Elucidation of the 3D structure of grape seed 7S globulin and its interaction with malvidin 3-glucoside: a molecular modeling approach

Francisco Chamizo-González, Belén Gordillo*, and Francisco J. Heredia.

Food Colour and Quality Laboratory, Facultad de Farmacia, Universidad de Sevilla, 41012 Sevilla, Spain

Francisco Chamizo-González: fchamizo@us.es

Belén Gordillo: bgordillo@us.es

Francisco J. Heredia: heredia@us.es

* Corresponding author:

Belén Gordillo

Food Colour & Quality Lab., Área de Nutrición y Bromatología. Facultad de Farmacia. Universidad de Sevilla. 41012-Sevilla, Spain

Tel.: +34 955420938

e-mail: bgordillo@us.es

ABSTRACT

Plant proteins are biopolymers with interesting technological applications for the food industry due to their ability to interact with phenolic compounds such as anthocyanins. The 3D structure of the 7S globulin from grape seed was elucidated for the first time using a homology model. The constructed 3D model showed that grape seed 7S globulin is rich in α-helices and β-sheets stabilized by six disulfide bridges. The interaction with the major grape anthocyanin malvidin-3-glucoside was also assessed by Docking and Molecular Dynamic simulation. Theoretical results demonstrated that 7S globulin interacts with Mv3glc through hydrogen, alkyl and π-alkyl bonds and the flavylium cation is oriented towards a hydrophobic region of the protein, being protected from hydration. Results provide valuable insights for understanding the mechanisms involved in the molecular interaction of grape anthocyanins with grape seed proteins that could be relevant to use them as potential color protecting agents in food industry applications.

Keywords: Grape seed; 7S globulin; Molecular modeling; Docking; Anthocyanin.

1. INTRODUCTION

Plant proteins have gained great interest for the food industry as alternative biopolymers to animal-based proteins and synthetic counterparts due to their bio- and techno-functional properties, as well as to safety, health, and sustainability considerations (Paramita, Panyoyai, & Kasapis, 2020). In addition to their nutritional value, proteins are used in food applications to modulate textural and sensory attributes such as viscosity, gelation, elasticity, plasticity, emulsification, aroma, flavor, or color (among others) contributing to food quality (Tomadoni, Capello, Ayala, Valencia & Gutierrez, 2020).

The capability of proteins to modulate these properties is highly dependent on their structural features, their ability to form specific three-dimensional (3D) configurations/conformations, and the interaction with other molecules present in food matrices. The elucidation of the plant protein morphology and the molecular understanding of the protein-ligand interactions is, therefore, crucial to define their potential uses and technological applications. Variations in the molecular features and surface properties due to different proportion of the α-helices and β-sheets, number of disulfide bridges, binding cavities, hydrophobicity, accessibility of amino acid residues (mainly prolyl residues and proline repeats) are important factors for understanding the protein structure-function relationships (Yildirim-Elikoglu & Erdem, 2018).

A variety of plant proteins from soybean, pea, lentil, wheat gluten, rice or potato have demonstrated high technological value for their utilization in wine fining related to their capability to interact with phenolic compounds (Marangon, Vincenzi, & Curioni, 2019). It has been described that most of these protein-phenolic interactions occurs by non-covalent forces (hydrogen bonds, hydrophobic and ionic interactions, and van der Waals forces), which depend on the protein and phenolic structures and the medium conditions

(Ozdal et al., 2013; Ulrih, 2017). Regarding the protein structure, Granato, Piano, Nasi, Ferranti, Iametti, & Bonomi (2010), confirmed that plant-based proteins having different surface hydrophobicities showed different affinity and selectivity in binding different forms of phenolics. These variations imply that plant proteins of different origin may be more or less effectives as stabilizing agents of food components relevant to the sensorial properties (bitterness, aroma or color).

Besides the aforementioned sources of plant proteins, special attention has been paid recently toward the use of protein fining agents endogenous to grapes as seed storage proteins. However, despite it promising technological value, the three-dimensional structure of grape seed proteins and the mechanism of the molecular interaction with specific wine phenolics, is still unknown. At this respect, Gianazza et al., (1989) and Gazzola,Vincenzi, Gastaldon, Tolin, Pasini, & Curioni (2014) reported that one of the main and most abundant storage protein of grape seed endosperm (*Vitis vinifera*) is the 7S globulin. In this sense, the assessment of the molecular mechanism involved in the interaction between the grape seed 7S globulin and the color form of grape anthocyanins (flavylium cation) could be of interest to the food industry in general and to the winemaking sector in particular; as has been reported for other animal proteins (Fu, Belwal, He, Xu, Li, & Luo, 2020).

For this purpose, firstly, the 3D structure of the grape seed 7S globulin and other surface properties should be well understood.

In general, the three-dimensional molecular structure of a protein can be experimentally resolved by X-ray or NMR crystallography or a theoretically obtained by computational techniques such as homology modeling (Salmaso & Moro, 2018). In this way, computational chemistry are innovative tools for the development of theoretical models that allow the characterization of the three-dimensional structures of the protein and the

prediction of molecular interactions with other compounds of interest and therefore deepen technological functionality.

Molecular docking is one of the molecular coupling tools that aim to predict the predominant binding model of a ligand with a protein of known three-dimensional structure (Morris & Lim-Wilby, 2008). Molecular docking methods considers several possible conformations and orientations of the ligand within the protein binding. By using scoring functions, it searches for high-dimensional spaces and correctly ranks candidate dockings binding models. For this reason, the elucidation of the three-dimensional molecular structure of a target protein is a necessary condition to achieve a successful docking method, as the docking itself (Khatoon, Pandey, & Prajapati, 2017).

Molecular coupling is mainly composed of two stages: conformation/orientation search and a scoring function, which associates a score with each predicted pose (Huang & Zou, 2010). The sampling process must effectively search for the conformational space described by the free energy landscape, where the energy, in the coupling, is approximated by the scoring function. The scoring function should be able to associate the native bound conformation with the global minimum of the energy hypersurface.

Molecular dynamics (MD) is a type of molecular computational simulation that allows the behavior or evolution of a system (physical, chemical or biological) to be analyzed over time. It is a method to generate the trajectories of a system composed of N particles by direct numerical integration of the equations of Newtons of motion, with specifications of an interatomic interaction potential of adequate initial and boundary conditions. MD is a modeling and simulation method at the atomistic level when the particles in question are the atoms that make up the material or system of study.

The aim of this study is to establish, through a computational study, the theoretical structure of the grape seed 7S globulin from *Vitis vinifera* and to obtain a first approach

110 that can demonstrate its interaction with the major anthocyanin of red grapes and wines
111 in order to provide new insights to potential applications of grape seed proteins.

112 **2. MATERIALS AND METHODS**

113 **2.1. Protein template search and selection of grape seed 7S globulin**

114 The amino acid sequence of the 7S globulin basic from grape seed *(Vitis vinifera)* was
115 obtained from the UNIPROTKB database (Entry: A0A438KKJ2). Grape seed 7S
116 globulin was subjected to a sequence similarity search using the servers PSI-BLAST
117 (Altschul, Gish, Miller, Myers, & Lipman, 1990), HHpred (Söding, Biegert, & Lupas,
118 2005) and Phyre (Kelley & Sternberg, 2009). This procedure allows finding the most
119 similar protein sequence to use as a template for the homology modeling. Results of the
120 three servers were compared and the best template was selected based on the following
121 quality parameters: percentage of sequence, percentage of sequence identity, percentage
122 of coverage and the E value.

123 **2.2. Homology modeling of the grape seed 7S globulin**

124 Homology modeling was used to construct the three-dimensional (3D) structure of the
125 grape seed 7S globulin. An extracellular dermal glycoprotein from carrot (EDGP,
126 glycoside hydrolase) was selected as template among the structures available at the
127 Protein Database (PDB ID: 3VLA; resolution 0.95Å) for the model building, according
128 to Yoshizawa, Shimizu, Hirano, Sato, & Hashimoto (2012). The 7S globulin protein
129 sequence was aligned with the 3VLA protein sequence using the software Clustal W
130 (Thompson, Higgins, & Gibson 1994). Then, the protein sequence alignment was used
131 as input file to create the 3D protein model of the grape seed 7S globulin by using the
132 SWISS-MODEL tool provided by the EXPASY server (Waterhouse et al., 2018).
133 The validation parameters used for assessing the accuracy of the predicted model and its
134 stereo chemical properties were the Global Model Quality Estimate (GMQE)

(Waterhouse et al., 2018) and the Qualitative Model Energy Analysis (QMEAN) (Benkert, Biasini, & Schwede, 2010). The resulting GMQE score (expressed as values between 0 and 1) indicates the expected precision of the built model with the alignment and template, and the target coverage. GMQE scores close to 1 indicates greater reliability of the built model. QMEAN is a composite scoring function assessing the major geometrical aspects of protein structures. Together with the QMEAN, the QMEAN-Z score provides an estimation of the conservation degree of the structural characteristics obtained in the model on a global scale (Benkert et al., 2010). It indicates if the QMEAN score obtained in the model would be comparable to the expected values of similarly sized experimental structures. QMEAN Z scores close to 0 indicate good agreement between the predicted model structure and similarly sized experimental structures while scores lower or around to -4.0 indicate low-quality models.

In addition, Procheck, Verify3D, and Errat servers were used to assess the quality of the 3D obtained structure.

2.3. Model refinement and validation structure of the grape seed 7S globulin

The energy minimization of the obtained model was performed using the Gromacs 5.0.7 software (Hess, Kutzner, Van der Spoel, & Lindahl 2008) with the force field OPLS-AA.

Grape seed 7S globulin was solvated in 1.0 nm cubic boxes by using single point charge water molecules, which were then replaced with counterions for electroneutrality. The minimization of energy was carried out in 50.000 interaction steps (Lemkul, 2018). The minimized energy structure was evaluated using the Procheck, Verify-3D and Errat servers.

2.4. Protein-Ligand interactions by Docking and Molecular Dynamic (MD) simulation

Docking studies were performed to predict the putative binding of malvidin-3-O-glucoside (Mv3glc, the major grape anthocyanin) as ligand to the grape seed 7S globulin constructed model. AutoDock Vina software (Trott & Olson, 2010) was used for the analysis.

The initial Mv3glc structure was obtained from the PubChem data base and the 2D and 3D structure was constructed with the Avogadro software (Bolton,Wang, Thiessen, & Bryant, 2008).

The 2D structure of Mv3glc was obtained from the PubChem database (Kim et al., 2018) and was optimized to a minimum of energy using the MMFF9 force field. The structure's hydrogens were then adjusted for a pH of 3.5. This was done with Avogadro software (Hanwell, Curtis, Lonie, Vandermeersch, Zurek, & Hutchison, 2012).

Autodock Tools (Morris et al., 2009) was used to generate PDBQT files of receptor and ligands from their traditional PDB files to be used for docking in the AutoDock Vina.

PDBQT file is an extended PDB format of coordinate file that includes atomic partial charges. Hydrogen atoms were added to the macromolecule and partial atomic charges were calculated. A grid box with size of $56 \times 78 \times 54$ points was used in the configuration file of the Autodock Vina software to cover the entire protein.

The grid box was centered at the coordinate of X: -9.285, Y: 45.567, Z: 4.049. The positions of the protein atoms were kept fixed and the torsion angle of the glycosidic bond of the ligand was rotated until the rigid docking in software AutoDock Vina allowed the favorable docking. Other docking parameters were set to default. The docking results from AutoDock Vina were validate using Autodock 4 docking software (Morris et al., 2009).

183 The binding free energy value (ΔG) obtained from the docking simulation was used to
184 calculate the equilibrium dissociation constant (K_i) of the complex protein-ligand using
185 the following Equation:

186 Eq. (1) $K_i = e^{\Delta G/RT}$

187 Where, ΔG is the binding energy (cal/mol), R is the gas constant (1.986 cal/mol*K), T is
188 the temperature (298K), and *e* is the Euler number.

189 To evaluate the stability of the interaction between the Mv3glc and the *Vitis vinifera* 7S
190 globulin, molecular dynamic (MD) simulations were carried out using the Gromacs
191 software. Mv3glc topology file was obtained from the CGenFF server
192 (Vanommeslaeghe et al., 2010)

193 The protein-ligand complex was solvated into a box of explicit single point charge
194 (SPC) water molecules and simulated using periodic boundary conditions (PBC) and
195 particle mesh Ewald sum (PME) to improve electrostatic interactions. System power
196 was minimized using 1000 steps. Two equilibrium stages of 100 ps each one were
197 performed to reach the optimal conditions of the pressure and temperature. The
198 reference pressure and temperature were 1 bar and 300 K (GROMACS 5.0.7 package).
199 After the two balancing phases completed, the desired temperature and pressure was
200 adjusted by the system. Then, a MD simulation was run 10 ns with a 2-fs lics time step
201 algorithm. The resulting trajectory was analyzed using GROMACS earnings. The Root
202 means square deviation parameter (RMSD) was used to calculate how much the
203 position of the ligand varies relative to the protein during the simulation time.

204 **2.5. Analysis of the grape seed 7S globulin structure and interaction studies**

205 Different softwares were used to visualize and analyze the results obtained in the
206 protein modeling and in the docking and molecular dynamics studies. UCSF Chimera
207 software was used to illustrate the 3D models (Pettersen et al., 2004). Two-dimensional

208 (2D) illustrations of Mv3glc sites interacting with the 7S globulin amino acids were
209 made using the Discovery Studio.

210 **3. RESULTS AND DISCUSSION**

211 **3.1. Model building of grape seed 7S globulin protein**

212 The homology modeling method was applied to construct a 7S globulin protein atomic
213 resolution model from the primary amino acid sequence and the crystal structure of a
214 defined homologous protein. The primary grape seed 7S globulin protein sequence was
215 sent to the PSI-BLAST (Altschul et al., 1990), HHpred (Söding et al., 2005) and Phyre
216 (Kelley & Sternberg, 2009) servers to identify the possible homologous structure to use
217 as a template for the homology modeling.

218 Table 1 shows the results of the percentages of identity and coverage, as well as the E
219 values for the different possible homologous protein structures to grape seed 7S
220 globulin protein. Possible homologous protein structures given by the serves included a
221 *Daucus carota* glucoside hydrolase (3VLA and 3VLB types), conglutin gamma from
222 lupine seeds (7S globulin type), 7S globulin from soybean and from lupine seeds,
223 endothiapepsin from *Cryphonectria parasitica*, and cathepsin D from *Rattus*
224 *norvegicus*. Based on the highest percentage of identity and coverage obtained by the
225 three servers, the crystal structure of *Daucus carota* glucoside hydrolase type A (3VLA)
226 was chosen as the most homologous suitable template to the grape seed 7S globulin.
227 The results of the percentage of identity given by the serves ranged from 68 to 69% and
228 the percentages of coverage from 93 to 100%.

229 Other valuable information given by the three servers indicate that 3VLA form *Daucus*
230 *carota* protein is rich in alpha and beta helices, which increase its reliability as a
231 homologous template for the grape seed 7S globulin.

In addition, Yoshizawa et al., (2012) confirmed that the structure of *Daucus carota* glucoside hydrolase (3VLA) was similar to 7S globulin from soybean (Bg7S).

It was observed that the first 23 amino acids of the grape seed 7S globulin protein sequence were not modeled in the structure, which could indicate that they act as signal peptides (SP). SP are short chain of peptides located at the N-terminal of proteins, which provides information for the protein secretion. In both prokaryotic and eukaryotic cells, proteins are allowed entering to the secretory pathway only if they contain specific addressing signal such as SP. In most cases, SP is a transient extension of the amino N-terminal of the protein, which is removed by peptidase signals once its targeting function has been carried out. To confirm this hypothesis, the amino acid sequence of the *Vitis vinifera* 7S globulin was analyzed by the server Signal P-5.0, which allows identifying SP in diverse classes of vegetal proteins, according to Almagro Armenteros et al., (2019). Results showed that the first amino acid of the primary structure of 7S globulin protein was alanine, with a probability of approximately 95%, as shown in Fig. S1. These results confirmed, therefore, that the short chain containing the 23 amino acids could be considered a SP of the grape seed 7S globulin.

Secondary structure alignment of the grape seed 7S globulin and 3VLA showed consensus structure consisted of 7 α-helices and 23 β-sheet. The rest of the regions contained random coils, structural mismatches, and sequence gap (Fig. S2). Therefore, 3VLA was used as a valid template for the construction of the 3D model of *Vitis vinifera* grape seed 7S globulin. The coordinate file for the atoms was obtained from the Protein Data Bank at the Brookhaven National Laboratory (PDB ID: 3VLA). The alignment search results against the PDB database of various servers and the Swiss model using 3VLA as template generated the model. Swiss model performs a comparative protein modeling by satisfying the spatial constraints of alignment with a

related structure (Waterhouse et al., 2018). The results for the evaluation of the generated model was carried out based on the following parameters: GMQE, QMEAN, Procheck, Verify3D and Errat. Results indicate that the GMQE score for the grape seed 7S globulin was 0.81 and the QMEAN -1.67, indicating good structural similarity between the two proteins.

The Procheck parameter provides a detailed evaluation on the stereochemistry of the conformation of the main chain. It generates a graph of conformational angles of each residue: φ angle (rotation around the N-Cα bond) and ψ angle (around the Cα-C bond of the same Cα atom, ψ) and a complete list of the residues. The results revealed that 90.1% of the residues were in favored regions, 7.7% in allowed regions, 0.9% in generous regions and 0.6% in disallowed regions.

Verify3D provided a value of 88.5%, which indicates a good validation of the protein structure. Therefore, there was a good compatibility of the 3D structure model with the amino acid sequence.

Errat verifies the structure of the protein by detecting local errors based on the statistics of unbound atomic interactions and comparing them with statistics of highly refined structures to suggest an overall quality factor. Results showed a factor of 89.7%.

The Procheck, Verify3D and Errat scores of the selected model are within the acceptable range. Therefore, the model obtained by the Swiss Model server was reliable and used for the refinement of the *Vitis vinifera* grape seed 7S globulin structure.

3.2. Model refinement and energy minimization

The preliminary model generated by the Swiss Model is characterized by a high energy level (-367217.7 kJ /mol). This is because the generated model contained unfavorable bond lengths, bond angles, torsion angles, and contacts. Thus, the model was subjected to an energy minimization and geometry optimization using the steepest descent

algorithm Gromacs (OPLS-AA force field). The energy minimization step is necessary to obtain a stable protein model, which provides a 3D conformation closer to its native state with effective functionality. Collisions and stearic stresses were reduced by relaxing close contacts in the geometric chain without significantly changing the overall structure (Messaoudi, Belguith, & Hamida, 2013).

Table 2 shows the results of the energy minimization and geometry optimization process. The energy minimization reduced the potential energy of the grape seed 7S globulin model to -1409676.87 kJ/mol, which indicate better model packaging quality. On the other hand, the Verify3D score for the 7S globulin model did not change significantly but the model had a satisfactory quality value of 89.3%.

The score indicated that 90.6% of the total 7S globulin amino acids were correctly positioned. Errat score did not improve significantly after the energy minimization, which indicate that local errors within the geometry in the model were largely unrepaired.

A model with an Errat score greater than 50% is considered a high-quality model. The Ramachandran (Procheck) plot also did not change significantly. Procheck analysis showed several residues located in the non-permitted region of the Ramachandran plot (Fig. 2), which were Ser 231, Arg96, Ser196, Ser417 and Arg52. The analysis of Procheck indicated that 90.6% of the amino acids were in the most favored region, 8.2% in the additional allowed region, and 0.3% in the generously allowed region.

Residues in the favored region were free of steric interference, while some relaxation of steric interference was allowed in the allowed region. The residues that were in the not allowed region contained the error in their structures. Therefore, based on the aforementioned results, it can be confirmed that the 3D constructed model of the grape seed 7S globulin was satisfactory.

3.3. Structural analysis of grape seed 7S globulin protein

The 7S globulin 3D model was determined using homology model computational techniques (Fig. 1). It was made up of a single chain that has α helices and β sheet and is roughly divided into a central slit.

The 7S globulin model adopted all the structural characteristics of the EDGP, the model protein selected for the structural design. In order to mathematically verify its similarity, the 7S globulin model was superimposed with the EDGP structure and a mean square deviation (RMSD) of 0.161 was obtained. This result indicates that the 7S globulin model is an accurate presentation of the experimental structure. Theoretical results indicated that 7S globulin has six disulfide bonds, which could help to stabilize the tertiary structure. Especially interesting are the Cys92-Cys116 and Cys426-Cys204 bonds. It seems that they are important for the stabilization of the protein because it links α helices and β sheet.

EDGP protein has been found to be similar to other plant proteins such as wheat TAXI-IA and soybean Bg7S. These two proteins are also largely stabilized by six disulfide bonds, which support the results found for the grape seed 7S globulin. Although the patterns of disulfide bridge formation are similar between EDGP and soybean Bg7S, they differ from those of wheat TAXI-IA (Yoshizawa et al., 2012).

In addition, different physico-chemical parameters of the grape seed 7S globulin were obtained by the ProtScale server from the Expasy server. Results showed that grape seed 7S globulin has a molecular weight of 43.7 KDa. Likewise, Gazzola et al., (2014) reported the existence of a 43 KDa monomeric protein that showed a high homology to the 7S globulin from legume seeds. Results were also in accordance with the weight of soybean basic globulin 7S described by Yoshizawa et al., (2012). Also, 7S globulin has a stability index that classifies it as stable.

The isoelectric point (pI) is the pH where the protein is not charged and therefore the pH at which 7S globulin could show a minimum of solubility. The computed pI value for the *Vitis vinifera* grape seed 7S globulin was 8.96 (pI>7), which indicate it basic nature. The computed isoelectric point (pI) is a useful parameter to develop and optimize purification methods of such as protein. Extinction coefficient calculations were also performed, which is useful to have an estimation for spectrophotometric measurements. The extinction coefficient obtained was expressed as units of $M^{-1}cm^{-1}$, at 280 nm measured in water. The results obtained show that when all the pairs of Cys residues are considered they form cystines. In this case, the value obtained was 29630 $M^{-1}cm^{-1}$ and if it is assumed that all Cys residues are reduced.

The lifetime was 10 h, this parameter is the time (in vitro) that it takes for the protein to disappear once it is synthesized by the cell.

The aliphatic index is the relative volume occupied by the aliphatic side chains that allows establishing a relationship with the thermostability of the protein. Also, ProtScale server made possible the knowledge of an average value of hydropathy (GRAVY) that indicates the possibility for 7S globulin to establish interactions with water. This finding suggests that 7S globulin may be highly hydrated in media aqueous. In this sense, the lower the general index of hydropathy, the greater the possibility of establishing this kind of interaction (Kyte & Doolittle, 1982). For this parameter, the value obtained was 0.110.

ProtScale hydrophobic analysis (Fig. S3 A) revealed that 7S globulin exhibited two maximum hydrophobicity at amino acids 11 and 12 and a minimum at amino acid 399. Finally, for flexibility, the Bhaskaran scale was used, which gives flexibility values for each of the amino acids, with variations between 0.300 and 0.540 (Gasteiger,

Hoogland, Gattiker, Wilkins, Appel & Bairoch, 2005). The high flexibility values suggest that they are flexible molecules (Fig. S3 B).

3.4. Docking and Molecular Dynamic of binding and subsite structure analysis

It was performed to predict non-covalent binding of the 7S globulin receptor (protein) with malvidin-3-O-glucoside (ligand) and to determine the structure of the most favorable binding site. Molecular coupling was performed using AutoDock Vina (Trott & Olson, 2010), a software that uses a method to produce the best ligand conformations that have the lowest possible binding energies. The 3D structure of the grape seed 7S globulin obtained from the homology modeling was used as the receptor molecule. And the Malvidin-3-O-glucoside (Mv3glc) as ligand (Fig. S4).

For the docking, it is necessary to prepare the files of the protein, the ligand and the conditions to carry out the interaction. The protein and ligand PDB files were edited using the AutoDock Tools software (Morris et al., 2009)

For the 7S globulin, non-protein molecules such as water and ions were removed and Kollman charges and polar hydrogens were added to the protein structure. These hydrogen atoms are important because they participate significantly in the interaction of the ligand. For the ligand, the glycosidic bond was considered rotary and partial atomic charges in the form of Gasteiger charges were added to the structure. The edited 7S globulin PDB files and ligands were saved as PDBQT files as they will be the files needed to run the AutoDock Vina simulation. Finally, grid box was designed to define an interaction space of the ligand with the protein. In our case, dimensions were made to cover all the protein. The grid box was centered at the coordinate of X: -9.285, Y: 45.567, Z: 4.049. Nine resulting ligand conformations were generated in the protein environment after each run in the AutoDock Vina software. The results are showed in Table S1.

381 The lowest energy conformation was chosen as the most favorable conformation at the
382 binding site. The coupling that resulted in the lowest binding energy (−7.0 kcal/mol),
383 indicating a high binding affinity. Fig. 3A shows the most favorable conformation of
384 Mv3glc when interacts with *Vitis vinifera* grape seed 7S globulin, being stabilized
385 mainly by hydrogen bonds. The results were corroborated with the coupling results of
386 Autodock 4 (Morris et al., 2009) that produced a minimal energy conformation identical
387 to that obtained by Autodock Vina. Therefore, it validates the Autodock Vina docking
388 results. The equilibrium constant was also calculated through equation 1, and the result
389 was $K_i = 7.3x10^{-3}$. Stănciuc et al., (2017) carried out experimental studies on the
390 interaction of anthocyanin compounds and different types of proteins such as β-
391 lactoglobulin, which reported affinity constant values for a temperature of 278 K on the
392 order of 10^{-6}. This fact supports a good approximation through theoretical calculation.

393 The binding of Mv3glc with the grape seed 7S globulin is illustrated as 2D drawings in
394 Fig. 3B. The side chain atoms of the residue Arg155, Ser153 and Ser26 form hydrogen
395 bonds with the hydroxyl group of the glucose (G). The ring A interacts through
396 hydrogen bonds with the Ser139 residue in addition to establishing other interactions
397 such as π-alkyl and alkyl. Ring B interacts primarily with residues of Arg150 and Thr25
398 via hydrogen bonds. Moreover, due to the aforementioned interactions, it was observed
399 that the C ring was oriented towards Val151, an amino acid having hydrophobic
400 characteristics. This finding is interesting since the oxygen in the C ring, being in a
401 hydrophobic region, would be protected from hydration (Fig. 3C). Protecting the
402 hydration of the flavyllium cation by a protein could represent a way of color protection
403 similar to the copigmentation phenomenon, which suggest potential technological
404 applications for the grape seed 7S globulin in the food industry.

The coupled ligand remained stable at the 7S globulin active site throughout the simulation (Fig. S5). To assess the deviation of the structure, the RMSD was calculated during the simulation. A stable interaction between the protein and the pigment was observed since it reached rapid equilibrium approximately from 0.2 ns, from this value the RMSD values remain between approximately 0.5 and 0.65 (Fig. 4) and it exhibited low levels of RMSD.

4. CONCLUSIONS

In this study, the structure of 7S globulin from *Vitis vinifera* grape seed was obtained for the first time from the amino acid sequence. The results show that 7S globulin is made up of α and β-sheet helices and has a great structural similarity with other plant globulin proteins such as soybean 7S globulin. The parameters that evaluate the quality of the protein morphology improved when the energy of the protein was minimized. In addition, the chemical interaction between the grape seed 7S globulin and malvidin-3-O-glucoside was demonstrated using docking techniques and molecular dynamics. Results showed that in this interaction, the flavylium cation is oriented towards a hydrophobic region of the protein, being protected from hydration. The findings open a new path of experimental studies focused on the protection of the color of anthocyanins, which could have a special interest in food industry applications.

Acknowledgments

Authors thank the assistance of the technical staff of Biology Service (SGI, Universidad de Sevilla, Spain).

Funding sources

This research was financially supported by the Ministerio de Economía y Competitividad, Spain, Gobierno de España (Project AGL2017-84793-C2 and FPI grant PRE2018-087184)

430

431 **Conflict of interest statement (before references)**

432 Authors declare no conflict of interest.

433 **References**

434 Almagro Armenteros, J. J., Tsirigos, K. D., Sønderby, C. K., Petersen, T. N., Winther,
435 O., Brunak, S., von Heijne, G., & Nielsen, H. (2019). SignalP 5.0 improves signal
436 peptide predictions using deep neural networks. *Nature Biotechnology*, *37*(4), 420–
437 423. https://doi.org/10.1038/s41587-019-0036-z

438 Altschul, S. F., Gish, W., Miller, W., Myers, E. W., & Lipman, D. (1990). Basic local
439 alignment search tool. *J Mol Biol*, *215*(4), 3–4.

440 Benkert, P., Biasini, M., & Schwede, T. (2010). Toward the estimation of the absolute
441 quality of individual protein structure models. *Bioinformatics*, *27*(3), 343–350.
442 https://doi.org/10.1093/bioinformatics/btq662

443 Bolton, E. E., Wang, Y., Thiessen, P. A., & Bryant, S. H. (2008). *Chapter 12 -*
444 *PubChem: Integrated Platform of Small Molecules and Biological Activities* (R. A.
445 Wheeler & D. C. B. T.-A. R. in C. C. Spellmeyer (eds.); Vol. 4, pp. 217–241).
446 Elsevier. https://doi.org/https://doi.org/10.1016/S1574-1400(08)00012-1

447 Fu, X., Belwal, T., He, Y., Xu, Y., Li, L., & Luo, Z. (2020). Interaction and binding
448 mechanism of cyanidin-3-O-glucoside to ovalbumin in varying pH conditions: A
449 spectroscopic and molecular docking study. *Food Chemistry*, *320*(February).
450 https://doi.org/10.1016/j.foodchem.2020.126616

451 Gasteiger, E., Hoogland, C., Gattiker, A., Wilkins, M. R., Appel, R. D., & Bairoch, A.
452 (2005). Protein identification and analysis tools on the ExPASy server. In *The*
453 *proteomics protocols handbook* (pp. 571–607). Springer.

454 Gianazza, E., Tedesco, G., Villa, P., Scienza, A., Cargnello, G., Righetti, P. G., &

455 Osnaghi, A. (1989). Characterization of the major proteins from Vitis vinifera
456 seeds. *Plant Science*, *62*(1), 73–81. https://doi.org/https://doi.org/10.1016/0168-
457 9452(89)90191-X
458 Hess, B., Kutzner, C., van der Spoel, D., & Lindahl, E. (2008). GROMACS 4:
459 Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular
460 Simulation. *Journal of Chemical Theory and Computation*, *4*(3), 435–447.
461 https://doi.org/10.1021/ct700301q
462 Huang, S.-Y., & Zou, X. (2010). Advances and challenges in protein-ligand docking.
463 *International Journal of Molecular Sciences*, *11*(8), 3016–3034.
464 Kelley, L. A., & Sternberg, M. J. E. (2009). Protein structure prediction on the Web: a
465 case study using the Phyre server. *Nature Protocols*, *4*(3), 363–371.
466 https://doi.org/10.1038/nprot.2009.2
467 Khatoon, N., Pandey, R. K., & Prajapati, V. K. (2017). Exploring Leishmania secretory
468 proteins to design B and T cell multi-epitope subunit vaccine using
469 immunoinformatics approach. *Scientific Reports*, *7*(1), 1–12.
470 https://doi.org/10.1038/s41598-017-08842-w
471 Kyte, J., & Doolittle, R. F. (1982). A simple method for displaying the hydropathic
472 character of a protein. *Journal of Molecular Biology*, *157*(1), 105–132.
473 https://doi.org/10.1016/0022-2836(82)90515-0
474 Lemkul, J. (2018). From proteins to perturbed Hamiltonians: A suite of tutorials for the
475 GROMACS-2018 molecular simulation package [article v1. 0]. *Living Journal of*
476 *Computational Molecular Science*, *1*(1), 5068.
477 Marangon, M., Vincenzi, S., & Curioni, A. (2019). Wine fining with plant proteins.
478 *Molecules*, *24*(11), 2186.
479 Messaoudi, A., Belguith, H., & Hamida, J. Ben. (2013). Homology modeling and

480 virtual screening approaches to identify potent inhibitors of VEB-1 β-lactamase.
481 *Theoretical Biology and Medical Modelling*, *10*(1), 22.
482 Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S.,
483 & Olson, A. J. (2009). AutoDock4 and AutoDockTools4: Automated docking with
484 selective receptor flexibility. *Journal of Computational Chemistry*, *30*(16), 2785–
485 2791. https://doi.org/10.1002/jcc.21256
486 Morris, G. M., & Lim-Wilby, M. (2008). Molecular docking. In *Molecular modeling of*
487 *proteins* (pp. 365–382). Springer.
488 Ozdal, T., Capanoglu, E., & Altay, F. (2013). A review on protein-phenolic interactions
489 and associated changes. *Food Research International*, *51*(2), 954–970.
490 https://doi.org/10.1016/j.foodres.2013.02.009
491 Paramita, V. D., Panyoyai, N., & Kasapis, S. (2020). Molecular Functionality of Plant
492 Proteins from Low- to High-Solid Systems with Ligand and Co-Solute.
493 *International Journal of Molecular Sciences*, *21*(7), 2550.
494 https://doi.org/10.3390/ijms21072550
495 Pettersen, E. F., Goddard, T. D., Huang, C. C., Couch, G. S., Greenblatt, D. M., Meng,
496 E. C., & Ferrin, T. E. (2004). UCSF Chimera--a visualization system for
497 exploratory research and analysis. *Journal of Computational Chemistry*, *25*(13),
498 1605–1612. https://doi.org/10.1002/jcc.20084
499 Salmaso, V., & Moro, S. (2018). Bridging molecular docking to molecular dynamics in
500 exploring ligand-protein recognition process: An overview. *Frontiers in*
501 *Pharmacology*, *9*, 923.
502 Söding, J., Biegert, A., & Lupas, A. N. (2005). The HHpred interactive server for
503 protein homology detection and structure prediction. *Nucleic Acids Research*,
504 *33*(suppl_2), W244–W248. https://doi.org/10.1093/nar/gki408

505 Stănciuc, N., Turturică, M., Oancea, A. M., Barbu, V., Ioniţă, E., Aprodu, I., &
506 Râpeanu, G. (2017). Microencapsulation of Anthocyanins from Grape Skins by
507 Whey Protein Isolates and Different Polymers. *Food and Bioprocess Technology*,
508 *10*(9), 1715–1726. https://doi.org/10.1007/s11947-017-1938-8
509 Thompson, J. D., Higgins, D. G., & Gibson, T. J. (1994). CLUSTAL W: improving the
510 sensitivity of progressive multiple sequence alignment through sequence
511 weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids*
512 *Research*, *22*(22), 4673–4680. https://doi.org/10.1093/nar/22.22.4673
513 Tomadoni, B., Capello, C., Valencia, G. A., & Gutiérrez, T. J. (2020). Self-assembled
514 proteins for food applications: A review. *Trends in Food Science and Technology*,
515 *101*, 1-16. https://doi.org/10.1016/j.tifs.2020.04.015
516 Trott, O., & Olson, A. J. (2010). AutoDock Vina: improving the speed and accuracy of
517 docking with a new scoring function, efficient optimization, and multithreading.
518 *Journal of Computational Chemistry*, *31*(2), 455–461.
519 https://doi.org/10.1002/jcc.21334
520 Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian,
521 E., Guvench, O., Lopes, P., & Vorobyov, I. (2010). CHARMM general force field:
522 A force field for drug-like molecules compatible with the CHARMM all-atom
523 additive biological force fields. *Journal of Computational Chemistry*, *31*(4), 671–
524 690.
525 Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer,
526 F. T., de Beer, T. A. P., Rempfer, C., Bordoli, L., Lepore, R., & Schwede, T.
527 (2018). SWISS-MODEL: homology modelling of protein structures and
528 complexes. *Nucleic Acids Research*, *46*(W1), W296–W303.
529 https://doi.org/10.1093/nar/gky427

Yildirim-Elikoglu, S., & Erdem, Y. K. (2018). Interactions between milk proteins and polyphenols: Binding mechanisms, related changes, and the future trends in the dairy industry. *Food Reviews International*, *34*(7), 665–697.

Yoshizawa, T., Shimizu, T., Hirano, H., Sato, M., & Hashimoto, H. (2012). Structural basis for inhibition of xyloglucan-specific endo-β-1,4-glucanase (XEG) by XEG-protein inhibitor. *The Journal of Biological Chemistry*, *287*(22), 18710–18716. https://doi.org/10.1074/jbc.M112.350520

542 **FIGURE CAPTIONS**

543 **Fig. 1**. 7S globulin homology model. The model presents α helices in red and β chain in
544 blue.

545 **Fig. 2.** Ramachandran plot of the 7S globulin homology model.

546 **Fig. 3.** A) More favorable 3D conformation of malvidin 3-O-glucoside (Mv3glc) inside
547 of 7S *Vitis vinifera* globulin. B) 2D illustration of malvidin-3-O-glucoside (Mv3glc)
548 interacting with 7S globulin residues showing the major interactions. C) 3D illustration
549 of the flavylium cation orientation towards to the hydrophobic region of 7S *Vitis*
550 *vinifera* globulin.

551 **Fig. 4**. 7S globulin molecular dynamics simulations. RMSD of 7S globulin backbone
552 coupled with malvidin-3-O-glucoside (Mv3glc).

553