nature chemistry

REVIEW ARTICLE

PUBLISHED ONLINE: 25 APRIL 2016 | DOI: 10.1038/NCHEM.2479

Counting on natural products for drug design

Tiago Rodrigues[1], Daniel Reker[1], Petra Schneider[1,2] and Gisbert Schneider[1]*

Natural products and their molecular frameworks have a long tradition as valuable starting points for medicinal chemistry and drug discovery. Recently, there has been a revitalization of interest in the inclusion of these chemotypes in compound collections for screening and achieving selective target modulation. Here we discuss natural-product-inspired drug discovery with a focus on recent advances in the design of synthetically tractable small molecules that mimic nature's chemistry. We highlight the potential of innovative computational tools in processing structurally complex natural products to predict their macromolecular targets and attempt to forecast the role that natural-product-derived fragments and fragment-like natural products will play in next-generation drug discovery.

Natural products and their intricate molecular frameworks offer medicinal chemists a range of uncharted chemotypes for the discovery of chemical probes and drugs[1–7]. Natural products often feature biologically relevant molecular scaffolds and pharmacophore patterns that have evolved as preferred ligand–protein binding motifs[8–12]. Therefore, natural products have long been explored as invaluable sources of inspiration for drug design, with particular effectiveness in cancerous and infectious diseases[13–16]. For example, rosuvastatin, a blockbuster drug for the treatment of high cholesterol, mimics the pharmacophore of the natural product mevastatin from the fungus *Penicillium citrinum* but contains fewer chiral centres (Fig. 1). Focused natural-product-inspired compound libraries may transfer, at least in part, pharmacologically relevant features to synthetically more tractable small molecules, thereby potentially improving the biological activity of these synthetic small molecules. Thus, it is no surprise to see pharmaceutical drug discovery programmes benefitting from the incorporation of natural-product-derived fragments into development pipelines. An analysis of drugs that have been approved by the US Food and Drug Administration since 1939 reveals the persistent use of natural-product-derived fragments in medicinal chemistry (Fig. 2b). We anticipate molecular design in chemical biology and medicinal chemistry to be further propelled by systematically exploring the chemical space of natural products[17,18]. Recent technological advances support this optimistic view, specifically for the discovery of biosynthetic gene clusters producing bioactive secondary metabolites[19], the development of bioinformatic methods for identifying such biosynthetic gene clusters in genome sequences and predicting the chemical structures of their products[20], and the synthesis of natural products and derivatives through the manipulation of biosynthetic enzymes[21].

Natural products are often perceived as chemically complex and differing from synthetic drug-like molecules in many regards[22]. Pioneering studies in the early 2000s revealed that these products contain a much larger fraction of sp^3-hybridized bridgehead atoms and chiral centres compared with synthetic small molecules[1,23]. Natural products also present a lower nitrogen but higher oxygen content on average. Nature apparently favours aliphatic over aromatic rings, with only 38% of the known natural products containing arene systems. Approximately 50% of the structurally resolved natural products in the Dictionary of Natural Products database do not have synthetic counterparts, and only approximately 20% of the ring systems that are present in natural products can also be found in trade drugs[1,19]. As current medicinal chemistry practice strives to abandon planar molecules and emphasize three-dimensionality for compound library design, natural products clearly offer innovative ring systems with suitable geometries for spatial side-chain positioning[24], which has recently been highlighted by the discovery of pharmacologically relevant macromolecular targets of the sesquiterpene englerin A (ref. 25).

Over the past two decades, fragment-based drug discovery has emerged as an alternative concept to circumvent some of the drawbacks of the traditional early drug discovery paradigm, which has all too often resulted in clinical attrition caused by toxicity and lack of efficacy[26]. Aided by diverse biophysical techniques and steered by ligand efficiency metrics[27], fragment-based drug discovery aims to optimize low-molecular-weight compounds that exert the desired biological activity into potent lead structure candidates through stepwise molecule growing and linking[28]. Natural-product-derived and bioactivity-oriented ring systems appear to be particularly suited for fragment growth into such ligand-efficient and selective new chemical entities[29,30]. Recent progress in utilizing natural-product-derived fragments to computationally infer the biomolecular targets and activities of natural products[31], together with the longstanding success of fragment-based *de novo* design[32], advocates for the concepts and methodology of modern computer-assisted drug design to support and guide these developments. Here, we review recent advances at the interface between theoretical and practical medicinal chemistry and promote the concept of computationally inspired natural product research.

Natural products as starting points for lead discovery

Natural-product-inspired synthetic compounds can provide feasible and innovative solutions to address important and enduring drug design challenges. For example, Hergenrother and colleagues reported a deoxynybomycin-inspired fragment-like chemical entity with potent activity against *Staphylococcus aureus* and improved aqueous solubility[33]. Infected mice that were treated with compound **1** (Fig. 3a) showed significantly longer survival rates relative to untreated animals. Interestingly, some of the investigated compounds would hardly fit the often-too-stringent medicinal chemistry criteria for drug leads. Although more thorough physicochemical and pharmacological profiling is imperative to validate compound **1** as a probe or lead structure, its success clearly suggests

[1]Swiss Federal Institute of Technology (ETH), Department of Chemistry and Applied Biosciences, Institute of Pharmaceutical Sciences, Vladimir-Prelog-Weg 4, 8093 Zürich, Switzerland. [2]inSili.com LLC, Segantinisteig 3, 8049 Zürich, Switzerland. *e-mail: gisbert.schneider@pharma.ethz.ch

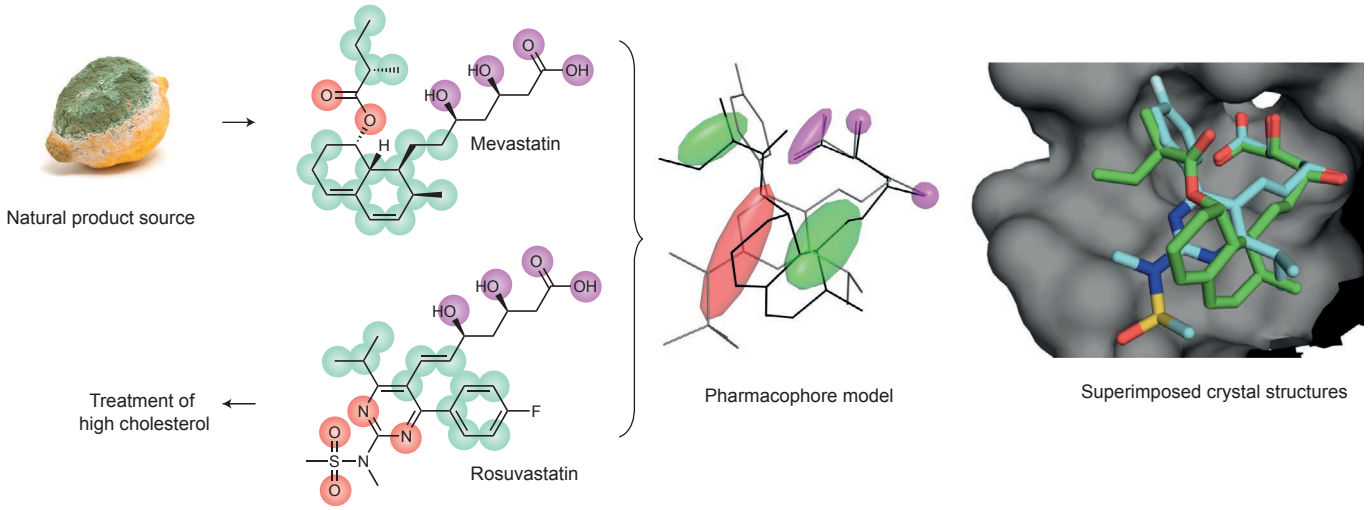


Figure 1 | Rosuvastatin is a natural-product-inspired drug. The drug rosuvastatin was developed as a synthetically accessible mimetic of the natural product mevastatin, which is produced by the fungus *Penicillium citrinum*. The two compounds possess a consensus pharmacophore but different chemical frameworks or 'scaffolds'. Lipophilic pharmacophore points are highlighted in green, hydrogen-bond acceptors in red and joint hydrogen-bond acceptors and donors in magenta[135]. The superimposed crystal-structure models of the two compounds (rosuvastatin in blue and mevastatin in green) reveal essentially identical interactions with the target enzyme 3-hydroxy-3-methylglutaryl-coenzyme A reductase. The surface of the enzyme pocket represents the solvent-accessible surface of the amino acid residues surrounding rosuvastatin (several amino acid residues are not shown to ensure the complete visibility of the ligand structures). The field of computational pharmacophore modelling has a long tradition in medicinal chemistry and provides methods of choice for understanding and accessing the pharmacological effects of bioactive natural products. Photograph: Ursula Alter/Getty.

that the rigorous application of decision guidelines, such as the popular 'rule of five'[34] or 'rule of three'[35], may hinder the discovery of biologically useful chemical entities[36,37]. Many of the existing guidelines have been developed with properties and structures of synthetic compounds in mind. Evidently, the excessive use of hard cut-offs for property rules and the occurrence of certain chemical structures should be avoided given their scientific and/or statistical weakness, specifically in the natural product context[38,39]. In fact, 18% of the natural products from the Dictionary of Natural Products database violate the 'rule of five' in at least two criteria (31% of Traditional Chinese Medicine Database, 15% of DrugBank, 9% of the ChEMBL database), which suggests the careful application of such guidelines not only for synthetic compounds, but also for natural products in particular[40].

Non-natural spirotryprostatin B analogues have recently been shown to act either as inhibitors of the p53–MDM2 protein–protein interaction or lead to cell cycle arrest in the G2/M phase by disrupting tubulin polymerization[41]. A concise and highly enantioselective synthesis of 3,3′-pyrrolidinyl spirooxindoles via asymmetric 1,3-dipolar cycloaddition afforded the desired compounds featuring

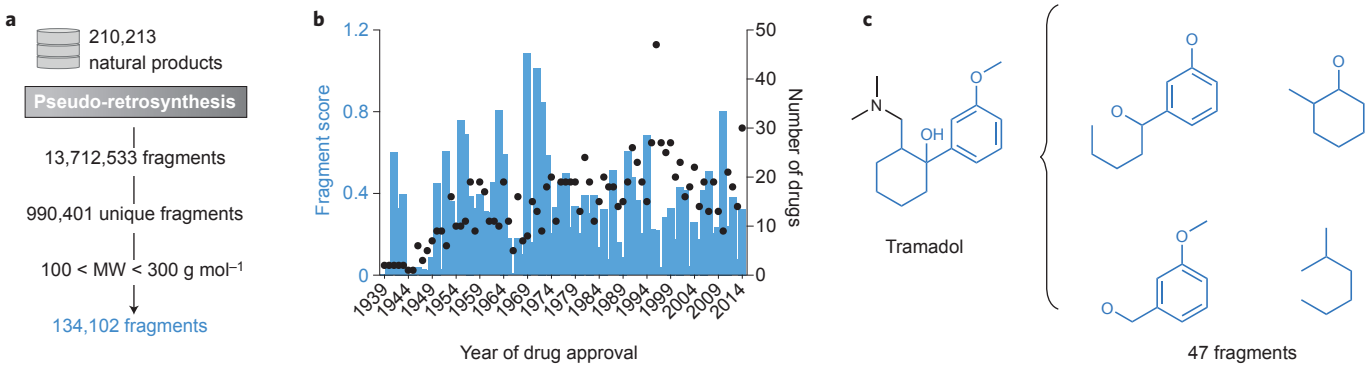


Figure 2 | Natural products and their computationally generated fragments as inspiration for drug discovery. a, Virtual natural product fragmentation. The Dictionary of Natural Products database contains 210,213 natural products. It yields 134,102 unique fragment-sized (molecular weight (MW) between 100 and 300 g mol^{-1}) substructures when applying a computational pseudo-retrosynthetic scheme (RECAP method)[136]. Many of these fragments are contained in approved drugs. The fragments can also be used as starting points for computational *de novo* drug design, which has become feasible with modern fragment-based molecular design software. **b**, Average number of natural-product-derived fragments in approved drugs. Natural products have always inspired medicinal chemistry. The histogram (blue bars) shows the molecular weight-corrected fragment count per drug molecule (fragment score = number of matching fragments / molecular weight). The black dots correspond to the annual number of US Food and Drug Administration-approved drugs (www.fda.gov) with a match of the active ingredient found in the DrugBank database (www.drugbank.ca). **c**, Natural-product-derived fragments in a synthetic drug. For example, the chemical structure of tramadol, a synthetic opioid analgesic[137], contains a total of 47 natural-product-derived fragments. Four of the 47 fragment structures are shown. Of note, the tertiary amide (drawn in black) is the only part of tramadol not covered by the fragment matches. Forty-seven divided by tramadol's molecular weight (263.4 g mol^{-1}) gives a relatively low normalized fragment score of 0.02. The drug Ultram contains tramadol as active ingredient. It was first approved by the US Food and Drug Administration in 1995. For the 27 drugs initially approved in 1995, we computed an average fragment score of 0.22 (compare with panel **b**).

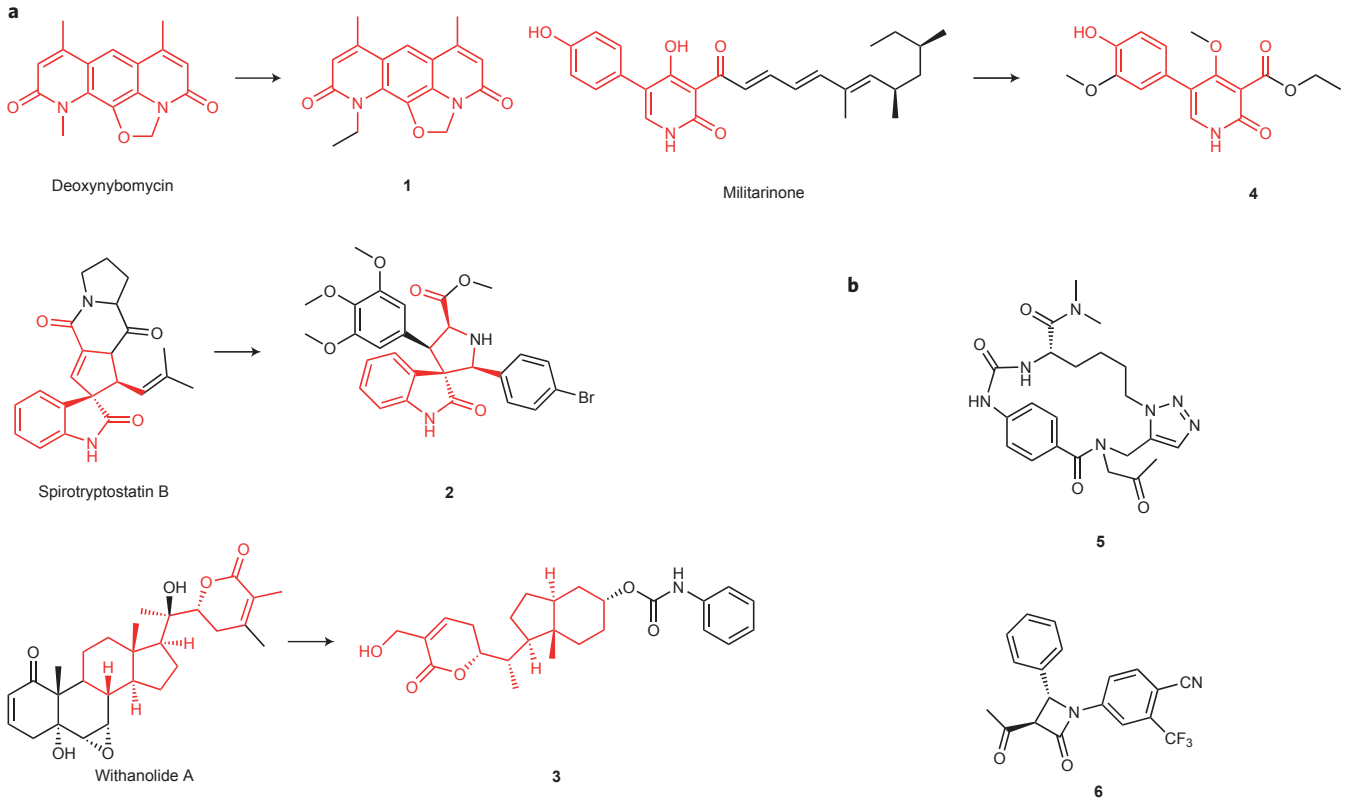


Figure 3 | Natural-product-inspired synthetic compounds with potent pharmacological effects. a, Red-coloured substructures highlight the structural commonalities of the natural templates and their synthetic mimetics. Appropriate chemical decoration of the natural-product-derived fragments and scaffolds leads to the desired pharmacological activities. Compound **1** is an antibacterial agent for combating *Staphylococcus aureus*. Compounds **2** and **3** are lead structures for the development of cytostatic and Smoothened protein modulators, respectively, whereas **4** is an inducer of neurite outgrowth. **b**, Compounds **5** and **6** were obtained as natural product-like chemical structures by diversity- and activity-oriented strategies, respectively.

one all-carbon quaternary centre and three tertiary stereocentres. *In vitro* profiling of the obtained focused library led to the identification of **2** as a low-micromolar cytostatic agent. Similarly, a biology-oriented synthesis programme on the *trans*-hydrindane dehydro-δ-lactone motif of withanolides led to the assembly of a focused library with enriched activity against the Smoothened receptor of Hedgehog signalling (for example, **3**)[42]. Following a similar strategy, Gademann and co-workers truncated militarinone and derivatized the selected natural product fragment to identify **4** as a potent inducer of neurite outgrowth[43].

Macrocyclic natural products often entail strongly organized conformations leading to pronounced target affinity without a significant loss of entropy on binding[44]. However, their structural complexity renders straightforward synthetic derivatization and optimization difficult, but evidence shows a large body of investigational and approved macrocyclic drugs for 'difficult targets', clearly offering exciting opportunities for drug discovery[45]. Diversity-oriented synthesis towards structurally diverse chemotypes is a viable enabling tool for effectively exploring natural product chemical space[46]. Recently, Spring and co-workers designed and prepared such a macrocycle library (for example, **5**, Fig. 3b) covering the shapes of known macrocyclic natural products. Several azide-derived aza-ylides were identified as a productive chemical coupling handle in aza-Wittig reactions, and macrocyclization was generally achieved via copper-catalysed cycloaddition and ruthenium-catalysed enyne metathesis[47]. Nelson and co-workers mimicked biosynthetic selection by prioritizing chemical reactions yielding natural-product-like compounds[48]. This activity-directed synthesis concept is based on assaying reaction products at increasingly low concentrations, as the key driver for the design of the subsequent reaction array. As such, lactam **6** was disclosed as a potent full agonist of the androgen receptor, suggesting that such an approach may be efficiently expanded to other target families. Pursuing a related approach, Aubé and co-workers devised synthetically tractable small molecule libraries inspired by the architecture of alkaloids[49]. Exploring different chemistries and attachment points for scaffold decoration, they obtained a total of 686 structurally diverse chemical entities. Importantly, these compounds merge physicochemical features from both drugs and natural products, thereby potentially facilitating further development.

These selected examples underscore the value of natural products as biologically pre-validated structures and as a potential source of chemical matter for inspired medicinal chemistry.

Computer-assisted design

Chemistry-driven approaches to developing natural-product-derived bioactive agents benefit from computational tools that help to rationalize the development of molecules with nature-inspired physicochemical properties. The concept of molecular scaffolds or frameworks connects the worlds of natural and synthetic compounds. More specifically, fragment-like natural products with innovative scaffolds may be used as starting points for chemical biology and medicinal chemistry programmes with great promise, as recently exemplified by Berg and co-workers through the development of natural-product-inspired synthetically accessible anticancer compounds[29,50]. To that end, software was developed to handle natural-product-derived scaffolds and identify areas in natural product chemical space that are correlated with certain bioactivity profiles[51]. A key feature of these tools is the visualization of distributions of natural products and synthetic

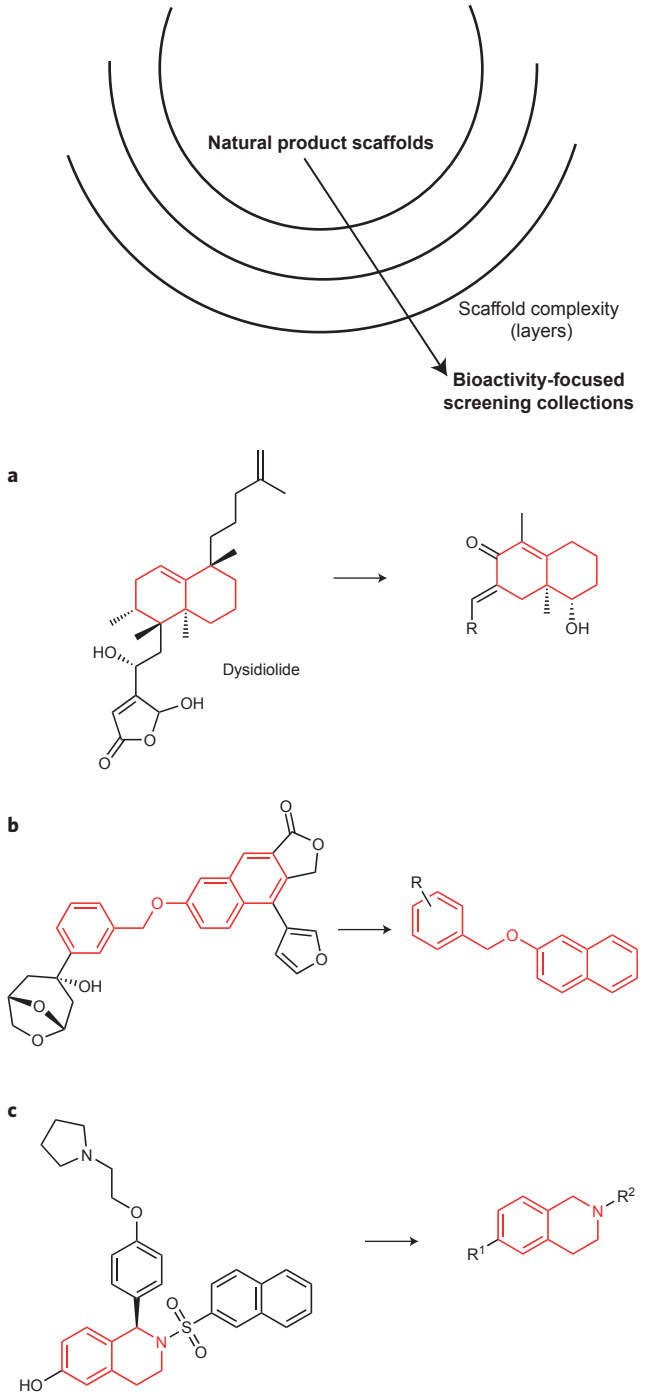


Figure 4 | Computer-assisted design of small molecules from natural product templates by scaffold simplification. a–c, Structurally complex frameworks of natural products can be manually or computationally dissected into simpler and synthetically more accessible fragment-like scaffolds (compare with Fig. 2a). These core structures borrow unique conformation-determining and physicochemical features from the natural product templates and may be used for the exploration of bioactivity-relevant chemical space. The general design strategy is outlined by the concentric circles: step-wise reduction of the structural complexity of the mother compound (natural product) to virtual fragments leads to small, chemically attractive scaffolds. Such a compound library based on the octahydronaphthalene core of dysidiolide afforded 11β-hydroxysteroid dehydrogenase type 1 inhibitors with high hit rates (**a**). The 2-(benzyloxy)naphthalene core of a naphtahlenic lignan lactone was used to design 5-lipoxygenase inhibitors (**b**), and isoquinoline-derived entities afforded inhibitors of the oestrogen receptor alpha (**c**).

bioactive compounds (sometimes referred to as 'chemography'), which facilitates the identification of promising molecular scaffolds for further exploration[52,53]. Waldmann and co-workers have appreciably contributed to the available technology and reported numerous case studies advocating the potential of scaffold-based compound design[54–58]. The structural classification of natural products offers a principle for organizing the structural chemical diversity that is embodied in natural products[8]. In short, this algorithm constructs a so-called scaffold tree for a given natural product by the stepwise simplification of the underlying molecular framework. Thereby, an efficient and intuitive mapping of complex structures can be obtained, allowing the strategic selection of ring systems for synthesis. One can expect a gradual loss of bioactivity with increasing simplification of the original natural product scaffolds. By pointing to key structural features that are correlated with a given bioactivity trait, this computational concept is of particular relevance for the design of synthetically tractable natural-product-like compound libraries.

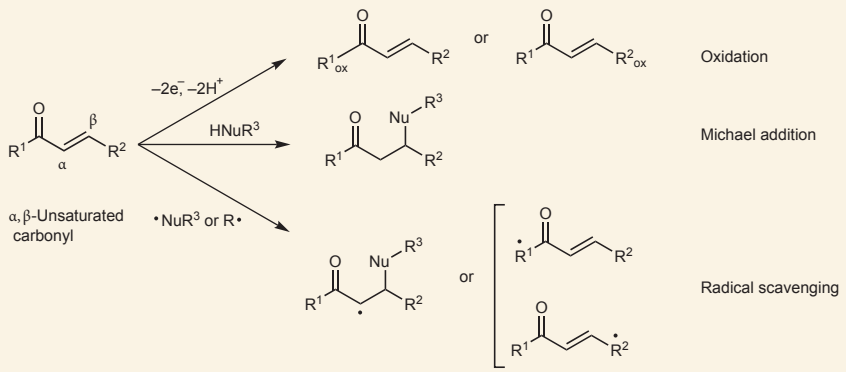
As a proof-of-concept, Koch *et al.* designed and experimentally validated a focused compound library that was built from the marine sesterterpenoid dysidiolide[8]. A total of 19% of the compounds presented the desired inhibitory activity against 11β-hydroxysteroid dehydrogenase (Fig. 4a). Of note, high receptor subtype selectivity was found across these screening hits. As seen from a drug design perspective, potency was achieved via the introduction of potentially 'problematic' groups, such as Michael acceptors (Box 1), which are not present in the original natural product. Importantly, the observed selectivity profile suggests no artefact readout. While an isolated example may not provide solid proof, the approach seems to have led to an increased hit rate compared with that of high-throughput screening[59]. Building on the same scaffold extraction concept, Wetzel *et al.* published the Scaffold Hunter[60] software tool in 2009, introducing the computational de-convolution of structurally intricate natural products as a key technique for generating virtual scaffold trees. This method facilitates the visualization of complex bioactivity data and promotes intuitive navigation in chemical space, with the goal of identifying simple ring systems exhibiting similar activity as the parent compound. This rationale was successfully used to identify both inhibitors and activators of pyruvate kinase. The focused screen yielded a considerably greater hit rate compared with that of straightforward high-throughput screening. In a larger-scale study, Renner *et al.* reported the hierarchical bioactivity-guided analysis of natural products via scaffold trees as a means to afford selective compounds for 5-lipoxygenase and oestrogen receptors with minimal scaffold identity to the natural product template (Fig. 4b,c)[61]. Starting from a scaffold hierarchy rooted at a seven-membered cyclic molecular framework and annotated activities for structurally less-complex intermediate products, the authors identified simplified, synthetically accessible fragments with inhibitory activity against 5-lipoxygenase and oestrogen receptors with improved ligand efficiency compared with the larger mother ring system. Building on this prolific approach, a recent comprehensive analysis of natural-product-derived fragments has been disclosed to inspire the design of new chemical entities[62]. It will be essential to accept that such natural-product-derived fragments often display only weak-to-moderate inhibitory potencies. Here, the driving metric of 'chemical beauty' is high ligand efficiency with the aim to identify suitable fragments for further optimization[63].

Predicting the macromolecular targets of natural products

The deployment of natural-product-inspired scaffolds in medicinal chemistry and drug discovery has traditionally been hampered by the lack of factual knowledge about the target engagement of the majority of the currently known natural products. Discovery projects in which the macromolecular targets are defined and known have repeatedly led to straightforward natural product

Box 1 | Natural products containing α,β-unsaturated carbonyls — a liability for drug design?

α,β-Unsaturated carbonyl groups are the most frequent potentially reactive substructures of natural products. One-sixth of the currently known natural products contain this generic chemical signature. It is controversial as to whether these compounds may serve as leads for drug design because their potential Michael acceptor function can lead to unwanted side effects, such as cell-damage or cytotoxicity. However, α,β-unsaturated carbonyls also have the potential to act as radical scavengers, enable covalent ligand binding to target proteins, or act as antioxidants, for example, by thiol trapping. Direct interaction with enzymes through α,β-unsaturated carbonyls can be desirable. For example, carcinogenesis can be blocked by a variety of substances that can induce (activate) certain enzymes, for example, quinone reductase and glutathione *S*-transferase activity, which then inhibits the 'reactive electrophilic forms of carcinogens'[138]. These abilities can be attributed to substances with Michael-accepting features. There have been several attempts to calculate the reactivity of α,β-unsaturated carbonyls, specifically high Michael acceptor reactivity, which is linked to compound toxicity[139]. It remains difficult to predict the actual *in vivo* functionality of these substructures, although Michael receptor functionality can be tuned by chemical optimization. One such approach is to vary the side chains of α-substituted enones ('α-modenones')[140], exploiting the fact that more electron-deprived bonds are more reactive. Spectroscopic methods support chemical optimization. For example, ^{13}C NMR shift analysis of the α and β carbons of the double bond captures the strength of its Michael-accepting capabilities[141].



derivatization. Innovative target prediction tools are now available to help identify the macromolecular receptors and potential off-target liabilities of drugs or lead-like entities and may help to identify the bioactivities of natural products[64–71]. The parallel virtual screening of an array of potential targets offers an affordable and attractive solution to unveiling putative ligand–receptor interactions by both ligand- and receptor-based methods[72]. For example, by relying on inverse molecular docking, cyclooxygenase-2 and peroxisome proliferator-activated receptor gamma were successfully identified as targets of meranzin, which exhibits potency comparable to that of indomethacin and rosiglitazone[73]. In a similar fashion, Rollinger and co-workers have productively applied three-dimensional pharmacophore models to interrogate target engagement by natural-product-derived structures[74,75]. These authors virtually screened 16 secondary metabolites from *Ruta graveolens* with a set of 2,208 pharmacophore models, resulting in the identification and experimental validation of the natural product arborinine as a human rhinovirus coat protein inhibitor and rutamarin as a cannabinoid-2 receptor ligand[76]. Elaboration on an acetylcholinesterase pharmacophore model also resulted in the identification of morphinan and isoquinolines as moderate inhibitors[77]. Potent partial agonists of proliferator-activated receptor gamma were identified among neolignans: dieugenol, tetrahydrodieugenol and magnolol[78]. Similarly, honokiol was identified as a weak non-adipogenic partial proliferator-activated receptor gamma agonist[79], providing opportunities for development. Although fragment-like natural products may serve as starting points for immediate further elaboration, the use of such weak effectors in chemical biology requires extensive prior validation[80].

The software PASS[81,82] is one of the early attempts to qualitatively predict thousands of biological activities from the two-dimensional chemical structure using molecular fragment descriptors[83]. This software was successfully applied to a set of more than 90 marine sponge alkaloids to identify anti-tumoural activity for approximately 80% of these natural products. Ninety-three individual components of St John's wort were analysed for probable pharmacological effects. Several compounds were correctly predicted for known side effects and cytochrome P450 modulation. Moreover, possible pharmacotherapeutic applications were identified for further experimental validation. Numerous target prediction tools have been developed to analyse synthetic drug-like molecules[84,85]. The majority of these methods rely on (i) chemical substructure similarity for target inference from reference drugs with known targets, (ii) explicit ligand–receptor docking, or (iii) genomic sequence and pathway information[86–88]. Therefore, it is no surprise that these approaches largely fail to predict targets for structurally intricate natural products because of their often rather different molecular constitution compared with that of synthetic drugs, computationally elusive receptor-relevant conformational ensembles[89], and lack of genomic and metabolomic reference data.

In contrast to explicitly substructure-based computational methods, the SPiDER target prediction software relies on a dual molecular representation in terms of topological pharmacophores and physicochemical properties[90,91]. Specifically, this software tackles the challenge of performing predictions in the absence of structural similarity for *de novo*-designed chemical entities. The idea is to find similarities between natural products and synthetic drugs on the level of pharmacophore features and computed properties. The algorithm clusters compounds with self-organizing maps built from molecular descriptors[92] that are suited for scaffold hopping between natural products and their synthetic mimetics[52]. This software was fruitfully employed for the identification of G-protein-coupled receptor ligands[90] and antiplasmodial kinase inhibitors[93]. We recently adapted this algorithm to predict the targets of structurally complex natural products. The antiproliferative macrocyclic archazolid A served as a challenging example, motivated by its profound structural dissimilarity to drug-like small molecules. Target identification for archazolid A was achieved through an analysis

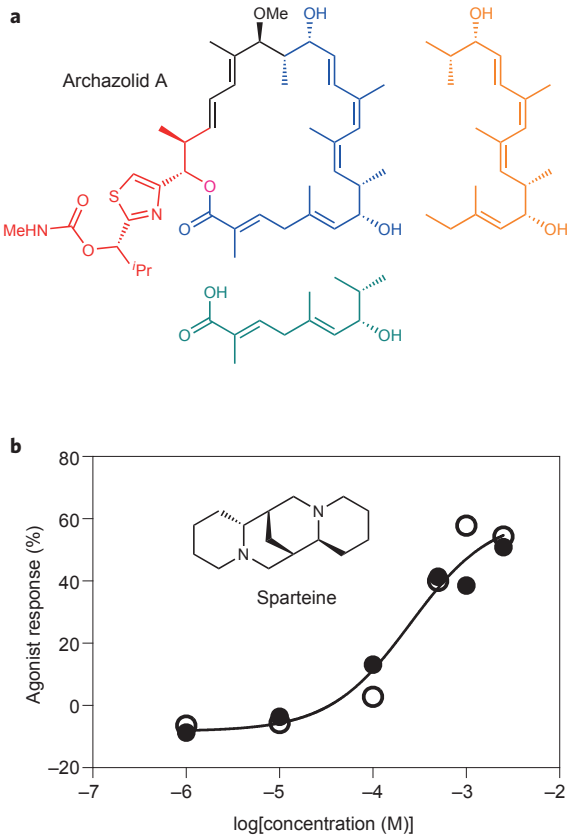


Figure 5 | Fragment-based target identification with the SPiDER software. Computer-assisted predictions of macromolecular drug targets help to effectively prioritize screening compound libraries, 'de-orphan' phenotypic screening hits, provide a means for drug re-purposing and complement experimental proteomic methods for target identification, among many other potential applications. Ligand-centric prediction tools do not require structural knowledge of the drug targets and represent an alternative to receptor-based approaches. The ligand-based computer tool SPiDER performs target predictions on the basis of topological pharmacophore and physicochemical molecular representations. Targets are inferred from pharmacologically active compounds with known targets by computing statistical similarity indices between query molecules and the reference compounds. **a**, The structure of archazolid A and its virtual fragments from which synthetic small-molecule counterparts with known macromolecular targets were identified. Through natural product fragmentation *in silico*, a link was established between the complex macrolide structure and drugs with known targets. **b**, SPiDER identified the fragment-like natural alkaloid sparteine as a functional partial agonist of the kappa opioid receptor (EC_{50} = 245±107 μM, n = 2; assayed at Cerep, France). The target predictions can guide biochemical natural product screening, limit the number of experiments and help save precious material.

of its computationally generated fragments (Fig. 5a). Strikingly, several proteins that were involved in the recognition and processing of arachidonic acid were predicted with high confidence and experimentally confirmed. For example, archazolid A inhibited 5-lipoxygenase and prostaglandin E_2 synthase 1, and the farnesoid X receptor was partially activated by archazolid A with nanomolar potency. Importantly, the activities that were found for archazolid A were at least comparable to those of these macromolecular targets' endogenous ligands and may contribute to the polypharmacological profile underlying the potent antiproliferative activity of archazolid A (ref. 31). This algorithm was equally used to unveil dual target activity of the fragment-like natural products graveolinine and isomacroin[94].

The aforementioned studies were motivated by the fact that natural-product-derived fragments resulted in more confidently predicted targets compared with their larger, structurally more complex mother compounds. To assess whether this effect is also observed for unprocessed fragment-sized natural products, we applied the original target prediction algorithm to all of the natural products from the Dictionary of Natural Products database (Fig. 6a). In fact, a total of 64,650 natural products (31%) are fragment-like (molecular weight between 100 and 300 g mol^{-1}). The SPiDER software was able to confidently (P value < 0.01) predict targets for 23,340 of those (36%), compared with only 31,556 (22%) of the 145,623 natural products with a molecular weight exceeding 300 g mol^{-1}. We observed that macromolecular targets of fragment-like natural products are as predictable as for natural-product-derived fragments (1.4 and 1.3 confidently predicted targets per compound, respectively), while larger natural products only achieved 0.6 confident target predictions on average. Drug-like molecules from DrugBank received 2.1 confidently predicted targets on average, which may suggest the greater target 'promiscuity' of synthetic drugs compared with natural products. The most frequently predicted targets of natural products are summarized in Fig. 6b,c. Ion channels, G-protein-coupled receptors and certain enzyme classes were preferably suggested by the software, which is perfectly in line with known preferences of bioactive natural products[95]. Such predictions may thus have the potential to guide multi-target approaches for natural-product-inspired drug discovery[96,97].

To further challenge the concept of target prediction for fragment-like natural products through a comparison with drug-like small molecules, we selected sparteine as a natural tool compound for a prospective experiment (Fig. 5b). In addition to weakly inhibiting p38α mitogen-activated protein kinase, sparteine was previously shown to bind to the muscarinic and nicotinic receptors[98]. Applying SPiDER, we predicted these two targets among the top three predictions. The kappa opioid receptor was listed as the second-most-confidently predicted target. Binding and functional assays confirmed the prediction and revealed sparteine as a ligand-efficient fragment for further development (ligand efficiency = 0.30).

This prospective analysis further attested to the use of advanced software for the discovery of the macromolecular targets of fragment-like natural compounds. Computational target prediction for complex natural products has become tangible and may facilitate the target-centric discovery of synthetic compounds with bioactivity-validated natural scaffolds. Despite the attractiveness of said tools, it should be kept in mind that only previously studied targets can be predicted, and in most cases, the target rank order should be interpreted as a qualitative index for hit identification, which is only poorly if at all correlated with the actual ligand potency. Moreover, the accuracy of these computational tools heavily relies on the molecular representation and reference database(s) that are used. Thus, we understand that the main value of target prediction for natural products lies in the rationalization and prioritization of compound screening campaigns, potentially unveiling metabolic[99] and pharmacological ligand–target networks[100].

Are natural products 'privileged' in drug discovery?

The scaffolds of natural products are repeatedly considered 'privileged' in drug discovery[101–104]. Computational methods should be able to evaluate this concept from a molecular perspective. Several studies have aimed at computing fragment-based 'natural-product-likeness' and related scoring systems, which assist medicinal chemists in designing screening compound collections[105]. Here, we investigated whether natural products might represent 'frequent hitters' — molecules that show activity in multiple biochemical and biological assays. We introduced this term more than a decade ago, motivated by the observation of the same screening hits being independently pursued in several parallel drug discovery projects[106]. In fact, there

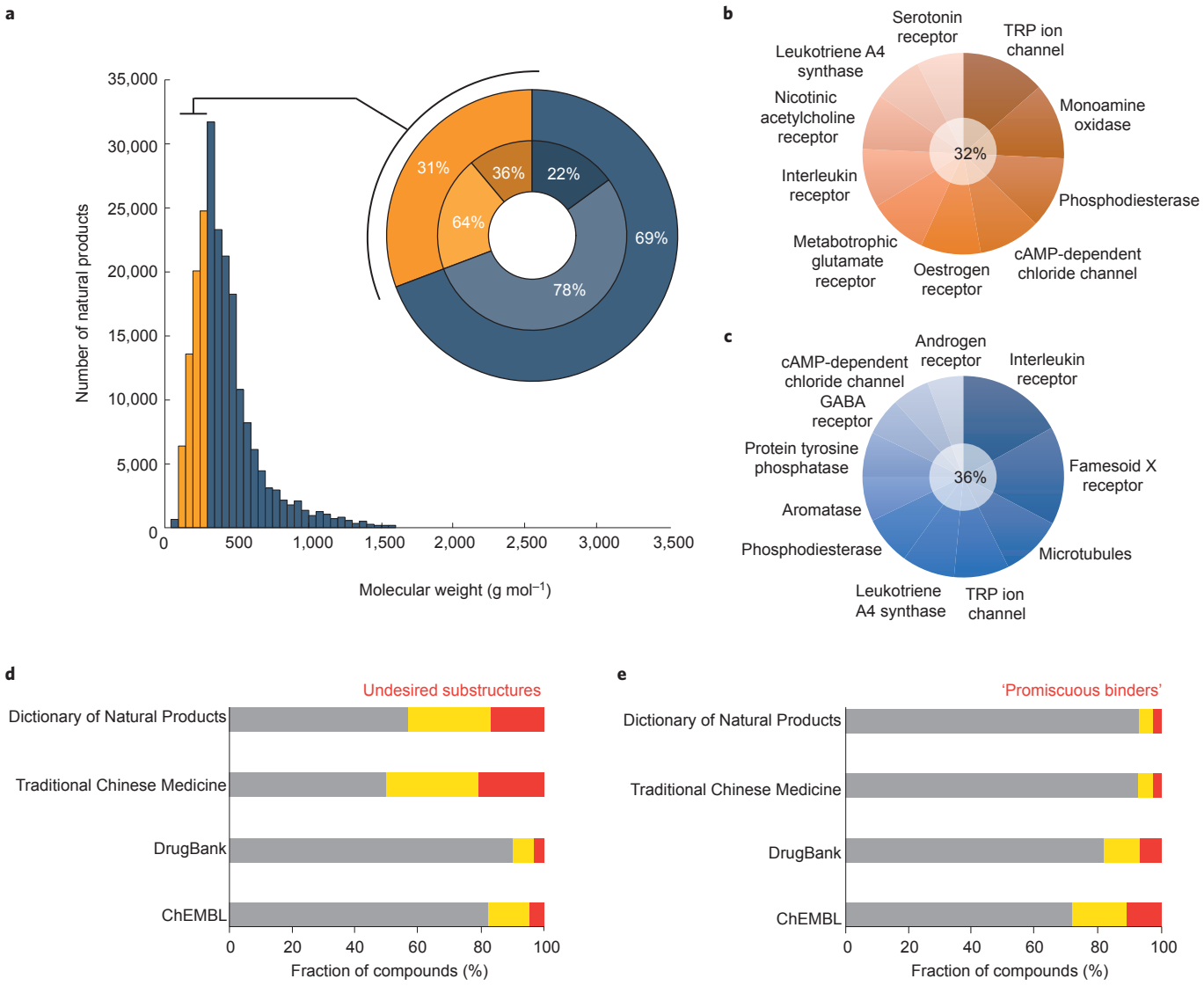


Figure 6 | Target predictions for natural products and synthetic bioactive molecules. a, Proportion of fragment-like natural products that are contained in the Dictionary of Natural Products database and their target predictability. The histogram shows the molecular weight distribution of natural products and highlights the range that is considered fragment-like (orange bars). A total of 64,650 structures (31%) of the natural products are fragment-sized. The inner circle of the pie chart compares the number of these small natural products (dark orange, 23,340 structures, 36%) with the larger ones (dark blue, 31,556 structures, 22%) for which confident predictions were obtained. Light colours indicate the ratio of compounds for which no confident target prediction could be obtained. The predictions were made with the SPiDER software[90]. **b**, The most frequently predicted macromolecular targets and target families for fragment-sized natural products (100 ≤ molecular weight < 300 g mol^{-1}). The top 10 targets account for 32% of the acquired predictions. **c**, The most frequently predicted macromolecular targets and target families for large natural products (molecular weight ≥ 300 g mol^{-1}). The top 10 targets account for 36% of the acquired predictions. **d**, Prevalence of undesired substructures in compound collections. The natural products from the Dictionary of Natural Products database (210,272 compounds) and a collection of Traditional Chinese Medicine Database (60,543 compounds) receive more substructure alerts than the mostly synthetic compounds stored in DrugBank (1,091 compounds) and ChEMBL (351,200 compounds). Grey = no alert, yellow = one alert, red = more than one alert. **e**, 'Promiscuous binder' prediction[121]. The coloured bars correspond to the percentage of the respective database. Overall, multi-target binding seems to be more likely for synthetic compounds than for natural products. Grey = pseudo-probability < 50%, yellow = 50% ≤ pseudo-probability < 90%, red = pseudo-probability ≥ 90%. TRP = transient receptor potential, GABA = gamma-aminobutyric acid.

are basically two causes of frequent-hitter behaviour: (i) unspecific reactivity caused by reactive and unstable functional groups, solubility issues and colloidal aggregation under assay conditions that lead to false-positive readouts in bioactivity screening[107,108]; and (ii) specific, reversible, concentration-dependent multi-target binding ('promiscuous binders'). Promiscuous binders in particular can be of great value for the design of bioactive tool compounds and lead structures. Recently, the rediscovery of pan-assay interference compounds has revitalized this discussion[109]. It has become common to computationally flag potential false positives by checking for the presence of certain reactive functional groups or undesirable structural elements[110,111]. To date, natural products as subsets of pharmaceutically interesting compounds have rarely been scrutinized for their frequent-hitter potential. Duan *et al.* studied the colloidal aggregation in a limited set of natural products present in traditional Chinese medicines. The data suggest that micromolar activities should be treated with some scepticism and carefully analysed in light of potential artefactual assay readouts[112]. For the sake of our study, we investigated four compound sets for frequent-hitter properties (Table 1): (i) 1,091 approved drugs from DrugBank[113],

Table 1 | Selection of natural product databases.

Database	Natural product entries	Content	Website
Dictionary of Natural Products	210,213	Natural products described in the literature	http://dnp.chemnetbase.com
Traditional Chinese Medicine	32,364	Natural products from herbs and animals, as listed in Chinese medical texts and dictionaries	http://tcm.cmu.edu.tw
SuperNatural	325,508	Natural products with toxicity and target prediction	http://bioinformatics.charite.de/supernatural/
ChEMBL	>75,000	Chemical structures from more than 48,000 publications	http://www.ebi.ac.uk/chembl/
MarinLit	26,490	Marine natural products, compiled from journal articles	http://pubs.rsc.org/marinlit/

(ii) the ChEMBL database[114], (iii) the Dictionary of Natural Products database and (iv) the Traditional Chinese Medicine Database[115].

As a first analysis, we employed three popular and widely accepted flagging lists to determine the fraction of potentially undesired functional groups and substructures: (i) 20 reactive substructures collected by Rishton[116], (ii) 55 reactive and undesired substructures defined by Hann *et al.*[117] and (iii) 75 substructure flags suggested by Baell and Halloway[118]. The results are presented in Fig. 6d. The compiled substructure flags clearly separate the small synthetic compounds from the natural products. There is a much greater fraction of compounds in the natural product databases with at least one substructure alert. Almost every second natural product contains a potentially undesired substructural element, whereas fewer than 10% of the approved drugs bear such a feature. Michael acceptor groups represent the most abundant flag within all of the analysed compound sets, meaning that nearly every sixth natural product contains this potentially reactive substructure (Table 2). This result agrees with a recent analysis by Villoutreix and co-workers, who compiled a consolidated flag list and applied it to 778 drugs[119]. Judging from a purely structural perspective, natural products contain more reactive substructural elements than synthetic drugs. It might therefore be wise to consider this particular aspect of natural products for the assessment of high-throughput screening results, hit prioritization and hit-to-lead progression. Importantly, not every potential reactive group actually represents a liability[120]. The rigorous blind elimination of flagged natural products from hit lists and screening collections would certainly be prejudicial. Depending on the macromolecular target and therapeutic indication in mind, a reactivity-based molecular mechanism might actually be desirable (Box 1).

The second part of our analysis attempts to estimate the ability of frequent hitters to interact with multiple targets and exhibit promiscuous binding potential (Fig. 6e). For this purpose, we employed a pharmacophore-based machine-learning model that was trained on the distinction between compounds with multiple substructure flags and compounds without such substructure alerts but that are known to bind to multiple target classes[121]. The result of this second analysis suggests that, despite the high prevalence of potentially undesired substructures in natural products, (i) synthetic drugs have a greater potential to interact with multiple targets, and (ii) target class promiscuity is a feature of synthetic compounds but less common for natural products. This outcome agrees with the observation that natural products have the lowest propensity to bind to a set of 100 different target proteins compared with commercial small molecules[122]. Moreover, the greater number of confidently predicted macromolecular targets of synthetic compounds is related to the known polypharmacological nature of synthetic drugs[123], which certainly warrants the further investigation of natural product polypharmacology, for example, by advanced pharmacophore and chemogenomic network modelling[124]. From this computational perspective, one might therefore conclude that natural products and natural-product-derived fragments are indeed 'privileged' for drug discovery and design.

Conclusion and outlook

The natural-product-inspired design of tool compounds and drug leads has always played an important role in the chemical sciences[125]. The re-emerging interest in natural products as sources for innovative drug discovery may benefit from modern chem- and bioinformatics approaches[126–128]. While advanced chemical synthesis techniques for straightforward derivatization or the manual truncation of selected moieties in natural products have become common practice[129–132], we are witnessing a growing interest in computational molecular design methods applied to natural-product-derived fragments. Several pioneering contributions have been reported recently, with computational approaches showcased as complementary drivers to further propel and steer the efficient exploration of natural products for drug discovery. We found that natural products might in fact be suited for selective target modulation, although they often contain perceived potentially reactive or undesirable functional groups and substructure moieties. Whether this detail actually influences natural-product-inspired hit-to-lead progression depends on the particular discovery project. In parallel, the systematic discovery of the macromolecular targets of natural products and their fragments has become tangible due to the availability of predictive computational tools, which will undoubtedly

Table 2 | Potential 'false positive' flags (undesired substructures) in compound databases.

Substructure flag	ChEMBL (351,200 compounds)	Dictionary of Natural Products (210,272 compounds)	Traditional Chinese Medicine (60,543 compounds)
Michael acceptor	4% (14,983)	17% (35,253)	14% (8,755)
Aliphatic ester	2% (8,104)	9% (18,776)	7% (4,148)
Aliphatic methylene chain ($n \geq 7$)	1% (5,267)	6% (12,940)	3% (1,739)
Epoxide or thioepoxide or aziridine	0% (527)	5% (10,618)	6% (3,679)
Aldehyde	0% (1,419)	3% (7,036)	6% (3,425)
Saponin derivative	0% (70)	3% (6,939)	5% (2,939)
Aliphatic ketone	1% (4,252)	3% (6,890)	5% (3,051)
Quinone	0% (1,231)	2% (5,070)	3% (1,871)

support and facilitate discovery programmes. Importantly, with our increasing knowledge of pharmacologically active metabolites of natural products[133], computational tools for predicting both the structures and the pharmacokinetic and pharmacodynamic properties of these metabolites will help prioritize natural-product-derived lead structure candidates[99]. Given that several of these software tools can either be licensed or are openly available to the community, we foresee increased interest in related paradigms by both academic research groups and pharmaceutical companies. Specifically, we anticipate that these tools may be deployed to probe traditionally challenging signalling pathways (for example, involving protein–protein interactions)[134] or prototyping ligands for novel disease-relevant drug targets. However, it may be wise not to put all of our eggs in one basket. We still need to fully explore the domains of application and understand the limitations of natural-product-inspired computational molecular design and target prediction. By maintaining some healthy scepticism, we feel that the time is ripe for (re)discovering and exploiting computational natural product analysis for chemical biology and drug discovery.

Received 20 June 2015; accepted 12 February 2016; published online 25 April 2016

References

1. Lee, M. L. & Schneider, G. Scaffold architecture and pharmacophoric properties of natural products and trade drugs: application in the design of natural product-based combinatorial libraries. *J. Comb. Chem.* **3,** 284–289 (2001).
2. Du, L. *et al.* Crowdsourcing natural products discovery to access uncharted dimensions of fungal metabolite diversity. *Angew. Chem. Int. Ed.* **53,** 804–809 (2014).
3. Bauer, A. & Brönstrup, M. Industrial natural product chemistry for drug discovery and development. *Nat. Prod. Rep.* **31,** 35–60 (2014).
4. Brown, D. G., Lister, T. & May-Dracka, T. L. New natural products as new leads for antibacterial drug discovery. *Bioorg. Med. Chem. Lett.* **24,** 413–418 (2014).
5. Scheepstra, M. *et al.* A natural-product switch for a dynamic protein interface. *Angew. Chem. Int. Ed.* **53,** 6443–6448 (2014).
6. Hattum, H. v. & Waldmann, H. Biology-oriented synthesis: harnessing the power of evolution. *J. Am. Chem. Soc.* **136,** 11853–11859 (2014).
7. Harvey, A. L., Edrada-Ebel, R. & Quinn, R. J. The re-emergence of natural products for drug discovery in the genomics era. *Nature Rev. Drug Discov.* **14,** 111–129 (2015).
8. Koch, M. A. *et al.* Charting biologically relevant chemical space: a structural classification of natural products (SCONP). *Proc. Natl Acad. Sci. USA* **102,** 17272–17277 (2005).
9. Eschenbrenner-Lux, V., Küchler, P., Ziegler, S., Kumar, K. & Waldmann, H. An enantioselective inverse-electron-demand imino Diels-Alder reaction. *Angew. Chem. Int. Ed.* **53,** 2134–2137 (2014).
10. Zimmermann, T. J. *et al.* Biology-oriented synthesis of a tetrahydroisoquinoline-based compound collection targeting microtubule polymerization. *ChemBioChem* **14,** 295–300 (2013).
11. Narayan, R., Bauer, J. O., Strohmann, C., Antonchick, A. P. & Waldmann, H. Catalytic enantioselective synthesis of functionalized tropanes reveals novel inhibitors of hedgehog signaling. *Angew. Chem. Int. Ed.* **52,** 12892–12896 (2013).
12. Lowe, D. B. Drug discovery: combichem all over again. *Nature Chem.* **6,** 851–852 (2014).
13. Molinari, G. Natural products in drug discovery: present status and perspectives. *Adv. Exp. Med. Biol.* **655,** 13–27 (2009).
14. Koehn, F. E. & Carter, G. T. The evolving role of natural products in drug discovery. *Nature Rev. Drug Discov.* **4,** 206–220 (2005).
15. Harvey, A. L. Natural products in drug discovery. *Drug Discov. Today* **13,** 894–901 (2008).
16. Patridge, E., Gareiss, P., Kinch, M. S. & Hoyer, D. An analysis of FDA-approved drugs: natural products and their derivatives. *Drug Discov. Today* **21,** 204–207 (2015).
17. Basu, S. & Waldmann, H. Polymer supported synthesis of a natural product-inspired oxepane library. *Bioorg. Med. Chem.* **22,** 4430–3333 (2014).
18. Huigens, R. W. III *et al.* A ring-distortion strategy to construct stereochemically complex and structurally diverse compounds from natural products. *Nature Chem.* **5,** 195–202 (2013).
19. Keller, N. P. Translating biosynthetic gene clusters into fungal armor and weaponry. *Nature Chem. Biol.* **11,** 671–677 (2015).
20. Medema, M. H. & Fischbach, M. A. Computational approaches to natural product discovery. *Nature Chem. Biol.* **11,** 639–648 (2015).
21. Kim, E., Moore, B. S. & Yoon, Y. J. Reinvigorating natural product combinatorial biosynthesis with synthetic biology. *Nature Chem. Biol.* **11,** 649–659 (2015).
22. Morrison, K. C. & Hergenrother, P. J. Natural products as starting points for the synthesis of complex and diverse compounds. *Nat. Prod. Rep.* **31,** 6–14 (2014).
23. Henkel, T., Brunne, R. M., Müller, H. & Reichel, F. Statistical investigation into the structural complementarity of natural products and synthetic compounds. *Angew. Chem. Int. Ed.* **38,** 643–647 (1999).
24. Ertl, P., Jelfs, S., Muhlbacher, J., Schuffenhauer, A. & Selzer, P. Quest for the rings. *In silico* exploration of ring universe to identify novel bioactive heteroaromatic scaffolds. *J. Med. Chem.* **49,** 4568–4573 (2006).
25. Akbulut, Y. *et al.* (–)-Englerin A is a potent and selective activator of TRPC4 and TRPC5 calcium channels. *Angew. Chem. Int. Ed.* **54,** 3787–3791 (2015).
26. Schuster, D., Laggner, C. & Langer, T. Why drugs fail — a study on side effects in new chemical entities. *Curr. Pharm. Des.* **11,** 3545–3559 (2005).
27. Hopkins, A. L., Keseru, G. M., Leeson, P. D., Rees, D. C. & Reynolds, C. H. The role of ligand efficiency metrics in drug discovery. *Nature Rev. Drug Discov.* **13,** 105–121 (2014).
28. Murray, C. W. & Rees, D. C. The rise of fragment-based drug discovery. *Nature Chem.* **1,** 187–192 (2009).
29. Grabowski, K., Baringhaus, K. H. & Schneider, G. Scaffold diversity of natural products: inspiration for combinatorial library design. *Nat. Prod. Rep.* **25,** 892–904 (2008).
30. Barelier, S. *et al.* Increasing chemical space coverage by combining empirical and computational fragment screens. *ACS Chem. Biol.* **9,** 1528–1535 (2014).
31. Reker, D. *et al.* Revealing the macromolecular targets of complex natural products. *Nature Chem.* **6,** 1072–1078 (2014).
32. Reutlinger, M., Rodrigues, T., Schneider, P. & Schneider, G. Multi-objective molecular *de novo* design by adaptive fragment prioritization. *Angew. Chem. Int. Ed.* **53,** 4244–4248 (2014).
33. Parkinson, E. I. *et al.* Deoxynybomycins inhibit mutant DNA gyrase and rescue mice infected with fluoroquinolone-resistant bacteria. *Nature Commun.* **6,** 6947 (2015).
34. Lipinski, C. A. Drug-like properties and the causes of poor solubility and poor permeability. *J. Pharmacol. Toxicol. Meth.* **44,** 235–249 (2000).
35. Congreve, M., Carr, R., Murray, C. & Jhoti, H. A rule of three for fragment-based lead discovery? *Drug Discov. Today* **8,** 876–877 (2003).
36. Baell, J., Congreve, M., Leeson, P. & Abad-Zapatero, C. Ask the experts: past, present and future of the rule of five. *Future Med. Chem.* **5,** 745–752 (2013).
37. Ntie-Kang, F., Lifongo, L. L., Judson, P. N., Sippl, W. & Efange, S. M. How "drug-like" are naturally occurring anti-cancer compounds? *J. Mol. Model.* **20,** 2069 (2014).
38. Kenny, P. W. & Montanari, C. A. Inflation of correlation in the pursuit of drug-likeness. *J. Comput. Aided Mol. Des.* **27,** 1–13 (2013).
39. Johnstone, C. Medicinal chemistry matters — a call for discipline in our discipline. *Drug Discov. Today* **17,** 538–543 (2012).
40. Quinn, R. J. *et al.* Developing a druglike natural product library. *J. Nat. Prod.* **71,** 464–468 (2008).
41. Antonchick, A. P. *et al.* Highly enantioselective synthesis and cellular evaluation of spirooxindoles inspired by natural products. *Nature Chem.* **2,** 735–740 (2010).
42. Svenda, J. *et al.* Biology-oriented synthesis of a withanolide-inspired compound collection reveals novel modulators of hedgehog signalling. *Angew. Chem. Int. Ed.* **54,** 5596–5602 (2015).
43. Schmid, F., Jessen, H. J., Burch, P. & Gademann, K. Truncated militarinone fragments identified by total chemical synthesis induce neurite outgrowth. *Med. Chem. Comm.* **4,** 135–139 (2013).
44. DeLorbe, J. E., Clements, J. H., Whiddon, B. B. & Martin, S. F. Thermodynamic and structural effects of macrocyclic constraints in protein–ligand interactions. *ACS Med. Chem. Lett.* **1,** 448–452 (2010).
45. Giordanetto, F. & Kihlberg, J. Macrocyclic drugs and clinical candidates: What can medicinal chemists learn from their properties? *J. Med. Chem.* **57,** 278–295 (2014).
46. Zhang, J. *et al.* Diversity-oriented synthesis of *Lycopodium* alkaloids inspired by the hidden functional group pairing pattern. *Nature Commun.* **5,** 4614 (2014).
47. Beckmann, H. S. *et al.* A strategy for the diversity-oriented synthesis of macrocyclic scaffolds using multidimensional coupling. *Nature Chem.* **5,** 861–867 (2013).
48. Karageorgis, G., Warriner, S. & Nelson, A. Efficient discovery of bioactive scaffolds by activity-directed synthesis. *Nature Chem.* **6,** 872–876 (2014).
49. McLeod, M. C. *et al.* Probing chemical space with alkaloid-inspired libraries. *Nature Chem.* **6,** 133–140 (2014).

50. Elumalai, N., Berg, A., Natarajan, K., Scharow, A. & Berg, T. Nanomolar inhibitors of the transcription factor STAT5b with high selectivity over STAT5a. *Angew. Chem. Int. Ed.* **54,** 4756–4763 (2015).
51. Larsson, J., Gottfries, J., Muresan, S. & Backlund, A. ChemGPS-NP: tuned for navigation in biologically relevant chemical space. *J. Nat. Prod.* **70,** 789–794 (2007).
52. Miyao, T., Reker, D., Schneider, P., Funatsu, K. & Schneider, G. Chemography of natural product space. *Planta Med.* **81,** 429–435 (2015).
53. Reutlinger, M. & Schneider, G. Nonlinear dimensionality reduction and mapping of compound libraries for drug discovery. *J. Mol. Graph. Model.* **34,** 108–117 (2012).
54. Bon, R. S. & Waldmann, H. Bioactivity-guided navigation of chemical space. *Acc. Chem. Res.* **43,** 1103–1114 (2010).
55. Lachance, H., Wetzel, S., Kumar, K. & Waldman, H. Charting, navigating, and populating natural product chemical space for drug discovery. *J. Med. Chem.* **55,** 5989–6001 (2012).
56. Takayama, H. *et al.* Discovery of inhibitors of the Wnt and Hedgehog signaling pathways through the catalytic enantioselective synthesis of an iridoid-inspired compound collection. *Angew. Chem. Int. Ed.* **52,** 12404–12408 (2013).
57. Dakas, P. Y. *et al.* Discovery of neuritogenic compound classes inspired by natural products. *Angew. Chem. Int. Ed.* **52,** 9576–9581 (2013).
58. Voigt, T. *et al.* A natural product inspired tetrahydropyran collection yields mitosis modulators that synergistically target CSE1L and tubulin. *Angew. Chem. Int. Ed.* **52,** 410–414 (2013).
59. Bennani, Y. L. Drug discovery in the next decade: innovation needed ASAP. *Drug Discov. Today* **16,** 779–792 (2011).
60. Wetzel, S. *et al.* Interactive exploration of chemical space with Scaffold Hunter. *Nature Chem. Biol.* **5,** 581–583 (2009).
61. Renner, S. *et al.* Bioactivity-guided mapping and navigation of chemical space. *Nature Chem. Biol.* **5,** 585–592 (2009).
62. Over, B. *et al.* Natural-product-derived fragments for fragment-based ligand discovery. *Nature Chem.* **5,** 21–28 (2013).
63. Bickerton, G. R., Paolini, G. V., Besnard, J., Muresan, S. & Hopkins, A. L. Quantifying the chemical beauty of drugs. *Nature Chem.* **4,** 90–98 (2012).
64. Nickel, J. *et al.* SuperPred: update on drug classification and target prediction. *Nucleic Acids Res.* **42,** W26–W31 (2014).
65. Keiser, M. J. *et al.* Predicting new molecular targets for known drugs. *Nature* **462,** 175–181 (2009).
66. Keiser, M. J. *et al.* Relating protein pharmacology by ligand chemistry. *Nature Biotechnol.* **25,** 197–206 (2007).
67. Gfeller, D. *et al.* SwissTargetPrediction: a web server for target prediction of bioactive small molecules. *Nucleic Acids Res.* **42,** W32–W38 (2014).
68. Reutlinger, M., Rodrigues, T., Schneider, P. & Schneider, G. Combining on-chip synthesis of a focused combinatorial library with computational target prediction reveals imidazopyridine GPCR ligands. *Angew. Chem. Int. Ed.* **53,** 582–585 (2014).
69. Antolin, A. A., Jalencas, X., Yelamos, J. & Mestres, J. Identification of pim kinases as novel targets for PJ34 with confounding effects in PARP biology. *ACS Chem. Biol.* **7,** 1962–1967 (2012).
70. Hopkins, A. L. Network pharmacology: the next paradigm in drug discovery. *Nature Chem. Biol.* **4,** 682–690 (2008).
71. Lounkine, E. *et al.* Large-scale prediction and testing of drug activity on side-effect targets. *Nature* **486,** 361–367 (2012).
72. Rollinger, J. M. Accessing target information by virtual parallel screening — the impact on natural product research. *Phytochem. Lett.* **2,** 53–58 (2009).
73. Do, Q. T. *et al.* Reverse pharmacognosy: identifying biological properties for plants by means of their molecule constituents: application to meranzin. *Planta Med.* **73,** 1235–1240 (2007).
74. Vuorinen, A., Nashev, L. G., Odermatt, A., Rollinger, J. M. & Schuster, D. Pharmacophore model refinement for 11β-hydroxysteroid dehydrogenase inhibitors: search for modulators of intracellular glucocorticoid concentrations. *Mol. Inf.* **33,** 15–25 (2014).
75. Bauer, J. *et al.* Discovery of depsides and depsidones from lichen as potent inhibitors of microsomal prostaglandin E2 synthase-1 using pharmacophore models. *ChemMedChem* **7,** 2077–2081 (2012).
76. Rollinger, J. M. *et al. In silico* target fishing for rationalized ligand discovery exemplified on constituents of *Ruta graveolens*. *Planta Med.* **75,** 195–204 (2009).
77. Schuster, D. *et al.* Morphinans and isoquinolines: acetylcholinesterase inhibition, pharmacophore modeling, and interaction with opioid receptors. *Bioorg. Med. Chem.* **18,** 5071–5080 (2010).
78. Fakhrudin, N. *et al.* Computer-aided discovery, validation, and mechanistic characterization of novel neolignan activators of peroxisome proliferator-activated receptor gamma. *Mol. Pharmacol.* **77,** 559–566 (2010).
79. Atanasov, A. G. *et al.* Honokiol: a non-adipogenic PPARγ agonist from nature. *Biochim. Biophys. Acta* **1830,** 4813–4819 (2013).
80. Arrowsmith, C. H. *et al.* The promise and peril of chemical probes. *Nature Chem. Biol.* **11,** 536–541 (2015).
81. Lagunin, A., Stepanchikova, A., Filimonov, D. & Poroikov, V. PASS: prediction of activity spectra for biologically active substances. *Bioinformatics* **16,** 747–748 (2000).
82. Lagunin, A., Filimonov, D. & Poroikov, V. Multi-targeted natural products evaluation based on biological activity prediction with PASS. *Curr. Pharm. Des.* **16,** 1703–1717 (2010).
83. Filimonov, D., Poroikov, V., Borodina, Y. & Gloriozova, T. Chemical similarity assessment through multilevel neighbourhoods of atoms: definition and comparison with the other descriptors. *J. Chem. Inf. Comput. Sci.* **39,** 666–670 (1999).
84. Chen, X. *et al.* Drug–target interaction prediction: databases, web servers and computational models. *Brief. Bioinform.* http://dx.doi.org/10.1093/bib/bbv066 (2015).
85. Liu, X. *et al.* Predicting targeted polypharmacology for drug repositioning and multi-target drug discovery. *Curr. Med. Chem.* **20,** 1646–1661 (2013).
86. Pahikkala, T. *et al.* Toward more realistic drug–target interaction predictions. *Brief. Bioinform.* **16,** 325–337 (2015).
87. Mousavian, Z. & Masoudi-Nejad, A. Drug–target interaction prediction via chemogenomic space: learning-based methods. *Expert Opin. Drug Metab. Toxicol.* **10,** 1273–1287 (2014).
88. Yamanishi, Y. Chemogenomic approaches to infer drug-target interaction networks. *Meth. Mol. Biol.* **939,** 97–113 (2013).
89. Rupp, M. *et al.* Machine learning estimates of natural product conformational energies. *PLoS Comput. Biol.* **10,** e1003400 (2014).
90. Reker, D., Rodrigues, T., Schneider, P. & Schneider, G. Identifying the macromolecular targets of *de novo*-designed chemical entities through self-organizing map consensus. *Proc. Natl Acad. Sci. USA* **111,** 4067–4072 (2014).
91. Schneider, G., Reker, D., Rodrigues, T. & Schneider, P. Coping with polypharmacology by computational medicinal chemistry. *Chimia* **68,** 648–653 (2014).
92. Reutlinger, M. *et al.* Chemically advanced template search (CATS) for scaffold-hopping and prospective target prediction for 'orphan' molecules. *Mol. Inf.* **32,** 133–138 (2013).
93. Reker, D. *et al.* Deorphaning pyrrolopyrazines as potent multi-target antimalarial agents. *Angew. Chem. Int. Ed.* **53,** 7079–7084 (2014).
94. Rodrigues, T., Reker, D., Kunze, J., Schneider, P. & Schneider, G. Revealing the macromolecular targets of fragment-like natural products. *Angew. Chem. Int. Ed.* **54,** 10516–10520 (2015).
95. Imming, P. Molecular targets of natural drug substances: idiosyncrasies and preferences. *Planta Med.* **76,** 1794–1801 (2010).
96. Koeberle, A. & Werz, O. Multi-target approach for natural products in inflammation. *Drug Discov. Today* **19,** 1871–1882 (2014).
97. Geldenhuys, W. J. & Van der Schyf, C. J. Rationally designed multi-targeted agents against neurodegenerative diseases. *Curr. Med. Chem.* **20,** 1662–1672 (2013).
98. Schmeller, T., Sauerwein, M., Sporer, F., Wink, M. & Muller, W. E. Binding of quinolizidine alkaloids to nicotinic and muscarinic acetylcholine receptors. *J. Nat. Prod.* **57,** 1316–1319 (1994).
99. Kirchmair, J. *et al.* Predicting drug metabolism: experiment and/or computation? *Nature Rev. Drug Discov.* **14,** 387–404 (2015).
100. Gertsch, J. Botanical drugs, synergy, and network pharmacology: forth and back to intelligent mixtures. *Planta Med.* **77,** 1086–1098 (2011).
101. Mak, L. *et al.* Anti-cancer drug development: computational strategies to identify and target proteins involved in cancer metabolism. *Curr. Pharm. Des.* **19,** 532–577 (2013).
102. de Sá Alves, F. R., Barreiro, E. J. & Fraga, C. A. From nature to drug discovery: the indole scaffold as a 'privileged structure'. *Mini Rev. Med. Chem.* **9,** 782–793 (2009).
103. Venkatraman, P. Specificity in the interaction of natural products with their target proteins — a biochemical and structural insight. *Mini Rev. Med. Chem.* **10,** 540–549 (2010).
104. Mishra, B. B. & Tiwari, V. K. Natural products: an evolving role in future drug discovery. *Eur. J. Med. Chem.* **46,** 4769–4807 (2011).
105. Jayaseelan, K. V., Moreno, P., Truszkowski, A., Ertl, P. & Steinbeck, C. Natural product-likeness score revisited: an open-source, open-data implementation. *BMC Bioinform.* **13,** 106 (2012).
106. Roche, O. *et al.* Development of a virtual screening method for identification of "frequent hitters" in compound libraries. *J. Med. Chem.* **45,** 137–142 (2002).
107. Coan, K. E. & Shoichet, B. K. Stoichiometry and physical chemistry of promiscuous aggregate-based inhibitors. *J. Am. Chem. Soc.* **130,** 9606–9612 (2008).
108. Seidler, J., McGovern, S. L., Doman, T. N. & Shoichet, B. K. Identification and prediction of promiscuous aggregating inhibitors among known drugs. *J. Med. Chem.* **46,** 4477–4486 (2003).

109. Baell, J. & Walters, M. A. Chemistry: chemical con artists foil drug discovery. *Nature* **513,** 481–483 (2014).
110. Petrova, T., Chuprina, A., Parkesh, R. & Pushechnikov, A. Structural enrichment of HTS compounds from available commercial libraries. *Med. Chem. Comm.* **3,** 571–579 (2012).
111. Walters, W. P., Stahl, M. T. & Murcko, M. A. Virtual screening — an overview. *Drug Discov. Today* **3,** 160–178 (1998).
112. Duan, D., Doak, A. K., Nedyalkova, L. & Shoichet, B. K. Colloidal aggregation and the *in vitro* activity of traditional Chinese medicines. *ACS Chem. Biol.* **10,** 978–988 (2015).
113. Wishart, D. S. *et al.* DrugBank: a knowledgebase for drugs, drug actions and drug targets. *Nucleic Acids Res.* **36,** D901–D906 (2008).
114. Bento, A. P. *et al.* The ChEMBL bioactivity database: an update. *Nucleic Acids Res.* **42,** D1083–D1090 (2014).
115. Chen, C. Y. TCM Database@Taiwan: the world's largest traditional Chinese medicine database for drug screening *in silico*. *PLoS ONE* **6,** e15939 (2011).
116. Rishton, G. M. Reactive compounds and *in vitro* false positives in HTS. *Drug Discov. Today* **2,** 382–384 (1997).
117. Hann, M. *et al.* Strategic pooling of compounds for high-throughput screening. *J. Chem. Inf. Comput. Sci.* **39,** 897–902 (1999).
118. Baell, J. B. & Holloway, G. A. New substructure filters for removal of pan assay interference compounds (PAINS) from screening libraries and for their exclusion in bioassays. *J. Med. Chem.* **53,** 2719–2740 (2010).
119. Lagorce, D., Sperandio, O., Baell, J. B., Miteva, M. A. & Villoutreix, B. O. FAF-Drugs3: a web server for compound property calculation and chemical library design. *Nucleic Acids Res.* **43,** W200–W207 (2015).
120. Jost, C., Nitsche, C., Scholz, T., Roux, L. & Klein, C. D. Promiscuity and selectivity in covalent enzyme inhibition: a systematic study of electrophilic fragments. *J. Med. Chem.* **57,** 7590–7599 (2014).
121. Schneider, P., Röthlisberger, M., Reker, D. & Schneider, G. Spotting and designing promiscuous ligands for drug discovery. *Chem. Commun.* **52,** 1135–1138 (2016).
122. Clemons, P. A. *et al.* Small molecules of different origins have distinct distributions of structural complexity that correlate with protein-binding profiles. *Proc. Natl Acad. Sci. USA* **107,** 18787–18792 (2010).
123. Kell, D. B. Finding novel pharmaceuticals in the systems biology era using multiple effective drug targets, phenotypic screening and knowledge of transporters: where drug discovery went wrong and how to fix it. *FEBS J.* **280,** 5957–5980 (2013).
124. Rollinger, J. M., Stuppner, H. & Langer, T. Virtual screening for the discovery of bioactive natural products. *Prog. Drug Res.* **65,** 213–249 (2008).
125. Harvey, A. L., Edrada-Ebel, R. & Quinn, R. J. The re-emergence of natural products for drug discovery in the genomics era. *Nature Rev. Drug Discov.* **14,** 111–129 (2015).
126. Cragg, G. M., Grothaus, P. G. & Newman, D. J. New horizons for old drugs and drug leads. *J. Nat. Prod.* **77,** 703–723 (2014).
127. Schuster, D. & Wolber, G. Identification of bioactive natural products by pharmacophore-based virtual screening. *Curr. Pharm. Des.* **16,** 1666–1681 (2010).
128. Jacoby, E. & Mozzarelli, A. Chemogenomic strategies to expand the bioactive chemical space. *Curr. Med. Chem.* **16,** 4374–4381 (2009).
129. Hübel, K., Lessmannm, T. & Waldmann, H. Chemical biology — identification of small molecule modulators of cellular activity by natural product inspired synthesis. *Chem. Soc. Rev.* **37,** 1361–1374 (2008).
130. Kumar, K. & Waldmann, H. Synthesis of natural product inspired compound collections. *Angew. Chem. Int. Ed.* **48,** 3224–3242 (2009).
131. Hirai, G. Mimicking/extracting structure and functions of natural products: synthetic approaches that address unexplored needs in chemical biology. *Chem. Rec.* **15,** 445–456 (2015).
132. Thomas, G. L. & Johannes, C. W. Natural product-like synthetic libraries. *Curr. Opin. Chem. Biol.* **15,** 516–522 (2011).
133. Gupta, P. K., Barone, G., Gurley, B. J., Fifer, E. K. & Hendrickson, H. P. Hydrastine pharmacokinetics and metabolism after a single oral dose of goldenseal (*Hydrastis canadensis*) to humans. *Drug Metab. Dispos.* **43,** 534–552 (2015).
134. Bier, D., Thiel, P., Briels, J. & Ottmann, C. Stabilization of protein–protein interactions in chemical biology and drug discovery. *Prog. Biophys. Mol. Biol.* **119,** 10–19 (2015).
135. Klenner, A., Hartenfeller, M., Schneider, P. & Schneider, G. 'Fuzziness' in pharmacophore-based virtual screening and *de novo* design. *Drug Discov. Today Technol.* **7,** e237–e244 (2010).
136. Lewell, X. Q., Judd, D. B., Watson, S. P. & Hann, M. M. RECAP — retrosynthetic combinatorial analysis procedure: a powerful new technique for identifying privileged molecular fragments with useful applications in combinatorial chemistry. *J. Chem. Inf. Comput. Sci.* **38,** 511–522 (1998).
137. Kusari, S. *et al.* Tramadol — a true natural product? *Angew. Chem. Int. Ed. Engl.* **53,** 12073–12076 (2014).
138. Talalay, P., De Long, M. J. & Prochaska, H. J. Identification of a common chemical signal regulating the induction of enzymes that protect against chemical carcinogenesis. *Proc. Natl Acad. Sci. USA* **85,** 8261–8265 (1988).
139. Aptula, A. O. & Roberts, D. W. Mechanistic applicability domains for nonanimal-based prediction of toxicological end points: general principles and application to reactive toxicity. *Chem. Res. Toxicol.* **19,** 1097–1105 (2006).
140. Amslinger, S. The tunable functionality of α,β-unsaturated carbonyl compounds enables their differential application in biological systems. *ChemMedChem* **5,** 351–356 (2010).
141. Avonto, C. *et al.* An NMR spectroscopic method to identify and classify thiol-trapping agents: revival of Michael acceptors for drug discovery? *Angew. Chem. Int. Ed.* **50,** 467–471 (2011).

Acknowledgements

We thank K.-H. Altmann for invaluable advice and discussion. This study was financially supported by the OPO Foundation, Zürich.

Author contributions

All authors contributed equally.

Additional information

Reprints and permissions information is available online at www.nature.com/reprints. Correspondence and requests for materials should be addressed to G.S.

Competing financial interests

P.S. and G.S. are the founders of inSili.com LLC, Zürich. All the other authors declare no competing financial interests.