ARTICLE

Received
00th January
20xx,

Structure Optimization of Lipopeptide Assemblies for Aldol Reactions in an Aqueous Medium

Bruna M. Soares,[a] Pedro T. Sodré,[a] Andrea M. Aguilar,[b] Barbara B. Gerbelli,[a] Juliane N.B.D. Pelin,[a] Karina B. Argüello,[a] Emerson R. Silva,[c] Marcelo A. Farias,[d] Rodrigo V. Portugal,[d] Carsten Schmuck,†[e] Maurício D. Coutinho-Neto,[a] Wendel A. Alves.*[a]

Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Four amphiphilic peptides were synthesized, characterized, and evaluated regarding their efficiency in the catalysis of direct aldol reactions in water. The lipopeptides differ by having a double lipid chain and a guanidinium pyrrole group functionalizing one Lys side chain. All the samples are composed of the amino acids L-proline (P), L-arginine (R), or L-lysine (K) functionalized with the cationic guanidiniocarbonyl pyrrole unit (GCP), L-tryptophan (W), and L-glycine (G), covalently linked to one or two long aliphatic chains, leading to surfactant-like designs with controlled proline protonation state and different stereoselectivity. Critical aggregation concentrations (*cac*) were higher in the presence of the GCP group, suggesting that self-assembly depends on charge distribution along the peptide backbone. Cryogenic Transmission Electron Microscopy (Cryo-TEM) and Small Angle X-ray Scattering (SAXS) showed a rich polymorphism including spherical, cylindrical, and bilayer structures. Molecular dynamics simulations performed to assess the lipopeptide polymorphs revealed an excellent agreement with core-shell arrangements derived from SAXS data and provided an atomistic view of the changes incurred by modifying head groups and lipid chains. The resulting nanostructures behaved as excellent catalysts for aldol condensation reactions, in which superior conversions (>99%), high diastereoselectivities (ds = 94:6), and enantioselectivities (*ee* = 92%) were obtained. Our findings contribute to elucidate the effect of nanoscale organization of lipopeptide assemblies in the catalysis of aldol reactions in an aqueous environment.

†This article is dedicated to the memory of Professor Carsten Schmuck.

Introduction

Investigations on self-assembling properties of amphiphilic peptides have emerged in recent years as an active and diversified field involving either fundamental or application-oriented research in biotechnology and materials science. These materials exhibit intrinsic multi-level organization, and this structural multiplicity generates several types of intra- and intermolecular interactions, making them extremely attractive as a model for the study of self-organization phenomena occurring in either living organisms or catalytic process.[1-7]

Aldol reactions are among the most important and investigated transformations in organic synthesis because a new C-C bond and a β-hydroxy carbonyl structure are formed from two carbonyl compounds.[8] Asymmetric aldol reactions are characterized by the formation of new stereogenic carbons, increasing the reactions' selectivity levels.[9]

The efficiency of peptides containing L-proline as an organocatalyst in aldol condensation reactions has been known for over the past half-century.[10-12] This methodology was revisited in 2000 when this amino acid was proved to be very efficient at promoting organocatalyzed aldol reactions.[13, 14] These studies are important milestones for the development of new organocatalysts containing proline and the corresponding application in several organic reactions.[15, 16]

Previous studies have shown that the supramolecular hydrophobic-amphiphilic approach is responsible for the high

[a] *Centro de Ciências Naturais e Humanas, Universidade Federal do ABC, 09210-580, Santo André, Brazil. E-mail: wendel.alves@ufabc.edu.br*
[b] *Instituto de Ciências Ambientais, Químicas e Farmacêuticas, Universidade Federal de São Paulo, 09972-270, Diadema, Brazil.*
[c] *Departamento de Biofísica, Universidade Federal de São Paulo, 04023-062, São Paulo, Brazil.*
[d] *Brazilian Nanotechnology National Laboratory, CNPEM, 13083-970, Campinas, Brazil.*
[e] *Institute for Organic Chemistry, University of Duisburg-Essen, 45117, Essen, Germany.*

Electronic Supplementary Information (ESI) available: [details of any supplementary information available should be included here]. See DOI: 10.1039/x0xx00000x

efficiency of L-proline-derived peptides as organocatalysts in asymmetric aldol reactions.[17-22] Proline-based compounds were also shown to be catalysts in water with the aid of polymer or surfactant.[23, 24]

Recently, we studied the effect of substitution of the ester by an amide linkage between the hydrophobic lipid chain and the PRW tripeptide headgroup (where, P: L-proline, R: L-arginine, W: L-tryptophan).[25] Both lipopeptides have been shown to self-assemble into micelles above a critical aggregation concentration, detected through intrinsic tryptophan fluorescence or using an added fluorophore probe (pyrene, the fluorescence of which is sensitive to the hydrophobic environment). Aldol reactions using cyclohexanone and *p*-nitrobenzaldehyde showed that PRW-NH-C_{16} has an excellent *anti/syn* diastereomeric ratio (93:7) and a high enantiomeric excess (*ee*) of 88%. These values are better than those observed for PRW-O-C_{16} under the same conditions (*anti/syn* = 91:9 and *ee* = 71%)[19]. Here, the linker group had an influence that is potentially due to differences in the local conformation around the catalytic site and/or the altered polarization of the amide *vs.* ester linkage. Also, for both systems, low correlations were observed in the absence of water. This effect can be associated with the lower organization of the lipopeptide molecules due to weakening of hydrophobic interactions that drive self-assembly into micelles. We have also found that the L-proline residue influences the lipopeptide self-assembly, considering RWG and PRWG residues attached to one or two aliphatic octadecyls chains.[25] The organocatalyzed aldol efficiency was increased by water availability, achieving almost 95% conversion and excellent diastereoselectivity (93:7). Also, it was observed an enhancement of conversion ratio upon the growth of proline-lipopeptide content, which formed micelles with a core radius of 3 nm and shell thickness of 2 nm.

In the present work, we carried out an experimental and theoretical study on the influence of replacing arginine residues by a synthetic arginine mimetic, the guanidiniumcarbonyl pyrrole unit (GCP), in PRWG-C_{18} and PRWG-$(C_{18})_2$ lipopeptides (Figure 1). GCP is an efficient binding site for oxoanions in polar solutions.[26] Owing to its weak basicity (pK_a = 7), the only partially protonated state within the peptide composition efficiently reduces charge repulsion while enabling strong interactions between side chains, thus assisting stabilization of aggregates.[27] We synthesized two series of lipopeptides, in both presence and absence of GCP groups, and studied their self-assembly and their effects on directly catalyzing asymmetric aldol reactions of cyclohexanone and *p*-nitrobenzaldehyde in water.

Molecular dynamics simulations was used previously[19] to assist characterization of PRW-O-C_{16} lipopeptide structure and understand its role on the pre-concentration of reactants. Here, we have extended these analyses to understand the differences between the distinct micelles produced from self-assembling PRW-O-C_{16}, PRW-NH-C_{16}, PRWG-C_{18} (**1**), and PRWG-$(C_{18})_2$ (**2**). We have also investigated the effect of adding a Lys-GCP group over the structural properties of the micelle (compounds **3** and **4**). Moreover, quantum mechanical calculations were carried to investigate a possible role of the micro-environment provided by the micellar arrangement at controlling proline protonation.

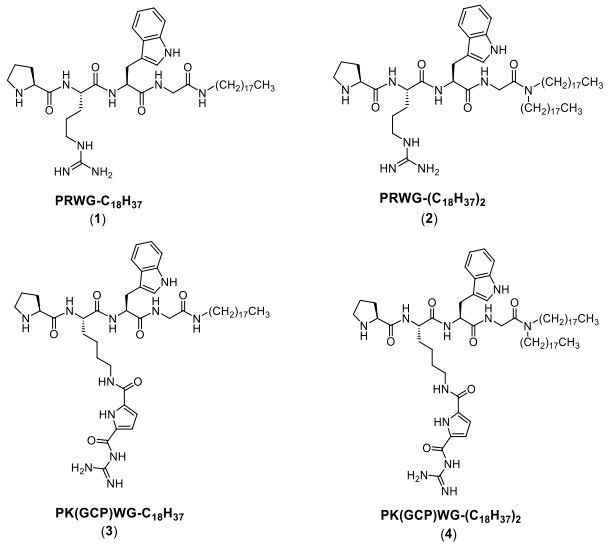


Figure 1. Molecular structures of the lipopeptides PRWG-$C_{18}H_{37}$ (**1**), PRWG-$(C_{18}H_{37})_2$ (**2**), PK(GCP)WG-$C_{18}H_{37}$ (**3**) and PK(GCP)WG-$(C_{18}H_{37})_2$ (**4**).

Experimental

Lipopeptide synthesis

All the experimental procedures for synthesizing the lipopeptides (**1**-**4**) are described in detail in the Support Information. Also, the mass spectroscopy spectra can be seen in Figure S1.

General information

All solutions were prepared with water purified by the Direct-Q System, Millipore, with a resistivity of 18.2 mΩ·cm^{-1} (at 25 °C) and TOC below 10 ppb. The reagents and solvents used were all of the analytical purity.

Fluorescence spectroscopy

Lipopeptide solid was firstly dissolved in 20 µL of HFIP (1,1,1,3,3,3-hexafluoro-2-propanol) and then water solutions (1.0 x10^{-7} to 0.1 wt%) were prepared. After one day of rest, the fluorescence of the systems was evaluated irradiated with λ_{exc} = 280 nm, and the emission spectrum were investigated in the range 300 nm $\leq \lambda_{em} \leq$ 460 nm. Emission spectra were obtained using fluorescence spectrophotometer, Cary Eclypse (Varian), model and quartz cuvette, SUPRASIL 10 x 10 mm, (Hellma Analytics).

Cryogenic Transmission Electronic Microscopy (Cryo-TEM)

Cryo-TEM images of 1 wt% of the lipopeptides were performed. The assays were made in Vitrobot Mark IV (Thermo Fisher Scientific, USA) instrument, using *blot time:* 1.0 or 2.5 s; *blot force*: -5 s, *blot wait*: 20 s, with temperature and humidity of 22 °C and 100%, respectively. For this, 3.0 µL of the sample was added in copper grids (Lacey Carbon Type A) with 300 mesh (TedPella, USA, which is submitted to glow discharge procedure in the equipment EasiGlow, Pelco, USA), using the negative charge. The grids were firstly immersed in liquid ethane, and then they were kept in liquid nitrogen (-173 °C) until the moment of analysis. The images were obtained by transmission electron microscope model TALOS F200C (Thermos Fisher Scientific, USA), operating at 200 kV, with CMOS Ceta 16M camera. The pixel size for the images obtained varying from 0.161nm to 0,413nm. (Thermo Fisher Scientific, USA). Image J program was used to calculate the size of particles.

Small Angle X-ray Scattering (SAXS)

SAXS experiments were carried out at the Brazilian Nanotechnology National Laboratory (Campinas, Brazil) on a Pilatus 300K detector (Dectris). The 2D scattering data were integrated using the program FIT2D.[28] The integrated intensity is displayed as a function of the momentum transfer modulus q, where q is defined as $q = 4\pi\sin(\theta)/\lambda$, which 2θ is the scattering angle and λ is the radiation wavelength. The measured q range was 0.0150 Å^{-1} ≤ q ≤ 0.45 Å^{-1} with radiation energy of 8 keV (λ =1.5498 Å), using sample holders for liquid samples available on the beamline at room temperature. Ten frames of 60 seconds each were recorded during the flow to avoid radiation damage. Afterward, the data treatment was performed by subtracting the solvent normalized intensity from the normalized intensity of the particles in the solution, using the Fit2D software, for the data fitting was used SASFit[29] program.

Aldol reactions

The catalytic efficiency of the lipopeptides was monitored by the direct aldol reaction using cyclohexanone and *p*-nitrobenzaldehyde. The reactions were performed using different catalyst amounts of 1.0, 2.5, and 5.0 mol%. Considering the 5 mol% catalyst solutions, 100 µL of cyclohexanone (0.96 mmol, 12 equivalents), 3.0 mg of catalyst, 12.0 mg of *p*-nitrobenzaldehyde (80 µmol, 1 equivalent), and 200 µL of water (2x cyclohexanone volume) were used. Firstly, a small amount of HFIP was added to the solid lipopeptide to break the secondary structures, and after the water was included. The systems were kept on the cupboard for one day to evaporate all the HFIP and promote the lipopeptide self-assembly. Then the aldol reagents were added, and the solutions were stirred at room temperature for two days. The mixtures were extracted with ethyl acetate four times *via* centrifugation at 4000 rpm for 5 min. The organic phase was then removed on a desiccator under a high vacuum, and the samples were solubilized using deuterated chloroform. NMR measurements using a (^{1}H) Bruker Ultrashield Plus 300 instrument were performed at 300 MHz. The yield and diastereomer *anti:syn* ratio were calculated using the NMR spectra obtained, for which tetramethylsilane (TMS) was used as a reference. The enantiomeric excess (*ee*) values were determined by HPLC on a Chiralpak AD-H column.

Molecular dynamics simulation

All molecular dynamics simulations were focused on spherical micelle systems. We have modelled all lipopeptides using the AMBER family of force fields. The alkyl tails were modelled using Lipid14[30,31] and Lipid17[32] force fields while the peptide part was modelled using the ff14SB force field.[33] The peptide-lipid boundary was modelled using GAFF[34] parameters for amides. The PRW-O-C_{16} ester bond connecting the peptide and alkyl chain employed the same parameters used in LIPID11 and LIPID14. We have used the standard RESP [35] charge scheme implemented in Ambertools[36, 37] to derive charges for the alkyl section using an acetyl cap. We employed the Hartree-Fock method and a 6-31G* basis set in all RESP charge derivations. Tip3P[38] water and Cl^- counter ions were added to the system to ensure neutrality. Note that Cl^- was used instead of trifluoroacetic acid (TFA) as the counter ion in our simulations, for simplicity sake. TFA is the experimental counter ion used in all lipopeptide systems.

Packmol[39, 40] was used to provide initial configurations for all micellar systems. Initial conditions for systems (**1**), (**2**), and (**4**) were selected to packing into a spherical object with a final diameter of circa 7 to 8 nm after minimization. The alkyl chains were packed into an inner sphere, with the peptide part pointing outwards. System (**3**) was packed into a cylinder with a diameter of circa 60 Å. Ions and water were added to the network using leap from Ambertools with a shell of water of 1.2 nm. We have used it in all simulated systems 160 lipopeptide units. The geometric parameters that control packing for PRWG-C_{18} and PRWG-$(C_{18})_2$ are close enough to PRW-O-C_{16}. Hence, we opted to perform our simulations using the same packing number as in our previous work[19] to simplify comparisons between the different systems.

Six simulation systems were considered, namely systems PRWG-C_{18} (**1**), PRWG-$(C_{18})_2$ (**2**), PK(GCP)WG$C_{18}H_{37}$ (**3**), PK(GCP)WG-$(C_{18}H_{37})_2$ (**4**), PRW-N-C_{16} and PRW-O-C_{16}. PRW-O-C_{16} has been previously investigated experimentally and theoretically.[19] However, due to small changes in the employed protocol and simulation time, we have duplicated our original simulation for comparison purposes.

The simulation consisted of the following steps: system minimization, *NVT* equilibration, *NPT* equilibration, and *NPT* production. Particular attention should be paid to the initial minimization and equilibration steps, as these systems typically need to relax their initial configuration to remove close contacts in the structure. Hence, our initial NVT equilibration procedure restricts the lipidic group of 2 kcal/$Å^2$ to ensure the system's stability during the heating phase to 300K for 200 ps. Next, a short restrained 300 ps NPT step at P=1ATM is performed, followed by 5 ns of unrestrained molecular dynamics at T=300K and P = 1 atm. A production of 100ns is performed for each of the simulated systems using the GPU implementation of AMBER [41, 42]. A Berendsen[43] barostat with a 1.0 ps relaxation time and pressure equals 1 atm, and the Langevin[44] thermostat with a 2 ps^{-1} collision frequency and a timestep of 0.02 fs were used. The algorithm Shake was used to constrain the Hydrogen to their equilibrium positions.[45] All structural properties were analyzed using the CCPTRAJ[46] program from AMBERTOOLS. Visualization was performed using VMD.[47] We have computed radial distribution functions (RDFs or g(r)), the radius of gyration (ROG), and hydrogen bonds (HBs) for all systems aiming to understand the differences between their supra-molecular arrangements.

Quantum Mechanical Calculations

Quantum mechanical (QM) calculations were performed in three distinct environments (water, cyclohexanone, and N-heptane) to evaluate the deprotonation ΔG and pKa for the amino-terminal of L-proline and trifluoroacetic acid. In all calculations, an N-methyl-amide group was used instead of the normal carboxylic acid, as it is a closer analog to the N-terminal of a short peptide. Results are summarized in Table S4. ΔG and pKa results used a proton solvation energy in the water of -265.9 kcal·mol^{-1} and the procedure suggested by Rossini.[48, 49] We assumed that the solvated proton is in pure water in all calculations reflecting the equilibrium state in a micro heterogeneous system in analogy to a procedure employed by Silva.[50] We have used the ma-def2-TZVP basis set,[51, 52] and B3LYP functional[53-55] with empirical dispersion corrections[56, 57] as implemented in Orca code version 4.2.1[58] in all calculations. Solvation effects were treated using the implicit solvent SMD (solvation model density) method from Truhlar and collaborators.[59] Computed results used bare as well as micro-solvated molecules using two waters in a procedure analog to the one suggested by Pliego and Riveros.[60] More specifically, in the micro-solvation calculations, two explicit water molecules were added in hydrogen bond configurations to each one of the solute molecules.

Results and Discussion

The self-assembly of lipopeptides was initially evaluated by steady-state fluorescence spectroscopy, by investigating the intensity of the intrinsic tryptophan fluorescence to obtain critical aggregation concentrations (*cac*). Emission curves from the present study are shown in Figure S2, whereas the behavior of fluorescence intensity as a function of lipopeptide concentration is shown in Figure 2.

The emission of indole may be blue-shifted if the group is buried within a hydrophobic cavity, and its emission may shift to longer wavelengths (red-shift) when the hydrogen bonding and the subsequent exposure of the tryptophan residue to the aqueous phase increases. Both peptides **3** and **4** showed fluorescence emission at longer wavelengths (> 350 nm), likely revealing GCP residues' influence in the self-assembly, favoring the high water content and closer packing of the lipopeptides, as shown by molecular dynamic simulation. Also, compound **3** showed an emission band at ~440 nm, characteristic of fluorescence resonance energy transfer (FRET), since the GCP/indole separation distances are

shorter than the typical Förster distance (9-16 Å). In this case, GCP absorption at around 300 nm overlaps with the emission of tryptophan, producing FRET at 440 nm (see Figure SX).

Critical concentrations were obtained from the intercept curves were determined to be (3.4 ± 0.5) x 10^{-4} wt% for (**1**) and (**2**), (1.2 ± 0.5) x 10^{-3} wt% for (**3**) and (2.5 ± 0.5) x 10^{-3} wt% for (**4**). This enhancement suggests higher hydrophilicity of sequences (**3**) and (**4**) than other compounds, which could be attributed to the GCP group charge, promoting a higher interaction with the water medium.

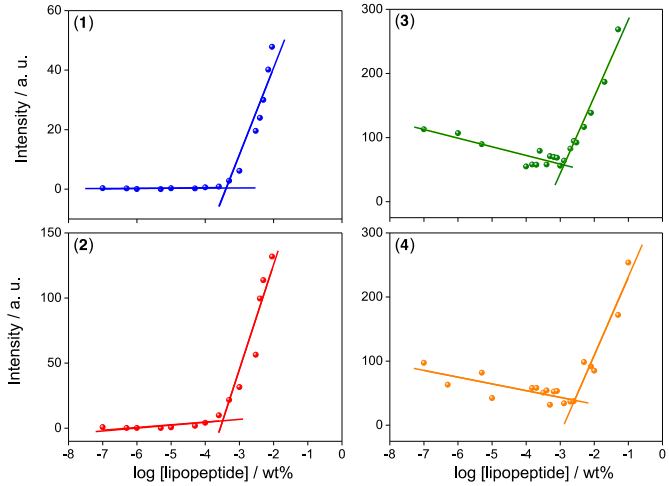


Figure 2. Fluorescence data used to estimate the CAC of lipopeptides **1-4**. The intersection between the straight lines fitted to the data from different fluorescence regimes provided the CAC estimate.

In Figure 3, SAXS curves from aqueous solutions prepared at different concentrations show scattering profiles characterized by oscillations along the q-range, indicating the presence of monodisperse nanoparticles with regular morphology. For some compositions and concentrations, it is possible to observe an abrupt decay at the low angle region, indicating the presence of interference peaks associated with interparticle correlations.[61] The datasets were firstly fitted by using the indirect Fourier transformation (IFT) method, which allows for determining the pair distance distribution function *p*(r). P(r) curves shown in Figure S3 are characterized by bimodal profiles, which indicate core-shell assemblies in solution.[62]

Considering the IFT approach used above, we have fitted our data according to spherical and cylindrical shells form factors to obtain quantitative measurements on characteristic sizes of nanoparticles in solution (Figures 3a and 3b). Interparticle correlations have been described by using a hard-sphere structure factor in the model. In the case of samples prepared with (**4**), a bilayer form factor[63] has been added to account for the presence of planar assemblies in the solution (Figure S4e). The values of the best fitting parameters are summarized in Table 1. Data from solutions prepared with (**1**) could be satisfactorily fitted using a spherical core-shell form factor convoluted with a hard-sphere structure factor. In the most diluted sample (0.12 wt%), we found the presence of aggregates with a core radius (R_{in}) of 1.4 nm and a shell (R_{out}) radius of 2.2 nm (Figure 3a). It is interesting to note that the entire data range could be adjusted with a single type of nanostructure, reinforcing the consistency of the model. For the intermediate concentration (0.50wt%), the form factor also consists of the spherical core-shell type, with the same R_{in} of the lowest concentration but with the R_{out} equal to 2.3 nm, indicating that growth of nanoparticles is observed in the medium. In this case, the curve profile with absence of intensity minimums shows a higher polydispersity, which is expected since more peptides in the medium and, consequently, more particles in equilibrium (Figure 3b).

Moreover, it is observed that only the particle shell has increased with increasing concentrations. It can be an indication of a new conformation from a more compact to a more elongated state. This could be attributed to the fact that we increase the interaction with the arginine lateral group promoting such behavior. It was impossible for the more concentrated sample (1.00 wt%) to make a good fitting using a simple spherical model (Figure S3e).

A decrease of the hard-sphere repulsion radius (RHS) verified upon increasing concentration suggests a higher interaction between the particles in solution.

The curves related to lipopeptide (**2**) for the most diluted sample curve (0.12 wt%) also indicated spherical shell aggregates (Figure 3a), with R_{in} of 1.8 nm and an external radius of 2.8 nm. The profile of the curve is consistent with the presence of regular particles with low polydispersity. For the intermediate concentration sample (0.50 wt%), the same size found at low concentration was observed (Figure 3b). For the more concentrated system (1.0 wt%), the curve shape indicates greater polydispersity with the presence of larger micelles. In this case, the internal radius is 1.8 nm, and the external radius is 2.8 nm (Figure S3e).

SAXS data are significantly different for peptides (**3**) and (**4**), likely revealing the influence of GCP residues in the self-assembly. For the most diluted sample prepared with (**3**), one observes that the low-q range exhibits a smooth decay scaling with q^{-1}, indicating the presence of cylindrical structures (Figure 3a). To fit these data, we used the cylinder core-shell form factor. The structural parameters obtained for this concentration were R_{in} = 1.5 nm and R_{out} = 2.7 nm.

For the intermediate concentration sample (0.50 wt%) again, a cylindrical model was used, increasing the size of R_{in} and R_{out} (Figure 3b).

Data from lipopeptide (**4**) are shown in Figure 3a. The most diluted sample curve (0.12 wt%) profile is accurately described by spherical shell form factors revealing low polydispersity and an internal radius of 1.6 nm. Samples at an intermediate concentration (0.50 wt%) are also described by the same factor presenting the same value for the inner radius, but the external radius increases by 0.8 nm (Figure 3b). For the more concentrated conditions (1.00 wt%), data fitting was carried out by adding two components corresponding to a bilayer and spherical form factor. This form factor has been used successfully to fit the form factor of lipopeptide nanotape and nanosheet structures (Figure S3e).[64, 65] and here it revealed bilayer that our samples contain planar structures with thickness equal to 7.2 nm, and spherical micelles with an internal radius of 1.2 nm and an external radius of 2.6 nm.

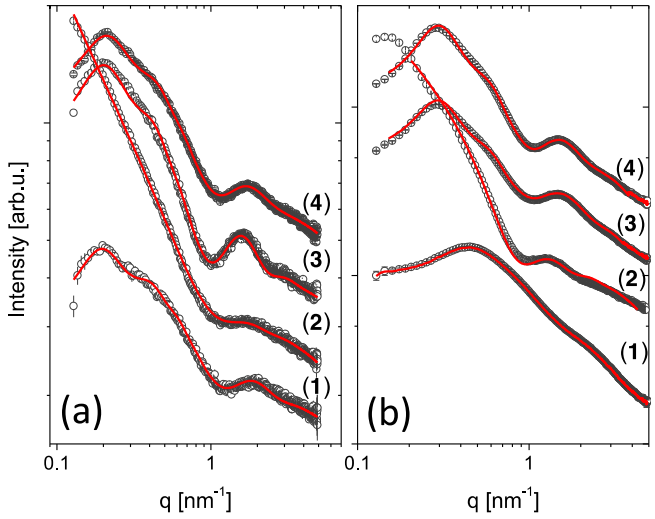


Figure 3. SAXS data from lipopeptide solutions (a)0.12wt% and (b)0.5wt%. The solid lines represent the fitting to the experimental data using a spherical shell (**1**) and (**2**), cylindrical shell (**3**), and bilayer form factors (**4**).

A noteworthy point of samples prepared with lipopeptide (**4**) is the presence of interference peaks at dilute preparations, indicating that interparticle correlations are present even at low concentrations. These findings suggest that nanoparticles, in this case, carry higher charge densities at their outer surfaces, enhancing repulsion between self-assemblies in solution. Such an effect is likely a consequence of the GCP moiety present at the headgroup of lipopeptide (**4**).

A recent study from our group investigated the influence of proline residues in mixtures containing the lipopeptides (**1**) and (**2**).[25] In that case, a decrease of nanoparticle thickness upon an increase of proline-lipopeptide content corroborated with the experimental data obtained in this work. As proline amino acid has a rigid side chain, its presence in lipopeptide constructs can increase the hydrophobicity and form more compact systems.

Furthermore, these results showed that the GCP charge can change the self-assembly process of the lipopeptides in water, in which polymorphism can influence hydrophilicity and compression effects. To shed light on the atomistic interactions that characterize the different structures, we have performed molecular dynamics simulations for (**1**) and (**2**). After that, we compared it to previously investigated systems, namely PWR-O-C_{16} and PWR-NH-C_{16}.[19]

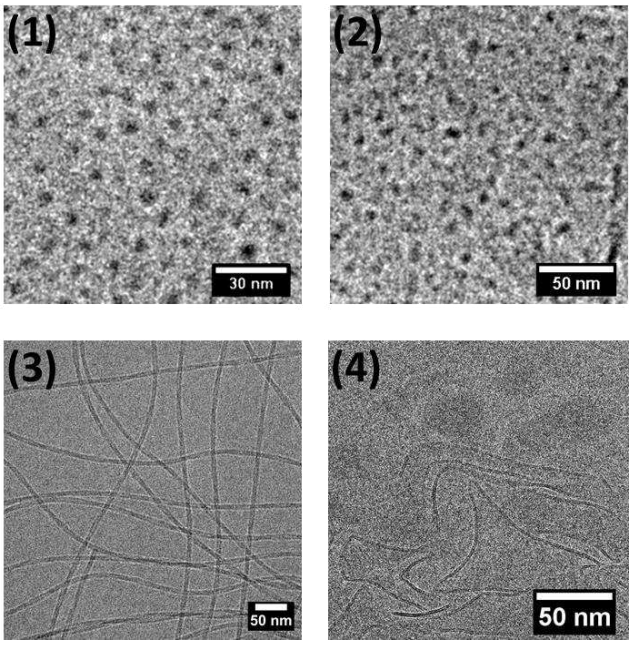


Figure 4. Cryo-TEM images of 1 wt% water solutions of **1**-**4**. Lipopeptide self-assembly in (**1**) and (**2**) micelles spherical, (**3**) micelles cylindrical, and (**4**) planar structures.

Cryo-TEM experiments were performed to assess the morphology in the real space of lipopeptides (**1**-**4**) above the *cac*. Using by Image-J program, we determine the size of particles obtained. Figure 4 shows circular nanoparticles, with diameters of 7.4 ± 0.3 nm and 9.4 ± 0.4 nm, respectively, for (**1**) and (**2**). System (**3**) is characterized by an elongated anisotropic structure of the filament type with an average cross-section of 6.4 ± 0.5 nm and length in micrometer order. However, the system (**4**) planar (bilayer) structures with 3.0 ± 0.5 nm thickness were found. The structures and sizes obtained in

these images corroborate the models proposed in the SAXS modeling, giving greater robustness to the results found.

Molecular dynamics simulation results for (**1**) and (**2**) revealed stable spherical micellar systems for the whole simulation time in water. Our simulations show a well-defined hydrophobic core for all systems investigated in agreement with SAXS results. Radial distribution functions using the C1 alkyl carbon atom (C1 meaning carbon 1 from the alkyl residue) as a reference for the micelle center are shown in Figures S6 and S7. These distributions present the average distances for the different peptides from the micelle center.

The computed radial distribution functions (RDFs) show a similar layered structural organization of the residues with respect to the micellar center for all systems. As reported previously for PRW-O-C_{16}, arginine and catalytic proline RDFs peak at about the same distance from the micelle center and comprise the hydrophilic pair. At the same time, tryptophan and glycine are closer to the hydrophobic core. For (**1**), C1-C18 g(r), where C18 is the last carbon atom of the alkyl chain, peaks at 2.33 nm while the g(r) for the C1-Pro pair peaks at 3.15 nm. We have used the backbone N, C, and O atoms to define the amino acid positions when computing RDFs. For PRWG-$(C_{18})_2$, the peak values for C1-C18 and C1-Pro g(r)s are 2.32 nm and 3.17 nm, illustrating the similarities between these two systems. RDF between the C1 alkyl carbon and water oxygen atoms revealed a drop in water content at about 4.5 nm from the hydrophobic core. At the peak of the proline g(r) for (**1**) (at r=3.15 nm), the C1-Wat(O) distribution has about 61% of its bulk value (g(r)=0.61), pointing to an intermediate environment in terms of water content where catalysis takes place. Interestingly, for the PRWG-$(C_{18})_2$ micelle (**2**), the value for the C1-Wat(O) distribution at the peak of the proline g(r) has only 52% of its bulk value, indicating a stronger hydrophobic nature for this system.

The radius of gyration values (ROG) shown in Table S2 for proline can be compared to SAXS results from Table 1. The total radio (R_{in} + R_{out}) determined experimentally agree with the ROG for the last residue, proline, being 3.78 nm and 4.22 nm for (**1**) and (**2**), respectively. This can be compared to 3.6 nm and 4.5 nm from SAXS results from Table 1. This result supports the choice for the aggregation number used to construct the micellar model systems. Despite having similar RDFs, the larger ROG values for (**2**) result from its bulkier hydrophobic core.

RDFs for PRW-O-C_{16} and PRW-NH-C_{16} systems, shown in Figure S8, present a similar pattern. Amino acids proline and arginine form a hydrophilic pair, while tryptophan distribution peaks close to the hydrophobic alkyl core for both systems. For PRW-NH-C_{16}, C1-C16 g(r), where C16 is the last carbon atom of the alkyl chain, peaks at 2.25 nm while the g(r) for the C1-Pro pair peaks at 2.86. nm. These values for the PRW-O-C_{16} system are 1.85 and 2.67 nm, respectively. The value for the C1-Wat(O) distribution at the peak of the proline g(r) is 0.53 for both systems (53% of the bulk value). ROG results for these two systems exhibit similar values. ROG for the last residue, proline, for PRW-O-C_{16} and PRW-C_{16} are 3.65 nm and 3.56 nm, respectively. This is close to the result for PRWG-C_{18} (**1**) of 3.78 nm and not far from 4.22 nm for PRWG-$(C_{18})_2$ (**2**) despite the presence of larger alkyl chains and different amino acid. The presence of glycine and a longer alkyl chain for (**1**), or two long alkyl chains (**2**), helps stabilize the hydrophobic core. This stabilization leads to smaller fluctuation values for the ROG, as shown in Table S2. Taken as a whole, these results indicate that all spherical micellar systems provide similar reaction environments for the aldol reaction to occur. However, (**1**) and (**2**) compounds are more stable and tightly packed than PRW-O-C_{16} and PRW-NH-C_{16} systems, and that increase in packing factor translates to better catalysis. When comparing PRW-O-C_{16} and PRW-NH-C_{16}, the added stability of PRW-NH-C_{16}, manifested as smaller fluctuations in ROGs (Table S3), correlates with higher conversion rates.

Table 1. Summary of structural parameters obtained from least-square fitting of SAXS data. R_{in}: inner radius; R_{out}: outer radius; t: thickness of bilayers, Σ_{in}: standard deviation of Gaussian function used to describe the apolar region and Σ_{out}: polar region. RHS: hard-sphere radius (effective distance between particles) and F_P: volume fraction occupied by hard spheres.

		Spherical shell form factor		Hard-Sphere structure fator		Cylinder shell form fator			Bilayer form factor .	
Compounds	Conc. (mg/ml)	R_{In} (nm)	R_{out} (nm)	RHS (nm)	FP	R_{In} (nm)	R_{out} (nm)	t (nm)	Σ_{out}	Σ_{in}
	1.25	1.4	2.2	15	0.14	—	—	—	—	—
(**1**)	5	2.2	2.3	6	0.13	—	—	—	—	—
	10	—	—	—	—	—	—	—	—	—
	1.25	1.8	2.8	14	0.12	—	—	—	—	—
	5	1.7	2.5	10	0.19	—	—	—	—	—
(**2**)	10	1.8	3.7	7	0.14	—	—	—	—	—
	1.25	—	—	—	—	1.5	2.7	—	—	—
(**3**)	5	—	—	—	—	1.8	3.5	—	—	—

	10	—	—	—	—	—	—	—	—	—
	1.25	1.6	2.5	13	0.12	—	—	—	—	—
	5	1.6	3.3	10	0.19	—	—	—	—	—
(**4**)	10	1.2	3.5	7	0.19	—	—	7.2	2.2	2.0

For simplification, systems (**3**) and (**4**) were modeled as a cylinder and a spherical body with the same aggregation number as (**1**) and (**2**). As a result, a direct comparison is not possible to SAXS data. However, intrinsic quantities such as radial distribution functions, area per head group, and alkyl chain volume can offer comparison data that help quantify the differences incurred by adding a bulkier head group such as Lys-GCP.

Radial distribution functions for systems (**3**) and (**4**) (shown in Figure S8) exhibit a similar layered pattern as systems (**1**) and (**2**). However, these systems show a prolate ellipsoid shape for the aggregation number used in the simulations (more akin to a cylinder), resulting in broader RDF distributions. System (**3**), however, reveals a more compact distribution compared to all others (as expected for a strongly prolate object). Interestingly, compounds (**3**) and (**4**) exhibit large differences in water content distribution (as shown in RDFs in Figure S9). At the peak of the proline g(r) for (**3**) (at r= 2.46 nm), the value for the C1-Wat(O) g(r) is 0.62 while for (**4**) this value is 0.48 at r=0.48. Therefore, despite the differences in packing and head groups, the content of water at the most probable position of the catalytic proline is remarkably similar for systems (**1**) and (**3**) (62% *vs.* 61% of the bulk value) and systems (**2**) and (**4**) (52% *vs.* 48% of the bulk value). This points to the strong influence of the alkyl tail on determining the environment where one expects reaction to take place.

Following the work of Tanford[66], Israelachvili,[67] Mitchel, and Ninham,[68] we have computed the packing parameter for the systems with one and two C_{18} alkyl chains. The packing parameter is defined as $p = v/(l_c \times a_0)$ where a_0 is the area per head group, v is alkyl chain volume, and l_c is the maximum carbon chain length. Alkyl chain volume computed from radial distribution functions obtained from molecular dynamic simulations produced 530 Å^3, 1040 Å^3, 410 Å^3, and 950 Å^3 per lipopeptide for systems (**1**) through (**4**), respectively. Note that the addition of GCP increases the assembly compactness. Assuming spherical structures and, using l_c computed through the Tanford formula (l_c= 1,5+1.265n_c, where n_c is the number of carbons in the alkyl chain), these volumes translate to areas per lipid of 58.3 Å^2, 91.4 Å^2, 49.2 Å^2,86.0 Å^2, and packing parameters of 0.37, 0.46, 0.34, and 0.46, respectively. Therefore, from conventional geometric criteria, systems (**1**) and (**3**) could still pack as spherical objects ($p \approx 1/3$), while systems (**2**) and (**4**) would pack either as a cylinder or bilayer shapes. However, Connolly's surface area[69] analysis from the same trajectories revealed much larger solvent-accessible areas from the alkyl chains than expected. These areas were 125 Å^2, 237 Å^2, 140 Å^2, and 267 Å^2 for systems (**1**) through (**4**), respectively. These larger areas result from the steric and electrostatic repulsion present due to the bulky head groups that distort the apolar surface. That increase in area per head group and extended configurations of the alkyl chains for systems (**2**) ($l_c \approx$ R, the core micelle radius) allowed for the size and curvature needed for packing into a spherical object.

The inclusion of an even bulkier head group for systems (**3**) and (**4**) resulted in comparatively smaller volumes and larger solvent-accessible areas, indicating a preference towards flatter surfaces such as a cylinder, vesicle, or bilayer. This occurs because the typical length of the alkyl chain is reduced in these systems due to closer packing of the alkyl chains (length being a fraction of l_c) as seen during molecular dynamics and on SAXS data. The increase in the solvent accessible area also correlates with the more hydrophilic nature of the GCP group compared to arginine, resulting in RDFs with higher water content closer to the hydrophobic core. The more significant number of hydrogen bonds allowed by the Lys-GCP group can increase water content and closer packing.

Aldol reactions

The catalytic activity of lipopeptides (**1**-**4**) in the aldol reaction was investigated by promoting the model reaction between *p*-nitrobenzaldehyde and cyclohexanone at room temperature in water, and the results (Figures S5 and S6) are summarized in Table 2. All performed reactions furnished excellent conversion for *anti* aldol adduct (>99%) and high levels of stereoselectivity using lipopeptides concentrations equal to 5.0 and 2.5 mmol% (*ds* 93;7; *ee* 90-92%). However, a slight decrease in the stereoselectivity level was

observed when the catalysts concentrations decreased to 1.0 mmol%, as shown in Table 2.

We recently demonstrated that the appropriate packing of the amphiphilic peptide assemblies can enhance stereoselectivity in aldol reactions.[REF] Furthermore, as it is known that the water and acid additives are indispensable for the high diastereo- and enantioseloectivities in these reactions.[REF] The excellent yields with high enantiomeric excesses obtained for all lipopeptides (**1**-**4**), suggesting that the self assembled nanostructures stabilizes the reactants and intermediates in the interface of aqueous reaction medium. The nanostructures will also concentrate the reactants in the required orientation [REF] giving high levels of conversion and stereoselectivity.

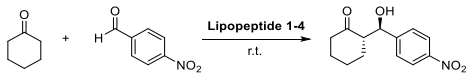
Molecular dynamics results indicate at least three distinct environments in the micellar systems, namely the apolar core, an intermediate hydrophilic region and the bulk. These different environments can affect the enamine formation equilibrium in aqueous solutions [REF].[70, 71] Trifluoroacetic acid (TFA) is present in all experimental setups as counter ions that stabilize the micelles. Therefore, we investigate the effects of the micelle environment on the protonation equilibrium promoted by TFA using Density Functional Theory calculations augmented with an implicit description of the solvent. Three different solvents, n-heptane, cyclohexanone, and water were employed to model the core, intermediate, and exterior micelle environments, respectively. The particular interest are the micelle middle region results modeled with cyclohexanone, since molecular dynamics indicate this region as the one where most of proline's reactive encounters (between cyclohexanone and proline) should happen.

Computed pK_a values in water using the micro solvation approach resulted in -0.56 and 10.02 for TFA and proline (amino group), respectively. These results compare very well to experimental ones of 0.3 and 10.6 [72], with errors of less than one pK_a unit. Moreover, we compute the environment's influence on the direct proton transfer from protonated L-proline to TFA. We hypothesize that TFA can act as a conjugate base in a proton shuttle mechanism involved in enamine formation in analogy with the Houk-List mechanism.[70, 71]

Results for the proton transfer reaction are reported in Table S4. As expected, the reaction is considerably endergonic in pure water, with a ΔG of 14.41 kcal·mol^{-1}. The computed value for the same reaction in cyclohexanone shifts this value down to 4.25 kcal·mol^{-1}. In this case, the micro-solvation approach is closer to the micro-heterogeneous environment present in the intermediate region of the micelles, where a substantial amount of water is still available, but one expects a lower dielectric medium when compared to bulk water. In n-heptane, the proton transfer reaction becomes strongly exergonic, with a ΔG of -41.12 kcal·mol^{-1}. Therefore, the closed to the micelle center, the easier it is for the proton transfer reaction to occur, favoring the aldol product formation.

Our results indicate that in regions of the micelle that are proline-rich (the intermediate realm, modeled as cyclohexanone), the proton transfer reaction is strongly enhanced compared to bulk solvent. Thus, we conclude that the micellar environment is critical in facilitating the initial steps of enamine formation by enabling the proton abstraction from proline by TFA conjugated base. We reason that, under micellar conditions, TFA possibly participates in a proton shuttle mechanism in enamine formation.

Table 2. Comparative results of the aldol reactions between *p*-nitrobenzaldehyde and cyclohexanone organocatalyzed by the lipopeptides **1**-**4**.[a]



Entry	*Lipopeptides*	*Catalyst (mol%)*	*Conv.*[b] *(%)*	*anti:syn*[c]	*ee*[d] *(%)*
1	(**1**)	5.0	>99	93:7	91
2	(**2**)	5.0	>99	94:6	93
3	(**3**)	5.0	>99	93:7	91
4	(**4**)	5.0	>99	93:7	92
5	(**1**)	2.5	>99	93:7	90
6	(**2**)	2.5	>99	93:7	90
7	(**3**)	2.5	>99	93:7	90
8	(**4**)	2.5	>99	94:6	92
9	(**1**)	1.0	>99	92:8	78
10	(**2**)	1.0	>99	93:7	85
11	(**3**)	1.0	>99	93:7	86
12	(**4**)	1.0	>99	93:7	88

[a] Reaction conditions: cyclohexanone (12 equivalent), *p*-nitrobenzaldehyde (1 equivalent), peptide (1.0 – 5.0 mol%) and water (2x cyclohexanone volume), at room temperature, for 2 days.
[b] Conversion was determined by NMR analyses of *anti* aldol adduct and referred to the combined yield of diastereomers.
[c] Diastereomeric *anti*:*syn* ratios were determined by ^{1}H NMR analysis of the crude product.
[d] Determined by chiral HPLC analysis of the *anti*-isomer.

Kinetics experiments were also performed to understand the dependence of the *anti*-aldol product conversion varying the time (0.5 – 6 h). Figure 5 shows that, initially, better conversion results were observed for the one lipid chain lipopeptides, reaching 33 % and 30% for (**1**) and (**3**), respectively. During that time, an enhancement of the sample (**4**) conversion was observed, which

obtained almost 100% after 6 h reaction, together with (**1**). However, as seen by Table 2, all samples presented excellent enantioselectivity and diastereoselectivity after two days' reaction, a condition characterized by total aldehyde consumption.

The linear curves slope by using pseudo-first-order kinetic velocity constant (*k*), considering the excess of cyclohexanone. For the (**1**-**4**) systems were found (12.1 ± 0.3) h^{-1}, (11.9 ± 0.3) h^{-1}, (8.6 ± 0.4) h^{-1} and (13.6 ± 0.2) h^{-1}, respectively. These results showed a polymorphism dependence on the aldol reaction efficiency, in which more spherical and micellar aggregates performed better conversions than the cylindrical shape. RDF in the micelle's structures observed higher water molecules access in the hydrophobic region, justifying the reaction efficiency. Also, the charge distribution in the molecule is an important parameter to be considered once higher charge content at the shell promoted more hydrophilicity comportment, as observed by SAXS, and, consequently, better aldol efficiency was verified for the system (**4**) due to the confinement of the water and the hydrophilic portion at the materials core, with the addition of the cyclohexanone excess.

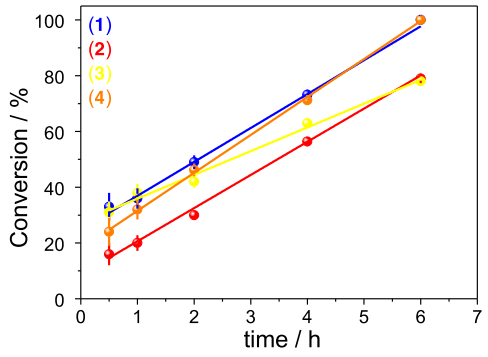


Figure 5. Comparison of the anti-aldol conversion, for systems (**1**-**4**) (2.5 mol%), during the time (0.5 h to 6 h), considering the reaction conditions: cyclohexanone (12 equivalent), *p*-nitrobenzaldehyde (1 equivalent), and water (2x cyclohexanone volume), at room temperature.

It is observed that lipopeptides with one aliphatic chain (micellar morphology, (**1**), and cylindrical, (**3**) initially present the best conversion rates (from 0.5 h to 1 h). During the reaction (2 h), lipopeptide (**4**) (which has populations of both spherical micelles and bilayers) gains prominence over the others. At the end of 6 h of reaction, it is together with micellar lipopeptide (**1**)**,** presenting the highest conversions. Therefore, cylindrical morphology from (**3**) favors only the beginning of the reaction (up to 1 h). After that time, there is a less pronounced increase in conversion, probably due to the packing factor. In cylindrical morphology, the packing is denser, providing a slower diffusion of the aldehyde within the confined system. On the other hand, for spherical micelles' morphology, diffusion is easier, leading to higher conversion rates over the reaction time. As shown in Table S1 of the SI, in 48 h, all reactions with the different lipopeptides have already reached a steady state with maximum conversion.

Conclusions

In this work, we studied the self-assembly process and the effect of the polymorphism of lipopeptides containing the PRW tripeptide attached to one or two lipid chains and in the absence or presence of the GCP group. These materials presented an enhancement of the *cac* from 3.4 x 10^{-4} wt% to 2.5 x 10^{-3} wt%, for the sequences with the GCP, suggesting a higher stronger packing of the material. The systems were characterized by β-sheet secondary structures. By Cryo-TEM and SAXS were observed spherical nanostructures for systems (**1**), (**2**), while system (**3**) presented cylindrical shape and (**4**) bilayer structure, indicating the presence of GCP group chances drastically the self-assembly of the systems. We saw that all systems act as excellent catalysts for the aldol reactions by catalytic assays, using *p*-nitrobenzaldehyde and cyclohexanone. However, the change of the hydrophobicity of the system by the lipid chain and/or by the GCP group promoted an enhancement of the packing factor for compounds (**3**) and (**4**) as shown by SAXS data and molecular dynamics simulations.

Results from molecular dynamics simulations paint an atomistic view of the internal micellar structure that offers explanations over the influence of different alkyl chains and amino acids on their network. Radial distribution functions revealed a stable layered structural organization of the residues to the micellar center for all systems. The presence of GCP lead to more tightly packed structures in the models investigated. Moreover, differences in the water content observed close to the proline position correlate with larger initial (0.5 and 1h) anti-aldol conversion percentages. We propose, based on Density Functional Calculations, that the micellar environment is critical in facilitating the initial steps in enamine formation by enabling the proton abstraction from proline by TFA. We reason that, under micellar conditions, TFA possibly participates in a proton shuttle mechanism in enamine formation. It is essential to mention that other micellar effects, such as controlling reagent

and product diffusion, can be at play, resulting in the observed differences in enantiomeric excess.

Conflict of interest

The authors declare no competing financial interests.

Author Contributions

Conceptualization, W.A.A., A.M.A., C.S. and M.D.C.N.; Lipopeptide synthesis and aldol reactions, B.M.S. and A.M.A.; Fluorescence measurements, K.B.A. and J.N.B.D.P.; Cryo-TEM, M.A.F. and R.V.P.; Molecular dynamics simulations and Quantum Mechanical Calculations, P.T.S. and M.D.C.N.; SAXS experiments, E.R.S., B.B.G. and W.A.A.; Formal analysis, W.A.A., B.M.S., A.M.A., E.R.S., B.B.G. and M.D.C.N.; Writing—original draft preparation, B.M.S. and B.B.G.; Writing—review and editing, W.A.A., A.M.A., E.R.S. and M.D.C.N.; Supervision, W.A.A.; Funding acquisition, W.A.A. All authors have read and agreed to the published version of the manuscript.

Acknowledgements

This work was supported by FAPESP (grant no. 2017/02317-2) and CNPq (grant no. 304389/2019-6). INCT in Bioanalytics (FAPESP grant no. 2014/50867-3 and CNPq grant no. 465389/2014-7) is kindly acknowledged for the grants. J.N.B.D.P. and B.B.G. acknowledge FAPESP (projects number 2015/20446-9 and 2018/05888-3) for a doctoral and post-doctoral fellowship, respectively. The staff at LNNano and LNLS are kindly recognized for invaluable help and access to TEM (project number TEM-23340) and SAXS Beamline (SAXS1- 20180103) facilities. The authors are grateful to the Multiuser Central Facilities at UFABC and UNIFESP.

References

1. R. Mammadov, A. B. Tekinay, A. Dana and M. O. Guler, *Micron (Oxford, England : 1993)*, 2012, **43**, 69-84.
2. A. Dehsorkhi, V. Castelletto and I. W. Hamley, *J. Pept. Sci.*, 2014, **20**, 453-467.
4. I. W. Hamley, *Soft Matter*, 2011, **7**, 4122-4138.
5. J. Pelin, C. J. C. Edwards-Gayle, V. Castelletto, A. M. Aguilar, W. A. Alves, J. Seitsonen, J. Ruokolainen and I. W. Hamley, *ACS Appl. Mater. Interfaces*, 2020, **12**, 13671-13679.
6. V. Castelletto, C. J. C. Edwards-Gayle, I. W. Hamley, J. N. B. D. Pelin, W. A. Alves, A. M. Aguilar, J. Seitsonen and J. Ruokolainen, *ACS Appl. Bio Mater.*, 2019, **2**, 3639-3647.
7. J. N. B. D. Pelin, B. B. Gerbelli, B. M. Soares, A. M. Aguilar and W. A. Alves, *Catal. Sci. Technol.,* 2019, **9**, 4304-4313.
8. G. Bredig and P. Fiske, *Biochem. Z*, 1912, **46**.
9. Y. Yamashita, T. Yasukawa, W.-J. Yoo, T. Kitanosono and S. Kobayashi, *Chem. Soc. Rev*, 2018, **47**, 4388-4480.
10. Z. G. Hajos and D. R. Parrish, *J. Org. Chem.*, 1974, **39**, 1615-1621.
11. U. Eder, G. Sauer and R. Weichert, *Angew. Chem., Int. Ed.*, 1971, **10**, 496-497.
12. C. Pidathala, L. Hoang, N. Vignola and B. List, *Angew. Chem., Int. Ed*, 2003, **42**, 2785-2788.
13. B. List, R. A. Lerner and C. F. Barbas, *J. Am. Chem. Soc.*, 2000, **122**, 2395-2396.
14. K. A. Ahrendt, C. J. Borths and D. W. C. MacMillan, *J. Am. Chem. Soc.*, 2000, **122**, 4243-4244.
15. A. M. Aguilar, Bruna M., J. N. B. D. Pelin and B. B. Gerbelli W. A. Alves, In: *Peptide-based Biomaterials*, ed. M. O. Guler, The Royal Society of Chemistry, 2021, ch. 5. 126-173.
16. A. J. Metrano, A. J. Chinn, C. R. Shugrue, E. A. Stone, B. Kim and S. J. Miller, *Chem. Rev.*, 2020, **120**, 11479-11615.
17. F. Rodríguez-Llansola, J. F. Miravet and B. Escuder, *Chem. Commun.*, 2009, 7303-7305.
18. F. Rodríguez-Llansola, B. Escuder and J. F. Miravet, *J. Am. Chem. Soc.*, 2009, **131**, 11478-11484.
19. B. M. Soares, A. M. Aguilar, E. R. Silva, M. D. Coutinho-Neto, I. W. Hamley, M. Reza, J. Ruokolainen and W. A. Alves, *Phys. Chem. Chem. Phys.*, 2017, **19**, 1181-1189.
20. V. Castelletto, C. J. C. Edwards-Gayle, I. W. Hamley, J. N. B. D. Pelin, W. A. Alves, A. M. Aguilar, J. Seitsonen and J. Ruokolainen, *ACS Appl. Bio Mater.*, 2019, **2**, 3639-3647.
21. M. Z. Shao, Q. X. Jin, L. Zhang and M. H. Liu, *Chem. Lett.,* 2012, **41**, 1349-1350.
22. O. Zozulia, M. A. Dolan and I. V. Korendovych, *Chem. Soc. Rev*., 2018, **47**, 3621-3639.
23. P. R. Arivalagan and Y. Zhao, *Org. Biomol. Chem*., 2015, **13**, 770-775.
24. N. Mase, N. Noshiro, A. Mokuya and K. Takabe, *Adv. Synth. Catal*., 2009, **351**, 2791-2796.
25. J. N. B. D. Pelin, C. J. C. Edwards-Gayle, A. M. Aguilar, A. Kaur, I. W. Hamley and W. A. Alves, *Soft Matter*, 2020, **16**, 4615-4624.
26. C. Schmuck, *Chem. Eur. J.*, 2000, **6**, 709-718.

27. M. Li, M. R. Stojkovic, M. Ehlers, E. Zellermann, I. Piantanida and C. Schmuck, Angew. Chem., Int. Ed., 2016, 55, 13015-13018.
28. A. P. Hammersley, *J. Appl. Crystallogr.*, 2016, **49**, 646-652.
29. I. Bressler, J. Kohlbrecher and A. F. Thünemann, *J. Appl. Crystallogr.*, 2015, **48**, 1587-1598.
30. A. A. Skjevik, B. D. Madej, R. C. Walker and K. Teigen, *J. Phys. Chem. C,* 2012, **116**, 11124-11136.
31. C. J. Dickson, B. D. Madej, A. A. Skjevik, R. M. Betz, K. Teigen, I. R. Gould and R. C. Walker, *J. Chem. Theory Comput.,* 2014, **10**, 865-879.
32. I. R. Gould, Skjevik A.A., Dickson, C.J., Madej, B.D., Walker, R.C., 2018, in prep.
33. J. A. Maier, C. Martinez, K. Kasavajhala, L. Wickstrom, K. E. Hauser and C. Simmerling, *J. Chem. Theory Comput.*, 2015, **11**, 3696-3713.
34. J. M. Wang, R. M. Wolf, J. W. Caldwell, P. A. Kollman and D. A. Case, *J. Comput. Chem.*, 2004, **25**, 1157-1174.
35. C. I. Bayly, P. Cieplak, W. D. Cornell and P. A. Kollman, *J. Phys. Chem.,* 1993, **97**, 10269-10280.
36. D. A. Case, T. E. Cheatham, T. Darden, H. Gohlke, R. Luo, K. M. Merz, A. Onufriev, C. Simmerling, B. Wang and R. J. Woods, *J. Comput. Chem.*, 2005, **26**, 1668-1688.
37. D.A. Case, I.Y. Ben-Shalom, S.R. Brozell, D.S. Cerutti, T.E. Cheatham, III, V.W.D. Cruzeiro, T.A. Darden, R.E. Duke, D. Ghoreishi, M.K. Gilson, H. Gohlke, A.W. Goetz, D. Greene, R Harris, N. Homeyer, Y. Huang, S. Izadi, A. Kovalenko, T. Kurtzman, T.S. Lee, S. LeGrand, P. Li, C. Lin, J. Liu, T. Luchko, R. Luo, D.J. Mermelstein, K.M. Merz, Y. Miao, G. Monard, C. Nguyen, H. Nguyen, I. Omelyan, A. Onufriev, F. Pan, R. Qi, D.R. Roe, A. Roitberg, C. Sagui, S. Schott-Verdugo, J. Shen, C.L. Simmerling, J. Smith, R. Salomon- Ferrer, J. Swails, R.C. Walker, J. Wang, H. Wei, R.M. Wolf, X. Wu, L. Xiao, D.M. York and P.A. Kollman (2018), AMBER 2018, University of California, San Francisco.
38. P. Dinér, M. Nielsen, M. Marigo and K. A. Jorgensen, *Angew. Chem. Int. Ed.*, 2007, **46**, 1983-1987.
39. L. Martinez, R. Andrade, E. G. Birgin and J. M. Martinez, *J. Comput. Chem.*, 2009, **30**, 2157-2164.
40. J. M. Martinez and L. Martinez, *J. Comput. Chem.*, 2003, **24**, 819-825.
41. A. W. Gotz, M. J. Williamson, D. Xu, D. Poole, S. Le Grand and R. C. Walker, *J. Chem. Theory Comput.*, 2012, **8**, 1542-1555.
42. R. Salomon-Ferrer, D. A. Case and R. C. Walker, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2013, **3**, 198-210.
43. H. J. C. Berendsen, J. P. M. Postma, W. F. Vangunsteren, A. Dinola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684-3690.
44. R. J. Loncharich, B. R. Brooks and R. W. Pastor, *Biopolymers*, 1992, **32**, 523-535.
45. J. P. Ryckaert, G. Ciccotti and H. J. C. Berendsen, *J. Comput. Phys.*, 1977, **23**, 327-341.
46. D. R. Roe and T. E. Cheatham, *J. Chem. Theory Comput.*, 2013, **9**, 3084-3095.
47. W. Humphrey, A. Dalke and K. Schulten, *J. Mol. Graphics Modell.*, 1996, **14**, 33-38.
48. E. Rossini and E. W. Knapp, *J. Comput. Chem*, 2016, **37**, 2163-2164.
49. Z. Szakacs and B. Noszal, *J. Math. Chem.*, 1999, **26**, 139-155.
50. A. Galano, A. Perez-Gonzalez, R. Castaneda-Arriaga, L. Munoz-Rugeles, G. Mendoza-Sarmiento, A. Romero-Silva, A. Ibarra-Escutia, A. M. Rebollar-Zepeda, J. R. Leon-Carmona, M. A. Hernandez-Olivares and J. R. Alvarez-Idaboy, *J. Chem. Inf. Model.* 2016, **56**, 1714-1724.
51. F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297-3305.
52. J. J. Zheng, X. F. Xu and D. G. Truhlar, *Theor. Chem. Acc.*, 2011, **128**, 295-305.
53. P. J. Stephens, F. J. Devlin, C. F. Chabalowski and M. J. Frisch, *J. Phys. Chem.*, 1994, **98**, 11623-11627.
54. A. D. Becke, *J. Phys. Chem.*, 1993, **98**, 5648-5652.
55. C. T. Lee, W. T. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785-789.
56. S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Phys. Chem.,* 2010, **132,** 154104-154118.
57. S. Grimme, S. Ehrlich and L. Goerigk, *J. Chem. Comput.*, 2011, **32**, 1456-1465.
58. F. Neese, *Wiley Interdiscip. Rev.: Comput. Mol. Sci.*, 2018, **8,** e1327.
59. A. V. Marenich, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem.B*, 2009, **113**, 6378-6396.
60. C. P. Kelly, C. J. Cramer and D. G. Truhlar, *J. Phys. Chem. A*, 2006, **110**, 2493-2499.
61. J. S. Pedersen, *Adv. Colloid Interface Sci.*, 1997, **70**, 171-210.
62. D. I. Svergun and M. H. J. Koch, *Rep. Prog. Phys.*, 2003, **66**, 1735-1782.

63. G. Pabst, M. Rappolt, H. Amenitsch and P. Laggner, *Phys. Rev. E*, 2000, **62**, 4000-4009
64. V. Castelletto, R. M. Gouveia, C. J. Connon and I. W. Hamley, *Faraday Discuss.*, 2013, **166**, 381-397.
65. V. Castelletto and I. W. Hamley, *Methods Mol. Biol.*, 2018, **1777**, 3-21.
66. C. Tanford, *J. Phys. Chem.*, 1972, **76**, 3020-&.
67. J. Israelachvli, *Biochim. Biophys. Acta*, 1977, **469**, 221-225.
68. D. J. Mitchell and B. W. Ninham, *J. Chem. Soc., Faraday Trans.*, 1981, **77**, 601-629.
69. M. L. Connolly, *J. Appl. Crystallogr.,* 1983, **16**, 548-558.
70. F. R. Clemente and K. N. Houk, *Angew. Chem., Int. Ed*, 2004, **43**, 5766-5768.
71. A. Armstrong, R. A. Boto, P. Dingwall, J. Contreras-Garcia, M. J. Harvey, N. J. Mason and H. S. Rzepa, *Chem. Sci.*, 2014, **5**, 2057-2071.
72. G. C. Fonger, *Toxicology*, 1995, **103**, 137-145.