Automated Search For The Low-lying Energy Isomers of Rhamnolipids and Related Organometallic Complexes

Leticia Dobler,[a,b,c] and Ricardo R. Oliveira*[c]

February, 2022

Rhamnolipids (RMLs) are a widely studied biosurfactant due to its high biodegradability, low toxicity, and environmentally friendly production, which has the bacterium Pseudomonas aeruginosa as producer and renewable source of feedstock. However, the knowledge of the structure-property relationship of several RMLs congeners is imperative for the design of a high efficient application. Aiming to a better understanding of RMLs at a molecular level, we performed an automated search for low energy structures of the most abundant congeners, namely, Rha-C10, Rha-C10-C10, Rha-Rha-C10 and Rha-Rha-C10-C10 and their respective congener with two carbon atoms (C2) at the side chain. Besides that, selected neutral metal complexes were also considered. We thus performed a metadynamics search followed by DFT optimizations of selected geometries. In addition, molecular dynamics simulations were also applied. Our results show a plethora of low energy structures for each congener. Although several of them have internal hydrogen bonds, this interaction alone cannot explain the stability. Moreover, geometries in closed conformation were always more stable than "open" ones and the effect of chain length on the geometry was found to be more prominent between Rha-Rha-C2-C2 and Rha-Rha-C10-C10. Finally, the energy differences between open and closed conformations for K+, Ni2+, Cu2+ and Zn2+ complexes were found to be 23.5 kcal mol−1, 62.8 kcal mol−1, 24.3 kcal mol−1 and 41.6 kcal mol−1, respectively, indicating a huge structural reorganization after the complex formation.

1 Introduction

Biosurfactant is a widely studied class of surfactants with great technological potential exhibiting interesting properties such as biodegradability and low toxicity.[10] Among all biosurfactants, the most studied ones are the rhamnolipids (RMLs),[3] a group of molecules which are produced on an industrial scale using the gram-negative bacterium Pseudomonas aeruginosa.[11] These molecules can be separated in four major subgroups: (i) mono-rhamno-mono-lipid (Rha-Cm), being composed by one rhamnose (Rha) molecule and β-hydroxyalkanoyl-β-hydroxyalkanoic acid; (ii) mono-rhamno-di-lipid (Rha-Cn-Cn), composed by one rhamnose and two β-hydroxyalkanoyl-β-hydroxyalkanoic acid; (iii) di-rhamno-mono-lipid (Rha-Rha–Cm), composed by two rhamnoses and one β-hydroxyalkanoyl-β-hydroxyalkanoic acid; and (iv) di-rhamno-di-lipid (Rha-Rha-Cn-Cn), composed by two rhamnoses and two β-hydroxyalkanoyl-β-hydroxyalkanoic acid (Figure 1). In general, RML production by Pseudomonas generates a complex mixture of congeners that strongly depends on the strain, medium composition, temperature, oxygenation and time of cultivation.[11] Mono-rhamnolipids are especially studied, due its higher surface activity and effective diffusion coefficient.[12]

Due its high surface activity and low toxicity,[13] RMLs have been applied in different fields such as food industry,[13] cosmetics,[14] pharmaceutical industry,[15] and environmental studies. Besides that, RMLs have been proposed as good mediator in oil spills,[16] in oil remediation,[17] as flushing agents for organic pollution[18] and for heavy metals and rare earth elements[19] remediation in water[20] and soils.[21,22]

Contamination of soils and water by heavy metals is an environmental issue causing long-term problems to living beings.[23] It is well known that the removal efficiency of heavy metals by RMLs depends on pH, concentration, exposure time
and soil composition, and it is possible that others conditions could affect the removal activity. \textsuperscript{[14]} This complexity hamper the comparisons between results in different works, and the exact definition of the best remediation condition is complicated to obtain.\textsuperscript{[15]} Besides that, remediation efficiency do not follow the Ochoa-Loza’s stabilization series\textsuperscript{[16]} indicating that a heavy metal removal mechanism is more complicated than complex formation.

Although \textit{in silico} investigations have limitations due to the great complexity that an \textit{in situ} environment has, it is possible to elucidate possible structures of RML aggregates, micelles or metal complexes. Because of the high complexity and computational cost, there are only a few works exploring the structure of RML systems, the most of them using force field methods.

By the means of classical molecular dynamics, Munusamy and co-workers performed an investigation of RMLs aggregates formation at the interface and inside the solution bulk.\textsuperscript{[17, 18]} Their findings suggest that RMLs form aggregates from 7 to 95 molecules.\textsuperscript{[19]} The same group also conclude that the presence of a second rhamnose do not interfere in the aggregation formation when those occur inside the bulk, but it has influence when the process is located at the interface.\textsuperscript{[20]}

It was also demonstrated that nonionic and anionic aggregates present different stability and geometries over the water surface.\textsuperscript{[21, 22]} Moreover, it has been shown that Rha-C\textsubscript{10}-C\textsubscript{10} aggregates are stable in the presence of long chain alkanes in water and can be used to remove hydrocarbons from aqueous solutions.\textsuperscript{[23, 24]} Besides that, dirhamnolipid can generate bigger molecular aggregates than monorhamnolipid in n-alkanes solutions.\textsuperscript{[25, 26]} Unfortunately, no work on molecular dynamics simulations have reported metal complexes with RMLs.

As far as we know, there is only one work about RML molecules using quantum chemistry methods. Kornii and co-workers\textsuperscript{[27]} performed numerical analyses of the geometric parameters and electronic structures of RML complexes based on PM7 semi-empirical method and density functional theory (DFT) applying B3LYP hybrid functional.\textsuperscript{[28]} The analysis focused on calcium, zinc, aluminum, and copper atoms.

In order to investigate the structure of different RMLs congeners and selected metal complexes at quantum chemistry level, a series of computations were performed with the extended tight binding (xtb) method\textsuperscript{[29]} combined with density functional theory (DFT) and polarizable continuum method (PCM) to simulate water as solvent. The selected congeners are Rha-C\textsubscript{10}, Rha-C\textsubscript{10}-C\textsubscript{10}, Rha-Rha-C\textsubscript{10} and Rha-Rha-C\textsubscript{10}-C\textsubscript{10} because they are the most abundant products by \textit{Pseudomonas aeruginosa}.\textsuperscript{[31]}

![Figure 1](image)

**Figure 1** The main four subgroups of rhamnolipids: mono-rhamno-monolipids (Rha-C\textsubscript{m}), mono-rhamno-dilipids (Rha-C\textsubscript{m}-C\textsubscript{n}), di-rhamno-monolipids (Rha-Rha-C\textsubscript{m}) and di-rhamno-dilipids (Rha-Rha-C\textsubscript{n}-C\textsubscript{m}).

### 2 Computational Details

Initial conformations for Rha-C\textsubscript{2}, Rha-C\textsubscript{2}-C\textsubscript{2}, Rha-Rha-C\textsubscript{2} and Rha-Rha-C\textsubscript{2}-C\textsubscript{2} were obtained from geometry optimizations applying the extended tight binding (xtb) method\textsuperscript{[32]} at GFN2-xtB level.\textsuperscript{[33, 34]} Initially, only chains with length of 2 carbon atoms (C\textsubscript{2}) were considered due to the computational limitations. In order to obtain the low energy conformers, Root-Mean-Square-Deviation (RMSD) based metadynamics\textsuperscript{[35]} were performed using the CREST (Conformer-Rotamer Ensemble Sampling Tool) searching approach\textsuperscript{[36]} starting from the optimized geometries at same level of theory. The global optimization was performed at 298.15 K applying the Generalized Born model with the solvent-accessible Surface Area (GBSA) for water as solvent.\textsuperscript{[37]} The energy threshold for conformational search was set to 60 kcal mol\textsuperscript{-1}. In the final ensemble, the most representative structures were selected by a combination of principle component analysis (PCA) and k-means clustering algorithm with dihedral angles as descriptors.\textsuperscript{[38, 39]}

From the final ensemble, structures were selected avoiding an extensive analysis throughout rotational isomers (rotamers). To this end, Density Functional Theory (DFT) geometry optimizations and hessian computations were performed using the composite method based on the r\textsuperscript{2}SCAN-3c (“Swiss army knife”) functional.\textsuperscript{[40]} Only the first 50 structures obtained after the clustering procedure were considered except for Rha-Rha-C\textsubscript{2}-C\textsubscript{2}. In this case, the first 100 structures were used in order to obtained a greater structural diversity. Furthermore, all minimum energy structures were arranged in ascending standard enthalpy order for $T = 298.15$ K ($H_0^{298}$). All DFT and xtb computations were performed using ORCA 5.1\textsuperscript{[41]} and xTB 6.4.
2 COMPUTATIONAL DETAILS

Figure 2 Low energy molecular structures of Rha-C2 (top panel), Rha-C2-C2 (middle panel) and Rha-Rha-C2 (bottom panel) obtained at r²SCAN-3c level of theory. Relative enthalpies (kcal mol⁻¹) are in parentheses. The C atoms are in grey, H atoms are in white and O atoms are in red.

The aliphatic chain length influence were analyzed applying molecular dynamics starting from "open" (linear) conformations of Rha-C₁₀, Rha-C₁₀-C₁₀, Rha-Rha-C₁₀ and Rha-Rha-C₁₀-C₁₀ at GFN2-xTB level. Side chains of 10 carbon atoms (C₁₀) were chosen because is more abundant among RML molecules. The NTV ensemble was applied with Berendsen thermostat integration step of 0.1 fs, T = 298.15 K and 500 ps of total trajectory time. Low energy structures along the trajectory were selected and optimized at r²SCAN-3c level. The same protocol was also applied for Rha-C₂, Rha-C₂-C₂, Rha-Rha-C₂ and Rha-Rha-C₂-C₂. It is worth mentioning that the same global minimum structure were obtained, in other words, molecular dynamics and CREST approaches are equivalent for
the exploration of chemical space. Finally, a comparison between chain lengths with two (C_2) and ten (C_{10}) carbon atoms were done. In order to improve discussions, aliphatic (side) chains and rhamnose (Rha) groups were numbered according to Figure S01 in the Electronic Supplementary Information (ESI).

Finally, molecular geometries of neutral metal complexes formed with K, Ni, Cu and Zn were also investigated by the same molecular dynamics protocol. These metals were chosen because they are abundant and it is possible to analyse the influence of different number of RML molecules and spin multiplicities. For K and Zn, the neutral complexes are in the singlet state and for Cu and Ni the multiplicities are doublet and triplet, respectively. For this propose, only the Rha-C_2-C_2 was considered because Rha-C_{10}-C_{10} is the most abundant RML but C_{10} chain length was reduced to C_2 due the computational limitations. The DFT step was performed applying the PBE0 functional \textsuperscript{56,57} with D3BJ dispersion correction \textsuperscript{58,59} and def2-SVP \textsuperscript{60} basis set except for the metal atoms where the def2-TZVP \textsuperscript{60} was used instead. This level of theory is referred as PBE0-D3/def2-SVP;def2-TZVP(M) where M stands for metal. A similar approach was recent applied for the study of organometallic complexes. \textsuperscript{61}

\section{Results and discussion}

For the Rha-C_2 system, the most stable isomer was RC2-1 (Figure 2 top panel) featuring a hydrogen bond between carboxylic and second hydroxyl (OH\textsuperscript{2}) groups from rhamnose (Rha) ring. RC2-2 is above RC2-1 only by 1.37 kcal mol\textsuperscript{-1} where the hydrogen bond was cleavage by carboxylic group rotation. In the case of RC2-3, the hydrogen bond was formed with OH\textsuperscript{4} and is 0.89 kcal mol\textsuperscript{-1} and 2.26 kcal mol\textsuperscript{-1} above in energy with respect to RC2-2 and RC2-1, respectively.

The effect of the number of side chains through ester bond formation can be analyzed studying the structures obtained for Rha-C_2-C_2. First, RC2C2-1 has a similar conformation when compared to RC2-1, i.e., a hydrogen bond between carboxylic and OH\textsuperscript{2} is present (Figure 2 middle panel). Munusamy and co-workers found a distance of 3.2 Å between the oxygen atoms (d_{O...O}) that participate in the hydrogen bond in Rha-C_{10}-C_{10} aggregates.\textsuperscript{10} This value is larger than the one obtained in RC2C2-1, d_{O...O} = 2.70 Å. Also, the distance between hydrogen and oxygen atoms (d_{H...O}) in RC2C2-1 is 1.71 Å. This discrepancy can be attributed to the difference in the environment and level of theory. Also, the authors found that 20% of Rha-C_{10}-C_{10} are in this geometric arrangement defined as "folded conformation".\textsuperscript{10} For the second isomer, RC2C2-2, two hydrogen bonds are present. The first one between the carboxyl and OH\textsuperscript{4} and the second between hydroxyl (carboxylic) group and oxygen atom in Rha cycle. This structure is 2.02 kcal mol\textsuperscript{-1} above RC2C2-1. The next isomer (RC2C2-3) is 3.57 kcal mol\textsuperscript{-1} above RC2C2-1 without internal hydrogen bonds. These results indicate that hydrogen bond formation is not a criteria for high stability.

Focusing in the influence of the number of Rha groups, the lowest energy structure of Rha-Rha-C_2 (RRC2-1) is formed by a hydrogen bond between the carboxylic group and OH\textsuperscript{4} in the Rha1 ring (Figure 2 bottom panel) with d_{H...O} = 1.804 Å. When this bond is formed with the OH\textsuperscript{4} in the Rha2 ring, the enthalpy rises to 3.10 kcal mol\textsuperscript{-1}. Furthermore, in (RRC2-3), there are two internal hydrogen bonds, one between the two Rha cycles (forming OH\textsuperscript{3}-OH\textsuperscript{4} hydrogen bond) and the another one is connecting the Rha1 with the carboxylic group. This isomer is 6.79 kcal mol\textsuperscript{-1} above RRC2-1. These results illustrate that there are more structural diversity when the num-
Results and Discussion

Figure 4 Optimized structures of K(ROO) (left panel) and Ni(ROO)$_2$ (right panel) complexes in open (top) and closed conformation (bottom) at PBE0-D3/def2-SVP,def2-TZVP(M) level of theory. The K atom is in purple and Ni atom is in light green. The energy is associated to the stabilization process due to the changing in conformation.

For the analysis of combined effects of the number of chains and Rha cycles, Rha-Rha-C$_2$-C$_2$ was also considered. In Figure 3, two views of low energy structures are present. In the most stable one, RRC2C2-1 (Figure 3, top panel), OH$^1$ in the Rha1 cycle is forming a hydrogen bond with the carboxylic group. In the RRC2C2-2 (Figure 3, middle panel), which is only 0.89 kcal mol$^{-1}$ above, has a complex geometry pattern with several hydrogen bond formation containing two bonds between Rha cycles and carboxylic group. Luft and co-workers found that “closed conformations” are preferred for dirhamnolipids at the air-water interface in agreement with the structures obtained for Rha-Rha-C$_2$-C$_2$ (Figure 3). Also the authors found that Rha2 has a stronger interaction with carboxylic group than Rha1. But from our results both Rha rings contributes to the stabilization. For example, in RRC2C2-1, the interaction between carboxylic and rhamnose groups occur through Rha2 but in RRC2C2-2 through Rha1 with just a very small difference of 0.89 kcal mol$^{-1}$. Moreover, they found $d_{H-O} = 2.6$ Å while for RRC2C2-1 slightly bigger, 2.75 Å. Besides that, we obtained $d_{H-O} = 1.791$ Å for the same isomer. The next one is RRC2C2-3, also exhibiting several hydrogen bonds but in this case, only the Rha2 group is contributing, indicating that the number of Rha cycles increases the geometric complexity. This isomer is 2.65 kcal mol$^{-1}$ above RRC2C2-1.

From Figure S02 in the ESI (top panels) it is possible to see that the chain length has a small impact on the local structure of Rha-C$_2$ and Rha-Rha-C$_2$. However, in the Rha-C$_{10}$ molecule, there is no internal hydrogen bond and Rha-Rha-C$_{10}$ has a slightly bigger hydrogen bond than Rha-Rha-C$_2$ with $d_{H-O} = 1.864$ Å. The difference increases when a second chain is considered as in the case of Rha-C$_{10}$-C$_{10}$ (Figure S02, bottom panel). Again, we obtained a bigger $d_{H-O}$ (of 1.780 Å) when compared with the C$_2$ congener. A more drastic change is observed for Rha-Rha-C$_{10}$-C$_{10}$ due to the increase in the molecular complexity when two rhamnose rings combined with two side chains are present (Figure S03). For example, in this case a greater difference in the hydrogen bond was obtained with

Figure 4 Optimized structures of K(ROO) (left panel) and Ni(ROO)$_2$ (right panel) complexes in open (top) and closed conformation (bottom) at PBE0-D3/def2-SVP,def2-TZVP(M) level of theory. The K atom is in purple and Ni atom is in light green. The energy is associated to the stabilization process due to the changing in conformation.

For the analysis of combined effects of the number of chains and Rha cycles, Rha-Rha-C$_2$-C$_2$ was also considered. In Figure 3, two views of low energy structures are present. In the most stable one, RRC2C2-1 (Figure 3, top panel), OH$^1$ in the Rha1 cycle is forming a hydrogen bond with the carboxylic group. In the RRC2C2-2 (Figure 3, middle panel), which is only 0.89 kcal mol$^{-1}$ above, has a complex geometry pattern with several hydrogen bond formation containing two bonds between Rha cycles and carboxylic group. Luft and co-workers found that “closed conformations” are preferred for dirhamnolipids at the air-water interface in agreement with the structures obtained for Rha-Rha-C$_2$-C$_2$ (Figure 3). Also the authors found that Rha2 has a stronger interaction with carboxylic group than Rha1. But from our results both Rha rings contributes to the stabilization. For example, in RRC2C2-1, the interaction between carboxylic and rhamnose groups occur through Rha2 but in RRC2C2-2 through Rha1 with just a very small difference of 0.89 kcal mol$^{-1}$. Moreover, they found $d_{H-O} = 2.6$ Å while for RRC2C2-1 slightly bigger, 2.75 Å. Besides that, we obtained $d_{H-O} = 1.791$ Å for the same isomer. The next one is RRC2C2-3, also exhibiting several hydrogen bonds but in this case, only the Rha2 group is contributing, indicating that the number of Rha cycles increases the geometric complexity. This isomer is 2.65 kcal mol$^{-1}$ above RRC2C2-1.

From Figure S02 in the ESI (top panels) it is possible to see that the chain length has a small impact on the local structure of Rha-C$_2$ and Rha-Rha-C$_2$. However, in the Rha-C$_{10}$ molecule, there is no internal hydrogen bond and Rha-Rha-C$_{10}$ has a slightly bigger hydrogen bond than Rha-Rha-C$_2$ with $d_{H-O} = 1.864$ Å. The difference increases when a second chain is considered as in the case of Rha-C$_{10}$-C$_{10}$ (Figure S02, bottom panel). Again, we obtained a bigger $d_{H-O}$ (of 1.780 Å) when compared with the C$_2$ congener. A more drastic change is observed for Rha-Rha-C$_{10}$-C$_{10}$ due to the increase in the molecular complexity when two rhamnose rings combined with two side chains are present (Figure S03). For example, in this case a greater difference in the hydrogen bond was obtained with
3 RESULTS AND DISCUSSION

- Figure 5 Optimized structures of Cu(ROO)$_2$ (left panel) and Zn(ROO)$_2$ (right panel) complexes in open (top) and closed conformation (bottom) at PBE0-D3/def2-SVP:def2-TZVP(M) level of theory. The Cu atom is in dark blue and Zn atom is in dark green.

- For the Ni(ROO)$_2$, the stabilization energy due to the formation of the "closed conformation" is relatively high, 62.8 kcal mol$^{-1}$. In this conformation, the local structure is a distorted octahedral geometry with the Ni–O distances in between 2.246 and 1.992 Å in contrast with the "open conformation" featuring a local square planar structure. Furthermore, in the closed form, one of the carboxylic groups is contributing to the coordination with only one oxygen atom (see Figure 5).

- Considering now the formation of Cu complex, Cu(ROO)$_2$, the open conformation also present a local square planar geometry. However, the closed conformation present a distorted square planer motif with Cu–O distances ranging from 1.916 to 1.933 Å. In this case, only one oxygen atom in both carboxylic groups are contributing to the coordination. The stabilization is very near to the potassium case, 23.5 kcal mol$^{-1}$ (Figure 5 left panel). The last complex is Zn(ROO)$_2$ formed by complexation with zinc. The starting conformation exhibit a local tetrahedral molecular geometry but in the closed one a disphenoidal geometry is present. The Zn–O distances are in between 2.406 and 1.941 Å. Again, only one oxygen atom in each carboxylic groups are contributing to the coordination. The energy difference between open and closed conformation is 41.6 kcal mol$^{-1}$. In all RML-metal complexes studied in this work, a huge structural reorganization is observed between open and closed molecular conformations.
4 Conclusion

A automated search throughout the potential energy surface of representative rhamnolipid congeners reveals a plethora of low energy structures. Remarkably, CREST and molecular dynamics protocols leads to the same global minimum structures indicating that both strategies are equivalent. The complexity of the molecular geometry increases with respect to the number of side chains, rhamnose rings and chain length. Although several low energy structures exhibit internal hydrogen bonds, this interaction alone is not a criteria for high stability. For example, the main differences between the two lowest energy structures of Rha-C$_2$ (RC2-1 and RC2-2) is the formation of one internal hydrogen bond, but the energy difference is only 1.37 kcal mol$^{-1}$. Moreover, geometries in closed conformation are always more stable than open conformation ones. Furthermore, the effect of chain length on the geometry is more prominent between Rha-Rha-C$_2$-$\ldots$C$_2$ and Rha-Rha-C$_{10}$. This result also indicate that the number of rhamnose rings and side chains induce differences among congeners with different side chain lengths.

Concerning on the formation of RML-metal complexes, all global minimum are in the closed conformation. The difference between the open and closed conformations of K(ROO)$_2$, Ni(ROO)$_2$, Cu(ROO)$_2$ and Zn(ROO)$_2$ are 23.5 kcal mol$^{-1}$, 62.8 kcal mol$^{-1}$, 24.3 kcal mol$^{-1}$ and 41.6 kcal mol$^{-1}$, respectively. Consequently, a huge structural reorganization, which was observed along the molecular dynamics trajectories, is necessary to form the complexes in its lowest energy structures. Finally, by applying a combination of xTB and DFT methods, this study reveals the richness of molecular structure of RML systems.

Conflicts of Interest

There are no conflicts to declare.

Keywords

rhamnolipids · density functional theory · organometallic complex · global minimum search

Acknowledgements

The authors acknowledge the National Laboratory for Scientific Computing (LNCC/MCTI, Brazil) for providing HPC resources of the SDumont supercomputer, which have contributed to the research results reported within this paper (http://sdumont.lncc.br). Also, the brazilian research support foundations FAPERJ, CAPES and CNPq for the financial support.

References

REFERENCES


