Novel scaffold of natural compound eliciting sweet taste revealed by machine learning.

Cédric Bouysset [1], Christine Belloir [2], Serge Antonczak [1], Loïc Briand [2], Sébastien Fiorucci [1 *]

[1] Université Côte d'azur, CNRS, Institut de Chimie de Nice UMR7272, 06108 Nice, France
[2] INRAE, CNRS, Université de Bourgogne-Franche Comté, AgroSup Dijon, Centre des Sciences du Goût et de l'Alimentation, 21000 Dijon, France

*Corresponding author: sebastien.fiorucci@univ-cotedazur.fr

Abstract

Sugar replacement is still an active issue in the food industry. The use of structure-taste relationships remains one of the most rational strategy to expand the chemical space associated to sweet taste. A new machine learning model has been setup based on an update of the SweetenersDB and on open-source molecular features. It has been implemented on a freely accessible webserver. Cellular functional assays show that the sweet taste receptor is activated *in vitro* by a new scaffold of natural compounds identified by the *in silico* protocol. The newly identified sweetener belongs to the lignan chemical family and opens a new chemical space to explore.

Keywords

Sweet taste, machine learning, natural compounds, sweetener, sweet taste receptor

Introduction

Consumer interest in natural high potency sweeteners has grown spectacularly in recent years, fueled by concerns about sugar overconsumption and the use of artificial additives in foods. There are three main strategies to reduce sugar intake: an abrupt reduction of sugar without substitution, the use of flavor materials to modify sweet taste perception and the use of alternative sweeteners. Though many low-calorie sweeteners are known, only few of them are used by the food industry (Belloir, Neiers, & Briand, 2017). The search of novel intense sweeteners, possessing the same chemosensory profile as sucrose, remains open and challenging.

All sweet tasting compounds are detected by a single heterodimeric G protein-coupled receptor composed of T1R2 and T1R3 subunits expressed at the surface of taste buds (Li et al., 2002; Nelson et al., 2001). However, no experimental 3D-structure of the T1R2/T1R3 sweet taste receptor is available and ligand-based approaches such as Structure Activity Relationship (SAR), are relevant to establish a link between the structure of a compound and its sweet taste. From original studies of Edna W. Deutsch & Corwin Hansch (Deutsch & Hansch, 1966), followed a year later by Robert S. Shallenberger & Terry E. Acree

(Shallenberger & Acree, 1967) to recent structure-taste relationship models (Achary, Toropova, & Toropov, 2019; Arnoldi, Bassoli, Merlini, & Ragg, 1991; Barker, Hattotuwagama, & Drew, 2002; Bassoli et al., 2001; Chéron, Casciuc, Golebiowski, Antonczak, & Fiorucci, 2017; Drew et al., 1998; Rojas, Tripaldi, & Duchowicz, 2016; Spillane & McGlinchey, 1981; Spillane et al., 2000, 1996; Spillane, McGlinchey, Muircheartaigh, & Benson, 1983; Spillane & Sheahan, 1989; Tuwani, Wadhwa, & Bagler, 2019; Van Der Heijden, Brussel, & Peer, 1979; Vepuri, Tawari, & Degani, 2007; Walters, 2006; Zheng, Chang, Xu, Xu, & Lin, 2019), the quest to understand the molecular features underlying sweet taste perception is still active. In this study, we present the first online tool able to predict sweet taste perception based on a machine learning protocol. We have updated and curated the previous database of 316 sweet compounds (SweetenersDB) and added new applicability domain metrics to assess the robustness of the predictions. A novel scaffold of natural sweetener, belonging to the lignan chemical family, that have never been annotated as sweet have been identified and experimentally validated.

Materials and Methods

Data preparation

Based on our previous work (Chéron et al., 2017), the database of sugars and sweeteners (Figure S1), named SweetenersDB, was curated and updated with missing compounds (Ruiz-Aceituno, Hernandez-Hernandez, Kolida, Moreno, & Methven, 2018). Each compound was labelled with a relative sweetness value, corresponding to a measure of the sweet taste intensity relative to sucrose. Relative sweetness is defined as the concentration ratio between a sucrose solution and a solution of sweetener perceived with the same intensity. The relative sweetness of each compound was transformed in logarithmic scale for easier manipulation, and it will be later referred to as logSw. For compounds that were already present in the database, we updated the SMILES (Simplified Molecular Input Line Entry System) to isomeric SMILES in order to differentiate stereoisomers. When the information on stereocenters was not available, we either regrouped the stereoisomers in a single entry with their average logSw value if the logSw difference was lower than 0.2, or we discarded both compounds. The resulting dataset consisted of 316 compounds in SweetenersDB (Table S1). The machine learning protocol was applied to two datasets of interest : 4796 natural compounds (Table S2) extracted from the SuperNatural II database and the phyproof catalogue from PhytoLab, already pre-screened by our previous model (Chéron et al., 2017).

Every compound in the datasets were collected as SMILES strings and sanitized with RDKit (Landrum et al., 2018). To assess the importance of predicting protonation states, the major microspecies of each compound was also determined with ChemAxon cxcalc tool (ChemAxon, 2018) at physiological salivary pH (pH=6.5). Structures were then standardized using the "standardizer" (EMBL-EBI, 2017) Python package: salts are removed from the structure, and a set of around 30 structure-normalization rules are applied to each molecular graph to cover most of tautomerization reactions. 0D, 1D and 2D descriptors were computed using Dragon v6.0.38 (Talete srl, 2014), RDKit (Landrum et al., 2018), Mordred (Moriwaki, Tian, Kawashita, & Takagi, 2018), and ChemoPy (Cao, Xu, Hu, & Liang, 2013). Descriptors from the three latter packages were regrouped as "open-source" descriptors. For each of these two descriptors sets, the initial number of features was reduced by removing those that could not be calculated for a molecule, as well as near-constant features (two or less unique values), features with a standard deviation below 0.001, and features with a correlation greater than 0.95. The resulting datasets consisted of

635 descriptors for the Dragon dataset, and 506 features for the "open-source" dataset. To avoid any model bias due to overfitting, the number of features used by the model is a hyperparameter that has been optimized.

The updated SweetenersDB was split in training and test sets using a Sphere Exclusion clustering algorithm. Dragon descriptors were chosen for this procedure: they were normalized between 0 and 1, and the clustering was initiated from the compound that is closest to the center of the dataset in the descriptor hyperspace. 64 diverse compounds (20.3%) were selected for the test set, leaving 252 compounds in the training set (Figure 1, Table S1). The chemical space was mapped using a t-distributed Stochastic Neighbor Embedding (t-SNE) analysis. t-SNE was performed with the scikit-learn python package (v0.20.2) (Pedregosa et al., 2011) using default parameters (perplexity of 30, early exaggeration of 12, learning rate of 200 and 1000 iterations) except for the embedding initialization which was done with principal component analysis.

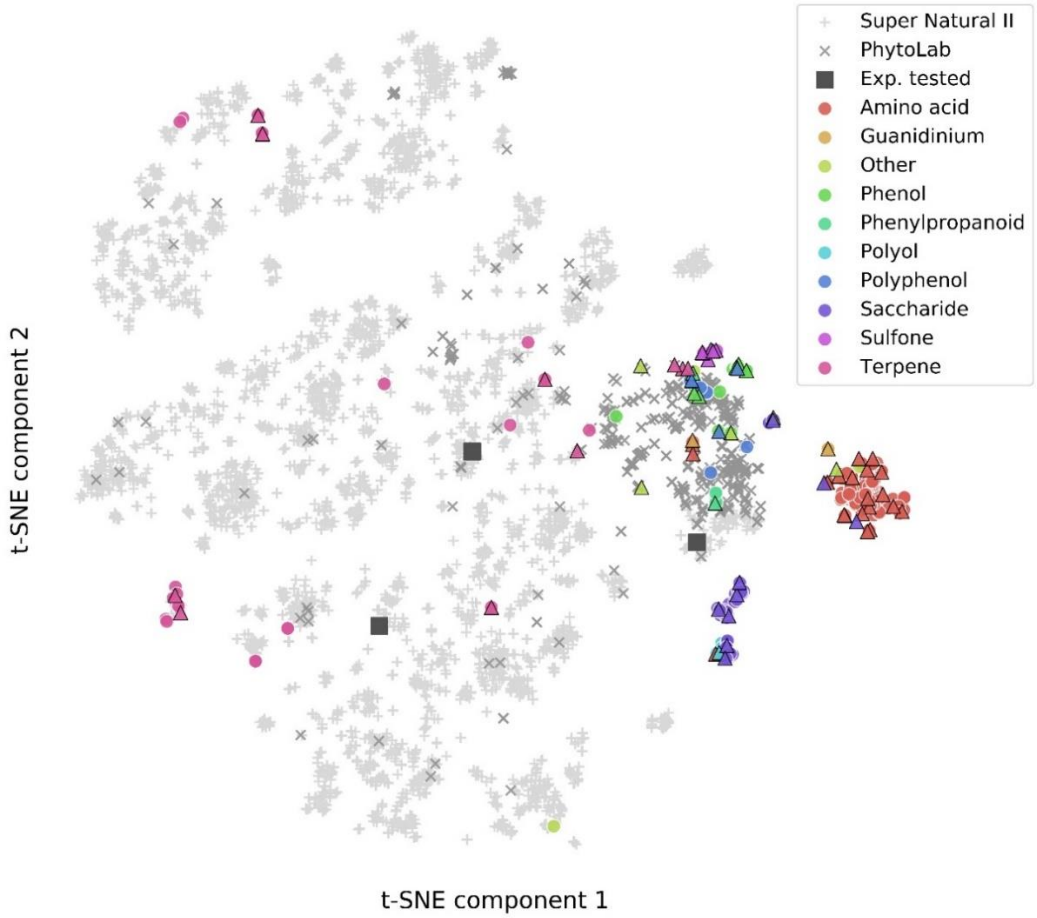

Figure 1: *Representation of the SweetenersDB chemical space based on a t-SNE dimensionality reduction method. Known sweet chemical families in the training and test set are represented by circle and triangles respectively. Light and dark grey data points represent natural compounds that were predicted as intensely sweet (logSw ≥ 2) by both our previous and current models (Table S2). Grey squares represent natural molecules experimentally tested in the present study.*

Machine-learning model for sweetness prediction

Several regression algorithms from the python package scikit-learn were evaluated: Random Forest, Support Vector Machine (SVM), Adaptative Boosting with a Decision Tree base estimator (AdaBoost Tree), and k-Nearest Neighbors. Five-fold cross validation was performed with hyperparameter tuning using a grid search. The workflow for each cross-validation fold was as follow: standardization of descriptors, feature selection, and model training. Selection of descriptors was done by keeping a given percentile of the highest ranked descriptors based on their Mutual Information with our endpoint. The optimal percentile of features was tuned as a parameter of the Grid Search.

Once optimal hyperparameters were found for each model, final models were trained using the full training dataset. Their predictive performance was evaluated based on criteria previously defined by Golbraikh and Tropsha (Golbraikh & Tropsha, 2002). For the "Dragon" models, only the SVM model did not pass all criteria, and for the "open source" model, only the AdaBoost Tree passed all criteria. In both cases, the AdaBoost Tree model was selected as the best performing model, using 32 descriptors for the "Dragon" model, and 51 descriptors for the "open source" model (Figure S2 and Table S4). A summary of their performances is reported in the results section (Table 1) and detailed in supporting information (Table S3).

In addition to training and validating several models for sweetness prediction, a web server implementing the "open-source" model was developed and is freely available at the following address: http://chemosimserver.unice.fr/predisweet/

Other chemoinformatics solutions are available but none of them has been implemented on a webserver. For instance, the e-Sweet platform (Zheng et al., 2019) is based on a consensus model of various machine learning protocols. The database used to train and test their model is very similar to the database used to setup Predisweet and e-Sweet performs as well as our model (R^2 on the test set is in the same range [0.75-0.78] for both solutions). Recently a new functionality to predict sweetness has been implemented on the BitterSweet webserver (Tuwani et al., 2019). The performance of BitterSweet is comparable to e-Sweet and Predisweet (R^2 of 0.72 on our test set) but the protocol is still unpublished, and seven molecules of the test set has not been considered as sweet.

Webserver interface

The user is asked for one or several molecules which can either be drawn directly on the chemical structure editor Ketcher or inputted as a simple text query or file in the SMILES format. The workflow (Figure 2) followed by query compounds is the same as used during model development. First, a molecule is generated from the SMILES string with RDKit to assess its sanity. The structure is then standardized using the "standardizer" Python module. The 51 molecular descriptors selected during model development are computed and standardized based on the training set transformations. The descriptors are passed to the AdaBoost Tree model in order to predict the logSw. Finally, the quality of each prediction is assessed based on three metrics, namely the applicability, reliability, and decidability domains (Hanser, Barber, Marchaland, & Werner, 2016). The applicability domain indicates if the compound is within the descriptor range of the training set and its score is computed using a convex hull approach. The reliability domain highlights the density of information around the compound. The reliability score is calculated by counting the number of molecules from the training set that are inside a sphere centered on the query. The decidability domain shows the confidence in the prediction that was made. The decidability score is based on the weights

of each decision tree that compose the AdaBoost model. It is computed by summing the weights of decision trees that made a prediction close to the model prediction and dividing it by the sum of all weights.

Each molecule is indexed in the database with its InChIKey, which avoids making predictions for the same molecule twice. For a seamless user experience, the name of each molecule is retrieved by querying PubChem with the pubchempy Python package, and a 2D representation of the compound is generated with RDKit.

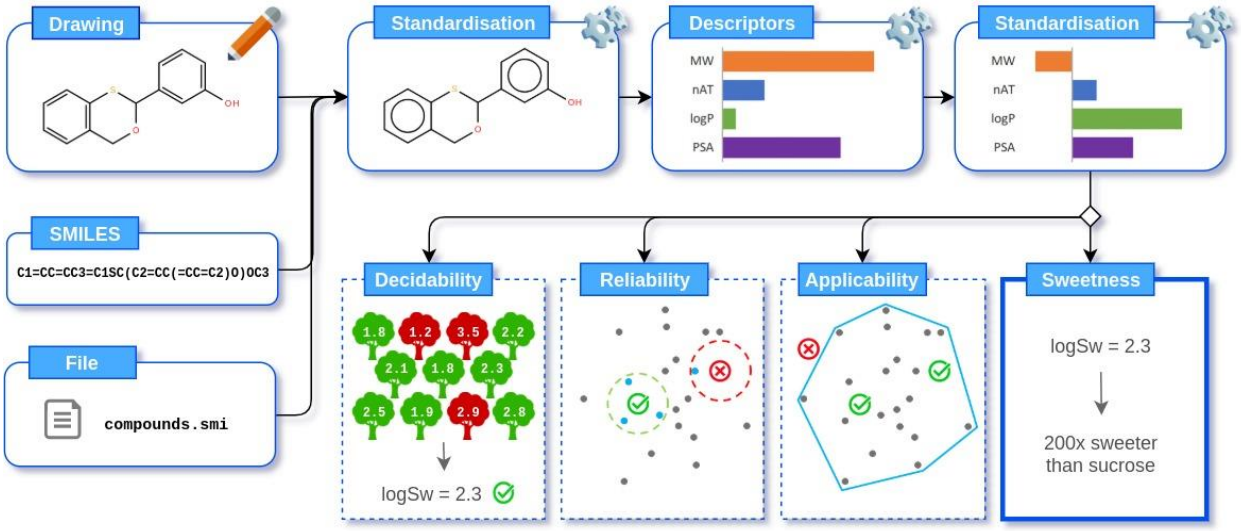

Figure 2: *Workflow followed by each molecule submitted to the webserver.*

Functional expression of the human sweet taste receptor

In order to validate the sweetness of the three natural compounds, we employed a cell-based expression system for the human T1R2/T1R3 sweet taste receptor as previously described (Poirier et al., 2012; Sigoillot et al., 2018). Briefly, the cDNAs coding human T1R2 and T1R3 subunits were cloned into pcDNA3 and pcDNA4 expression plasmids, respectively. HEK293T cells stably expressing Gα16gust44 and T1R3 were seeded at a density of 0.4 $\times 10^6$ cells per well into 96-well black walled, clear bottom microtiter plates (Falcon) in high-glucose DMEM supplemented with 2 mM GlutaMAX, 10% dialyzed foetal bovine serum, penicillin/streptomycin, G418 (400 µg/mL) and zeocin (250 µg/mL) at 37 °C and 6.3% CO_2, in a humidified atmosphere. Twenty-four hours later, HEK293T-Gα16gust44-T1R3 cells were transiently transfected with pcDNA3-T1R2 (120ng/well) with Lipofectamine 2000. Calcium signal of mock-transfected cells (HEK293T Gα16gust44 cells stably expressing T1R3 transfected with pcDNA3 empty vector) were always measured in parallel and compared. Twenty-four hours after transfection, the cells were loaded for 1 hour at 37°C with the calcium indicator Fluo4-AM (Molecular Probes) diluted in C1 buffer (130 mM NaCl, 5 mM KCl, 10 mM Hepes pH 7.4, 2 mM $CaCl_2$) in the presence of pluronic acid (0.025%, w/v) and probenecid (2.5 mM). After washing with C1 buffer, cells were stimulated with a range of sweet tasting compounds. The fluorescence intensity was measured for 90 seconds (excitation 488 nm, emission 510 nm) into an automated fluorimetric FlexStation®3 Multi-Mode microplate reader. The change in fluorescence upon stimulus application were averaged, mock-substracted and baseline-corrected. The EC_{50} values were calculated using SigmaPlot software by nonlinear regression using the function:

$$f(x) = min + \frac{max - min}{1 + \left(\frac{x}{EC_{50}}\right)^{-\text{Hillslope}}}$$

Chemicals

All tested compounds (arctiin, ginsenoside Rd and jujuboside A, Figure 3) were purchased from Phytolab GmbH & Co. KG, with the exception of sucralose obtained from Sigma-Aldrich. All the compounds were dissolved first in DMSO (100 mM in 100% DMSO), and then diluted with the C1 buffer solution; except for sucralose, which was dissolved in the C1 buffer solution directly.

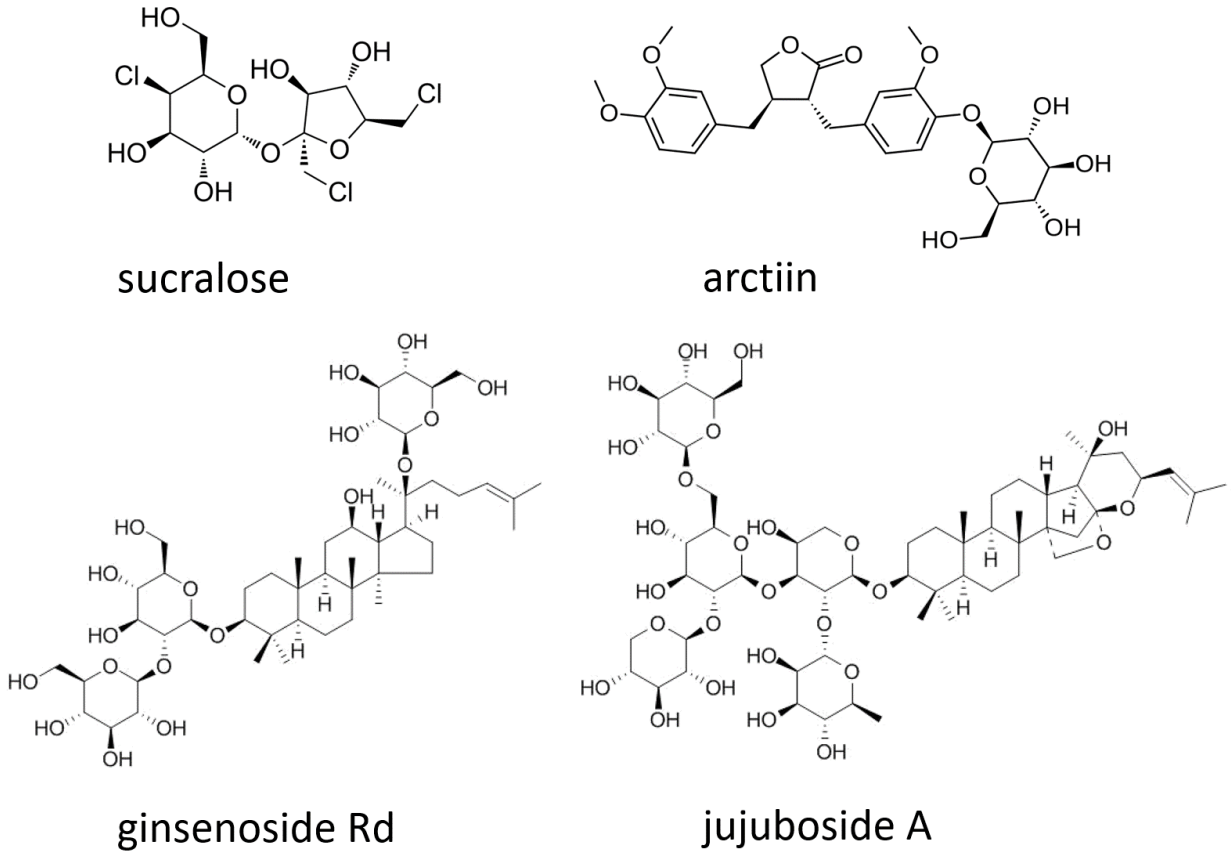

***Figure 3**: Structure of the tested compounds*

Results and discussion

New machine-learning model based on open-source features

The performance of the Open-source and Dragon models has been compared. Both models show good predictivity on the test set according to state of the art QSAR rules (Table 1). Slightly more than 90% of the test set are predicted with an absolute error lower than a log unit (Figure S3). The models are less

accurate for high sweetness values since they have been trained with less information for highly potent sweeteners. Improving the quality of the machine learning model would then requires i) expanding the chemical diversity of sweet compounds and ii) a larger database of *in vivo* and *in vitro* experiments. A threshold of LogSw larger than 2 has then been chosen to minimize false positive predictions prior *in vitro* validation. Since similar performance have been obtained for both models, the open-source version have been implemented on a webserver, freely accessible at the following address: http://chemosimserver.unice.fr/predisweet/. Another model has been set up with descriptors calculated at salivary pH to assess the effect of the protonation state on the model performance. Even though more than a quarter of the molecules had different descriptor values between the default and the salivary pH dataset, there was no significant difference in terms of performance. The protonation assessment step thus has been skipped in the final protocol. We emphasize that the model has not been trained to predict bitter taste and we envision to include this feature in a future work. Additionally, any QSAR model has a field of application that clearly defines the boundaries within which the model should be used, usually referred to as the applicability domain. We've implemented three different metrics to explicitly inform the user whether the model and its prediction can be trusted for a particular query molecule.

Table 1: Performance of the models according to Golbraikh and Tropsha rules. (Golbraikh & Tropsha, 2002)

Rules	Open-source model	Dragon model
$R^2 > 0.6$	0.74	0.75
$Q^2 > 0.5$	0.84	0.79
$\lvert R^2 - R_0^2\rvert / R^2 < 0.1$	0.02	0.05
$0.85 <= k <= 1.15$	0.93	0.90
$\lvert R_0^2 - R'^2_0\rvert < 0.3$	0.07	0.12

Identification of a new sweet scaffold

A large database of natural compounds has been virtually screened to identify new putative sweeteners. The analysis of the resulting sweet chemical space of ~4800 natural compounds shows that it does not fully overlap the chemical space of known sweeteners (Figure 1). It suggests that a large part of the natural chemical space remains unexplored. We have finally selected three natural compounds that have been tested for their ability to activate the human sweet taste receptor T1R2/T1R3 expressed in HEK cells, as previously reported (Poirier et al., 2012). As a negative control, HEK293T Gα16gust44 cells stably expressing T1R3 were mock-transfected with the empty expression vector to control for T1R2-independent non-specific signals. In addition to a LogSw value higher than 2, the price and the commercial availability were two important criteria in the compound choice. Two of them, Jujuboside A and Ginsenisode Rd, belong to the triterpene chemical family. The third one, arctiin, possesses a lignan scaffold. As shown in Figure 4b, application of arctiin on T1R2/T1R3-expressing cells evoked calcium responses in a dose-dependent manner, while no fluorescence signals where observed with mock transfected cells. The half-maximal effective concentrations (EC_{50}) of arctiin was 2.5 ± 0.4 mM. As a control, we determined the concentration-

response curve for the high-intensity sucralose (Figure 4a) leading to an EC_{50} value of 87 ± 13 µM, in agreement with reported values (Assadi-Porter et al., 2010; Masuda et al., 2012; Servant et al., 2010). In contrast, jujuboside A and ginsenisode Rd showed detectable activity on the T1R2/T1R3 receptor, but only at the highest tested concentration (Figure 4c and d) precluding establishment of complete dose–response curve and calculation of EC_{50} values. This concentration used was the maximum one that did not induce any side effects on mock transfected cells.

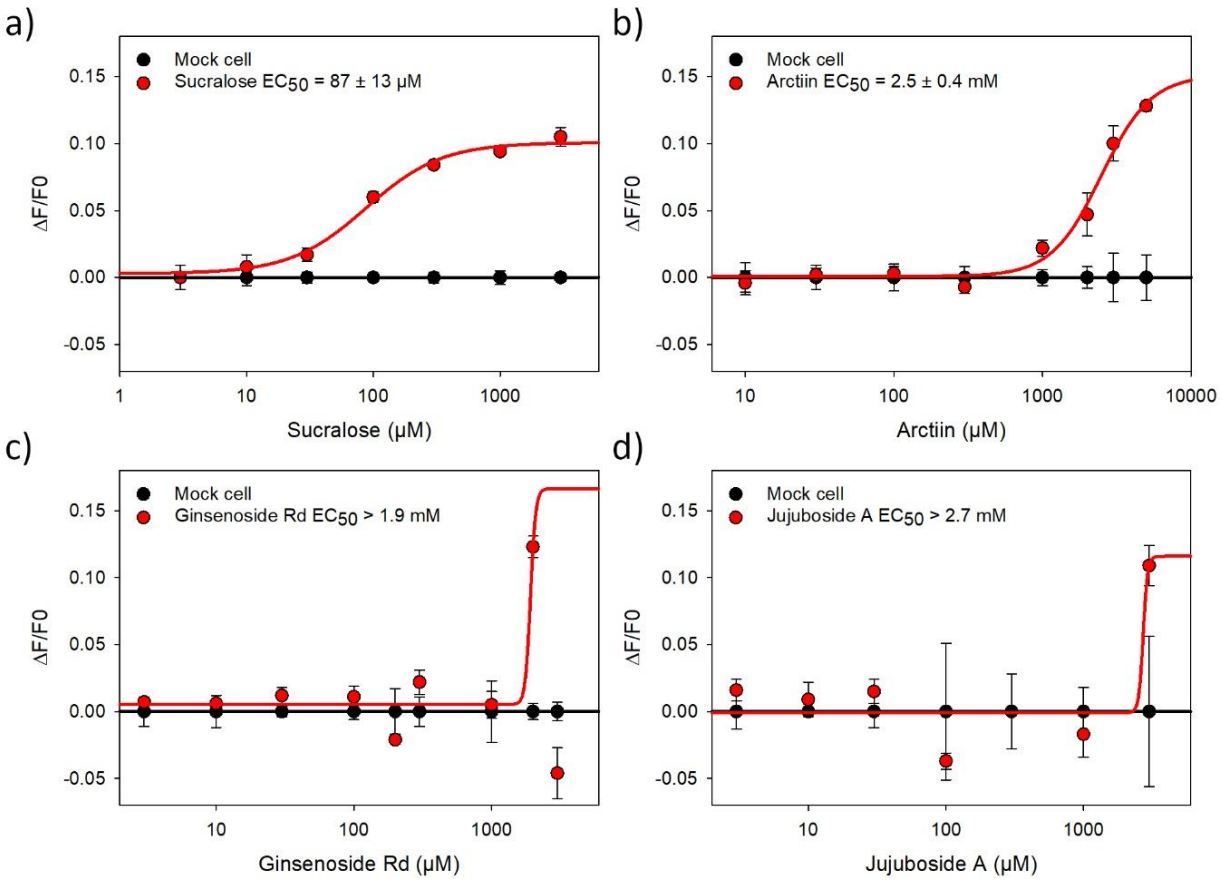

Figure 4*: Response of the human sweet taste receptor to the three natural compounds identified by the machine learning protocol and sucralose used as a control. Dose-response curves of T1R2/T1R3-expressing cells (red curve) and mock-transfected cells (black curve). All concentrations were measured in triplicate and each experiment was repeated at least 2 times.*

Conclusion

In this study we have used machine learning to predict novel agonists of the sweet taste receptor. An AdaBoost Tree model was setup based on open-source chemical features optimized on a curated database of 316 known sweet agents (SweetenersDB) and implemented on a freely available webserver. The virtual screening of a large database of natural compounds identified thousands of putative sweeteners, of which three were selected for *in vitro* functional assays of the human sweet taste receptor and dose-response analyses. Among them, we identified arctiin as a novel agonist of the T1R2/T1R3 sweet taste receptor with

an EC_{50} value of 2.5±0.4mM. It belongs to the lignan chemical family, polyphenols found in plants, of which epi-lyoniresinol has already been annotated as slightly sweet by sensory analyses (Cretin et al., 2015; Marchal, Cretin, Sindt, Waffo-Téguo, & Dubourdieu, 2015). As numerous natural sweeteners, arctiin might also possess bitter taste but it would require additional experiments out of the scope of the present study to assess its aftertaste. Nevertheless, our results confirm that the lignan chemical family opens a new chemical space for the search of new sweet agents and machine learning is a fruitful approach in this context.

Acknowledgements

This work was supported by the French Ministry of Higher Education and Research [PhD Fellowship], by GIRACT (Geneva, Switzerland) [9th European PhD in Flavor Research Bursaries for first year students] and the Gen Foundation (Registered UK Charity No. 1071026) [a charitable trust which principally provides grants to students/researchers in natural sciences, in particular food sciences/technology]. We also benefited from funding from the French government, through the UCAJEDI "Investments in the Future" project managed by the ANR grant No. ANR-15-IDEX-01.

References

Achary, P. G. R., Toropova, A. P., & Toropov, A. A. (2019). Combinations of graph invariants and attributes of simplified molecular input-line entry system (SMILES) to build up models for sweetness. *Food Research International*, *122*, 40–46. https://doi.org/10.1016/j.foodres.2019.03.067

Arnoldi, A., Bassoli, A., Merlini, L., & Ragg, E. (1991). Isovanillyl sweeteners. Synthesis, conformational analysis, and structure–activity relationship of some sweet oxygen heterocycles. *J. Chem. Soc., Perkin Trans. 2*, (9), 1399–1406. https://doi.org/10.1039/P29910001399

Assadi-Porter, F. M., Maillet, E. L., Radek, J. T., Quijada, J., Markley, J. L., & Max, M. (2010). Key Amino Acid Residues Involved in Multi-Point Binding Interactions between Brazzein, a Sweet Protein, and the T1R2-T1R3 Human Sweet Receptor. *Journal of Molecular Biology*, *398*(4), 584–599. https://doi.org/10.1016/j.jmb.2010.03.017

Barker, J. S., Hattotuwagama, C. K., & Drew, M. G. B. (2002). Computational studies of sweet-tasting molecules. *Pure and Applied Chemistry*, *74*(7), 1207–1217. https://doi.org/10.1351/pac200274071207

Bassoli, A., Drew, M. G. B., Hattotuwagama, C. K., Merlini, L., Morini, G., & Wilden, G. R. H. (2001). Quantitative Structure-Activity Relationships of Sweet Isovanillyl Derivatives. *Quantitative Structure-Activity Relationship*, *20*(1), 3–16. https://doi.org/10.1002/1521-3838(200105)20:1<3::AID-QSAR3>3.0.CO;2-H

Belloir, C., Neiers, F., & Briand, L. (2017). Sweeteners and sweetness enhancers. *Current Opinion in Clinical Nutrition and Metabolic Care*, *20*(4), 279–285. https://doi.org/10.1097/MCO.0000000000000377

Cao, D. S., Xu, Q. S., Hu, Q. N., & Liang, Y. Z. (2013). ChemoPy: Freely available python package for computational biology and chemoinformatics. *Bioinformatics*, *29*(8), 1092–1094. https://doi.org/10.1093/bioinformatics/btt105

ChemAxon. (2018). *Calculator Plugins*. Retrieved from http://www.chemaxon.com

Chéron, J. B., Casciuc, I., Golebiowski, J., Antonczak, S., & Fiorucci, S. (2017). Sweetness prediction of natural compounds. *Food Chemistry*, *221*, 1421–1425. https://doi.org/10.1016/j.foodchem.2016.10.145

Cretin, B. N., Sallembien, Q., Sindt, L., Daugey, N., Buffeteau, T., Waffo-Teguo, P., … Marchal, A. (2015). How stereochemistry influences the taste of wine: Isolation, characterization and sensory

evaluation of lyoniresinol stereoisomers. *Analytica Chimica Acta*, *888*, 191–198. https://doi.org/10.1016/j.aca.2015.06.061

Deutsch, E. W., & Hansch, C. (1966). Dependence of relative sweetness on hydrophobic bonding [22]. *Nature*, Vol. 211, p. 75. https://doi.org/10.1038/211075a0

Drew, M. G. B., Wilden, G. R. H., Spillane, W. J., Walsh, R. M., Ryder, C. A., & Simmie, J. M. (1998). Quantitative Structure−Activity Relationship Studies of Sulfamates RNHSO $_3$ Na: Distinction between Sweet, Sweet-Bitter, and Bitter Molecules. *Journal of Agricultural and Food Chemistry*, *46*(8), 3016–3026. https://doi.org/10.1021/jf980095c

EMBL-EBI. (2017). *standardiser*. Retrieved from https://github.com/flatkinson/standardiser

Golbraikh, A., & Tropsha, A. (2002). Beware of q^2! *Journal of Molecular Graphics and Modelling*, *20*(4), 269–276. https://doi.org/10.1016/S1093-3263(01)00123-1

Hanser, T., Barber, C., Marchaland, J. F., & Werner, S. (2016). Applicability domain: towards a more formal definition. *SAR and QSAR in Environmental Research*, *27*(11), 893–909. https://doi.org/10.1080/1062936X.2016.1250229

Landrum, G., Kelley, B., Tosco, P., sriniker, gedeck, NadineSchneider, … Avery, P. (2018, April 20). *rdkit/rdkit: 2018_03_1 (Q1 2018) Release*. https://doi.org/https://doi.org/10.5281/zenodo.1222070

Li, X., Staszewski, L., Xu, H., Durick, K., Zoller, M., & Adler, E. (2002). Human receptors for sweet and umami taste. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(7), 4692–4696. https://doi.org/10.1073/pnas.072090199

Marchal, A., Cretin, B. N., Sindt, L., Waffo-Téguo, P., & Dubourdieu, D. (2015). Contribution of oak lignans to wine taste: Chemical identification, sensory characterization and quantification. *Tetrahedron*, *71*(20), 3148–3156. https://doi.org/10.1016/j.tet.2014.07.090

Masuda, K., Koizumi, A., Nakajima, K., Tanaka, T., Abe, K., Misaka, T., & Ishiguro, M. (2012). Characterization of the Modes of Binding between Human Sweet Taste Receptor and Low-Molecular-Weight Sweet Compounds. *PLoS ONE*, *7*(4), e35380. https://doi.org/10.1371/journal.pone.0035380

Moriwaki, H., Tian, Y. S., Kawashita, N., & Takagi, T. (2018). Mordred: A molecular descriptor calculator. *Journal of Cheminformatics*, *10*(1). https://doi.org/10.1186/s13321-018-0258-y

Nelson, G., Hoon, M. A., Chandrashekar, J., Zhang, Y., Ryba, N. J. P., & Zuker, C. S. (2001). Mammalian sweet taste receptors. *Cell*, *106*(3), 381–390. https://doi.org/10.1016/S0092-8674(01)00451-2

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., … Duchesnay, É. (2011). Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Poirier, N., Roudnitzky, N., Brockhoff, A., Belloir, C., Maison, M., Thomas-Danguin, T., … Briand, L. (2012). Efficient Production and Characterization of the Sweet-Tasting Brazzein Secreted by the Yeast Pichia pastoris. *Journal of Agricultural and Food Chemistry*, *60*(39), 9807–9814. https://doi.org/10.1021/jf301600m

Rojas, C., Tripaldi, P., & Duchowicz, P. R. (2016). A New QSPR Study on Relative Sweetness. *International Journal of Quantitative Structure-Property Relationships*, *1*(1), 78–93. https://doi.org/10.4018/ijqspr.2016010104

Ruiz-Aceituno, L., Hernandez-Hernandez, O., Kolida, S., Moreno, F. J., & Methven, L. (2018). Sweetness and sensory properties of commercial and novel oligosaccharides of prebiotic potential. *Lwt*, *97*(April), 476–482. https://doi.org/10.1016/j.lwt.2018.07.038

Servant, G., Tachdjian, C., Tang, X. Q., Werner, S., Zhang, F., Li, X., … Karanewsky, D. S. (2010). Positive allosteric modulators of the human sweet taste receptor enhance sweet taste. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(10), 4746–4751. https://doi.org/10.1073/pnas.0911670107

Shallenberger, R. S., & Acree, T. E. (1967). Molecular theory of sweet taste [16]. *Nature*, Vol. 216, pp. 480–482. https://doi.org/10.1038/216480a0

Sigoillot, M., Brockhoff, A., Neiers, F., Poirier, N., Belloir, C., Legrand, P., … Briand, L. (2018). The

Crystal Structure of Gurmarin, a Sweet Taste–Suppressing Protein: Identification of the Amino Acid Residues Essential for Inhibition. *Chemical Senses*, *43*(8), 635–643. https://doi.org/10.1093/chemse/bjy054

Spillane, W. J., & McGlinchey, G. (1981). Structure—activity studies on sulfamate sweeteners II: Semiquantitative structure-taste relationship for sulfamate (rnhso 3−) sweeteners—the role of R. *Journal of Pharmaceutical Sciences*, *70*(8), 933–935. https://doi.org/10.1002/jps.2600700826

Spillane, W. J., McGlinchey, G., Muircheartaigh, I., & Benson, G. A. (1983). Structure–activity studies on sulfamate sweetners III: Structure–taste relationships for heterosulfamates. *Journal of Pharmaceutical Sciences*, *72*(8), 852–856. https://doi.org/10.1002/jps.2600720804

Spillane, W. J., Ryder, C. A., Curran, P. J., Wall, S. N., Kelly, L. M., Feeney, B. G., & Newell, J. (2000). Development of structure–taste relationships for sweet and non-sweet heterosulfamates†. *Journal of the Chemical Society, Perkin Transactions 2*, (7), 1369–1374. https://doi.org/10.1039/b002482l

Spillane, W. J., Ryder, C. A., Walsh, M. R., Curran, P. J., Concagh, D. G., & Wall, S. N. (1996). Sulfamate sweeteners. *Food Chemistry*, *56*(3), 255–261. https://doi.org/10.1016/0308-8146(96)00022-2

Spillane, W. J., & Sheahan, M. B. (1989). Semi-quantitative and quantitative structure–taste relationships for carboand hetero-sulphamate ($RNHSO_3^-$) sweeteners. *J. Chem. Soc., Perkin Trans. 2*, (7), 741–746. https://doi.org/10.1039/P29890000741

Talete srl. (2014). *Dragon (Software for Molecular Descriptor Calculation)*.

Tuwani, R., Wadhwa, S., & Bagler, G. (2019). BitterSweet: Building machine learning models for predicting the bitter and sweet taste of small molecules. *Scientific Reports*, *9*(1), 1–13. https://doi.org/10.1038/s41598-019-43664-y

Van Der Heijden, A., Brussel, L. B. P., & Peer, H. G. (1979). Quantitative structure-activity relationships (QSAR) in sweet aspartyl dipeptide methyl esters. *Chemical Senses*, *4*(2), 141–152. https://doi.org/10.1093/chemse/4.2.141

Vepuri, S. B., Tawari, N. R., & Degani, M. S. (2007). Quantitative structure-activity relationship study of some aspartic acid analogues to correlate and predict their sweetness potency. *QSAR and Combinatorial Science*, *26*(2), 204–214. https://doi.org/10.1002/qsar.200530191

Walters, D. E. (2006). Analysing and predicting properties of sweet-tasting compounds. In *Optimising Sweet Taste in Foods* (pp. 283–291). https://doi.org/10.1533/9781845691646.3.283

Zheng, S., Chang, W., Xu, W., Xu, Y., & Lin, F. (2019). e-Sweet: A Machine-Learning Based Platform for the Prediction of Sweetener and Its Relative Sweetness. *Frontiers in Chemistry*, *7*(JAN), 35. https://doi.org/10.3389/fchem.2019.00035