Molecular Therapy
Nucleic Acids
Original Article

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

Farkhad Maksudov,[1,4] Evgenii Kliuchnikov,[1,4] Daniel Pierson,[2,4] M.L. Ujwal,[3] Kenneth A. Marx,[1,3] Arani Chanda,[2] and Valeri Barsegov[1,3]

[1]Department of Chemistry, University of Massachusetts, Lowell, MA 01854, USA; [2]Technical Operations, Sarepta Therapeutics, Cambridge, MA 02142, USA; [3]Inciton, Inc., Andover, MA 01854, USA

Elucidating the structure-function relationships for therapeutic RNA mimicking phosphorodiamidate morpholino oligonucleotides (PMOs) is challenging due to the lack of information about their structures. While PMOs have been approved by the US Food and Drug Administration for treatment of Duchenne muscular dystrophy, no structural information on these unique, charge-neutral, and stable molecules is available. We performed circular dichroism and solution viscosity measurements combined with molecular dynamics simulations and machine learning to resolve solution structures of 22-mer, 25-mer, and 30-mer length PMOs. The PMO conformational dynamics are defined by the competition between non-polar nucleobases and uncharged phosphorodiamidate groups for shielding from solvent exposure. PMO molecules form non-canonical, partially helical, stable folded structures with a small 1.4- to 1.7-nm radius of gyration, low count of three to six base pairs and six to nine base stacks, characterized by -34 to -51 kcal/mol free energy, -57 to -103 kcal/mol enthalpy, and -23 to -53 kcal/mol entropy for folding. The 4.5- to 6.2-cm^3/g intrinsic viscosity and Huggins constant of 4.5–9.9 are indicative of extended and aggregating systems. The results obtained highlight the importance of the conformational ensemble view of PMO solution structures, thermodynamic stability of their non-canonical structures, and concentration-dependent viscosity properties. These principles form a paradigm to understand the structure-properties-function relationship for therapeutic PMOs to advance the design of new RNA-mimic-based drugs.

INTRODUCTION

Single-stranded nucleic acids have been studied for decades, first in attempts to understand the nature of structural variability in both naturally occurring DNA and RNA, then later in efforts to begin mimicking the functionality of specific single-stranded RNAs in order to understand their roles in normal cellular functions as well as disease etiology and progression.[1,2] More recent attempts to utilize canonical RNA for an array of therapeutic purposes (gene editing, replacement, addition, and inhibition) were challenged by a number of their undesirable properties, which included immune activation propensity, susceptibility to nuclease digestion in the cellular environment, and unfavorable physicochemical properties inhibiting transport into cells, thereby requiring the development of specialized cellular delivery technologies.[3] To overcome these limitations, a number of backbone modifications have been developed recently, resulting in the current diverse array of backbone platform replacement approaches that have led to a number of successful nucleic acid therapeutics being approved for clinical use.[4]

Phosphorodiamidate morpholino oligonucleotides (PMOs) (Figure 1) are a subclass of antisense oligonucleotides (ASOs) that have the canonical nucleic acid backbone replaced by morpholino rings connected by phosphorodiamidate linkages.[5–7] They are single-stranded DNA analogs that have favorable biophysical properties, including high solubility in an aqueous medium, low to no metabolic degradation primarily due to their neutral charge, and high duplex stability, which make this class of compounds useful as potential therapeutics.[8–14] PMOs have been approved by the US Food and Drug Administration (FDA) for the treatment of Duchenne muscular dystrophy (DMD) since 2016,[15–17] and they have been designed to target Marburg, Ebola,[18,19] Picornaviruses, and other viruses,[8,20] along with bacterial targets and have been developed as anti-cancer agents.[21] For treatment of DMD, PMOs can be designed to target regions of the dystrophin pre-mRNA to allow skipping of a targeted exon, and restoration of the mRNA reading frame allowing for translation of a shortened, yet functional, dystrophin protein. PMOs can be designed to bind to complementary sequences in target mRNA by Watson-Crick base pairing to effectively block translation through an RNase H-independent steric blockade.[22] Furthermore, PMOs can also be effective as antiviral agents, since the formation of a

---

Received 27 October 2022; accepted 10 February 2023;
https://doi.org/10.1016/j.omtn.2023.02.007.

[4]These authors contributed equally

Correspondence: Arani Chanda, Technical Operations, Sarepta Therapeutics, Cambridge, MA 02142, USA.
E-mail: achanda@sarepta.com
Correspondence: Valeri Barsegov, Department of Chemistry, University of Massachusetts, Lowell, MA 01854, USA.
E-mail: valeri_barsegov@uml.edu

A

Exon45: 5’-CAA TGC CAT CCT GGA GTT CCT G-3’ 22-mer 4A, 7C, **5G**, 6T (23%G)
Exon53: 5’-GTT GCC TCC GGT TCT GAA GGT GTT C-3’ 25-mer 2A, 6C, **8G**, 9T (32%G)
Exon51: 5’-CTC CAA CAT CAA GGA AGA TGG CAT TTC TAG-3’ 30-mer 10A, 7C, **6G**, 7T (20%G)

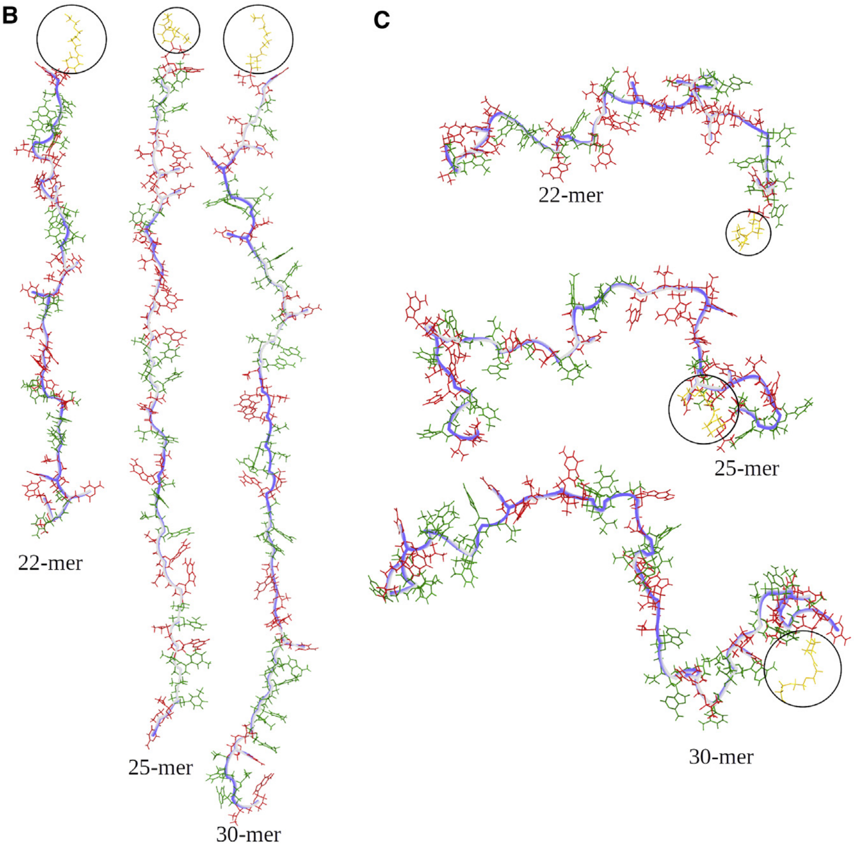

Figure 1. 22-mer, 25-mer, and 30-mer PMO structures

Shown for three therapeutic PMOs, Exon45 (22-mer PMO), Exon53 (25-mer PMO), and Exon51 (30-mer PMO), are the primary sequence (A) and the reference structures: initial extended conformations (B) and unfolded conformations (C). The PMOs represent therapeutic oligomers complementary to the different indicated exon transcripts in the human dystrophin gene. Each sequence is shown with its base composition, the total amount of nucleobases, i.e., adenine, A; cytosine, C; guanine, G; and thymine, T, and the relative amount of G bolded (percentage). The conformers are shown in licorice representation (sticks) and in twister representation (blue line) describing the backbone. The MPG linker is shown in orange (within the black circle). The A and T bases are shown in green color, whereas C and G bases are shown in red.

PMO:mRNA duplex may effectively block translation of the viral RNA genome, thereby inhibiting viral replication.[8] PMOs have been shown to resist the activity of a variety of enzymes present in biological fluids, including nucleases, proteases, esterases, and hydrolases.[23] In addition, due to their uncharged backbone and overall neutral charge, interactions of PMO with cellular proteins are minimal, helping to limit therapeutic side effects.[24,25] Therefore, these attractive properties make PMOs very different from DNA, RNA, and other ASOs in terms of their solution structures, roles in therapeutic activities, and interactions with other biological moieties.

While conformational transitions in biomolecules (DNA, RNA, and proteins) in general, and RNA molecules in particular, have been studied extensively, very little is known about dynamic structural transitions in PMOs, whose backbone structure is distinctly different from canonical nucleic acids. There is no information about the specific PMO structures that exist in aqueous solution, and there are no experimentally resolved crystal or NMR structures available for PMO oligomers. To date, only one study of a related but distinctly different morpholino backbone has been published on the oligomers’ physical properties (melting temperature, circular dichroism (CD), thermodynamic parameters).[26] As a physical technique, CD spectroscopy has proved very useful over the decades in characterizing low-resolution aspects of the solution structures of nucleic acids, especially the presence of structured, helical regions of both double- and single-stranded nucleic acids.[27] Importantly, it is possible to calculate CD spectra theoretically using structures from the all-atom molecular dynamics (MD) simulations[28,29] in order to interpret the experimental CD profiles, but this can only be performed if PMOs’ solution structures are known. However, the lack of any known high-resolution structures makes this comparison problematic. This is in addition to the lack of atomic force fields available that are required for a computational exploration of PMO’s dynamic structural properties.

To overcome these multiple limitations, we carried out for the first time combined experimental and computational studies of the dynamic structural properties of PMOs, in which structural inputs from the all-atom MD simulations have been used to inform experimental measurements. We, first, calculated atomic partial charges and developed the atomic force field for those atoms and atomic groups forming the morpholino ring and the phosphorodiamidate group in the PMO backbone. These efforts enabled us to then carry out microsecond-long all-atom MD simulations using PMOs of therapeutics-based size composed of 22 nucleobases (22-mer), 25 nucleobases (25-mer), and 30 nucleobases (30-mer) as important study systems. These three oligomers’ sequences are complementary to exon 45, exon 53, and exon 51, respectively, of the dystrophin gene pre-mRNA transcript.[4] By directly correlating the CD spectra measured experimentally with those calculated theoretically based on the output from the all-atom MD simulations, we resolved the ensemble of PMO conformer structures that exist in an aqueous

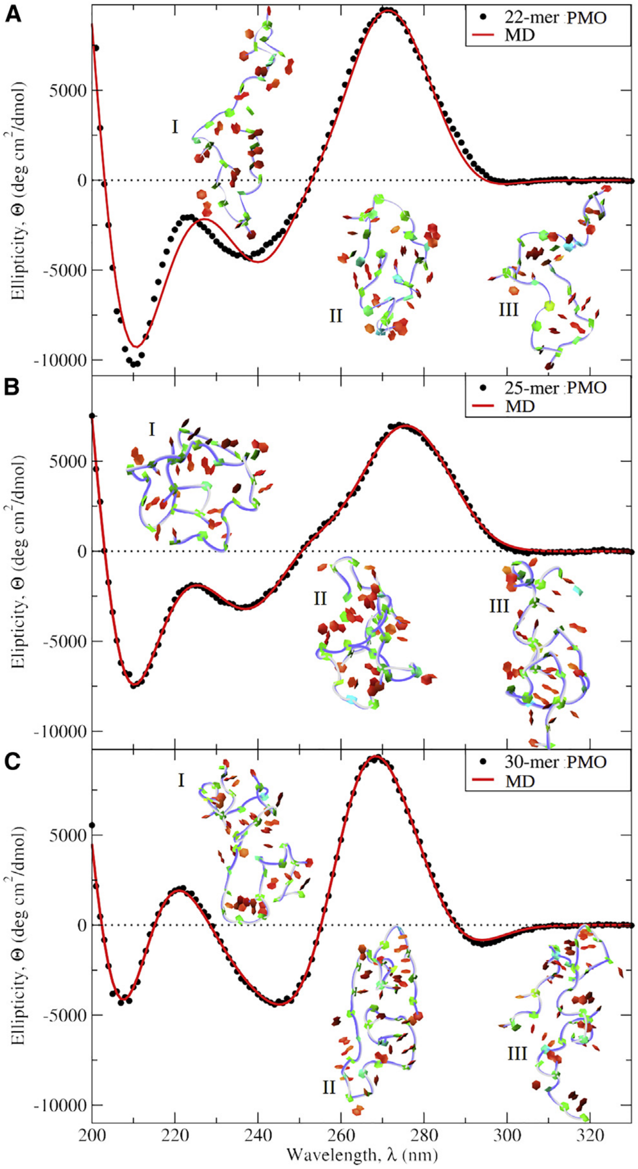

Figure 2. Theoretical reconstruction of CD spectra for 22-mer, 25-mer, and 30-mer PMOs

Superposed are the CD profiles for the 22-mer (A), 25-mer (B), and 30-mer (C) PMOs obtained experimentally (black solid lines) and theoretically (red dashed lines). The snapshots of PMO structures generated *in silico* numbered I–III, which correspond to the most representative, highest-weight-solution conformations contributing to the ensemble average CD spectra, are shown in twister representation (blue line going through backbone) and in PaperChain representation (for nucleic bases).

solution for these three therapeutic PMOs at room temperature. These structures were further cross-validated by their use in calculating the concentration-dependent profiles of viscosity of PMO solutions, which approximate very well the experimental viscosity measurements in the dilute and semi-dilute solution regimes. Viscosity characterization is an important physical property of these PMO oligomers to understand solution aggregation since these properties are directly relevant to the stable and injectable formulation aspects of PMO use as therapeutics and in biotechnology applications as RNA-mimic drugs.[30,31] The multiple points of agreement we have obtained between the results of experiments and simulations have enabled us to calculate the hydrodynamic properties of PMO solutions, including the intrinsic viscosity and Huggins constant,[32] and the thermodynamic state functions: the entropy, enthalpy, and free-energy changes associated with PMO folding in solution. Importantly, the trend observed for concentration-dependent viscosity validated the simulation approach as a predictive guide to therapeutic delivery by confirming the suitability of intermediate PMO concentration (50 mg/mL) for therapeutic use. The novel structural insights gained here, coupled with detailed information about the PMOs' molecular properties and thermodynamic parameters, significantly advances the current understanding of the solution structure-function relationship for PMOs. Taken together, the results obtained can be useful in the rational design of a new generation of RNA-mimic drugs.

RESULTS

22-mer, 25-mer, and 30-mer PMOs

The sequences of nucleotides forming the 22-mer, 25-mer, and 30-mer PMOs, complementary to exons 45, 53, and 51, respectively, of the dystrophin gene pre-mRNA, are displayed in Figure 1, which shows the total number of nucleobases adenine, cytosine, guanine, and thymine along with the percent of G in each sequence, which varied between 20% and 32%. All PMOs contain a polyethylene glycol at the 5′ end of the molecule. Guanine-rich sequences are known to form preferred secondary structures,[33] but since there are no more than two guanines next to each other in any of the PMO structures, those secondary structures are not expected to form in either the 22-mer, 25-mer, or 30-mer PMOs.

CD spectra

We carried out CD spectroscopic measurements on the 22-mer, 25-mer, and 30-mers to explore the structure of these PMOs in aqueous solution. Because the CD spectra obtained for the uncharged PMO molecules reconstituted and diluted in water and in Dulbecco's phosphate buffered saline (DPBS) were very similar (data not shown), in this study we describe only the CD spectra for the PMO molecules in pure aqueous solution. The higher-order structures of PMOs were assessed by CD analysis in the UV wavelength region (ca. 200–340 nm). The spectra of ellipticity Θ as a function of wavelength λ for PMOs are displayed in Figure 2. The most significant band at $\lambda \approx 270$ nm was observed in all PMO samples, and overall the intensities and exact peak and trough positions are only slightly different for individual PMOs. The addition of up to 1 M urea and 1 M LiCl did not result in any changes to these peak features (see Figure S4B). These stable CD spectral features indicate the overall stability of the PMOs' solution structures' chirality properties.

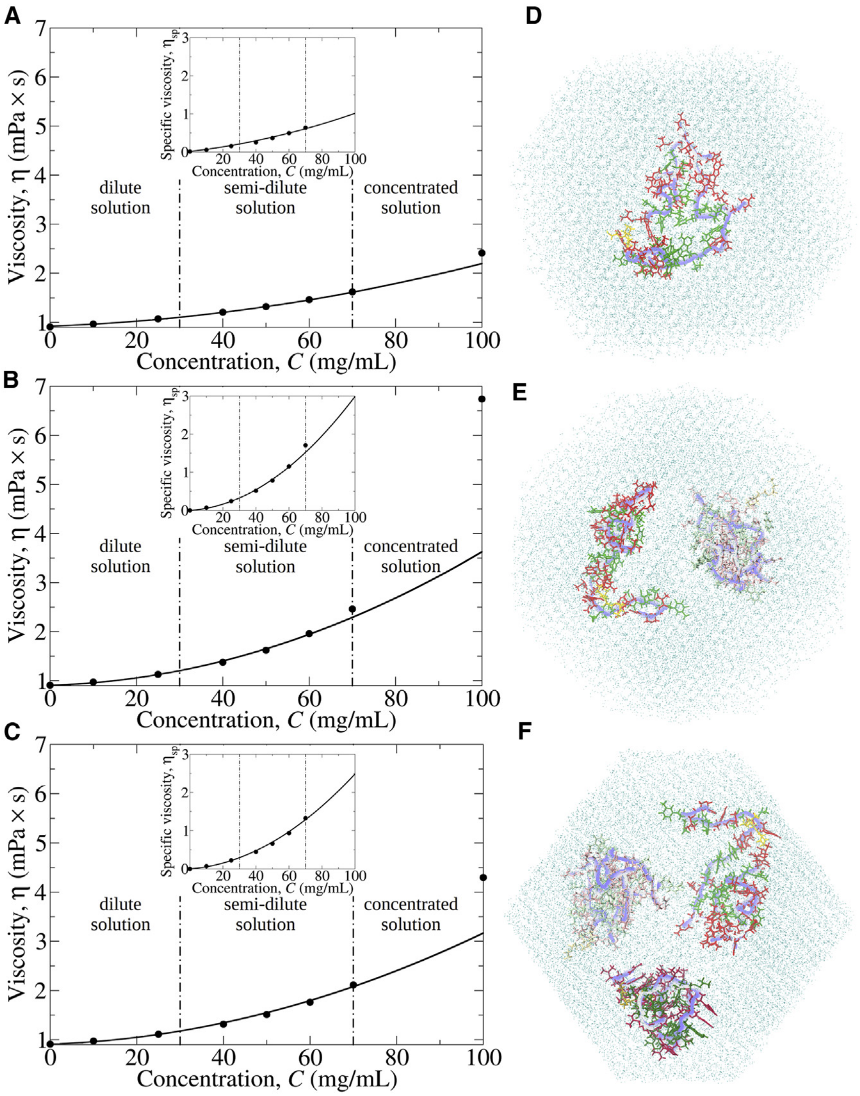

Figure 3. Theoretical reconstruction of viscosity profiles for 22-mer, 25-mer, and 30-mer PMOs

Superposed are the 25°C viscosity η versus concentration C profiles for the 22-mer (A), 25-mer (B), and 30-mer (C) PMOs obtained experimentally (black data points) and theoretically (black solid line). The insets show the profiles of specific viscosity $\eta_{sp} = \frac{\eta - \eta_0}{\eta_0}$ (η_0 is the viscosity of a pure solvent). Vertical dashed-dotted lines represent the borderline between the regions of dilute, semi-dilute, and concentrated solutions. (D–F) Representative snapshots from the all-atom MD simulations for the 25-mer PMO in the simulation box representing different solution concentrations: C = 35 g/L (D), C = 75 g/L (E), and C = 100 g/L (F). The conformers are shown in licorice representation (sticks) and in twister representation (blue line) describing the backbone. MPG linker is shown in orange, A and T (C and G) bases are shown in green (red).

Viscosity-concentration profiles

We measured solution viscosity η for the 22-mer, 25-mer, and 30-mer PMOs as a function of their solution concentration (by mass) C. The number of nucleotides in a PMO sequence was not found to affect the viscosity of these molecules significantly at or below C = 25-mg/mL concentrations. The viscosities of C = 25-mg/mL concentrations of 22-mer, 25-mer, and 30-mer at 25°C temperature were similar, being equal to η = 1.07, 1.13, and 1.11 mPa/s, respectively. However, at higher concentrations (range 35–100 mg/mL), the solution viscosity of the 22-mer was lower compared with the 25-mer and 30-mer (Figure 3). Viscosity measurements for C = 100-mg/mL PMO solutions at 25°C temperature gave η = 2.41 mPa/s, 6.74 mPa/s, and 4.30 mPa/s for the 22-mer, 25-mer, and 30-mer PMOs, respectively. These measurements revealed that the viscosity of PMO solutions is positively correlated with the concentration and somewhat with the length of PMOs. This becomes evident for PMO concentrations above C = 70 mg/mL, where the dependence of η on C becomes strongly nonlinear. This finding is indicative of intermolecular interactions (dimerization, trimerization, etc.) that occur in all three PMOs at higher solution concentrations. Similar observations for the dependence of η on C have been reported, e.g., for monoclonal antibodies, where intermolecular interactions resulting in the formation of networks of antibody molecules and higher-order structures were implicated.[34]

Dynamic structural transitions in PMO sequences

To provide a structural basis for interpretation of the experimental CD spectra, we turned to computational molecular modeling. The use of MD simulations is predicated upon the availability of an accurate force field for the molecular system in question. Although there are force fields for nucleic acids (DNA and RNA), there are no force fields available for PMOs. However, PMOs are RNA-like molecules, in which the canonical five-member ribose ring is replaced with the six-member morpholino ring and the canonical phosphate is replaced by the phosphorodiamidate linkage. Therefore, we used RNA force field bsc0$_{\chi OL3}$ with improved torsion angles[35–38] as the basis, and we extended it to account for atoms not described in bsc0$_{\chi OL3}$ using GAFF.[39] Selection of the dominant conformation of the morpholino ring, calculation of atomic charges in the morpholino ring and in the phosphorodiamidate group, development of the atomic force filed parameters for MD simulations of PMOs, and reconstruction of the initial PMO structures are all described in detail in the supplemental information.

For each PMO system, we generated 15 independent 1-μs-long MD simulation runs (total time of 15 μs) using the extended conformations as initial structures (see Figure 1). To generate different

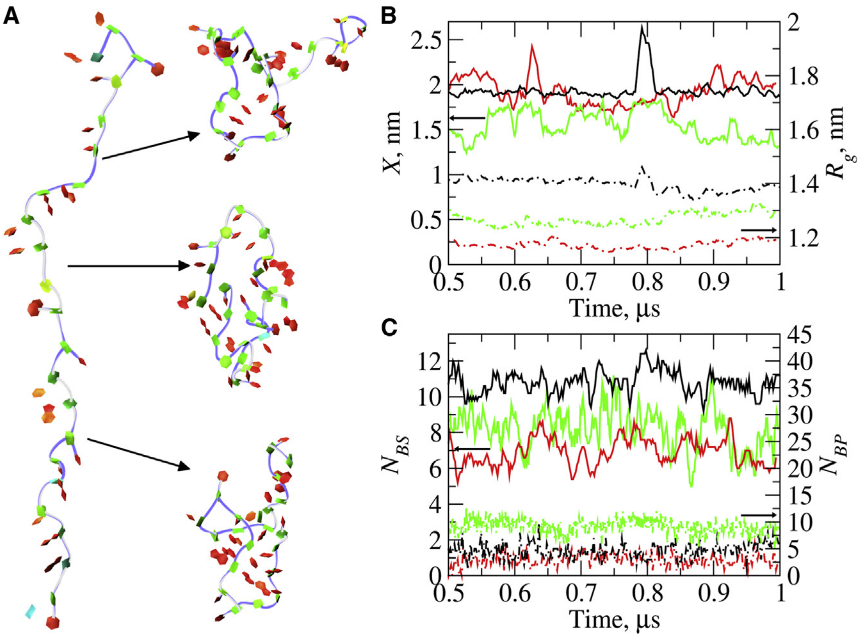

Figure 4. Conformational transitions and dynamic structural properties of therapeutic 22-mer, 25-mer, and 30-mer PMOs

(A) Extended conformation of the 22-mer PMO (see Figure 1) and three representative partially folded conformations formed in the course of three independent MD runs. The extended and partially folded conformations are displayed in the twister representation (for the backbone; blue line) and in the PaperChain representation (for nucleic bases). (B and C) Time profiles of the structural properties of the 22-mer (red curves), 25-mer (green curves), and 30-mer (black curves) PMOs from structure snapshots extracted from these simulations at every 5-ns time interval. (B) Evolution of the end-to-end distance X (solid curves; left y axis) and radius of gyration R_g (dashed curves; right y axis) from a 1-μs MD simulation run. (C) Number of base-stacking interactions N_{BS} (solid curves; left y axis) and base-pairing interactions N_{BP} (dashed curves; right y axis).

dissimilar initial extended conformations for each of the 22-mer, 25-mer, and 30-mer PMO molecules, with different orientations of bases along the backbone but with the same end-to-end distance X, we ran 30-ns MD simulations with the first and last P-atoms constrained (supplemental methods). Starting from these extended conformations, the PMO molecules formed the lower energy stable compact conformations in the first 100–200 ns of MD simulations (Figure 4A). Videos S1 and S2 exemplify, respectively, the folding transitions and conformational fluctuations for the 22-mer PMO.

We explored the dynamic structural properties of the 22-mer, 25-mer, and 30-mer PMO molecules by profiling the end-to-end distance X, the radius of gyration R_g, and the number of base pairs N_{BP} and base stacks N_{BS} (supplemental methods). While X and R_g carry information about the overall spatial distribution of atoms in PMO molecules (tertiary structure), N_{BP} and N_{BS} reflect their secondary structure characteristics. For example, for an extended conformation of the 22-mer (Figure 4A), N_{BP} might be expected to be very small ($N_{BP} \approx 0$), whereas N_{BS} could take on large values. By contrast, for the potential double-stranded helical arrangement in the 22-mer, N_{BP} and N_{BS} could be largely due to the formation of the base-pairing and base-stacking interactions. The typical time profiles of X, R_g, N_{BP}, and N_{BS} are displayed in Figures 4B and 4C, which show that, while these characteristics fluctuate about their average values, they do not significantly change with time for any PMOs modeled. The end-to-end distance varies between $X = 1.2$ nm and 2.6 nm, and the radius of gyration fluctuates around $R_g = 1.3$–1.5 nm for the 22-mer, 25-mer, and 30-mer PMOs (Figure 4B). We did not observe the reverse transition from the more compact (folded) structures to the more extended (unfolded) structures for any of the PMOs studied on the 1-μs timescale. This would have been reflected in large-amplitude changes in both X and R_g. Hence, the results obtained demonstrate that folded PMO molecules form stable structures in aqueous solution.

Time-dependent profiles of N_{BP} and N_{BS} revealed only a limited number of base pairs and base stacks formed. Indeed, N_{BP} fluctuates averaging 3.3 ± 0.9 pairs for the 22-mer, 8.9 ± 2.7 for the 25-mer, and 5.4 ± 1.7 for the 30-mer; N_{BS} also fluctuates, averaging 6.5 ± 3.6 for the 22-mer, 8.7 ± 1.5 for the 25-mer, and 11 ± 2.1 for the 30-mer (Figure 4C). This means that none of the PMOs form significant sections of a typical canonical nucleic acid duplex structure, as in the case of RNAs. This can also be gleaned from the structure snapshots of the 22-mer PMO (Figure 4A). Indeed, to form canonical RNA-like structures, the 22-mer, 25-mer, and 30-mer PMOs would have to fold and be stabilized by a majority of their bases occurring in the form of base pairs and roughly by twice as many base stacks. This is not the case for the 22-mer, 25-mer, and 30-mer PMOs. Hence, due to a lower count of base pairs and base stacks, the secondary structures of PMOs largely deviate from the canonical folded double-stranded RNA-like duplex array (Figure 4A).

Theoretical modeling of CD profiles

Next, we turned to theoretical modeling of the CD spectra using the output from the all-atom MD simulations for the 22-mer, 25-mer, and 30-mer PMOs to compare with experiments (Figure 2). To select structurally different conformations, we employed the eRMSD measure of structure similarity,[40,41] which varies between ~1.2 and ~2.0 for all PMO molecules. For each 22-mer, 25-mer, and 30-mer PMO, we selected the structures that differ in their eRMSD values by ΔeRMSD >0.003. These structures were then used to calculate the theoretical CD spectral profiles for the PMO molecules. An example of theoretical CD profiles for different conformations of the 22-mer PMO is presented in Figure S3. We found that the existence of

Table 1. Solution conformations of therapeutic 22-mer PMO for Exon45, 25-mer PMO for Exon53, and 30-mer PMO for Exon51

22-mer PMO										
	w	X, nm	R_g, nm	N_{BP}	N_{BS}	SASA, Å^2	$[\eta]$, cm^3/g	ΔG, kcal/mol	$T\Delta S$, kcal/mol	ΔH, kcal/mol
I	0.23	1.75	1.76	3	4	4,665	5.35	−58	−19	−75
II	0.22	0.81	1.31	3	8	4,368	3.88	−18	−28	−46
III	0.16	1.66	1.50	4	4	4,912	4.72	−23	−21	−44
25-mer PMO										
	w	X, nm	R_g, nm	N_{BP}	N_{BS}	SASA, Å^2	$[\eta]$, cm^3/g	ΔG, kcal/mol	$T\Delta S$, kcal/mol	ΔH, kcal/mol
I	0.19	2.1	1.25	6	12	4,803	5.8	−71	−43	−114
II	0.17	2.5	1.22	5	6	4,602	3.9	−33	−46	−79
III	0.16	3.9	1.37	6	11	5,002	4.5	−55	−32	−87
30-mer PMO										
	w	X, nm	R_g, nm	N_{BP}	N_{BS}	SASA, Å^2	$[\eta]$, cm^3/g	ΔG, kcal/mol	$T\Delta S$, kcal/mol	ΔH, kcal/mol
I	0.25	3.7	1.60	12	7	5,651	5.7	−53	−60	−113
II	0.13	1.7	1.59	5	12	5,794	5.7	−54	−58	−112
III	0.12	2.6	1.71	3	5	6,299	6.7	−51	−34	−85

Shown for each principal solution conformer I–III are the equilibrium population w, end-to-end distance X, radius of gyration R_g, number of base pairs N_{BP}, numbers of base stacks N_{BS}, SASA, intrinsic viscosity $[\eta]$, and thermodynamic state functions (changes in free energy ΔG, entropy $T\Delta S$, and enthalpy ΔH) for folding of PMOs.

multiple different 22-mer PMO conformations (see the inset in Figure 2A) gives rise to distinctly different CD profiles.

We used non-linear regression to fit the ensemble of theoretical CD spectral curves to the experimental CD spectrum (see section "materials and methods"). This helped us to resolve the most relevant solution structures and evaluate their weights in the statistical ensemble. A similar non-linear regression approach was used successfully in our previous study to model the experimental small-angle X-ray scattering (SAXS) data.[42] Briefly, for each i-th structure type, i = 1, 2, …, N – total number of structures, a CD profile $\theta_i(\lambda)$ is calculated. Next, the average theoretical profile $\Theta_{th}(\lambda)$ is constructed using a weighted superposition, $\Theta_{th}(\lambda) = \sum_i w_i \theta_i(\lambda)$ (see Equation 1), where w_i is the statistical weight (population weight) for the i-th structure type. We used the mean squared error (MSE) as a penalty function (see Equation 2) and population weights $w_1, w_2, \ldots, w_N$ ($w_{tot} = \sum_i w_i = 1$) as regression coefficients (see section "materials and methods"). For all three PMOs, the gradient descent algorithm converged after 20,000 iterations. Applying this approach has enabled us to describe solution structures for the 22-mer, 25-mer, and 30-mer PMOs, which have the largest statistical weights in the ensemble of conformations. By discarding the structures with small weights <0.05 (i.e., less than 5% of the total population w_{tot}), we identified the most important conformations, which account for ∼85%–90% (majority) of all solution conformations. The first three most important (most populated) conformations I–III for the 22-mer, 25-mer, and 30-mer PMO (displayed in Figure 2) are characterized in Table 1, which also lists the population weights corresponding to these conformations. The ensemble average theoretical CD profiles $\Theta_{th}(\lambda)$ for the 22-mer, 25-mer, and 30-mer PMOs are compared with the experimental CD spectra in Figures 2A–2C, which also show the most important solution conformation types I–III.

Next, we explored the dynamic structural properties of conformation types I–III of the 22-mer, 25-mer, and 30-mer PMOs in an aqueous solution. The end-to-end distance X varies between 0.81 and 1.75 nm for the 22-mer, between 2.1 and 3.9 nm for the 25-mer, and between 1.7 and 3.7 nm for the 30-mer PMO. The radius of gyration R_g varies between 1.31 and 1.76 nm for the 22-mer, between 1.22 and 1.37 nm for the 25-mer, and between 1.59 and 1.71 nm for the 30-mer PMO (Table 1). The number of base pairs N_{BP} varies between three and four for the 22-mer, between five and six for the 25-mer, and between three and twelve for the 30-mer PMO. Finally, the number of base stacks N_{BS} varies between four and eight for the 22-mer, between six and twelve for the 25-mer, and between five and twelve for the 30-mer PMO (Table 1). The solvent-accessible surface area (SASA), which quantitates the degree to which a molecule is exposed to water, varies between 4,368 Å^2 and 4,912 Å^2 for the 22-mer, between 4,602 Å^2 and 5,002 Å^2 for the 25-mer, and between 5,651 Å^2 and 6,299 Å^2 for the 30-mer PMO (Table 1).

Theoretical reconstruction of viscosity-concentration profiles for PMOs

Next, we modeled experimental solution viscosity data for the 22-mer, 25-mer, and 30-mer PMOs (Figure 3; profiles of specific viscosity η_{sp} are displayed in the insets). In the Einstein formula for the average solution viscosity η in the dilute regime (Equation 3), the slope for the linear dependence of the reduced viscosity η/η_0 on the solution concentration C, $\eta/\eta_0 = 1 + [\eta]C$, is given by the intrinsic viscosity $[\eta]$, which is the inverse concentration of a molecule in its pervaded volume averaged over the ensemble of molecular conformations.[32] First, we used the first 10 most populated conformations identified in the all-atom MD simulations for the 22-mer, 25-mer, and 30-mer PMOs (Table 1) and the HYDROPRO package[43] to calculate the values of intrinsic viscosity for these conformers,

Table 2. Dynamic molecular properties and thermodynamic state functions for the three therapeutic 22-mer, 25-mer, and 30-mer PMOs

Parameter	22-mer PMO	25-mer PMO	30-mer PMO
X, nm	1.6 ± 0.5	3.2 ± 1.0	2.8 ± 0.9
R_g, nm	1.48 ± 0.19	1.4 ± 0.1	1.7 ± 0.1
N_{BP}	3.3 ± 0.9	5.9 ± 1.1	6.4 ± 4.2
N_{BS}	6.5 ± 3.6	8.3 ± 3.2	8.7 ± 3.0
SASA, Å^2	4,538 ± 248	4,819 ± 158	5,945 ± 341
$[\eta]$, cm^3/g	4.5 ± 0.6	4.7 ± 0.6	6.2 ± 0.8
k_H	4.5	9.9	3.8
ΔG, kcal/mol	−34/−46 (−37 ± 23)	−51/−51 (−58 ± 20)	−50/−69 (−53 ± 33)
$T\Delta S$, kcal/mol	−23/−46 (−24 ± 7)	−38/−57 (−33 ± 14)	−53/−68 (−48 ± 20)
ΔH, kcal/mol	−57/−92 (−62 ± 38)	−89/−108 (−79 ± 60)	−103/−137 (−105 ± 60)

Shown are the following ensemble average quantities (and standard deviations) determined from the principal solution conformers for each PMO: the end-to-end distance X, radius of gyration R_g, total number of base pairs N_{BP}, total number of base stackings N_{BS}, SASA, intrinsic viscosity $[\eta]$, and Huggins constant k_H. Also shown are the changes in free energy ΔG, entropy $T\Delta S$ (at $T = 300$K), and enthalpy ΔH for ensemble average PMO solution structures. These quantities were calculated using the five or six most populated structures, which account for ~90% of the total population. ΔG, $T\Delta S$, and ΔH were calculated using the PMOs' unfolded structures and extended structures (Figure 1B; separated by a slash) as the reference states (Figure 1C). Also shown are the average values and standard deviations of ΔG, $T\Delta S$, and ΔH extracted from the histograms displayed in Figure S7 using the unfolded structures as the reference states (shown in parentheses). The Huggins constant k_H is obtained from the fit of the Einstein formula given by Equation 3 into the experimental data points.

which are shown in Table 1 for the most important conformations I–III. The intrinsic viscosity varies for different conformations ($\{[\eta]_i\}$) for the same PMO as well as for different PMOs: it varies between 3.88 and 5.35 cm^3/g for the 22-mer, between 3.9 and 5.8 cm^3/g for the 25-mer, and between 5.7 and 6.7 cm^3/g for the 30-mer PMO (Table 1). Next, we calculated the theoretical ensemble average intrinsic viscosity, $[\eta] = [\eta]_{th} = \sum_i w_i [\eta]_i$, for the 22-mer, 25-mer, and 30-mer PMOs using their population weights ($\{w_i\}$). The average values of $[\eta]$ for the 22-mer, 25-mer, and 30-mer PMOs, shown in Table 2, vary from 4.5 cm^3/g for the 22-mer to 4.7 cm^3/g for the 25-mer and to 6.2 cm^3/g for the 30-mer PMOs. We used the obtained values of $[\eta]$ for the 22-mer, 25-mer, and 30-mer PMOs to predict the dependence of η on C in the dilute solution regime; i.e., $\eta = \eta_0(1 + [\eta]C)$. In these calculations, we set the solvent (water) viscosity to $\eta_0 = 0.909$ mPa × s as estimated from the y intercepts in the experimental viscosity profiles. Figure 3 shows excellent agreement between the experimental data points and theoretical profiles of η for all PMOs studied. These results also confirm our findings regarding the ensemble of PMO structures existing in an aqueous solution and validate our computational modeling approach.

Next, we used the full Einstein equation, $\eta = \eta_0(1 + [\eta]C + k_H[\eta]^2C^2)$, to describe the experimental data in the dilute and semi-dilute solution regimes. By performing non-linear fitting to the experimental viscosity data, we estimated the Huggins constant k_H for the 22-mer, 25-mer, and 30-mer PMOs, which is equal to 4.5 for the 22-mer, 9.9 for the 25-mer, and 3.8 for the 30-mer (Table 2). The theoretical curves of η versus C deviate from the experimental data points in the concentrated solution regime for all PMOs, especially for the 25-mer and 30-mer PMOs (Figure 3). To elucidate the reason for this disagreement, we prepared three model systems for the 25-mer PMO, which correspond to 35, 75, and 100 mg/mL concentration of this PMO in aqueous solution. Structure snapshots displayed in Figure 3 show that, while for 35 mg/mL concentration there is only one 25-mer PMO molecule per water box of ~430 nm^3 volume (Figure 3D), for higher 75- and 100-mg/mL concentrations there are, respectively, two and three 25-mer PMO molecules in the water box (Figures 3E and 3F), separated by ~2.8 and ~2.0 nm. These distances are close to the 2.6-nm size of 25-mer PMO molecules, estimated to be equal to twice the average radius of gyration ($2R_g$) of this PMO (Table 2). Therefore, at a higher (>75–100 mg/mL) concentration, the 25-mer PMO molecule is expected to aggregate in an aqueous solution.

We obtained similar results for the smaller 22-mer and for the larger 30-mer PMOs. At higher 75- and 100-mg/mL concentrations, the 22-mer PMO molecules are separated by 2.3 and 1.1 nm, respectively, and the 30-mer PMO molecules are separated by 3.5 and 1.6 nm, respectively. These values are comparable with the 22-mer and 30-mer PMOs' respective average sizes of 3.0–3.4 ($2R_g$) (Table 2). Hence, at higher 75- to 100-mg/mL concentration, the 22-mer and 30-mer PMOs also tend to aggregate, forming oligomers (dimers, trimers, etc.) This concentration-dependent aggregation explains the differences between the theoretical results obtained based on the perturbative expansion of η in the powers of C truncated at the quadratic term ($\sim C^2$) (Equation 3) and the upward-deviating experimental data for the dependence of η on C for the 22-mer, 25-mer, and 30-mer PMOs (Figure 3).

Resolving thermodynamic state functions for PMOs

Next, we focused on the energetics of conformational transitions in PMOs. We used the results of MD simulations for the 22-mer, 25-mer, and 30-mer PMOs at $T = 300$ K to gather thermodynamic information underlying structure alterations in these PMO molecules (see section "materials and methods"). The simple MM/GBSA method enabled us to analyze many thousands of conformations of the 22-mer, 25-mer, and 30-mer PMOs, in order to resolve the entire probability distributions of thermodynamic state functions, ΔH, $T\Delta S$, and ΔG. For the most important folded conformations I–III and unfolded (reference) structures (Figure 1C), the values of ΔH for folding vary between −44 and −75 kcal/mol for the 22-mer, between −79 and −114 kcal/mol for the 25-mer, and between −85 and −113 kcal/mol for the 30-mer PMOs. The values of $T\Delta S$ for folding range from −19 to −28 kcal/mol for the 22-mer, from −32 to −46 kcal/mol for the 25-mer, and from −34 to −60 kcal/mol for the 30-mer PMOs; the values of ΔG for folding vary between −18 and −58 kcal/mol for the 22-mer, between −33 and −71 kcal/mol for the 25-mer, and between −51 and −54 kcal/mol for the 30-mer PMOs (Table 1).

The histogram-based estimates of the probability distributions of values of thermodynamic state functions, ΔH, $T\Delta S$, and ΔG, for the folding of the 22-mer, 25-mer, and 30-mer PMOs, are displayed in Figure S7, which shows that these quantities are broadly distributed, especially ΔH and ΔG. Indeed, values of ΔH range from -143 to -13 kcal/mol for the 22-mer, from -164 to 50 kcal/mol for the 25-mer, and from -187 to 33 kcal/mol for the 30-mer PMOs (Figure S7); values of ΔG vary between 2 and -75 kcal/mol for the 22-mer, between 3 and -80 kcal/mol for the 25-mer, and between 5 and -82 kcal/mol for the 30-mer PMOs (Figure S7). Variations in $T\Delta S$ are smaller, ranging between -38 and -12 kcal/mol for the 22-mer, between -54 and -11 kcal/mol for the 25-mer, and between -80 and -15 kcal/mol for the 30-mer PMOs (Figure S7).

Finally, we calculated the average values of ΔH, $T\Delta S$, and ΔG for PMO folding using the unfolded PMO structures as the reference state (Figure 1C) and then the extended structures as the reference states (Figure 1B). For the folded reference structures, we obtained $\Delta H = -57$, -89, and -103 kcal/mol for the 22-mer, 25-mer, and 30-mer PMOs, respectively; $T\Delta S = -23$, -38, and -53 kcal/mol for the 22-mer, 25-mer, and 30-mer PMOs, respectively; and $\Delta G = -34$, -51, and -50 kcal/mol for the 22-mer, 25-mer, and 30-mer PMOs, respectively (Table 2). For the extended reference structures, we obtained $\Delta H = -92$, -108, and -137 kcal/mol for the 22-mer, 25-mer, and 30-mer PMOs, respectively; $T\Delta S = -46$, -57, and -68 kcal/mol for the 22-mer, 25-mer, and 30-mer PMOs, respectively; and $\Delta G = -46$, -51, and -69 kcal/mol for the 22-mer, 25-mer, and 30-mer PMOs, respectively (Table 2).

DISCUSSION

PMOs play an increasingly important role in the development of ASO-based approaches to drug discovery and have led to the approval of several nucleic acid therapeutics for clinical use.[4] However, the structure-function relationships for RNA-mimicking PMOs, or other ASO-based RNA mimics, remain elusive due mainly to the lack of any information about their sequence-dependent solution structures. To the best of our knowledge, there are no X-ray crystallography data or NMR structures of PMO molecules available to date. This lack of knowledge hampers the much-needed development of PMO-based applications of rational drug design approaches.

In this study, we have overcome this limitation—the lack of PMO solution structure knowledge—by carrying out combined experimental and computational studies for a range of PMO study systems of increasing size, from the 22-mer PMO, to the 25-mer PMO, and to the 30-mer PMO (Figure 1). This is, to the best of our knowledge, the first systematic study of the structural, energetic, and hydrodynamic properties of a range of PMOs at the atomic level of detail. Computational molecular modeling continues to play an important role in the molecular-level understanding of the properties of biomolecules (DNA, RNA, and proteins) in general and RNA and their derivatives in particular.[26,44] We carried out CD spectroscopic measurements on the 22-mer, 25-mer, and 30-mer PMOs. The CD technique has been used in numerous previous studies to help elucidate the chiral/helical structures of short-chain oligonucleotides. These studies demonstrated that canonical single-stranded nucleic acid bases stack with differing levels of long-range order in aqueous solution as a result of their hydrophobic base-facing stacking interactions.[45] In addition, CD spectroscopy has contributed a great deal to our understanding of the contribution of helices, bulges, loops, and base mismatches to the overall tertiary structures observed in canonical RNAs,[45] as well as to the characteristic right-handed B-form duplex DNA structures and A-form duplex DNA and RNA structures[45] observed in neutral aqueous solutions of moderate salt concentration.

Considering that PMOs are charge neutral, it is expected that secondary structures for these molecules are primarily governed by the balance between the hydrophobic and hydrophilic interactions in the backbone and in the bases, which are different from canonical polyanionic nucleic acids. However, in the absence of any information about the secondary structure of PMOs (or any other neutral ASOs), meaningful interpretation of CD spectra for the 22-mer, 25-mer, and 30-mer PMOs remains difficult. Experimentally, we found that all three PMO systems share similar broad features, exhibiting a large positive peak at $\lambda \approx 270$ nm and a small peak at $\lambda \approx 220$ nm, with a trough at $\lambda \approx 240$ nm and negative peak at $\lambda \approx 210$ nm. In addition to shifts in the peak maxima and trough, the large positive peak at $\lambda \approx 270$ nm decreases when the 22-mer and 25-mer PMOs are compared, and increases when the 25-mer and 30-mer PMOs are compared (Figure 2). However, overall, these three similar CD spectra share the chiral features of canonical RNA oligonucleotide CD spectra that possess an A-type of base conformation with a significant degree of ordered right-handed base stacking.[46] The small peak at $\lambda \approx 220$ nm increases for the 22-mer, to the 25-mer, and to the 30-mer PMOs, but an important difference is that, for the 22-mer and 25-mer, the ellipticity Θ is negative, while for the 30-mer it is positive (Figure 2). Finally, a negative peak around $\lambda \approx 210$ nm is present in all three PMOs' spectra. This lowest-wavelength negative peak has been shown to be due to more localized electrons in stacked bases in single-stranded RNAs.[47] Further interpretation of CD spectral curves requires the examination of actual solution structures, which we carry out below.

We profiled the dependence of viscosity of the PMO-derived aqueous solutions on the PMO concentrations for the 22-mer, 25-mer, and 30-mer PMOs (Figure 3). We found that, overall, the dependence of the PMO solution viscosity η on PMO concentration C is roughly linear in the 0- to 70-mg/mL range of solution concentration, and is strongly non-linear for higher concentration (Figure 3). This higher-concentration non-linearity is undoubtedly due to formation of higher-order structures and intermolecular networks mediated by interactions between multiple PMO molecules in close proximity.[34] Few data are available for the viscosity of a range of higher RNA/DNA concentrations and their potential higher-order structures.[48] However, for PMOs in the lower-concentration linear region of the η versus C profiles, formation of ordered chiral secondary structure in PMOs in their monomeric forms is determined primarily by the interplay between the hydrophobic and hydrophilic interactions, and not by any formal charge interactions. Again, the lack of any prior solution structure information about PMOs makes it difficult to understand and unambiguously interpret the results of solution viscosity measurements. However, the trend observed for concentration-dependent viscosity confirmed the suitability of using the intermediate concentration (50 mg/mL) therapeutically for PMOs and highlighted the importance of simulations in providing a predictive guide for therapeutics delivery and formulation decisions. In the matter of therapeutics delivery, it is worth noting that the intermediate concentration (50 mg/mL) studied here is significantly diluted with 0.9% saline prior to patient administration via intravenous (i.v.) infusion; therefore, its viscosity could have no direct impact on efficacy. However, viscosity is a key parameter from the standpoint of manufacturability. The viscosity of the solution is critical for consistent performance of downstream unit operations such as filtration and filling, as well as storage and stability.

The results obtained thus far indicate that the experimental CD spectra (Figure 2) and solution viscosity versus concentration data (Figure 3) require meaningful solution structure-based PMO modeling for their interpretation. This is a formidable task due to the possible existence of an entire manifold of conformational subpopulations of any given PMO in an aqueous solution. Because there is no structural information available, we turned to computational molecular modeling to discover PMOs' solution structure populations. We carried out quantum chemistry calculations of atomic partial charges (see Figures S1 and S2) and derived the Amber force field-based molecular mechanics parameters (Tables S1 and S2). These efforts have enabled us to explore the micro-second conformational dynamics of the 22-mer, 25-mer, and 30-mer PMO molecules in an aqueous solution (Figure 4 for 22-mer PMO).

We explored the PMOs' conformers molecular characteristics (see Table 1), including metrics of the secondary structure content (N_{BP} and N_{BS}), the surface area of PMO molecular conformations accessible to water (SASA), measures of the overall distribution of PMO atoms in three-dimensional space (X and R_g), as well as hydrodynamic characteristics ($[\eta]$ and k_H). We obtained the average values of N_{BP}, N_{BS}, X, R_g, SASA, and $[\eta]$ for the 22-mer, 25-mer, and 30-mer PMOs (Table 2). We used the average intrinsic viscosity $[\eta]$ for the PMOs obtained from the modeling of the experimental CD spectra (Table 2) to predict the dependence of PMO solution viscosity η on concentration C. This can be viewed as an example of physics-based cross-validation. A very good agreement between the theoretical profiles of $\eta = \eta_0(1 + [\eta]C)$ and the experimental data points in the dilute solution regime (Figure 3) were obtained, which enabled us to calculate Huggins constants for the PMOs (Table 2). The structural characteristics (R_g, N_{BP}, N_{BS}, and SASA) and hydrodynamic quantities ($[\eta]$ and k_H) of PMOs all increase with the number of bases (system size) for the 22-mer, 25-mer, and 30-mer (Table 2). However, this increase is not monotonic: X and k_H are larger for the smaller 25-mer PMO compared with the larger 30-mer PMO (Table 2). The clearest interpretation of these property trends is that they largely depend upon the PMOs' morpholino nucleotide base composition and sequence, and not only its size. Another point worth considering is that the Huggins constant reveals something important about all of the PMOs. For the three PMOs, k_H changes from 4.5 for the 22-mer, to 9.9 for the 25-mer, and to 3.8 for the 30-mer. Elevated values of this magnitude are known to be exhibited by larger systems that can be globular but are not particularly flexible (Bovine Serum Albumin), that are more rigid and extended in character (polystyrene sulfonate), and that exhibit aggregation behavior (Folch-Pi protein, silica rods) to name a few properties (see Table 1 in Pamies et al.[49]). These k_H values for the PMOs suggest that the ensembles of conformers we have identified possess viscosity behavior that is driven by a relative lack of flexibility, being somewhat extended in shape and being prone to aggregation. This is especially true for the 25-mer with the highest value of $k_H = 9.9$. In fact, this also agrees with our observation of the 25-mer having the largest X value and especially its viscosity behavior in the concentrated solution regime, where it exhibits the largest deviation from ideal behavior and is, therefore, most prone to aggregation of all the three PMOs (Figure 3).

We attempted to carry out the direct MD simulation-based modeling of experimental CD spectra of the 22-mer, 25-mer, and 30-mer PMO molecules at elevated temperatures, but we arrived at poor fits (data not shown), which are indicative of insufficient sampling of the conformational subpopulations for these PMOs. This finding implies the existence of a rugged free-energy landscape for the folding of PMOs, in which local minima are separated by large energy barriers. Therefore, we employed temperature replica exchange MD (T-REMD; supplemental information) simulations widely used in the conformational sampling of RNA tetranucleotides, protein unfolding, and other applications.[50–53] Because of the conformational exchange, low-temperature conformations can surmount high energy barriers by switching to higher temperatures; as a result, the sampling efficiency improves. The theoretical CD profiles constructed based on the output from T-REMD simulations are compared with the experimental CD spectra for the 55°C and 85°C in Figure S4, which shows good agreement between the calculated and experimental CD spectra. For both calculated and experimental spectra, there is a decrease in the intensity of all positive and negative peaks with increasing temperature, as well as shifts in the wavelength maxima. These features resemble the behavior determined experimentally for canonical mRNA.[46] The intensity decreases can be understood in terms of temperature-dependent decreasing order in the population ensemble of conformers that would inevitably decrease the overall chirality of the conformers, irrespective of the chromophore or its molecular location. The intensity decreases, especially at the highest wavelength, appear gradual, suggesting the absence of energetic cooperativity as the conformers disorder. This is expected given the lack of any regular long-range order evident visually in the most populated conformers we identified (Figures 2 and 5). Of the wavelength shifts with increasing temperature, the easiest to understand is the $\lambda \approx 268$ nm peak at 25°C, shifting up to $\lambda \approx 273$ nm at 55°C, then to $\lambda \approx 282$ nm at 85°C. This region of the spectrum is solely determined by electronic transitions in the bases.

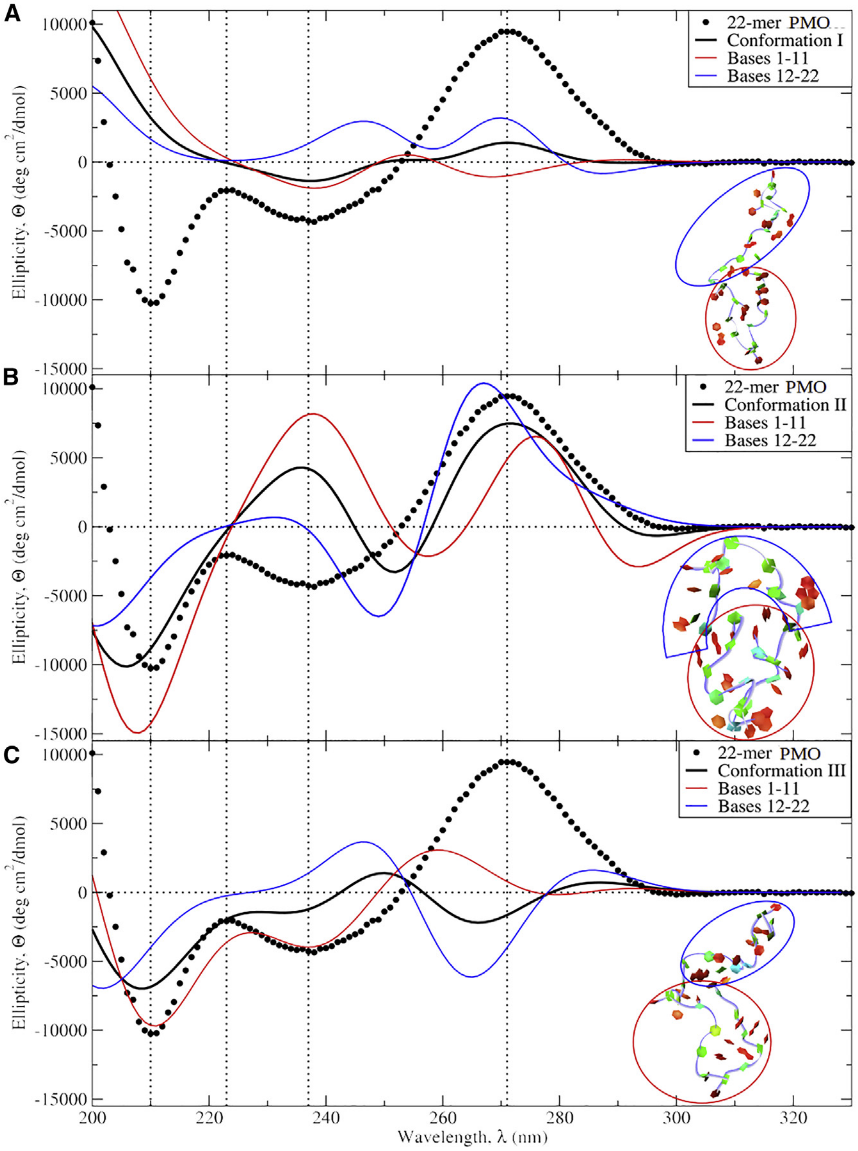

Figure 5. Interpretation of CD spectra for conformers I, II, and III for 22-mer PMO

Shown are the theoretical CD profiles (black solid lines) for conformer I (A), conformer II (B), and conformer III (C). Each of these profiles has been broken down into CD spectra calculated separately for the two halves of each conformer, shown structurally in the panels' insets, for bases 1–11 (red solid lines) and for bases 12–22 (blue solid lines). The experimental CD profile for 22-mer PMO in water is shown via black data points. The four vertical dotted lines are placed at wavelengths corresponding to peak maxima and trough minima to aid in the visual comparison of all the CD curves.

Taken together, the results obtained highlight the following important aspects of PMO-based RNA mimics: (1) the utility of the conformer ensemble perspective on folding and conformational transitions in PMOs, (2) the predominance of non-canonical single-stranded RNA-like structures of PMOs, (3) the importance of morpholino nucleotide base composition and sequence of PMOs, and (4) the thermodynamic stability of compact folded PMO structures due to a rugged free-energy landscape. In the remainder of this paper, we discuss these important aspects of PMO molecules in more detail.

PMO conformational ensemble perspective

Utilizing the ensemble average solution conformations of the 22-mer, 25-mer, and 30-mer PMOs, theoretical CD spectra were produced that correspond well to their experimental CD spectra (Figures 2A–2C and S3). We would not be able to achieve this high level of accuracy at describing the experimental CD profiles (and viscosity-concentration profiles) had the correct PMO conformations not been generated in the course of the MD simulations. Thus validated, the results of the MD simulations demonstrate the existence of rapidly interconverting subpopulations of distinctly different conformations, in which the nucleobases compete with the phosphorodiamidate groups and morpholino rings of the backbone for shielding from their exposure to water molecules. The ensemble average population of solution conformers is a weighted superposition of a large number of conformationally different subpopulations. Rather than the structural paradigm coming from protein chemistry, where a single structure (or few structures) represents the native state (folded basin), our findings illuminate the ensemble of conformers view of PMO solution structure, which has been demonstrated for canonical single-stranded RNAs in a number of cases.[54]

The individual CD spectra for the highest-weight solution structure conformers for the 22-mer PMO, contributing to the total theoretical CD spectrum that agrees with the experiment, are in fact quite distinct from one another (Figure S3). In the λ = 230- to 290-nm wavelength region, which corresponds to molar ellipticity resulting from electronic transitions for more delocalized electrons in bases, there are conformers that have completely opposite chirality behavior. For the 22-mer PMO conformers with $w = 0.22$ (green) and $w = 0.13$ (orange), their CD spectra resemble that of the right-handed helicity reminiscent of B-form DNA, while, for the conformers with $w = 0.13$ (purple) and $w = 0.16$ (magenta), their spectra resemble the left-handed base helicity (Figure S3).[27] In the shorter λ = 200- to 230-nm wavelength region, corresponding to molar ellipticities from electronic transitions for more localized electrons in the bases, a similar behavior is observed. For three 22-mer PMO conformers (in Figure S3), the molar ellipticities are quite positive, suggesting an overall right-handed helicity in the chromophore arrangement within the backbone, while, for the other three conformers, they possess a pronounced negative trough region, suggesting an overall left-handed helicity occurring in their base chromophores.[27] To add to the complexity of this picture, for the w = 0.22 conformer with right-handed helicity exhibited in the base origin higher-wavelength region, there is a left-handed base origin helicity exhibited as a negative ellipticity trough in the lower wavelength region, perhaps due to different bases in this structure having differing helical arrangements.

Adding further to the complexity of interpreting individual conformers' calculated CD spectra are the following observations. We calculated the CD spectra for fragments of conformers from the 22-mer resulting from breaking each conformer into two fragments of equal size. These are shown along with the conformer's fragments schematically in Figure 5. What is clear for the three conformers examined is the fact that, even within a given conformer, the two fragments' CD spectra look very different; this is especially true for conformers I and III. Therefore, it appears that few regular structural features occur in the ensemble solution structure populations of PMOs. One final point is worth mentioning. When the 30-mer CD was examined experimentally with solution concentrations of either urea or LiCl added separately at up to 1 M, no alterations to the CD spectra were observed (Figure S4B). Urea breaks hydrogen bonds and is widely used experimentally to totally remove all secondary structures in canonical nucleic acids, and LiCl is known to remove G-quartet types of structures and other non-Watson-Crick base pairing and encourage duplex formation instead.[55] Therefore, the lack of any spectra changes with the addition of these two solutes in the CD experiments suggests that little canonical base-paired hydrogen-bonded secondary structure or exotic G-quartet or other non-Watson-Crick-paired structures exist or can be encouraged to form in the solution PMO ensemble of conformations. This picture is in agreement with our simulation-based view of the PMO ensemble conformations: that they lack any substantial conformer stabilizing a sizable canonical nucleic acid secondary structure as well as being devoid of exotic G-quartet or other non-Watson-Crick base-pairing types of structures.

Non-canonical structure of PMO molecules

3Dynamic competition between the phosphorodiamidate groups and morpholino rings of the backbone with the bases for shielding from their exposure to solvent (water) molecules gives rise to the formation of non-canonical structures. These are quite different from the canonical arrangements of base pairs and base stacks that characterize single-stranded RNA molecules. Indeed, to form canonical RNA-like structures, the 22-mer, 25-mer, and 30-mer PMOs would have to be stabilized by a majority of its bases occurring in the form of base pairs and roughly by twice as many base stacks. This is clearly not the case for PMOs, because, in all three PMOs, the charged canonical phosphate groups (as in RNA) are replaced in the PMO backbone with the neutral phosphorodiamidate groups. As a result, polar but uncharged phosphorodiamidate groups compete with polar nucleobases for shielding from solvent exposure in PMOs, and the pattern of base-pairing and base-stacking interactions found in typical RNAs is not energetically favorable. Hence, owing to a lower count of base pairs and base stacks, the secondary structure of PMO molecules largely deviates from the canonical double-stranded RNA-like folded duplex array, thus favoring the formation of the non-canonical partially helical arrangement of morpholino nucleotides (Figure 4A).

Interestingly, we found that, as a consequence of the polar nucleobases being partially exposed to water solvent, a relatively large number of bases are capable of forming intermolecular (rather than intramolecular) base-pairing interactions. This relatively large fraction of bases observed on the periphery of the PMO's ensemble structures suggests that they are poised structurally to interact with other sequences of complementary nucleic acids, such as in pre-mRNAs in the cellular environment, where their specific interactions can be designed to result, for instance, in targeted exon skipping to produce desired therapeutic effects.[9–14] Another important consequence of the solvent exposure of polar nucleobases' hydrogen-bonding moieties is that the PMO molecules "self-interact" in concentrated solutions. Indeed, the results displayed for the 25-mer PMO indicate that at higher 75- to 100-mg/mL concentration, this PMO molecule tends to aggregate, i.e., form dimers and trimers, in aqueous solutions (Figure 3). We obtained similar results for the smaller 22-mer and for the larger 30-mer PMOs (data not shown).

Importance of morpholino nucleotide composition of PMO sequences

Given the uncharged backbone of PMOs, they are not expected to have the same interaction energetics with water that canonical nucleic acids do, where a significant fraction of bases participate in base-paired double-helical structured regions sequestered from the water environment, as well as forming loops and bulges.[48] What we have seen from the ensemble of conformers present in the oligomers' solution populations is that a sizable fraction of bases stack on the exterior of individual conformers, strongly indicating that the PMO conformers' structures result from a competition between the bases and the uncharged backbone for their being shielded from interacting with water molecules. Therefore, when we examine in detail the base pairing and base-stacking properties of the conformers, some interesting facts emerge. In Figure 6A, we present an averaged view of the different PMO conformer populations' pyrimidine/purine base-stacking picture in the different PMOs. To begin with, it is worth noting that the three PMOs have percentage-purine base compositions of 41% for the 22-mer, 40% for the 25-mer, and 53% for the 30-mer PMOs. Purines being a larger base than pyrimidines (nine- versus six-ring atoms), their stacking might be expected to produce a greater stacking enthalpy contributing to overall stability. We see that the 30-mer has a far greater number of purine-purine stacks than other stacking types compared with the 22-mer and 25-mer PMOs (∼120% more), well in excess of its somewhat larger percentage-purine fraction relative to the other two PMOs' purine-purine stacks (∼31% more).

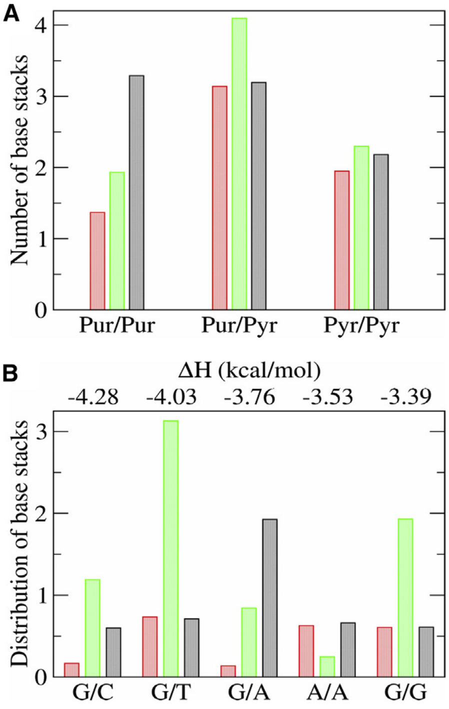

Figure 6. Distribution of base-stacking interactions in 22-mer, 25-mer, and 30-mer PMO molecules

(A) The ensemble average number of base stacks, shown for the 22-mer (red bars), 25-mer (green bars), and 30-mer PMOs (black bars), are broken down into different types of interacting bases: purine interacting with purine, pyrimidine stacked with purine, and pyrimidine stacked with a pyrimidine. (B) The ensemble average number of base stacks shown broken down into different nearest-neighbor pairs of interacting bases, including guanine (G), cytosine (C), thymine (T), and adenine (A) (bottom *x* axis), sorted in a descending order of their interaction enthalpies ΔH of single-stranded base stacks coming from the nearest-neighbor model[56] (top *x* axis).

Viewed from a somewhat different perspective, we decided to examine the three PMOs' base-stacking distributions relative to the types of possible base-stacking pairs (Figure 6B) sorted in the descending order of the enthalpies ΔH of canonical single-stranded DNA base stacks derived in Chakraborty et al.[56] using the nearest-neighbor model.[57,58] The bottom *x* axis shows only the five strongest types (of 10 total) of interacting stacked pairs arrayed from left to right and the top *x* axis shows their corresponding enthalpy values. We present these results for the three PMOs' conformer ensemble-averaged properties in Figure 6B, which shows large variations in the distributions between the three PMOs. The 22-mer has low values for most of the five interaction pairs, while the 25-mer has quite large values. In agreement with our previous purine-purine pair observation of the percentage-purine-enriched 30-mer, its G/A pair value is high compared with the other two PMOs. When we examine the relative occurrence of the strongest interacting pairs shown in Figure 6B with the weakest five pairs (data not shown), we observe that there are clear differences. We tabulated for each PMO the average number of base stacks for the five largest enthalpy interacting pairs and compared them with the lowest five enthalpy interacting pairs. For the 25-mer, there is a significantly greater population of the five largest enthalpy interacting pair types (1.14) compared with the value for the lowest enthalpy interacting pair types (0.52). Among the five largest enthalpy interacting pair types are all three purine-purine pairs (G/G, G/A, and A/A), in agreement with our discussion of the 25-mer above. For the 30-mer, the same trend is evident but not nearly to the same extent (0.91 versus 0.83, respectively), while, for the 22-mer, the trend reverses (0.45 versus 0.89, respectively). These data suggest that the PMO molecule's sequence and base composition have an effect on its ability to form base stacks and of which pair type, but not always of the maximum energetic magnitude. Therefore, we find that the competition between bases and uncharged backbone to being shielded from interaction with water molecules always forces a given conformer to adopt a compromise structure that does not necessarily maximize the possible energy derivable from base-stacking enthalpies.

These findings can also be used to explain the spectral signatures in CD spectra at higher temperature (Figure S4). It is known that the two-ring purines contain electrons that are more delocalized in lower-energy molecular orbitals than the single-ring pyrimidine bases, with electronic transitions occurring at higher wavelengths where purines have their absorption maxima.[59] As a result, their helicity or chirality properties are expected to be dominant over pyrimidines at these higher wavelengths. Therefore, the resulting shift upward in the wavelength maximum with the increasing temperature can be easily understood in terms of purines remaining preferentially in helical structures longer than pyrimidines as the temperature is increased. This is exactly what our analysis of base-stacking distributions suggests (Figure 6); namely, that the three purine-purine stacking pairs, G/G, G/A, and A/A, have the third to fifth highest out of 10 values of ΔH contributing to stability, and that the 30-mer ensemble-averaged conformer properties have the greatest fraction of their stacked bases in the purine-purine category, strongly suggesting that they will exhibit the higher-temperature CD helical properties at increasing wavelength maximum characteristic of their most stable purine-purine base stacks. This observed increase in the higher-wavelength CD maximum with increasing temperature, as we observed in Figure S4, is a common feature of mixed-sequence nucleic acids, due to the similar effect of purine-purine base stacks being preferentially more stable and retaining greater helical character at higher temperatures.[59]

Thermodynamic stability of compact folded PMO structures

The large negative values of the free energy of folding for the 22-mer, 25-mer, and 30-mer PMOs, $\Delta G = -34$ to -51 kcal/mol (Table 2; and -37 to -58 kcal/mol from the histograms in Figure S7) we have obtained demonstrate that folded PMO molecules form somewhat stable structures in aqueous solution. The probability distributions of all three thermodynamic state functions ΔH, $T\Delta S$, and ΔG for all the PMO molecules are nearly symmetric, but quite broad, especially the distributions of ΔH and ΔG (Figure S7). The standard deviations and the average values are comparable; i.e., $\Delta H = -62 \pm 38$ kcal/mol and $\Delta G = -37 \pm 23$ kcal/mol (22-mer), $\Delta H = -79 \pm 60$ kcal/mol and $\Delta G = -58 \pm 20$ kcal/mol (25-mer), and $\Delta H = -105 \pm 60$ kcal/mol and $\Delta G = -52 \pm 33$ kcal/mol (30-mer); see Table 2. These numbers point to the large variability of these energies across the ensemble of PMO solution structures and the diversity of solution conformations in terms of the energies of PMOs' intramolecular interactions and their interactions with water solvent. For all PMOs studied, the probability distributions of the PMOs' folding entropy ($T\Delta S$) are much narrower than the probability distributions of the PMOs' folding enthalpy (ΔH); see Figure S7. This means that there exist a large number of alternative base-pairing and base-stacking arrangements, which corresponds to variable enthalpy, for a given fixed conformation of a PMO's backbone. Video S3 exemplifies this interesting aspect of the conformational dynamics of PMOs in solution.

For all PMOs studied, the folding entropy $T\Delta S$ is negative; i.e., $T\Delta S = -23$ kcal/mol (for 22-mer), -38 kcal/mol (for 25-mer), and -53 kcal/mol (for 30-mer); see Table 2. This entropy decrease for a PMO molecule (i.e., PMO plus water solvent) for the folded conformations compared with the unfolded conformations implies reduced structural heterogeneity of folded PMO molecules. We also estimated the difference in solvation of PMO molecules by water in the folded state versus the unfolded state. The average area of solvent-accessible PMO surface (SASA) in the folded state is 4,538 $Å^2$ (for 22-mer), 4,819 $Å^2$ (for 25-mer), and 5,945 $Å^2$ (for 30-mer); see Table 2. The average SASA for the unfolded state comes to 5,700 $Å^2$ (for 22-mer), 6,235 $Å^2$ (for 25-mer), and 7,169 $Å^2$ (for 30-mer), and so, when PMOs fold, the area of PMO molecules exposed to solvent decreases by 1,162 $Å^2$ (for 22-mer), 1,416 $Å^2$ (for 25-mer), and 1,224 $Å^2$ (for 30-mer). Assuming that the solvation energy (enthalpy) is proportional to SASA, folding of PMO molecules in aqueous solution results in reduced solvation and lowering of solvation energies.

In summary, we have determined, for the first time, the solution conformer structures and a number of novel physical properties of the little-studied PMO molecules, for three examples that are already approved for use in the clinic as therapeutics. We have identified the existence of ensembles of stable solution conformers for the individual PMOs that are validated by being used to accurately determine their measured CD spectra and viscosity properties. The conformational dynamics we identified are defined by the competition between the non-hydrogen-bonding faces of nucleobases and the uncharged phosphorodiamidate groups for shielding from solvent exposure. As a result, PMO molecules form non-canonical, partially helical, yet stable folded structures with a small 1.4- to 1.7-nm radius of gyration, and three to six base pairs and six to nine base stacks, the latter largely confined to the molecules' exterior. The free-energy landscape for PMO folding is rugged with -31 to -42 kcal/mol free energy, -52 to -80 kcal/mol enthalpy, and -21 to -38 kcal/mol entropy of folding. The estimated 4.5- to 6.2-cm^3/g intrinsic viscosity and 4.5–9.9 Huggins constant are indicative of low-flexibility, somewhat extended, and aggregating systems. The latter property is perhaps facilitated by the exterior, solution pendant stacked bases we identified in our ensembles of solution conformers, allowing adjacent PMOs to interact easily through intermolecular base pairing and/or base stacking. The existence of these exterior-oriented stacked bases, in contrast to mostly interior-oriented bases in canonical RNAs, allows PMOs to more readily interact with their target cellular RNAs, representing a potentially unique mechanism of action for these RNA-mimic oligonucleotides. The atomic force field for PMOs developed in this work can now be used in computational studies focusing on duplex formation of PMOs with complementary RNAs, in order to explore mechanism of action for RNA-mimic oligonucleotides.

To conclude, we have gathered a sufficient corpus of knowledge to understand solution structures and behavior of PMO sequences. This can be used to formulate semi-quantitative rules that will greatly aid the drug discovery process, including sequence selection criteria (e.g., exact sequence and limit on G content) and the high-throughput screening approaches. Our findings highlight the importance of the conformational ensemble perspective of PMOs' solution structures, the thermodynamic stability of the non-canonical structures of PMOs, and the specific morpholino nucleotides composition and sequence of PMOs. The elucidation of these fundamental physicochemical principles, for the first time, now allows investigators to better understand the solution structure-properties-function relationship for PMOs to greatly facilitate the rational design of a new generation of RNA-mimic-based drugs. They also contribute important predictive capabilities to formulation and delivery considerations, such as how concentration-dependent viscosity properties affecting aggregation can be obtained through calculations.

MATERIALS AND METHODS

Synthesis of PMOs

Synthesis of PMOs comprising 22 nucleobases (22-mer), 25 nucleobases (25-mer), and 30 nucleobases (30-mer) (Figure 1) from the dystrophin gene (*DMD*) were performed using solid phase synthetic methodology described elsewhere.[5,60–62] All PMOs were characterized including liquid chromatography-mass spectrometry (LC-MS) for identity and high-pressure liquid chromatography (HPLC) for overall purity. Other novel methods for the synthesis of PMOs have been described in the literature.[63,64]

CD spectroscopy

CD studies were carried out using a Chirascan Q100 Circular Dichroism Spectrometer with Pro-Data Viewer v.4.7.0.194 data analysis software (KBI Biopharma, Louisville, CO). Initially, both cell path length and PMO concentration were evaluated to determine the optimal testing parameters. The concentration of the sample was adjusted based on Beer's law to maintain an optimal absorbance signal of ~0.8 AU. It was noted that the cell pathlength does not have a significant impact on the CD spectrum noise level, and hence a 1-cm pathlength cell was used for testing with a target concentration of 0.05 mg/mL. The raw CD spectra were buffer subtracted, baseline corrected, and normalized to the mean residue molar ellipticity. Analysis was performed at 20°C (Figure 2), except for the 30-mer, where additional data were collected at higher temperatures.

Viscosity measurements

Experimental viscosity values were assessed for the 22-mer, 25-mer, and 30-mer PMOs. Concentrated stock solutions of each PMO were formulated by dissolving lyophilized drug substance material into both DPBS and distilled water. The concentration of the stock solutions was verified using UV-visible (UV-vis) spectroscopy. A set of samples with concentrations ranging from 0 to 100 mg/mL PMO were made from the stock solutions via serial dilution. The samples were analyzed for viscosity at 25°C on a RheoSense VROC Initium Model #INI-H-1000. Analysis conditions were set to a controlled sample flow rate of 1,000 μL/min using the Initium high-pressure E02 chip.

Construction of PMO structures

Construction of the atomic models of PMOs composed of 22 nucleobases (22-mer), 25 nucleobases (25-mer), and 30 nucleobases (30-mer) was carried out using the Visual Molecular Dynamics (VMD) package.[65] The atomic coordinates of the six-member morpholino ring in the chair conformation were obtained from Caleman et al.[66] We attached guanine (G), cytosine (C), thymine (T), and adenine (A) nucleobases to the morpholino ring to obtain morpholino nucleotides; i.e., the canonical nucleotides having the five-member ribose ring substituted with the six-member morpholino ring. The connection between the morpholino ring and any of the nucleobases is through the covalent bond C1′−N1 as in canonical DNA or RNA nucleotides (Figures S1 and S2B). The linker (PMO) oligomer backbone was formed by connecting morpholino nucleotides through phosphorodiamidate groups. The morpholino polyethyleneglycol (MPG) linker was attached to the 5′ end of each sequence through the phosphorodiamidate group (Figures S1 and S2A).

MD simulations

Atomic partial charges and force field parameters were derived as described in the supplemental information. The all-atom MD simulations were carried out as described in the previous studies.[53,67] Each PMO study system (22-mer, 25-mer, and 30-mer PMOs) was solvated in an octahedron water box, containing a PMO wrapped with ∼8,000 water molecules (∼280 nm^3 volume), ∼8,600 water molecules (300 nm^3 volume), and ∼10,000 water molecules (∼360 nm^3 volume), respectively. These simulation setups correspond to ∼50 mg/mL PMO mass concentration. PMOs are uncharged oligomeric molecules, and so counterions were not included. Each PMO system was energy minimized using the steepest descent algorithm[68] over 10,000 steps, and then using the conjugate gradient method over 5,000 steps (with 50-kcal/mol energy restraint imposed on all solute atoms). Heating of each PMO study system was carried out from 0 to 300 K within 50 ps. Next, 100 ps of restrained MD simulations (with 0.05-kcal/mol energy constraint placed on all solute atoms) were performed for each PMO to equilibrate the system. Finally, 1-μs-long unrestrained production MD simulation runs in water at $T = 300$ K were carried out for each PMO study system using the CUDA version of pmemd[69] in the GPU-accelerated[68,70] AMBER 20 package.[71] The output from MD simulations was analyzed as described in supplemental methods.

Theoretical calculation of CD spectra

For each conformer and for each model system (22-mer, 25-mer, and 30-mer PMO), we calculated theoretical CD spectral profiles using the matrix method described elsewhere.[27–29] In the matrix method, one calculates the interactions between the various electronic excitations to determine the values of rotational strength. This yields a CD profile; that is, a set of values for the rotational strength for each electronic transition as a function of the wavelength λ, $\theta(\lambda)$.[72] The methodology for the theoretical calculation of a CD spectrum is implemented in the DichroCalc software.[72,73]

Resolving PMO solution conformations

The step-by-step implementation of the non-linear regression algorithm for theoretical reconstruction of the average theoretical CD spectrum Θ_{th} is the following. Step 1: assign a random weight (i.e., ensemble population) w_i to each CD profile θ_i representing the i-the conformation of a PMO in question in the ensemble of N conformations. The PMO conformations selected should have large negative value of the free energy for folding. Step 2: form a weighted superposition,

$$\Theta_{\text{th}}(\lambda) = w_1\theta_1(\lambda) + w_2\theta_2(\lambda) + \ldots + w_N\theta_N(\lambda) \tag{1}$$

in order to obtain the theoretical spectrum $\Theta_{th}(\lambda)$. Step 3: calculate the MSE,

$$MSE = \frac{1}{m}\sum_{j=1}^{m}\left[\Theta_j(\lambda_j) - \Theta_{th\,j}(\lambda_j)\right]^2 \tag{2}$$

by comparing the values of experimental CD data point Θ_j and theoretical prediction $\Theta_{th\,j}$ for all values of the wavelength λ_j, $j = 1, 2, \ldots, m$ ($m = 131$ is total number of data points). Step 4: minimize the MSE by varying the populations w_i for all $i = 1, 2, \ldots, N$ using the stochastic gradient descent algorithm. The five to seven conformations with the largest weights ≥ 0.05 were selected for further analysis and modeling.

Theoretical reconstruction of PMO solution viscosity versus concentration profiles

The solution conformations of the PMO molecule in question obtained from the numerical fit of theoretical CD curves to the experimental CD spectra were used to reconstruct solution viscosity profiles. For each i-th PMO conformation, $i = 1, 2, \ldots, n$ (n is the total number of PMO conformations from the modeling of CD spectra), the intrinsic viscosity η_i values were calculated using the hydrodynamic volume, which was estimated with the HYDROPRO package.[74] The ensemble average intrinsic viscosity $[\eta]$ was calculated as $[\eta] = \sum_{i=I}^{n} w_i\eta_i$, where the total population adds up to unity; i.e., $\sum_{i=I}^{n} w_i = 1$. The profiles of PMO solution viscosity η versus concentration C were reconstructed using the Einstein formula[75]:

$$\eta = \eta_0\left(1 + [\eta]C + k_H[\eta]^2C^2\right) \tag{3}$$

where η_0 is solvent viscosity and k_H is the Huggins coefficient k_H. The profiles of specific viscosity η_{sp} were obtained using the formula: $\eta_{sp} = \frac{\eta - \eta_0}{\eta_0}$.

Thermodynamic state functions

The Gibbs free energy (ΔG), enthalpy (ΔH), and entropy (ΔS) of PMO folding were calculated for each j-th conformer observed in equilibrium MD simulations for 22-mer, 25-mer, and 30-mer PMO, where $j = 1, 2, \ldots, N$ (N is the total number of conformations for a PMO), using the extended structures (Figure 1B) and unfolded structures of PMOs (Figure 1C) as the reference states. To calculate ΔG, ΔH, and ΔS, we utilized the molecular mechanics/generalized born surface area (MM/GBSA) method[76] implemented in the MMPBSA.py program.[77] The enthalpy of a state H is estimated using the following equation:

$$H = E_{int} + E_C + E_{vdW} + E_p + E_{np} \tag{4}$$

In Equation 4, the first three terms are standard molecular mechanics potentials, which describe the bond length potential, bond angle potential, and dihedral angle potential (included in E_{int}), electrostatic interaction potential (E_C), and van der Waals interaction potential (E_{vdW}). In Equation 4, E_p and E_{np} are the polar and non-polar contributions to the solvation free energies, respectively; E_p is obtained by using the generalized Born (GB) model, and E_{np} is estimated using the solvent-accessible surface area (SASA). Although computationally expensive approaches to the entropy calculations exist,[78] in this study, we estimated the entropy of a state S using the normal mode analysis.[79] Changes in thermodynamic state functions ΔH (enthalpy) and $T\Delta S$ (entropy) were estimated by taking the difference between the enthalpies and entropies for a reference state (extended or unfolded conformation; see Figures 1B and 1C) and the j-th conformation, i.e., $\Delta H_j = H_j - H_{ref}$ and $\Delta S_j = S_j - S_{ref}$, and ΔG_j was then calculated as $\Delta G_j = \Delta H_j - T\Delta S_j$ (for T = 300 K). The procedure for creating the reference structures for the calculation of thermodynamic state functions is described in supplemental information in more detail.

DATA AVAILABILITY

All data are available from the corresponding authors upon reasonable request and included in the main text and supplemental information.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.omtn.2023.02.007.

ACKNOWLEDGMENTS

We acknowledge helpful discussions with Willow R. DiLuzio. This work was supported by Sarepta Therapeutics, USA and in part by NSF grant MCB-2027530 (to V.B.).

AUTHOR CONTRIBUTIONS

F.M.: formal analysis, methodology, visualization, investigation, writing − original draft. E.K.: formal analysis, methodology, visualization, investigation, writing − original draft. D.P.: formal analysis, investigation, visualization, writing − original draft. M.L.U.: methodology, writing − original draft. K.A.M: conceptualization, formal analysis, investigation, methodology, supervision, validation, visualization, writing − original draft. A.C.: conceptualization, formal analysis, investigation, methodology, supervision, validation, visualization, writing − original draft. V.B.: conceptualization, formal analysis, investigation, methodology, supervision, validation, visualization, writing − original draft.

DECLARATION OF INTERESTS

D.P. and A.C. are employees of Sarepta Therapeutics Inc. and may own stock/options in the company.

REFERENCES

1. Rohs, R., West, S.M., Liu, P., and Honig, B. (2009). Nuance in the double-helix and its role in protein-DNA recognition. Curr. Opin. Struct. Biol. *19*, 171–177.
2. Rohs, R., Jin, X., West, S.M., Joshi, R., Honig, B., and Mann, R.S. (2010). Origins of specificity in protein-DNA recognition. Annu. Rev. Biochem. *79*, 233–269.
3. Laing, C., and Schlick, T. (2011). Computational approaches to RNA structure prediction, analysis, and design. Curr. Opin. Struct. Biol. *21*, 306–318.
4. Kulkarni, J.A., Witzigmann, D., Thomson, S.B., Chen, S., Leavitt, B.R., Cullis, P.R., and van der Meel, R. (2021). The current landscape of nucleic acid therapeutics. Nat. Nanotechnol. *16*, 630–643.
5. Summerton, J., and Weller, D. (1997). Morpholino antisense oligomers: design, preparation, and properties. Antisense Nucleic Acid Drug Dev. *7*, 187–195.
6. Moulton, H.M., Hase, M.C., Smith, K.M., and Iversen, P.L. (2003). HIV Tat peptide enhances cellular delivery of antisense morpholino oligomers. Antisense Nucleic Acid Drug Dev. *13*, 31–43.
7. Xiong, Y., McQuistan, T.J., Stanek, J.W., Summerton, J.E., Mata, J.E., and Squier, T.C. (2018). Detection of unique Ebola virus oligonucleotides using fluorescently-labeled phosphorodiamidate morpholino oligonucleotide probe pairs. Anal. Biochem. *557*, 84–90.
8. Nan, Y., and Zhang, Y.J. (2018). Antisense phosphorodiamidate morpholino oligomers as novel antiviral compounds. Front. Microbiol. *9*, 750.
9. Lim, K.R.Q., Maruyama, R., and Yokota, T. (2017). Eteplirsen in the treatment of Duchenne muscular dystrophy. Drug Des. Dev. Ther. *11*, 533–545.
10. Alfano, L.N., Charleston, J.S., Connolly, A.M., Cripe, L., Donoghue, C., Dracker, R., Dworzak, J., Eliopoulos, H., Frank, D.E., Lewis, S., et al. (2019). Long-term treatment with eteplirsen in nonambulatory patients with Duchenne muscular dystrophy. Medicine (Baltim.) *98*, e15858.
11. Servais, L., Mercuri, E., Straub, V., Guglieri, M., Seferian, A.M., Scoto, M., Leone, D., Koenig, E., Khan, N., Dugar, A., et al. (2022). Long-term safety and efficacy data of golodirsen in ambulatory patients with Duchenne muscular dystrophy amenable to exon 53 skipping: a first-in-human, multicenter, two-part, open-label, phase 1/2 trial. Nucleic Acid Ther. *32*, 29–39.
12. Frank, D.E., Schnell, F.J., Akana, C., El-Husayni, S.H., Desjardins, C.A., Morgan, J., Charleston, J.S., Sardone, V., Domingos, J., Dickson, G., et al. (2020). Increased dystrophin production with golodirsen in patients with Duchenne muscular dystrophy. Neurology *94*, e2270–e2282.
13. Scaglioni, D., Catapano, F., Ellis, M., Torelli, S., Chambers, D., Feng, L., Beck, M., Sewry, C., Monforte, M., Harriman, S., et al. (2021). The administration of antisense oligonucleotide golodirsen reduces pathological regeneration in patients with Duchenne muscular dystrophy. Acta Neuropathol. Commun. *9*, 7–17.
14. Wagner, K.R., Kuntz, N.L., Koenig, E., East, L., Upadhyay, S., Han, B., and Shieh, P.B. (2021). Safety, tolerability, and pharmacokinetics of casimersen in patients with D uchenne muscular dystrophy amenable to exon 45 skipping: a randomized, double-blind, placebo-controlled, dose-titration trial. Muscle Nerve *64*, 285–292.
15. Drug, F.D.A. (2016). Approval Package: Exondys 51 Injection (Eteplirsen). Published online.
16. Drug, F.D.A. (2019). Approval Package: Vyondys 53 (Golodirsen). Published online.
17. Drug, F.D.A. (2021). Approval Package: AMONDYS 45.
18. Warfield, K.L., Swenson, D.L., Olinger, G.G., Nichols, D.K., Pratt, W.D., Blouch, R., Stein, D.A., Aman, M.J., Iversen, P.L., and Bavari, S. (2006). Gene-specific

countermeasures against Ebola virus based on antisense phosphorodiamidate morpholino oligomers. PLoS Pathog. *2*, e1.

19. Iversen, P.L., Warren, T.K., Wells, J.B., Garza, N.L., Mourich, D.V., Welch, L.S., Panchal, R.G., and Bavari, S. (2012). Discovery and early development of AVI-7537 and AVI-7288 for the treatment of Ebola virus and Marburg virus infections. Viruses *4*, 2806–2830.
20. Warren, T.K., Shurtleff, A.C., and Bavari, S. (2012). Advanced morpholino oligomers: a novel approach to antiviral therapy. Antivir. Res. *94*, 80–88.
21. Howard, J.J., Sturge, C.R., Moustafa, D.A., Daly, S.M., Marshall-Batty, K.R., Felder, C.F., Zamora, D., Yabe-Gill, M., Labandeira-Rey, M., Bailey, S.M., et al. (2017). Inhibition of Pseudomonas aeruginosa by peptide-conjugated phosphorodiamidate morpholino oligomers. Antimicrob. Agents Chemother. *61*. e01938–e01916.
22. Summerton, J. (1999). Morpholino antisense oligomers: the case for an RNase H-independent structural type. Biochim. Biophys. Acta *1489*, 141–158.
23. Hudziak, R.M., Barofsky, E., Barofsky, D.F., Weller, D.L., Huang, S.B., and Weller, D.D. (1996). Resistance of morpholino phosphorodiamidate oligomers to enzymatic degradation. Antisense Nucleic Acid Drug Dev. *6*, 267–272.
24. Moulton, J.D., and Jiang, S. (2009). Gene knockdowns in adult animals: PPMOs and vivo-morpholinos. Molecules *14*, 1304–1323.
25. Hagedorn, P.H., Hansen, B.R., Koch, T., and Lindow, M. (2017). Managing the sequence-specificity of antisense oligonucleotides in drug discovery. Nucleic Acids Res. *45*, 2262–2282.
26. Golyshev, V.M., Abramova, T.V., Pyshnyi, D.V., and 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.
27. Johnson, W.C. (1996). Determination of the conformation of nucleic acids by electronic CD. In Circular Dichroism and the Conformational Analysis of Biomolecules (Springer Science & Business Media), pp. 433–468.
28. Micsonai, A., Wien, F., Kernya, L., Lee, Y.H., Goto, Y., Réfrégiers, M., and Kardos, J. (2015). Accurate secondary structure prediction and fold recognition for circular dichroism spectroscopy. Proc. Natl. Acad. Sci. USA *112*, E3095–E3103.
29. Chin, S.L., Lu, Q., Dane, E.L., Dominguez, L., McKnight, C.J., Straub, J.E., and Grinstaff, M.W. (2016). Combined molecular dynamics simulations and experimental studies of the structure and dynamics of poly-amido-saccharides. J. Am. Chem. Soc. *138*, 6532–6540.
30. Groves, M.J. (1989). Parenteral Technology Manual: An Introduction to Formulation, Production and Quality Aspects of Parenteral Products (Interpharm Press).
31. Overcashier, D.E., Chan, E.K., and Hsu, C.C. (2006). Technical considerations in the development of prefilled syringes for protein products. Am. Pharmaceut. Rev. *9*, 77–83.
32. Doi, M., and Edwards, S.F. (1988). In The Theory of Polymer Dynamics, *Vol 73* (oxford university press).
33. Sen, D., and Gilbert, W. (1988). Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. Nature *334*, 364–366.
34. Bravo-Anaya, L.M., Pignon, F., Martínez, F.A.S., and Rinaudo, M. (2016). Rheological properties of DNA molecules in solution: molecular weight and entanglement influences. Polymers *8*, 279.
35. Pérez, A., Marchán, I., Svozil, D., Sponer, J., Cheatham, T.E., III, Laughton, C.A., and Orozco, M. (2007). Refinement of the AMBER force field for nucleic acids: improving the description of α/γ conformers. Biophys. J. *92*, 3817–3829.
36. Zgarbová, M., Otyepka, M., Šponer, J., Mládek, A., Banáš, P., Cheatham, T.E., III, and 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. Theor. Comput. *7*, 2886–2902.
37. Zgarbová, M., Šponer, J., Otyepka, M., Cheatham, T.E., III, Galindo-Murillo, R., and Jurečka, P. (2015). Refinement of the sugar-phosphate backbone torsion beta for AMBER force fields improves the description of Z-and B-DNA. J. Chem. Theor. Comput. *11*, 5723–5736.
38. Cornell, W.D., Cieplak, P., Bayly, C.I., Gould, I.R., Merz, K.M., Ferguson, D.M., Spellmeyer, D.C., Fox, T., Caldwell, J.W., and Kollman, P.A. (1995). A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. J. Am. Chem. Soc. *117*, 5179–5197.
39. Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A., and Case, D.A. (2004). Development and testing of a general amber force field. J. Comput. Chem. *25*, 1157–1174.
40. Bottaro, S., Bussi, G., Pinamonti, G., Reißer, S., Boomsma, W., and Lindorff-Larsen, K. (2019). Barnaba: software for analysis of nucleic acid structures and trajectories. Rna *25*, 219–231.
41. Bottaro, S., Di Palma, F., and Bussi, G. (2014). The role of nucleobase interactions in RNA structure and dynamics. Nucleic Acids Res. *42*, 13306–13314.
42. Kumar, R., Maksudov, F., Kononova, O., Marx, K.A., Barsegov, V., and Singh, B.R. (2020). Botulinum endopeptidase: SAXS experiments and MD simulations reveal extended solution structures that account for its biochemical properties. J. Phys. Chem. B *124*, 5801–5812.
43. García De La Torre, J., Huertas, M.L., and Carrasco, B. (2000). Calculation of hydrodynamic properties of globular proteins from their atomic-level structure. Biophys. J. *78*, 719–730.
44. Šponer, J., Bussi, G., Krepl, M., Banáš, P., Bottaro, S., Cunha, R.A., Gil-Ley, A., Pinamonti, G., Poblete, S., Jurečka, P., et al. (2018). RNA structural dynamics as captured by molecular simulations: a comprehensive overview. Chem. Rev. *118*, 4177–4338.
45. Sosnick, T.R. (2001). Characterization of tertiary folding of RNA by circular dichroism and urea. Curr. Protoc. Nucleic Acid Chem. *Chapter 11*. Unit 11.5-15.
46. Kloczewiak, M., Banks, J.M., Jin, L., and Brader, M.L. (2022). A biopharmaceutical perspective on higher-order structure and thermal stability of mRNA vaccines. Mol. Pharm. *19*, 2022–2031.
47. Gray, D.M., Liu, J.J., Ratliff, R.L., and Allen, F.S. (1981). Sequence dependence of the circular dichroism of synthetic double-stranded RNAs. Biopolymers *20*, 1337–1382.
48. Koenig, V.L., Carrier, W.L., and Rahn, R.O. (1974). Viscosity studies on dna and the observation of double-stranded and single-stranded breaks in a 40% dmso-phosphate buffer system. Int. J. Biochem. *5*, 601–611.
49. Pamies, R., Hernández Cifre, J.G., del Carmen López Martínez, M., and García de la Torre, J. (2008). Determination of intrinsic viscosities of macromolecules and nanoparticles. Comparison of single-point and dilution procedures. Colloid Polym. Sci. *286*, 1223–1231.
50. Ahmad, M., Gu, W., and Helms, V. (2008). Mechanism of fast peptide recognition by SH3 domains. Angew. Chem., Int. Ed. Engl. *47*, 7626–7630.
51. Li, W., Zhang, J., Wang, J., and Wang, W. (2008). Metal-coupled folding of Cys2His2 zinc-finger. J. Am. Chem. Soc. *130*, 892–900.
52. Garcia, A.E., and Paschek, D. (2008). Simulation of the pressure and temperature folding/unfolding equilibrium of a small RNA hairpin. J. Am. Chem. Soc. *130*, 815–817.
53. Bottaro, S., Bussi, G., Kennedy, S.D., Turner, D.H., and Lindorff-Larsen, K. (2018). Conformational ensembles of RNA oligonucleotides from integrating NMR and molecular simulations. Sci. Adv. *4*, eaar8521.
54. Nagy, G., Igaev, M., Jones, N.C., Hoffmann, S.V., and Grubmüller, H. (2019). SESCA: predicting circular dichroism spectra from protein molecular structures. J. Chem. Theor. Comput. *15*, 5087–5102.
55. Kejnovská, I., Kypr, J., and Vorlícková, M. (2003). Circular dichroism spectroscopy of conformers of (guanine+ adenine) repeat strands of DNA. Chirality *15*, 584–592.
56. Chakraborty, D., Hori, N., and Thirumalai, D. (2018). Sequence-dependent three interaction site model for single-and double-stranded DNA. J. Chem. Theor. Comput. *14*, 3763–3779.
57. SantaLucia, J., Allawi, H.T., and Seneviratne, P.A. (1996). Improved nearest-neighbor parameters for predicting DNA duplex stability. Biochemistry *35*, 3555–3562.
58. SantaLucia, J., Jr., and Hicks, D. (2004). The thermodynamics of DNA structural motifs. Annu. Rev. Biophys. Biomol. Struct. *33*, 415–440.
59. Kypr, J., Kejnovská, I., Renčiuk, D., and Vorlícková, M. (2009). Circular dichroism and conformational polymorphism of DNA. Nucleic Acids Res. *37*, 1713–1725.
60. Summerton, J., and Weler, D. (1993). Uncharged Morpolino-Based Polymers Having Phosphorous Containing Chiral Intersubunit Linkages. Published online.

61. Weler, D., and Hassinger, J. (2009). Oligonucleotide Analogs Having Cationic Intersubunit Linkages. Published online.

62. Sinha, S., Kundu, J., and Ghosh, U. (2021). Synthesis of Fmoc-Protected Morpholino Monomers and Their Use in the Synthesis of Morpholino Oligomer.

63. Bhadra, J., Kundu, J., Ghosh, K.C., and Sinha, S. (2015). Synthesis of phosphorodiamidate morpholino oligonucleotides by H-phosphonate method. Tetrahedron Lett. *56*, 4565–4568.

64. Langner, H.K., Jastrzebska, K., and Caruthers, M.H. (2020). Synthesis and characterization of thiophosphoramidate morpholino oligonucleotides and chimeras. J. Am. Chem. Soc. *142*, 16240–16253.

65. Humphrey, W., Dalke, A., and Schulten, K. (1996). VMD: visual molecular dynamics. J. Mol. Graph. *14*, 33–38.

66. Caleman, C., Van Maaren, P.J., Hong, M., Hub, J.S., Costa, L.T., and Van Der Spoel, D. (2012). Force field benchmark of organic liquids: density, enthalpy of vaporization, heat capacities, surface tension, isothermal compressibility, volumetric expansion coefficient, and dielectric constant. J. Chem. Theor. Comput. *8*, 61–74.

67. Harikrishna, S., and Pradeepkumar, P.I. (2017). Probing the binding interactions between chemically modified siRNAs and human argonaute 2 using microsecond molecular dynamics simulations. J. Chem. Inf. Model. *57*, 883–896.

68. Salomon-Ferrer, R., Götz, A.W., Poole, D., Le Grand, S., and Walker, R.C. (2013). Routine microsecond molecular dynamics simulations with AMBER on GPUs. 2. Explicit solvent particle mesh Ewald. J. Chem. Theor. Comput. *9*, 3878–3888.

69. Darden, T., York, D., and Pedersen, L. (1993). Particle mesh Ewald: an N• log (N) method for Ewald sums in large systems. J. Chem. Phys. *98*, 10089–10092.

70. Le Grand, S., Götz, A.W., and Walker, R.C. (2013). Speed without compromise—a mixed precision model for GPU accelerated molecular dynamics simulations. Comput. Phys. Commun. *184*, 374–380.

71. Case, D.A., Belfon, K., Ben-Shalom, I., Brozell, S.R., Cerutti, D., Cheatham, T., Cruzeiro, V.W.D., Darden, T., Duke, R.E., Giambasu, G., et al. (2020). Amber 2020.

72. Bulheller, B.M., Rodger, A., and Hirst, J.D. (2007). Circular and linear dichroism of proteins. Phys. Chem. Chem. Phys. *9*, 2020–2035.

73. Bulheller, B.M., and Hirst, J.D. (2009). DichroCalc—circular and linear dichroism online. Bioinformatics *25*, 539–540.

74. Ortega, A., Amorós, D., and García de la Torre, J. (2011). Prediction of hydrodynamic and other solution properties of rigid proteins from atomic-and residue-level models. Biophys. J. *101*, 892–898.

75. Zuev, Y.F., Litvinov, R.I., Sitnitsky, A.E., Idiyatullin, B.Z., Bakirova, D.R., Galanakis, D.K., Zhmurov, A., Barsegov, V., and Weisel, J.W. (2017). Conformational flexibility and self-association of fibrinogen in concentrated solutions. J. Phys. Chem. B *121*, 7833–7843.

76. Hou, T., Wang, J., Li, Y., and Wang, W. (2011). Assessing the performance of the MM/PBSA and MM/GBSA methods. 1. The accuracy of binding free energy calculations based on molecular dynamics simulations. J. Chem. Inf. Model. *51*, 69–82.

77. Miller, B.R., III, McGee, T.D., Jr., Swails, J.M., Homeyer, N., Gohlke, H., and Roitberg, A.E. (2012). py: an efficient program for end-state free energy calculations. J. Chem. Theor. Comput. *8*, 3314–3321.

78. Schlitter, J. (1993). Estimation of absolute and relative entropies of macromolecules using the covariance matrix. Chem. Phys. Lett. *215*, 617–621.

79. McQuarrie, D.A. (2000). Statistical Mechanics (Sterling Publishing Company).

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$_{\chi \mathrm{OL3}}$ 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$_{\chi \mathrm{OL3}}$ 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$_{\chi \mathrm{OL3}}$, 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 $\rho_{jk} <$ 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

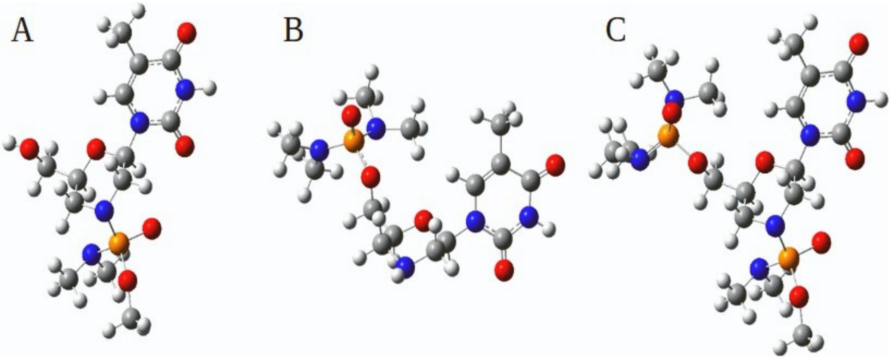

Figure S1. Structures used in RESP calculations of atomic partial charges for PMOs: Shown are the phosphorodiamidate morpholino (PM) nucleotide structures used for calculating partial charges (see Materials and Methods in the main text): PM with the phosphate group attached to the 3'-end (panel **A**), 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.

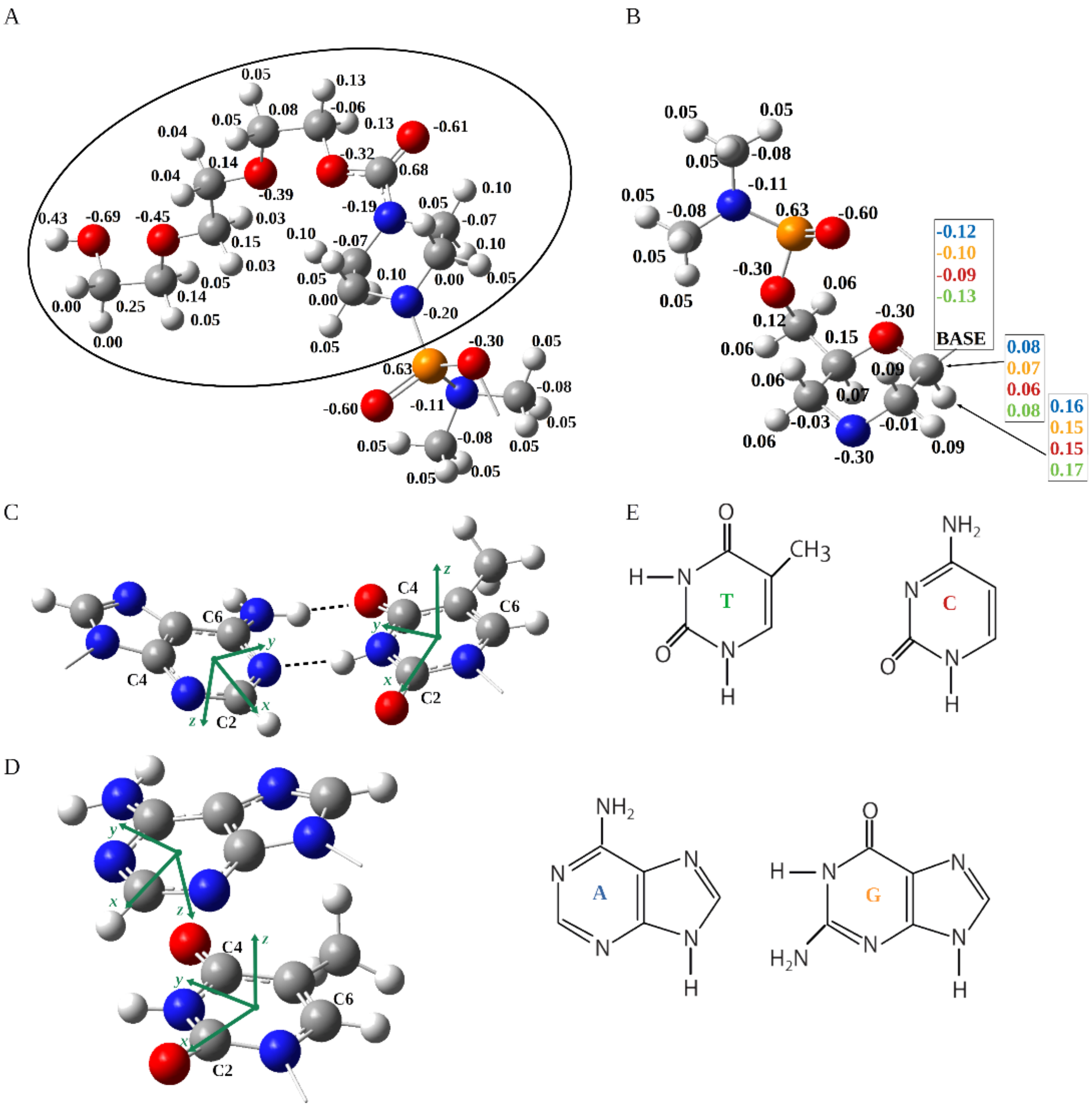

Figure S2. Atomic partial charges for PMO backbone structure: Panel **A**: Partial charges for the 5'-end (MPG) linker (circled). Panel **B**: Partial charges for the non-terminal PMO residues. The charges on the atoms of nucleic acid bases used are the same as in the bsc0$_{\chi \mathrm{OL3}}$ force 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.

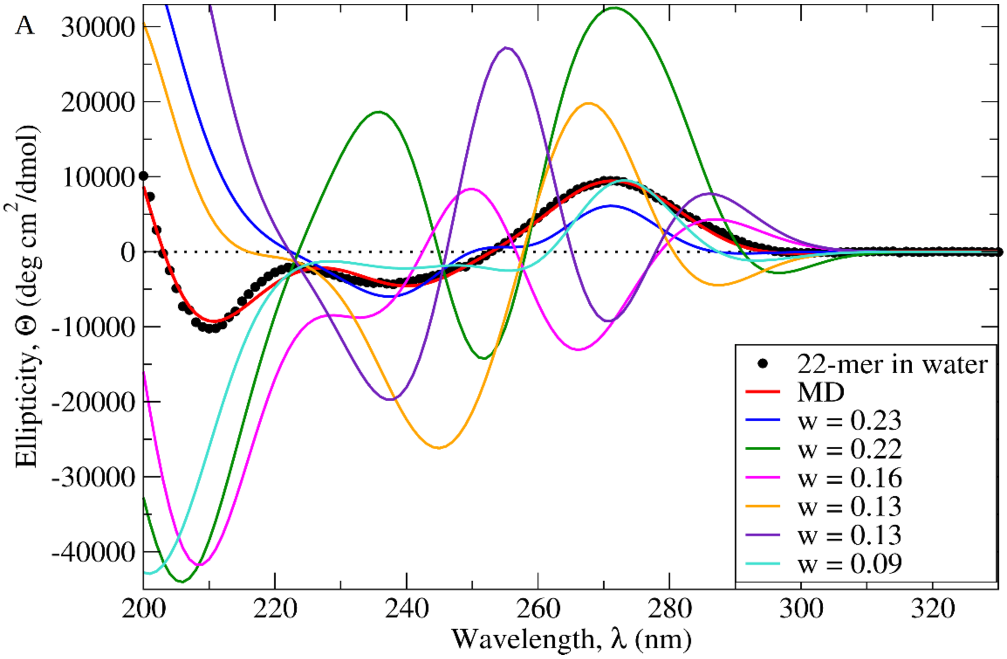

Figure S3. Circular Dichroism spectra for 22-mer PMO: Theoretical CD spectra for the six most important (most populated) conformations I-VI generated *in silico* for the 22-mer PMO complementary to Exon 45 (see Figures 2 and 4 in main text for conformers' structures) with the corresponding weights (equilibrium 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).

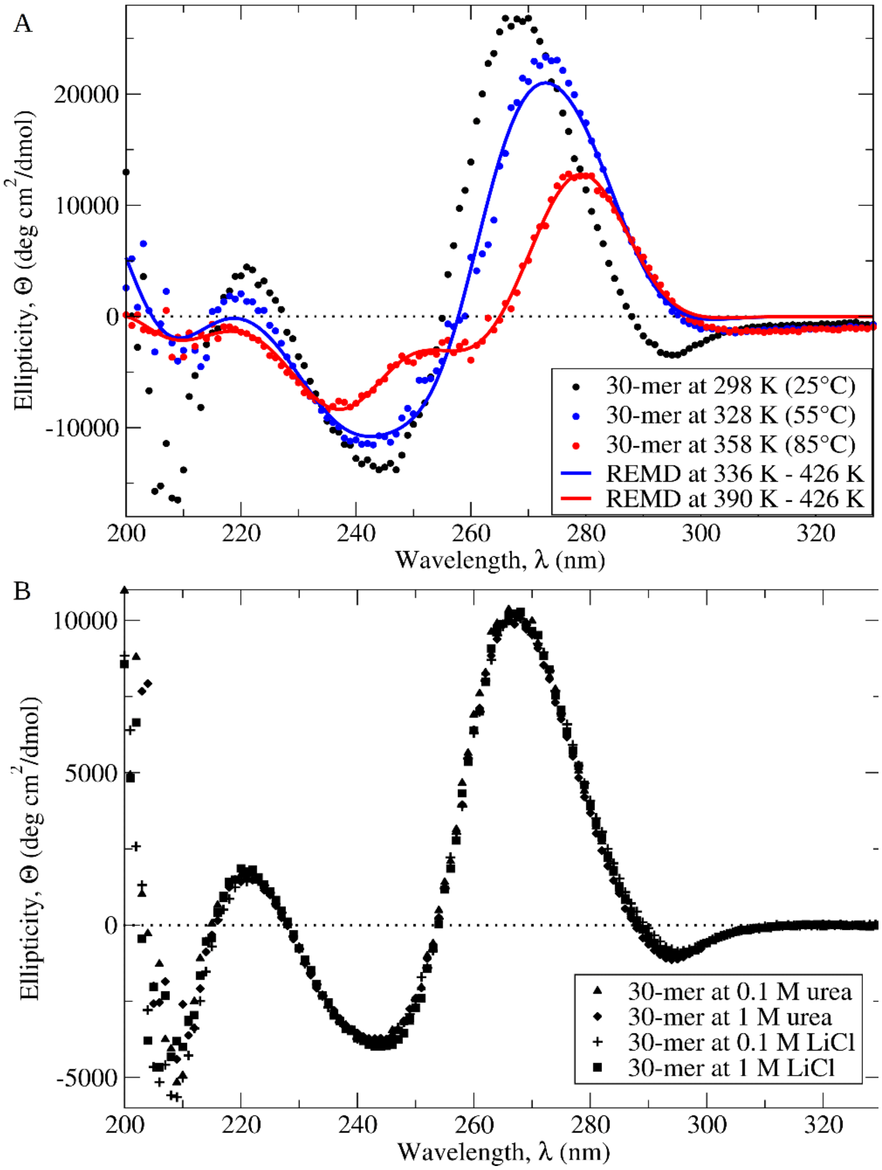

Figure S4. Circular Dichroism spectra for 30-mer PMO: Panel **A**: Experimental CD spectra for the 30-mer PMO complementary to Exon51 determined at three different temperatures 25ºC, 55ºC and 85ºC (color denotation is explained in *the inset*). Experimental data points are superposed with theoretical 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*).

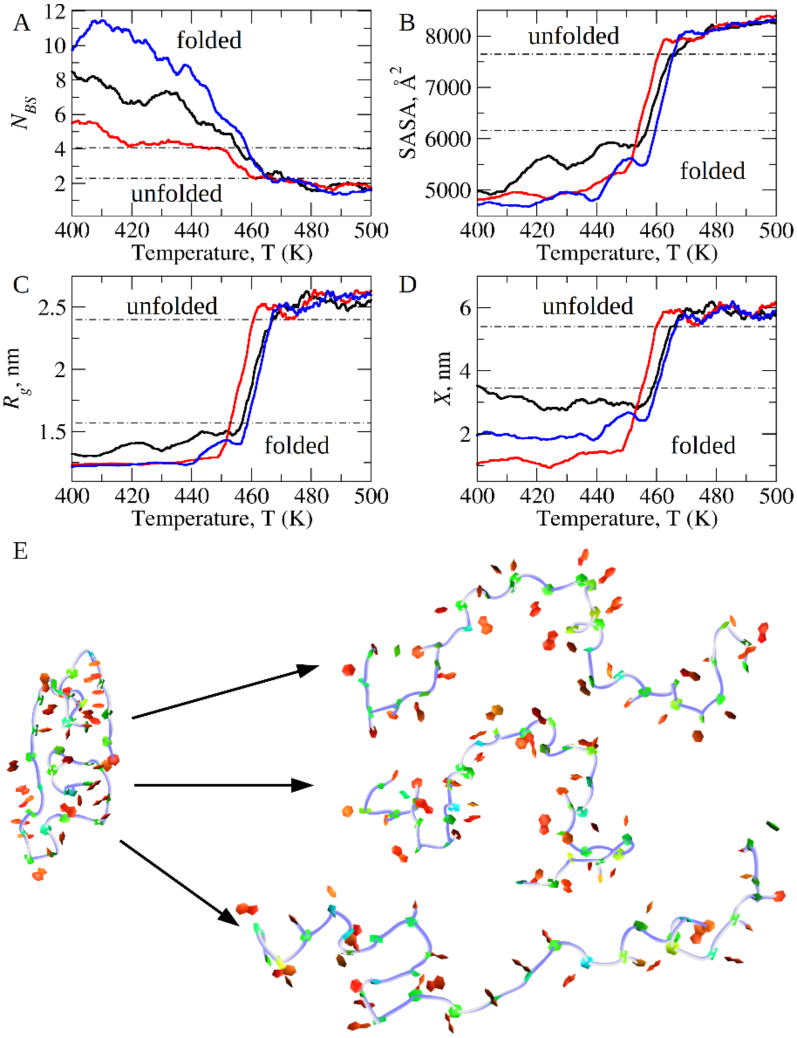

Figure S5. Dynamic structural properties and structure alterations associated with thermal unfolding of 30-mer PMO: Panels **A-D** display temperature-dependent profiles of the secondary and tertiary structural properties of the 30-mer from three independent 2.5-µs MD simulation runs of gradual heating 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).

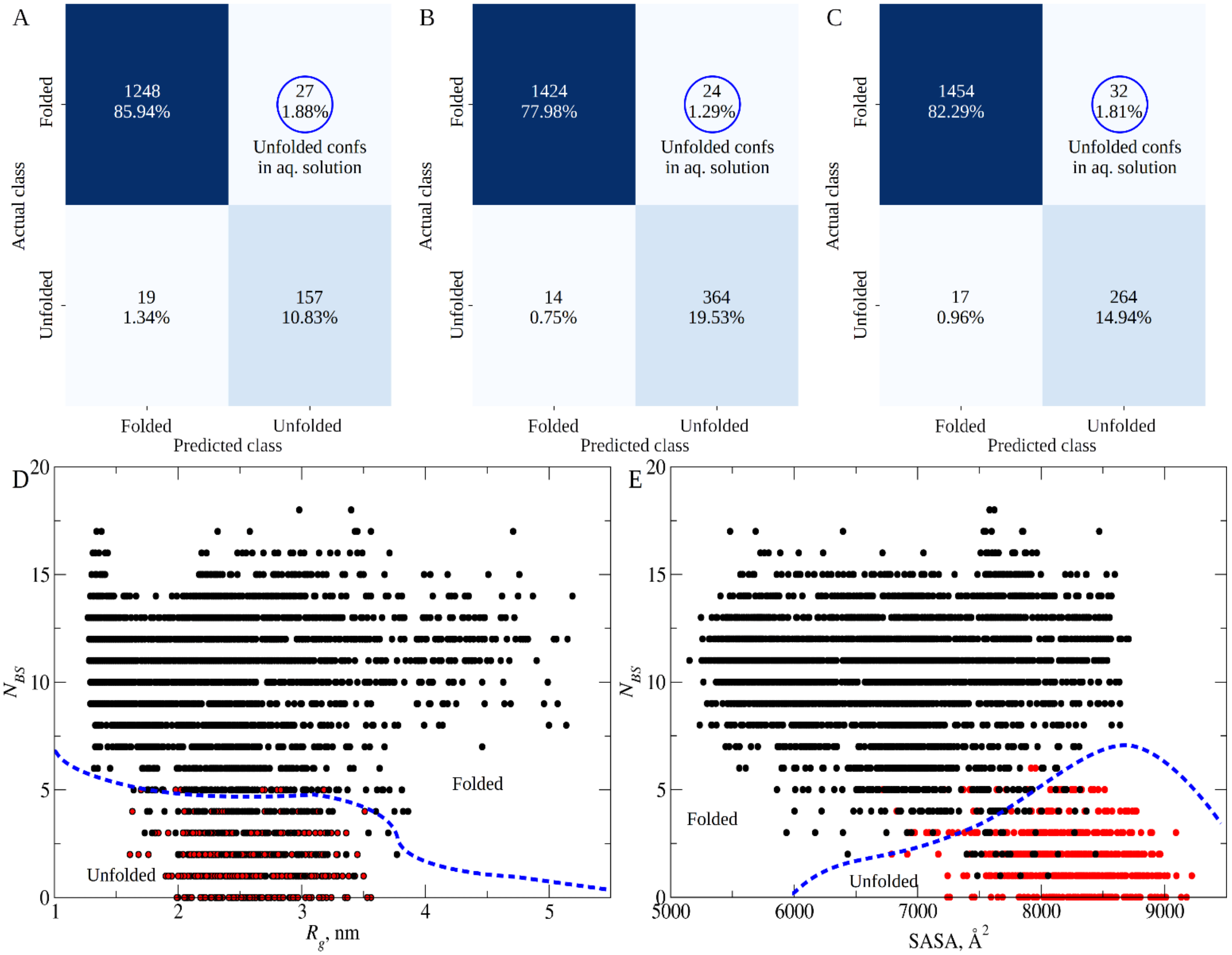

Figure S6. Support Vector Machines based analysis of conformations for 22-mer, 25-mer, and 30-mer PMOs: Confusion matrices describing the number of predictors made using the SVM technique compared to every prediction's true class for the 22-mer ($D_{test.22}$; panel **A**), 25-mer ($D_{test.25}$; panel **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).

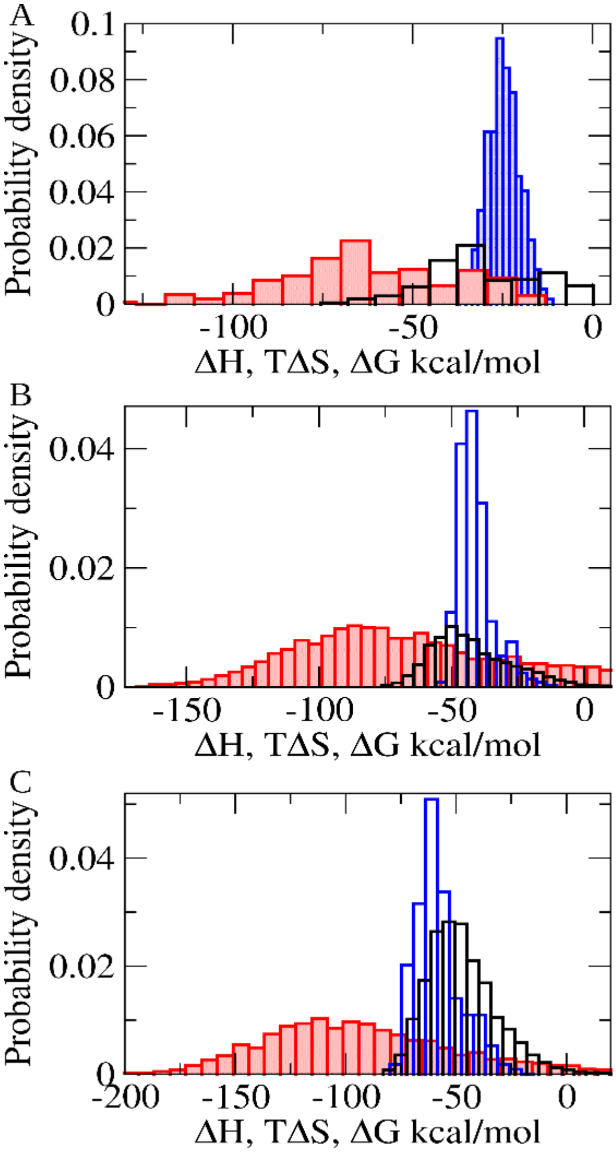

Figure S7. Histograms of thermodynamic state functions for 22-mer, 25-mer, and 30-mer PMOs: Shown are the histogram-based estimates of the probability distributions for the values of enthalpy change ΔH (red bars), entropy change $T\Delta S$ (blue bars) and free energy change ΔG (black bars) for folding 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.