

SuperNatural 3.0—a database of natural products and natural product-based derivatives

Kathleen Gallo, Emanuel Kemmler, Andrian Goede¹, Finnja Becker, Mathias Dunkel, Robert Preissner and Priyanka Banerjee^{1*}

Institute of Physiology and Science-IT, Charite - Universitätsmedizin Berlin, corporate member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Philippstrasse 12, 10115 Berlin, Germany

Received September 15, 2022; Revised October 07, 2022; Editorial Decision October 15, 2022; Accepted November 02, 2022

ABSTRACT

Natural products (NPs) are single chemical compounds, substances or mixtures produced by a living organism - found in nature. Evolutionarily, NPs have been used as healing agents since thousands of years and still today continue to be the most important source of new potential therapeutic preparations. Natural products have played a key role in modern drug discovery for several diseases. Furthermore, following consumers' increasing demand for natural food ingredients, many efforts have been made to discover natural low-calorie sweeteners in recent years. SuperNatural 3.0 is a freely available database of natural products and derivatives. The updated version contains 449 058 natural compounds along with their structural and physicochemical information. Additionally, information on pathways, mechanism of action, toxicity, vendor information if available, drug-like chemical space prediction for several diseases as antiviral, antibacterial, antimalarial, anticancer, and target specific cells like the central nervous system (CNS) are also provided for the natural compounds. The updated version of the database also provides a valuable pool of natural compounds in which potential highly sweet compounds are expected to be found. The possible taste profile of the natural compounds was predicted using our published VirtualTaste models. The SuperNatural 3.0 database is freely available via http://bioinf-applied.charite.de/supernatural_3, without any login or registration.

INTRODUCTION

Historically, natural products (NP) are known to be a rich source of compounds for drug discovery (1). Natural prod-

ucts have been acting as a source of therapeutic agents for many years and have shown beneficial effects. Well-known examples include the anticancer agent paclitaxel (sold as Taxol), originally extracted from the Pacific yew tree (2); the heart medicine digoxin (3), extracted from the foxglove plant; and aspirin, derived from a precursor found in the leaves of the willow tree, and used for its health benefits for thousands of years (4). Typically, such natural compounds have better bioavailability than synthetic compounds. Natural products are important in the cosmetics, food, and nutrition industries because of their many beneficial properties and positive commercial aspects (4). A broad range of natural products is used in cosmetics preparations, skin-care such as treatment of dryness, treatment of eczema and acne, as well as antioxidant, anti-inflammatory, anti-ageing, hair care products such as hair growth promoters, hair colour, scalp complaints like dandruff, and skin protection, and toiletry preparations (5). As the incidence of overweight and diabetic people has increased worldwide, there has been a great demand for new alternative low or no-calorie sweeteners for dietetic and diabetic purposes. Consequently, the search for sugar substitutes of natural sources has led to the discovery of a number of substances that possess an intensely sweet taste or taste-modifying properties. Over a hundred plant materials have been found to taste sweet because they contain large amounts of sugars and/or polyols or other sweet constituents (6). Drug discovery using natural products is a challenging task for designing new leads. Research in drug discovery needs to develop robust and viable lead molecules, which step forward from a screening hit to a drug candidate through structural elucidation and structure identification through gas chromatography–mass spectrometry (GC–MS), nuclear magnetic resonance (NMR) and high-performance liquid chromatography (HPLC). The development of new computational tools and databases has revolutionised the discovery and screening of natural products and their activity (1). The SuperNatural database, an open-accessed database was first published in 2006 (7), and the second update

*To whom correspondence should be addressed. Tel: +49 30 450 528 505; Fax: +49 30 450 540 955; Email: priyanka.banerjee@charite.de
Present address: Priyanka Banerjee, Institute of Physiology, Charité-University Medicine Berlin, Philippstrasse 12, 10115, Berlin, Germany.

(SuperNatural II) was published in 2015 (8). The SuperNatural database has successfully supported researchers from the wide scientific research community and allowed performing research in screening new molecules and activities to establish natural products as a significant source of drug discovery. It has been successfully supported the identification of lead structures (9,10). Here, we present the updated version of the SuperNatural database - SuperNatural 3.0. The SuperNatural 3.0 database of freely available natural compounds aggregates information from several sources and published literature. The database also includes information on reliable chemical suppliers who offer natural products. We have collected information on over 449 058 unique compounds and made them available as downloadable files. The database has been thoroughly curated using several curation criteria and a confidence score has been assigned to the compounds based on the annotation levels. All natural compounds having information on taxonomy or vendors information and having linked to at least three freely available NP database has been assigned with a confidence score of 1. The confidence score of 0.5 is assigned to compounds with no taxonomy information but linked with at least one NP database besides SuperNatural database. The database also contains information on physicochemical properties, toxicity class, mechanism of action (MoA), therapeutic pathways, focussed-targeted library, taste-related information and disease indications. Furthermore, we have also analysed and screened potential inhibitors from the natural product chemical space to support the COVID-19 drug discovery. We hope the SuperNatural 3.0 will facilitate the unmet need for utilisation of natural products for the benefit of humankind and the development of new lead for drug discovery.

MATERIALS AND METHODS

Software implementation

The data is stored in a relational MySQL database, which is hosted on the Charité IT system. For handling chemical information in the database, the Python package RDKit (<http://www.rdkit.org/>) and ChemAxon (<https://chemaxon.com>) software were used.

The website back-end consists of a lab-based LAMP (Linux/Apache/MySQL/PHP) server, with PHP serving as the back-end language. The database connection is established through the MySQL interface and front-end data delivery through a mixture of Html from submission responses and AJAX requests.

Website functionalities are implemented using Javascript and, in extension, its plugin jQuery (<https://jquery.com/>). Additionally, the CSS-Framework Bootstrap 4 (<https://getbootstrap.com/>) is used. Tables on the website were created with the jQuery plugin DataTables (<https://datatables.net>) and the absolute sorting extension (<https://datatables.net/plug-ins/sorting/absolute>). For the chemistry interface, the JavaScript library ChemDoodle Web components (<https://web.chemdoodle.com/>) was used. The usage of a JavaScript-capable browser is essential, and the server was tested on the most recent version of Google Chrome and Mozilla Firefox.

Database functionalities

Search strategy. There are four different options when searching for natural compounds as shown in Figure 1:

- **Search by name/IDs:** The name/id search requires either the compound name, a Supernatural ID or a natural product supplier ID. When using the name search option, all relevant entries will be displayed in a selectable dropdown menu.
- **Search by properties:** The property search function provides various filters, which can be applied to identify a compound of interest. These include ranges for the molecular weight, the topological polar surface area (TPSA), logP, hydrogen bond acceptor (HBA), hydrogen bond donor (HBD), types of bonds (rotatable bonds, amide bonds), types of atoms (hetero atoms, heavy atoms), stereocenters, and different ring types (aromatic, saturated and aliphatic). Additionally, properties such as oral toxicity value (LD50)-based toxicity classes are also reported for the natural compounds. The toxicity class of each compound is predicted using the ProTox-II method (11).
- **By similarity:** For the search for similar compounds in the database, there are four ways for users to submit a small molecule structure, the simplest being to enter the PubChem (12) name. Additionally, a SMILES (Simplified Molecular-Input Line-Entry System) string of the compound can be entered, or a standard molecule file uploaded. Once a molecular structure is uploaded, or a name/SMILES string entered, the corresponding molecular structure will be displayed in the integrated ChemDoodle structure viewer (13), where additional modifications such as deletion or addition of atoms or substructures can be done. Alternatively, it is possible to draw a molecular structure from scratch using the provided drawing tools. Subsequently, the finished structure is translated into an ECFP 4 molecular fingerprints (14) and compared to all-natural compounds in the database. The Tanimoto coefficient is used as measure of similarity, with a Tanimoto similarity (15) of one indicating that the natural compound from the database and the queried structure are identical.
- **By substructure:** The search by substructures works completely analogue, the only difference being that instead of a complete molecular structure, a substructure of interest is entered, and the database is queried for all-natural compounds that contain the specified substructure with a Tanimoto similarity of 1.0 (1 being the most similar, and 0 dissimilar).

Pathways. The evaluations for the pathways are based on the different types of data attributes extracted from the current version of the ChEMBL database (ChEMBL_29) (14); a table containing all human UNIPROT-IDs were mapped with to the ChEMBL datasets, and its relations to different databases like, e.g.KEGG and HGNC, taken from UNIPROT (16). Optimal binders were filtered according to Peon et al. (17), considering a combination of binding strength, IC₅₀/EC₅₀ values and confidence scores. Data containing information on pathways for human dis-

Please select a specific indication from the choices below. As a result, all SuperNatural compounds that score higher than 0.8 for the selected indication will be displayed.

Disease Indications

Indication: Multiple myeloma
Confidence limit: 0.80 1.0 e.g. 0.7-0.9
Find Reset

Associated human targets of CHEMBL55415:

| Target Uniprot ID | Activity in nm | Target Name | Target gene |
|-------------------|----------------|---|-------------|
| P43250 | 6 | G-protein-coupled receptor kinase 6 | GRKB_HUMAN |
| P51649 | 7 | Succinate-semialdehyde dehydrogenase, mitochondrial | SSDH_HUMAN |
| PODU86 | 9 | Alpha-amylase 1A | AMY1A_HUMAN |
| P14618 | 12 | Pyruvate kinase PKM | KPYM_HUMAN |
| Q9Y3R4 | 55 | Sialidase-2 | NEUR2_HUMAN |
| Q13332 | 600 | Receptor-type tyrosine-protein phosphatase S | PTPRS_HUMAN |

Top 100 active compounds against main COVID-19 protease

| SuperNatural ID | Name | Confidence |
|-----------------|--------------|------------|
| SN00013745 | Scutellarein | 0.9 |

Enriched human pathways of CHEMBL55415:

| Pathway ID | Pathway name | e-value | p-value |
|------------|---|----------|-----------|
| hsa00310 | Lysine degradation | 0.982279 | 0.0145565 |
| hsa04144 | Endocytosis | 0.999986 | 0.039779 |
| hsa05202 | Transcriptional misregulation in cancer | 1 | 0.0623158 |
| hsa04934 | Cushing syndrome | 1 | 0.071725 |
| hsa04062 | Chemokine signaling pathway | 1 | 0.0869136 |
| hsa01100 | Metabolic pathways | 1 | 0.391268 |

Showing 1 to 6 of 6 entries

For a natural compound For a molecular structure For a human target For a non-human target

Mechanism of Actions

Please fill only one field for searching!

Supernatural ID: SN0176761 e.g. SN0034589
Compound Name: scutellarein e.g. Aspirin
Find Reset

Please note: Calculations may take up to 10 seconds!

Organoleptic properties

Confidence: 7-0.9
Taste: Bitter Salty Sour Sweet Umami
Find R

By name/ids By properties By similarity By substructure

Compound Search

Please fill only one field for searching!

Supernatural ID: SN0176761 e.g. SN0034589
Compound Name: scutellarein
Scutellarein
Scutellarein 4'-methyl Ether
Supply-Code: e.g. NP-010664
Find Reset

For a natural compound For a molecular structure For a human pathway

Pathways

Please fill only one field for searching!

Supernatural ID: scutellarein e.g. SN0034589
Compound Name: Scutellarein
Scutellarein 4'-methyl Ether
Find Reset

Focused Targeted Libraries

Target library: Anti-Bacterial Anti-Cancer Anti-Malaria Central nervous system targets Viral targets
Tanimoto similarity:
Please note: Calc

SuperNatural 3.0 database

Figure 1. The overall design of the search function modules in SuperNatural 3.0 database. The database provides seven main function modules: (i) 'Compound Search' module; (ii) 'Mechanism of Actions' module; (iii) 'Pathways' module; (iv) 'Target Library' module; (v) 'Disease Indication' module; (vi) 'COVID-19 virtual screening' module; (vii) 'Organoleptic properties' module.

eases and infections and the selected pathways and human genes was derived from KEGG (18) (<https://www.genome.jp/kegg/>). Similarly, information about druggable genes from IDG was extracted. The table for the druggable genes was enriched by additional 43 genes that are targets of approved drugs which are themselves small molecules, taken from the Therapeutic Target database TTD (19). Using KEGG mapper a table was derived containing all information between KEGG pathways related to diseases and human genes. More information on the pathways mapping and evaluation of the confidence score can be found in the FAQ section of the webserver under the section 'Pathways'.

Mechanism of action. The prediction of the mechanism of action can be done starting either from a molecular structure, which is entered by the user or identified via name/ID, or a human protein target. It is based on the ChEMBL 29 database (20), which was filtered and standardised so that only highly accurate direct interactions between small molecules and human proteins are considered. For all remaining molecular structures, Morgan fingerprints of length 1024 were calculated using the Python library RDKit. When a search structure is entered, the corresponding fingerprint is calculated and compared to the ChEMBL structures. As a result, the five most similar small-molecule structures (according to Tanimoto coefficient) are chosen, and their interactions displayed, including UniProt (16) information. Similarly, entering a human target of interest will identify the ten most active ChEMBL compounds (defined in nanomolar) for the target and sub-

sequently display the five most similar natural compounds for each of the ten identified interactions.

Targeted libraries. SuperNatural database 3.0 provides focussed targeted libraries for several diseases like anti-cancer, anti-bacterial, antiviral, antimalaria and diseases focussed on the central nervous system (CNS). These are based on molecular fingerprint similarity (21) and matched molecular pair analysis (22) (MMPA) of the SuperNatural database 3.0. to that of the focussed libraries extracted from the Life Chemicals screening set (<https://lifechemicals.com/downloads>). Based on this analysis, the database currently contains 2137 compounds with 70% and above confidence for antibacterial activity. Similarly, 1473 compounds with anticancer activity, 2552 compounds with antiviral activity, 118 with antimalarial activity and 94 with CNS activity. The targeted classes and respective compounds can be searched using different thresholds as explained in the FAQ section of the website.

COVID-19 main protease inhibitors chemical space. The 2019 novel Coronavirus disease (COVID-19) is caused by the Severe Acute Respiratory Syndrome Coronavirus-2 (SARS-CoV-2) virus (23). The SARS-CoV-2 main protease (Mpro, also name 3CLpro) has been a major interest concerning the drug discovery for COVID-19 disease (24). Computationally evaluated main protease inhibitors from the SuperNatural 3.0 database were predicted using two different prediction methods:

- a) *Machine learning method*: With the ongoing SARS-CoV-2 pandemic there is an urgent need for the discovery of a treatment for the coronavirus disease (COVID-19). Machine learning models can assist drug discovery through the prediction of the best compounds based on previously published data. Herein, we have ENSEMBLE machine learning methods (11) to develop predictive models from recent SARS-CoV-2 *in vitro* inhibition data (20). The models were evaluated on several performance measures and have achieved above 80% accuracy, sensitivity, specificity and AUC-ROC both on 10-fold cross-validation and external validation sets (see Statistics section on the webserver). Around 12648 natural compounds were predicted to be active inhibitors of the main protease with a confidence score of 0.9 and above. The supernatural compounds along with predicted inhibitory strength against the main-protease can be searched under the tab 'COVID-19' using difference confidence intervals.
- b) *Molecular docking-based virtual screening*: The molecular docking protocol (9) was established using AutoDock, software (25). The crystal structure of the COVID-19 main protease in complex with an inhibitor N3, with the identification number 6lu7, was downloaded from the PDB protein data bank (<https://www.rcsb.org/>). Due to computational capacity limitations, the decision was made to reduce the search space, based on a similarity search of the known N3 inhibitor and the natural compounds from the SuperNatural 3.0 database. A KNIME (26) workflow was used to conduct the similarity search using circular fingerprints (Morgan fingerprints) and a threshold of above 0.75 Tanimoto Coefficient. The docking is evaluated based on the binding affinity (here equal to docking score) which is determined by the binding energy. The binding energy values of the 1078 docked compounds range between -10 kcal/mol and -3.7 kcal/mol. The distribution of the docking scores is visualized (see Statistics). The vast majority of compounds obtained binding energy between -8.5 kcal/mol and -7.5 kcal/mol. Considering that lower binding energy is generally associated with favourable docking, the threshold was set to smaller than -9.0 kcal/mol. The predicted SuperNatural compounds inhibitor space can be searched via the function 'COVID-19' on the website.

Organoleptic properties. Sweet and bitter taste properties are found in most classes of natural compounds and a close relationship is found in many structural categories. Furthermore, sourness is also one of the important organoleptic properties of natural compounds. The supernatural 3.0 database compounds were used as input structure for the taste prediction (sweet, bitter and sour) using the VirtualTaste web server (27) which is based on the machine learning methods described in the published paper (28). Almost, 170 265 compounds were predicted to be bitter, and 31 803 compounds were predicted to be sweet, with a confidence score of at least 0.7 (see Statistics tab on the webserver).

Additionally, new machine learning methods were developed to predict salty and umami tastes, using a logis-

tic regression model from the Python library scikit-learn (<https://scikit-learn.org/stable/>). A train/test split and additionally 10-fold cross-validation on the training set was performed to evaluate the models. Cross-validation for the salty (/umami) model achieved an average score of 92.2% (95.8%), and validation on the external test set was 95.6% (94.6%) correct, with an AUC score of 99.2% (97.6%). 1888 compounds were predicted to be salty and 384 785 were predicted as sour and 45 934 compounds to be umami with a confidence of at least 0.7.

The natural products compounds and taste properties can be searched under the tab 'Taste' using different tastes (sweet, bitter, sour, umami, salty) and confidence scores as search parameters.

Prediction of activity against specific disease indications

The prediction of the association of natural compounds with disease indications is based on information extracted from the Therapeutic Target Database (31). Specific indications were identified by their ICD 11 code and all associated compounds were used as training set for the development of machine learning models (32). Hereby, indications lacking sufficient information (<15 associated compounds) were excluded, leaving 80 ICD 11 categories for which a model was built. For each indication, a number of different machine learning models were tested and evaluated regarding their performance, including logistic regression, linear discriminant analysis, *k* nearest neighbors, decision tree, support vector machines, gaussian naïve bayes and random forests. Model accuracy was evaluated using 10-fold cross-validation and the best performing model was chosen for each indication. Thirty-two models achieved a performance better than 0.6 and were included in the database. On the webpage, searching for a specific indication displays a result table including both the performance of the specific machine learning model (evaluated via 10-fold cross-validation) and the individual score of the natural compounds achieved with this model.

Application case

As an example, natural compound Scutellarein is used to explain the functionalities of the database. Scutellarein which is extracted from the perennial herb *Scutellaria lateriflora* and is known to possess anticancer potential (29). The user can search the compound 'scutellarein' via name under compound search or via SuperNatural ID (SN0176761) or via structure search. The query result page will include information on the compounds, structure, toxicity class, physicochemical properties, and vendor information. Additionally, confidence score of 1 for the annotation level is also shown. Further information on source organisms and link to the taxonomy data is also provided in a table. Using the MOA function and the ID information on related target can be retrieved such as G protein-coupled receptor kinase 6, Succinate-semialdehyde dehydrogenase, mitochondrial, Alpha-amylase 1A. Similarly, using the pathways search, enriched human pathways information on the compounds can be achieved including the enrichment scores and e-values for the respective pathways. Via the COVID-19 tab, the compound shows the potential to be an inhibitor

of the main protease of the SARs-CoV2 with a confidence score of 0.9. Scutellarein has been also found as one of the polar compounds as the herbal remedies of coronavirus (30), MERS, or SARS (10). Furthermore, the taste of the compound which is bitter can be accessed using the ‘Taste’ tab of the webserver. The disease indication for scutellarein suggest active role in diseases like multiple myeloma with accuracy score of 80%. Studies has reported Scutellarein to be selectively targeting multiple myeloma cells and only cytotoxic to the malignant cells compared to healthy cells (33).

Downloads

The SuperNatural 3.0 database not only provides extended ‘Search’ options with different properties of natural compounds but also offers the possibility to download all searched results as tables and data sets in a user-friendly manner. Users can customize the compound download list through advanced search and manual selection. The download formats are available as pdf, CSV, excel, sdf.

To facilitate the natural product-based drug discovery pipeline and virtual screening protocol, the entire dataset is made available for bulk download as a CSV file via the following link (http://bioinf-applied.charite.de/supernatural_3/subpages/faq.php/#10).

Conclusion and future directions

Natural products and their analogues have been known to made major contributions in drug discovery especially for cancer and infectious diseases. Nevertheless, natural product-based drug discovery also presents challenges in terms of technical screening, characterisation and optimization. Improved genome analysis, bioengineering and computational analysis are addressing such challenges and opening up new opportunities and novel information. Here, we present the updated version of the SuperNatural database (3.0), which aims to support the virtual screening process of the NP-based drug discovery. Besides the drug discovery research, SuperNatural 3.0 also provides information on several aspects such as taste-related information for the food industry. The current version of the database contains around 449058 natural compounds along with their structural and physicochemical information. Furthermore, information on toxicity class, mechanism of actions and related pathways/molecular targets are also provided for the respective compounds. Addressing the COVID-19 pandemic, the database provides a COVID-19 drug discovery chemical space for the prediction and identification of natural compounds as potential inhibitors of the SARs-CoV-2 main protease. In future, we expect the SuperNatural database to grow continuously with extensive data deposition, curation and resource integration. We envisage that the SuperNatural database will be able to grow as a comprehensive natural product-based repository for the users, vendors and researchers and leading to the re-emergence of natural products for drug discovery.

After publication, the SuperNatural 3.0 database will provide a platform for the users and vendors to submit new data or report corrections on the existing dataset. Re-

searchers involved in natural-product based research are encouraged to submit new data (MOL or SDF format) and activities with publication reference.

DATA AVAILABILITY

Super Natural 3.0 database is freely available to all via http://bioinf-applied.charite.de/supernatural_3, without any login or registration. To facilitate the natural product-based drug discovery pipeline and virtual screening protocol, the entire dataset is made available for bulk download as a CSV file via the following link (http://bioinfapplied.charite.de/supernatural_3/subpages/faq.php/#10).

ACKNOWLEDGEMENTS

We thank the students of the Structural Bioinformatics Group at Charité for testing the SuperNatural 3.0 database functionalities.

FUNDING

Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) as part of the clinical research unit [CRU339]: Food allergy and tolerance (FOOD@) [428445448, 428447634]; German Research Foundation (DFG) and the Open Access Publication Fund of Charité – Universitätsmedizin Berlin; Deutsche Forschungsgemeinschaft/International [DFG TRR295]; BMBF funded by SimLeap project [13GW0226].

Conflict of interest statement. None declared.

REFERENCES

1. Harvey, A.L., Edrada-Ebel, R. and Quinn, R.J. (2015) The re-emergence of natural products for drug discovery in the genomics era. *Nat. Rev. Drug Discov.*, **14**, 111–129.
2. Expósito, O., Bonfill, M., Moyano, E., Onrubia, M., Mirjalili, M.H., Cusidó, R.M. and Palazón, J. (2009) Biotechnological production of taxol and related taxoids: current state and prospects. *Anticancer. Agents Med. Chem.*, **9**, 109–121.
3. Iyer, S.S., Gensollen, T., Gandhi, A., Oh, S.F., Neves, J.F., Collin, F., Lavin, R., Serra, C., Glickman, J., de Silva, P.S.A. *et al.* (2018) Dietary and microbial oxazoles induce intestinal inflammation by modulating aryl hydrocarbon receptor responses. *Cell*, **173**, 1123–1134.
4. Desborough, M.J.R. and Keeling, D.M. (2017) The aspirin story - from willow to wonder drug. *Br. J. Haematol.*, **177**, 674–683.
5. Kaliyadan, F., Al Dhafiri, M. and Aatif, M. (2021) Attitudes toward organic cosmetics: a cross-sectional population-based survey from the middle east. *J. Cosmet. Dermatol.*, **20**, 2552–2555.
6. Lê, K.-A., Robin, F. and Roger, O. (2016) Sugar replacers: from technological challenges to consequences on health. *Curr. Opin. Clin. Nutr. Metab. Care*, **19**, 310–315.
7. Dunkel, M., Fullbeck, M., Neumann, S. and Preissner, R. (2006) SuperNatural: a searchable database of available natural compounds. *Nucleic Acids Res.*, **34**, D678–D83.
8. Banerjee, P., Erehman, J., Gohlke, B.-O., Wilhelm, T., Preissner, R. and Dunkel, M. (2015) Super Natural II – a database of natural products. *Nucleic Acids Res.*, **43**, D935–D939.
9. Abel, R., Paredes Ramos, M., Chen, Q., Pérez-Sánchez, H., Coluzzi, F., Rocco, M., Marchetti, P., Mura, C., Simmaco, M., Bourne, P.E. *et al.* (2020) Computational prediction of potential inhibitors of the main protease of SARS-CoV-2. *Front. Chem.*, **8**, 590263.
10. Attia, Y.A., Alagawany, M.M., Farag, M.R., Alkhatib, F.M., Khafaga, A.F., Abdel-Moneim, A.-M.E., Asiry, K.A., Mesalam, N.M., Shafi, M.E., Al-Harathi, M.A. *et al.* (2020) Phytogetic products and phytochemicals as a candidate strategy to improve tolerance to coronavirus. *Front. Vet. Sci.*, **7**, 573159.

11. Banerjee,P., Eckert,A.O., Schrey,A.K. and Preissner,R. (2018) ProTox-II: a webserver for the prediction of toxicity of chemicals. *Nucleic Acids Res.*, **46**, W257–W263.
12. Kim,S., Chen,J., Cheng,T., Gindulyte,A., He,J., He,S., Li,Q., Shoemaker,B.A., Thiessen,P.A., Yu,B. *et al.* (2019) PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res.*, **47**, D1102–D1109.
13. Burger,M.C. (2015) ChemDoodle web components: HTML5 toolkit for chemical graphics, interfaces, and informatics. *J. Cheminform.*, **7**, 35.
14. Rogers,D. and Hahn,M. (2010) Extended-connectivity fingerprints. *J. Chem. Inf. Model.*, **50**, 742–754.
15. Bajusz,D., Rácz,A. and Héberger,K. (2015) Why is tanimoto index an appropriate choice for fingerprint-based similarity calculations?*J. Cheminform.*, **7**, 20.
16. UniProt Consortium,T. (2018) UniProt: the universal protein knowledgebase. *Nucleic Acids Res.*, **46**, 2699.
17. Peón,A., Naulaerts,S. and Ballester,P.J. (2017) Predicting the reliability of drug-target interaction predictions with maximum coverage of target space. *Sci. Rep.*, **7**, 3820.
18. Kanehisa,M., Furumichi,M., Sato,Y., Ishiguro-Watanabe,M. and Tanabe,M. (2021) KEGG: integrating viruses and cellular organisms. *Nucleic Acids Res.*, **49**, D545–D551.
19. Qin,C., Zhang,C., Zhu,F., Xu,F., Chen,S.Y., Zhang,P., Li,Y.H., Yang,S.Y., Wei,Y.Q., Tao,L. *et al.* (2014) Therapeutic target database update 2014: a resource for targeted therapeutics. *Nucleic Acids Res.*, **42**, D1118–D1123.
20. Gaulton,A., Hersey,A., Nowotka,M., Bento,A.P., Chambers,J., Mendez,D., Motowop,P., Atkinson,F., Bellis,L.J., Cibrián-Uhalte,E. *et al.* (2017) The ChEMBL database in 2017. *Nucleic Acids Res.*, **45**, D945–D954.
21. Cereto-Massagué,A., Ojeda,M.J., Valls,C., Mulero,M., Garcia-Vallvé,S. and Pujadas,G. (2015) Molecular fingerprint similarity search in virtual screening. *Methods*, **71**, 58–63.
22. Griffen,E., Leach,A.G., Robb,G.R. and Warner,D.J. (2011) Matched molecular pairs as a medicinal chemistry tool. *J. Med. Chem.*, **54**, 7739–7750.
23. V'kovski,P., Kratzel,A., Steiner,S., Stalder,H. and Thiel,V. (2021) Coronavirus biology and replication: implications for SARS-CoV-2. *Nat. Rev. Microbiol.*, **19**, 155–170.
24. Zhang,L., Lin,D., Sun,X., Curth,U., Drosten,C., Sauerhering,L., Becker,S., Rox,K. and Hilgenfeld,R. (2020) Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α -ketoamide inhibitors. *Science*, **368**, 409–412.
25. Trott,O. and Olson,A.J. (2010) AutoDock vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.*, **31**, 455–461.
26. Beisken,S., Meinel,T., Wiswedel,B., de Figueiredo,L.F., Berthold,M. and Steinbeck,C. (2013) KNIME-CDK: Workflow-driven cheminformatics. *BMC Bioinf.*, **14**, 257.
27. Fritz,F., Preissner,R. and Banerjee,P. (2021) VirtualTaste: a web server for the prediction of organoleptic properties of chemical compounds. *Nucleic Acids Res.*, **49**, W679–W684.
28. Banerjee,P. and Preissner,R. (2018) BitterSweetForest: a random forest based binary classifier to predict bitterness and sweetness of chemical compounds. *Front. Chem.*, **6**, 93.
29. Shi,X., Chen,G., Liu,X., Qiu,Y., Yang,S., Zhang,Y., Fang,X., Zhang,C. and Liu,X. (2015) Scutellarein inhibits cancer cell metastasis in vitro and attenuates the development of fibrosarcoma in vivo. *Int. J. Mol. Med.*, **35**, 31–38.
30. Liu,H., Ye,F., Sun,Q., Liang,H., Li,C., Li,S., Lu,R., Huang,B., Tan,W. and Lai,L. (2021) Scutellaria baicalensis extract and baicalin inhibit replication of SARS-CoV-2 and its 3C-like protease in vitro. *J. Enzyme Inhib. Med. Chem.*, **36**, 497–503.
31. Wang,Y., Zhang,S., Li,F., Zhou,Y., Zhang,Y., Wang,Z., Zhang,R., Zhu,J., Ren,Y., Tan,Y. *et al.* (2020) Therapeutic target database 2020: enriched resource for facilitating research and early development of targeted therapeutics. *Nucleic Acids Res.*, **48**, D1031–D1041.
32. Harrison,J.E., Weber,S., Jakob,R. and Chute,C.G. (2021) ICD-11: an international classification of diseases for the twenty-first century. *BMC Med. Inform. Decis. Mak.*, **21**, 206.
33. Shi,L., Wu,Y., Lv,D.L. and Feng,L. (2019) Scutellarein selectively targets multiple myeloma cells by increasing mitochondrial superoxide production and activating intrinsic apoptosis pathway. *Biomed. Pharmacother.*, **109**, 2109–2118.