

SOFTWARE NOTE

MonteCarbo: A software to generate and dock multifunctionalized ring molecules

Santiago Alonso-Gil 

Department of Structural and Computational Biology, Max F. Perutz Laboratories, University of Vienna, Vienna, Austria

Correspondence

Santiago Alonso-Gil, Department of Structural and Computational Biology, Max F. Perutz Laboratories, University of Vienna, Dr.-Bohr-Gasse 9, Wien 1030, Austria.
Email: santiago.alonso.gil@univie.ac.at

Funding information

H2020 Marie Skłodowska-Curie Actions, Grant/Award Number: 847548

Abstract

MonteCarbo is an open-source software to construct simple 5-, 6-, and 7-membered ring multifunctionalized monosaccharides and nucleobases and dock them into the active site of carbohydrate-active enzymes. The core bash script executes simple orders to generate the Z-matrix of the neutral molecule of interest. After that, a Fortran90 code based on a pseudo-random number generator (Monte Carlo method) is executed to assign dihedral angles to the different rotamers present in the structure (ring and rotating functional groups). The program also has a generalized internal coordinates (GIC) implementation of the Cremer and Pople puckering coordinates ring. Once the structures are generated and optimized, a second code is ready to execute in serial the docking of multiple conformers in the active site of a wide family of enzymes.

KEYWORDS

carbohydrates, conformations, docking, drug design, Monte Carlo

1 | INTRODUCTION

The symbiosis between experiments and computation is key, in the XXI century, for the development of new substrates and inhibitors in enzymology.¹⁻⁴ In glycobiology, understanding the catalytic mechanism, how the natural monosaccharides change their shape in the active site of the enzymes, is crucial to develop and design new drugs for glyco-related illnesses.⁵ Acarbose and miglitol for type 2 diabetes⁶ and swainsonine for cancer treatment⁷ are good examples of drugs acting as glycosidase inhibitors. In general, the most extensively used molecules for this purpose are derivatives of piperidine (N-based rings), thiane (S-based rings), and sulfolane (SO₂-based rings).⁸⁻¹¹

Apart from medical purposes, the design of new in silico monosaccharides is a fundamental step to predict the reaction mechanism of carbohydrate-active enzymes. First, sulfur-based derivatives usually mimic natural substrates' shapes for glycosidases, showing the conformation of the Michaelis complex of the catalyzed reaction.^{12,13} Second, gluco-, manno-, and galacto-imidazole are used as transition

state mimics for glycoside hydrolases, showing the conformation of the most energetic state along the hydrolysis.¹⁴ Finally, fluoro-derivative monosaccharides are used to activate the glycosyl-enzyme intermediate (or covalent intermediate) formation in retaining glycosidases. This information is key to decipher the conformation of the sugar moiety after the glycosylation step of the reaction.^{15,16} However, experiments show that thio-sugars do not adopt the same conformations as the natural ligand.¹⁷⁻¹⁹ In these cases, where the conformation of the mimic will not be the same as the natural ligand, it can be advantageous to quickly produce a large number of conformations for consideration.

Following the prescription reported by French and Brady in 1990,²⁰ the main limitations connected with proper modeling of carbohydrates are:

- a. *Hydrogen bonding*: saccharides are polyalcohol molecules. The oxygen atoms present in the ring and OH groups are H-bond acceptors for the alcohol groups' hydrogen atoms. A proper description

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2021 The Author. *Journal of Computational Chemistry* published by Wiley Periodicals LLC.

of the H-bond interactions is determinant for the reliability of the models.

- b. *Anomeric effects*: carbohydrates present a particular sensitivity in the asymmetric centers where a carbon atom is connected to two electronegative oxygen atoms. Depending on the medium (solvent or enzymatic environment), the bond length of the glycosidic bond, C—O bonds and torsions in the vicinity of the anomeric carbon are altered depending on the ring's conformation and several interactions between the sugar and the medium.
- c. *Multiple minimum problems*: during the geometry and energy minimization of our model (using a force field or quantum mechanics), the optimization will show a local minimum. This structure could not be a global minimum due to the many torsional degrees of freedom of a simple monosaccharide.
- d. *Comparison with experiments*: the matching between a single model and experiments with carbohydrates is not always possible. Monosaccharides present one or more conformations in solution that differ in their intramolecular interactions and may affect the results of, for instance, an NMR spectra.
- e. *Conformational analysis*: natural carbohydrates are formed by 5- and 6-membered rings. This kind of structure presents a complex conformational free energy surface. Twenty different canonical conformations can describe Furanose-like systems (10 envelopes—E—and 10 twisted—T—conformers) and 38 canonical conformations represent pyranose-like structures (2 chairs—C—, 6 boats—B—, 6 skew-boats—S—, 12 envelopes—E—, and 12 half-chairs—H).²¹ Furthermore, every rotamer (alcohol group, CH₂OH, φ , and ϕ dihedral angle in oligosaccharides—Figure 1) and its interaction with other rotamers can affect the model's stability pattern.

Improvements in computational capabilities since the publication of the French and Brady article have turned some of the limitations described in the article into simple problems solved in a matter of minutes. A recently published review by Scherbinina and Toukach shows state-of-the-art carbohydrates structure databases and computational techniques to obtain reliable models.²² Also, French and

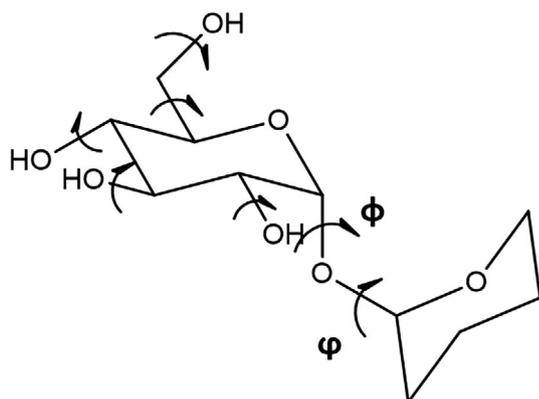


FIGURE 1 Simplified representation of a disaccharide and the rotamers present on it

Johnson reported a review about the most insightful works in the matter of modeling carbohydrates.²³

In the present work, we want to focus on applying Monte Carlo (MC) techniques²⁴ in the conformational study of carbohydrates. This method's primary basis is to generate random changes in a saccharide structure to search a possible conformation with lower energy and repeat this procedure hundreds of thousands or millions of times until ensuring a reliable sampling of the molecule's conformational space. To show some examples, in 1993, Peters and coworkers applied this technique for the conformational analysis of four disaccharides focusing on the random sampling of the exocyclic dihedral angles (CH₂OH, φ , and ϕ —Figure 1),²⁵ Dowd et al. in 2011 also add the OH groups to the Peters approach to study the opened- and closed-ring forms of carbohydrates²⁶ and in 2017, Zhang and collaborators combined MC and torsion-angle molecular dynamics simulations for oligo- and polysaccharides.²⁷

Many computational approaches were used to construct conformational free and potential energy surfaces for ring molecules.^{28–31} In this article, we present an MC-based code called *MonteCarbo*. Its principal function is to generate conformers of multi-functionalized 5-, 6-, or 7-membered ring molecules. Afterward, the program can perform docking calculations with them into the active site of several glycosidases to test their substrate/inhibitor mimic capabilities. While the previous MC-based studies addressed only the exocyclic dihedral angles, our approach increased the versatility of such methods to include the ring's puckering as a random variable. With this cheap-and-fast approximation, we firmly believe that *MonteCarbo* will become a powerful tool in the field of drug design.

2 | METHODS

2.1 | Theoretical background

2.1.1 | Cremer and Pople puckering coordinates

In 1975, Cremer and Pople developed a mathematical expression to describe the conformational space of N -membered rings.³² Given a ring structure with N atoms, $N - 3$ puckering coordinates are needed to describe the whole conformational space. For cases where N is odd, the expressions for the puckering amplitude(s) q_m and phase angle(s) ϕ_m are defined as:

$$q_m \cdot \cos\phi_m = \left(\frac{2}{N}\right)^{\frac{1}{2}} \sum_{j=1}^N z_j \cdot \cos\left(\frac{2\pi m(j-1)}{N}\right) = B_m \quad (1)$$

$$q_m \cdot \sin\phi_m = -\left(\frac{2}{N}\right)^{\frac{1}{2}} \sum_{j=1}^N z_j \cdot \sin\left(\frac{2\pi m(j-1)}{N}\right) = A_m \quad (2)$$

Applying for $m = 2, 3, \dots, (N - 1)/2$. For systems with an even N , the last puckering coordinate is defined as:

$$q_{\frac{N}{2}} = N^{-\frac{1}{2}} \sum_{j=1}^N (-1)^{j-1} \cdot z_j \quad (3)$$

where z_j are the normalized coordinates of the N atoms of the ring using their geometrical center as the origin of coordinates.

For a consistent description of the Cremer and Pople coordinates, the ring members are tagged from 1 to N . In this work, the anomeric carbon is indicated by the number 1 and the heteroatom by the number N (X in Figures 2–4). The rest of the members have to be connected consecutively. In case that the ring has no heteroatom, the selection of the first and last atoms will depend on the user's choice.

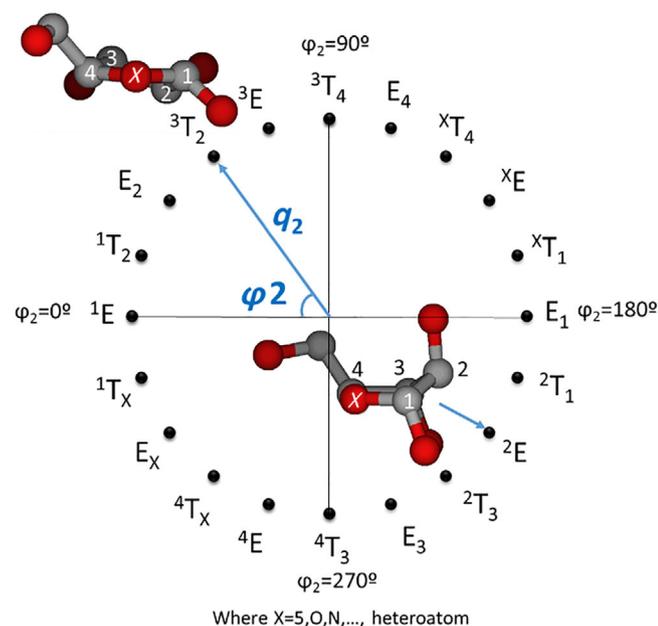


FIGURE 2 Conformational space for five-membered rings

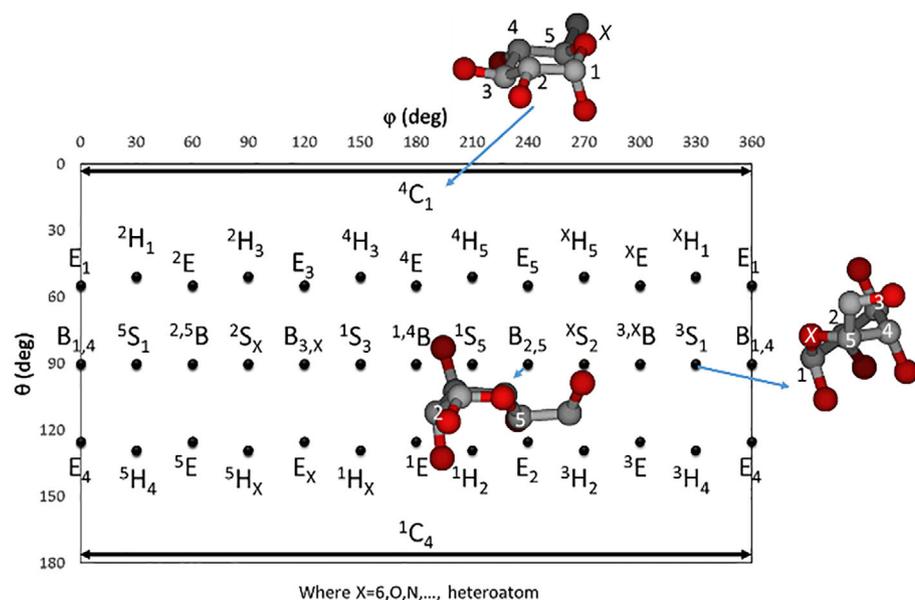


FIGURE 3 Mercator representation of the conformational space for six-membered rings

The conformational space of a 5-membered ring formed by 20 canonical conformers is described by q_2 and ϕ_2 (Figure 2). Envelops—E—have four atoms on-the-plane and one only atom above or below the plane. Twists—T—have three coplanar atoms and two consecutive atoms on opposite sides of the plane.³²The three-dimensional (3-D) conformational space of 6-membered rings is defined by q_2 , q_3 and ϕ_2 . However, the scientific community uses the Q , ϕ , and θ polar coordinates and project them into a Mercator representation (Figure 3). More details about the terminology and symbolism of the different 38 conformers are already described by IUPAC in reference 33. The complex $(q_2, q_3, \phi_2, \phi_3)$ 4-Dimensional conformational space for 7-membered rings can be simplified to a 3-D representation and divided into three (ϕ_2, ϕ_3) planes at $q_3 = 0.6$ (Twist-Chair/Chair plane), $q_3 = 0.0$ (Twist-Boat/Boat plane) and $q_3 = 0.4$ (Sofa/Twist-Sofa/Sofa-Boat).³⁴ For clarity and practical reasons, the $q_3 = 0.4$ plane is not depicted in Figure 4.

It is worth mentioning that the $q_3 = 0$ plane presents a sharp distribution (Figure 4, down) where the twist-boat (TB)—green strings—and the boat (B)—gray strings—conformations are described by a given value of ϕ_2 while a change in ϕ_3 does not affect to the structure (more details in Supporting Information).

In case the bond distances and angles are known, $N - 3$ endocyclic dihedral angles are necessary to construct and define a specific conformation for an N -membered ring. In this article's following point, this geometrical property will be used to establish a random selection of conformers.

2.2 | Simulation algorithm

2.2.1 | GICs-based puckering code for Gaussian 16

The main idea of this work is to develop an algorithm to pick a random conformation from a group of structures. This follows a three-stepped

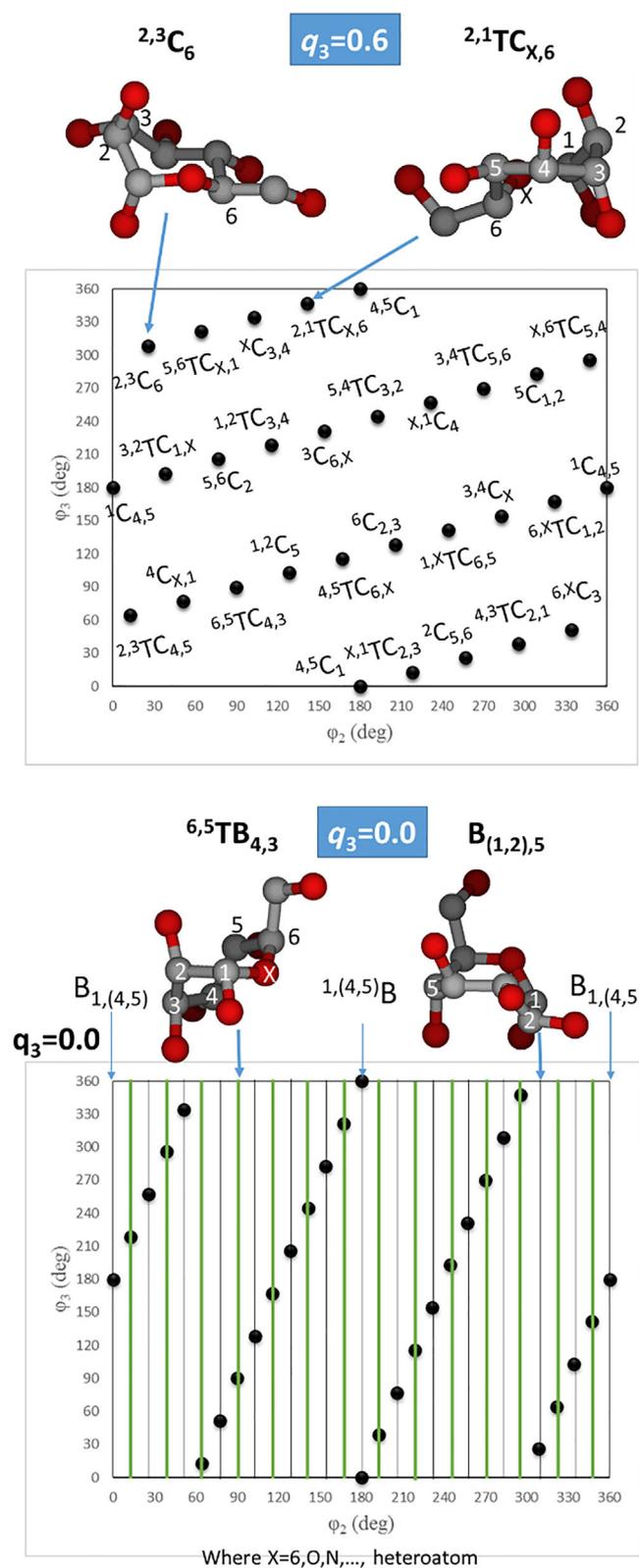


FIGURE 4 Conformational subspaces for seven-membered rings at $q_3 = 0.6$ (up) and $q_3 = 0.0$ (down)

pathway: generating the conformers by changing their puckering coordinates, creating a database with the endocyclic structural information, and selecting randomly the array of endocyclic dihedral angles corresponding to a unique conformation. For the first step, we

present a strategy based on performing scan calculations with the Gaussian 16 software.³⁵ The main initial problem was that Cremer and Pople's mathematical expressions are not implemented as generalized internal coordinates (GICs) in the quantum mechanics code. However, the last version of Gaussian includes adding and defining homemade GICs using the most common mathematical operators.

Following the recipe described in reference 31 and starting from the x , y and z (Cartesian) coordinates of the N atoms of the ring, the code calculates the center of geometry (XCntr, YCntr, and ZCntr functions). It recalculates the N atoms' new coordinates using the center of geometry as the origin of coordinates. After that, employing simple mathematical operators and the function *SQRT*, the code obtains the values of A_m , B_m , and $q_{N/2}$ (for $N = 6$) defined by Equations (1)–(3).

At the time to get the values of q_m and ϕ_m for $N = 5$ and $N = 7$ and Q , ϕ , and θ for $N = 6$, one technical problem appeared: Gaussian does not have the function *arctan* defined in its code and its necessary to use it to get the puckering phases:

$$q_m = \sqrt{A_m^2 + B_m^2} \quad (4)$$

$$\phi_m = \arctan\left(\frac{A_m}{B_m}\right) \quad (5)$$

By a trigonometric relationship between the *arctan* and the *arccos* functions, we can transform the Equation (5) into:

$$\phi_m = \arctan\left(\frac{A_m}{B_m}\right) = \arccos\left(\frac{B_m}{\sqrt{A_m^2 + B_m^2}}\right) \quad (6)$$

However, this conversion only defines the interval $[0, \pi]$ of ϕ_m . To solve this problem, we observed that the function A_m is antisymmetrical at $\phi_m = 0$. So, to define ϕ_m in the interval $[0, 2\pi]$, we use the following expression:

$$\phi_m = \frac{A_m}{\sqrt{A_m^2 + \varepsilon}} \arccos\left(\frac{B_m}{\sqrt{A_m^2 + B_m^2}}\right) + 180 \quad (7)$$

where $\varepsilon = 10^{-6}$ avoids a division by zero when $A_m = 0$.

In the case of $N = 6$, the polar coordinates are calculated as follows:

$$Q = \sqrt{q_2^2 + q_3^2} > 0 \quad (8)$$

$$\phi_m = \frac{A_m}{\sqrt{A_m^2 + \varepsilon}} \arccos\left(\frac{B_m}{q_2}\right) + 180 \in [0, 2\pi] \quad (9)$$

$$\theta = \arccos\left(\frac{q_3}{\sqrt{q_2^2 + q_3^2}}\right) \in [0, \pi] \quad (10)$$

One of the limitations of the mathematical interface of Gaussian to define GICs is the absence of periodicity. For this reason, the phase

puckering coordinates present problems when their values are close to 0 or 2π . Although we could be able to explore around 99% of the conformational potential energy surfaces of 5-, 6-, and 7-membered rings. The reader can find these codes in the files *puckN.gic* (for $N = 5, 6$, and 7) of the *MonteCarbo* distribution.

2.2.2 | Quantum-mechanical calculations: obtaining the structural parameters for MonteCarbo libraries

All the quantum-mechanics calculations were performed with Gaussian 16³⁵ employing Density Functional Theory.³⁶ By ensuring a proper equilibrium between a good H-bond description, accuracy, and time-economy,³⁷ the calculations were performed with the B97-2 functional³⁸ and the def2-SVP basis set.³⁹ Scan optimizations were performed using the *tight* criteria for convergence.

For collecting distances, angles and, more importantly, endocyclic dihedral angles, scan calculations of saturated and nonsubstituted 5-, 6-, and 7-membered rings were performed using the previously GICs-based implementation of the puckering coordinates. The chosen molecules had the $C_NH_{2N}X$ formula where $N = 5, 6$, and 7 and $X = CH_2, O, S, SO_2, NH, BH$, and PH (Figure 5, more details in Figures S1–S21). The evolution of the endocyclic dihedral angles is available in the *N-x-H-y-X-D#.txt* files of the *MonteCarbo* distribution ($x = 5, 6, 7$; $y = 0, 1$; $X = Oxy, S, SO_2, N, B, P$, and $\# = 1\dots, N - 3$). For instance, the cyclopentane information is in the *N-5-H-0-D1.txt* and *N-5-H-0-D2.txt* files.

Also, for the proper construction of functional groups attached to the rings, several molecular models formed by a ring and the substitute groups were optimized (list of available functional groups and structural details in Supporting Information).

The distances, angles, and exocyclic dihedral angles are available in the files with extension **.var* of the *MonteCarbo* distribution.

2.2.3 | Random number generator

MonteCarbo generates a z-matrix Gaussian input with the structural information of the molecule of study. However, there are at least $N - 3$ endocyclic dihedral angles that will change from one structure

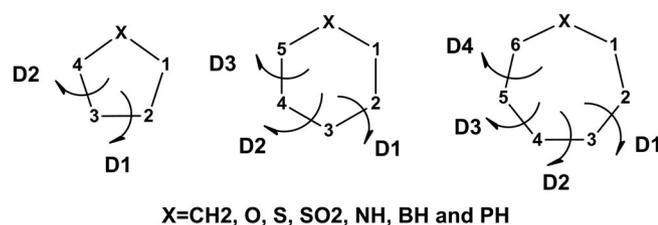


FIGURE 5 Schematic representation of the studied rings and the definition of the main endocyclic dihedral angles $D\#$, where $\# = 1\dots, N - 3$

to another. If the molecule contains rotamers like $-OH, -CH_2OH$, and so on, the number of random degrees of freedom increases.

Inspired by Vilaseca and coworkers' work,⁴⁰ we developed a *Fortran 90* code based on a pseudo-random number generator. This code generates $N - 3$ pseudo-random numbers that take $N - 3$ endocyclic dihedral angles. After that, it generates one pseudo-random number in the range of $[-180^\circ, 180^\circ]$ per rotamer. Finally, the *MonteCarbo* script implements the given dihedral angles into the file where the variables of the z-matrix are defined. In addition to that, for rings with NH and PH, the equatorial/axial position of the H-N/H-P hydrogen requires an extra random number that can be -120° or 120° .

The *rangen*.f90* files are available and open-source, ready for changes, and corrections in the *MonteCarbo* folder.

2.2.4 | MonteCarbo script

MonteCarbo is a bash script that generates a z-matrix Gaussian input model of a neutral, multi-functionalized (5, 6, or 7)-membered ring molecule. The script requires some information as an input to construct the model and the random replicas (Figure 6).

The script combines the information and power of the *N-x-H-y-X-D#.txt*, *rangen*.f90* and *puckN.gic* files to develop a *calc.gjf* input file. Depending on the user's selection, the final output of the required structures can be a Gaussian input, a PDB or an XYZ file (Open Babel⁴¹ is required for the conversions).

The main limitation of *MonteCarbo* is that in extreme multi-functionalization cases with voluminous groups, the code generates structures with steric hindrance or overlapping. For instance, in case our molecule presents two neighbor CH_2OH groups, the generation of random conformers of it will lead into some structures where the OH groups overlap or cross the same point in the space.

The code is free to download in <https://github.com/drsalonsogil/montecarbo> and a *README* file is available with further information.

2.2.5 | MCdock: testing the substrate/inhibitor role of the monosaccharide in glycosidases

MCdock is another bash script that prepares the generated and/or optimized structures¹ by *MonteCarbo* to be docked in the active site of a wide group of glycoside hydrolases. For that purpose, the program requires the execution of the *prepare_ligand* code of AutoDock 4.2. package⁴² (present in MGLTools) and the *vina* code of the AutoDock Vina package.⁴³ The workflow of *MCdock* is shown in Figure 7. The repository contains a *receptors* folder where several families of glycosidases are classified depending on the name of the sugar that hydrolyses: ARABINO-FRUCTO, FUCO, GALACTO, GALNAC-GLCNAC, GLUCO, IDURONIC-SIALIC, MANNO, RHAMNO, and XYLO. The user is free to create new folders and generate the *pdbqt* and configuration files for the receptor of interest. Otherwise, AutoDock Vina will not find a receptor or the coordinates where the ligand has to be docked.

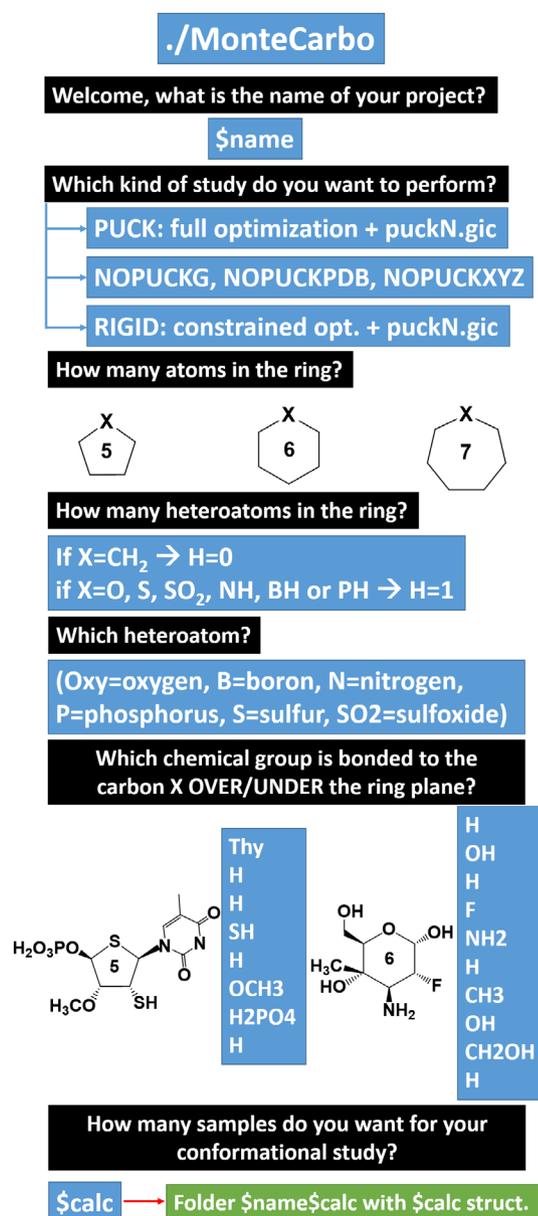


FIGURE 6 Workflow for the execution of the *MonteCarbo* script. The authors strongly recommend using an *input.dat* file to perform parallel jobs and increase the projects' efficiency (more details in the README file). Depending on the user's will, *MonteCarbo* will generate \$calc Gaussian Z-matrix input, \$calc PDB or \$calc XYZ coordinates files (\$calc is the number of conformers requested by the user)

3 | APPLICATIONS

Following the work-flows of Figures 6 and 7, we have executed *MonteCarbo* in the generation of three ring molecules: 2-hydroxy-tetrahydrofuran, α -D-glucose and α -D-glycero-D-idoseptanoside. In the last example, after an initial geometrical optimization, we have also applied *MCdock* to analyze the binding properties between the 7-membered ring sugar and mannosidases (details about input files and geometries are on pages S88–S100 of Figures S22–S26).

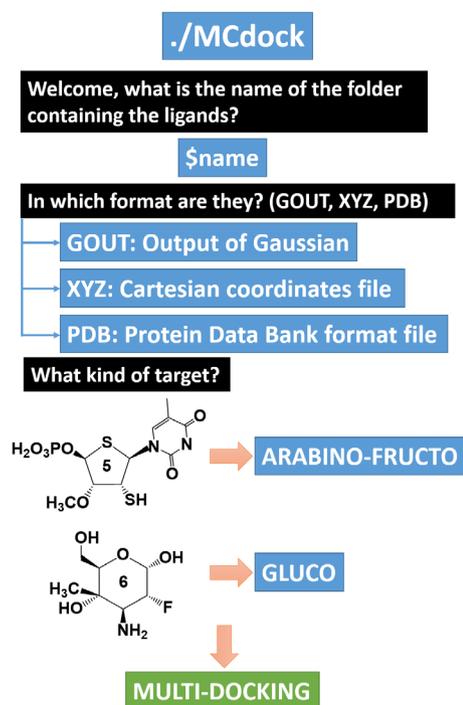


FIGURE 7 Workflow for the execution of the *MCdock* script (more details in the README file). This process' output(s) will be the output obtained by the multiple AutoDock Vina calculations⁴³

3.1 | Simple case: 2-hydroxy-tetrahydrofuran

After generating 500 conformers of 2-OH-C₄H₇O using *MonteCarbo* and representing the overlap between the different obtained structures with PyMOL² the result is shown in Figure 8. Using a simple case, we can easily observe how the code chooses between different conformers of a 5-membered ring and the different orientations of the hydrogen atom of the hydroxyl group present in the molecule. Furthermore, due to the conversion from z-matrix to Cartesian coordinates, the structure's first carbon is always in the origin of coordinates. The second atom is still at the same distance (and at the same position). The other ring members form a continuous rainbow due to the proximity between the structures over the conformational energy surface.

3.2 | Test case: α -D-glucose

As performed with the previous simple case, 500 conformers of α -D-glucose were generated executing *MonteCarbo* and the resulting overlap is represented in Figure 9. Compared with the previous structure, we observe a hydroxyl group whose oxygen remains in the same position in the center of the image. Its hydrogen position takes a continuous of positions due to the random assignment of the H—O—C—C dihedral angle between -180° and 180° by the pseudo-random number generator.

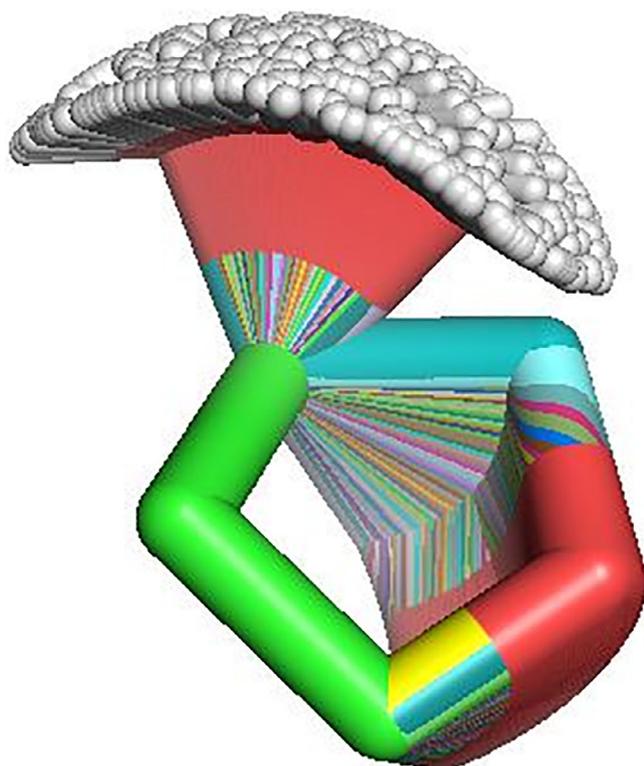


FIGURE 8 Overlap of 500 structures of the 2-hydroxy-tetrahydrofuran molecule generated by *MonteCarbo*

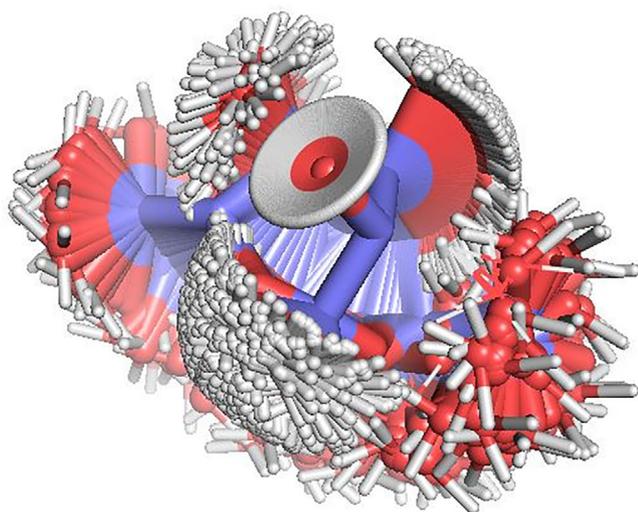


FIGURE 9 Overlap of 500 structures of α -D-glucose generated by *MonteCarbo*

3.3 | Docking: a 7-membered ring mimics mannose

Inspired by the work of Pecuh, Ernst et al. where a crystal structure of mannose-specific bacterial lectin FimH in complex with a septanoside is reported,⁴⁴ 50 structures of α -D-glycero-D-idoseptanoside were generated and minimized (B97-2/def2-SVP level) in the gas-phase, using tight criteria. Frequencies were

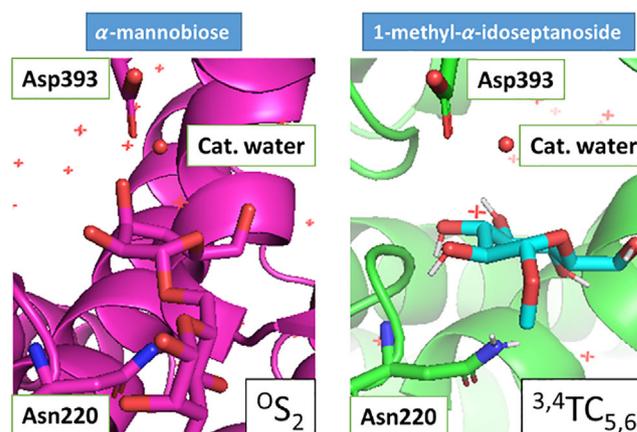


FIGURE 10 Representation of the active site of the GH125 α -mannosidase in presence of the natural substrate (α -mannobiose, PDB 5M7I18) and 1-methyl- α -D-glycerol-D-idoseptanoside (docking calculation)

calculated to ensure the structures were located in minima over the conformational energy surface.

In the experiment with the septanoside, a ${}^{3,4}TC_{5,6}$ conformation is observed (PDB 5CGB), while the experiment with mannose showed a 4C_1 conformation (PDB 4BUQ).⁴⁵ Furthermore, a computational analysis of the 1-hydroxymethyl- α -D-glycero-D-idoseptanoside shows the ${}^{3,4}TC_{5,6}$ conformation as the most stable conformation.⁴⁶ Also, after analyzing the 50 structures of the hydrolyzed α -D-glycero-D-idoseptanoside, the most stable conformation is the ${}^{3,4}TC_{5,6}$ (more details in Supporting Information). Then, we can conclude that the FimH enzyme recognizes the sugar in the most stable conformation without changing its shape. But what happens if we try to dock the septanoside in a mannosidase?

Using *MCdock*, we docked a ${}^{3,4}TC_{5,6}$ structure of hydrolyzed α -D-glycero-D-idoseptanoside, 1-methyl- α -D-glycero-D-idoseptanoside, and dimethyl-1,X- α -D-glycero-D-idoseptanoside, for X = 2, 3, 4, 5, and 7, in the active site of nine mannosidases (six α - and three β -mannosidases). The OCH₃ groups were included to avoid nonexisting H-bond interactions in presence of an oligosaccharide. In Figure 10, comparing the structure of the GH125 glycosidase in the presence of the natural substrate and the septanoside indicates the mimic power of the artificial sugar. It correlates the OS_2 conformation of the 6-membered rings with the ${}^{3,4}TC_{5,6}$ of the 7-membered rings. The full report about the docking calculations is available in Supporting Information. Those calculations found four suitable GH-septanoside complexes where the 7-membered ring in a ${}^{3,4}TC_{5,6}$ conformation mimics the 6-membered ring of the natural substrates in an OS_2 conformation (GH38, GH76, GH92, and GH125, Figure S27).

4 | CONCLUSIONS

MonteCarbo is an easy-to-use computation-friendly software able to model and to dock multi-functionalized monosaccharides. Being an

open-source package, the code-user interaction is permitted, and, depending on the circumstances of the research, *MonteCarbo* and *MCdock* can change and evolve in the proper direction.

We have demonstrated the power of the provided codes in terms of quick-and-cheap structure generation and the relevance of the obtained results testing new substrates and inhibitors for carbohydrate-active enzymes.

As a limitation, the program does not have any internal mechanism to decipher if a structure will be physically reliable or to avoid chemical changes during the optimization process. These processes require a postanalysis to confirm and delete incorrect configurations.

ACKNOWLEDGMENTS

Special thanks to Vendula Brozkova, the creative mind behind the name of the program. Also, to Prof. Bojan Zagrovic for his support and scientific orientation. Santiago Alonso-Gil acknowledges the VIP² fellowship 2020 funded by the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No. 847548.

ORCID

Santiago Alonso-Gil  <https://orcid.org/0000-0001-7596-8161>

ENDNOTES

- ¹ The scientific community strongly recommend that docking calculations are performed with optimized ligands. *MonteCarbo* has the RIGID option keeping the conformation fixed during the optimization, in case the conformation of interest is not a minimum in the conformational potential energy surface.
- ² The PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC.

REFERENCES

- [1] Y. Zhang, A. N. A. Aryee, B. K. Simpson, *Curr. Opin. Food Sci.* **2020**, 31, 63.
- [2] B. Da'adoosh, K. Kaito, K. Miyashita, M. Sakaguchi, A. Goldblum, *PLoS Comput. Biol.* **2020**, 16, 3.
- [3] I. Abdelli, F. Massani, S. B. Brikci, S. Ghalem, *J. Biomol. Struct. Dyn.* **2020**, 1. <https://www.tandfonline.com/doi/full/10.1080/07391102.2020.1763199>.
- [4] M. W. van der Kamp, A. J. Mulholland, *Nat. Prod. Rep.* **2008**, 25, 1001.
- [5] A. P. Montgomery, K. Xiao, X. Wang, D. Skropeta, H. Yu, in *Advances in Protein Chemistry and Structural Biology*, Vol. 109 (Ed: T. Karabencheva-Christova), Elsevier, Amsterdam **2017**, p. 25.
- [6] F. A. van de Laar, P. L. B. J. Lucassen, K. P. Akkermans, E. H. van de Lisdonk, G. E. H. M. Rutten, C. van Weel, *Cochrane Database Syst. Rev.* **2005**, 2, 1465.
- [7] T. M. Wrodnigg, A. J. Steiner, B. J. Ueberbacher, *Anti-Cancer Agents Med. Chem.* **2008**, 8, 77.
- [8] A. Wadood, M. Ghufuran, A. Khan, S. S. Azam, M. Jelani, R. Uddin, *Int. J. Biol. Macromol.* **2018**, 111, 82.
- [9] A. Ueda, J. Pi, Y. Makura, M. Tanaka, J. Uenishi, *RSC Adv.* **2020**, 10, 9730.
- [10] V. Ulgar, J. G. Fernández-Bolaños, M. Bols, *J. Chem. Soc., Perkin Trans.* **2002**, 1, 1242.
- [11] I. Dragutan, V. Dragutan, A. Démonceau, *RSC Adv.* **2012**, 2, 719.
- [12] A. J. Thompson, J. Dabin, J. Iglesias-Fernandez, A. Ardevol, Z. Dinev, S. J. Williams, O. Bande, A. Siriwardena, C. Moreland, T. Hu, D. K. Smith, H. J. Gilbert, C. Rovira, G. J. Davies, *Angew. Chem. Int. Ed.* **2012**, 54, 5378.
- [13] J.-Y. Zou, G. J. Kleywegt, J. Stahlberg, H. Driguez, W. Nerinckx, M. Claeysens, A. Koivula, T. T. Teeri, T. A. Jones, *Structure.* **1999**, 7, 1035.
- [14] S. Alonso-Gil, *J. Carbohydrate Chem.* **2020**, 39, 175.
- [15] K. Syson, C. E. M. Stevenson, A. M. Rashid, G. Saalbach, M. Tang, A. Tuukkanen, D. I. Svergun, S. G. Withers, D. M. Lawson, S. Bornemann, *Biochemistry.* **2014**, 53, 2494.
- [16] T. M. Gloster, S. Roberts, V. M. Ducros, G. Perugino, M. Rossi, R. Hoos, M. Moracci, A. Vasella, G. J. Davies, *Biochemistry.* **2004**, 43, 6101.
- [17] K. J. Gregg, W. F. Zandberg, J. H. Hehemann, G. E. Whitworth, L. Deng, D. J. Vocadlo, A. B. Boraston, *J. Biol. Chem.* **2011**, 286, 15586.
- [18] S. Alonso-Gil, A. Males, P. Z. Fernandes, S. J. Williams, G. J. Davies, C. Rovira, *J. Am. Chem. Soc.* **2017**, 139, 1085.
- [19] Y. Zhu, M. D. L. Suits, A. Thompson, S. Chavan, Z. Dinev, C. Dumon, N. Smith, K. W. Moremen, Y. Xiang, A. Siriwardena, S. J. Williams, H. J. Gilbert, G. J. Davies, *Nat. Chem. Biol.* **2010**, 6, 125.
- [20] A. D. French, J. W. Brady, *Comp. Model. Carbohydrate Mol. ACS Symp. Series*; American Chemical Society, Washington, DC, **1990**; Chapter 1, pp. 1–19.
- [21] R. Rao, *Conformation of Carbohydrates*, Taylor & Francis, Ed **1998**.
- [22] S. I. Scherbinina, P. V. Toukach, *Int. J. Mol. Sci.* **2020**, 21, 7702.
- [23] A. D. French, G. P. Johnson, in *The Plant Cell Wall. Methods in Molecular Biology*, Vol. 2149 (Ed: Z. Popper), Humana, New York **2020**, p. 513.
- [24] N. Metropolis, S. Ulam, *J. Am. Stat. Assoc.* **1949**, 44, 335.
- [25] T. Peters, B. Meyer, R. Stuike-Prill, R. Somorjai, J.-R. Brisson, *Carbohydr. Res.* **1993**, 238, 49.
- [26] M. K. Dowd, D. E. Kiely, J. Zhang, *Carbohydr. Res.* **2011**, 346, 1140.
- [27] W. Zhang, S. C. Howell, D. W. Wright, A. Heindel, X. Qiu, J. Chen, J. E. Curtis, *J. Mol. Graph. Modell.* **2017**, 73, 179.
- [28] H. B. Mayes, L. J. Broadbelt, G. T. Beckham, *J. Am. Chem. Soc.* **2014**, 136, 1008.
- [29] M. Sega, E. Autieri, F. Pederiva, *J. Chem. Phys.* **2009**, 130, 225102.
- [30] G. J. Davies, A. Planas, C. Rovira, *Acc. Chem. Res.* **2012**, 45, 308.
- [31] L. Wang, B. J. Berne, *J. Chem. Phys.* **2018**, 149, 072306.
- [32] D. Cremer, J. A. Pople, *J. Am. Chem. Soc.* **1975**, 97, 1354.
- [33] IUPAC and IUBMB, *Pure & Appl. Chem.* **1996**, 68, 1919.
- [34] I. K. Boessenkool, J. C. A. Boeyens, *J. Cryst. Mol. Struct.* **1980**, 10, 11.
- [35] M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, G. A. Petersson, H. Nakatsuji, X. Li, M. Caricato, A. V. Marenich, J. Bloino, B. G. Janesko, R. Gomperts, B. Mennucci, H. P. Hratchian, J. V. Ortiz, A. F. Izmaylov, J. L. Sonnenberg, D. Williams-Young, F. Ding, F. Lipparini, F. Egidi, J. Goings, B. Peng, A. Petrone, T. Henderson, D. Ranasinghe, V. G. Zakrzewski, J. Gao, N. Rega, G. Zheng, W. Liang, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, K. Throssell, J. A. Montgomery Jr., J. E. Peralta, F. Ogliaro, M. J. Bearpark, J. J. Heyd, E. N. Brothers, K. N. Kudin, V. N. Staroverov, T. A. Keith, R. Kobayashi, J. Normand, K. Raghavachari, A. P. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, J. M. Millam, M. Klene, C. Adamo, R. Cammi, J. W. Ochterski, R. L. Martin, K. Morokuma, O. Farkas, J. B. Foresman, D. J. Fox, *Gaussian 16, Revision C.01*, Gaussian, Inc, Wallingford, CT **2016**.
- [36] a) H. Hohenberg, W. Kohn. *Phys. Rev.* **1964**, 136, 3B, B864; b) W. Kohn, L. J. Sham. *Phys. Rev.* **1965**, 140, 4A, A1133.
- [37] L. Rao, H. Ke, G. Fu, X. X. Y. Yan, *J. Chem. Theory Comput.* **2009**, 5, 86.
- [38] P. J. Wilson, T. J. Bradley, D. J. Tozer, *J. Chem. Phys.* **2001**, 115, 9233.
- [39] F. Weigend, R. Ahlrichs, *Phys. Chem. Chem. Phys.* **2005**, 7, 3297.

- [40] E. Vilaseca, A. Isvoran, S. Madurga, I. Pastor, J. L. Garcés, F. Mas, *Phys. Chem. Chem. Phys.* **2011**, *13*, 7396.
- [41] N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, G. R. Hutchison, *Aust. J. Chem.* **2011**, *3*, 33 The Open Babel Package, Version 2.3.1 <http://openbabel.org>.
- [42] G. M. Morris, R. Huey, W. Lindstrom, M. F. Sanner, R. K. Belew, D. S. Goodsell, A. J. Olson, *J. Comp. Chem.* **2009**, *16*, 2785.
- [43] O. Trott, A. J. Olson, *J. Comp. Chem.* **2010**, *31*, 455.
- [44] C. P. Sager, B. Fiege, P. Zihlmann, R. Vannam, S. Rabbani, R. P. Jakob, R. C. Preston, A. Zalewski, T. Maier, M. W. Pecuh, B. Ernst, *Chem. Sci.* **2018**, *9*, 646.
- [45] G. Roos, A. Wellens, M. Touaibia, N. Yamakawa, P. Geerlings, R. Roy, L. Wyns, J. Bouckaert, *ACS Med. Chem. Lett.* **2013**, *4*, 1085.
- [46] M. P. DeMatteo, N. L. Snyder, M. Morton, D. M. Baldisseri, C. M. Hadad, M. W. Pecuh, *J. Org. Chem.* **2005**, *70*, 24.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: S. Alonso-Gil, *J Comput Chem* **2021**, *42*(21), 1526. <https://doi.org/10.1002/jcc.26559>