Medicine in Drug Discovery 9 (2021) 100081

Contents lists available at ScienceDirect

Medicine in Drug Discovery

journal homepage: https://www.journals.elsevier.com/medicine-in-drug-discovery

Research Paper

Peptides in chemical space

Alice Capecchi, Jean-Louis Reymond *

Department of Chemistry and Biochemistry, University of Bern, Freiestrasse 3, 3012 Bern, Switzerland

ARTICLE INFO

Article history:
Received 17 December 2020
Accepted 11 January 2021
Available online 15 January 2021

Keywords:
peptides
chemical space
machine learning
genetic algorithm
molecular fingerprints
data visualization

ABSTRACT

Peptides, defined as sequences of amino acids up to approximately 50 residues in length, represent an extremely large reservoir of potentially bioactive compounds, referred to here as the peptide chemical space. Recent advances in computer hardware and software have led to a wide application of computational methods to explore this chemical space. Here, we review different in silico approaches including structure-based design, genetic algorithms, and machine learning. We also review the use of molecular fingerprints to sample virtual libraries and to visualize the peptide chemical space. Finally, we present an overview of the known peptide chemical space in form of an interactive map representing 40,531 peptides collected from eleven open-access peptide and peptide-containing databases, accessible at https://tm.gdb.tools/map4/peptide_databases_tmap/. These peptides are displayed as TMAP (Tree-Map) according to their molecular fingerprint similarity computed using MAP4, a MinHashed atom pair fingerprint well suited to analyze large molecules.

1. Introduction

Here, we define the peptide chemical space as the ensemble of all amino acid sequences up to approximately 50 residues in length. The relevance of this chemical space for medicine is evidenced by a large number of therapeutic peptides, in particular hormones and analogs such as insulin (1) or the recently FDA-approved bremelanotide (2–4). Despite its size, the peptide chemical space can be precisely defined through a list of amino acid building blocks, usually the proteinogenic amino acids, and the length and topologies of the peptide chains that are considered, which may be linear, cyclic, or branched.

Following the invention of solid-phase peptide synthesis (SPPS) and recombinant methods in molecular biology, a number of experimental approaches have been developed to search the peptide chemical space for compounds binding to a specific molecular target by synthesizing and testing large combinatorial libraries (5). More recently, advances in computer hardware and software have made it possible to select bioactive peptides by computational methods, thereby focusing experimental evaluation to a selected set of test sequences, as well as to develop a global understanding of the peptide chemical space by comparing all known bioactive peptides with each other (6–10). In this review, we summarize recent advances in computational peptide design. We classified computational design approaches as follows: a) structure-based design, where 3D-modeling of the site of action guides the selection of test peptides; b) GA (genetic algorithms), which select test peptides by iterative cycles of mutations and fitness selection; c) ML (machine learning) methods, which exploit information on known bioactive peptides to propose new ones; and d) molecular fingerprints, which focus on calculated molecular similarity between peptide structures to enable a focused sampling (Figure 1). Finally, we present an overview of the currently know chemical space of bioactive peptides in form of an interactive chemical space map.

2. Structure-based design

If the 3D structure of the targeted site of action of the desired peptide is known in advance, one can select potentially bioactive peptides by modeling their interactions with this site using docking and molecular dynamics. This approach has been historically the first method to design bioactive peptides computationally and has been extensively reviewed (11–15). A recent example of this approach is the computational design of miniprotein inhibitors of the SARS-CoV-2 spike protein ACE2 (Angiotensin-converting enzyme 2) interaction stopping the viral entry into cells by Cao et al. (16). The inhibitors were designed in silico using two different strategies. First, a library of peptide sequences was designed using the Rosetta software (17) to incorporate the ACE2 helix responsible for the most interaction with the spike protein RDB (Receptor Binding Domain). Second, a set of sequences was designed from scratch though large large-scale de novo design of small helical scaffolds, followed by RIF (Rotamer Interaction Field) docking with the spike protein RBD (receptor binding domain), where RIF docking has the peculiarity of considering multiple conformations of the binding pocket (Figure 1A) (18).

* Corresponding author.
E-mail address: jean-louis.reymond@dcb.unibe.ch (J.-L. Reymond).

http://dx.doi.org/10.1016/j.medidd.2021.100081

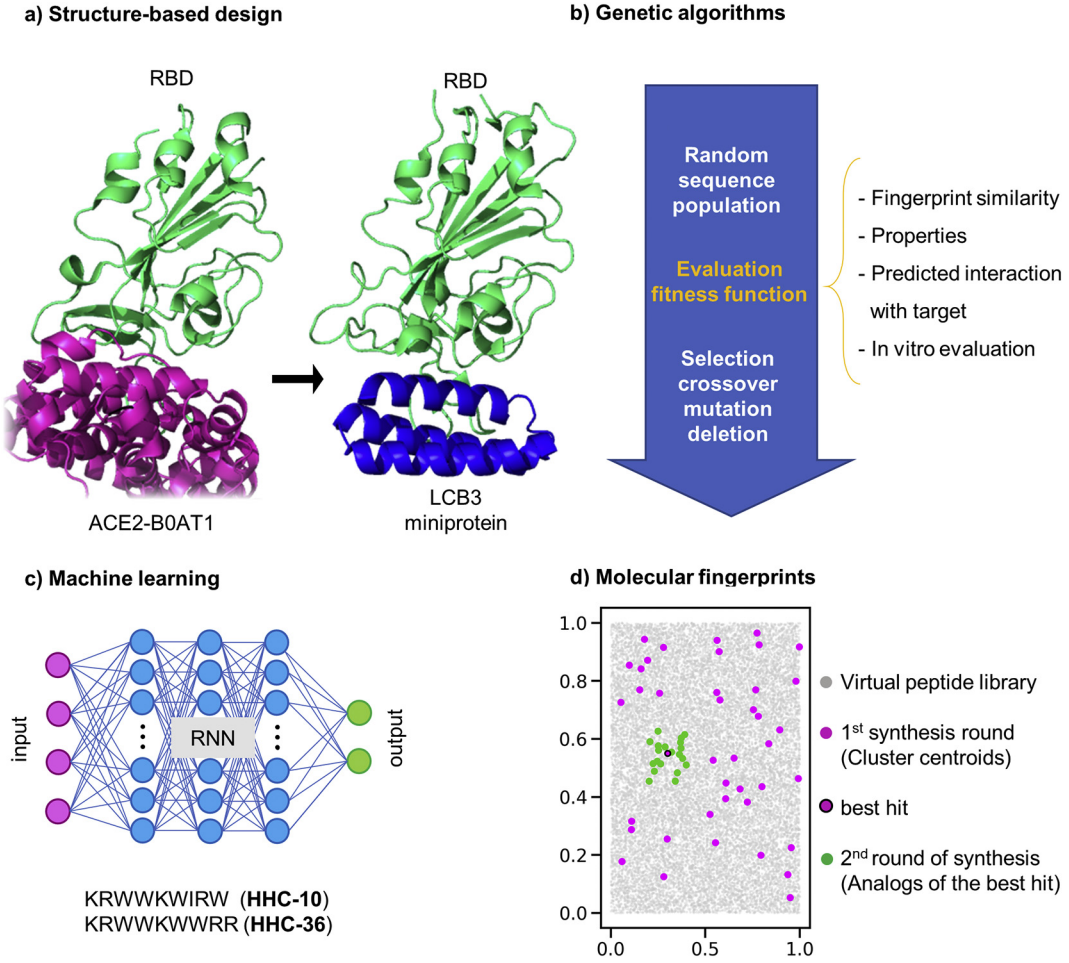

Figure 1. (A) The SARS-CoV-2 spike protein RBD (Receptor Binding Domain, green) binds ACE2 (Angiotensin-converting enzyme 2, pink), causing the virus to enter the cell. The miniprotein LCB3 (blue) was designed to bind the RBD (green) and inhibit the interaction between RBD and ACE2. (B) Genetic algorithm workflow schematic and classification based on the fitness function. (C) Example of an RNN (Recurrent Neural Network) architecture and sequences of the antimicrobial HHC-10 and HHC-36 discovered with this approach. (D) schematic representation of the sampling of a virtual library using molecular fingerprints.

Other recent examples of structure-based peptide design include the discovery of cyclic peptides with high binding affinity to diverse influenza strains through modeling based on antibody loops by Sevy et al. (14), and the design of stapled peptides that activate the VapC complex of the *Mycobacterium tuberculosis* and lead to the arrest of bacterial cell growth by Kang et al. (19). Structure-based design sometimes simply aims to identify peptides that mimic the structure of a known bioactive peptide. A recent example of this approach is the design of peptides that assemble into cross-α amyloid-like structures by Zhang et al. (20).

3. Genetic algorithms

A GA is a search algorithm inspired by the evolution theory, which optimizes a population of solutions toward a given goal through iterative cycles of mutations and selection of the fittest solutions using a fitness function (21). If the solutions searched by the algorithm are set to be peptide sequences, GAs can be applied to find novel peptides (Figure 1B).

The fitness function of a peptide GA can be based on calculated properties. In 2003, Teixido et al. used a GA to identify peptides capable of crossing the BBB (blood-brain barrier) with a fitness function based on a set of descriptors comprising molecular weight, length, amphiphilicity, isoelectric point, LogP, secondary structure, presence of aromatic and positive residues, potential hydrogen bonds, and the nature of C- and N-termini. The ideal set of values for these descriptors was derived from a statistical analysis of the experimental data on peptide-BBB permeability (22). More recently, Beltran and Brizuela used mean hydrophobicity, helical hydrophobic moment, net charge, and isoelectric point to design selective cationic antibacterial peptides (23). In another recent example of GA guided by properties, Port et al. optimized a guava antimicrobial peptide using a fitness function based on the ratio between hydrophobic moment and α-helix propensity (24).

Predicted protein–peptide interactions can also be used as fitness function of a GA. In 2011, Knapp et al. optimized peptides for major histocompatibility complex binding using a GA and the consensus of five different binding prediction methods in its fitness function (25). More recently, King et al. discovered an α-conotoxin analog with optimal binding to the α3β2-nicotinic acetylcholine receptor using a GA and an AutoDock-based fitness function to guide their search (26).

The fitness function used to guide selection in a GA can also be estimated with ML property prediction. For example, Fjell et al. used the prediction of an artificial neural network to drive a GA toward active antimicrobial peptides (27). Additional ML approaches are discussed in the following sections. A further example of GA for peptide design exploiting molecular fingerprint similarity as fitness function is discussed below in Section 5.

It is also possible to use the feedback of experimental analysis to guide a GA, as exemplified by Yoshida et al., who combined supervised ML with in vitro testing as fitness function of a GA to optimize antimicrobial peptides (28). Another recent example is the work of Neuhaus et al. (29). Starting from known ACPs (anticancer peptides), these authors used a GA

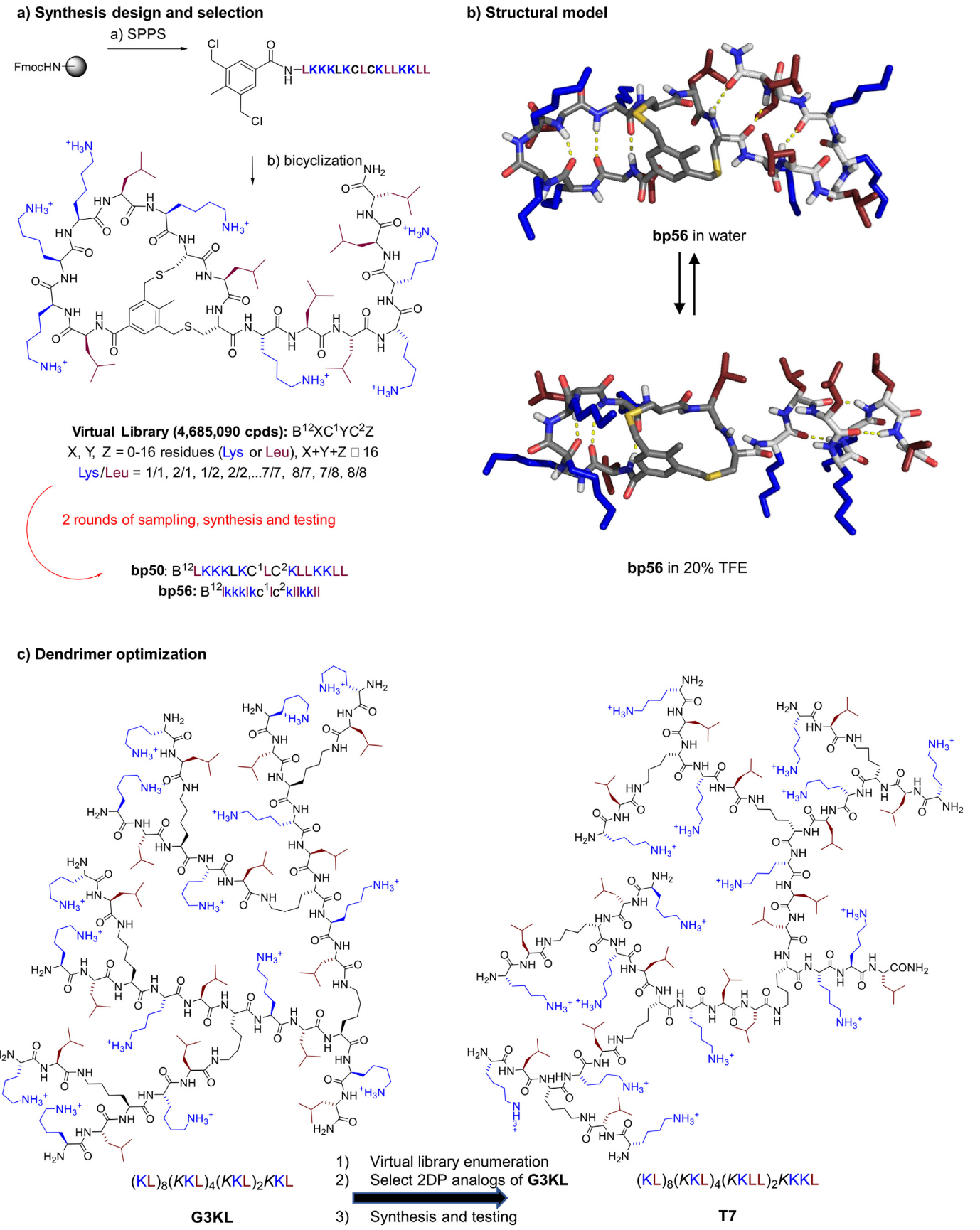

Figure 2. Molecular fingerprint-guided discovery of antimicrobial peptides. (A) Synthesis and virtual library design and selection of **bp56**. (B) Molecular dynamics studies of **bp56** in water with or without TFE (trifluoroethanol) to mimic the membrane environment reveals a dynamic conformation. (C) Optimization of antimicrobial peptide dendrimer **G3KL** by virtual library enumeration, nearest neighbor selection, synthesis, and testing.

coupled with in vitro testing to generate new ACPs with improved activity. They showed that both the interaction with the membrane and the degree of peptide dimerization were responsible for the anticancer activity.

4. Machine learning

ML approaches are used for two major tasks: property prediction and generation of new sequences. For property prediction, one uses supervised ML techniques, for which the task consists of mapping an input to a specific output (Figure 1C) (30–35). The input can be not only the peptide sequence itself but also descriptors, structure-based features, molecular fingerprints, or a combination of the previous. The output of the ML model is usually a label, such as active/inactive for a specific application. Property prediction by ML requires a large amount of highly curated data, highlighting the importance of manually curated peptide databases that collect sequence activity and toxicity.

The first example of this approach was reported in 2009 by Cherkasov et al. with the discovery of two tryptophan- and arginine-rich antimicrobial peptides, HHC-10 and HHC-36, which were more potent and shorter than similar arginine-rich peptides found in Nature such as indolicidin (Figure 1C) (36). These authors trained an artificial neural network classifier with 44 QSAR descriptors to discern between antimicrobial and nonantimicrobial peptides. They then used this trained neural network to classify each peptide in a virtual library of 100,000 random nonapeptides enriched with tryptophan, arginine, and lysine, as active or inactive. In a recent example of this approach (37), Timmons and Hewage showed that one can use supervised ML to train a neural network classifier to distinguish between hemolytic and nonhemolytic peptides at the example of peptides from the DAASPDB and the Hemolytic databases (38).

The second application of ML consists of training generative models to output new peptide sequences with specific characteristics. For example, Müller et al. recently reported an LSTM-RNN (long short-term memory recurrent neural network) capable of generating helical peptides with predicted antimicrobial activity (39). In a similar approach, Grisoni et al. trained an LSTM-RNN to generate α-helical cationic amphipathic sequences, and then fine-tuned it using 26 known ACPs. Twelve of the proposed sequences were synthesized and ten showed the expected membranolytic activity (40).

Classification and generative ML models can also be combined. For example, Tucs et al. recently reported a GAN (Generative Adversarial Network) to generate antimicrobial peptides (41). A GAN is a ML architecture composed of an ML generative model and a discriminator, which is generally an ML classifier, whereby both models are trained as a pair. The task of the generator is to generate sequences resembling known antimicrobial peptide sequences, while the task of the discriminator is to distinguish between potential antimicrobial peptides and random sequences.

5. Molecular fingerprints

We recently showed that one can discover bioactive peptides computationally in the absence of precise structural modeling by using molecular fingerprint comparisons. This approach is well-known in small molecule drug discovery (42), but still underexploited with peptides. We demonstrated the feasibility of this approach by discovering antimicrobial bicyclic peptides against the Gram-negative bacterium *Pseudomonas aeruginosa* and its biofilms (43).

To discover active bicyclic peptides, we used a shape and pharmacophore fingerprint called 2DP describing the relative positions of cationic and hydrophobic groups, an important parameter for the targeted membrane disruptive activity. The workflow comprised the following steps: 1) establishing a SPPS protocol for bicyclic peptides comprising nine variable positions; 2) enumerating a virtual library considering all possible combinations of lysine, leucine at the variable positions; 3) computing 2DP-fingerprint similarities between all pairs of bicyclic peptides and clustering the virtual library to sample the overall diversity of the virtual library; and 4) synthesizing and testing a small set of sampled bicyclic peptides. This approach led to the identification of a single active bicyclic peptide, which we then optimized by synthesizing and testing further analogs identified by 2DP-similarity searching in the virtual library (Figure 1D).

Table 1
Peptide and peptide-containing databases publicly available and downloadable in bulk

Name	Description	Size[a]	Web page	Ref.
PDB[b]	3-D structural data of large biomolecules	8805	https://www.rcsb.org/	(62)
SwissProt[c]	Sequences and functional information of peptides and proteins manually annotated	9129	https://www.uniprot.org/	(63)
SATPdb[d]	Therapeutic peptides	14,985	http://crdd.osdd.net/raghava/satpdb/	(64)
DBAASP[e]	Antimicrobial peptides	10,999	https://dbaasp.org/	(65)
DRAMP[f]	Antimicrobial peptides	3673	cpu-bioinfor.org	(66)
AVPdb[g]	Antiviral peptides	1801	http://crdd.osdd.net/servers/avpdb/	(67)
SPdb[h]	Signal peptides	2340	http://proline.bic.nus.edu.sg/spdb/	(68)
NeuroPedia[i]	Neuropeptides	392	http://proteomics.ucsd.edu/Software/NeuroPedia/	(69)
DADP[j]	Anuran defense peptides	743	http://split4.pmfst.hr/dadp/	(70)
Quorumpeps[k]	Quorum sensing peptides	243	http://quorumpeps.ugent.be/	(71)
AntiAngioPred[l]	Angiogenic peptides	197	http://clri.res.in/subramanian/tools/antiangiopred/index.html	(72)
Total of peptidic entries constituted by 2 to 50 natural amino acids		53,307		
Unique sequences collected across databases		40,531		

[a] Number of unique peptidic entries constituted by 2 to 50 natural amino acids.
[b] PBD = Protein Data Bank.
[c] SwissProt = peptide sequences from the Uni-Prot database.
[d] SATPdb = Structurally Annotated Therapeutic Peptides database.
[e] DBAASP = Database of Antimicrobial Activity and Structure of Peptides.
[f] DRAMP = Data Repository of Antimicrobial Peptides.
[g] AVPdb = Antiviral Peptide database.
[h] SPdb = Signal Peptide database.
[i] NeuroPedia = Neuropeptides database and spectral library.
[j] DADP = Database of Anuran Defense Peptides.
[k] Quorumpeps = Quorumpeps database.
[l] AntiAngioPred = Server for Prediction of Anti-Angiogenic Peptides.

The virtual library of the above proof-of-concept experiment only comprised 6230 different bicyclic peptides. In a subsequent project, we applied the same approach to a differently designed and much larger virtual library of 4.7 million bicyclic peptides and identified the cysteine bridged bicyclic peptide **bp50** and its D-enantiomer **bp56** as potent antimicrobial peptides against multidrug-resistant strains of *Acinetobacter baumannii* and *P. aeruginosa* (Figure 2A) (44). Note that membrane disruptive peptides are generally conformationally flexible and that this conformational flexibility is necessary for their activity (45). Indeed, molecular dynamics studies and CD spectra suggest that **bp56** exists in a dynamic equilibrium between a β-sheet conformation in water and a partially α-helical amphiphilic conformation in a membrane environment (Figure 2B).

Considering that our 2DP molecular fingerprint could be applied to any type of peptide chain topology, we further implemented this fingerprint-based approach to search for analogs of AMPD (antimicrobial peptide dendrimer) **G3KL**, which contains a highly ramified peptide chain (46). This AMPD kills a broad range of Gram-negative bacteria including multidrug-resistant clinical isolates by a membrane disruptive mechanism with almost

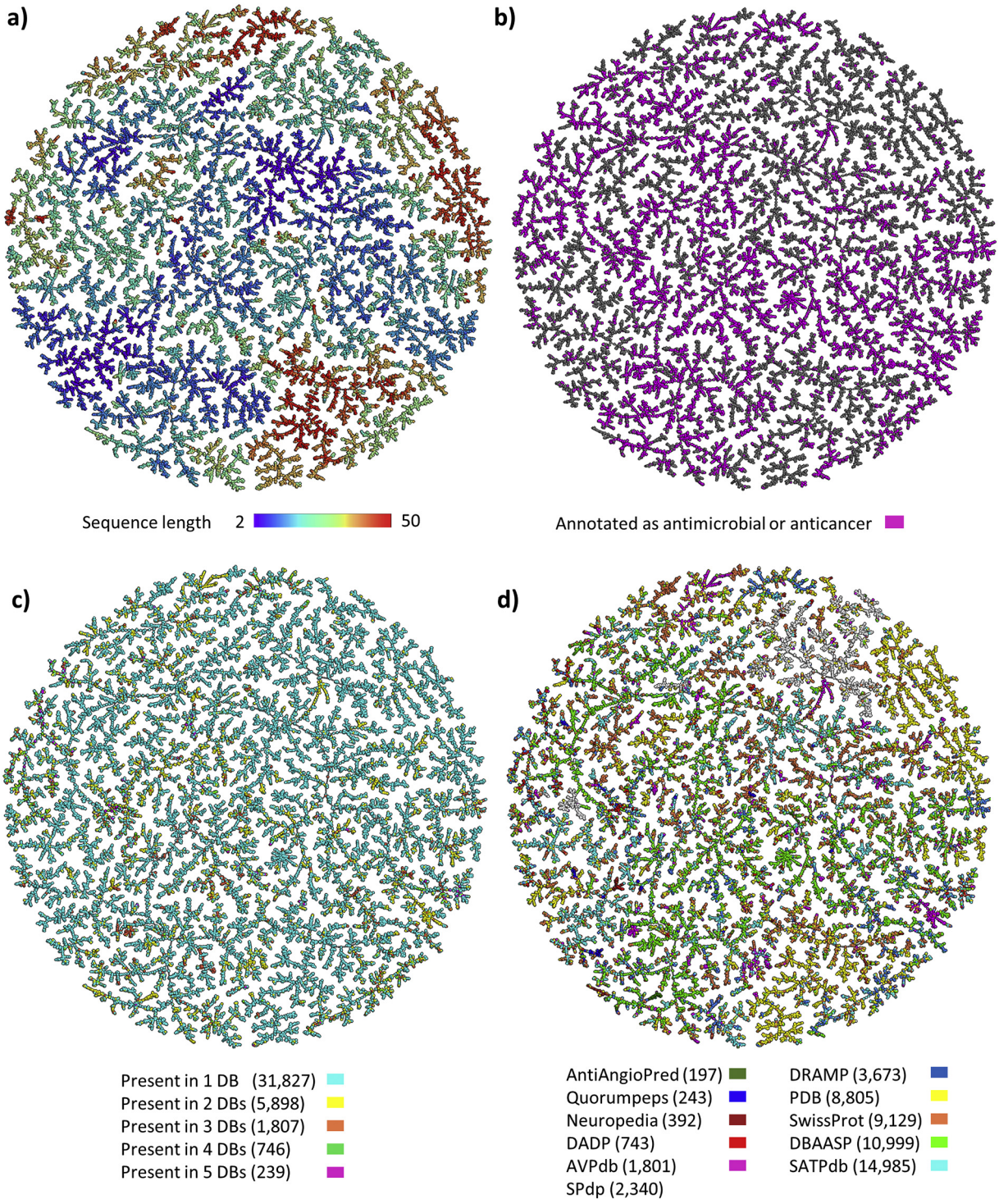

Figure 3. TMAP of the MAP4-encoded peptide databases space-colored according to (A) sequence length, (B) source database (DB, entries present in multiple databases were assigned to the smallest one), (C) occurrences across databases, (D) antimicrobial and anticancer activity. Further colors based on different activity criteria are available, and they can be found in the TMAP at https://tm.gdb.tools/map4/peptide_databases_tmap/.

no resistance (47–49). **G3KL** exhibits angiogenic as well as antibiofilm properties (50,51), and can act synergistically with small molecule antibiotics (52). By generating a virtual library of **G3KL** analogs and testing 2DP-nearest neighbors of **G3KL**, we identified AMPD **T7** exhibiting enhanced serum stability and a broader activity spectrum (Figure 2C) (53). Interestingly, AMPD **T7** corresponds to minor sequence changes at the dendrimer core compared to **G3KL**, which were thought to be negligible by design but turned out to have a major impact on antimicrobial activity.

We recently implemented molecular fingerprint similarity as a fitness function in a GA called PDGA (peptide design genetic algorithm) capable of generating peptides resembling any target molecule of choice. Because the molecular fingerprint can be computed for any molecule of interest, PDGA can generate peptide analogs of both peptides and nonpeptides. Furthermore, PDGA operates with diverse peptide topologies including linear, cyclic, or polycyclic peptides as well as peptide dendrimers. In a typical PDGA run, all generated sequences above a defined molecular fingerprint similarity threshold are identified as analogs. In a proof-of-principle computation, we showed that PDGA generates known analogs of the cyclic peptide tyrocidine A and peptide dendrimer **G3KL** (54).

6. Visualizing the peptide chemical space

The ability to compute similarities between peptides allows representing the peptide chemical space in the form of maps in which distances represent similarities. Such maps provide an overview that helps to perceive the structural diversity of peptides. In our first implementation of this approach, we created an interactive map of the Protein Data Bank chemical space based on computed 3D-shape similarities (55). However, this representation was only applicable to macromolecules with known 3D structures such as those in the Protein Data Bank.

To represent peptide structural diversity in a general context, we have used the molecular shape similarity fingerprint used above with PDGA to compute similarities between molecules featured in the non-Lipinski part of the ChEMBL and PubChem databases. These non-Lipinski subsets comprise 376,504 respectively 15,798,352 entries, 16% respectively 7% of which contain a dipeptide substructure (56). These similarity comparisons can be represented in interactive 3D maps displayed using Faerun (57), in which each molecule appears as a point color-coded by a property of choice, and its structure is displayed using Smilesdrawer (58).

More recently, we created a high-resolution molecular fingerprint called MAP4 useful to analyze diverse molecular classes spanning from small molecule drugs to metabolites, natural products, and macromolecules including peptides, oligonucleotides, and complex carbohydrates (59). The MAP4 fingerprint can be used in combination with the TMAP mapping tool (60) to create insightful representations of molecular databases, as recently shown for the case of the Natural Product Atlas (61).

For the present review, we have collected bioactive peptides from eleven publicly accessible databases that cover a wide range of size and scope (Table 1) (62–72). We considered 40,531 database entries corresponding to sequences of between 2 and 50 natural amino acids, calculated their SMILES representation using RDKit (73), and used this data to compute a TMAP based on the MAP4 fingerprint. This map organizes peptides by their size and sequence (Figure 3A, https://tm.gdb.tools/map4/peptide_databases_tmap/). The map colored by source activity type illustrates that the largest fraction of peptides in these databases (17,260 sequences, 43% of the total) are annotated as antimicrobial and anticancer, and stem from the DBAASP, DRAMP, AVPdb, and the antimicrobial and anticancer sections of the SATPdb (Figure 3B and online map). Color-coding the map by the number of databases in which a peptide is listed shows that most peptides (60%) occur only in one database, while 11% are present in two databases, 3.4% in three, 1.4% in four, and less than 1% in five databases (Figure 3C). Color-coding by database shows that several databases tend to cover specific regions of the peptide chemical space, which is not surprising for activity-specific databases such as SPdb in which sequences have limited diversity, but somewhat surprising for the peptides retrieved from PDB (Figure 3D).

7. Conclusion and outlook

In this review, we presented computational approaches to explore the peptide chemical space. Structure-based designs are well-suited when detailed information exists on the targeted site of action. On the other hand, GAs have broader applicability since they can be used to design peptide sequences even if the targeted activity is not defined by a structure but more generally by a set of properties. ML methods are similarly broad in their applicability but require a large number of known peptides with documented activity to enable model training. Finally, molecular fingerprints can be used to guide the sampling of large virtual peptide libraries as well as the optimization of known actives. Molecular fingerprints furthermore allow to compute maps that facilitate a global understanding of the peptide chemical space. Most interestingly, GAs, ML, and molecular fingerprint-based approaches are possible without detailed knowledge of the peptide 3D-structure and allow to explore diverse peptide chain topologies, also incorporating nonnatural amino acids. Such computational methods can play an enabling role in expanding the reach of peptides for therapeutic applications.

CRediT author statement

AC and JLR conceptualized and wrote the paper.

Funding source

Swiss National Science Foundation, Grant no. 200020_178998.

Declaration of competing interest

The authors declare that they do not have any conflicts of interest.

Acknowledgments

This work was supported financially by the Swiss National Science Foundation, Grant no. 200020_178998.

References

[1] Sakula A, Langerhans Paul. A centenary tribute. J R Soc Med. 1847-1888;81(1988):414–5.
[2] Kingsberg SA, Clayton AH, Portman D, Williams LA, Krop J, Jordan R, et al. Bremelanotide for the treatment of hypoactive sexual desire disorder. Obstet Gynecol. 2019;134:899–908. https://doi.org/10.1097/AOG.0000000000003500.
[3] Al Shaer D, Al Musaimi O, Albericio F, de la Torre BG. 2019 FDA TIDES (peptides and oligonucleotides) harvest. Pharmaceuticals. 2020;13:40. https://doi.org/10.3390/ph13030040.
[4] de la Torre BG, Albericio F. The Pharmaceutical Industry in 2019. An analysis of FDA drug approvals from the perspective of molecules. Molecules. 2020;25:745. https://doi.org/10.3390/molecules25030745.
[5] Lam KS. Affinity selection and sequencing. Nat Chem Biol. 2019;15:320–1. https://doi.org/10.1038/s41589-019-0253-2.
[6] Fjell CD, Hiss JA, Hancock REW, Schneider G. Designing antimicrobial peptides: form follows function. Nat Rev Drug Discov. 2012;11:37–51. https://doi.org/10.1038/nrd3591.
[7] Mansbach RA, Travers T, McMahon BH, Fair JM, Gnanakaran S. Snails in silico: a review of computational studies on the Conopeptides. Mar Drugs. 2019;17:145. https://doi.org/10.3390/md17030145.
[8] Torres MDT, Sothiselvam S, Lu TK, de la Fuente-Nunez C. Peptide design principles for antimicrobial applications. J Mol Biol. 2019. https://doi.org/10.1016/j.jmb.2018.12.015.
[9] Mulligan VK. The emerging role of computational design in peptide macrocycle drug discovery. Expert Opin Drug Discovery. 2020;15:833–52. https://doi.org/10.1080/17460441.2020.1751117.
[10] Lin X, Li X, Lin X. A review on applications of computational methods in drug screening and design. Molecules. 2020;25:1375. https://doi.org/10.3390/molecules25061375.
[11] Nikiforovich GV. Computational molecular modeling in peptide drug design. Int J Pept Protein Res. 1994;44:513–31. https://doi.org/10.1111/j.1399-3011.1994.tb01140.x.
[12] Lee AC-L, Harris JL, Khanna KK, Hong J-H. A comprehensive review on current advances in peptide drug development and design. Int J Mol Sci. 2019;20:2383. https://doi.org/10.3390/ijms20102383.

[13] Han Y, Král P. Computational design of ACE2-based peptide inhibitors of SARS-CoV-2. ACS Nano. 2020;14:5143–7. https://doi.org/10.1021/acsnano.0c02857.
[14] Sevy AM, Gilchuk IM, Brown BP, Bozhanova NG, Nargi R, Jensen M, et al. Computationally designed cyclic peptides derived from an antibody loop increase breadth of binding for influenza variants. Structure. 2020;28:1114–23 e4 https://doi.org/10.1016/j.str.2020.04.005.
[15] Korendovych IV, DeGrado WF. De novo protein design, a retrospective. Q Rev Biophys. 2020;53. https://doi.org/10.1017/S0033583519000131.
[16] Cao L, Goreshnik I, Coventry B, Case JB, Miller L, Kozodoy L, et al. De novo design of picomolar SARS-CoV-2 miniprotein inhibitors. Science. 2020;370:426–31. https://doi.org/10.1126/science.abd9909.
[17] Leman JK, Weitzner BD, Lewis SM, Adolf-Bryfogle J, Alam N, Alford RF, et al. Macromolecular modeling and design in Rosetta: recent methods and frameworks. Nat Methods. 2020;17:665–80. https://doi.org/10.1038/s41592-020-0848-2.
[18] Dou J, Vorobieva AA, Sheffler W, Doyle LA, Park H, Bick MJ, et al. De novo design of a fluorescence-activating β-barrel. Nature. 2018;561:485–91. https://doi.org/10.1038/s41586-018-0509-0.
[19] Kang S-M, Moon H, Han S-W, Kim D-H, Kim BM, Lee B-J. Structure-based De novo Design of *M. tuberculosis* VapC-activating stapled peptides. ACS Chem Biol. 2020;15:2493–8. https://doi.org/10.1021/acschembio.0c00492.
[20] Zhang S-Q, Huang H, Yang J, Kratochvil HT, Lolicato M, Liu Y, et al. Designed peptides that assemble into cross-α amyloid-like structures. Nat Chem Biol. 2018;14:870–5. https://doi.org/10.1038/s41589-018-0105-5.
[21] Holland JH. Genetic algorithms. Sci Am. 1992;267:66–73.
[22] Teixidó M, Belda I, Roselló X, González S, Fabre M, Llorá X, et al. Development of a genetic algorithm to design and identify peptides that can cross the blood-brain barrier. QSAR Comb Sci. 2003;22:745–53. https://doi.org/10.1002/qsar.200320004.
[23] Beltran JA, Brizuela CA. Design of selective cationic antibacterial peptides: a multiobjective genetic algorithm approach. 2016 IEEE Congr. Evol. Comput. CEC. Vancouver, BC, Canada: IEEE; 2016. p. 484–91. https://doi.org/10.1109/CEC.2016.7743833.
[24] Porto WF, Irazazabal L, Alves ESF, Ribeiro SM, Matos CO, Pires ÁS, et al. In silico optimization of a guava antimicrobial peptide enables combinatorial exploration for peptide design. Nat Commun. 2018;9:1490. https://doi.org/10.1038/s41467-018-03746-3.
[25] Knapp B, Giczi V, Ribarics R, Schreiner W. PeptX: using genetic algorithms to optimize peptides for MHC binding. BMC Bioinform. 2011;12:241. https://doi.org/10.1186/1471-2105-12-241.
[26] King MD, Long T, Andersen T, McDougal OM. Genetic algorithm managed peptide mutant screening: optimizing peptide ligands for targeted receptor binding. J Chem Inf Model. 2016;56:2378–87. https://doi.org/10.1021/acs.jcim.6b00095.
[27] Fjell CD, Jenssen H, Cheung WA, Hancock REW, Cherkasov A. Optimization of antibacterial peptides by genetic algorithms and cheminformatics. Chem Biol Drug Des. 2011;77:48–56. https://doi.org/10.1111/j.1747-0285.2010.01044.x.
[28] Yoshida M, Hinkley T, Tsuda S, Abul-Haija YM, McBurney RT, Kulikov V, et al. Using evolutionary algorithms and machine learning to explore sequence space for the discovery of antimicrobial peptides. Chem. 2018;4:533–43. https://doi.org/10.1016/j.chempr.2018.01.005.
[29] Neuhaus CS, Gabernet G, Steuer C, Root K, Hiss JA, Zenobi R, et al. Simulated molecular evolution for anticancer peptide design. Angew Chem Int Ed. 2019;58:1674–8. https://doi.org/10.1002/anie.201811215.
[30] Basith S, Manavalan B, Shin TH, Lee G. Machine intelligence in peptide therapeutics: a next-generation tool for rapid disease screening. Med Res Rev. 2020;40:1276–314. https://doi.org/10.1002/med.21658.
[31] Aranha MP, Spooner C, Demerdash O, Czejdo B, Smith JC, Mitchell JC. Prediction of peptide binding to MHC using machine learning with sequence and structure-based feature sets. Biochim Biophys Acta BBA - Gen Subj. 1864;2020:129535. https://doi.org/10.1016/j.bbagen.2020.129535.
[32] Zhang YP, Zou Q. PPTPP: a novel therapeutic peptide prediction method using physicochemical property encoding and adaptive feature representation learning. Bioinformatics. 2020;36:3982–7. https://doi.org/10.1093/bioinformatics/btaa275.
[33] Lee EY, Fulan BM, Wong GCL, Ferguson AL. Mapping membrane activity in undiscovered peptide sequence space using machine learning. Proc Natl Acad Sci. 2016;113:13588–93. https://doi.org/10.1073/pnas.1609893113.
[34] Spänig S, Heider D. Encodings and models for antimicrobial peptide classification for multi-resistant pathogens. BioData Min. 2019;12:7. https://doi.org/10.1186/s13040-019-0196-x.
[35] Plisson F, Ramírez-Sánchez O, Martínez-Hernández C. Machine learning-guided discovery and design of non-hemolytic peptides. Sci Rep. 2020;10:16581. https://doi.org/10.1038/s41598-020-73644-6.
[36] Cherkasov A, Hilpert K, Jenssen H, Fjell CD, Waldbrook M, Mullaly SC, et al. Use of artificial intelligence in the Design of Small Peptide Antibiotics Effective against a broad Spectrum of highly antibiotic-resistant superbugs. ACS Chem Biol. 2009;4:65–74. https://doi.org/10.1021/cb800240j.
[37] Timmons PB, Hewage CM. HAPPENN is a novel tool for hemolytic activity prediction for therapeutic peptides which employs neural networks. Sci Rep. 2020;10:10869. https://doi.org/10.1038/s41598-020-67701-3.
[38] Gautam A, Chaudhary K, Singh S, Joshi A, Anand P, Tuknait A, et al. Hemolytik: a database of experimentally determined hemolytic and non-hemolytic peptides. Nucleic Acids Res. 2014;42:D444–9. https://doi.org/10.1093/nar/gkt1008.
[39] Müller AT, Hiss JA, Schneider G. Recurrent neural network model for constructive peptide design. J Chem Inf Model. 2018;58:472–9. https://doi.org/10.1021/acs.jcim.7b00414.
[40] Grisoni F, Merk D, Consonni V, Hiss JA, Tagliabue SG, Todeschini R, et al. Scaffold hopping from natural products to synthetic mimetics by holistic molecular similarity. Commun Chem. 2018;1:1–9. https://doi.org/10.1038/s42004-018-0043-x.
[41] Tucs A, Tran DP, Yumoto A, Ito Y, Uzawa T, Tsuda K. Generating ampicillin-level antimicrobial peptides with activity-aware generative adversarial networks. ACS Omega. 2020;5:22847–51. https://doi.org/10.1021/acsomega.0c02088.
[42] Scior T, Bender A, Tresadern G, Medina-Franco JL, Martinez-Mayorga K, Langer T, et al. Recognizing pitfalls in virtual screening: a critical review. J Chem Inf Model. 2012;52:867–81. https://doi.org/10.1021/ci200528d.
[43] Bonaventura ID, Jin X, Visini R, Probst D, Javor S, Gan B-H, et al. Chemical space guided discovery of antimicrobial bridged bicyclic peptides against *P. aeruginosa* and its biofilms. Chem Sci. 2017;8:6784–98. https://doi.org/10.1039/C7SC01314K.
[44] Bonaventura ID, Baeriswyl S, Capecchi A, Gan B-H, Jin X, Siriwardena TN, et al. An antimicrobial bicyclic peptide from chemical space against multidrug resistant gram-negative bacteria. Chem Commun. 2018;54:5130–3. https://doi.org/10.1039/C8CC02412J.
[45] Nguyen LT, Haney EF, Vogel HJ. The expanding scope of antimicrobial peptide structures and their modes of action. Trends Biotechnol. 2011;29:464–72. https://doi.org/10.1016/j.tibtech.2011.05.001.
[46] Stach M, Siriwardena TN, Kohler T, van Delden C, Darbre T, Reymond JL. Combining topology and sequence design for the discovery of potent antimicrobial peptide dendrimers against multidrug-resistant *P. aeruginosa*. Angew Chem Int Ed Engl. 2014;53:12827–31. https://doi.org/10.1002/anie.201409270.
[47] Pires J, Siriwardena TN, Stach M, Tinguely R, Kasraian S, Luzzaro F, et al. In vitro activity of the novel antimicrobial peptide dendrimer G3KL against multidrug-resistant *A. baumannii* and *P. aeruginosa*. Antimicrob Agents Chemother. 2015;59:7915–8. https://doi.org/10.1128/aac.01853-15.
[48] Gan B-H, Siriwardena TN, Javor S, Darbre T, Reymond J-L. Fluorescence imaging of bacterial killing by antimicrobial peptide dendrimer G3KL. ACS Infect Dis. 2019;5:2164–73. https://doi.org/10.1021/acsinfecdis.9b00299.
[49] Jeddou FB, Falconnet L, Luscher A, Siriwardena T, Reymond J-L, van Delden C, et al. Adaptive and mutational responses to peptide dendrimer antimicrobials in *P. aeruginosa*. Antimicrob Agents Chemother. 2020;64. https://doi.org/10.1128/AAC.02040-19.
[50] Abdel-Sayed P, Kaeppeli A, Siriwardena T, Darbre T, Perron K, Jafari P, et al. Antimicrobial dendrimers against multidrug-resistant *P. aeruginosa* enhance the angiogenic effect of biological burn-wound bandages. Sci Rep. 2016;6:1–10. https://doi.org/10.1038/srep22020.
[51] Han X, Liu Y, Ma Y, Zhang M, He Z, Siriwardena TN, et al. Peptide dendrimers G3KL and TNS18 inhibit *P. aeruginosa* biofilms. Appl Microbiol Biotechnol. 2019;103:5821–30. https://doi.org/10.1007/s00253-019-09801-3.
[52] Gan B-H, Cai X, Javor S, Köhler T, Reymond J-L. Synergistic effect of Propidium iodide and small molecule antibiotics with the antimicrobial peptide dendrimer G3KL against gram-negative Bacteria. Molecules. 2020;25:5643. https://doi.org/10.3390/molecules25235643.
[53] Siriwardena TN, Capecchi A, Gan B-H, Jin X, He R, Wei D, et al. Optimizing Antimicrobial Peptide Dendrimers in Chemical Space. Angew Chem Int Ed. 2018;57:8483–7. https://doi.org/10.1002/anie.201802837.
[54] Capecchi A, Zhang A, Reymond J-L. Populating chemical space with peptides using a genetic algorithm. J Chem Inf Model. 2019. https://doi.org/10.1021/acs.jcim.9b01014.
[55] Jin X, Awale M, Zasso M, Kostro D, Patiny L, Reymond JL. PDB-explorer: a web-based interactive map of the protein data bank in shape space. BMC Bioinform. 2015;16:339. https://doi.org/10.1186/s12859-015-0776-9.
[56] Capecchi A, Awale M, Probst D, Reymond J-L. PubChem and ChEMBL beyond Lipinski. Mol Inform. 2019. https://doi.org/10.1002/minf.201900016.
[57] Probst D, Reymond J-L, Wren J. FUn: a framework for interactive visualizations of large, high-dimensional datasets on the web. Bioinformatics. 2018;34:1433–5. https://doi.org/10.1093/bioinformatics/btx760.
[58] Probst D, Reymond J-L. SmilesDrawer: parsing and drawing SMILES-encoded molecular structures using client-side JavaScript. J Chem Inf Model. 2018;58:1–7. https://doi.org/10.1021/acs.jcim.7b00425.
[59] Capecchi A, Probst D, Reymond J-L. One molecular fingerprint to rule them all: drugs, biomolecules, and the metabolome. J Chem. 2020;12:43. https://doi.org/10.1186/s13321-020-00445-4.
[60] Probst D, Reymond J-L. Visualization of very large high-dimensional data sets as minimum spanning trees. J Chem. 2020;12:12. https://doi.org/10.1186/s13321-020-0416-x.
[61] Capecchi A, Reymond J-L. Assigning the origin of microbial natural products by chemical space map and machine learning. Biomolecules. 2020;10:1385. https://doi.org/10.3390/biom10101385.
[62] Berman HM, Westbrook J, Feng Z, Gilliland G, Bhat TN, Weissig H, et al. The protein data bank. Nucleic Acids Res. 2000;28:235–42.
[63] T.U. Consortium. UniProt: a worldwide hub of protein knowledge. Nucleic Acids Res. 2019;47:D506–15. https://doi.org/10.1093/nar/gky1049.
[64] Singh S, Chaudhary K, Dhanda SK, Bhalla S, Usmani SS, Gautam A, et al. SATPdb: a database of structurally annotated therapeutic peptides. Nucleic Acids Res. 2016;44:D1119–26. https://doi.org/10.1093/nar/gkv1114.
[65] Pirtskhalava M, Gabrielian A, Cruz P, Griggs HL, Squires RB, Hurt DE, et al. DBAASP v.2: an enhanced database of structure and antimicrobial/cytotoxic activity of natural and synthetic peptides. Nucleic Acids Res. 2016;44:D1104–12. https://doi.org/10.1093/nar/gkv1174.
[66] Kang X, Dong F, Shi C, Liu S, Sun J, Chen J, et al. DRAMP 2.0, an updated data repository of antimicrobial peptides. Sci Data. 2019;6. https://doi.org/10.1038/s41597-019-0154-y.
[67] Qureshi A, Thakur N, Tandon H, Kumar M. AVPdb: a database of experimentally validated antiviral peptides targeting medically important viruses. Nucleic Acids Res. 2014;42:D1147–53. https://doi.org/10.1093/nar/gkt1191.
[68] Choo KH, Tan TW, Ranganathan S. SPdb – a signal peptide database. BMC Bioinform. 2005;6:249. https://doi.org/10.1186/1471-2105-6-249.

[69] Kim Y, Bark S, Hook V, Bandeira N. NeuroPedia: neuropeptide database and spectral library. Bioinform Oxf Engl. 2011;27:2772–3. https://doi.org/10.1093/bioinformatics/btr445.
[70] Novković M, Simunić J, Bojović V, Tossi A, Juretić D. DADP: the database of anuran defense peptides. Bioinform Oxf Engl. 2012;28:1406–7. https://doi.org/10.1093/bioinformatics/bts141.
[71] Wynendaele E, Bronselaer A, Nielandt J, D'Hondt M, Stalmans S, Bracke N, et al. Quorumpeps database: chemical space, microbial origin and functionality of quorum sensing peptides. Nucleic Acids Res. 2013;41:D655–9. https://doi.org/10.1093/nar/gks1137.
[72] Ettayapuram Ramaprasad AS, Singh S, Gajendra PSR, Venkatesan S. AntiAngioPred: a server for prediction of anti-angiogenic peptides. PLoS One. 2015;10:e0136990. https://doi.org/10.1371/journal.pone.0136990.
[73] RDKit. https://www.rdkit.org/. (accessed September 25, 2018).