



**HAL**  
open science

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

Cédric Bouysset, Christine Belloir, Serge Antonczak, Loïc Briand, Sébastien Fiorucci

### ► **To cite this version:**

Cédric Bouysset, Christine Belloir, Serge Antonczak, Loïc Briand, Sébastien Fiorucci. Novel scaffold of natural compound eliciting sweet taste revealed by machine learning. *Food Chemistry*, 2020, 324, pp.126864. <10.1016/j.foodchem.2020.126864>. <hal-02547525>

**HAL Id: hal-02547525**

**<https://univ-cotedazur.hal.science/hal-02547525v1>**

Submitted on 22 Aug 2022

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY-NC 4.0 - Attribution - Non-commercial use - International License

# 1 Novel scaffold of natural compound eliciting sweet taste 2 revealed by machine learning.

3  
4 Cédric Bouysset<sup>1</sup>, Christine Belloir<sup>2</sup>, Serge Antonczak<sup>1</sup>, Loïc Briand<sup>2</sup>, Sébastien Fiorucci<sup>1\*</sup>

5  
6 <sup>1</sup> Université Côte d'azur, CNRS, Institut de Chimie de Nice UMR7272, 06108 Nice, France

7 <sup>2</sup> INRAE, CNRS, Université de Bourgogne-Franche Comté, AgroSup Dijon, Centre des Sciences du Goût  
8 et de l'Alimentation, 21000 Dijon, France

9  
10 \*Corresponding author: [sebastien.fiorucci@univ-cotedazur.fr](mailto:sebastien.fiorucci@univ-cotedazur.fr)

## 11 Abstract

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

## 19 Keywords

20 Sweet taste, machine learning, natural compounds, sweetener, sweet taste receptor  
21

## 22 Introduction

23 Consumer interest in natural high potency sweeteners has grown spectacularly in recent years, fueled by  
24 concerns about sugar overconsumption and the use of artificial additives in foods. There are three main  
25 strategies to reduce sugar intake: an abrupt reduction of sugar without substitution, the use of flavor  
26 materials to modify sweet taste perception and the use of alternative sweeteners. Though many low-  
27 calorie sweeteners are known, only few of them are used by the food industry (Belloir, Neiers, & Briand,  
28 2017). The search of novel intense sweeteners, possessing the same chemosensory profile as sucrose,  
29 remains open and challenging.

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

36 (Shallenberger & Acree, 1967) to recent structure-taste relationship models (Achary, Toropova, &  
37 Toropov, 2019; Arnoldi, Bassoli, Merlini, & Ragg, 1991; Barker, Hattotuwigama, & Drew, 2002; Bassoli  
38 et al., 2001; Chéron, Casciuc, Golebiowski, Antonczak, & Fiorucci, 2017; Drew et al., 1998; Rojas,  
39 Tripaldi, & Duchowicz, 2016; Spillane & McGlinchey, 1981; Spillane et al., 2000, 1996; Spillane,  
40 McGlinchey, Muircheartaigh, & Benson, 1983; Spillane & Sheahan, 1989; Tuwani, Wadhwa, & Bagler,  
41 2019; Van Der Heijden, Brussel, & Peer, 1979; Vepuri, Tawari, & Degani, 2007; Walters, 2006; Zheng,  
42 Chang, Xu, Xu, & Lin, 2019), the quest to understand the molecular features underlying sweet taste  
43 perception is still active.

44 In this study, we present the first online tool able to predict sweet taste perception based on a machine  
45 learning protocol. We have updated and curated the previous database of 316 sweet compounds  
46 (SweetenersDB) and added new applicability domain metrics to assess the robustness of the predictions.  
47 A novel scaffold of natural sweetener, belonging to the lignan chemical family, that have never been  
48 annotated as sweet have been identified and experimentally validated.

49

## 50 Materials and Methods

### 51 Data preparation

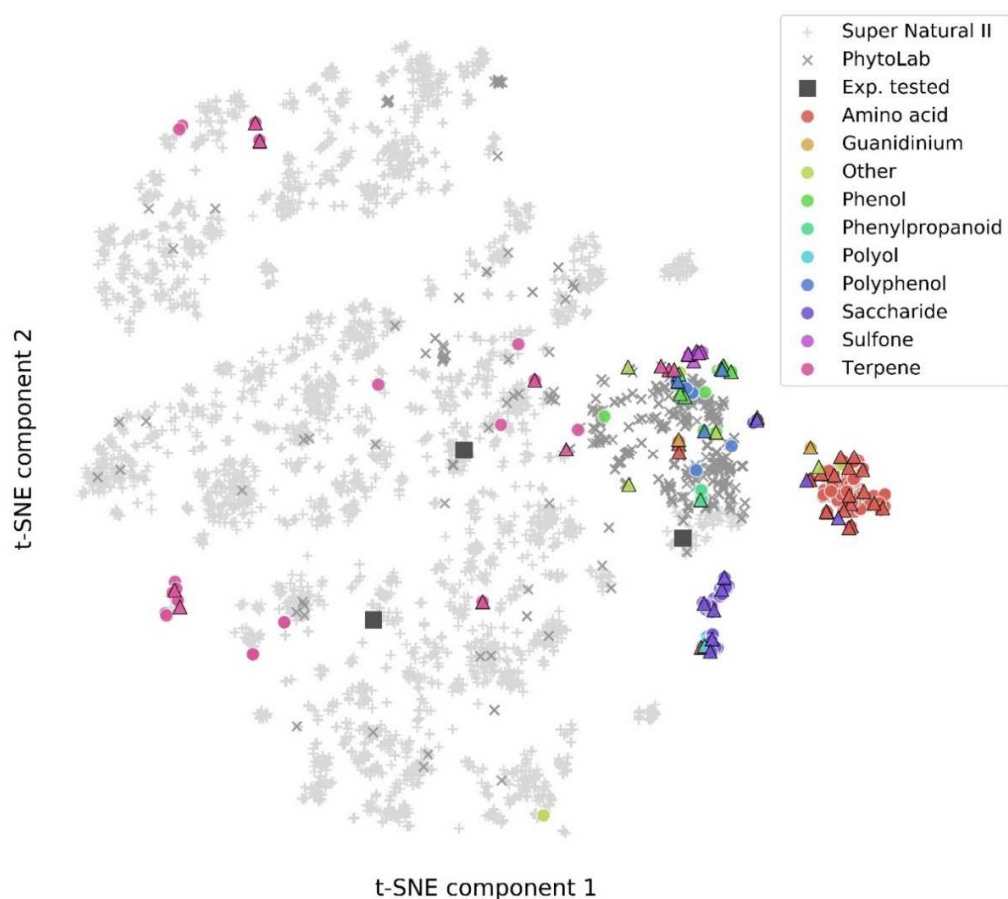
52 Based on our previous work (Chéron et al., 2017), the database of sugars and sweeteners (Figure S1),  
53 named SweetenersDB, was curated and updated with missing compounds (Ruiz-Aceituno, Hernandez-  
54 Hernandez, Kolida, Moreno, & Methven, 2018). Each compound was labelled with a relative sweetness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

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

77 deviation below 0.001, and features with a correlation greater than 0.95. The resulting datasets consisted  
78 of 635 descriptors for the Dragon dataset, and 506 features for the “open-source” dataset. To avoid any  
79 model bias due to overfitting, the number of features used by the model is a hyperparameter that has been  
80 optimized.

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

90



91  
92 **Figure 1:** Representation of the SweetenersDB chemical space based on a t-SNE dimensionality  
93 reduction method. Known sweet chemical families in the training and test set are represented by circle  
94 and triangles respectively. Light and dark grey data points represent natural compounds that were  
95 predicted as intensely sweet ( $\log Sw \geq 2$ ) by both our previous and current models (Table S2). Grey  
96 squares represent natural molecules experimentally tested in the present study.

97

## 98 Machine-learning model for sweetness prediction

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

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

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

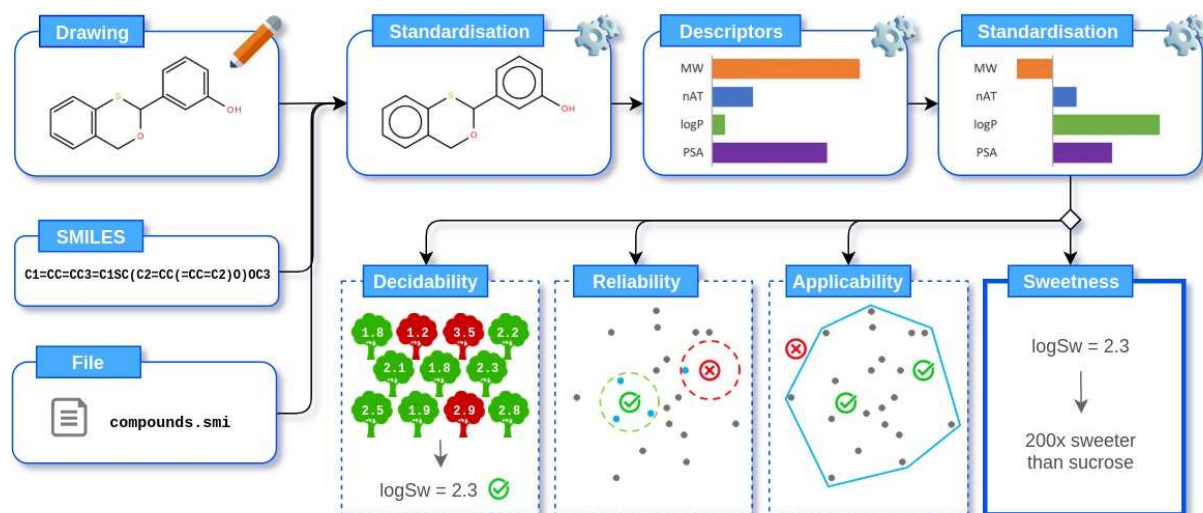
117 Other chemoinformatics solutions are available but none of them has been implemented on a webserver.  
118 For instance, the e-Sweet platform (Zheng et al., 2019) is based on a consensus model of various machine  
119 learning protocols. The database used to train and test their model is very similar to the database used to  
120 setup Predisweet and e-Sweet performs as well as our model ( $R^2$  on the test set is in the same range [0.75-  
121 0.78] for both solutions). Recently a new functionality to predict sweetness has been implemented on the  
122 BitterSweet webserver (Tuwani et al., 2019). The performance of BitterSweet is comparable to e-Sweet  
123 and Predisweet ( $R^2$  of 0.72 on our test set) but the protocol is still unpublished, and seven molecules of  
124 the test set has not been considered as sweet.

125

## 126 Webserver interface

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

140 made. The decidability score is based on the weights of each decision tree that compose the AdaBoost  
 141 model. It is computed by summing the weights of decision trees that made a prediction close to the model  
 142 prediction and dividing it by the sum of all weights.  
 143 Each molecule is indexed in the database with its InChIKey, which avoids making predictions for the  
 144 same molecule twice. For a seamless user experience, the name of each molecule is retrieved by querying  
 145 PubChem with the pubchempy Python package, and a 2D representation of the compound is generated  
 146 with RDKit.  
 147



148  
 149 **Figure 2:** Workflow followed by each molecule submitted to the webserver.  
 150

## 151 Functional expression of the human sweet taste receptor

152 In order to validate the sweetness of the three natural compounds, we employed a cell-based expression  
 153 system for the human T1R2/T1R3 sweet taste receptor as previously described (Poirier et al., 2012;  
 154 Sigoillot et al., 2018). Briefly, the cDNAs coding human T1R2 and T1R3 subunits were cloned into  
 155 pcDNA3 and pcDNA4 expression plasmids, respectively. HEK293T cells stably expressing Gα16gust44  
 156 and T1R3 were seeded at a density of  $0.4 \times 10^6$  cells per well into 96-well black walled, clear bottom  
 157 microtiter plates (Falcon) in high-glucose DMEM supplemented with 2 mM GlutaMAX, 10% dialyzed  
 158 foetal bovine serum, penicillin/streptomycin, G418 (400 µg/mL) and zeocin (250 µg/mL) at 37 °C and  
 159 6.3% CO<sub>2</sub>, in a humidified atmosphere. Twenty-four hours later, HEK293T-Gα16gust44-T1R3 cells were  
 160 transiently transfected with pcDNA3-T1R2 (120ng/well) with Lipofectamine 2000. Calcium signal of  
 161 mock-transfected cells (HEK293T Gα16gust44 cells stably expressing T1R3 transfected with pcDNA3  
 162 empty vector) were always measured in parallel and compared. Twenty-four hours after transfection, the  
 163 cells were loaded for 1 hour at 37°C with the calcium indicator Fluo4-AM (Molecular Probes) diluted in  
 164 C1 buffer (130 mM NaCl, 5 mM KCl, 10 mM Hepes pH 7.4, 2 mM CaCl<sub>2</sub>) in the presence of pluronic  
 165 acid (0.025%, w/v) and probenecid (2.5 mM). After washing with C1 buffer, cells were stimulated with a  
 166 range of sweet tasting compounds. The fluorescence intensity was measured for 90 seconds (excitation  
 167 488 nm, emission 510 nm) into an automated fluorimetric FlexStation®3 Multi-Mode microplate reader.  
 168 The change in fluorescence upon stimulus application were averaged, mock-subtracted and baseline-

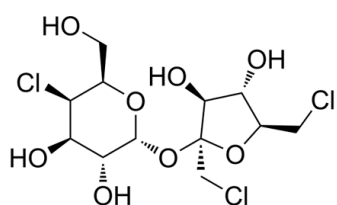
169 corrected. The EC<sub>50</sub> values were calculated using SigmaPlot software by nonlinear regression using the  
170 function:

171 
$$f(x) = \min + \frac{\max - \min}{1 + \left(\frac{x}{EC_{50}}\right)^{-\text{Hillslope}}}$$

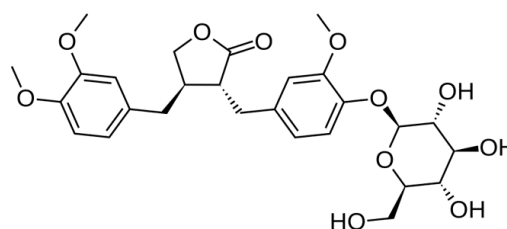
172

### 173 Chemicals

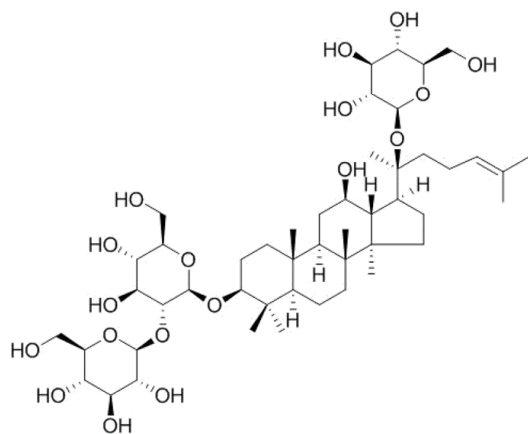
174 All tested compounds (arctiin, ginsenoside Rd and jujuboside A, Figure 3) were purchased from Phytolab  
175 GmbH & Co. KG, with the exception of sucralose obtained from Sigma-Aldrich. All the compounds were  
176 dissolved first in DMSO (100 mM in 100% DMSO), and then diluted with the C1 buffer solution; except  
177 for sucralose, which was dissolved in the C1 buffer solution directly.  
178



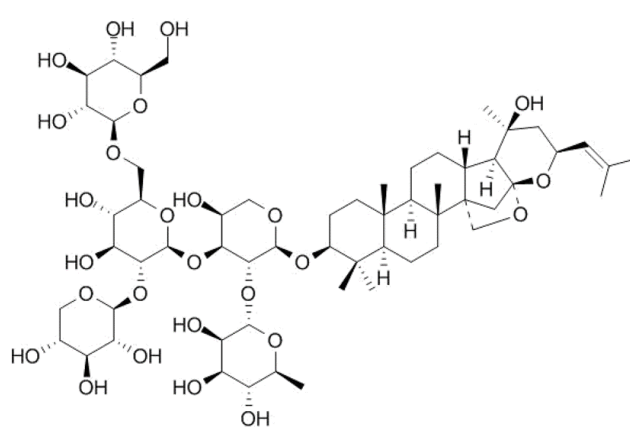
sucralose



arctiin



ginsenoside Rd



jujuboside A

179

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

### 181 Results and discussion

182 New machine-learning model based on open-source features

183 The performance of the Open-source and Dragon models has been compared. Both models show good  
184 predictivity on the test set according to state of the art QSAR rules (Table 1). Slightly more than 90% of

185 the test set are predicted with an absolute error lower than a log unit (Figure S3). The models are less  
 186 accurate for high sweetness values since they have been trained with less information for highly potent  
 187 sweeteners. Improving the quality of the machine learning model would then requires i) expanding the  
 188 chemical diversity of sweet compounds and ii) a larger database of *in vivo* and *in vitro* experiments. A  
 189 threshold of LogSw larger than 2 has then been chosen to minimize false positive predictions prior *in*  
 190 *vitro* validation. Since similar performance have been obtained for both models, the open-source version  
 191 have been implemented on a webserver, freely accessible at the following address:  
 192 <http://chemosimserver.unice.fr/predisweet/>. Another model has been set up with descriptors calculated at  
 193 salivary pH to assess the effect of the protonation state on the model performance. Even though more than  
 194 a quarter of the molecules had different descriptor values between the default and the salivary pH dataset,  
 195 there was no significant difference in terms of performance. The protonation assessment step thus has  
 196 been skipped in the final protocol. We emphasize that the model has not been trained to predict bitter taste  
 197 and we envision to include this feature in a future work. Additionally, any QSAR model has a field of  
 198 application that clearly defines the boundaries within which the model should be used, usually referred to  
 199 as the applicability domain. We've implemented three different metrics to explicitly inform the user  
 200 whether the model and its prediction can be trusted for a particular query molecule.

202 Table 1: Performance of the models according to Golbraikh and Tropsha rules. (Golbraikh & Tropsha,  
 203 2002)

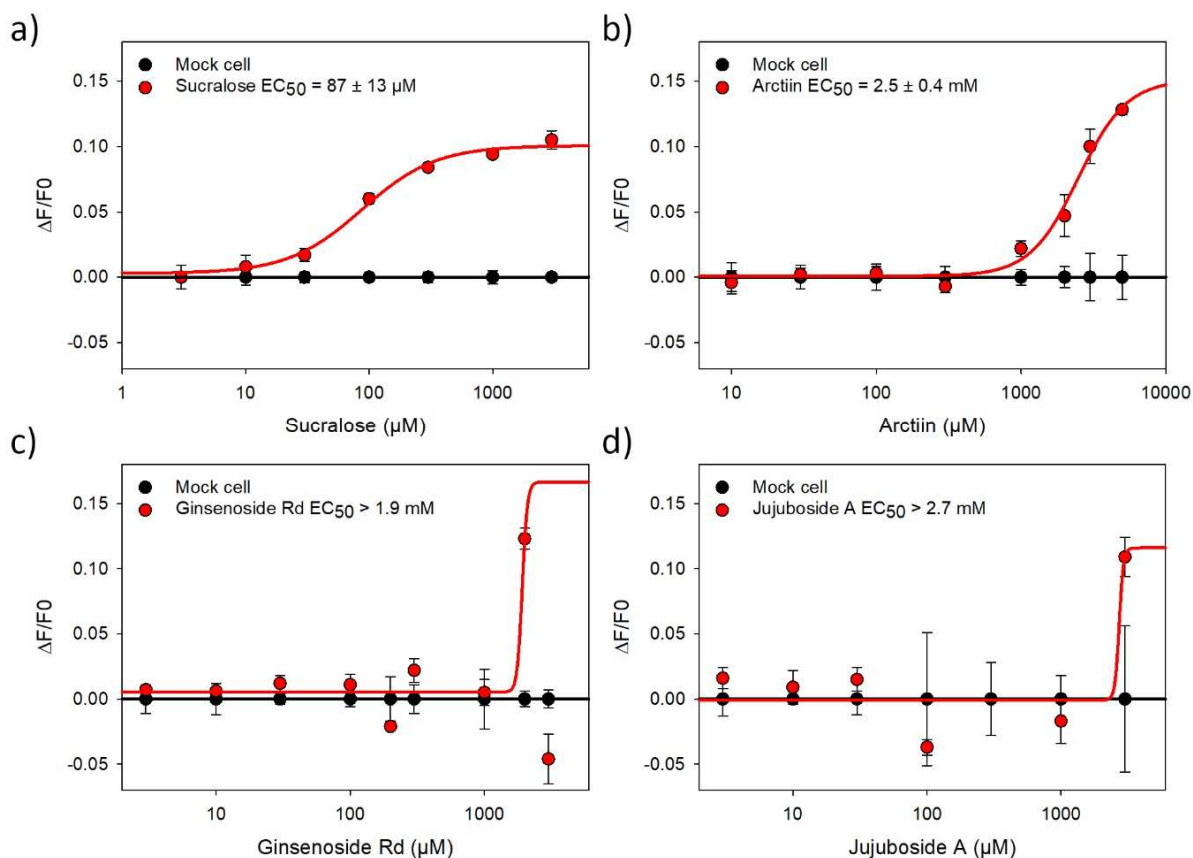
Rules	Open-source model	Dragon model
$R^2 > 0.6$	0.74	0.75
$Q^2 > 0.5$	0.84	0.79
$ R^2 - R_0^2 /R^2 < 0.1$	0.02	0.05
$0.85 \leq k \leq 1.15$	0.93	0.90
$ R_0^2 - R_0'^2  < 0.3$	0.07	0.12

204

## 205 Identification of a new sweet scaffold

206 A large database of natural compounds has been virtually screened to identify new putative sweeteners.  
 207 The analysis of the resulting sweet chemical space of ~4800 natural compounds shows that it does not  
 208 fully overlap the chemical space of known sweeteners (Figure 1). It suggests that a large part of the  
 209 natural chemical space remains unexplored. We have finally selected three natural compounds that have  
 210 been tested for their ability to activate the human sweet taste receptor T1R2/T1R3 expressed in HEK  
 211 cells, as previously reported (Poirier et al., 2012). As a negative control, HEK293T Ga16gust44 cells  
 212 stably expressing T1R3 were mock-transfected with the empty expression vector to control for T1R2-  
 213 independent non-specific signals. In addition to a LogSw value higher than 2, the price and the  
 214 commercial availability were two important criteria in the compound choice. Two of them, Jujuboside A  
 215 and Ginsenisode Rd, belong to the triterpene chemical family. The third one, arctiin, possesses a lignan  
 216 scaffold. As shown in Figure 4b, application of arctiin on T1R2/T1R3-expressing cells evoked calcium  
 217 responses in a dose-dependent manner, while no fluorescence signals were observed with mock

218 transfected cells. The half-maximal effective concentrations ( $EC_{50}$ ) of arctiin was  $2.5 \pm 0.4$  mM. As a  
 219 control, we determined the concentration-response curve for the high-intensity sucralose (Figure 4a)  
 220 leading to an  $EC_{50}$  value of  $87 \pm 13$   $\mu$ M, in agreement with reported values (Assadi-Porter et al., 2010;  
 221 Masuda et al., 2012; Servant et al., 2010). In contrast, jujuboside A and ginsenoside Rd showed  
 222 detectable activity on the T1R2/T1R3 receptor, but only at the highest tested concentration (Figure 4c and  
 223 d) precluding establishment of complete dose-response curve and calculation of  $EC_{50}$  values. This  
 224 concentration used was the maximum one that did not induce any side effects on mock transfected cells.  
 225



226  
 227 **Figure 4:** Response of the human sweet taste receptor to the three natural compounds identified by the  
 228 machine learning protocol and sucralose used as a control. Dose-response curves of T1R2/T1R3-  
 229 expressing cells (red curve) and mock-transfected cells (black curve). All concentrations were measured  
 230 in triplicate and each experiment was repeated at least 2 times.  
 231

## 232 Conclusion

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

238 response analyses. Among them, we identified arctiin as a novel agonist of the T1R2/T1R3 sweet taste  
239 receptor with an EC<sub>50</sub> value of 2.5±0.4mM. It belongs to the lignan chemical family, polyphenols found  
240 in plants, of which epi-lyoniresinol has already been annotated as slightly sweet by sensory analyses  
241 (Cretin et al., 2015; Marchal, Cretin, Sindt, Waffo-Téguo, & Dubourdieu, 2015). As numerous natural  
242 sweeteners, arctiin might also possess bitter taste but it would require additional experiments out of the  
243 scope of the present study to assess its aftertaste. Nevertheless, our results confirm that the lignan  
244 chemical family opens a new chemical space for the search of new sweet agents and machine learning is a  
245 fruitful approach in this context.  
246

## 247 Acknowledgements

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

## 255 References

- 256 Achary, P. G. R., Toropova, A. P., & Toropov, A. A. (2019). Combinations of graph invariants and  
257 attributes of simplified molecular input-line entry system (SMILES) to build up models for  
258 sweetness. *Food Research International*, 122, 40–46. <https://doi.org/10.1016/j.foodres.2019.03.067>
- 259 Arnoldi, A., Bassoli, A., Merlini, L., & Ragg, E. (1991). Isovanillyl sweeteners. Synthesis,  
260 conformational analysis, and structure–activity relationship of some sweet oxygen heterocycles. *J.*  
261 *Chem. Soc., Perkin Trans. 2*, (9), 1399–1406. <https://doi.org/10.1039/P29910001399>
- 262 Assadi-Porter, F. M., Maillet, E. L., Radek, J. T., Quijada, J., Markley, J. L., & Max, M. (2010). Key  
263 Amino Acid Residues Involved in Multi-Point Binding Interactions between Brazzein, a Sweet  
264 Protein, and the T1R2-T1R3 Human Sweet Receptor. *Journal of Molecular Biology*, 398(4), 584–  
265 599. <https://doi.org/10.1016/j.jmb.2010.03.017>
- 266 Barker, J. S., Hattotuagama, C. K., & Drew, M. G. B. (2002). Computational studies of sweet-tasting  
267 molecules. *Pure and Applied Chemistry*, 74(7), 1207–1217.  
268 <https://doi.org/10.1351/pac200274071207>
- 269 Bassoli, A., Drew, M. G. B., Hattotuagama, C. K., Merlini, L., Morini, G., & Wilden, G. R. H. (2001).  
270 Quantitative Structure-Activity Relationships of Sweet Isovanillyl Derivatives. *Quantitative*  
271 *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](https://doi.org/10.1002/1521-3838(200105)20:1<3::AID-QSAR3>3.0.CO;2-H)
- 272 Belloir, C., Neiers, F., & Briand, L. (2017). Sweeteners and sweetness enhancers. *Current Opinion in*  
273 *Clinical Nutrition and Metabolic Care*, 20(4), 279–285.  
274 <https://doi.org/10.1097/MCO.0000000000000377>
- 275 Cao, D. S., Xu, Q. S., Hu, Q. N., & Liang, Y. Z. (2013). ChemoPy: Freely available python package for  
276 computational biology and chemoinformatics. *Bioinformatics*, 29(8), 1092–1094.  
277 <https://doi.org/10.1093/bioinformatics/btt105>
- 278 ChemAxon. (2018). *Calculator Plugins*. Retrieved from <http://www.chemaxon.com>
- 279 Chéron, J. B., Casciuc, I., Golebiowski, J., Antonczak, S., & Fiorucci, S. (2017). Sweetness prediction of  
280 natural compounds. *Food Chemistry*, 221, 1421–1425.  
281

282 <https://doi.org/10.1016/j.foodchem.2016.10.145>

283 Cretin, B. N., Sallembien, Q., Sindt, L., Daugey, N., Buffeteau, T., Waffo-Teguo, P., ... Marchal, A.

284 (2015). How stereochemistry influences the taste of wine: Isolation, characterization and sensory

285 evaluation of lyoniresinol stereoisomers. *Analytica Chimica Acta*, 888, 191–198.

286 <https://doi.org/10.1016/j.aca.2015.06.061>

287 Deutsch, E. W., & Hansch, C. (1966). Dependence of relative sweetness on hydrophobic bonding [22].

288 *Nature*, Vol. 211, p. 75. <https://doi.org/10.1038/211075a0>

289 Drew, M. G. B., Wilden, G. R. H., Spillane, W. J., Walsh, R. M., Ryder, C. A., & Simmie, J. M. (1998).

290 Quantitative Structure–Activity Relationship Studies of Sulfamates RNHSO<sub>3</sub> Na: Distinction

291 between Sweet, Sweet-Bitter, and Bitter Molecules. *Journal of Agricultural and Food Chemistry*,

292 46(8), 3016–3026. <https://doi.org/10.1021/jf980095c>

293 EMBL–EBI. (2017). *standardiser*. Retrieved from <https://github.com/flatkinson/standardiser>

294 Golbraikh, A., & Tropsha, A. (2002). Beware of q<sup>2</sup>! *Journal of Molecular Graphics and Modelling*,

295 20(4), 269–276. [https://doi.org/10.1016/S1093-3263\(01\)00123-1](https://doi.org/10.1016/S1093-3263(01)00123-1)

296 Hanser, T., Barber, C., Marchaland, J. F., & Werner, S. (2016). Applicability domain: towards a more

297 formal definition. *SAR and QSAR in Environmental Research*, 27(11), 893–909.

298 <https://doi.org/10.1080/1062936X.2016.1250229>

299 Landrum, G., Kelley, B., Tosco, P., sriniker, gedec, NadineSchneider, ... Avery, P. (2018, April 20).

300 *rdkit/rdkit: 2018\_03\_1 (Q1 2018) Release*. <https://doi.org/https://doi.org/10.5281/zenodo.1222070>

301 Li, X., Staszewski, L., Xu, H., Durick, K., Zoller, M., & Adler, E. (2002). Human receptors for sweet and

302 umami taste. *Proceedings of the National Academy of Sciences of the United States of America*,

303 99(7), 4692–4696. <https://doi.org/10.1073/pnas.072090199>

304 Marchal, A., Cretin, B. N., Sindt, L., Waffo-Tégou, P., & Dubourdieu, D. (2015). Contribution of oak

305 lignans to wine taste: Chemical identification, sensory characterization and quantification.

306 *Tetrahedron*, 71(20), 3148–3156. <https://doi.org/10.1016/j.tet.2014.07.090>

307 Masuda, K., Koizumi, A., Nakajima, K., Tanaka, T., Abe, K., Misaka, T., & Ishiguro, M. (2012).

308 Characterization of the Modes of Binding between Human Sweet Taste Receptor and Low-

309 Molecular-Weight Sweet Compounds. *PLoS ONE*, 7(4), e35380.

310 <https://doi.org/10.1371/journal.pone.0035380>

311 Moriwaki, H., Tian, Y. S., Kawashita, N., & Takagi, T. (2018). Mordred: A molecular descriptor

312 calculator. *Journal of Cheminformatics*, 10(1). <https://doi.org/10.1186/s13321-018-0258-y>

313 Nelson, G., Hoon, M. A., Chandrashekar, J., Zhang, Y., Ryba, N. J. P., & Zuker, C. S. (2001).

314 Mammalian sweet taste receptors. *Cell*, 106(3), 381–390. [https://doi.org/10.1016/S0092-](https://doi.org/10.1016/S0092-8674(01)00451-2)

315 [8674\(01\)00451-2](https://doi.org/10.1016/S0092-8674(01)00451-2)

316 Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... Duchesnay, É. (2011).

317 Scikit-learn: Machine Learning in Python. *Journal of Machine Learning Research*, 12, 2825–2830.

318 Poirier, N., Roudnitzky, N., Brockhoff, A., Belloir, C., Maison, M., Thomas-Danguin, T., ... Briand, L.

319 (2012). Efficient Production and Characterization of the Sweet-Tasting Brazzein Secreted by the

320 Yeast *Pichia pastoris*. *Journal of Agricultural and Food Chemistry*, 60(39), 9807–9814.

321 <https://doi.org/10.1021/jf301600m>

322 Rojas, C., Tripaldi, P., & Duchowicz, P. R. (2016). A New QSPR Study on Relative Sweetness.

323 *International Journal of Quantitative Structure-Property Relationships*, 1(1), 78–93.

324 <https://doi.org/10.4018/ijqspr.2016010104>

325 Ruiz-Aceituno, L., Hernandez-Hernandez, O., Kolida, S., Moreno, F. J., & Methven, L. (2018).

326 Sweetness and sensory properties of commercial and novel oligosaccharides of prebiotic potential.

327 *Lwt*, 97(April), 476–482. <https://doi.org/10.1016/j.lwt.2018.07.038>

328 Servant, G., Tachdjian, C., Tang, X. Q., Werner, S., Zhang, F., Li, X., ... Karanewsky, D. S. (2010).

329 Positive allosteric modulators of the human sweet taste receptor enhance sweet taste. *Proceedings of*

330 *the National Academy of Sciences of the United States of America*, 107(10), 4746–4751.

331 <https://doi.org/10.1073/pnas.0911670107>

332 Shallenberger, R. S., & Acree, T. E. (1967). Molecular theory of sweet taste [16]. *Nature*, Vol. 216, pp.

333 480–482. <https://doi.org/10.1038/216480a0>

334 Sigoillot, M., Brockhoff, A., Neiers, F., Poirier, N., Belloir, C., Legrand, P., ... Briand, L. (2018). The  
335 Crystal Structure of Gurmarin, a Sweet Taste–Suppressing Protein: Identification of the Amino Acid  
336 Residues Essential for Inhibition. *Chemical Senses*, 43(8), 635–643.  
337 <https://doi.org/10.1093/chemse/bjy054>

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

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

344 Spillane, W. J., Ryder, C. A., Curran, P. J., Wall, S. N., Kelly, L. M., Feeney, B. G., & Newell, J. (2000).  
345 Development of structure–taste relationships for sweet and non-sweet heterosulfamates †. *Journal of*  
346 *the Chemical Society, Perkin Transactions 2*, (7), 1369–1374. <https://doi.org/10.1039/b002482i>

347 Spillane, W. J., Ryder, C. A., Walsh, M. R., Curran, P. J., Concagh, D. G., & Wall, S. N. (1996).  
348 Sulfamate sweeteners. *Food Chemistry*, 56(3), 255–261. [https://doi.org/10.1016/0308-](https://doi.org/10.1016/0308-8146(96)00022-2)  
349 [8146\(96\)00022-2](https://doi.org/10.1016/0308-8146(96)00022-2)

350 Spillane, W. J., & Sheahan, M. B. (1989). Semi-quantitative and quantitative structure–taste relationships  
351 for carboand hetero-sulphamate (RNHSO<sub>3</sub><sup>-</sup>) sweeteners. *J. Chem. Soc., Perkin Trans. 2*, (7), 741–  
352 746. <https://doi.org/10.1039/P29890000741>

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

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

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

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

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

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

368