Characterizing Bitterness: Identification of Key Structural Features and Development of a Classification Model

Sarah Rodgers*

Unilever Food and Health Research Institute, Olivier van Noortlaan 120, 3133 AT Vlaardingen, The Netherlands

Robert C. Glen and Andreas Bender[†]

Unilever Centre for Molecular Science Informatics, Chemistry Department, University of Cambridge, Lensfield Road, Cambridge CB2 1EW, United Kingdom

This work describes the first approach in the development of a comprehensive classification method for bitterness of small molecules. The data set comprises 649 bitter and 13 530 randomly selected molecules from the MDL Drug Data Repository (MDDR) which are analyzed by circular fingerprints (MOLPRINT 2D) and information-gain feature selection. The feature selection proposes substructural features which are statistically correlated to bitterness. Classification is performed on the selected features via a naïve Bayes classifier. The substructural features upon which the classification is based are able to discriminate between bitter and random compounds, and thus we propose they are also functionally responsible for causing the bitter taste. Such substructures include various sugar moieties as well as highly branched carbon scaffolds. Cynaropicrine contains a number of the substructural features found to be statistically associated with bitterness and thus was correctly predicted to be bitter by our model. Alternatively, both promethazine and saccharin contain fewer of these substructural features, and thus the bitterness in these compounds was not identified. Two different classes of bitter compounds were identified, namely those which are larger and contain mainly oxygen and carbon and often sugar moieties, and those which are rather smaller and contain additional nitrogen and/or sulfur fragments. The classifier is able to predict 72.1% of the bitter compounds. Feature selection reduces the number of false-positives while also increasing the number of false negatives to 69.5% of bitter compounds correctly predicted. Overall, the method presented here presents both one of the largest databases of bitter compounds presently available as well as a relatively reliable classification method.

1. INTRODUCTION

Taste is the most important factor influencing the food people choose to eat.[1] The flavoring of molecules assists in the identification of nutritionally rich foods and rejection of potentially damaging foods. Humans perceive five different taste modalities: sweet, sour, salt, bitter, and umami. Sweet, salt, and umami tastes signal the presence of beneficial ingredients and thus have a positive association. Sweet compounds tend to contain a lot of energy in the form of carbohydrates, whereas salty foods are likely to be mineral-rich. Umami, described as a 'savory' taste,[2] assists in the identification of foods with a high protein content. Acids, that form as foods spoil, are perceived as sour, serving as a warning that food may no longer be fresh. Bitterness is associated with toxicity, and many bitter compounds are derived from plants, where they have evolved to produce these compounds as a chemical defense against being eaten.[3] In general, therefore, bitter foods are rejected; however, a slight bitter taste is sometimes favored—in particular today where the risk associated with food bought at the supermarket is very small.

Taste is mediated through receptors on the tongue; salt and sour compounds operate via membrane ion channels,[4,5] while the other tastes are mediated via G-protein coupled receptors.[6,7] Bitterness is the most complex of the tastes: there are many more bitter components than are found for the other taste components and also more receptors to which they can bind. Of the 25 bitter receptors that have been sequenced and cloned,[8,9] only seven have ligands reported in the scientific literature.[10–13] Therefore, the majority of bitter receptors are orphans. The range of molecules capable of producing a bitter taste is very large and diverse, including proteins, amino acids, salts, fats, flavonoids, glucosides, and terpenes to name but a few.

Many bitter tastants have health benefits, for example isothiocyanates from broccoli protect against cancer.[14] In addition, a large number of oral pharmaceuticals produce a bitter taste, for example indinavir, a protease inhibitor used in the treatment of human immunodeficiency virus (HIV) infection.[15] However, the mechanism by which bitter compounds exert their taste is poorly understood. Taste information for many compounds is unknown, and, to our knowl-

* Corresponding author phone: +44(0)1509644436; fax: +44(0)-1509645576; e-mail: Sarah.Rodgers@AstraZeneca.com. Current address: AZ R&D Charnwood, Bakewell Road, Loughborough, Leicestershire LE11 5RH, U.K.
† Current address: Lead Discovery Informatics, Novartis Institutes for BioMedical Research Inc., 250 Massachusetts Ave, Cambridge, MA 02139.

edge, there are no publicly available bitter databases. To gain the maximum benefit from these compounds, bitterness must first be identified and then managed. The ability to predict bitterness in novel ingredients for which the taste status is unknown would be highly valuable; researchers can be alerted to the presence of molecules that will potentially produce a bitter taste. Effective masking strategies, developed through an improved understanding of bitter taste perception, could then be employed to ensure the product is acceptable to consumers.

For peptides, such a prediction tool exists, based on the observation that hydrophobic peptides tend to be bitter.[16] Research concerning the prediction of bitterness in small molecules tends to be limited to the identification of quantitative structure−activity relationships. For example, Spillane et al.[17] used space-fill models of atomic structures to predict the bitter taste of 10 benzenesulfamates. An alternative study linked chemical properties, including amounts of proteins, fats, and salts and gas chromatography spectra, of a number of samples of cheeses to their bitter taste.[18] Both studies produced good correlations between structural properties and bitter taste, but in particular the structural diversity covered by the data sets is very limited and cannot be applied to small molecules in general.

The classification of bitter molecules described here is based on a much larger data set than has been used before, comprising a total of 649 bitter molecules compiled from the literature. This database is by no means complete, thus limiting somewhat the scope of the classification model. However, as far as we are aware, the database is much larger than any other collection of bitter molecules and includes a diverse range of compounds. It therefore provides a suitable starting point for developing a general classification tool for bitterness prediction in small molecules, the first of its kind.

Section 2 summarizes the algorithm as well as the data set. Section 3 presents the results, while section 4 provides a full discussion. The conclusions follow in the final section 5.

2. MATERIAL AND METHODS

(a) Database Compilation of Bitter and Nonbitter Compounds. A 2D database of bitter molecules was constructed from an extensive scan of the scientific literature and patents, by conducting searches on "*bitterness*" in the Food Science and Technology Abstracts,[19] BIOSIS,[20] Derwent World Patents Index,[21] and databases of internal reports at Unilever. Where possible, structures of bitter compounds were extracted from SciFinder.[22] Alternatively, structures were searched for within the Dictionary of Natural Products (DNP)[23] and the World Drug Index (WDI),[24] or they were constructed using ChemDraw.[25] Duplicates, salts, peptides, and ions were removed from the data set. The database was further reduced by the exclusion of synthetic analogues, resulting in a total of 649 bitter molecules. This is an updated version of the database of bitter compounds used in an earlier study,[26] which provides an overview of the structural groups present in the database. Short peptides that were originally included in the database were omitted since their amide backbones would overly dominate the classification model. Alternatively one could compile a similar relative frequency of amide bonds in the inactive data set (see below).
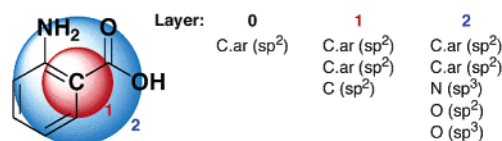

Figure 1. Representation of features via MOLPRINT 2D circular fingerprints and Sybyl mol2 atom typing.

Threshold values for perceived bitterness, which are available for a large proportion of the molecules in the database, were not considered in these analyses due to the variability of the studies from which the molecules were obtained and thus the unreliability of the data.

Both bitter and nonbitter molecules are required to build a classification model for bitterness. Ideally, a database of (experimentally tested) nonbitter molecules would be employed, but there are none publicly available. A generic set of molecules were however required for this methodology to discriminate between substructural features (circular fingerprints) found proportionally in greater abundance in bitter molecules. To approximate this, 13 530 molecules were selected randomly from the MDL Drug Data Repository (MDDR)[27] and were defined as the inactive set (where bitter is active). The reason to choose the MDDR database instead of other databases was the assumption that it more closely populates 'bioactive chemical space', part of which can be seen to represent the 'chemical space of bitterness'. All molecules contained in the MDDR are able to bind at receptors (or other macromolecules involved in biological processes), which is certainly not the case for a random collection of compounds. All duplicates, salts, ions, and known bitter molecules were removed. While many compounds, such as alkaloids, are known to confer a bitter taste, the assumption that a compound selected randomly from this data set is nonbitter will result in some false-negatives in the data set. This is also a common problem in many retrospective virtual screening sets derived e.g. from the MDDR (where no information about definite *inactivity* of compounds is given). Still, the assumption that the majority of compounds are inactive (nonbitter) should hold in the majority of cases.

An additional smaller data set was also compiled which (as opposed to the 649 compounds Unilever data set) could be made public. The 33 structures employed in this smaller set are depicted in Figures 2 and 3, and they are also available as Supporting Information, along with 254 compounds which represent the inactive (nonbitter) structures of this data set.

(b) Classification Method Employed. The classification is based on circular fingerprints,[28] information-gain feature selection, and the naïve Bayesian classifier and relies on the hypothesis that compounds sharing structural similarity are likely to have similar taste properties.[29] It was shown to compare favorably with other classification methods[30] when used to classify a recently published[31] database of compounds having a range of activities from the MDL Drug Data Report (MDDR).[32]

MOLPRINT 2D fingerprints (aka Atom Environments) were used to provide a 2D representation of the chemical structures. Sybyl[33] mol2 atom types are employed for this process. An individual fingerprint is calculated for each atom in the molecule, considering those atoms up to 2 bonds from
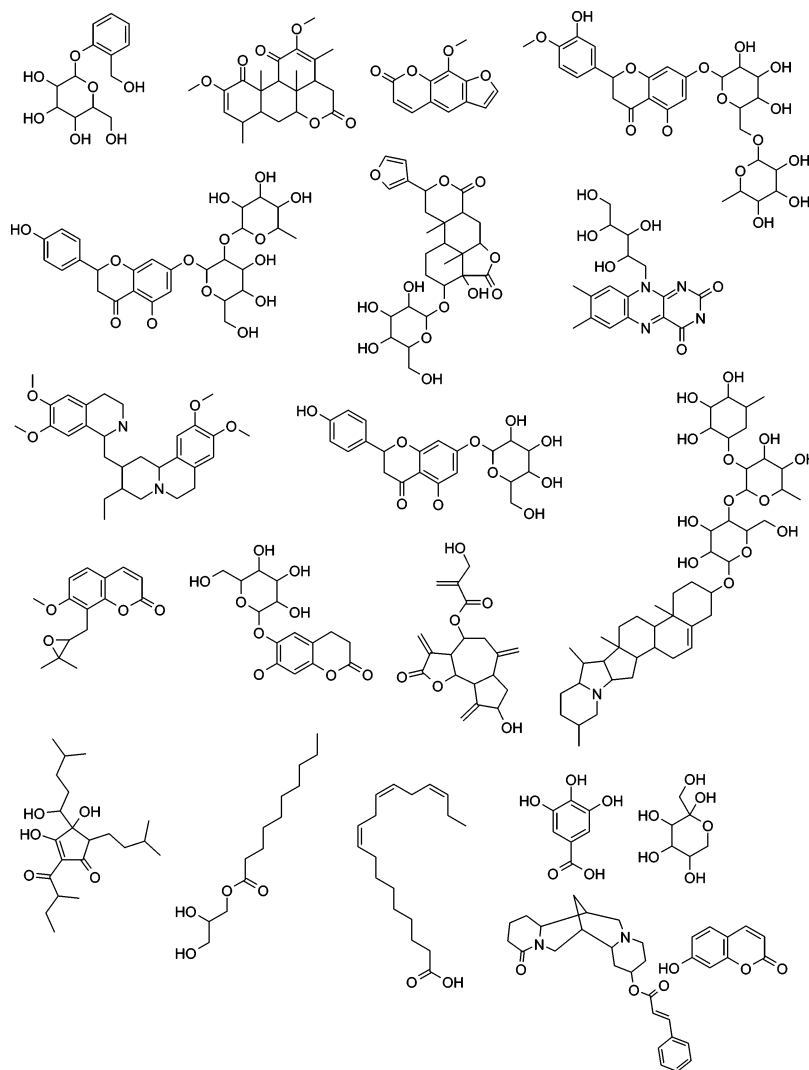
KEY FEATURES OF BITTERNESS

J. Chem. Inf. Model., Vol. 46, No. 2, 2006 **571**

Figure 2. Public data set of 33 bitter compounds, out of a total number of 649 (in-house) bitter structures compiled. Structures shown here contain mainly the elements carbon and oxygen.
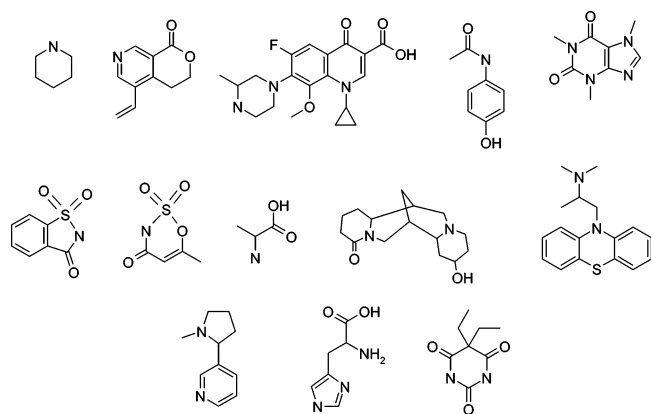

Figure 3. Public data set of 33 bitter compounds, out of a total number of 649 (in-house) bitter structures compiled. Structures shown here contain, in additional to carbon and oxygen, also sulfur and nitrogen moieties.

the central atom (Figure 1). The molecular fingerprint thus consists of the individual atom fingerprints of all the heavy atoms in the structure. Subsequently, feature selection was employed to identify those features that were best able to discriminate between the bitter and random compounds. This was achieved by calculating the information gain associated with each MOLPRINT 2D fingerprint. The information

entropy, S, may first be calculated

$$S = \sum_{i=1,2} p \log_2 p \tag{1}$$

where p is the probability that a randomly selected molecule of the whole data set belongs to each of the defined classes 1 and 2 (which is equivalent to the relative data set sizes). The information gain, I, of each individual feature is then given by

$$I = S - \sum_v \frac{|D_v|}{|D|} S_v \tag{2}$$

where S_v is the information entropy in data subset v; $|D|$ is the total number of datapoints; and $|D_v|$ is the number of datapoints in subset v. A higher information gain is related to a better separation between bitter and random structures. Finally, a naïve Bayesian classifier was used to build a classification model for bitterness. The Bayesian classifier is considered naïve because it assumes that the variables are independent of each other. (However, it performs surprisingly well even where this is not the case.) The classifier is trained with a data set of feature vectors, F, and their associated

Table 1. Classification Predictions for the Data Set Containing 649 Bitter and 13 530 Nonbitter Compounds in a 5-Fold Cross-Validation

	true positive	false negative	true negative	false positive	positive correct	negative correct
no feature selection	468	181	12037	1493	72.10%	89.00%
feature selection	451	198	12514	1016	69.50%	92.50%

Table 2. Classification Predictions for the (Public) Data Set Containing 33 Bitter and 265 Nonbitter Compounds in a 5-Fold Cross-Validation

	true positive	false negative	true negative	false positive	positive correct	negative correct
no feature selection	23	10	129	125	69.70%	50.79%
feature selection	19	14	184	70	57.58%	72.44%

known classes, CL; class 1 and class 2 (CL$_1$ and CL$_2$) are the bitter and nonbitter data set. The Bayesian classifier predicts the class that a new feature belongs to as the one with the highest probability of $P(CL_v|F)$:

$$\frac{P(CL_1|F)}{P(CL_2|F)} = \frac{P(CL_1)}{P(CL_2)} \Pi_i \frac{P(f_i|CL_1)}{P(f_i|CL_2)} \qquad (3)$$

The $P(f|CL)$ on the right-hand side of the equation denotes the relative frequencies of individual features which are multiplied to give the final estimate of the molecule belonging to the active (bitter) vs the inactive (nonbitter) class, respectively. This method was described earlier,[28,30] where a full account of the method may be found.

The database of bitter compounds represents a general activity group, within which reside several subactivity groups (i.e. all molecules are bitter but exert bitterness via 25 different receptors and thus—at least—25 mechanisms). However, as described above, the majority of bitter receptors are orphans, and in many cases where ligands of receptors have been identified only very few are known for each target receptor. While it is acknowledged that the data set may not sample the chemical space of bitter molecules exhaustively, the ability of the Bayes classifier to subsume multiple models should be beneficial in the model generation step.

(c) Computational Details. Five-fold cross-validation of the model was performed, and all the predictions reported in the following are cross-validation results. Calculations were performed employing no feature selection and selection of the number of features which is present in the smaller of the two data sets (here, the bitter data set). While feature selection could be seen as a parameter that needs to be optimized, more recent research[35] has shown that in combination with appropriate Laplacian modification the optimum number of features selected is not that crucial, with the above parameter providing classification performance close to the global optimum.

3. RESULTS

Visual inspection of the data set of bitter compounds (the public data set of 33 compounds is given in Figures 2 and 3) shows that at least two distinct subclasses of compounds can be identified. The first of the classes comprises compounds that are rather large and contain mainly carbon, oxygen, and sugar-like fragments (Figure 2). Compounds drawn from the second class are, on average, rather small and contain (additional) nitrogen- and sulfur-based fragments. While no precise relationship between structure and function can be presented here, the implications of this observation were shown to profoundly influence classification results described later (see Discussion section).
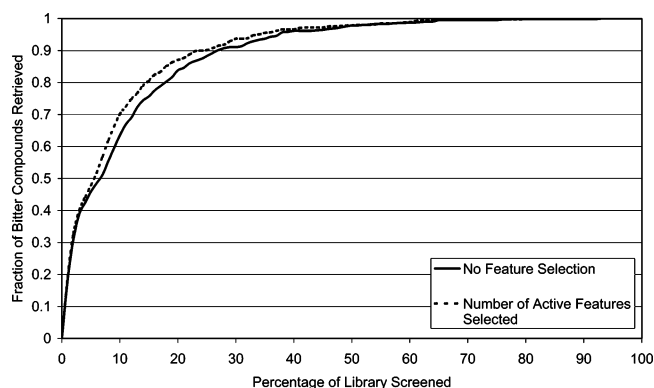
Figure 4. Cumulative recall plots for the 5-fold cross-validation of 649 bitter and 13 530 nonbitter compounds. At 20% of the library screened about 80% of all bitter compounds are retrieved.

Classification results for both the large data set containing 649 bitter and 13 530 nonbitter structures and the public one containing 33 and 265 compounds, respectively, are given in Tables 1 and 2. In the large data set (numbers given in Table 1) and without employing feature selection, 72.1% of bitter compounds are correctly classified. If feature selection is employed to select the number of features present in the smaller (bitter) data set, bitter compounds are predicted correctly in only 69.5% of the cases (−2.6%). The proportion of nonbitter (or random) compounds assigned nonbitter status is also included in the table for reference, but these values will not be discussed here since the true bitterness status of these compounds is unknown.

On the smaller data set (numbers given in Table 2) and without employing feature selection, 69.7% of bitter compounds are correctly classified. If feature selection is employed to select the number of features present in the smaller (bitter) data set, bitter compounds are predicted correctly in only 57.6% of the cases (−12.1%).

In both cases the tendency was observed that feature selection improved classification results to a small extent, due to a smaller number of false positives, while at the same time increasing the number of false negatives. The results also illustrate the importance of a data set which samples chemical space sufficiently, which seems not to be given for the public data set of 33 bitter structures (since it shows far inferior performance, compared to the larger data set).

The recall plot from the 5-fold cross-validation is shown in Figure 4. Approximately 80% of all bitter structures are correctly identified where 20% of the database is retrieved. The influence of feature selection on compound retrieval is small, but given the 'vertical readout' of the plot between about 10% and 20% of structures retrieved continuously about 5% more active compounds are correctly identified.

KEY FEATURES OF BITTERNESS

J. Chem. Inf. Model., Vol. 46, No. 2, 2006 **573**

Table 3. Molecular Features Possessing Greatest Information Gain with Respect to Bitterness[a]

Fragment Number	Fragment	Frequency Bitter Dataset	Frequency Nonbitter Dataset	Information Gain	Relative Frequency Bitter Dataset	Relative Frequency Nonbitter Dataset
1		120	317	0.0139	18.49%	2.34%
2		76	76	0.0139	11.71%	0.56%
3		116	342	0.0125	17.87%	2.53%
4		126	508	0.0106	19.41%	3.75%
5		147	713	0.0106	22.65%	5.27%
6		105	352	0.0102	16.18%	2.60%
7		231	1818	0.0097	35.59%	13.44%
8		80	196	0.0095	12.33%	1.45%
9		24	5	0.0062	3.70%	0.04%
10		30	23	0.0059	4.62%	0.17%

[a] Mainly fragments resulting from sugarlike moieties and certain (highly branched) scaffold patterns are selected by the information-gain feature selection step.

Tests conducted on the inactive set confirm that the section of the MDDR selected for the test and training sets has very little effect on the retrieval of bitter compounds. The inactive set is much larger than the active set and thus is unlikely to strongly influence the retrieval of bitter compounds since its feature distribution varies less with the particular compounds contained in each of the set.

In each model, a large number of compounds from the MDDR were predicted to be bitter in absolute numbers but not in relative numbers. This follows from the different database sizes, with the nonbitter data set being about 15 times as large as the bitter one. Therefore, false positive predictions are more likely to occur. A limiting factor in the tests conducted here is the inactive data set, which, due to the nature of the database, is likely to contain bitter molecules. It is impossible to determine which of the inactive compounds may also be bitter because in general, they have not been tested for bitterness. Thus, a reasonable proportion of the *false positives* may in fact be bitter.

The combination of MOLPRINT 2D fingerprints, information-gain feature selection, and a naïve Bayesian classifier has been shown here to be sufficiently reliable, correctly classifying approximately 70% of the bitter compounds in our full database and 64% in our smaller data set. Examination of the features selected to be significant for bitterness can provide information about which structural features are important in conferring the bitter taste. This is considered in the discussion that follows.

4. DISCUSSION

The 10 features selected by the information-gain feature selection which are most significant for bitterness are shown in Table 3, along with absolute and relative frequencies and the information gain associated with each feature.

As mentioned briefly in the Material and Methods section, it can be seen that the knowledge of the classifier about bitterness of structures is dominated by the "carbon/oxygen" group of compounds (shown in Figure 2). Apart from fragments which can easily be associated with sugar moieties frequently present in Figure 2 (such as features at the top three positions in Table 3), surprisingly also characteristic
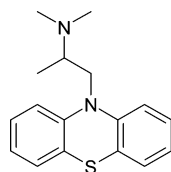
Figure 5. Promethazine, an antihistamine, is predicted to belong to the group of nonbitter compounds (by a small margin, according to the likelihoods obtained by the Bayes classifier).
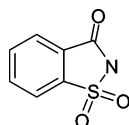

Figure 6. Saccharin, an artificial sweetener. Its bitterness was not detected by the method presented here.

scaffold fragments can also be identified. Examples are quaternary carbon centers (fragment number four) and highly branched carbon centers (fragments nine and ten) which are characteristic for substituted steroid scaffolds. More specifically, fragment nine corresponds to carbon atom number eight of a steroid scaffold. While it may well be possible that bitter compounds more often contain steroid core structures than bioactive compounds in general, it has to be admitted that this finding may also simply be a result of the limited (though relatively large) data set used.

According to the selection of characteristic features it can be anticipated that certain classes of bitter compounds cannot be predicted reliably. This is indeed the case, as illustrated on the structures shown in Figures 5 and 6, which are promethazine and saccharin.

Promethazine belongs to the group of bitter compounds dominated by nitrogen-containing scaffolds, which represents the smaller of the two subgroups in both the larger data set and the publicly available set of bitter compounds containing 33 structures. In the model presented here, the features contained in promethazine are found not to be characteristic for bitter compounds, accordingly the compound is classified as a false negative. (Although it should be noted that the 'score' of this compound is only slightly below 0, namely about −2, on a scale from about +100 to −100, where positive scores indicate bitterness of compounds).

Compounds are not restricted to displaying only one taste modality, a number of sweet compounds are also bitter. Saccharin activates the hTAS2R43 and hTAS2R44 bitter receptors,[11] and it is possibly the sulfonamide group that is important for this receptor−ligand association. It is not a potently bitter compound and generally needs to be present in large amounts to be perceived as bitter.[12] Saccharin is unlikely to be one of the main ligands for hTAS2R43 and hTAS2R44 as it is only weakly bitter and also is a synthetic compound. However, saccharin is capable of some interaction with the bitter receptors and is therefore of interest. In the model presented here saccharin was classified as a nonbitter compound. This example of saccharin illustrates not only a false-negative prediction of the model due to the areas of chemical space it covers but also that the annotation of bioactivity (here bitterness) is often not possible in a clear-cut manner.

The bitter molecules that were correctly assigned bitterness status contained features selected by the information-gain feature selection. According to the information gains calcu-
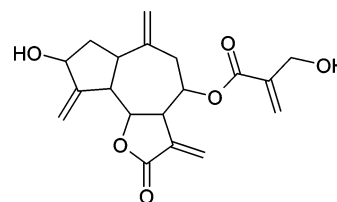

Figure 7. Cynaropicrine. This compound was correctly identified as belonging to the bitter class of compounds.

lated as well as the relative frequencies of features (Table 3), it appears that individually these features do not provide a lot of information, but collectively they can help to effectively identify the bitter compounds from a random test set. This is in agreement with earlier work, where it was found that only the *collection* of features employed for classification via a Bayes classifier provides a sensible model for inhibitors of HIV protease.[34] The structure in Figure 7 was correctly classified as being a bitter compound with a very high score, and it represents a structure of the group of oxygen-containing bitter compounds. As many of the structures which belong to this class also possess sugar moieties, the relationship between the features responsible for— seemingly—fundamentally different taste modalities such as sweetness and bitterness would be a very interesting area for future research.

The database of bitter compounds considered here has previously been analyzed in order to identify chemical groups specific to bitter compounds when compared with a data set of random compounds.[26] The circular fingerprints employed here are limited to a radius of two bonds and thus are much smaller than the substructural groups identified in earlier work. In addition, the structures discovered earlier are more diverse and cover more structural groups than the atom environments. Hence the two methods can be seen as complementary: Circular fingerprints "work collectively" to select those molecules with the greatest likelihood of being bitter and provide a reliable classification method. Previous substructural analysis on the other hand focused on identifying substructural groups that indicate bitterness instead of trying primarily to build a classification method.

It may, however, be appropriate to compare the two approaches in terms of how effectively they perform in identifying the bitter molecules. In a previous study, 50% of the database of bitter molecules were effectively associated with a significantly bitter substructure. A significantly bitter structure was defined as one in which at least 20% of the molecules containing it are bitter (where the proportion of bitter compounds in the whole database was 3%). The Bayesian classifier presented here was able to identify approximately 70% of the bitter molecules from the full database. No direct comparison should be made between both methods due to the different databases and different computational routines employed (cross-validation vs classification of the whole data set).

The different substructures selected by each method as significant for bitterness can explain the differences in bitter molecules identified: the features shown in Table 3 cover solely oxygen and carbon-containing fragments, thereby missing some of the less populated areas of 'bitter chemical space'. The earlier study based on a maximum common substructure (MCS) approach was able to select also

KEY FEATURES OF BITTERNESS

J. Chem. Inf. Model., Vol. 46, No. 2, 2006 **575**

nitrogen-containing features. On the other hand, if compared to previous work, more reliable classification of both bitter and nonbitter structures was presented in the current work.

Therefore, both approaches seem to be superior in one respect: While maximum common substructure-based approaches present features which are more easily interpretable (since they are larger) and more diverse, the approach presented here shows improved classification accuracy.

The classification model developed here is able to effectively predict the bitterness in 70% of the bitter molecules in our database, while also predicting many random compounds to be so. This reflects the complexity of the problem and how little is known about the structures that cause bitterness. It is clear that alternative methods, including the type of descriptor used and the method of descriptor selection, may be required to retrieve the remaining 30%. Other descriptions of the molecules could include 3D conformations, physicochemical properties, and shape indices. The model developed here provides a starting point, or a benchmark, in the classification of bitter molecules.

By identifying compounds that share structural similarity with known bitter compounds, we have developed a method to aid in the identification of bitter compounds. However, this is not a bitterness classifier per se, as all of the specific structural features required to produce a bitter taste are not known. This work, and previous analyses, have identified some substructures that can be associated with bitterness, but there are likely to be many more. In addition, there are many known bitter compounds for which our work has identified no significant substructures.

5. CONCLUSION AND OUTLOOK

This study has described the first general classification of bitter molecules on the basis of their 2D structure. MOLPRINT 2D circular fingerprints, information-gain based feature selection, and the naïve Bayesian classifier were in a cross-validation study able both to highlight possible bitter substructures and to provide a first classification model for bitter compounds. The classifier is able to predict 72.1% of the bitter compounds correctly. Feature selection reduces the number of false-positives while also increasing the number of false negatives (69.5% and 92.5%) of compounds correctly predicted.

A classification model for bitterness as described in this work is relevant for predicting bitterness in compounds for which the taste status is unknown. This will be particularly useful at the early stages of food product development where ingredients are being formulated. In cases where a food is known to be bitter, the classification model can be used to identify which of the ingredients is most likely to be responsible for the bitter taste. It will also be useful in determining the likelihood that certain key ingredients, such as functional foods, are likely to give a bitter taste. Alerting developers to the presence of bitter compounds will allow them to apply relevant masking strategies, such as adding compounds that will either block the bitter receptor or bind to the bitter compound to prevent its interaction with the receptor, to prevent the perception of the bitter molecule by the consumer.

MOLPRINT 2D, comprising fingerprinting, feature selection, and classification algorithms, is freely available from http://www.cheminformatics.org/.

Supporting Information Available: Thirty-three structures depicted in Figures 2 and 3 and 254 compounds which represent the inactive (nonbitter) structures of this data set. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES AND NOTES

(1) Glanz, K.; Basil, M.; Maibach, E.; Goldberg, J.; Snyder, D. Why Americans eat what they do: Taste, nutrition, cost, convenience, and weight control concerns as influences on food consumption. *J. Am. Diet. Assoc.* **1998**, *98*, 1118−1126.

(2) Ninomiya, K. Umami: A Universal Taste. *Food Rev. Int.* **2002**, *18*, 23−38.

(3) Meyerhof, W. Elucidation of mammalian bitter taste. *Rev. Physiol. Bioch. P.* **2005**, in press. DOI: 10.1007/s10254-005-0041.

(4) Heck, G. L.; Mierson, S.; DeSimone, J. A. Salt taste transduction occurs through an amiloride-sensitive sodium transport pathway. *Science* **1984**, *223*, 403−405.

(5) Kinnamon, S. C.; Dionne, V. E.; Beam, K. G. Apical localization of K^+ channels in taste cells provides the basis for sour taste transduction. *P. Natl. Acad. Sci. U.S.A.* **1988**, *85*, 7023−7027.

(6) Wong, G. T.; Gannon, K. S.; Margolskee, R. F. Transduction of bitter and sweet taste by gustducin. *Nature* **1996**, *381*, 796−800.

(7) Chaudhari, N.; Landin, A. M.; Roper, S. D. A metabotropic glutamate receptor variant functions as a taste receptor. *Nature Neurosci.* **2000**, *3*, 113−119.

(8) Adler, E.; Hoon, M. A.; Mueller, K. L.; Chandrashekar, J.; Ryba, N. J. P.; Zuker, C. S. A novel family of mammalian taste receptors. *Cell* **2000**, *100*, 693−702.

(9) Chandrashekar, J.; Mueller, K. L.; Hoon, M. A.; Adler, E.; Feng, L.; Guo, W.; Zuker, C. S.; Ryba, J. P. T2Rs function as bitter taste receptors. *Cell* **2000**, *100*, 703−711.

(10) Bufe, B.; Hofmann, T.; Krautwurst, D.; Raguse, J.-D.; Meyerhof, W. The human TAS2R16 receptor mediates bitter taste in response to β-glucopyranosides. *Nat. Genet.* **2002**, *32*, 397−401.

(11) Kuhn, C.; Bufe, B.; Winnig, M.; Hofmann, T.; Frank, O.; Behrens, M.; Lewtschenko, T.; Slack. J. P.; Ward, C. D.; Meyerhof, W. Bitter Taste Receptors for Saccharin and Acesulfame K. *J. Neurosci.* **2004**, *24*, 10260−10265.

(12) Pronin, A. N.; Tang, H.; Connor, J.; Keung, W. Identification of Ligands for Two Human Bitter T2R Receptors. *Chem. Senses* **2004**, *29*, 583−593.

(13) Behrens, M.; Brockhoff, A.; Kuhn, C.; Bufe, B.; Winnig, M.; Meyerhof, W. The human taste receptor hTAS2R14 responds to a variety of different bitter compounds. *Biochem. Bioph. Res. Commun.* **2004**, *319*, 479−485.

(14) Zhang, Y.; Talalay, P.; Cho, C.-G.; Posner, G. H. A major inducer of anticarcinogenic protective enzymes from broccoli: Isolation and elucidation of structure. *Proc. Natl. Acad. Sci. U.S.A.* **1992**, *89*, 2399−2403.

(15) Schiffman, S. S.; Zervakis, J.; Heffron, S.; Heald, A. E. Effect of Protease Inhibitors on the Sense of Taste. *Nutrition* **1999**, *15*, 767−772.

(16) Ney, K. H. Bitterness of Peptides: Amino Acid Composition and Chain Length. *ACS Symp. Ser.* **1979**, *115*, 149−173.

(17) Spillane, W. J.; Feeney, B. G.; Coyle, C. M. Further studies on the synthesis and tastes of monosubstituted benzenesulfamates. A semiquantitative structure-taste relationship for the meta-compounds. *Food Chem.* **2002**, *79*, 15−22.

576 *J. Chem. Inf. Model., Vol. 46, No. 2, 2006*

RODGERS ET AL.

(18) Carpino, S.; Acree, T. E.; Barbano, D. M.; Licitra, G.; Siebert, K. J. Chemometric Analysis of Ragusano Cheese Flavor. *J. Agric. Food Chem.* **2002**, *50*, 1143−1149.

(19) FSTA, Food Science and Technology Abstracts, IFIS Publishing: Shinfield, 2003. Database available from Ovid Technologies Inc. at URL http://www.ovid.com/site/index.jsp.

(20) BIOSIS Previews/RN(R), Biological Abstracts Inc.: Philadelphia, PA, 2003. Database is available from STN at URL http://www.stn-international.de/.

(21) WPIDS, Derwent World Patents Index Subscribers file, Thomson Scientific: Philadelphia, 2003. Database is available from STN at URL http://www.stn-international.de/.

(22) SciFinder, 2004 edition, software available from the American Chemical Society at URL http://www.cas.org/.

(23) DNP, Dictionary of Natural Products, version 8.1, 2004. Database is available from Chapman and Hall/CRC Press at URL http://www.chemnetbase.com/.

(24) WDI, World Drug Index, 2004. Database is available from Derwent Publications Ltd. at URL http://thomsonderwent.com/.

(25) ChemDraw, 2001, version 7.0, software is available from Cambridge-Soft at URL http://products.cambridgesoft.com/.

(26) Rodgers, S.; Busch, J.; Peters, H.; Christ-Hazelhof, E. Building a Tree of Knowledge: Analysis of Bitter Molecules. *Chem. Senses* **2005**, *30*, 547−557.

(27) MDL Drug Data Report; MDL ISIS/HOST software, MDL Information Systems, Inc.

(28) Bender, A.; Mussa, H. Y.; Glen, R. C. Molecular Similarity Searching Using Atom Environments, Information-Based Feature Selection, and a Naïve Bayesian Classifier. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 170−178.

(29) Bender, A.; Glen, R. C. Molecular similarity: a key technique in molecular informatics. *Org. Biomol. Chem.* **2004**, *2*, 3204−3218.

(30) Bender, A.; Mussa, H. Y.; Glen, R. C. Similarity Searching of Chemical Databases Using Atom Environment Descriptors (MOLPRINT 2D): Evaluation of Performance. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1708−1718.

(31) Hert, J.; Willett, P.; Wilton, D. J.; Acklin, P.; Azzaoui, K.; Jacoby, E.; Schuffenhauer, A. Comparison of Fingerprint-Based Methods for Virtual Screening Using Multiple Bioactive Reference Structures. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1177−1185.

(32) MDL Drug Data Report, MDL ISIS/HOST software, MDL Information Systems, Inc.: San Leandro, CA.

(33) Clark, R. D.; Cramer, R. D.; Van Opdenbosch, N. Validation of the General Purpose Tripos 5.2 Force Field. *J. Comput. Chem.* **1989**, *10*, 982−1012.

(34) Klon, A. E.; Glick, M.; Davies J. D. Application of Machine Learning To Improve the Results of High-Throughput Docking Against the HIV-1 Protease. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 2216−2224.

(35) Bender, A. Studies on Molecular Similarity. Ph.D. Thesis, University of Cambridge, 2005.