mVOC: a database of microbial volatiles

Marie Chantal Lemfack[1], Janette Nickel[2,3], Mathias Dunkel[2], Robert Preissner[2] and Birgit Piechulla[1,*]

[1]University of Rostock, Institute of Biological Sciences, Rostock 18059, Germany, [2]Charité—University Medicine Berlin, Structural Bioinformatics Group, Institute of Physiology & Experimental Clinical Research Center, Berlin 13125, Germany and [3]Charité—University Medicine Berlin, Division of General Pediatrics, Department of Pediatric Oncology and Hematology, Berlin 13353, Germany

ABSTRACT

Scents are well known to be emitted from flowers and animals. In nature, these volatiles are responsible for inter- and intra-organismic communication, e.g. attraction and defence. Consequently, they influence and improve the establishment of organisms and populations in ecological niches by acting as single compounds or in mixtures. Despite the known wealth of volatile organic compounds (VOCs) from species of the plant and animal kingdom, in the past, less attention has been focused on volatiles of microorganisms. Although fast and affordable sequencing methods facilitate the detection of microbial diseases, however, the analysis of signature or fingerprint volatiles will be faster and easier. Microbial VOCs (mVOCs) are presently used as marker to detect human diseases, food spoilage or moulds in houses. Furthermore, mVOCs exhibited antagonistic potential against pathogens *in vitro*, but their biological roles in the ecosystems remain to be investigated. Information on volatile emission from bacteria and fungi is presently scattered in the literature, and no public and up-to-date collection on mVOCs is available. To address this need, we have developed mVOC, a database available online at http://bioinformatics. charite.de/mvoc.

INTRODUCTION

Microorganisms are universal in the biosphere. They are often found in large quantities and diverse compositions (microbiome). For example, there are more microorganisms ($\sim$2 kg) than human cells in humans, and most of them are essential and useful for the human host vitality (1). Bacteria are also dominant inhabitants of the leaf surfaces (10^7 cells/cm^2) and they are prominent in the soil, e.g. 1 g of soil contains $\sim$10^{11} microbial cells (2).

It is well known that microbes produce a diversity of natural compounds, e.g. antibiotics. Interestingly, the small molecular mass substances released by microorganisms were often overlooked, partially due to the lack of appropriate absorption and detection technologies. Many of these small molecules ($<$300 Da) exhibit high-vapour pressures and low boiling points, and, together with a lipophilic character, these features support volatility.

In the past decade, research on microbial smells experienced a renaissance owing to their global appearance. Some examples of bacterial volatile emissions are mentioned here. Undoubtedly, prominent malodorous volatiles are produced by microorganisms during the process of putrefaction (e.g. amines, sulphur compounds, indole and ammonia) (3), whereas the aromas of wines, sauerkraut, cheese and other milk product fermentations are usually recognized as pleasant by human noses (e.g. acids, alcohols and esters). The earthy and muddy smell of wet forest soils is due to the emission of the volatile geosmin released by *Streptomyces* species (4–6). Microbiologists typically recognize the characteristic smell of indole from *Escherichia coli*. The human microbial flora at any given anatomical site is relatively specifically accompanied by a typical volatile organic compound (VOC) profile (e.g. oral and breath malodour, smell of sputum VOCs, gases released by the gut, sweat and sebum smell and foot odour). The VOC mixture of breath originates from more than one source within the respiratory system (e.g. tongue, oropharynx and bronchioles), and respiratory disorders can result in odorous gases being expelled into the air, which can be useful for diagnostic purposes (3). For example, to detect *Mycobacterium tuberculosis*, methyl nicotinate showed promising results to be used as a non-invasive and rapid diagnostic tool, or the emission of 2-nonanone of Pseudomonas aeruginosa VOCs may be used as *in vivo* marker to detect lung infections (7). Freshly secreted sweat is sterile, but due to biotransformation by

*To whom correspondence should be addressed. Tel: +49 381 498 6130; Fax: +49 381 498 6132; Email: birgit.piechulla@uni-rostock.de

The authors wish it to be known that, in their opinion, the first two authors should be regarded as Joint First Authors.

microorganisms (aerobic coryneforms, propionibacteria and Micrococcaceae), odoriferous VOCs are produced. Another example is wound infection: when a wound is formed, it is a new ecological niche favourable to microbial growth. Research is ongoing to develop a 'wound sniffing' device that discriminates between wound-related and -unrelated volatiles. Volatiles, particularly off-flavours, are also fingerprints to screen systematically for spoiled foodstuff or to identify hidden microbial growth in buildings (8,9). Furthermore, in the past few years, the interest of researchers studying the effects of microbial VOCs (mVOCs) on plants has become increasingly evident, showing that some rhizobacteria release a blend of volatile components that promote growth of *Arabidopsis thaliana* (10), whereas others inhibit or are toxic for plants (11–14). It is also well documented that mVOCs are important factors in mediating specific microbial interactions; in fact, intra- or inter-species interactions between bacteria and fungi in the soil result in morphological and phenotypical alterations of the receiving organism (15). There are numerous instances of mVOCs being closely associated with insect feeding behaviours, and other microbial volatiles are also known as powerful repellants (16). Moreover, some volatile compounds such as higher alcohols (2-methyl-1-butanol, 3-methyl-1-butanol and isobutanol) can be used as biofuels (17); however, considering the fact that the natural microbial production rates are too low to support industrial production, metabolic engineering is widely used to improve the production (17,18, http://dx.doi.org/10.5772/52050).

These are just a few examples of the up to now known 349 bacteria and 69 fungi that are volatile emitters. Taking in considerations that to date ~10 000 microbial species are described and at least a million are expected to exist on earth, the VOC profiles of a surprisingly small number of microorganisms were investigated so far. The VOC spectra of microorganisms are species-specific and can be simple or complex (19). The qualitative and quantitative composition of an mVOC profile is variable, depending on growth conditions (temperature, oxygen availability, pH), the carbon source availability and the age of the culture (3,20–24). Ultimately, the volatile emission profile is a consequence of specific metabolic activities of the particular microorganism.

Considering the importance and the central roles of mVOCs in our biosphere, our objective was the establishment of a database of microbial volatiles for public use. Here, we present for the first time a user-friendly compilation of the microbial volatiles extracted from the literature. Originally, only mVOCs were filed using the Pubchem ID as an essential criterion. During the literature search, it turned out that many mVOCs have not received a Pubchem ID, but might be biologically relevant. Therefore, these compounds were also included into the database. Mixtures of mVOCs are composed of various chemical classes, e.g. low molecular weight fatty acids and their derivatives (hydrocarbons, alcohols, aldehydes and ketones), terpenoids, aromatic compounds, nitrogen containing compounds and volatile sulphur compounds (15,25). To date, ~1000 volatiles are filed in the mVOC database. References are given to each bacterial and

fungal strain or isolate presented in the database. The user interface offers several search options, for instance, by species name, Pubchem ID, structure, molecular weight and logP value. Online upload is possible, allowing a timely incorporation of new data sets, which is expected to happen progressively in this fast-growing research field.

MATERIALS AND METHODS

The data were acquired by an extensive literature search available on PubMed (http://www.ncbi.nlm.nih.gov/pubmed). Most of the information on mVOCs was found in ~20 journals and the full-text of ~100 articles was yielding most of the results compiled in the mVOC database. The literature was manually screened by biochemists. To update mVOC in the future, literature will continuously be screened and data will be checked by a chemist or biochemist before entering them manually to the database.

For conducting the similarity search of the compounds, the chemoinformatics package MyChem (http://mychem.sourceforge.net/) is integrated into the database. It enables the analysis and conversion of chemical data using OpenBabel (http://openbabel.org/) functionality. For the purpose of calculating the Tanimoto coefficient (26), it is obligatory to assign fingerprints to the compounds. This step is also performed by the MyChem package. Thereto, OpenBabel uses the Daylight theory for fingerprints (http://www.daylight.com/dayhtml/doc/theory/theory.finger.html). For the similarity determination between the compound of interest and the compounds of the mVOC database, the Tanimoto coefficient is the measure of choice:

$$\text{Tanimoto coefficient}_{A,B} = \frac{AB}{A+B-AB}$$

Bits of the binary fingerprint vectors were set to one in compound A and compound B as well as bits were set to one in both compounds and used for the calculation. The values calculated by the Tanimoto coefficient range between 0 and 1, where 1 indicates similar structures and 0 means that no similarity is found between the fingerprint representations of the molecules. Referring to the 'similarity property principle' (27), compounds that are structurally similar should exhibit a similar biological function. Nevertheless, small structural modifications can change the biological activity of the molecules dramatically (28). However, a Tanimoto coefficient >0.85 implies that the compared compounds may have a similar biological activity (29).

As an applet for sketching compounds for the 'Structure Search' and 'Add a new mVOC' function, the open-source web-component ChemDoodle is implemented on the mVOC website. ChemDoodle is also used for a 3D visualization of the mVOC structure. ChemDoodle guarantees smooth usage on different platforms.

For retrieving Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways, a similarity search between the 846 mVOCs and the 200 000 compounds from the SuperTarget database (30) was carried out. This step was conducted to obtain information on synthesis/degradation

of the mVOCs as well as potential target interactions. Compounds were considered when the Tanimoto coefficient of the similarity pair between volatile and SuperTarget compound was at least 0.85. Those compounds were mapped onto the pathway maps. Additionally, mVOCs were mapped to KEGG compounds and are also displayed on the pathway maps. The database features pathways of species available in the database. The mapped pathways are visualized by web service.

The mVOC database is implemented as a relational database on a MySQL server. Php and javascript have been used to build the website. Web access is enabled by Apache HTTP Server 2.

RESULTS

With a number of 846 compounds and 5431 synonyms, which are assigned to 349 bacterial and 69 fungi species, mVOC is the first online database containing information about mVOCs and their emitting organisms.

Search options

The database provides several possibilities to search for compounds (Figure 1). On the one hand, a form ('browse mVOC') is available, and the user can choose to search by PubChem-ID, name or molecular formula. In addition, compounds can be searched by selecting a
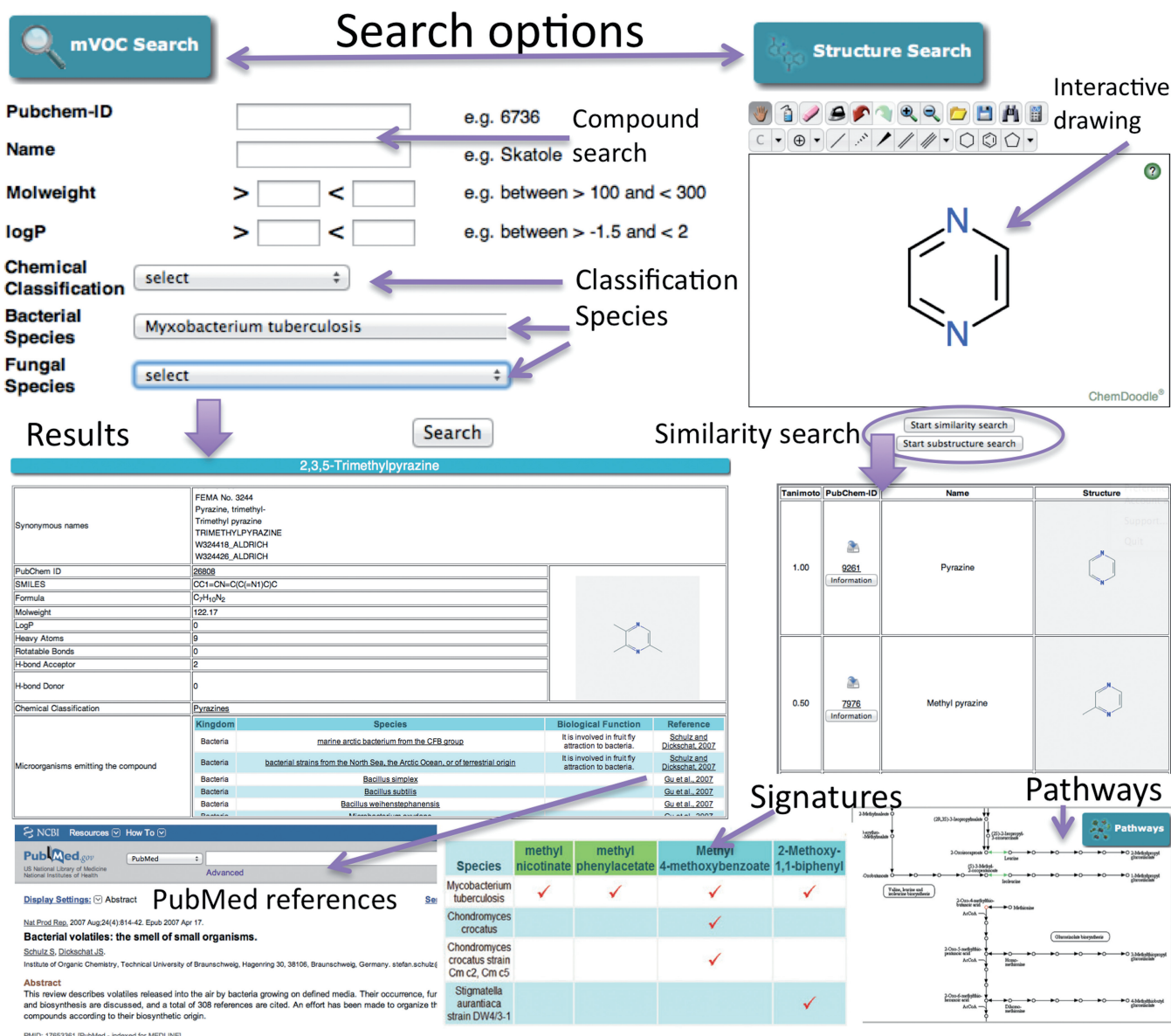
Figure 1. The mVOC database offers different search options. mVOC search: a general search form for mVOCs based on PubChem ID, name, several molecular properties as well as species. The result table is directly retrieved. Structure search: interactively drawing a structure and performing a structure or substructure search. The result table shows volatile compounds similar to the search entry (similarity search) or volatile compounds including a substructure that is similar to the search entry (substructure search). By clicking on 'Information', one will be directed to the result table of the mVOC. Signatures: the signature table shows all species emitting the same compounds as the chosen species. Compounds emitted by just one species are highlighted in green. KEGG pathways: cutout of the 2-oxocarboxylic acid metabolism pathway.

range of properties like molecular weight, logP values or by specific chemical groups (e.g. alkenes), by species, by the microorganisms' kingdom and by a combination of these parameters. Furthermore, compounds can be searched by structural similarity ('Structure Search') including substructure search. For this purpose, a molecular structure is necessary to be known. The compound of interest is screened against the database by calculating the Tanimoto coefficient between the composed compound and all compounds of the mVOC database. Finally, the user can draw a compound with the embedded ChemDoodle interface (http://www.chemdoodle.com) or upload its MOL file. In addition to that, a possibility of browsing the mVOC database is given under the form 'browse mVOC'. The database can be browsed by initial letters or chemical groups.

Search results

The resulting report of 'browse mVOC' shows information about the compounds including name, synonyms, Pubchem-ID and structural information of the mVOC of interest (Figure 1). Additionally, microorganisms emitting the compound, the effect of the compound on other organisms, the respective methods for retrieving the compounds and the corresponding references are displayed. The search results for the 'Structure Search' are represented in order of similarity with information given about the calculated Tanimoto coefficient, PubChem-ID, name and 2D structure. The button 'Information' provides detailed information about the mVOCs that are similar to the query compound.

Biological interpretation

The website features the use of KEGG pathway maps (http://www.kegg.jp/) through Web service. KEGG pathway maps supply knowledge about metabolic pathways as well as compound target interactions and offer a possibility for biological interpretation (Figure 1). Compounds of the mVOC database are mapped onto the pathways showing information about metabolic pathways providing an opportunity for further analysis. Moreover, an investigation of medical effects is also possible. A link to gene or gene clusters responsible for mVOC production will be included in future versions of this database.

Another important feature is a 'signature table' of an organism of choice (Figure 1). After selecting a species from the bacterial or fungal species dropdown menu from 'browse mVOC', a 'signature' button is available on top of the result page. The 'signature table' plots the emitted mVOCs of the chosen species compared with all microbial species, which emit these mVOCs. The table shows the uniqueness of the compounds, which is, for example, important for distinguishing between (more or less pathogenic) species.

Database extension

To enlarge the mVOC database, an upload function 'Add new mVOCs' is included (Figure 1). The user can upload a compound by drawing its structure with the ChemDoodle application. After uploading the compound, it will be verified by biochemists, and after being proofed as mVOC, it will be included into the database. Users are also encouraged to contact the authors when new volatile spectra are ready to be uploaded.

DISCUSSION

Microbes make up the majority of the world's biomass; their numbers and diversity greatly surpass those of all other organisms (31). Microbial chemical ecology is an important part of our life, and the analysis of the microbiome and unravelling its physiology including immune response, metabolism as well as pathology are future goals (32). Although sequencing becomes cheaper, an analysis of volatiles will always be faster and less invasive. Therefore, the mVOC database is an indispensable platform for this burgeoning field of microbial volatiles. Interest is particularly focussed on the identification of 'signature volatiles' of human, animal and plant pathogenic species (33). Based on these results, new possibilities for using diagnostic tools can be considered. The application of volatile antibiotics can also be envisioned.

REFERENCES

1. Bosch,T.C. and McFall-Ngai,M.J. (2011) Metaorganisms as the new frontier. *Zoology*, **114**, 185–190.
2. Torsvik,V. and Ovreas,L. (2002) Microbial diversity and function in soil: from genes to ecosystems. *Curr. Opin. Microbiol.*, **5**, 240–245.
3. Thorn,R.M. and Greenman,J. (2012) Microbial volatile compounds in health and disease conditions. *J. Breath Res.*, **6**, 024001.
4. Buttery,R. and Garibaldi,J. (1976) Geosmin and methylisoborneol in garden soil. *J. Agric. Food Chem.*, **24**, 1246–1247.
5. Wnorowski,A.U. (1992) Tastes and odours in the aquatic environment: a review. *Water SA*, **18**, 203–215.
6. Citron,C.A., Gleitzmann,J., Laurenzano,G., Pukall,R. and Dickschat,J.S. (2012) Terpenoids are widespread in Actinomycetes: a correlation of secondary metabolism and genome data. *Chembiochem*, **13**, 202–214.
7. Shestivska,V., Spanel,P., Dryahina,K., Sovova,K., Smith,D., Musılek,M. and Nemec,A. (2012) Variability in the concentrations of volatile metabolites emitted by genotypically different strains of *Pseudomonas aeruginosa*. *J. Appl. Microbiol.*, **113**, 701–713.
8. Casalinuovo,L.A., Di Pierro,D., Coletta,M. and Di Francesco,P. (2006) Application of electronic noses for disease diagnosis and food spoilage detection. *Sensors*, **6**, 1428–1439.

9. Kuske,M., Romain,A.C. and Nicolas,J. (2005) Microbial volatile organic compounds as indicators of fungi. Can an electronic nose detect fungi in indoor environments? *Build. Environ.*, **40**, 824–831.

10. Ryu,C.M., Farag,M.A., Hu,C.H., Reddy,M.S., Wie,H.X., Pare,P.W. and Kloepper,J.W. (2003) Bacterial volatiles promote growth in Arabidopsis. *Proc. Natl Acad. Sci. USA*, **100**, 4927–4932.

11. Vespermann,A., Kai,M. and Piechulla,B. (2007) Rhizobacterial volatiles affect the growth of fungi and *Arabidopsis thaliana*. *Appl. Environ. Microbiol.*, **73**, 5639–5641.

12. Kai,M. and Piechulla,B. (2010) Impact of volatiles of the rhizobacteria *Serratia odorifera* on the moss *Physcomitrella* patens. *Plant Signal Behav.*, **5**, 1–3.

13. Kai,M., Crespo,E., Cristescu,S.M., Harren,F.J.M., Francke,W. and Piechulla,B. (2010) *Serratia odorifera*: analysis of volatile emission and biological impact of volatile compounds on *Arabidopsis thaliana*. *Appl. Microbiol. Biotechnol.*, **88**, 965–976.

14. Wenke,K., Kai,M. and Piechulla,B. (2009) Belowground volatiles facilitate interactions between plant roots and soil organisms. *Planta*, **231**, 499–506.

15. Effmert,U., Kalderás,J., Warnke,R. and Piechulla,B. (2012) Volatile mediated interactions between bacteria and fungi in the soil. *J. Chem. Ecol.*, **38**, 665–703.

16. Davis,T.S., Crippen,T.L., Hofstetter,R.W. and Tomberlin,J.K. (2013) Microbial volatile emissions as insect semiochemicals. *J. Chem. Ecol.*, **39**, 687–806.

17. Blombach,B. and Eikmanns,B.J. (2011) Current knowledge on isobutanol production with *Escherichia coli Bacillus subtilis* and *Corynebacterium glutamicum*. *Bioeng. Bugs*, **2**, 346–350.

18. Ruffing,M. (2013) Metabolic engineering of hydrocarbon biosynthesis for biofuel production. In: Fang,Z. (ed.), *Liquid, Gaseous and Solid Biofuels - Conversion Techniques*. InTech, Rijeka, Croatia.

19. Kai,M., Effmert,U., Berg,G. and Piechulla,B. (2007) Volatiles of bacterial antagonists inhibit mycelial growth of the plant pathogen *Rhizoctonia solani*. *Arch. Microbiol.*, **187**, 351–360.

20. Börjesson,T., Stöllman,U. and Schnürer,J. (1992) Volatile metabolites produced by six fungal species compared with other indicators of fungal growth on cereal grains. *Appl. Environ. Microbiol.*, **58**, 2599–2605.

21. Fischer,G., Schwalbe,R., Möller,M. and Ostrowski,R. (1999) Species-spezific production on microbial volatile organic compounds (MVOC) by airborne fungi from a compost facility. *Chemosphere*, **39**, 795–810.

22. Bruce,A., Wheatley,R.E., Humphris,S.N., Hackett,C.A. and Florence,M.E.J. (2000) Production of volatile organic compounds by *Trichoderma* in media containing different amino acids and their effect on selected wood decay fungi. *Holzforschung*, **54**, 481–486.

23. Blom,D., Fabbri,C., Connor,E.C., Schiestl,F.P., Klauser,D.R., Boller,T., Eberl,L. and Weisskopf,L. (2011) Production of plant growth modulating volatiles is widespread among rhizosphere bacteria and strongly depends on culture conditions. *Environ. Microbiol.*, **13**, 3047–3058.

24. Weise,T., Kai,M., Gummesson,A., Troeger,A., Von Reuß,S., Piepenborn,S., Kosterka,F., Sklorz,M., Zimmermann,R., Francke,W. *et al.* (2012) Volatile organic compounds produced by the phytopathogenic bacterium *Xanthomonas campestris* pv. *vesicatoria* 85-10. *Beilstein J. Org. Chem.*, **8**, 579–596.

25. Schulz,S. and Dickschat,J.S. (2007) Bacterial volatiles: the smell of small organisms. *Nat. Prod. Rep.*, **24**, 814–842.

26. Willett,P., Barnard,J.M. and Downs,G.M. (1998) Chemical similarity searching. *J. Chem. Inf. Comput. Sci.*, **38**, 983–996.

27. Johnson,M.A. and Maggiora,G.M. (eds), (1990) *Concepts and Applications of Molecular Similarity*. John Willey and Sons, New York, p. 393.

28. Eckert,H. and Bajorath,J. (2007) Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches. *Drug Discov. Today*, **27**, 225–233.

29. Patterson,D.E., Cramer,R.D., Ferguson,A.M., Clark,R.D. and Weinberger,L.E. (1996) Neighborhood behavior: a useful concept for validation of 'molecular diversity' descriptors. *J. Med. Chem.*, **39**, 3049–3059.

30. Hecker,N., Ahmed,J., von Eichborn,J., Dunkel,M., Macha,K., Eckert,A., Gilson,M.K., Bourne,P.E. and Preissner,R. (2012) SuperTarget goes quantitative: update on drug-target interactions. *Nucleic Acids Res.*, **40**, D1113–D1117.

31. Romeo,J.T. (2013) Preface: microbial chemical ecology. *J. Chem. Ecol.*, **39**, 807–808.

32. Blaser,M., Bork,P., Fraser,C., Knight,R. and Wang,J. (2013) The microbiome explored: recent insights and future challenges. *Nat. Rev. Microbiol.*, **11**, 213–217.

33. Syhre,M. and Chambers,S.T. (2008) The scent of *Mycobacterium tuberculosis*. *Tuberculosis*, **88**, 317–323.