OMTN, Volume *31*

Supplemental information

Therapeutic phosphorodiamidate morpholino oligonucleotides: Physical properties, solution structures, and folding thermodynamics

Farkhad Maksudov, Evgenii Kliuchnikov, Daniel Pierson, M.L. Ujwal, Kenneth A. Marx, Arani Chanda, and Valeri Barsegov

Supplemental Material

Supplemental Methods

Atomic partial charges and force field parameters: For each PMO study system, we calculated atomic partial charges separately for the following three structure fragments: i) with phosphorodiamidate group attached to the 3'-end and capping the HO atomic group attached to the O5' atom (mimicking the PMO 5'-end; see Figure S1A); ii) for phosphorodiamidate group attached to the 5'-end and capping the HN atomic group attached to the N3' atom (mimicking the PMO 3'-end; Figure S1B); and iii) for phosphorodiamidate group attached to both 5'- and 3'-ends (Figure S1C). These fragments were reconstituted as described in the previous section. The procedure of Restrained Electro Static Potential (RESP) charge fitting to obtain atomic charges has been described[1], and was implemented in the RED server (RESP ESP charge Derive Server[2]). We used this method to derive atomic partial charges for each of the four "morpholino nucleotides" separately. Next, for each atom of the morpholino ring we averaged charges over these different partial charge calculations, so the atomic charges on the morpholino ring would be the same (except for the C1' and H1' atoms) for any base, as is the case for the ribose ring in the canonical bsc0$_{\chi OL3}$ force field for RNA (see Refs. 35-38 in the main text). We also averaged partial charges on atoms in the phosphorodiamidate group. The values of bond angles and dihedral angle parameters were adopted from the bsc0$_{\chi OL3}$ force field for atom types for those atoms which comprise the morpholino ring, except for the N3'-atom. Parameters related to atom types of the atoms in the phosphorodiamidate group that are not described in bsc0$_{\chi OL3}$, such as the N3' atom in the morpholine ring, and connections between them (i.e. covalent bonds, bond angles, and dihedral angles), were developed using the general Amber force field GAFF (see Ref. (39) in the main text; see also Tables S1 and S2).

Quantum chemistry calculations and force field development for PMOs: Here we describe in detail i) the selection of the dominant conformation of the morpholino ring, ii) the calculation of atomic charges in the morpholino ring and in the phosphorodiamidate group, iii) the development of the atomic force filed parameters for MD simulations of PMOs, and iv) the reconstruction of the initial PMO structures.

"Chair" versus "boat" conformations of morpholino – A 6-atom morpholino ring, forming a cyclic structure, might exist in either the "chair" or "boat" conformation. We carried out quantum chemistry calculations to determine the free-energy minimum state, and to estimate the free-energy difference between the "boat" and "chair" conformations. Atomic coordinates of the morpholino ring in the "chair" conformation were obtained from Ref.[3], and coordinates for the "boat" conformation were derived from those for the "chair" conformation by changing the rotation angle N3'-C2'-C1'-H1' with GaussView 5.0[4]. Next, we performed optimization of the two ring geometries, using Hartree-Fock (HF) theory with the 6-31G* basis set and the Gaussian 16 package[5], to compare the energies of these conformers. The free energy was found to be lower for the "chair" conformation by $\Delta E = 6.7$ kcal/mol, which agrees with a previous study of morpholino geometries[6]. Because the characteristic temperature for the transition from the "chair" conformation to the "boat" conformation $T_{ch} = {\Delta E}/{k_B} \approx 3{,}372$ K is much higher than room temperature ($T = 300$K), in the all-atom MD simulations we set all morpholino rings to be in the free-energy minimum "chair" conformation.

Calculation of atomic partial charges – Next, we created the atomic structures of PMO systems (Figure 1 in the main part). In a PMO sequence, each "morpholino nucleotide" is connected with the

phosphorodiamidate group either through the 3'-end, through the 5'-end, or both the 3'-end and the 5'-end (Supplemental Figure S1). We created and utilized these three fragments in the calculation of atomic partial charges in the morpholino ring and phosphorodiamidate groups. We used the RESP method[1] in conjunction with HF theory and the 6-31G* basis set. Because this is a standard choice for the level of theory used in the GAFF[7] and bsc0$_{\chi OL3}$ force fields (Refs. 35-38 in the main text), our calculations are consistent with these previous derivations of atomic partial charges. Several *ab initio* calculations were carried out for each fragment and for each of the guanine (G), cytosine (C), thymine (T), and adenine (A) bases; 3 runs per nucleobase (total of 36 runs). The partial charges for atoms in the morpholino ring and in the phosphorodiamidate group were averaged over these runs. Two exceptions were charges carried by the C1' and H1' atoms, which depend upon the attached nucleobase (Supplemental Figure S2). All atomic charges for nucleobases were obtained from bsc0$_{\chi OL3}$. Since the negatively charged O-atom within the canonical phosphate group is replaced by the uncharged -N(CH$_3$)$_2$ group, the total charge for all non-terminal "morpholino nucleotides" (Supplemental Figure S2) was set to zero. The Morpholino polyethyleneglycol (MPG) linker was constructed as described in Supplemental Methods.

Force field parameters: All-atom types for atoms in the phosphorodiamidate group, morpholino ring, and MPG linker were determined using the Antechamber package[8]. The molecular mechanical parameters, such as the equilibrium bond distances (r_0), spring constants for covalent bonds (k_b), equilibrium bond angles (θ_0), and spring constants for bending of bond angles (k_θ), as well as torsional angle parameters (magnitude associated with torsion energy ($V_n/2$), phase offset (γ) and periodicity (n)), were obtained from the bsc0$_{\chi OL3}$ force field for the atom types described therein. Parameters for atom types, which are not described by bsc0$_{\chi OL3}$, were generated based on GAFF[7] (Supplemental Tables S1 and S2).

PMO structures: We created the topology files for each of the "morpholino nucleotides" and for the MPG linker. Next, we linked "morpholino nucleotides" through the phosphorodiamidate groups and added the MPG linker at the 5'-end of each of the target sequences (see Figure 1 in the main part). These structures were used in all the MD simulations described in the main part.

Creating reference structures for calculation of thermodynamic state functions: To generate the extended initial (reference) conformations for each of the 22-mer, 25-mer, and 30-mer PMO molecules, we ran short (~30 ns) all-atom MD simulations with the first and last P-atoms in these molecules constrained. We selected 54 extended conformations for the 22-mer, 59 conformations for the 25-mer, and 63 extended conformations for the 30-mer PMO. These were used to generate the average reference enthalpy (H_{ref}) and reference entropy (S_{ref}) for the extended initial conformations for the 22-mer, 25-mer, and 30-mer PMOs shown in Figure 1B in the main part. To generate the unfolded initial (reference) conformations of PMO oligomers, we performed the all-atom MD simulations of thermal unfolding MD simulations using the Generalized Born (GB) model of implicit solvation[9] implemented in pmemd[10]. In these simulations, the solution conformations obtained from the numerical fit of the theoretical CD profiles to the experimental CD spectra were used as initial conformations. Each of the 22-mer, 25-mer, and 30-mer PMO molecules were gradually heated from 300 K to 500 K over a 2.5-μs time interval to make them unfold (see Figure S5). The numerical output from these MD simulations of thermal unfolding was then used in conjunction with the Support Vector Classifier approach to Machine Learning to identify the unfolded conformations for the 22-mer, 25-mer, and 30-mer PMOs (see Figure S6). We selected 54 unfolded conformations for the 22-mer, 59 conformations for the 25-mer, and 63 unfolded conformations for the 30-mer PMO. These conformations were used to generate the average reference enthalpy (H_{ref}) and reference entropy (S_{ref}) for the unfolded initial conformations for the 22-mer, 25-mer, and 30-mer PMOs shown in Figure 1C in the main part.

Analysis of MD simulation output: The numerical output from MD simulations for PMOs (coordinate and energy files) was used in data analyses and modeling. The end-to-end distance X was calculated as the distance between the P-atoms of the first and last "morpholino nucleotides". The radius of gyration R_g was calculated using the coordinates of all atoms, $R_g = (\sum_q^Q m_p \boldsymbol{r}_p^2 / \sum_q^Q m_p)^{1/2}$, where m_p is the mass of atom p and $\boldsymbol{r}_p$ is the position of atom p, relative to the center of mass of the molecule. The Solvent Accessible Surface Area (SASA) was estimated using the LCPO algorithm[11] implemented in the CPPTRAJ module[12] in AmberTools20[13]. The total number of base pairs and number of base stackings were calculated using Barnaba software[14]. The structure schematic for the calculation of the numbers of base pairs and base stackings is shown in Figure S2. Bases were classified as stacked if ($|z_{kj}|$ and $|z_{jk}|$ > 2Å) and (ρ_{kj} or ρ_{jk} < 2.5Å) and ($|\theta_{kj}|$ < 40°). Here, $\rho_{ij} = \sqrt{x_{kj}^2 + y_{kj}^2}$, where the x- and y-axes are in the plane of the base (x_{kj} and y_{kj} are the distances between the centers of mass of the two bases along the x- and y-axes, respectively) and the z-axis is normal to the xy-plane, z_{kj} is the distance between the centers of mass of the two bases, and θ_{kj} is the angle between the normal vectors of the two bases (Figure S2; Ref.[14]). All the non-stacked bases are considered to be base-paired if $|\theta_{kj}|$ < 60° and there exists at least one hydrogen bond (H-bond) between k-th and j-th bases (Figure S2 in the main part). We assume that the H-bond D–H…A between the hydrogen donor atom (D) and acceptor atom (A) is formed if the donor–acceptor distance d_{DA} is less than the 3.3Å cutoff and the bond angle is larger than the 140° cutoff (Ref.[14]). To avoid selecting similar conformations we used the eRMSD measure of structural similarity (Ref.[15]) implemented in the Barnaba software[14]. Briefly, the eRMSD is a contact map-based distance metric, with the addition of a number of features that make it suitable for the comparison of structures of nucleic acids. We used eRMSD to prescreen the output from the all-atom MD simulations for 22-mer, 25-mer, and 30-mer PMOs and to discard similar structures from subsequent data analysis.

Replica Exchange Molecular Dynamics: Temperature Replica Exchange Molecular Dynamics (T-REMD) is an enhanced sampling technique[16]. T-REMD simulations were performed with the Generalized Born (GB) model of implicit solvation[9] implemented in pmemd[10]. A total of 10 independent replicas running at different temperatures (300 K, 318 K, 336 K, 354 K, 372 K, 390 K, 408 K, 426 K, 444 K, and 462 K) were simulated in parallel. After a 1 ns time interval, replica exchange was attempted between neighboring replicas, i.e. between 300 K and 318 K replicas, between 318 K and 336 K replicas, between 336 K and 354 K replicas, etc. Successful exchange probability p_{ij} was determined by the i-th and j-th replica pair's temperature and energy differences based on the Metropolis criterion: $p_{ij} = \min\{1, \exp[-(\beta_i - \beta_j)(E_j - E_i)]\}$, where $\beta_i = 1/k_B T_i$ and $\beta_j = 1/k_B T_j$, and T_i, T_j and E_i, E_j are temperatures and potential energies of the i-th and j-th replicas (k_B is a Boltzmann constant). Conformations were extracted from T-REMD trajectories every 5 ns for data analysis.

We employed T-REMD simulations to theoretically model CD spectra for the 30-mer PMO at higher T = 328-358 K range because non-linear regression and numerical fitting of the ensemble average theoretical profiles to the experimental CD spectra for 30-mer PMOs in the 328-358 K temperature range resulted in poor fits and large associated values of MSE (results not shown). This points to insufficient sampling of statistical ensembles of the 30-mer PMO conformations at elevated temperatures. A total of 10 replicas were simulated starting from their extended conformation as initial structures, each for 750 ns, in the temperature range between 300 K and 462 K. Due to replicas rapidly exchanging between different temperatures, changes in the kinetic energy and subsequent energy

redistribution between the kinetic and potential energy forms lead to changes in the potential energy of the system, which is why multiple unfolding-refolding transitions are observed. We observed a large number of unfolding-refolding transitions for the 30-mer PMOs. Hence, our use of T-REMD resulted in the covering of a much larger conformational space than the brute force use of conventional MD simulations.

Classification of unfolded conformations of PMOs: To generate the unfolded conformations for the 22-mer, 25-mer, and 30-mer PMO molecules, we carried out 4 independent 2.5-μs long MD simulations of thermal unfolding, using the first 4 most populated conformations as initial structures (first 3 of them are shown in Figure 2 in the main part). In these simulations, each of the 22-mer, 25-mer, and 30-mer PMOs were gradually heating them from 300 K to 500 K. The numerical output from these MD simulation runs (coordinate and energy files) were then used to identify molecular properties that best characterize the unfolding transitions in the 22-mer, 25-mer, and 30-mer PMO molecules at high temperatures (order parameters). The temperature-dependent profiles of the end-to-end distance X, radius of gyration R_g, number of base stacks N_{BS}, and Solvent Accessible Surface Area (SASA), extracted as described in Materials and Methods (see main text) shows the sigmoidal shape characteristic of a phase transition (Figure S5). This has helped us to select the unfolded conformations of the 22-mer, 25-mer, and 30-mer PMO molecules to form the training set for the Machine Learning based classification of the ensemble of conformations for these PMOs. Specifically, we selected the conformations for which the values of X, R_g, N_{BS}, and SASA are in the upper 90% of their maximum values (see horizontal dashed lines in Figure S5A-D).

For each PMO system, we constructed the dataset containing the values of X, R_g, N_{BS}, SASA and eRMSD. These quantities are described in detail in Materials and Methods (see main part). The numerical output containing a manifold of various conformations for the PMO molecules, obtained from the equilibrium MD simulations in an aqueous solution at 300 K, were used to form datasets $D_{eq,22}$, $D_{eq,25}$ and $D_{eq,30}$ for the 22-mer, 25-mer, and 30-mer PMOs, respectively. The numerical output data for PMOs' conformations obtained from the MD simulations of thermal unfolding were used to form datasets $D_{unf,22}$, $D_{unf,25}$ and $D_{unf,30}$ for the 22-mer, 25-mer, and 30-mer PMOs, respectively. For each PMO system, these datasets were combined into datasets D_{22}, D_{25} and D_{30}, with two classes labeled 'Folded' and 'Unfolded'. These combined datasets were randomly separated into two halves forming the training sets ($D_{train,22}$, $D_{train,25}$ and $D_{train,30}$) and the test sets ($D_{test,22}$, $D_{test,25}$ and $D_{test,30}$). Next, we employed the Support Vector Machines (SVM) method to perform data classification, i.e. separation of PMOs' structures, generated in MD simulations, into the folded conformations ('Folded' class) and the unfolded conformations ('Unfolded' class). The implementation of the SVM algorithm was based on Python scikit-learn package[17].

SVM was implemented with the linear, polynomial (degree up to 5), sigmoidal, and radial basis function (RBF) kernels. Using each kernel, we extracted the confusion matrices, which label the class chosen for each prediction against the true (known) class. Each matrix compares the predictions to their correct classes, and, for each matrix entry, we report the number of predictions and the percentage of predictions per total number of data points in a sample. Using the obtained confusion matrices, we calculated the number of false negatives (FN), i.e. the conformations from the MD simulations of thermal unfolding which were classified as 'Folded' class, and the number of false positives (FP), i.e. the conformations from 'Folded' class which were classified as the unfolded conformations. The latter are the target structures. These were identified the best with the RBF kernel giving the smallest FN number but the largest FP number. Figure S5 shows the confusion matrices for SVM obtained with the RBF kernel for the 22-mer, 25-mer, and 30-mer PMOs applied to the test sets $D_{test,22}$, $D_{test,25}$ and

$D_{test,30}$. Since the test sets contain only a half of the unfolded structures generated at 300 K, next, we applied these pre-trained SVM models to the data sets $D_{eq,22}$, $D_{eq,25}$ and $D_{eq,30}$, in order to extract all the unfolded PMOs' conformations that would exclude conformations from the MD simulations of thermal unfolding. Therefore, we identified 54 conformations for the 22-mer, 59 conformations for the 25-mer, and 63 conformations for the 30-mer PMO. The results obtained for the classification of conformations of the 30-mer PMO into the 'Folded' class and 'Unfolded' class are displayed in Figure S5, which shows the two-dimensional projections of the separating hypersurface, i.e. N_{BS} vs. R_g and N_{BS} vs. SASA.

Supplemental Movies

Video S1. Folding of the 22-mer PMO: The movie shows the conformational transition of the 22-mer PMO from the extended state to a collapsed state (folded state) as observed in a 150-ns MD simulation run at T = 300 K. The MD run was carried out in explicit water (cyan transparent spheres). The PMO molecule is shown in the Licorice representation (sticks) and in the Twister representation for the backbone (blue line). The MPG-linker is shown in orange, A and T bases are shown in green, and C and G bases are shown in red. The length of the movie is 25 s (the movie is played ~167,000,000 times slower than the computational experiment).

Video S2. Conformational dynamics of the 22-mer PMO: The movie shows conformational fluctuations of the 22-mer in the folded state as observed in a 1-μs MD simulation run at T = 300 K. The MD run was carried out in explicit water (cyan transparent spheres). The PMO molecule is shown in the Licorice representation (sticks) and in the Twister representation for the backbone (blue line). The MPG-linker is shown in orange, A and T bases are shown in green, and C and G bases are shown in red. The length of the movie is 16 s (the movie is played ~16,000,000 times slower than the computational experiment).

Video S3. Conformational dynamics of the 30-mer PMO: The movie shows conformational fluctuations of the 30-mer in the folded state as observed in a selected 400 ns of a 1-μs MD simulation run at T = 300 K. The MD run was carried out in explicit water (cyan transparent spheres). The PMO molecule is shown in Twister representation (for the backbone; blue line) and in PaperChain representation (for nucleic bases). The movie shows different base-pairing and base-stacking arrangements, which correspond to variable enthalpy, for a given fixed conformation of a PMO's backbone. The length of the movie is 15 s (the movie is played ~37,500,000 times slower than the computational experiment).

Supplemental Table

Table S1. Molecular Mechanical parameters for covalent bonds and bond angles in PMOs: Shown for each covalent bond are the equilibrium covalent bond distance r_0 and spring constant k_b, and for each bond angle the equilibrium bond angle θ_0 and bending angle spring constant k_a. These force field parameters were determined for the PMOs comprising the oligonucleotide backbone explored in this study (see Material and Methods in the main text). The following atom types are described: CT – sp^3 hybridized carbon with 4 explicit substituents; NT – sp^3 hybridized nitrogen with 4 explicit substituents; H1 – hydrogen on aliphatic carbon with 1 electron-withdrawing group; OS – ester oxygen; O2 – phosphate oxygen; H – amide hydrogen; P – phosphorus in phosphate group; and HC – aliphatic carbon.

Bond	k_b, kcal/mol/Å	r_0, Å
CT-NT	326	1.465
CT-CT	310	1.526
CT-H1	340	1.090
CT-OS	320	1.410
P-O2	525	1.480
P-OS	230	1.610
NT-H	434	1.010
NT-P	367	1.670
Angle	k_θ, kcal/mol/rad²	θ_0, deg
NT-CT-CT	80.0	111.20
NT-CT-HC	49.5	109.88
CT-CT-HC	50.0	109.50
H1-CT-H1	35.0	109.50
CT-NT-CT	50.0	109.50
CT-NT-H	50.0	109.50
CT-CT-OS	50.0	109.50
OS-CT-H1	50.0	109.50
CT-OS-CT	60.0	109.50
P-OS-CT	100.0	120.50
CT-CT-CT	40.0	109.50
CT-NT-H	47.4	109.29
CT-NT-P	78.4	119.86
NT-P-O2	42.9	114.64
NT-P-OS	43.9	102.23
NT-P-NT	42.7	103.37

Table S2. Molecular Mechanical parameters for torsion angles in PMOs: Shown for each torsion angle are the number of bond paths and the magnitude of torsion energy $V_n/2$, the phase offset γ, and the periodicity of torsion n. Atom X denotes any atom of the atom type described in the caption to Table S1.

Torsion angle	no. of paths	$V_n/2$, kcal/mol	γ, deg	n
X–CT–CT–X	9	1.4	0.0	3.0
X–CT–NT–X	6	1.8	0.0	3.0
CT–NT–P–O2	1	3.0	180.0	-2.0
CT–OS–P–NT	3	2.4	0.0	2.0
CT–CT–NT–P	6	1.8	0.0	3.0
CT–NT–P–OS	6	18.0	180.0	2.0
CT–NT–P–NT	6	18.0	180.0	2.0
X–CM–P–X	6	8.8	180.0	2.0

Supplemental Figures

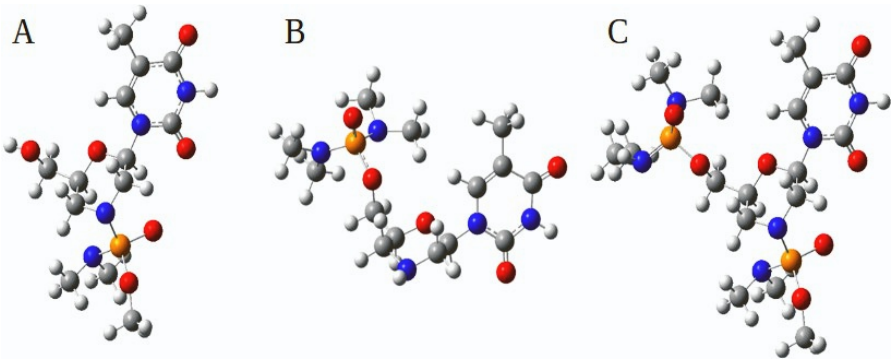

Figure S1. Structures used in RESP calculations of atomic partial charges for PMOs: Shown are the phosphorodiamidate morpholino (PM) nucleotide structures used for calculating partial charges (see Materials and Methods in the main text): PM with the phosphate group attached to the 3'-end (panel **A**), attached to the 5'-end (panel **B**), and attached to both the 3'-end and the 5'-end (panel **C**). Color denotation: the C-atoms are shown in grey; the N-atoms are in blue, the O-atoms are in red, the P-atoms are in orange, and the H-atoms are shown in white color.

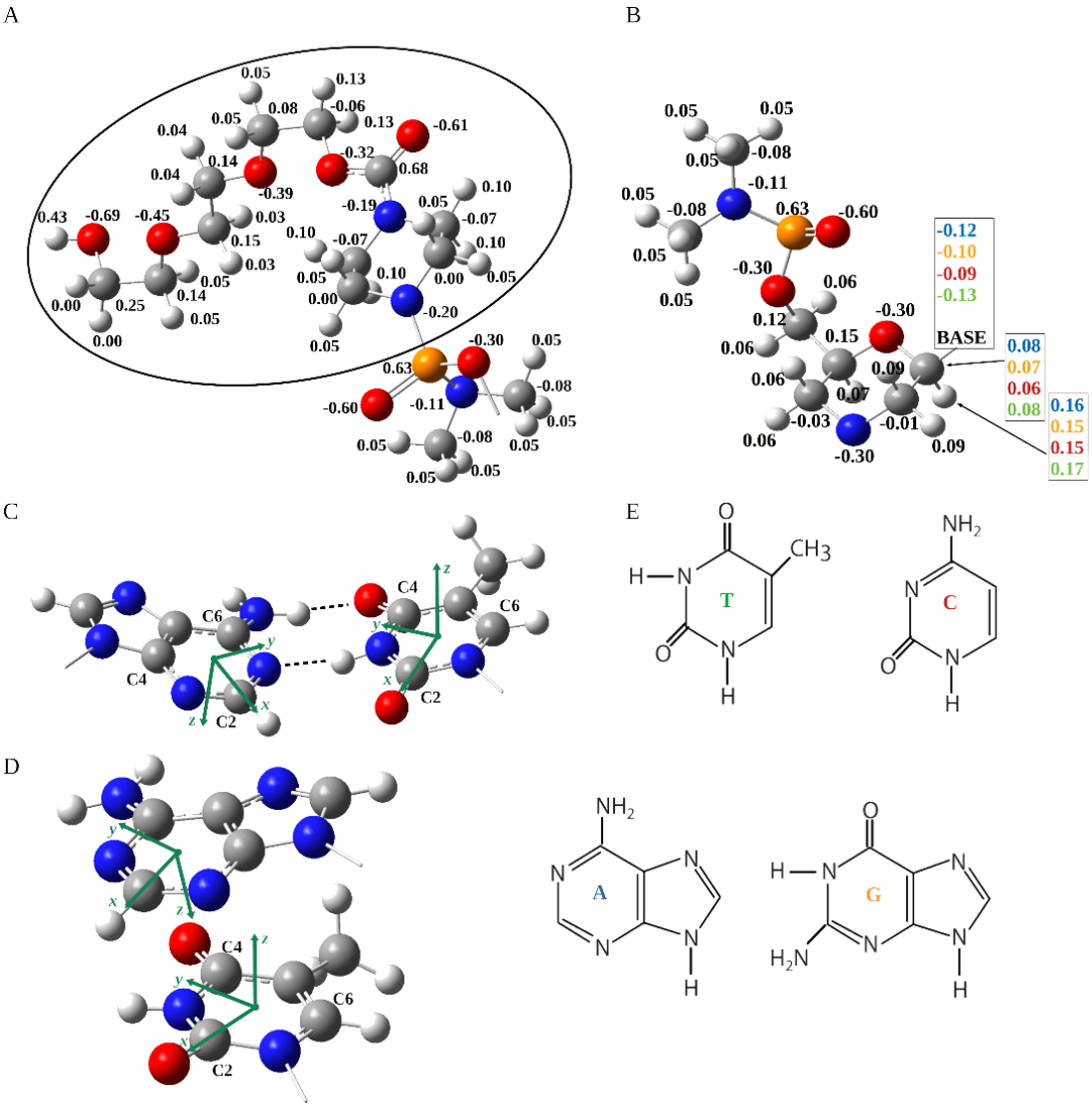

Figure S2. Atomic partial charges for PMO backbone structure: Panel **A**: Partial charges for the 5'-end (MPG) linker (circled). Panel **B**: Partial charges for the non-terminal PMO residues. The charges on the atoms of nucleic acid bases used are the same as in the bsc0$_{\chi OL3}$ force field, as well as partial charges carried by C1' and H1' atoms. All charges for morpholino ring and phosphorodiamidate groups were calculated using the RESP method (see Supplemental Methods in main text). For each of the four bases the total charge is shown in different colors for each base: for A in blue, for C in red, for G in orange, and for T in green. Panel **C**: Local coordinate systems for purines and pyrimidines used in the calculation of base pairing and base stacking interactions (Supplemental Methods in main text). The center of the base ring atoms C2, C4, and C6 represents the origin of the coordinate system as shown. The *x*-and *y*-axes lie in the plane of the base while the *z*-axis lies normal to the *xy*-plane. The *x*-axis is pointed in the C2-atom direction, and the *y*-axis is pointed toward the C4-atom (for C and U) or toward the C6-atom (for A and G). Two bases are forming a base pair via hydrogen bonds represented as dashed black lines. Panel **D**: Same as Panel **C**, but bases are forming base stacking interaction. Panel **E**: Structures of four ribonucleic acid bases used in the computational modeling of PMO molecules.

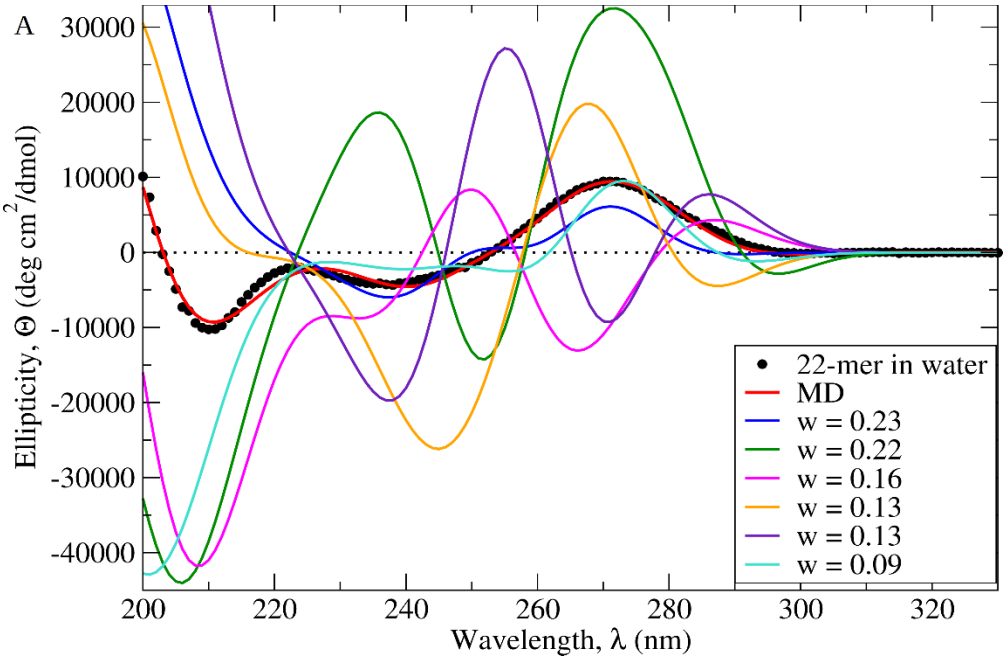

Figure S3. Circular Dichroism spectra for 22-mer PMO: Theoretical CD spectra for the six most important (most populated) conformations I-VI generated *in silico* for the 22-mer PMO complementary to Exon 45 (see Figures 2 and 4 in main text for conformers' structures) with the corresponding weights (equilibrium populations) shown in *the inset*. The experimental data points are displayed as black data points (average spectrum), and the theoretical profiles are shown using differently colored curves (the average theoretical spectrum from the MD simulations is shown as a red curve).

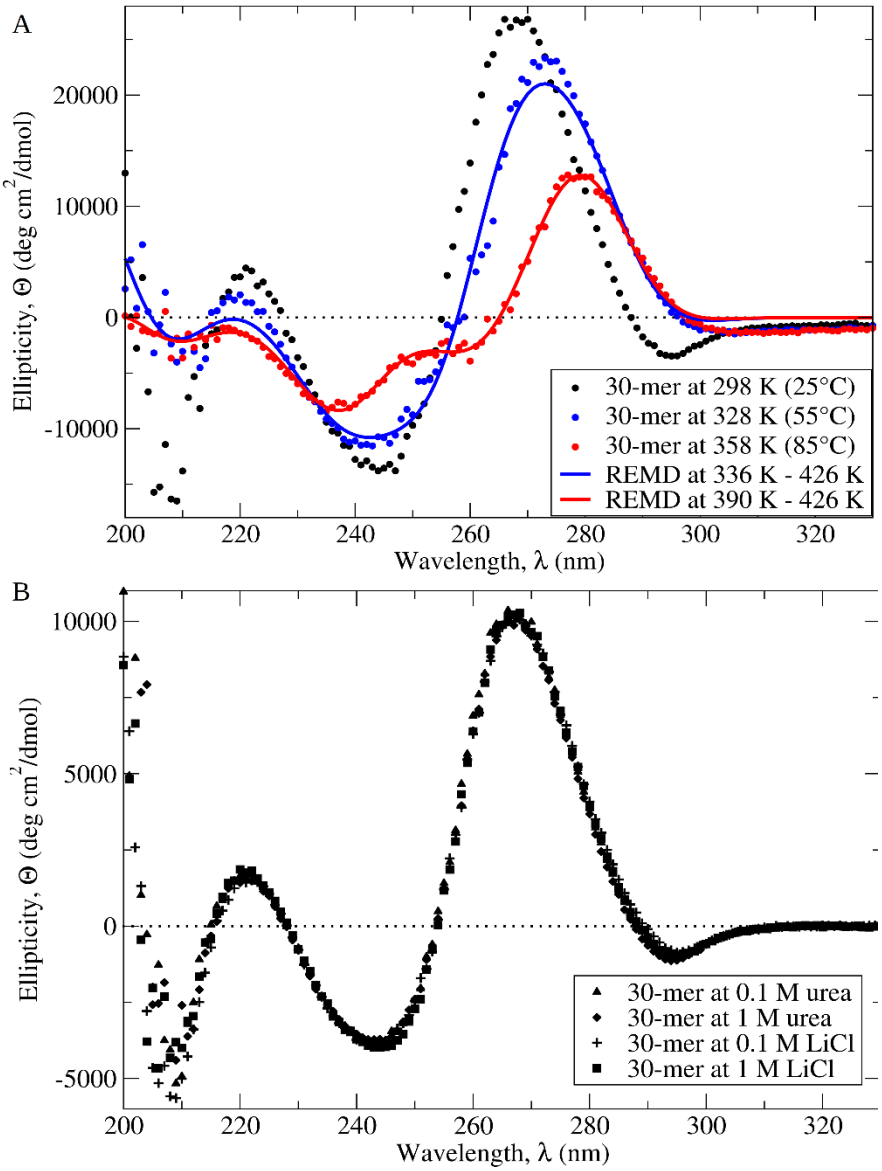

Figure S4. Circular Dichroism spectra for 30-mer PMO: Panel **A**: Experimental CD spectra for the 30-mer PMO complementary to Exon51 determined at three different temperatures 25ºC, 55ºC and 85ºC (color denotation is explained in *the inset*). Experimental data points are superposed with theoretical CD profiles (solid lines) for 30-mer PMO. The theoretical CD curves modeling experimental CD spectra at 55ºC were constructed using conformers obtained from T-REMD simulations in the 336K-426K temperature range and non-linear fitting (see Materials and Methods in the main text). The experimental CD spectrum at 85°C was modelled using conformers from T-REMD simulations in the 390K-426K temperature range. Panel **B**: Experimental CD spectra for the 30-mer PMO complementary to Exon51 corresponding to 0.1 M and 1 M urea solution, and 0.1 M and 1 M LiCl solution (color denotation is explained in *the inset*).

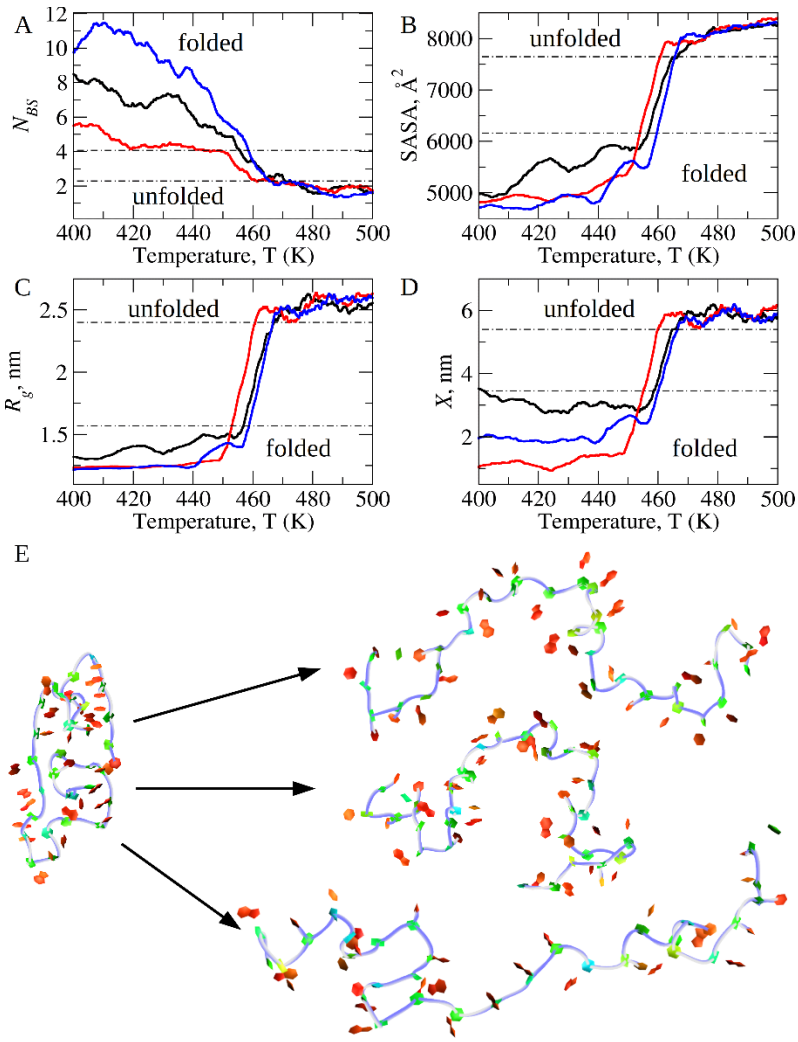

Figure S5. Dynamic structural properties and structure alterations associated with thermal unfolding of 30-mer PMO: Panels **A-D** display temperature-dependent profiles of the secondary and tertiary structural properties of the 30-mer from three independent 2.5-µs MD simulation runs of gradual heating of this PMO from 300 K to 500 K: number of base-stacking interactions N_{BS} (panel **A**), Solvent Accessible Surface Area (SASA; panel **B**), radius of gyration R_g (panel **C**), and end-to-end distance X (panel **D**). Only the results of 400 K – 500 K portion of the thermal unfolding experiments are shown for clarity. Horizontal dashed-dotted lines represent the regions corresponding to the folded and unfolded conformations. Panel **E** shows an example of the initial folded conformation for the thermal unfolding simulation (structure II from Figure 2C in the main next; left) and three representative unfolded conformations formed in the course of 3 independent MD runs (right). The folded and extended conformations are displayed in the Twister representation (backbone; blue line) and in PaperChain representation (bases).

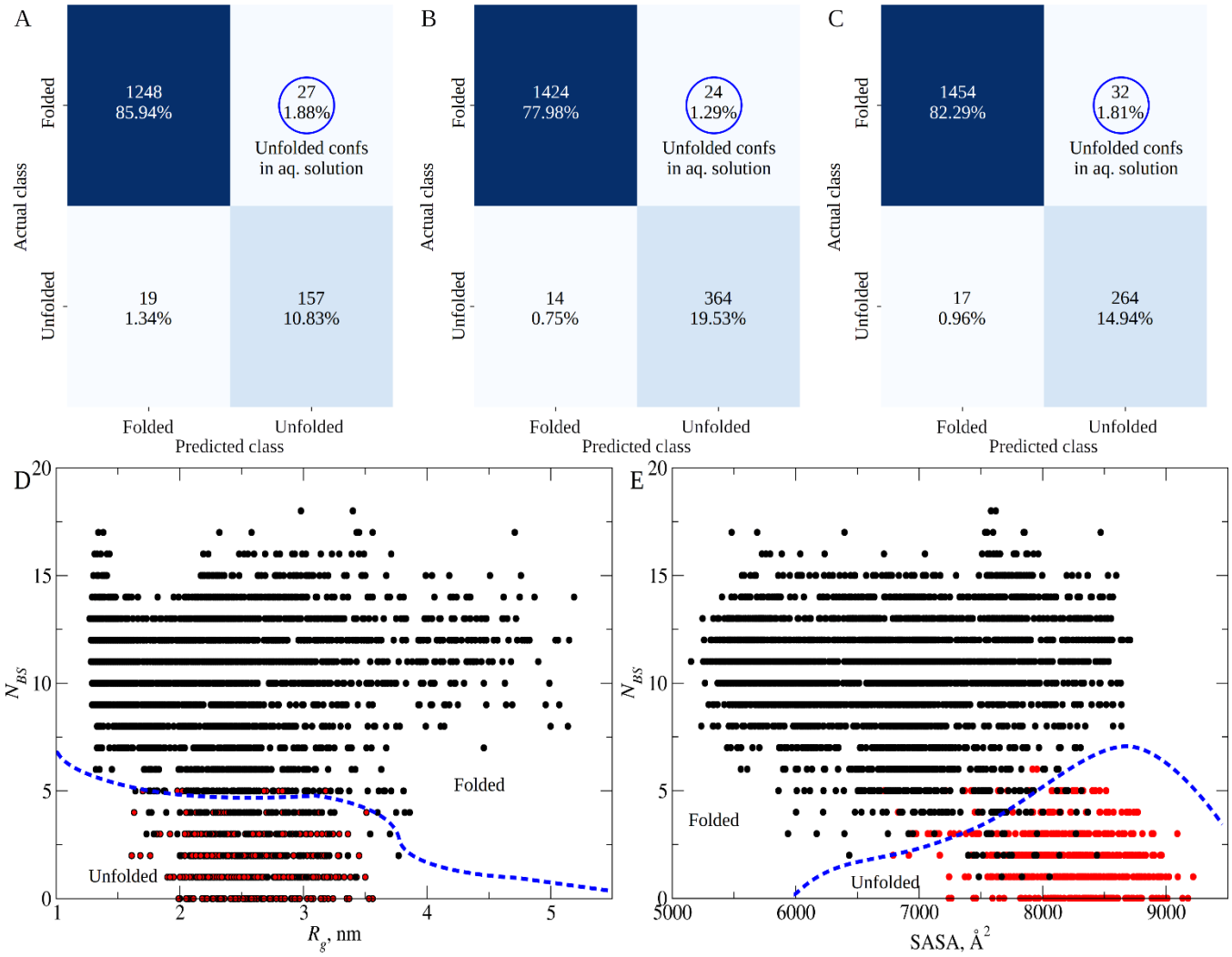

Figure S6. Support Vector Machines based analysis of conformations for 22-mer, 25-mer, and 30-mer PMOs: Confusion matrices describing the number of predictors made using the SVM technique compared to every prediction's true class for the 22-mer ($D_{test.22}$; panel **A**), 25-mer ($D_{test.25}$; panel **B**) and 30-mer ($D_{test.30}$; panel **C**) PMOs. The values in the center of each box indicate the number of predictions and the percent of predictions made for a certain class compared to the overall number of true class samples. The darker color of the cells corresponds to the larger amount of data points. Also shown are the scatter plots of a number of base-stacking interactions N_{BS} vs. radius of gyration R_g (panel **D**) and vs. Solvent Accessible Surface Area (SASA; panel **E**) for the 30-mer PMO. The black data points correspond to the 30-mer PMO's conformations extracted from the equilibrium MD simulations in an aqueous solution at 300 K, while the red data points represent the conformations observed in the MD simulations of thermal unfolding. The blue dashed lines represent the two-dimensional projection of an optimal hypersurface obtained using the Support Vector Classifier, which separates the conformations that belong to the 'Folded' class and 'Unfolded' class. The black data points under the border line which belong to the 'Unfolded' class are the initial unfolded (reference) structures used in data analysis (see Figure 1C in the main text).

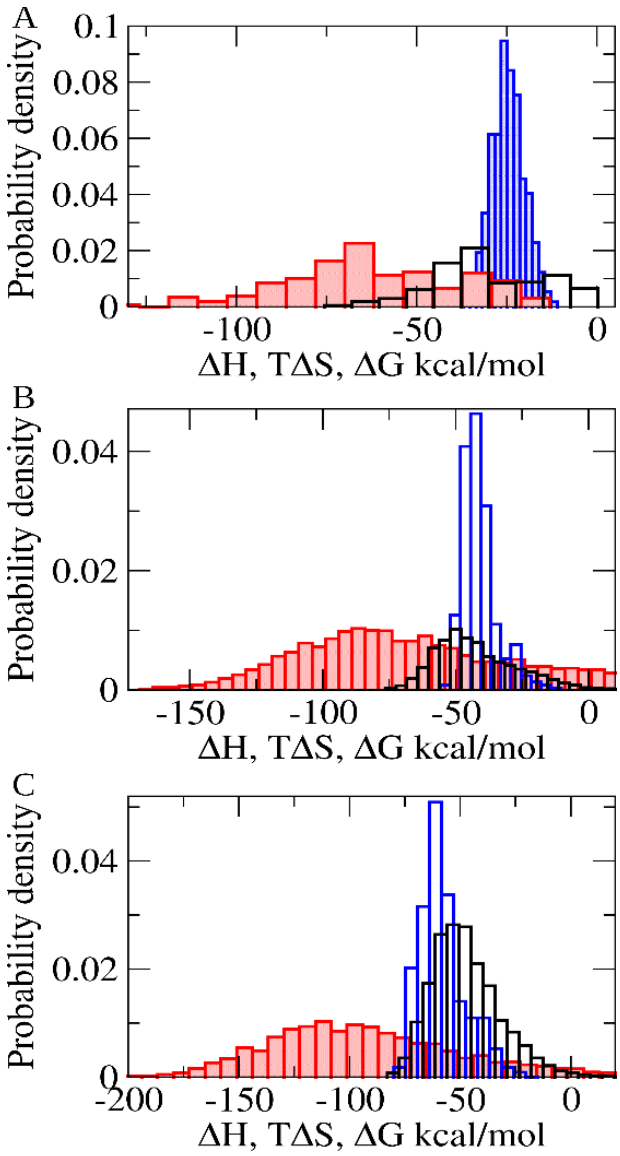

Figure S7. Histograms of thermodynamic state functions for 22-mer, 25-mer, and 30-mer PMOs: Shown are the histogram-based estimates of the probability distributions for the values of enthalpy change ΔH (red bars), entropy change $T\Delta S$ (blue bars) and free energy change ΔG (black bars) for folding of the 22-mer PMO (panel **A**), 25-mer PMO (panel **B**) and 30-mer PMO (panel **C**), obtained using the unfolded solution structures as the reference states. The distributions were sampled using the output from the equilibrium MD simulations at T = 300 K temperature (see Materials and Methods in the main text).

Supplemental references:

1. Cieplak P, Cornell WD, Bayly C, Kollman PA. Application of the multimolecule and multiconformational RESP methodology to biopolymers: Charge derivation for DNA, RNA, and proteins. *J Comput Chem*. 1995;16(11):1357-1377.
2. Vanquelef E, Simon S, Marquant G, et al. RED Server: a web service for deriving RESP and ESP charges and building force field libraries for new molecules and molecular fragments. *Nucleic Acids Res*. 2011;39:W511-W517.
3. Caleman C, Van Maaren PJ, Hong M, Hub JS, Costa LT, Van Der Spoel D. Force field benchmark of organic liquids: density, enthalpy of vaporization, heat capacities, surface tension, isothermal compressibility, volumetric expansion coefficient, and dielectric constant. *J Chem Theory Comput*. 2012;8(1):61-74.
4. Dennington R, Keith T, Millam J, others. GaussView, version 5. Published online 2009.
5. Frisch MJ, Trucks GW, Schlegel HB, et al. Gaussian 16 rev. *Gaussian Inc Wallingford, CT, USA*. Published online 2016.
6. Xie M, Zhu G, Hu Y, Gu H. Conformations of morpholine in liquid and adsorbed on gold nanoparticles explored by Raman spectroscopy and theoretical calculations. *J Phys Chem C*. 2011;115(42):20596-20602.
7. Wang J, Wolf RM, Caldwell JW, Kollman PA, Case DA. Development and testing of a general amber force field. *J Comput Chem*. 2004;25(9):1157-1174.
8. Wang J, Wang W, Kollman PA, Case DA. Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model*. 2006;25(2):247-260.
9. Qiu D, Shenkin PS, Hollinger FP, Still WC. The GB/SA continuum model for solvation. A fast analytical method for the calculation of approximate Born radii. *J Phys Chem A*. 1997;101(16):3005-3014.
10. Darden T, York D, Pedersen L. Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *J Chem Phys*. 1993;98(12):10089-10092.
11. Weiser J, Shenkin PS, Still WC. Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO). *J Comput Chem*. 1999;20(2):217-230.
12. Roe DR, Cheatham III TE. PTRAJ and CPPTRAJ: software for processing and analysis of molecular dynamics trajectory data. *J Chem Theory Comput*. 2013;9(7):3084-3095.
13. Case DA, Belfon K, Ben-Shalom I, et al. Amber 2020. Published online 2020.
14. Bottaro S, Bussi G, Pinamonti G, Reißer S, Boomsma W, Lindorff-Larsen K. Barnaba: software for analysis of nucleic acid structures and trajectories. *Rna*. 2019;25(2):219-231.
15. Bottaro S, Di Palma F, Bussi G. The role of nucleobase interactions in RNA structure and dynamics. *Nucleic Acids Res*. 2014;42(21):13306-13314.
16. Sponer J, Bussi G, Krepl M, et al. RNA structural dynamics as captured by molecular simulations: a comprehensive overview. *Chem Rev*. 2018;118(8):4177-4338.

17. Pedregosa F, Varoquaux G, Gramfort A, et al. Scikit-learn: Machine learning in Python. *J Mach Learn Res*. 2011;12:2825-2830.