ResearchGate

See discussions, stats, and author profiles for this publication at: https://www.researchgate.net/publication/275717327

Promiscuity and selectivity of bitter molecules and their receptors

ARTICLE *in* BIOORGANIC & MEDICINAL CHEMISTRY · APRIL 2015

Impact Factor: 2.79 · DOI: 10.1016/j.bmc.2015.04.025 · Source: PubMed

CITATION
1

READS
73

2 AUTHORS:

Antonella Di Pizio
Hebrew University of Jerusalem
12 PUBLICATIONS **11** CITATIONS
SEE PROFILE

Masha Y Niv
Hebrew University of Jerusalem
59 PUBLICATIONS **1,146** CITATIONS
SEE PROFILE

Available from: Masha Y Niv
Retrieved on: 14 January 2016

Bioorganic & Medicinal Chemistry xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Bioorganic & Medicinal Chemistry

journal homepage: www.elsevier.com/locate/bmc

Promiscuity and selectivity of bitter molecules and their receptors

Antonella Di Pizio, Masha Y. Niv *

Institute of Biochemistry, Food Science and Nutrition, Robert H Smith Faculty of Agriculture Food and Environment, The Hebrew University, Rehovot 76100, Israel
Fritz Haber Center for Molecular Dynamics, The Hebrew University, Jerusalem 91904, Israel

ARTICLE INFO

Article history:
Received 5 February 2015
Revised 4 April 2015
Accepted 8 April 2015
Available online xxxx

Keywords:
Chemosensory
Chemical senses
Molecular recognition
GPCR
Canonical binding site
Orthosteric
Modeling

ABSTRACT

Bitter taste is essential for survival, as it protects against consuming poisonous compounds, which are often bitter. Bitter taste perception is mediated by bitter taste receptors (TAS2Rs), a subfamily of G-protein coupled receptors (GPCRs). The number of TAS2R subtypes is species-dependent, and varies from 3 in chicken to 50 in frog. TAS2Rs present an intriguing case for studying promiscuity: some of the receptors are still orphan, or have few known agonists, while others can be activated by numerous, structurally dissimilar compounds. The ligands also vary in the repertoire of TAS2Rs that they activate: some bitter compounds are selective toward a single TAS2R, while others activate multiple TAS2Rs. Selectivity/promiscuity profile of bitter taste receptors and their compounds was explored by a chemoinformatic approach. TAS2R-promiscuous and TAS2R-selective bitter molecules were found to differ in chemical features, such as *A*log*P*, E-state, total charge, number of rings, globularity, and heavy atom count. This allowed the prediction of bitter ligand selectivity toward TAS2Rs. Interestingly, while promiscuous TAS2Rs are activated by both TAS2R-promiscuous and TAS2R-selective compounds, almost all selective TAS2Rs in human are activated by promiscuous compounds, which are recognized by other TAS2Rs anyway. Thus, unique ligands, that may have been the evolutionary driving force for development of selective TAS2Rs, still need to be unraveled.

© 2015 Elsevier Ltd. All rights reserved.

1. Introduction

Bitter taste is one of the basic taste modalities and is essential for rejecting potentially harmful substances.[1] The detection of the structurally diverse naturally occurring bitter compounds, as well as of synthetic bitter compounds, is mediated by the TAS2R[†] family of G-protein coupled receptors (GPCRs)[‡].[2] The number of TAS2R genes, as well as the fraction of pseudogenes, is species-dependent, indicating gene expansions and contractions during evolution.[3] Number of subtypes is varying widely, humans have 25 TAS2Rs, frogs 50 TAS2Rs, and chicken just 3 TAS2Rs.[4] TAS2Rs are expressed in many extra-oral tissues and are thought to have multiple physiological roles,[5] suggesting potential existence of endogenous ligands and the need for detecting an even larger repertoire of diverse ligands than previously thought.

The number of TAS2R agonists (the molecules that activate TAS2Rs and elicit bitter taste sensation) is estimated by thousands.[6] So far, we have gathered structures of close to 700 bitter compounds in the BitterDB database,[7] based on entries in Merck index, Fenaroli book of flavors and several publications, in which the bitter taste of the molecule was indicated. Importantly, for some of the bitter ligands, the association with particular TAS2Rs was established by in vitro assays, usually using calcium imaging,[6] and this information can also be accessed via the BitterDB (http://bitterdb.agri.huji.ac.il).[7]

In human, the 25 TAS2Rs represent about 4% of the GPCRs. Because of low sequence similarity of TAS2Rs with other GPCRs, their classification is ambiguous: TAS2Rs were grouped with the frizzled receptors[8] or considered a distinct family.[9] However, TAS2Rs are typically considered as class A.[10–12]

Intriguingly, some bitter taste receptors have a broad range of chemically diverse ligands, while others are narrowly tuned.[4,6,7] Promiscuity of proteins is highly abundant in nature and is increasingly investigated because of its implication in many areas of applied biology.[13] Elucidation of how selectivity and promiscuity are achieved within the binding pocket of proteins may contribute to the drug development process and enable rational manipulation of proteins toward binding of drugs.

The molecular properties influencing the pharmacological selectivity/promiscuity profile of GPCRs may be essential for understanding the molecular recognition processes and the side effects associated with many drugs targeting these receptors. Recently, Levit et al.[14] have focused on class A GPCRs with available

* Corresponding author.
E-mail address: masha.niv@mail.huji.ac.il (M.Y. Niv).
† TAS2R, sometimes abbreviated as T2R, bitter taste receptor.
‡ GPCRs, G-protein coupled receptors.

http://dx.doi.org/10.1016/j.bmc.2015.04.025
0968-0896/© 2015 Elsevier Ltd. All rights reserved.

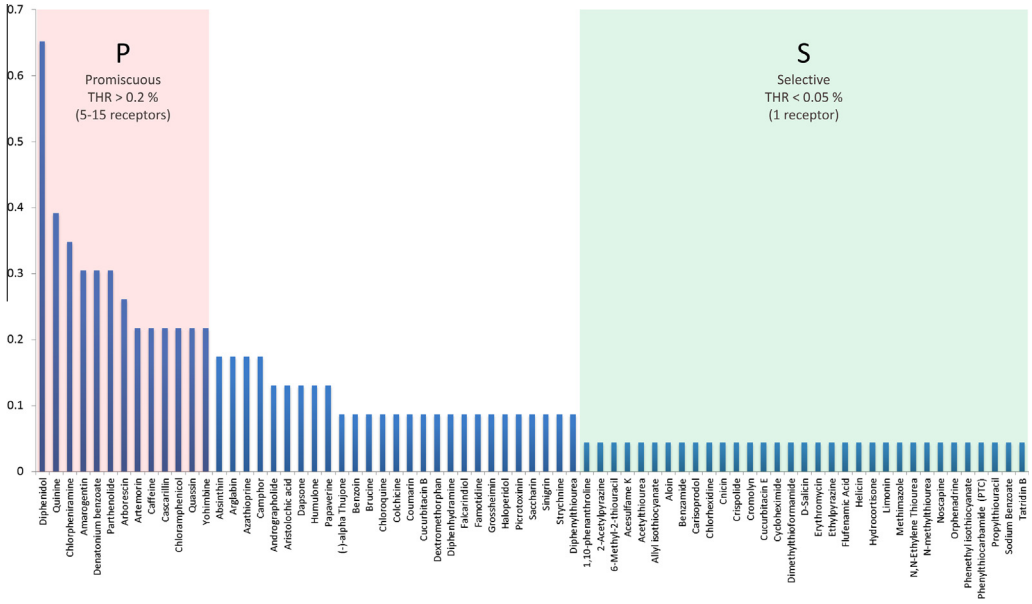


Figure 1. THR distribution of Set 1. Set 2 compounds are highlighted in red (promiscuous) and green (selective).

experimental structures and identified features of the orthosteric (canonical) binding site that correlate with the number and diversity of antagonists. Specifically, the number of unique scaffolds (a measure of the chemical variability) of antagonists was shown to be in correlation with the binding site exposure and hydrophobicity, and in negative correlation with the number of hydrogen bond donors in the binding site.

Ligands, as well as receptors, can exhibit promiscuity: many compounds interact with many different proteins. Studying the promiscuity (or polypharmacology) of ligands[15] has some practical advantages: while receptors are a given in the biological system, the ligands can be modified synthetically, and their pharmacological profile can (at least in theory) be tailor-made. Indeed, the classical idea of *'selective ligands for single molecular targets'* is starting to give way to the polypharmacology paradigm which is based on *'the promiscuous modulation of several molecular targets'*.[16–18] With the progress in GPCR structure determination and molecular modeling, more information becomes available, and GPCR ligand selectivity is being closely investigated. Shonberg et al.,[19] reviewing the current state of solved GPCR structures, highlighted ligand–receptor interactions in the binding pocket that can contribute to design of GPCR ligands with better affinity and subtype selectivity. A focused analysis of the subtype-selective compounds within the aminergic GPCRs has been carried out by Michino et al.[20] In particular, binding of ligands to the orthosteric binding site, to the secondary binding pocket, or concomitantly to both was discussed. The second and third binding scenarios were suggested as best strategies to be exploited for the optimization of existing lead compounds.

Since no experimental structures of TAS2Rs are available yet, iterative combination of homology modeling, site-directed mutagenesis and calcium imaging assays is used to unravel the ability of TAS2Rs to accommodate dissimilar bitter agonists.[7] Usages of both common and ligand-specific sub-pockets within the orthosteric binding site emerge as typical strategies.[21,22]

The current study focuses on the selectivity and promiscuity of bitter ligands and TAS2Rs. By applying a chemoinformatic approach, we aim to highlight chemical properties of TAS2R-promiscuous and TAS2R-selective compounds, and to investigate bitter selectivity of ligands in relation to selectivity and promiscuity of bitter receptors.

2. Results and discussion

A set of 104 bitter compounds was comprehensively tested by Meyerhof et al. on all human bitter taste receptors (hTAS2Rs)[§] under the same assay conditions.[6] Since the agonistic activity of these compounds toward hTAS2Rs occurs at different concentrations, our dataset, referred to as Set 1 hereafter, contains only 73 compounds that elicited TAS2R activation at concentration of 300 μM or lower. As can be seen in Figure 1, Set 1 includes both promiscuous and selective compounds toward hTAS2Rs. Compound promiscuity can be defined as THR[¶] (target hit-rate) parameter: the number of targets hit at a specific concentration divided by the number of targets tested.[23] We consider as targets hTAS2Rs only, and define compounds with THR $\geqslant$ 0.2 as TAS2R-promiscuous (P) and compounds with THR $\leqslant$ 0.05 as TAS2R-selective (S). We will refer to P and S compounds without the intermediate promiscuity compounds as Set 2 (Supplementary data Table S1).

A subset of the compounds in Set 1 were recently profiled also against the bitter taste receptors in chicken (ggTAS2Rs)[‖], turkey (mgTAS2Rs)[††], zebra finch (tgTAS2Rs)[‡‡], and six representative bitter taste receptors out of the total 50 in frog (xtTAS2Rs)[§§].[4] Among these, applying a cut-off of 300 μM (see Methods) results in 25 compounds, referred to as Set 3 (Supplementary data Table S1). Chicken, turkey, zebra finch and frog receptors were activated by 14, 8, 2 and 21 compounds respectively.

Notably, the TAS2R-selectivity/promiscuity profile seems to hold across species (Fig. 2A). This is not the general rule, as discussed by Hopkins et al.[16] It will be interesting to follow cross-species profiles as repertoires of TAS2R of additional species become available in the future. On the other hand, if 'target' definition for THR is extended to include any target (and not just TAS2R), the promiscuity profile changes dramatically (Fig. 2B).

The fact that selectivity and promiscuity of compounds toward TAS2Rs does not generalize to other targets (Fig. 2B) could be due

§ hTAS2R, human bitter taste receptor.
¶ THR, target hit-rate parameter.
‖ ggTAS2R, chicken (*Gallus gallus*) bitter taste receptor.
†† mgTAS2R, turkey (*Meleagris gallopavo*) bitter taste receptor.
‡‡ tgTAS2R, zebra finch (*Taeniopygia guttata*) bitter taste receptor.
§§ xtTAS2R, frog (*Xenopus tropicalis*) bitter taste receptor.

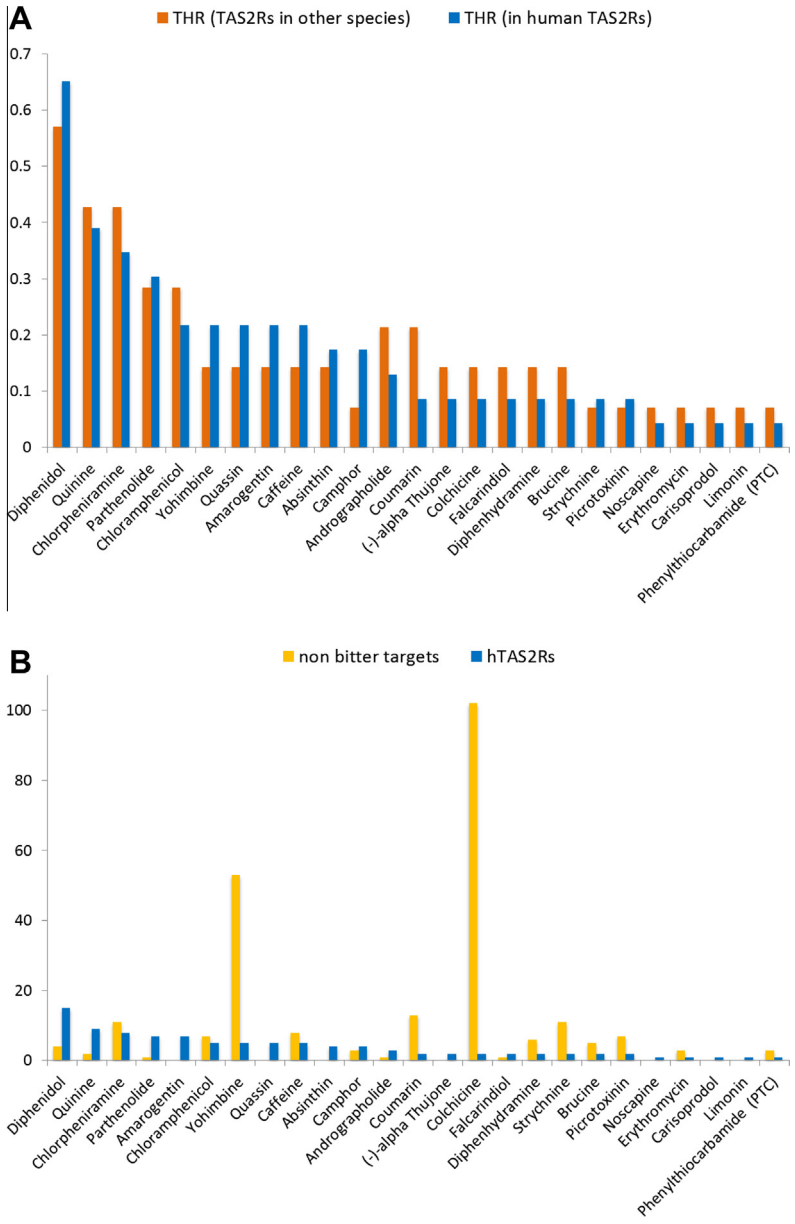
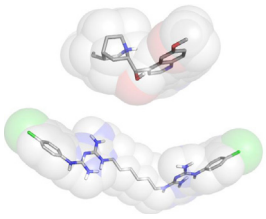


Figure 2. (A) THR of bitter receptors in humans (blue bars) and other species (dark orange bars) toward Set 3; (B) number of receptors reported to interact with Set 3 compounds: hTAS2Rs (blue bars), non-bitter targets (yellow bars; data were taken from ZINC Database,[24] BindingDB[25] and IUPHAR Database,[26] considering all kind of targets).

to several reasons: (1) The physiologically relevant range of bitter compounds, as present in food and accessible to taste receptors expressed in the oral cavity, is micromolar.[27] In vitro activation of TAS2R-expressing cells ranges from 10 nm for flufenamic acid to sub-millimolar concentrations.[6] While more potent compounds may very well exist, they are yet unknown.[7] On the other hand, the concentrations that are relevant for potential drugs are usually in the nanomolar or sub-nanomolar ranges, and compounds with micromolar activities are rarely reported. Thus the relevant concentration ranges and the available data for TAS2Rs and for general protein targets are different. (2) The existing data for various compounds may be biased by the pharmaceutical interest they present. Some of the compounds could be tested against many targets, while others could be hardly studied at all. To test these hypotheses, a sub-group of Set 3 compounds that were screened, together with other 638 different compounds, against more than a hundred of human targets,[28] have been analyzed. Focusing on these compounds, we saw that the promiscuity toward bitter receptors differs from promiscuity toward other targets. This observation still holds when considering only compounds that are reported to have micromolar (1–100 μM) binding affinities (Supplementary data Fig. S1). These data indicate that there might be a real difference between promiscuity toward TAS2Rs versus promiscuity toward non-TAS2R targets. The robustness and significance of this finding remains to be explored.

2.1. TAS2R-promiscuous and TAS2R-selective compounds

In order to investigate the differences in chemical structure profiles between TAS2R-promiscuous and TAS2R-selective compounds (highlighted in red and green, respectively, in Fig. 1), a chemoinformatic analysis was carried out on Set 2.



Quinine	Promiscuous THR= 0.39	Heavy Atoms: 24	Globularity: 0.85
Chlorhexidine	Selective THR= 0.04	Heavy Atoms: 34	Globularity: 0.71

Figure 3. Illustration of globularity and heavy atom content in examples of promiscuous and selective bitter compounds.

Descriptors that are commonly used in structure–activity relationship (SAR)[¶¶] and virtual screening campaigns for drug discovery and development were calculated for Set 2 molecules. These descriptors include: molecular weight (MW) and heavy atom count, as measures of size; *A*log*P* (octanol/water partition coefficient) to evaluate lipophilicity; hydrogen bond donors (HBD), hydrogen bond acceptors (HBA), polar surface area (PSA), total charge, and electrotopological index (E-state) accounting for polarity and distribution of charge; rotatable bonds (RB), accounting for flexibility; and descriptors aimed to quantify the complexity of molecules, such as Fsp3 (number of sp^3 hybridized carbons/total carbon count), number of chiral centers, ring and aromatic ring counts; and finally volume and globularity as descriptors of three dimensional molecular shape.

A commonly used machine learning technique, nominal logistic classification, was used to develop a classifier of bitter compounds using the descriptors mentioned above.

Even though no single descriptor correlates with THR very strongly (see Supplementary data Fig. S2), we were able to obtain a classifier that can discriminate between two groups, TAS2R-selective and TAS2R-promiscuous compounds. In particular, the significant descriptors contributing to the model (see Eq. 1 below) are: *A*log*P*, E-state, total charge, number of rings and globularity (in inverse correlation with compound selectivity), and heavy atom count (in positive correlation with compound selectivity).

$$S = \frac{1}{1 + e^{\{-[68.4 - (1.4 * A\ \log\ P) - (0.6 * \text{E-state}) + (3.1 * \text{Heavy atoms}) - (3.1 * \text{Rings}) - (8.1 * \text{Charge}) - (61.4 * \text{Globularity})]\}}} \quad (1)$$

where *S* is the probability of selectivity.

This finding means that the lower the lipophilicity of the molecule (*A*log*P*), the higher is the probability for the molecule to be TAS2R-selective. Lipophilicity is generally recognized as an important determinant of ligand selectivity, and previous analyses found a similar correlation,[29] though lipophilicity alone is not generally useful to evaluate or optimize the selectivity of individual compounds.[30]

Higher number of rings contributes to lower probability of bitter ligand selectivity, and this is also not surprising: the number of rings itself usually correlates with lipophilicity,[30,31] and was previously found in correlation with compound promiscuity.[32]

The negative contribution of total charge to probability of selectivity seems to conflict with the trend of the *A*log*P*, since charge usually indicates polarity. But it has been demonstrated that the large majority of promiscuous compounds is positively charged, usually due to ionizable amines.[30,32,33] We note that TAS2R-promiscuous compounds in Set 2 are positively charged (see Supplementary data Fig. S3C), and indeed the positive atoms in Set 2 are mostly charged amines (correlation of 0.74 between number of positive atoms and number of positive amines).

[¶¶] SAR, structure–activity relationship.

Table 1
Confusion matrix

	Predicted selective	Predicted promiscuous
Actual selective	30	3
Actual promiscuous	4	9

E-state index combines both molecular electronic and topological information into a single quantitative description,[34] and bears a negative contribution to ligand selectivity probability.

The molecular size is usually considered an important determinant for promiscuity, as a surrogate for molecular complexity, but analyses of different datasets produce inconsistent results: both positive and inverse correlation between selectivity and molecular size were found.[29] In our analysis, selectivity increases with increasing size, such that bigger molecules (higher heavy atom numbers) are more likely to be TAS2R-selective.

Interestingly, the globularity emerged as one of the most influential descriptors, negatively contributing to selectivity probability (parameter estimate of −61.42 in the logistic model). Globularity descriptor is calculated as $(4\pi r^2)/(\text{SASA})$, where r is the radius of a sphere with a volume equal to the molecular volume; and SASA is the solvent accessible surface area. The values typically range between 0.75 for flatter and 0.95 for more globular compounds (QikProp, version 3.9, Schrödinger, LLC, New York, NY, 2014). In Set 2, globularity anticorrelates with heavy atom numbers (correlation of −0.63, see Supplementary data Fig. S2), meaning that small molecules are more likely to be globular. These small and globular molecules are also more likely to be TAS2R-promiscuous than large flat molecules. Figure 3 furnishes an illustrative example of compounds with similar polarity (i.e., total charge and *A*log*P*), but different THR values, with promiscuous, small and more globular quinine, and the selective, large and relatively flat chlorhexidine.

The performance of the logistic model (Eq. 1) is attested by the confusion matrix: the model is able to correctly assign 30/33 selective molecules and 9/13 promiscuous molecules (in Table 1). There are few false negatives, most of them due to incorrect assignment of promiscuous ligands as selective ones.

Figure 4 shows the probability values computed by our models: by default the compounds are considered selective if the probability of selectivity is higher than 0.5 (which is the same as the probability of promiscuity lower than 0.5) and promiscuous if the probability of selectivity is lower than 0.5 (which is the same as the probability of promiscuity is higher than 0.5). We note that the prediction of selectivity has a higher sensitivity since the true positives are more than the false negatives ($SE_{selectivity}$: 0.91,

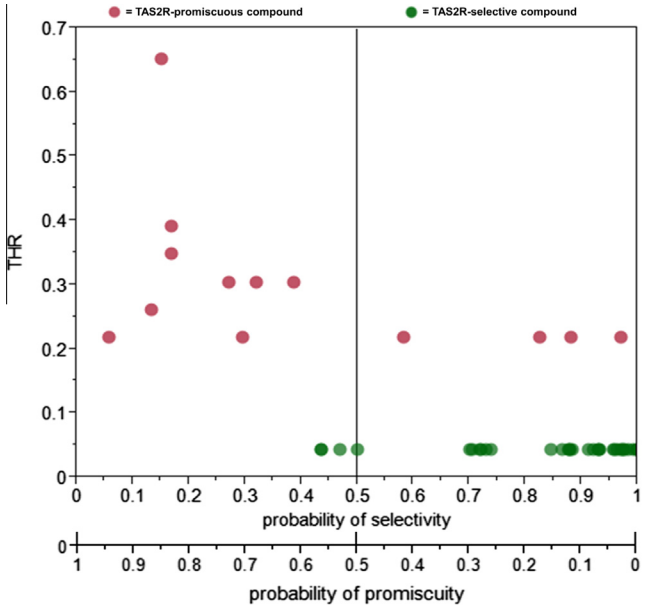


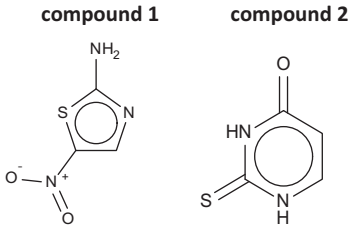
Figure 4. Selectivity and promiscuity prediction versus THR. TAS2R-selective compounds are shown in green, TAS2R-promiscuous compounds in red.

$SP_{selectivity}$: 0.69), whereas the prediction of promiscuity has higher specificity because of fewer false positives ($SE_{promiscuity}$: 0.69, $SP_{promiscuity}$: 0.88). According to the specific research goal, it is possible to modify the cut-off to increase the specificity or the sensitivity in the direction of promiscuity or selectivity prediction.

Recently, Ji et al.,[35] by investigating the activation of almost all human bitter receptors in vitro, identified novel bitter compounds, which were not a part of Sets 1–3. Applying the activity cut-off of 300 μM to the Ji set, 10 compounds were selected as validation set; 8 are selective, activating just one receptor or at most two, while compounds **1** and **2** (Fig. 5) are more promiscuous, having THR of 0.18 and 0.23, respectively. According to our definition of promiscuity (THR ⩾ 0.2), compound **1** should not be considered a TAS2R-promiscuous compound, and **2** is just above the threshold. Our model is able to capture the difference of these two compounds from the others: the prediction of promiscuity is 0.42 for **1** and 0.43 for **2**, while the 8 selective compounds are indeed predicted as selective (0.8–1 probability of selectivity). In Supplementary data the validation set compounds with chemical structures, THR and predicted promiscuity are reported (Supplementary data Table S2).

Encouraged by the excellent performance of the classifier on an external set, we next applied the classifier to molecules from BitterDB,[7] that have not yet been screened against all human TAS2Rs. Principal component analysis (PCA) of the BitterDB Set is dominated by PC1 (the larger loadings correspond to MW and PSA) and PC2 (the larger loadings are Fsp3 and number of aromatic rings), that together account for 61% of the BitterDB variability (Wiener et al., in preparation). Compounds from BitterDB that are promiscuous (dark red), predicted to be promiscuous (light red), selective (dark green) and predicted to be selective (light green) are shown on the PCA map (Fig. 6).

The predicted selective and promiscuous compounds span throughout the bitter space. It should be noted that both the initial dataset and the predicted set have more selective than promiscuous compounds (∼400 predicted selective vs ∼200 predicted promiscuous compounds). This could be either the typical ratio, or we might be witnessing a biasing in prediction due to the imbalanced set (13 promiscuous vs 33 selective) used to develop the predictor. To test the training bias, the selective compound set was randomly split into two subsets of sizes that match the size



Figuer 5. Structures of compounds 1 and 2 in the validation set.

of the promiscuous training set. The models generated with the new sets were used to predict the promiscuity and selectivity of BitterDB compounds, and confirmed that the number of predicted selective compounds is higher than number of predicted promiscuous compounds (see Supplementary data). Therefore, so far the experimental data from both Meyerhof et al.[6] (used for training set) and Ji et al.[35] (used for validation) have shown that it is easier to find TAS2R-selective compounds than TAS2R-promiscuous ones.

It is therefore interesting to zoom-in into regions in the chemical space that show a higher concentration of molecules predicted to be TAS2R-promiscuous. We highlight molecules in three classes in these mostly-promiscuous regions: we find opioids, barbiturates and amino ester anesthetics in the class 1; ephedrines and saccharin derivatives in class 2; and flavonoids in class 3 (see Tables S3–S5 for the complete list of molecules). Further investigation may shed light on physiological importance of a molecule being TAS2R-promiscuous.

2.2. Promiscuous and selective TAS2Rs

The 25 human TAS2Rs have varying receptive range,[6,7] with the number of known agonists varying from 31 to zero (for orphan TAS2Rs) when considering Set 1, or from 16 to zero when considering Set 3. Analyzing Set 3, varying receptive range is also apparent for the frog receptors tested by Behrens et al.;[4] on the other hand, all tested zebra finch receptors are selective, both turkey receptors show medium level of promiscuity, and all three chicken receptors seem to recognize numerous compounds. To compare the receptive ranges of TAS2Rs in the different species, we used Set 3 compounds and defined the promiscuity index (PI)[‖‖], that is, the number of bitter compounds that activate the receptor divided by the total number of molecules. Another measure of receptor promiscuity that can be useful, since it represents the diversity of the ligands set,[14] is the number of unique scaffolds (NUS)[†††] for each receptor divided by the total number of NUS, that is, PI_{NUS}.

Figure 7 illustrates that the correlation between the PI and PI_{NUS} is high: where there are more ligands, there is also a higher number of unique scaffolds.

Previous analysis on class A GPCR promiscuity identified positive correlation of promiscuity toward antagonists with binding site exposure and hydrophobicity, and inverse correlation with the area of hydrogen bond donors in the binding site.[14] This predictive model was successfully applied to several A GPCRs using the newly published X-ray structures, or using high quality homology models. However, performing a similar analysis on bitter taste receptors presents some challenges. Since no structure is yet available for TAS2Rs, one needs to rely on homology models built on very distantly related templates. Due to limitations related to loops modeling, extracellular loop 2 (ECL2)[‡‡‡] structure is not reliable and thus needs to be deleted.[36] Furthermore, only a few bitter taste

[‖‖] PI, promiscuity index.
[†††] NUS, number of unique scaffolds.
[‡‡‡] ECL2, extracellular loop 2.

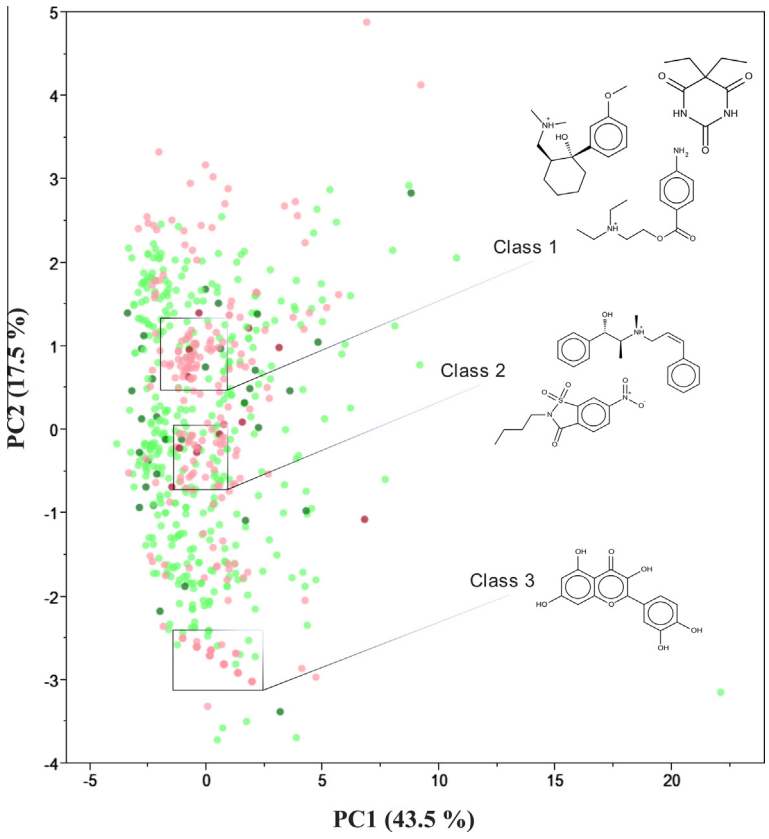


Figure 6. PCA of bitter compounds, colored by predicted selectivity: selective and promiscuous compounds are shown in green and red, respectively; predicted selective and predicted promiscuous compounds in light green and light red, respectively. PC1 and PC2 explain 43.5% and 17.5% of variation, respectively.

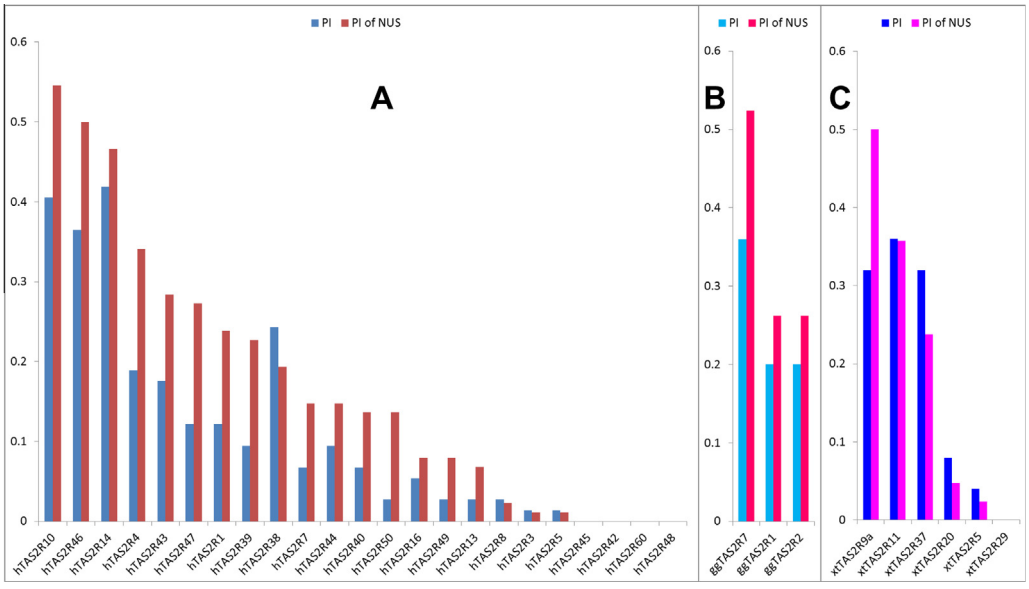


Figure 7. Target promiscuity indexes of human (A), chicken (B) and frog (C) bitter receptors toward Set 3 compounds.

antagonists are known, and thus we concentrate on agonists, rather than on antagonists.

Here we attempted to account for molecular determinants of TAS2Rs that influence the levels of promiscuity toward their agonists. We considered as promiscuous the receptors with $PI_{NUS} \geqslant 0.5$, and as selective those with $PI_{NUS} \leqslant 0.2$. The resulting logistic model is described in Supplementary data. The TAS2R descriptors identified as significant, or close to significant, are the binding site surface, the area of hydrogen bond donors and the hydrophobic content of the binding site. This is in agreement with the promiscuity descriptors arising from model of class A promiscuity toward antagonists.[14] The surface seems to be the most prominent property distinguishing between selective and promiscuous TAS2Rs. Because of the few observations (6 promiscuous and 13

selective or orphan bitter taste receptors) and because of the uncertainness due to the homology modeling, the logistic model (see Supplementary data) cannot be considered very reliable. As more data on compound activities and better structural templates for TAS2R will become available, a more reliable model predicting promiscuity of TAS2Rs based on structural features will be likely obtained.

2.3. Are selective TAS2Rs activated by selective compounds?

It is interesting to analyze the relation between the selectivity and promiscuity of ligands to selectivity and promiscuity of their receptors. Many studies have been undertaken on receptors,[37] transporters[38] and enzymes[39] to investigate ligand and protein promiscuity, and in particular their connection with protein function and its evolution over time. Proteins interacting with different substrates result in multiple biological functions (also called 'generalist' proteins), while selective proteins are thought to have evolved from the more promiscuous forms to fulfill novel and more specified functions (and, thus, defined as 'specialist' proteins).[40,41]

Broadly tuned and narrowly tuned bitter receptors could be associated with 'generalist' and 'specialist' receptors, respectively.[4] If selective TAS2Rs evolved to recognize unique ligands, which are not recognized by other TAS2Rs, we would expect that TAS2R-selective compounds activate selective TAS2Rs. To see if this is really the case, Set 1 has been analyzed.

Figure 8 shows the matrix of receptor selectivity versus compound selectivity. It illustrates that promiscuous receptors (i.e., hTAS2R14, 10 and 46) as well as selective receptors (i.e., hTAS2R13, 49, 50, 3 and 5) are activated by compounds of varying selectivity (from promiscuous (red) to less promiscuous (yellow) and selective (light green). The same is true for selective receptors, which are activated by both promiscuous and selective compounds. In this analysis, hTAS2R38 is activated by many, mainly selective compounds: this is due to the high abundance of the isothiocyanate scaffold (and consequently of selective hTAS2R38 agonists) in Set 1.[42]

From the side of the ligands, promiscuous compounds can activate both promiscuous and selective receptors; selective compounds mainly activate promiscuous receptors.

1,10-Phenantroline and cromolyn are the only selective compounds that activate only selective receptors, hTAS2R5 and hTAS2R49, respectively. Cromolyn is a prescription drug used for treatment of asthma,[43] which is a khellin analog (from *Ammi visnaga*, better known as Khella). Interestingly, its mechanism of action is not fully understood. Notably, TAS2Rs are considered as novel targets for treatment of respiratory diseases[44] and hTAS2R49 seems to be involved in the chronic rhinosinusitis.[45] Thus, some of the therapeutic actions of cromolyn may be due to its action on a bitter taste receptor. On the other hand, the evolution of bitter taste receptor hTAS2R49 might be related to its physiological role in the upper airways. hTAS2R49, and other selective TAS2Rs, which are activated by a single, naturally occurring, and physiologically beneficial compound, might have evolved to recognize it. This is in agreement with the idea that the existence of narrow- and broad-tuned receptors could be related to their functional specialization. hTAS2R5 and hTAS2R49 are the only specialized TAS2Rs activated by selective bitter compounds, while this does not seem to be the case for all other selective TAS2Rs. Thus we suggest that the relevant unique compounds of the selective TAS2Rs are yet to be discovered. An alternative hypothesis is that the driving force for development of selective TAS2Rs is due to their ability to be activated by bitter compounds that inactivate some other TAS2Rs, as shown by Brockhoff et al.[46] Identification of additional bitter compounds and characterization of their pharmacology are required in order to further explore this complex question.

3. Conclusions

The chemoinformatic analysis presented here aims to answer fundamental questions about ligand–receptor recognition and the selectivity/promiscuity profile of bitter taste receptors.

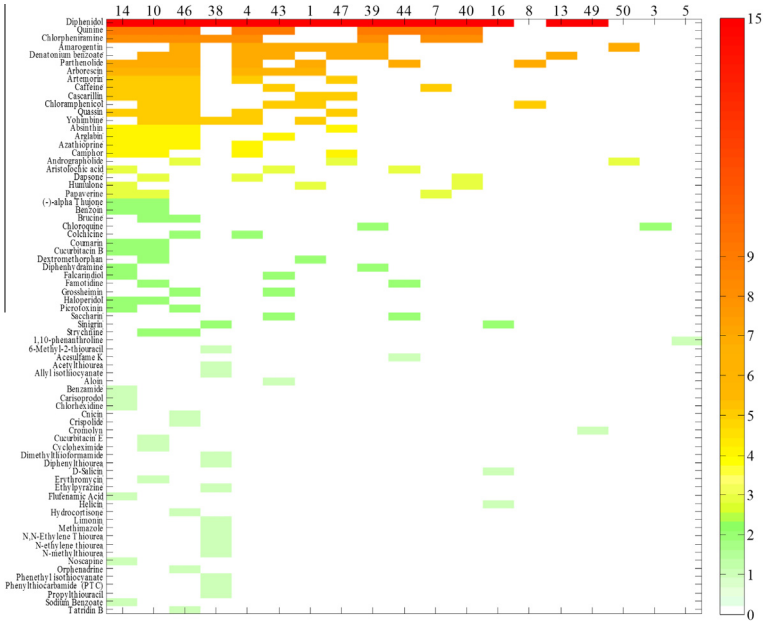


Figure 8. Target selectivity versus compound selectivity. Compounds (*y*-axis) are ranked by the number of receptors they can activate (most promiscuous in the top, most selective in the bottom) and receptors (*x*-axis, only the number of TAS2Rs is reported) by the number of compounds they can be activated by (from maximal on the left to minimal on the right). Color map indicates the compound promiscuity (the cell is colored in white if the compound cannot activate the receptor, from light green to red if can activate from 1 to 15 receptors).

The lack of structural data limits the possibility to fully investigate the structural determinants of TAS2R-promiscuity, but the properties of the binding site previously identified as relevant for A GPCR promiscuity[14] correlate also with the TAS2R-promiscuity.

Through physicochemical descriptors of the bitter compounds, we have developed a classifier that is able to discriminate between TAS2R-selective and TAS2R-promiscuous compounds. Promiscuous bitter compounds do not seem to be promiscuous toward other targets, but the ligand properties that contribute to TAS2R-promiscuity are in general agreement with properties found in previous analysis carried out on large database of drugs and drug candidates.

The logistic model for TAS2R selectivity versus promiscuity was validated on a small external set and applied to BitterDB tastants. Interestingly, predicted TAS2R-promiscuous compounds cluster in sub-regions of the chemical space.

The ability to predict TAS2R-selectivity of a ligand is an important outcome, since selective targeting of individual bitter taste receptor is required in order to disentangle the complex roles the bitter taste receptors family members may have.

According to the knowledge we have so far, selective bitter taste receptors can accommodate both promiscuous and selective compounds. Therefore the specific chemical classes identified as potential promiscuous compounds could be screened on orphan bitter taste receptors and potentially lead to their de-orphanization. In addition, unique ligands, that may have been the evolutionary driving force for development of selective TAS2Rs, still need to be unraveled.

4. Methods

All structures shown in the text and in Supplementary data Tables were realized with JChem for Excel (JChem for Excel 6.1.0.180, 2013, ChemAxon (http://www.chemaxon.com)).

4.1. Analysis of ligands

Compounds used in our analysis were all downloaded from the BitterDB[7] in sdf format. The 3D structures, including stereoisomers, tautomers and protonation states at pH 7.0 ± 0.5 were generated with LigPrep (version 2.9, Schrödinger, LLC, New York, NY, 2014) as implemented in Maestro (version 9.8, Schrödinger, LLC, New York, NY, 2014).

Particularly, the following sets of molecules were used:

Set 1: 73 compounds which are a subset of the 104 compounds tested on cells expressing individual hTAS2Rs,[6] and include only compounds that activate at least one of the human TAS2Rs at concentration <300 μM or lower (this cut-off allows to remove from our analysis molecules with very low affinity, but retaining information from molecules within the physiologically relevant (high micromolar) ranges).

Set 2: 46 compounds selected from Set 1 according to their THR: 13 promiscuous compounds (THR ⩾ 0.2) and 33 selective compounds (THR ⩽ 0.05). To calculate the THR, hTAS2R9 and hTAS2R41 were excluded from the count of total number of tested targets because of experimental problems encountered with them: the use of a nonfunctional hTAS2R9 variant (A187V) prevented from detecting cognate bitter chemicals;[6] hTA2R41 is not activated by any compounds in Meyerhof et al.,[6] but could be activated by chloramphenicol in other studies.[47] Set 2 compounds are shown in Supplementary data Table S1.

Set 3: 25 compounds with the relative fluorescence changes ($\Delta F/F$) >0.3 at concentration of 300 μM or lower, when applied to cells expressing ggTAS2R, mgTAS2R, tgTAS2R and xtTAS2R receptors.[4] Set 3 compounds are shown in Supplementary data Table S1.

Validation Set: 10 most potent compounds from Ji et al.[35] To calculate the THR of these compounds, hTAS2R9 and hTAS2R41 were excluded from the count of total number of tested targets as before. hTAS2R48 was excluded because of its failed expression in Ji et al.[35] Compounds in the validation set are shown in Supplementary data Table S2.

Set 2 was used to build the logistic model for classifying bitter compounds into selective and promiscuous. The following molecular descriptors were calculated with Canvas (version 1.9, Schrödinger, LLC, New York, NY, 2014):[48] MW, heavy atom count, *A*log*P*, HBD, HBA, PSA, total charge, E-state, RB, chiral centers, aromatic rings, ring count, globularity, volume. A protocol in Discovery Studio 4.0 (Accelrys) was used for Fsp3 calculation.

Statistical analyses were carried out in the JMP Pro statistical software package (version 10.0.0; SAS Institute Inc., NC, USA). Logistic classification analysis was used to determine the ligand descriptors that discriminate between selectivity and promiscuity.

Descriptors were selected according to their statistical significance, calculated by the likelihood ratio test. The likelihood ratio tests evaluate if the specified model is significantly better than a model without a given effect. If a test shows significance, then the higher order model is justified. The minimum AICc (the Corrected Akaike's Information Criterion) was used as the model selection criterion. AIC is a well-known way to balance the model fit against the model complexity (see Eq. 2).[49]

$$\mathrm{AIC} = -2\ \log(L) + 2k \tag{2}$$

where L is the maximized value of the likelihood function for the estimated model, and k is the number of parameters.

The Corrected Akaike's Information Criterion is a bias-corrected version of the AIC, that takes into consideration the sample size, and is given by Eq. 3:[50]

$$\mathrm{AICc} = -2\ \log(L) + 2k + \frac{2k(k+1)}{(n-k-1)} \tag{3}$$

where n is the number of observations.

Smaller values of AICc indicate better model fits as assessed by the likelihood function. The AICc penalizes the number of parameters, since it includes a penalty that is an increasing function of the number of estimated parameters. The penalty discourages overfitting (indeed increased number of descriptors tends to improve the model fits), considering the value of n.

The AICc of our model is 54.06 (the maximal value corresponds to 74.93 when using all 15 descriptors). In Table 2 the parameter estimates and probabilities are shown.

Volume, with the lowest influence in the model (parameter estimate of −0.03), is slightly above the significance level of 0.05. Distributions of ligand descriptors in the classifier were extrapolated with the distribution module and correlations between them were computed using the Multivariate module (Supplementary data Figs. S2 and S3).

The sensitivity (SE)[§§§] and specificity (SP)[¶¶¶] of the model were evaluated by the Eqs. 4 and 5:

$$\mathrm{SE} = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{4}$$

$$\mathrm{SP} = \frac{\mathrm{TN}}{\mathrm{TN} + \mathrm{FP}} \tag{5}$$

where TP, TN, FP and FN represent the number of true positives, true negatives, false positives, and false negative, respectively. The default cut-off of 0.5 was used to calculate both $\mathrm{SE}_{\mathrm{selectivity}}$ (0.91) and $\mathrm{SP}_{\mathrm{selectivity}}$ (0.69), and $\mathrm{SE}_{\mathrm{promiscuity}}$ (0.69) and $\mathrm{SP}_{\mathrm{promiscuity}}$ (0.88).

§§§ SE, sensitivity.
¶¶¶ SP, specificity.

Table 2
Parameter estimates

Term	Estimate	Std error	ChiSquare	*p* values
*A*log*P*	−1.4495748	0.608134	7.07	0.0078*
E-state	−0.6289038	0.2715418	8.39	0.0038*
Heavy atom count	3.13926921	1.4664225	6.93	0.0085*
Ring count	−3.0935651	1.3070589	9.24	0.0024*
Total charge	−8.0890955	3.3658643	10.07	0.0015*
Globularity	−61.420756	30.525853	6.06	0.0139*
Volume	−0.0375157	0.0217711	3.66	0.0557

Chi-Square is the Likelihood-ratio Chi-square test for the hypothesis in which all regression parameters are zero. The Likelihood-ratio Chi-square tests are calculated as twice the difference of the log-likelihoods between the full model and the model constrained by the hypothesis to be tested (the model without the effect).
p value is the probability of obtaining a greater Chi-square value by chance alone if the specified model fits no better than the model without the effect.
* Statistically significant at the 0.05 level.

The model was validated with an external set of molecules (validation set); the structures and the promiscuity prediction of these molecules are reported in Supplementary data Table S2.

4.2. Analysis of receptors

Sequence alignment of TAS2Rs and the GPCR templates in the active conformation, β2 adrenergic receptor and M2 muscarinic receptor, PDB IDs: 3SN6[51] and 4MQS,[52] respectively, was performed with PROMALS[53] and manually adjusted according to previous alignments.[7,54]

ggTAS2Rs, xtTAS2Rs tested in Behrens et al.,[4] and hTAS2Rs were analyzed. hTAS2R9 and hTAS2R41 were excluded (see above).

Homology modeling and side chain refinement were performed with Prime (version 3.5, Schrödinger, LLC, New York, NY, 2014). Each structure was prepared using the Protein Preparation Wizard implemented in Maestro (version 9.8, Schrödinger, LLC, New York, NY, 2014).

The orthosteric binding site was defined using a 4 Å sphere centered at $W88^{3.32}$ and $V89^{3.33}$ (TAS2R10 numbering),[54] and was subjected to SiteMap[55] analysis (version 3.1, Schrödinger, LLC, New York, NY, 2014). SiteMap was run in the default mode. Calculated surfaces representing the hydrophobic, donor, acceptor, hydrophilic (donor + acceptor) regions, and the total area of the binding site were used as descriptors. The relative percentage of hydrophobic (M/L/I/V), aromatic (W/F/Y), polar (S/T/N/Q/H), basic (K/R), acidic (D/E), and small (A/C/G) binding site content were calculated for each receptor and used as sequence-based descriptors.

With the logistic classification analysis in JMP Pro, as described above, we tried to correlate the descriptors of the binding site with the PI_{NUS} of TAS2Rs. The NUS in each set was computed using the Bemis–Murcko scaffold decomposition protocol implemented in Canvas.[56] The results of this analysis are shown in Supplementary data.

Acknowledgments

Fellowship from the Nutrigenomics and Functional Foods Research Center at the Institute of Biochemistry, Food Science and Nutrition (The Hebrew University) to A.D.P., the German Research Foundation DFG (ME 1024/8-1) and the Israel Science Foundation (No. 432/12) grants to M.Y.N. are gratefully acknowledged. M.Y.N. participates in the European COST Action CM1207 (GLISTEN). The authors thank Dr. Dizza Bursztyn for consultation on statistical analysis and Dr. Maik Behrens and Ayana Wiener for stimulating discussions.

Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.bmc.2015.04.025.

References and notes

1. Glendinning, J. I. *Physiol. Behav.* **1994**, *56*, 1217.
2. Adler, E.; Hoon, M. A.; Mueller, K. L.; Chandrashekar, J.; Ryba, N. J. P.; Zuker, C. S. *Cell* **2000**, *100*, 693.
3. Dong, D.; Jones, G.; Zhang, S. *BMC Evol. Biol.* **2009**, *9*, 12.
4. Behrens, M.; Korsching, S. I.; Meyerhof, W. *Mol. Biol. Evol.* **2014**, *31*, 3216.
5. Foster, S. R.; Roura, E.; Thomas, W. G. *Pharmacol. Ther.* **2014**, *142*, 41.
6. Meyerhof, W.; Batram, C.; Kuhn, C.; Brockhoff, A.; Chudoba, E.; Bufe, B.; Appendino, G.; Behrens, M. *Chem. Senses* **2010**, *35*, 157.
7. Wiener, A.; Shudler, M.; Levit, A.; Niv, M. Y. *Nucleic Acids Res.* **2012**, *40*, D413.
8. Fredriksson, R.; Lagerstrom, M. C.; Lundin, L. G.; Schioth, H. B. *Mol. Pharmacol.* **2003**, *63*, 1256.
9. Horn, F.; Bettler, E.; Oliveira, L.; Campagne, F.; Cohen, F. E.; Vriend, G. *Nucleic Acids Res.* **2003**, *31*, 294.
10. Matsunami, H.; Montmayeur, J. P.; Buck, L. B. *Nature* **2000**, *404*, 601.
11. Nordstrom, K. J.; Sallman Almen, M.; Edstam, M. M.; Fredriksson, R.; Schioth, H. B *Mol. Biol. Evol.* **2011**, *28*, 2471.
12. Di Pizio, A.; Niv, M. Y. *Israel J. Chem.* **2014**, *54*, 1205.
13. Nobeli, I.; Favia, A. D.; Thornton, J. M. *Nat. Biotechnol.* **2009**, *27*, 157.
14. Levit, A.; Beuming, T.; Krilov, G.; Sherman, W.; Niv, M. Y. *J. Chem. Inf. Model.* **2014**, *54*, 184.
15. Keiser, M. J.; Setola, V.; Irwin, J. J.; Laggner, C.; Abbas, A. I.; Hufeisen, S. J.; Jensen, N. H.; Kuijer, M. B.; Matos, R. C.; Tran, T. B.; Whaley, R.; Glennon, R. A.; Hert, J.; Thomas, K. L. H.; Edwards, D. D.; Shoichet, B. K.; Roth, B. L. *Nature* **2009**, *462*, 175.
16. Hopkins, A. L.; Mason, J. S.; Overington, J. P. *Curr. Opin. Struct. Biol.* **2006**, *16*, 127.
17. Hopkins, A. L. *Nat. Chem. Biol.* **2008**, *4*, 682.
18. Peters, J. U. *J. Med. Chem.* **2013**, *56*, 8955.
19. Shonberg, J.; Kling, R. C.; Gmeiner, P.; Löber, S. *Bioorg. Med. Chem.* **2014**.
20. Michino, M.; Beuming, T.; Donthamsetti, P.; Newman, A. H.; Javitch, J. A.; Shi, L. *Pharmacol. Rev.* **2015**, *67*, 198.
21. Brockhoff, A.; Behrens, M.; Niv, M. Y.; Meyerhof, W. *Proc. Natl. Acad. Sci. U.S.A.* **2010**, *107*, 11110.
22. Born, S.; Levit, A.; Niv, M. Y.; Meyerhof, W.; Behrens, M. *J. Neurosci.: Official J. Soc. Neurosci.* **2013**, *33*, 201.
23. Azzaoui, K.; Hamon, J.; Faller, B.; Whitebread, S.; Jacoby, E.; Bender, A.; Jenkins, J. L.; Urban, L. *ChemMedChem* **2007**, *2*, 874.
24. Irwin, J. J.; Sterling, T.; Mysinger, M. M.; Bolstad, E. S.; Coleman, R. G. *J. Chem. Inf. Model.* **2012**, *52*, 1757.
25. Liu, T.; Lin, Y.; Wen, X.; Jorissen, R. N.; Gilson, M. K. *Nucleic Acids Res.* **2007**, *35*, D198.
26. Sharman, J. L.; Benson, H. E.; Pawson, A. J.; Lukito, V.; Mpamhanga, C. P.; Bombail, V.; Davenport, A. P.; Peters, J. A.; Spedding, M.; Harmar, A. J.; Nc, I.; Alexander, S. P. H.; Bonner, T. I.; Catterall, W. A.; Christopoulos, A.; Davenport, A. P.; Dollery, C. T.; Enna, S.; Harmar, A. J.; Kaibuchi, K.; Kanai, Y.; Laudet, V.; Neubig, R. R.; Ohlstein, E. H.; Peters, J. A.; Pin, J. P.; Ruegg, U.; Spedding, M.; Wright, M. W.; du Souich, P. *Nucleic Acids Res.* **2013**, *41*, D1083.
27. Keast, R. S. J.; Roper, J. *Chem. Senses* **2007**, *32*, 245.
28. Ganter, B.; Snyder, R. D.; Halbert, D. N.; Lee, M. D. *Pharmacogenomics* **2006**, *7*, 1025.
29. Tarcsay, A.; Keseru, G. M. *J. Med. Chem.* **2013**, *56*, 1789.
30. Peters, J.-U.; Hert, J.; Bissantz, C.; Hillebrecht, A.; Gerebtzoff, G.; Bendels, S.; Tillier, F.; Migeon, J.; Fischer, H.; Guba, W.; Kansy, M. *Drug Discov. Today* **2012**, *17*, 325.
31. Ritchie, T. J.; Macdonald, S. J. F.; Young, R. J.; Pickett, S. D. *Drug Discov. Today* **2011**, *16*, 164.
32. Leeson, P. D.; Springthorpe, B. *Nat. Rev. Drug Discov.* **2007**, *6*, 881.
33. Lovering, F. *Medchemcomm* **2013**, *4*, 515.
34. Hall, L. H.; Kier, L. B. *J. Chem. Inf. Model.* **1995**, *35*, 1039.
35. Ji, M.; Su, X.; Su, X.; Chen, Y.; Huang, W.; Zhang, J.; Gao, Z.; Li, C.; Lu, X. *Chem. Biol. Drug Des.* **2014**, *84*, 63.
36. Yarnitzky, T.; Levit, A.; Niv, M. Y. *Curr. Opin. Drug Discov. Dev.* **2010**, *13*, 317.
37. Eick, G. N.; Colucci, J. K.; Harms, M. J.; Ortlund, E. A.; Thornton, J. W. *PLoS Genet.* **2012**, *8*, e1003072.
38. Brill, S.; Sade-Falk, O.; Elbaz-Alon, Y.; Schuldiner, S. *J. Mol. Biol.* **2015**, *427*, 468.
39. Pandya, C.; Farelli, J. D.; Dunaway-Mariano, D.; Allen, K. N. *J. Biol. Chem.* **2014**, *289*, 30229.
40. Khersonsky, O.; Roodveldt, C.; Tawfik, D. S. *Curr. Opin. Chem. Biol.* **2006**, *10*, 498.
41. Khersonsky, O.; Tawfik, D. S. *Annu. Rev. Biochem.* **2010**, *79*, 471.
42. Behrens, M.; Gunn, H. C.; Ramos, P. C. M.; Meyerhof, W.; Wooding, S. P. *Chem. Senses* **2013**, *38*, 475.
43. Godfrey, S.; Balfour-Lynn, L.; Konig, P. *J. Pediatr.* **1975**, *87*, 465.
44. Liggett, S. B. *Expert Opin. Ther. Targets* **2013**, *17*, 721.
45. Mfuna Endam, L.; Filali-Mouhim, A.; Boisvert, P.; Boulet, L.-P.; Bosse, Y.; Desrosiers, M. *Int. Forum Allergy Rhinol.* **2014**, *4*, 200.

46. Brockhoff, A.; Behrens, M.; Roudnitzky, N.; Appendino, G.; Avonto, C.; Meyerhof, W. *J. Neurosci.: Official J. Soc. Neurosci.* **2011**, *31*, 14775.
47. Thalmann, S.; Behrens, M.; Meyerhof, W. *Biochem. Biophys. Res. Commun.* **2013**, *435*, 267.
48. Duan, J.; Dixon, S. L.; Lowrie, J. F.; Sherman, W. *J. Mol. Graph. Model.* **2010**, *29*, 157.
49. Akaike, H. *IEEE Trans. Autom. Control* **1974**, *19*, 716.
50. Sugiura, N. *Commun. Stat.—Theory Methods* **1978**, *7*, 13.
51. Rasmussen, S. G.; DeVree, B. T.; Zou, Y.; Kruse, A. C.; Chung, K. Y.; Kobilka, T. S.; Thian, F. S.; Chae, P. S.; Pardon, E.; Calinski, D.; Mathiesen, J. M.; Shah, S. T.; Lyons, J. A.; Caffrey, M.; Gellman, S. H.; Steyaert, J.; Skiniotis, G.; Weis, W. I.; Sunahara, R. K.; Kobilka, B. K. *Nature* **2011**, *477*, 549.
52. Kruse, A. C.; Ring, A. M.; Manglik, A.; Hu, J.; Hu, K.; Eitel, K.; Hubner, H.; Pardon, E.; Valant, C.; Sexton, P. M.; Christopoulos, A.; Felder, C. C.; Gmeiner, P.; Steyaert, J.; Weis, W. I.; Garcia, K. C.; Wess, J.; Kobilka, B. K. *Nature* **2013**, *504*, 101.
53. Pei, J.; Kim, B.-H.; Tang, M.; Grishin, N. V. *Nucleic Acids Res.* **2007**, *35*, W649.
54. Born, S.; Levit, A.; Niv, M. Y.; Meyerhof, W.; Behrens, M. *J. Neurosci.* **2013**, *33*, 201.
55. Halgren, T. *Chem. Biol. Drug Des.* **2007**, *69*, 146.
56. Bemis, G. W.; Murcko, M. A. *J. Med. Chem.* **1996**, *39*, 2887.