SuperSweet—a resource on natural and artificial sweetening agents

Jessica Ahmed, Saskia Preissner, Mathias Dunkel, Catherine L. Worth, Andreas Eckert and Robert Preissner*

Charité – Universitätsmedizin Berlin, Institute of Physiology, Structural Bioinformatics Group, Lindenberger Weg 80, 13125 Berlin, Germany

Received August 15, 2010; Revised September 21, 2010; Accepted September 24, 2010

ABSTRACT
A vast number of sweet tasting molecules are known, encompassing small compounds, carbohydrates, D-amino acids and large proteins. Carbohydrates play a particularly big role in human diet. The replacement of sugars in food with artificial sweeteners is common and is a general approach to prevent cavities, obesity and associated diseases such as diabetes and hyperlipidemia. Knowledge about the molecular basis of taste may reveal new strategies to overcome diet-induced diseases. In this context, the design of safe, low-calorie sweeteners is particularly important. Here, we provide a comprehensive collection of carbohydrates, artificial sweeteners and other sweet tasting agents like proteins and peptides. Additionally, structural information and properties such as number of calories, therapeutic annotations and a sweetness-index are stored in SuperSweet. Currently, the database consists of more than 8000 sweet molecules. Moreover, the database provides a modeled 3D structure of the sweet taste receptor and binding poses of the small sweet molecules. These binding poses provide hints for the design of new sweeteners. A user-friendly graphical interface allows similarity searching, visualization of docked sweeteners into the receptor etc. A sweetener classification tree and browsing features allow quick requests to be made to the database. The database is freely available at: http://bioinformatics.charite.de/sweet/.

INTRODUCTION
There are three major compounds of life: proteins; lipids and carbohydrates. The perception of sweet taste, mainly associated with advantageous food, has had an important evolutionary influence on different physiological regulation mechanisms. During human development, sugar was always luxury. In 1885 Constantin Fahlberg produced the first artificial sweetener, saccharin, and the scientific establishment was surprised by its extreme sweetness (1). Significant to this discovery was the fact that sweet taste became affordable to poor people. Following the commercial success of artificial sweeteners, a battle between the sugar and sweetener industries began (2). Saccharin was claimed to be carcinogenic in rats (3). However, it was later shown that saccharin is neither toxic nor carcinogenic in normal amounts (4), yet its reputation remains tarnished. Today, the replacement of sugar and other carbohydrates with artificial sweeteners in food is common (5) and is a general approach to prevent cavities (6,7), obesity and associated diseases such as diabetes and hyperlipidemia (8,9).

Currently, the sweet taste receptor, which is a heterodimer of two transmembrane proteins (T1R2 and T1R3) and has several different binding sites, has not been crystallized and is therefore unavailable in the Protein Data Bank (PDB) (10). Such a structure is crucial to elucidating how both small sweeteners and molecules as large as proteins bind and activate the sweet taste receptor (11). In the meantime, modeling studies can provide vital clues to these mechanisms (12). The understanding of compounds binding to the receptor is of relevance not only for the development of new artificial sweeteners but also for improving our understanding of known sweet molecules and what makes them ‘sweet’.

The first publicly available carbohydrate database was CarbBank (13), where users are able to search for carbohydrate structures, sub-structures and non-carbohydrate substituents. Wilhelm von der Lieth established the SweetDB (14), a web-based interface for glycoscientists, which was the basis for further carbohydrate tools collected in the Glycosciences portal (15) and the Glycome-DB (16) that comprises 35,000 carbohydrate sequences with a variety of query options.

*To whom correspondence should be addressed. Tel: +49 30 450 540 755; Fax: +49 30 450 540 955; Email: robert.preissner@charite.de


© The Author(s) 2010. Published by Oxford University Press. This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (http://creativecommons.org/licenses/by-nc/2.5), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.
There are also a number of databases that deal with glycans. GlycoBase (17) and GlycoExtractor (18) are databases that assist with interpreting high-performance liquid chromatography-glycan profiles. Tyrian Diagnostics used text-mining to develop the GlycoSuiteDB (19), which stores over 7650 glycan structures extracted from 740 papers. The Glycoconjugate Data Bank (20) provides a special tool for N-glycan primary structure verification. The connection to metabolic pathways is provided by KEGG-Glycan (21).

Although, there are a number of resources available with relation to carbohydrates, they are lacking with respect to sweetness and sweeteners. SuperSweet aims to integrate knowledge about the structure of sugars and sweetening agents with receptor binding poses, chemical properties and additional information like sweetness, approval, origin, therapeutic effect and metabolism.

THE DATABASE

The SuperSweet database was developed for researchers and dieticians and offers a user-friendly interface with helpful examples and FAQs. Currently, SuperSweet comprises more than 8000 carbohydrates, proteins, D-amino acids and artificial (synthesized) sweeteners, which were retrieved from the literature and different pre-existing data sources like Pubchem (22) and the PDB. Similarity searches extended and completed the sweetening agent data set. Besides information about the physicochemical properties of the sweet compounds, the database also offers information about the number of calories, the 3D structure, therapeutic annotations and, if detectable, the sweetness of the molecule. Structural information is available and displayed for each sweet molecule and sweet protein in the database. Moreover, the domain containing the small molecule active site of the sweet receptor was homology modeled and provided in SuperSweet (Figure 1). The small molecules were docked into the modeled binding site and the poses are also stored in the database.

There are different options for browsing through the database and for retrieving the data. First, the data can be retrieved by name, physicochemical properties or properties such as calories and sweetness. Secondly, the user is allowed to upload or draw a molecule using the Marvin Sketch plugin (http://www.chemaxon.com). The query structure is compared with the entries of SuperSweet and the results presented in a table comprising molecules and a Tanimoto coefficient expressing their similarity. Thirdly, a sweet-tree is available in SuperSweet that allows the fast and easy selection of a group of sweet tasting molecules (Table 1). Here, the user can find for example, all sweet tasting proteins, or peptides or small molecules like Flavonoids. Finally, SuperSweet offers a browse section, which provides an easy way to access the SuperSweet entries by choosing different categories of molecules based on properties.

Figure 1. Homology model of the sweet taste receptor with the sweetening agent Stevioside docked. The T1R2 protomer is displayed in cartoon format and the T1R3 protomer is displayed in wireframe format with a solvent accessible surface rendered (1.2Å probe radius) in transparent yellow. The stevioside molecule is displayed in spacefill format and colored according to atom type. Stevioside was docked into the open protomer, T1R3, as the closed protomer, T1R2, is too small to host large sweeteners. The pocket in the closed protomer is situated on the opposite side to that with stevioside bound—the approximate position is indicated by a black star. The cysteine-rich and transmembrane domains are shown schematically and are not to scale.

MATERIALS AND METHODS

Data acquisition

The sweet tasting molecules were extracted from the literature and publicly available databases like Pubchem, the PDB and MonoSaccharideDB and were filtered using different terms like ‘sweetening agents’. In the next step the data set was extended by using similarity search methods.

Homology modeling of the sweet taste receptor

A model of the large extracellular domain of the sweet taste receptor, which contains two binding sites for small molecular-weight sweeteners, was built using homology modeling. The sweet taste receptor is a class C (or metabotropic) G-protein coupled receptor and exists as a heterodimer (consisting of T1R2 and T1R3) (Figure 1). PSI-BLAST searches (23) of the PDB revealed that T1R2 and T1R3 share ~25% sequence identity with the metabotropic glutamate receptors and are in close proximity to one another on the phylogentic tree of class C.
GPCRs (24). As we wanted to use the homology model of the sweet taste receptor for docking studies, it was important that we chose a template structure that is in an active (open-closed) conformation, preferably with a natural ligand bound. Accordingly, an active form of metabotropic glutamate receptor 1 (mGluR1) was used as the template (PDB code: 1EWK) (25), which also had the highest sequence coverage and the highest resolution (2.2Å) compared to other crystal structures of activated glutamate receptors. A multiple sequence alignment was created using MUSCLE (26). The alignment can be downloaded from the SuperSweet website. Homology modeling was carried out using Modeller (27) in Accelrys Discovery Studio 2.5. T1R2 was built using chain A of 1EWK (closed form) and T1R3 using chain B (open form) (12). The large insertions in T1R2 and T1R3 compared to the template

---

**Table 1. Organization of the Sweet-tree**

<table>
<thead>
<tr>
<th>Main class</th>
<th>Subclass</th>
<th>Examples</th>
</tr>
</thead>
<tbody>
<tr>
<td>Carbohydrates</td>
<td>1) Monosacharides,</td>
<td><img src="image1" alt="Image" /> 1) Dextrose</td>
</tr>
<tr>
<td></td>
<td>2) Disaccharides,</td>
<td><img src="image2" alt="Image" /> 2) Lactose</td>
</tr>
<tr>
<td></td>
<td>3) Polysaccharide,</td>
<td><img src="image3" alt="Image" /> 3) Inulin</td>
</tr>
<tr>
<td></td>
<td>4) Sugar alcohols</td>
<td></td>
</tr>
<tr>
<td>Peptides</td>
<td>1) Amino acids,</td>
<td><img src="image4" alt="Image" /> 1) D-Tryptophane</td>
</tr>
<tr>
<td></td>
<td>2) Proteins</td>
<td><img src="image5" alt="Image" /> 2) Curculin</td>
</tr>
<tr>
<td>Small molecules</td>
<td>1) Acesulfames,</td>
<td><img src="image6" alt="Image" /> 1) Acesulfam-K</td>
</tr>
<tr>
<td></td>
<td>2) Alditols,</td>
<td><img src="image7" alt="Image" /> 2) Sorbitol</td>
</tr>
<tr>
<td></td>
<td>3) Alitame/aclames,</td>
<td><img src="image8" alt="Image" /> 3) Alitame</td>
</tr>
<tr>
<td></td>
<td>4) Aspartames,</td>
<td><img src="image9" alt="Image" /> 4) Aspartame</td>
</tr>
<tr>
<td></td>
<td>5) Cumarins,</td>
<td><img src="image10" alt="Image" /> 5) Cumarin</td>
</tr>
<tr>
<td></td>
<td>6) Cyclamat-like</td>
<td><img src="image11" alt="Image" /> 6) Cyclamic acid</td>
</tr>
<tr>
<td></td>
<td>7) Dulcines,</td>
<td><img src="image12" alt="Image" /> 7) Dulcin</td>
</tr>
<tr>
<td></td>
<td>8) Flavonoids,</td>
<td><img src="image13" alt="Image" /> 8) Neohesperidin-dihydrochalcone</td>
</tr>
<tr>
<td></td>
<td>9) Guanidine</td>
<td><img src="image14" alt="Image" /> 9) Guanidine</td>
</tr>
<tr>
<td></td>
<td>10) Carbonate,</td>
<td>10) Carbonate</td>
</tr>
<tr>
<td></td>
<td>11) Nucleotides,</td>
<td>11) Nucleotides</td>
</tr>
<tr>
<td></td>
<td>12) Saccharines,</td>
<td>12) Saccharin</td>
</tr>
<tr>
<td></td>
<td>13) Saponins,</td>
<td>13) Saponins</td>
</tr>
<tr>
<td></td>
<td>14) Steviol</td>
<td>14) Steviol</td>
</tr>
<tr>
<td></td>
<td>15) Glycosides,</td>
<td>15) Glycosides</td>
</tr>
<tr>
<td></td>
<td>16) Terpens,</td>
<td>16) Terpens</td>
</tr>
<tr>
<td></td>
<td>17) Others</td>
<td>17) Others</td>
</tr>
</tbody>
</table>

The first column presents the three main classes: carbohydrates, peptides and small molecules. The second column shows the subclasses of the main class. The third column shows some examples of the subclasses.

---

were removed from the final model. Lastly, side-chain clashes were removed and the structure was minimized by carrying out 100 steps of both steepest descent and conjugant gradient minimization.

**Generation of the binding poses of the small molecules**

Docking of the small compounds into the homology modeled receptor was done using the docking program GOLD 4.1.1 (28). In order to define the binding site of the sweet taste receptor, the template structure (mGluR1 containing a glutamate bound to each chain) was superimposed onto the homology model of the sweet taste receptor and the glutamate molecules copied over to the homology model. The binding sites of the sweet taste receptor were then defined by using the glutamate molecules as reference ligands; all atoms within 5Å of the glutamate molecule formed the binding sites for the docking experiments. For each small molecule, 100 docking runs were performed. A previous docking study showed that the sweet taste receptor’s active site in the closed protomer is too small to host some of the larger synthetic sweeteners and is only able to host four compounds out of those tested: saccharin, alitame, aspartame and 6-Cl-tryptophan (12). Experimental work has shown that aspartame and neotame bind to the T1R2 subunit (29). In accordance with these findings, we therefore docked molecules with a molecular weight >400 kDa into T1R3 (open form) and all other molecules into the pockets of both T1R2 and T1R3. The resulting docking poses were then ranked using the GoldScore fitness function. The best scoring docking pose for each molecule can be viewed using a Jmol applet and the respective structure files are also available for download.

**Conformer generation**

For the small sweetening molecules, conformers were generated using the Accelrys tool. For each small molecule 20 conformers are stored and available for download on the website (30).

**Similarity search**

For the similarity search in the SuperSweet database, we implemented a bit vector ‘structural fingerprint’, which encodes the chemical and topological characteristics of a molecule. The fingerprint was pre-calculated for the small molecules of SuperSweet and is also calculated for the query structure, in order to compare it to the database entries. Open Babel implements four different fingerprints (FP2, FP3, FP4 and MACCS). Fingerprint 2 (FP2) is widely used for the comparison of small molecules and is path-based and indexes linear molecules up to seven atoms. However, this fingerprint is not ideal for use in SuperSweet due to its inability to distinguish between different ring structures and therefore between carbohydrates. To overcome this problem, we implemented a combinatorial fingerprint of fingerprint 2 and fingerprint 4. Fingerprint 4 is based on a set of SMARTS patterns and also considers functional groups. For similarity searching the Tanimoto coefficient is used, which gives values in the range of zero (fingertips have no bits in common) to identical (all bits the same).

**Server**

SuperSweet is designed as a relational database on a MySQL server. Additionally, the MyChem package (http://mychem.sourceforge.net/) is installed to provide a complete set of functions for handling chemical data within MySQL. Most of the functions used by MyChem depend upon Open Babel (http://openbabel.sourceforge.net/). The structural fingerprint is implemented in Open Babel. To allow the upload or drawing of a query structure, the Marvin Sketch plugin (http://www.chemaxon.com) was installed. For the visualization of the 3D structures Jmol (http://www.jmol.org/) was installed. The website is built with PHP and web access is enabled via Apache HTTP Server 2.2.

**EXAMPLE OF USE**

Searching for a natural sweetening agent (search field ‘Origin’) with molecular weight between 800 and 900 and sweetness above 200 returns Stevioside. Steviol glycosides like Rebauudioside A or Rubusoside are non-caloric sweeteners that are found, for instance, in sweet Chinese tea (Rubus suavissimus) and Stevia rebaudiana (31). These compounds are of research interest because advantageous effects were observed regarding cancer and blood pressure (32). These effects seem to be the result of binding to other membrane proteins (33). Clicking on the protein icon in the results table shows the docking pose of Stevioside to the sweet receptor (see Figure 1). The structure of Stevioside, the computed conformers, the best docking pose and the modeled receptor structure are downloadable from the website.

More information on Steviol glycosides can be found using the ‘Sweet-tree’ or by performing a search using the field ‘Compound name’ on the ‘Property search’ page. Clicking on the similarity search icon delivers the top 10 similar compounds.

**CONCLUSION AND FURTHER DIRECTIONS**

SuperSweet compiles information on natural and artificial sweetening agents including their properties such as 3D structure, origin, sweetness, approval, calories etc. and provides hypotheses on their binding to the receptor.

Homology modeling provides a useful means of generating 3D conformations of proteins where experimental structures are not available. For this work, we generated models of the sweet taste receptor using mGluR1. The sequence identity between the receptors is rather low and is within the twilight zone of protein sequence alignments, which makes homology modeling more difficult (34). The quality of our homology model may also be affected by the fact that mGluR1 exists as a homodimer, whereas the sweet taste receptor is a heterodimer. These facts have implications for our docking experiment results due to the strong dependence
of docking results on the accuracy of protein structure, especially in the binding site (35). Although GOLD has consistently been shown to be among the best performing docking algorithms in terms of the accuracy of docking poses, it is less able to distinguish the most native-like pose from all of the generated docking poses (36–38). In SuperSweet we have only made the highest scoring pose available for each docked compound and therefore the accuracy of these docking solutions should be considered in light of the aforementioned limitations in homology models and in silico docking.

Unlike the metabotropic glutamate receptors, the sweet taste receptor is predicted to have multiple binding sites (29): (i) two cavities which correspond to the Glu hosting cavities of mGluR1; (ii) a secondary binding site on the surface of the receptor for sweet proteins that corresponds to the wedge model (39) and (iii) overlapping binding sites in the seven transmembrane helix domain for the agonist cyclamate (40) and the inverse agonist lactisole (41). For this work, we only performed docking experiments to the binding sites corresponding to the Glu hosting cavities of mGluR1. These two sites differ in size due to the receptor model being in an open-closed conformation; the binding site in T1R3 is much larger as it is in an open conformation, whereas T1R2 is in a closed conformation. Therefore, we only docked compounds with a molecular weight > 400 kDa into T1R3 (open form) and all other compounds into the pockets of both T1R2 and T1R3. In addition, the existence of other binding sites or alternative binding mechanisms cannot be excluded (12). Compared to mGluR1, the additional diversity of compounds binding to the sweet taste receptor and the existence of additional binding sites therefore adds further complexity to in silico docking experiments to the sweet taste receptor.

One of the future goals of SuperSweet is the integration of sugars and sweetening agents into biochemical pathway maps (including PubMed references) to better understand their different ways of metabolism and their impact on metabolic diseases and to foresee possible risk factors. After improving the similarity search and inclusion of pharmacophore searching to find new putative sweetening agents, a sweetness prediction tool is planned to be implemented. We plan to perform additional text-mining in order to obtain information on the number of calories, sweetness and therapeutic effects of sweet compounds where missing. Docking poses of the sweet proteins to the sweet taste receptor are also planned to be integrated. Another interesting aspect would be the comparison of sweet taste perception with characteristics of sour and bitter taste perception, which is a problem in the development of artificial sweeteners.

AVAILABILITY

The SuperSweet database is freely available under the url: http://bioinformatics.charite.de/sweet/ and will be updated regularly.

ACKNOWLEDGEMENTS

The authors thank D. Kuzman and A. Chefai for manual curation of the sweetener properties and A.L. Wölke for help with the ‘Sweet tree’.

FUNDING

Deutsche Forschungsgemeinschaft (SFB 449): Investitionsbank Berlin (IBB); Deutsche Krebshilfe; Bundesministerium für Bildung und Forschung (BMBF); European Union (EU). Funding for open access charge: DFG, BMBF and EU.

Conflict of interest statement. None declared.

REFERENCES


