.A., Liao J., Roberts B.L., Li X., Yang K., Stevens C. M., Tang W., Tang W. 2020. Chemical synthesis and biological application of modified oligonucleotides. *Bioconjug. Chem.* **31**, 1213–1233.
5. Pyshnyi D.V., Lomzov A.A., Pyshnaya I.A., Ivanova E.M. 2006. Hybridization of the bridged oligonucleotides with DNA: Thermodynamic and kinetic studies. *J. Biomol. Struct. Dyn.* **23**, 567–579.
6. Banerjee D., Tateishi-Karimata H., Ohyama T., Ghosh S., Endoh T., Takahashi S., Sugimoto N. 2020. Improved nearest-neighbor parameters for the stability of RNA/DNA hybrids under a physiological condition. *Nucleic Acids Res.* 1–13.
7. Sugimoto N., Nakano S., Katoh M., Matsumura A., Nakamuta H., Ohmichi T., Yoneyama M., Sasaki M. 1995. Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry.* **34**, 11211–11216.
8. Hoshika S., Leal N. A., Kim M.-J., Kim M.-S., Karalkar N. B., Kim H.-J., Bates A. M., Watkins N.E., SantaLucia H.A., Meyer A.J., DasGupta S., Piccirilli J.A., Ellington A.D., SantaLucia J., Georgiadis M.M., Benner S.A. 2019. Hachimoji DNA and RNA: A genetic system with eight building blocks. *Science.* **363**, 884–887.
9. Watkins N.E., SantaLucia J. 2005. Nearest-neighbor thermodynamics of deoxyinosine pairs in DNA duplexes. *Nucleic Acids Res.* **33**, 6258–6267.
10. McTigue P.M., Peterson R.J., Kahn J.D. 2004. Sequence-dependent thermodynamic parameters for locked nucleic acid (LNA)–DNA duplex formation. *Biochemistry.* **43**, 5388–5405.
11. Owczarzy R., You Y., Groth C.L., Tataurov A.V. 2011. Stability and mismatch discrimination of locked nucleic acid–DNA duplexes. *Biochemistry.* **50**, 9352–9367.
12. Lesnik E.A., Freier S.M. 1995. Relative thermodynamic stability of DNA, RNA, and DNA:RNA hybrid duplexes: Relationship with base composition and structure. *Biochemistry.* **34**, 10807–10815.
13. Giesen U., Kleider W., Berding C., Geiger A., Ørum H., Nielsen P.E. 1998. A formula for thermal stability ($T(m)$) prediction of PNA/DNA duplexes. *Nucleic Acids Res.* **26**, 5004–5006.

14. von Ahlsen N., Wittwer C.T., Schütz E. 2001. Oligonucleotide melting temperatures under PCR conditions: Nearest-neighbor corrections for Mg^{2+} , deoxynucleotide triphosphate, and dimethyl sulfoxide concentrations with comparison to alternative empirical formulas. *Clin. Chem.* **47**, 1956–1961.
15. Armacost K.A., Riniker S., Cournia Z. 2020. Novel directions in free energy methods and applications. *J. Chem. Inf. Model.* **60**, 1–5.
16. Kollman P.A., Massova I., Reyes C., Kuhn B., Huo S., Chong L., Lee M., Lee T., Duan Y., Wang W., Donini O., Cieplak P., Srinivasan J., Case D.A., Cheatham T.E. 2000. Calculating structures and free energies of complex molecules: Combining molecular mechanics and continuum models. *Acc. Chem. Res.* **33**, 889–897.
17. Tsui V., Case D.A. 2000. Molecular dynamics simulations of nucleic acids with a generalized born solvation model. *J. Am. Chem. Soc.* **122**, 2489–2498.
18. Rastelli G., Rio A.Del., Degliesposti G., Sgobba M. 2009. Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA. *J. Comput. Chem.* **32**.
19. Kumar S., Rosenberg J.M., Bouzida D., Swendsen R.H., Kollman P.A. 1992. The weighted histogram analysis method for free-energy calculations on biomolecules: 1. The method. *J. Comput. Chem.* **13**, 1011–1021.
20. Lee T.-S., Radak B.K., Huang M., Wong K.-Y., York D.M. 2014. Roadmaps through free energy landscapes calculated using the multidimensional vFEP approach. *J. Chem. Theory Comput.* **10**, 24–34.
21. Williams-Noonan B.J., Yuriev E., Chalmers D.K. 2018. Free energy methods in drug design: prospects of “alchemical perturbation” in medicinal chemistry. *J. Med. Chem.* **61**, 638–649.
22. Sakuraba S., Asai K., Kameda T. 2015. Predicting RNA duplex dimerization free-energy changes upon mutations using molecular dynamics simulations. *J. Phys. Chem. Lett.* **6**, 4348–4351.
23. Lee H.-C., Hsu W.-C., Liu A.-L., Hsu C.-J., Sun Y.-C. 2014. Using thermodynamic integration MD simulation to compute relative protein–ligand binding free energy of a GSK3 β kinase inhibitor and its analogs. *J. Mol. Graph. Model.* **51**, 37–49.
24. Wang E., Sun H., Wang J., Wang Z., Liu H., Zhang J.Z.H., Hou T. 2019. End-point binding free energy calculation with MM/PBSA and MM/GBSA: Strategies and applications in drug design. *Chem. Rev.* **119**, 9478–9508.
25. Gilson M.K., Honig B. 1988. Calculation of the total electrostatic energy of a macromolecular system: Solvation energies, binding energies, and conformational analysis. *Proteins Struct. Funct. Bioinform.* **4**, 7–18.
26. Srinivasan J., Cheatham T.E., Cieplak P., Kollman P.A., Case D.A. 1998. Continuum solvent studies of the stability of DNA, RNA, and phosphoramidate–DNA helices. *J. Am. Chem. Soc.* **120**, 9401–9409.
27. Brooks B.R., Janežic D., Karplus M. 1995. Harmonic analysis of large systems: 1. Methodology. *J. Comput. Chem.* **16**, 1522–1542.
28. Choudhary M.I., Shaikh M., Tul-Wahab A., Ur-Rahman A. 2020. *In silico* identification of potential inhibitors of key SARS-CoV-2 3CL hydrolase (Mpro) via molecular docking, MMGBSA predictive binding energy calculations, and molecular dynamics simulation. *PLoS One.* **15**, e0235030.
29. Wright D.W., Hall B.A., Kenway O.A., Jha S., Cooney P.V. 2014. Computing clinically relevant binding free energies of HIV-1 protease inhibitors. *J. Chem. Theory Comput.* **10**, 1228–1241.
30. Lomzov A.A., Vorobjev Y.N., Pyshnyi D.V. 2015. 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.* **119**, 15221–15234.
31. Yesudas J.P., Blinov N., Dew S.K., Kovalenko A. 2015. Calculation of binding free energy of short double stranded oligonucleotides using MM/3D-RISM-KH approach. *J. Mol. Liq.* **201**, 68–76.
32. Golyshev V.M., Abramova T.V., Pyshnyi D.V., Lomzov A.A. 2019. Structure and hybridization properties of glycine morpholine oligomers in complexes with DNA and RNA: Experimental and molecular dynamics studies. *J. Phys. Chem. B.* **123**, 10571–10581.
33. Golyshev V.M., Pyshnyi D.V., Lomzov A.A. 2021. Effects of phosphoryl guanidine modification of phosphate residues on the structure and hybridization of oligodeoxyribonucleotides. *J. Phys. Chem. B.* **125**, 2841–2855.
34. Suresh G., Priyakumar U.D. 2015. Inclusion of methoxy groups inverts the thermodynamic stabilities of DNA–RNA hybrid duplexes: A molecular dynamics simulation study. *J. Mol. Graph. Model.* **61**, 150–159.
35. Siri Wong K., Chuichay P., Saen-oon S., Suparpprom C., Vilaivan T., Hannongbua S. 2008. Insight into why pyrrolidiny peptide nucleic acid binding to DNA is more stable than the DNA–DNA duplex. *Biochem. Biophys. Res. Commun.* **372**, 765–771.
36. Lomzov A.A., Gorelov V.V., Golyshev V.M., Abramova T.V., Pyshnyi D.V. 2015. 139 Analysis of structure and thermodynamics of modified DNA duplexes using molecular dynamics simulation. *J. Biomol. Struct. Dyn.* **33**, 90–91.
37. Xia T., SantaLucia J., Burkard M.E., Kierzek R., Schroeder S. J., Jiao X., Cox C., Turner D.H. 1998. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson–Crick base pairs. *Biochemistry.* **37**, 14719–14735.
38. Case D.A., Babin V., Berryman J.T., Betz R.M., Cai Q., Cerutti D.S., Cheatham T.E., Darden T.A., Duke R.E., Gohlke H., Goetz A.W., Gusarov S., Homeyer N., Janowski P., Kaus J., et al. 2014. AMBER 14, University of California, San Francisco. <https://doi.org/10.13140/rg.2.2.17892.37766>
39. Pettersen E.F., Goddard T.D., Huang C.C., Couch G.S., Greenblatt D.M., Meng E.C., Ferrin T.E. 2004. UCSF Chimera: A visualization system for exploratory research and analysis. *J. Comput. Chem.* **25**, 1605–1612.
40. Pérez A., Marchán I., Svozil D., Sponer J., Cheatham T.E., Laughton C.A., Orozco M. 2007. Refinement of the AMBER force field for nucleic acids: Improving the description of α/γ conformers. *Biophys. J.* **92**, 3817–3829.
41. Zgarbová M., Otyepka M., Šponer J., Mládek A., Banáš P., Cheatham T.E., Jurečka P. 2011. Refinement

- of the Cornell et al. nucleic acids force field based on reference quantum chemical calculations of glycosidic torsion profiles. *J. Chem. Theory Comput.* **7**, 2886–2902.
42. Lu X.-J. 2003. 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* **31**, 5108–5121.
 43. Lu X.-J., Olson W.K. 2008. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc.* **3**, 1213–1227.
 44. Gyi J.I., Conn G.L., Lane A.N., Brown T. 1996. Comparison of the thermodynamic stabilities and solution conformations of DNA–RNA hybrids containing purine-rich and pyrimidine-rich strands with DNA and RNA duplexes. *Biochemistry.* **35**, 12538–12548.
 45. Arnott S., Chandrasekaran R., Millane R.P., Park H.S. 1986. DNA–RNA hybrid secondary structures. *J. Mol. Biol.* **188**, 631–640.
 46. Hall K.B., McLaughlin L.W. 1991. Thermodynamic and structural properties of pentamer DNA–DNA, RNA–RNA and DNA–RNA duplexes of identical sequence. *Biochemistry.* **30**, 10606–10613.
 47. Lomzov A.A., Kupryushkin M.S., Dyudeeva E.S., Pyshnyi D.V. 2021. A comparative study of the hybridization of phosphoryl guanidine oligonucleotides with DNA and RNA. *Russ. J. Bioorg. Chem.* **47** (2), 460–468.
 48. Cross C.W., Rice J.S., Gao X. 1997. Solution structure of an RNA–DNA hybrid duplex containing a 3'-thioformacetal linker and an RNA A-tract. *Biochemistry.* **36**, 4096–4107.
 49. Zimmerman S.B., Pfeiffer B.H. 1981. A RNA–DNA hybrid that can adopt two conformations: an X-ray diffraction study of poly(rA)–poly(dT) in concentrated solution or in fibers. *Proc. Natl. Acad. Sci. U. S. A.* **78**, 78–82.
 50. Zhou J., Gregurick S.K., Krueger S., Schwarz F.P. 2006. Conformational changes in single-strand DNA as a function of temperature by SANS. *Biophys. J.* **90**, 544–551.
 51. Chen H., Meisburger S.P., Pabit S.A., Sutton J.L., Webb W.W., Pollack L. 2012. Ionic strength-dependent persistence lengths of single-stranded RNA and DNA. *Proc. Natl. Acad. Sci. U. S. A.* **109**, 799–804.
 52. Isaksson J., Acharya S., Barman J., Cheruku P., Chattopadhyaya J. 2004. Single-stranded adenine-rich DNA and RNA retain structural characteristics of their respective double-stranded conformations and show directional differences in stacking pattern. *Biochemistry.* **43**, 15996–16010.
 53. Chakraborty K., Mantha S., Bandyopadhyay S. 2013. Molecular dynamics simulation of a single-stranded DNA with heterogeneous distribution of nucleobases in aqueous medium. *J. Chem. Phys.* **139** (7), 075103.
 54. Sen S., Nilsson L. 2001. MD Simulations of homomorphous PNA, DNA, and RNA single strands: Characterization and comparison of conformations and dynamics. *J. Am. Chem. Soc.* **123**, 7414–7422.
 55. Owczarzy R., You Y., Moreira B.G., Manthey J.A., Huang L., Behlke M.A., Walder J.A. 2004. Effects of sodium ions on DNA duplex oligomers: Improved predictions of melting temperatures. *Biochemistry.* **43**, 3537–3554.
 56. SantaLucia J. 1998. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. U. S. A.* **95**, 1460–1465.
 57. Nonin S., Leroy J.-L., Gueron M. 1995. Terminal base pairs of oligodeoxynucleotides: Imino proton exchange and fraying. *Biochemistry.* **34**, 10652–10659.
 58. Smith L.G., Tan Z., Spasic A., Dutta D., Salas-Estrada L.A., Grossfield A., Mathews D.H. 2018. Chemically accurate relative folding stability of RNA hairpins from molecular simulations. *J. Chem. Theory Comput.* **14**, 6598–6612.
 59. Shen L., Johnson T.L., Clugston S., Huang H., Butenhof K.J., Stanton R.V. 2011. Molecular dynamics simulation and binding energy calculation for estimation of oligonucleotide duplex thermostability in RNA-based therapeutics. *J. Chem. Inf. Model.* **51**, 1957–1965.

Translated by A. Levina