JOURNAL OF CHEMICAL INFORMATION AND MODELING

ARTICLE

pubs.acs.org/jcim

Molecular Dynamics Simulation and Binding Energy Calculation for Estimation of Oligonucleotide Duplex Thermostability in RNA-Based Therapeutics

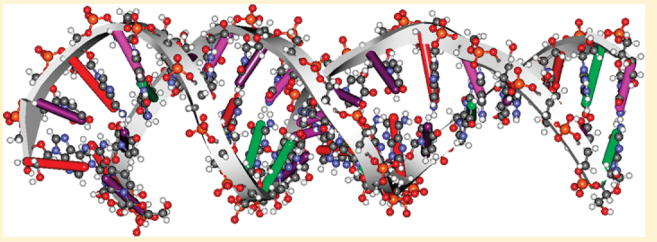
Lingling Shen,[†,§] Theresa L. Johnson,[†,¶] Susan Clugston,[†,£] Hongwei Huang,[‡] Kenneth J. Butenhof,[‡] and Robert V. Stanton*[,†]

[†]Pfizer, 620 Memorial Drive, Cambridge, Massachusetts 02139, United States

[‡]Accelrys, 200 Wheeler Road, South Tower, Second Floor, Burlington, Massachusetts 01803, United States

Supporting Information

ABSTRACT: For oligonucleotide-based therapeutics, a thorough understanding of the thermodynamic properties of duplex formation is critical to developing stable and potent drugs. For unmodified small interfering RNA (siRNA), DNA antisense oligonucleotide (AON) and locked nucleic acid (LNA), DNA/LNA modified oligonucleotides, nearest neighbor (NN) methods can be effectively used to quickly and accurately predict duplex thermodynamic properties such as melting point. Unfortunately, for chemically modified olignonucleotides, there has been no accurate prediction method available. Here we describe the potential of estimating melting temperature (T_m) for nonstandard oligonucleotides by using the correlation of the experimental T_m with the calculated duplex binding energy (BE) for oligonucleotides of a given length. This method has been automated into a standardized molecular dynamics (MD) protocol through Pipeline Pilot (PP) using the CHARMm component in Discovery Studio (DS). Results will be presented showing the correlation of the predicted data with experiment for both standard and chemically modified siRNA and AON.



INTRODUCTION

With the many recent discoveries in human genetics and mechanisms of gene regulation, the promise of harnessing these pathways for powerful therapeutics continues to grow. In particular, RNA interference (RNAi) and antisense therapies are being developed by several companies with drugs recently advancing into phase II and III clinical trials.[1] First recognized by Fire et al. in 1998, RNAi uses small interfering RNA (siRNA) through an evolutionally conserved pathway to enable knockdown (KD) of target genes through cleavage of their mRNA.[2] With the correct sequence selection, this can be done with a high level of specificity. Knockdown of the mRNA level leads to a subsequent decrease of the encoded protein, although protein KD is highly dependent on the level of message KD and the turnover rate of the protein. In contrast to double-stranded siRNAs, antisense oligonucleotides (AONs) are single strands of DNA or DNA analogues complementary to a target gene's mRNA sequence. AONs typically prevent translation through recruitment of RNaseH (an endonuclease) which then degrade the target mRNA, although other mechanisms such as steric blocking are also possible.

One challenge for identifying potential siRNA and AON therapies is selection of potent candidate sequences from the thousands of possibilities available in a normal length gene. The design is further complicated by the potential to vary sequence length and introduce chemical modifications to specific oligonucleotides leading to billions of possibilities. An essential parameter which can aid in the design of siRNA and AON compounds is the thermostability of the oligonucleotide duplex (for siRNA, the duplex thermostability referred to here is that of the two strands of the siRNA itself, while for AON, it is the compound with its complementary portion of mRNA).[3] While accurate computational methods exist for unmodified and LNA-modified oligonucleotides, this is not true for oligonucleotides containing other nonstandard chemical modifications to the oligonucleotides.[4−8] The lack of a predictive model for T_m raises challenges for oligonucleotide design, as modified oligonucleotides are increasingly being used in siRNA, AON, and immune stimulatory therapies.[9] In siRNA and AON, oligonucleotide modifications are typically intended to improve thermostability, biostability, and cellular uptake. As an example, modifications at the 2′-position of pentose sugars in siRNA help to stabilize the siRNA by increasing its persistence as compared with unmodified siRNA because of its increased serum and intracellular serum stability.[10] Also, locked nucleic acids (LNA) are conformationally restricted oligonucleotide derivatives that greatly enhance the binding affinity of DNA and RNA duplexes with higher

Received: March 24, 2011
Published: June 25, 2011

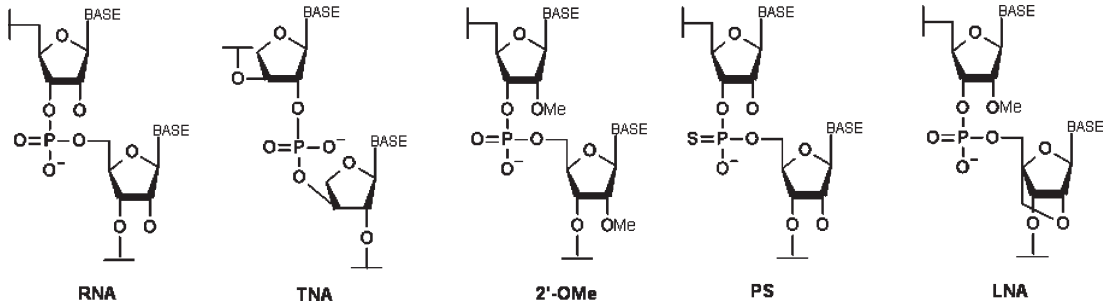


Figure 1. Chemical structures of RNA unmodified and modified oligonucleotide.

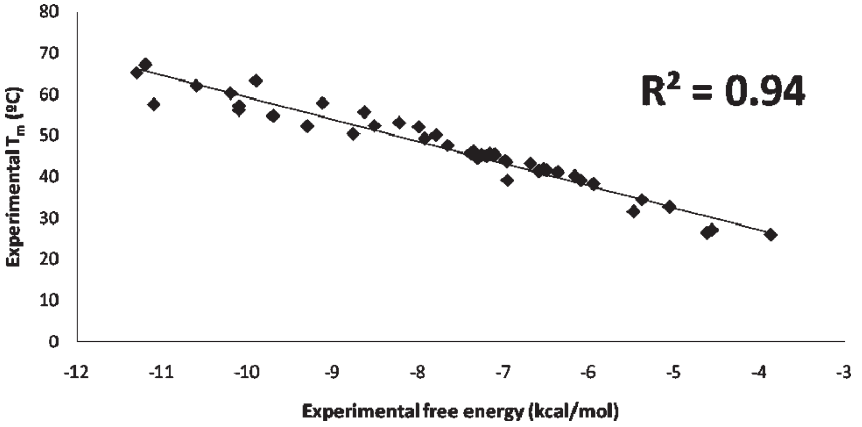


Figure 2. Correlation between experimental T_m and experimental free energy of the unmodified nucleotide data set.

duplex T_m.[11] (Figure 1 shows the structure of 2′-OMe and LNA modifications) With the success of LNA, 2′OMe, 2′MOE, and other chemical modifications, the identification of additional chemistries remains an active area of research. Computational models of nucleotides and oligonucleotides play an important role in vetting the many possibilities for chemical modifications and selecting a limited number for synthesis. The need for careful selection is enhanced by the expense of these materials, as the synthetic routes are lengthy (15−20 steps) and gram quantities may be necessary to fully test their properties.

In the remainder of this paper, we present a method to approximate the T_m for unmodified/natural and chemically modified oligonucleotide duplexes. For an unmodified (i.e., standard RNA or DNA nucleotides) oligonucleotide duplex, the NN method can be effectively utilized to predict the T_m. However, for novel oligonucleotides where the data does not exist to generate NN thermodynamic parameters, an alternative approach is necessary. Here we explore the use of the correlation between T_m and the free energy of complexation to approximate the binding energy. This correlation is demonstrated using the experimental data for 45 standard complementary RNA duplexes from a study by Freier et al.[12] in Figure 2 and 28 α-L-threofuranosyl-(3′→2′) oligonucleotides (TNA, Figure 1) from Eschenmoser et al.[13] in Figure 3. As can be seen from these plots, the correlation between T_m and complexation free energy is nearly linear with R^2 of 0.94 and 0.90, respectively. This correlation opens up the possibility of approximating T_m by calculating the duplex binding energy. An automated MD method for this calculation is presented here along with its validation against experimental data sets for modified and unmodified oligonucleotide duplexes.[14,15]

METHODS

A simple estimation of the free energy of complexation $\Delta G_{complex}$ is that it can be approximated by the binding energy.

$$BE = H_{duplex} - (H_{strand\text{-}1} + H_{strand\text{-}2})$$

where H represents the total (internal) energy from molecular mechanics for the individual components subtracted from the internal energy of the system in the complex. This approach completely ignores the contribution of entropy as well as solvation. Corrections for entropic effects can be incorporated using a wide range of sampling methods.[16,17] These methods have most commonly been applied to protein−ligand complex formation and might be equally applicable to the oligonucleotide system with sense−antisense chain interactions. However, unlike protein−ligand interactions, the change in the distribution of conformational states due to binding is expected to be significant for both chains of the oligonucleotide. In the current system under consideration, ignoring differences in conformations, there are three distinct solvated species, the antisense strand (A), the sense strand (S), and the duplex (AS). The interaction of these species can be illustrated by the following two steps:

$$A^* + S^* \Leftrightarrow A^*S^* \ (\Delta G_{bind} \text{ for static optimal conformers})$$

where A* and S* denote the conformation required for optimal binding. The Gibbs energy of binding can thus be split into a conformational term ΔG_{conf} and a static term ΔG_{stat}

$$\Delta G_{bind} = \Delta G_{conf} + \Delta G_{stat}$$

dx.doi.org/10.1021/ci200141j |J. Chem. Inf. Model. 2011, 51, 1957–1965

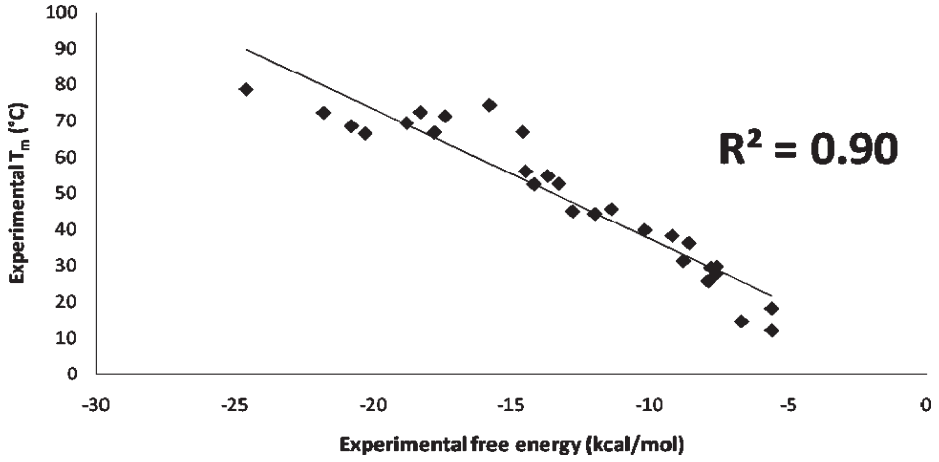


Figure 3. Correlation between experimental T_m and experimental free energy of the TNA modified nucleotide data set.

which can be further expanded to

$$\Delta G_{bind} = \Delta G_{conf} + \Delta W + \Delta G_{int} - T\Delta S$$

where ΔW represents the change in the potential of mean force for A* and S* due to solute−solvent interactions, ΔG_{int} is the interaction free energy between the oligonucleotide strands in the duplex, and $T\Delta S$ is the change in nonelectrostatic entropy determined at temperature T.[18,19]

This approach was judged to be computationally prohibitive for use in a rapid method to predict T_m for multiple oligonucleotides because of the reliance on an accurate computational estimation of entropy. Therefore, we make the approximation of estimating the binding energy by using the free energy. We seek to estimate the average binding energy from the relation

$$BE_{averaged} = \langle G_{AS} - (G_A + G_S) \rangle$$

where the three free energies G_{AS}, G_A, and G_S are each calculated for conformations sampled from a 100 ps window at the end of a molecular dynamics trajectory that has reached equilibrium. In this approach we forego a computational estimation of the effects of entropy which may become significant when predicting T_m for different chain lengths. However, for duplexes of identical (or similar) length, it is reasonable to assume that entropic effects do not vary substantially with sequence and thus will not adversely affect the correlation between average BE and T_m.

To summarize, the basic strategy of this method is to generate an initial 3D conformation from the sequence, use molecular dynamics to generate an equilibrated dynamic system from which a sample of equilibrated duplex conformers can be obtained, and calculate the duplex average BE using stable conformers.

Duplex Modeling and Simulation Setup. After the 3D A-type helical structures were created from the sequences using nucleotide builder in Discovery Studio, each duplex was immersed in a box containing pre-equilibrated TIP3P water molecules, with each side of the box extending at least 10 Å away from any solute atom.[20] All water molecules within a distance of 2 Å of any solute atom were removed, and sodium chloride molecules were randomly added to the water box to reach a final concentration of 0.15 M.

For modified oligonucleotides, the necessary bond, angle, and dihedral parameters were derived from standard DNA and RNA charmm27 parameters. The details of modified charmm27 force field for use with the chemically modified oligonucleotides in our research are as follows: For 2′-OMe-modified oligonucleotide, the parameters required for naturally occurring RNA residue modifications were supplemented with the results of Raviprasad Aduri et al.[21,22,24] In standard charmm27 RNA, the 2′ charge group (O2′-H) is neutral, and this was maintained so that the new O2′-CH_3 group was balanced to have an overall charge of zero. The force field parameters were well parametrized in charmm27 except for those parameters that were involved in the modified O2′ atom. The parameters for these interactions were adapted from the PARAM99 Amber force field used by Raviprasad Aduri. The 2′-OMe parameters used to supplement the charmm27 force field are listed in the Supporting Information, while for LNA, parameters were adopted from Pande and Nilsson.[23,24]

Molecular Dynamics Simulation Protocol. As part of the design of T_m prediction protocol, we were interested in studying the relation between the binding energy (as defined previously) and the T_m. Therefore, all simulations in this study were run with weak harmonic distance restraints (force constant of 5 kcal/mol/Å^2) imposed on the Watson−Crick hydrogen bonds present in the 5′ and 3′ termini base pairs. The system was subjected to equilibration involving solvent minimization using 5000 steps of steepest descent and 5000 steps of conjugate gradient while holding the solute phosphate atoms restrained to their initial positions by means of a harmonic force constant of 10 kcal/mol/Å^2. Then, while maintaining the positional restraints on the solute, the solvent was heated from 50 to 300 K in a 20 ps MD phase using a constant pressure of 1 atm. The solvated system was then equilibrated for 10 ps followed by an unrestrained (except for the hydrogen bonding restraints on the terminal base pairs) simulation at a constant temperature of 300 K and a constant pressure of 1 atm for 500 ps. Based on longer MD trajectory analysis of each modified and unmodified sample oligonucleotides used in the study, stable conformations are reproducibly found in the unconstrained 400−500 ps production range (Figure 4). Therefore, we used last 100 ps of the production simulation (time range of 400−500 ps) to calculate the BE of the two strands by subtracting the total energy of the sum of the two strands from that of the duplex. This was done using the Generalized Born with simple Switching (GBSW) implicit solvation method in all three systems (i.e., each single strand and the duplex). Finally the average BE was calculated over these 400 frames. The entire procedure has been automated into DS, and the BE calculation PP protocol was adapted for use with oligonucleotides (Figure 5).

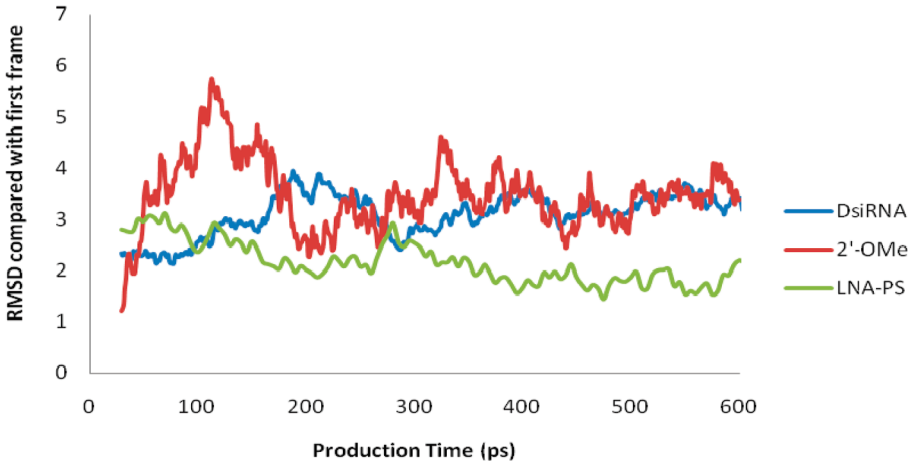


Figure 4. Trajectory analysis for select oligonucleotide duplexes.

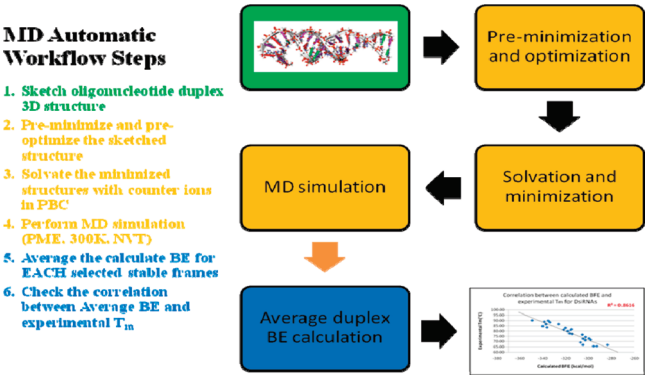


Figure 5. Automated MD workflow is oligonucleotide duplex thermostability estimation.

All molecular dynamics simulations were run using CHARMm through the use of DS components.[25,26] SHAKE constraints were applied on all covalent bonds involving hydrogen atoms, allowing an iteration time step of 1 fs.[27] The leapfrog algorithm was used to integrate the equations of motion using periodic boundary conditions on an orthrombic box of TIP3P water. Constant pressure was maintained using the Langevin piston method with a piston of mass 400 amu and a collision frequency of 200 ps^{-1}, coupled to a temperature bath of 300 K. The system was maintained at 300 ± 10 K, once heated, by scaling the velocities accordingly. Electrostatic interactions were evaluated using particle mesh ewald (PME). A constant dielectric constant of 1 was used for the oligomer, and a dielectric of 80 was used for the implicit solvent as recommended for GBSW and the charmm27 force field. The atom-based nonbonded interactions were truncated beyond 12 Å using a force shift approach. The nonbonded lists were maintained for pairs within a distance of 14 Å and updated heuristically whenever an atom had moved more than 1 Å since the last update. Coordinates were saved every 1 ps for further analysis except for the final 400−500 ps of simulation where coordinates were saved every 0.25 ps. The methodology for its computational procedure design and workflow integration and customization is elaborated in Supporting Information.[28,29] The approximate run time is 15 h per duplex calculation using four standard CPUs.

Table 1. Detailed Information of Olignucleotide Duplexes Which Have 3D structure Available in the PDB Databank and Reported Experimental T_m

PDB ID	length	modification	T_m(°C)
2G92	12	ribo-difluorotoluyl nucleotide	53.2
157D	12	non-Watson−Crick base pair	58.2
259D	8	poly CG	74.1
1PBM	6	2′-OMe	43.0
1AL5	12	none	59.5
1RNA	14	poly UA	43.0
1YYK	12	from ribonuclease structure	65.0

Spectrophotometric T_m Measurements. The T_ms of the oligonucleotides not reported elsewhere were measured using a UV spectrophotometric method. They were performed in dual single beam mode on a Cary100 Spectrophotometer equipped with a multicell changer and Peltier temperature controller. Each sample was diluted to 5 μM in phosphate-buffered saline (PBS, Ambion) pH 7.4 with 0.1 mM EDTA (Ambion). A 140 μL amount of sample was placed in each cuvette (4 × 2 × 10 mm, Varian 66-100149-00). A reference cuvette with 400 μL buffer and the temperature probe were used to monitor and record the temperature of the solution during the run. The absorbance was monitored at 260 nm as the temperature of the sample was increased from 25 to 95 °C and then as it was decreased from 95 to 25 °C, with a 3 min incubation at the end of each ramp. An interval of 0.25 °C and ramp rate of 0.5 °C was used. This was repeated for a total of four temperature ramps. The T_m was determined by performing a first derivative of the raw data with an interval of 0.3 and filter size of 5, and the peak was identified in the Cary Win UV Thermal Application software. The peak from the first derivative corresponded to the temperature at which half the duplex was separated (the T_m). The results of the four T_m measurements were then averaged.

RESULTS AND DISCUSSION

Initially, a set of duplexes with known structures from the PDB was selected to demonstrate the correlation of melting temperature with binding energy. X-ray and NMR structures were used to eliminate the error associated with generating a three-dimensional (3D)

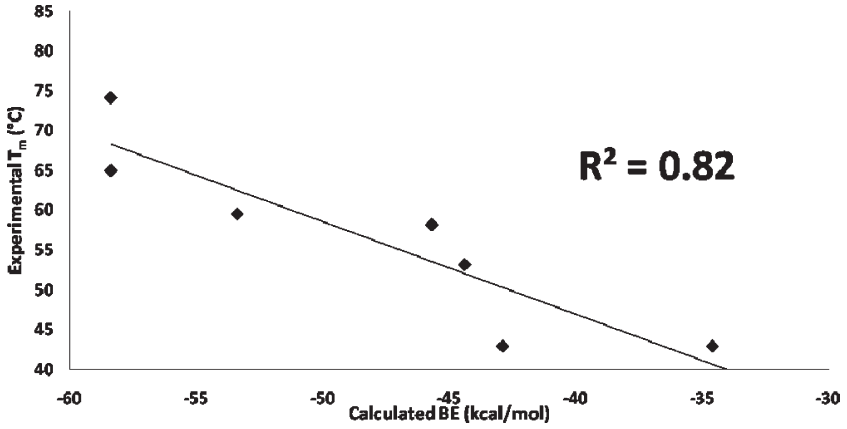


Figure 6. Least squares regression correlation between experimental T_m and calculated BE of the PDB data set.

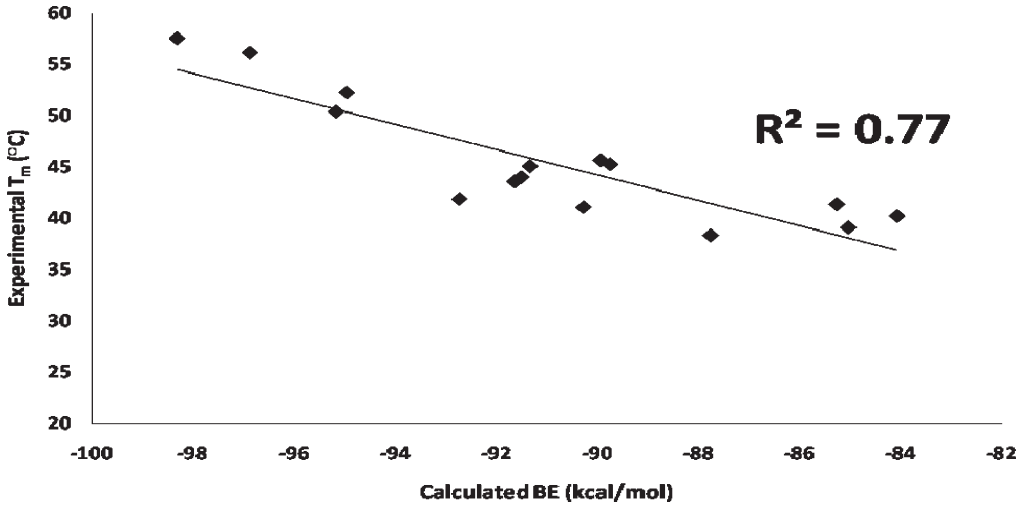


Figure 7. Least squares regression correlation between experimental T_m and calculated BE of the data set from the 8mer subset used originally in the nearest-neighbor model.

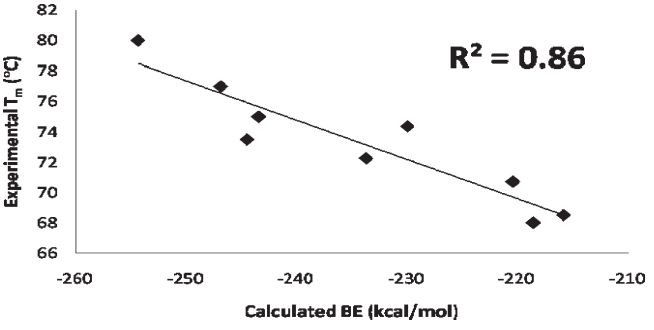


Figure 8. Least squares regression correlation between experimental T_m and calculated BE of the data set from the in-house siRNA 21mers.

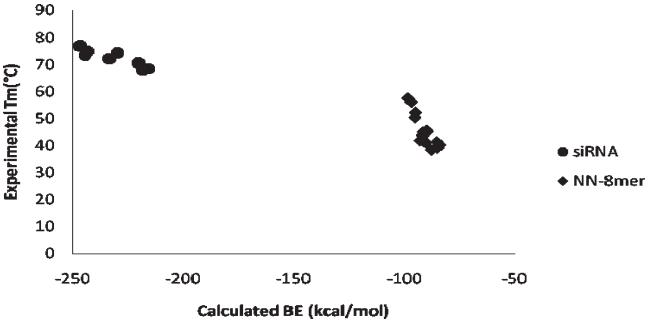


Figure 9. The correlation between experimental T_m and calculated BE from the 8mer subset used originally in the nearest-neighbor model and in-house siRNA 21mers.

conformation. Seven PDB structures of RNA duplexes (Table 1) were identified with a range of lengths (6−12 nucleotides) and chemical modifications (1PBM is completely 2′-OMe modified and 2G92 has a 2′,4-difluorotoluyl ribonuclease at the 4 position from the 5′ end of one of the strands).[30−37] The PDB duplexes were minimized (no MD run), and their binding energies were calculated from this single structure. As shown in Figure 6, it is reassuring that a solid correlation with an R^2 of 0.82 is achieved for this fairly diverse set of seven structures. A second data set of 15 RNA 8-mer duplexes from Freier et al. and 9 in-house siRNA was used to investigate errors that might be introduced through the creation of the 3D conformation and use of the binding energy determined from the MD ensemble (sequences for the

dx.doi.org/10.1021/ci200141j |J. Chem. Inf. Model. 2011, 51, 1957–1965

compounds are included in the Supporting Information). The compounds were all run through the MD protocol described earlier with the BE calculated from the ensemble of the last 400 structures saved over 100 ps. As shown in Figures 7 and 8, the BE to T_m correlation gives an R^2 of 0.77 for the 8-mer data and 0.86 for the 21-mers; this demonstrates that the nontrivial task of generating the 3D structures from sequence can be done systematically well enough to support these calculations. Figure 9 demonstrates the strong correlation between oligo length and T_m (or BE). This length dependence will need to be taken into account when trying to predict the T_m of de novo sequences due to the large BE differences for sequences of different lengths. It is important to note that within a simulation, the variability of the BE is relatively small across conformations with a STDEV of ~10 kcal/mol.

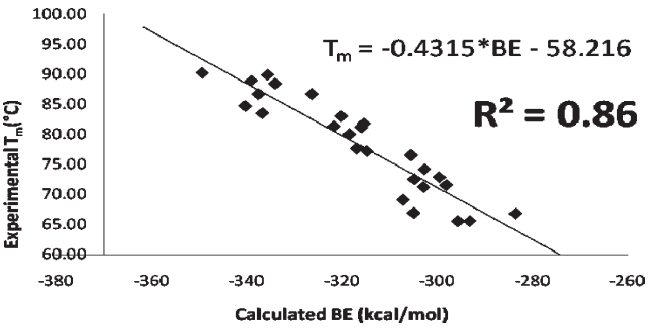


Figure 10. Least squares regression correlation between experimental T_m and calculated BE of the in-house GL2 firefly luciferase DsiRNA data set.

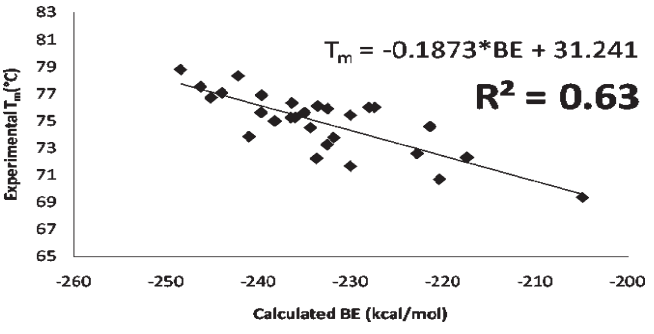


Figure 11. Least squares regression correlation between experimental T_m and calculated BE of the in-house GL2 firefly luciferase 2′-OMe-modified siRNA data set.

Next we investigated the effects of DNA and chemically modified nucleotide insertions on the overall BE to T_m correlation. Several data sets were examined including the following: (1) a set of 26 dicer substrates (DsiRNA) (27 sense strand RNA with a combination of 23 RNA and 2 DNA in the antisense strand), (2) 29 2′-OMe-modified siRNA (21 nucleotides in a standard 19 complementary bases with 2 nucleotide overhangs on each end, (3) 21 LNA-modified 16-mers with phosphothioate[24] linkages, (4) a set of 4 21mer siRNA (as in the 2′-OMe data set) but modified with an internal Pfizer proprietary nucleoside.

The DsiRNA data set was included to examine the effect of length on the T_m correlation. Dicer substrates are speculated to have improved potency as triggers of RNA interference, particularly when used at low dose, due to effects relating to linkage of Dicer processing and RISC loading.[38] The 2′-OMe set of compounds allow the examination of effects caused by both modification position and number of modifications. Due to their increased serum and intracellular serum stability, methoxy modifications to the 2′-OH group of the ribose ring (2′-OMe) have been intensively researched as possible candidates to inhibit gene expression. The Pfizer nucleotide was included to demonstrate errors which might be introduced into the calculation because of the need for force field parametrization.

The LNA-PS AON data set was taken from a study of Frieden et al. in which they were looking at the effect of gap size on the antisense activity of LNA gapmers.[39] Antisense gapmers are

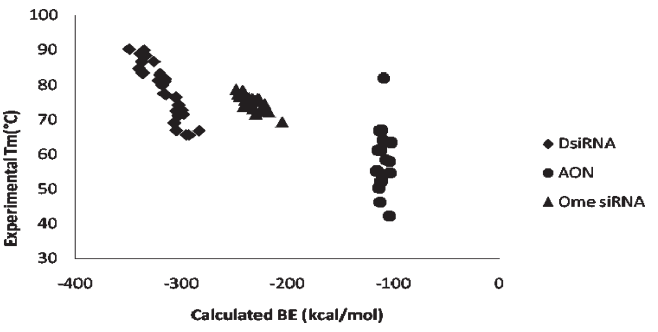


Figure 13. The correlation between experimental T_m and calculated BE from the in-house GL2 firefly luciferase DsiRNA data set, 2′-OMe-modified siRNA data set, and the literature LNA-PS-modified AON data set.

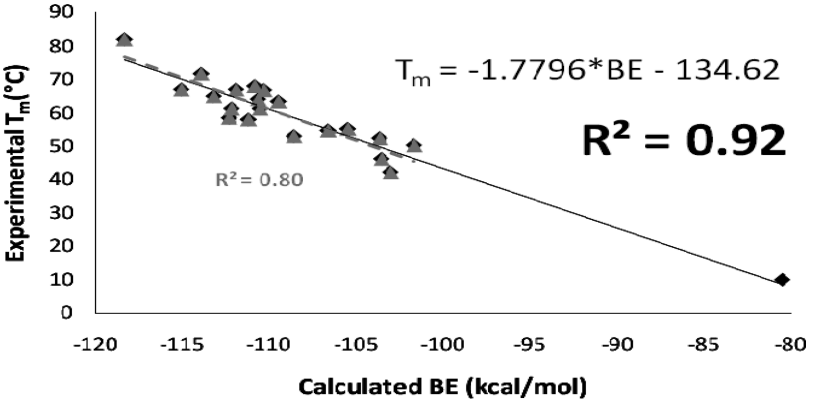


Figure 12. Least squares regression correlation between experimental T_m and calculated BE of the in-house GL2 firefly luciferase LNA-PS-modified AON data set. (the gray triangle data points given R^2 =0.80).

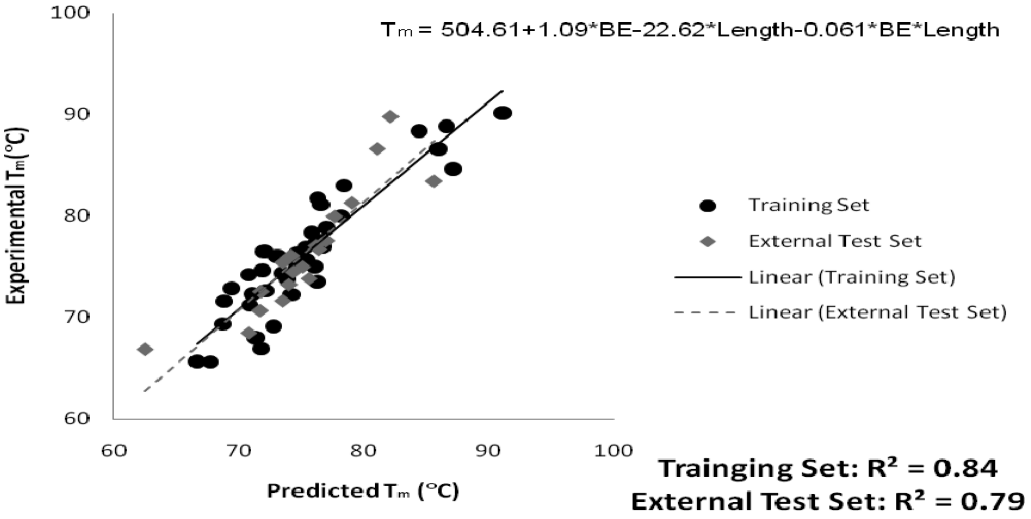


Figure 14. Standard curve for in-house data generated using a general linear multiple regression correlation between experimental T_m and calculated BE from 41 data in the training set (the black round points) generated from in-house DsiRNA data set, siRNA data set, and 2′-OMe data set. The rest of in-house data (20, the gray diamond points) are served as external test set for validation purposes.

designed with LNA or another nuclease-resistant nucleoside placed at the 5′ and 3′ 'wing' regions, leaving a "gap" of DNA to allow RNaseH recruitment. The wing regions serve the dual purpose of not only improving the exonuclease stability but also improving the T_m. The LNA modification uses a 2′-O, 4′-C methylene bridge on the ribose ring to restrict the conformational flexibility and produces an RNA-like C3′-endo conformation to the sugar moiety of the oligonucleotide. This greatly enhances the binding affinity of the LNA with both DNA, resulting in an approximate 5 °C increase in T_m with each LNA insertion. The PS linkage, where a nonbridging oxygen on the phosphate linkage is replaced with a sulfur atom, is typically included to increase the nuclease stability of the compounds although some authors have reported that it may have a role in cell penetration. In our evaluation, a literature data set of 21 LNA and PS-modified 16mers from 3 different parental AON sequences were selected for evaluation.

As can be seen in Figures 10−12, overall the BE to T_m correlation is maintained even with different nucleotide modification patterns. R^2 values of 0.86, 0.63, and 0.92 (0.80 with the exclusion of the lowest T_m) are seen for the DsiRNA, 2′-OMe, and LNA-PS AON gapmer data sets, respectively. Again, as seen in Figure 13, the importance of sequence length appears as a dominant factor, not allowing correlation between data sets with compounds of different lengths. This length dependence is related to entropic differences which are not accounted for within our protocol. One possible explanation for the wide range of R^2 values lies in the experimental T_m range of each data set. The ranges are 25 °C, 10 °C, and 40 °C for the DsiRNA, the 2′-OMe siRNA, and the LNA-PS AON data set, respectively. Generally, the narrower the response range (T_m in our case), the larger the predicted error, and this can be seen in our data. For example, the 2′-OMe siRNA data set has the smallest T_m range among the three and the largest potential for error. Additionally, customized charmm27 force field parameters that were adopted from the literature for our chemically modified data sets are likely to impact the accuracy of the calculation. Because the parameters for the modified residues are not as robust as for the standard residues, it is not surprising to see better correlation between the calculated BE and the experimental T_m for the DsiRNA data set.[40,41]

In practical application, the computational estimation of T_m for novel chemical modifications and insertion patterns remains challenging because of the difficulty of accurate force field parametrization and the need to create a standard T_m vs BE curve. However, the expense of the computational experiment can be easily justified in comparison with the difficulties of preparing the materials. Nucleotide synthesis is typically lengthy, and to properly evaluate a novel nucleotide, a gram or more of material must be made for at least two if not all four of the bases. The nucleotides must then be incorporated into sequences, purified, and duplexed. Because of the sequence dependence of many properties, multiple sequences need to be made and tested to get a true understanding of its properties. Overall the synthetic process can take several months, making thorough theoretical evaluations of the many potential options critical.

Given the dependence of the BE vs T_m correlation on length and nucleoside chemistry, it is only possible to create a qualitative relation when all of the variables are considered simultaneously as is the case when evaluating novel structures. Here, we divided all of the in-house data sets examined thus far, including various lengths and chemical modifications, into a training set (41 data) and test set (20 data) based on the experimental T_m range. In Figure 14 we show a general linear multiple regression model that was fit to the training set and validated using the external test set with R^2 of 0.84 and 0.79, respectively. The data is fit to a line which includes length as variable along with binding energy.

$$T_m = 504.61 + (1.09 \times \mathrm{BE}) - (22.62 \times \mathrm{length}) - (0.061 \times \mathrm{BE} \times \mathrm{length})$$

While the relation is only qualitative it should surfice for in silico virtual screening of duplex melting temperatures to aid in the study of modified nucleotide insertion patterns as well as rank ordering potential novel nucleotide modifications.

Overall we have shown that an accurate T_m can be predicted from the strand binding energy of a molecular dynamics ensemble for a short oligonoucleotide duplex. While highly length dependent the correlation is maintained across multiple chemical modifications and insertion patterns. Standard curves can therefore be generated for use with evaluation of potential novel

nucleotides to predict their effect on overall strand T_m, as well as the effects of various modification patterns on strand stability. While force field parametrization of novel nucleosides remains challenging this technique remains an important tool to evaluate and prioritize potential modifications, given the enormous time and effort necessary to synthetically develop the compounds.

CONCLUSION

The protocol we have developed is an automated protocol using CHARMm MD simulations and BE calculation to estimate the thermodynamic properties for unmodified and chemically modified oligonucleotide duplexes. Our simulations and evaluations have shown that the calculated BE correlates well with the experimental T_m. It can be used to predict the T_m for both unmodified and chemically modified siRNA and AON in a length-dependent manner. Although a full understanding of thermostability has not yet been accomplished because of the length-dependent correction that must be determined, our process does provide insights into selecting and prioritizing novel oligonucleotide modifications for further study prior to time-consuming synthesis in siRNA and AON therapeutics.

ASSOCIATED CONTENT

Supporting Information. Complete sequence information for the compounds, the 2′-OMe charmm27 force field parameters, and the methodology for the computational workflow design, integration, and customization. This material is available free of charge via the Internet at http://pubs.acs.org.

AUTHOR INFORMATION

Corresponding Author

*Tel: +1 617 5513274. Fax: +1 617 5513117. E-mail: robert.stanton@pfizer.com.

Present Addresses

[§]Novartis Inc., Building 8 Lane 898 Halei Rd., Pudong New Area, Shanghai 201203, China.

[¶]EMD Serono Research Institute, 45A Middlesex Turnpike, Billerica, MA 01821.

[£]Cubist Pharmaceuticals, 65 Hayden Ave., Lexington, MA 02421.

ACKNOWLEDGMENT

The authors thank Jeremy Little for assistance in synthesizing oligonucleotides and Jodi Shaulsky for assistance in BE calculation PP script.

REFERENCES

(1) Melnikova, I. RNA-based therapeutics. *Nat. Rev. Drug Discovery* **2007**, *9*, 863–864.

(2) Fire, A.; Xu, S.; Montgomery, M. K.; Kostas, S. A.; Driver, S. E.; Mello, C. C. Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans. *Nature* **1998**, *391*, 806–811.

(3) Weiss, B.; Davidkova, G.; Zhou, L. W. Antisense RNA gene therapy for studying and modulating biological processes. *Cell Mol. Life Sci.* **1999**, *55*, 334–358.

(4) Xia, T.; SantaLucia, J., Jr.; Burkard, M. E.; Kierzek, R.; Schroeder, S. J.; Jiao, X.; Cox, C.; Turner, D. H. Thermodynamic parameters for an expanded nearest-neighbor model for formation of RNA duplexes with Watson-Crick base pairs. *Biochemistry* **1998**, *37*, 14719–14735.

(5) Owczarzy, R.; Vallone, P. M.; Gallo, F. J.; Paner, T. M.; Lane, M. J.; Benight, A. S. Predicting sequence-dependent melting stability of short duplex DNA oligomers. *Biopolymers* **1997**, *44*, 217–239.

(6) SantaLucia, J., Jr. A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl. Acad. Sci. U.S.A.* **1998**, *95*, 1460–1465.

(7) Sugimoto, N.; Nakano, S.; Katoh, M.; Matsumura, A.; Nakamuta, H.; Ohmichi, T.; Yoneyama, M.; Sasaki, M. Thermodynamic parameters to predict stability of RNA/DNA hybrid duplexes. *Biochemistry* **1995**, *34*, 11211–11216.

(8) McTigue, P. M.; Peterson, R. J.; Kahn, J. D. Sequence-dependent thermodynamic parameters for locked nucleic acid (LNA)−DNA duplex formation. *Biochemistry* **2004**, *43*, 5388–5405.

(9) Zhang, H. Y.; Du, Q.; Wahlestedt, C.; Liang, Z. RNA Interference with chemically modified siRNA. *Curr. Top. Med. Chem.* **2006**, *6*, 893–900.

(10) Chiu, Y. L.; Rana, T. M. siRNA function in RNAi: a chemical modification analysis. *RNA* **2003**, *9*, 1034–1048.

(11) Vester, B.; Wengel, J. LNA (locked nucleic acid): high-affinity targeting of complementary RNA and DNA. *Biochemistry* **2004**, *43*, 13233–13241.

(12) Freier, S. M.; Kierzek, R.; Jaeger, J. A.; Sugimoto, N.; Caruthers, M. H.; Neilson, T.; Turner, D. H. Improved free-energy parameters for predictions of RNA duplex stability. *Proc. Natl. Acad. Sci. U.S.A.* **1986**, *83*, 9373–9377.

(13) Schöning, K.-U.; Scholz, P.; Wu, X.; Guntha, S.; Delgado, G.; Krishnamurthy, R.; Eschenmoser, A. The α-L-Threofuranosyl-(3′→2′)-oligonucleotide System ('TNA'): Synthesis and Pairing Properties. *Helv. Chim. Acta* **2002**, *85*, 4111–4153.

(14) Cheatham, T. E., III; Kollman, P. A. Molecular dynamics simulation of nucleic acids. *Annu. Rev. Phys. Chem.* **2000**, *51*, 435–471.

(15) Jha, S.; Coveney, P. V.; Laughton, C. A. Force field validation for nucleic acid simulations: comparing energies and dynamics of a DNA dodecamer. *J. Comput. Chem.* **2005**, *26*, 1617–1627.

(16) Chipot, C., Pohorille, A., Ed.; Free Energy Calculations: Theory and Applications in Chemistry and Biology. *Theor. Chem. Acc.* **2008**, *121*, 105-106.

(17) Gilson, M. K.; Zhou, H. X. Calculation of protein-ligand binding affinities. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 21–42.

(18) Gilson, M. K.; Given, J. A.; Bush, B. L.; McCammon, J. A. The statistical-thermodynamic basis for computation of binding affinities: a critical review. *Biophys. J.* **1997**, *72*, 1047–1069.

(19) Olson, M. A. Calculations of free-energy contributions to protein-RNA complex stabilization. *Biophys. J.* **2001**, *81*, 1841–1853.

(20) Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983**, *79*, 926–935.

(21) Foloppe, N.; MacKerell, J. A. D. All-atom empirical force field for nucleic acids: I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.* **2000**, *21*, 86–104.

(22) MacKerell, A. D.; Banavali, N. K. All-atom empirical force field for nucleic acids: II. Application to molecular dynamics simulations of DNA and RNA in solution. *J. Comput. Chem.* **2000**, *21*, 105–120.

(23) Pande, V.; Nilsson, L. Insights into structure, dynamics and hydration of locked nucleic acid (LNA) strand-based duplexes from molecular dynamics simulations. *Nucleic Acids Res.* **2008**, *36*, 1508–1516.

(24) Aduri, R.; Psciuk, B. T.; Saro, P.; Taniga, H.; Schlegel, H. B.; SantaLucia, J. AMBER Force Field Parameters for the Naturally Occurring Modified Nucleosides in RNA. *J. Chem. Theory Comput.* **2007**, *3*, 1464–1475.

(25) Brooks, B. R.; Bruccoleri, R. E.; Olafson, B. D.; States, D. J.; Swaminathan, S.; Karplus, M. CHARMM: A Program for Macromolecular Energy, Minimization, and Dynamics Calculations. *J. Comput. Chem.* **1983**, *4*, 187–217.

(26) Brooks, B. R.; Brooks, C. L., III; Mackerell, A. D.; Nilsson, L.; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C.; Boresch, S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S.; Gao, J.;

Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J.; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z.; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X.; Yang, W.; York, D. M.; Karplus, M. CHARMM: The Biomolecular simulation Program. *J. Comput. Chem.* **2009**, *30*, 1545–1615.

(27) Ryckaert, J.-P.; Ciccotti, G.; Berendsen, H. J. C. Numerical integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of n-Alkanes. *J. Comput. Phys.* **1977**, *23*, 327–341.

(28) *Pipeline Pilot*, version 6.1.5; Accelrys: San Diego, CA, 2009.

(29) *Disocvery Studio*, version 2.5; Accelrys: San Diego, CA, 2009.

(30) Xia, J.; Noronha, A.; Toudjarska, I.; Li, F.; Akinc, A.; Braich, R.; Frank-Kamenetsky, M.; Rajeev, K. G.; Egli, M.; Manoharan, M. Gene silencing activity of siRNAs with a ribo-difluorotoluyl nucleotide. *ACS Chem. Biol.* **2006**, *1*, 176–183.

(31) Li, F.; Pallan, P. S.; Maier, M. A.; Rajeev, K. G.; Mathieu, S. L.; Kreutz, C.; Fan, Y.; Sanghvi, J.; Micura, R.; Rozners, E.; Manoharan, M.; Egli, M. Crystal structure, stability and in vitro RNAi activity of oligoribonucleotides containing the ribo-difluorotoluyl nucleotide: insights into substrate requirements by the human RISC Ago2 enzyme. *Nucleic Acids Res.* **2007**, *35*, 6424–6438.

(32) Teplova, M.; Minasov, G.; Tereshko, V.; Inamati, G. B.; Cook, P. D.; Manoharan, M.; Egli, M. Crystal structure and improved antisense properties of 2′-O-(2-methoxyethyl)-RNA. *Nat. Struct. Biol.* **1999**, *6*, 535–539.

(33) Haeberli, P.; Berger, I.; Pallan, P. S.; Egli, M. Syntheses of 4′-thioribonucleosides and thermodynamic stability and crystal structure of RNA oligomers with incorporated 4′-thiocytosine. *Nucleic Acids Res.* **2005**, *33*, 3965–3975.

(34) Rozners, E.; Moulder, J. Hydration of short DNA, RNA and 2′-OMe oligonucleotides determined by osmotic stressing. *Nucleic Acids Res.* **2004**, *32*, 248–254.

(35) Egli, M.; Minasov, G.; Tereshko, V.; Pallan, P. S.; Teplova, M.; Inamati, G. B.; Lesnik, E. A.; Owens, S. R.; Ross, B. S.; Prakash, T. P.; Manoharan, M. Probing the influence of stereoelectronic effects on the biophysical properties of oligonucleotides: comprehensive analysis of the RNA affinity, nuclease resistance, and crystal structure of ten 2′-O-ribonucleic acid modifications. *Biochemistry* **2005**, *44*, 9045–9057.

(36) Egli, M.; Pallan, P. S. Insights from crystallographic studies into the structural and pairing properties of nucleic acid analogs and chemically modified DNA and RNA oligonucleotides. *Annu. Rev. Biophys. Biomol. Struct.* **2007**, *36*, 281–305.

(37) Dock-Bregeon, A. C.; Chevrier, B.; Podjarny, A.; Moras, D.; deBear, J. S.; Gough, G. R.; Gilham, P. T.; Johnson, J. E. High resolution structure of the RNA duplex [U(U-A)6A]2. *Nature* **1988**, *335*, 375–378.

(38) Amarzguioui, M.; Rossi, J. J. Principles of Dicer substrate (D-siRNA) design and function. *Methods Mol. Biol.* **2008**, *442*, 3–10.

(39) Frieden, M.; Christensen, S. M.; Mikkelsen, N. D.; Rosenbohm, C.; Thrue, C. A.; Westergaard, M.; Hansen, H. F.; Orum, H.; Koch, T. Expanding the design horizon of antisense oligonucleotides with alpha-L-LNA. *Nucleic Acids Res.* **2003**, *31*, 6365–6372.

(40) MacKerell, A. D., Jr.; Banavali, N.; Foloppe, N. Development and current status of the CHARMM force field for nucleic acids. *Biopolymers* **2000**, *56*, 257–265.

(41) Priyakumar, U. D.; Mackerell, A. D., Jr. Atomic detail investigation of the structure and dynamics of DNA.RNA hybrids: a molecular dynamics study. *J. Phys. Chem. B.* **2008**, *112*, 1515–1524.