Development of a QSAR model to predict protein-flavor binding in protein-rich food systems

Food Chemistry

Barallat-Pérez, Cristina; Hollebrands, Boudewijn; Janssen, Hans Gerd; Martins, Sara; Fogliano, Vincenzo et al

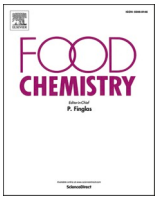
https://doi.org/10.1016/j.foodchem.2024.142268

Food Chemistry 467 (2025) 142268

Contents lists available at ScienceDirect

Food Chemistry

journal homepage: www.elsevier.com/locate/foodchem



Development of a QSAR model to predict protein-flavor binding in protein-rich food systems

Cristina Barallat-Pérez [a,*,1], Boudewijn Hollebrands [b,c,**,1], Hans-Gerd Janssen [b,c], Sara Martins [d], Vincenzo Fogliano [a,***], Jos Hageman [e], Teresa Oliviero [a]

[a] Department of Agrotechnology and Food Sciences, Wageningen, The Netherlands
[b] Unilever Foods Innovation Centre-Hive, Wageningen, The Netherlands
[c] Wageningen University & Research, Laboratory of Organic Chemistry, Wageningen, The Netherlands
[d] AFB International EU, Oss, The Netherlands
[e] Biometris, Applied Statistics, Wageningen University & Research, Wageningen, The Netherlands.

ARTICLE INFO

Keywords:
Commercial plant-based proteins
Flavor compounds
Protein-flavor binding
Quantitative structure-activity relationship
Prediction
Physicochemical properties

ABSTRACT

Protein-flavor binding is a common challenge in food formulation. Prediction models provide a time-, resource-, and cost-efficient way to investigate how the structural and physicochemical properties of flavor compounds affect this binding mechanism. This study presents a Quantitative Structure-Activity Relationship model derived from five commercial plant-based proteins and thirty-three flavor compounds. The results showed that protein-flavor binding is primarily influenced by the structure and physicochemical properties of the flavor compound, with the protein source having a minor contribution. In addition to hydrophobicity, topological, electronic, and geometrical descriptors significantly contribute to the observed protein-flavor binding. The Random Forest model demonstrated a strong correlation between predicted and experimental values ($Q^2 = 0.93$) and a high predictive ability for a validation set of flavors and proteins not previously used ($Q^2 = 0.88$). The prediction model developed holds promise for customizing flavor combinations and streamlining product design, thereby, optimizing efficiency while reducing the risk of flavor overdose.

1. Introduction

The effort of a balanced global food system led to a shift toward more sustainable protein sources, with a particular emphasis on plant-based alternatives (Schreuders et al., 2019). The most common animal replacements are plant proteins (PP) derived from soybeans (*Glycine* max) and peas (*Pisum sativum* L.) due to their excellent techno-functional properties such as water-holding, gelling, fat-absorbing, and emulsifying capacities (Kyriakopoulou et al., 2019). The need to find alternative protein sources to feed the global population and develop protein-rich protein food products (> 20 % protein content) boosted the use of a more diverse offer of pulses, where lupins (*Lupinus angustifolius* L.), faba beans (*Vicia faba* L.), and lentils (*Lens culinaris* L.) have recently gained significant attention (Kyriakopoulou et al., 2019).

Unfortunately, plant-derived proteins present undesirable off-flavors, which can negatively impact consumer acceptance (Wang et al., 2024; Xiang et al., 2023). Actions to improve the food flavor profile include adding flavorings (*i.e.,* flavor compounds). In nature, flavor compounds encompass various chemical classes like aldehydes, ketones, esters, alcohols, and terpenes, each with unique molecular structures and physicochemical properties. When added to protein-based food matrices, flavor compounds interact with- and may bind to proteins, reducing flavor perception. These interactions can be reversible or non-reversible (Anantharamkrishnan et al., 2020; Wongprasert et al., 2024), influenced by structural, thermodynamic, and physicochemical characteristics, including unsaturation, spatial configuration,

* Corresponding author at: Department of Agrotechnology and Food Sciences, Wageningen, The Netherlands.
** Corresponding author at: Unilever Foods Innovation Centre-Hive, Wageningen, The Netherlands.
*** Corresponding author at: Department of Agrotechnology and Food Sciences, Wageningen, The Netherlands.
E-mail addresses: cristina1.barallatperez@wur.nl (C. Barallat-Pérez), boudewijn.hollebrands@unilever.com (B. Hollebrands), hans-gerd.janssen@unilever.com (H.-G. Janssen), smartins@afbinternational.com (S. Martins), vincenzo.fogliano@wur.nl (V. Fogliano), jos.hageman@wur.nl (J. Hageman), teresa.oliviero@wur.nl (T. Oliviero).
[1] These authors equally contributed as the first author of the Manuscript.

https://doi.org/10.1016/j.foodchem.2024.142268
Received 12 July 2024; Received in revised form 19 November 2024; Accepted 24 November 2024
Available online 29 November 2024

alkyl chain type, functional group position, chain length, hydrophobicity, and water solubility (Ammari & Schroen, 2018; Guo et al., 2024; Li et al., 2024; Semenova et al., 2002; Wei et al., 2024).

Analytical measurements such as equilibrium dialysis (Damodaran & Kinsella, 1981), static headspace gas chromatography–mass spectrometry (GC–MS) (Wang & Arntfield, 2015), high-performance liquid chromatography (Li et al., 2000), solid phase microextraction (Gkionakis et al., 2007), and atmospheric pressure chemical ionization-mass spectrometry (Viry et al., 2018) have been used to quantify flavor binding. Although no single method is capable of yielding a complete, quantitative picture of the flavor binding phenomenon, GC–MS is the most widely employed technique for studying binding (Reineccius, 2010). Over the past few decades, mathematical methods and prediction models have been implemented as a powerful and complementary approach to GC–MS studies for quantifying binding interactions (Temthawee et al., 2020; Wongprasert et al., 2024). In practice, prediction models result in time-, resource-, and cost-efficient methods compared to conventional experimental laboratory work. Some examples are models based on best-fit partial least-squares regression (Tan & Siebert, 2008) and computational tools. Insightful approaches are the development of models based on Quantitative Structure-Activity Relationships (QSAR) or molecular docking (Bi et al., 2022; Tromelin & Guichard, 2004). Based on the fact that both the flavor as well as the protein's structural and physicochemical features are key in the binding phenomenon, QSAR models seem the optimal tool to model flavor partitioning in complex protein solutions (Tromelin & Guichard, 2004).

Up to now, the application of these prediction models has primarily focused on a narrow and specific range of flavor compounds (*e.g.,* aldehydes) (Snel et al., 2023). There is limited understanding regarding a broader and more varied flavor dataset. Similarly, most of these prediction models have investigated a limited selection of food ingredients, and thus, the use of commercial food proteins remains relatively unexplored. The food industry employs protein isolates and concentrates as meat and dairy replacers for practical reasons. Protein isolates vary in composition and consist of multiple, non-uniform types of protein fractions (Sadeghi et al., 2023). Furthermore, working with protein isolates offers a more practical approach to daily food applications.

In this study, it is hypothesized that a modeling approach could predict flavor binding in commercial plant protein-based model systems for diverse flavor compounds and reveal the key physicochemical and configurational properties of the flavor compounds, determining the binding mechanism. For this purpose, five different PP (soy protein isolate (SPI), pea protein isolate (PPI), lupin protein isolate (LPI), faba protein isolate (FPI), and lentil protein concentrate (LPC)) were tested with thirty-three flavor compounds belonging to seven chemical classes (*i.e.,* aldehydes, ketones, alcohols, lactones, pyrazines, furans, and sulfur-type compounds). Besides their significant role in flavoring applications, offering a range of sweet, fruity, floral, smoky, citrus, and fresh notes within the food and beverage industry, the selection criteria for these flavor compounds included both their physicochemical properties and structural configurations, highlighting similarities and differences, including the number of double bonds, spatial configuration, type and position of the functional group, chain length, and hydrophobicity. While studying a single protein's structure is key to understanding its flavor-binding role, it may not translate well to practical food applications. Relying solely on isolated protein fractions to meet texture and appearance requirements seems impractical. Therefore, commercial SPI and PPI were chosen due to their widespread acceptance as alternatives to meat and dairy products in the food industry, following the rationale behind previous work from our group (Barallat-Pérez et al., 2023). Additionally, the need to find alternative protein sources addressing the global food demand has driven interest in a broader variety of pulses such as FPI, LPI, and LPC. Furthermore, an independent set of five flavor compounds (*p*-anisaldehyde, ethyl octanoate, methyl salicylate, 3-methyl-2,4-nonadione, and *delta*-dodecalactone) was tested in combination with Whey Protein Isolate (WPI), and Bovine Serum Albumin (BSA) as controls to verify the accuracy of the model. The experimental assessment of protein-flavor binding was done by using static headspace (HS) GC–MS. A QSAR model was trained and validated to develop a relationship between the experimental determined protein-flavor binding and the physicochemical- and configuration properties. This model could assist food developers in crafting customized flavor profiles by leveraging the most important physicochemical and configurational properties, thereby reducing the need for excessive flavoring addition in (plant-based) product formulations. The distinctiveness of this work is based on the utilization of commercial food proteins that are well-suited for real food applications, offering a more realistic approach to predicting protein-flavor binding.

2. Materials and methods

2.1. Flavor compounds

A summary of the chosen flavors, their Canonical Simplified Molecular Input Line Entry System (SMILES) codes, and molecular and physicochemical properties (*i.e.,* chemical structure, LogP, molecular weight, vapor pressure, solubility, and boiling point) can be found in Supplementay Table S1 (Kim et al., 2023; Wishart et al., 2022; The Good Scents Company Information System, 2021). Unsaturation, spatial configuration, alkyl chain type, position of the functional group, chain length, and hydrophobicity were the selection criteria for the flavor compounds to be included in the study (Barallat-Pérez et al., 2023; Bi et al., 2022; Guo et al., 2024; Li et al., 2024; Wei et al., 2024).

Hexanal, heptanal, *trans*-2-heptenal, 4-*cis*-heptenal, *trans-trans*-2,4-heptadienal, octanal, *trans*-2-octenal, nonanal, *trans*-2-nonenal, *trans*-2-*cis*-6-nonadienal, *trans-trans*-2,6-nonadienal, 2,4-dimethylbenzaldehyde, *trans*-decenal, 2,3-butadione, 2-hexanone, 2-heptanone, 2-octanone, 6-methylheptan-2,4-dione, 2-nonanone, 1-pentanol, 1-penten-3-ol, 1-heptanol, 2-octanol, 1-octen-3-ol, creosol, 1-nonanol, linalool, dimethyl disulfide, 2,5-dimethylpyrazine, methyl propyl disulfide, β-ionone, α-ionone, and 2-pentylfuran, were purchased from Sigma-Aldrich (Zwijndrecht, the Netherlands) and all had a purity of ≥95 %.

2.2. Protein materials

Commercial plant proteins were provided by different suppliers. Soy Protein Isolate SUPRO® XT219D IP (SPI) was obtained from Solae (St. Louis, Missouri, USA); Pea Protein Isolate S85F (PPI) was purchased from Roquette (Lestrem, France); Faba Protein Isolate 90-C-EU (FPI) was acquired from AGT (Waalwijk, The Netherlands); Lupin Protein Isolate 10,600 (LPI) was purchased from ProLupin (Grimmen, Germany), and Lentil Protein Concentrate VITESSENCE® Pulse 2550 37403F00 (LPC) was obtained from Ingredion (Westchester, Illinois, USA). The two commercial proteins used to verify the model were: Whey Protein Isolate BiPro® (WPI) which was supplied by Davisco International (Le Sueur, Minnesota, USA), and Bovine Serum Albumin Purified Protein was purchased from Sigma-Aldrich (>98 %) (BSA). Typically, commercial proteins undergo multiple processing steps aimed at purifying and optimizing functional properties. Protein content varied from 51.9 %–97.6 %.

Supplementay Table S2 shows manufacturer-specified values such as fat, carbohydrate, and protein content for all PP, WPI, and BSA. Proteins were selected based on their nutritional composition, such as low-fat and high-protein content and prevalence in plant-based food substitutes (*e.g.,* high-protein-based beverages). To minimize the potential for variation in the results, the protein batches were kept in the dark, in sealed bags, and stored in a controlled environment with a cool temperature of 10–15 °C and low humidity.

2.3. Other chemicals

$Na_2HPO_4 \bullet 7H_2O$, $NaH_2PO_4 \bullet 2H_2O$, Na_2HPO_4, 8-anilino-1-

naphthalenesulfonate, chloroform (99.8 %), and methanol were analytical grade and purchased from Sigma-Aldrich. Pierce™ BCA ($HO_2CC_9H_5N_2$) assay kits were acquired from Thermo Fisher Scientific Inc., (Waltham, Massachusetts, USA) and contained albumin standard ampules (2 mg/mL, 10 × 1 mL containing bovine serum albumin in 0.9 % saline and 0.05 % sodium azide), and two BCA™ (bicinchoninic acid) reagents: A) Na_2CO_3, $NaHCO_3$, $(HO_2CC_9H_5N)_2$ and $C_4H_4Na_2O_6$ in 0.1 M NaOH; B) 4 % $CuSO_4 \bullet 5H_2O$ (25 mL).

2.4. Preparation of food flavor stock solutions

The chosen flavor compounds were individually prepared as described by Barallat-Pérez et al., 2023. To overcome the practical difficulties of preparing accurate flavor stock solutions and dealing with volatility (evaporation) and solubility challenges, additional precautions were taken. Individual flavor stock solutions were prepared in 100 mL sodium phosphate buffer solution (pH 7.0, 50 mM) at an initial concentration of 10 mg/L (Sigma-Aldrich) using amber glass bottles (Pyrex®, Thermo Fisher Scientific Inc.) closed with screw caps. The flavor stock solutions were subjected to an ultrasonic water bath treatment (25 kHz, ultrasonic time 100 %) (Elma Schmidbauer GmbH, Germany) for a duration of 1 h at a temperature of 20 °C to achieve homogenization of the solution. The stock solutions were repeatedly prepared in triplicate and stored under refrigeration conditions (3–5 °C).

2.5. Preparation of food protein solutions

Food protein solutions were prepared following the protocol proposed by Wang & Arntfield, 2015 and adapted from Barallat-Pérez et al., 2023. Specifically, the selected proteins (SPI, PPI, FPI, LPI, LPC, BSA, and WPI) were individually prepared at an initial concentration of 2 (w/v) % in a sodium phosphate buffer (pH 7.0, 50 mM). To address the solubility challenge known in PP, the proteins were gradually added to the solution while stirring, allowing their hydration in water. Subsequently, they were subjected to vortexing for a duration of 10–20 s (at 3200 rpm, Genie II, Genie™, Sigma-Aldrich) and subsequently transferred to an ultrasonic water bath (Elma Schmidbauer GmbH, Singen, Germany) for 20 min at a temperature of 20 °C, to break down protein clusters and ensure homogenization of the solutions. Mild heating (20 °C) aided in dissolving the proteins. Subsequently, the protein solutions underwent multiple rounds of vortexing (3200 rpm) lasting between 10 and 20 s each to guarantee a uniform mixture distribution. Lastly, visual verification ensured that no clumps remained.

2.6. preparation of the gas chromatography-mass spectrometry samples

Gas chromatography-mass spectrometry samples were prepared following the protocol proposed by Wang & Arntfield, 2015 and adapted from Barallat-Pérez et al., 2023. A 20 mL GC–MS vial was utilized to add 1 mL of the 2 (w/v) % solution of each protein, which was subsequently followed by adding 1 mL of flavor stock solution. Consequently, a protein solution with a final concentration of 1 (w/v) % and 5 mg/L flavor concentration was obtained. Subsequently, the vials were closed with screw caps (19 mm silicone PTFE SUPELCO®) and placed in a water bath shaker (SW22, Julabo GmbH, Seelbach, Germany) at 30 °C, 125 rpm for 3 h before headspace analysis (HS). The samples were prepared and analyzed in triplicate.

2.7. Binding measurement and calculation

Protein-flavor binding was measured using headspace GC–MS analysis utilizing an Agilent-7890 A GC instrument coupled with an Agilent 5975C triple-axis detector MS (Agilent, Amstelveen, the Netherlands). The GC operated in split mode at a 1:10 split ratio with a split flow of 8 mL/min. The samples underwent incubation and agitation for 14 min at a temperature of 40 °C, following a protocol established by Wang & Arntfield, 2015 and accordingly modified by Barallat-Pérez et al., 2023. Subsequently, 1 mL of the sample's HS was introduced into the system. The experiment utilized a DB-WAX column (20 m*180 μm*0.3 μm), operated at a constant flow of 0.8 mL/min. Column temperature was programmed at a rate of 40 °C/min to 240 °C. The mass spectrometer operated at 70 eV electron ionization and a mass range spanning from 35 to 200 Da. MassHunter Quantitative Analysis software (MSD ChemStation F.01.03.2357) was used for the quantitation of the flavor. Furthermore, the NIST Mass Spectrometry Library (InChI Library v.105) was employed to provide chemical and physical data on the chosen flavor compounds. Flavor compounds were studied individually to prevent potential flavor-flavor competition for protein binding sites. A quantitative measure for flavor binding to the proteins was obtained utilizing the Eq. (1) presented by Wang & Arntfield, 2015.

$$Binding\ (\%) = \left(1 - \frac{HS\,1 - HS\,2}{HS\,3}\right) x\,100 \quad (1)$$

HS_1 represents the (headspace) abundance, which refers to the peak response of the flavored aqueous protein solution. HS_2 indicates the abundance in the headspace when the flavor is absent, just the protein solution. Meanwhile, HS_3 reflects the abundance in the headspace in the absence of protein, therefore, just the given response from the flavor compound. The main ions and retention times of each flavor were determined using each flavor compound accordingly.

2.8. QSAR modeling

A graphical outline of the different steps followed for the QSAR modeling is shown in Fig. 1. The experimental binding values (Supplementay Table S3) were used as response variables to describe protein-flavor binding, and a set of descriptors was used as explanatory variables in a QSAR model. The descriptors were collected from the literature, experimentally determined, or calculated using the SMILES code of the flavor compounds. A Boruta algorithm was applied to select the important descriptors for the Random Forest (RF) regression model (Kursa et al., 2010). The effectiveness of the prediction model was evaluated by both a "Leave-One-Out Cross-validation" (LOOCV) and by assessment of an external validation set.

To assess the significance of each descriptor for the RF regression model, the importance value of each descriptor was determined and compared with randomized descriptor values following the Boruta algorithm procedure (Kursa et al., 2010). The procedure was applied 128 times, and in each iteration, the dataset (Supplementay Table S3 found in Supplementary Material) was randomly divided into training and test sets using a 4:1 ratio. The descriptor importance values for both original- and randomized descriptors were determined using the VarImp function from the R-package Caret (Kuhn, 2008). Then the descriptors were only considered for the model that exhibited a determined importance value significantly higher than the maximum observed importance value among the randomized descriptors. Response plots (Fig. 3B) were generated to visualize the relationship between individual descriptors and protein-flavor binding (response). Using 10-fold cross-validation, predictions were made by varying the values of the selected descriptor from its observed minimum to maximum across 100 evenly spaced sampling points, while keeping the values of all other descriptors fixed at their calculated averages.

2.8.1. Molecular descriptors

Molecular descriptors included flavor and protein features. For the selected flavor compounds, SMILES were obtained from the PubChem database (Kim et al., 2023). SMILES is a standardized notation for representing chemical structures in a computer-friendly format. It uses a character string to uniquely denote each structure, allowing for easy exchange and manipulation of chemical information. The list of SMILES codes was submitted to the online chemical database (OCHEM) (Sushko

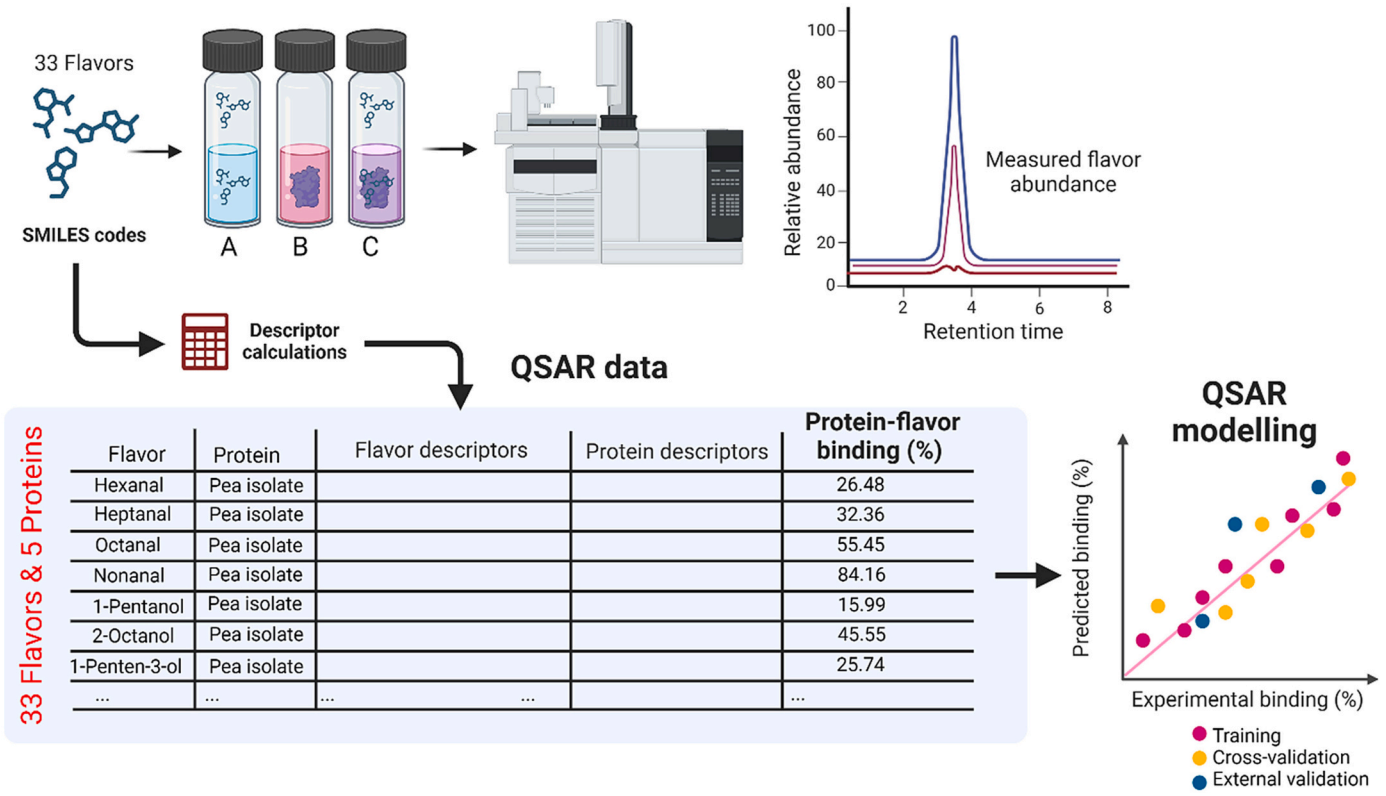


Fig. 1. Schematic overview of the protein-flavor prediction model. GC–MS measurements experimentally determined the binding between thirty-three flavor compounds and five commercial plant-based proteins, and a set of molecular descriptors was selected, calculated, and measured for QSAR predictive modeling.

et al., 2011) and the molecular structures were energy-minimized using BALLOON optimization (Vainio & Johnson, 2007). Supplementay Table S4 provides a full definition of the descriptors according to the OCHEM CDK manual. A set of in total 309 2D and 3D molecular descriptors was calculated using the descriptors calculator tool (The Chemistry Development Kit version 2.8) (Steinbeck et al., 2003). The LogP, vapor pressure, solubility, and topological polar surface area values of the flavors were obtained from the PubChem database.

The selection of protein descriptors used in this study was based on experimental laboratory analysis, prior literature and expert knowledge, and open-access availability. The protein descriptors included surface hydrophobicity, protein solubility, particle size, zeta potential, polydispersity index, isoelectric point (pI), and nutritional information.

The surface hydrophobicity of PP was experimentally determined following the approach of Li-Chan et al. (Li-Chan et al., 1984). Likewise, the protein solubility was experimentally determined using the BCA assay during the surface hydrophobicity experiments (pH 7.0). The BCA assay was used to determine the final concentration of soluble protein. The solubility was calculated using the following Eq. (2):

$$Solubility\ (\%) = \left(\frac{C\ protein,s}{C\ product^{*}W\ protein}\right) x\ 100 \qquad (2)$$

With *Cprotein,s*, being the final soluble protein concentration determined via the BCA assay, *Cproduct* the concentration of isolate in the solution before centrifuging, and *Wprotein* the mass fraction of protein in the PP.

Particle size, zeta potential (surface charge), and polydispersity (heterogeneity) index were measured using a Zetasizer Ultra (Malvern Instruments Ltd., Worcestershire, UK) (Kew et al., 2021) using quartz disposable cuvettes (Hellma, Müllheim, Germany). Protein solutions were first diluted to a concentration of 0.1 wt% and subsequently subjected to filtration through a 0.22 μm syringe filter (PTFE Syringe filters, PerkinElmer, Shelton, Connecticut, USA) for particle size measurement *via* dynamic light scattering.

The refractive index of the protein solution was set at 1.5 with an absorption of 0.001. The samples were equilibrated for two minutes at 25 °C and analyzed using backscattering technology at a detection angle of 173°. The measurements were performed in triplicate.

The pI values of the PP were obtained from literature. Values ranged from 4.0 to 5.0 (Lee et al., 2021; Shrestha et al., 2021; Tiong et al., 2024; Verfaillie et al., 2023).

2.8.2. Data preprocessing

Several preprocessing steps were performed before model development. First, 110 non-varying molecular descriptors were removed from the total of 328 descriptors in the dataset. The experimentally determined protein-flavor binding constant was used as the dependent variable. All training data was centered and scaled to unit variance before the training of the models. RF was used for training the models, unless otherwise stated, and was implemented in RStudio (version 2021.09.0 build 351, Boston, Massachusetts, USA) using the rf function from the R-package Caret (version 6.0–94) (Kuhn, 2008). The hyperparameters were optimized by a grid search and the final optimized model had a cost and sigma value of 14 and 0.01, respectively.

2.8.3. Validation of the prediction model

The model was validated through LOOCV, using only the selected descriptors outlined above. To avoid introducing validation bias, all data points related to a single flavor compound in combination with each protein were deliberately excluded during each cross-validation repetition. This approach ensured that each flavor compound was systematically left out during cross-validation. In each cross-validation iteration, the remaining dataset was used for training and consequently used to predict the values that were excluded. Subsequently, the values (Q^2) were computed by comparing the predicted values with the observed ones. Q^2 represents a measure of predictive accuracy (Andini et al.,

2021; Hageman et al., 2017).

2.8.4. External validation

For external validation, the protein-flavor binding of a new set of flavor compounds (*p*-anisaldehyde, ethyl octanoate, methyl salicylate, 3-methyl-2,4-nonadione, and *delta*-dodecalactone) and commercial proteins (WPI and BSA), was experimentally determined and predicted using ensemble modeling. The physicochemical properties and spatial configuration of the selected flavor molecules can be found in Supplementay Table S5 (The Good Scents Company Information System, 2021). To predict the binding values, the flavor and protein descriptors were collected using the procedures previously described (see section 2.8.1). The protein-flavor binding values of the validation compounds were predicted (Supplementay Table S6) using an ensemble of the models employed during LOOCV. The thirty-three models were trained on distinct training sets and were used to predict the protein-flavor binding of each validation compound. This approach resulted in thirty-three prediction values for each protein-flavor combination. These predictions were used to calculate average protein-flavor binding values and intervals indicating variability from the validation procedure.

3. Results and discussion

3.1. Protein-flavor binding mechanism: An overview

The binding mechanism of flavor compounds to commercial plant-based proteins is shown in Fig. 2. Fig. 2A provides a visual and selective representation of one of these protein-flavor interactions (*e.g.,* lupin +2-nonanone) to comprehend the experimentally determined binding mechanism better. In flavored protein-containing samples (Fig. 2A), relative headspace abundance is significantly diminished compared to protein-free samples due to the protein-flavor binding mechanism.

Fig. 2B shows a heat map reflecting the extent of protein-flavor interaction between the thirty-three flavor compounds and the five commercial plant-based proteins. Notably, binding responses widely

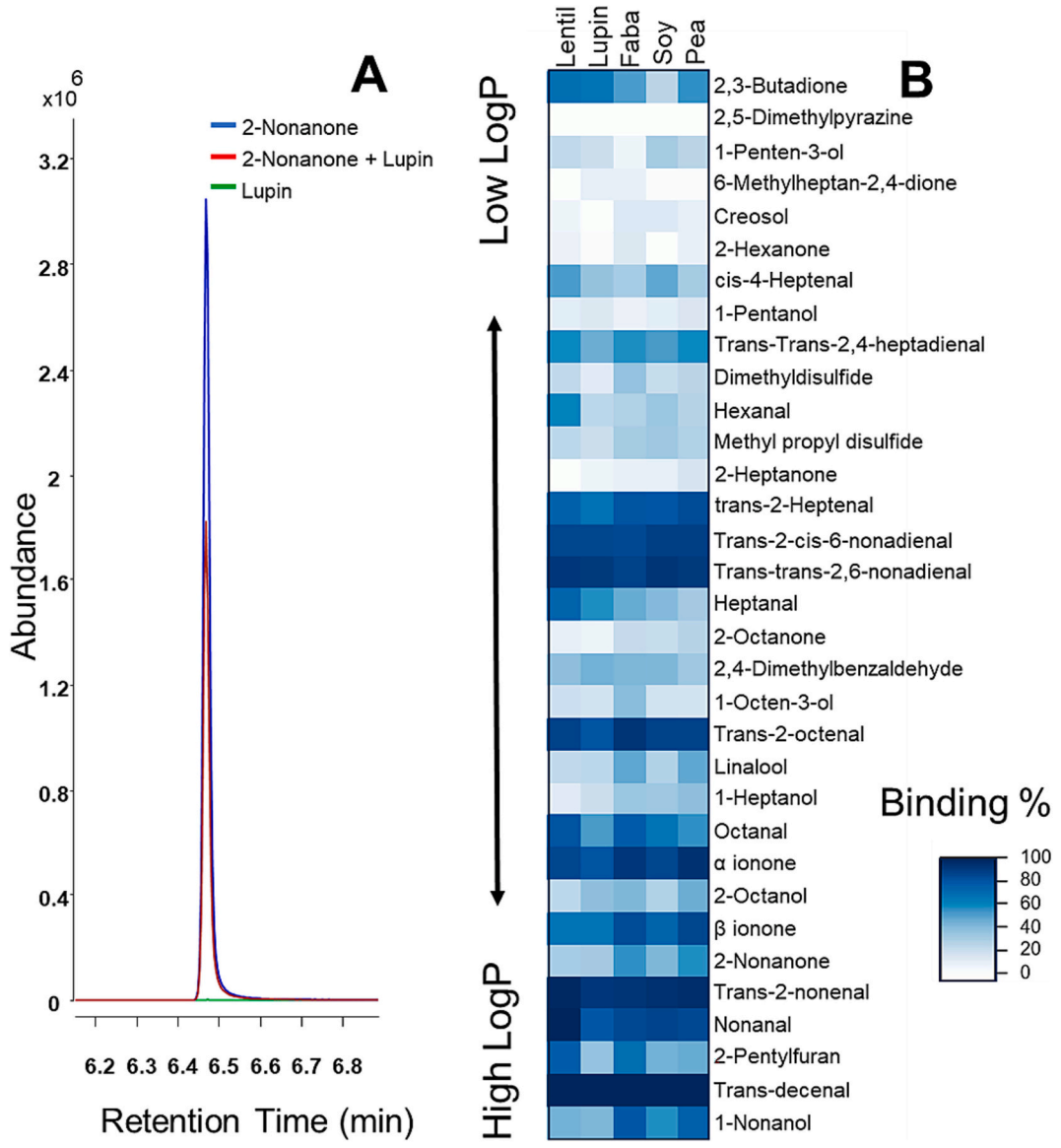


Fig. 2. *(A) Measured abundance of 2-nonanone with and without lupin protein (LPI); (B) Correlation plot shows an overview of the experimentally determined protein-flavor binding values of five commercial plant-based proteins (rows) and thirty-three flavor compounds (columns) using Eq. (1). The flavors are listed in increasing order of their octanol-water partitioning coefficient (LogP). The color intensity scales varied from white to dark blue indicating the level of protein-flavor binding.* (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

vary across flavor compounds and, to a lesser extent, between protein sources. The experimentally determined binding values can be found in Supplementay Table S3.

As expected, binding is, to some extent, influenced by the hydrophobicity of the flavor compound (Fig. 2B). Flavors with high binding values are visualized by the dark blue color and are mainly grouped at the bottom of the correlation plot where the flavors are arranged by their LogP values, which increase with the chain length (Bi et al., 2022; Su et al., 2021). Closer observation of the data shows binding <1 % for the rather hydrophilic flavor compound 2,5-dimethylpyrazine (Supplementay Table S3) (LogP = 0.6). Other hydrophilic flavor compounds within the data set also showed limited binding. Moreover, these exhibited high variability in the binding levels across the different protein sources (top part of Fig. 2A). For example, the experimentally determined protein-flavor binding of 2-hexanone (LogP = 1.4) ranges from <1 % to 15 % (Supplementay Table S3), where the lowest value is for binding with LPI, and the highest value is for binding with FPI. Similarly, binding values for 1-penten-3-ol (LogP = 1.1) vary from 6 % to 32 % (Supplementay Table S3) where the lowest value is for binding with FPI, and the highest value is for binding with SPI. On the contrary, low variation in the experimental binding was observed for flavors with higher binding affinity, *e.g., trans*-2-nonenal (LogP = 3.1), where values ranged from 90 % to 97 % (Supplementay Table S3). An additional parameter to consider is the accuracy of the experimental data. Determining the protein-flavor binding accurately of hydrophilic flavors is more difficult than that of the more hydrophobic species. Hydrophilic flavors often show low protein-flavor binding affinities and large experimental errors. Viry et al. (Viry et al., 2018), Snel et al. (Snel et al., 2023), and Wei et al. (Wei et al., 2024) explained the low protein binding values of hydrophilic flavor compounds by a “pushing-out (size-exclusion) effect”. This phenomenon leads to the expulsion of small and hydrophilic flavor compounds from the solution, pushing them into the headspace (Snel et al., 2023; Viry et al., 2018). Wei et al., 2024 described this effect as a “salting-out effect”, where the protein in solution lowers the surface tension of the dispersion, enhancing flavor release to the headspace.

From the modeling perspective, the non-uniform variance of errors, known as heteroskedasticity, is a complication that may impact the reliability of the predicted data obtained from the QSAR models.

This work was designed with a strong focus on practical applications, recognizing that entirely flavor-free plant materials are not yet available. As a result, the binding capacity may be slightly compromised due to the naturally occurring flavors in the plant proteins. However, it offers a more realistic scenario for day-to-day food applications.

3.2. Key descriptors involved in the protein-flavor binding mechanism

Reflecting on the first hypothesis, a QSAR study was performed to determine the most important properties that participate in the protein-flavor binding. The RF model was selected to predict flavor binding to proteins because of its high accuracy, robustness, and low likelihood of overfitting data. For selection, only twenty-eight descriptors of the initial 328 descriptors consistently performed better than those of repeatedly randomized descriptor data. In essence, only twenty-eight descriptors are here related to protein-flavor binding, which is influenced by flavor structure and physicochemical properties. On the contrary, protein-related descriptors such as particle size, zeta potential, surface hydrophobicity, isoelectric point, and the residual fat content (up to 0.09 (w/v) % in solution), exhibited minimal influence on the protein-flavor binding mechanism.

From these twenty-eight candidate descriptors, twelve were selected for the prediction model. Their calculated importance scores are plotted in Fig. 3A. Descriptors significantly impacting protein-flavor binding include constitutional, geometrical, hybrid, electronic, and topological descriptor classes (Toppur & Jaims, 2021) (Fig. S1). A full definition of the descriptors according to the OCHEM CDK manual can be found in Appendix, Supplementary Material, Supplementay Table S4.

As shown also in Fig. 2, protein-flavor binding strongly depends on flavor hydrophobicity (Bi et al., 2022; Su et al., 2021). Hence, it is not surprising that the constitutional descriptors ALogP2, ALogP, and XLogP, each referring to hydrophobicity albeit slightly different, demonstrated the most pronounced importance scores within the protein-flavor model, as illustrated in Fig. 3A.

Besides constitutional, the hybrid and topological descriptors like ECCEN, BCUTc-1I, BCUTp-1I, and MDEC-22 also contribute to the protein-flavor binding model (Fig. 3A). These descriptors provide information about the proximity of atoms, distances between them, connectivity within a molecule, and overall molecular spatial relationships (Toppur & Jaims, 2021) and help to describe the flavor compounds' functional groups and the locations within the molecular structures. Results showed that flavor structure (location of the functional group, spatial configuration) and physicochemical properties (LogP) were primary contributors to protein-flavor binding being aligned with previous scientific studies (Damodaran & Kinsella, 1981; Guo et al., 2024; Kühn et al., 2006; Li et al., 2024; Semenova et al., 2002; Wei et al., 2024; Zhou & Cadwallader, 2006). For example, carbonyl group displacement along the molecule enhanced binding significantly in commercial protein isolate-based systems (Damodaran & Kinsella, 1981). The positioning of the keto group within the inner structure of the molecule leads to steric hindrance, causing a reduction in the interaction's free energy.

In addition, the descriptors WNSA-3, PNSA-3, and RNCS define the flavor compounds' surface area and partial charge (Supplementay Table S4) (Toppur & Jaims, 2021) and show here as well significant contributions to protein-flavor binding. For instance, the charge density of a carbonyl flavor compound is comparatively greater than that of an ester (Ayed et al., 2014). Consequently, carbonyl-type flavor compounds exhibit a higher degree of retention than esters.

The response plots (Fig. 3B) illustrate the contribution to protein-flavor binding (response contribution) when only the input values of the twelve selected descriptors for the model change. Three main patterns were observed: First, an inverse response between polarizability (BCUTp-1I) and partial charge (BCUTc-1I) is observed; Low polarizability or high partial charge seemed to result in higher binding values. Second, remarkably similar patterns are seen for ECCEN and MDEC-22, providing connectivity information and molecular distance between C, N, and O. Third, a comparable pattern was found between WNSA-3, PNSA-3, and RNCS, which relates to the charge distribution of the flavor compounds (Supplementay Table S4) (Toppur & Jaims, 2021).

Confirming the initial hypothesis, the results clearly indicate that the properties of the flavor compound—such as topology, geometry, and hydrophobicity—primarily dictate protein-flavor binding. Aldehydes exhibited the highest affinity, followed by sesquiterpenoids and alcohols, whereas ketones showed the weakest binding (Supplementay Table S3).

Under the tested conditions, the protein-related descriptors and protein sources (SPI, PPI, LPI, LPC, and FPI) showed here minimum impact on the model's performance (Fig. 3 and Fig. S1). These results are consistent with previously published data by Barallat-Pérez et al., 2023 and Snel et al., 2023 who studied flavor binding to commercial food proteins where flavor binding proved to be mainly dependent on the flavor molecule and its physicochemical properties such as spatial configuration, absence/presence of double bonds, and location of the functional group.

In light of the previous information and the additional data presented in Fig. 2 and Fig. 3A, it becomes evident that the interaction between proteins and flavors extends beyond hydrophobicity. Additional factors related to configuration properties and charge distribution also play a crucial role in the observed binding phenomenon, complementing each other in a way that should not be ignored.

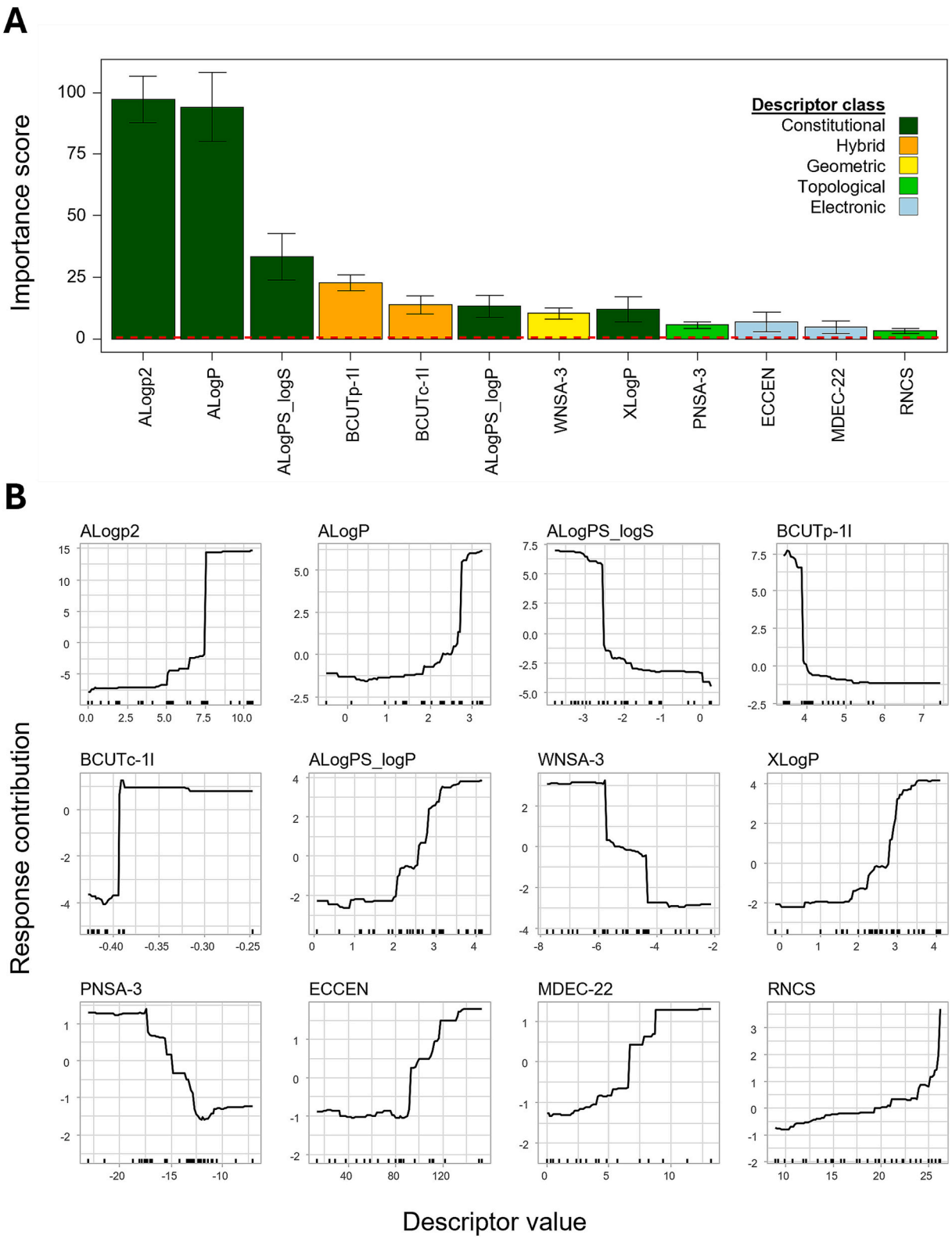


Fig. 3. Ranking, categorization, and response plots of selected descriptors for the QSAR model predicting protein-flavor binding. (A) High importance scores indicate a strong relationship between molecular descriptors (constitutional, hybrid, geometric, topological, and electronic) and protein-flavor binding. Randomization of the descriptor values and replication of the models ($n = 128$) was applied to determine which descriptors perform significantly better than the best result of randomized descriptor data (red line). The error bars indicate the 99 % confidence interval. (B) The response plots show the contribution to protein-flavor binding (Response contribution) when only the value (Descriptor value) changes of the selected descriptor. The rug marks show how descriptor values were distributed in the used dataset. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.3. Predicting protein-flavor binding

Reflecting on the second hypothesis, the prediction of the protein-flavor binding was conducted employing RF using the twelve selected descriptors. For this purpose, a "leave-one-flavor-out" cross-validation approach was used. Individual data points corresponding to a particular flavor within the dataset were methodically excluded one by one, and the model was further trained using the remaining data points. Subsequently, the performance of the model was assessed by making predictions for the excluded data points. This process was repeated for each data point in the dataset. Fig. 4 shows the correlation between the predicted and experimentally determined protein-flavor binding values. The $Q^2 = 0.93$ indicated a strong correlation between predicted and experimental values (Fig. 4). A more detailed overview can be found in Supplementary Material Fig. S2.

From Fig. 4 it is evident that the protein source does not strongly affect flavor binding. None of the protein sources showed significantly larger differences between the predicted and experimental values. These results are in line with the outcome of the descriptor ranking and selection illustrated in Fig. 3A (showing no significant contribution of protein-related descriptors) and with literature information (Barallat-Pérez et al., 2023; Snel et al., 2023).

3.4. Prediction of validation compounds

Training a QSAR model that can predict all possible combinations of proteins and flavors is not feasible. The number of flavor compounds in nature is enormous, thus, model restrictions must be validated and defined for the structural domain and response space. To prove the model's applicability, protein-flavor binding was predicted and compared to experimentally determined data for compounds not in the training and testing data set (Supplementay Table S1). The validation compounds, such as *p*-anisaldehyde, methyl salicylate, and 3-methyl-2,4-nonanedione, shared functional groups with our training set. Ethyl octanoate and *delta*-dodecalactone, often employed in food flavors, were also added to the data set. Supplementary Material, Supplementay Table S5, shows the physicochemical and structural properties of the compounds used in the validation.

The flavor binding values of the validation compounds were predicted by calculation of the twelve descriptor values for each flavor in the validation set. These calculated values were used as input variables for the previously trained model. The ensemble model is comprised of thirty-three different prediction models as a result of the LOOCV and was applied to predict the flavor binding of the validation compounds. For each model, the data of one flavor compound was left out and with the remaining data, the model was trained. Fig. 5 shows the predicted and experimental values of the validation compounds.

Methyl salicylate, 3-methy-2,4-nonadione, and *p*-anisaldehyde are more hydrophilic and volatile than ethyl octanoate and *delta*-dodecalactone (Supplementay Table S5 and S6) and show lower binding (Fig. 5). Additionally, methyl salicylate, 3-methy-2,4-nonadione, and *p*-anisaldehyde are characterized by a spherical-spatial configuration, while ethyl octanoate and *delta*-dodecalactone have a linear-shape structure. The steric effect from a spherical structure might hinder the binding to the available hydrophobic pockets (Anantharamkrishnan et al., 2020; Guo et al., 2024; Wongprasert et al., 2024).

Longer chain-length and linear molecules, such as ethyl octanoate and *delta*-dodecalactone, consistently displayed a pattern in their response plots (Fig. S3). This consistency indicated that they tend to have higher MDEC values, which provide information about atom connectivity along the molecule, closely related to higher binding affinity.

Despite the manufacturing and processing history that plant protein isolates and concentrates often go through and the possible remaining traces of fat and carbohydrates present, no large differences in flavor binding affinity were found between the pure protein (BSA) and the isolate (WPI) (fat levels <1, Supplementay Table S2) as seen in Fig. 5. Furthermore, given the current studied system and the selected validation compounds, the proposed prediction model showed high predictive ability ($Q^2 = 0.88$). Therefore, the obtained knowledge repeatedly confirms the minor role of the protein source on the extent of protein-flavor binding.

Using five verification flavor compounds *p*-anisaldehyde, ethyl octanoate, methyl salicylate, 3-methyl-2,4-nonadione, and *delta*-dodecalactone may not fully address all flavor chemical classes and physicochemical properties existing in nature. However, it builds an initial understanding of the protein-flavor binding mechanism and strengthens the potential of the current prediction model. Follow-up studies are suggested to expand to more diverse flavor compound datasets and more diverse compositional and experimental conditions to cover broader food applications.

4. Conclusions

The uniqueness of the present work lies in the use of a blend of

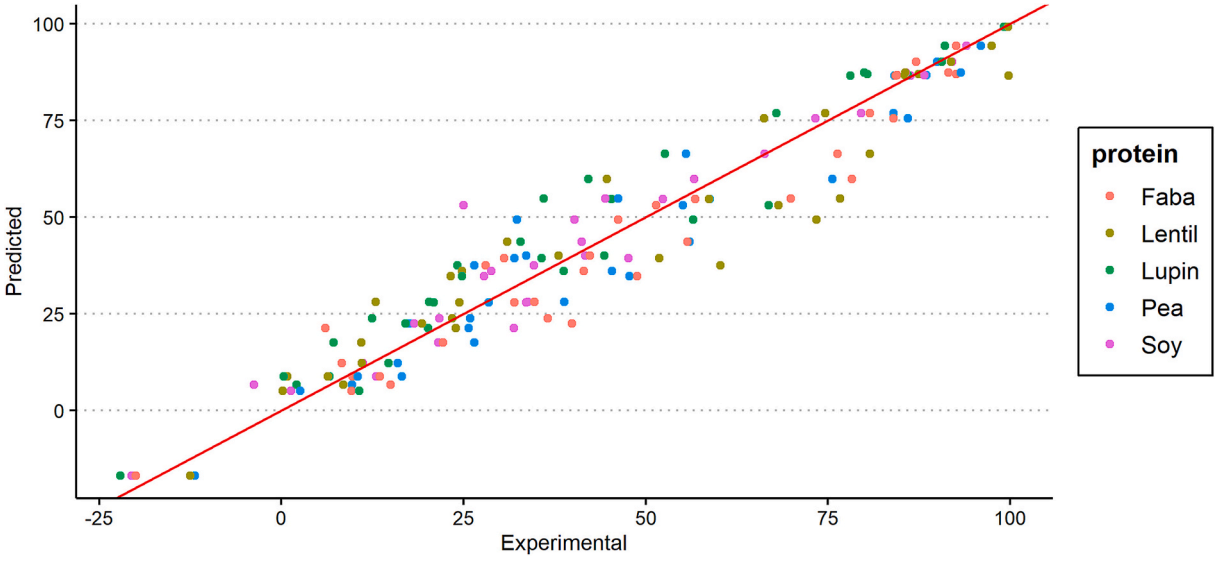


Fig. 4. Random Forest (RF) model depicting predicted *vs.* experimental protein-flavor binding values for the studied flavors and commercial protein isolates and concentrates.

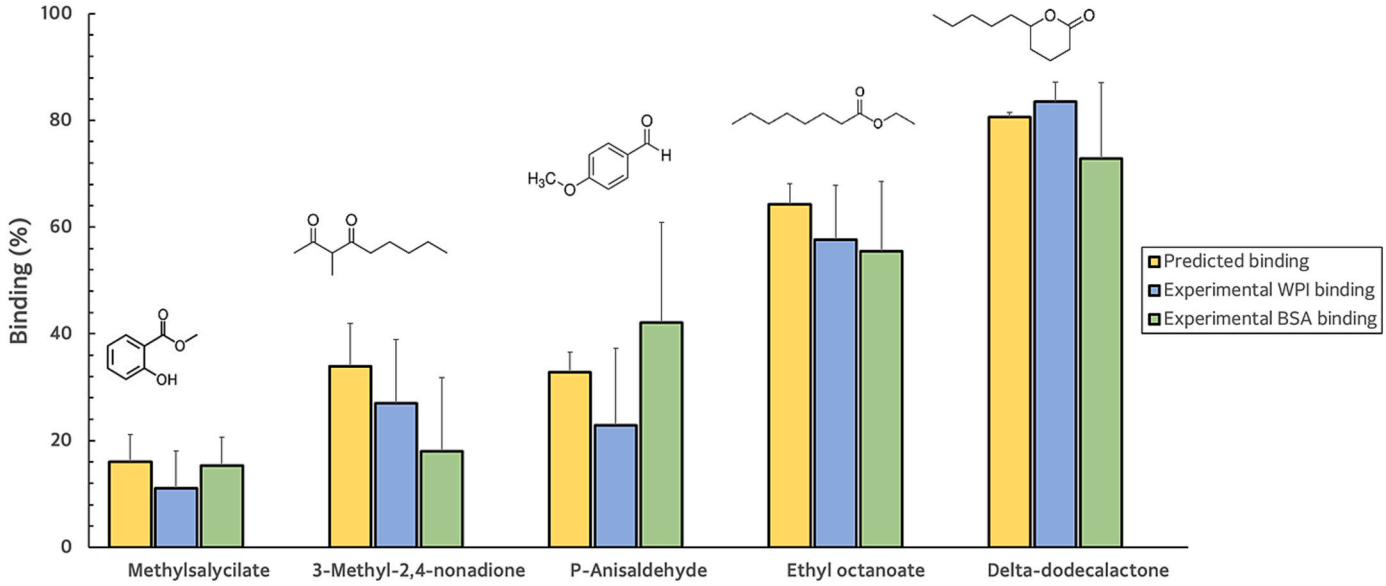


Fig. 5. Predicted *vs.* experimental binding data for p-anisaldehyde, ethyl octanoate, methyl salicylate, 3-methyl-2,4-nonadione, and delta-dodecalactone in Whey Protein Isolate (WPI), and Bovine Serum Albumin (BSA). Results are expressed as mean ± 2*stdev. Binding was calculated using Eq. (1).

different food protein fractions, suited for real food applications and thus, representing a practical approach to predict protein-flavor binding. A modeling approach was hypothesized to predict flavor binding in commercial protein-based model systems for various flavor compounds and to reveal the key physicochemical and configurational properties of the flavor compounds that determine the binding mechanism. The results unequivocally demonstrate a strong correlation between predicted and experimental values, as demonstrated by the Random Forest model highlighting that protein-flavor binding is primarily dictated by the flavor compound itself under the here researched conditions. Beyond hydrophobicity, topological, electronic, and geometrical descriptors complementarily contribute to the observed protein-flavor binding.

The obtained results have the potential to expand the current understanding of protein-flavor interactions serving as a first step toward developing time-, cost- and resource-efficient methods for predicting flavor binding in protein-rich systems and optimizing flavor formulas (less flavor dosing) based on their structure and physicochemical properties.

Although the applied model system in this study may not fully capture the complexity of real-world foods, the prediction model offers valuable preliminary insights. It provides a straightforward method for predicting flavor binding in protein-rich aqueous systems, initially focusing on single-component food systems using commercial protein isolates. Acknowledging the complexity of real food systems, which involve mixtures of flavors, sugar, salt, fat, and extensive processing conditions (heat, pH, ionic strength), further research is needed to explore the broader application of the presented model reflecting more practical food scenarios.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.foodchem.2024.142268.

CRediT authorship contribution statement

Cristina Barallat-Pérez: Writing – original draft, Methodology, Investigation, Conceptualization. **Boudewijn Hollebrands:** Writing – original draft, Visualization, Methodology, Investigation. **Hans-Gerd Janssen:** Writing – review & editing, Visualization, Supervision, Project administration. **Sara Martins:** Conceptualization. **Vincenzo Fogliano:** Writing – review & editing, Resources, Project administration, Funding acquisition. **Jos Hageman:** Writing – review & editing, Visualization, Supervision. **Teresa Oliviero:** Writing – review & editing, Visualization, Supervision, Project administration.

Declaration of generative AI and AI-assisted technologies in the writing process

During the preparation of this work, the author(s) used ChatGPT (version 3.5) to improve the readability and language of the Manuscript. After using this tool/service, the author(s) reviewed and edited the content as needed and take(s) full responsibility for the content of the publication.

Declaration of competing interest

Boudewijn Hollebrands and Hans-Gerd Janssen are employed by Unilever, a multinational company active in foods and home and personal care products. The authors report there are no competing interests to declare.

Data availability

Data will be made available on request.

Acknowledgements

The authors express their gratitude to Herrald Steenbergen and Oscar Dofferhoff for their assistance with the headspace measurements, and to Juliana Villasante for her help during the particle size measurements.

References

Ammari, A., & Schroen, K. (2018). Flavor retention and release from beverages: A kinetic and thermodynamic perspective. *Journal of Agricultural and Food Chemistry, 66*(38), 9869–9881. https://doi.org/10.1021/acs.jafc.8b04459

Anantharamkrishnan, V., Hoye, T., & Reineccius, G. A. (2020). Covalent adduct formation between flavor compounds of various functional group classes and the model protein β-Lactoglobulin. *Journal of Agricultural and Food Chemistry, 68*(23), 6395–6402. https://doi.org/10.1021/acs.jafc.0c01925

Andini, S., Araya-Cloutier, C., Lay, B., Vreeke, G., Hageman, J., & Vincken, J.-P. (2021). QSAR-based physicochemical properties of isothiocyanate antimicrobials against gram-negative and gram-positive bacteria. *LWT, 144*, Article 111222. https://doi.org/10.1016/j.lwt.2021.111222

Ayed, C., Lubbers, S., Andriot, I., Merabtine, Y., Guichard, E., & Tromelin, A. (2014). Impact of structural features of odorant molecules on their retention/release

behaviours in dairy and pectin gels. *Food Research International, 62*, 846–859. https://doi.org/10.1016/j.foodres.2014.04.050

Barallat-Pérez, C., Janssen, H.-G., Martins, S., Fogliano, V., & Oliviero, T. (2023). Unraveling the role of flavor structure and physicochemical properties in the binding phenomenon with commercial food protein isolates. *Journal of Agricultural and Food Chemistry, 71*(50), 20274–20284. https://doi.org/10.1021/acs.jafc.3c05991

Bi, S., Pan, X., Zhang, W., Ma, Z., Lao, F., Shen, Q., & Wu, J. (2022). Non-covalent interactions of selected flavors with pea protein: Role of molecular structure of flavor compounds. *Food Chemistry, 389*, Article 133044. https://doi.org/10.1016/j.foodchem.2022.133044

Damodaran, S., & Kinsella, J. E. (1981). Interaction of carbonyls with soy protein: Thermodynamic effects. *Journal of Agricultural and Food Chemistry, 29*(6), 1249–1253. https://doi.org/10.1021/jf00108a037

Gkionakis, G. A., David Anthony Taylor, K., Ahmad, J., & Heliopoulos, G. (2007). The binding of the flavour of lactones by soya protein, amino acids and casein. *International Journal of Food Science and Technology, 42*(2), 165–174. https://doi.org/10.1111/j.1365-2621.2006.01184.x

Guo, Y., Gong, Q., Sun, F., Cheng, T., Fan, Z., Huang, Z., Liu, J., Guo, Z., & Wang, Z. (2024). Interaction mechanism of pea proteins with selected pyrazine flavors: Differences in alkyl numbers and flavor concentration. *Food Hydrocolloids, 147*, Article 109314. https://doi.org/10.1016/j.foodhyd.2023.109314

Hageman, J. A., Engel, B., de Vos, R. C. H., Mumm, R., Hall, R. D., Jwanro, H., … van Eeuwijk, F. A. (2017). Robust and confident predictor selection in metabolomics. In S. Datta, & B. J. A. Mertens (Eds.), *Statistical analysis of proteomics, metabolomics, and Lipidomics data using mass spectrometry* (pp. 239–257). Springer International Publishing. https://doi.org/10.1007/978-3-319-45809-0_13.

Kew, B., Holmes, M., Stieger, M., & Sarkar, A. (2021). Oral tribology, adsorption and rheology of alternative food proteins. *Food Hydrocolloids, 116*, Article 106636. https://doi.org/10.1016/j.foodhyd.2021.106636

Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., … Bolton, E. E. (2023). PubChem 2023 update. *Nucleic Acids Research, 51*(D1), D1373–D1380. https://doi.org/10.1093/nar/gkac956

Kühn, J., Considine, T., & Singh, H. (2006). Interactions of Milk proteins and volatile flavor compounds: Implications in the development of protein foods. *Journal of Food Science, 71*(5). https://doi.org/10.1111/j.1750-3841.2006.00051.x

Kuhn, M. (2008). Building predictive models in R using the caret package. *Journal of Statistical Software, 28*(5), 1–26. https://doi.org/10.18637/jss.v028.i05

Kursa, M. B., Jankowski, A., & Rudnicki, W. R. (2010). Boruta–a system for feature selection. *Fundamenta Informaticae, 101*(4), 271–285. https://doi.org/10.3233/FI-2010-288

Kyriakopoulou, K., Dekkers, B., & Van Der Goot, A. J. (2019). Plant-based meat analogues. In Galanakis (Ed.), *Sustainable meat production and processing* (pp. 103–126). Elsevier. https://doi.org/10.1016/B978-0-12-814874-7.00006-7.

Lee, H. W., Lu, Y., Zhang, Y., Fu, C., & Huang, D. (2021). Physicochemical and functional properties of red lentil protein isolates from three origins at different pH. *Food Chemistry, 358*, Article 129749. https://doi.org/10.1016/j.foodchem.2021.129749

Li, Y., Xu, J., Sun, F., Guo, Y., Wang, D., Cheng, T., Xu, M., Wang, Z., & Guo, Z. (2024). Spectroscopy combined with spatiotemporal multiscale strategy to study the adsorption mechanism of soybean protein isolate with meat flavor compounds (furan): Differences in position and quantity of the methyl. *Food Chemistry, 451*, Article 139415. https://doi.org/10.1016/j.foodchem.2024.139415

Li, Z., Grün, I. U., & Fernando, L. N. (2000). Interaction of vanillin with soy and dairy proteins in aqueous model systems: A thermodynamic study. *Journal of Food Science, 65*(6), 997–1001. https://doi.org/10.1111/j.1365-2621.2000.tb09406.x

Li-Chan, E., Nakai, S., & Wood, D. F. (1984). Hydrophobicity and solubility of meat proteins and their relationship to emulsifying properties. *Journal of Food Science, 49*(2), 345–350. https://doi.org/10.1111/j.1365-2621.1984.tb12418.x

Reineccius, G. (2010). Instrumental methods of analysis. In *, 4. Proceedings of the Society for Analytical Chemistry* (pp. 229–265). https://doi.org/10.1002/9781444317770.ch9

Sadeghi, R., Colle, M., & Smith, B. (2023). Protein composition of pulses and their protein isolates from different sources and in different isolation pH values using a reverse phase high performance liquid chromatography method. *Food Chemistry, 409*, Article 135278. https://doi.org/10.1016/j.foodchem.2022.135278

Schreuders, F. K. G., Dekkers, B. L., Bodnár, I., Erni, P., Boom, R. M., & Van Der Goot, A. J. (2019). Comparing structuring potential of pea and soy protein with gluten for meat analogue preparation. *Journal of Food Engineering, 261*, 32–39. https://doi.org/10.1016/j.jfoodeng.2019.04.022

Semenova, M. G., Antipova, A. S., Belyakova, L. E., Polikarpov, Y. N., Wasserman, L. A., Misharina, T. A., … Golovnya, R. V. (2002). Binding of aroma compounds with legumin. III. Thermodynamics of competitive binding of aroma compounds with 11S globulin depending on the structure of aroma compounds. *Food Hydrocolloids, 16*(6), 573–584. https://doi.org/10.1016/S0268-005X(02)00019-X

Shrestha, S., Hag, L. V. T., Haritos, V. S., & Dhital, S. (2021). Lupin proteins: Structure, isolation and application. *Trends in Food Science and Technology, 116*, 928–939. https://doi.org/10.1016/j.tifs.2021.08.035

Snel, S. J. E., Pascu, M., Bodnár, I., Avison, S., Van Der Goot, A. J., & Beyrer, M. (2023). Flavor-protein interactions for four plant protein isolates and whey protein isolate with aldehydes. *LWT, 185*, Article 115177. https://doi.org/10.1016/j.lwt.2023.115177

Steinbeck, C., Han, Y., Kuhn, S., Horlacher, O., Luttmann, E., & Willighagen, E. (2003). The chemistry development kit (CDK): An open-source Java library for chemo- and bioinformatics. *Journal of Chemical Information and Computer Sciences, 43*(2), 493–500. https://doi.org/10.1021/ci025584y

Su, K., Brunet, M., Festring, D., Ayed, C., Foster, T., & Fisk, I. (2021). Flavour distribution and release from gelatine-starch matrices. *Food Hydrocolloids, 112*, Article 106273. https://doi.org/10.1016/j.foodhyd.2020.106273

Sushko, I., Novotarskyi, S., Körner, R., Pandey, A. K., Rupp, M., Teetz, W., … Tetko, I. V. (2011). Online chemical modeling environment (OCHEM): Web platform for data storage, model development and publishing of chemical information. *Journal of Computer-Aided Molecular Design, 25*(6), 533–554. https://doi.org/10.1007/s10822-011-9440-2

Tan, Y., & Siebert, K. J. (2008). Modeling bovine serum albumin binding of flavor compounds (alcohols, aldehydes, esters, and ketones) as a function of molecular properties. *Journal of Food Science, 73*(1). https://doi.org/10.1111/j.1750-3841.2007.00591.x

Temthawee, W., Panya, A., Cadwallader, K. R., & Suppavorasatit, I. (2020). Flavor binding property of coconut protein affected by protein-glutaminase: Vanillin-coconut protein model. *LWT, 130*, Article 109676. https://doi.org/10.1016/j.lwt.2020.109676

The Good Scents Company Information System. (2021). https://www.thegoodscentscompany.com/.

Tiong, A. Y. J., Crawford, S., Jones, N. C., McKinley, G. H., Batchelor, W., & Van 'T Hag, L.. (2024). Pea and soy protein isolate fractal gels: The role of protein composition, structure and solubility on their gelation behaviour. *Food Structure, 40*, Article 100374. https://doi.org/10.1016/j.foostr.2024.100374

Toppur, B., & Jaims, K. J. (2021). Determining the best set of molecular descriptors for a toxicity classification problem. *RAIRO - Operations Research, 55*(5), 2769–2783. https://doi.org/10.1051/ro/2021134

Tromelin, A., & Guichard, E. (2004). 2D- and 3D-QSAR models of interaction between flavor compounds and beta-Lactoglobulin using catalyst and Cerius2. *QSAR and Combinatorial Science, 23*(4), 214–233. https://doi.org/10.1002/qsar.200430859

Vainio, M. J., & Johnson, M. S. (2007). Generating conformer ensembles using a multiobjective genetic algorithm. *Journal of Chemical Information and Modeling, 47*(6), 2462–2474. https://doi.org/10.1021/ci6005646

Verfaillie, D., Janssen, F., Van Royen, G., & Wouters, A. G. B. (2023). A systematic study of the impact of the isoelectric precipitation process on the physical properties and protein composition of soy protein isolates. *Food Research International, 163*, Article 112177. https://doi.org/10.1016/j.foodres.2022.112177

Viry, O., Boom, R., Avison, S., Pascu, M., & Bodnár, I. (2018). A predictive model for flavor partitioning and protein-flavor interactions in fat-free dairy protein solutions. *Food Research International, 109*, 52–58. https://doi.org/10.1016/j.foodres.2018.04.013

Wang, K., & Arntfield, S. D. (2015). Binding of selected volatile flavour mixture to salt-extracted canola and pea proteins and effect of heat treatment on flavour binding. *Food Hydrocolloids, 43*, 410–417. https://doi.org/10.1016/j.foodhyd.2014.06.011

Wang, Y., Tuccillo, F., Niklander, K., Livi, G., Siitonen, A., Pöri, P., … Katina, K. (2024). Masking off-flavors of faba bean protein concentrate and extrudate: The role of in situ and in vitro produced dextran. *Food Hydrocolloids, 150*, Article 109692. https://doi.org/10.1016/j.foodhyd.2023.109692

Wei, M., Liu, X., Xie, P., Han, A., Lei, Y., Yang, X., Liu, Y., Zhang, S., & Sun, B. (2024). Investigating the interactions between selected heterocyclic flavor compounds and beef myofibrillar proteins using SPME-GC–MS, spectroscopic, and molecular docking approaches. *Journal of Molecular Liquids, 403*, Article 124878. https://doi.org/10.1016/j.molliq.2024.124878

Wishart, D. S., Guo, A., Oler, E., Wang, F., Anjum, A., Peters, H., … Gautam, V. (2022). HMDB 5.0: The human metabolome database for 2022. *Nucleic Acids Research, 50*(D1), D622–D631. https://doi.org/10.1093/nar/gkab1062

Wongprasert, T., Mathatheeranan, P., Chen, X., Vilaivan, T., Suriya, U., Rungrotmongkol, T., & Suppavorasatit, I. (2024). Molecular interactions by thermodynamic and computational molecular docking simulations of selected strawberry esters and pea protein isolate in an aqueous model system. *LWT, 198*, Article 115964. https://doi.org/10.1016/j.lwt.2024.115964

Xiang, L., Jiang, B., Xiong, Y. L., Zhou, L., Zhong, F., Zhang, R., … Xiao, Z. (2023). Beany flavor in pea protein: Recent advances in formation mechanism, analytical techniques and microbial fermentation mitigation strategies. *Food Bioscience, 56*, Article 103166. https://doi.org/10.1016/j.fbio.2023.103166

Zhou, Q., & Cadwallader, K. R. (2006). Effect of flavor compound chemical structure and environmental relative humidity on the binding of volatile flavor compounds to dehydrated soy protein isolates. *Journal of Agricultural and Food Chemistry, 54*(5), 1838–1843. https://doi.org/10.1021/jf052269d