

Accurately Modeling the Conformational Preferences of Nucleosides

Mihai Burai Patrascu, Elise Malek-Adamian, Masad J. Damha, and Nicolas Moitessier*

Department of Chemistry, McGill University, 801 Sherbrooke Street West, Montreal, Quebec H3A 0B8, Canada

ABSTRACT: Sugar puckering of nucleosides impacts nucleic acid structures; hence their biological function. Similarly, nucleoside-based therapeutics may adopt different conformations affecting their binding affinity, DNA incorporation, and excision rates. As a result, significant efforts have been made to develop nucleoside analogues adopting specific conformations to improve bioactivity and pharmacokinetic profiles of the corresponding nucleoside-containing drugs. Understanding and ultimately predicting these conformational preferences would significantly help in the design of more effective structures. We report herein a computational study based on hybrid QM/MM umbrella sampling simulations that allow the accurate prediction of the sugar conformational preferences of chemically modified nucleosides in solution. Moreover, we pair these simulations with natural bond orbital (NBO) analysis to gain key insights into the role of substituents in the conformational preferences of these nucleosides.

Chemically modified oligonucleotides have generated interest as promising therapeutic agents since the advent of small interfering RNA and antisense technologies. These oligomers are chemically modified at the nucleotide components to increase their stability, bioavailability, and overall pharmacokinetic properties.¹ Alone, these modified nucleotides also serve as unnatural analogues for the treatment of viral infections and cancer (Figure 1a). Examples are Zidovudine (1),² Lamivudine (2),³ and Emtricitabine (3),⁴ inhibitors of the HIV's reverse transcriptase, and Gemcitabine (4), a chemotherapeutic.⁵ Enzyme inhibitors and other small-molecule drugs often owe their activity to their shape and chemical features. High-throughput computational methods to investigate and/or guide the design of this class of small-molecule drugs, such as docking methods,⁶ often consider drug and even macromolecular target flexibility. In the field of chemically modified oligonucleotides, the problem is more complex since a subtle structural change in a nucleotide may have a profound effect on the duplex's shape and stability and requires more time-consuming methods.⁷ Nucleoside building blocks can be tailored to adopt different conformations (sugar puckers), which, in turn, affect duplex structure, pairing affinity, and ultimately biological activity.⁸ In general, sugars that adopt the C3'-endo conformation (North, Figure 1b) demonstrate increased binding affinity toward complementary RNA strands.⁹ Electronegative 2'-substituents, such as 2'-methoxyethyl and 2'-F (Figure 1c), impart stereoelectronic effects that favor the North conformation.⁸ In contrast, the 2'-FANA modification has been shown to favor the South/ East conformation, but it has still been shown to increase binding to complementary RNA targets.¹⁰

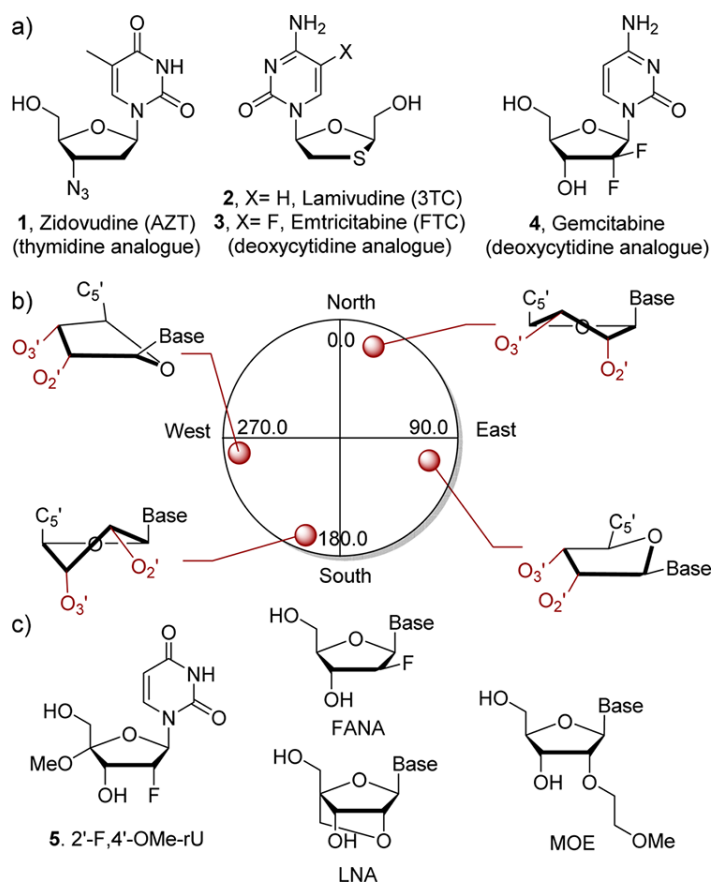


Figure 1. (a) Nucleoside analogues used as drugs. (b) Conformational characterization of the ribose pucker.¹¹ (c) 2'-F,4'-OMe-rU (5)¹² and clinically relevant nucleoside analogues.

The puckering of the ribose (or deoxy-ribose) ring is in large part controlled by hyperconjugation effects. To model these molecules, one can consider molecular mechanics (MM)-derived methods, although their accuracy for this class of molecules must be demonstrated and/or improved.¹³ One can also envision quantum mechanical (QM) techniques, although investigating the dynamic equilibrium between the different conformations and the conformational ensemble may be time-consuming. We proposed to find a suitable computational technique that would accurately describe the conformation of nucleosides in solution. Herein we report a benchmark study performed on a compound

we recently synthesized ¹² (5, Figure 1c) and the identification of a methodology for accurately predicting the conformational preferences of various carbohydrates and nucleosides.

Computational investigations of furanoside puckering reported somewhat conflicting information. On one hand, density functional theory (DFT) calculations were used with carbohydrates such as α - and β -D-aldopentofuranosides.¹⁴ On the other hand, Roy et al.¹⁵ reported the use of a special molecular dynamics (MD) technique - umbrella sampling - advocating for its use as a principal method of investigation of sugar puckering in saccharides. They also showed that, compared to a QM/MM approach using semiempirical methods such as self-consistent charge density functional tight binding (SCC-DFTB), a pure MM approach using the GLYCAM parameter set yielded more accurate results.^{16,17} By contrast, others showed that, among various semiempirical QM methods, SCC-DFTB was the one that produced sugar pucker closest to those obtained by high-level calculations.^{18,19} While some methods arrive at apparently contradictory results, the systems under scrutiny were significantly different. For example, electron-withdrawing substituents (e.g., fluorine) on sugar rings affect sugar puckering due to hyperconjugation effects. While these effects can be described by QM methods, they are poorly modeled by MM methods, which are not specifically parametrized for these effects.¹³

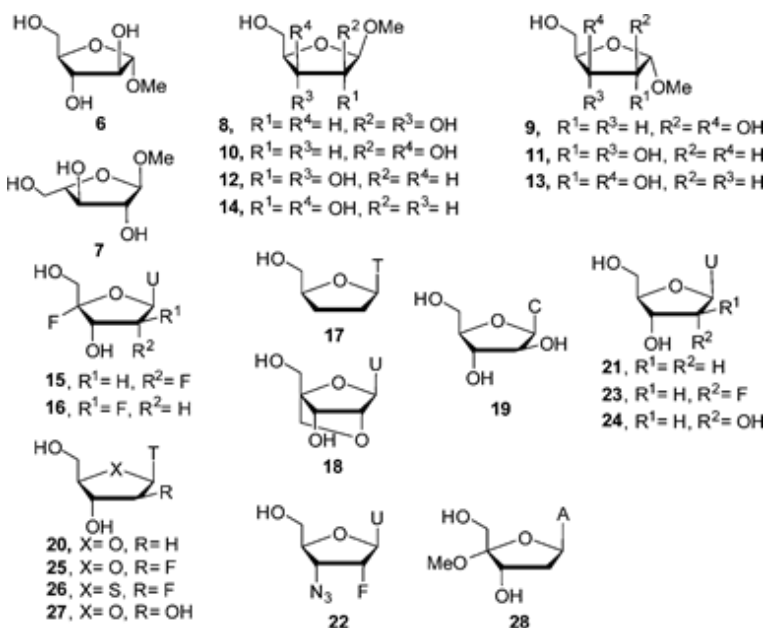
In order to make an informed decision, we started a benchmark study. We first assessed DFT calculations on a set of idealized envelope conformations of nucleoside 5.¹² While the experimental data on 5 show a N/S ratio of 87:13, the DFT calculations yielded a N/S ratio >100:1 (Figure S1), which we attributed to the improper modeling of the water effect. We then switched our attention to MM calculations in water, which showed an erroneous N/S

ratio favoring the South region (Figure S2), likely underestimating the hyperconjugation effects. Since DFT and MM calculations proved to be unsuitable for this type of system, we turned to hybrid QM/MM umbrella sampling simulations. Umbrella sampling ensures that the conformational space is effectively explored. While the ligand is treated in a QM fashion, the explicit solvent is treated with MM. We decided to test various semiempirical methods for the QM region, namely RM1, PM3, PM6 (previously found to be more accurate than PM7 on carbohydrates),²⁰ PM3CARB1, and SCC-DFTB. Timewise, this approach performed similarly to the pure MM approach. A comparison of the PMF curves obtained with RM1, PM3, PM6, PM3CARB1, and SCC-DFTB (Figure S2) showed that only the latter correctly assigned the global minimum in the North region, while finding a local minimum in the South conformation. We computed a N/S ratio of 84:16, similar to the experimental N/S ratio of 87:13. To verify the accuracy of our modeling, we predicted the NMR spectrum of **5** using a DFT/gauge-invariant atomic orbital approach, obtaining an RMSE of 0.34 ppm between the predicted and experimental NMR spectra (Supporting Information (SI)), which is within the reported error for this method but too large to accurately predict N/S ratios.²¹ While these preliminary results provided us with confidence in our modeling, testing this method on a single compound was not sufficient to make assertions about its accuracy; thus, we assembled a set of various monosaccharides (6–14) and nucleosides (1, 5, 15–28) published in the literature that could serve as a validation set (Chart 1 and Chart S1).

We then verified whether the QM/MM approach using SCC-DFTB would yield N/S ratios and conformers close to experimental ones (Figure 2). Crystal structure data²² were

available for some monosaccharides; thus, we were also able to compare our lowest-energy conformations to those found in the crystal structures (Figure 3). The experimental N/S ratios were obtained using the program PSEUROT²³ from NMR spectroscopy data (Tables S6 and S7).²⁴ The available experimental data are thus indirect and are the result of some computations with the associated error. Henceforth the experimental data obtained with PSEUROT shall be referred to as pseudo-experimental data. In fact, we even found conflicting reports about the pseudo-experimentally determined N/S ratio of 6.^{14,16}

Chart 1. Compounds Subjected to QM/MM Umbrella Sampling Simulations



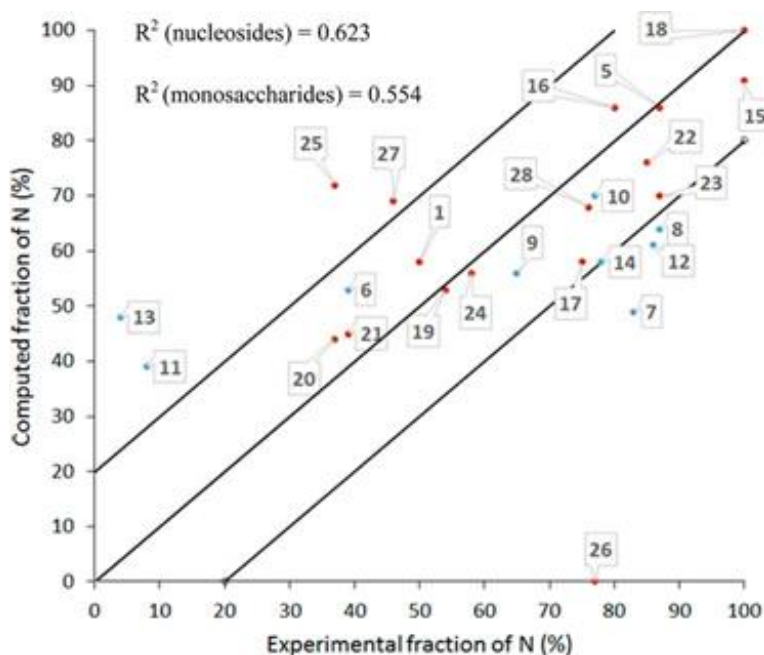


Figure 2. Comparison between predicted and pseudo-experimental %N for monosaccharides (blue) and nucleosides (red).

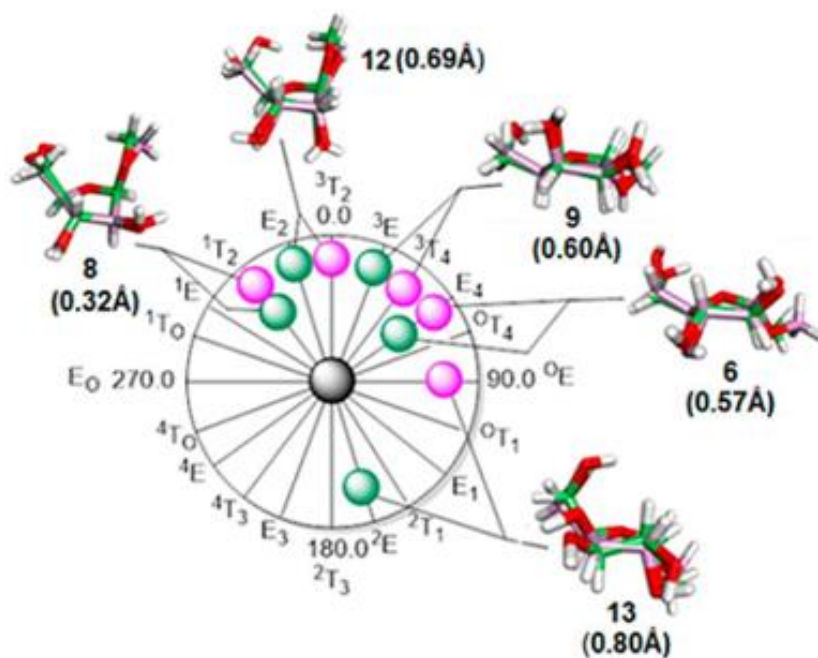


Figure 3. Superposition of the crystal structures and lowest-in-energy predicted conformers of 6, 8, 9, 12, and 13 (pink, computed structure; green, crystal structure). RMSD values are given in parentheses.

Overall the data in Figure 2 suggest that SCC-DFTB performs well when predicting the N/S ratios for monosaccharides; moreover, when compared to the crystal structure conformers, all our computed structures had a heavy atom position RMSD $< 1 \text{ \AA}$ (Figure 3). The computed structures, with the exception of 12, show no intramolecular hydrogen bonding, consistent with the crystal structures.^{22a} The differences between the computed and crystal structures arose from a strong intermolecular H-bonding network present in the crystal structure but broken in solution, as recently reported.²⁵ We then analyzed our predicted N/S ratios. As mentioned earlier, analysis of the computed N/S ratio for 6 proved to be difficult, since the pseudo-experimental data are conflicting. Nonetheless, the fact that our computed lowest energy conformer (SI) matched that obtained in the crystal structure gave us confidence that our ratios were accurate. This assertion is supported by the good agreement between the pseudo-experimental and computed data (Figure 2), with three exceptions: 7, 11, and 13. The computed ring pucker distribution for 7 showed a concentrated ring population in the SE/S, SW/ NW regions as well as in the N region. This behavior was described by Evdokimov et al.^{22b} in their furanoside crystal structure analysis where they showed that furanosides preferentially adopt conformations where the ring substituents have nominal eclipsing, namely in the SE/SW ($P = 160^\circ$ to 200°) and NW/N ($P = 340^\circ$ to 20°) areas, and avoid regions of maximum eclipsing ($P = 90^\circ$ and 270°). We understand that crystal structures and solution structures may be different, but we believe that the agreement between models and experiments is an additional validation of our modeling. Similarly, the computed data for 11 and 13 shows a large population of SE and NE in addition to S and N, while PSEUROT assumes a 2-state equilibrium.

The computed N/S ratios for the nucleosides shown in Figure 2 indicate that our method predicts their conformational preferences, with a few exceptions (nucleosides 25–27), with ratios close to those derived from experimental NMR data. Since we also wanted to understand why our method did not work on all the nucleosides, we decided to run a natural bond orbital (NBO) analysis on all the nucleosides, which makes it possible to quantify the anomeric and gauche effects, allowing for the accurate computation of the energetic contribution of these effects toward the overall puckering. Furthermore, NBO analysis allows the computation of the molecular orbitals of every nucleoside, effectively offering important insight into potential intramolecular H-bonding (SI).

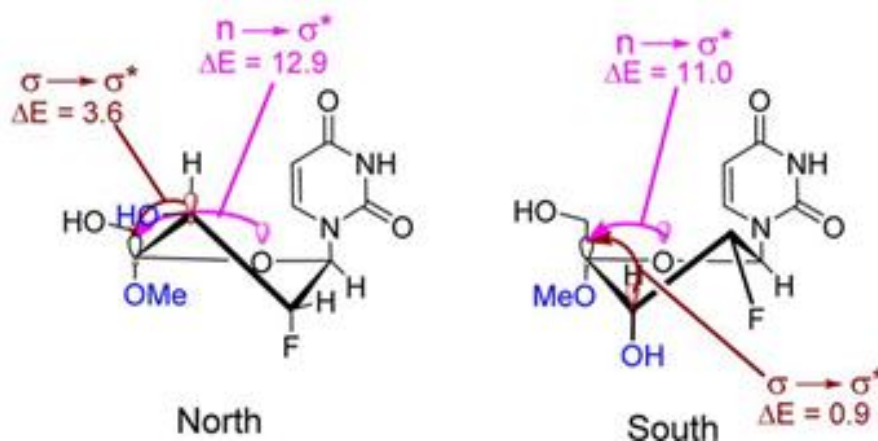


Figure 4. Anomeric and gauche effects observed for the North and South structures of 5. All relative energies are given in kcal/mol. Anomeric effects ($nO4' \rightarrow \sigma^*C4'OMe$) are described by pink arrows, while hyper-conjugation effects ($\sigma C3'H \rightarrow \sigma^*C4'OMe$) are described by brown arrows.

The anomeric effect found in 5 ($nO4' \rightarrow \sigma^*C4'OMe$, Figure 4) in the case of the North pucker is more pronounced than in the South pucker (with a computed energy difference of 1.9 kcal/mol) due to the position of the anomeric oxygen above the plane for the N conformation and within the ring plane for the S conformation. The anomeric effect is not

the sole contributor to the sugar puckering preference - several gauche effects also play important roles in this respect. For example, the $\sigma\text{C3'H}\rightarrow\sigma^*\text{C4'OMe}$ (Figure 4) and $\sigma\text{C3'H}\rightarrow\sigma^*\text{C2'F}$ gauche effects are prevalent in the North conformation, while the $\sigma\text{C3'C4'}\rightarrow\sigma^*\text{C2'F}$ and $\sigma\text{C2'H}\rightarrow\sigma^*\text{C3'O}$ gauche effects are predominant in the South conformation. We also looked into the possibility of stabilizing intramolecular H-bonding and found that three H-bonds were observed for the North conformation and only one for the South conformation.¹² Interestingly, SCC-DFTB was shown to somewhat underestimate H-bond strengths;²⁶ thus, the effects of the intramolecular H-bonds might be far greater than presented here. Following this analysis, we tried to understand why computations with 2'-araF-T (25) failed to yield a theoretical ratio close to the experimental one. For almost all nucleosides (except 25), we expected a 2-state equilibrium (with two minima distinguishable in the North and South regions). A report by Watts and Damha²⁷ showed that 2'-fluoroarabinonucleosides can adopt a nstate ($n = 2, 3$, and 4) equilibrium, with potential minima in the N, NE, SE, and S regions of the pseudo-rotational circle. While conformational analysis of 25 has only been described once, it is not unreasonable to assume that it may in fact adopt a 3- or 4-state equilibrium in solution. To understand whether 25 could indeed adopt such an equilibrium, we plotted the PMF curve and puckering distributions (SI). The low energy barrier (~ 0.5 kcal/mol) for the N-E transition advances the possibility of a 3-state equilibrium. Significant puckering populations are present in the NE, SE, and S regions of the pseudo-rotational circle, making the 2-state equilibrium hypothesis unreliable. The NBO analysis we performed on the lowest-energy conformations of 25 revealed that the anomeric effect $\text{nO4'}\rightarrow\sigma^*\text{C4'OMe}$ was slightly more

pronounced in the NE conformation ($\Delta E = +0.3$ kcal/mol). The gauche effects were overall prevalent in the NE conformation, although the energy differences were somewhat small (SI). The $\sigma_{C3'C4'} \rightarrow \sigma^*_{C2'F}$ effect was significantly larger for the NE conformation ($\Delta E = +4.0$ kcal/mol), most likely allowed by the better orbital overlap when compared to the SE conformation. We then turned to nucleoside 26. Compared to all the other nucleosides investigated, 26 contains a sulfur atom, instead of the ring oxygen, which changes the stereoelectronic effects governing the sugar puckering. While introduction of this atom in the ring pushes the conformation of the nucleoside to the North region, our method suggests a primarily South conformation. Interestingly, a report by Petraglia and Corminboeuf highlights the lack of proper parametrization of sulfur in the SCC-DFTB method,²⁸ which led to non-covalent interactions being poorly described by the current SCC-DFTB. As non-covalent interactions (especially polarization effects) play a significant role in sugar puckering, the failure of our method with 26 could likely be attributed to the lack of proper sulfur parametrization. When nucleoside 27 is compared to nucleoside 19, it is seen that the two differ only in the identity of the base; molecular orbital analysis of both 19 and 27 showed a pronounced intramolecular H-bonding network (SI). These stabilizing interactions contribute to the overall puckering shift toward the North region. Furthermore, NBO analysis revealed that the anomeric effect is very slightly in favor of the South conformation, while all the hyperconjugation effects are dominant in the North conformation. The results are very similar to those for 19, where the anomeric effect is also dominant in the South conformation ($\Delta E = -0.6$ kcal/mol), while the gauche effects are predominant in the North conformation. The last nucleoside that initially presented a

deviation from the experimental value was nucleoside 28 (exp.: 91:9, pred.: 68:32). An electron-withdrawing substituent in the 4' position is expected to shift the sugar puckering toward the North conformation, due to the more pronounced anomeric effect. To understand where this discrepancy comes from, we decided to verify the experimental results associated with this nucleoside in the reported SI. According to our calculations (based on the data provided in the original report²⁹), the N/S ratio (76:24) is closer to our predictions. This is in line with our expectations, since the lack of a strong electron-withdrawing substituent in the 2' position affects the N/S ratios in favor of the South conformation. This is also supported by NBO analysis, which shows a significant hyperconjugation effect ($\sigma_{C2'H} \rightarrow \sigma^*_{C3'O}$) in favor of the South conformation ($\Delta E = -3.5$ kcal/mol). Moreover, NBO analysis revealed one intramolecular H-bond for both the North and South structures (SI). Throughout these investigations, we have observed that neither pure DFT nor pure MM methods we tested could reproduce the conformational behavior of non-natural nucleosides in solution. While DFT accounts for hyperconjugation effects, solvent effects such as H-bonds in water can only be reproduced by MM techniques (explicit solvent molecules). A hybrid QM/MM method was the solution to investigate the conformational preferences of a variety of nucleosides and modified monosaccharides. This method yielded reliable structures that closely matched the crystal structures for the modified monosaccharides, as well as N/S ratios close to experimental ones for both the monosaccharides and nucleosides. Moreover, our method provides key insights into the role of the substituents on the sugar ring of nucleosides and the finely tuned control of sugar puckering by hyperconjugation effects and intramolecular H-bonding. However, the method we describe here also has limitations, such as improper sulfur parametrization and

underestimated strengths of H-bonds. Nonetheless, we believe that if the system of interest is a non-natural nucleoside, one can obtain reliable structures and N/S ratios (if only for qualitative description of one's system before attempting synthesis) by using a QM/MM approach involving SCCDFTB. Finally, if resources permit, one can attempt the analysis of a non-natural nucleoside with DFTB3,26 which appears to describe H-bonds better.

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.7b07436.

QM/MM and NBO protocol; DFT, MM, and QM/MM benchmark study; relevant electronic effects and intramolecular H-bonding in nucleosides; PMFs and puckering distributions for all nucleosides and monosaccharides (PDF)

AUTHOR INFORMATION

Corresponding Author

*nicolas.moitessier@mcgill.ca

ORCID

Masad J. Damha: 0000-0002-4458-1623

Nicolas Moitessier: 0000-0001-6933-2079

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

We acknowledge the Natural Sciences and Engineering Research Council of Canada (NSERC) for financial support to N.M. and M.J.D. as well as Compute Canada for their generous CPU time allocations.

REFERENCES

- (1) Khvorova, A.; Watts, J. K. *Nat. Biotechnol.* 2017, 35, 238.
- (2) Broder, S. *Antiviral Res.* 2010, 85, 1.
- (3) Soudeyns, H.; Yao, X. I.; Gao, Q.; Belleau, B.; Kraus, J. L.; Nguyen-Ba, N.; Spira, B.; Wainberg, M. A. *Antimicrob. Agents Chemother.* 1991, 35, 1386.
- (4) Liotta, D. C.; Painter, G. R. *Acc. Chem. Res.* 2016, 49, 2091.
- (5) Plunkett, W.; Huang, P.; Gandhi, V. *Anti-Cancer Drugs* 1995, 6, 7.
- (6) Moitessier, N.; Englebiennne, P.; Lee, D.; Lawandi, J.; Corbeil, C. R. *Br. J. Pharmacol.* 2008, 153, S7.
- (7) Yurenko, Y. P.; Zhurakivsky, R. O.; Ghomi, M.; Samijlenko, S. P.; Hovorun, D. M. *J. Phys. Chem. B* 2007, 111, 9655.
- (8) Deleavey, G. F.; Damha, M. J. *Chem. Biol.* 2012, 19, 937.
- (9) Kawasaki, A. M.; Casper, M. D.; Freier, S. M.; Lesnik, E. A.; Zounes, M. C.; Cummins, L. L.; Gonzalez, C.; Cook, P. D. *J. Med. Chem.* 1993, 36, 831.
- (10) (a) Trempe, J.-F.; Wilds, C. J.; Denisov, A. Y.; Pon, R. T.; Damha, M. J.; Gehring, K. *J. Am. Chem. Soc.* 2001, 123, 4896. (b) Damha, M. J.; Wilds, C. J.; Noronha, A.; Brukner, I.; Borkow, G.; Arion, D.; Parniak, M.A. *J. Am. Chem. Soc.* 1998, 120, 12976. (c) Kalota, A.; Karabon, L.; Swider, C. R.; Viazovkina, E.; Elzagheid, M.; Damha, M. J.; Gewirtz, A. *M. Nucleic Acids Res.* 2006, 34, 451.
- (11) Li, F.; Sarkhel, S.; Wilds, C. J.; Wawrzak, Z.; Prakash, T. P.; Manoharan, M.; Egli, M. *Biochemistry* 2006, 45, 4141.
- (12) Malek-Adamian, E.; Guenther, D. C.; Matsuda, S.; Martinez-Montero, S.; Zlatev, I.; Harp, J.; Burai Patrascu, M.; Foster, D. J.; Fakhoury, J.; Perkins, L.; Moitessier, N.;

- Manoharan, R. M.; Taneja, N.; Bisbe, A.; Charisse, K.; Maier, M.; Kallanthottathil, R. G.; Egli, M.; Manoharan, M.; Damha, M. J. *J. Am. Chem. Soc.* 2017, accepted.
- (13) Liu, Z.; Pottel, J.; Shahamat, M.; Tomberg, A.; Labute, P.; Moitessier, N. *J. Chem. Inf. Model.* 2016, 56, 788.
- (14) Houseknecht, J. B.; Lowary, T. L.; Hadad, C. M. *J. Phys. Chem. A* 2003, 107, 5763.
- (15) Islam, S. M.; Richards, M. R.; Taha, H. A.; Byrns, S. C.; Lowary, T. L.; Roy, P.-N. *J. Chem. Theory Comput.* 2011, 7, 2989.
- (16) Islam, S. M.; Roy, P.-N. *J. Chem. Theory Comput.* 2012, 8, 2412.
- (17) Kirschner, K. N.; Yongye, A. B.; Tschampel, S. M.; González-Outeiriño, J.; Daniels, C. R.; Foley, L. B.; Woods, R. J. *J. Comput. Chem.* 2008, 29, 622.
- (18) Barnett, C. B.; Naidoo, K. J. *J. Phys. Chem. B* 2010, 114, 17142.
- (19) Jaiyong, P.; Bryce, R. A. *Phys. Chem. Chem. Phys.* 2017, 19, 15346.
- (20) http://openmopac.net/PM7_accuracy/SACCHARIDES.html
(accessed August 28, 2017).
- (21) Pankratyev, E. Y.; Tulyabaev, A. R.; Khalilov, L. M. *J. Comput. Chem.* 2011, 32, 1993.
- (22) (a) Evdokimov, A.; Gilboa, A. J.; Koetzle, T. F.; Klooster, W. T.; Schultz, A. J.; Mason, S. A.; Albinati, A.; Frolow, F. *Acta Crystallogr., Sect. B* 2001, 57, 213. (b) Evdokimov, A. G.; Martin, J. M. L.; Kalb, A. J. *J. Phys. Chem. A* 2000, 104, 5291. (c) Evdokimov, A.; Kalb, A. J.; Koetzle, T. F.; Klooster, W. T.; Martin, J. M. L. *J. Phys. Chem. A* 1999, 103, 744.
- (23) De Leeuw, F. A. A. M.; Altona, C. J. *Comput. Chem.* 1983, 4, 428.
- (24) Taha, H. A.; Richards, M. R.; Lowary, T. L. *Chem. Rev.* 2013, 113, 1851.

- (25) Rocheleau, S.; Pottel, J.; Huskić, I.; Moitessier, N. *Eur. J. Org.Chem.* 2017, 2017, 646.
- (26) (a) Gaus, M.; Cui, Q.; Elstner, M. *J. Chem. Theory Comput.* 2011, 7, 931. (b) Yang, Y.; Yu, H.; York, D.; Cui, Q.; Elstner, M. *J. Phys. Chem. A* 2007, 111, 10861.
- (27) Watts, J. K.; Damha, M. J. *Can. J. Chem.* 2008, 86, 641.
- (28) Petraglia, R.; Corminboeuf, C. *J. Chem. Theory Comput.* 2013, 9, 3020.
- (29) Petrová, M.; Páv, O.; Buděšínský, M.; Zborníková, E.; Novák, P.; Rosenbergová, Š.; Pačes, O.; Liboska, R.; Dvořáková, I.; Šimák, O.; Rosenberg, I. *Org. Lett.* 2015, 17, 3426.