

# Molecular Similarity for Drug Discovery, Target Prediction and Chemical Space Visualization

Jean-Louis Reymond\*

**Abstract:** Similar drug molecules often have similar properties and activities. Therefore, quantifying molecular similarity is central to drug discovery and optimization. Here I review computational methods using molecular similarity measures developed in my group within the interdisciplinary network NCCR TransCure investigating the physiology, structural biology and pharmacology of ion channels and membrane transporters. We designed a 3D molecular shape and pharmacophore comparison algorithm to optimize weak and unselective inhibitors by scaffold hopping and discovered potent and selective inhibitors of the ion channels TRPV6 and TRPM4, of endocannabinoid membrane transport, and of the divalent metal transporters DMT1 and ZIP8. We predicted off-target effects by combining molecular similarity searches from different molecular fingerprints against target annotated compounds from the ChEMBL database. Finally, we created interactive chemical space maps reflecting molecular similarities to facilitate the selection of screening compounds and the analysis of screening results. These different tools are available online at <https://gdb.unibe.ch/tools/>.

**Keywords:** Chemical space · Drug discovery · Molecular similarity · Polypharmacology · Virtual screening



**Jean-Louis Reymond** is Professor of Chemistry and Chemical Biology at the University of Bern, Switzerland, and was Deputy Director of the NCCR TransCure (2010–2022). He studied Chemistry and Biochemistry at the ETH Zurich and obtained his PhD at the University of Lausanne on natural products synthesis (1989). After a Post-Doc and Assistant Professorship at the Scripps Research Institute, he joined the

University of Bern (1997). His current research focuses on chemical space exploration methods combining data science, artificial intelligence, organic synthesis, and biological assays to discover new bioactive small molecules and peptides, with applications in membrane transporters modulators, kinase inhibitors, antimicrobial compounds and transfection reagents (<https://gdb.unibe.ch>). He is the author of >300 scientific publications and reviews.

## 1. Introduction

The biological activity of small molecule drugs primarily depends on their complementarity to their targets, usually proteins, following the well-known lock-and-key principle, implying that similar drug molecules often have similar biological activities.<sup>[1,2]</sup> A broad variety of computational methods can be used to predict either the complementarity of potential drug molecules to their protein target, typically by docking, or the molecular similarity between potential drug molecules, most often using molecular fingerprints.<sup>[3–5]</sup> Although only moderately reliable, such predictions are extremely useful in drug discovery to focus experiments on small sets of test compounds, for example by selecting them from commercial catalogs of millions of screening compounds such as those accessible at the ZINC database,<sup>[6]</sup> and to guide hit and lead optimization by analog design.<sup>[7]</sup> Computational models can also predict physico-chemical<sup>[8]</sup> and ADME-Tox<sup>[9,10]</sup> properties as well as possible off-targets,<sup>[11–13]</sup> typically through multiple com-

parisons with drug molecules of known properties and activities as collected in databases such as ChEMBL and PubChem.<sup>[14,15]</sup> Here I review molecular similarity measures developed in my group for virtual screening, target prediction and chemical space visualization, and how they were used to discover and optimize inhibitors of ion channels and membrane transporters within the NCCR TransCure project.

## 2. Drug Discovery

Our virtual screening approach for TransCure was designed in the context of identifying a potent and selective inhibitor of TRPV6, a calcium channel initially discovered by Matthias Hediger and a putative cancer target for which no structural information was available, but for which several weak inhibitors ( $IC_{50} \sim 100 \mu M$ ) had been described.<sup>[16–20]</sup> Our hypothesis was that these weak inhibitors must be partly correct but their scaffold was probably not the right one. Among the many options to compute molecular similarity,<sup>[2]</sup> we focused on molecular shape and pharmacophores, which is one of the most useful concepts in drug discovery because it enables scaffold-hopping and the discovery of new compound series.<sup>[21,22]</sup> Considering the weak inhibitors available as seed molecules, this approach appeared as a good option to identify different and potentially better scaffolds for this target.

The gold standard for molecular shape and pharmacophore comparison, then and now, is the ROCS score developed by the software company Open Eye.<sup>[23]</sup> The ROCS algorithm searches for the best overlay between 3D-molecular models of a seed and a query molecule in different conformations, quantifying the overlay of molecular surfaces either in terms of pure shape, or in terms of pharmacophore by assigning properties to the molecular surface. We developed a simpler version of this approach comparing only the lowest energy conformers of seed and query predicted by the 3D-builder CORINA<sup>[24]</sup> based on a modified version of the ligand overlap score.<sup>[25]</sup> Our extended ligand overlap score (xLOS) was calculated from the relative positions of hydrophobic, hydrogen-bond acceptor and hydrogen-bond donor atoms as three separate categories and was minimized by optimizing the relative position of seed and query. Benchmarking using a reference set<sup>[26]</sup> showed that xLOS performed as good as ROCS and better than a 3D-pharmacophore fingerprint<sup>[27]</sup> or a simple Daylight binary

\*Correspondence: Prof. Dr. J.-L. Reymond, E-mail: [jean-louis.reymond@unibe.ch](mailto:jean-louis.reymond@unibe.ch)  
Department of Chemistry, Biochemistry and Pharmaceutical Sciences,  
University of Bern, Freiestrasse 3, CH-3012 Bern, Switzerland

substructure fingerprint<sup>[5]</sup> in a virtual screening benchmark, and performed significant scaffold-hopping (Fig. 1).<sup>[28]</sup>

Starting with the chemical structure of five weak TRPV6 inhibitors described in the literature,<sup>[18]</sup> we used xLOS similarity to select 133 test compounds in the catalog of ~800,000 drug-like molecules of a commercial provider. The activity assays were performed in the group of Matthias Hediger by monitoring calcium uptake using a fluorogenic calcium dye in a stably transfected HEK293 cell line expressing TRPV6,<sup>[29]</sup> and revealed five hit compounds with  $IC_{50}$  ~20  $\mu$ M. In particular inhibitor **8** ( $IC_{50}$  = 31  $\mu$ M) from a (4-phenylcyclohexyl)piperazine scaffold had a high xLOS similarity to the prolinol type seed inhibitor **1** ( $IC_{50}$  = 90  $\mu$ M). Further optimization by purchasing an additional 90 analogs, again selected by xLOS similarity, confirmed the activity of this scaffold, and provided a first potent inhibitor, which turned out to be most active as the 1,4-*cis* cyclohexane diastereomer *cis*-**11a** ( $IC_{50}$  =  $1.0 \pm 0.11$   $\mu$ M). This diastereoselective inhibition was confirmed with additional analogs obtained by synthesis, which led to the optimized inhibitor *cis*-**22a** ( $IC_{50}$  =  $0.32 \pm 0.12$   $\mu$ M).<sup>[30]</sup> Further studies focusing on electrophysiology and on introducing structural features from the natural product capsaicin to mitigate off-target effects led to the photoswitchable *Z*-**9e** ( $IC_{50}$  =  $1.7 \pm 0.4$   $\mu$ M) only active as the *Z*-isomer,<sup>[31]</sup> and to the more potent and selective TRPV6 inhibitor **30G** ( $IC_{50}$  =  $0.082 \pm 0.004$   $\mu$ M) (Fig. 2).<sup>[32]</sup> A collaboration with Alexander Sobolevsky and his

group allowed to solve X-ray and cryo-EM structures of several TRPV6 inhibitor complexes including *cis*-**22a** (PDB 7K4B) and **30G** (PDB 7K4D). These different structures showed that our inhibitors plug the open pore of TRPV6 and convert the channel into a nonconducting, inactive state.<sup>[33]</sup>

In a similar approach in collaboration with Hugues Abriel, we searched for selective inhibitors of TRPM4, a monovalent cation channel involved in heart diseases and cancer.<sup>[34,35]</sup> We tested xLOS analogs of known cation channel blockers such as flufenamic acid<sup>[36]</sup> using a fluorescent sodium uptake assay for TRPM4 in transfected HEK293 cells. These studies led to the identification of **CBA** ( $IC_{50}$  =  $1.5 \pm 0.1$   $\mu$ M) and **NBA** ( $IC_{50}$  =  $0.4 \pm 0.3$   $\mu$ M) as the first potent and selective TRPM4 inhibitors,<sup>[37,38]</sup> with interesting species specific differences in their activities.<sup>[39]</sup> We have also used xLOS in virtual screening campaigns to optimize inhibitors of the iron transporter DMT1 such as the pyrazolyl-pyrimidine **3** ( $IC_{50}$  = 0.17  $\mu$ M),<sup>[40,41]</sup> however, in this case detailed studies showed that the literature seed compound as well as our inhibitor **13** ( $IC_{50}$  =  $1.1 \pm 0.01$   $\mu$ M) acted indirectly by iron chelation, leaving bis-thioureas such as **2** ( $IC_{50}$  = 0.29  $\mu$ M) as the only class of potent inhibitors for DMT1 as confirmed by X-ray crystallographic studies.<sup>[42,43]</sup>

X-ray crystallographic studies were also critical in confirming the binding mode of Aurora A kinase inhibitors **6** ( $IC_{50}$  = 2.3  $\mu$ M, PDB 4ZS0) and **9** ( $IC_{50}$  =  $2 \pm 0.5$  nM, PDB 4ZTR) discovered by an xLOS driven screening campaign.<sup>[44]</sup> In this study, we found

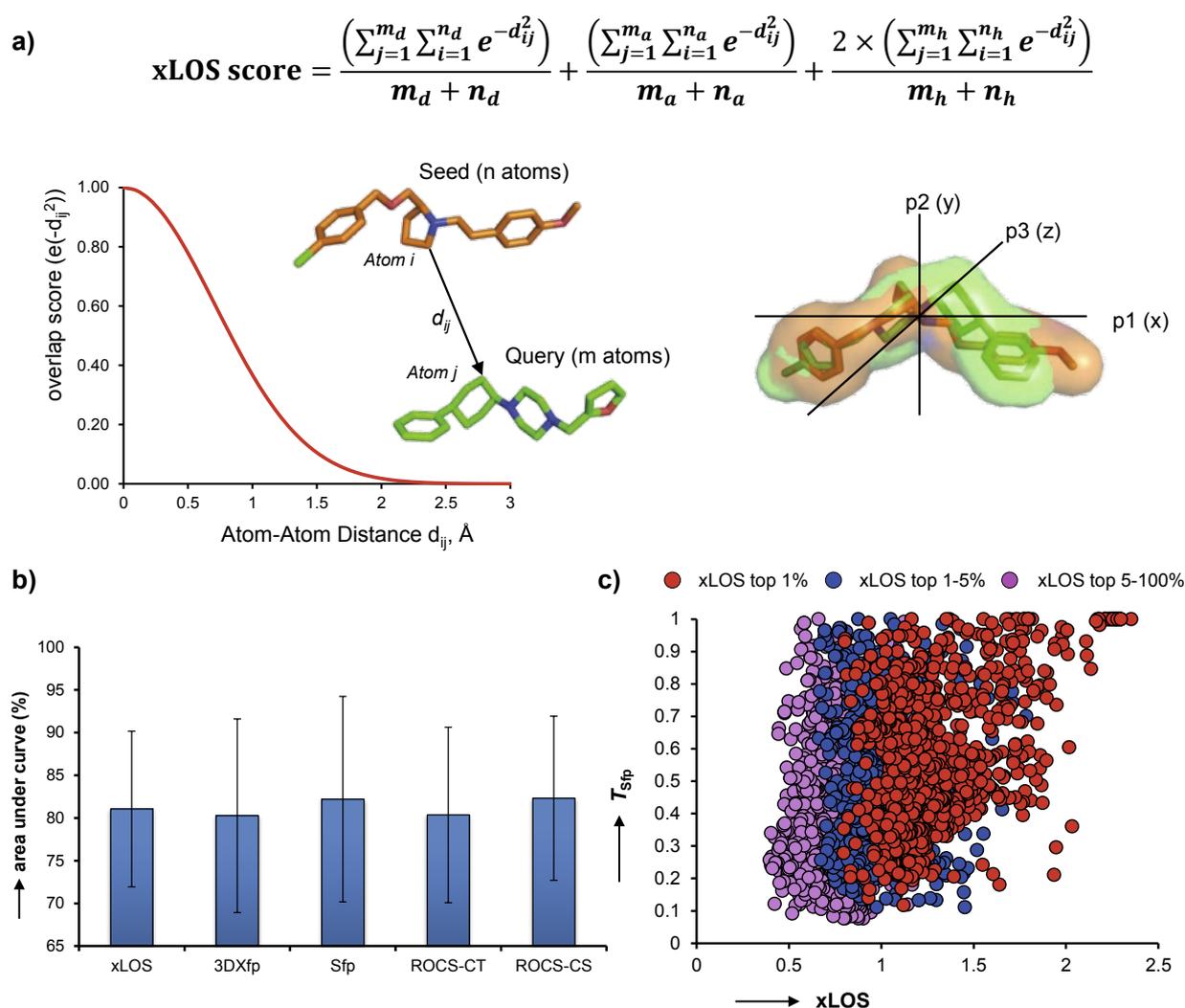


Fig. 1. The ligand overlap score xLOS for similarity between a seed and a query molecule. a) Equation and principle of comparing seed and query molecule by scoring atom-pair distances using xLOS. b) Benchmarking study comparing xLOS with ROCS, a Daylight substructure fingerprint (Sfp) and a 3D-pharmacophore fingerprint (3DXfp). c) Scaffold hopping of xLOS is demonstrated in the DUD-E benchmarking set by the fact that the xLOS score (x-axis) is only very weakly correlated with the Tanimoto similarity of substructure fingerprint (y-axis).

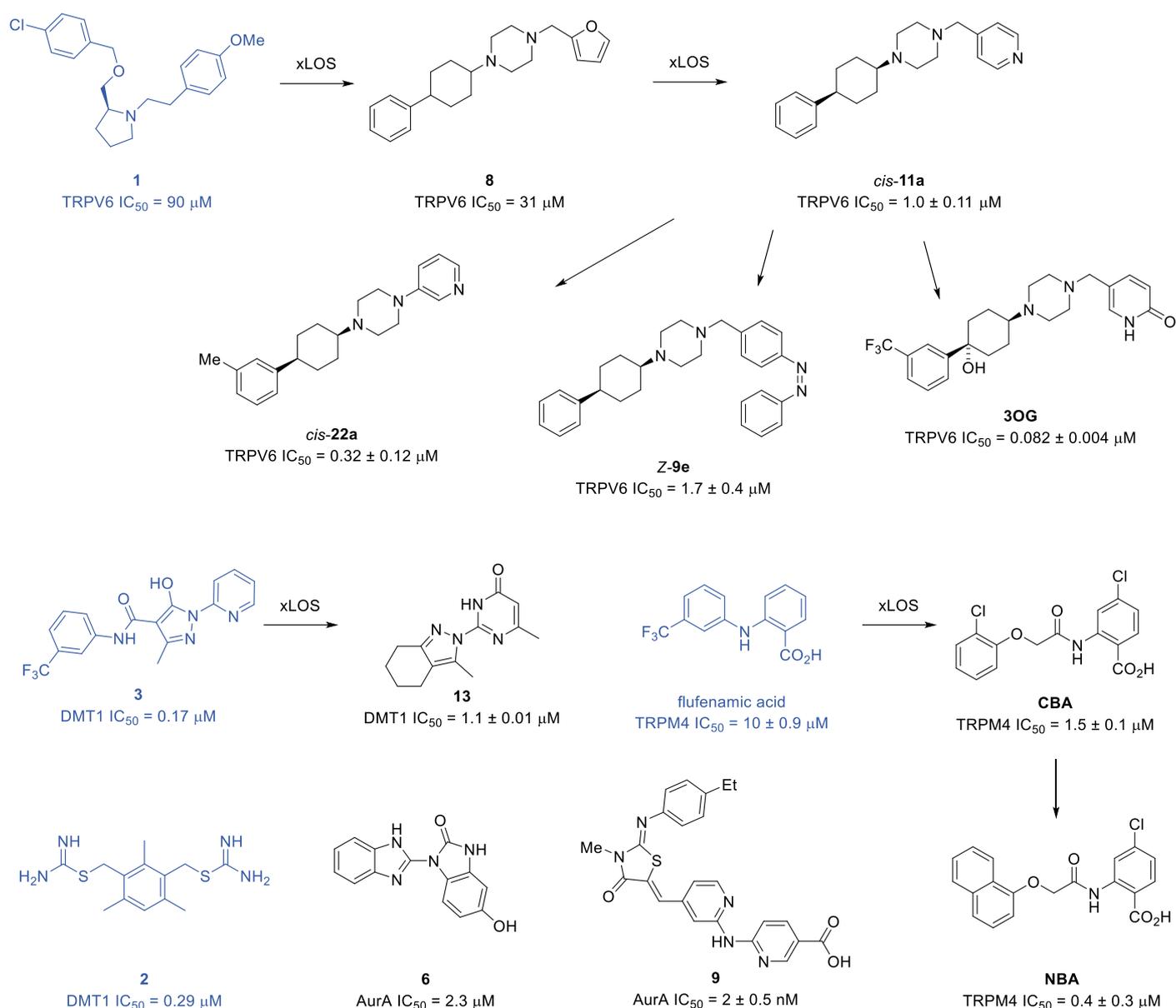


Fig. 2. Seed compounds (blue) and inhibitors discovered in screening programs initiated by virtual screening using the shape and pharmacophore similarity algorithm xLOS. Compound numbering of the original publications was preserved.

that the thiazolidinone nucleus of the potent and highly selective inhibitor **9** did not exhibit any significant thiol reactivity despite being flagged as a potential pan-assay interference substructure (PAINS).<sup>[45]</sup> A similar lack of thiol reactivity and a very low toxicity were observed during the preclinical development of a thiazolidinone series identified in collaboration with Jürg Gertsch in an xLOS-driven screening campaign searching for inhibitors of endocannabinoid transport,<sup>[46,47]</sup> which is being further developed by Synendos Therapeutics.<sup>[48]</sup>

### 3. Target Prediction

To be reasonably useful, small molecule modulators should not only be potent on their intended target, but also show good selectivity against other targets, including possible close isoforms as well as safety panels containing known problematic side activities.<sup>[49]</sup> This polypharmacology is usually unintended but very frequent and must be addressed as soon as possible in the course of a discovery project. Thanks to the availability of open access databases such as ChEMBL listing millions of compounds and their associated bioactivity data,<sup>[14]</sup> one can use molecular similarity to compare any hit compound with known bioactive molecules and deduce the probability of off-target

effects. Following early reports such as PASS,<sup>[50,51]</sup> SEA<sup>[52]</sup> and OCEAN,<sup>[53]</sup> numerous approaches exploiting various statistical methods have been reported, a few of which have been made accessible online.<sup>[12,13,54]</sup>

Most target prediction methods exploit compound–activity databases by measuring molecular similarity using a substructure molecular fingerprint such as the extended connectivity fingerprint ECFP4, originally developed in the 1960's as the Morgan fingerprint.<sup>[55,56]</sup> In our own approach to the problem, we used nearest-neighbor searches combining ECFP4 with several other molecular fingerprints comparing molecules by composition<sup>[57]</sup> and molecular shape and pharmacophores,<sup>[58]</sup> as well as fusion fingerprints.<sup>[59]</sup> The system, called polypharmacology browser (PPB),<sup>[60]</sup> searched for similarity of a query molecule to ChEMBL compounds annotated with any of 4,613 possible bioactivities on single protein targets, cell lines and organisms. When tested with our TRPV6 inhibitor *cis-22a*, PPB successfully predicted five of twelve experimentally verified off-targets, and none of twelve experimentally verified inactive off-targets, which was comparable to the performance of several other off-target prediction tools available online (Fig. 3a).

We later reported a second, modified polypharmacology browser PPB2, which used the activity data of 344,163 ChEMBL

compounds annotated with 1,720 different single protein targets.<sup>[61]</sup> In this second version, we used only three different fingerprints but investigated several models for target prediction including nearest neighbor (NN) searches with the different fingerprints, Naïve-Bayes (NB) machine learning and a deep neural network (DNN). The recall and precision of the off-target prediction was strongly influenced by the ECfp4 similarity of the query molecule to the target annotated ChEMBL compound, as well as by the target class. Among the different models, the DNN as well as a combination of NN and NB machine learning performed best. When tested with off-targets of *cis-22a*, PPB2 performed significantly better than PPB, with the DNN performing best by being able to predict eight of the twelve active off-target and none of the inactive ones (Fig. 3b).

Additional insight into the performance of PPB2 compared to other online prediction tools was provided by analyzing the details of a target identification project for a highly cytotoxic triazine series (IC<sub>50</sub> ~20 nM) identified in a phenotypic assay for angiogenesis inhibition and initially thought to target kinases.<sup>[62]</sup> In the absence of any measurable activity in whole kinome profiling, we followed a suggestion of PPB2 that this compound might inhibit

lymphosphatidic acid acyl transferase beta (LPAAT-β), a prediction which was verified experimentally. Interestingly, the key molecular similarity comparison in this project was made by XFP, a molecular shape and pharmacophore similarity fingerprint,<sup>[58]</sup> and was missed by other target prediction models.

#### 4. Visualizing Chemical Space

During applied projects such as those performed in the NCCR TransCure, the selection of test compounds guided by virtual screening engages substantial time and resources and determines project outcome. To enable better decision making in this critical early project phase, we have developed interactive tools to perform rapid similarity searches in large databases such as ZINC using various molecular fingerprints, and to visualize the content of the source database or the selected subsets in form of chemical space maps. In both cases our interactive tools display chemical structures, which is essential to evaluate and compare molecules. With such tools, various options for virtual screening can be considered and possible sets of test molecules can be compared against each other and against the entire source database.

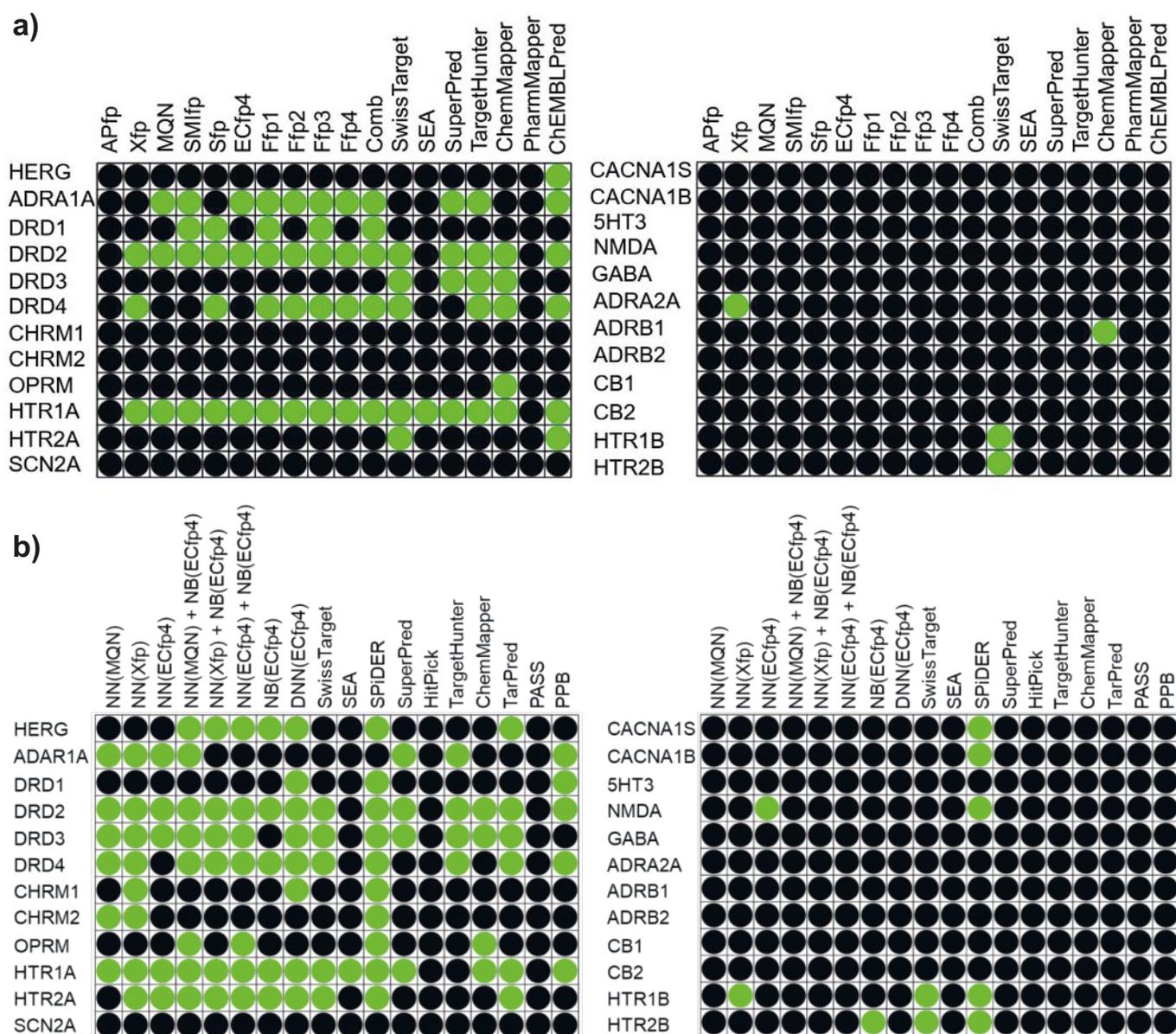


Fig. 3. Predicting off-targets of the TRPV6 inhibitor *cis-22a* (structure shown in Fig. 2) by molecular similarity. (a) eight experimentally verified active (left panel,  $\geq 50\%$  inhibition at  $10\ \mu\text{M}$ ) and inactive (right panel,  $< 50\%$  inhibition at  $10\ \mu\text{M}$ ) targets and active (green) predictions by various molecular fingerprints in PPB (first eight columns) and other online prediction tools. (b) Same as (a) for PPB2.

Our fastest and most versatile molecular similarity search tool is the multi-fingerprint browser for the ZINC database.<sup>[63]</sup> This tool offers a choice of four different molecular fingerprints to retrieve a list of analogs of any seed molecule, and to cluster the results by similarity. By contrast to typical online similarity searches such as those at the website of ChEMBL and PubChem which use only a single, substructure-based similarity method, our search tool additionally offers fuzzy comparisons using Molecular Quantum Numbers (MQN), a set of 42 descriptors counting different types of structural features,<sup>[57]</sup> and the SMILES fingerprint SMIPF, which counts the occurrence of 34 different characters in the SMILES notation of molecules.<sup>[64]</sup> Such fuzzy searches allow very substantial scaffold hopping<sup>[21]</sup> and can enable surprising and counter-intuitive analog discoveries.<sup>[65,66]</sup>

To visualize collections of molecules, often referred to as chemical spaces,<sup>[67,68]</sup> we produce 2D- or 3D-maps organized by molecular fingerprint similarity where each point corresponds to one or several molecules. Principal component analysis (PCA) is well suited to produce maps from MQN and SMIPF fingerprints, for which the first two principal components usually cover more than 70% of data variability. For binary substructure fingerprints such as ECFP4 which have very high dimensionality, we produce the maps by applying PCA to an N-dimensional secondary molecular fingerprint obtained by computing similarities to N different reference compounds using the original fingerprint, which results in a strong dimensionality reduction.<sup>[69,70]</sup> PCA and similarity PCA maps provide nice overviews of large molecular databases such as the generated databases (GDBs)<sup>[71,72]</sup> PubChem,<sup>[73]</sup> or the RCSB Protein DataBank,<sup>[74]</sup> and can be inspected with interactive Java applets in 2D<sup>[75]</sup> or 3D,<sup>[38,76,77]</sup> or with the application Fearun.<sup>[78]</sup> Fearun can display millions of datapoints in 3D within a web-browser, and represents molecules in millisecond without any server communication using the drawing program Smilesdrawer.<sup>[79]</sup> This visualization was adapted for a virtual reality 3D-chemical space map of DrugBank, a database containing thousands of small molecule drugs.<sup>[80]</sup>

Applying PCA and other dimensionality reduction methods such as t-SNE<sup>[81]</sup> and UMAP<sup>[82]</sup> to high-dimensional molecular fingerprint datasets usually creates images with uneven datapoint density, which hides the bulk of the data in dense clusters and draws attention to outliers. To overcome this limitation, we have reported the tree-map (TMAP), a dimensionality reduction method which organizes high-dimensional datasets in a 2D-tree connecting approximate nearest neighbors, such that the density of points is homogeneous over the entire map. TMAP largely outperforms UMAP in terms of computation speed, maximum dataset size, and preservation of nearest-neighbor relationships, and has been exemplified with datasets from computer science, physics, chemistry, biology, and literature.<sup>[83]</sup> For molecular datasets, we usually compute TMAPs using MinHashed molecular fingerprints encoding circular substructures (MHFP6)<sup>[84]</sup> or pairs of circular substructures and the topological distances between them (MAP4).<sup>[85]</sup> These molecular fingerprints outperform binary molecular fingerprints such as ECFP4 in bioactivity benchmarks<sup>[86]</sup> and allow fast determination of approximate nearest neighbors for TMAP construction. As for the PCA maps discussed above, TMAPs are displayed interactively in a web-browser using Faerun.

TMAPs representing various high-dimensional datasets are accessible at <https://tmap.gdb.tools/> and <https://tm.gdb.tools/map4/>. TMAPs can provide unsuspected insights into large datasets, as recently exemplified by the analysis of natural products, for which the TMAP showed that they separate in different groups according to their origin as from plants, bacteria or fungi, which implies that the origin of any natural product can be inferred from its molecular structure (<https://np-svm-map4.gdb.tools/>).<sup>[87,88]</sup> TMAPs provide useful project overviews, as exemplified here for a discovery project in collaboration with Matthias Hediger, targeting the divalent metal transporter ZIP8, a transporter implicated

in various diseases but for which no previous pharmacology was available.<sup>[89–91]</sup> In this project, we purchased a set of diverse fragment-like<sup>[92]</sup> molecules belonging to the fragment subset FDB-17<sup>[93]</sup> of the generated database GDB-17.<sup>[94]</sup> A limited screening using a transiently transfected cell line was sufficient to identify a hit series exemplified by the tricyclic indole (S)-3 (IC<sub>50</sub> = 17.2 ± 3.8 μM). A subsequent result analysis using TMAP showed that the screening set used for this discovery was not truly representative of the original FDB17 database and represented a selection typical of commercial fragment series, and provided useful hints for further screening efforts (Fig. 4).<sup>[95]</sup>

## 5. Conclusion and Outlook

This review presented methods developed in my research group to use molecular similarity for drug discovery, target prediction, and chemical space visualization and how they were used in NCCR TransCure projects for which no structural information was available on the protein targets. Transforming our computations into successful experiments depended on a well-organized pipeline to verify hit compound purity, structure and activity, on consistent follow-up studies, on the intrinsic hit-rate of the selected protein targets, and unavoidably also on chance. We currently keep improving our methods to address further challenges such as the discovery of bioactive peptides<sup>[95–98]</sup> and reaction informatics.<sup>[99–101]</sup>

## Acknowledgements

This work was supported financially by the Swiss National Science Foundation and was made possible by the many talented students who dedicated their ideas, calculations and experiments to the chemical space adventure within the interdisciplinary collaborative environment of the NCCR TransCure.

Received: October 4, 2022

- [1] C. G. Wermuth, *Drug Discovery Today* **2006**, *11*, 348, <https://doi.org/10.1016/j.drudis.2006.02.006>.
- [2] G. Maggiora, M. Vogt, D. Stumpfe, J. Bajorath, *J. Med. Chem.* **2014**, *57*, 3186, <https://doi.org/10.1021/jm401411z>.
- [3] P. Willett, *Drug Discovery Today* **2006**, *11*, 1046, <https://doi.org/10.1016/j.drudis.2006.10.005>.
- [4] P. Kolb, R. S. Ferreira, J. J. Irwin, B. K. Shoichet, *Curr. Opin. Biotechnol.* **2009**, *20*, 429, <https://doi.org/10.1016/j.copbio.2009.08.003>.
- [5] T. Scior, A. Bender, G. Tresadern, J. L. Medina-Franco, K. Martinez-Mayorga, T. Langer, K. Cuanalo-Contreras, D. K. Agrafiotis, *J. Chem. Inf. Model.* **2012**, *52*, 867, <https://doi.org/10.1021/ci200528d>.
- [6] J. J. Irwin, K. G. Tang, J. Young, C. Dandarchuluun, B. R. Wong, M. Khurelbaatar, Y. S. Moroz, J. Mayfield, R. A. Sayle, *J. Chem. Inf. Model.* **2020**, <https://doi.org/10.1021/acs.jcim.0c00675>.
- [7] S. Mignani, J. Rodrigues, H. Tomas, R. Jalal, P. P. Singh, J.-P. Majoral, R. A. Vishwakarma, *Drug Discovery Today* **2018**, *23*, 605, <https://doi.org/10.1016/j.drudis.2018.01.010>.
- [8] C. A. Lipinski, F. Lombardo, B. W. Dominy, P. J. Feeney, *Adv. Drug Deliv. Rev.* **1997**, *46*, 3, [https://doi.org/10.1016/s0169-409x\(00\)00129-0](https://doi.org/10.1016/s0169-409x(00)00129-0).
- [9] I. V. Tetko, P. Bruneau, H.-W. Mewes, D. C. Rohrer, G. I. Poda, *Drug Discovery Today* **2006**, *11*, 700, <https://doi.org/10.1016/j.drudis.2006.06.013>.
- [10] I. Sushko, E. Salmina, V. A. Potemkin, G. Poda, I. V. Tetko, *J. Chem. Inf. Model.* **2012**, *52*, 2310, <https://doi.org/10.1021/ci300245q>.
- [11] A. Koutsoukas, R. Lowe, Y. KalantarMotamedi, H. Y. Mussa, W. Klaffke, J. B. O. Mitchell, R. C. Glen, A. Bender, *J. Chem. Inf. Model.* **2013**, *53*, 1957, <https://doi.org/10.1021/ci300435j>.
- [12] A. Mayr, G. Klambauer, T. Unterthiner, M. Steijaert, J. K. Wegner, H. Ceulemans, D.-A. Clevert, S. Hochreiter, *Chem. Sci.* **2018**, *9*, 5441, <https://doi.org/10.1039/c8sc00148k>.
- [13] M. Awale, J. L. Reymond, *Methods Mol. Biol.* **2019**, *1888*, 255, [https://doi.org/10.1007/978-1-4939-8891-4\\_15](https://doi.org/10.1007/978-1-4939-8891-4_15).
- [14] A. Gaulton, A. Hersey, M. Nowotka, A. P. Bento, J. Chambers, D. Mendez, P. Mutowo, F. Atkinson, L. J. Bellis, E. Cibrián-Uhalte, M. Davies, N. Dedman, A. Karlsson, M. P. Magariños, J. P. Overington, G. Papadatos, I. Smit, A. R. Leach, *Nucleic Acids Res.* **2017**, *45*, D945, <https://doi.org/10.1093/nar/gkw1074>.
- [15] A. Gindulyte, B. A. Shoemaker, B. Yu, J. He, J. Zhang, J. Chen, L. Zaslavsky, P. A. Thiessen, Q. Li, S. He, S. Kim, T. Cheng, E. E. Bolton, *Nucleic Acids Res.* **2018**, *47*, D1102, <https://doi.org/10.1093/nar/gky1033>.

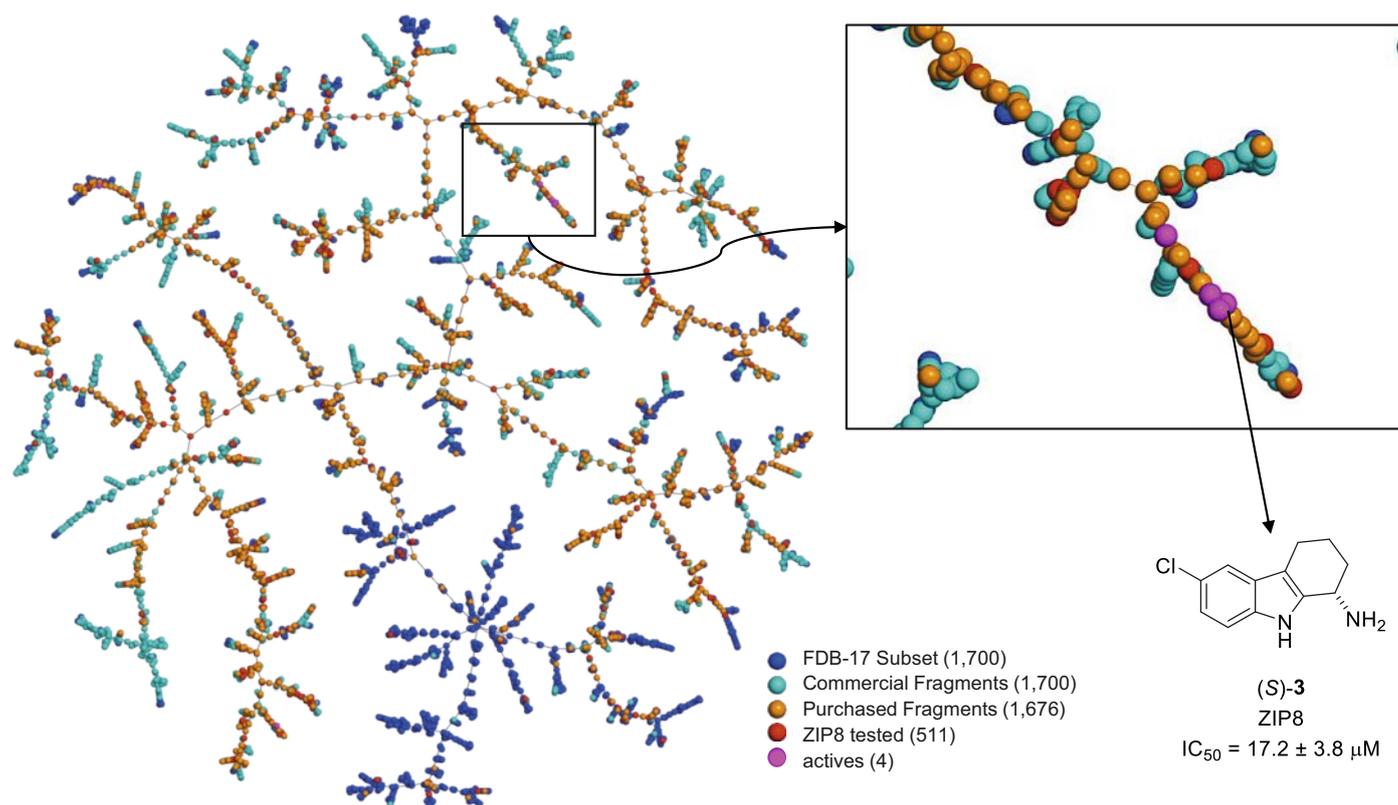


Fig. 4. TMAP visualization of a screening project leading to the discovery of (S)-3 as the first inhibitor of the divalent metal transporter ZIP8. The TMAP illustrates the relative chemical space coverage of the tested fragment set in the chemical space covered by fragments from different sources. The interactive map is accessible at <https://gdb.unibe.ch/tools/>.

- [16] J. B. Peng, X. Z. Chen, U. V. Berger, P. M. Vassilev, H. Tsukaguchi, E. M. Brown, M. A. Hediger, *J. Biol. Chem.* **1999**, *274*, 22739, <https://doi.org/10.1074/jbc.274.32.22739>.
- [17] L. Yue, J. B. Peng, M. A. Hediger, D. E. Clapham, *Nature* **2001**, *410*, 705, <https://doi.org/10.1038/35070596>.
- [18] W. F. McCalmont, T. N. Heady, J. R. Patterson, M. A. Lindenmuth, D. M. Haverstick, L. S. Gray, T. L. Macdonald, *Bioorg. Med. Chem. Lett.* **2004**, *14*, 3691, <https://doi.org/10.1016/j.bmcl.2004.05.011>.
- [19] K. A. Bolanz, M. A. Hediger, C. P. Landowski, *Mol. Cancer Ther.* **2008**, *7*, 271, <https://doi.org/10.1158/1535-7163.mct-07-0478>.
- [20] C. Fecher-Trost, P. Weissgerber, U. Wissenbach, *Handb. Exper. Pharmacol.* **2014**, *222*, 359, [https://doi.org/10.1007/978-3-642-54215-2\\_14](https://doi.org/10.1007/978-3-642-54215-2_14).
- [21] G. Schneider, W. Neidhart, T. Giller, G. Schmid, *Angew. Chem., Int. Ed. Engl.* **1999**, *38*, 2894, [https://doi.org/10.1002/\(SICI\)1521-3773\(19991004\)38:19<2894::AID-ANIE2894>3.0.CO;2-F](https://doi.org/10.1002/(SICI)1521-3773(19991004)38:19<2894::AID-ANIE2894>3.0.CO;2-F).
- [22] A. Nicholls, G. B. McGaughey, R. P. Sheridan, A. C. Good, G. Warren, M. Mathieu, S. W. Muchmore, S. P. Brown, J. A. Grant, J. A. Haigh, N. Nevins, A. N. Jain, B. Kelley, *J. Med. Chem.* **2010**, *53*, 3862, <https://doi.org/10.1021/jm900818s>.
- [23] P. C. Hawkins, A. G. Skillman, A. Nicholls, *J. Med. Chem.* **2007**, *50*, 74, <https://doi.org/10.1021/jm0603365>.
- [24] J. Sadowski, J. Gasteiger, *Chem. Rev.* **1993**, *93*, 2567, <https://doi.org/10.1021/cr00023a012>.
- [25] E. A. Kennewell, P. Willett, P. Ducrot, C. Luttmann, *J. Comput.-Aided Mol. Des.* **2006**, *20*, 385, <https://doi.org/10.1007/s10822-006-9072-0>.
- [26] M. M. Mysinger, M. Carchia, J. J. Irwin, B. K. Shoichet, *J. Med. Chem.* **2012**, *55*, 6582, <https://doi.org/10.1021/jm300687e>.
- [27] M. Awale, X. Jin, J. L. Reymond, *J. Cheminf.* **2015**, *7*, 3, <https://doi.org/10.1186/s13321-014-0051-5>.
- [28] C. Simonin, M. Awale, M. Brand, R. van Deursen, J. Schwartz, M. Fine, G. Kovacs, P. Hafliger, G. Gyimesi, A. Sithampari, R. P. Charles, M. A. Hediger, J. L. Reymond, *Angew. Chem., Int. Ed. Engl.* **2015**, *54*, 14748, <https://doi.org/10.1002/anie.201507320>.
- [29] N. Montalbetti, A. Simonin, M. G. Dalghi, G. Kovacs, M. A. Hediger, *J. Biomol. Screen.* **2014**, *19*, 900, <https://doi.org/10.1177/1087057114521663>.
- [30] C. Humer, S. Lindinger, A. L. Carrel, C. Romanin, C. Höglinger, *Biomolecules* **2022**, *12*, 804, <https://doi.org/10.3390/biom12060804>.
- [31] M. R. Cunha, R. Bhardwaj, S. Lindinger, C. Butorac, C. Romanin, M. A. Hediger, J.-L. Reymond, *ACS Med. Chem. Lett.* **2019**, *10*, 1341, <https://doi.org/10.1021/acsmchemlett.9b00298>.
- [32] M. R. Cunha, R. Bhardwaj, A. L. Carrel, S. Lindinger, C. Romanin, R. Parise-Filho, M. A. Hediger, J.-L. Reymond, *RSC Med. Chem.* **2020**, *11*, 1032, <https://doi.org/10.1039/D0MD00145G>.
- [33] R. Bhardwaj, S. Lindinger, A. Neuberger, K. D. Nadezhdin, A. K. Singh, M. R. Cunha, I. Derler, G. Gyimesi, J.-L. Reymond, M. A. Hediger, C. Romanin, A. I. Sobolevsky, *Science Adv.* **2020**, *6*, eabe1508, <https://doi.org/10.1126/sciadv.abe1508>.
- [34] C. Wang, K. Naruse, K. Takahashi, *Cells* **2018**, *7*, <https://doi.org/10.3390/cells7060062>.
- [35] A. Borgström, C. Peinelt, P. Stoklosa, *Biomolecules* **2021**, *11*, 229, <https://doi.org/10.3390/biom11020229>.
- [36] K. E. Gardam, J. E. Geiger, C. M. Hickey, A. Y. Hung, N. S. Magoski, *J. Neurophysiol.* **2008**, *100*, 38, <https://doi.org/10.1152/jn.90265.2008>.
- [37] L. C. Ozthathil, C. Delalande, B. Bianchi, G. Nemeth, S. Kappel, U. Thomet, D. Ross-Kaschitza, C. Simonin, M. Rubin, J. Gertsch, M. Lochner, C. Peinelt, J. L. Reymond, H. Abriel, *Br. J. Pharmacol.* **2018**, *175*, 2504, <https://doi.org/10.1111/bph.14220>.
- [38] C. Delalande, M. Awale, M. Rubin, D. Probst, L. C. Ozthathil, J. Gertsch, H. Abriel, J.-L. Reymond, *Eur. J. Med. Chem.* **2019**, *166*, 167, <https://doi.org/10.1016/j.ejmech.2019.01.048>.
- [39] P. Arullampalam, B. Preti, D. Ross-Kaschitza, M. Lochner, J.-S. Rougier, H. Abriel, *Front. Pharmacol.* **2021**, *12*, <https://doi.org/10.3389/fphar.2021.712354>.
- [40] J. A. Cadieux, Z. Zhang, M. Mattice, A. Brownlie-Cutts, J. Fu, L. G. Ratkay, R. Kwan, J. Thompson, J. Sanghara, J. Zhong, Y. P. Goldberg, *Bioorg. Med. Chem. Lett.* **2012**, *22*, 90, <https://doi.org/10.1016/j.bmcl.2011.11.069>.
- [41] N. Montalbetti, A. Simonin, C. Simonin, M. Awale, J.-L. Reymond, M. A. Hediger, *Biochem. Pharmacol.* **2015**, *96*, 216, <https://doi.org/10.1016/j.bcp.2015.05.002>.
- [42] Z. Zhang, V. Kodumuru, S. Sviridov, S. Liu, M. Chafeev, S. Chowdhury, N. Chakka, J. Sun, S. J. Gauthier, M. Mattice, L. G. Ratkay, R. Kwan, J. Thompson, A. B. Cutts, J. Fu, R. Kamboj, Y. P. Goldberg, J. A. Cadieux, *Bioorg. Med. Chem. Lett.* **2012**, *22*, 5108, <https://doi.org/10.1016/j.bmcl.2012.05.129>.
- [43] C. Manatschal, J. Pujol-Giménez, M. Poirier, J.-L. Reymond, M. A. Hediger, *Rutzler, eLife* **2019**, *8*, e51913, <https://doi.org/10.7554/eLife.51913>.
- [44] F. Kilchmann, M. J. Marcaida, S. Kotak, T. Schick, S. D. Boss, M. Awale, P. Gönczy, J.-L. Reymond, *J. Med. Chem.* **2016**, *59*, 7188, <https://doi.org/10.1021/acs.jmedchem.6b00709>.
- [45] S. J. Capuzzi, E. N. Muratov, A. Tropsha, *J. Chem. Inf. Model.* **2017**, *57*, 417, <https://doi.org/10.1021/acs.jcim.6b00465>.

- [46] A. Chicca, S. Nicolussi, R. Bartholomäus, M. Blunder, A. Aparisi Rey, V. Petrucci, I. del C. Reynoso-Moreno, J. M. Viveros-Paredes, M. Dalghi Gens, B. Lutz, H. B. Schiöth, M. Soeberdt, C. Abels, R.-P. Charles, K.-H. Altmann, J. Gertsch, *Proc. Natl. Acad. Sci. U. S. A.* **2017**, *114*, E5006, <https://doi.org/10.1073/pnas.1704065114>.
- [47] R. Bartholomäus, S. Nicolussi, A. Baumann, M. Rau, A. C. Simão, J. Gertsch, K.-H. Altmann, *ChemMedChem* **2019**, *14*, 1590, <https://doi.org/10.1002/cmde.201900390>.
- [48] J.-L. Reymond, S. Nicolussi, J. Gertsch, US10774077B2, **2020**.
- [49] S. Bendels, C. Bissantz, B. Fasching, G. Gerebtzoff, W. Guba, M. Kansy, J. Migeon, S. Mohr, J.-U. Peters, F. Tillier, R. Wyler, C. Lerner, C. Kramer, H. Richter, S. Roberts, *J. Pharmacol. Toxicol. Methods* **2019**, *99*, 106609, <https://doi.org/10.1016/j.vascn.2019.106609>.
- [50] A. Lagunin, A. Stepanchikova, D. Filimonov, V. Poroikov, *Bioinformatics* **2000**, *16*, 747, <https://doi.org/10.1093/bioinformatics/16.8.747>.
- [51] P. V. Pogodin, A. A. Lagunin, D. A. Filimonov, V. V. Poroikov, *SAR and QSAR Environ. Res.* **2015**, *26*, 783, <https://doi.org/10.1080/1062936x.2015.1078407>.
- [52] M. J. Keiser, B. L. Roth, B. N. Armbruster, P. Ernsberger, J. J. Irwin, B. K. Shoichet, *Nat. Biotechnol.* **2007**, *25*, 197, <https://doi.org/10.1038/nbt1284>.
- [53] P. Czodrowski, W.-G. Bolick, *J. Chem. Inf. Model.* **2016**, *56*, 2013, <https://doi.org/10.1021/acs.jcim.6b00067>.
- [54] D. Gfeller, A. Grosdidier, M. Wirth, A. Daina, O. Michielin, V. Zoete, *Nucleic Acids Res.* **2014**, *42*, W32, <https://doi.org/10.1093/nar/gku293>.
- [55] H. L. Morgan, *J. Chem. Doc.* **1965**, *5*, 107, <https://doi.org/10.1021/c160017a018>.
- [56] D. Rogers, M. Hahn, *J. Chem. Inf. Model.* **2010**, *50*, 742, <https://doi.org/10.1021/ci100050t>.
- [57] K. T. Nguyen, L. C. Blum, R. van Deursen, J.-L. Reymond, *ChemMedChem* **2009**, *4*, 1803, <https://doi.org/10.1002/cmde.200900317>.
- [58] M. Awale, J. L. Reymond, *J. Chem. Inf. Model.* **2014**, *54*, 1892, <https://doi.org/10.1021/ci500232g>.
- [59] N. Salim, J. Holliday, P. Willett, *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 435, <https://doi.org/10.1021/ci025596j>.
- [60] M. Awale, J. L. Reymond, *J. Cheminf.* **2017**, *9*, 11, <https://doi.org/10.1186/s13321-017-0199-x>.
- [61] M. Awale, J.-L. Reymond, *J. Chem. Inf. Model.* **2019**, *59*, 10, <https://doi.org/10.1021/acs.jcim.8b00524>.
- [62] M. Poirier, M. Awale, M. A. Roelli, G. T. Giuffredi, L. Ruddigkeit, L. Evensen, A. Stooss, S. Calarco, J. B. Lorens, R.-P. Charles, J.-L. Reymond, *ChemMedChem* **2019**, *14*, 224, <https://doi.org/10.1002/cmde.201800554>.
- [63] M. Awale, J. L. Reymond, *Nucleic Acids Res.* **2014**, *42*, W234, <https://doi.org/10.1093/nar/gku379>.
- [64] J. Schwartz, M. Awale, J.-L. Reymond, *J. Chem. Inf. Model.* **2013**, *53*, 1979, <https://doi.org/10.1021/ci400206h>.
- [65] L. C. Blum, R. van Deursen, S. Bertrand, M. Mayer, J. J. Bürgi, D. Bertrand, J.-L. Reymond, *J. Chem. Inf. Model.* **2011**, *51*, 3105, <https://doi.org/10.1021/ci200410u>.
- [66] J. J. Bürgi, M. Awale, S. D. Boss, T. Schaer, F. Marger, J. M. Viveros-Paredes, S. Bertrand, J. Gertsch, D. Bertrand, J. L. Reymond, *ACS Chem. Neurosci.* **2014**, *5*, 346, <https://doi.org/10.1021/cn4002297>.
- [67] R. S. Bohacek, C. McMartin, W. C. Guida, *Med. Res. Rev.* **1996**, *16*, 3, [https://doi.org/10.1002/\(SICI\)1098-1128\(199601\)16:1<3::AID-MED1>3.0.CO;2-6](https://doi.org/10.1002/(SICI)1098-1128(199601)16:1<3::AID-MED1>3.0.CO;2-6).
- [68] A. Mullard, *Nature* **2017**, *549*, 445, <https://doi.org/10.1038/549445a>.
- [69] J. L. Medina-Franco, G. M. Maggiora, M. A. Giulianotti, C. Pinilla, R. A. Houghten, *Chem. Biol. Drug Des.* **2007**, *70*, 393, <https://doi.org/10.1111/j.1747-0285.2007.00579.x>.
- [70] M. Awale, J. L. Reymond, *J. Chem. Inf. Model.* **2015**, *55*, 1509, <https://doi.org/10.1021/acs.jcim.5b00182>.
- [71] L. C. Blum, R. van Deursen, J. L. Reymond, *J. Comput.-Aided Mol. Des.* **2011**, *25*, 637, <https://doi.org/10.1007/s10822-011-9436-y>.
- [72] L. Ruddigkeit, L. C. Blum, J. L. Reymond, *J. Chem. Inf. Model.* **2013**, *53*, 56, <https://doi.org/10.1021/ci300535x>.
- [73] R. van Deursen, L. C. Blum, J. L. Reymond, *J. Chem. Inf. Model.* **2010**, *50*, 1924, <https://doi.org/10.1021/ci100237q>.
- [74] X. Jin, M. Awale, M. Zasso, D. Kostro, L. Patiny, J. L. Reymond, *BMC Bioinformatics* **2015**, *16*, 339, <https://doi.org/10.1186/s12859-015-0776-9>.
- [75] M. Awale, R. van Deursen, J. L. Reymond, *J. Chem. Inf. Model.* **2013**, *53*, 509, <https://doi.org/10.1021/ci300513m>.
- [76] M. Awale, J. L. Reymond, *J. Cheminf.* **2016**, *8*, 25, <https://doi.org/10.1186/s13321-016-0138-2>.
- [77] M. Awale, D. Probst, J. L. Reymond, *J. Chem. Inf. Model.* **2017**, *57*, 643, <https://doi.org/10.1021/acs.jcim.6b00690>.
- [78] D. Probst, J.-L. Reymond, *Bioinformatics* **2018**, *34*, 1433, <https://doi.org/10.1093/bioinformatics/btx760>.
- [79] D. Probst, J.-L. Reymond, *J. Chem. Inf. Model.* **2018**, *58*, 1, <https://doi.org/10.1021/acs.jcim.7b00425>.
- [80] D. Probst, J.-L. Reymond, *J. Chem. Inf. Model.* **2018**, *58*, 1731, <https://doi.org/10.1021/acs.jcim.8b00402>.
- [81] L. Maaten, G. Hinton, *J. Mach. Learn. Res.* **2008**, *9*, 2579.
- [82] L. McInnes, J. Healy, J. Melville, arXiv:1802.03426 [cs, stat] **2018**.
- [83] D. Probst, J.-L. Reymond, *J. Cheminf.* **2020**, *12*, 12, <https://doi.org/10.1186/s13321-020-0416-x>.
- [84] D. Probst, J.-L. Reymond, *J. Cheminf.* **2018**, *10*, 66, <https://doi.org/10.1186/s13321-018-0321-8>.
- [85] A. Capecchi, D. Probst, J.-L. Reymond, *J. Cheminf.* **2020**, *12*, 43, <https://doi.org/10.1186/s13321-020-00445-4>.
- [86] S. Riniker, G. A. Landrum, *J. Cheminf.* **2013**, *5*, 26, <https://doi.org/10.1186/1758-2946-5-26>.
- [87] A. Capecchi, J.-L. Reymond, *Biomolecules* **2020**, *10*, 1385, <https://doi.org/10.3390/biom10101385>.
- [88] A. Capecchi, J.-L. Reymond, *J. Cheminf.* **2021**, *13*, 82, <https://doi.org/10.1186/s13321-021-00559-3>.
- [89] H. Fujishiro, S. Himeno, *Biol. Pharm. Bull.* **2019**, *42*, 1076, <https://doi.org/10.1248/bpb.b18-00637>.
- [90] S. Jenkitkasemwong, C.-Y. Wang, B. Mackenzie, M. D. Knutson, *Biomaterials* **2012**, *25*, 643, <https://doi.org/10.1007/s10534-012-9526-x>.
- [91] C.-Y. Wang, S. Jenkitkasemwong, S. Duarte, B. K. Sparkman, A. Shawk, B. Mackenzie, M. D. Knutson, *J. Biol. Chem.* **2012**, *287*, 34032, <https://doi.org/10.1074/jbc.M112.367284>.
- [92] A. K. S. Romasanta, P. van der Sijde, I. Hellsten, R. E. Hubbard, G. M. Keseru, J. van Muijlwijk-Koezen, I. J. P. de Esch, *Drug Discovery Today* **2018**, *23*, 1596, <https://doi.org/10.1016/j.drudis.2018.05.004>.
- [93] R. Visini, M. Awale, J. L. Reymond, *J. Chem. Inf. Model.* **2017**, *57*, 700, <https://doi.org/10.1021/acs.jcim.7b00020>.
- [94] L. Ruddigkeit, R. van Deursen, L. C. Blum, J. L. Reymond, *J. Chem. Inf. Model.* **2012**, *52*, 2864, <https://doi.org/10.1021/ci300415d>.
- [95] J. Pujol-Giménez, M. Poirier, S. Bühlmann, C. Schuppisser, R. Bhardwaj, M. Awale, R. Visini, S. Javor, M. A. Hediger, J.-L. Reymond, *ChemMedChem* **2021**, *16*, 3306, <https://doi.org/10.1002/cmde.202100467>.
- [96] A. Capecchi, J.-L. Reymond, *Med. Drug Discovery* **2021**, *9*, 100081, <https://doi.org/10.1016/j.medidd.2021.100081>.
- [97] A. Capecchi, X. Cai, H. Personne, T. Köhler, C. van Delden, J.-L. Reymond, *Chem. Sci.* **2021**, *12*, 9221, <https://doi.org/10.1039/D1SC01713F>.
- [98] E. Zakharova, M. Orsi, A. Capecchi, J.-L. Reymond, *ChemMedChem* **2022**, *e202200291*, <https://doi.org/10.1002/cmde.202200291>.
- [99] P. Schwaller, D. Probst, A. C. Vaucher, V. H. Nair, D. Kreutter, T. Laino, J.-L. Reymond, *Nat. Mach. Intell.* **2021**, *3*, 144, <https://doi.org/10.1038/s42256-020-00284-w>.
- [100] D. Kreutter, P. Schwaller, J.-L. Reymond, *Chem. Sci.* **2021**, *12*, 8648, <https://doi.org/10.1039/D1SC02362D>.
- [101] D. Probst, P. Schwaller, J.-L. Reymond, *Digital Discovery* **2022**, *1*, 91, <https://doi.org/10.1039/D1DD00006C>.

#### License and Terms



This is an Open Access article under the terms of the Creative Commons Attribution License CC BY 4.0. The material may not be used for commercial purposes.

The license is subject to the CHIMIA terms and conditions: (<https://chimia.ch/chimia/about>).

The definitive version of this article is the electronic one that can be found at <https://doi.org/10.2533/chimia.2022.1045>