Economic and Human Features for Energy and Environmental Indicators: A Tool to Assess Countries' Progress towards Sustainability

CHEMICAL ENGINEERING TRANSACTIONS
VOL. 102, 2023
A publication of AIDIC
The Italian Association of Chemical Engineering
Online at www.cetjournal.it
Guest Editors: Laura Piazza, Mauro Moresi, Francesco Donsì

ISBN 979-12-81206-01-4; **ISSN** 2283-9216
DOI: 10.3303/CET23102048

Machine Learning aided Molecular Modelling of Taste to Identify Food Fingerprints

Lorenzo Pallante[a], Marco Cannariato[a], Fosca Vezzulli[b], Marta Malavolta[c], Milena Lambri[b], Marco A. Deriu[a,*]

[a]Polito[BIO]MedLab, Department of Mechanical and Aerospace Engineering, Politecnico di Torino, Torino, Italy
[b]Department for Sustainable Food Process, Università Cattolica del Sacro Cuore, Piacenza, Italy
[c]Faculty of Computer and Information Science, University of Ljubljana, 1000 Ljubljana, Slovenia
marco.deriu@polito.it

Nature has developed fascinating mechanisms for selecting and monitoring nutrients through refined systems for food intake and uptake. One of the most important is the sense of taste. Taste is an emergent property involving a complex network of multilevel biological interactions beginning with the activation of specific protein receptors as a consequence of interaction with food molecules. In this context, crucial information about the mechanisms underlying the functioning of taste can be obtained by using molecular mechanistic modelling and machine learning tools borrowed from the field of drug design and the study of structural biology and protein biophysics. The ultimate goal is to develop predictive models capable of studying the intricate connection of molecular, sub-cellular and cellular phenomena underlying the complex biological mechanisms that regulate the relationships between food constituents and perceived taste. Artificial intelligence-driven digital tools for taste prediction and the study of molecular features of the interaction between food molecules and taste receptors have been recently developed by our group. Such tools are the operating engines of the decision support tool developed during the VIRTUOUS project (https://virtuoush2020.com).
In this work, these tools were used to generate molecular fingerprints of coffee starting from its chemical composition. Through methods that integrate molecular modelling techniques and machine learning, molecules extracted from coffee were characterized in terms of binding affinity, specificity, and selectivity toward bitter receptors. The targeting ability of coffee-extracted molecules for human TAS2Rs was studied with an atomistic resolution to obtain a virtual fingerprint that links the molecular structure of food ingredients with their bitter profile. The study fits within the digital transition vision that leverages modelling and computational approaches to develop decision-supporting tools for developing solutions in the areas of nutrition, health and the modern food industry.

1. Introduction

Taste is a multi-layered sensory experience, encompassing the recognition of flavours, which stem from the combination of stimuli from the olfactory, gustatory, and trigeminal systems. The sense of taste is crucial in regulating food consumption as it enables the evaluation of a food's nutritional content and safety, thus preventing the ingestion of hazardous or poisonous substances. (Roper, 2017). Sweet, umami, bitter, sour, and salty are the five basic taste sensations and each is linked to a specific bodily function. Despite the common association of bitter taste with unpleasant flavours and potentially harmful substances like spoiled food or toxins, not all bitter-tasting compounds, e.g. coffee, unsweetened cocoa, and untreated olives, are harmful or unappetizing. From a molecular point of view, bitter taste receptors belong to the taste 2 receptor family (TAS2Rs) of the G protein-coupled receptors (GPCRs) (Chandrashekar et al., 2000). Their structure includes a concise N-terminal located outside of the cell, a C-terminal situated inside the cell, and seven helices that cross the cell membrane (known as transmembrane helices or TMDs). These helices are linked by unstructured domains including three intracellular loops (ICLs) and three extracellular loops (ECLs) (Zhang et al., 2017). Interestingly, the most conserved domain among the bitter taste receptors and other GPCRs belonging to similar

Paper Received: 14 February 2023; Revised: 26 March 2023; Accepted: 7 July 2023
Please cite this article as: Pallante L., Cannariato M., Vezzulli F., Malavolta M., Lambri M., Deriu M.A., 2023, Machine Learning Aided Molecular Modelling of Taste to Identify Food Fingerprints, Chemical Engineering Transactions, 102, 283-288 DOI:10.3303/CET23102048

families is constituted by the TMD bundle: this structure forms the receptor structural core, participates in ligand recognition, and mediates the transduction of information (Di Pizio et al., 2016). Despite each bitter taste receptor only having a unique binding cleft for bitter compounds, the TAS2R receptor diversity permits the detection of an extensive range of bitter-tasting substances (Meyerhof et al., 2009). Studies on protein structure-function relationships have consistently demonstrated that the binding pocket for TAS2R receptors is situated on the extracellular surface of the transmembrane (TM) bundle, between TMs III, V, VI, and VII. This location is typical for class A GPCRs. Additionally, the most extended loop on the extracellular side of the receptor, known as ECL2, plays a critical role in the activation of TAS2R and binding with ligands (Born et al., 2013; Nowak et al., 2018). The composition of the binding site varies greatly among TAS2Rs, allowing for different agonist-specific interaction patterns (Born et al., 2013). Bitter taste receptors are stimulated by an extensive diversity of chemicals. This ability to recognize a broad spectrum of chemical compositions is accomplished through various interaction mechanisms between the various ligands within the binding site (Brockhoff et al., 2010). These receptors are categorized into two groups: promiscuous, which can be triggered by a high number of chemically distinct ligands, and selective, which are only activated by a few specific compounds (Meyerhof et al., 2009). Similarly, also bitter ligands can target several TAS2Rs or only specific receptors (Di Pizio and Niv, 2015). Investigating the promiscuity of both ligands and receptors is essential, especially when predicting protein-ligand interactions. The primary benefit of this research is the opportunity to rationally design ligands, planning their chemical structure to meet specific requirements. In the field of precision nutrition, the prediction of receptor-ligand interactions can be leveraged to generate food fingerprints that can be associated with quality and perceived taste. These fingerprints could then be used for the rational design of foods and in quality controls. This work employs molecular modelling and machine learning-driven tools to identify coffee fingerprints relating chemical composition to selective interaction with the human bitter taste receptor system.

2. Material and Methods

2.1 Molecular Modelling of Ligand-Bitter Receptor interactions

The three-dimensional structures of the 25 human bitter taste receptors were obtained from publicly available resources or built in-house using homology modelling. 23 bitter taste receptors were downloaded from the BitterDB (https://bitterdb.agri.huji.ac.il/dbbitter.php) (Dagan-Wiener et al., 2019), whereas the remaining two, i.e. TAS2R45 and TAS2R48, were obtained by homology modelling using the same workflow used for the BitterDB structures, with the structure of β2 adrenergic receptor (PDB ID: 3SN6) as a template. The structure of the TAS2R14 was first created and then used as a template for the TAS2R45 and TAS2R48 structures. MEDELLER (Kelm et al., 2010) software was employed to perform the homology modelling steps. The models of the 25 human bitter receptors were evaluated by using the PROCHECK suite (Laskowski et al., 1993).

The protein-membrane complex was built using CHARMM-GUI (Wu et al., 2014). A homogeneous bilayer membrane composed of POPC (16:0/18:1 acyl chains) was created by filling a rectangular box of 8 nm x 8 nm in x and y coordinates, and around 11 nm along the z direction. TIP3P water model was used to solvate all systems, then Na^+ and Cl^- ions were added to neutralize the systems and at a physiological concentration of 0.15 M. Parameters of the protein/membrane complex were generated by CHARMM-GUI using the CHARMM36 force field (Vanommeslaeghe et al., 2010). Molecular dynamics simulations were carried out for the 25 receptors using GROMACS 2020 (Abraham et al., 2015) following the CHARMM-GUI suggested workflow. The production MD simulation without position restraints was carried out in an NPT ensemble with a Nose-Hoover thermostat (Nosé, 1984) and Parrinello-Rahman barostat (Parrinello and Rahman, 1981) for 200 ns and a timestep of 2 fs. For each TAS2R, the last 50 ns of the 3 MD replicas were concatenated before performing a cluster analysis using the single linkage algorithm, the root-mean-squared deviation (RMSD) between atoms forming the binding pocket, and a cutoff of 0.15 nm. The centroids of the obtained clusters were considered for the following docking simulations, discarding the ECL2 as it could reduce binding pocket accessibility (Xu et al., 2022). Residues forming the binding pocket were derived from the experimental structure of human TAS2R46 in complex with strychnine (PDB 7XP6): the binding pocked has been defined as the residues within 0.4 nm of the ligand atoms (Xu et al., 2022). The corresponding residues in the remaining TAS2Rs were then obtained after performing a multiple sequence alignment, using Molecular Operating Environment (MOE) software.

Before predicting the poses of the investigated chemicals inside the receptors' binding pockets with molecular docking, shown in the following paragraph, the TAS2R configurations were pre-processed by SPORES software (ten Brink and Exner, 2009), while 3D coordinates and the protonation state of the ligands were obtained using MOE. Molecular docking was carried out by PLANTS software (Korb et al., 2009), using $PLANTS_{CHEMPLP}$ scoring function, search speed 1, 20 aco ants, and 25 aco evaporation factor. The binding site has been defined as a sphere of radius 1.25 nm centred in the centre of mass of residues forming the binding pocket. For each receptor, the docking procedure was performed 100 times, keeping the highest-affinity pose for each run. The best-scoring poses were clustered based on the RMSD of ligand heavy atoms, and the centroid with the highest

affinity was selected. The interactions between the docked ligand and the receptor were then identified and characterised using PLIP software (Adasme et al., 2021).

2.2 Machine learning-driven bitter predictor informed by molecular modelling

The coffee fingerprint predictor is obtained by the composition of 25 ML binary classifiers, each representing a TAS2R receptor, which predicts the binding event (ligand-receptor binding=1; ligand-receptor not bound=0). The coffee chemical composition is represented as a set of ligands potentially targeting TAS2R receptors. Therefore, the coffee fingerprint prediction will be a binary matrix matching the coffee chemical composition with the affinity towards the 25 TAS2R receptors.
For each above-mentioned TAS2R model, the ML classifier was developed using, as model features, quantities obtained by docking simulations (e.g., docking score, ligand-receptor h-bond type, etc..). The VirtuousSweetBitter (Maroni et al., 2022) database, which contains 1996 molecules with experimentally-known tastes (1043 bitter and 953 sweet), has been used for ML training and testing for validation with experimental data.
Then, the predictor was deployed to identify the fingerprint of a specific type of coffee, recently characterized by (Vezzulli et al., 2022), and here represented as a set of ligands collected in the so-called Coffee Dataset.
For each TAS2R, the procedure sketched in Figure 1 was followed.

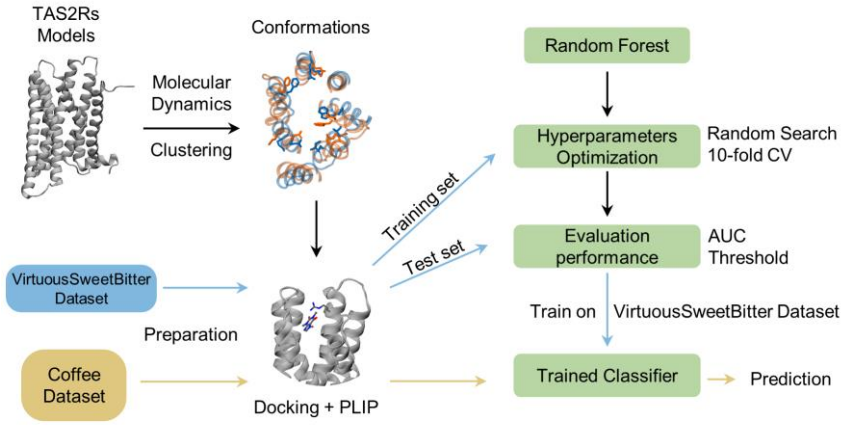

Figure 1. Schematic representation of the workflow followed to build the Molecular Modelling informed classifiers able to predict the targeting ability of small molecules over the bitter taste receptors.

First, the VirtuousSweetBitter Dataset was docked on each bitter taste receptor configuration. From the docking analysis, two continuous features, namely the docking score (D-tot) and the docking score normalized on the number of ligand heavy atoms (D-hev), were derived together with a set of binarized features representing the set of unique ligand-protein interactions identified by the PLIP software (Adasme et al., 2021). The two continuous features (D-tot, D-hev) were normalized on a scale from 0 to 1. The number of total features varies according to specific ligand-receptor interactions and is comprised of between 115 and 172. Then, a Random Forest classifier was built as follows. First, hyperparameters, such as the number of trees and the maximum number of splits, were optimized through a random search (100 repetitions) using 80% of the VirtuousSweetBitter Dataset as a training set. For each set of parameters, a 10-fold cross-validation has been performed and the average area under the receiver operating characteristic (ROC) curve (AUC) has been used to evaluate the classifier performance. After the optimization of the hyperparameters, the classifier was trained on the training set, the performances were evaluated on the test set (20% of VirtuousSweetBitter Dataset) in terms of AUC, and the best threshold was identified as the one maximizing the difference between the true positive rate and the false positive rate. Finally, the obtained classifier was used to predict the targeting ability of molecules in the Coffee Dataset over the 25 human bitter taste receptors. Feature importance analysis was performed through the impurity-based importance of the Random Forest classifier.

3. Results

TAS2R conformational dynamics and related properties have been analyzed at the structural equilibrium (reached in the last 50 ns of all the performed MD simulations). For each TAS2R model, equilibrium trajectories from MD replicas were clustered based on structural similarity (i.e., RMSD clustering) of the receptor binding

pocket. Centroids of most sampled receptor conformations have been used for molecular docking simulations. Through an iterative docking process on the VirtuousSweetBitter Dataset against TAS2R models, the best metrics, among those proposed by PLANTS, were identified to maximize accuracy in discriminating sweet from bitter compounds. The best scoring metric was the “number of heavy atoms (D-hev)”. Moreover, the “total docking score (D-tot)” as it does not contain information about the structure of the compound was also considered. The distribution of D-tot is similar for sweet and bitter compounds, while the distributions of D-hev show that bitter molecules have a less negative score, meaning that they are generally smaller (Figure 2A).
The same scores have been then deployed for the Coffee Dataset, which was characterized, in general, by a tendency towards less negative D-tot and more negative D-hev compared to considered bitter compounds (Figure 2B). For the Coffee Dataset, the D-hev distributions are highly asymmetrical, with tails extending towards low scores (absolute value).

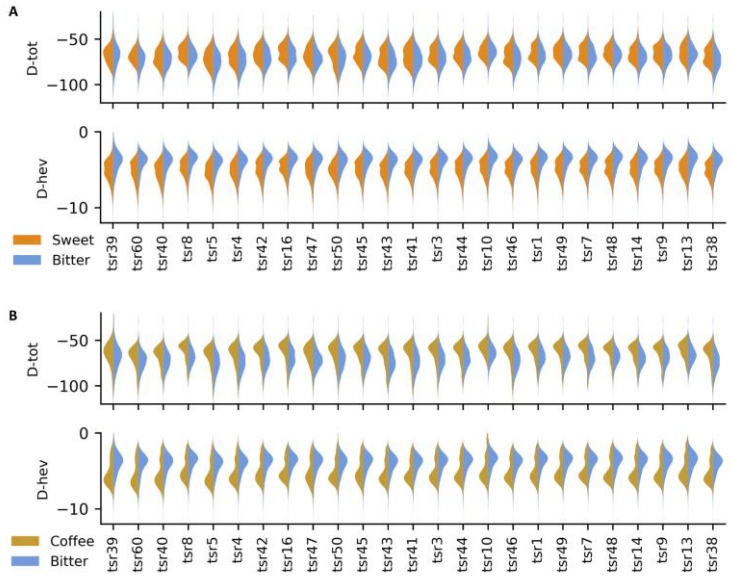

Figure 2. (A) Distribution of the total docking scores (D-tot) and the docking scores normalized on the number of heavy atoms (D-hev) across the 25 bitter taste receptors for the compounds in the VirtuousSweetBitter Dataset (A) and in the Coffee Dataset (B). For the VirtuousSweetBitter Dataset, the taste of the molecules is highlighted in orange and blue for sweet and bitter tastes respectively (A), whereas the scores for the Coffee Dataset are compared to the ones of the bitter molecules within the VirtuousSweetBitter Dataset (B).

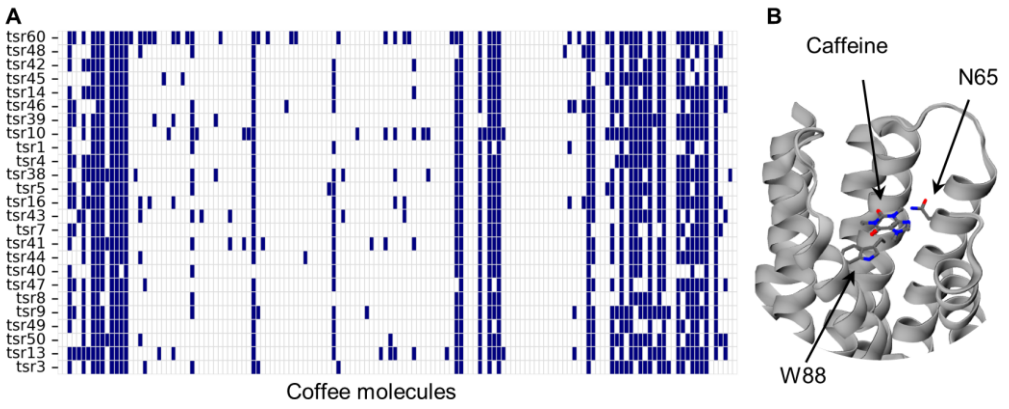

Figure 3. (A) Coffee Fingerprint: is a matrix where each box identifies a possible interaction between a bitter taste receptor and a molecule (Coffee Dataset). A blue box indicates a favourable, whereas a white box denotes that the interaction is unlikely. (B) Visual rendering of caffeine binding pose in TAS2R46.

A random forest classifier was constructed for each TAS2R by optimizing the hyperparameters through 10-fold cross-validation. Feature importance analysis showed that the most important feature for all TAS2Rs is D-hev, followed by D-tot. Interestingly, the PLIP software analysis showed that the most important interactions were different among the bitter receptors. While relevant electrostatic interactions were identified in some TAS2Rs, a predominance of hydrophobic interactions was observed in other receptors. The inclusion of these interaction features enhanced the discrimination performance of the classifiers compared with classification based on docking score alone. The AUC obtained on the VirtuousSweetBitter test set (for all TAS2R classifiers) was in the range of 0.81-0.86, showing that the molecular modelling-based classification could discriminate very well between sweeteners and bitterants. The validated trained classifiers were also deployed to predict the fingerprint of the Coffee Dataset. In detail, the fingerprint predictors identified specific interactions between TAS2Rs and coffee molecular compounds (Figure 3A). Also, the analysis allowed to characterize the ligand-receptor binding pose features with an atomic resolution (Figure 3B). According to the predictions, 100 out of the 144 molecules from the coffee analysis could effectively bind at least one TAS2R receptor. Moreover, this analysis pinpointed differences in the promiscuity and selectivity of the investigated chemicals: 31 coffee-derived compounds were able to bind a single bitter taste receptor, whereas other 26 more than 20 TAS2Rs.

4. Discussion

This work addresses the prediction of the coffee bitter taste profile starting from its chemical composition, by integrating molecular and machine learning-driven modelling. In more detail, we developed a ML predictor of bitterness and validated it with experimental data (as made in Maroni et al., 2022) obtaining AUCs above 0.8. In contrast to previous works in the field, the present predictor takes in input features related to the ligand-receptor interactions. Therefore, learning is based primarily on the crucial types of protein-ligand interaction that determine bitterness recognition. Assuming that a compound, recognized as bitter, binds at least one bitter receptor, the output of the predictor can be used to estimate a potential bitter signature. This signature was estimated by using molecular docking methods to comparatively characterize the binding of those compounds, identified as bitter, toward the various existing bitter receptors. In this sense, machine learning and molecular modeling are coupled: the former is used to screen ligand activity based on the molecular characteristics of the ligand-receptor interaction, and the latter uses molecular docking to perform an affinity study of a predicted bitter ligand toward all bitter receptors. The analysis of the docking scores in the VirtuousSweetBitter Dataset showed that the normalization over the number of heavy atoms increases the discrimination of non-bitter/bitter compounds (Figure 2A), therefore the dimension of the ligand and its relative affinity towards TAS2Rs play a crucial role in determining its taste. Moreover, integrating the docking scores with information on the protein-ligand interactions resulted in a substantial increase in the classification performances, which implies that the nature of the interactions might determine the receptor targeting. The TAS2R-specific classifiers were used to predict the taste of coffee metabolites and pinpoint the main interactions between coffee-extracted chemicals and TAS2Rs (Figure 3A). It is worth mentioning that this analysis highlighted that caffeine interacts with the TAS2R46 with a pi-stacking interaction with residue W88 (Figure 3B), in line with previous evidence from a structure of the human TAS2R46 complexed with strychnine (Xu et al., 2022). Interestingly, the developed models are promising to also gain insights into the promiscuity/selectivity nature of analyzed receptors and compounds. This aspect will be in future assessed and refined using data from previous literature (Dagan-Wiener et al., 2019) and expanded with new experimental data regarding the specific TAS2R-ligand interactions.

5. Conclusions

The present work presents a proof of concept in the field of machine learning based informed by molecular modelling for applications in the field of nutrition. Through the analysis of protein-ligand interactions, it is possible to predict the bitter taste profile of compounds extracted from coffee, highlighting their bitter/non-bitter nature. With this methodology, it may be possible to predict the bitter footprint of food ingredients to help the analysis of their taste and quality and the development of solutions in the areas of nutrition, health, and the food industry.

Acknowledgements. The present work has been developed as part of the VIRTUOUS project (https://virtuoush2020.com), funded by the European Union's Horizon 2020 research and innovation program under the Marie Sklodowska-Curie-RISE Grant Agreement No 872181.

References

Abraham, M.J., Murtola, T., Schulz, R., Páll, S., Smith, J.C., Hess, B., Lindah, E., 2015. Gromacs: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. SoftwareX.

Adasme, M.F., Linnemann, K.L., Bolz, S.N., Kaiser, F., Salentin, S., Haupt, V.J., Schroeder, M., 2021. PLIP 2021: expanding the scope of the protein–ligand interaction profiler to DNA and RNA. Nucleic Acids Research 49, W530–W534.

Born, S., Levit, A., Niv, M.Y., Meyerhof, W., Behrens, M., 2013. The human bitter taste receptor TAS2R10 is tailored to accommodate numerous diverse ligands. Journal of Neuroscience 33, 201–213.

Brockhoff, A., Behrens, M., Niv, M.Y., Meyerhof, W., 2010. Structural requirements of bitter taste receptor activation. Proceedings of the National Academy of Sciences of the United States of America 107, 11110–11115.

Chandrashekar, J., Mueller, K.L., Hoon, M.A., Adler, E., Feng, L., Guo, W., Zuker, C.S., Ryba, N.J.P., 2000. T2Rs Function as Bitter Taste Receptors. Cell 100, 703–711.

Dagan-Wiener, A., Di Pizio, A., Nissim, I., Bahia, M.S., Dubovski, N., Margulis, E., Niv, M.Y., 2019. BitterDB: taste ligands and receptors database in 2019. Nucleic Acids Research 47, D1179–D1185.

Di Pizio, A., Levit, A., Slutzki, M., Behrens, M., Karaman, R., Niv, M.Y., 2016. Comparing Class A GPCRs to bitter taste receptors, in: Biophysical Methods in Cell Biology. Elsevier Ltd, pp. 401–427.

Di Pizio, A., Niv, M.Y., 2015. Promiscuity and selectivity of bitter molecules and their receptors. Bioorganic & Medicinal Chemistry 23, 4082–4091.

Kelm, S., Shi, J., Deane, C.M., 2010. MEDELLER: Homology-based coordinate generation for membrane proteins. Bioinformatics 26, 2833–2840.

Korb, O., Stützle, T., Exner, T.E., 2009. Empirical Scoring Functions for Advanced Protein−Ligand Docking with PLANTS. J. Chem. Inf. Model. 49, 84–96.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. Journal of Applied Crystallography.

Maroni, G., Pallante, L., Di Benedetto, G., Deriu, M.A., Piga, D., Grasso, G., 2022. Informed classification of sweeteners/bitterants compounds via explainable machine learning. Current Research in Food Science 5, 2270–2280.

Meyerhof, W., Batram, C., Kuhn, C., Brockhoff, A., Chudoba, E., Bufe, B., Appendino, G., Behrens, M., 2009. The molecular receptive ranges of human TAS2R bitter taste receptors. Chemical Senses 35, 157–170.

Nosé, S., 1984. A unified formulation of the constant temperature molecular dynamics methods. The Journal of Chemical Physics 81, 511–519.

Nowak, S., Di Pizio, A., Levit, A., Niv, M.Y., Meyerhof, W., Behrens, M., 2018. Reengineering the ligand sensitivity of the broadly tuned human bitter taste receptor TAS2R14. Biochimica et Biophysica Acta - General Subjects 1862, 2162–2173.

Parrinello, M., Rahman, A., 1981. Polymorphic transitions in single crystals: A new molecular dynamics method. Journal of Applied Physics 52, 7182–7190.

Roper, S.D., 2017. Taste: Mammalian Taste Bud Physiology, in: Reference Module in Neuroscience and Biobehavioral Psychology. Elsevier, pp. 887–893.

ten Brink, T., Exner, T.E., 2009. Influence of Protonation, Tautomeric, and Stereoisomeric States on Protein−Ligand Docking Results. J. Chem. Inf. Model. 49, 1535–1546.

Vanommeslaeghe, K., Hatcher, E., Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., Mackerell, A.D., 2010. CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. Journal of Computational Chemistry.

Vezzulli, F., Rocchetti, G., Lambri, M., Lucini, L., 2022. Metabolomics Combined with Sensory Analysis Reveals the Impact of Different Extraction Methods on Coffee Beverages from Coffea arabica and Coffea canephora var. Robusta. Foods 11, 807.

Wu, E.L., Cheng, X., Jo, S., Rui, H., Song, K.C., Dávila-Contreras, E.M., Qi, Y., Lee, J., Monje-Galvan, V., Venable, R.M., Klauda, J.B., Im, W., 2014. CHARMM-GUI membrane builder toward realistic biological membrane simulations. Journal of Computational Chemistry.

Xu, W., Wu, L., Liu, S., Liu, X., Cao, X., Zhou, C., Zhang, J., Fu, Y., Guo, Y., Wu, Y., Tan, Q., Wang, L., Liu, J., Jiang, L., Fan, Z., Pei, Y., Yu, J., Cheng, J., Zhao, S., Hao, X., Liu, Z.-J., Hua, T., 2022. Structural basis for strychnine activation of human bitter taste receptor TAS2R46. Science 377, 1298–1304.

Zhang, Haonan, Qiao, A., Yang, D., Yang, L., Dai, A., de Graaf, C., Reedtz-Runge, S., Dharmarajan, V., Zhang, Hui, Han, G.W., Grant, T.D., Sierra, R.G., Weierstall, U., Nelson, G., Liu, W., Wu, Y., Ma, L., Cai, X., Lin, G., Wu, X., Geng, Z., Dong, Y., Song, G., Griffin, P.R., Lau, J., Cherezov, V., Yang, H., Hanson, M.A., Stevens, R.C., Zhao, Q., Jiang, H., Wang, M.-W., Wu, B., 2017. Structure of the full-length glucagon class B G-protein-coupled receptor. Nature 546, 259–264.