Supplementary Information :

Multivariate Analysis of Peptide-Driven Nucleation and Growth of Au Nanoparticles

Kacper J. Lachowski [a,c,*], Kiran Vaddi [a], Nada Y. Naser [a], François Baneyx [a], Lilo D Pozzo [a,b,c,*]

0.1 Materials

Pyrene (98.2 % purity) was obtained from Sigma-Aldrich and used without further purification.

0.2 UV-Vis Spectra

Note that although Z2M6I contains an outlier in the 0 mM HEPES condition shown in Figure S1 J, the use of GFT on the transformed spectra would have treated this sample as noise.

0.3 Determination of Lipidated Peptide Critical Micelle Concentration

The pyrene stock was prepared by first preparing a 0.5 mM stock solution of pyrene in ethanol. The ethanol solution was diluted with deionized water to reach a pyrene concentration of 0.146 mg/mL. Note that the final ethanol concentration was 0.144 vol.%. Peptide stock solutions were prepared at $5 \times 10^{-1}\ wt.\%$ and $5 \times 10^{-2}\ wt.\%$. Samples were prepared using an Opentrons OT-2 liquid handling robot. Solutions were prepared by transferring peptide, pyrene stock, and additional water into Caplugs Evergreen untreated 96-well microplates (black, flat bottom). The concentration of peptide in the samples was varied from 0.04 mM to 3 mM. Samples were mixed by pipetting up and down after addition of water, and were then characterized using a BioTek Synergy H1 microplate reader. Samples were excited at 334 nm and emission was obtained at 372 (**peak 1**) and 384 nm (**peak 3**).

Pyrene remains dissolved in samples where the peptide concentration is too low for formation of micelles. Above the critical micelle concentration (CMC), pyrene adsorbs to the peptide lipid tails inside micelles. The resulting change in the polarity of pyrene's environment leads to changes in its fluorescence emission. The ratio between emission **peaks 3** and **1** as a function of peptide concentration is used to estimate the critical micelle formation [1, 2, 3]. The CMC is estimated from the intersection of two lines in the plot of the ratio of the two fluorescence peaks as

[a] *Department of Chemical Engineering, University of Washington, Seattle, WA.*
[b] *Department of Material Science and Engineering, University of Washington, Seattle, WA.*
[c] *Molecular Engineering and Sciences Institute, University of Washington, Seattle, WA.*
[*] *Correspondence to KJL (lachok@uw.edu) and LDP (dpozzo@uw.edu).*

a function of the logarithm of peptide concentration. The order of the resulting CMC values are MZ2 > MZ2R > PZ2.

0.4 Au Reduction Kinetics Assay

Samples were prepared by first adding all reagents except $HAuCl_4$ to well in a 48-well plate. Then the latter was added to each well, the solution was pipetted up and down, and the samples were moved into the plate reader for measurement.

1 Functional Data analysis

In this section, we describe the mathematical details of our framework.

1.1 Functional Principal Component Analysis

Similar to standard PCA used in the multivariate analysis of high-dimensional vectors, functional principal component analysis (FPCA) aims to represent each data point (a function) in terms of its variation from the mean [4, 5]. One advantage of using PCA-based representation of the data is that dimensionality can be reduced to represent *types* or *modes* of variation. This is especially useful for UV-Vis spectra where the expected variations are low in number such as peak position, amplitude, etc. The goal in PCA is to obtain dominant (or principal) variations of the data (from the mean) by solving for a set of weight vectors that maximize variance along different components called *principal components*. Let d discrete evaluations of a function $f(t)$ represented as f_{ij} for $j = 1, 2, 3, \ldots, d$ be the i^{th} data point of the functional data $\mathcal{F}$ for $i = 1, 2, 3, \ldots, N$. Similar to multi-variate PCA for vector-like data, FPCA assumes an underlying linear model for the variations of $\mathcal{F}$:

$$f = \mu_f + \sum_{j=1}^{c} \beta_j w_j \tag{1}$$

where μ_f is the mean of samples f_i and $\beta_j, j \in \{1, 2, \ldots, c\}$ is the *learnable principal coefficients* corresponding to an orthogonal *basis function* w_j or *principal components* of the linear model. It can be shown that a solution to PCA problem of fitting a linear model to data such that cross-correlation is minimized in the new orthogonal basis w_j is given by orthogonal decomposition of covariance matrix [6]. The covariance of the data $V(s, t)$ can now be estimated using:

$$V(s,t) = \frac{1}{N-1} \sum_{i}^{N} (f_i(s) - \mu_f)(f_i(t) - \mu_f)$$

The linear model of $\mathcal{F}$ along any given component $f = \mu + \beta_j * w_j$ highlights a particular type of variations of sample x_i along a component w_j. Following the terminology of multi-variate PCA, the values β_j are called *scores* for the optimized

principal components w_j. We illustrate usage of functional PCA using an example case study below where the samples from $\mathcal{F}$ are a set of functions $f(t) = c\sin(2\pi t)$ for $c \in [0, 1]$ and $t \in [0, 1]$ as shown in Figure S6. The samples have only one mode of variation (i.e. one dimension of functional variation) in the amplitude.

Because the FPCA is a (linear) generative model, we can draw samples from it by picking arbitrary principal component scores. Figure S7 depicts the only principal component along with the mean function (in red solid line) and a percentage indicating the total amount of variation accounted for by each component. The grey gradient solid lines in Figure S7 represent linear spaced samples along $\beta_j \in [-1, 1]$ from the trained FPCA model for the functional data in Figure S6. We observe that for any given sample draw $f = \mu_f + \beta_1 w_1$ from the FPCA model, the value α changes the peak amplitude which is the only variation present in $\mathcal{F}$.

In this paper, we use a class of FPCA (called joint-FPCA) introduced in [7] that can efficiently decouple the variations along x and y axis of one dimensional functional data. We refer interested readers to [7] and [5, Chapter 8.9] as the details are beyond the scope of this paper.

1.2 Graph Fourier Transforms (GFT)

Graph Fourier Transforms (GFT) [8, 9] offer a way to decompose signals indexed by nodes/vertices into a weighted combination of different *frequencies* on a given graph structure. Given a spatial embedding of a graph, the frequencies correspond to different oscillations of a signal. Mathematically, the GFT defines an irreducible representation of signal space $\mathcal{S}$ with the basis given by the eigenfunctions of the graph Laplacian matrix L representing variations of signals defined on the graph vertices. The graph Laplacian is defined as $L := D - A$ where D is the degree matrix and A is the adjacency matrix. Given a signal $S : V \to \mathbb{R}^d$, the Laplacian operator L computes the signal variation from local average:

$$(Ls)_i = \left(\frac{1}{D_{ii}} \sum_{ij \in E} s_j \right) - s_i$$

Using the spectral theorem for symmetric matrices [10], we can obtain a orthogonal basis for the linear operator $L_{n \times n}$ in terms of eigenfunctions $\phi_0, \phi_1, \ldots, \phi_n$ for a graph with n nodes. Motivated by interpretation of Fourier basis as the eigenfunctions of the second derivative operator on a circle, the Laplace eigenfunctions are associated with *oscillations* on an irregular domain such as a graph. GFT $\hat{S}$ in the eigenbasis of L for signal S can be obtained using:

$$\hat{S}(k) = \sum_{i \in V} S(i)\phi_k \tag{2}$$

Figure S8 below shows an example using GFT on a circle where different eigenfunctions of a circle (represented as a graph) are shown using a color code.

We recover the standard Fourier basis in one-dimensions in Figure S9 by plotting the eigenbasis in Figure S8 with vertex index (i.e. θ). Any signal defined over the vertices of a circle can then be represented as a sum of different ϕ_i in (see Figure S8) multiplied by a scalar using Equation (2) analogous to one-dimensional Fourier representation. Once the graph signal is represented using the Fourier transform, we can then obtain a flattened vector where each dimension correspond to a combination of the original signal dimension and frequency. A distance between two Fourier transformed graph signals $\hat{S}_i, \hat{S}_j$ can then be computed using:

$$d = 1 - \frac{(\hat{S}_i - \overline{\hat{S}_i}) \cdot (\hat{S}_j - \overline{\hat{S}_j})}{||(\hat{S}_i - \overline{\hat{S}_i})||_2 ||(\hat{S}_j - \overline{\hat{S}_j})||_2} \tag{3}$$

where $\bar{\cdot}$ and $||\cdot||_2$ represent the mean and norm of the signal. Two signals are considered identical if the distance d is zero and the similarity decreases with increase in d.

1.3 USAXS Scattering Profiles and AUTORG Results

The AUTORG results for lipidated and non-lipidated peptides are presented in Figure S12 and Figure S13, respectively. The lipidated and non-lipidated peptide-$HAuCl_4$ assemblies are of similar size across all of the studied conditions. However, the size of the lipidated variants appears to vary less as a function of the peptide:$HAuCl_4$ ratio when compared with the non-lipidated variants. Note in Figure S13 that the data for peptides which are lacking methionine showed similar scattering behavior to the Z2 peptide. We interpret this to mean that peptide-$HAuCl_4$ interactions that are responsible for the formation of the observed aggregates are not due to methionine-$HAuCl_4$ complexation.

1.4 SAXS

MZ2 (myristoylated RMRMKMK) was dissolved at a concentration of 4.2 mM in ultra pure water. The sample was flowed into a quartz capillary (1 mm outer diameter, Charles Supper Company, Natick, MA). The capillary was loaded inside a Kratky-type SAXS instrument (SAXSess, Anton Paar, Graz, Austria) with a Cu $K\alpha$ source ($\lambda = 1.54$ Å). The slit-smeared scattering data was fit using a slit-smeared cylinder model using SASview (SasView v.5.0.1, www.sasview.org) [11, 12, 13]. The resulting radius of the fitted cylinder model was 5.2 nm. The length could not be determined from the scattering data due to the lack of a Guinier region in the q-range of the instrument. Therefore we can only conclude that the length of the extended, self-assembled MZ2 structures was in excess of 60 nm.

Table 1: List of items used per plate

Supplier	Material and Dimensions	Catalog Number	Part Name	Qty.
Formlabs	Clear	RS-F2-GPCL-04	Center piece/main body	1
McMaster-Carr	Aluminum	9140T109	Cover plate source	1
McMaster-Carr	Aluminum	9140T109	Cover plate detector	1
McMaster-Carr	Kapton (0.0010" thick)	2271K401	Windows	2
McMaster-Carr	Buna-N (1.5 mm wide, 9 mm ID)	9262K126	O-rings	96
McMaster-Carr	Stainless steel M3 (8 mm Long, 0.5 mm Pitch)	90666A104	Screws	18
MiSUMi	Brass Cadmium-Free (M3x0.5 mm)	FB-305580	Threaded insert	9

1.5 USAXS Sample Holder Assembly

The liquid sample holder used for USAXS experiments is a sandwiched well-plate style assembly which has the same profile as a standard well plate. Files and instructions for assembly have been shared online [14]. The process of assembly is as follows:

1. Insert o-rings into center piece face which will be facing the X-ray source.
2. Place Kapton sheet and then the aluminum cover piece which will be facing the source on top of the o-rings.
3. Insert screws and fasten in the order specified in the image.
4. Turn over the plate and insert o-rings.
5. Load liquid samples manually or with a liquid-handling system. Suggested volume is 450 μL.
6. Place second Kapton sheet and aluminum cover plate on top, and fasten using the same order of screws.

1.6 Opentrons Protocol Details

Code to control the OT2 is available online [15]. The protocol uses an Opentrons P300 GEN2 and P50 GEN1. Deck layout is shown in Figure S16.

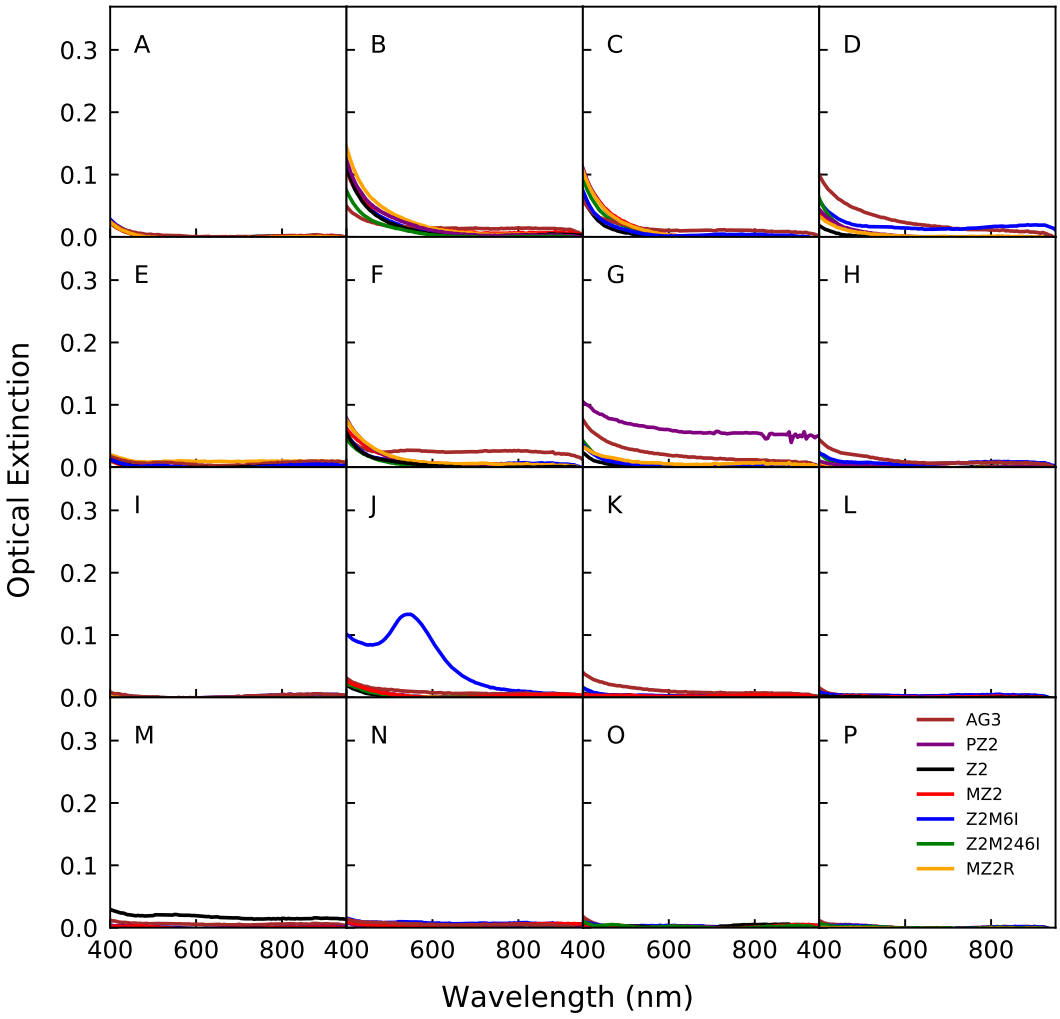

Figure S1: UV-Vis spectra of Au Nanoparticles prepared in the presence of different peptides as indicated by the color labeling. HEPES concentration is 0 mM, $HAuCl_4$ concentration decreases from top to bottom (0.2, 0.117, 0.0684, 0.04 mM), and peptide concentration increases from left to right (0, 0.04, 0.0894, 0.2 mM).

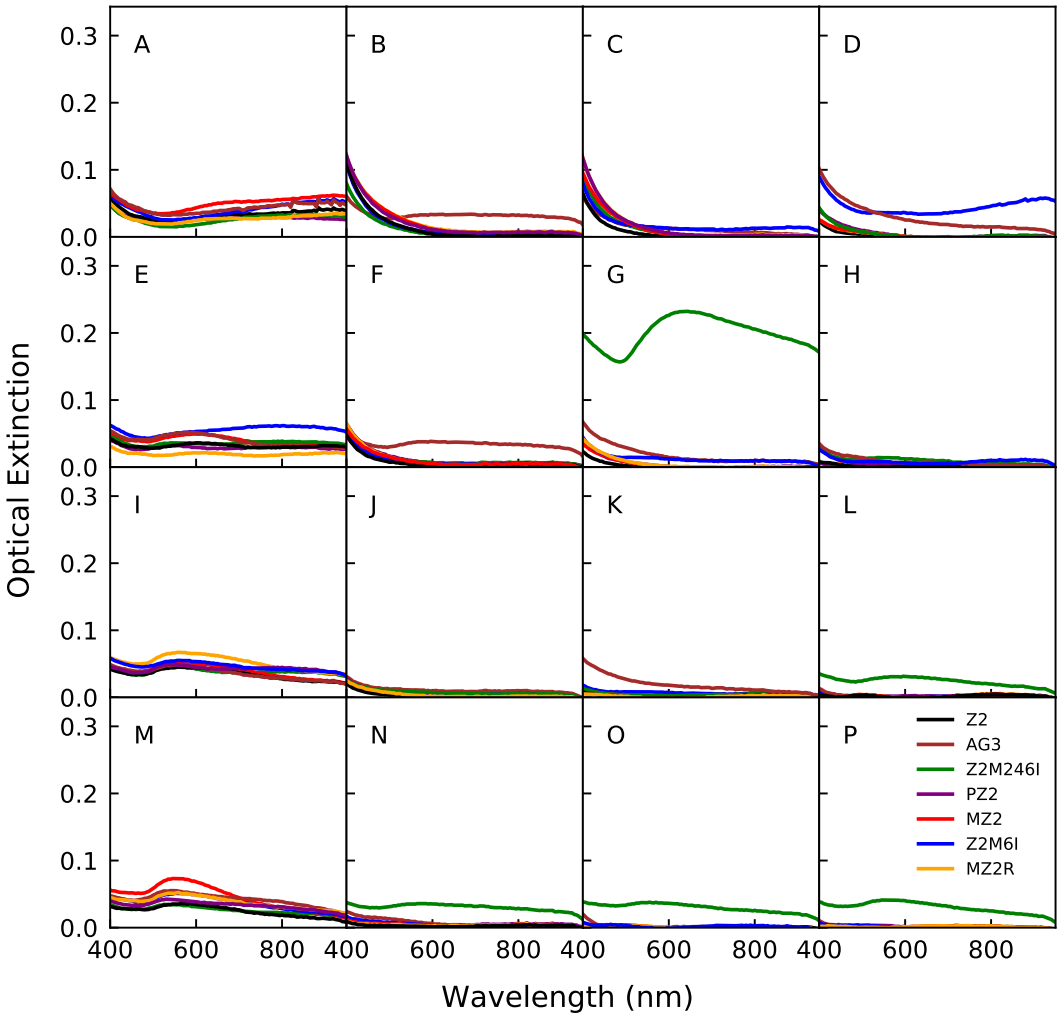

Figure S2: UV-Vis spectra of Au Nanoparticles prepared in the presence of different peptides as indicated by the color labeling. HEPES concentration is 0.01 mM, $HAuCl_4$ concentration decreases from top to bottom (0.2, 0.117, 0.0684, 0.04 mM), and peptide concentration increases from left to right (0, 0.04, 0.0894, 0.2 mM).

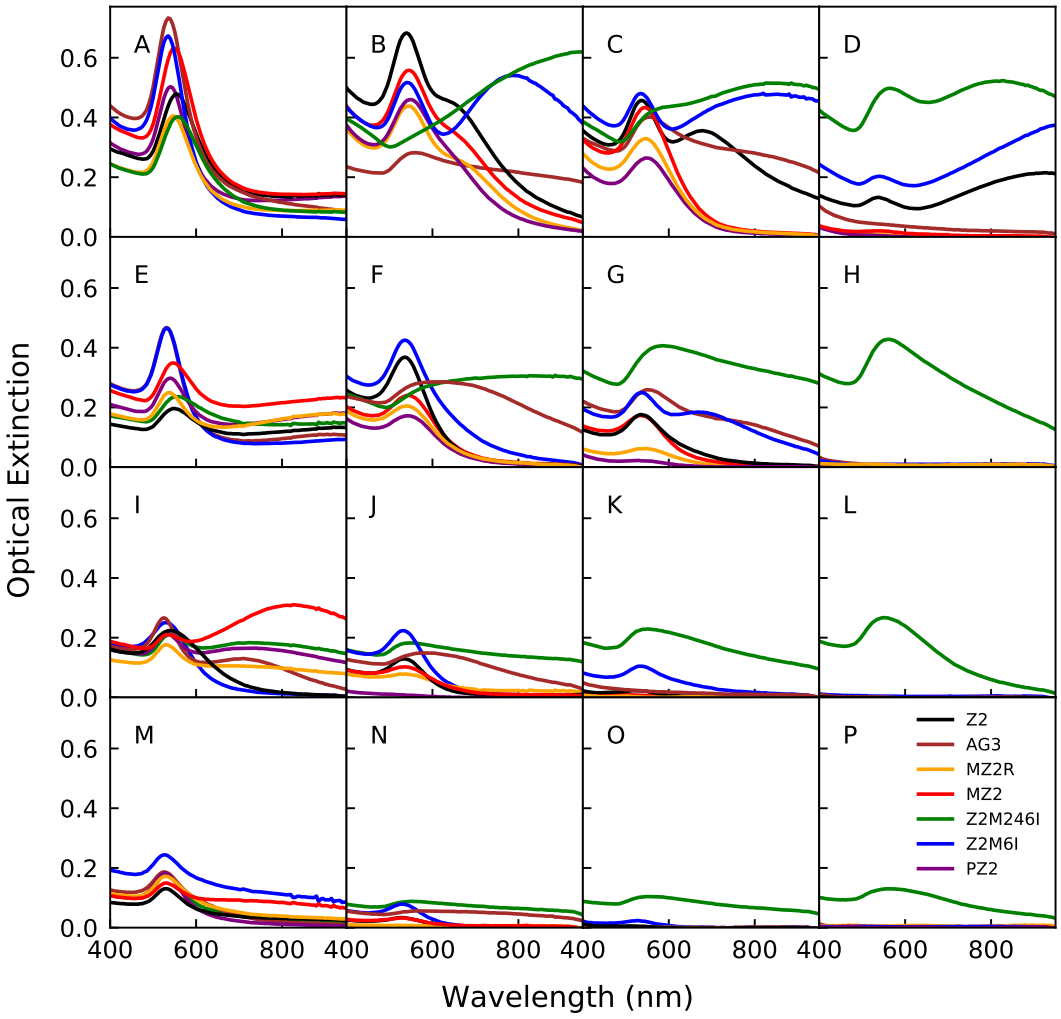

Figure S3: UV-Vis spectra of Au Nanoparticles prepared in the presence of different peptides as indicated by the color labeling. HEPES concentration is 0.1 mM, $HAuCl_4$ concentration decreases from top to bottom (0.2, 0.117, 0.0684, 0.04 mM), and peptide concentration increases from left to right (0, 0.04, 0.0894, 0.2 mM).

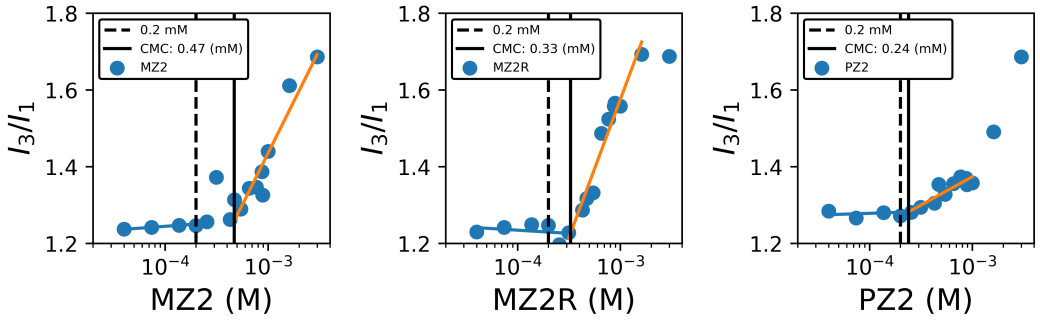

Figure S4: Ratio of fluorescence intensity at 384 to 372 nm resulting from excitation at 334 nm as a function of peptide concentration. Adsorption of pyrene to micelles leads to a change in the fluorescence spectra which is used to estimate the CMC by finding intersection of two lines fitted to the data. The CMC of MZ2, MZ2R, and PZ2 were estimated to be 0.47 mM, 0.33 mM, and 0.24 mM, respectively.

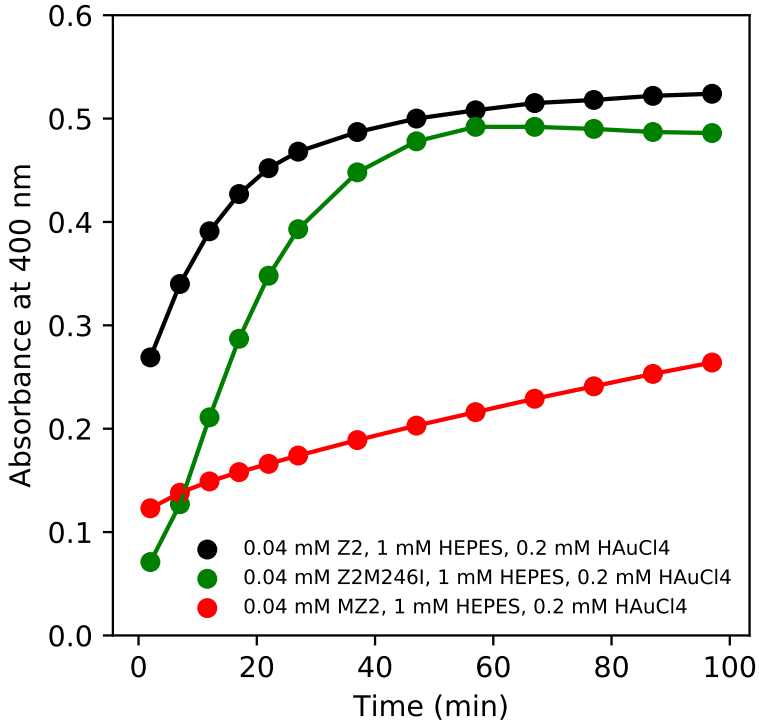

Figure S5: Absorbance at 400 nm as a function of time (minutes) for Z2, Z2M246I, and MZ2.

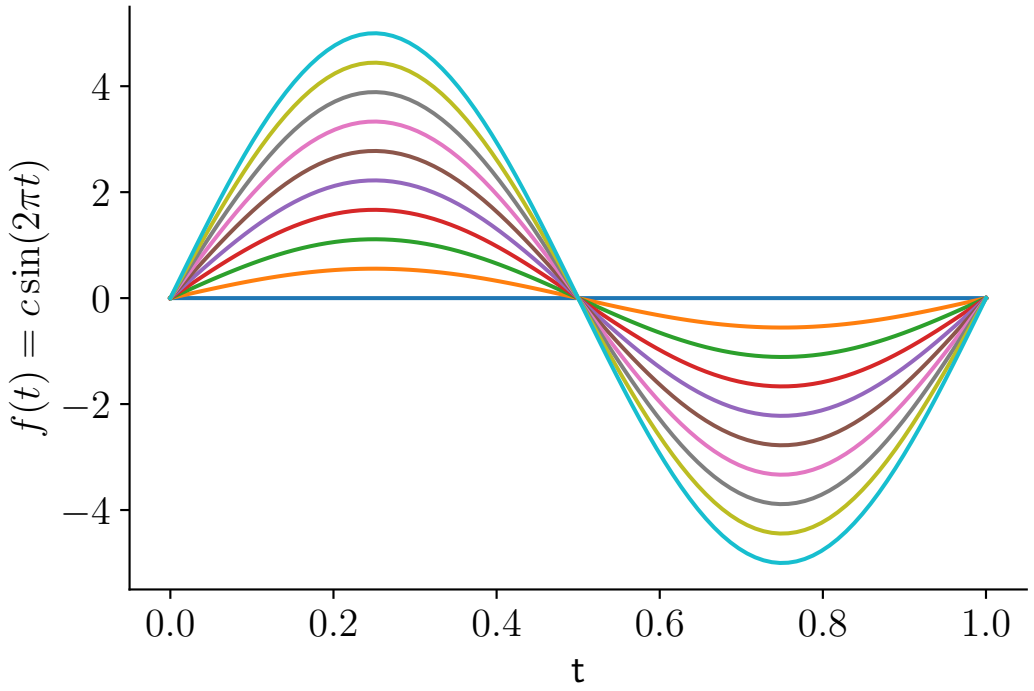

Figure S6: Synthetic data generated with $f(t) = c\sin(2\pi t)$ with $c \in [0, 1]$

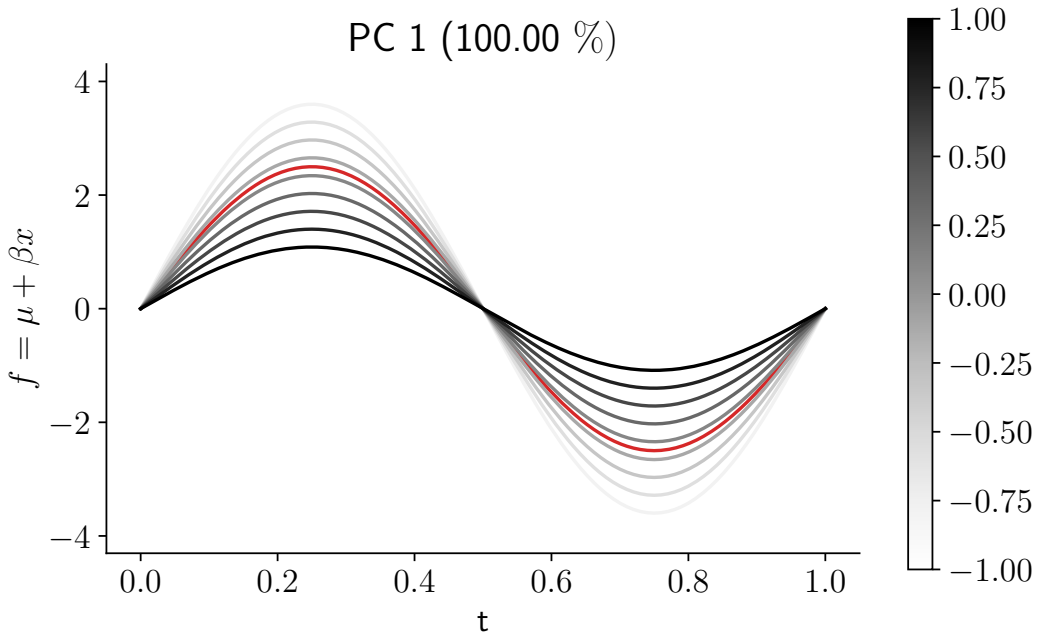

Figure S7: Samples generated from the Principal component 1 of FPCA using Equation (1) with mean of the data in Figure S6 shown in the solid red line.

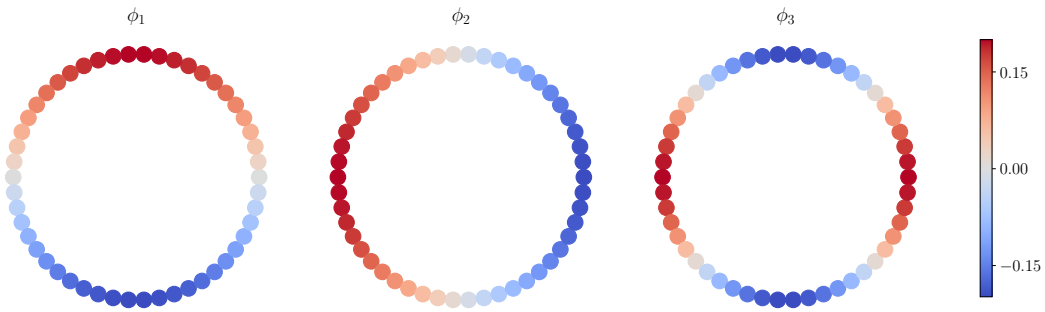

Figure S8: Eigen-functions ϕ_i for $i = 1, 2, 3$ for a circle

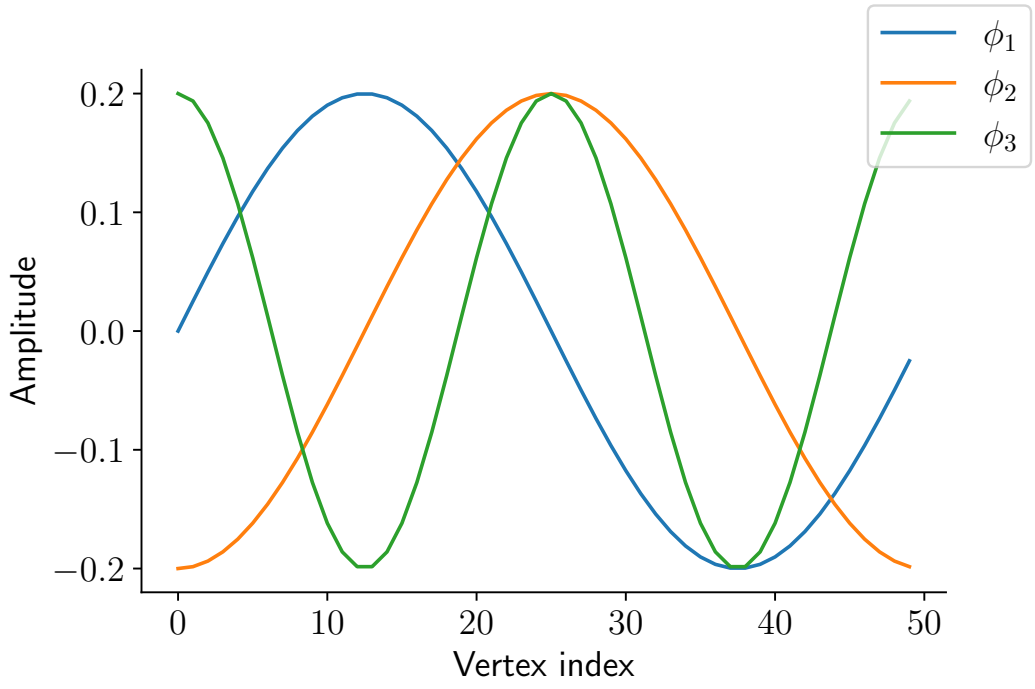

Figure S9: Eigen-functions of the circle graph in Figure S8 visualized as one dimensional functions over vertex index

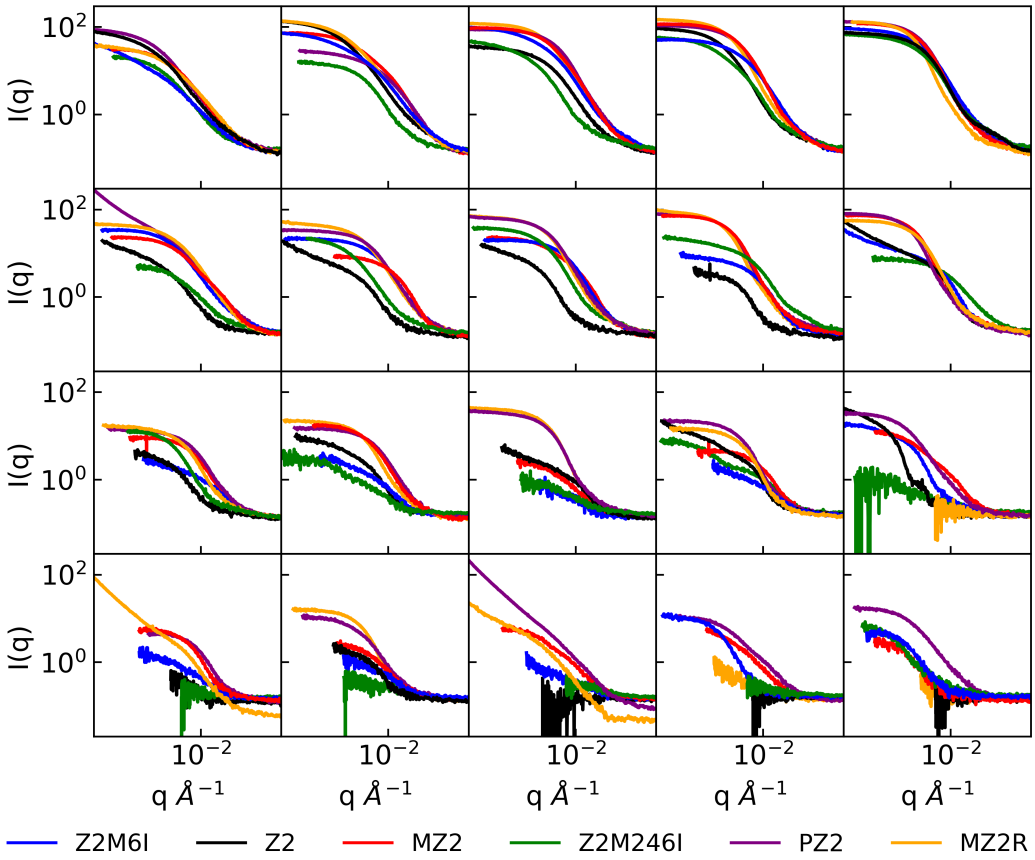

Figure S10: Smeared USAXS data. Sample compositions: 0 mM HEPES; peptide left to right 0.040, 0.060, 0.089, 0.134, and 0.200 mM; $HAuCl_4$ top to bottom 0.200, 0.117, 0.068, and 0.040 mM.

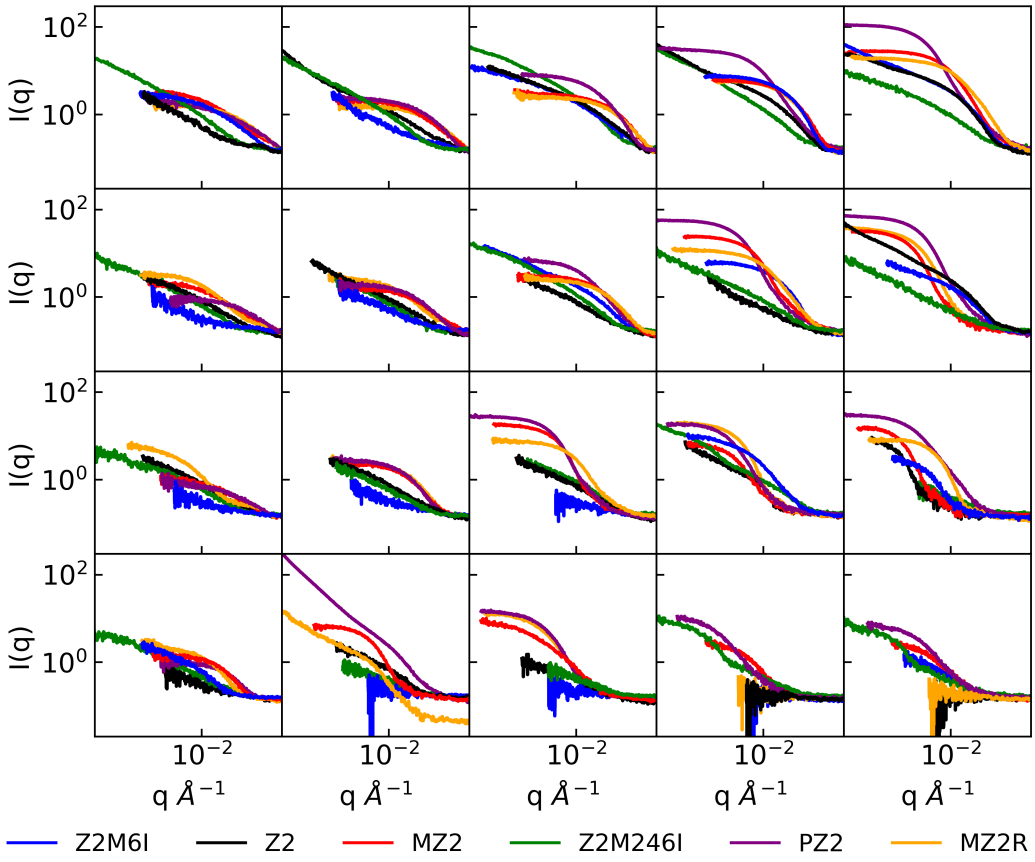

Figure S11: Smeared USAXS data. Sample compositions: 1 mM HEPES; peptide left to right 0.040, 0.060, 0.089, 0.134, and 0.200 mM; $HAuCl_4$ top to bottom 0.200, 0.117, 0.068, and 0.040 mM.

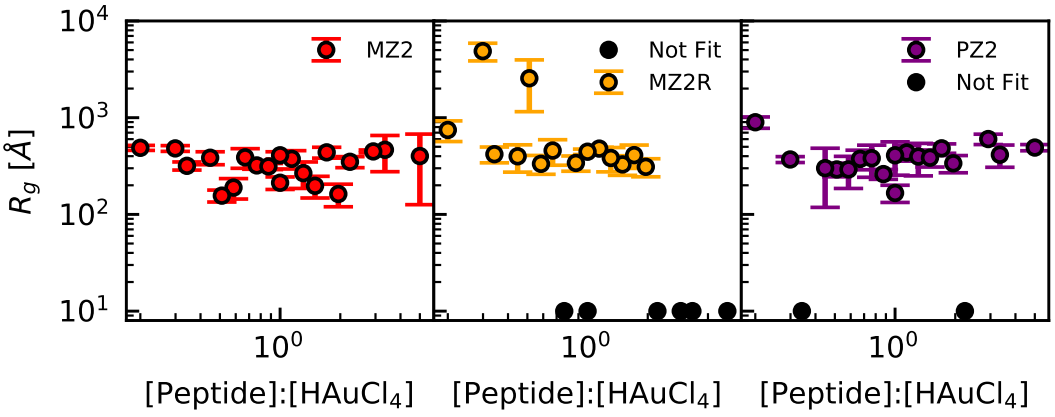

Figure S12: Radius of gyration results obtained from AUTORG plotted as a function of peptide:$HAuCl_4$ molar ratio for all lipidated peptides and a HEPES concentration of 0 mM. Error bars indicate standard deviation. Profiles which did not contain a Guinier region that could be fitted were marked using black and plotted at a fixed R_g of 10 Å.

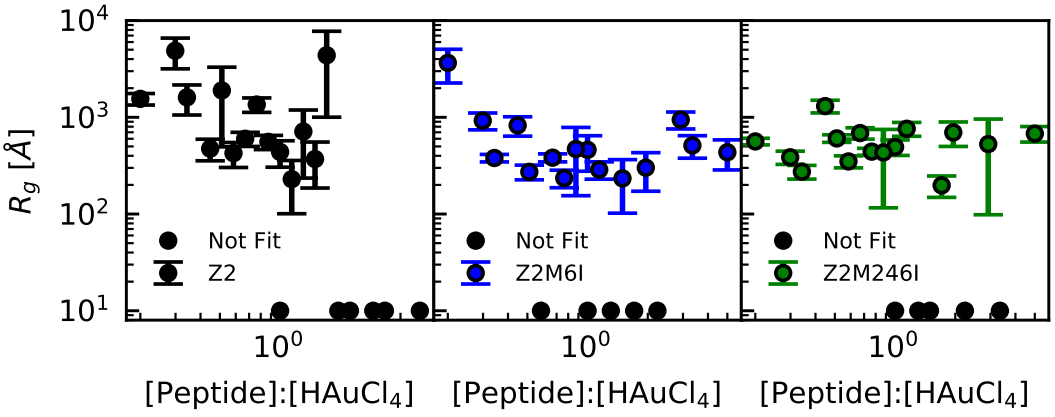

Figure S13: Radius of gyration results obtained from AUTORG plotted as a function of peptide:$HAuCl_4$ molar ratio for non-lipidated peptides (except AG3) and a HEPES concentration of 0 mM. Error bars indicate standard deviation. Profiles which did not contain a Guinier region that could be fitted were marked using black and plotted at a fixed R_g of 10 Å.

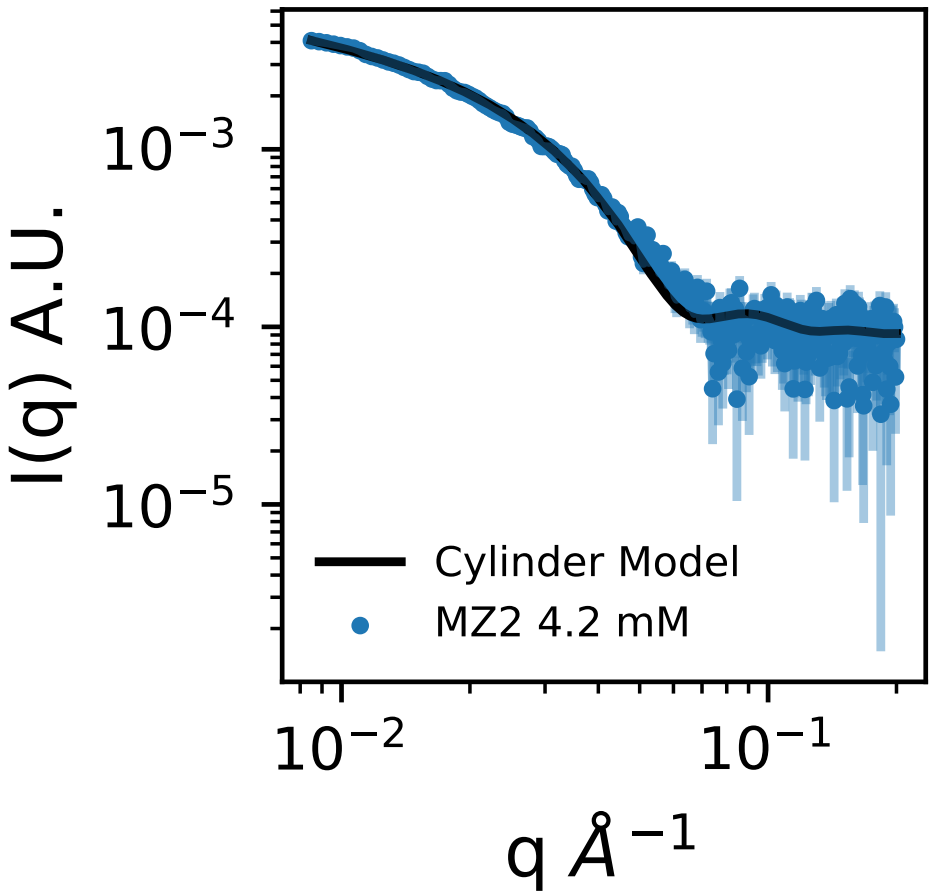

Figure S14: Smeared small-angle scattering of MZ2 dissolved in water (4.2 mM) fit with a cylinder model (radius = 5.2 nm).

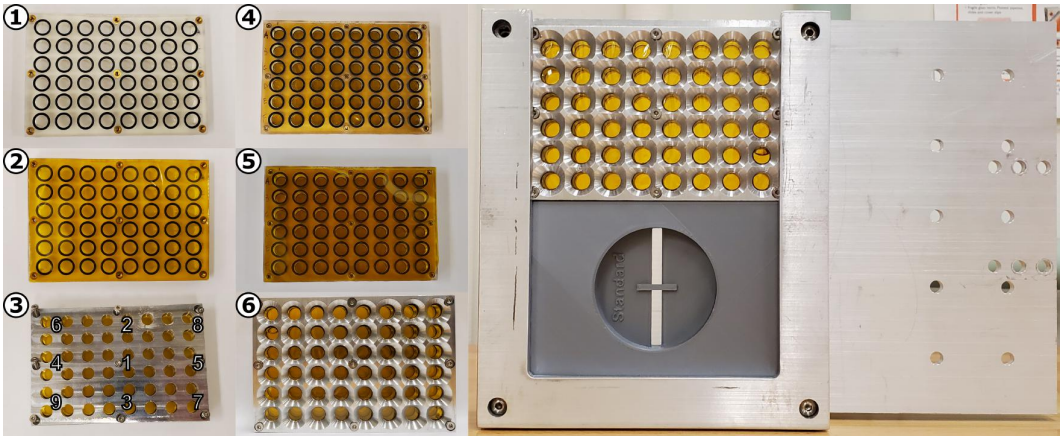

Figure S15: Process of assembling USXAXS sample holder, and the frame which holds up to two plates during experiments. A 3D printed mock cartridge is used to hold silver behenate for USAXS calibration.

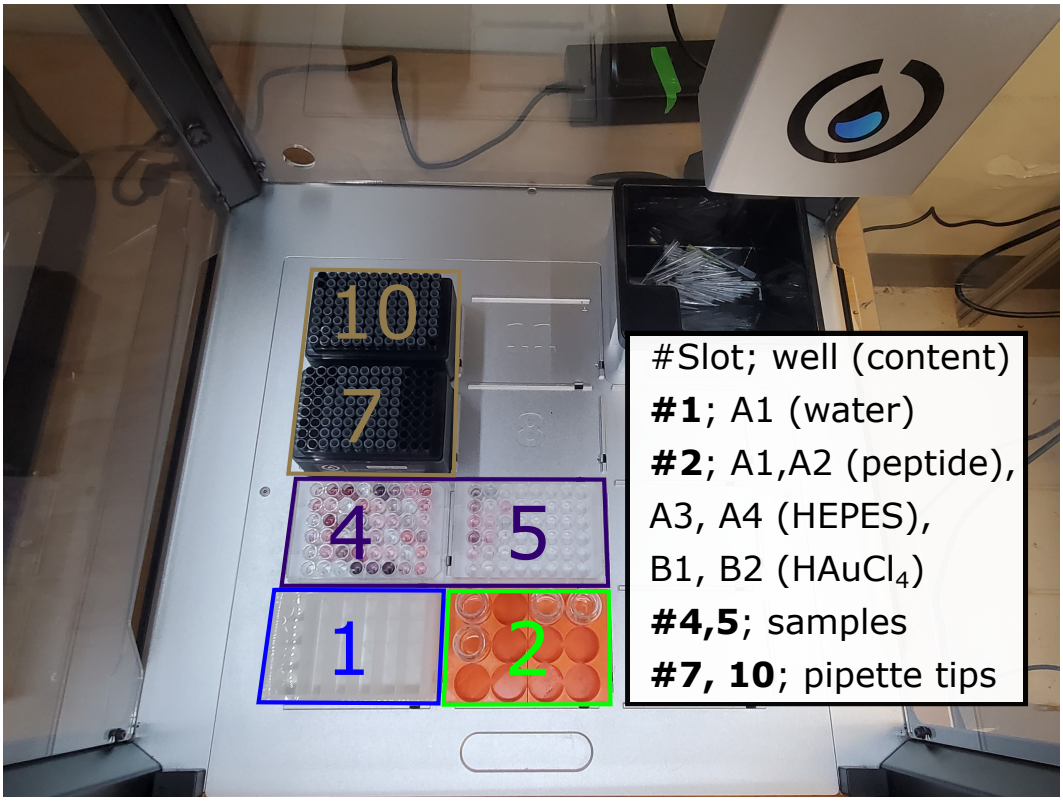

Figure S16: Opentrons OT-2 deck layout: 1, Agilent 6 column reservoir, water; 2, stocks in 20 mL scintillation vials seated in 3D printed vial holder; 4 and 5, 48 well microplates; 7 and 10, 300 μL pipette tips. Stock distribution in slot 2: A1, peptide; A3 and A4, HEPES; B1, $HAuCl_4$

References

[1] K. Kalyanasundaram and J. K. Thomas. Environmental Effects on Vibronic Band Intensities in Pyrene Monomer Fluorescence and Their Application in Studies of Micellar Systems. *Journal of the American Chemical Society*, 99(7):2039–2044, 1977.

[2] Jessica A. Hutchinson, Ian W. Hamley, Juan Torras, Carlos Alemán, Jani Seitsonen, and Janne Ruokolainen. Self-Assembly of Lipopeptides Containing Short Peptide Fragments Derived from the Gastrointestinal Hormone PYY 3–36 : From Micelles to Amyloid Fibrils. *The Journal of Physical Chemistry B*, 123(3):614–621, jan 2019.

[3] Lucas Piñeiro, Mercedes Novo, and Wajih Al-Soufi. Fluorescence emission of pyrene in surfactant solutions. *Advances in Colloid and Interface Science*, 215:1–12, 2015.

[4] JO Ramsay and BW Silvermann. Functional data analysis. springer series in statistics, 1998.

[5] Anuj Srivastava and Eric P Klassen. *Functional and shape data analysis*, volume 1. Springer, 2016.

[6] Jonathon Shlens. A tutorial on principal component analysis. *arXiv preprint arXiv:1404.1100*, 2014.

[7] J Derek Tucker, Wei Wu, and Anuj Srivastava. Generative models for functional data using phase and amplitude separation. *Computational Statistics & Data Analysis*, 61:50–66, 2013.

[8] Benjamin Ricaud, Pierre Borgnat, Nicolas Tremblay, Paulo Gonçalves, and Pierre Vandergheynst. Fourier could be a data scientist: From graph fourier transform to signal processing on graphs. *Comptes Rendus Physique*, 20(5):474–488, 2019.

[9] Aliaksei Sandryhaila and José MF Moura. Discrete signal processing on graphs: Graph fourier transform. In *2013 IEEE International Conference on Acoustics, Speech and Signal Processing*, pages 6167–6170. IEEE, 2013.

[10] Paul R Halmos. What does the spectral theorem say? *The American Mathematical Monthly*, 70(3):241–247, 1963.

[11] Lars Onsager. the Effects of Shape on the Interaction of Colloidal Particles. *Annals of the New York Academy of Sciences*, 51(4):627–659, 1949.

[12] Gérard Fournet. Étude théorique et expérimentale de la diffusion des rayons X par les ensembles denses de particules. *Bulletin de la Société française de Minéralogie et de Cristallographie*, 74(1):37–172, 1951.

[13] Jan Skov Pedersen. Analysis of small-angle scattering data from colloids and polymer solutions: Modeling and least-squares fitting. *Advances in Colloid and Interface Science*, 70(1-3):171–210, 1997.

[14] Lilo Pozzo, Kacper Lachowski, Sage Scheiwiller, and Edwin Antonio. pozzo-research-group/Automation-Hardware: v0.1.0 - Initial release for documenting liquid sample holder on Zenodo, January 2022.

[15] Kacper J. Lachowski, Kiran Vaddi, Nada Y. Naser, François Baneyx, and Lilo D. Pozzo. Gold Nanoparticles Synthesized in the Presence of Peptides - UV-Vis Spectra, Fluorescence, USAXS, Electron Microscopy, February 2022.