

Molecular modeling of the interface of an egg yolk protein-based emulsion

Original

Molecular modeling of the interface of an egg yolk protein-based emulsion / Ferrari, Marco; Handgraaf, Jan-Willem; Boccardo, Gianluca; Buffo, Antonio; Vanni, Marco; Marchisio, Daniele L.. - In: PHYSICS OF FLUIDS. - ISSN 1070-6631. - ELETTRONICO. - 34:2(2022), p. 021903. [10.1063/5.0079883]

Availability:

This version is available at: 11583/2955762 since: 2022-03-02T11:56:11Z

Publisher:

AIP

Published

DOI:10.1063/5.0079883

Terms of use:

This article is made available under terms and conditions as specified in the corresponding bibliographic description in the repository

Publisher copyright

AIP postprint/Author's Accepted Manuscript e postprint versione editoriale/Version of Record

(Article begins on next page)

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0079883

Revised manuscript

Molecular modeling of the interface of an egg yolk protein-based emulsion

Marco Ferrari,¹ Jan-Willem Handgraaf,² Gianluca Boccardo,¹ Antonio Buffo,^{1, a)} Marco Vanni,¹ and Daniele L. Marchisio¹

¹⁾*Department of Applied Science and Technology, Politecnico di Torino, Corso Duca degli Abruzzi 24, 10129, Torino, Italy*

²⁾*Siemens Industry Software Netherlands B.V., Galileiweg 8, 2333 BD, Leiden, The Netherlands*

(Dated: 4 January 2022)

Many food emulsions are stabilized by functional egg yolk biomolecules, which act as surfactants at the oil/water interface. Detailed experimental studies on egg yolk emulsifying properties have been largely hindered due to the difficulty in isolating individual chemical species. Therefore, this work presents a molecular model of an oil/water interfacial system where the emulsifier is one of the most surface-active proteins from the egg yolk low-density lipoproteins (LDL), the so-called Apovitellenin I. Dissipative Particle Dynamics (DPD) was here adopted in order to simulate large systems over long time-scales, when compared with full-atom molecular dynamics (MD). Instead of a manual assignment of the DPD simulation parameters, a fully-automated coarse-graining procedure was employed. The molecular interactions used in the DPD system were determined by means of a parameter calibration based on matching structural data from atomistic Molecular Dynamics (MD) simulations. Despite of the little availability of experimental data, the model was designed to test the most relevant physical properties of the protein investigated. Protein structural and dynamics properties obtained via MD and DPD were compared highlighting advantages and limits of each molecular technique. Promising results were achieved from DPD simulations of the oil/water interface. The proposed model was able to properly describe the protein surfactant behavior in terms of interfacial tension decrease at increasing protein surface concentration. Moreover, the adsorption time of a free protein molecule was estimated and, finally, an LDL-like particle adsorption mechanism was qualitatively reproduced.

^{a)}Author to whom correspondence should be addressed: antonio.buffo@polito.it

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

1 I. INTRODUCTION

2 Food emulsions are made of a continuous water phase, a disperse phase with a high content
3 of oil, and a surfactant that stabilizes the oil drops.¹⁻⁵ The droplet size distribution (DSD)
4 is the most important property of the emulsion since the structure, stability, taste, and
5 color of the final product depend on the DSD.¹⁻⁵ The DSD in turn depends on the emulsion
6 composition, the type of process and the operating conditions under which the production
7 process operates.⁶ The production of emulsions is based on mixing the ingredients and
8 applying a suitable mechanical energy to the emulsion for promoting droplet formation and
9 breakage, in order to reach the desired DSD. A typical mixing process is composed by two
10 steps: first, the ingredients (mainly egg yolk, vinegar, oil, water, salt) are mixed together in
11 large stirred vessels at moderate rotational speed; then, this premixed emulsion is fluxed into
12 a high-shear device, commonly a cone mill mixer, where the oil droplets undergo breakage
13 until the final size distribution is reached.³⁻⁵ This last step is crucial to fine-tune the DSD,
14 in order to determine the properties of the final product.

15 Many food emulsions are stabilized by surface-active biopolymers that adsorb on the
16 droplet surface and form protective coatings.¹ Some of these functional molecules are integral
17 components of more complex food ingredients used in food products (e.g., egg yolk, milk, and
18 flour).^{1,2} Although the egg yolk is recognized as one of the most widely employed emulsifiers
19 for both industrial and home-made food emulsion preparation,¹ many issues need to be
20 addressed, especially the adsorption mechanism of egg yolk proteins at oil-water interface
21 and their emulsifier behaviour.⁷ Indeed, the egg yolk is a complex system with different
22 structural levels consisting in non-soluble protein aggregates (granules) in suspension in
23 a clear yellow fluid (plasma) that contains low-density lipoproteins (LDLs) and soluble
24 proteins.⁷ Experimental research concerning the emulsifying properties of egg yolk proteins
25 has been hindered by the difficulties in extracting individual components from the complex
26 matrix, therefore, they are less amenable to detailed study by being less readily available in
27 pure form.⁸⁻¹⁰

28 During the emulsification process, the interfacial properties between disperse and contin-
29 uous phases play an essential role in the formation and the stabilization of the oil droplets.^{1,2}
30 Therefore, it is important to have a fundamental understanding of the factors that influ-
31 ence the type, concentration, interactions, and arrangement of surface-active molecules at

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

32 interfaces.^{1,2} Computer modeling techniques can greatly enhance the comprehension of the
33 way the molecules organize themselves in a liquid.^{11–14} Molecular simulations can provide
34 valuable insight into the relationship between molecular properties and structural organiza-
35 tion that are relevant for a better understanding of the behavior of food emulsions, including
36 the miscibility/immiscibility of liquids, the formation of surfactant micelles, the adsorption
37 and displacement of emulsifiers at interfaces, the transport of nonpolar molecules through
38 aqueous phases, the conformation and flexibility of biopolymers in solution, polymer inter-
39 actions, and the formation of gels.^{15–24} The first step in a molecular simulation is to define
40 the characteristics of the molecules involved (e.g., size, shape, flexibility, and polarity) and
41 the nature of the intermolecular pair potentials that act between them, making a number
42 of simplifying assumptions as a compromise between the model reliability and a reasonable
43 computational time.²⁵ A collection of these molecules is arbitrarily distributed within a box
44 that represents a certain region of space, and the change in the conformation and/or orga-
45 nization of the molecules is then monitored as they are allowed to interact with each other.
46 Depending on the simulation technique used, one can obtain information about the evolution
47 of the structure with time and/or about the equilibrium structure of the molecular ensem-
48 ble. The most commonly used computer simulation techniques in this context are the Monte
49 Carlo approach and Molecular Dynamics (MD). In these models the involved molecules can
50 be described with all their atomistic details or some of them can be coarse-grained, as in
51 Dissipative Particle Dynamics (DPD).^{19,26–30}

52 Many molecular modeling studies of food structures were carried out employing the afore-
53 mentioned approaches.¹⁹ The adsorption of flexible proteins (β -casein³¹ and a proteinlike
54 heteropolymer³²) at an oil-water interface was studied by means of Monte Carlo simula-
55 tions. On the other hand, the majority of MD studies on protein adsorption at fluid inter-
56 faces have been on globular proteins using both all-atom and coarse-grained models, with
57 few studies on unstructured intrinsically disordered proteins.^{33–40} Few works have been car-
58 ried out on protein models via coarse-grained DPD technique, although this approach allows
59 the simulation of large systems over relatively long-time scales with respect to full-atomistic
60 studies.^{28,29,41} DPD uses simplified soft potentials and coarse-grained representations of mod-
61 eled structures.^{27–29} In contrast to MD, in DPD systems the intended physical properties are
62 determined by means of parameter calibration. One of the most popular method of calibra-
63 tion is based on mapping onto Flory–Huggins theory.²⁹ Another approach is to couple DPD

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

64 with MD simulations to calibrate models by matching the structural data from the atomistic
65 simulations.⁴²⁻⁴⁴ Previous DPD studies investigated the adsorption of semi-flexible rod-like
66 objects,⁴⁵ conformation changes⁴⁶ or the folding of small proteins.⁴⁷ However, all computer
67 molecular techniques have been successfully employed in modeling of interfacial systems and
68 in the calculation of the surface tension when an amphiphilic non-protein molecule act as a
69 surfactant.⁴⁸⁻⁵¹ Moreover, DPD is well-suited for modeling of multi-component systems such
70 as emulsions, and it has been used in a number of studies to look at the effect of adsorbing
71 molecules on the stability of oil or water droplets in emulsions.^{19,52-54} These have mainly
72 been carried out on hydrocarbon oil emulsions with synthetic copolymers as the adsorbing
73 molecules, but the methodology and the general results are relevant also for food emulsions.

74 The main goal of the present work is to model an oil/water interfacial system where
75 the emulsifier is one of the most surface-active proteins from the egg yolk LDL, in order to
76 provide new insights into physics of the food emulsion production process. Despite of the
77 little availability of experimental data, the model was designed to test the most relevant
78 physical properties of such a protein by means of the DPD approach in which the parameter
79 calibration is based on MD simulations. Instead of a manual assignment, a fully *automated*
80 coarse-graining procedure was employed to the molecules involved in the ternary system,
81 assuming a flexible, disordered structure for the protein. Promising results were obtained
82 in terms of both equilibrium and dynamic properties of the egg-yolk protein. Finally, the
83 adsorption mechanism of a LDL-like particle is also qualitatively reproduced.

84 This paper is structured as follows: in Section II the molecular description of the studied
85 system is presented; the molecular techniques here used are briefly introduced in Section
86 III; the model development and calibration are explained in Section IV together with all the
87 simulation details; Section V shows the relevant results of systems investigated and, finally,
88 in Section VI the main conclusions are reported.

89 II. MOLECULAR DESCRIPTION OF THE MACROSCOPIC SYSTEM

90 The first step in the development of the molecular model for an egg yolk protein-based
91 emulsion is to identify the chemical species to be simulated and to define the characteristics
92 of the molecules involved at the interface. The basic components of the system under
93 investigation are three: the triglyceride with three monounsaturated oleic acid residues

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

94 which stands for the oil phase, the protein Apovitellenin I coming from the egg yolk LDL
95 and, finally, water. In this Section a general description of the macroscopic system to be
96 modeled is provided, together with the adopted simplifications.

97 An example of a food emulsion where the egg yolk is widely used as an emulsifier is
98 mayonnaise. This is a stable liquid-liquid emulsion with a high content of the dispersed oil
99 phase. In this work a regular mayonnaise with around 70% of fat content¹ is considered and
100 the experimental work of Dubbelboer *et al.*³ is used as a reference to identify the ingredients
101 of the mayonnaise, especially the molecules to play a primary role at the oil/water interface.
102 It is important to highlight that also in this work the dispersed phase consists of the soybean
103 oil, while the chemical species that act as surfactants are derived from the egg yolk. These
104 two components characterize the specific type of mayonnaise studied, therefore a further
105 description of the vegetable oil and the egg yolk used in the production of the food emulsion
106 is presented in order to correctly select the molecules to be modeled.

107 Regarding the dispersed phase, a fully refined soybean oil is employed in which the
108 triglyceride molecules are present with a concentration larger than 99%.⁵⁵ Triglycerides are
109 tri-esters consisting of a glycerol bound to three fatty acid molecules. Based on the number of
110 double bonds and the chain length, the fatty acids occurring in triglycerides of the soybean oil
111 are saturated, monounsaturated and polyunsaturated with 16 or 18 carbon atoms according
112 to an internal distribution.⁵⁵ For the sake of simplicity, here homotriglycerides are taken
113 into account where the three fatty acids are identical (without an internal distribution). In
114 particular, the triglyceride molecules with three monounsaturated oleic acid residues (18
115 carbon atoms for chain) will be modeled as the representative of the oil phase, instead of
116 hydrocarbons as it was done in previous DPD works on similar emulsions.⁵²⁻⁵⁴ It should be
117 noted that the protein adsorption to different hydrophobic materials may cause differences
118 in the conformation of the adsorbed molecule; in this sense our simplification may have
119 an impact that it is difficult to quantify. That being said, it is known that the modeling
120 of a simpler hydrocarbon–water system instead of a triglyceride–water system might not
121 necessarily lead to realistic results,⁵⁶ therefore a triglyceride–water system was modeled in
122 this work.

123 The second fundamental component in the mayonnaise production is the hen egg yolk.
124 It is mainly composed of two fractions – plasma and granules – which are natural nano-
125 and micro-assemblies. Plasma contains a large quantity of lipids structured as low-density

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

lipoproteins (LDLs), whereas granules are mainly composed of proteins aggregated in micrometric assemblies.⁷ Assuming a pH equal to 3.8 for the mayonnaise,⁵⁷ plasma proteins represent about 2/3 of oil–water interface in acidic conditions (at all ionic strengths).⁷ Previous works have shown that LDLs are likely to play primary roles in the formation and stabilization of egg yolk-based emulsions.^{7,58–61} Consequentially, LDLs are considered to contribute mainly to yolk emulsifying properties.⁷ LDLs are spherical nanoparticles (17–60 nm) with a lipid core of triglycerides and cholesterol esters in a liquid state surrounded by a monofilm of phospholipids and apoproteins.^{7,62–67} The LDL adsorption mechanism at the oil-water interface was investigated by several works.^{7,67–71} In fact, LDLs serve as vectors of surfactant constituents (proteins and phospholipids) that could not be soluble in water until they reach the interface. The adsorption of apoproteins and phospholipids at the interface lead to the formation of a film that stabilize the emulsion.⁶⁹ Therefore, both apoproteins and phospholipids are essential to understand the interfacial properties of egg yolk LDLs. The protein identified as Apovitellenin I is considered to be the most surface-active, among the apoproteins contained in LDL.^{64,67} Due to its structure and composition, which combines amphipathic character and flexibility, Apovitellenin I shows a great capacity to adsorb at the oil–water interface in emulsions.⁶⁷ In LDL, Apovitellenin I is mostly present as a homodimer, thus containing two identical polypeptide chains of 82 amino acid residues which are linked by a single disulfide bond at the cysteine residue.^{64,67} The sequence of the mature protein is available in the UniProtKB database⁷² under the accession number P02659 (www.uniprot.org/uniprot/P02659). However, the detailed 3D structure and other physico-chemical information of Apovitellenin I are not available in the literature to the best of authors' knowledge, increasing the complexity of its modeling approach. The presence of salts, small surfactant molecules (phospholipids) or other additives is here neglected since only the emulsifying capacity of the considered egg yolk LDL protein is investigated. Furthermore, the pH of the system is kept constant and equal to 3.8. The molecular model of the oil-water interface is then described in the following sections.

III. THEORETICAL BACKGROUND

In this Section only the main basic concepts of the standard Dissipative Particle Dynamics (DPD) method are presented, while a further detailed description of both MD and DPD

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

156 techniques can be found in the literature^{25,27-29,73,74} and in the Supplementary Material.

157 DPD is a stochastic mesoscale particle model that it has been devised to allow the sim-
 158 ulation of the dynamics of mesoscopic particles. Unlike classic Molecular Dynamics, each
 159 DPD particle i , called bead, represents a molecular cluster (a molecule fragment or a group
 160 of solvent molecules) rather than an individual atom. The major difference between MD
 161 and DPD, apart from the coarse-grained nature of the molecules, is the nature of the forces
 162 between them. The force acting on each bead i contains three parts: the conservative,
 163 dissipative, and stochastic (random) forces, each of which is pairwise additive. Here the
 164 conservative force felt by bead i includes: 1) contributions from repulsive interactions with
 165 surrounding beads; 2) contributions due to the springs connecting bead i to other beads in
 166 the same molecule; and 3) contributions due to angle bending interactions. The repulsive
 167 force \mathbf{F}_{ij}^r , which is modeled as a soft repulsion between beads i and j , is defined as follows:

$$\mathbf{F}_{ij}^r = \begin{cases} a_{ij}(1 - r_{ij}/r_c)\hat{\mathbf{r}}_{ij} & \text{if } r_{ij} \leq r_c \\ 0 & \text{if } r_{ij} > r_c \end{cases}, \quad (1)$$

168 where $r_{ij} = |\mathbf{r}_i - \mathbf{r}_j|$ is the distance between beads i and j at positions \mathbf{r}_i and \mathbf{r}_j respectively,
 169 and $\hat{\mathbf{r}}_{ij} = (\mathbf{r}_i - \mathbf{r}_j)/r_{ij}$ is the direction between the two beads. The parameters a_{ij} are
 170 the DPD interaction parameters defined for each bead pair, while r_c stands for the cutoff
 171 distance. For the system investigated in this work, their definition will be given in the
 172 Section IV B and they will be here used as fitting parameters for the calibration of the DPD
 173 model. The adjacent beads are constrained with permanent lengths and angular bonds. In
 174 this study, the bonds were modeled using harmonic spring quadratic potentials given as:

$$U_{ij}^S = k_S(r_{ij} - l_H)^2, \quad (2)$$

$$U_{ijk}^A = k_A(\theta_{ijk} - \theta_H)^2, \quad (3)$$

176 where l_H and θ_H are the equilibrium lengths and angles for beads i , j and k . The stiffness
 177 of the length and angular bond constraints is defined by the values of k_S and k_A .

178 As it is customary in DPD, the quantities here reported have to be considered reduced
 179 (dimensionless) and the scaling factors for the main properties (mass, length, time, energy)
 180 will be explained in Section IV C. Finally, it is important to point out that the coarse-
 181 graining of the molecular structures and the soft interactions allow larger systems to be
 182 modeled over significantly longer times than with (atomistic scale) molecular modeling,^{41,74}

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

183 thus allowing the dynamics of mesoscopic systems to be followed over relevant time scales
184 as well as length scales.

185 IV. MODELING DETAILS

186 In order to consider both the complex composition of the emulsion and the equilibration
187 time required by macro-molecules to re-arrange at interfaces, the DPD approach is employed
188 in which the parameter calibration is based on MD simulations. Next sections will present
189 the setup of MD simulations, the DPD model development in which both the coarse-graining
190 procedure and the calibration of parameters are explained and, finally, definitions of the main
191 physical properties investigated here.

192 A. MD simulations

193 The purpose of all-atom MD simulations is to use their results to calibrate the DPD
194 parameter set. Only MD simulations of one protein molecule in bulk phases (water or oil)
195 were performed rather than the entire ternary interfacial system due to the size of the latter
196 which would require excessive computational time. An initial guess of both protein and
197 triglyceride structures was manually made from scratch via a molecule editor. In particular,
198 Figure 2a shows the all-atom protein model. It can be clearly seen the disulfide bond linking
199 two identical polypeptide chains. Furthermore, the N- and C- terminal amino acid residues
200 and, if applicable, the functional group of side chains were protonated or deprotonated by
201 comparing their corresponding pK_a with the pH of the solution.⁷⁵ Thus, at pH 3.8 the
202 net charge of the protein homodimer results equal to 16 e and the protein molecular mass
203 M is 18675.6 Da. MD simulations were performed using the OPLS-AA force field,^{76,77}
204 while water was described by the TIP3P water model.⁷⁸ A cutoff of 7.5 Å was used for
205 long-range interactions, and both electrostatic and van der Waals interactions were handled
206 using a smooth particle mesh Ewald summation method (SPME).⁷⁹ For the protein and
207 the triglyceride, first 20-ps simulation in vacuum with a time step of 1 fs was performed
208 on the single molecule to relax its initial structure. Before solvation with water or oil,
209 the protein was centered in a rectangular box with a minimum distance of any part of
210 the molecule defined to be at least 1 nm from box walls in order to satisfy the minimum

211 image convention when using periodic boundary conditions. According to the reproduced
212 environment, the box was filled with respectively 15994 water or 325 triglyceride molecules,
213 plus 16 Cl^- counterions to ensure the electroneutrality of the system. Thus, the resulting
214 MD box contains a total of 50694 or 56987 atoms in the case of protein in water or oil bulk
215 respectively. After a simple energy minimization to ensure that the system had no steric
216 clashes or inappropriate geometry, a 0.5-ns NPT (i.e., constant number of particles, pressure,
217 and temperature) equilibration simulation at ambient pressure (1 atm) and temperature (298
218 K) was performed. Pressure and temperature were fixed using the Berendsen barostat and
219 thermostat⁸⁰ and the Verlet algorithm was used to integrate the equations of motion with an
220 increased time step of 2 fs. To verify that the system was at the equilibrium, the fluctuations
221 in the temperature, pressure, density, and potential energy were monitored. In particular,
222 the average density reached during the last 0.2 ns of equilibration simulation was equal
223 to 1059.57 and 921.85 kg/m^3 respectively for the protein in water and in oil system, both
224 with fluctuations in the 0.1%. Finally, NVT (i.e., constant number of particles, volume,
225 and temperature) production simulations ranging from 2 to 6 ns were performed to collect
226 statistically averaged results by saving particle trajectories every 250 time steps.

227 B. Coarse-graining procedure and parameter calibration

228 The main steps of the DPD model development are summarized in a schematic diagram
229 in Figure 1, in which each stage is explained in this Section.

231 The first step toward a realistic DPD molecular model is to obtain the coarse-grained (CG)
232 representation of the molecules together with their full parameter set of both inter- and intra-
233 molecular interactions. For this scope, the Automated Fragmentation and Parametrization
234 (AFP) method is used and here a very brief introduction to this approach is provided. For
235 a fully detailed discussion on it, the reader can refer to the work of Fraaije *et al.*⁸¹.

236 Starting from their fully atomistic representations, the molecules involved in the investi-
237 gated system are fragmented according to a scoring function, through a simulated annealing
238 function that cuts through bonds; the optimal bond fission pattern is preserved and the
239 fragments are stored. The scoring function is here defined as:

$$S = \left(1 - \frac{V}{V_0}\right)^2, \quad (4)$$

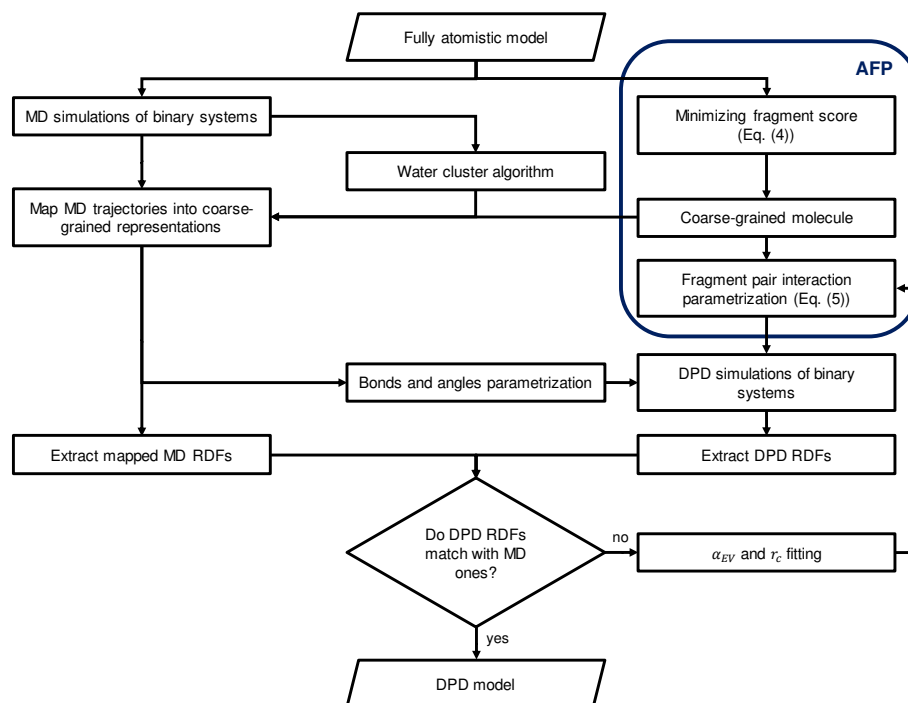


FIG. 1. Schematic diagram of the main stages followed in this work to develop the DPD model. See Section IV B for details of each step.

240 where V is the volume of the fragment and V_0 is the reference volume of a cluster of three
 241 water molecules in its lowest energy conformation (i.e., the reference volume used here is
 242 equal to 67.7 \AA^3 as in the original AFP work⁸¹). In this approach the molecule-unique frag-
 243 mentation is used in order to preserve as much as possible of the properties of the molecule.
 244 This means that the fragments are not database-unique, as is customary in coarse-grained
 245 simulations, but completely specific to a given molecule. By applying this fragmentation
 246 technique, the triglyceride molecule and the homodimer Apovitellenin I are comprised of
 247 20 and 500 beads respectively, while each water bead corresponds to three atomistic water
 248 molecules. In particular, Figure 2 shows the all-atom (a) and the corresponding coarse-
 249 grained (b) representation of the protein molecule.

251 In the AFP framework, the interaction DPD parameter a_{ij} is split into two contributions,

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

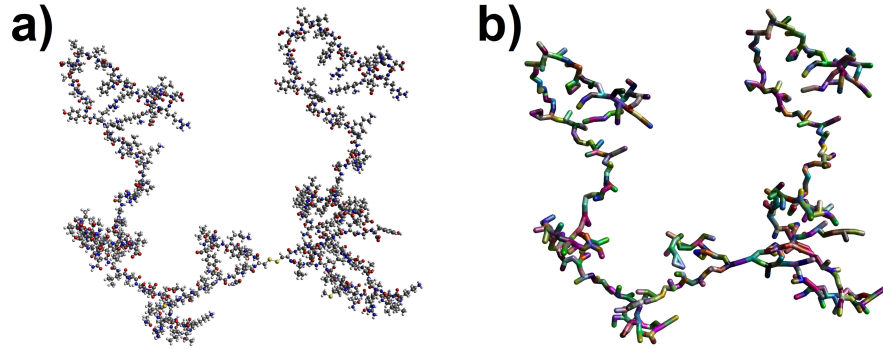


FIG. 2. All-atom (a) and corresponding coarse-grained (b) model obtained via AFP of Apovitelenin I. DPD beads are represented by colored fragments, highlighting the bond fission pattern.

one from the excluded volume and the second from the residual interactions:

$$a_{ij} = \alpha_{EV}v_i v_j + \alpha_{res}\sqrt{v_i v_j}\beta\Delta G_{res,ij}, \quad (5)$$

where $v_i = V_i/V_0$ is the scaled molecular volume of fragment i , $\beta = 1/k_bT$, α_{EV} and α_{res} represent two global adjustable parameters and $\Delta G_{res,ij}$ is the residual Gibbs energy of mixing of a *hypothetical* equimolar mixture of fragments i and j . The Gibbs energy of mixing was calculated through COSMO-RS calculations,^{82,83} using the charge envelope of the fragments (the so-called sigma profiles). The COSMO charge envelope is here computed via a modified version of AM1,⁸⁴⁻⁸⁶ using atomic partial charges derived from the charge equilibration (QEq) method.⁸⁷ By definition the residual Gibbs energy of mixing between identical fragments is zero, i.e., $\Delta G_{res,ii} = 0$, thus it follows trivially that a_{ii} is reduced only to the excluded volume contribution and, in particular, for water bead self-interaction $a_{ww} = \alpha_{EV}$. It is also important to point out here that the bead-size effect is taken into account in the definition of DPD a_{ij} parameter given in Eq. (5) by considering the fragment volume scaled with respect to the reference volume, V_0 , of a cluster of three water molecules. This allows to consider a constant DPD base unit of length, h , for all fragments irrespective of size or composition. As in the original AFP work,⁸¹ here the value of h is assumed equal to 7.65 Å as the yardstick for length in DPD approach. This value corresponds to five 3-mer water clusters per cell of size h^3 , or, in terms of the DPD dimensionless unit system, this corresponds to a density of 5 for water under ambient conditions. The soft-core repulsion

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

270 potential employed here is devoid of the short-range Lennard-Jones divergence. Also, the
271 typical long-range electrostatic Coulomb term is avoided completely, through using the close-
272 contact electrostatic interaction of the COSMO model. Both interactions are therefore
273 replaced by a soft repulsive potential that is local, with a length scale limited to the cutoff,
274 r_c . Hence, in AFP approach the fragment-specific chemical information is condensed in only
275 one parameter: the DPD a parameter. The magnitude of the repulsion (not the spatial
276 extension) is modified depending on the volume of the underlying molecular fragment, and
277 residual interactions. In order to map the characteristics of the atomistic models into the
278 DPD system, MD simulations of protein in water and oil bulks were used to extract molecular
279 characteristics such as radial distribution functions as well as the distributions of lengths
280 and angles for molecules bonded with length and angular bonds. To make MD and DPD
281 models physically comparable, it is necessary to map atomistically detailed trajectories into
282 their corresponding coarse-grained representations considering a length scale factor, h , to
283 convert atomistic coordinates and MD box dimensions into a CG model. When dealing with
284 the triglyceride and the protein in which their fragmentation information has been already
285 well-defined through the AFP approach, the mapped MD trajectories of such molecules
286 are easily determined by replacing the fully atomistic coordinates with the centre-of-mass
287 positions of provided molecular fragments. However, in the case of atomistic water models,
288 where the water particles move independently, their CG representation has to be dynamically
289 identified. Therefore, a clustering method is required to enable the mapping of multiple
290 water molecules into a single CG bead. Here, the water molecules clustering algorithm
291 proposed by Pieczywek, Płaziński, and Zdunek⁸⁸ was employed, which is based on a step-
292 wise iterative nearest neighbour search algorithm. The number of water molecules per bead
293 in all clusters is kept constant and equal to the degree of coarse-graining employed here,
294 i.e., a 3 to 1 CG ratio, corresponding to the number of clustering steps performed for each
295 simulation time frame. This represents the major advantage compared to other approach
296 where, instead, the total number of beads in the system have to be provided,⁸⁹ leading to
297 some issues converging with the desired number of equally sized clusters. Very briefly, as the
298 algorithm initialization, a grid of fixed-size cubes was superimposed onto the MD simulation
299 box and initial positions of bead centers were generated by randomly choosing coordinates
300 of water molecules from the first time frame. For each step of the algorithm, an iterative
301 search for the unique nearest water molecule was carried out in the area adjacent to the unit

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

302 cell in which the coarse-grained bead is located. The unique nearest water molecule was
303 defined by means of the Euclidean distance from the center-of-mass of a CG bead. When
304 all of the CG beads had the same number of molecules assigned to them (equal to the CG
305 ratio), the algorithm finished and the positions of the beads were updated by calculating
306 the center-of-mass of the molecular clusters. Hence, for each MD simulation time frame, the
307 water molecules were divided into equally sized groups based on their proximity.

308 The mapped MD trajectories were used to extract radial distribution functions (RDFs)
309 of coarse-grained molecules. Thus, using the AFP method as a basis, a further DPD param-
310 eter calibration was carried out by using the MD RDFs as reference curves to be compared
311 with those extracted from DPD simulations. Since the RDF is solely determined by the
312 conservative force,⁹⁰ the repulsion force coefficients was adjusted to match MD and DPD
313 RDFs. As the specific fragment pair interactions were defined in Eq. (5), the global ad-
314 justable parameters which serve to define the mutual repulsive interaction between *all* the
315 beads belonging to a single type of molecule can be used to calibrate the DPD model. In
316 particular, α_{EV} and the cutoff distance, r_c , were used as fitting parameters, while for all the
317 fragment pairs the DPD-sigma parameter was set to the standard value of 3.0^{29} and α_{res}
318 was kept equal to 6.1 as in the original AFP work.⁸¹ Therefore, from both MD and DPD
319 simulations of protein in water and in oil bulk, only RDFs referring to all beads belonging
320 to water, oil, and protein were extracted and the results of the calibration are presented and
321 discussed in Section V. Obviously, from simulations of the binary systems only water-water,
322 oil-oil, water-protein, and oil-protein interactions can be exactly calibrated. However, the
323 remaining interactions, i.e., oil-water and protein-protein, must be determined to build the
324 DPD model of the ternary system. In particular, the oil-water α_{EV} value was obtained by
325 simply fitting the experimental interfacial tension between purified soybean oil and water,⁵¹
326 found to be equal to 31-32 mN/m and independent on the presence of salt.⁹¹ For the protein-
327 protein repulsive interaction, the same α_{EV} value of water-protein was arbitrarily chosen as
328 a first guess. This value could be of paramount importance since the self-protein interaction
329 may effect the structural configuration of the protein as well as equilibrium and dynamics
330 properties of the ternary system. The study of protein-protein interactions needs therefore
331 a deeper insight, which could be the scope of future works.

332 The parametrization of intra-molecular interactions (bonds and angles) of CG molecules
333 was also based on MD simulations. The basic concept is to construct the distribution

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

334 function of each of these quantities from atomistic model simulations. By using again the
335 molecular fragment information obtained via AFP within the atomistic MD trajectories,
336 the distribution functions of bond lengths and bending angles were calculated based on the
337 center of the coarse-grained fragments. Then, a robust and fast approach when dealing
338 with hundreds of bond and angle interaction types generated from the automated coarse-
339 graining procedure employed in this work (AFP) is to derive parameters from distributions
340 directly,^{43,92,93} instead of fitting each bond-stretching and bending angle potential obtained
341 from Boltzmann inversion with a harmonic approximation.⁹⁴ When assuming a harmonic
342 bond potential (Eq. (2)), the resulting distribution is a Gaussian that can be equated with
343 the distribution of the bonds. It follows that the equilibrium bond length, l_H , is simply
344 the average of the distribution and the bond constant, k_S , can be expressed in terms of the
345 standard deviation of that distribution.^{43,92,93} For angles, the same would hold for harmonic
346 potentials (Eq. (3)), except that the angle is bounded between 0° and 180° . This means
347 that the distribution for a purely harmonic potential will not be a Gaussian, but rather a
348 Gaussian that is cut off at 180° . However, a reasonable procedure is to simply take the angle
349 where the distribution is maximal and treating that as if it were the average, equating it to
350 the equilibrium angle, θ_H . Taking the standard deviation to calculate the angle potential
351 strength, k_A , also is reasonable.⁴³ It is important to point out that this procedure is not able
352 to capture multiple maxima and/or minima in bond and angle distributions from atomistic
353 MD simulations.⁴³ Without a further modification, bonded interaction parameters directly
354 derived from MD distributions can be used in DPD simulations by using a shorter time step
355 than that typically used in DPD works (i.e., $\Delta t = \mathcal{O}(0.01)^{29}$). In fact, the exact replication
356 of the MD structures required the strength of bonds to become too large for relatively long
357 time step, resulting in unstable simulations.⁴¹ Therefore, in order to preserve the distance
358 and angular bond characteristics, a dimensionless time step of $\Delta t = 0.001$ was used to
359 integrate the DPD equations of motion.⁸⁸

360 C. DPD simulation parameters

361 To avoid using excessively large or small numbers and to simplify the calculations, DPD
362 systems were usually scaled by arbitrarily chosen base units. As it was already discussed
363 in the previous subsection, the conversion factor $h = 7.65 \text{ \AA}$ was here employed as base

364 unit of length. The mass of one water bead consisting of three water molecules equal to
 365 8.974×10^{-26} kg, was used as the base mass unit. Both MD and DPD simulations were
 366 performed at ambient temperature (298 K), giving $k_b T = 4.11 \times 10^{-21}$ J used as the base
 367 unit for energy, where k_b is the Boltzmann constant. The base time unit τ was estimated by
 368 evaluating the diffusion coefficient. This is computed from both MD and DPD simulations
 369 by using the standard mean-squared displacement (MSD) method through the well-known
 370 Einstein relation.²⁵ By defining the scaling factor $S = D_{W,Exp}/D_{W,DPD} = 7.63 \times 10^{-9}$ m²/s,
 371 where $D_{W,Exp}$ and $D_{W,DPD}$ are respectively the experimental water self-diffusion coefficient
 372 at ambient conditions and the simulated one via DPD, the base unit used to convert the
 373 reduced DPD time into real unit reads as follows:

$$\tau = \frac{h^2}{S} \approx 77 \text{ ps.} \quad (6)$$

Therefore, the real protein diffusion coefficient computed from DPD simulations was simply determined by multiplying the simulated value for the scaling factor, S .⁹⁵ Since no experimental measurement is available in the literature, the protein diffusion D computed via MD and DPD were compared with three correlations proposed for the prediction of protein diffusion coefficients in free solution, based on the molecular weight M (Eq. (7a)⁹⁶), on the radius of gyration R_g (Eq. (7b)⁹⁷), and on both the molecular weight and the radius of gyration of the protein (Eq. (7c)⁹⁸), respectively:

$$D = 8.34 \times 10^{-8} \left(\frac{T}{\eta M^{1/3}} \right), \quad (7a)$$

$$D = 5.78 \times 10^{-8} \left(\frac{T}{\eta R_g} \right), \quad (7b)$$

$$D = 6.85 \times 10^{-8} \left(\frac{T}{\eta \sqrt{M^{1/3} R_g}} \right), \quad (7c)$$

374 where η is the solvent viscosity, i.e., 0.894 and 50 cP at 25 °C for water⁹⁹ and for soybean
 375 oil,¹⁰⁰ respectively.

376 Several DPD simulation configurations were investigated in this work. In order to match
 377 the coarse-grained characteristics from MD simulations, the binary systems were reproduced
 378 using DPD. The MD box was scaled according to the length conversion factor h and one
 379 CG protein molecule was located at its center. According to the binary environment, the
 380 box was then filled with water beads or oil CG molecules to obtain the overall DPD density
 381 $\rho = 5$. The DPD simulations were performed with an equilibration period of 10^5 steps, then

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

382 followed by a production phase of 10^6 steps, saving particle trajectories every 250 steps.
 383 Once DPD parameters have been calibrated as explained in the previous subsection, two
 384 DPD configurations of the interfacial system were carried out in order to study the equilib-
 385 rium properties at increasing protein interface concentration c_i and the protein adsorption
 386 at the oil/water interface. Both initial configurations consisted of a central water phase
 387 segregated by two oil phases, thus forming two planar interfaces in equidistant yz -planes.
 388 The 50/50 oil-to-water bead ratio was kept constant for all DPD simulations and both the
 389 number of water beads and oil CG molecules was adjusted to keep the same overall DPD
 390 density of 5 when the protein molecules were also added in the DPD box. The equilib-
 391 rium simulations were conducted with increasing protein interface concentration c_i , which is
 392 simply calculated by multiplying the number of the protein molecules at each interface for
 393 the protein molecular mass M , divided for the constant interface yz -area expressed in real
 394 units. The protein molecules were initially located at the oil-water interface to make sure
 395 that both interfaces contain the same number at equilibrium in order to perform averages
 396 on both interfaces. For equilibrium DPD simulations, the box was an orthorhombic cell of
 397 reduced size $L_x \times L_y \times L_z$, where $L_y = L_z = 32$ and L_x was properly adjusted up to 52 based
 398 on the protein molecule number to allow both interfaces to be independent. Simulations
 399 were run for 2.5×10^5 equilibration steps and for a production period of 10^6 steps, saving
 400 time frame data for post-processing every 500 steps. Here the interfacial tension, σ_{DPD} ,
 401 was computed by integrating the difference between normal and tangential stress across the
 402 interface separating the segregated components.¹⁰¹ Thus, if the normal to the interface lies
 403 along the x -direction, the interfacial tension is deduced from the local components of the
 404 pressure tensor:

$$\sigma_{\text{DPD}} = \frac{1}{2} \int (p_N^* - p_T^*) dx = \frac{1}{2} \int \left(p_{xx}^* - \frac{1}{2} (p_{yy}^* + p_{zz}^*) \right) dx, \quad (8)$$

405 where p_N^* and p_T^* are the normal and tangential components of the pressure tensor profile
 406 in reduced DPD units. The factor 1/2 before the integral sign is due to the presence of two
 407 symmetric interfaces in the DPD simulation box when using periodic boundary conditions.
 408 Since the oil droplets of a food emulsion have a diameter of the order of microns,³ it is reason-
 409 able to neglect the curvature effect when modeling the interfacial system at the nano-scale,
 410 thus allowing to use the above formula, valid for planar geometry only.¹⁰¹ The conversion of
 411 σ_{DPD} to real units operates as follows: $\sigma_{\text{calc}} = \frac{k_b T}{h^2} \sigma_{\text{DPD}}$. The quantity σ_{calc} can be directly

412 compared with experimentally measured interfacial tension. The free protein adsorption at
413 the oil/water interface was also studied by locating one protein molecule in the center of
414 an orthorhombic DPD box $L_x \times L_y \times L_z$, where $L_y = L_z = 20$ and L_x was ranged from 40
415 to 56 in order to properly increase the mutual initial distance between the protein center
416 and the interface. In addition, the adsorption at the oil/water interface was tested for an
417 LDL-like particle configuration by initially creating a small droplet of 15 oil CG molecules
418 surrounded by one protein molecule. These latter DPD simulations were performed with
419 2×10^5 equilibration steps and a production period of up to 4×10^6 steps, saving simulation
420 time frames every 500 steps to check if the protein adsorption has taken place.

421 Apart from the water cluster algorithm, which was performed in the MATLAB environment,⁸⁸
422 all MD and DPD simulation setup, runs, and post-processing analyzes were conducted within
423 the CULGI software package,¹⁰² together with all other tools and algorithms employed in
424 this work.

425 V. RESULTS AND DISCUSSION

426 The results of the DPD model calibration explained in Section IV B are shown in Figure
427 3, where the distance is expressed in real units, and in Table I. Using the MD RDFs as
428 references, the DPD RDFs were adjusted in order to best match curve heights and shapes
429 by calibrating both α_{EV} and r_c of molecule bead pairs. These two terms define both the
430 magnitude (via Eq. (5)) and the spatial extension of the repulsive force (Eq. (1)). Typically,
431 in standard DPD the cutoff value also represents the base unit of length and, therefore, is
432 often set equal to 1 in dimensionless unit.²⁹ In contrast, here the dimensionless value of r_c
433 resulting from fitting the first peaks of RDF curves shown in Figure 3 was found to be equal
434 to 0.7. Hence, the cutoff, r_c , and the length factor, h , were decoupled in order to assure
435 both the constant DPD number density of 5 and the repulsive force calibration. The results
436 of α_{EV} fitting are summarized in Table I. Although the oil-water α_{EV} turned out to be
437 substantially smaller than all the others in Table I, the overall repulsion between water and
438 oil beads was properly reproduced due to the two contributions in Eq. (5) and a cutoff, r_c ,
439 equal to 1 in this specific case, in which a sophisticated calibration was not needed.

442 The molecular model is tested and the main findings are presented here, paying a particu-
443 lar attention to verify the emulsifying behaviour of Apovitellenin I at the oil/water interface.

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

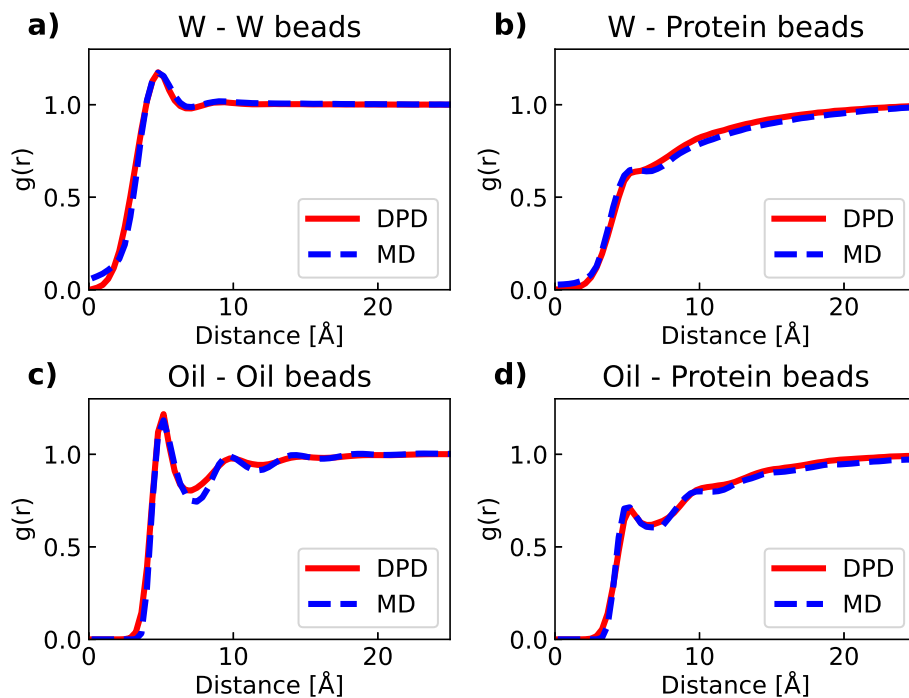


FIG. 3. Results of the DPD parameter calibration of water-water (a), water-protein (b), oil-oil (c), and oil-protein (d) interactions based on matching RDFs of the mapped MD reference model (dashed blue line) with corresponding RDFs extracted from DPD simulations (solid red line).

444 First, preliminary structural and dynamic quantities of the protein are estimated by per-
 445 forming both MD and DPD simulations of one protein molecule in bulk phases. Then, the
 446 DPD simulation results of the ternary system are discussed in terms of both equilibrium
 447 and dynamic aspects.

448 Table II reports End-to-End distance and radius of gyration mean values and standard
 449 deviations of Apovitellenin I in water and oil bulks computed via MD and DPD simulations.
 450 The MD values were averaged over the simulation time, meanwhile 10 independent DPD
 451 simulations with the same initial configuration were carried out from which the reported
 452 values are extrapolated by computing their respective arithmetically averaged frequency
 453 distributions. It is important to recall that Apovitellenin I is modeled here as a homodimer,

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

TABLE I. Values of the global parameter α_{EV} used in Eq. (5) to define the mutual repulsion between all the beads belonging to water, oil, and protein in the DPD model of this work. The cutoff distance, r_c , is equal to 0.7 unless otherwise specified.

α_{EV}	W	Oil beads	Protein beads
W	25 ^a	-	-
Oil beads	8.5 ^b	100	-
Protein beads	40	100	40 ^c

^a Exactly corresponding to a_{ww} .

^b Value obtained by fitting experimental interfacial tension between soybean oil and water,⁹¹ with a cutoff distance, r_c , equal to 1.

^c Arbitrarily chosen equal to the water-protein value.

454 so the two polypeptide chains are labeled as 1 and 2 in Table II where the End-to-End
 455 distance is that between the N-terminal and the C-terminal of each chain, while the protein
 456 radius of gyration refers to the homodimer itself. By looking at mean values reported in
 457 the Table II, it can be noticed that a good accordance between the two molecular technique
 458 is achieved. The largest differences are only related to the chain 1 End-to-End distance
 459 and the radius of gyration of the protein in water environment. The MD radius of gyration
 460 data suggest that the protein is more compact in water than in oil environment, while an
 461 opposite trend is detected via DPD. Another considerable dissimilarity regards the standard
 462 deviation values calculated with the two techniques. Both MD and DPD were able to
 463 identify a smaller error of the respective quantity in oil than in water bulk meaning a less
 464 flexible protein structure in the former environment than in the latter. However, all the DPD
 465 standard deviations are significantly higher than those obtained via MD. This might be due
 466 to two main reasons. First, combining distributions from independent DPD simulations into
 467 a single arithmetically averaged distribution involves that the variance of the averaged one is
 468 always at least as large as the minimum of the variances of input distributions.¹⁰³ Secondly,
 469 the soft potential applied in the DPD force field can provide less steric hindrance compared
 470 to the Lennard-Jones potential used in MD. Moreover, the higher variation in DPD than
 471 MD may be related to the lack of additional bond constraints for intra-protein molecular
 472 interaction^{46,104} in the present DPD framework, thus assuming a completely flexible nature

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

TABLE II. End-to-End distance and radius of gyration mean values and standard deviations of Apovitellenin I in water and oil bulk phases computed via MD and DPD simulations.

			MD	DPD ^a
Apovitellenin I in Water	End-to-End distance [Å]	Chain 1	50.46 ± 2.93	62.06 ± 18.84
		Chain 2	69.84 ± 2.82	65.87 ± 18.37
	Radius of gyration [Å]		24.98 ± 0.50	35.67 ± 5.26
Apovitellenin I in Oil	End-to-End distance [Å]	Chain 1	57.22 ± 0.96	58.38 ± 14.59
		Chain 2	64.49 ± 0.49	63.39 ± 14.20
	Radius of gyration [Å]		27.04 ± 0.13	29.39 ± 2.84

^a The reported values are extrapolated from respective frequency distributions arithmetically averaged over 10 independent simulations.

473 of Apovitellenin I without a specific secondary structure. This latter explanation can be
 474 also given to the opposite trend of the mean value of the protein radius of gyration reported
 475 by means of MD and DPD in the two bulk phases.

477 Table III shows the comparison of diffusion coefficient values, D , of Apovitellenin I in
 478 water and oil bulk calculated by means of three correlations found in the literature (Eq.
 479 (7)^{96–98}) and computed from MD and DPD simulations. MD protein radius of gyration in the
 480 respective solution reported in Table II are used in expressions based on such a property (Eqs.
 481 (7b) and (7c)). Table III also reports the diffusion errors in terms of ranges of variability.
 482 In particular, the accuracy of correlation results was taken from the corresponding previous
 483 works,^{96–98} meanwhile MD and DPD uncertainties were directly estimated from simulations.
 484 As it can be seen, both correlation and simulation results show a difference in the protein
 485 diffusion coefficient of at least one order of magnitude between the water and oil solution.
 486 The larger diffusion coefficient in water than in oil is mostly likely due to the larger oil
 487 viscosity than the water one that can be responsible of the limited mobility of Apovitellenin
 488 I in oil phase. By comparing the results for water environment, MD and DPD give a
 489 remarkable agreement between them although all the correlations indicate a slightly higher
 490 value. On the other hand, the accordance on simulation results is relatively lost when
 491 dealing with oil bulk, but the DPD value is noticeably close to those predicted via empirical

TABLE III. Comparison of diffusion coefficient values of Apovitellenin I in water and oil bulk as predicted by three correlations (Eq. (7)) and as computed from MD and DPD simulations.

$D \times 10^{-12}$ [m ² /s]	Correlation results			MD	DPD ^a
	Eq. (7a) ⁹⁶	Eq. (7b) ⁹⁷	Eq. (7c) ⁹⁸		
Apovitellenin I in Water	82.3 – 127.2	65.7 – 89.0	80.6 – 97.0	22.7 – 24.0	20.9 – 26.1
Apovitellenin I in Oil	1.47 – 2.27	1.10 – 1.45	1.40 – 1.65	0.296 – 0.297	1.97 – 2.92

^a Averaged on 10 independent simulations.

492 correlations. It is also important to highlight here that the diffusion coefficient of proteins
 493 in solution computed by molecular simulation techniques tends to be underestimated when
 494 compared to the true value.¹⁰⁵ That being said, although it is really hard to validate the
 495 data reported in Tables II and III without experimental evidence, it is possible to affirm
 496 that molecular modeling techniques lead to very reasonable results.

498 Let us move now on the discussion of the ternary system made by oil, water and pro-
 499 tein via DPD simulations. In order to study the equilibrium properties of such a system,
 500 the starting configuration of the DPD box consists of two symmetrical interfaces due to
 501 the periodic boundary conditions applied in the three directions. Figure 4 shows the equi-
 502 librated DPD boxes representing the oil-water interface where Apovitellenin I acts as the
 503 surfactant at increasing protein surface concentrations and by highlighting the planar inter-
 504 faces. Figure 5 reports profiles of the number density of oil, water and protein (i) and stress
 505 profiles (difference between normal and tangential pressures, $p_N^* - p_T^*$) (ii) along the nor-
 506 malized x -direction normal to the interfaces at increasing protein interface concentrations
 507 corresponding to those of Figure 4 (a, b, and c). The dashed lines represent the interface
 508 position in the initial DPD configuration. It points out the initial phase separation and
 509 the resulting mutual interpenetration of each component at equilibrium. The profile plots
 510 show the symmetry of the equilibrated ternary system and define the interfacial region that
 511 contains the protein layer and the bulk region that lies between the interfaces. As it can be
 512 seen in Figures 5 a.i), b.i) and c.i), the most interesting result is that the protein molecules
 513 penetrate the water bulk to a much larger extent than the oil bulk, especially at higher
 514 interface protein concentrations. As expected by looking at Table I, this is mostly likely due
 515 to the higher overall repulsion between protein and oil than that between protein and wa-

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

516 ter. By looking at Figures 5 a.ii), b.ii) and c.ii), the mechanical equilibrium of the system is
517 reached in both oil and water phases since the stress profiles fluctuate with small oscillations
518 around zero in the bulk regions. As a consequence, the local contribution to the interfacial
519 tension is located only at the interfaces, with an increase in the stress in the protein region.
520 Therefore, the accuracy of the interfacial tension calculation is achieved. In order to avoid
521 size effects along x -axis and allow both interfaces to be independent, the bulk phases must
522 be large enough to reach the mechanical equilibrium by increasing the L_x dimension as the
523 number of protein molecules increases keeping the interface yz -area constant.

525 Figure 6 reports the trend of the protein layer thickness (a), the protein mean radius of
526 gyration, $\langle R_{g,Protein} \rangle$ (b), and, finally, the interfacial tension (c) as a function of the interface
527 concentration of Apovitellenin I. Three independent DPD runs were carried out and the av-
528 eraged values are shown together with the corresponding standard deviations. Error bars are
529 generally smaller than symbols indicating a high reproducibility of the current DPD model.
530 The most remarkable result is the interfacial tension decrease as the protein interface con-
531 centration increases. This trend clearly evidences the capability of Apovitellenin I to behave
532 as a surfactant. As expected, the minimum value of the interfacial tension is reached at the
533 saturation of the interface, which does no longer allow direct interactions between oil and
534 water. As shown in Figure 6c, the saturation is obtained at the protein interface concentra-
535 tion equal to 3.0-3.5 mg/m², where the interfacial tension ranges between 8 and 10 mN/m.
536 The maximum protein coverage (about 3.0 mg/m²) of the present system is in line with
537 that observed in an experimental work where the oil-in-water emulsion stabilized by flexible
538 proteins (caseins) was studied.¹⁰⁶ Moreover, Dauphas *et al.*⁶⁹ reported that the equilibrium
539 interfacial tension for the oil-water interface with adsorbed LDL film at pH 3 is 9.5 mN/m,
540 which is markedly consistent with our result. It is also important to highlight that, when
541 no protein molecules are added, the interfacial tension between water and oil phase mod-
542 eled as homotriglycerides is accurately reproduced in agree with the experimental value.^{1,91}
543 $\langle R_{g,Protein} \rangle$ (Figure 6b) is computed from the mean value of the protein R_g distribution, fur-
544 ther averaged over 3 DPD simulations. Therefore, $\langle R_{g,Protein} \rangle$ provides information about
545 the conformation and packing of protein molecules at the interface. At low concentration,
546 the protein radius of gyration is higher than its corresponding DPD value in both bulk situ-
547 ations (see Table II). This can indicate that, when very few protein molecules are absorbed
548 at the oil-water interface, they assume a more elongated conformation than that in water or

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

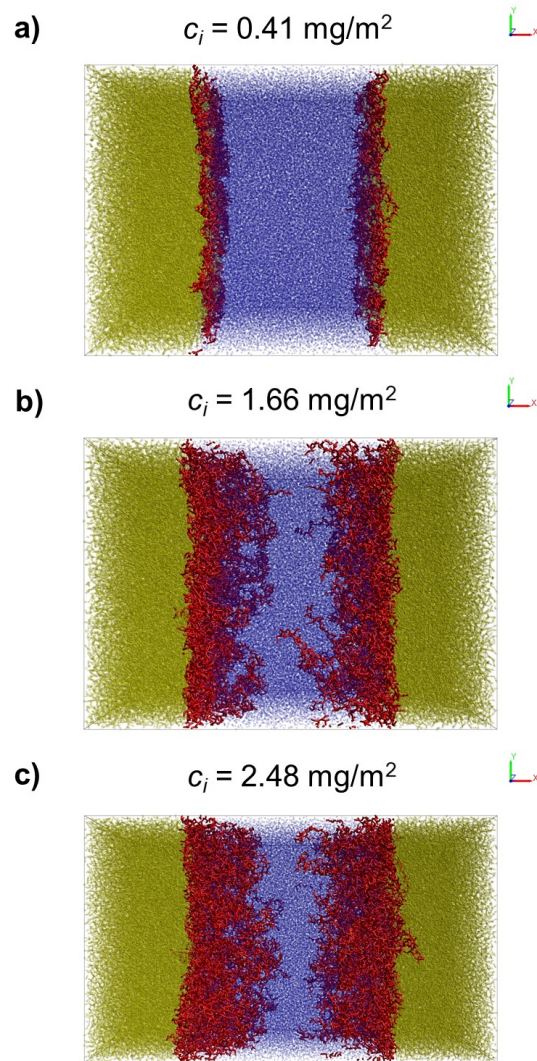


FIG. 4. Snapshots of equilibrated DPD boxes of the interface between oil (yellow) and water (blue) where Apovitellenin I (red) acts as the surfactant at increasing protein interface concentration, c_i (a, b, and c).

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

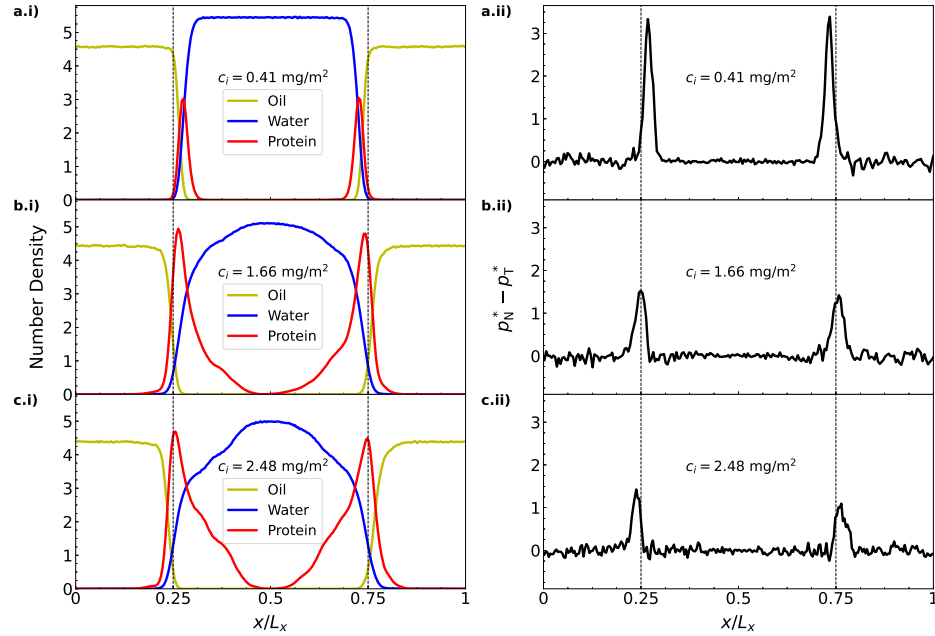


FIG. 5. Profiles of the number density of oil, water and protein (i) and of the difference between normal and tangential pressures, $p_N^* - p_T^*$, (ii) along the normalized x -direction normal to the interfaces at increasing protein interface concentrations (a, b, and c).

549 oil solution. Meanwhile, at increasing protein concentration, the mean radius of gyration of
 550 Apovitellenin I at the interface decreases to a stable value and becomes comparable to that
 551 in free solution. Thus, the packing mode of protein molecules at interface can be considered
 552 similar to that observed in bulk phases, when the protein interface concentration is high.
 553 Regarding the thickness of the protein layer (Figure 6a), it is directly derived from the width
 554 of the protein density profile along the x -direction normal to the interface surface (see Fig-
 555 ures 5i for reference). As expected, the protein layer thickness increase from 2 to 13 nm as
 556 the protein interface concentration increases until the saturation of the interface where the
 557 maximum and stable value for the thickness is reached. Fang and Dalglish¹⁰⁶ reported that
 558 the adsorbed layer of casein molecules at the maximum coverage of the oil-water interface
 559 was about 10 nm thick so that the protein molecules protrude further into the solution, as
 560 also shown in this work (Figures 4 and 5i). Moreover, previous works^{107,108} found that the

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

561 interfacial layer surrounding oil droplets in mayonnaise have an average thickness of around
562 14 nm, which is comprised of surface-active proteins and lecithin-protein granules from egg
563 yolk. Those findings are reasonably in accordance with our results. It is also straightforward
564 to point out here that the emulsifier behaviour of only one LDL apoprotein is tested since it
565 is identified as one of the most surface-active. LDL phospholipids may also have an effect on
566 the interfacial tension of LDL-based emulsion by a further decrease of its saturation value.

568 In order to study the adsorption of Apovitellenin I at the oil-water interface, DPD simu-
569 lations of a box containing two equidistant interface and one free protein molecule initially
570 located in the center of the water phase were carried out. So, the protein diffusion from the
571 aqueous environment towards the oil-water interface is investigated as represented in Figure
572 7, where an illustrative example shows the three main steps of the protein adsorption mech-
573 anism . First, the protein moves to the interface (a), then a portion of the molecule initiates
574 the protein adsorption (b) and, after a certain time, Apovitellenin I is totally adsorbed at the
575 oil-water interface (c). Apparently, there is no specific reason for the protein to be prefer-
576 ably adsorbed at the right rather than at the left interface as the two sides are symmetrical.
577 Moreover, the protein desorption has not been observed meaning that the adsorption process
578 is most likely irreversible as also reported in previous experimental works.^{7,67} To estimate
579 the time required by a protein molecule to be fully absorbed as a function of its distance
580 from the oil-water interface, multiple DPD simulations were performed by increasing the
581 box size in the x -direction normal to the interfaces and the results are summarized in Figure
582 8. Since the oil-to-water bead ratio is kept constant and the protein molecule is placed in
583 the center of the water phase at the beginning of the simulation (see Figure 7 for reference),
584 the abscissa of Figure 8 represents the initial distance between the geometric center of the
585 protein molecule and the oil-water interface. The y-coordinate of Figure 8 expresses the time
586 elapsed from the start of the simulation to the moment in which the protein molecule is to-
587 tally adsorbed at one of the interfaces and it is estimated by visual inspection of simulation
588 time frames. As also done in Figure 6, for each point three independent DPD simulations
589 were carried out from which the mean value and the standard deviation were extracted.
590 Although the error bars are relatively large, a linear trend passing through the origin of the
591 axes can be identified in the range of investigated distances. The slope of $0.978 \text{ ns}/\text{\AA}$ can
592 be considered as an estimation of the required time of a liberated Apovitellenin I molecule
593 to be totally adsorbed at a free interface as a function of their mutual distance.

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0079883

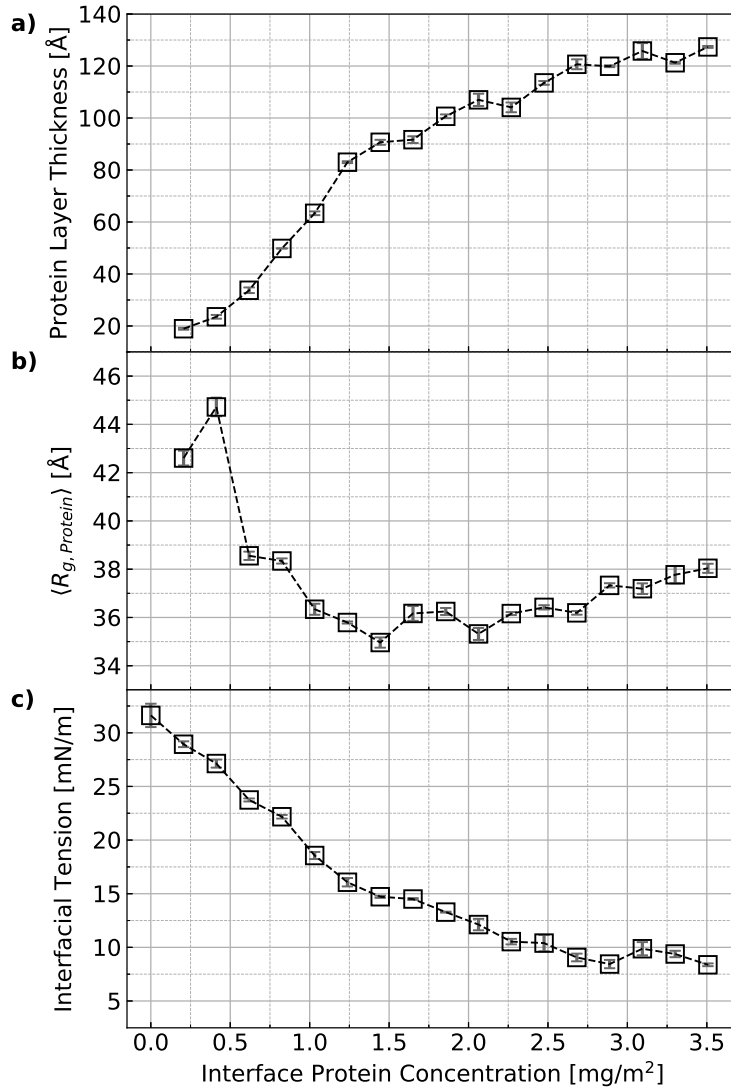


FIG. 6. Protein layer thickness (a), protein mean radius of gyration, $\langle R_{g, Protein} \rangle$ (b), and interfacial tension (c) as a function of the interface concentration of Apovitellenin I. Error bars are estimated from three independent DPD simulations.

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

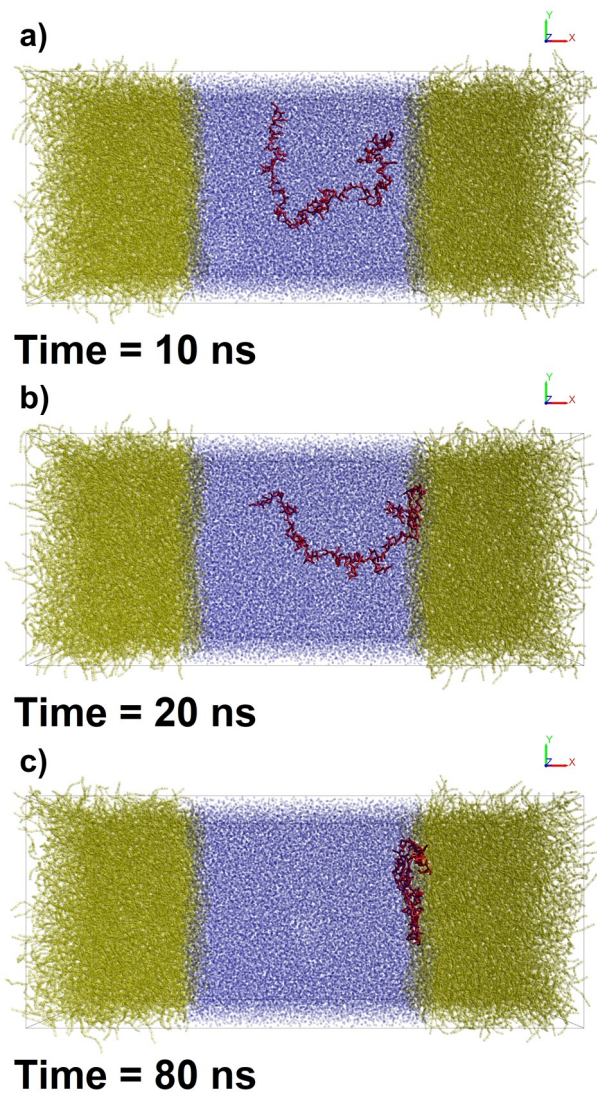


FIG. 7. Snapshots of the DPD simulation showing an illustrative example of the adsorption process of Apovitellenin I (one free molecule in red) at the interface between oil (yellow) and water (blue). The most significant steps of the adsorption mechanism are successively represented in a, b, and c.

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

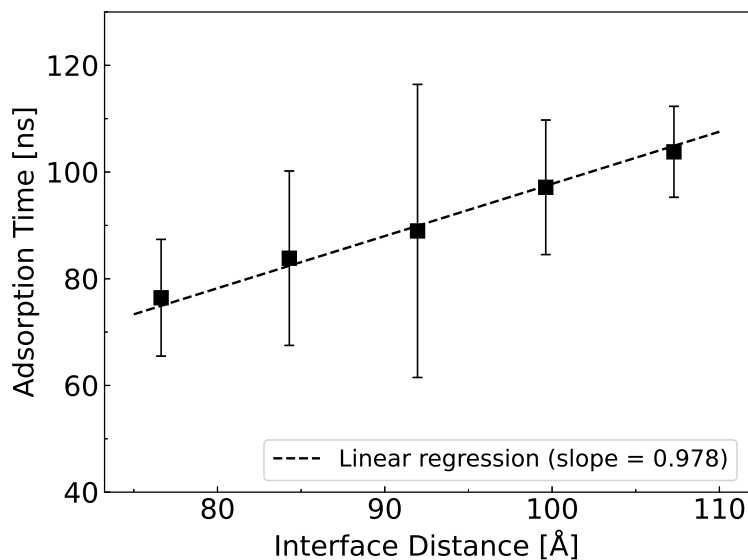


FIG. 8. Trend of the time required by one free molecule of Apovitellenin I to be fully adsorbed at the oil-water interface as a function of the initial distance between the protein geometric center and the oil-water interface. Error bars are estimated from three independent DPD simulations.

595 As already stated, LDL particles act as vectors of surfactant constituents (e.g., Apovitel-
596 lenin I) that could not be soluble in water until they reach the interface. Therefore, a DPD
597 simulation of a LDL-like particle with a lipid core surrounded by one molecule of Apovitel-
598 lenin I was performed and the adsorption mechanism at the oil-water interface was tested.
599 Although it is clear that this structure is far from being a realistic representation of a LDL
600 particle, surprisingly the adsorption process proposed by Anton⁷ is qualitatively reproduced
601 as it can be seen in Figure 9 (Multimedia view). Indeed, first the LDL-like particle diffuses in
602 the water bulk (a) until the protein situated on the particle surface comes into contact with
603 the interface causing the unfolding of the LDL-like particle (b). Thus, the protein molecule
604 initiates the LDL-like particle disruption by its anchorage at the oil-water interface. Then,
605 the neutral lipids are released from the particle core and merge with the oil phase, while
606 the protein molecule adsorbs at the interface (c). Since the system dimensions of Figure
607 9 (Multimedia view) are the same of those represented in Figure 7, a general comparison
608 can be made between two configurations, namely the liberated protein and the LDL-like

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

609 particle. In particular, the adsorption time of the LDL-like particle is significantly higher
610 than that of the free protein. This can be intended as a greater stability of Apovitellenin I
611 when surrounding the LDL-like particle rather than as a free molecule, also confirming that
612 the liberated protein is supposed to be almost insoluble in water. Finally, it is important to
613 remark that the representation of the LDL-like particle here presented must be considered
614 qualitative, since both LDL size and its specific composition, namely including also the lipid
615 distribution of the LDL core and all surfactant components situated on the LDL surface
616 (e.g., phospholipids and other apoproteins), were not considered in the analysis.

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

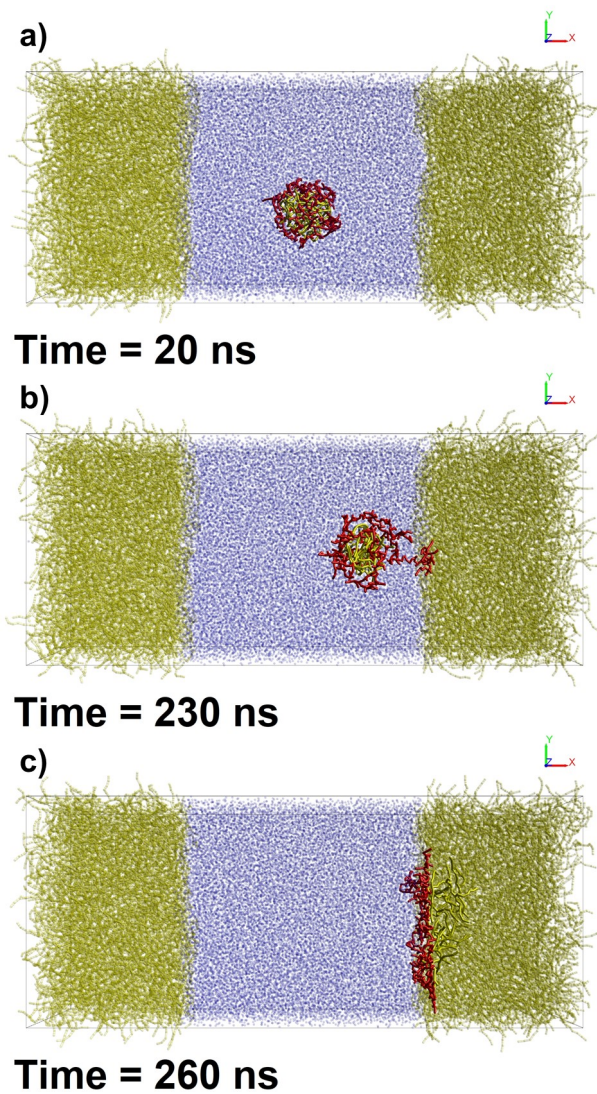


FIG. 9. Snapshots of the DPD simulation showing the adsorption process of a LDL-like particle with a lipid core (bright yellow) surrounded by one molecule of Apovitellenin I (red) at the interface between oil (yellow) and water (blue). The most significant steps of the adsorption mechanism are successively represented in a, b, and c (Multimedia view).

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

618 VI. CONCLUSIONS

619 Although egg yolk is widely used as an emulsifier in many food emulsion preparations,
620 little experimental research on emulsifying properties of its individual components has been
621 carried out since their extraction and isolation from the egg yolk complex matrix turned out
622 to be difficult. Hence, this work focuses on the molecular model of an oil/water interface
623 stabilized by one of the most surface-active protein of egg yolk LDLs, called Apovitellenin
624 I. In order to take into account the system size, composition and the equilibration time
625 needed by macro-molecules to re-arrange at interfaces, the molecular modeling technique
626 here proposed is the Dissipative Particle Dynamics approach. Once the chemical species
627 were determined, especially the biomolecule that should act as a surfactant at the oil/water
628 interface, an automated coarse-graining procedure was carried out on the molecules involved
629 in the ternary system. In DPD systems the intended physical properties are determined by
630 means of a parameter calibration, which was here based on coupling DPD with all-atom
631 Molecular Dynamics simulations of a single protein molecule in two different solvents, water
632 and oil. Thus, both inter- and intra-molecular interactions employed in the DPD system
633 are solely determined by matching the structural data from the atomistic simulations. The
634 model was designed to test the most relevant physical properties of the protein studied,
635 especially its emulsifier behavior. The results of MD and DPD simulations are compared in
636 terms of protein structural and dynamics properties (radius of gyration, end-to-end distance,
637 and diffusion coefficient), showing a good agreement between the two molecular techniques.
638 Then, the oil-water interface system was simulated via the DPD technique. In particular, the
639 present molecular modeling approach was able to properly describe the protein surfactant
640 behavior by interfacial tension decrease at increasing protein surface concentration. The
641 protein density profile, layer thickness, and adsorption time at the oil-water interface were
642 also investigated, giving reasonable results in line with experimental evidence of similar pro-
643 tein systems. In addition, the adsorption mechanism of an LDL-like particle is qualitatively
644 reproduced. The modeling method here presented shows how computer molecular simula-
645 tions can greatly help in the comprehension of food emulsion behavior and, in general, offer
646 the advantage of estimating properties that are difficult to measure experimentally.

647 These results are encouraging and could be a starting point to explore the role of other
648 surfactant molecules from egg yolk with an analogous molecular modeling method. More-

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

649 over, the main findings of this work together with non-equilibrium studies at the meso-scale
650 will pave the way for a better understanding of the breakage and coalescence events of the
651 oil droplets occurring in the food emulsion preparation. This information can be eventu-
652 ally transferred to a computational fluid dynamics study coupled to a population balance
653 model thus achieving a complete, general, and multi-scale digital twin of the food emulsion
654 production process.

655 SUPPLEMENTARY MATERIAL

656 See Supplementary Material for a further description of MD and DPD techniques used
657 in this work.

658 ACKNOWLEDGMENTS

659 This work was carried out in the context of the VIMMP project (www.vimmp.eu), where
660 the entire workflow will contribute to populate a marketplace for generic multiscale and
661 multiphysics simulations. The VIMMP project has received funding from the European
662 Union's Horizon 2020 Research Innovation Programme under Grant Agreement n. 760907.
663 We thank Dr. Piotr Pieczywek (Institute of Agrophysics, Polish Academy of Sciences,
664 Doświadczalna 4, 20-270, Lublin, Poland) for sharing with us the MATLAB code for the
665 water cluster algorithm employed in this work. We greatly appreciate the suggestion of a
666 reviewer to include Figure 1.

667 CONFLICT OF INTEREST

668 The authors have no conflicts to disclose.

669 DATA AVAILABILITY

670 The data that support the findings of this study are openly available in Zenodo at <http://doi.org/10.5281/zenodo.5703247>, reference number 109.
671

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

672 **REFERENCES**

- 673 ¹D. McClements, *Food Emulsions: Principles, Practice, and Techniques* (CRC Press, Boca
674 Raton, FL, 2005).
- 675 ²S. Friberg, J. Sjoblom, and K. Larsson, *Food Emulsions* (CRC Press, Boca Raton, FL,
676 2003).
- 677 ³A. Dubbelboer, J. Janssen, H. Hoogland, E. Zondervan, and J. Meuldijk, "Pilot-scale
678 production process for high internal phase emulsions: Experimentation and modeling,"
679 *Chemical Engineering Science* **148**, 32–43 (2016).
- 680 ⁴A. Dubbelboer, *Towards optimization of emulsified consumer products : modeling and
681 optimization of sensory and physicochemical aspects*, PhD dissertation, Technische Uni-
682 versiteit Eindhoven, Department of Chemical Engineering and Chemistry (2016).
- 683 ⁵S. Maindarkar, A. Dubbelboer, J. Meuldijk, H. Hoogland, and M. Henson, "Prediction
684 of emulsion drop size distributions in colloid mills," *Chemical Engineering Science* **118**,
685 114–125 (2014).
- 686 ⁶P. Walstra, "Principles of emulsion formation," *Chemical Engineering Science* **48**, 333–349
687 (1993).
- 688 ⁷M. Anton, "Egg yolk: structures, functionalities and processes," *Journal of the Science
689 of Food and Agriculture* **93**, 2871–2880 (2013).
- 690 ⁸D. G. Dalgleish, "Food emulsions," in *Emulsions and emulsion stability*, edited by
691 J. Sjoblom (Marcel Dekker Inc., 1996) p. 287–325.
- 692 ⁹R. W. Burley, "Isolation and properties of a low-molecular-weight protein (apovitellenin
693 I) from the high-lipid lipoprotein of emu egg yolk," *Biochemistry* **12**, 1464–1470 (1973).
- 694 ¹⁰R. W. Burley, "Studies on the Apoproteins of the Major Lipoprotein of the Yolk of Hen's
695 Eggs I. Isolation and Properties of the Low-molecular-weight Apoproteins." *Australian
696 Journal of Biological Sciences* **28**, 121–132 (1975).
- 697 ¹¹W. Norde, *Colloids and Interfaces in Life Sciences and Bionanotechnology* (CRC Press,
698 Boca Raton, FL, 2011).
- 699 ¹²C.-A. Palma, M. Cecchini, and P. Samorì, "Predicting self-assembly: from empirism to
700 determinism," *Chemical Society Reviews* **41**, 3713–3730 (2012).
- 701 ¹³A. J. Stone, *The Theory of Intermolecular Forces* (Oxford University Press, Oxford, U.K.,
702 2013).

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0079883

- 703 ¹⁴A. D. Lavino, M. Ferrari, A. A. Barresi, and D. Marchisio, “Effect of different good sol-
704 vents in flash nano-precipitation via multi-scale population balance modeling-CFD cou-
705 pling approach,” *Chemical Engineering Science* **245**, 116833 (2021).
- 706 ¹⁵S. R. Euston, “Computer simulation of proteins: adsorption, gelation and self-
707 association,” *Current Opinion in Colloid & Interface Science* **9**, 321–327 (2004).
- 708 ¹⁶L. A. Pugaloni, E. Dickinson, R. Ettelaie, A. R. Mackie, and P. J. Wilde, “Competitive
709 adsorption of proteins and low-molecular-weight surfactants: computer simulation and
710 microscopic imaging,” *Advances in Colloid and Interface Science* **107**, 27–49 (2004).
- 711 ¹⁷D. Zahn, “On the role of the solvent in biosystems: atomistic insights from computer
712 simulations,” *Frontiers in Bioscience-Landmark* **14**, 3586–3593 (2009).
- 713 ¹⁸E. Dickinson, “Structure and rheology of colloidal particle gels: Insight from computer
714 simulation,” *Advances in Colloid and Interface Science* **199-200**, 114–127 (2013).
- 715 ¹⁹S. Euston, “14 - Modelling and computer simulation of food structures,” in *Food Mi-
716 crostructures*, Woodhead Publishing Series in Food Science, Technology and Nutrition,
717 edited by V. Morris and K. Groves (Woodhead Publishing, 2013) pp. 336–385.
- 718 ²⁰A. Jusufi, “Molecular simulations of self-assembly processes of amphiphiles in dilute so-
719 lutions: the challenge for quantitative modelling,” *Molecular Physics* **111**, 3182–3192
720 (2013).
- 721 ²¹J.-W. Handgraaf and F. Zerbetto, “Molecular dynamics study of onset of water gelation
722 around the collagen triple helix,” *Proteins: Structure, Function, and Bioinformatics* **64**,
723 711–718 (2006).
- 724 ²²A. D. Lavino, N. Di Pasquale, P. Carbone, and D. L. Marchisio, “A novel multiscale model
725 for the simulation of polymer flash nano-precipitation,” *Chemical Engineering Science*
726 **171**, 485–494 (2017).
- 727 ²³A. D. Lavino, P. Carbone, and D. Marchisio, “MARTINI coarse-grained model for poly- ϵ -
728 caprolactone in acetone-water mixtures,” *The Canadian Journal of Chemical Engineering*
729 **98**, 1868–1879 (2020).
- 730 ²⁴A. D. Lavino, L. Banetta, P. Carbone, and D. L. Marchisio, “Extended Charge-On-
731 Particle Optimized Potentials for Liquid Simulation Acetone Model: The Case of Ace-
732 tone–Water Mixtures,” *The Journal of Physical Chemistry B* **122**, 5234–5241 (2018).
- 733 ²⁵D. Frenkel and B. Smit, *Understanding Molecular Simulation*, 2nd ed. (Academic Press,
734 Inc., USA, 2001).

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

- 735 ²⁶C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models*, 2nd ed.
736 (Wiley, New York, 2004).
- 737 ²⁷P. J. Hoogerbrugge and J. M. V. A. Koelman, “Simulating Microscopic Hydrodynamic
738 Phenomena with Dissipative Particle Dynamics,” *Europhysics Letters (EPL)* **19**, 155–160
739 (1992).
- 740 ²⁸P. Español and P. Warren, “Statistical Mechanics of Dissipative Particle Dynamics,”
741 *Europhysics Letters (EPL)* **30**, 191–196 (1995).
- 742 ²⁹R. D. Groot and P. B. Warren, “Dissipative particle dynamics: Bridging the gap between
743 atomistic and mesoscopic simulation,” *The Journal of Chemical Physics* **107**, 4423–4435
744 (1997).
- 745 ³⁰N. Lauriello, J. Kondracki, A. Buffo, G. Boccardo, M. Bouaifi, M. Lisal, and D. Marchisio,
746 “Simulation of high Schmidt number fluids with dissipative particle dynamics: Parameter
747 identification and robust viscosity evaluation,” *Physics of Fluids* **33**, 073106 (2021).
- 748 ³¹E. Dickinson and S. R. Euston, “Monte Carlo simulation of colloidal systems,” *Advances
749 in Colloid and Interface Science* **42**, 89–148 (1992).
- 750 ³²R. E. Anderson, V. S. Pande, and C. J. Radke, “Dynamic lattice Monte Carlo simulation
751 of a model protein at an oil/water interface,” *The Journal of Chemical Physics* **112**,
752 9167–9185 (2000).
- 753 ³³G. Dalkas and S. R. Euston, “Molecular simulation of protein adsorption and conformation
754 at gas-liquid, liquid-liquid and solid-liquid interfaces,” *Current Opinion in Colloid &
755 Interface Science* **41**, 1–10 (2019).
- 756 ³⁴D. Zare, K. M. McGrath, and J. R. Allison, “Deciphering β -Lactoglobulin Interactions
757 at an Oil-Water Interface: A Molecular Dynamics Study,” *Biomacromolecules* **16**, 1855–
758 1861 (2015).
- 759 ³⁵D. Zare, J. R. Allison, and K. M. McGrath, “Molecular Dynamics Simulation of
760 β -Lactoglobulin at Different Oil/Water Interfaces,” *Biomacromolecules* **17**, 1572–1581
761 (2016).
- 762 ³⁶D. L. Cheung, “Adsorption and conformations of lysozyme and α -lactalbumin at a water-
763 octane interface,” *The Journal of Chemical Physics* **147**, 195101 (2017).
- 764 ³⁷D. L. Cheung, “Conformations of Myoglobin-Derived Peptides at the Air-Water Inter-
765 face,” *Langmuir* **32**, 4405–4414 (2016).

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

- 766 ³⁸S. R. Euston, “Molecular Dynamics Simulation of Protein Adsorption at Fluid Interfaces:
767 A Comparison of All-Atom and Coarse-Grained Models,” *Biomacromolecules* **11**, 2781–
768 2787 (2010).
- 769 ³⁹S. R. Euston, P. Hughes, M. A. Naser, and R. E. Westacott, “Comparison of the Adsorbed
770 Conformation of Barley Lipid Transfer Protein at the Decane-Water and Vacuum-Water
771 Interface: A Molecular Dynamics Simulation,” *Biomacromolecules* **9**, 1443–1453 (2008).
- 772 ⁴⁰S. R. Euston, P. Hughes, M. A. Naser, and R. E. Westacott, “Molecular Dynamics
773 Simulation of the Cooperative Adsorption of Barley Lipid Transfer Protein and cis-
774 Isocohumulone at the Vacuum-Water Interface,” *Biomacromolecules* **9**, 3024–3032 (2008).
- 775 ⁴¹F. Sepehr and S. J. Paddison, “Dissipative Particle Dynamics interaction parameters from
776 *ab initio* calculations,” *Chemical Physics Letters* **645**, 20–26 (2016).
- 777 ⁴²H. Lei, B. Caswell, and G. E. Karniadakis, “Direct construction of mesoscopic models
778 from microscopic simulations,” *Physical Review E* **81**, 026704 (2010).
- 779 ⁴³W. Tschöp, K. Kremer, J. Batoulis, T. Bürger, and O. Hahn, “Simulation of poly-
780 mer melts. I. Coarse-graining procedure for polycarbonates,” *Acta Polymerica* **49**, 61–74
781 (1998).
- 782 ⁴⁴A. Vishnyakov and A. V. Neimark, “Self-Assembly in Nafion Membranes upon Hydration:
783 Water Mobility and Adsorption Isotherms,” *The Journal of Physical Chemistry B* **118**,
784 11353–11364 (2014).
- 785 ⁴⁵K. Patterson, M. Lisal, and C. M. Colina, “Adsorption behavior of model proteins on
786 surfaces,” *Fluid Phase Equilibria* **302**, 48–54 (2011).
- 787 ⁴⁶A. Vishnyakov, D. S. Talaga, and A. V. Neimark, “DPD Simulation of Protein Confor-
788 mations: From α -Helices to β -Structures,” *The Journal of Physical Chemistry Letters* **3**,
789 3081–3087 (2012).
- 790 ⁴⁷K. Okuwaki, H. Doi, K. Fukuzawa, and Y. Mochizuki, “Folding simulation of small pro-
791 teins by dissipative particle dynamics (DPD) with non-empirical interaction parameters
792 based on fragment molecular orbital calculations,” *Applied Physics Express* **13**, 017002
793 (2019).
- 794 ⁴⁸M. Ndao, F. Goujon, A. Ghoufi, and P. Malfreyt, “Coarse-grained modeling of the
795 oil–water–surfactant interface through the local definition of the pressure tensor and in-
796 terfacial tension,” *Theoretical Chemistry Accounts* **136**, 21 (2017).

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

- 797 ⁴⁹A. Ghoufi, P. Malfreyt, and D. J. Tildesley, "Computer modelling of the surface tension
798 of the gas-liquid and liquid-liquid interface," *Chemical Society Reviews* **45**, 1387-1409
799 (2016).
- 800 ⁵⁰A. Khedr and A. Striolo, "DPD Parameters Estimation for Simultaneously Simulating
801 Water-Oil Interfaces and Aqueous Nonionic Surfactants," *Journal of Chemical Theory
802 and Computation* **14**, 6460-6471 (2018).
- 803 ⁵¹A. Maiti and S. McGrother, "Bead-bead interaction parameters in dissipative particle
804 dynamics: Relation to bead-size, solubility parameter, and surface tension," *The Journal
805 of Chemical Physics* **120**, 1594-1601 (2004).
- 806 ⁵²S.-l. Lin, M.-y. Xu, and Z.-r. Yang, "Dissipative particle dynamics study on the mesostruc-
807 tures of n-octadecane/water emulsion with alternating styrene-maleic acid copolymers as
808 emulsifier," *Soft Matter* **8**, 375-384 (2012).
- 809 ⁵³F. Alvarez, E. A. Flores, L. V. Castro, J. G. Hernández, A. López, and F. Vázquez, "Dis-
810 sipative Particle Dynamics (DPD) Study of Crude Oil-Water Emulsions in the Presence
811 of a Functionalized Co-polymer," *Energy & Fuels* **25**, 562-567 (2011).
- 812 ⁵⁴L. Rekvig, B. Hafskjold, and B. Smit, "Molecular Simulations of Surface Forces and Film
813 Rupture in Oil/Water/Surfactant Systems," *Langmuir* **20**, 11583-11593 (2004).
- 814 ⁵⁵E. G. Perkins, "Chapter 2 - composition of soybeans and soybean products," in *Practical
815 Handbook of Soybean Processing and Utilization*, edited by D. R. Erickson (AOCS Press,
816 1995) pp. 9-28.
- 817 ⁵⁶J. Leaver and D. G. Dalgleish, "Variations in the binding of β -casein to oil-water interfaces
818 detected by trypsin-catalysed hydrolysis," *Journal of Colloid and Interface Science* **149**,
819 49-55 (1992).
- 820 ⁵⁷R. Xiong, G. Xie, and A. Edmondson, "Modelling the pH of mayonnaise by the ratio of
821 egg to vinegar," *Food Control* **11**, 49-56 (2000).
- 822 ⁵⁸M. Anton and G. Gandemer, "Composition, Solubility and Emulsifying Properties of
823 Granules and Plasma of Egg Yolk," *Journal of Food Science* **62**, 484-487 (1997).
- 824 ⁵⁹J. N. Dyer-Hurdon and I. A. Nnanna, "Cholesterol Content and Functionality of Plasma
825 and Granules Fractionated from Egg Yolk," *Journal of Food Science* **58**, 1277-1281 (1993).
- 826 ⁶⁰M. Le Denmat, M. Anton, and V. Beaumal, "Characterisation of emulsion properties
827 and of interface composition in O/W emulsions prepared with hen egg yolk, plasma and
828 granules," *Food Hydrocolloids* **14**, 539-549 (2000).

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

- 829 ⁶¹V. Martinet, V. Beaumal, M. Dalgarrondo, and M. Anton, “Emulsifying properties and
830 adsorption behavior of egg yolk lipoproteins (LDL and HDL) in oil-in-water emulsions,”
831 in *Food emulsions and dispersions*, edited by M. Anton (Research Signpost, Trivandrum,
832 2002) p. 103–116.
- 833 ⁶²W. Cook and W. Martin, “Egg lipoproteins,” in *Structural and Functional Aspects of*
834 *Lipoproteins in Living Systems*, edited by E. Tria and A. Scanu (Academic Press, London,
835 1969) p. 579–615.
- 836 ⁶³R. W. Burley and W. H. Cook, “Isolation and composition of avian egg yolk granules and
837 their constituent α - and β -lipovitellins,” *Canadian Journal of Biochemistry and Physiol-*
838 *ogy* **39**, 1295–1307 (1961).
- 839 ⁶⁴P. Jolivet, C. Boulard, V. Beaumal, T. Chardot, and M. Anton, “Protein components
840 of low-density lipoproteins purified from hen egg yolk,” *Journal of Agricultural and Food*
841 *Chemistry* **54**, 4424–4429 (2006).
- 842 ⁶⁵P. Jolivet, C. Boulard, T. Chardot, and M. Anton, “New Insights into the Structure of
843 Apolipoprotein B from Low-Density Lipoproteins and Identification of a Novel YGP-like
844 Protein in Hen Egg Yolk,” *Journal of Agricultural and Food Chemistry* **56**, 5871–5879
845 (2008).
- 846 ⁶⁶R. J. Evans, D. H. Bauer, S. L. Bandemer, S. B. Vaghefi, and C. J. Flegal, “Structure of
847 egg yolk very low density lipoprotein. polydispersity of the very low density lipoprotein
848 and the role of lipovitellenin in the structure,” *Archives of Biochemistry and Biophysics*
849 **154**, 493–500 (1973).
- 850 ⁶⁷M. Anton, V. Martinet, M. Dalgarrondo, V. Beaumal, E. David-Briand, and H. Rabes-
851 ona, “Chemical and structural characterisation of low-density lipoproteins purified from
852 hen egg yolk,” *Food Chemistry* **83**, 175–183 (2003).
- 853 ⁶⁸V. Martinet, P. Saulnier, V. Beaumal, J.-L. Courthaudon, and M. Anton, “Surface
854 properties of hen egg yolk low-density lipoproteins spread at the air–water interface,”
855 *Colloids and Surfaces B: Biointerfaces* **31**, 185–194 (2003).
- 856 ⁶⁹S. Dauphas, V. Beaumal, A. Riaublanc, and M. Anton, “Hen egg yolk low-density lipopro-
857 teins film spreading at the air-water and oil-water interfaces,” *Journal of Agricultural and*
858 *Food Chemistry* **54**, 3733–3737 (2006).
- 859 ⁷⁰S. Dauphas, V. Beaumal, P. Gunning, A. Mackie, P. Wilde, V. Vié, A. Riaublanc, and
860 M. Anton, “Structures and rheological properties of hen egg yolk low density lipopro-

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

- 861 tein layers spread at the air–water interface at pH 3 and 7,” *Colloids and Surfaces B:*
862 *Biointerfaces* **57**, 124–133 (2007).
- 863 ⁷¹S. Dauphas, V. Beaumal, P. Gunning, A. Mackie, P. Wilde, V. Vié, A. Riaublanc, and
864 M. Anton, “Structure modification in hen egg yolk low density lipoproteins layers between
865 30 and 45mN/m observed by AFM,” *Colloids and Surfaces B: Biointerfaces* **54**, 241–248
866 (2007).
- 867 ⁷²The UniProt Consortium, “Uniprot: the universal protein knowledgebase in 2021,” *Nu-*
868 *cleic Acids Research* **49**, D480–D489 (2020).
- 869 ⁷³M. Allen and D. Tildesley, *Computer Simulation of Liquids*, 2nd ed. (Oxford University
870 Press, Oxford, U.K., 2017).
- 871 ⁷⁴E. Moeendarbary, T. Y. Ng, and M. Zangeneh, “Dissipative particle dynamics: Intro-
872 duction, methodology and complex fluid applications — A review,” *International Journal*
873 *of Applied Mechanics* **01**, 737–763 (2009).
- 874 ⁷⁵D. R. Lide, ed., *CRC Handbook of Chemistry and Physics*, 85th ed. (CRC press, Boca
875 Raton, FL, 2005).
- 876 ⁷⁶W. L. Jorgensen and J. Tirado-Rives, “The OPLS [optimized potentials for liquid simula-
877 tions] potential functions for proteins, energy minimizations for crystals of cyclic peptides
878 and crambin,” *Journal of the American Chemical Society* **110**, 1657–1666 (1988).
- 879 ⁷⁷W. L. Jorgensen, D. S. Maxwell, and J. Tirado-Rives, “Development and Testing of
880 the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic
881 Liquids,” *Journal of the American Chemical Society* **118**, 11225–11236 (1996).
- 882 ⁷⁸W. L. Jorgensen, J. Chandrasekhar, J. D. Madura, R. W. Impey, and M. L. Klein,
883 “Comparison of simple potential functions for simulating liquid water,” *The Journal of*
884 *Chemical Physics* **79**, 926–935 (1983).
- 885 ⁷⁹U. Essmann, L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen, “A
886 smooth particle mesh Ewald method,” *The Journal of Chemical Physics* **103**, 8577–8593
887 (1995).
- 888 ⁸⁰H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola, and J. R. Haak,
889 “Molecular dynamics with coupling to an external bath,” *The Journal of Chemical Physics*
890 **81**, 3684–3690 (1984).
- 891 ⁸¹J. G. E. M. Fraaije, J. van Male, P. Becherer, and R. Serral Gracià, “Coarse-grained mod-
892 els for automated fragmentation and parametrization of molecular databases,” *Journal of*

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0079883

- 893 Chemical Information and Modeling **56**, 2361–2377 (2016).
- 894 ⁸²M. Diedenhofen and A. Klamt, “COSMO-RS as a tool for property prediction of IL
895 mixtures—A review,” *Fluid Phase Equilibria* **294**, 31–38 (2010).
- 896 ⁸³A. Klamt, *COSMO-RS: From Quantum Chemistry to Fluid Phase Thermodynamics and*
897 *Drug Design* (Elsevier, Amsterdam, 2005).
- 898 ⁸⁴A. Jakalian, B. L. Bush, D. B. Jack, and C. I. Bayly, “Fast, efficient generation of
899 high-quality atomic charges. AM1-BCC model: I. Method,” *Journal of Computational*
900 *Chemistry* **21**, 132–146 (2000).
- 901 ⁸⁵A. Jakalian, D. B. Jack, and C. I. Bayly, “Fast, efficient generation of high-quality
902 atomic charges. AM1-BCC model: II. Parameterization and validation,” *Journal of Com-*
903 *putational Chemistry* **23**, 1623–1641 (2002).
- 904 ⁸⁶P. C. Petris, P. Becherer, and J. G. E. M. Fraaije, “Alkane/Water Partition Coefficient
905 Calculation Based on the Modified AM1 Method and Internal Hydrogen Bonding Sam-
906 pling Using COSMO-RS,” *Journal of Chemical Information and Modeling* **61**, 3453–3462
907 (2021).
- 908 ⁸⁷A. K. Rappe and W. A. Goddard, “Charge equilibration for molecular dynamics simula-
909 tions,” *The Journal of Physical Chemistry* **95**, 3358–3363 (1991).
- 910 ⁸⁸P. M. Pieczywek, W. Płaziński, and A. Zdunek, “Dissipative particle dynamics model
911 of homogalacturonan based on molecular dynamics simulations,” *Scientific Reports* **10**,
912 14691 (2020).
- 913 ⁸⁹K. R. Hadley and C. McCabe, “On the Investigation of Coarse-Grained Models for Water:
914 Balancing Computational Efficiency and the Retention of Structural Properties,” *The*
915 *Journal of Physical Chemistry B* **114**, 4590–4599 (2010).
- 916 ⁹⁰E. Zohravi, E. Shirani, and A. Pischevar, “Influence of the conservative force on transport
917 coefficients in the DPD method,” *Molecular Simulation* **44**, 254–261 (2018).
- 918 ⁹¹A. G. Gaonkar, “Effects of salt, temperature, and surfactants on the interfacial tension
919 behavior of a vegetable oil/water system,” *Journal of Colloid and Interface Science* **149**,
920 256–260 (1992).
- 921 ⁹²D. Reith, H. Meyer, and F. Müller-Plathe, “Mapping Atomistic to Coarse-Grained Poly-
922 mer Models Using Automatic Simplex Optimization To Fit Structural Properties,” *Macro-*
923 *molecules* **34**, 2335–2345 (2001).

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883

- 924 ⁹³Y. Li, B. C. Abberton, M. Kröger, and W. K. Liu, “Challenges in Multiscale Modeling
925 of Polymer Dynamics,” *Polymers* **5**, 751–832 (2013).
- 926 ⁹⁴V. Agrawal, G. Arya, and J. Oswald, “Simultaneous Iterative Boltzmann Inversion for
927 Coarse-Graining of Polyurea,” *Macromolecules* **47**, 3378–3389 (2014).
- 928 ⁹⁵J. G. E. M. Fraaije, J. van Male, P. Becherer, and R. Serral Gracià, “Calculation
929 of Diffusion Coefficients through Coarse-Grained Simulations Using the Automated-
930 Fragmentation-Parametrization Method and the Recovery of Wilke–Chang Statistical
931 Correlation,” *Journal of Chemical Theory and Computation* **14**, 479–485 (2018).
- 932 ⁹⁶M. E. Young, P. A. Carroad, and R. L. Bell, “Estimation of diffusion coefficients of
933 proteins,” *Biotechnology and Bioengineering* **22**, 947–955 (1980).
- 934 ⁹⁷M. T. Tyn and T. W. Gusek, “Prediction of diffusion coefficients of proteins,” *Biotech-
935 nology and Bioengineering* **35**, 327–338 (1990).
- 936 ⁹⁸L. He and B. Niemeyer, “A Novel Correlation for Protein Diffusion Coefficients Based on
937 Molecular Weight and Radius of Gyration,” *Biotechnology Progress* **19**, 544–548 (2003).
- 938 ⁹⁹J. Kestin, M. Sokolov, and W. A. Wakeham, “Viscosity of liquid water in the range -8°C
939 to 150°C,” *Journal of Physical and Chemical Reference Data* **7**, 941–948 (1978).
- 940 ¹⁰⁰F. C. Magne and E. L. Skau, “Viscosities and Densities of Solvent-Vegetable Oil Mix-
941 tures,” *Industrial & Engineering Chemistry* **37**, 1097–1101 (1945).
- 942 ¹⁰¹J. H. Irving and J. G. Kirkwood, “The statistical mechanical theory of transport processes.
943 iv. the equations of hydrodynamics,” *The Journal of Chemical Physics* **18**, 817–829 (1950).
- 944 ¹⁰²Culgi B.V., The Netherlands, “The Chemistry Unified Language Interface (CULGI),”
945 www.culgi.com (2020), version 13.0.0.
- 946 ¹⁰³T. P. Hill and J. Miller, “How to combine independent data sets for the same quantity,”
947 *Chaos: An Interdisciplinary Journal of Nonlinear Science* **21**, 033102 (2011).
- 948 ¹⁰⁴E. K. Peter, K. Lykov, and I. V. Pivkin, “A polarizable coarse-grained protein model
949 for dissipative particle dynamics,” *Physical Chemistry Chemical Physics* **17**, 24452–24461
950 (2015).
- 951 ¹⁰⁵J. Wang and T. Hou, “Application of molecular dynamics simulations in molecular prop-
952 erty prediction II: Diffusion coefficient,” *Journal of Computational Chemistry* **32**, 3505–
953 3519 (2011).
- 954 ¹⁰⁶Y. Fang and D. G. Dalgleish, “Dimensions of the Adsorbed Layers in Oil-in-Water Emul-
955 sions Stabilized by Caseins,” *Journal of Colloid and Interface Science* **156**, 329–334 (1993).

This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0079883

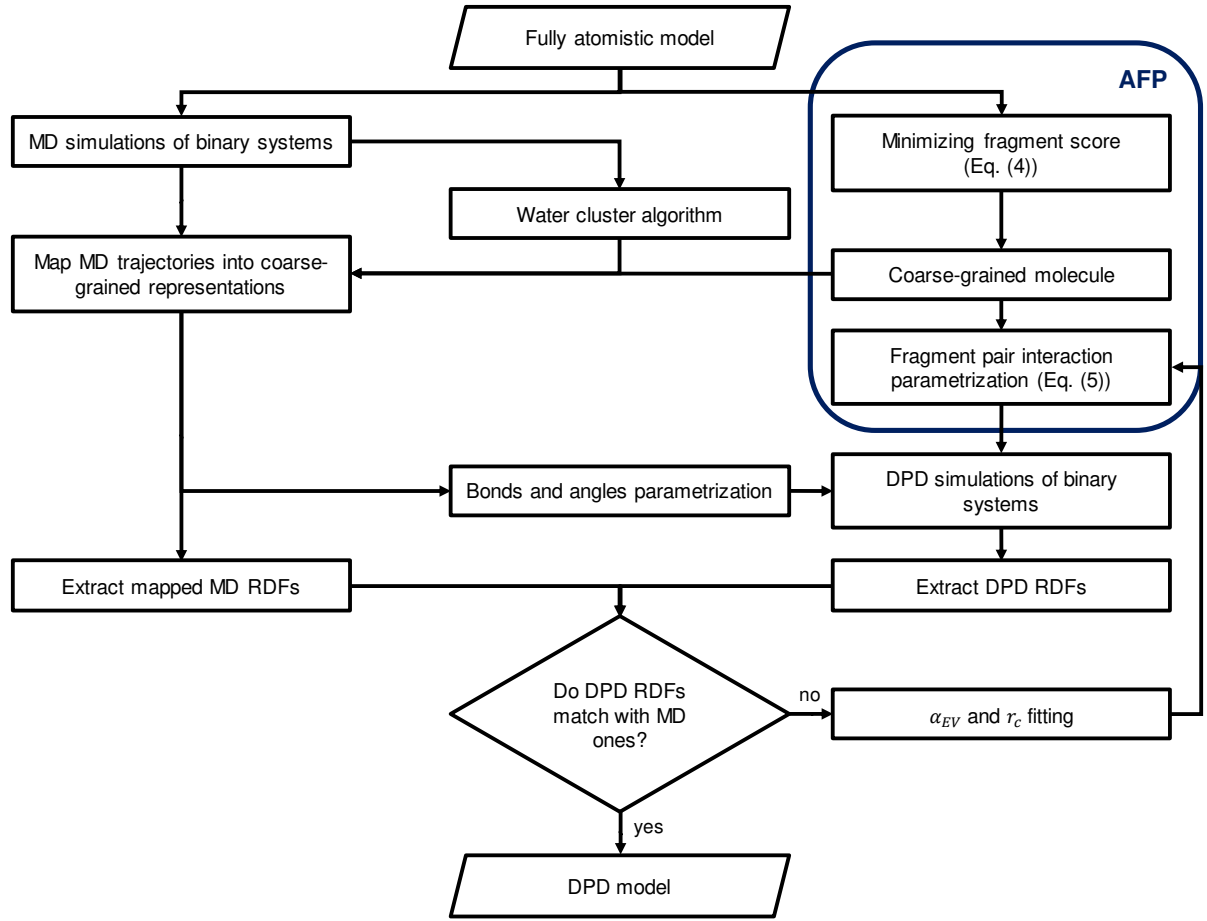
⁹⁵⁶ ¹⁰⁷L. Ford, R. Borwankar, R. Martin, and D. Holcomb, “Dressings and sauces,” in *Food*
⁹⁵⁷ *Emulsions*, edited by S. Friberg and K. Larsson (Marcel Dekker, New York, 1997) 3rd
⁹⁵⁸ ed., pp. 361–412.

⁹⁵⁹ ¹⁰⁸M. Langton, E. Jordansson, A. Altskär, C. Sørensen, and A.-M. Hermansson, “Mi-
⁹⁶⁰ crostructure and image analysis of mayonnaises,” *Food Hydrocolloids* **13**, 113–125 (1999).

⁹⁶¹ ¹⁰⁹M. Ferrari, J.-W. Handgraaf, G. Boccardo, A. Buffo, M. Vanni, and D. L. Marchisio,
⁹⁶² “Dataset for “Molecular modeling of the interface of an egg yolk protein-based emulsion” ,”
⁹⁶³ Zenodo (2021), Dataset, <https://doi.org/10.5281/zenodo.5703247>.

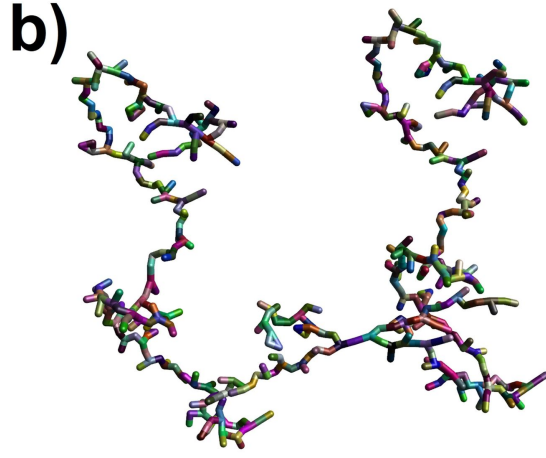
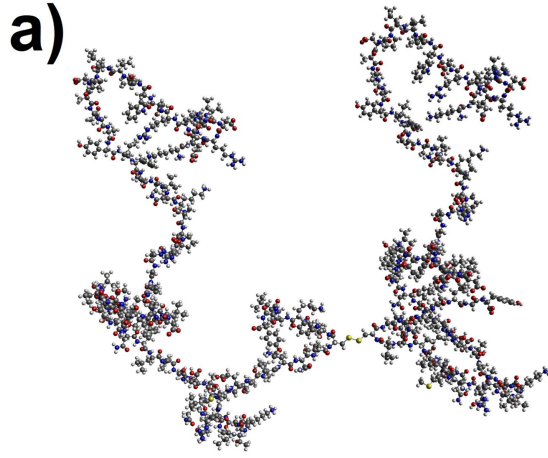
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883



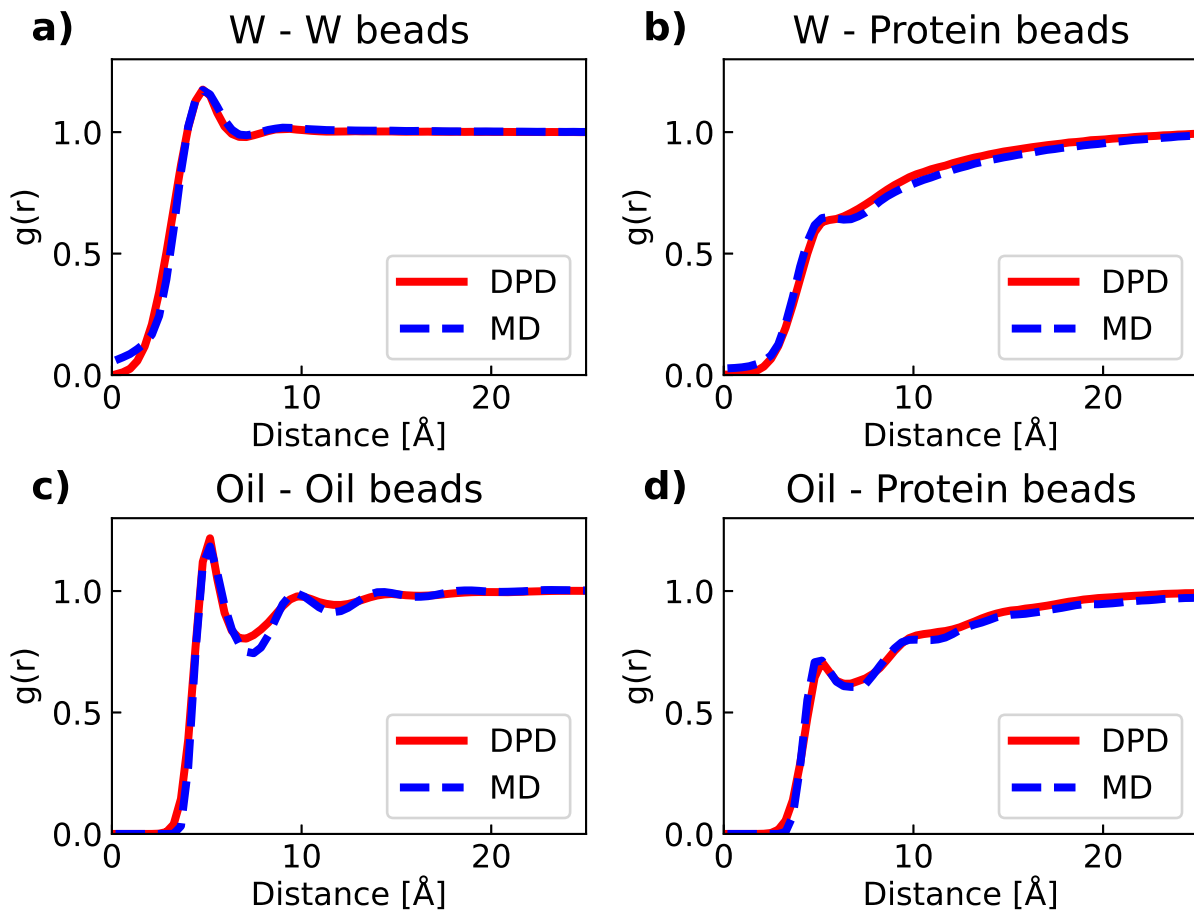
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883



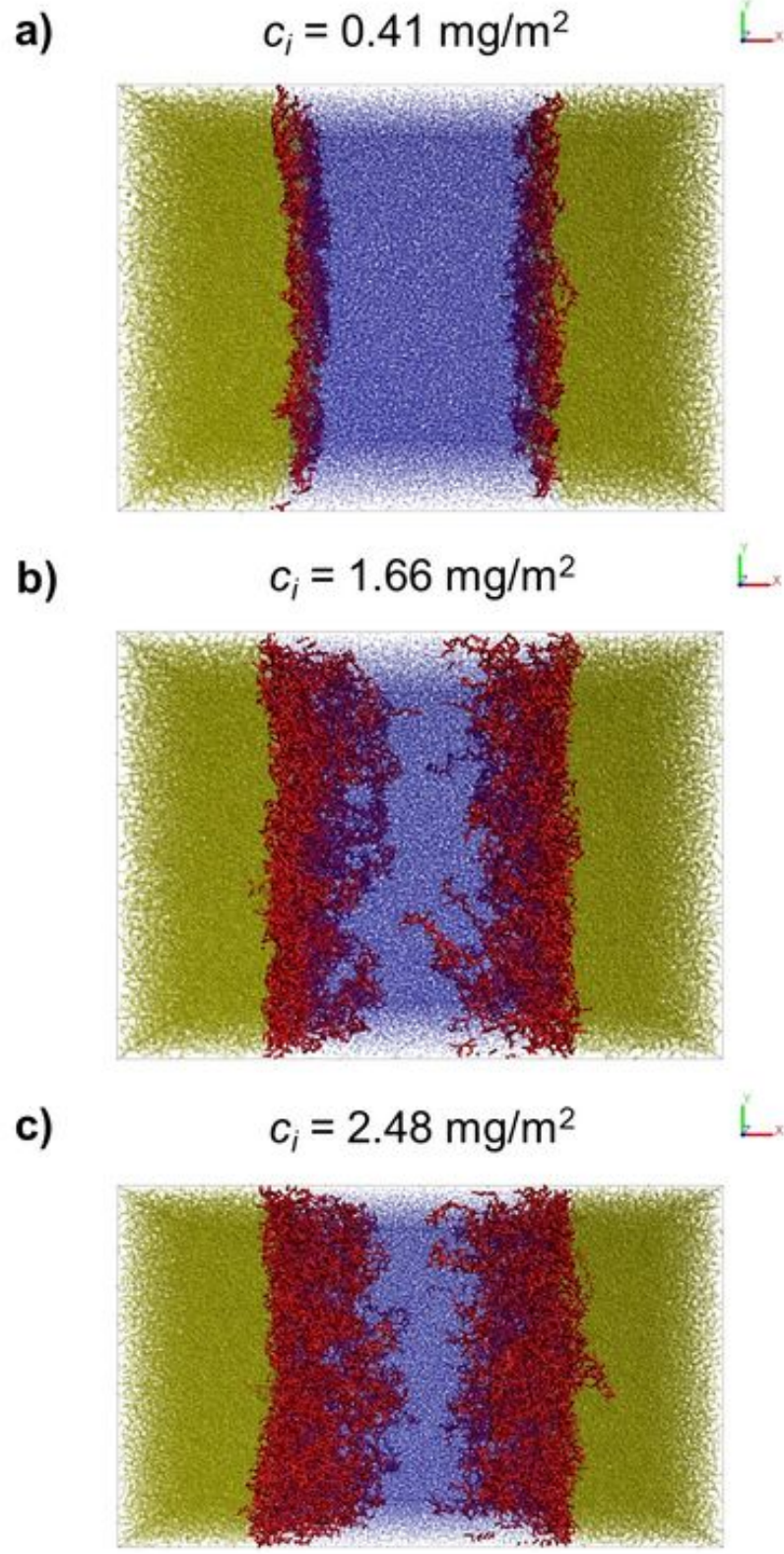
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0079883



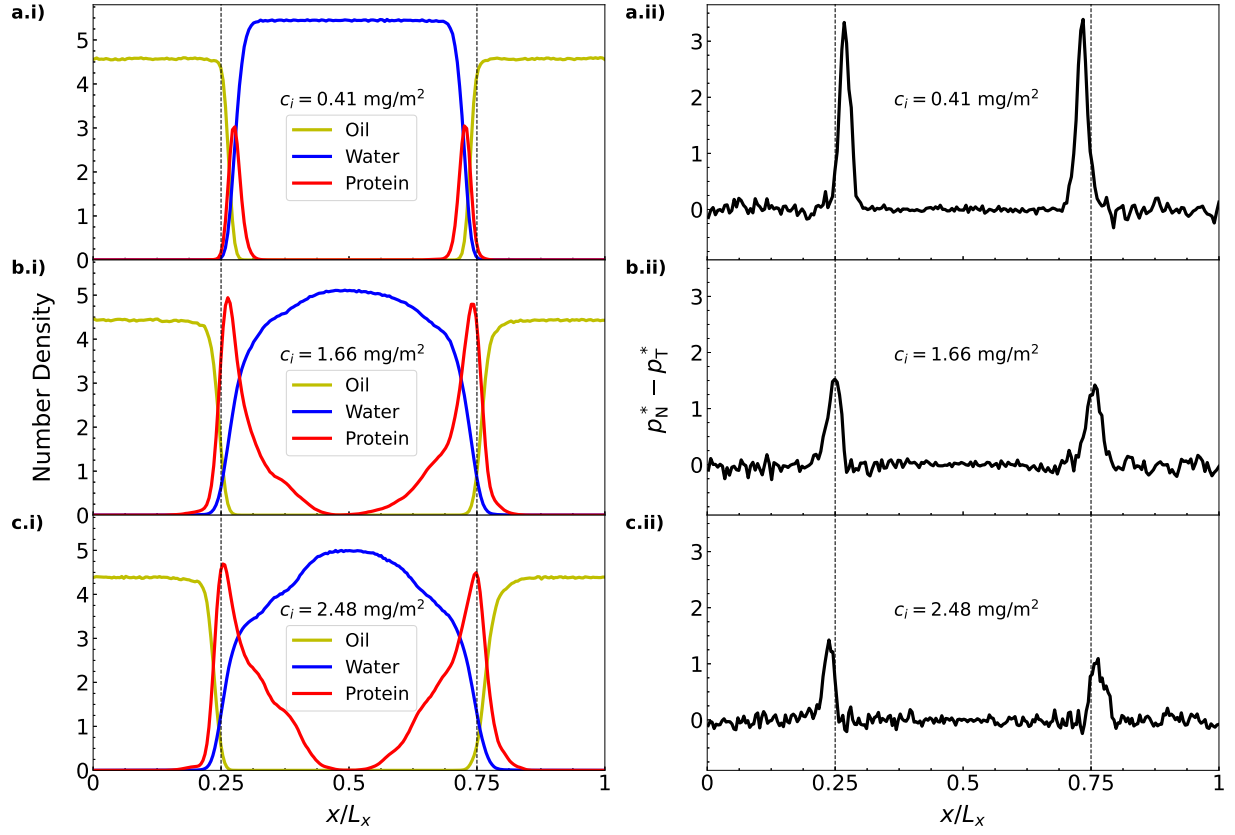
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883



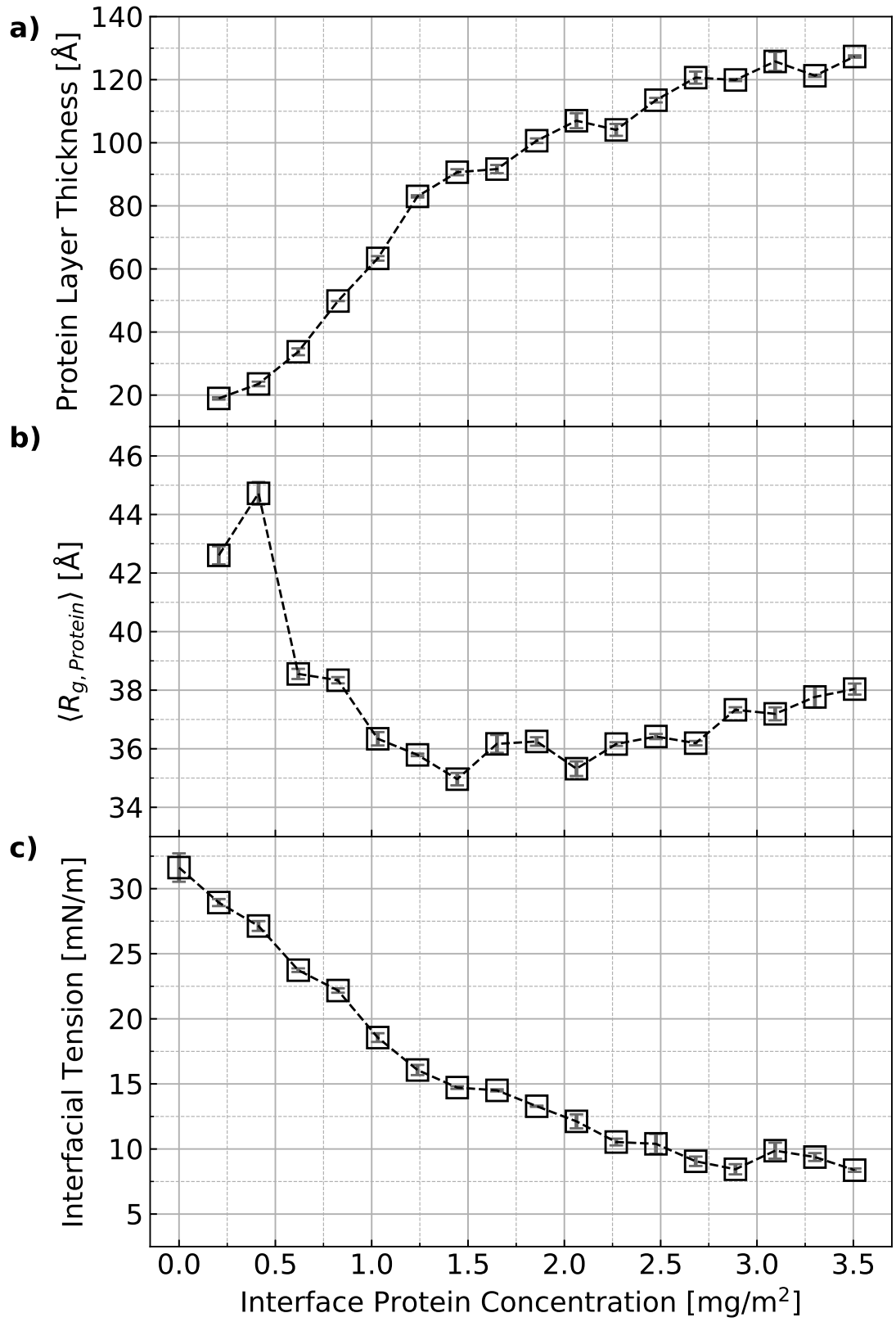
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883



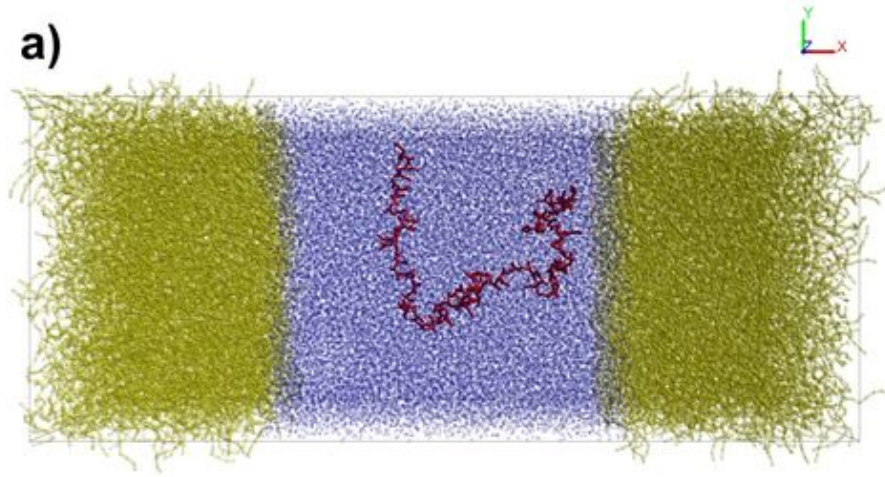
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0079883

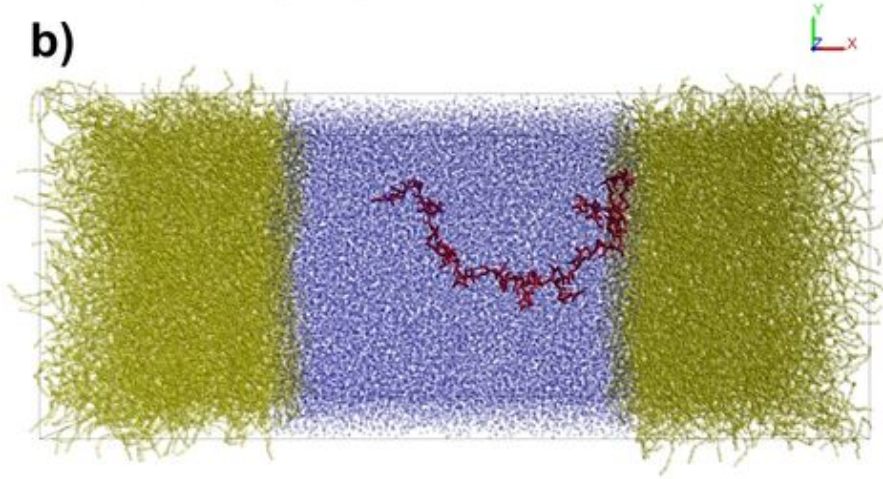


This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

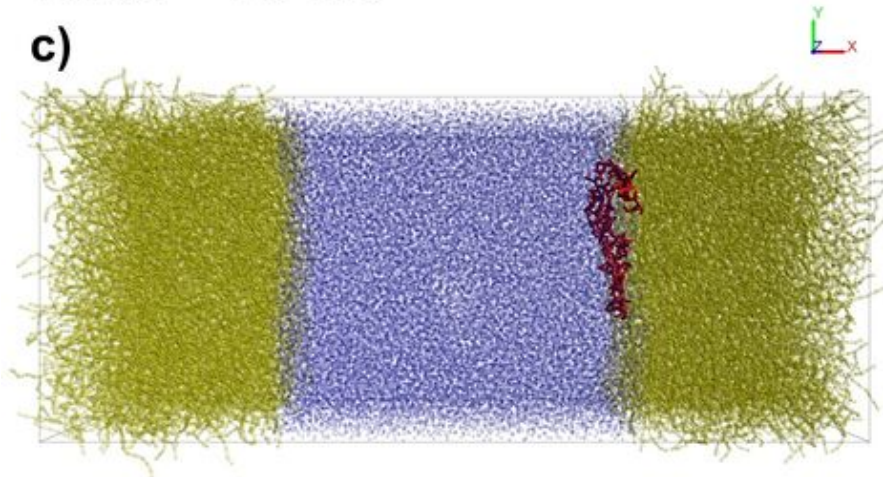
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883



Time = 10 ns



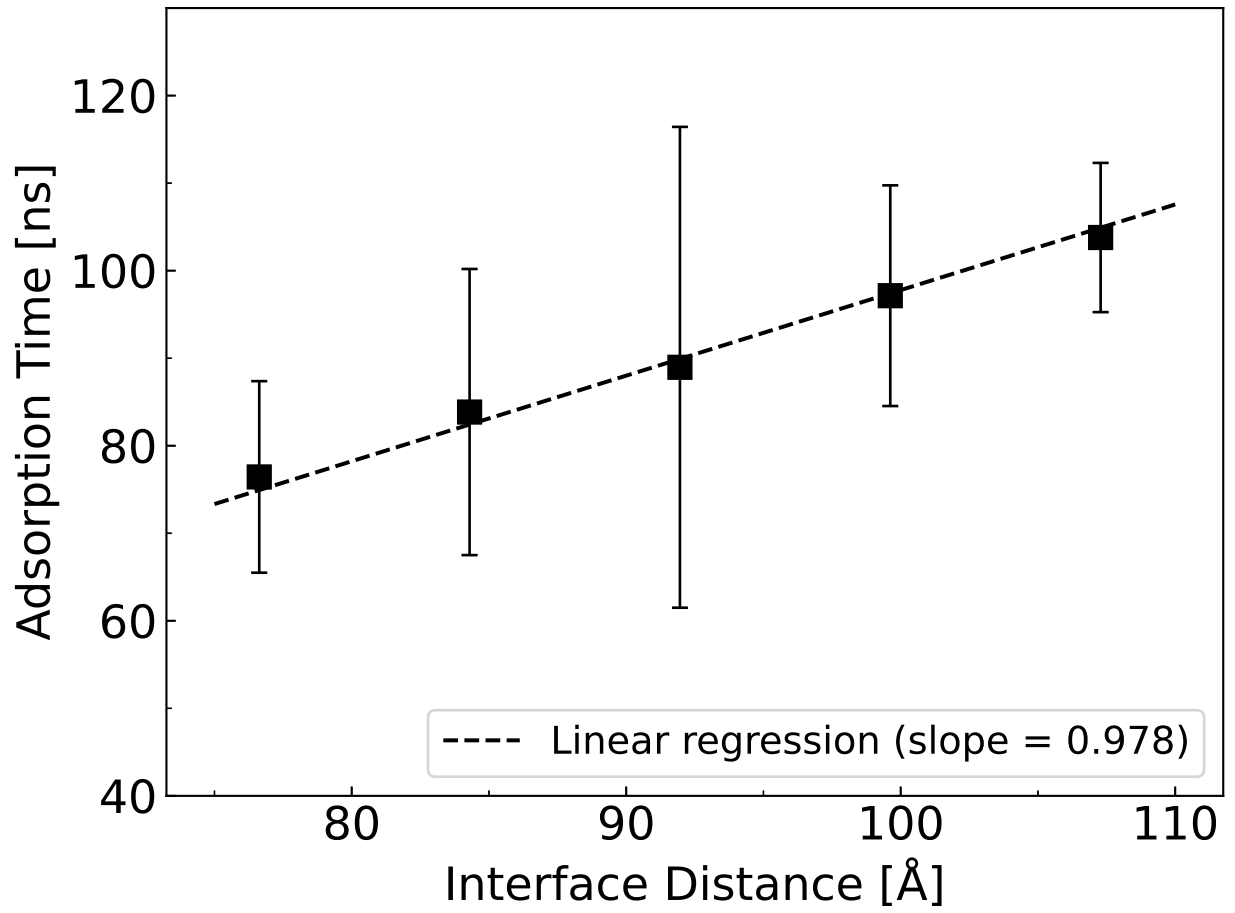
Time = 20 ns



Time = 80 ns

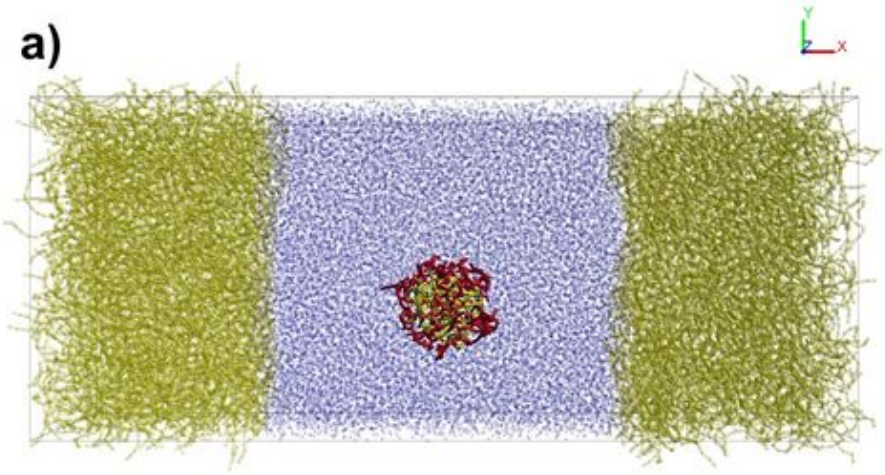
This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

PLEASE CITE THIS ARTICLE AS DOI: 10.1063/5.0079883

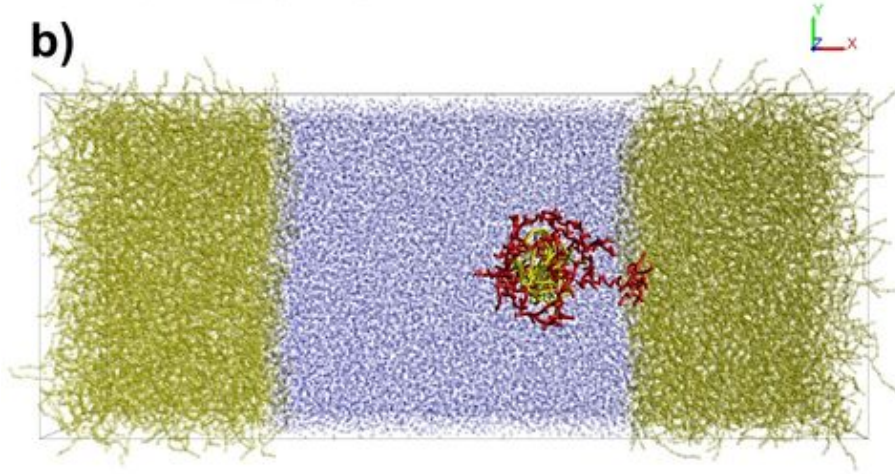


This is the author's peer reviewed, accepted manuscript. However, the online version of record will be different from this version once it has been copyedited and typeset.

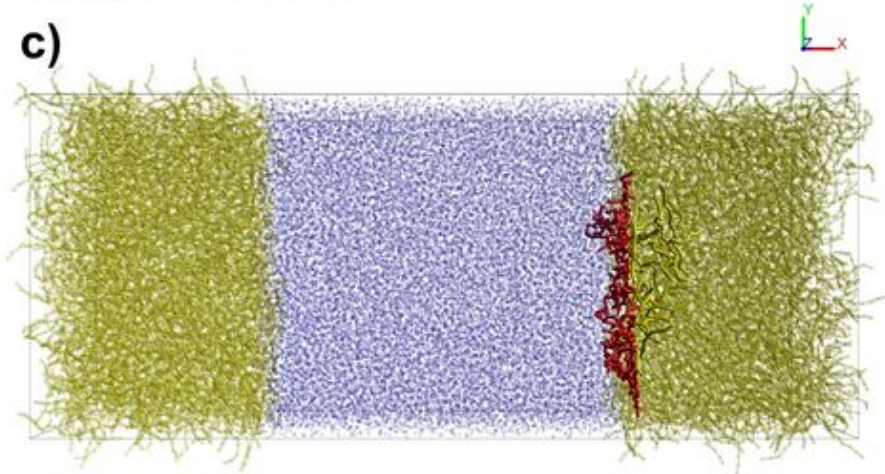
PLEASE CITE THIS ARTICLE AS DOI: 10.1063/1.50079883



Time = 20 ns



Time = 230 ns



Time = 260 ns