

# StreptomeDB 3.0: an updated compendium of streptomycetes natural products

Aurélien F.A. Moumbock<sup>1,†</sup>, Mingjie Gao<sup>1,†</sup>, Ammar Qaseem<sup>1,†</sup>, Jianyu Li<sup>1</sup>, Pascal A. Kirchner<sup>1</sup>, Bakoh Ndingkokhar<sup>2</sup>, Boris D. Bekono<sup>3</sup>, Conrad V. Simoben<sup>4</sup>, Smith B. Babiaka<sup>5</sup>, Yvette I. Malange<sup>5</sup>, Florian Sauter<sup>1</sup>, Paul Zierrep<sup>1</sup>, Fidele Ntie-Kang<sup>4,5,6</sup> and Stefan Günther<sup>1,\*</sup>

<sup>1</sup>Institute of Pharmaceutical Sciences, Albert-Ludwigs-Universität Freiburg, Hermann-Herder-Straße 9, D-79104 Freiburg, Germany, <sup>2</sup>Department of Organic Chemistry, University of Yaoundé I, P. O. Box 812, Yaoundé, Cameroon, <sup>3</sup>Department of Physics, Higher Teacher Training College, University of Yaoundé I, P. O. Box 47, Yaoundé, Cameroon, <sup>4</sup>Department of Pharmaceutical Chemistry, Martin-Luther-Universität Halle-Wittenberg, Wolfgang-Langenbeck Straße 4, D-06120 Halle (Saale), Germany, <sup>5</sup>Department of Chemistry, University of Buea, P. O. Box 63, Buea, Cameroon and <sup>6</sup>Institute of Botany, Technische Universität Dresden, Zellescher Weg 20b, D-01217 Dresden, Germany

Received September 14, 2020; Revised September 22, 2020; Editorial Decision September 23, 2020; Accepted September 24, 2020

## ABSTRACT

Antimicrobial resistance is an emerging global health threat necessitating the rapid development of novel antimicrobials. Remarkably, the vast majority of currently available antibiotics are natural products (NPs) isolated from streptomycetes, soil-dwelling bacteria of the genus *Streptomyces*. However, there is still a huge reservoir of streptomycetes NPs which remains pharmaceutically untapped and a compendium thereof could serve as a source of inspiration for the rational design of novel antibiotics. Initially released in 2012, StreptomeDB (<http://www.pharmbioinf.uni-freiburg.de/streptomedb>) is the first and only public online database that enables the interactive phylogenetic exploration of streptomycetes and their isolated or mutasynthesized NPs. In this third release, there are substantial improvements over its forerunners, especially in terms of data content. For instance, about 2500 unique NPs were newly annotated through manual curation of about 1300 PubMed-indexed articles, published in the last five years since the second release. To increase interoperability, StreptomeDB entries were hyperlinked to several spectral, (bio)chemical and chemical vendor databases, and also to a genome-based NP prediction server. Moreover, predicted pharmacokinetic and toxicity profiles were added. Lastly, some recent real-world use cases of StreptomeDB are highlighted, to illustrate its applicability in life sciences.

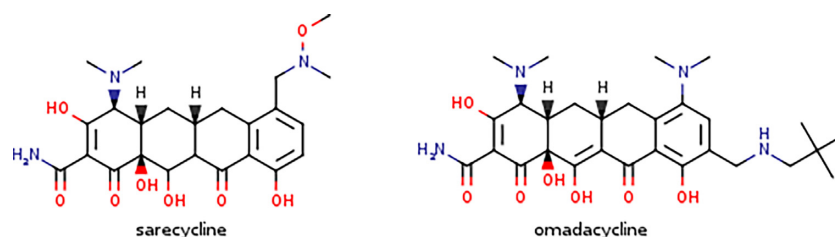
## INTRODUCTION

Microbial pathogens are gradually becoming less sensitive or even fully resistant to the currently available antimicrobials, typically as a result of genetic mutations or the misuse of antimicrobials (1). Consequently, the World Health Organization has classified antimicrobial resistance as an emerging global health threat (2), necessitating the rapid development of new antimicrobials. A recent literature survey by Newmann and Cragg shows that more than half of the antibacterials newly approved over the last four decades (3) are either natural products (NPs) or their synthetic derivatives, collectively known as antibiotics. Remarkably, the vast majority of them stem from streptomycetes, filamentous Gram-positive soil-dwelling bacteria of the genus *Streptomyces* (4). One of the most recent success stories includes the market approval by the US Food and Drug Administration (FDA) in 2018 of two tetracycline-derived third generation antibiotics: sarecycline and omadacycline (Figure 1), displaying narrow- and broad-spectrum antimicrobial activities, respectively (5,6).

Importantly, there is still a huge reservoir of streptomycetes NPs possessing atypical scaffolds, whose therapeutic potentials have not yet been fully investigated. In addition to their broad-spectrum antimicrobial activities, streptomycetes NPs have been reported to display various other types of activity including antiparasitic, anticancer, and anti-inflammatory activities (7). Given the recent advances in computer-aided drug discovery and multi-omics approaches to harness NPs (8,9), a compendium of streptomycetes NPs could serve as a source of inspiration for the rational design of new drugs of various classes.

\*To whom correspondence should be addressed. Tel: +49 761 203 4871; Fax: +49 761 203 97769; Email: stefan.guenther@pharmazie.uni-freiburg.de

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



**Figure 1.** Chemical structures of two recently FDA-approved tetracycline-derived antibiotics.

Initially released in 2012, StreptomeDB is the first public online database enabling the interactive phylogenetic exploration of streptomycetes and their isolated or mutasynthesized NPs (10). Other databases containing microbial NPs include the Natural Products Atlas (11), Antibase (12), and MarinLit (13). StreptomeDB has evolved over the years (version 2.0 (14) was released in 2015) to become a streptomycetes hub for MS/NMR structural dereplication of newly isolated NPs, substructure searching and similarity searching.

Considering the resurgence in NP-inspired drug discovery and the vast number of published articles describing the isolation and/or mutasynthesis of NPs from streptomycetes, we were prompted to update StreptomeDB. In this third release, there are substantial improvements over its forerunners, especially in terms of data content. The retrieval of relevant PubMed-indexed articles and their full-text curation, as well as scaffold decomposition, phylogenetic tree construction and NMR/MS spectral data prediction, were performed in the same manner as described for previous releases (10,14). However, to increase interoperability, StreptomeDB entries were hyperlinked to several other public online resources. Moreover, computationally predicted drug absorption, distribution, metabolism, excretion, and toxicity (ADMET) profiles were added as new features. After summarizing the growth of StreptomeDB over the years, we shall describe recent developments and highlight some recent real-world use cases exemplifying the applicability of StreptomeDB.

## GROWTH OF THE DATABASE

Notable increments in various StreptomeDB releases are summarized in Table 1. About 2500 NPs and 700 producer streptomycetes strains were newly annotated from about 1300 relevant PubMed-indexed articles published in the last five years, since the second release. Accordingly, the number of compound–organism, compound–biosynthesis route and compound–activity relationships have practically doubled, and meanwhile there is a 34% increase in the number of unique scaffolds since the last release. Overall, these statistics reflect the constantly increasing interest of researchers toward the isolation of bioactive molecules from *Streptomyces* species.

## RECENT DEVELOPMENTS

### Hyperlinks to other online resources

There is a plethora of experimental (bio)chemical data on streptomycetes and their NPs that is available in a number of disparate public databases, which was aggregated

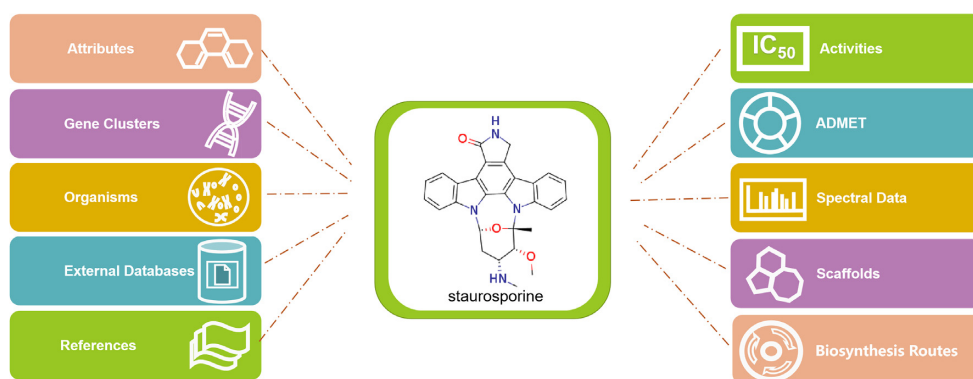
**Table 1.** The growth of the StreptomeDB over the years

	StreptomeDB		
Release number	1.0	2.0	3.0
Publication year	2012	2015	2020
Number of compounds	2444	4040	6524
Number of unique scaffolds	–*	4680	6262
Number of organisms (including strains)	1985	2584	3302
Number of compound–organism relationships	4341	6717	10 912
Number of compound–biosynthesis route relationships	307	731	1392
Number of compound–activity relationships	1036	3813	6850
Number of compounds with predicted NMR data	–	3989	6507
Number of compounds with predicted MS data	–	1945	4943
Number of compounds with predicted ADMET properties	–	–	6524
Number of referenced articles	4544	5486	6754

\* not yet implemented.

in StreptomeDB to facilitate their access to the research community. Also, we added hyperlinks to chemical vendor databases, which may foster the purchase and further investigations on streptomycetes NPs. Accordingly, when available, PubChem Compound ID (CID) (15) and CAS Registry Number (CAS RN<sup>®</sup>) were included. Further, the retrieved CIDs were mapped to several other online resources via the UniChem cross-reference platform (16). UniChem uses the InChI to easily compare various molecular structure encodings of the same compound from different sources in order to establish molecular uniqueness.

These online resources can be subdivided into three categories. Firstly, spectral databases: NMRShiftDB (17) and GNPS (18), containing NMR and MS data, respectively. Secondly, biochemical databases: ChEMBL (19), PDB (20), DrugBank (21), KEGG (22), ChEBI (23), BindingDB (24), ACToR (25), MetaboLights (26), and EPA CompTox Dashboard (27) collectively contain bioactivity data, metabolic data, toxicity data, biosynthetic pathways and protein–ligand cocrystal structures, etc. Thirdly, chemical vendor databases: MolPort (<https://www.molport.com/>), Selleck (<https://www.selleckchem.com/>), and ZINC (28). It is worth mentioning that both PubChem (15) and ChEBI (23) also contain links to chemical vendors websites. For GNPS (18) which is not integrated to UniChem, the relevant IDs were extracted from a downloadable JSON file (<https://ccms-ucsd.github.io/GNPSDocumentation/api/>).



**Figure 2.** Knowledge graph of the StreptomeDB compound card with staurosporine as example.

Nowadays, computational methods that bridge the gap between the genomic and chemical space are instrumental in genomics-driven NP discovery (29–31). Within the framework of the ‘Genes to Metabolites’ approach, we hyperlinked streptomycetes strains to the SeMPI web server (32), a genome-based pipeline we developed, that leverages published genomic data to identify encoding biosynthetic gene clusters of isolated NPs and predicts the putative structures of related NPs that could be synthesized by polyketide synthases of type I (32), and nonribosomal peptide synthetases (unpublished data).

### Predicted ADMET properties

Among other physicochemical descriptors, we predicted ADMET properties of all StreptomeDB molecules using the pkCSM web server (33). The prediction of various pharmacokinetic classes can provide a better understanding of the parameters that assist in the design of orally available, safe and efficacious drugs. A total of thirty ADMET descriptors were computed for each molecule. These include (i) adsorption: water solubility (log mol/l), Caco-2 permeability (log Papp in  $10^{-6}$  cm/s), human intestinal absorption (% absorbed), skin permeability (log Kp), P-gp substrate binding, P-gp I inhibition, and P-gp II inhibition; (ii) distribution: human VDss (log L/kg), human unbound fraction (Fu), blood brain barrier permeability (log BB), and central nervous system permeability (log PS); (iii) metabolism: CYP2D6 substrate binding, CYP3A4 substrate binding, CYP1A2 inhibition, CYP2C19 inhibition, CYP2C9 inhibition, CYP2D6 inhibition, and CYP3A4 inhibition; (iv) excretion: total clearance (log ml/min/kg), and renal OCT2 substrate binding; and (v) toxicity: AMES toxicity, maximum human dose (log mg/kg/day), hERG I inhibition, hERG II inhibition, oral rat acute toxicity (LD50, mol/kg), oral rat chronic toxicity (LOAEL, log mg/kg.bw/day), hepatotoxicity, skin sensitization, *Tetrahymena pyriformis* toxicity (log  $\mu$ g/l), and minnow toxicity (log mM).

### Overview of the StreptomeDB compound card

The depiction of 2D chemical structures displayed on compound cards (as well as the scaffold browser) of the StreptomeDB web-user interface (WUI) were updated by applying a new version of the command-line tool

Molconverter (Marvin 20.18.0, 2020, ChemAxon, <https://chemaxon.com/>). All new features, namely the hyperlinks and predicted ADMET properties, were added to the compound card. A knowledge graph of the StreptomeDB compound card is shown in Figure 2.

### USE CASES

StreptomeDB WUI and its downloadable data have widely been used in numerous real-world use cases. These can broadly be classified in four categories: (i) data integration in bioinformatics pipelines and other databases, (ii) substructure searching and virtual screening, (iii) structural dereplication, and (iv) comparative data analysis and benchmarking. In the following, we shall highlight some of the most recent real-world use cases, performed since the second release.

#### Data integration in bioinformatics pipelines and other databases

The chemical structures of StreptomeDB molecules were used as training data of bioinformatics pipelines such as the NP-likeness scorers, NaPLoS (34) and NP-Scout (35), and the peptidic NP dereplicators, VarQuest (36) and NPS (37). With regards to database integration, the non-ribosomal peptide and macrolactone subsets of StreptomeDB were incorporated into Norine (38) and MacrolactoneDB (39), respectively. Moreover, all StreptomeDB molecules were incorporated in non-redundant NP datasets, such as NPASS (40) and COCONUT (41).

#### Substructure searching and virtual screening

In order to identify novel tryptophan 6-halogenase enzymes, Lee *et al.* used 6-chloroindole as substructure to query StreptomeDB and retrieved related NPs and their producer organisms (42). From the results, the authors selected and purchased the *Streptomyces albus* N-16041 strain, from which they later on characterized a novel tryptophan 6-halogenase enzyme and identified its key catalytic residues for regioselective aromatic halogenation.

StreptomeDB molecules have also been used in both ligand- and structure-based virtual screening. On the one hand, Dias *et al.* developed quantitative structure–activity



relationship (QSAR) models for the prediction of antibacterial activity against methicillin-resistant *Staphylococcus aureus* (MRSA) (43). The authors screened StreptomeDB molecules against the best QSAR model and identified 212 hits with predicted minimum inhibitory concentration below the threshold of 5 mM. From these hits, 112 are annotated as antibiotics in StreptomeDB, with four of them as anti-MRSA agents. On the other hand, Kalhor *et al.* docked StreptomeDB molecules to the GyrA subunit of the DNA gyrase enzyme to identify antibiotics that can overcome fluoroquinolone resistance in wild-type and mutated GyrA (44). Three angucycline-possessing hits were identified as potential binders of both wild-type and mutated GyrA.

### Structural dereplication

Mehetre *et al.* matched the experimental MS data of new isolates from *Streptomyces* sp. GH176 to the predicted MS data of StreptomeDB and they dereplicated the chemical structures of abyssomicin I and terpentecin (45). Similarly, Arn *et al.* used the StreptomeDB WUI to dereplicate the structure of dihydropicromycin, isolated from *Streptomyces* sp. W367A (46).

### Comparative data analysis and benchmarking

In a couple of studies (47,48), the druglikeness of StreptomeDB molecules was profiled and compared with that of other NP databases including NANPDB (48) and NuBBE<sub>DB</sub> (49), and also with that of the DrugBank (21) dataset of experimental/approved drugs. We equally used StreptomeDB molecules as a benchmark dataset for the aforementioned SeMPI web server (32).

### CONCLUSION

Streptomyces harbor unique and structurally diverse NPs that can serve as leads for the development of various new drugs, notably against antibiotic-resistant bacterial infections. StreptomeDB (<http://www.pharmbioinf.uni-freiburg.de/streptomedb>) is the first and only public compendium solely dedicated to streptomyces and their NPs. In this third release, it has grown to contain about 6500 NPs originating from about 3300 producer streptomyces strains. Moreover, new features were added to StreptomeDB to enhance its usability. These new features include hyperlinks to a number of diverse public online resources. Additionally, we improved the physicochemical descriptions of database molecules by including predicted ADMET properties, and the WUI was updated to enhance user-friendliness. The highlighted real-world use cases clearly demonstrate its wide range of applicability in life sciences. With this increase in the number of entries and in information for each entry, it is hoped that StreptomeDB would become more useful.

### ACKNOWLEDGEMENTS

We thank the many users of StreptomeDB for their feedback.

### FUNDING

Baden-Württemberg Foundation [BWST.WSF-043]; China Scholarship Council [201908080143]; German Academic Exchange Service [91653768, 91611788, 57516652]; German Research Foundation [RTG 1976]. Funding for open access charge: University of Freiburg and German Research Foundation [RTG 1976].

*Conflict of interest statement.* None declared.

### REFERENCES

- Andersson, D.I., Balaban, N.Q., Baquero, F., Courvalin, P., Glaser, P., Gophna, U., Kishony, R., Molin, S. and Tønjum, T. (2020) Antibiotic resistance: turning evolutionary principles into clinical reality. *FEMS Microbiol. Rev.*, **44**, 171–188.
- World Health Organization ed. (2014) In: *Antimicrobial Resistance: Global Report on Surveillance*. World Health Organization, Geneva.
- Newman, D.J. and Cragg, G.M. (2020) Natural products as sources of new drugs over the nearly four decades from 01/1981 to 09/2019. *J. Nat. Prod.*, **83**, 770–803.
- Xia, H., Zhan, X., Mao, X.-M. and Li, Y.-Q. (2020) The regulatory cascades of antibiotic production in *Streptomyces*. *World J. Microbiol. Biotechnol.*, **36**, 13.
- Zhan, G., Critchley, I., Lin, L.-Y. and Alvandi, N. (2019) Microbiological profile of sarecycline, a novel targeted spectrum tetracycline for the treatment of *Acne vulgaris*. *Antimicrob. Agents Chemother.*, **63**, e01297-18.
- Honeyman, L., Ismail, M., Nelson, M.L., Bhatia, B., Bowser, T.E., Chen, J., Mechiche, R., Ohemeng, K., Verma, A.K., Cannon, E.P. *et al.* (2015) Structure-activity relationship of the aminomethylcyclines and the discovery of omadacycline. *Antimicrob. Agents Chemother.*, **59**, 7044–7053.
- Sivalingam, P., Hong, K., Pote, J. and Prabakar, K. (2019) Extreme environment *Streptomyces*: potential sources for new antibacterial and anticancer drug leads? *Int. J. Microbiol.*, **2019**, 5283948.
- Moumbock, A.F.A., Li, J., Mishra, P., Gao, M. and Günther, S. (2019) Current computational methods for predicting protein interactions of natural products. *Comput. Struct. Biotechnol. J.*, **17**, 1367–1376.
- Hug, J.J., Bader, C.D., Remskar, M., Cirnski, K. and Müller, R. (2018) Concepts and methods to access novel antibiotics from *Actinomycetes*. *Antibiot. Basel Switz.*, **7**, 44.
- Lucas, X., Senger, C., Erxleben, A., Grüning, B.A., Döring, K., Mosch, J., Flemming, S. and Günther, S. (2013) StreptomeDB: a resource for natural compounds isolated from *Streptomyces* species. *Nucleic Acids Res.*, **41**, D1130–D1136.
- van Santen, J.A., Jacob, G., Singh, A.L., Aniebok, V., Balunas, M.J., Bunsko, D., Neto, F.C., Castaño-Espriu, L., Chang, C., Clark, T.N. *et al.* (2019) The Natural Products Atlas: an open access knowledge base for microbial natural products discovery. *ACS Cent. Sci.*, **5**, 1824–1833.
- Laatsch, H. (2014) In: *The Natural Compound Identifier Antibase Software*. John Wiley & Sons, Hoboken.
- Blunt, J. and Munro, M. (2003) In: *MarinLit. A Database of the Literature on Marine Natural Products for Use on a Macintosh Computer Prepared and Maintained by the Marine Chemistry Group*. Department of Chemistry, University of Canterbury, Canterbury.
- Klementz, D., Döring, K., Lucas, X., Telukunta, K.K., Erxleben, A., Deubel, D., Erber, A., Santillana, I., Thomas, O.S., Bechthold, A. *et al.* (2016) StreptomeDB 2.0—an extended resource of natural products produced by streptomyces. *Nucleic Acids Res.*, **44**, D509–D514.
- 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.
- Chambers, J., Davies, M., Gaulton, A., Papadatos, G., Hersey, A. and Overington, J.P. (2014) UniChem: extension of InChI-based compound mapping to salt, connectivity and stereochemistry layers. *J. Cheminformatics*, **6**, 43.
- Steinbeck, C. and Kuhn, S. (2004) NMRShiftDB—compound identification and structure elucidation support through a free community-built web database. *Phytochemistry*, **65**, 2711–2717.

18. Wang, M., Carver, J.J., Phelan, V.V., Sanchez, L.M., Garg, N., Peng, Y., Nguyen, D.D., Watrous, J., Kapono, C.A., Luzzatto-Knaan, T. *et al.* (2016) Sharing and community curation of mass spectrometry data with Global Natural Products Social Molecular Networking. *Nat. Biotechnol.*, **34**, 828–837.
19. Mendez, D., Gaulton, A., Bento, A.P., Chambers, J., De Veij, M., Félix, E., Magariños, M.P., Mosquera, J.F., Mutowo, P., Nowotka, M. *et al.* (2019) ChEMBL: towards direct deposition of bioassay data. *Nucleic Acids Res.*, **47**, D930–D940.
20. Burley, S.K., Bertram, H.M., Bhikadiya, C., Bi, C., Chen, L., Di Costanzo, L., Christie, C., Dalenberg, K., Duarte, J.M., Dutta, S. *et al.* (2019) RCSB Protein Data Bank: biological macromolecular structures enabling research and education in fundamental biology, biomedicine, biotechnology and energy. *Nucleic Acids Res.*, **47**, D464–D474.
21. Wishart, D.S., Feunang, Y.D., Guo, A.C., Lo, E.J., Marcu, A., Grant, J.R., Sajed, T., Johnson, D., Li, C., Sayeeda, Z. *et al.* (2018) DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.*, **46**, D1074–D1082.
22. Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. and Morishima, K. (2017) KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.*, **45**, D353–D361.
23. Hastings, J., Owen, G., Dekker, A., Ennis, M., Kale, N., Muthukrishnan, V., Turner, S., Swainston, N., Mendes, P. and Steinbeck, C. (2016) ChEBI in 2016: improved services and an expanding collection of metabolites. *Nucleic Acids Res.*, **44**, D1214–D1219.
24. Gilson, M.K., Liu, T., Baitaluk, M., Nicola, G., Hwang, L. and Chong, J. (2016) BindingDB in 2015: a public database for medicinal chemistry, computational chemistry and systems pharmacology. *Nucleic Acids Res.*, **44**, D1045–D1053.
25. Judson, R.S., Martin, M.T., Egeghy, P., Gangwal, S., Reif, D.M., Kothiy, P., Wolf, M., Cathey, T., Transue, T., Smith, D. *et al.* (2012) Aggregating data for computational toxicology applications: the U.S. Environmental Protection Agency (EPA) Aggregated Computational Toxicology Resource (ACToR) system. *Int. J. Mol. Sci.*, **13**, 1805–1831.
26. Haug, K., Cochrane, K., Nainala, V.C., Williams, M., Chang, J., Jayaseelan, K.V. and O'Donovan, C. (2020) MetaboLights: a resource evolving in response to the needs of its scientific community. *Nucleic Acids Res.*, **48**, D440–D444.
27. Williams, A.J., Grulke, C.M., Edwards, J., McEachran, A.D., Mansouri, K., Baker, N.C., Patlewicz, G., Shah, I., Wambaugh, J.F., Judson, R.S. *et al.* (2017) The CompTox Chemistry Dashboard: a community data resource for environmental chemistry. *J. Cheminformatics*, **9**, 61.
28. Sterling, T. and Irwin, J.J. (2015) ZINC 15—ligand discovery for everyone. *J. Chem. Inf. Model.*, **55**, 2324–2337.
29. Mounbock, A.F., Simoben, C.V., Wessjohann, L., Sippl, W., Günther, S. and Ntie-Kang, F. (2017) Computational studies and biosynthesis of natural products with promising anticancer properties. In: Badria, F.A. (ed). *Natural Products and Cancer Drug Discovery*. Intech Open, Rijeka, pp. 257–285.
30. Lee, N., Hwang, S., Kim, J., Cho, S., Palsson, B. and Cho, B.-K. (2020) Mini review: genome mining approaches for the identification of secondary metabolite biosynthetic gene clusters in *Streptomyces*. *Comput. Struct. Biotechnol. J.*, **18**, 1548–1556.
31. Mounbock, A.F., Ntie-Kang, F., Akone, S.H., Li, J., Gao, M., Telukunta, K.K. and Günther, S. (2019) An overview of tools, software, and methods for natural product fragment and mass spectral analysis. *Phys. Sci. Rev.*, **4**, 9.
32. Zierp, P.F., Padilla, N., Yonchev, D.G., Telukunta, K.K., Klementz, D. and Günther, S. (2017) SeMPI: a genome-based secondary metabolite prediction and identification web server. *Nucleic Acids Res.*, **45**, W64–W71.
33. Pires, D.E.V., Blundell, T.L. and Ascher, D.B. (2015) pkCSM: predicting small-molecule pharmacokinetic and toxicity properties using graph-based signatures. *J. Med. Chem.*, **58**, 4066–4072.
34. Sorokina, M. and Steinbeck, C. (2019) NaPLeS: a natural products likeness scorer-web application and database. *J. Cheminformatics*, **11**, 55.
35. Chen, Y., Stork, C., Hirte, S. and Kirchmair, J. (2019) NP-Scout: machine learning approach for the quantification and visualization of the natural product-likeness of small molecules. *Biomolecules*, **9**, 43.
36. Gurevich, A., Mikheenko, A., Shlemov, A., Korobeynikov, A., Mohimani, H. and Pevzner, P.A. (2018) Increased diversity of peptidic natural products revealed by modification-tolerant database search of mass spectra. *Nat. Microbiol.*, **3**, 319–327.
37. Tagirdzhanov, A.M., Shlemov, A. and Gurevich, A. (2019) NPS: scoring and evaluating the statistical significance of peptidic natural product-spectrum matches. *Bioinform. Oxf. Engl.*, **35**, i315–i323.
38. Flissi, A., Ricart, E., Campart, C., Chevalier, M., Dufresne, Y., Michalik, J., Jacques, P., Flahaut, C., Lisacek, F., Leclère, V. *et al.* (2020) Norine: update of the nonribosomal peptide resource. *Nucleic Acids Res.*, **48**, D465–D469.
39. Zin, P.P.K., Williams, G.J. and Ekins, S. (2020) Cheminformatics analysis and modeling with MacrolactoneDB. *Sci. Rep.*, **10**, 6284.
40. Zeng, X., Zhang, P., He, W., Qin, C., Chen, S., Tao, L., Wang, Y., Tan, Y., Gao, D., Wang, B. *et al.* (2018) NPASS: natural product activity and species source database for natural product research, discovery and tool development. *Nucleic Acids Res.*, **46**, D1217–D1222.
41. Sorokina, M. and Steinbeck, C. (2020) Review on natural products databases: where to find data in 2020. *J. Cheminformatics*, **12**, 20.
42. Lee, J., Kim, J., Kim, H., Kim, E.-J., Jeong, H.-J., Choi, K.-Y. and Kim, B.-G. (2020) Characterization of a tryptophan 6-halogenase from *Streptomyces albus* and its regioselectivity determinants. *Chembiochem Eur. J. Chem. Biol.*, **21**, 1446–1452.
43. Dias, T., Gaudêncio, S.P. and Pereira, F. (2018) A computer-driven approach to discover natural product leads for methicillin-resistant *Staphylococcus aureus* infection therapy. *Mar. Drugs*, **17**, 16.
44. Kalhor, H., Sadeghi, S., Marashiyani, M., Kalhor, R., Aghaei Gharehbolagh, S., Akbari Eidgahi, M.R. and Rahimi, H. (2020) Identification of new DNA gyrase inhibitors based on bioactive compounds from *Streptomyces*: structure-based virtual screening and molecular dynamics simulations approaches. *J. Biomol. Struct. Dyn.*, **38**, 791–806.
45. Gajanan, T., Vinodh, J.S., Bhushan, B.B., Desai, D., Santhakumari, B., Mahesh, S.D. and Syed, G.D. (2019) Bioactivities and molecular networking-based elucidation of metabolites of potent actinobacterial strains isolated from the Unkeshwar geothermal springs in India. *RSC Adv.*, **9**, 9850–9859.
46. Arn, F., Frasson, D., Krosalakova, I., Rezzonico, F., Pothier, J.F., Riedl, R. and Sievers, M. (2020) Isolation and identification of *Actinomycetes* strains from Switzerland and their biotechnological potential. *Chimia*, **74**, 382–390.
47. Chen, Y., Garcia de Lomana, M., Friedrich, N.-O. and Kirchmair, J. (2018) Characterization of the chemical space of known and readily obtainable natural products. *J. Chem. Inf. Model.*, **58**, 1518–1532.
48. Ntie-Kang, F., Telukunta, K.K., Döring, K., Simoben, C.V., A Mounbock, A.F., Malange, Y.I., Njume, L.E., Yong, J.N., Sippl, W. and Günther, S. (2017) NANPDB: a resource for natural products from Northern African sources. *J. Nat. Prod.*, **80**, 2067–2076.
49. Pilon, A.C., Valli, M., Dametto, A.C., Pinto, M.E.F., Freire, R.T., Castro-Gamboa, I., Andricopulo, A.D. and Bolzani, V.S. (2017) NuBBE<sub>DB</sub>: an updated database to uncover chemical and biological information from Brazilian biodiversity. *Sci. Rep.*, **7**, 7215.