

# Generating Bioactive Natural Product-inspired Molecules with Machine Intelligence

Petra Schneider, Karl-Heinz Altmann, and Gisbert Schneider\*

**Abstract:** The computer-assisted design of new chemical entities has made a leap forward with the development of machine learning models for automated molecule generation. The overarching goal of this conceptual approach is to augment the creativity of medicinal chemists with a machine intelligence. In this Perspective we highlight prospective applications of “de novo” drug design and target prediction, aiming to generate natural product-inspired bioactive compounds from scratch. A virtual chemist transforms pharmacologically active natural products into new, easily synthesizable small molecules with desired properties and activity. Computational activity prediction and automated compound generation offer the possibility to systematically transfer the wealth of pharmaceutically active natural products to synthetic small molecule drug discovery. We present selected prospective examples and dare a forecast into the future of natural product-inspired drug discovery.

**Keywords:** Artificial intelligence · Drug discovery · Machine learning · Natural product



**Petra Schneider** received a doctorate in chemistry from the Technical University Berlin. She worked as a research scientist at Roche Diagnostics and Roche Pharmaceuticals in Basel. As a member of the Roche Global Chemistry Team, she introduced the term ‘frequent hitter’ to drug discovery. She co-founded several startup companies and is the CEO of inSili.com LLC, Zurich. Her current research at ETH

Zurich focuses on chemical databases, virtual compound screening, and applied machine learning for pharmaceutical discovery. As a member of the ETH President’s staff she manages institutional relationships and public-private partnerships.



**Karl-Heinz Altmann** has been a Professor of Pharmaceutical Sciences at the ETH Zurich since 2003. He studied chemistry at the University of Mainz, and holds a PhD degree in organic chemistry from the University of Basel. After post-doctoral research at Cornell University and the University of Lausanne, he joined Ciba-Geigy’s Central Research Laboratories in Basel. In 1997 he moved to Oncology

Research within Novartis Pharma AG, Basel. Prior to joining ETH, he was the acting Global Head of Chemistry of the Novartis Institutes for BioMedical Research (NIBR). Prof. Altmann’s research interests are at the interface between chemistry and biology, with a particular focus on the chemical synthesis and the biological and pharmacological profiling of biologically active natural products and their synthetic and semi-synthetic analogs. He is a co-founder of Tolremo Therapeutics AG. His research contributions received the Novartis Leading Scientist Award and the Paul-Ehrlich Prize of the Société Chimie Thérapeutique, France.



**Gisbert Schneider** is a full professor of Computer-Assisted Drug Design at the ETH Zurich and the Director of the Singapore-ETH Centre. His research group develops and applies learning methods for integration into the pharmaceutical sciences. He graduated in biochemistry from the Freie Universität Berlin. After several post-doctoral positions he joined Roche Pharmaceuticals in Basel and received

Habilitation from the Albert-Ludwigs University Freiburg. Prior to joining ETH in 2010, he held the Beilstein Endowed Chair of Chem- and Bioinformatics at Goethe-University Frankfurt. He is an elected Fellow of the University of Tokyo and a co-founder of several start-up companies. Prof. Schneider coined the term “scaffold hopping” and pioneered molecular design with artificial intelligence. His research contributions received the Herman Skolnik Award, the Proust Institute – Overton and Meyer Award for New Technologies in Drug Discovery, the Gmelin-Beilstein Commemorative Coin, and the Falling Walls Breakthrough of the Year Award.

## 1. Introduction

Revitalizing the concept of natural product-inspired drug discovery presents a unique opportunity for future healthcare. Evidently, natural products have traditionally inspired chemists and filled the industry pipelines.<sup>[1]</sup> Approximately half of the new drugs approved by the U.S. Food and Drug Administration (FDA) are of natural origin or natural product-derived synthetic compounds.<sup>[2]</sup> For example, marine natural products are of particular interest as anticancer and antiviral agents, with several approved drugs.<sup>[3]</sup> Similarly, certain edible algae have recently been identified as sources of potential anti-obesity agents.<sup>[4]</sup> However, despite their unrivalled appeal for pharmaceutical discovery as a source of inspiration, natural products as drugs have to some degree fallen into disgrace in the medicinal chemistry community because of supply problems (ecological sustainability), partially lengthy and costly total syntheses (economical sustainability), and an often

\*Correspondence: Prof. Dr. G. Schneider, E-mail: gisbert@ethz.ch, Dept. Chemistry and Applied Sciences, ETH Zurich, CH-8093 Zurich, Switzerland

unknown molecular mode of action (scientific sustainability).<sup>[5]</sup> Advances in computer-assisted drug discovery, spurred in part by stellar method development in both machine learning (as a subdomain of ‘artificial intelligence’) and biotechnology,<sup>[6,7]</sup> bear promise to overcome some of these limitations.<sup>[8,9]</sup> In this Perspective we highlight the ability of ligand-based machine learning approaches to generate new chemical structures from scratch by molecular ‘*de novo*’ design, and explain how these tools can be used to obtain natural product-inspired drug-like chemical entities (Fig. 1). Selected retrospective and prospective applications illustrate the feasibility of this innovative drug design concept as an amalgamation of traditional and modern medicinal chemistry.

## 2. Molecular *de novo* Design

Molecular ‘scaffold hopping’ aims to identify pairs of molecules that have markedly different chemical structures but share a certain function of interest, *e.g.* binding to the same macromolecular target.<sup>[10]</sup> *De novo* structure generation provides such sets of structurally diverse molecules that have certain features in common.<sup>[11]</sup> In fact, automated *de novo* molecule construction enables medicinal chemists to computationally access a virtually infinite chemical space. Scaffold hopping by *de novo* design seems particularly suited for finding structurally novel hit and lead compounds.<sup>[12]</sup> The approach ideally complements virtual chemical library screening,<sup>[13]</sup> but requires synthesizing the computer-generated molecules.

Molecular design with machine intelligence includes both rule-based and rule-free approaches.<sup>[14]</sup> Rule-based methods use sets of molecular building blocks and chemical transformations (*e.g.* virtual reaction schemes) for molecule construction. The DOGS (Design of Genuine Structures) algorithm belongs to this class of methods.<sup>[15]</sup> In contrast, rule-free ‘generative’ methods sample new molecules from a learned statistical distribution of the training data, which usually requires a large set of known molecules.<sup>[16]</sup> Most of the contemporary generative approaches build on deep neural networks.<sup>[17]</sup> Recurrent neural networks with long short-term memory as ‘chemical language models’,<sup>[18]</sup> variational autoencoders,<sup>[19]</sup> generative adversarial networks,<sup>[20]</sup> graph neural networks,<sup>[21]</sup> and various other deep learning architectures,<sup>[22]</sup> have been proposed for this purpose. Rule-free and rule-based methods alike have been successfully employed for prospective small molecule drug design, resulting in new bioactive compounds.<sup>[23]</sup> Importantly, both contemporary *de novo* methods deliver synthetically feasible molecular designs, thereby overcoming a stigma of earlier approaches.<sup>[24]</sup> In contrast to generative models, rule-based methods do not require large sets of training data. These methods are applicable when only a single active molecule (the ‘template’ molecule for ligand-based *de novo* design) or a model of the binding pocket (for structure-based *de novo* design) is known.

### 2.1 Generating Natural Product-inspired Molecules with Machine Intelligence

Computational *de novo* design is increasingly employed for automated small molecule construction, starting from mostly synthetic small molecules as design templates or as training data for

the underlying machine learning models. The first applications of generative deep learning (chemical language model) to natural products were published in 2018.<sup>[25]</sup> Several novel nuclear hormone receptor modulators with micromolar to nanomolar activity were obtained. Selected recent examples of prospective studies aiming to ‘scaffold hop’ from natural products to isofunctional synthetic molecules are compiled in Fig. 2. Of note, in each case, only a very small number of known bioactive natural products (one to six molecules) served as design template(s) for rule-based compound construction (Fig. 2a–c), or deep generative models (Fig. 2d), respectively.

Two independent design runs with DOGS software, taking either valerenic acid (**1**) or dehydroabietic acid (**2**) as design template, resulted in the identical computer-generated molecule **3**.<sup>[26]</sup> This observation highlights the scaffold hopping capability of rule-based molecular design. Target prediction with SPiDER software<sup>[27]</sup> predicted retinoid X receptor (RXR) agonism for the two natural products and the *de novo* molecule. Tetrahydroindole derivative **3** was obtained in a microwave-assisted two-step synthesis. Full dose-response analysis revealed low micromolar potency of the natural product templates and *de novo* generated compound **3** on all three RXR subtypes.

In an attempt to morph structurally more intricate natural products to small synthetic mimetics DOGS software was applied to the marine natural product marinopyrrole A (**4**), which led to computer generated molecule **5**.<sup>[28]</sup> Compound **5** was obtained in a two-step synthetic route, as suggested by the design software. Target prediction suggested cyclooxygenase-1 (COX-1) inhibition as a hitherto unknown activity of the anticancer agent marinopyrrole A and the *de novo* molecule. *In vitro* testing confirmed compounds **4** ( $IC_{50} = 7.7 \pm 3.9 \mu M$ ) and **5** ( $IC_{50} = 0.10 \pm 0.05 \mu M$ ) as direct COX-1 inhibitors. Of note, molecule **5** inhibited COX-1 with low-nanomolar potency in platelets ( $IC_{50} = 0.009 \pm 0.001 \mu M$ ). This *de novo* designed natural product mimetic preferentially inhibits the biosynthesis of COX-1-derived products in human platelets and monocytes. Compound **5** behaved similar to indomethacin with regard to COX-1 inhibition in platelets and showed greater COX-1 selectivity. The unique binding mode of this *de novo* generated compound was confirmed by X-ray structure determination of the ligand-enzyme complex. This example demonstrates the applicability of ligand-based molecular design to practical natural product-inspired medicinal chemistry.

In another study, the structurally complex natural product (–)-englerin A (**6**), a known inhibitor of transient receptor potential (TRP) cation channels, served as design template for DOGS.<sup>[29]</sup> Using two different computational scoring methods (pharmacophore-based and shape-based), two different *de novo* generated molecules were prioritized for synthesis. Compounds **7** and **8** were afforded in 3-step and 2-step synthetic protocols, respectively, as suggested by the software. Activity determination confirmed the natural product and the computer-generated molecules as potent inhibitors of TRPM8 ( $K_i = 0.2–0.3 \mu M$ ). Importantly, the chemical constitution of the natural product templates used for rule-based *de novo* design served as the only reference information for automated ligand-based molecule construction. This computational approach might, therefore, prove particularly useful in “low-data”

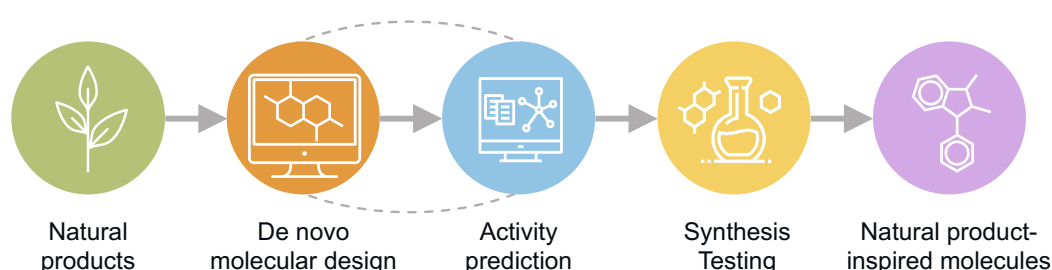


Fig. 1. A natural product-inspired *de novo* molecular design process. Certain machine learning methods integrate the chemical construction (design) and evaluation (activity prediction) of the molecular design into a single computational model (dashed line).

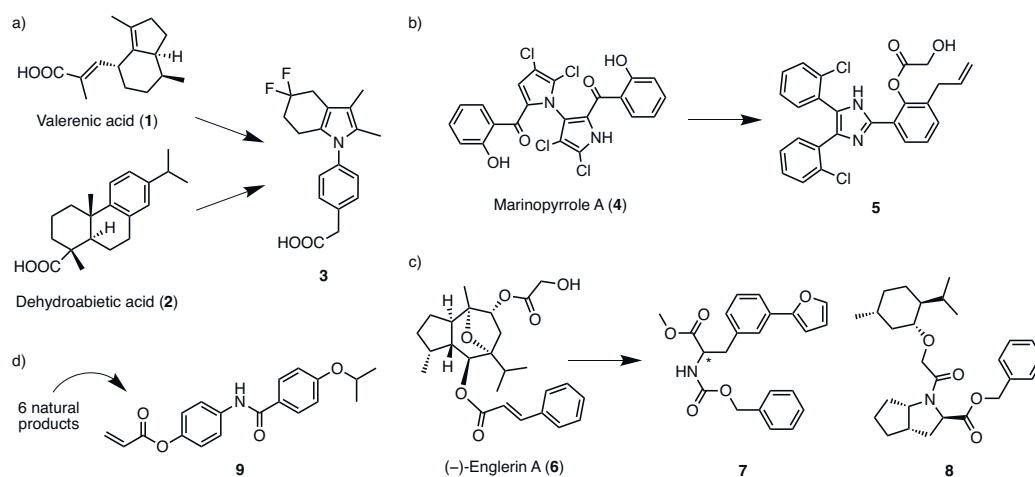


Fig. 2. Examples of natural product-inspired *de novo* design. The natural product templates (left side of the arrows) and the synthesized computer-generated compounds (right side of the arrows) represent scaffold-hops. Compound construction in (a–c) was performed with rule-based DOGS software. A generative chemical language model was used in example (d). Bioactivity: a) retinoid X receptor agonist (RXR $\alpha$  EC<sub>50</sub> = 0.97±0.07  $\mu$ M), b) cyclooxygenase-1 inhibitor (COX-1 IC<sub>50</sub> = 0.10±0.05  $\mu$ M), c) transient receptor potential melastatin 8 ion channel (TRPM8 K<sub>b</sub> = 0.2  $\mu$ M), d) *pan*-RXR agonist (RXR $\alpha$  EC<sub>50</sub> = 29±5  $\mu$ M, RXR $\beta$  EC<sub>50</sub> = 27±1  $\mu$ M, RXR $\gamma$  EC<sub>50</sub> = 19.1±0.1  $\mu$ M).

situations that are restrictive for *de novo* drug design with data-hungry generative deep learning.

Using a generative approach (chemical language model) for natural-product inspired molecular design, computer-generated compound **9** was obtained by 2-step synthesis, and experimentally confirmed as a weak *pan*-RXR partial agonist (EC<sub>50</sub> = 20–30  $\mu$ M, 6–10 fold receptor activation).<sup>[30]</sup> The importance of this result lies in the fact that only six natural products (known RXR agonists) were used to bias the generative chemical language model, and no explicit target prediction was performed. The machine intelligence implicitly captured the structural requirements of RXR agonists and ranked the *de novo* generated molecules.

## 2.2 Case Study: Designing Nucleoside Analogs with a Chemical Language Model

In light of the urgent need for novel antiviral therapeutics, *de novo* design may play a decisive role for future rapid hit and lead identification. To provide another example of generative *de novo* design, we targeted RNA-dependent RNA polymerase (RdRp) of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2),<sup>[31]</sup> aiming to obtain ideas for new SARS-CoV-2 RdRp inhibitors. For this purpose, we trained a generative chemical language model.<sup>[32]</sup> In the first step, a generic model was obtained by learning the chemical syntax of a large set of known, bioactive druglike compounds. In the second step, this model was biased toward nucleoside analogues acting as SARS-CoV-2 RdRp inhibitors (Fig. 3a).<sup>[33]</sup> New molecules were sampled from this finetuned model (Fig. 3b).

These *de novo* generated structures contain several well-known substructures of RdRp inhibitors, but also some innovative moieties. This set of computer-generated molecules, like all molecular *de novo* designs, should serve as inspiration rather than elaborated lead structures because of certain limitations of the approach. In this particular example, no background information about nucleoside interaction in RNA and the mechanism of RdRp inhibition was considered during neural network training. Neither target selectivity, pharmacokinetic and -dynamic properties, nor the synthesizability of the designs were explicitly considered. Consequently the suggested molecules will benefit from careful checking by human experts and other computational tools. The selected designs then have to be synthesized and tested before any claim of pharmacological activity can be made. Similar considerations should be taken into account whenever employing molecular *de novo* design methods.

## 3. De-orphaning the Targets of Bioactive Natural Products with Machine Intelligence

Targeted natural product-inspired drug discovery requires the knowledge of the pharmacologically relevant macromolecular binding partners of the natural products. Numerous software tools are available for target and activity prediction, ranging from structure-based (*e.g.* docking) to ligand-based (*e.g.* substructure-, pharmacophore-, shape-based) methods.<sup>[34]</sup> No method is perfect but all have their individual sweet spot. One of the most successful and widely applied tools is TIGER (Target Inference Generator), which has proven applicable to natural products.<sup>[35]</sup> The TIGER algorithm works on the two-dimensional chemical structure (chemical constitution) of the ligand and does not take the target structure into account. It is thus applicable to a wide range of targets and ligands. Most target prediction tools, including TIGER, were developed using small molecule reference data. Their prediction accuracy typically suffers when applied to larger natural product structures, *e.g.* macrocycles or peptides. Aiming to partially alleviate this issue, one can virtually dissect the large natural product into smaller portions and perform target predictions for the resulting “drug-sized” fragments.<sup>[36]</sup> Fig. 4 shows three such examples of new target identification with TIGER. Resveratrol (**9**) is a small natural product, for which estrogen receptor beta antagonism was predicted and experimentally confirmed (K<sub>i</sub> = 0.4  $\mu$ M).<sup>[37]</sup> For the medium-sized anticancer depsipeptide dolicolide (**10**) the software revealed prostanoid E receptor EP3 antagonism (IC<sub>50</sub> = 16 ± 7 nM, K<sub>B</sub> = 6 nM).<sup>[38]</sup> For the polyketide archazolid A, a known inhibitor of V-ATPase, farnesoid X receptor (FXR) agonist, EC<sub>50</sub> = 0.2  $\mu$ M and other hitherto unknown targets could be identified.<sup>[36]</sup>

Aside from providing a straightforward access to target and activity prediction for large natural products, fragment-based prediction sometimes points to the most important function-conveying substructural moieties (magenta colored parts in Fig. 4), which can be useful for chemical derivatization and guided optimization.

### 3.1 Revealing Targets of Complex Natural Products: A Prospective Study

The marine natural product (–)-zampanolide **20** is a microtubule-stabilizing antiproliferative macrolide from the Togan sponge *Cacospongia mycofijensis* (Fig. 5).<sup>[39]</sup> Its total synthesis was achieved in 2012,<sup>[40]</sup> and several structural analogues have been synthetically obtained ever since.<sup>[41]</sup> The unmodified natural product (**12**) potently inhibits the growth of different human cancer cells *in vitro* with nanomolar IC<sub>50</sub> values, whereas its synthetic de-

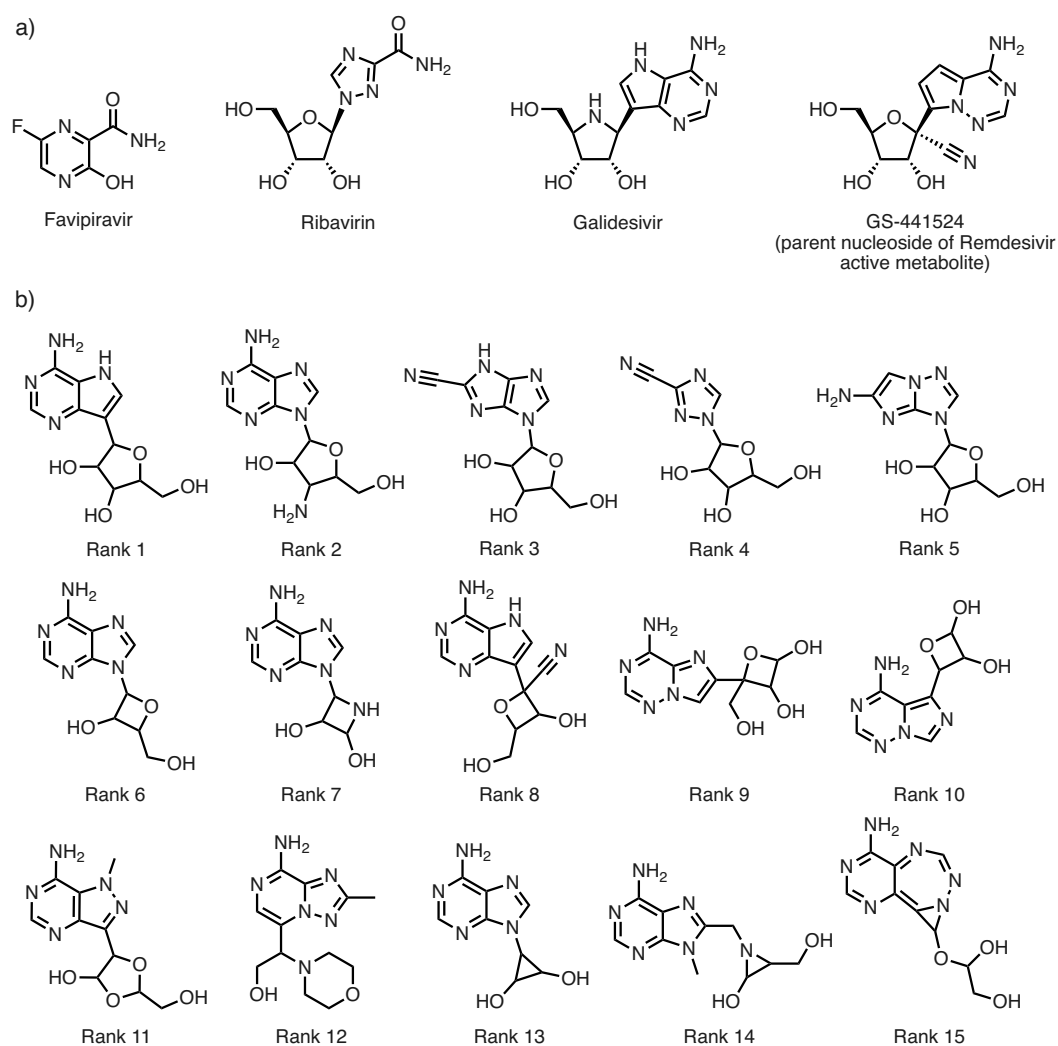


Fig. 3. Template compounds effective against SARS-CoV-2 RdRp (a), and a choice of *de novo* generated nucleoside analogs and mimetics (b). The computer-generated molecules are ranked according to their topological pharmacophore similarity to the template ensemble (decreasing similarity from Rank 1 to Rank 15).

relative desTHP(-)-zampanolide (**13**) is >100-fold less potent.<sup>[40]</sup> In contrast to **12**, **13** shows no measurable binding to stabilized microtubules up to concentrations of 25  $\mu\text{M}$  (unpublished data). We took this information as a starting point to identify potential human drug targets of compound **13**. Based solely on the two-dimensional structure of compound **13**, TIGER software (version 16.10) suggested cholecystinin receptor B (TIGER score = 90), N-formyl peptide receptor 2 FPR2 (TIGER score = 24.5), and prostanoid receptor EP3 (TIGER score = 22.3) as the most confidently predicted targets. The suggestion of FPR2 seemed particularly interesting because of its vital role in cell differentiation and chemotaxis.<sup>[42]</sup> Experimental validation corroborated compound **13** as a partial agonist of FPR2 *in vitro* (75% max. receptor activation,  $\text{EC}_{50} = 5 \pm 1 \mu\text{M}$ , Fig. 6).<sup>[43]</sup> This prospective example positively advocates the use of machine learning models for ligand-based target prediction and ligand ‘de-orphaning’.

#### 4. Conclusions and Outlook

Computational *de novo* molecular design has proven value for generating reasonable chemical structures. These computer-generated molecules provide working hypotheses for chemical synthesis and experimental testing. Concluding from several successful prospective applications to natural products, this approach seems suited for finding synthetic analogues and mimetics of bioactive natural products. Both rule-based and rule-free machine learning methods for *de novo* structure generation can be used for this purpose. Remaining challenges are the structural complexity of certain natural products and the functional relevance of certain substructural elements, e.g. carbohydrate moieties, among others.

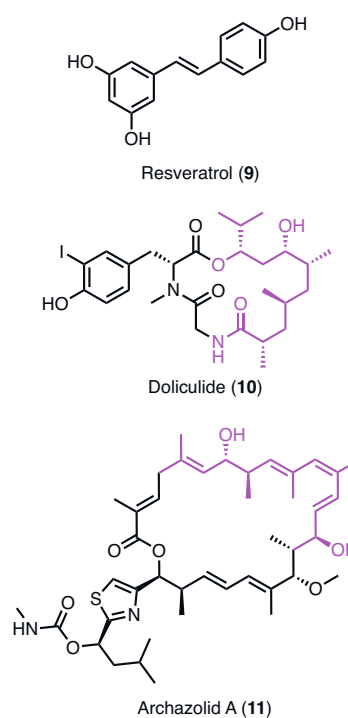


Fig. 4. Examples of pharmacologically active natural products for which new molecular targets were identified with a machine learning model. The magenta-coloured parts of dolicolide and archazolid A are relevant for the predicted target interactions.



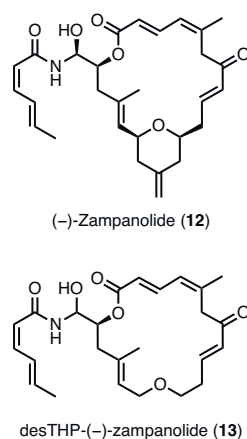


Fig. 5. Structures of the antiproliferative marine macrolide (-)-zampanolide and its synthetic derivative desTHP(-)-zampanolide.

Also, important activity-determining structural features, *e.g.*, stereocenters, are insufficiently accounted for by the currently available ligand-based *de novo* design methods.

Recently, there have been attempts to integrate computational molecule construction and activity prediction into a single generative model. These computational models, if proven successful in diverse prospective settings, will be able to automatically generate new chemical entities with desired properties and biological activity, without the need for explicit activity prediction. Certain deep learning approaches have already been shown applicable in this regard.<sup>[44]</sup> Geometric deep learning techniques allow to explicitly consider three-dimensional chemical features for model building.<sup>[21]</sup> However, this approach has not been applied to natural products yet. A further area of development is the combination of deep learning with a rule-based molecule construction process, thereby allowing chemists to explore the virtual chemical space that is accessible with the molecular building blocks that are readily available in the laboratory.<sup>[45]</sup> In the near future we expect to see fully automated laboratories implementing design-make-test-optimize cycles,<sup>[46]</sup> thereby enabling the full exploration of the structural and functional diversity of natural products for *de novo* design. According to the World Health Organization, infectious diseases represent a leading global public health threat in the 21<sup>st</sup> century.<sup>[47]</sup> At the same time, populations are growing and ageing owing to successes against infections. This situation paradoxically raises the risk of developing chronic diseases.<sup>[48]</sup> The proposed drug design approach could enable a renaissance of natural product-inspired pharmaceutical research by amalgamating modern medicinal chemistry with generative artificial intelligence. This matchless setting will alleviate some of the limitations of traditional natural product-based drug discovery, and harness the healing power of naturally evolved solutions for future synthetic medicines.

#### Acknowledgements

This research was supported by the Swiss National Science Foundation (grant no. 205321\_182176 to G.S.).

#### Conflict of Interest Statement

P.S. and G.S. declare a potential financial conflict of interest as in their role as founders of inSili.com LLC and consultants to the pharmaceutical industry.

Received: April 13, 2022

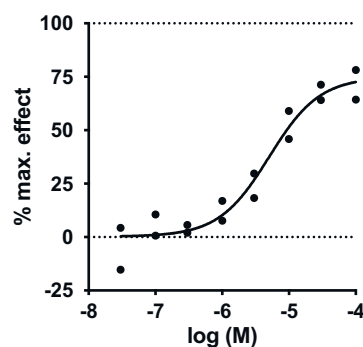


Fig. 6. Activation of N-formyl peptide receptor 2 (FPR2) by desTHP(-)-zampanolide (13). Data shown from  $n = 2$  independent experiments.

- [1] A. L. Harvey, *Drug Discov. Today* **2008**, *13*, 894, <https://doi.org/10.1016/j.drudis.2008.07.004>
- [2] E. Patridge, P. Gareiss, M. S. Kinch, D. Hoyer, *Drug Discov. Today* **2015**, *21*, 204, <https://doi.org/10.1016/j.drudis.2015.01.009>.
- [3] a) U. Lindequist, *Biomol. Ther.* **2016**, *24*, 561, <https://doi.org/10.4062/biomolther.2016.181>; b) P. Shinde, P. Banerjee, A. Mandhare, *Expert Opin. Ther. Pat.* **2019**, *29*, 283, <https://doi.org/10.1080/13543776.2019.1598972>.
- [4] C. Wan-Loy, P. Siew-Moi, *Mar. Drugs* **2016**, *14*, pii: E222, <https://doi.org/10.3390/md14120222>.
- [5] a) G. Molinari, *Adv. Exp. Med. Biol.* **2009**, *655*, 13, ; b) A. G. Atanasov, S. B. Zotchev, V. M. Dirsch, International Natural Product Sciences Taskforce, C. T. Supuran, *Nat. Rev. Drug Discov.* **2021**, *20*, 200, <https://doi.org/10.1038/s41573-020-00114-z>.
- [6] a) E. N. Muratov, R. Amaro, C. H. Andrade, N. Brown, S. Ekins, D. Fourches, O. Isayev, D. Kozakov, J. L. Medina-Franco, K. M. Merz, T. I. Oprea, V. Poroiikov, G. Schneider, M. H. Todd, A. Varnek, D. A. Winkler, A. V. Zakharov, A. Cherkasov, A. Tropsha, *Chem. Soc. Rev.* **2021**, *50*, 9121, <https://doi.org/10.1039/D0CS01065K>; b) X. Yang, Y. Wang, R. Byrne, G. Schneider, S. Yang, *Chem. Soc. Rev.* **2019**, *119*, 10520, <https://doi.org/10.1021/acs.chemrev.8b00728>.
- [7] A. L. Harvey, R. Edrada-Ebel, R. J. Quinn, *Nat. Rev. Drug Discov.* **2015**, *14*, 111, <https://doi.org/10.1038/nrd4510>.
- [8] N. E. Thomford, D. A. Senthane, A. Rowe, D. Munro, P. Seele, A. Maroyi, K. Dzobo, *Int. J. Mol. Sci.* **2018**, *19*, 1578, <https://doi.org/10.3390/ijms19061578>.
- [9] T. Rodrigues, D. Reker, P. Schneider, G. Schneider, *Nat. Chem.* **2016**, *8*, 531, <https://doi.org/10.1038/nchem.2479>.
- [10] a) G. Schneider, W. Neidhart, T. Giller, G. Schmid, *Angew. Chem. Int. Ed.* **1999**, *38*, 2894, [https://doi.org/10.1002/\(SICI\)1521-3773\(19991004\)38:19%3C2894::AID-ANIE2894%3E3.0.CO;2-F](https://doi.org/10.1002/(SICI)1521-3773(19991004)38:19%3C2894::AID-ANIE2894%3E3.0.CO;2-F); b) Y. Hu, D. Stumpfe, J. Bajorath, *J. Med. Chem.* **2016**, *59*, 4062, <https://doi.org/10.1021/acs.jmedchem.5b01746>.
- [11] G. Schneider, U. Fechner, *Nat. Rev. Drug Discov.* **2005**, *4*, 649, <https://doi.org/10.1038/nrd1799>.
- [12] Book: G. Schneider (ed.), *De Novo Molecular Design*, Wiley-VCH, Weinheim, New York, **2013**, <https://doi.org/10.1002/9783527677016>.
- [13] a) J. Lyu, S. Wang, T. E. Balius, I. Singh, A. Levit, Y. S. Moroz, M. J. O'Meara, T. Che, E. Algaa, K. Tolmacheva, A. A. Tolmachev, B. K. Shoichet, B. L. Roth, J. J. Irwin, *Nature* **2019**, *566*, 224, <https://doi.org/10.1038/s41586-019-0917-9>; b) F. Gentile, J. C. Yaacoub, J. Gleave, M. Fernandez, A. T. Ton, F. Ban, A. Stern, A. Cherkasov, *Nat. Protoc.* **2022**, *17*, 672, <https://doi.org/10.1038/s41596-021-00659-2>.
- [14] a) P. Schneider, W. P. Walters, A. T. Plowright, N. Sieroka, J. Listgarten, R. A. Goodnow Jr, J. Fisher, J. M. Jansen, J. S. Duca, T. S. Rush, M. Zentgraf, J. E. Hill, E. Krutoholow, M. Kohler, J. Blaney, K. Funatsu, C. Luebkeermann, G. Schneider, *Nat. Rev. Drug Discov.* **2020**, *19*, 353, <https://doi.org/10.1038/s41573-019-0050-3>; b) G. Schneider, *Nat. Mach. Intell.* **2019**, *1*, 128, <https://doi.org/10.1038/s42256-019-0030-7>.
- [15] M. Hartenfeller, H. Zettl, M. Walter, M. Rupp, F. Reisen, E. Proschak, S. Weggen, H. Stark, G. Schneider, *PLoS Comput Biol.* **2012**, *8*, e1002380, <https://doi.org/10.1371/journal.pcbi.1002380>.
- [16] a) Q. Vanhaelen, Y. C. Lin, A. Zhavoronkov, *ACS Med. Chem. Lett.* **2020**, *11*, 1496, <https://doi.org/10.1021/acsmchemlett.0c00088>; b) Book: T. Jebara, *Machine Learning: Discriminative and Generative*, Kluwer Academic,

- Dordrecht, 2004; c) Book: D. Foster, *Generative Deep Learning: Teaching Machines to Paint, Write, Compose, and Play*, O'Reilly Media, Sebastopol, 2019.
- [17] a) H. Chen, O. Engkvist, Y. Wang, M. Olivecrona, T. Blaschke, *Drug Discov. Today* **2018**, *23*, 124, <https://doi.org/10.1016/j.drudis.2018.01.0391>; b) A. Zhavoronkov, Y. A. Ivanenkov, A. Aliper, M. S. Veselov, V. A. Aladinskiy, A. V. Aladinskaya, V. A. Terentiev, D. A. Polykovskiy, M. D. Kuznetsov, A. Asadulaev, Y. Volkov, A. Zholus, R. R. Shayakhmetov, A. Zhebrak, L. I. Minaeva, B. A. Zagribelnyy, L. H. Lee, R. Soll, D. Madge, L. Xing, T. Guo, A. Aspuru-Guzik, *Nat. Biotechnol.* **2019**, *37*, 1038, <https://doi.org/10.1038/s41587-019-0224-x>; c) B. Sanchez-Lengeling, A. Aspuru-Guzik, *Science* **2018**, *361*, 360, <https://doi.org/10.1126/science.aat2663>; d) D. C. Elton, Z. Boukouvalas, M. D. Fuge, P. W. Chung, *Mol. Syst. Des. Eng.* **2019**, *4*, 828, <https://doi.org/10.1039/C9ME00039A>.
- [18] a) M. H. S. Segler, T. Kogej, C. Tyrchan, M. P. Waller, *ACS Cent. Sci.* **2018**, *4*, 120, <https://doi.org/10.1021/acscentsci.7b00512>; b) F. Grisoni, G. Schneider, *Chimia* **2019**, *73*, 1006, <https://doi.org/10.2533/chimia.2019.1006>
- [19] T. Blaschke, M. Olivecrona, O. Engkvist, J. Bajorath, H. Chen, *Mol. Inf.* **2018**, *37*, 1700123, <https://doi.org/10.1002/minf.201700123>.
- [20] a) E. Putin, A. Asadulaev, Y. Ivanenkov, V. Aladinskiy, B. Sanchez-Lengeling, A. Aspuru-Guzik, A. Zhavoronkov, *J. Chem. Inf. Model.* **2018**, *58*, 1194, <https://doi.org/10.1021/acs.jcim.7b00690>; b) E. Lin, C. H. Lin, H. Y. Lane, *Molecules* **2020**, *25*, 3250, <https://doi.org/10.3390/molecules25143250>.
- [21] a) Y. Li, L. Zhang, Z. Liu, *J. Cheminform.* **2018**, *10*, 33; b) P. Pogány, N. Arad, S. Genway, S. D. Pickett, *J. Chem. Inf. Model.* **2019**, *59*, 1136, <https://doi.org/10.1021/acs.jcim.8b00626>; c) K. Atz, F. Grisoni, G. Schneider, *Nat. Mach. Intell.* **2021**, *3*, 1023, <https://doi.org/10.1038/s42256-021-00418-8>.
- [22] a) M. Popova, O. Isayev and A. Tropsha, *Sci. Adv.* **2018**, *4*, eaap7885, <https://doi.org/10.1126/sciadv.aap7885>; b) B. Sattarov, I. I. Baskin, D. Horvath, G. Marcou, E. J. Bjerrum, A. Varnek, *J. Chem. Inf. Model.* **2019**, *59*, 1182, <https://doi.org/10.1021/acs.jcim.8b00751>; c) Y. Xu, K. Lin, S. Wang, L. Wang, C. Cai, C. Song, L. Lai, J. Pei, *Future Med. Chem.* **2019**, *11*, 567, <https://doi.org/10.4155/fmc-2018-0358>.
- [23] a) G. Schneider, D. E. Clark, *Angew. Chem. Int. Ed.* **2019**, *58*, 10792; b) V. D. Mouchlis, A. Afantitis, A. Serra, M. Fratello, A. G. Papadiamantis, V. Aidinis, I. Lynch, D. Greco, G. Melagraki, *Int. J. Mol. Sci.* **2021**, *22*, 1676, <https://doi.org/10.3390/ijms22041676>.
- [24] a) M. Hartenfeller, G. Schneider, *Methods Mol. Biol.* **2011**, *672*, 299, [https://doi.org/10.1007/978-1-60761-839-3\\_12](https://doi.org/10.1007/978-1-60761-839-3_12); b) W. Gao, C. W. Coley, *J. Chem. Inf. Model.* **2020**, *60*, 5714, <https://doi.org/10.1021/acs.jcim.0c00174>; c) G. Bhisetti, C. Fang, *Methods Mol. Biol.* **2022**, *2390*, 409, [https://doi.org/10.1007/978-1-0716-1787-8\\_17](https://doi.org/10.1007/978-1-0716-1787-8_17).
- [25] D. Merk, L. Friedrich, F. Grisoni, G. Schneider, *Mol. Inf.* **2018**, *37*, 1700153, <https://doi.org/10.1002/minf.201700153>.
- [26] D. Merk, F. Grisoni, L. Friedrich, E. Gelzinyte, G. Schneider, *J. Med. Chem.* **2018**, *61*, 5442, <https://doi.org/10.1021/acs.jmedchem.8b00494>.
- [27] D. Reker, T. Rodrigues, P. Schneider, G. Schneider, *Proc. Natl. Acad. Sci. U. S. A.* **2014**, *111*, 4067, <https://doi.org/10.1021/acs.jmedchem.8b00494>.
- [28] L. Friedrich, G. Cingolani, Y. H. Ko, M. Iaselli, M. Miciaccia, M. G. Perrone, K. Neukirch, V. Bobinger, D. Merk, R. K. Hofstetter, O. Werz, A. Koeberle, A. Scilimati, G. Schneider, *Adv. Sci.* **2021**, *8*, e2100832, <https://doi.org/10.1002/adv.202100832>.
- [29] a) L. Friedrich, T. Rodrigues, C. S. Neuhaus, P. Schneider, G. Schneider, *Angew. Chem. Int. Ed.* **2016**, *55*, 6789, <https://doi.org/10.1002/anie.201601941>; b) L. Friedrich, R. Byrne, A. Treder, I. Singh, C. Bauer, T. Gudermann, M. Mederos Y Schnitzler, U. Storch, G. Schneider, *ChemMedChem* **2020**, *15*, 566, <https://doi.org/10.1002/cmdc.202000017>.
- [30] D. Merk, F. Grisoni, L. Friedrich, G. Schneider, *Commun. Chem.* **2018**, *1*, 68, <https://doi.org/10.1038/s42004-018-0068-1>.
- [31] a) G. Li, E. De Clercq, *Nat. Rev. Drug Discov.* **2020**, *19*, 149, <https://doi.org/10.1038/d41573-020-00016-0>; b) M. Wang, R. Cao, L. Zhang, X. Yang, J. Liu, M. Xu, Z. Shi, Z. Hu, W. Zhong, G. Xiao, *Cell Res.* **2020**, *30*, 269, <https://doi.org/10.1038/s41422-020-0282-0>.
- [32] M. Moret, L. Friedrich, F. Grisoni, D. Merk, G. Schneider, *Nat. Mach. Intell.* **2020**, *2*, 171, <https://doi.org/10.1038/s42256-020-0160-y>.
- [33] T. P. Sheahan, A. C. Sims, S. Zhou, R. L. Graham, C. S. Hill, S. R. Leist, A. Schäfer, K. H. Dinnon III, S. A. Montgomery, M. L. Agostini, A. J. Pruijssers, J. D. Chapell, A. J. Brown, G. R. Bluemling, M. G. Natchus, M. Saindane, A. A. Kolykhalov, G. Painter, J. Harcourt, A. Tamin, N. J. Thornburg, R. Swanstrom, M. R. Denison, R. S. Baric, *bioRxiv* **2020**, preprint.
- [34] R. Byrne, G. Schneider, *Methods Mol. Biol.* **2019**, *1888*, 273, [https://doi.org/10.1007/978-1-4939-8891-4\\_16](https://doi.org/10.1007/978-1-4939-8891-4_16).
- [35] P. Schneider, G. Schneider, *Expert Opin. Drug Discov.* **2017**, *12*, 271, <https://doi.org/10.1080/17460441.2017.1274727>.
- [36] D. Reker, A. M. Perna, T. Rodrigues, P. Schneider, M. Reutlinger, B. Mönch, A. Koeberle, C. Lamers, M. Gabler, H. Steinmetz, R. Müller, M. Schubert-Zsilavec, O. Werz, G. Schneider, *Nat. Chem.* **2014**, *6*, 1072, <https://doi.org/10.1038/nchem.2095>.
- [37] P. Schneider, G. Schneider, *Angew. Chem. Int. Ed.* **2017**, *56*, 11520, <https://doi.org/10.1002/anie.201706376>.
- [38] G. Schneider, D. Reker, T. Chen, K. Hauenstein, P. Schneider, K. H. Altmann, *Angew. Chem. Int. Ed.* **2016**, *55*, 12408, <https://doi.org/10.1002/anie.201605707>.
- [39] J. J. Field, A. J. Singh, A. Kanakkanthara, T. Halafih, P. T. Northcote, J. H. Miller, *J. Med. Chem.* **2009**, *52*, 7328, <https://doi.org/10.1021/jm901249g>.
- [40] D. Zurwerra, F. Glaus, L. Betschart, J. Schuster, J. Gertsch, W. Ganci, K. H. Altmann, *Chem. Eur. J.* **2012**, *18*, 16868, <https://doi.org/10.1002/chem.201202553>.
- [41] C. P. Bold, M. Gut, J. Schürmann, D. Lucena-Agell, J. Gertsch, J. F. Díaz, K. H. Altmann, *Chem. Eur. J.* **2021**, *27*, 5936, <https://doi.org/10.1002/chem.202003996>.
- [42] P. J. G. Cussell, M. Gomez Escalada, N. G. N. Milton, A. W. J. Paterson, *Neural Regen. Res.* **2020**, *15*, 1191, <https://doi.org/10.4103/1673-5374.272566>.
- [43] The *in vitro* activity assays were performed by Eurofins Cerep (Celle l'Evescault, France, [www.eurofinsdiscoveryservices.com](http://www.eurofinsdiscoveryservices.com)) on a fee-for-service basis. The cAMP response upon human recombinant (CHO cells) receptor activation was measured, with fMLP as a positive control, as described in: K. Wenzel-Seifert, C. M. Hurt, R. Seifert, *J. Biol. Chem.* **1998**, *273*, 24181, <https://doi.org/10.1074/jbc.273.37.24181>.
- [44] a) M. Moret, M. Helmstädter, F. Grisoni, G. Schneider, D. Merk, *Angew. Chem. Int. Ed.* **2021**, *60*, 19477, <https://doi.org/10.1002/anie.202104405>; b) M. Moret, F. Grisoni, P. Katzberger, G. Schneider, *J. Chem. Inf. Model.* **2022**, *62*, 1199, <https://doi.org/10.1021/acs.jcim.2c00079>.
- [45] A. Button, D. Merk, J. A. Hiss, G. Schneider, *Nat. Mach. Intell.* **2019**, *1*, 307, <https://doi.org/10.1038/s42256-019-0067-7>.
- [46] a) G. Schneider, *Nat. Rev. Drug Discov.* **2018**, *17*, 97, <https://doi.org/10.1038/nrd.2017.232>; b) F. Grisoni, B. J. H. Huisman, A. L. Button, M. Moret, K. Atz, D. Merk, G. Schneider, *Sci. Adv.* **2021**, *7*, eabg3338, <https://doi.org/10.1126/sciadv.abg3338>.
- [47] a) World Health Organization, Geneva, Switzerland, <https://www.who.int/en/>; b) N. I. Nii-Trebi NI, *Biomed. Res. Int.* **2017**, *2017*, 5245021.
- [48] a) World Health Organization, Geneva, Switzerland, WHO 2014 Update, <https://www.who.int/mediacentre/factsheets/fs310/en/>; b) S. D. S. Diani, *Med. Hypotheses* **2018**, *113*, 30, <https://doi.org/10.1016/j.mehy.2018.02.011>.

#### 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.2022.396>