Electronic Supplementary Material (ESI) for Soft Matter.

Supplementary materials

Emulsifying properties of plant-derived polypeptide and their conjugates: A self-consistent-field calculation study of the impact of hydrolysis

Yue Ding[*a,b], Adem Zengin[c], Weiwei Cheng[a], Libo Wang[a] and Rammile Ettelaie[*b]

[a] *College of Food and Bioengineering, International Joint Laboratory of Food Processing and Quality Safety Control of Henan Province, Henan University of Science and Technology, Luoyang 471000, PR China.*

[b] *Food Colloids Group, School of Food Science and Nutrition, University of Leeds, Woodhouse Lane, Leeds LS2 9JT, UK.*

[c] *Department of Food Engineering, Faculty of Engineering, Sakarya University, Serdivan Sakarya, Turkey.*

* Corresponding authors:

Rammile Ettelaie, Tel: (+44) 113 3432981, E-mail: r.ettelaie@leeds.ac.uk
Yue Ding, Tel: (+86) 0379 64282342, E-mail: dingyue@haust.edu.cn

S.1 Outline of the SCF calculations

In this section, we briefly outline our Self-Consistent-Field (SCF) calculation and its implementation, based on the scheme of Scheutjens and Fleer,[1, 2] as used in the present work here. A more detailed and extensive account of the theory can be found in many excellent books, articles and reviews,[3-8] , as well as several of our own previous publications.[9, 10]

A prerequisite to performing SCF calculations is the derivation of a free energy functional. This provides the free energy of the system for any given variation of the density profiles of various monomer species belonging to chains, solvent, and other molecules (e.g. ions) present in the solution. This task is not a trivial one, though it is now quite well established.[6, 11, 12] It involves a statistical mechanics averaging of the position of all the monomers on the chains that lead to a given set of density profile variations $\{\phi_k^\alpha(\underline{r})\}$, where $\phi_k^\alpha(\underline{r})$ is the density of monomers of kind α, belonging to chains of type k, at a location $\underline{r}$. In essence, starting from a (Hamiltonian) function which specifies the energy of the system for a given conformation of chains, one derives a coarse-grained functional $F\left(\{\phi_k^\alpha(\underline{r})\}\right)$, providing the free energy of the system for any desired set of density variations $\{\phi^\alpha(\underline{r})\}$. Since in this work we are interested in the behaviour of chains in a gap between two approaching droplets, we initially only consider the region between two infinite parallel plates. Additionally, as with any numerical scheme, it is necessary to discretise this space between the surfaces into a mesh or a suitable grid for the intended computation. We do this by dividing the gap into a set of layers parallel to the plates, having each a thickness a_o. Although a_o can be chosen to have any suitable value, by taking this to be the nominal size of a monomer (e.g. length of a peptide bond ~ 0.3 nm here), a more physical interpretation to the model is provided based on the Flory-Huggins lattice model of polymer solutions. The position $\underline{r}$ is now given in terms of the layer number, labelled i=1,2,3,…..L, extending from one plate to the other. The layers themselves are assumed to consist of lattice sites of size a_o (see **Fig. 1** in the main text for a schematic illustration). All grid sites need to be filled either by monomers belonging to chains, solvent molecules or ions. Likewise, no lattice site can be doubly occupied. This enforces the overall incompressibility of the solution, leading to

$$\sum_{k,\alpha}\phi_k^\alpha(i)=\sum_{k,\alpha}\Phi_k^\alpha=1 \quad , \qquad \text{(S1)}$$

where we have reserved the symbol Φ_k^α to represent the bulk value of ϕ_k^α outside the gap, far away from the interfaces. Note that with our definition and the particular choice of a_o, the volume fraction of each type of monomer α coincides with the number density of that monomer within each layer.

The free energy functional per unit area, when suitably discretised in the gap between the two planes, takes on the following form:

$$\frac{\Delta F(L)}{k_B T} = -\sum_{i=1}^{L}\sum_{k}\frac{1}{N_k}\sum_{\alpha}\left(\phi_k^{\alpha}(i)-\Phi_k^{\alpha}\right) - \sum_{i=1}^{L}\sum_{\alpha}\psi^{\alpha}(i)\sum_{k}\phi_k^{\alpha}(i)$$
$$+\frac{1}{2}\sum_{i=1}^{L}\sum_{k,m}\sum_{\alpha\neq\beta}\chi_{\alpha\beta}\left(\phi_k^{\alpha}(i)-\Phi_k^{\alpha}\right)\left(<\phi_m^{\beta}(i)>-\Phi_m^{\beta}\right) \qquad \text{(S2)}$$
$$+\frac{1}{2}\sum_{i=1}^{L}\psi_{el}(i)\sum_{k}\sum_{\alpha}q_{\alpha}\phi_k^{\alpha}(i) + \sum_{k}\sum_{\alpha}\chi_{s\alpha}\left[\phi_k^{\alpha}(1)+\phi_k^{\alpha}(L)\right]$$

where free energy F is expressed with reference to a uniform solution where no interfaces are present. In the above equation, T is the temperature, k_B the Boltzmann constant, $\psi_{el}(i)$ the electric potential in layer i, and q_α the charge of monomers of type α. The degree of polymerization of chains of type k is denoted as N_k, where this is equal to 1 for solvent molecules and ions in the solution. The short-range interactions between two monomer species are characterised by a set of Flory-Huggins χ-parameters, $\{\chi_{\alpha\beta}\}$. A negative value of $\chi_{\alpha\beta}$ indicates that the monomers belonging to groups α and β have a preference to be in contact with each other, while a positive value signifies an unfavourable interaction between the two. Similarly, the strength of interactions between surface and monomers of kind α is denoted by $\chi_{s\alpha}$. Since a monomer residue can interact with other monomers within the same layer, as well as those in the two adjacent layers, we define a layer-averaged value for each $\phi_k^{\alpha}(i)$. This is calculated as

$$<\phi_k^{\alpha}(i)>=\lambda_{-1}\phi_k^{\alpha}(i-1)+\lambda_0\phi_k^{\alpha}(i)+\lambda_1\phi_k^{\alpha}(i+1) \qquad \text{(S3)}$$

where $\lambda_{-1} = \lambda_1 = 1/6$ and $\lambda_0 = 4/6$, reflecting the possible number of neighbours in each layer (for our chosen cubic grid). The longer-range electrostatic interaction between charged groups is represented by the fourth term in equation (S2), where the electric potential ψ_{el} itself depends on the spatial distribution of the charged monomers and ions, as governed in accordance to Poisson's equation. With ψ_{el} expressed in units of (k_BT/e), length in units of (a_0) and charge density in units of (e/a_0^3), Poisson equation, in its discretised form used in our calculations, reads:

$$\psi_{el}(i+1)+\psi_{el}(i-1)-2\psi_{el}(i) = -\frac{4\pi l_B}{a_0}\sum_{k}\sum_{\alpha}q_{\alpha}\phi_k^{\alpha}(i) \qquad \text{(S4)}$$

where $l_B = e^2/(4\pi k_B T\varepsilon_0\varepsilon_r)$ in equation (S4) is the Bjerum length in the solvent (assumed to be water with a relative dielectric constant ε_r=80).

The mathematical process of averaging out the spatial positions of monomers, leading to the free energy functional in equation (S2), also introduces a set of auxiliary fields $\{\psi^{\alpha}(i)\}$ acting on each type of monomer α (as well as solvent molecules, ions, etc.), within every layer i. The physical interpretation of these fields can roughly be understood as follows. For any chosen variation of the density profiles across the gap, these fields project out that $\{\phi_k^{\alpha}(i)\}$,

when they are applied to an equivalent system in which all internal interactions (including electrostatic ones) have been switched off. Interestingly, one can prove that the set of density profiles $\{\phi_k^\alpha(i)\}$ and their corresponding fields $\{\psi^\alpha(i)\}$ leading to smallest value of $\Delta F(L)$ in equation (S2), satisfy the following relation:

$$\psi^\alpha(i) = \psi_h(i) + q^\alpha \psi_{el}(i) + \sum_k \sum_\beta \chi_{\alpha\beta} \left(< \phi_k^\beta > - \Phi_k^\beta \right) \qquad \text{(S5)}$$

Since the minimisation has to be conducted with due consideration for the restriction imposed by the incompressibility condition, i.e. Eq. (S1), this introduces a set of so called Lagrange multipliers for each layer *i*, associated with the constraints. These can be seen appearing as $\psi_h(i)$ in eq. (S5). Again, a simple and somewhat more physical interpretation is to consider them as a hard-core interaction, equally applying to all monomers in a given layer irrespective of their type, enforcing the incompressibility requirement expressed in Eq. (S1).

As indicated in the methodology section in the main text, the focus of SCF calculation is to determine the dominant density profiles that satisfy Eq. (S5) and thus minimises the free energy of the system. It was also outlined there, that this can be carried out using an iterative type procedure. Equation (S5) allows one to calculate fields $\{\psi^\alpha(i)\}$ for a given set of density profiles. To be able to conduct the iterative process successfully, one needs to be able to do the reverse and calculate the density profiles $\{\phi_k^\alpha(i)\}$ resulting from a given set of fields. A very efficient and fast method for performing this is based on first computing a set of segment probability functions $G_k^f(s,i)$ and $G_k^b(s,i)$. These are proportional to the probability that the first *s* monomer residues of a chain of type *k* will have a conformation such that the s^{th} monomer resides in the layer *i*. As protein fragments of interest in our study will generally have an asymmetric primary sequence, two such probabilities need to be defined, one for each end of the chain from which the *s* monomers are considered. The two are distinguished from each other by the use of the backwards and forwards superscripts "*f*" and "*b*" here. What makes the calculation of the segment probability functions simple is the connectivity of the chains. This means that if the s^{th} monomer is in layer *i* then $(s\text{-}1)^{\text{th}}$ residue must have been in one of the adjacent layers, *i*-1 *or i*+1, or is positioned in the same layer *i*. This condition at once leads to the following recursive relation for the segment probability functions:

$$G_k^f(s,i) = \frac{1}{6}\left[G_k^f(s-1,i) + 4G_k^f(s-1,i) + G_k^f(s-1,i) \right] \exp\left[-\psi^{t_f(s)}(i) \right] \qquad \text{(S6.a)}$$

for the forward, and

$$G_k^b(s,i) = \frac{1}{6}\left[G_k^b(s-1,i) + 4G_k^b(s-1,i) + G_k^b(s-1,i) \right] \exp\left[-\psi^{t_b(s)}(i) \right] \qquad \text{(S6.b)}$$

for the backward segment probability. In equation (S6), we have also defined two functions $t_f(s)$ and $t_b(s)$, which evaluate to the kind number to which the s^{th} monomer (counting from the appropriate chain end) belongs. Starting with $G_k^b(1,i) = \exp\left[-\psi^{t_b(s)}(i) \right]$ and $G_k^f(1,i) = \exp\left[-\psi^{t_f(s)}(i) \right]$, it is easy to see how through the use of the above recursive

relations, the full set of segment density functions for all s=1 to N_k for any chain of type k can quickly be computed. With the full set of segment probability functions now available, the density profile for any type of monomer can be obtained using the composition law[1, 3, 7]

$$\phi_k^\alpha(i) = \frac{\Phi_k^\alpha \sum_{s=1}^{N_k} G_k^f(s,i) G_k^b(N_k+1-s,i)\delta_{\alpha,t_f(s)}}{\exp\left[-\psi^\alpha(i)\right]} \qquad \text{(S7)}$$

where $\delta_{i,j}$ is the Kronecker delta function, evaluating to 1 if i=j and 0 otherwise.

Starting from an initial rough guess for $\{\phi_k^\alpha(i)\}$, the iterative process now substitutes the resulting fields from Eq. (S5) back into (S6) and (S7) to obtain an improved set of density profiles. These are used then in (S5) to obtain a new set of $\{\psi^\alpha(i)\}$ and the process repeats. Convergence is obtained when the values no longer change (to within a desired tolerance error) from one step to the next. The minimum value of the free energy can now be calculated from the computed $\{\phi_k^\alpha(i)\}$ and associated $\{\psi^\alpha(i)\}$ using Eq. (S2). Finally, the interaction potential induced between the two interfaces, resulting from the presence of chains is obtained as

$$V_{pl}(L) = \Delta F(L) - \Delta F(\infty) \qquad \text{(S8)}$$

where $\Delta F(\infty)$ is the value of the free energy when the two plane interfaces are sufficiently far not to influence each other, i.e. the presence of one plane does not affect the adsorption in the vicinity of the other. The above calculated interactions are for two flat interfaces, but can easily be converted to those for two spherical droplets, using the so called Derjaguin approximation[13]

$$V_{sp}(r) = \pi R \int_r^\infty V_{pl}(L)dL \qquad , \qquad \text{(S9)}$$

valid when the range of interaction is expected to be far smaller than the radius of droplets, R. The value of R used in our circulations was 0.5 μm (i.e. 1 μm sized droplets). However, it is easy to repeat the analysis for other values of R. In general, as seen from equation (S9), the interactions scale linearly with size.

The SCF calculations efficiently evaluate all the conformations of a chain in the field produced by its neighbouring molecules. Being a numerical calculation (as opposed to a simulation) it is considerably faster than simulation techniques such a molecular dynamics or Monte Carlo simulations. In particular, it is much easier to extract values for free energy changes and hence induced interaction potentials, using SCFC than any of these other techniques. Nonetheless, the method has a number of disadvantageous that should also be borne in mind. Firstly, the method is limited to studying equilibrium behaviour. In other words, no dynamical or transient phenomenon, such as speed of adsorption can be investigated using this method. Secondly, the calculations assume that all configuration that can be taken by a chain are available for it to adopt (i.e. the system is ergodic). This is very

often not the case if the protein is trapped in a long-lasting metastable state, as may be the case with highly folded globular proteins. Here we assume that the process of hydrolysis is sufficient to destroy most of the secondary or tertiary structure of the resulting polypeptides. Finally, as was eluded to further above, SCFC is a mean field type theory. That is to say that it assumes that the behaviour of the system is dominated by the most probable density profile in the gap between the two surfaces (i.e. the density profile variation that minimises the free energy). Any fluctuations about this most probable state are ignored. However, this last condition is often satisfied for densely covered surfaces.

S.2 Calculation of DH

In the idealized model mix system considered here, we only have two protein fragments present. Let us denote the number of monomer residues for each of the two polypeptides as N_1 and N_2, with chains having n_1 and n_2 remaining cleavable bonds, susceptible to further hydrolysis by enzyme, still unbroken in their primary structure. Then their presence in the bulk solution, in terms of their relative volume fractions Φ_1/Φ_2, has the following relationship with the probability p of getting a susceptible (to the chosen selective enzyme) peptide bond cleaved

$$\frac{\Phi_1}{\Phi_2} = \frac{N_1 p^2 (1-p)^{n_1}}{N_2 p^2 (1-p)^{n_2}} \tag{S10}$$

In deriving the above equation, we make the assumption that all the cleavable bonds (i.e. the C-terminal end of peptide bonds involving either Lysine or Arginine), have the same likely chance of being broken by trypsin. With this simplification in mind, then the probability p of getting a susceptible peptide bond broken at a given DH is as follows:

$$p = \frac{DH}{DH_{\max}} \tag{S11}$$

where DH_{max} is the maximum DH value that can be achieved by enzymatic hydrolysis of the chosen selective enzyme (trypsin here).

With both equation (S10) and (S11) at hand, DH can be estimated for any desired bulk volume fraction ratio in the solution between our two protein fragments, according to equation (S12) below:

$$DH = DH_{\max}\left(1 - \left[\frac{\Phi_1 N_2}{\Phi_2 N_1}\right]^{1/(n_1 - n_2)}\right) \tag{S12}$$

Now, the soybean β-conglycinin α' subunit considered in our study is made of 621 amino acid residues, of which only 84 can be cleaved by trypsin. Thus, the maximum DH achievable via trypsin hydrolysis is 84/(621-1) =13.55% (i.e. DH_{max}=13.55%). The two polymers studied in our mixed system, i.e. conjugated Glu^{93}-Arg^{302} and the shorter unreacted fragment

Met[322]-Lys[355], consist of 210 (or 260 with the attached hydrophilic part) and 34 monomers, respectively. That is $N_1 =$ 260 and $N_2 =$ 34. Furthermore, they contain 32 and 2 remaining cleavable sites according to their primary structure, i.e. $n_1 = 32$ and $n_2 = 2$. Thus, with these numbers, equation (3) become:

$$DH = 13.5\% \left(1 - \left[\frac{34\Phi_1}{260\Phi_2} \right]^{1/30)} \right) \qquad \text{(S13)}$$

Thus, when the two biopolymers are present on the hydrophobic surface in the ratios of 1:1, 1:3 and 1:5, the estimated DH for each case is calculated accordingly from equation (13), using the corresponding volume fraction ratio in bulk Φ_1/Φ_2. The bulk volume fraction and the DH values for each case are provided in the Table S2.1.

Table S2.1 - Estimated DH at a given volume fraction ratio of conjugated Glu[93]-Arg[302] and short unconjugated Met[322]-Lys[355]

Φ_1/Φ_2	estimated DH
1:30	2.25 %
1:100	2.69 %
1:150	2.84 %

S.3 Properties of other fragments

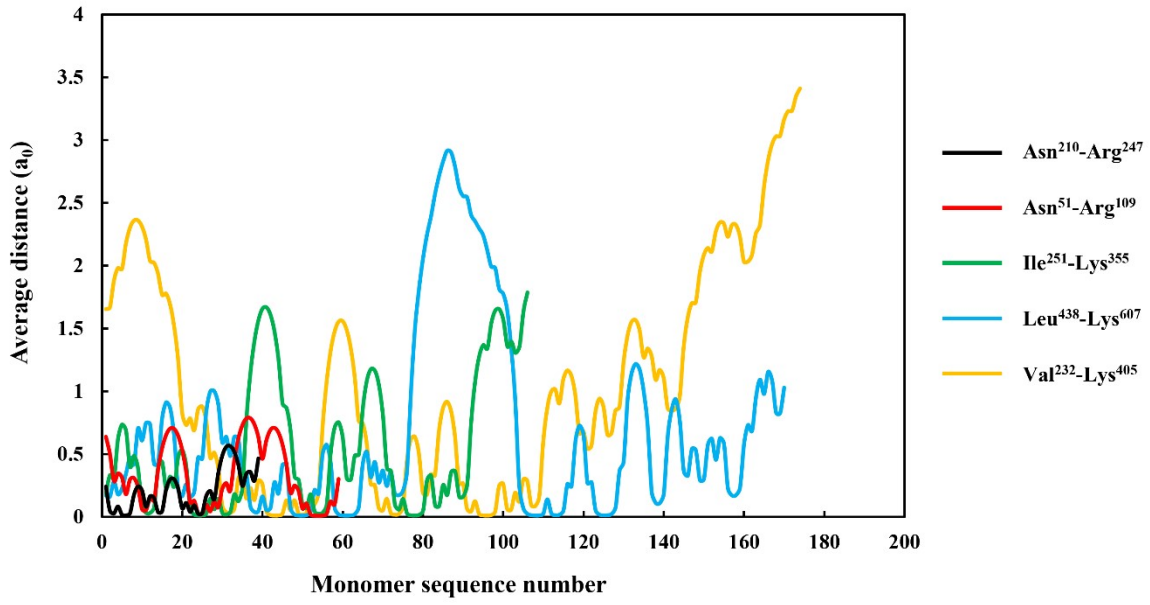

Fig. S.1 - The average distance of each monomer residue, making up the adsorbed soy-derived fragments (Asn[210]-Arg[247], Asn[51]-Arg[109], Ile[251]-Lys[355], Leu[438]-Lys[607] and Val[232]-Lys[405].), measured perpendicularly away from a hydrophobic surface. The distance (in units of monomer size), is plotted against the sequence number of

monomers, starting with the first monomer at *N*-terminus end of a polypeptide. Results were obtained at a background electrolyte volume fraction of 0.01 (roughly equals to 100 mM NaCl) and at a solution pH = 5.5.

The average conformations of five more protein fragments that can also arise from hydrolysis of soy protein by trypsin are provided here. The result show the average distance of each monomer of the fragment away from the surface. It is observed that roughly speaking, the fragments take three kinds of conformations, i.e. 1) lying flat on the surface, 2) resembling a di-block-like polymer, or 3) acting somewhat like a tri-block chain. In particular, shorter polypeptides tend to adopt flat configurations on the interface, while there is a higher chance of encountering a di-block or tri-block-like behaviour for larger fragments.

References

1. J. Scheutjens and G. J. Fleer, *Journal of Physical Chemistry*, 1979, **83**, 1619-1635.
2. J. Scheutjens and G. J. Fleer, *Macromolecules*, 1985, **18**, 1882-1900.
3. O. A. Evers, J. Scheutjens and G. J. Fleer, *Macromolecules*, 1990, **23**, 5221-5233.
4. O. A. Evers, J. Scheutjens and G. J. Fleer, *Journal of the Chemical Society-Faraday Transactions*, 1990, **86**, 1333-1340.
5. G. J. Fleer, *Adv. Colloid Interface Sci.*, 2010, **159**, 99-116.
6. G. J. Fleer, M. A. Cohen Stuart, J. M. H. M. Scheutjens, T. Cosgrove and B. Vincent, *Polymers at interfaces*, Chapman and Hall, London, 1993.
7. F. A. Leermakers, P. J. Atkinson, E. Dickinson and D. S. Horne, *Journal of Colloid Interface Science*, 1996, **178**, 681-693.
8. G. Fleer, M. C. Stuart and F. Leermakers, in *Fundamentals of Interface and Colloid Science*, ed. J. Lyklema, Academic Press, 2005, vol. 5, pp. 1.1-1.94.
9. R. Ettelaie, B. S. Murray and E. L. James, *Colloid Surf. B-Biointerfaces*, 2003, **31**, 195-206.
10. R. Ettelaie, N. Khandelwal and R. Wilkinson, *Food Hydrocolloids*, 2014, **34**, 236-246.
11. A. Y. Grosberg and A. R. Khokhlov, *Statistical Physics of Macromolecules* AIP press, New York, 1994.
12. I. M. Lifshitz, A. Y. Grosberg and A. R. Khokhlov, *Reviews of Modern Physics*, 1978, **50**, 683-713.
13. R. J. Hunter, *Foundation of colloid science*, Clarendon press, Oxford, 1987.