

VirtualTaste: a web server for the prediction of organoleptic properties of chemical compounds

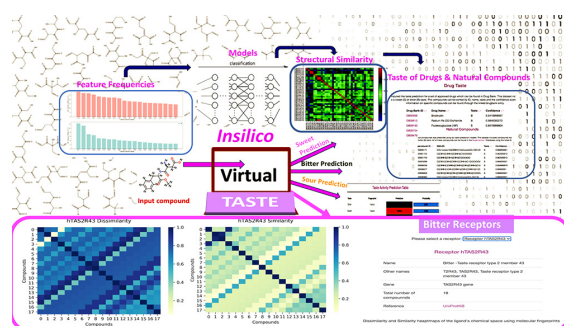
Franziska Fritz¹, Robert Preissner^{2,*} and Priyanka Banerjee^{1,*}¹Institute of Physiology, Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Philippstrasse 12, 10115, Berlin, Germany and²Institute of Physiology and Science-IT, Charité - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Philippstrasse 12, 10115, Berlin, Germany

Received March 17, 2021; Revised April 07, 2021; Editorial Decision April 08, 2021; Accepted April 09, 2021

ABSTRACT

Taste is one of the crucial organoleptic properties involved in the perception of food by humans. Taste of a chemical compound present in food stimulates us to take in food and avoid poisons. Bitter taste of drugs presents compliance problems and early flagging of potential bitterness of a drug candidate may help with its further development. Similarly, the taste of chemicals present in food is important for evaluation of food quality in the industry. In this work, we have implemented machine learning models to predict three different taste endpoints—sweet, bitter and sour. The VirtualTaste models achieved an overall accuracy of 90% and an AUC of 0.98 in 10-fold cross-validation and in an independent test set. The web server takes a two-dimensional chemical structure as input and reports the chemical's taste profile for three tastes—using molecular fingerprints along with confidence scores, including information on similar compounds with known activity from the training set and an overall radar chart. Additionally, insights into 25 bitter receptors are also provided via target prediction for the predicted bitter compounds. VirtualTaste, to the best of our knowledge, is the first freely available web-based platform for the prediction of three different tastes of compounds. It is accessible via <http://virtualtaste.charite.de/VirtualTaste/> without any login requirements and is free to use.

GRAPHICAL ABSTRACT



INTRODUCTION

Organoleptic properties such as taste and smell are very important for the evaluation of quality of products in the food and pharmaceutical industries (1). Taste is an important organoleptic property and is one of the crucial senses involved in the perception of food by humans (2). Taste is simulated when fundamental nutrients or harmful compounds, such as toxic molecules, activate specialized receptors located in taste buds. Taste of a chemical compound present in food stimulates us to take in nutrients and avoid poisons (3). The major gustatory receptors from the largest G protein-coupled receptor (GPCR) family in mammals are responsible for sensing taste molecules (4). The taste prediction of a compound is of considerable interest in the food industry. However, the science of taste and the molecules that initiate it, is no longer the exclusive domain of food research. Many active ingredients present in drugs taste bitter and thus are aversive to children as well as many adults. The bitterness of medicines presents compliance problems, and early flagging of potential bitterness of a drug candidate may help with its further development (5). Similarly, sour taste is the key element in the flavor profile of food acidulants (6). Understanding the chemistry and physiology of sour taste is

*To whom correspondence should be addressed. Tel: +49 30 450 528 505; Fax: +49 30 450 540 955; Email: priyanka.banerjee@charite.de
Correspondence may also be addressed to Robert Preissner. Email: robert.preissner@charite.de

Present address: Robert Preissner and Priyanka Banerjee, Institute of Physiology, Charité-University Medicine Berlin, Philippstrasse 12, Berlin, 10115 Berlin, Germany.

critical for efficient control of flavor in the formulation of acid and acidified foods (7). Sour taste is the aspect of flavor most commonly associated with acids, but they are also able to elicit non-sour taste characteristics such as bitterness (8). Therefore, efficient tools for predicting and masking the taste of ingredients are often sought after by pharmaceutical and food industries.

Chemoinformatic-based analysis of chemical features present in bitter and sweet compounds revealed significant information (9). The chemical structure of sweet tasting compounds is known to be incredibly diverse (10). The list of chemically diverse sweet tasting compounds is long and it includes structural classes like heterocyclics (saccharin, acesulfame K); amino acids (glycine, D-tryptophan), dipeptides (aspartame, neotame), sulfamates (cyclamate), halogenated sugars (sucralose), terpenes and terpene glycosides (hernandulcin, stevioside, rebaudiosides), polyols (sorbitol, maltitol, lactitol), urea derivatives (dulcin, superaspartame, suosan), oximes (perillartine) and nitroanilines (11). Additionally, a number of proteins are known to have a sweet taste (4). On the other hand, bitter agonists include plant-derived and synthetic compounds such as amides, peptides, heterocyclic compounds, glycosides, alkaloids, terpenoids, phenols and flavonoids (2). The sour taste receptors are triggered by acids, more specifically hydrogen ions (H^+) (6).

Traditionally, discovering the taste of compounds is done using the human taste panel or cell-based high throughput screening (9). This process is not only time-consuming and expensive but also laborious. Additionally, the use of sensory panellists is also challenging because of the potential toxicity related to the chemical as well as subjectivity of taste panellists (3). There is certainly a strong rationale to apply predictive models to the initial stages of the innovation pipeline, where a larger pool of compounds is available. Hence, computational models can provide significant alternatives to rapidly identify the taste of chemical compounds. Furthermore, with the increasing influx of chemicals, traditional discovery processes will be facilitated using the computational models and as a result the chemicals can be tested in a timely manner (12). Over the last decade, artificial intelligence and machine learning models played an ever-increasing role in understanding the physicochemical properties and activities of chemical compounds (13). These efforts have led to some significant prediction of novel activity endpoints and understanding of the mode of action of these compounds in the field of chemical research, drug discovery and also food research to some extent (9). However, it is prudent to admit that the models are only as good as the data they are based on; there is still no exception to the GIGO principle 'garbage in, garbage out'. Several computational methods have been published in the literature to either predict 'bitter taste' (14) or 'sweet taste' (15) of chemical compounds. The BitterSweetForest Model is the first published model which was developed to predict bitter and sweet taste, as well as bitter-sweet features present in a chemical compound (9).

In this study, we present the first online prediction server which predicts three different taste endpoints: sweet, bitter and sour. The VirtualTaste models for sweet, bitter and sour taste were validated by both internal 10-fold cross-

validation and external validation with performance evaluated by five different performance metrics: accuracy, sensitivity, specificity, area under the receiver operating curve (ROC-AUC) and F1-measure. The top-performing models were applied to predict the taste of approved drugs from DrugBank (16) and natural compounds from the Super-Natural II database (17). Furthermore, the similarity-based method was applied to predict the bitter receptors, which could provide insights into the mechanism of the predicted bitter taste for compounds.

MATERIALS AND METHODS

Software implementation

The VirtualTaste prediction platform was developed as an integrated, searchable for predicted compounds and machine-learning knowledge-based web server. The pre-predicted data for approved drugs and natural compounds are stored in a relational MySQL database. For similarity search using OpenBabel (<http://openbabel.org/>), data is stored in a structure-data file (SDF) format. The web server was developed using PHP7, Python, JavaScript and ChemDoodle Web components (<https://web.chemdoodle.com/>), an open source JavaScript library for the chemistry interface. RDKit package (<http://www.rdkit.org/>) was used for handling the storing and representation of the chemical data in the database and the webserver. The web server back-end is built using PHP and Python; web access is enabled via the Apache HTTP Server. Redis is employed for queuing and assessing the API requests. The server has been tested on the recent version of Mozilla Firefox, Google Chrome and Apple Safari.

Input and output

The VirtualTaste web server consists of several features—such as three different taste models (sweet, bitter and sour), target prediction for 25 bitter receptors and the predicted taste of approved drugs and natural compounds with higher confidence (above 75%). The user can submit small molecules in four different ways into the web server via the tab 'Prediction'. The molecule can be uploaded as a standard molecule file, molecule name, SMILES (Simplified Molecular-Input Line-Entry System) string of the compound or the user can simply sketch their molecule of interest. Optionally, the user may select different models or all models for prediction. The results are displayed in tabular format, and include the molecular structure with the physicochemical properties and the three most similar molecules from the training set to define the applicability domain of the VirtualTaste prediction models. The user can access the result in the results section or can download the important information in comma-separated values (csv) file format. These prediction results are also displayed as a radar plot comparing the average confidence score of the active compounds in the training set of each taste model, to that of the input compound (see Figure 1). Additionally, if the compound is predicted as bitter, the potential bitter receptors profile of the compound is also provided in the 'Target prediction' table.

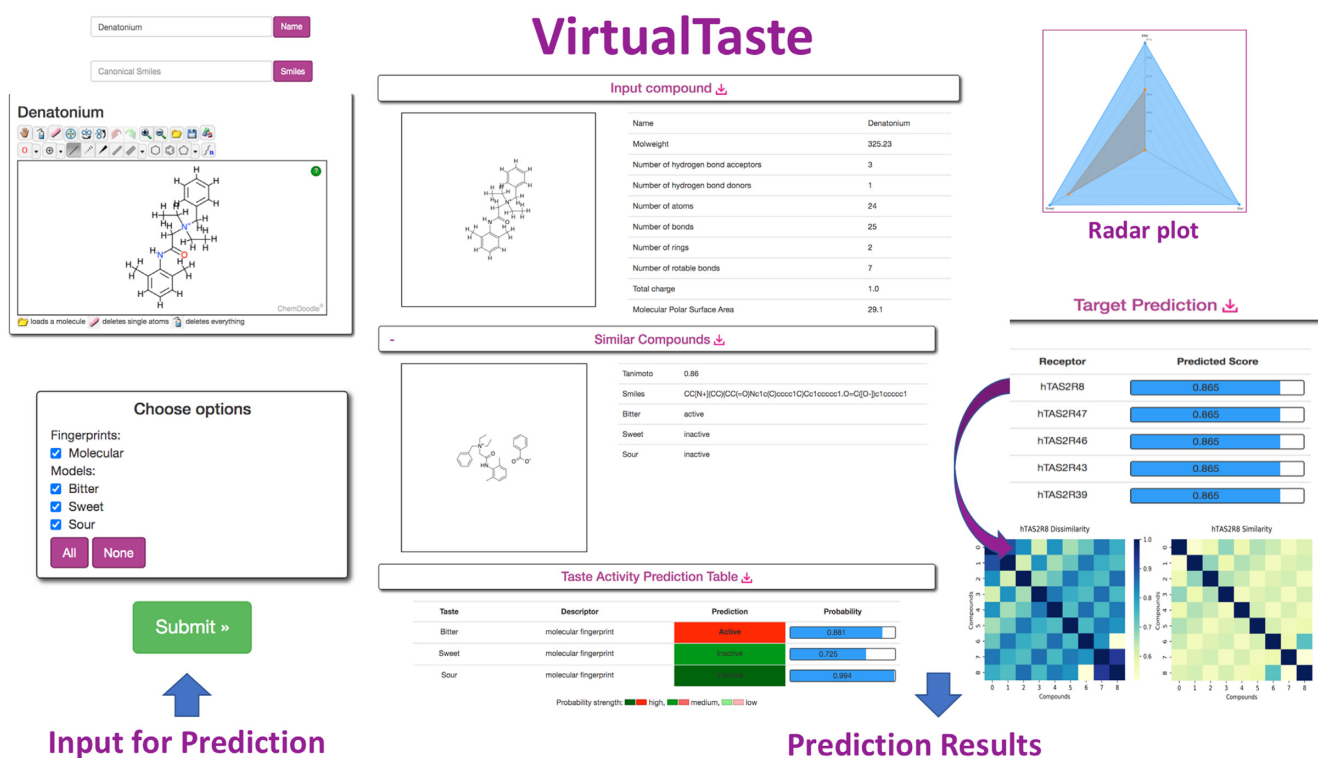


Figure 1. Illustration of an example compound (Denatonium) used as an application case. Denatonium is the input compound; the user can choose either a single or all endpoints for the prediction. In this case all taste endpoints were selected. The results displayed show the taste profile of the input compound. The result page also includes information on similar compounds, overall radar plot, and bitter receptor (target) prediction.

Datasets

The data was collected from literature sources and different publicly available databases (18). Ambiguous compounds, salts and mixtures, as well as entries classified as inconclusive were removed from the final dataset. The data were standardized using the RDkit node of the KN-IME analytics platform (19). For each of the three VirtualTaste models, compounds were divided into training and external validation sets, keeping the ratio of the actives (sweet/bitter/sour) and inactive (non-sweet/non-bitter/non-sour) constant (see S1). Brief descriptions on the respective datasets are as follows:

Sweet data: The sweet data was taken from the SuperSweet database (11) and from our published work BitterSweet-Forest (9). The total number of compounds in the training set is 1608 molecules, and the test set contained 403 molecules.

Bitter data: The bitter compound data was taken from the BitterDB (20) and from our published work BitterSweet-Forest (9). The total number of compounds in the training set is 1289 molecules, and the test set contained 323 molecules.

Sour data: The sour data was extracted from public databases (18) and manually curated from literature sources from the PubMed database (<https://pubmed.ncbi.nlm.nih.gov/>) (21). The dataset consists of 1214 training set molecules, and 133 test set molecules.

Bitter Receptor data: A dataset of diverse ligands that interact with the 25 human bitter taste receptors (TAS2Rs)

which belong to the superfamily of G-protein-coupled receptors (GPCRs) (8), was extracted from various publicly available databases (18,20) and from literature using a text mining approach (21). A total of 356 ligands were extracted which interacts with 25 human T2Rs.

Performance evaluation of the virtualtaste models

Each model was validated using 8 different sampling methods for 10-fold cross-validation (22). The data was split into 10-fold, keeping the ratio of active and inactive data constant as published in our previous work (22). Additionally, an external set was used for the evaluation of the predictive performance of each model. Each model was evaluated by the following performance metrics:

- **Prediction Accuracy** is defined as the ability of a model to correctly predict the total number of activities and inactives.
- **Sensitivity** is the trained model's ability to correctly predict the positive (taste) class.
- **Specificity** is defined as the trained model's ability to correctly predict the negative (non-taste) class.
- **The area under the curve (AUC) of a receiver operating characteristic (ROC) curve** is computed using the ROC curve which plots the true positive rate (sensitivity) against the false positive rate (1 – specificity) at different thresholds. The value of AUC ranges from 0.50 (random classifiers) to 1.00 (perfect classifiers) (22).

Table 1. Performance statistics for the VirtualTaste models applied to cross-validation and external validation sets

VirtualTaste models Data sampling method		VirtualSweet SMOTETC	VirtualBitter SMOTE VDM	VirtualSour AugRandOS
Cross-validation	<i>Prediction accuracy</i>	0.88	0.94	0.98
	<i>Sensitivity</i>	0.97	0.94	0.94
	<i>Specificity</i>	0.96	0.92	0.97
	<i>ROC-AUC</i>	0.99	0.97	0.97
	<i>F-measure</i>	0.87	0.94	0.98
External validation	<i>Prediction accuracy</i>	0.89	0.90	0.97
	<i>Sensitivity</i>	0.86	0.88	0.80
	<i>Specificity</i>	0.92	0.97	0.99
	<i>ROC-AUC</i>	0.95	0.96	0.99
	<i>F-measure</i>	0.88	0.88	0.84

- *F1 measure* is a measure of a test's accuracy and is defined as the weighted harmonic mean of the precision and recall of the test.

The performance statistics both on cross-validation as well as external validation for the top-performing VirtualTaste models are summarized in Table 1. The models achieved a prediction accuracy of 88% and above on both, cross-validation and external validation. The specificity and sensitivity of the VirtualTaste models are balanced and scored 90% and above, except for the sensitivity of VirtualSweet (86%), VirtualBitter (88%) and VirtualSour (80%) on the external validation set. The ROC-AUC values of all the three models is between 0.95 and 0.99. The *F1 measure* of all the models is higher than 0.84 (See Table 1).

VirtualTaste prediction models

The VirtualTaste models were developed using our previously published BitterSweetForest model (9) based on the Random Forest (RF) algorithm and eight different data sampling methods (23). The BitterSweetForest classification model gives a numerical estimate of the features and can produce interpretable models with low complexity. The RF classification algorithm performs classification using an ensemble method, which considers votes from multiple unbiased classifiers (decision trees) resulting in less scope for class bias and overfitting (24). The number of trees (ntree, settings: 100, 200, 300, 500, 1000) was used for training the models. The models were implemented using the Scikit-learn package (version 0.20) in Python (version 3.6.6) and the 10-fold cross-validation was used for the model optimization. The standard deviation and average of the accuracy and other parameters loss are computed (See Supplementary data). The MACCS and Morgan molecular fingerprints (<http://www.rdkit.org/>) were used and have shown an optimal performance for the prediction (Table 1). A detailed information on the construction of the models and evaluation can be found at the web server FAQ section as well as in the published work (22). More details on individual models and the features responsible for class predictions are provided in the 'Model Information' section of the web server.

VirtualSweet. The VirtualSweet prediction model predicts the sweet taste of a chemical compound. The model is based on the RF algorithm and the Synthetic Minority Over-Sampling Technique-using Tanimoto Coefficient

(SMOTETC) data sampling method (22). The active compounds in the training set contain sweet compounds and inactives were represented using bitter and tasteless compounds. The model achieved a prediction accuracy of 88% on 10-fold cross-validation and 89% on external validation. The AUC-ROC values of cross-validation and external validation are 0.95 and 0.99, respectively (Table 1).

VirtualBitter. The VirtualBitter prediction model predicts the bitter taste of a chemical compound. The model is based on the RF algorithm and the Synthetic Minority Over-Sampling Technique-using Value Difference Metric (SMOTEVDM) data sampling method (22). The active compounds in the training set contain bitter compounds and inactives were represented using sweet and tasteless compounds. The model achieved a prediction accuracy of 94% on 10-fold cross-validation and 90% on external validation. The AUC-ROC values of cross-validation and external validation are 0.97 and 0.96, respectively (Table 1).

VirtualSour. The VirtualSour prediction model predicts the sour taste of a chemical compound. Sour taste is influenced by pH and acids present in foods. Here, a data-driven machine-learning method based on a ligand-based approach is employed to predict the sour/non-sour compounds. The model is based on the RF algorithm and the Augmented Random Over Sampling (AugRandOS) data sampling method (22). The model achieved a prediction accuracy of 98% on 10-fold cross-validation and 97% on external validation. The AUC-ROC values of cross-validation and external validation are 0.97 and 0.99, respectively (Table 1).

Bitter receptors. The target prediction for the 25 human bitter receptors (hTAS2Rs) were performed using a similarity-based approach (25). It is observed that overall similarity-based approaches can outperform machine learning methods even with a low similarity threshold, especially in cases where there is less data (25). The respective receptors (targets) are predicted for a query molecule only if the query molecule is predicted as bitter by the Virtual-Bitter model. The target protein is predicted by computing pairwise similarity of the query molecule to the known ligands of that protein. The similarity is measured using the Tanimoto Coefficient (TC) (26) and molecular fingerprints. Prediction strength is defined computing maximum pairwise similarity between the query molecule to any of the

ligands of the protein (maxTC). More information on the receptors and respective chemical space similarity and dissimilarity heatmaps can be found at the 'Receptors' section of the web server.

Taste of medicines (approved drugs) and natural compounds

Medicines often taste bitter, thus are aversive to children and many adults. Often an unpleasant taste of drugs becomes a challenge in administering medicine to children (1). It is reported that in 90% of the cases, the drug taste and palatability were the biggest hindrance to treatment in this sensitive group (5). Hence, prediction of bitter taste of drugs and taste masking of the bitter drugs is important for better patient compliance, especially in paediatrics and geriatric population (27). On the other hand, identification of novel sweetening compounds from natural resources is an active research topic in the food industry. Understanding of the structure-taste relationships of natural compounds can help to expand the chemical space associated with both sweet and bitter tastes. To check the applicability domain of the VirtualTaste models, we applied our VirtualSweet and VirtualBitter models to predict the taste of chemical compounds from the approved drugs and the natural product chemical space. A total of 1969 compounds were collected from the DrugBank database (16) and were predicted using the VirtualTaste models. A total number of 1898 of approved drugs were predicted to taste bitter and 71 drugs as sweet in taste. The list of drugs and their taste class and confidence score is provided in a table under the section 'Drug Taste' on the web server. Additionally, IDs are linked to the DrugBank database to help the user access detailed information on the drug. Similarly, 326 000 compounds were extracted from the SuperNatural II database (17) and were predicted. A total of 3864 compounds were predicted to be bitter and 197 compounds as sweet with a confidence of 95% and above. These findings are similar to our published work (9). The list of natural compounds with SuperNatural II database ids, SMILES, VirtualTaste class and confidence is provided in the section 'Natural compounds' of the web server.

Application case

For a given input compound (as SMILES or name or user designed structure), in this case 'Denatonium', the VirtualTaste web server will predict a single or up to three different taste endpoints (sweet, bitter and sour) as specified by the user in the 'Prediction' tab. The result output will include information on the input molecule such as SMILES or Name, calculated physicochemical data, similar compounds present in the training set of the models along with TC values, and the compound's known active/inactive class. The taste activity prediction table displays information on the respective models (VirtualBitter, VirtualSweet, VirtualSour), descriptors and the predicted class (active or inactive) along with a confidence score for each endpoint. Using the radar plot, users can evaluate the strength of the prediction for a particular endpoint. This is done by comparing the predicted confidence score of the input compound to that of the average prediction confidence of the training set

molecules of that model. Additionally, if a compound is predicted as bitter (as in this case), the target prediction table will also include information on possible bitter receptor activity of that molecule along with the predicted score (see Figure 1). VirtualTaste also provides the user with downloadable files for similar compounds, taste activity prediction, and target prediction. All data is available for viewing and downloading as comma-separated values (csv) files by clicking the relevant links. For convenience, the results of all different result tables can be downloaded as separate individual output files and results are saved on the server during the entire session and can be retrieved by the active user by clicking the 'Results' tab in the home page.

CONCLUSIONS AND FUTURE UPDATES

In this work, we present a computational platform 'VirtualTaste' to predict three different tastes (sweet, bitter, sour) of chemical compounds. The models achieved a prediction *accuracy* of 88% and above both on cross-validation and external validation. The *specificity* and *sensitivity* of the VirtualTaste models are balanced and scored 90% and above, except for the *sensitivity* of VirtualSweet (86%), VirtualBitter (88%) and VirtualSour (80%) on external validation sets. The ROC-AUC values of all the three models is between 0.95 to 0.99. The *F1 measure* of all the models is higher than 0.84. When compared with other published models for taste prediction (14,15) all the models of the VirtualTaste web server performed from the range of comparatively good to better in some cases. Performance based comparison using measures like accuracy, sensitivity, specificity and AUC-ROC has been provided as S1 and S2. Additionally, the VirtualTaste web server also provides prediction for potential bitter receptors using a pairwise similarity-based approach. One of the major challenges of computing machine learning models is the availability of diverse and quality data. Therefore, a similarity-based method was used to predict 25 human bitter receptors.

Furthermore, the VirtualTaste models were applied to predict bitter and sweet tastes of the approved drugs and natural compounds, and this predicted data is provided in the 'DrugTaste' and 'Natural Compounds' sections of the web server. This will help the user to quickly look into the structure-taste relationship of the compounds without individually predicting it.

We hope that the VirtualTaste web server will help the experimental food chemist to predict compounds of three different tastes in a fast and easy way. Besides providing support for the basic taste chemistry research community, VirtualTaste also aims to help in the identification of novel sweet, bitter and sour tasting compound discoveries in the industry (28). Furthermore, identification of novel TAS2R agonists is important in the research related to inflammatory lung diseases like asthma and chronic obstructive pulmonary disease (COPD) (29). It is believed that understanding TAS2R receptor-agonists and their role in airway cells can help in therapy in obstructive airway diseases (30).

As an evolutionary step, VirtualTaste will focus on method development towards other organoleptic properties in the future such as scents. Furthermore, to maintain the high standard of the VirtualTaste web server, regular up-

dates will be executed, including addition of new models for the prediction of receptors.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

We thank the students of the Structural Bioinformatics Group at Charité for testing the VirtualTaste web server.

FUNDING

Funding for open access charge: Charité – University Medicine Berlin [KFO339];Deutsche Forschungsgemeinschaft/International [DFG TRR295].

Conflict of interest statement. None declared.

REFERENCES

- Mennella, J.A., Spector, A.C., Reed, D.R. and Coldwell, S.E. (2013) The bad taste of medicines: overview of basic research on bitter taste. *Clin. Ther.*, **35**, 1225–1246.
- Jagupilli, A., Howard, R., Upadhyaya, J.D., Bhullar, R.P. and Chelikani, P. (2016) Bitter taste receptors: Novel insights into the biochemistry and pharmacology. *Int. J. Biochem. Cell Biol.*, **77**, 184–196.
- Breslin, P.A.S. (2013) An evolutionary perspective on food and human taste. *Curr. Biol.*, **23**, R409–R418.
- Ahmad, R. and Dalziel, J.E. (2020) G protein-coupled receptors in taste physiology and pharmacology. *Front. Pharmacol.*, **11**, 587664.
- Milne, C.-P. and Bruss, J.B. (2008) The economics of pediatric formulation development for off-patent drugs. *Clin. Ther.*, **30**, 2133–2145.
- Li, H. and Liu, F. (2015) The chemistry of sour taste and the strategy to reduce the sour taste of beer. *Food Chem.*, **185**, 200–204.
- Ramos Da Conceicao Neta, E.R., Johanningsmeier, S.D. and McFeeters, R.F. (2007) The chemistry and physiology of sour taste—a review. *J. Food Sci.*, **72**, R33–R38.
- Meyerhof, W., Batram, C., Kuhn, C., Brockhoff, A., Chudoba, E., Bufe, B., Appendino, G. and Behrens, M. (2010) The molecular receptive ranges of human TAS2R bitter taste receptors. *Chem. Senses.*, **35**, 157–170.
- Banerjee, P. and Preissner, R. (2018) BitterSweetForest: a random forest based binary classifier to predict bitterness and sweetness of chemical compounds. *Front. Chem.*, **6**, 93.
- Karl, C.M., Wendelin, M., Lutsch, D., Schleining, G., Dürrschmid, K., Ley, J.P., Krammer, G.E. and Lieder, B. (2020) Structure-dependent effects of sweet and sweet taste affecting compounds on their sensorial properties. *Food Chem. X*, **7**, 100100.
- Ahmed, J., Preissner, S., Dunkel, M., Worth, C.L., Eckert, A. and Preissner, R. (2011) SuperSweet—a resource on natural and artificial sweetening agents. *Nucleic Acids Res.*, **39**, D377–D382.
- Kurczab, R., Smusz, S. and Bojarski, A. (2011) Evaluation of different machine learning methods for ligand-based virtual screening. *J. Cheminform.*, **3**, P41.
- Banerjee, P., Eckert, A.O., Schrey, A.K. and Preissner, R. (2018) ProTox-II: a webserver for the prediction of toxicity of chemicals. *Nucleic Acids Res.*, **46**, W257–W263.
- Dagan-Wiener, A., Nissim, I., Ben Abu, N., Borgonovo, G., Bassoli, A. and Niv, M.Y. (2017) Bitter or not? BitterPredict, a tool for predicting taste from chemical structure. *Sci. Rep.*, **7**, 12074.
- Zheng, S., Chang, W., Xu, W., Xu, Y. and Lin, F. (2019) e-Sweet: a machine-learning based platform for the prediction of sweetener and its relative sweetness. *Front. Chem.*, **7**, 35.
- Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z. *et al.* (2018) DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.*, **46**, D1074–D1082.
- Banerjee, P., Erehman, J., Gohlke, B.-O., Wilhelm, T., Preissner, R. and Dunkel, M. (2015) Super Natural II—a database of natural products. *Nucleic Acids Res.*, **43**, D935–D939.
- Gaulton, A., Bellis, L.J., Bento, A.P., Chambers, J., Davies, M., Hersey, A., Light, Y., McGlinchey, S., Michalovich, D., Al-Lazikani, B. *et al.* (2012) ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.*, **40**, D1100–D1107.
- Banerjee, P., Cebon, N., Dill, F., Gabriel, T., Kötter, T., Meinl, T., Ohl, P., Sieb, C., Thiel, K. and Wiswedel, B. (2008) KNIME: The Konstanz Information Miner. In: Preisach, C., Burkhardt, H., Schmidt-Thieme, L. and Decker, R. (eds). *Data Analysis, Machine Learning and Applications SE-38, Studies in Classification, Data Analysis, and Knowledge Organization*. Springer, Berlin, Heidelberg, pp. 319–326.
- Dagan-Wiener, A., Di Pizio, A., Nissim, I., Bahia, M.S., Dubovski, N., Margulis, E. and Niv, M.Y. (2019) BitterDB: taste ligands and receptors database in 2019. *Nucleic Acids Res.*, **47**, D1179–D1185.
- Sayers, E.W., Beck, J., Bolton, E.E., Bourexis, D., Brister, J.R., Canese, K., Comeau, D.C., Funk, K., Kim, S., Klimke, W. *et al.* (2021) Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.*, **49**, D10–D17.
- Banerjee, P., Dehnbostel, F.O. and Preissner, R. (2018) Prediction is a balancing act: importance of sampling methods to balance sensitivity and specificity of predictive models based on imbalanced chemical data sets. *Front. Chem.*, **6**, 362.
- Breiman, L. (2001) Random forests. *Mach Learn.*, **45**, 5–32.
- Mitchell, J.B.O. (2014) Machine learning methods in cheminformatics. *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, **4**, 468–481.
- Mathai, N. and Kirchmair, J. (2020) Similarity-based methods and machine learning approaches for target prediction in early drug discovery: performance and scope. *Int. J. Mol. Sci.*, **21**, 3585.
- Bajusz, D., Rácz, A. and Héberger, K. (2015) Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminform.*, **7**, 20.
- Douroumis, D. (2011) Orally disintegrating dosage forms and taste-masking technologies; 2010. *Expert Opin. Drug Deliv.*, **8**, 665–675.
- Nakagita, T., Taketani, C., Narukawa, M., Hirokawa, T., Kobayashi, T. and Misaka, T. (2020) Ibuprofen, a nonsteroidal anti-inflammatory drug, is a potent inhibitor of the human sweet taste receptor. *Chem. Senses.*, **45**, 667–673.
- Zhang, Y., Wang, X., Li, X., Peng, S., Wang, S., Huang, C.Z., Huang, C.Z., Zhang, Q., Li, D., Jiang, J. *et al.* (2017) Identification of a specific agonist of human TAS2R14 from Radix Bupleuri through virtual screening, functional evaluation and binding studies. *Sci. Rep.*, **7**, 12174.
- Sharma, P., Conaway, S. and Deshpande, D. (2021) Bitter taste receptors in the airway cells functions. In: *Handbook of Experimental Pharmacology*. Springer, Berlin, Heidelberg, pp. 1–25.