



Cite this: *Soft Matter*, 2020,  
16, 10122

## A spectroscopic and molecular dynamics study on the aggregation process of a long-acting lipidated therapeutic peptide: the case of semaglutide<sup>†</sup>

M. Venanzi, <sup>a</sup> M. Savioli, <sup>a</sup> R. Cimino, <sup>a</sup> E. Gatto, <sup>a</sup> A. Palleschi, <sup>a</sup> G. Ripani, <sup>a</sup> D. Cicero, <sup>a</sup> E. Placidi, <sup>b</sup> F. Orvieto<sup>c</sup> and E. Bianchi<sup>c</sup>

The aggregation properties of semaglutide, a lipidated peptide drug agonist of the Glucagon-like peptide 1 receptor recently approved for the treatment of type 2 diabetes, have been investigated by spectroscopic techniques (UV-Vis absorption, steady-state and time-resolved fluorescence, and electronic circular dichroism) and molecular dynamics simulations. We show that in the micromolar concentration region, in aqueous solution, semaglutide is present as monomeric and dimeric species, with a characteristic monomer-to-dimer transition occurring at around 20 μM. The lipid chain stabilizes a globular morphology of the monomer and dimer species, giving rise to a locally well-defined polar outer surface where the lipid and peptide portions are packed to each other. At very long times, these peptide clusters nucleate the growth of larger aggregates characterized by blue luminescence and a β-sheet arrangement of the peptide chains. The understanding of the oligomerization and aggregation potential of peptide candidates is key for the development of long acting and stable drugs.

Received 31st May 2020,  
Accepted 22nd July 2020

DOI: 10.1039/d0sm01011a

rsc.li/soft-matter-journal

## Introduction

Peptides represent a unique class of pharmaceutical compounds able to target protein–protein interactions where large and shallow surface areas are considered that cannot be easily targeted by conventional small molecule drugs.

Peptides are intrinsic signalling molecules for many physiological functions. In particular, in the case of the peptide G protein-coupled receptor (GPCR), the receptor/peptide hormone interaction site is quite large and requires specific conformational changes for signal transduction. Peptide therapeutics represent a unique opportunity for the cases where high selectivity and strong receptor affinity are required.<sup>1,2</sup> For this purpose novel strategies allow now to overcome the well-known liabilities of peptides, such as low metabolic stability, high clearance rates due to rapid renal filtration and low bioavailability.

In particular, different approaches allow for the modulation of the pharmacokinetic profile through the incorporation of

non-natural amino acids and, most importantly, with the derivatization/conjugation of half-life extension moieties, such as fatty acids,<sup>3</sup> or cholesterol,<sup>4,5</sup> which promote complexation with plasma proteins such as human serum albumin (HSA) or promote self-assembly to form larger structures. Peptides bound to HSA are sterically shielded from proteolytic degradation and the relatively large size of HSA (66 kDa) protects against rapid renal filtration.

The mechanism underlying the extended half-life of lipidated peptides involves both albumin binding and protraction of absorption by oligomerization. The first lipidated pharmaceuticals that were developed were analogues of the natural peptide hormones, insulin and Glucagon-like peptide 1 (GLP-1). The molecules differed in the type of fatty acids and the kind and length of hydrophilic spacers such as γGlu or PEG2 (8-amino-3,6-dioxaoctanoic acid) between the peptide sequence and the fatty acid, so as to modulate affinity to HSA, oligomerization and target receptor potency. Two successful examples for lipidated GLP-1 analogues are liraglutide,<sup>6</sup> and semaglutide.<sup>7</sup>

GLP-1 is a 30 amino acid hormone controlling the glucose levels by stimulating pancreatic secretion of insulin.<sup>8</sup> GLP-1 undergoes a self-limiting action, inhibiting the insulin production when the glucose reaches normal levels in the blood, therefore reducing hypoglycemic risks. Unfortunately, its residence time in the plasma is usually very short (1–2 minutes) due to fast enzymatic degradation by dipeptidyl-peptidase 4 (DPP-IV).<sup>9</sup>

Liraglutide is a once daily analogue that is approved for antidiabetic and obesity therapies. Starting from natural GLP-1,

<sup>a</sup> Dept. of Chemical Science and Technologies, University of Rome Tor Vergata,  
Via Ricerca Scientifica, 1, 00133 Rome, Italy. E-mail: venanzi@uniroma2.it

<sup>b</sup> Dept. of Physics, University of Rome 'La Sapienza', P.le Aldo Moro 5, 00185 Rome, Italy

<sup>c</sup> Peptide and Small Molecule R&D Department, IRBM SpA, via Pontina KM 30,600,  
00071 Pomezia, Italy

† Electronic supplementary information (ESI) available: Optical spectroscopy characterization of the SMG aggregation process (Fig. S1–S4; Tables T1 and T2); size distribution of SMG nanostructures imaged by AFM (Fig. S5–S7); and Pasternack modelling of SMG long-time aggregation. See DOI: 10.1039/d0sm01011a

extensive structure activity relationship studies led to: (a) the identification of most tolerable positions for fatty acid derivatization; (b) optimization of the spacer, such as  $\gamma$ Glu, to compensate for the lack of the acidic group when the fatty acid is amidated to a Lys side chain; and (c) length of the fatty acid, with the finding that only lipids above a certain length ( $>\text{C12}$ ) could translate to half-life longer than 10 hours (subcutaneous route).<sup>10</sup>

Liraglutide has a Lys to Arg substitution at position 34 while Lys(26) is derivatized on its side chain through a  $\gamma$ Glu spacer and a palmitoyl group. The presence of the C16 fatty acid promotes binding to HSA and protracted absorption, resulting in an extended half-life in humans (11–15 h) by subcutaneous administration.<sup>10</sup>

Further improvement in the use of fatty acids for conjugation to peptides as a half-life extending approach was recently achieved with the implementation of fatty diacids for derivatizing peptides with a long carbon chain ending at one terminus with a free carboxylate (C16-OH, C18-OH, and C20-OH). The presence of the carboxylate group on the fatty acid enables an improved interaction with HSA resulting in an extended half-life of one week in humans such as for semaglutide, a new GLP-1 analogue.<sup>7</sup>

The long retention times of lipidated peptides in the blood have been explained by enhanced resistance to renal secretion promoted either by the formation of micelle-like peptide nanostructures, or with the association to HSA. As the lipidated peptide is thought to bind HSA exclusively as a monomeric species, the two mechanisms appear to be competitive, as the binding to HSA subtracts peptide monomers to the aggregation equilibria leading to micellization.<sup>11,12</sup> Also self-association to form either amorphous aggregates or highly structured fibrillar species may limit their use, therefore careful inspection of this propensity should be part of the discovery process.<sup>13</sup>

The chemical structure of semaglutide in the following denoted as **SMG** for brevity is reported in Scheme 1. In **SMG**, an Ala residue in position 8 was substituted with a non-metabolic residue, *i.e.*  $\alpha$ -amino isobutyric acid (Aib), and the side chain of Lys(26) was derivatized by a spacer comprising two PEG2 groups, a  $\gamma$ -glutamyl residue, and a C18-OH lipid chain. The fat lipid chain is thought to increase the affinity of **SMG** to HSA, while the non-metabolic Aib residue is expected to increase its resistance to enzymatic degradation by DPP-IV.<sup>7</sup>

The aim of this contribution is to study the aggregation properties of **SMG** in aqueous solutions, determining its critical

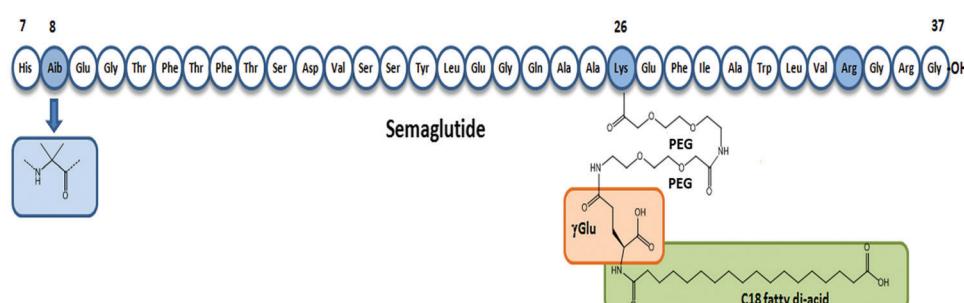
aggregation concentration and characterizing by spectroscopic methods and molecular dynamics simulations the formation of small supramolecular structures nucleating the growth of large **SMG** aggregates. The presence of two fluorescent amino acids, *i.e.* Trp (W) and Tyr (Y), in the **SMG** peptide chain allowed us to apply optical spectroscopy techniques to characterize the aggregation properties of **SMG**.

## Materials and methods

### Synthesis of semaglutide

The sequence His-Aib-Glu-Gly-Thr-Phe-Thr-Ser-Asp-Val-Ser-Ser-Tyr-Leu-Glu-Gly-Gln-Ala-Ala-Lys[(C18acid- $\gamma$ -Glu-PEG2-PEG2)]-Glu-Phe-Ile-Ala-Trp-Leu-Val-Arg-Gly-Arg-Gly-OH was synthesized by a stepwise solid phase assembly on a 2-chlorotriptyl chloride resin using the Fmoc-tBu strategy. The first amino acid Fmoc-Gly-OH (1 equivalent) was loaded on the resin by incubation in a DMF solution in the presence of 2 equivalents of *N,N*-diisopropyl-ethylamine. The sequence assembly was performed on a Symphony (Protein Technologies Inc.) synthesizer with Lys(26) incorporated as Fmoc-Lys(Dde)-OH so as to enable side chain derivatization in the solid phase with two PEG2 (8-amino-3,6-dioxaoctanoic acid) units,  $\gamma$ -glutamic acid and saturated fatty diacid C18-OH. The cleavage from the resin and side chain deprotection was performed by incubation with a TFA solution containing 5% phenol, 2.5% triisopropylsilane and 5% water. The crude peptide was precipitated in methyl-*tert*-butyl ether. The lyophilized crude peptide was purified by reverse phase chromatography on a C4 column (Reprosil, 200 Å, 5 µm, Dr Maisch GmbH) on a Waters HPLC system. The analysis was performed by ultra-high performance liquid chromatography-UV-mass spectrometry (UPLC-UV-MS) on a Waters Acuity UPLC system equipped with a Waters BEH130 C4 (2.1 × 100 mm, 1.7 µm, at 45 °C) column. Analysis was typically performed over 2, 5 or 10 min runs as required applying the linear gradients of acetonitrile in H<sub>2</sub>O, 0.1% TFA with a flow rate of 0.4 mL/min and UV detection at 214 nm. Mass analysis was performed on a Waters SQ or SQ2 detector with electrospray ionization in the positive ion detection mode. The scan range of the mass-to-charge ratio was 1000–3000.

UV-Vis absorption measurements were carried out using a Cary100 Scan spectrophotometer using asymmetric quartz cuvettes with 4/10 mm optical length at room temperature (25 °C). All the



**Scheme 1** Schematic view of the semaglutide (**SMG**) structure.

absorption spectra were corrected by subtracting the absorption spectrum of the buffer solution in the same wavelength range (200–350 nm). The concentration of freshly prepared solutions was obtained from the Lambert–Beer equation, assuming the molar extinction coefficient at  $\lambda = 280$  nm  $\varepsilon(280) = 7200 \text{ M}^{-1} \text{ cm}^{-1}$ , as a result of the sum of the Trp [ $\varepsilon(280) = 5600 \text{ M}^{-1} \text{ cm}^{-1}$ ] and Tyr [ $\varepsilon(280) = 1600 \text{ M}^{-1} \text{ cm}^{-1}$ ] molar extinction coefficients.<sup>14</sup> All spectroscopic experiments were carried out at pH 8 in phosphate buffer solutions to match the slightly alkaline environmental conditions of pancreatic juice.

Steady-state fluorescence experiments were carried out on a thermostatted (25 °C) 4/10 mm asymmetric quartz cuvette using a Fluoromax-4 spectrofluorometer (Horiba, Jobin Yvon) with single photon counting (SPC) detection. Fluorescence emission spectra were measured at excitation wavelengths of 280 and 295 nm, while the associated excitation spectra were measured at emission wavelengths of 350, 360 and 410 nm. 2D emission/excitation fluorescence maps were obtained by collecting 20 emission spectra recorded by varying the excitation wavelengths between 250 and 300 nm, and the emission wavelengths between 300 and 500 nm.

Fluorescence anisotropy measurements were carried out on the same apparatus equipped with automatically driven Glan–Thompson polarizers. The position of the excitation and emission polarizers was calibrated using a Ludox scattering solution. The anisotropy coefficient  $r$  was obtained by the equation:

$$r = \frac{I_{\parallel} - I_{\perp}}{I_{\parallel} + 2I_{\perp}} \quad (1)$$

where  $I_{\parallel}$  and  $I_{\perp}$  are the fluorescence intensities measured where the excitation and emission polarizers are oriented to each other in a parallel or perpendicular position, respectively.

Rotational correlation times were obtained through the Perrin equation for a spherical rotor<sup>15</sup>

$$\frac{r_0}{r} = 1 + \frac{\tau}{\theta} \quad (2)$$

where  $r_0$  is the limit anisotropy coefficient at  $t = 0$ , an intrinsic property of the fluorophore,  $\tau$  the fluorescence lifetime of the fluorophore, and  $\theta$  the rotational correlation time. The latter is connected to the hydrodynamic volume  $V_h$  of a spherical rotor through the Einstein–Smoluchowski equation:<sup>15</sup>

$$\theta = \frac{\eta V_h}{RT} \quad (3)$$

where  $\eta$  is the solution bulk viscosity.

The critical aggregation concentration (cac) of **SMG** was determined through the pyrene fluorescence assay, measuring the relative intensities of the first [ $S_2(v=0) \rightarrow S_0(v=0)$ ,  $\lambda_{\max} = 373$  nm] and third [ $S_2(v=0) \rightarrow S_0(v=1)$ ,  $\lambda_{\max} = 384$  nm] vibronic transitions.<sup>16</sup> The titration experiment was carried out adding at each step 25 µl of 66 µM **SMG** to a 1 µM pyrene solution (phosphate buffer, pH 8, 25 °C).

Time-resolved fluorescence experiments were carried out on a thermostatted (25 °C) asymmetric (4/10 mm) quartz cuvette with an EAI Spec-ps (Edinburgh Analytical Instruments) spectrophotometer with time-correlated SPC detection. Excitation at

295 and 344 nm was performed using two Hamamatsu diode lasers (pulse width 1 ns). Experimental fluorescence time decays were analyzed by an iterative deconvolution method, using a multiexponential fitting function, *i.e.*

$$I(t) = \sum_i \alpha_i \exp\left(-\frac{t}{\tau_i}\right) \quad (4)$$

where  $\alpha_i$  are the pre-exponential weighting factors, proportional to the abundance of the  $i$ th species at  $t = 0$ , and  $\tau_i$ , the lifetime of the  $i$ th time decay component.

Circular dichroism experiments were carried out using a Jasco J-1500 CD spectropolarimeter (Jasco International Co.) in the wavelength range from 190 to 250 nm, at 25 °C using an equipped Peltier thermostatted cell holder (PTC-510). Measurements were performed in a 0.1 cm quartz cuvette (Hellma, Mühlheim, Germany) using a scan speed of 50 nm min<sup>-1</sup>, a bandwidth of 1 nm, and an integration time of 1 s. Each spectrum was obtained as the average of 4 repeated scans.

Atomic force microscopy experiments were carried out in air at room temperature on a Veeco Multitip IIIa instrument (Santa Barbara, CA). The AFM measurements were carried out in tapping mode on films obtained by drop casting micromolar aqueous (phosphate buffer, pH 8) solutions of **SMG** on mica, and incubating for 18 hours in a dryer. For the AFM measurements, a Si super sharp tip, functionalized by carbon flake, was used (curvature 1 nm, elastic constant 5 N m<sup>-1</sup>, and resonance frequency 150 kHz).

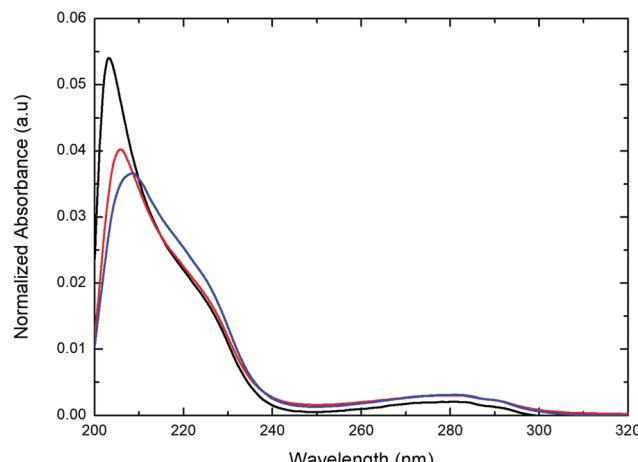
MD simulations of aggregates have been performed placing, in a  $5.13 \times 13.38 \times 8.43 \text{ nm}^3$  box, 4 **SMG** molecules, 18 145 water molecules, and 16 Na<sup>+</sup> counter-ions to neutralize the total charge. All the simulations have been performed using the gromos53a6-FF force field<sup>17</sup> (modified for the LysP chain) with the MD simulation package GROMACS16 MPI.<sup>18</sup> Time steps of 2 fs have been used for the equilibration and production runs. The equilibrium temperature (300 K) was controlled by velocity rescaling with a coupling constant of 0.6 ps. The pressure was controlled using an isotropic Berendsen barostat<sup>19</sup> with a coupling constant of 1 ps and a reference external pressure of 1 atm. Lennard-Jones long range interactions were treated with a cut-off radius of 1.4 nm and coulombic interactions were calculated by the Particle Mesh Ewald method.<sup>20</sup> Three replicas of each system have been performed for 30 ns long simulations with different starting configurations.

## Results

### Aggregation of **SMG** at early times

The absorption spectrum of **SMG** in the near UV region (Fig. 1) is characterized by the overlap of the  $\pi \rightarrow \pi^*$  ( $\lambda_{\max} = 205$  nm) and  $n \rightarrow \pi^*$  ( $\lambda_{\max} = 280$  nm) transitions of chromophores in the sequence (Scheme 1): the indole group of Trp(31) and the phenol group of Tyr(19).

From Fig. 1 it can be observed that, in the **SMG** concentration region comprised between 20 and 30 µM, a hypochromic effect and slight broadening of the  $\pi \rightarrow \pi^*$  transition take place,



**Fig. 1** Absorption spectra of **SMG** in aqueous solution (phosphate buffer, pH = 8). Black: 4  $\mu\text{M}$ ; red: 20  $\mu\text{M}$ ; and blue: 30  $\mu\text{M}$ . The spectra were normalized to unit area to emphasize differences in the shape.

as usually observed in aggregation phenomena promoted by the stacking of aromatic moieties.<sup>1,21,22</sup>

Fluorescence emission spectra of **SMG** were measured either at the excitation wavelengths of 280 nm, where both Trp and Tyr chromophores can be excited, and 295 nm, where Trp is selectively excited. Interestingly, the emission spectra obtained at the two excitation wavelengths closely overlap (ESI,† Fig. S1), showing in both cases wavelength emission maxima peaked at  $\lambda = 350$  nm. In agreement with this finding, 2D emission/excitation fluorescence maps of freshly prepared **SMG** solutions in the micromolar concentration region show the presence of a single emitting species, confirming the occurrence of a total Tyr\*  $\rightarrow$  Trp\* energy transfer (Fig. 2).

Fluorescence emission spectra, normalized by the absorbance at the excitation wavelength ( $\lambda_{\text{ex}} = 295$  nm), showed decreasing intensities upon increasing the **SMG** molar concentration along the series 4 > 30 > 20  $\mu\text{M}$  (ESI,† Fig. S2), strengthening the idea that an aggregation process takes place in the micromolar concentration region investigated, leading to the quenching of the Trp emission.

**Table 1** Lifetimes and pre-exponential weights obtained by iterative deconvolution of the **SMG** fluorescence time decay at different concentrations (phosphate buffer, pH 8,  $T = 25$  °C;  $\lambda_{\text{exc}} = 295$  nm,  $\lambda_{\text{em}} = 350$  nm)

[ <b>SMG</b> ] ( $\mu\text{M}$ )	$\alpha_1$	$\tau_1$ (ns)	$\alpha_2$	$\tau_2$ (ns)	$\alpha_3$	$\tau_3$ (ns)	$\langle \tau \rangle$ (ns)
4	0.36	1.0	0.50	3.4	0.14	6.5	3.0
20	0.38	1.4	0.45	3.5	0.16	6.7	3.2
30	0.38	1.2	0.45	3.5	0.17	6.9	3.2

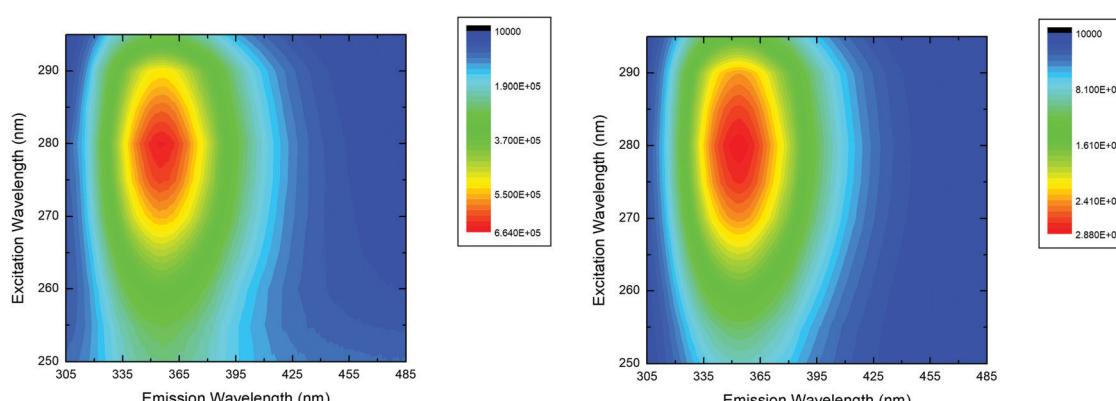
Time resolved fluorescence measurements were carried out exciting selectively the Trp fluorophore ( $\lambda_{\text{exc}} = 295$  nm,  $\lambda_{\text{em}} = 350$  nm) for **SMG** concentrations comprised between 4 and 30  $\mu\text{M}$ . The lifetimes and pre-exponential weights obtained from iterative deconvolution of the **SMG** fluorescence time decays are reported in Table 1.

The time-resolved fluorescence emission of single Trp proteins is generally accounted for by a biexponential decay, showing a shorter time component at around 0.5–1 ns and a longer lifetime at around 3.5 ns.<sup>23</sup> This behaviour was explained in terms of the rotamer model that assigns each lifetime to a specific conformer.<sup>24</sup> In some oligopeptides, the deconvolution of Trp fluorescence time decay required a third lifetime component lasting some 5 to 10 ns. This slow time decay was found in some constrained peptides in which the indole group is buried in a hydrophobic pocket or experiences restricted dynamics.<sup>25,26</sup>

In our case, three exponential time components were needed to adequately reproduce the Trp experimental decays (Table 1), independently of the peptide concentration (at least in the micromolar range investigated). This finding clearly suggests a certain conformational heterogeneity of Trp in **SMG**, most likely arising from the presence of monomer/aggregate equilibria. We are tempted to assign the slow time component of the Trp emission to a hydrophobic environment protecting the fluorophore from the solvent.

Steady-state fluorescence anisotropy measurements were carried out to determine the fluorescence anisotropy coefficient  $r$ , and, through the Perrin (eqn (3)) and Einstein–Smoluchowski (eqn (4)) equations, the rotational diffusion coefficient and hydrodynamic volume of **SMG** at micromolar concentrations.

In Fig. 3 we report the anisotropy coefficient  $r$  as a function of the temperature/viscosity ratio ( $T/\eta$ ) measured for **SMG**



**Fig. 2** 2D emission/excitation fluorescence maps of **SMG** aqueous solutions (phosphate buffer, pH 8,  $T = 25$  °C). Left: 4  $\mu\text{M}$  and right: 30  $\mu\text{M}$ .

water/glycerol 1 : 2 (v/v) solutions. The time decay parameters of **SMG** under the same experimental conditions are reported in the ESI<sup>†</sup> in Table T1.

It should be noted that the  $r$  vs.  $T/\eta$  behaviour is highly non-linear, suggesting some conformational flexibility of the region sampled by Trp. In the presence of segmental flexibility, the integrated anisotropy coefficient  $r$  can be described by the equation:<sup>15</sup>

$$r(\tau) = \frac{\alpha r_0}{1 + \left( \frac{1}{\theta_F} + \frac{1}{\theta_S} \right) \tau} + \frac{(1 - \alpha)r_0}{1 + \frac{\tau}{\theta_S}} \quad (5)$$

where  $\theta_F$  and  $\theta_S$  represent the fast and slow time components of the anisotropy time decay, associated with the local motion of the tryptophan residue and the global rotation of **SMG**, respectively; and  $\alpha$  and  $(1 - \alpha)$  are the relative weights of the two rotational components.

It should be noted that the anisotropy coefficient  $r$  is a function of  $\tau$ , the fluorescence lifetime of the fluorophore. Substituting for the two  $\theta_F$  and  $\theta_S$  time components by the Einstein-Smoluchowski equation, we obtain:

$$r(\tau) = \frac{\alpha r_0}{1 + \left( \frac{1}{V_F} + \frac{1}{V_S} \right) \frac{RT\tau}{\eta}} + \frac{(1 - \alpha)r_0}{1 + \frac{1}{V_S} \frac{RT\tau}{\eta}} \quad (6)$$

where  $V_F$  and  $V_S$  are the hydrodynamic volumes of the spherical rotors associated with the fast and slow components of the anisotropy decay. The data reported in Fig. 3 were fitted by eqn (6), providing an  $r_0$  value equal to 0.260. Interestingly, under these experimental conditions, *i.e.* high viscosity water/glycerol 1 : 2 solutions, the segmental flexibility accounts for around 30% of the overall rotational motion depolarizing the fluorescence anisotropy.

Fluorescence anisotropy coefficients measured at different **SMG** molar concentrations, *i.e.* 4, 20 and 30  $\mu\text{M}$ , in pH 8 aqueous solutions are reported in Table 2. Using the average time decays and the limit anisotropy coefficient ( $r_0$ ) determined above, and

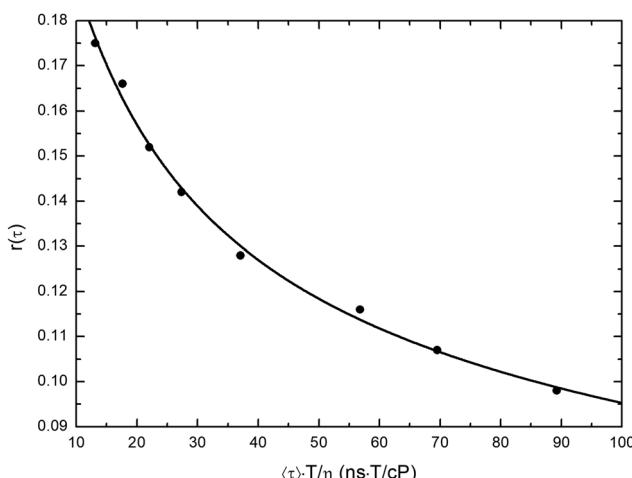


Fig. 3 Anisotropy coefficient  $r$  of 20  $\mu\text{M}$  **SMG** in water/glycerol 1 : 2 (v/v) as a function of the temperature/viscosity ratio.

Table 2 Anisotropy coefficients ( $r$ ), rotational correlation time ( $\Theta$ ), hydrodynamic volume ( $V_h$ ) and radii ( $r_h$ ) at different **SMG** molar concentrations (phosphate buffer, pH 8,  $T = 25^\circ\text{C}$ ;  $\lambda_{\text{exc}} = 295\text{ nm}$ ,  $\lambda_{\text{em}} = 350\text{ nm}$ )

[SMG] ( $\mu\text{M}$ )	$r$	$\Theta$ (ns)	$V_h$ ( $\text{nm}^3$ )	$r_h$ ( $\text{\AA}$ )
4	$0.072 \pm 0.005$	$1.15 \pm 0.06$	$5.3 \pm 0.3$	$10.8 \pm 0.1$
20	$0.089 \pm 0.001$	$1.67 \pm 0.06$	$7.7 \pm 0.3$	$12.3 \pm 0.1$
30	$0.096 \pm 0.001$	$1.87 \pm 0.06$	$8.6 \pm 0.4$	$12.7 \pm 0.1$

assuming that, under these low viscosity experimental conditions, the measured residual anisotropy is entirely ascribable to the **SMG** global rotation, the rotational correlation time ( $\theta$ ) and the hydrodynamic volume ( $V_h$ ) can be obtained by eqn (2) and (3), respectively. Hydrodynamic radii at different **SMG** concentrations, also reported in Table 2, were obtained by considering **SMG** as a rigid spherical rotor, *i.e.*

$$r_h = \sqrt[3]{\frac{3V_h}{4\pi}} \quad (7)$$

The increase of the hydrodynamic volume observed for increasing **SMG** concentrations suggests the formation of small oligomers at higher concentrations investigated.

The critical aggregation concentration (cac) of **SMG** in a phosphate buffer solution (pH 8) was obtained by the well-established pyrene assay. The latter is based on the dependence of the ratio between the intensities of the first and third vibronic components ( $I_1/I_3$ ) of the pyrene emission band on the environment polarity.<sup>16</sup> Specifically, when the pyrene is embedded in an apolar environment, the  $I_1/I_3$  ratio strongly decreases. Increasing amounts of **SMG** were therefore added to a 1  $\mu\text{M}$  pyrene aqueous solution, collecting the pyrene fluorescence emission spectra reported in the ESI<sup>†</sup> (Fig. S3). In Fig. 4 the ratio of the intensities of  $I_1$  and  $I_3$  vibronic bands, centred at  $\lambda_{\text{em}} = 373$  and 384 nm, respectively, is reported as a function of the **SMG** concentration (on a logarithmic scale).

The cac can be determined at the intersection point of the two straight lines associated with the concentration regions

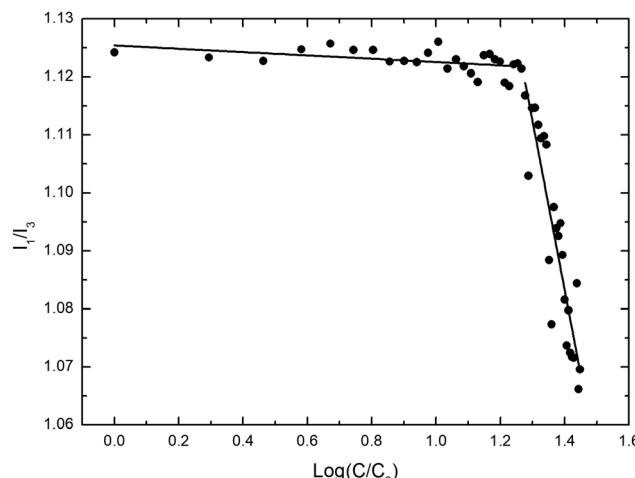


Fig. 4 Ratio of the fluorescence intensities of the first ( $\lambda_{\text{em}} = 373\text{ nm}$ ) and third ( $\lambda_{\text{em}} = 384\text{ nm}$ ) vibronic components as a function of the **SMG** concentration (log scale).

where the pyrene fluorophore is fully exposed to water molecules or it interacts with the peptide aggregate, respectively. From the data reported in Fig. 4, a cac value of 20.2  $\mu\text{M}$  can be obtained, in agreement with the UV-Vis absorption and fluorescence results reported above that showed a similar abrupt change in the same concentration region.

**Molecular dynamics simulation.** MD simulations have been carried out to investigate the initial steps of **SMG** aggregation. During the 30 ns-long MD simulations three different structures have been observed, namely **SMG** monomers, dimers and trimers (Fig. 5).

Larger structures were never observed during the sampled simulation time. During the simulation, the monomer showed a variety of structures featuring coil (57%), beta-sheet (16%), bend (18%), and turn (6%) conformations. This conformational landscape is essentially conserved in the dimer and trimer structures.

In the structures reported in Fig. 5, charged peptide side chains are fully exposed to the solvent, while the lipidated chains (LysP) appear folded and packed towards the peptide chains, and only partially accessible to water molecules. Noteworthy, the carboxylate groups of the lipid chains protrude towards the solvent exposed surface, stabilizing the oligomer predominant structure. Interestingly, in all the simulations the side-chain phenyl group of Phe(6) stays rather close to the indole group of Trp(25). Besides that, the side-chain groups of the aromatic residues appear to be tightly arranged in the

**Table 3** Rotational autocorrelation times ( $\Theta$ ), gyration ( $r_g$ ) and hydrodynamic ( $r_h$ ) radii of **SMG** monomers, dimers and trimers

System	$\Theta$ (ns)	$r_g$ (Å)	$r_h$ (Å)
Monomer	$1.12 \pm 0.25$	$9.9 \pm 0.5$	$10.4 \pm 0.3$
Dimer	$2.04 \pm 0.86$	$12.1 \pm 0.3$	$12.8 \pm 0.3$
Trimer	$5.38 \pm 2.35$	$21.0 \pm 2.0$	$17.0 \pm 0.1$

oligomerized structures, as also suggested by the observed hypsochromic effect shown by the UV absorption spectra and the Trp fluorescence emission quenching.

This finding clearly explains the multiexponential character of the Trp fluorescence time decay, and the occurrence of a long-lasting time component, most likely associated with Trp residues well embedded into the hydrophobic core of **SMG** aggregates.

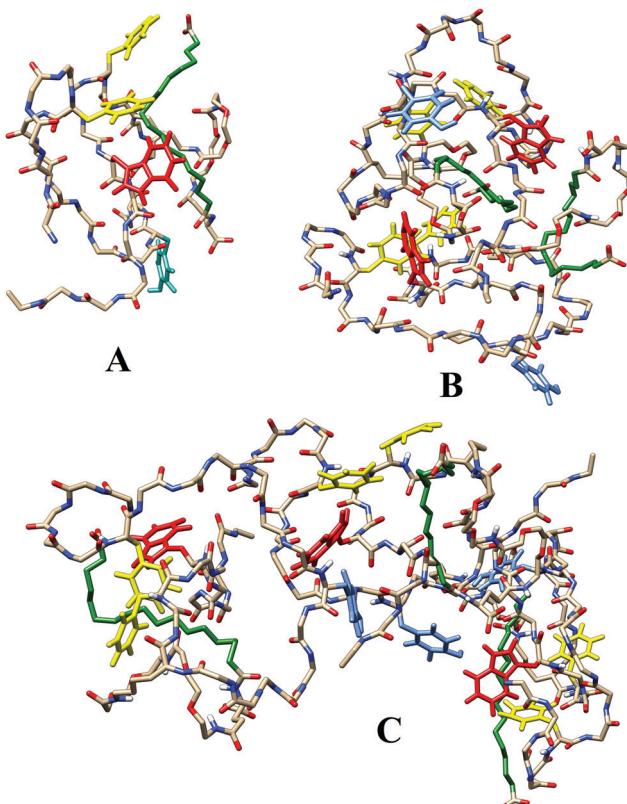
These findings allowed us to conclude that the fast folding of a single molecule is rapidly followed by the formation of **SMG** dimers and, sporadically, trimer oligomers. Gyration ( $r_g$ ) and hydrodynamic ( $r_h$ ) radii of the **SMG** monomer, dimer and trimer were obtained from GROMACS tools (Table 3). Specifically,  $r_h$  was obtained using eqn (8)<sup>20</sup>

$$r_h = \langle r_{ij}^{-1} \rangle^{-1} \quad (8)$$

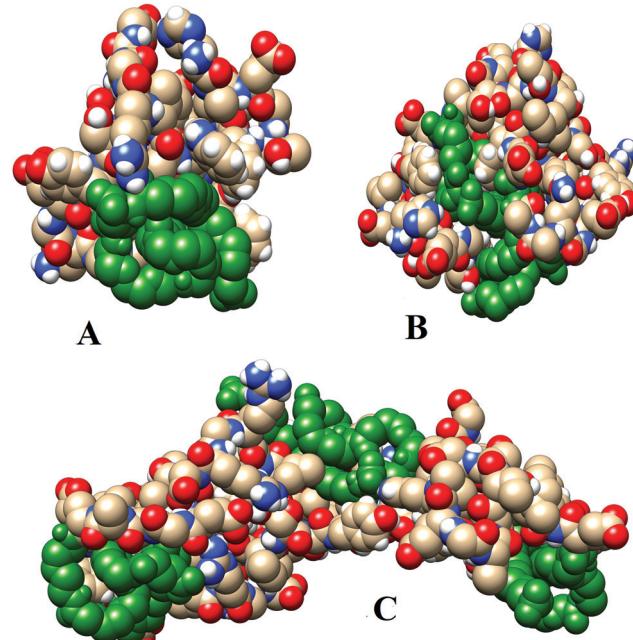
where  $i$  and  $j$  are indices running on the  $C^\alpha$  atoms and carbon atoms of LysP separated by two bonds.

Rotational autocorrelation times were obtained by the inverse of the angular velocity of the orientation vector, considering each aggregate as a rotating ellipsoid (Fig. 6).

The comparison between the calculated hydrodynamic radii with the experimental values reported in Table 2 indicates that, below the cac, **SMG** monomers, characterized by a  $r_h$  of 10.8



**Fig. 5** Molecular structures of the **SMG** monomer (A), dimer (B) and trimer (C) as provided by MD simulations. The aromatic groups are reported in red (Trp), yellow (Phe) and light blue (Tyr). LysP side chains are shown in green.



**Fig. 6** Filled-sphere representation of the **SMG** monomer (A), dimer (B) and trimer (C) from MD simulations. LysP side chains are shown in green.

(fluorescence anisotropy)  $-10.4 \text{ \AA}$  (MD) predominate, while, above the cac, dimeric clusters [ $r_h = 12.7$  (fluorescence anisotropy)  $-12.8 \text{ \AA}$  (MD)] prevail. These findings led us to conclude that: (i) at the critical aggregation concentration determined by the pyrene assay, *i.e.*  $20 \mu\text{M}$ , a monomer-to-dimer transition takes place, and (ii) the structures predicted by MD simulations faithfully reproduce the dimensions of the small clusters nucleating the growth of **SMG** aggregates.

The structures reported in Fig. 6 clearly highlight the role of the LysP lipid chain that, interacting with the peptide backbone, gives rise to a locally well-defined apolar outer region that protrudes from the peptide assembly.

### Long-time aggregation of SMG

Following the **SMG** aggregation process, we noticed that both the absorption and fluorescence spectra varied significantly day by day. In particular, the absorption spectra of a **SMG**  $20 \mu\text{M}$  solution showed a strong decrease and wide broadening of the  $\pi \rightarrow \pi^*$  and  $n \rightarrow \pi^*$  absorption bands associated with the indole (Trp) and phenol (Tyr) aromatic groups (ESI,† Fig. S4).

The associated fluorescence emission spectra initially showed the same effect, *i.e.* decreased intensities and broadening of the emission band peaked at around  $360 \text{ nm}$ . Besides that, at very long times (days), a new emission feature appears at longer wavelength, characterized by a fluorescence maximum at around  $410 \text{ nm}$  (Fig. 7A).

In Fig. 7B, the wavelength emission maxima at different times (days) of  $20 \mu\text{M}$  **SMG** aqueous solutions are reported. The solid line in Fig. 7B was obtained fitting the experimental data by the model of Pasternack *et al.*<sup>27</sup>

The latter describes aggregation as a fractal autocatalytic process, characterized by a time dependent rate constant. In this kinetic approach, the rate-determining step is the achievement of a critical concentration of small clusters, formed by  $m$  monomers, and nucleating the autocatalytic growth of large aggregates. In the Pasternack model, the catalytic rate constant,  $k_c$ , scales with the mean cluster size,  $s(t)$ , which in turn scales as a power law dependence on time, *i.e.*  $s(t) \sim t^n$ . The time dependent rate constant is therefore written as:

$$k(t) = k_0 + k_c(k_c t)^n \quad (9)$$

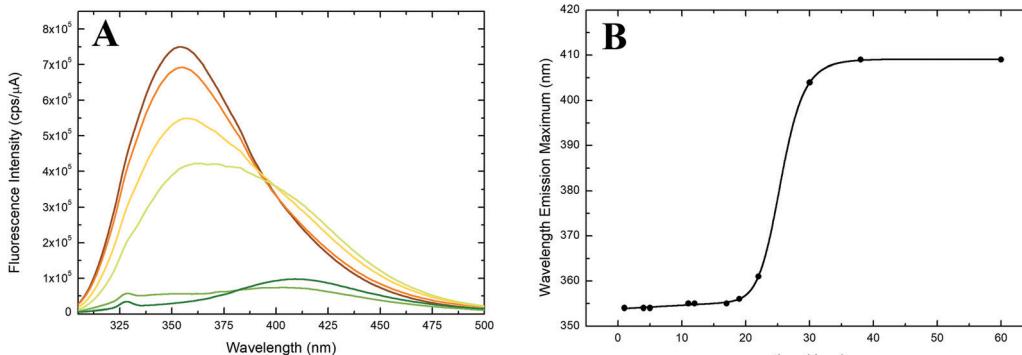


Fig. 7 (A) Emission spectra of **SMG**  $20 \mu\text{M}$  (phosphate buffer, pH 8) at different times ( $\lambda_{\text{exc}} = 295 \text{ nm}$ ). (B) Wavelength emission maxima of **SMG**  $20 \mu\text{M}$  at different times (days). The solid line was obtained fitting the experimental data by the model of Pasternack *et al.*<sup>27</sup>

to account for the non-catalytic ( $k_0$ ) and catalytic ( $k_c$ ) pathways. The integrated equation rate is readily written as:<sup>27</sup>

$$\frac{([M] - [M_i])}{([M_0] - [M_i])} = \left(1 + (m-1)\left\{k_0 t + (n+1)^{-1}(k_c t)^{n+1}\right\}\right)^{-\frac{1}{m-1}} \quad (10)$$

where  $[M]$ ,  $[M_0]$  and  $[M_i]$  are the monomer concentration at time  $t$ ,  $t = 0$ , and  $t = \infty$ , respectively.

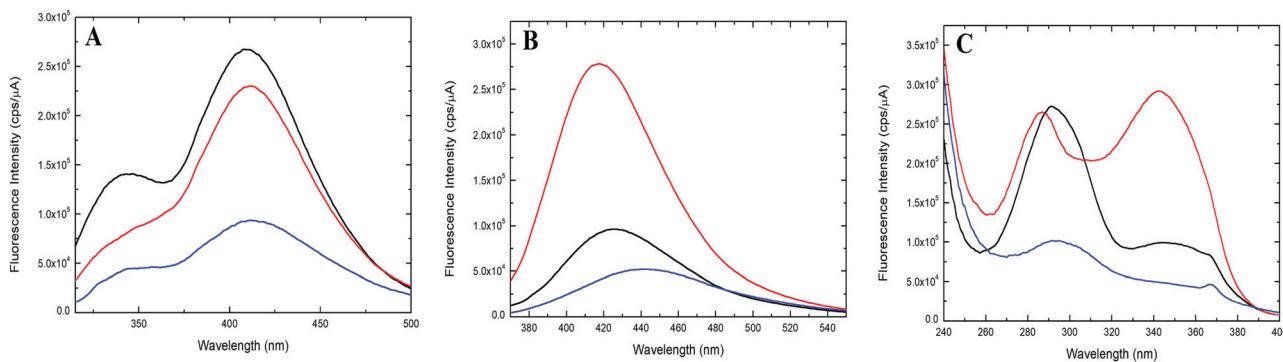
We adapted eqn (10) to fit the data reported in Fig. 7B, obtaining:  $k_0 = (1.5 \pm 0.4) \times 10^{-3} \text{ d}^{-1}$ ,  $k_c = (4.72 \pm 0.04) \times 10^{-2} \text{ d}^{-1}$ , and  $m = 2.1 \pm 0.6$ ,  $n = 14 \pm 3$ . Details of the fitting procedure are reported in the ESI.†

Notably, the value obtained for the  $m$  parameter, *i.e.* the number of monomers forming the clusters, closely parallels the results of MD simulations. The relatively high value of  $n$  indicates that a strongly cooperative process takes place, as usually found in hierarchical self-assembly.<sup>28</sup>

**SMG** aggregation should therefore occur through the following steps: (i) a fast oligomerization step leading to small peptide clusters (predominantly **SMG** dimers), followed by (ii) a very long lag time needed to reach a critical concentration, after which (iii) a fast and highly cooperative step involving the formation of mesoscopic (micrometric) structures, characterized by a faint blue luminescence, takes place.

In Fig. 8 we reported the fluorescence emission spectra of aged solutions of **SMG** (phosphate buffer, pH 8) at  $\lambda_{\text{exc}} = 295 \text{ nm}$  (Fig. 8A) and  $\lambda_{\text{exc}} = 350 \text{ nm}$  (Fig. 8B), together with the fluorescence excitation spectra at  $\lambda_{\text{em}} = 420 \text{ nm}$  (Fig. 8C).

It appears that the excitation of the indole group at  $\lambda_{\text{exc}} = 295 \text{ nm}$  gives rise to a complex emission band, characterized by a shorter wavelength component, peaked at  $330 \text{ nm}$ , typical of the tryptophan monomer emission, and a longer wavelength component, peaked at around  $410 \text{ nm}$ . The latter mainly originates from a characteristic absorption peaked at around  $\lambda_{\text{exc}} = 350 \text{ nm}$ , as proved by the intense emission obtained by excitation at this wavelength (Fig. 8B), and the excitation spectra reported in Fig. 8C. Fluorescence spectra of different **SMG** solutions clearly show that the intensity of the longer wavelength component increases in competition with the tryptophan monomer emission.



**Fig. 8** Fluorescence spectra of aged solutions of **SMG** (phosphate buffer, pH 8) at different times. (A) Emission spectra at  $\lambda_{\text{exc}} = 295 \text{ nm}$ ; (B) emission spectra at  $\lambda_{\text{exc}} = 350 \text{ nm}$ ; and (C) excitation spectra at  $\lambda_{\text{em}} = 420 \text{ nm}$ . The spectra were obtained by aged solutions of freshly prepared 4 (black), 20 (red) and 30  $\mu\text{M}$  (blue) **SMG** solutions.

Although this ‘exotic’ fluorescence has been found in several peptide systems, its origin at the moment is still fiercely debated. It has been proposed that this fluorescence originates from an extensive network of hydrogen bonds, like those realized in micrometric amyloid-like structures,<sup>29,30</sup> a quantum confinement effect in peptide nanostructures,<sup>31</sup> or semiconductor-like behaviour from micrometric structures.<sup>32</sup> Recently, Niyangoda *et al.* ascribed this blue luminescence to protein carbonyl groups.<sup>33</sup>

These findings led some researchers to propose this emission feature as a distinctive signal of the formation of amyloid structures, paving the way for the design of optical sensors capable of early diagnosis of neurodegenerative diseases.<sup>34</sup> However, the possibility of experimental artifacts due to diffused light contamination or oxidation of aromatic amino acids was recently advanced.<sup>35</sup> Formation of di-tyrosine cross-links was also shown to give rise to a weak fluorescence in the same region.<sup>36</sup>

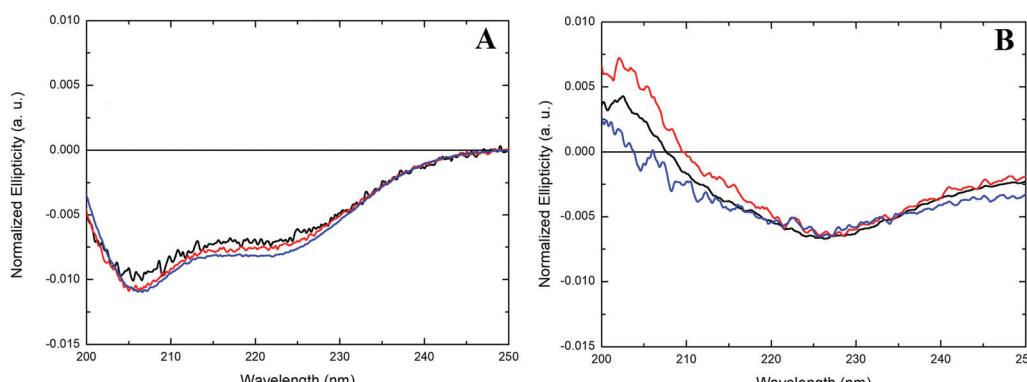
Time-resolved emission measurements carried out at  $\lambda_{\text{exc}} = 298$  and 344 nm and  $\lambda_{\text{em}} = 350$  nm and 420 nm in **SMG** aged solutions, showed that the blue luminescence reported in Fig. 8 takes place in the nanosecond time region (ESI,† Table T2), confirming the fluorescence nature of the emission. The recovered time decay parameters also suggest that the emitting excited state (or species) is beyond any doubt different from the species emitting in the UV region. Significantly, only a maximum of 10% of the signal can be ascribed to diffused light contamination.

Interestingly, the CD spectra of freshly prepared and aged **SMG** aqueous solutions show that aggregation is coupled with a secondary structure transition from a mixture of random coil and helical conformations (Fig. 9A), to a predominant  $\beta$ -sheet structure (Fig. 9B).

This finding strongly suggests that the aggregated species, giving rise to the intense fluorescence emission peaked at  $\lambda_{\text{em}} = 410 \text{ nm}$ , are characterized by a predominant  $\beta$ -sheet structure.

In contrast with this evidence, fluorescence anisotropy measurements carried out in **SMG** aged solutions with  $\lambda_{\text{exc}} = 298$  and 344 nm and  $\lambda_{\text{em}} = 420 \text{ nm}$  showed that the blue emission reported in Fig. 8 is characterized by a quite low anisotropy coefficient, independently of the peptide concentration ( $r = 0.043$  and 0.033 for 4 and 30  $\mu\text{M}$  **SMG** solutions, respectively). This finding clearly ruled out the diffused light origin of the observed emission, in agreement with the results of time resolved experiments, but also suggests that the species emitting in the blue region of the spectrum are of small size. However, in the absence of information on the orientation of the absorption and emission dipoles of the species responsible for this ‘exotic’ fluorescence, no final conclusion can be drawn from these results.

Further evidence strengthens the idea that **SMG** forms large aggregates at very long times. AFM imaging of freshly prepared 30  $\mu\text{M}$  **SMG** solutions deposited on mica by drop casting



**Fig. 9** Circular dichroism spectra of freshly prepared (A) and aged (B) **SMG** solutions. Black: 4  $\mu\text{M}$ ; red: 20  $\mu\text{M}$ ; and blue: 30  $\mu\text{M}$ .

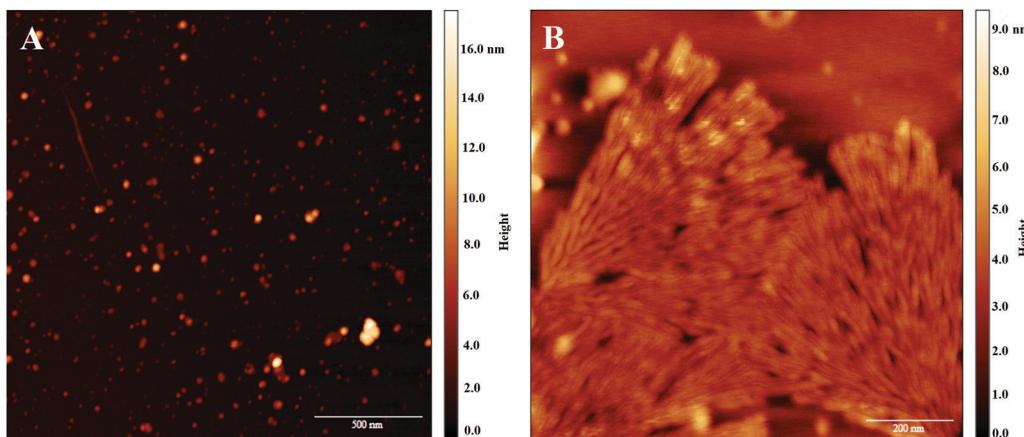


Fig. 10 Atomic force microscopy images of **SMG** nanostructures deposited on mica from freshly prepared (A) and aged (B) 30  $\mu\text{M}$  **SMG** solutions (phosphate buffer, pH 8).

revealed the formation of small globular structures showing diameters of about 10–20 nm (Fig. 10A). In the case of aged 30  $\mu\text{M}$  **SMG** solutions, Gorgonian-like dendrimeric structures growing from rod building blocks, and characterized by lengths of around 100 nm and widths of around 20 nm, can be easily spotted (Fig. 10B). Size distributions of the **SMG** nanostructures imaged by AFM measurements, and deposited on mica from freshly prepared and aged **SMG** solutions are reported in the ESI<sup>†</sup> (Fig. S5–S7). The observation of these fractal structures clearly enables recalling in the mind the interpretation of the aggregation process provided by the Pasternack model.

## Conclusions

The aggregation properties of semaglutide in the micromolar concentration region have been investigated by optical spectroscopy techniques and MD simulations. Under the applied experimental conditions, **SMG** features a very stable conformational landscape, characterized by a dynamic equilibrium between random coil and helical structures. MD simulations show that **SMG** attains a densely packed globular morphology, with the peptide charged groups exposed to the solvent, and the fatty acid chain closely packed to the peptide backbone, thereby protecting the protein core. In this oligomeric structure **SMG** would be easily cleared by renal filtration with the equilibrium towards a monomeric species concurring with HSA complexation and prolonged half-life.

**SMG** was found to be quite soluble in aqueous solutions, forming at early times monomer species and small oligomers (dimers). Fluorescence anisotropy measurements and MD simulations consistently predict correlation times and hydrodynamic volumes of the monomer and dimer species.

At very long times (weeks) compelling spectroscopic evidence indicates that an aggregation process steadily takes place, leading to the formation of an entirely new species, characterized by a faint blue luminescence signal, a  $\beta$ -sheet arrangement of the peptide chains, and, at the micrometric scale, Gorgonian-like morphologies. This aggregation process was shown to be highly

cooperative, as a result of the hierarchical character of peptide self-assembly, nucleated by the small clusters (dimers and more rarely trimers) formed at early times.

The understanding of the oligomerization and aggregation potential of peptide candidates is key for the development of long acting and stable peptide drugs, where binding to human serum albumin has the key role as a half-life extending strategy, and the propensity to fibril formation should be counteracted as part of the drug development optimization process.<sup>37</sup>

## Conflicts of interest

There are no conflicts to declare.

## Acknowledgements

The technical assistance of Miss Erika Lunetta and Miss Alice Morganti (University of Rome Tor Vergata) in the optical spectroscopy experiments is gratefully acknowledged. This project received funding from the European Union Horizon 2020 Research and Innovation Program under the Marie Skłodowska-Curie grant agreement no. 872233 (“PEPSA-MATE”).

## References

- 1 J. L. Lau and M. K. Dunn, *Bioorg. Med. Chem.*, 2018, **26**, 2700.
- 2 A. P. Davenport, C. C. G. Scully, C. de Graaf, A. J. H. Brown and J. J. Maguire, *Nat. Rev. Drug Discovery*, 2020, **19**, 389.
- 3 P. Kurtzhals, S. Havelund, I. Jonassen and J. Markussen, *J. Pharm. Sci.*, 1997, **86**, 1365; E. M. Bech, S. L. Pedersen and K. J. Jensen, *ACS Med. Chem. Lett.*, 2018, **9**, 577.
- 4 P. Ingallinella, E. Bianchi, N. A. Ladwa, Y. J. Wang, R. Hrin, M. Veneziano, F. Bonelli, T. J. Ketas, J. P. Moore, M. D. Miller and A. Pessi, *Proc. Natl. Acad. Sci. U. S. A.*, 2009, **106**, 5801.
- 5 A. Santoprete, E. Capitò, P. E. Carrington, A. Pocai, M. Finotto, A. Langella, P. Ingallinella, K. Zytko, S. Bufali,

- S. Cianetti, M. Veneziano, F. Bonelli, L. Zhu, E. Monteagudo, D. J. Marsh, R. Sinharoy, E. Bianchi and A. Pessi, *J. Pept. Sci.*, 2011, **17**, 270.
- 6 D. J. Drucker, A. Dritselis and P. Kirkpatrick, *Nat. Rev. Drug Discovery*, 2010, **9**, 267.
- 7 J. Lau, P. Bloch, L. Schäffer, I. Pettersson, J. Spetzler, J. Kofod, K. Madsen, L. B. Knudsen, J. McGuire, D. B. Steensgaard, H. M. Strauss, D. X. Gram, S. M. Knudsen, F. S. Nielsen, P. Thygesen, S. Reedtz-Runge and T. Kruse, *J. Med. Chem.*, 2015, **58**, 7370.
- 8 B. Manandhar and J. M. Ahn, *J. Med. Chem.*, 2015, **58**, 1020.
- 9 R. Meintein, B. Gallwitz and W. E. Schmidt, *Eur. J. Biochem.*, 1993, **214**, 829.
- 10 L. B. Knudsen, P. F. Nielsen, P. O. Huusfeldt, N. L. Johansen, K. Madsen, F. Z. Pedersen, H. Thogersen, M. Wilken and H. Agerso, *J. Med. Chem.*, 2000, **43**, 1664.
- 11 S. Janosch, C. Nicolini, B. Ludolph, C. Peters, M. Völkert, T. L. Hazlet, E. Gratton, H. Waldmann and R. Winter, *J. Am. Chem. Soc.*, 2004, **126**, 7496.
- 12 Y. Wang, A. Lomakin, S. Kanai, R. Alex, S. Belli, M. Donzelli and G. B. Benedek, *J. Controlled Release*, 2016, **241**, 25.
- 13 K. L. Zapadka, F. J. Becher, A. L. Gomes dos Santos and S. E. Jackson, *Interface Focus*, 2017, **7**, 20170030.
- 14 R. Lettieri, M. D'Abramo, L. Stella, A. La Bella, F. Leonelli, L. Giansanti, M. Venanzi and E. Gatto, *Spectrochim. Acta*, 2018, **195**, 84.
- 15 J. R. Lakowicz, *Principles of Fluorescence Spectroscopy*, Springer, Singapore, 3rd edn, 2006.
- 16 D. S. Karpovich and G. J. Blanchard, *J. Phys. Chem.*, 1995, **99**, 3951.
- 17 C. Oostenbrink, A. Villa, A. E. Mark and W. F. van Gunsteren, *Comput. Chem.*, 2004, **25**, 1656.
- 18 H. J. C. Berendsen, D. van der Spoel and R. van Drunen, *Comput. Phys. Commun.*, 1995, **91**, 43.
- 19 H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. Di Nola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684.
- 20 J. G. Kirkwood, *J. Polym. Sci.*, 1954, **12**, 1.
- 21 M. Caruso, E. Placidi, E. Gatto, C. Mazzuca, L. Stella, G. Bocchinifuso, A. Palleschi, F. Formaggio, C. Toniolo and M. Venanzi, *J. Phys. Chem. B*, 2013, **117**, 5448.
- 22 M. Caruso, E. Gatto, E. Placidi, G. Ballano, F. Formaggio, C. Toniolo, D. Zanuy, C. Aleman and M. Venanzi, *Soft Matter*, 2014, **10**, 2508.
- 23 J. M. Beechem and L. Brand, *Annu. Rev. Biochem.*, 1985, **54**, 43.
- 24 R. F. Chen, J. R. Knutson, H. Ziffer and D. Potter, *Biochemistry*, 1991, **30**, 5184.
- 25 K. Dönig, W. Beck, L. Konermann and F. Jähnig, *Biophys. J.*, 1997, **72**, 326.
- 26 C. P. Pan, P. L. Muiño, M. D. Barkley and P. R. Callis, *J. Phys. Chem. B*, 2011, **115**, 3245.
- 27 R. F. Pasternack, E. J. Gibbs, P. J. Collings, J. C. dePaula, L. C. Turzo and A. Terracina, *J. Am. Chem. Soc.*, 1998, **120**, 5873.
- 28 B. B. Gerbelli, S. V. Vassiliades, J. E. U. Rojas, J. N. B. D. Pelin, R. S. N. Mancini, W. S. G. Pereira, A. M. Aguilar, M. Venanzi, F. Cavalieri, F. Giuntini and W. A. Alves, *Macromol. Chem. Phys.*, 2019, **220**, 1900085.
- 29 A. Shukla, S. Mukherjee, S. Sharma, V. Agrawal, K. V. Radha Kishan and P. Guptasarma, *Arch. Biochem. Biophys.*, 2004, **428**, 144.
- 30 D. Pinotsi, L. Grisanti, P. Mahou, R. Gebauer, C. F. Kaminski, A. Hassanali and G. S. Kaminski Schierle, *J. Am. Chem. Soc.*, 2016, **138**, 3046.
- 31 N. Amdursky, M. Molotskii, D. Aronov, L. Adler-Abramovich, E. Gazit and G. Rosenman, *Nano Lett.*, 2009, **9**, 3111.
- 32 L. Laureana del Mercato, P. P. Pompa, G. Maruccio, A. Della Torre, S. Sabella, A. M. Tamburro, R. Cingolani and R. Rinaldi, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 18019.
- 33 C. Niyangoda, T. Miti, L. Breydo, V. Uversky and M. Muschol, *PLoS One*, 2017, **12**, e0176983.
- 34 F. T. S. Chan, G. S. Kaminski Schierle, J. R. Kumita, C. W. Bertoncini, C. M. Dobson and C. F. Kaminski, *Analyst*, 2013, **138**, 2156.
- 35 T. N. Tykhonova, N. R. Rovnyagina, A. Y. Zherebker, N. N. Sluchanko, A. A. Rubekina, A. S. Orekhov, E. N. Nikolaev, V. V. Fadeev, V. N. Uversky and E. A. Shirshin, *Arch. Biochem. Biophys.*, 2018, **651**, 13.
- 36 R. Aeschbach, R. Amadò and H. Neukom, *Biochim. Biophys. Acta*, 1976, **439**, 292.
- 37 I. W. Hamley, *Chem. Rev.*, 2017, **117**, 14015.