

Rotational Profiler: A Fast, Automated, and Interactive Server to Derive Torsional Dihedral Potentials for Classical Molecular Simulations

Victor H. Rusu, Denys E. S. Santos, Marcelo D. Poletto, Marcelo M. Galheigo, Antônio T. A. Gomes, Hugo Verli, Thereza A. Soares, and Roberto D. Lins*

 Cite This: *J. Chem. Inf. Model.* 2020, 60, 5923–5927

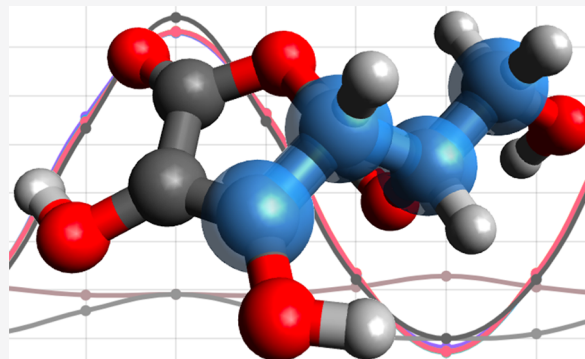
 Read Online

ACCESS |

 Metrics & More

 Article Recommendations

ABSTRACT: Rotational Profiler provides an analytical algorithm to compute sets of classical torsional dihedral parameters by fitting an empirical energy profile to a reference one that can be obtained experimentally or by quantum-mechanical methods. The resulting profiles are compatible with the functional forms in the most widely used biomolecular force fields (e.g., GROMOS, AMBER, OPLS, and CHARMM). The linear least-squares regression method is used to generate sets of parameters that best satisfy the fitting. Rotational Profiler is free to use, analytical, and force field/package independent. The formalism is herein described, and its usage, in an interactive and automated manner, is made available as a Web server at <http://rotprof.lncc.br>.



INTRODUCTION

Molecular simulations have reached maturity and are nowadays used not only to understand but also to predict molecular-level properties. Applications range from physics to chemistry and the very diverse biological fields (e.g., biophysics, proteomics, biological chemistry). Modern problems in biological chemistry typically fall within room and physiological temperatures. Therefore, the energies associated with these processes are on the order of tens of kJ mol^{-1} .¹ As a result, large biological complex phenomena (e.g., molecular interactions, protein folding, computational protein engineering, membrane remodeling) are best described by classical statistical mechanics.

In the above setting, a Hamiltonian is associated with a set of classical parameters (i.e., a force field) that can provide the potential energy of a given conformation of a molecular system. The parameters may come directly from experiments, from high-precision quantum-mechanical methods, or from fitting to experimental macroscopic quantities. Therefore, the quality of the results from classical simulations is intrinsically related to the ability of a force field to correctly describe a molecular property. The potential energy of a classical system is often expressed by a number of independent terms in an additive manner, which has been widely described in the literature in the last decades. Among the bonded terms, torsions involving two planes between four consecutive particles are responsible for the largest amplitude motions.

Small inaccuracies in the parametrization of the bonded terms may allow a molecular system to equivocally explore energetically forbidden regions of the phase space.^{2–4} As a consequence, the parametrization of torsional potentials is often revisited as an attempt to increase force field accuracy.^{5–9} All major biomolecular force fields use cosine functions to describe the energy profile of torsional dihedrals as the one shown in eq 1, where U^{tor} is the torsional potential energy; k_i is the force constant, θ is the dihedral angle, and n and δ are the multiplicity and phase shift, respectively.

$$U^{\text{tor}}(\theta; n_i, \delta_i, k_i) = \sum_i k_i [1 + \cos(n_i \theta - \delta_i)] \quad (1)$$

To allow a fast and accurate parametrization of torsional potentials, we developed, implemented, and deployed an algorithm as a Web server called Rotational Profiler. This service allows the fitting of a classical energy profile of a dihedral to a reference energy curve. A total of seven sets of parameters are generated (force constants, multiplicities, and phase shifts). Each set provides a different fitting, and the

Received: October 7, 2020

Published: November 19, 2020



resulting curves are plotted for comparison along with the reference curve. The corresponding sets can be interactively selected by the end user.

METHOD

In the search for parameters to correctly model dihedral potential function for molecules, one must be aware of the difference between the energy profile extracted from quantum calculation and simpler classical models. Very often, it is imperative to add potential functions to the classical models to rectify inaccuracies occurring due to the simplifications implied in the latter model.

This task starts with the evaluation of the function, $F(\theta)$, defined as the difference between the reference (e.g., quantum), $Q(\theta)$, and classical, $M(\theta)$, rotational profiles for a given molecule.

$$F(\theta) = Q(\theta) - M(\theta) \quad (2)$$

Because the rotational profiles evaluated through quantum and classical models are sinusoidal functions, the same can be said about $F(\theta)$. As a straightforward consequence, this function can also be defined as a summation of sinusoidal functions. Therefore, we use a set of sinusoidal profiles, $g_i(\theta; n_i, \delta_i)$, whose functional form follows that in the most widely used force field models (e.g., GROMOS, AMBER, OPLS, and CHARMM) to describe the profile $F(\theta)$.

$$g_i(\theta; n_i, \delta_i) = [1 + \cos(n_i\theta - \delta_i)] \quad (3)$$

The Rotational Profiler algorithm uses seven different sinusoidal functions, whose parameters, k_i , n_i , and δ_i , are optimized and can be chosen by the user, to describe the function $F(\theta)$.

$$F(\theta) = \sum_{i=1}^7 k_i g_i(\theta; n_i, \delta_i) + E(\theta) \quad (4)$$

The k_i parameters are constants that weigh up the contribution of each linearly independent function composing the summation. With this description, the main task of the program is to acquire the set of k_i parameters that lead to the minimum value for the error function, $E(\theta)$, i.e. that best fits the function $F(\theta)$.

Different approaches have been used to solve this problem.^{10–12} While fitting torsional potentials are not the rate limiting step of classical parametrization, global optimizers schemes add unnecessary computational cost and a level of uncertainty to the fitting. To overcome such issues, we have opted to address it through the algebraic approach of linear least-square regression, as it is a well-established methodology and allows for a fast and accurate solution as previously demonstrated by Hopkins and Roitberg.¹³ Applying it to eq 4, it is possible to calculate the set of k_i parameters that satisfy the optimization condition through the resolution of the linear system of equations exposed on eq 5.

$$\sum_{i=1}^7 k_i \sum_{\theta} g_i(\theta) g_j(\theta) = \sum_{\theta} g_j(\theta) F(\theta) \quad j \in [1, 7] \quad (5)$$

In matrix notation:

$$A \cdot K = C \quad (6)$$

$$A = [A_{ij}]_{7 \times 7} = \begin{bmatrix} A_{11} & \cdots & A_{17} \\ A_{21} & \cdots & A_{27} \\ \vdots & \ddots & \vdots \\ A_{71} & \cdots & A_{77} \end{bmatrix} \quad (7)$$

$$K = [k_i]_{7 \times 1} = \begin{bmatrix} k_1 \\ \vdots \\ k_7 \end{bmatrix} \quad (8)$$

$$C = [C_j]_{7 \times 1} = \begin{bmatrix} C_1 \\ \vdots \\ C_7 \end{bmatrix} \quad (9)$$

Where each term inside the A and C matrices are described as follows:

$$A_{ij} = \sum_{\theta} g_i(\theta) g_j(\theta) \quad (10)$$

$$C_j = \sum_{\theta} g_j(\theta) F(\theta) \quad (11)$$

To solve the matrix equation, the algorithm implemented in the Rotational Profiler code performs a matrix inversion operation on matrix A through the use of the Gauss–Jordan elimination.¹⁴ In this way, the vector of k_i parameters, which minimizes the function $E(\theta)$, is found as follows in the eq 11.

$$K = A^{-1}C \quad (12)$$

WEBSERVER USAGE AND EXAMPLES

Torsional potentials should be the last parameter to be derived, as small changes in electrostatics and van der Waals parameters may severely affect their behavior. Server usage require the user to provide a pair of files containing two columns each. The first column should list the dihedral angle values in degrees, while the second one should list the associated potential energy, in any unit, for the molecule to be parametrized. The first file should contain the reference energy profile that can be obtained by quantum mechanical calculations or experimental means. Using the molecular mechanics functional form and the associated parameters for the molecule to be parametrized, a single point energy calculation should be carried out for the same dihedral angle values to obtain the curve to be adjusted. It is recommended that the parameters for the scanned dihedral should be set to 0; otherwise, the generated parameters will have to be used in addition to the existing potentials. Because different force fields employ different 1–4 and sometimes higher-order scaling schemes, it guarantees that the generated potentials will correctly describe the dihedral within the chosen force field. Finally, it is important to ensure that the same energy unit is used in the reference and in the to be adjusted curve files. It is also recommended to set the minimum energy point to 0 in both QM and MM energy profiles and scale the remaining points accordingly, as it prevents numerical issues. Upon uploading the two files, the program will calculate the best fits. All parameters are optimized, and the resulting set (K_{ϕ_n} , m , and δ) is outputted for each one of the seven functions used to perform the linear least-square regression on the k_i coefficients (as shown in an

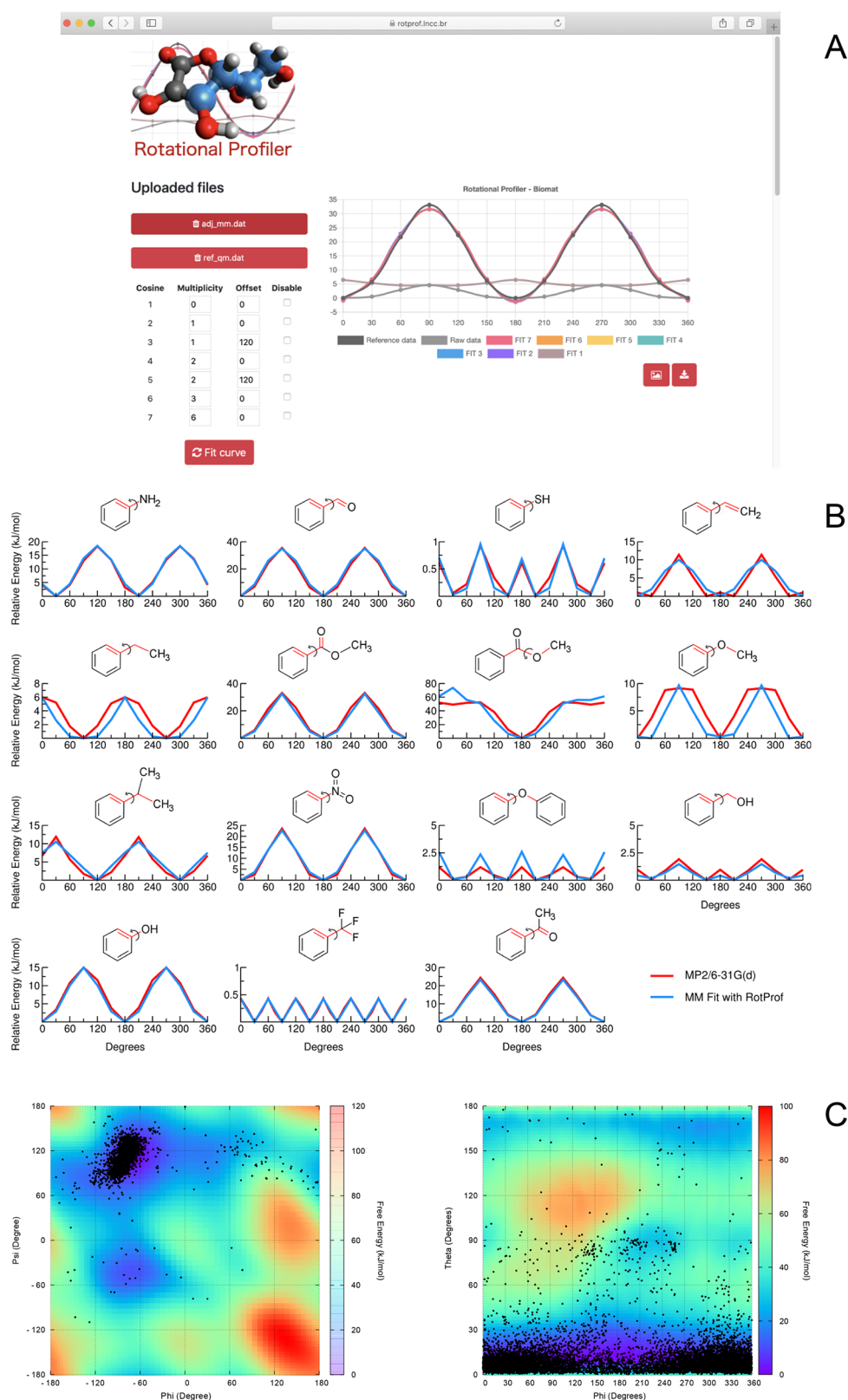


Figure 1. Results from fittings obtained by the Rotational Profiler algorithm. (A) Site page displaying the various fittings for a curve to a reference profile. (B) Classical mechanics torsional energy profiles for selected organic compounds upon fitting to their corresponding reference energy profile, calculated at MP2/6-31G(d) level of theory, as described by Poletto et al.¹⁷ (C) Superposition of the relative free energy landscape, as calculated by metadynamics, to a compilation of experimental values (black dots) for φ/ψ and θ/φ for β -D-GlcNac-(1 \rightarrow 4)- β -D-GlcNac (left panel) and D-GlcNac (α and β ; right panel), respectively. θ and φ dihedral angles are defined by Cremer–Pople coordinates.¹⁹ The torsional potentials used in this study were generated by the Rotational Profiler tool and implemented in the GROMOS 53A6_{GLYC} force field and described by Pol–Fachin et al.⁹ The free energy landscape and compilation of experimental data come from Meirelles et al.²⁰

example in Figure 1A). The user can choose the most appropriate fitting to be used, which should be added to the classical model, so that a similar description of relative conformational classical energies can be obtained (example shown in Figure 1B). It is worth noting that Rotational Profiler does not constrain the phase shifts, δ , to $0^\circ/180^\circ$. As comprehensively discussed and previously demonstrated by Hopkins and Roitberg,¹³ allowing the use of asymmetrical δ provides a closer fit to the reference curve at expense of portability, as associated potentials may not be transferable between stereoisomers.¹² Therefore, if the user is concerned about full transferability, one may consider using dihedral offset values equals to 0° or 180° . Because biomolecules are made of chiral building blocks and chirality plays a major role on their interactions, the current strategy has chosen to favor accuracy over portability. Therefore, the end user should be aware of this fact and carefully consider the appropriateness of model compound used for parametrization as well as to validate the parameters. That can be achieved by running a Monte Carlo, molecular dynamics and/or enhanced sampling simulations and checking the conformational distributions and energy barriers associated with the parametrized dihedral.

Parameterization of torsional potentials for an average biomolecule generally requires appropriate choice of model compounds that can best represent the several dihedrals in that molecule. When parametrizing multiple dihedrals in the same molecule, it is important to consider whether the behavior of these dihedrals is independent or if applying a potential along one torsion can perturb the calculated energy profile (minima and maxima points as well as rotational energy barriers) of another torsion. In addition, a dihedral can often be defined by more than one set of four atoms. Carbohydrate rings are typical examples of both cases and strategies must be devised to avoid overfitting. In these cases, one approach is to apply the calculated potentials on a single set of four atoms that do not interfere with the following plane and adjust all remaining potentials that define the specific dihedral to 0. This protocol has been used in the parametrization of carbohydrates in the GROMOS 53A6_{GLYC}.^{8,9} The quality of the force field was assessed by comparing available NMR data with the results of a series of 1 μ s unrestrained molecular dynamics and metadynamics simulations. The parametrization yielded proper description of ring pucker preferences for 16 aldohexopyranoses, the correct conformation of glycosidic linkages for 8 disaccharides and the relative conformational distributions of the hydroxymethyl group in galactose, glucose, and mannose anomers.⁹ In addition, simulations of 10 glycoproteins showed that this strategy was able to describe the glycosidic linkage preferences between amino acids and carbohydrates as well as NMR-derived protein–carbohydrate and carbohydrate–carbohydrate NOE signals.⁸

The algorithm herein presented has been previously used as a stand-alone version to parametrize hundreds of biomolecules, used in different packages and force fields.^{15–18} As expected, appropriate sampling is obtained when coupled to either deterministic or stochastic methods to generate molecular ensembles. Figure 1C showcases the conformational agreement between metadynamics calculations and experimental data for dihedrals of *N*-acetyl-glucosamines that had their potentials derived by this algorithm. The Rotational Profiler is now made widely and permanently available to the community through a portal at the Brazilian National Scientific Computing

Laboratory (LNCC) at the following address: <http://rotprof.lncc.br>.

AUTHOR INFORMATION

Corresponding Author

Roberto D. Lins – Aggeu Magalhães Institute, Oswaldo Cruz Foundation, Recife, Pernambuco 50740-465, Brazil;
orcid.org/0000-0002-3983-8025; Email: roberto.lins@cpqam.fiocruz.br

Authors

Victor H. Rusu – Swiss National Supercomputing Centre, Lugano, Ticino 6900, Switzerland
Denys E. S. Santos – Department of Fundamental Chemistry, Federal University of Pernambuco, Recife, Pernambuco 50740-640, Brazil
Marcelo D. Poletto – Department of General Biology, Federal University of Viçosa, Viçosa, Minas Gerais 36570-000, Brazil; orcid.org/0000-0001-9210-690X
Marcelo M. Galheigo – Brazilian National Scientific Computing Laboratory, Petrópolis, Rio de Janeiro 25651-075, Brazil
Antônio T. A. Gomes – Brazilian National Scientific Computing Laboratory, Petrópolis, Rio de Janeiro 25651-075, Brazil
Hugo Verli – Center for Biotechnology, Federal University of Rio Grande do Sul, Rio Grande do Sul 91500-970, Brazil; orcid.org/0000-0002-4796-8620
Thereza A. Soares – Department of Fundamental Chemistry, Federal University of Pernambuco, Recife, Pernambuco 50740-640, Brazil; orcid.org/0000-0002-5891-6906

Complete contact information is available at: <https://pubs.acs.org/10.1021/acs.jcim.0c01168>

Author Contributions

V.H.R. wrote the code and conceived the algorithm and mathematical formalism. T.A.S. and D.E.S.S. provided expertise with the equations and discussion. M.D.P. and H.V. conducted extensive testing and validation. M.M.G. and A.T.A.G. ported and deployed the code. R.D.L. conceived the idea and supervised code development. The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This work has been supported by grants from CNPq (425997/2018-9; 400738/2019-8), CAPES (88882.160170/2014-01), FACEPE (APQ-0346-2.09/19), FIOCRUZ (VPPIS-001-FIO-18-39), and INCT-FCx (465259/2014-6).

ABBREVIATIONS

MP2, Møller–Plesset perturbation theory of second order; GlcNac, *N*-acetyl-glucosamine

REFERENCES

(1) van Gunsteren, W. F.; Bakowies, D.; Baron, R.; Chandrasekhar, I.; Christen, M.; Daura, X.; Gee, P.; Geerke, D. P.; Glattli, A.; Hunenberger, P. H.; Kastenholz, M. A.; Oostenbrink, C.; Schenk, M.; Trzesniak, D.; van der Vegt, N. F.; Yu, H. B. *Biomolecular Modeling: Goals, Problems, Perspectives. Angew. Chem., Int. Ed.* **2006**, *45*, 4064–4092.

- (2) Stortz, C. A.; Johnson, G. P.; French, A. D.; Csonka, G. I. Comparison of Different Force Fields for the Study of Disaccharides. *Carbohydr. Res.* **2009**, *344*, 2217–2228.
- (3) Chandrasekhar, I.; Kastenholz, M.; Lins, R. D.; Oostenbrink, C.; Schuler, L. D.; Tieleman, D. P.; van Gunsteren, W. F. A Consistent Potential Energy Parameter Set for Lipids: Dipalmitoylphosphatidylcholine as a Benchmark of the GROMOS96 45a3 Force Field. *Eur. Biophys. J.* **2003**, *32*, 67–77.
- (4) Lins, R. D.; Hunenberger, P. H. A New Gromos Force Field for Hexopyranose-Based Carbohydrates. *J. Comput. Chem.* **2005**, *26*, 1400–1412.
- (5) Foley, B. L.; Tessier, M. B.; Woods, R. J. Carbohydrate Force Fields. *Wiley interdisciplinary reviews. Comput, Mol. Sci.* **2012**, *2*, 652–697.
- (6) Lopes, P. E. M.; Guvench, O.; MacKerell, A. D., Jr. Current Status of Protein Force Fields for Molecular Dynamics Simulations. *Methods Mol. Biol.* **2015**, *1215*, 47–71.
- (7) Paissoni, C.; Nardelli, F.; Zanella, S.; Curnis, F.; Belvisi, L.; Musco, G.; Ghitti, M. Critical Assessment of Force Fields Accuracy against Nmr Data for Cyclic Peptides Containing B-Amino Acids. *Phys. Chem. Chem. Phys.* **2018**, *20*, 15807–15816.
- (8) Pol-Fachin, L.; Verli, H.; Lins, R. D. Extension and Validation of the Gromos 53a6(Glyc) Parameter Set for Glycoproteins. *J. Comput. Chem.* **2014**, *35*, 2087–2095.
- (9) Pol-Fachin, L.; Rusu, V. H.; Verli, H.; Lins, R. D. Gromos 53a6glyc, an Improved Gromos Force Field for Hexopyranose-Based Carbohydrates. *J. Chem. Theory Comput.* **2012**, *8*, 4681–4690.
- (10) Betz, R. M.; Walker, R. C. Paramfit: Automated Optimization of Force Field Parameters for Molecular Dynamics Simulations. *J. Comput. Chem.* **2015**, *36*, 79–87.
- (11) Friederich, P.; Konrad, M.; Strunk, T.; Wenzel, W. Machine Learning of Correlated Dihedral Potentials for Atomistic Molecular Force Fields. *Sci. Rep.* **2018**, *8*, 2559.
- (12) Guvench, O.; MacKerell, A. D., Jr. Automated Conformational Energy Fitting for Force-Field Development. *J. Mol. Model.* **2008**, *14*, 667–679.
- (13) Hopkins, C. W.; Roitberg, A. E. Fitting of Dihedral Terms in Classical Force Fields as an Analytic Linear Least-Squares Problem. *J. Chem. Inf. Model.* **2014**, *54*, 1978–1986.
- (14) Aransay, J.; Divasón, J., Applications of the Gauss-Jordan Algorithm, Done Right, *arXiv*. **2014**, 1401.5910v2.
- (15) Arantes, P. R.; Pedebos, C.; Polêto, M. D.; Pol-Fachin, L.; Verli, H. The Lazy Life of Lipid-Linked Oligosaccharides in All Life Domains. *J. Chem. Inf. Model.* **2020**, *60*, 631–643.
- (16) Arantes, P. R.; Polêto, M. D.; John, E. B. O.; Pedebos, C.; Grisci, B. I.; Dorn, M.; Verli, H. Development of Gromos-Compatible Parameter Set for Simulations of Chalcones and Flavonoids. *J. Phys. Chem. B* **2019**, *123*, 994–1008.
- (17) Polêto, M. D.; Rusu, V. H.; Grisci, B. I.; Dorn, M.; Lins, R. D.; Verli, H. Aromatic Rings Commonly Used in Medicinal Chemistry: Force Fields Comparison and Interactions with Water toward the Design of New Chemical Entities. *Front. Pharmacol.* **2018**, *9*, 395.
- (18) Kirschner, K. N.; Lins, R. D.; Maass, A.; Soares, T. A. A Glycam-Based Force Field for Simulations of Lipopolysaccharide Membranes: Parametrization and Validation. *J. Chem. Theory Comput.* **2012**, *8*, 4719–4731.
- (19) Cremer, D.; Pople, J. A. General Definition of Ring Puckering Coordinates. *J. Am. Chem. Soc.* **1975**, *97*, 1354–1358.
- (20) de Meirelles, J. L.; Nepomuceno, F. C.; Peña-García, J.; Schmidt, R. R.; Pérez-Sánchez, H.; Verli, H. Current Status of Carbohydrates Information in the Protein Data Bank. *J. Chem. Inf. Model.* **2020**, *60*, 684–699.