

# Chemical Evolution of Natural Product Structure

Michael Grigalunas, Susanne Brakmann, and Herbert Waldmann\*



Cite This: *J. Am. Chem. Soc.* 2022, 144, 3314–3329



Read Online

ACCESS |



Metrics & More



Article Recommendations



Supporting Information

**ABSTRACT:** Natural products are the result of Nature's exploration of biologically relevant chemical space through evolution and an invaluable source of bioactive small molecules for chemical biology and medicinal chemistry. Novel concepts for the discovery of new bioactive compound classes based on natural product structure may enable exploration of wider biologically relevant chemical space. The pseudo-natural product concept merges the relevance of natural product structure with efficient exploration of chemical space by means of fragment-based compound development to inspire the discovery of new bioactive chemical matter through *de novo* combination of natural product fragments in unprecedented arrangements. The novel scaffolds retain the biological relevance of natural products but are not obtainable through known biosynthetic pathways which can lead to new chemotypes that may have unexpected or unprecedented bioactivities. Herein, we cover the workflow of pseudo-natural product design and development, highlight recent examples, and discuss a cheminformatic analysis in which a significant portion of biologically active synthetic compounds were found to be pseudo-natural products. We compare the concept to natural evolution and discuss pseudo-natural products as the human-made equivalent, i.e. the chemical evolution of natural product structure.

## INTRODUCTION

The importance of small molecules that can perturb biological systems in a controlled fashion is underpinned by their successful applications in chemical biology<sup>1</sup> and medicine.<sup>2</sup> Nevertheless, the continued exploration of biologically relevant chemical space for the discovery of new small molecules remains an essential task for the better understanding of biological processes and for the advancement of therapeutics. This is a challenging endeavor since the possible number of compounds that may qualify as bioactive small molecules is so vast that the time and resources needed to synthesize, let alone biologically evaluate, exceed reality.<sup>3</sup> Therefore, there is a need for methods and sources of inspiration that can help navigate through chemical space and focus on areas that are biologically relevant.<sup>4</sup>

Through evolution, Nature has been exploring chemical space which is reflected in secondary metabolites or natural products (NPs). Via enzymatic cascades, organisms produce NPs that can carry out specific biological functions either within and/or between organisms to give the producing organism a selective advantage. Based on the demands of the environment, i.e. selectivity pressures, NPs with different molecular scaffolds can be produced that modulate different targets directly or indirectly to provide the organism with a higher level of fitness. Therefore, NPs reflect organisms' adaptations to their environments through evolution and Nature's exploration of biologically relevant chemical space. Accordingly, NP structure can be considered biologically prevalidated.

NPs have been and continue to be a reservoir for successful molecular discovery programs;<sup>5</sup> however, NPs have been subjected to evolutionary constraints. Natural evolution itself is very slow and, in tandem with selection pressures, has resulted in NPs occupying only a fraction of theoretical NP-like

chemical space.<sup>6</sup> These constraints may have led to many known NP scaffolds being highly conserved and therefore the diversity of scaffolds readily accessible from Nature to be limited. Additionally, there are undoubtedly numerous biologically intriguing NPs that have not yet been identified. Therefore, solely relying on Nature as a source of biologically active molecules brings about limitations.

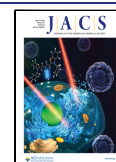
Attempts to override the limitations of Nature and expand beyond the chemical space occupied by NPs to provide new small molecules with unexpected or unprecedented bioactivities need to face the question how biologically relevant chemical space can be explored efficiently?<sup>7</sup>

Several chemical design strategies have been developed for the search of biologically active compounds.<sup>8</sup> Diversity-oriented synthesis (DOS) provides chemically diverse libraries enriched in variations of scaffolds, stereochemistry, and appendages to quickly probe larger amounts of chemical space.<sup>9</sup> A combinatorial approach utilizing DNA encoded libraries (DEL) has emerged as a method for the generation of relatively large libraries with chemical diversity.<sup>10</sup> While DOS and DEL can generate expansive libraries that can quickly explore diverse chemical space, more recently in tandem,<sup>11</sup> and have provided biologically valuable compounds,<sup>12,13</sup> much of the space explored may not necessarily be biologically relevant.

Several approaches prioritize exploring more focused areas of chemical space that are occupied by NPs of biological interest. Access to this space is facilitated by the design of more

Received: October 25, 2021

Published: February 21, 2022



synthetically tractable derivatives relative to the parent NP. Compounds resulting from function-oriented synthesis<sup>14,15</sup> have reduced complexity relative to the parent NP while those from dynamic retrosynthetic analysis proposed by Shenvi et al.<sup>16,17</sup> retain the intrinsic complexity of the parent NP while minimizing synthetic complexity. While both methods can yield compounds with similar or enhanced biological properties, the chemical and biological space explored is narrow and usually similar to the parent NP's.

Other design principles aim to provide molecules that maintain the biological relevance of NPs by employing logic derived from natural evolution.<sup>18,19</sup> Biology-oriented synthesis (BIOS) is based on the principle that proteins and NPs have coevolved to encompass only a small portion of chemical space.<sup>20–22</sup> While appendages of NPs can be highly diverse, the core scaffolds of NPs are conserved as are the corresponding binding pockets of proteins. Using this logic, BIOS employs hierarchical classification to identify simplified NP core structures that may retain biologically relevant characteristics.<sup>23</sup> Decoration of these scaffolds has led to several biologically active compound collections that retain relevance to NPs but are more synthetically tractable (Figure 1a); however, BIOS is limited both biologically and chemically because the core scaffolds are present in current NPs obtained through existing biosynthetic pathways.

In Nature, common NP intermediates can undergo different biosynthetic pathways to arrive at chemically and biologically diverse NPs as “end points”. The complexity to diversity approach employs “end point” NPs as divergent intermediates and chemically extends biosynthetic pathways by distorting the scaffolds of NPs through ring distortion reactions (Figure 1a).<sup>29–31</sup> The resulting compound collections have diverse scaffolds that are distinct from the parent NP but retain the complexity of and biological relevance to NPs; however, the concept may be limited by the finite number of possible starting points.

While all of these methods have been successful in providing compound libraries that are enriched in bioactivities, each has, to some extent, limitations that hinder wider exploration of biologically relevant chemical space. To expand beyond these limitations while still retaining relevance to NPs, we have developed the concept of pseudo-natural products (pseudo-NPs).<sup>19,32–34</sup> The principle merges the biological relevance of NPs with the rapid accessibility to diverse chemical space of fragment-based discovery. Through cheminformatic deconstruction of NPs, we have defined about 2,000 NP fragment groups that represent the chemical space defined by the structures of the NPs known at the time.<sup>35</sup> The pseudo-NP method aims for the *de novo* combination of NP fragments or NPs that are themselves fragment-sized<sup>35–38</sup> in a manner that is not observed in current NP structures (Figure 1b). This can be achieved through the mixing of different fragment combinations<sup>39</sup> and/or the combination of fragments in different orientations.<sup>40</sup> The new scaffolds retain the chemical and biological relevance of NPs but are not possible through known biosynthetic pathways, hence the name pseudo-NPs.

Notably, the biological goal of pseudo-NP design differs from the synthetic NP-hybridization strategy. The NP-hybridization combines pharmacophores of two NPs that have different but synergistic biological effects. The resulting chimeric molecules can have enhanced therapeutic value but most likely retain the mode of actions and/or biological targets of the two parent NPs.<sup>41–44</sup> In contrast, the new arrangements

of NP fragments in the resulting pseudo-NPs are intended to explore new areas of biologically relevant chemical space and have biological targets that are unrelated to the NP fragments and NPs from which they are derived.

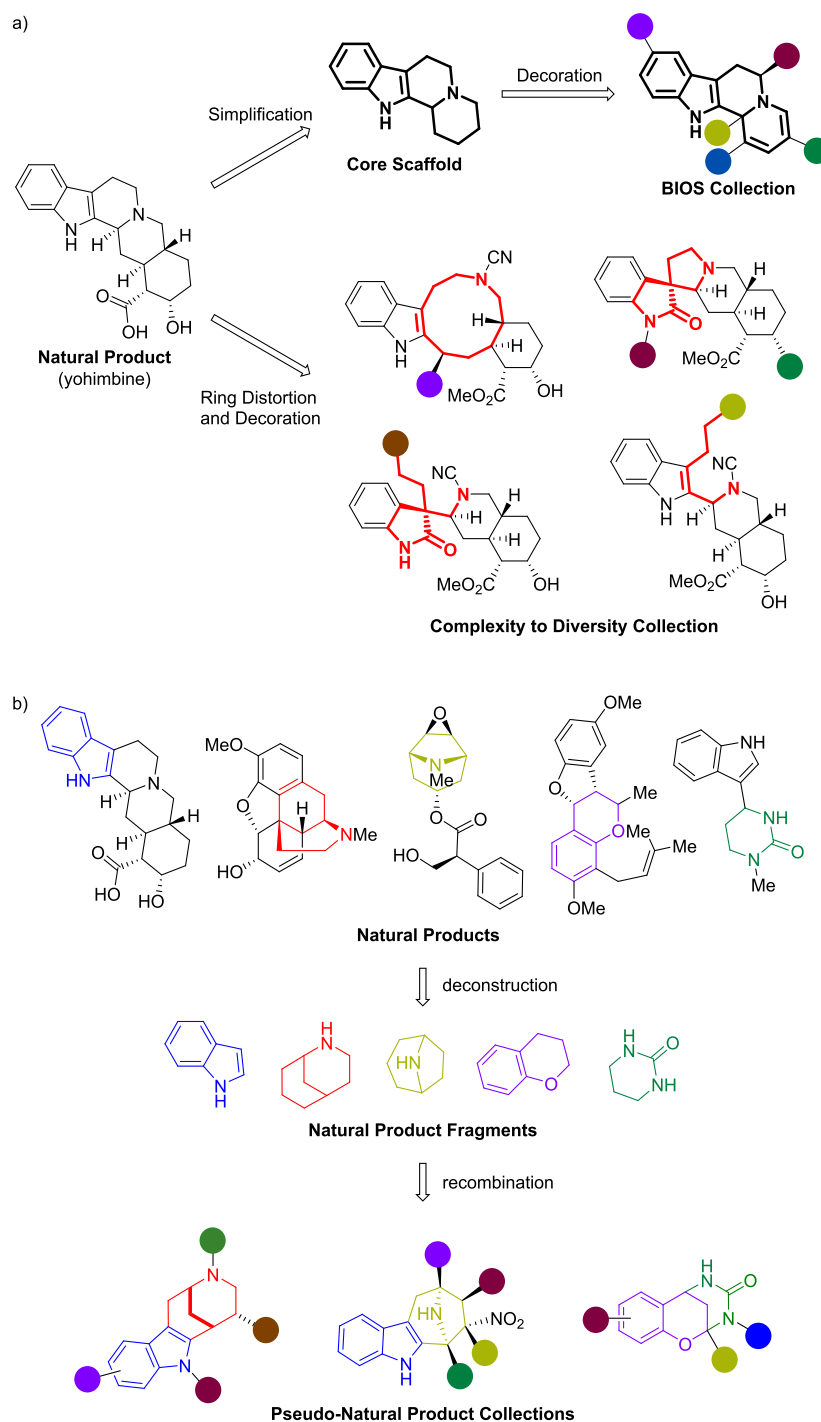
In this perspective, we discuss the design principles and biological evaluation of pseudo-NP collections as well as highlight recent examples. We discuss that, through cheminformatic analysis, pseudo-NPs have been made unintentionally for decades and comprise a significant portion of synthetically made bioactive small molecules. We then tie our studies of pseudo-NPs to the concept of chemical evolution, and argue that the method can be considered the manmade equivalent of NP structure evolution. Taken all together, we argue that the pseudo-NP concept is a valid principle for the wider exploration of biologically relevant chemical space.

## ■ DESIGN AND BIOLOGICAL EVALUATION OF PSEUDO-NATURAL PRODUCT COLLECTIONS

The general design principle of pseudo-NPs revolves around the combination of NP fragments to arrive at scaffolds that resemble NPs but are not possible through known biosynthetic pathways. Pseudo-NPs should share some chemical similarities with NPs in that they are both comprised of NP fragments and should therefore resemble the relatively high fraction of  $sp^3$  atoms and abundant stereogenicity observed in NP structure.<sup>45,46</sup> On the other hand, pseudo-NPs should have a degree of chemical distinguishability since fragment combinations and/or orientations in pseudo-NPs are distinct from NPs.

The combination partners should be derived from different biosynthetic origins and/or have different heteroatom content (O and N) to ensure the exploration of new chemical space. It should be considered that criteria typically employed by the “rule of three”<sup>47</sup> for fragments may not be entirely valid for NP fragments.<sup>48</sup> Therefore, NP-like fragments can be considered to have an AlogP < 3.5, a molecular weight between 120 and 350 Da,  $\leq 3$  hydrogen bond donors,  $\leq 6$  hydrogen bond acceptors, and  $\leq 6$  rotatable bonds.<sup>35</sup> Within these criteria, NP fragments can include fully synthetic fragments, NPs that are themselves fragment-sized, and products resulting from the distortion or fragmentation of NPs (Figure 2a).<sup>49</sup> Suitable NP fragments could have reactive handles that are amenable for the installation of additional NP fragments through complexity generating reactions. Alternatively, pseudo-NP scaffolds and the generation of fragments can be simultaneously constructed through intramolecular reactions.

Chemically diverse pseudo-NP collections can be designed by using different combination strategies (Figure 2b). Novel scaffolds can arise through the mixing and matching of different NP fragments. Additionally, combining similar fragments with different connectivity patterns can add further degrees of variability from known NP structures. Several fragment connectivity patterns are observed in Nature and can be translated to pseudo-NPs. These can be classified into two categories: connections with shared atom(s) and connections without shared atoms. Those with shared atoms include fusion spiro (one shared atom), fusion edge (two shared atoms), and fusion bridged (three or more shared atoms). Fusion patterns without shared atoms include but are not limited to monopodal, bipodal, tripodal, bridged bipodal, and bridged tripodal.<sup>33</sup> A further degree of scaffold diversification is possible by combining similar fragments with similar connectivity patterns but having different regioisomeric



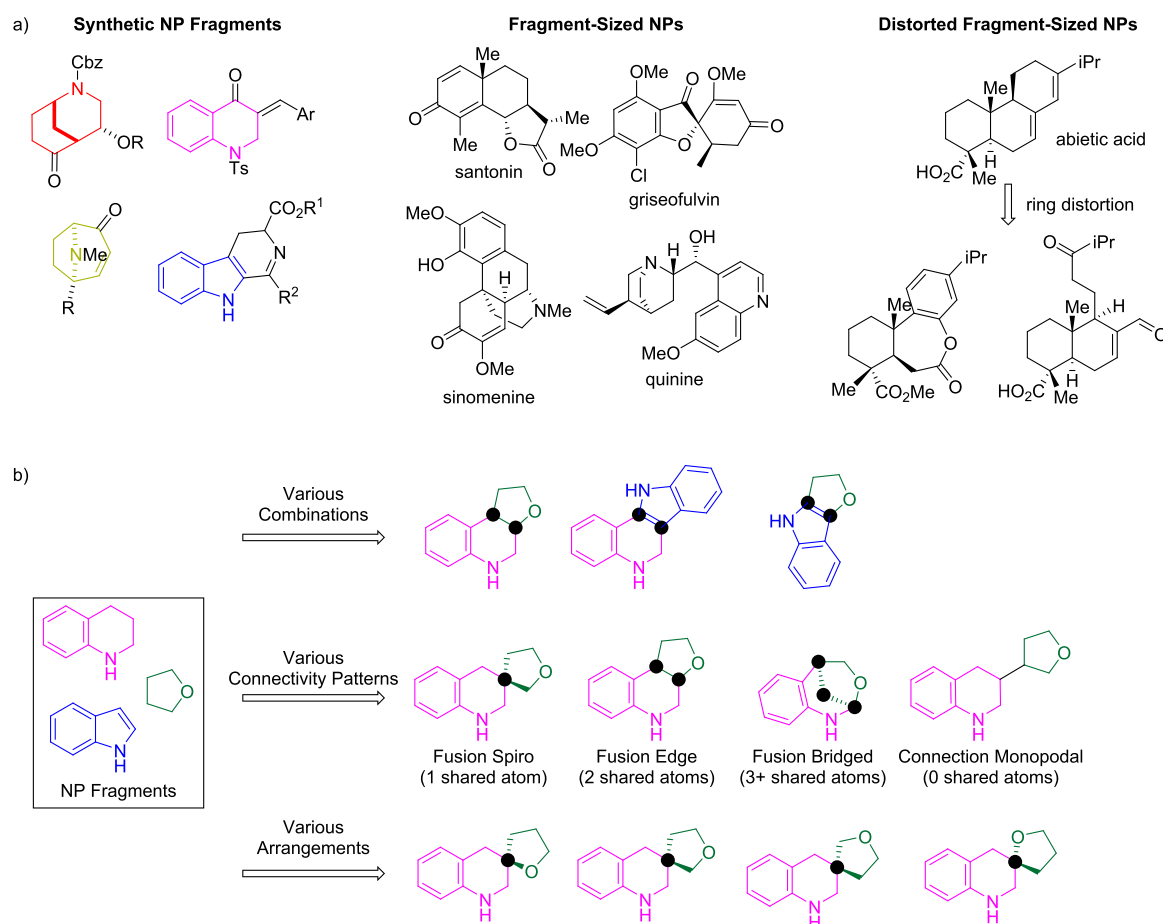
**Figure 1.** (a) Examples of biology-oriented synthesis<sup>24</sup> (BIOS) and the complexity to diversity<sup>25</sup> methods employing the natural product yohimbine as a starting point. (b) Examples of the pseudo-natural product (pseudo-NP) method by computational deconstruction of NPs to NP fragments and *de novo* recombination.<sup>26–28</sup>

arrangements. Taken together, the combination of different NP fragments with different connection types and/or different regioisomeric arrangements may lead to chemically and biologically diverse pseudo-NP collections.

Since their scaffolds are novel, pseudo-NPs may lead to new chemotypes for known biological targets or to the discovery of unknown biological targets. To efficiently probe broad areas of biological space, biological evaluation of pseudo-NPs should employ unbiased, target-agnostic assays.<sup>51</sup> We have successfully employed different cell-based phenotypic assays to

identify pseudo-NPs that affect important cellular processes or signaling cascades, including monitoring glucose uptake, autophagy, Wnt and Hedgehog signaling, T-cell differentiation, and induction of reactive oxygen species.

Beyond phenotypic assays, morphological profiling via the “Cell Painting Assay” can be used to evaluate compound-induced morphological changes in the entire cell.<sup>52,53</sup> In the Cell Painting Assay, images of treated cells are acquired through fluorescent microscopy followed by the extraction of several hundred features to provide a characteristic “finger-



**Figure 2.** (a) Possible fragment types for combination in pseudo-NP design: synthetic NP fragments, fragment-sized NPs, and ring distorted fragment-sized NPs.<sup>50</sup> (b) Fragment combination strategies to arrive at diverse pseudo-NPs including mixing fragment combinations, different connectivity patterns, and/or varying regioisomeric arrangements. The black dots indicate common atoms due to the fusion patterns.

print” that encompasses the cells’ morphological changes. Compound-induced morphological perturbation can be quantified as an “induction value” by calculating the number of significantly changed features relative to a DMSO control. Compounds that have suitable induction values are considered bioactive. The fingerprints can also be compared to reference compounds with known bioactivities to generate target/modes of action hypotheses or compared within and between compound collections for biological performance.

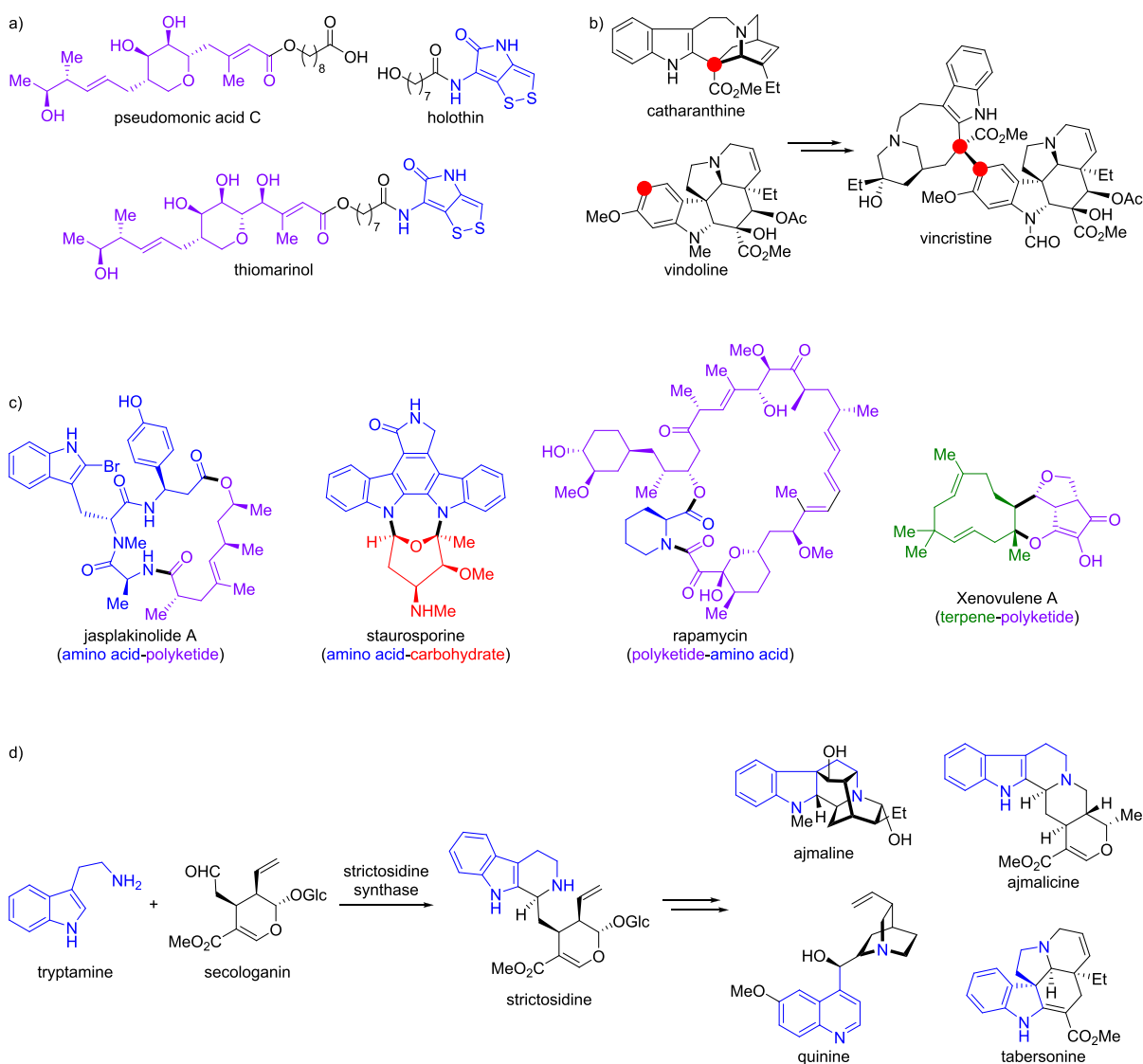
## ■ HYBRID NATURAL PRODUCT STRATEGIES ARE EMPLOYED BY NATURE

The concept of combining privileged NP scaffolds, i.e. fragment-sized NPs and/or NP fragments, to provide “hybrid NPs” has also been explored by Nature. Hybrid NPs may be characterized as the product of homo- or heterodimerizations of smaller NPs in which the monomeric NPs could be, but are not necessarily, from different biosynthetic origins.<sup>43</sup> Dimerized NPs are typically combined via a monopodal connection and retain resemblance to the core scaffolds of the monomeric NPs, such as thiomarinol and vincristine. The structure of thiomarinol resembles the combination of biosynthetic derivatives of the antibiotics pseudomonic acid C and holothin (Figure 3a).<sup>54,55</sup> While the monomers represent two separate classes of antibiotics, their incorporation into what is the scaffold of thiomarinol gives a more potent, broader spectrum antibiotic. This enhanced antibiotic activity may be explained

by the antibiotic properties of the two monomers being retained in thiomarinol, resulting in a synergistic effect. The antimetabolic drug vincristine is used for the treatment of leukemia and lymphomas<sup>56</sup> and is the combination of the indole alkaloids catharanthine and vindoline (Figure 3b).<sup>57</sup> Interestingly, catharanthine and/or vindoline induce mitotic arrest only at concentrations several orders of magnitude greater than vincristine.<sup>58</sup> The significant difference in potency of vincristine may be rationalized by a binding mode to tubulin that is not possible with solely the monomers.<sup>59</sup>

Other types of hybrid NPs fuse biosynthetically unrelated metabolic units (or fragments),<sup>60</sup> including but not limited to amino acids, carbohydrates, polyketides, and terpenoids, to form the hybrid NP core scaffold (Figure 3c).<sup>61–64</sup> These compounds offer diversity through different fragment combinations that biosynthetic pathways of a single origin cannot provide. After combination, biosynthetic cascades can further distort and fuse the original metabolic units into complex scaffolds that do not intuitively resemble the original pieces. Strictosidine is formed through the combination of the amino acid derivative tryptamine and the monoterpene glycoside secologanin and is the biosynthetic divergent intermediate for thousands of monoterpene alkaloids with complex multipodal fusion patterns (Figure 3d).<sup>65</sup>

The NP hybridization approach may be viewed as the combination of biosynthetic strategies Nature has compiled to produce new molecules. This strategy has been proven by



**Figure 3.** Naturally occurring hybrid natural products. (a) The hybrid natural product thiomarinol resembles the combination of pseudomonadic acid C and holothin. (b) The biosynthesis of vincristine includes the heterodimerization of the natural products catharanthine and vindoline. Points of connectivity are highlighted in red. (c) Selected hybrid natural products that are the combination of more than one metabolic unit of different biosynthetic origin. (d) Representative monoterpene alkaloids from the divergent biosynthetic intermediate strictosidine. Strictosidine is the combination of the amino acid derivative tryptamine and the monoterpene glycoside secologanin. Atoms from tryptamine are highlighted in blue.

evolution to successfully access compounds that explore new areas of biologically relevant chemical space that have enhanced or new biological properties relative to their fragments. Similarly, the pseudo-NP concept takes Nature's compilation of biologically prevalidated NP fragments, which results from these same biosynthetic pathways, and combines them in new arrangements not found in Nature.

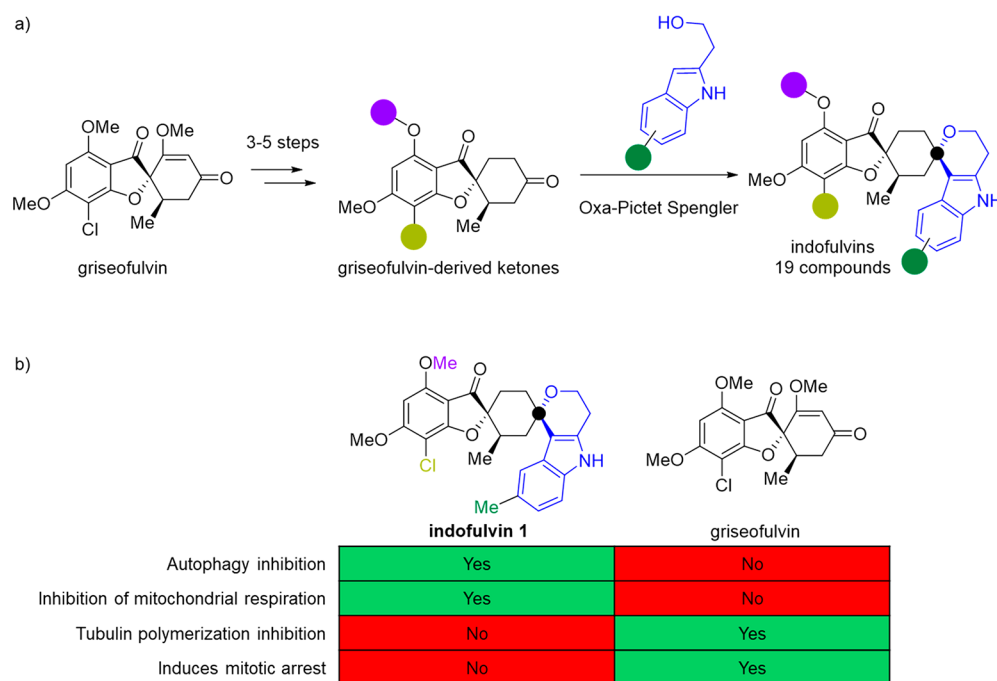
Where these two approaches diverge is in the preparation of new molecules. Nature synthesizes NP hybrids through enzymatic reactions that are summarized as biosynthetic pathways. Through evolution, new biosynthetic pathways can arise that may yield new NP hybrids. On the other hand, pseudo-NPs are designed by deconstruction of NPs to NP fragments through cheminformatic analyses. De novo recombinations are envisioned and synthesized via chemical reactions that may not occur in Nature.

Both methods may produce biologically relevant chemical matter; however, evolution of biosynthetic pathways in Nature to provide new molecules is slow while synthesis allows this

process to be accelerated and more flexible. By employing chemical synthesis, reaction pathways can be employed that are not possible through biosynthesis resulting in new combinations and connectivities of NP fragments. Therefore, the pseudo-NP concept does not necessarily compete with NP hybridization but rather complements it and explores areas of biologically relevant chemical space that are not possible through existing biosynthetic pathways.

## RECENT EXAMPLES OF PSEUDO-NATURAL PRODUCTS

**Indofulvins.** A recent example of the pseudo-NP concept combined the indole-containing fragment 4*H*-pyranoidole and the fragment-sized NP griseofulin to arrive at indofulvins (Figure 4a).<sup>66</sup> A collection of 19 indofulvins was obtained by employing an iso-oxa-Pictet Spengler (IOPS) reaction as the key step to combine the two fragments with a fusion spiro connection pattern. While the Pictet Spengler reaction is



**Figure 4.** (a) Synthesis of the indofulvin pseudo-NP collection through the combination of the fragment-sized NP griseofulvin and 4*H*-pyranoindole derivatives. (b) Biological comparisons of the pseudo-NP **indofulvin 1** and griseofulvin showing that the combination of fragments has led to new bioactivity of autophagy inhibition potentially through mitochondrial respiration while simultaneously losing the native tubulin-affecting bioactivity of griseofulvin. The black dot indicates a common atom due to the fusion pattern.

present in biosynthetic pathways, its iso-version is not currently in the biosynthetic repertoire, resulting in a novel arrangement of the 4*H*-pyranoindole fragment.

Biological evaluation of the collection in various phenotypic assays revealed that indofulvins are inhibitors of starvation-induced autophagy, with **indofulvin 1** being the most potent compound ( $IC_{50} = 0.82 \mu M$ ). While **indofulvin 1** inhibited autophagy and was validated at a protein level, neither griseofulvin nor the 4*H*-pyranoindole fragment of **indofulvin 1** affected autophagy (Figure 4b). Evaluation of **indofulvin 1** via the Cell Painting Assay revealed a high fingerprint similarity to oligomycin which targets mitochondrial respiration and led to the hypothesis that **indofulvin 1** may inhibit autophagy through a similar mode of action. In a Mito Stress Test,<sup>67</sup> **indofulvin 1** dose-dependently reduced the oxygen consumption rate and increased the rate of extracellular acidification and suggests that **indofulvin 1** may inhibit autophagy via modulation of mitochondrial function.

Griseofulvin is known to modulate tubulin and cause mitotic cellular arrest. To see if the inherent bioactivity is retained, griseofulvin was compared to **indofulvin 1** in a tubulin polymerization assay and in a cell-based assay measuring mitotic arrest. Griseofulvin showed dose-dependent activity in both assays whereas **indofulvin 1** was inactive and led to the conclusion that **indofulvin 1** does not retain the native bioactivity of griseofulvin.

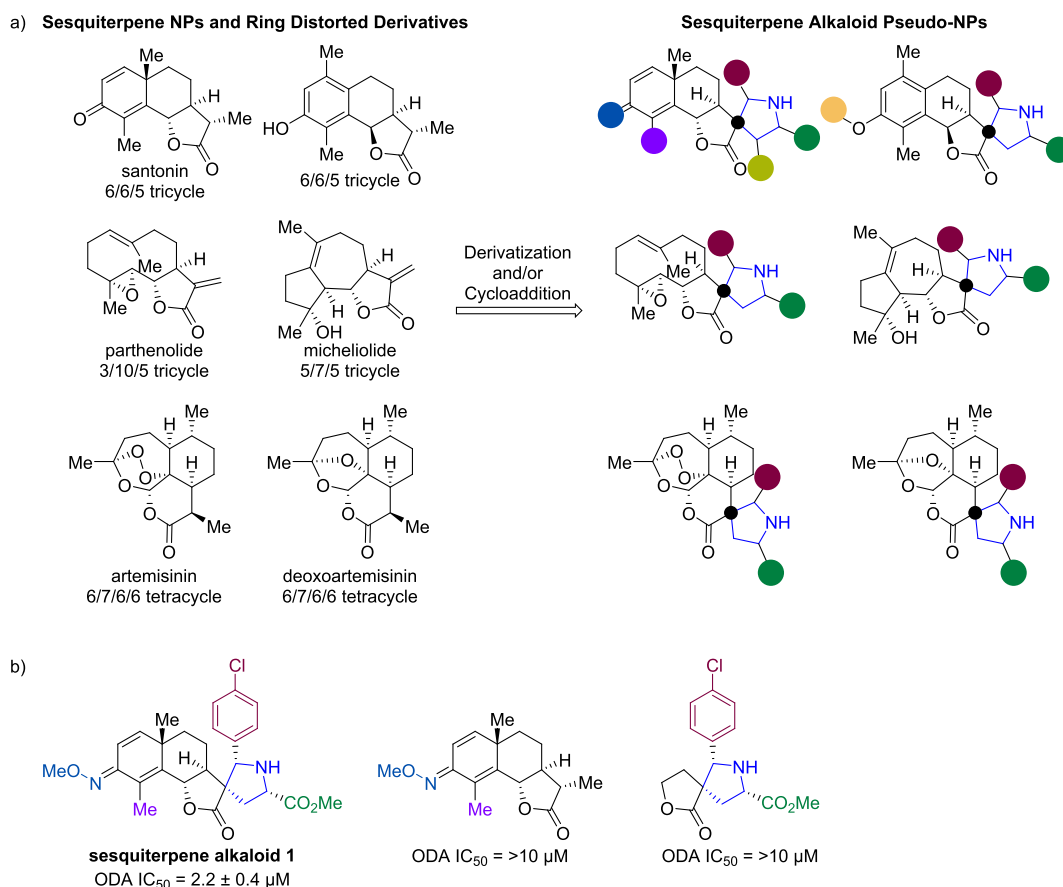
Overall, the indofulvin class represents an unprecedented chemotype for autophagy inhibition. This example indicates that the combination of fragments can provide new bioactivity while simultaneously losing the native bioactivity of an initial fragment and adds validation to the underlying principle of pseudo-NPs.

**Sesquiterpenoid Alkaloids.** A diverse collection of compounds was produced by merging the principles of

complexity to diversity and pseudo-NPs.<sup>68</sup> The strategy employed six fragment-sized sesquiterpenoid lactones that are either NPs or were obtained via ring distortion and can be considered to be biosynthetically related scaffolds but are yet chemically diverse (Figure 5a). The pseudo-NP principle was then applied by combining the sesquiterpenoid fragments with the biosynthetically unrelated alkaloid fragment pyrrolidine via a 1,3-dipolar cycloaddition to arrive at sesquiterpenoid alkaloid pseudo-NPs. The versatility of the cycloaddition allowed for access to various pyrrolidine diastereomers based on the ligand and conditions employed. In total, 89 chemically and stereogenically diverse sesquiterpenoid alkaloid pseudo-NPs were synthesized.

The morphological diversity of the collection was compared via the Cell Painting Assay. The Cell Painting Assay fingerprints of the sesquiterpenoid alkaloid pseudo-NPs and sesquiterpenoid lactone derivatives were clearly distinguishable, indicating that the addition of the pyrrolidine fragment had changed the bioactivity. Different pyrrolidine stereoisomers of dehydrosantonin-derived pseudo-NPs could be identified by the Cell Painting Assay, representing biological diversity through different diastereomers; however, when keeping the pyrrolidine fragment constant and comparing different sesquiterpenoid fragments, the results were not as clearly defined. While some compounds had a low chemical similarity, the similarity of their fingerprints in the Cell Painting Assay was relatively high. Nevertheless, the common pyrrolidine fragment was found not to be solely responsible for the high fingerprint similarities, as this fragment by itself did not share the similar fingerprint to the pseudo-NPs.

The library was also subjected to various phenotypic assays, and **sesquiterpene alkaloid 1** was identified as a new chemotype for the inhibition of Hedgehog-dependent osteoblast differentiation (Figure 5b). Interestingly, three



**Figure 5.** (a) Combination of sesquiterpene NPs and ring distorted derivatives with a pyrrolidine fragment to arrive at a chemically and stereochemically diverse collection of sesquiterpene alkaloid pseudo-NPs. (b) Identification of **sesquiterpene alkaloid 1** as an inhibitor in a Hedgehog-dependent osteoblast differentiation assay (ODA).<sup>69</sup> The individual Santonin-derived and pyrrolidine fragments were inactive in an ODA at  $10 \mu\text{M}$ . The black dots indicate common atoms due to the fusion patterns. The colored circles in **Figure 5a** represent points of derivatization and correlate to the substitutions in **Figure 5b**.

other diastereomers as well as the two partial structures of the active compound were inactive at  $10 \mu\text{M}$  indicating that it is the combination of fragments that gives rise to the pseudo-NP's bioactivity. Together, these results suggest that merging the complexity to diversity and pseudo-NPs concepts may be complementary and lead to chemically and biologically diverse compound classes that are not accessible through existing biosynthetic pathways. We have also successfully synthesized and biologically characterized several other pseudo-NPs<sup>33</sup> which are depicted below in **Figure 11**.

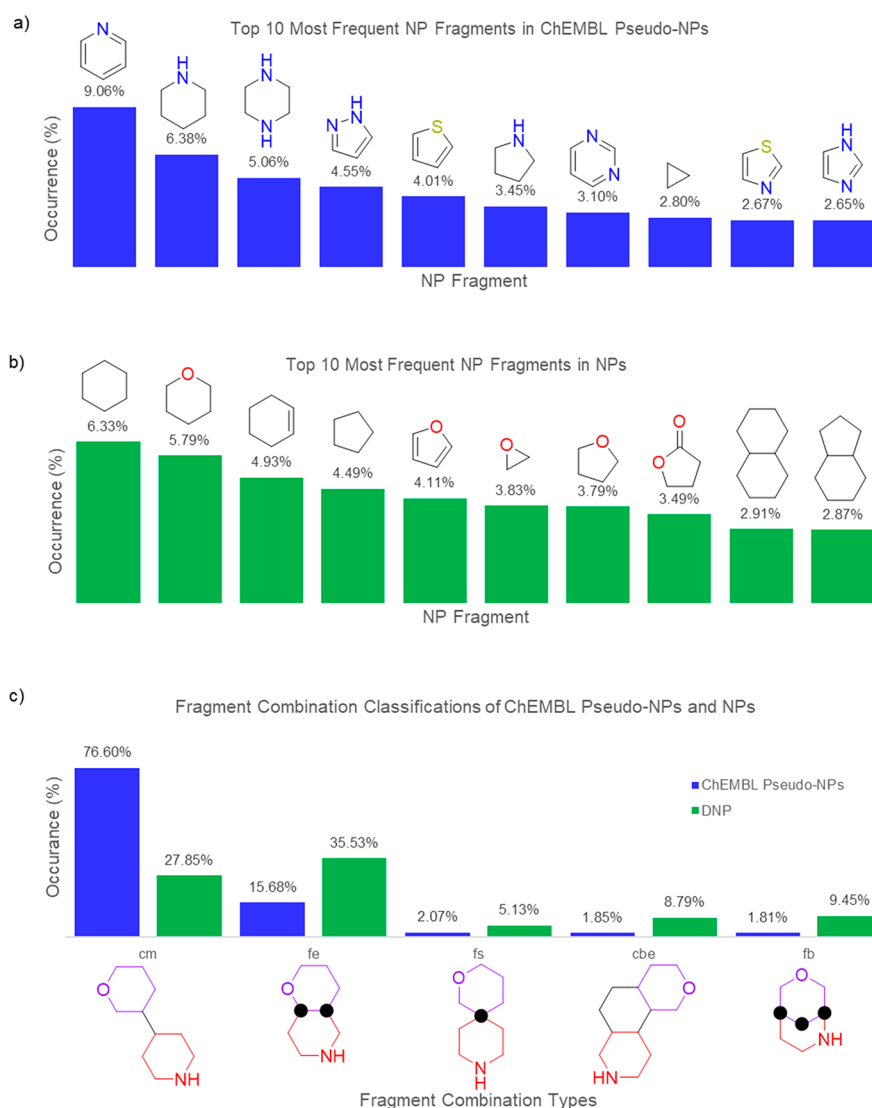
## ■ MANY SYNTHETIC BIOACTIVE COMPOUNDS ARE PSEUDO-NATURAL PRODUCTS

A pseudo-NP can be generally defined as a synthetic or biosynthetically engineered small molecule that contains two or more NP fragments in arrangements that are not found in Nature. Although the design principles for the pseudo-NP concept have only recently been outlined, many bioactive compound classes that qualify as pseudo-NPs may have been unintentionally synthesized over the years. In order to shed light on the frequency of pseudo-NPs, NP fragment combinations and connectivity patterns within the Dictionary of NPs (DNP), representing NPs, and synthetic, non-natural compounds in the ChEMBL database, representing synthetic bioactive compounds, were cheminformatically analyzed and compared.<sup>70</sup> NP fragments were defined from a previous

study<sup>35</sup> and further refined to 1,674 Murcko scaffolds. Due to its abundance in both NPs and synthetic compounds, the benzene fragment was excluded from these studies. NP fragment connectivities were also evaluated and resulted in the consideration of 18 plausible connectivity patterns; however, only the five most frequent connectivity patterns will be discussed here.

Analysis of curated DNP and ChEMBL databases revealed that both have a similarly high percentage of compounds that contain two or more NP fragments (DNP = 67%, ChEMBL = 61%). Furthermore, 344,394 (21%) compounds in the curated ChEMBL data set fit the definition of pseudo-NPs. These pseudo-NPs are predominantly comprised of 2–4 NP fragments that are combined by 5 different connectivity patterns. Overall, the ChEMBL pseudo-NPs are highly diverse and have 117,000 unique scaffolds with <100 members per scaffold.

Since there is a high proportion of pseudo-NPs in the ChEMBL database, a literature search was conducted to see their historical trends. It was found that pseudo-NPs have been made continuously for over 40 years, and that the percentage of pseudo-NPs in published structures per year has been increasing over time. These studies indicate that a significant portion of bioactive synthetic compounds are pseudo-NPs and have been synthesized without the pseudo-NP concept as a guiding principle, further adding validation to the overall concept.



**Figure 6.** Top ten most frequent NP fragments in (a) ChEMBL pseudo-NPs and (b) NPs.<sup>70</sup> (c) Occurrence of fragment combinations of ChEMBL pseudo-NPs (blue) and NPs (green).<sup>70</sup> Examples of connectivity types are depicted below the graph with the NP fragments in purple and red. The black dots indicate common atoms due to the fusion patterns. cm = connection monopodal; fe = fusion edge; fs = fusion spiro; cbe = connection bipodal edge; fb = fusion bridged.

### Comparison of Natural Product Fragments, Connectivity Patterns, and Molecular Properties in Synthetic Pseudo-Natural Products and Natural Products.

While pseudo-NPs frequently populate the ChEMBL database, the ChEMBL pseudo-NPs differ significantly from NPs for which NP fragments are combined and how they are combined. The most abundant (top 10) NP fragments observed in the ChEMBL pseudo-NPs are nitrogen- and sulfur-containing heterocycles, with more than half of them being aromatic (Figure 6a).<sup>70</sup> On the other hand, the DNP set is enriched in alicyclic and oxygen-containing ring systems with only one aromatic fragment in the top 10 (Figure 6b). Trends in NP fragment connection types are also different between the two sets. ChEMBL pseudo-NPs are dominated by the monopodal connectivity (cm, 77%) whereas NPs' connectivity types are more broadly distributed and are dominated by multibond connectivities (Figure 6c).

The molecular features also varied significantly between data sets (Table 1 and SI Figure 1). NPs are more enriched in oxygen content while ChEMBL pseudo-NPs have a higher

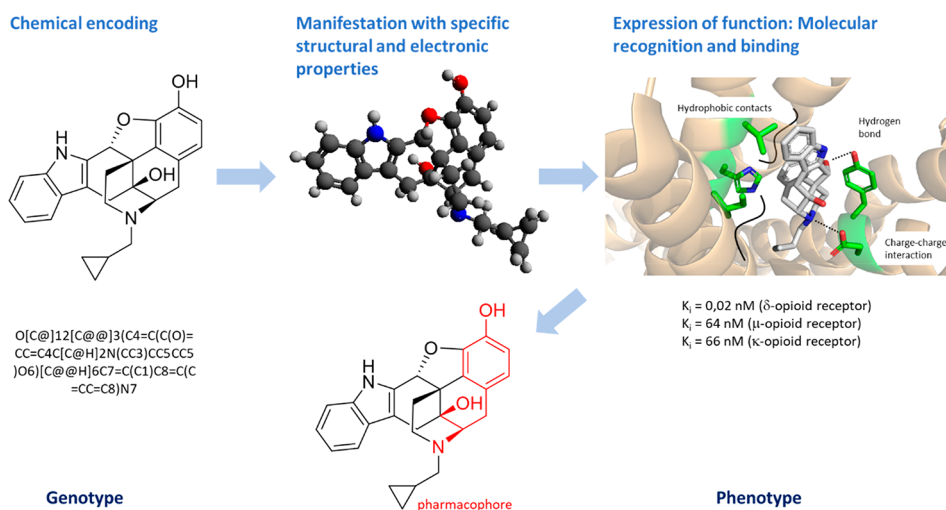
**Table 1. Molecular Features and NP-Likeness Scores of ChEMBL Pseudo-NPs and DNP Reference Sets<sup>a</sup>**

	ChEMBL Pseudo-NPs	DNP
Molecular Weight	412.77 ± 103.09	419.98 ± 146.39
Number of O Atoms	2.48 ± 1.77	5.75 ± 3.78
Number of N Atoms	4.11 ± 1.83	0.58 ± 1.27
fsp <sup>3</sup>	0.33 ± 0.18	0.63 ± 0.24
NP-Likeness Score	-1.10 ± 0.77	2.10 ± 0.86

<sup>a</sup>Values are averages of the data sets. An NP-likeness score<sup>71</sup> was used to compare atom connectivities. A higher NP score is more NP-like while a lower score is less NP-like. fsp<sup>3</sup> = fraction of sp<sup>3</sup> hybridized atoms. DNP = Dictionary of Natural Products. Plots for molecular feature distributions in both curated data sets can be found in the Supporting Information (SI Figure 1).

abundance of nitrogen atoms. Furthermore, the fraction of sp<sup>3</sup> hybridized atoms in NPs is significantly higher than ChEMBL pseudo-NPs. NP-likeness scores<sup>71</sup> of the two data sets were calculated in which a higher value indicates the set is more NP-like. Even though the ChEMBL pseudo-NPs are comprised of





**Figure 7.** Left, representation of naltrindole, an opioid receptor antagonist using standard encoding by molecular graphics or a SMILES code. Both codes represent the chemical information on the molecule and at the same time its genetic information, or its genotype, which may be inherited and mutated during chemical synthesis. Center, the three-dimensional representation of naltrindole puts functional groups in perspective; these may interact with complementary structures, e.g. in a receptor protein, and can (*a posteriori*) be associated with pharmacophore parts of the molecule. Right, naltrindole can recognize and bind to diverse opioid receptors via electrostatic interactions, hydrogen bonds, and hydrophobic interactions thereby exerting its function (here, inhibition). Protein structure: 4ej4.<sup>76</sup>

NP fragments, the NP-likeness score was significantly lower than that for the DNP data set. The differences in molecular features and NP-likeness scores may reflect the frequency with which fragments are used and how they are combined. The realization of these trends may aid in the rational design of future bioactivity-enriched pseudo-NP collections and is expanded upon in the Discussion section.

## ■ PSEUDO-NATURAL PRODUCTS AS A STRATEGY FOR THE CHEMICAL EVOLUTION OF NATURAL PRODUCT STRUCTURE

Evolutionary lines of thought have become part of the chemical mindset at the latest since the introduction of combinatorial synthetic principles for the discovery of new biologically active compounds and therapeutics. The power of *natural evolution*, however, reaches far beyond combinatorial reaction design and the extensive collection of molecules in libraries. Central to natural evolution is the fact that *information* of properties that determine an organism's fitness in a given environment is passed on to descendants and, by mutation and selection, continually extended and improved. We discuss here that this important idea can be applied to the design and improvement of pseudo-NP structures. To this end, a precise definition of the term "information" in a chemical context is required.

Natural evolution relies on *genetic information* which, on the molecular level, is contained in DNA or RNA and encoded by molecular symbols, i.e. nucleobases A, G, C, T (U). The totality of genetic information or hereditary makeup of an organism is called *genotype*. It is translated into a totality of characteristics, called *phenotype*, that results after expression of genetic information in the form of proteins. These enable a wide repertoire of functions because of their ability to fold into three-dimensional structures and having a greater spectrum of specific chemical properties than nucleic acids.

Molecules in general contain *chemical information*, and molecular structures are depicted (or *encoded*) by two-dimensional representations and sometimes by additional

codes such as SMILES. Additionally, a molecule can be described by a number of variables which include the microenvironment of every single atom such as the number of its valence electrons, bonds, non-H atoms, and specific symmetry properties. The chemical information on this molecule then results from the sum of these variables (for a detailed description, see Böttcher<sup>72</sup> and Demoret et al.<sup>73</sup>). The term "information", however, is of value only with the assignment of an information content or a *meaning*. In Nature, the meaning of a gene becomes evident after translation into a protein with the ability to perform a defined task, for example, as an enzyme. A protein will affect the quality of life, or *fitness*, of an organism which is continually evaluated by natural evolution. For small molecules, two-dimensional projections correspond to structures with spatial and electronic properties; these express a meaning based on the fact that they can interact (as a ligand, agonist, or antagonist) and react with other molecules. These *functions* may be observed e.g. with pharmacological or biochemical properties such as inhibitory activity ( $K_i$ ) or substrate quality ( $K_M$ ) and, similar to the function of a protein, be targeted by evolution—in Nature or experiment.<sup>74</sup> Evolution of small molecule structure (more specifically, natural product structure) and function during chemical synthesis, however, requires that the chemical information describing its properties and its potential to interact can be *inherited* to derivatives and undergo *mutation*, i.e. by modification of its molecular symbols (atoms, fragments, etc.). In this sense, chemical information is genetic information. In congruence with definitions for "genotype" and "phenotype" that were originally derived for a molecular description of evolution by M. Eigen,<sup>75</sup> the structure of a molecule can be viewed as its *genotype*, and the function it performs as its *phenotype* (for an example, see Figure 7).

From the medicinal chemist's point of view, the information contained in NP structures can also be described as *biologically relevant* because NPs exert effects on the reproduction rate of an organism that is evaluated during evolution. As explained above, genetic information can be inherited, mutated, and

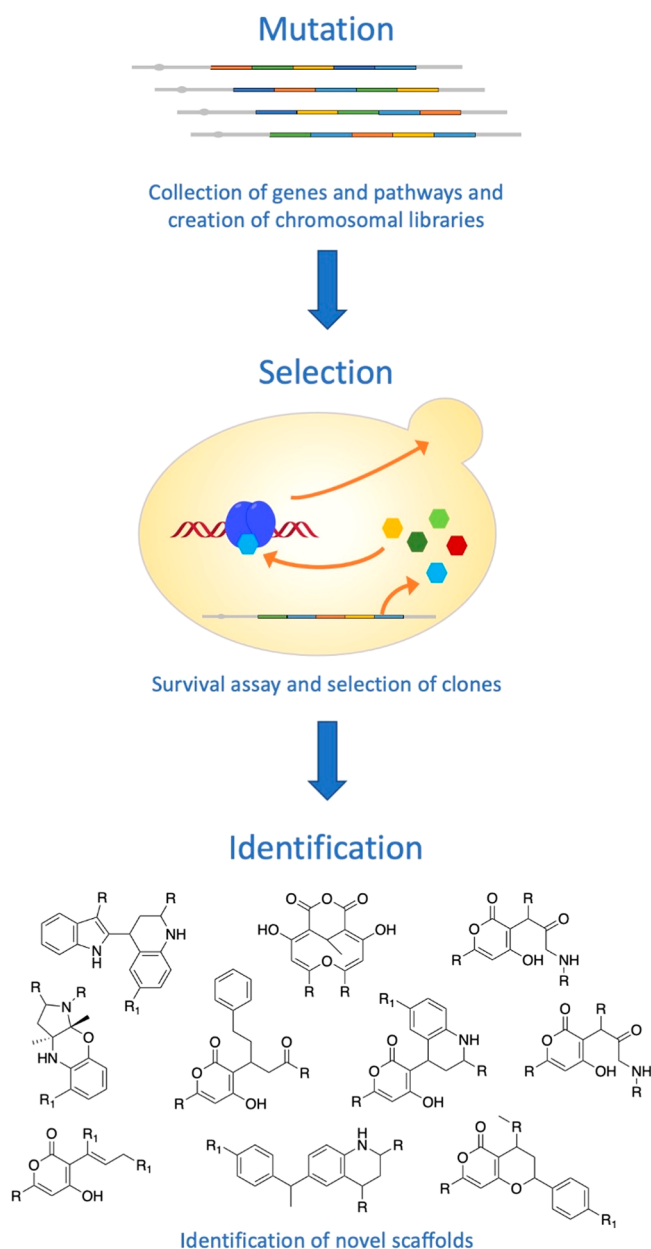
selected. In Nature, the term “mutation” refers to a range of mechanisms from substitution of a single monomeric symbol to recombination, duplication, or deletion of genes. Experimentally, these processes can be realized using individual or combinatorial synthetic and biosynthetic strategies.

An example for accessing previously unknown NPs that combine known fragments in ways not observed before was recently demonstrated in a landmark study that employed genetic material encoding enzymes and enzymatic cascades from a collection of natural sources (Figure 8).<sup>77</sup> Genes were recombined using DNA cloning techniques and expressed in a microbial host which then was cultivated in a “survival assay” (i.e., requiring the action of certain target proteins). As a result, 74 novel structures were identified of which more than 75% had not been described previously. Analysis revealed that some combinations of substructures emerged for the first time while other molecules showed known substructure combinations but with *new connectivities*. It can be concluded that the recombination of biosynthetic pathways, as shown in this study, leads to new linkages of the symbols (here, NP fragments) and thereby to the encoding of new information by new structures. A similar statement has been made by Shenvi for new encoding by chemical synthesis.<sup>73</sup>

Nature’s potential to mutate, recombine, and select that leads to hybrid NP structures as described above can be combined with insights into the informational background and biosynthetic mimicry of evolutionary processes. To meet this goal, NP fragments can be considered as building blocks with inheritable genetic information which can not only be combined according to “Nature’s state-of-the-art” but also by use of a repertoire of synthetic strategies. Thereby, new and unprecedented connectivities may be explored, which may not be possible solely based on natural (biosynthetic) reactions. This requires reaction mechanisms that reflect all possible electronic and geometric degrees of freedom of the respective fragments.

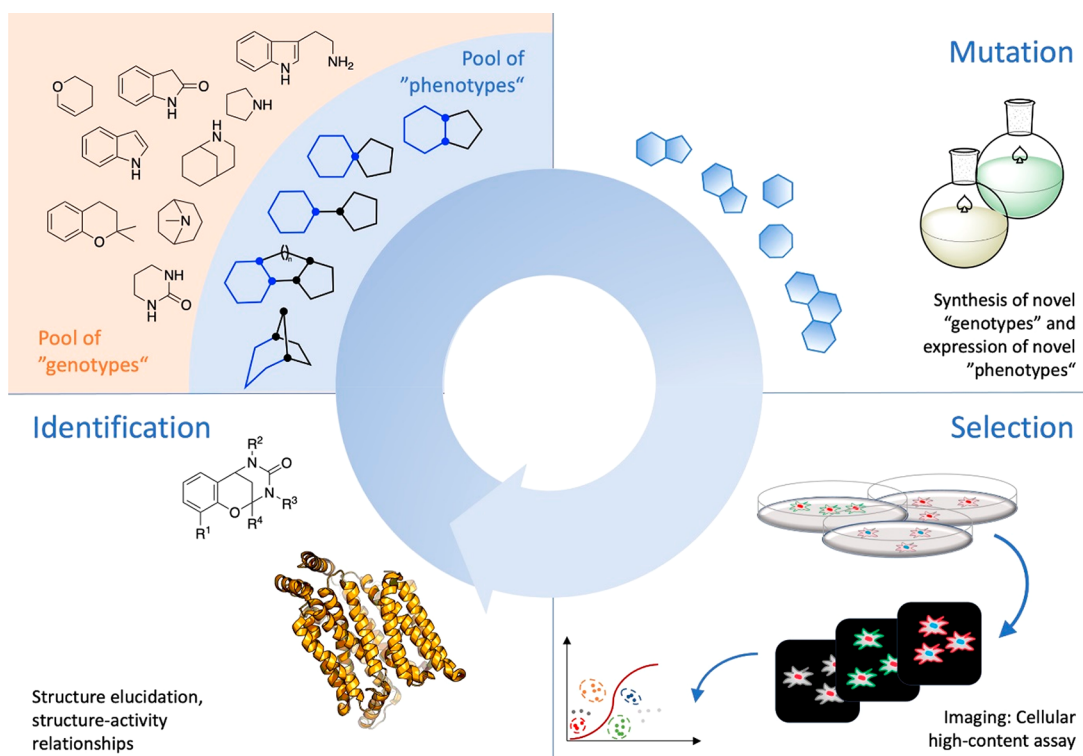
The pseudo-NP strategy requires that NP fragments (= inheritable genotypes) are varied by mutation (substitution, derivatization) and recombination (fragment assembly). In a combinatorial synthetic setting, the pool of fragments will be combined and result in a pool of PNPs with new three-dimensional structures that open up new possibilities to recognize, bind, and react (= new phenotypes). These can be tested for novel or desired bioactivities using appropriate target molecules or cellular systems then identified and categorized with regard to structure–activity-relationships, mode-of-action, and putative targets. The outcome of such a cycle of variation (mutation) and identification (selection) is, besides the new molecule(s), new chemical and biological information that can be used as input for another cycle (Figure 9). In conclusion, the complete iterative process of (1) designing, (2) preparing, and (3) biologically characterizing mimics the Darwinian view of evolution from “mutation–selection–amplification” and can be regarded as a chemical evolution of NP structure.

The immediate usefulness of this concept is demonstrated by the examples described above and those previously reported;<sup>33</sup> however, it may be possible to improve the iterative optimization strategy by greater involvement of evolutionary aspects by (1) including the target molecule of interest for the *direct selection* of active ligands or inhibitors and (2) allowing for reversible and competitive product formation in compound libraries. Dynamic combinatorial chemistry (DCC) provides for libraries capable of responding to target-

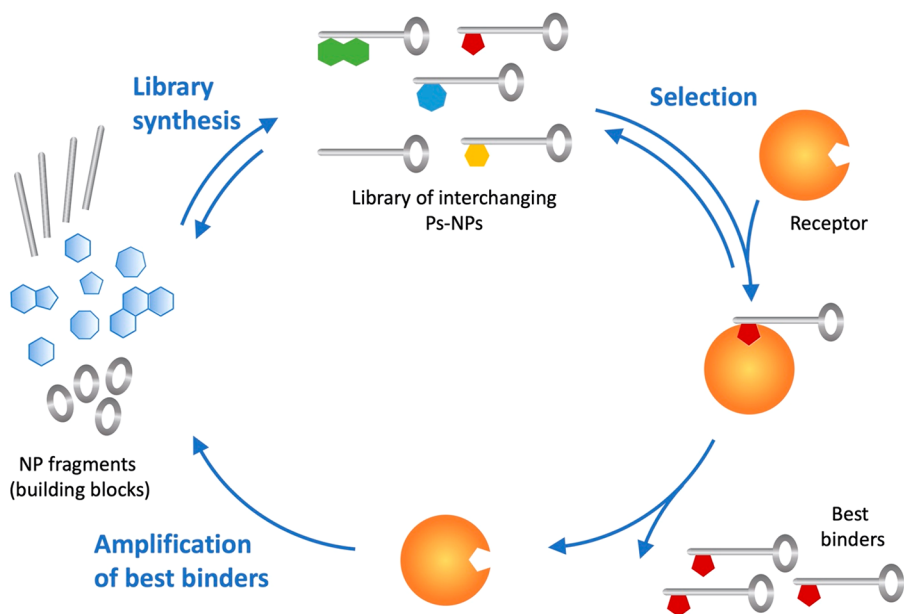


**Figure 8.** Scheme of the combinatorial approach developed by Evola.<sup>77</sup> Genetic information encoding biosynthetic pathways or single enzymes was collected from diverse natural sources and recombined using molecular cloning techniques. Expression of these genes and gene combinations in a microbial host allowed for the alteration or supplementation of existing pathways, thereby enabling the synthesis of new or modified natural products. Their presence was challenged in a cellular assay in which surviving clones showed altered or even “fitter” behavior. These clones were isolated, sorted according to selective criteria, and submitted to detailed analyses for identifying small molecules.

binding events by changing their composition, thus offering a chance to improve signals for detection. In the context of pseudo-NP synthesis, DCC will start from pools of soluble, reasonably stable NP fragments (building blocks) that react in a single compartment to form an adaptive library—provided that appropriate functional groups and reversible reaction types are available that give access to continually interchanging adducts. In the presence of a target, such a dynamic process allows for the *amplification* of active molecules within the



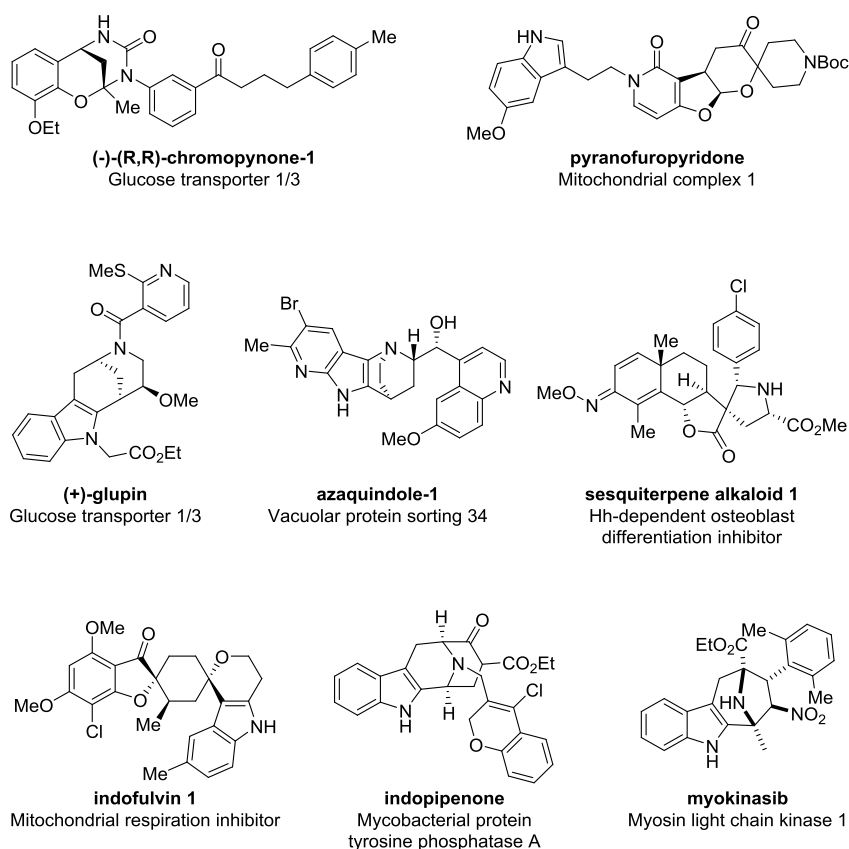
**Figure 9.** Focused chemical evolution. A “pool of genotypes” is synthesized starting from a set of NP fragments using synthetic strategies that allow for the formation of specific connectivities and thus, the expression of “phenotypes” which may be further derivatized. Due to structural properties such as the number of  $sp^3$ -hybridized atoms, stereogenic atoms, heteroatoms, and aromaticity, the phenotypes exhibit varying potentials to recognize and interact with proteins, e.g. enzymes. This will become evident upon application to cellular screening platforms which can be monitored for structural changes, e.g. by fluorescence imaging. A variety of data can be observed, combined, and sorted according to desired criteria. Small molecules, which cause a relevant change, are isolated, characterized, and submitted to studies of structure–activity relationships, thereby revealing new chemical information. The outcome is a starting point for successive rounds of mutagenizing synthesis, selection, and identification. For definitions of genotype, phenotype, evolution, and information in different contexts, please see the [Supporting Information](#).



**Figure 10.** In a dynamic library, potential ligands to a target of interest are synthesized using a set of building blocks (here, NP fragments) and reversible reaction types allowing for continuous interchange. Upon addition of a molecular receptor (usually a protein), the best binders are selected and thereby removed from the equilibrated pool, forcing the dynamic system to respond by resynthesis of this type of small molecule until the equilibrium is reached again.<sup>78</sup>

library, in the sense of a “self-screening process” that was already suggested by J.-M. Lehn (Figure 10).<sup>78</sup>

Although DCC has already become important for biomedical research, it is limited by low library diversity and



**Figure 11.** Structures of pseudo-natural products for which biological targets and/or biological modes of action have been determined.

missing tools for the analysis of large libraries.<sup>79</sup> A strategy which promises to circumvent these limitations is DNA-encoded library (DEL) synthesis which enables the generation and selection of large-scale dynamic libraries on the basis of DNA-mediated assembly, and facilitates identification of individual compounds by deciphering of a DNA tag which encodes building blocks and reaction steps. Substantial progress in this field<sup>80,81</sup> may also provide for a basis to dynamic combinatorial libraries of future pseudo-NPs.

## DISCUSSION

### Validation of the Pseudo-Natural Product Concept.

The pseudo-NP concept is intended to overcome limitations that are inherent to other NP-inspired molecular design principles and explore new areas of biologically relevant chemical space. The concept merges the biological relevance of NPs with the facile exploration of chemical space of fragment-based discovery by combining NP fragments to obtain scaffolds that retain relevance to NPs but are not possible through current biosynthetic pathways.

The pseudo-NP scaffolds can be made up of the combination of NP fragments that are synthetically made, fragment-sized NPs, or derived from NPs by ring distortion or degradation. Scaffold diversity can be achieved through combining different NP fragments, fusing fragments together with different connectivity patterns, and/or having fragments in different arrangements. We recently investigated whether different pseudo-NP fragment combination principles (as described in Figure 2b) can afford chemically and biologically diverse compound collections. The studies mixed and matched a small set of NP fragments in complementary arrangements<sup>39</sup>

as well as combined the same two NP fragments but with different connectivity patterns and/or fragment orientations.<sup>40</sup> Evaluation of the pseudo-NPs by cheminformatics and the Cell Painting Assay revealed that the compound sets were indeed chemically and biologically diverse and supports the hypothesis that different fragment combination strategies can lead to diverse compound collections.

Beyond morphological profiling, we have observed that pseudo-NP collections are enriched with bioactivities that affect therapeutically relevant processes and pathways (Figure 11). Interestingly, many of these examples have concluded that the bioactivity of the pseudo-NP is not due to either individual NP fragment by itself, but rather the combination and the orientation of the NP fragments. These unprecedented NP fragment combinations represent new chemotypes for the identified bioactivities in Figure 11, indicating that design guided by the pseudo-NP concept has led to the exploration of new chemical space while retaining biological relevance. In addition to our own studies, several other research groups have reported the synthesis of bioactive pseudo-NPs.<sup>82–85</sup>

While recent examples reported by us and others delineate the biological potential of pseudo-NP compound design, it should be noted that 21% of all synthetic bioactive molecules (as summarized in ChEMBL) are actually pseudo-NPs.<sup>70</sup> This shows that vast numbers of pseudo-NPs have been unintentionally synthesized without a guiding principle for decades. The frequent occurrence of bioactive pseudo-NPs supports the concept of pseudo-NPs as a design principle for the generation of novel bioactive small molecules.

From a conceptual perspective, the pseudo-NP workflow represents a chemical evolution of NP structures. Genotypes

encoded in biologically relevant NP fragments can undergo “mutations” by chemical alteration and recombination in unprecedented manners that may afford new phenotypes in the form of pseudo-NP scaffolds that can interact with biological systems. “Selection” of pseudo-NPs via biological assays may then lead to further iterations of the cycle to provide more optimal or “evolved” chemical and biological information.

As highlighted here and elsewhere,<sup>33,34</sup> the pseudo-NP concept has combined fragment-based discovery with the biological relevance of NPs to guide the design of biologically unprecedented compounds. The logic of the concept represents a human-made equivalent to natural evolution and can be viewed as the chemical evolution of NP structure. Taken all together, the pseudo-NP approach represents a validated design principle for the wider exploration of biologically relevant chemical space.

**Design and Prediction of Future Pseudo-Natural Product Classes.** Understanding the differences between how chemists and Nature select NP fragments for combination and how they are connected may help guide the design of future compound collections (Figure 6 and Table 1). Although ChEMBL pseudo-NPs contain NP fragments, their molecular features more closely aligned with drug-like compounds than NPs (i.e., low  $\text{fsp}^3$  and nitrogen-rich).<sup>86–89</sup> Accordingly, it is recommended that future pseudo-NP collections should not focus on combining solely nitrogen-rich and/or aromatic NP fragments, but deviate to incorporate a balance of saturated oxygen-containing and/or aliphatic NP fragments with a maximum total of four NP fragments. Furthermore, NP fragment connectivity patterns should not be so heavily enriched by monopodal connections (77% frequency in ChEMBL pseudo-NPs, Figure 6c). Employing a wider array of fusion patterns that are more frequent in NPs may inherently mimic the broader shape distribution observed in NPs relative to the generally rod- and/or disk-like shapes of synthetic compounds and may also lead to more diverse bioactivity profiles between collections.<sup>90–92</sup> Access to these desired features will require the use of existing and the development of new complexity-generating reactions in stereoselective and asymmetric fashions rather than the simple linear linkage of fragments. The resulting scaffolds will be enriched in stereochemical content and have high densities of chemical information.<sup>72</sup>

By selecting more  $\text{sp}^3$ -enriched and oxygen-containing NP fragments as combination partners and employing a broader distribution of fusion patterns, future pseudo-NP collections may more closely mimic properties Nature has instilled in NPs; i.e., they will be more NP-like.<sup>71</sup> However, novel arrangements of NP fragments will deviate from the current NP structure to afford compound classes that will explore new areas of chemical space while retaining the biological relevance of NPs.

Beyond having Nature guide pseudo-NP design, the choice of fragment combination may also be influenced by evaluating the biological diversity of current pseudo-NPs. If the arrangement of a common set of fragments results in diverse biological profiles, then the fragments could be classified as nondominating and may be good candidates for future combinations. Conversely, if pseudo-NPs with a common fragment share similar biological profiles, the fragment could be classified as dominating and may lead to redundant bioactivity profiles in future combinations.

We recently analyzed a set of pseudo-NPs comprised of common fragments via the Cell Painting Assay and were able

to classify fragments as nondominating or dominating.<sup>39</sup> These assignments were used to design new pseudo-NP compound classes that were correctly predicted to have unique and similar morphological profiles relative to the initial set of pseudo-NPs. Similar analyses may help guide the design of future NP fragment combinations to obtain biologically diverse compound collections that more efficiently explore biological space.

The initial design of the pseudo-NP concept is to employ NPs as starting points since they are biologically prevalidated compounds selected through natural evolution, as are their fragments.<sup>35</sup> However, it should be noted that biologically relevant chemical space is not occupied only by various combinations of NP fragments (i.e., NPs, NP derivatives, or pseudo-NPs). This is clear from the large number of bioactive molecules that contain or are exclusively composed of purely synthetic fragments.<sup>93,94</sup> These man-made fragments in bioactive compounds should therefore also be considered biologically relevant. The evolutionary algorithm of the pseudo-NP concept (Figure 9) may then be used in a broader sense to include not only NP fragments but fragments of all biologically relevant compounds. This may expand the exploration of the concept from evolutionarily relevant NP-like chemical space to more general biologically relevant chemical space.

## OUTLOOK

The pseudo-NP design criteria provide numerous opportunities for molecular discovery programs. Nevertheless, multidisciplinary investigations and advancements are needed to facilitate the exploration of the biologically relevant chemical space that can be accessed by pseudo-NP structure. From a cheminformatic perspective, a further understanding of the differences between how Nature and chemists have selected which NP fragments to combine and how to combine them may aid in the design of future compound collections. At the core of accessing new pseudo-NPs are novel fragment combinations through organic synthesis which will heavily rely on the use and development of complexity-generating methodologies envisioned by organic chemists. Furthermore, the identification of biological relevance of future pseudo-NPs will be underpinned by the continuous development of broad, high content assays and new techniques for target identification by biologists and chemical biologists. Through multidisciplinary collaboration and advancements, the design, synthesis, and biological evaluation of new pseudo-NPs may help navigate biologically relevant chemical space to provide compounds that will be beneficial to chemical biology and medicinal chemistry programs.

## ASSOCIATED CONTENT

### Supporting Information

The Supporting Information is available free of charge at <https://pubs.acs.org/doi/10.1021/jacs.1c11270>.

Computational details of the data provided in Table 1 and definitions of genotype, phenotype, evolution, and information in different contexts (PDF)

## AUTHOR INFORMATION

### Corresponding Author

Herbert Waldmann – Max-Planck-Institute of Molecular Physiology, 44227 Dortmund, Germany; Faculty of

Chemistry and Chemical Biology, TU Dortmund University, 44227 Dortmund, Germany; [orcid.org/0000-0002-9606-7247](https://orcid.org/0000-0002-9606-7247); Email: [herbert.waldmann@mpi-dortmund.mpg.de](mailto:herbert.waldmann@mpi-dortmund.mpg.de)

## Authors

**Michael Grigalunas** – Max-Planck-Institute of Molecular Physiology, 44227 Dortmund, Germany; [orcid.org/0000-0003-0419-8168](https://orcid.org/0000-0003-0419-8168)

**Susanne Brakmann** – Faculty of Chemistry and Chemical Biology, TU Dortmund University, 44227 Dortmund, Germany; [orcid.org/0000-0002-9906-3735](https://orcid.org/0000-0002-9906-3735)

Complete contact information is available at:  
<https://pubs.acs.org/10.1021/jacs.1c11270>

## Funding

Open access funded by Max Planck Society.

## Notes

The authors declare no competing financial interest.

## ACKNOWLEDGMENTS

We are grateful to the Max-Planck-Gesellschaft and TU Dortmund for making this research possible through institutional financial support. We are also thankful to José-Manuel Gally for contributions and helpful discussions regarding cheminformatic analyses of pseudo-natural products.

## REFERENCES

- (1) Kumar, K.; Waldmann, H. Nature Inspired Small Molecules for Chemical Biology. *Isr. J. Chem.* **2019**, *59*, 41.
- (2) Brown, D. G.; Wobst, H. J. A Decade of FDA-Approved Drugs (2010–2019): Trends and Future Directions. *J. Med. Chem.* **2021**, *64*, 2312.
- (3) Bohacek, R. S.; McMartin, C.; Guida, W. C. The Art and Practice of Structure-Based Drug Design: A Molecular Modeling Perspective. *Med. Res. Rev.* **1996**, *16*, 3.
- (4) Dobson, C. M. Chemical Space and Biology. *Nature* **2004**, *432*, 824.
- (5) Newman, D. J.; Cragg, G. M. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. *J. Nat. Prod.* **2020**, *83*, 770.
- (6) Pye, C. R.; Bertin, M. J.; Lokey, R. S.; Gerwick, W. H.; Linington, R. G. Retrospective Analysis of Natural Products Provides Insights for Future Discovery Trends. *Proc. Natl. Acad. Sci. U.S.A.* **2017**, *114*, 5601.
- (7) Bon, R. S.; Waldmann, H. Bioactivity-Guided Navigation of Chemical Space. *Acc. Chem. Res.* **2010**, *43*, 1103.
- (8) Grygorenko, O. O.; Volochnyuk, D. M.; Ryabukhin, S. V.; Judd, D. B. The Symbiotic Relationship Between Drug Discovery and Organic Chemistry. *Chem.—Eur. J.* **2020**, *26*, 1196.
- (9) Schreiber, S. L. Molecular Diversity by Design. *Nature* **2009**, *457*, 153.
- (10) Franzini, R. M.; Randolph, C. Chemical Space of DNA-Encoded Libraries. *J. Med. Chem.* **2016**, *59*, 6629.
- (11) Gerry, C. J.; Wawer, M. J.; Clemons, P. A.; Schreiber, S. L. DNA Barcoding a Complete Matrix of Stereoisomeric Small Molecules. *J. Am. Chem. Soc.* **2019**, *141*, 10225.
- (12) Gerry, C. J.; Schreiber, S. L. Recent Achievements and Current Trajectories of Diversity-Oriented Synthesis. *Curr. Opin. Chem. Biol.* **2020**, *56*, 1.
- (13) Reiher, C. A.; Schuman, D. P.; Simmons, N.; Wolkenberg, S. E. Trends in Hit-to-Lead Optimization Following DNA-Encoded Library Screens. *ACS Med. Chem. Lett.* **2021**, *12*, 343.
- (14) Wender, P. A.; Verma, V. A.; Paxton, T. J.; Pillow, T. H. Function-Oriented Synthesis, Step Economy, and Drug Design. *Acc. Chem. Res.* **2008**, *41*, 40.
- (15) Crane, E. A.; Gademann, K. Capturing Biological Activity in Natural Product Fragments by Chemical Synthesis. *Angew. Chem., Int. Ed.* **2016**, *55*, 3882.
- (16) Woo, S.; Shenvi, R. A. Natural Product Synthesis through the Lens of Informatics. *Acc. Chem. Res.* **2021**, *54*, 1157.
- (17) Huffman, B. J.; Shenvi, R. A. Natural Products in the “Marketplace”: Interfacing Synthesis and Biology. *J. Am. Chem. Soc.* **2019**, *141*, 3332.
- (18) Nelson, A.; Karageorgis, G. Natural Product-Informed Exploration of Chemical Space to Enable Bioactive Molecular Discovery. *RSC Med. Chem.* **2021**, *12*, 353.
- (19) Grigalunas, M.; Burhop, A.; Christoforow, A.; Waldmann, H. Pseudo-Natural Products and Natural Product-Inspired Methods in Chemical Biology and Drug Discovery. *Curr. Opin. Chem. Biol.* **2020**, *56*, 111.
- (20) Wetzel, S.; Bon, R. S.; Kumar, K.; Waldmann, H. Biology-Oriented Synthesis. *Angew. Chem., Int. Ed.* **2011**, *50*, 10800.
- (21) Karageorgis, G.; Waldmann, H. Guided by Evolution: Biology-Oriented Synthesis of Bioactive Compound Classes. *Synthesis (Stuttg.)* **2019**, *51*, 55.
- (22) Van Hattum, H.; Waldmann, H. Biology-Oriented Synthesis: Harnessing the Power of Evolution. *J. Am. Chem. Soc.* **2014**, *136*, 11853.
- (23) Koch, M. A.; Schuffenhauer, A.; Scheck, M.; Wetzel, S.; Casaulta, M.; Odermatt, A.; Ertl, P.; Waldmann, H. Charting Biologically Relevant Chemical Space: A Structural Classification of Natural Products (SCONP). *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 17272.
- (24) Dücker, H.; Pries, V.; Khedkar, V.; Menninger, S.; Bruss, H.; Bird, A. W.; Maliga, Z.; Brockmeyer, A.; Janning, P.; Hyman, A.; Grimme, S.; Schürmann, M.; Preut, H.; Hübel, K.; Ziegler, S.; Kumar, K.; Waldmann, H. Natural Product-Inspired Cascade Synthesis Yields Modulators of Centrosome Integrity. *Nat. Chem. Biol.* **2012**, *8*, 179.
- (25) Paciaroni, N. G.; Ratnayake, R.; Matthews, J. H.; Norwood, V. M.; Arnold, A. C.; Dang, L. H.; Luesch, H.; Huigens, R. W. A Tryptoline Ring-Distortion Strategy Leads to Complex and Diverse Biologically Active Molecules from the Indole Alkaloid Yohimbine. *Chem.—Eur. J.* **2017**, *23*, 4327.
- (26) Schneidewind, T.; Kapoor, S.; Garivet, G.; Karageorgis, G.; Narayan, R.; Vendrell-Navarro, G.; Antonchick, A. P.; Ziegler, S.; Waldmann, H. Phenotypic Discovery of Myokinasib, a Kinase Inhibitor with Unprecedented Chemotype Targeting Myosin Light Chain Kinase I. *Cell Chem. Biol.* **2019**, *26*, 512.
- (27) Karageorgis, G.; Reckzeh, E. S.; Ceballos, J.; Schwalfenberg, M.; Sievers, S.; Ostermann, C.; Pahl, A.; Ziegler, S.; Waldmann, H. Chromopyrones Are Pseudo Natural Product Glucose Uptake Inhibitors Targeting Glucose Transporters GLUT-1 and -3. *Nat. Chem.* **2018**, *10*, 1103.
- (28) Ceballos, J.; Schwalfenberg, M.; Karageorgis, G.; Reckzeh, E. S.; Sievers, S.; Ostermann, C.; Pahl, A.; Sellstedt, M.; Nowacki, J.; Carnero Corrales, M. A.; Wilke, J.; Laraia, L.; Tschapalda, K.; Metz, M.; Sehr, D. A.; Brand, S.; Winkhofer, K.; Janning, P.; Ziegler, S.; Waldmann, H. Synthesis of Indomorphan Pseudo-Natural Product Inhibitors of Glucose Transporters GLUT-1 and -3. *Angew. Chem., Int. Ed.* **2019**, *58*, 17016.
- (29) Motika, S. E.; Hergenrother, P. J. Re-Engineering Natural Products to Engage New Biological Targets. *Nat. Prod. Rep.* **2020**, *37*, 1395.
- (30) Morrison, K. C.; Hergenrother, P. J. Natural Products as Starting Points for the Synthesis of Complex and Diverse Compounds. *Nat. Prod. Rep.* **2014**, *31*, 6.
- (31) Huigens, R. W.; Morrison, K. C.; Hicklin, R. W.; Flood, T. A., Jr.; Richter, M. F.; Hergenrother, P. J. A Ring-Distortion Strategy to Construct Stereochemically Complex and Structurally Diverse Compounds from Natural Products. *Nat. Chem.* **2013**, *5*, 195.
- (32) Cremonnik, G. S.; Liu, J.; Waldmann, H. Guided by Evolution: From Biology Oriented Synthesis to Pseudo Natural Products. *Nat. Prod. Rep.* **2020**, *37*, 1497.

- (33) Karageorgis, G.; Foley, D. J.; Laraia, L.; Brakmann, S.; Waldmann, H. Pseudo Natural Products — Chemical Evolution of Natural Product Structure. *Angew. Chem., Int. Ed.* **2021**, *60*, 15705.
- (34) Karageorgis, G.; Foley, D. J.; Laraia, L.; Waldmann, H. Principle and Design of Pseudo-Natural Products. *Nat. Chem.* **2020**, *12*, 227.
- (35) Over, B.; Wetzel, S.; Grütter, C.; Nakai, Y.; Renner, S.; Rauh, D.; Waldmann, H. Natural-Product-Derived Fragments for Fragment-Based Ligand Discovery. *Nat. Chem.* **2013**, *5*, 21.
- (36) Crane, E. A.; Gademann, K. Capturing Biological Activity in Natural Product Fragments by Chemical Synthesis. *Angew. Chem., Int. Ed.* **2016**, *55*, 3882.
- (37) Vu, H.; Pedro, L.; Mak, T.; McCormick, B.; Rowley, J.; Liu, M.; Di Capua, A.; Williams-Noonan, B.; Pham, N. B.; Pouwer, R.; Nguyen, B.; Andrews, K. T.; Skinner-Adams, T.; Kim, J.; Hol, W. G. J.; Hui, R.; Crowther, G. J.; Van Voorhis, W. C.; Quinn, R. J. Fragment-Based Screening of a Natural Product Library against 62 Potential Malaria Drug Targets Employing Native Mass Spectrometry. *ACS Infect. Dis.* **2018**, *4*, 431.
- (38) Pascolutti, M.; Campitelli, M.; Nguyen, B.; Pham, N.; Gorse, A. D.; Quinn, R. J. Capturing Nature's Diversity. *PLoS One* **2015**, *10*, No. e0120942.
- (39) Grigalunas, M.; Burhop, A.; Zinken, S.; Pahl, A.; Gally, J.-M.; Wild, N.; Mantel, Y.; Sievers, S.; Foley, D. J.; Scheel, R.; Strohmman, C.; Antonchick, A. P.; Waldmann, H. Natural Product Fragment Combination to Performance-Diverse Pseudo-Natural Products. *Nat. Commun.* **2021**, *12*, 1883.
- (40) Liu, J.; Cremosnik, G. S.; Otte, F.; Pahl, A.; Sievers, S.; Strohmman, C.; Waldmann, H. Design, Synthesis, and Biological Evaluation of Chemically and Biologically Diverse Pyrroquinoline Pseudo Natural Products. *Angew. Chem., Int. Ed.* **2021**, *60*, 4648.
- (41) Decker, M. Hybrid Molecules Incorporating Natural Products: Applications in Cancer Therapy, Neurodegenerative Disorders and Beyond. *Curr. Med. Chem.* **2011**, *18* (10), 1464.
- (42) Shaveta; Singh, P.; Mishra, S. Hybrid Molecules: The Privileged Scaffolds for Various Pharmaceuticals. *Eur. J. Med. Chem.* **2016**, *124*, 500.
- (43) Tietze, L. F.; Bell, H. P.; Chandrasekhar, S. Natural Product Hybrids as New Leads for Drug Discovery. *Angew. Chem., Int. Ed.* **2003**, *42*, 3996.
- (44) Nepali, K.; Sharma, S.; Sharma, M.; Bedi, P. M. S.; Dhar, K. L. Rational Approaches, Design Strategies, Structure Activity Relationship and Mechanistic Insights for Anticancer Hybrids. *Eur. J. Med. Chem.* **2014**, *77*, 422.
- (45) Saldívar-González, F. I.; Valli, M.; Andricopulo, A. D.; Da Silva Bolzani, V.; Medina-Franco, J. L. Chemical Space and Diversity of the NuBBE Database: A Chemoinformatic Characterization. *J. Chem. Inf. Model.* **2019**, *59*, 74.
- (46) Chen, Y.; Kirchmair, J. Cheminformatics in Natural Product-Based Drug Discovery. *Mol. Inf.* **2020**, *39*, 2000171.
- (47) Congreve, M.; Carr, R.; Murray, C.; Jhoti, H. A "Rule of Three" for Fragment-Based Lead Discovery? *Drug Discovery Today* **2003**, *8*, 876.
- (48) Köster, H.; Craan, T.; Brass, S.; Herhaus, C.; Zentgraf, M.; Neumann, L.; Heine, A.; Klebe, G. A Small Nonrule of 3 Compatible Fragment Library Provides High Hit Rate of Endothiapepsin Crystal Structures with Various Fragment Chemotypes. *J. Med. Chem.* **2011**, *54*, 7784.
- (49) Pahl, A.; Waldmann, H.; Kumar, K. Exploring Natural Product Fragments for Drug and Probe Discovery. *Chimia (Aarau).* **2017**, *71*, 653.
- (50) Rafferty, R. J.; Hicklin, R. W.; Maloof, K. A.; Hergenrother, P. J. Synthesis of Complex and Diverse Compounds through Ring Distortion of Abietic Acid. *Angew. Chem., Int. Ed.* **2014**, *53*, 220.
- (51) Wang, Y.; Cornett, A.; King, F. J.; Mao, Y.; Nigsch, F.; Paris, C. G.; McAllister, G.; Jenkins, J. L. Evidence-Based and Quantitative Prioritization of Tool Compounds in Phenotypic Drug Discovery. *Cell Chem. Bio.* **2016**, *23*, 862.
- (52) Bray, M. A.; Singh, S.; Han, H.; Davis, C. T.; Borgeson, B.; Hartland, C.; Kost-Alimova, M.; Gustafsdottir, S. M.; Gibson, C. C.; Carpenter, A. E. Cell Painting, a High-Content Image-Based Assay for Morphological Profiling Using Multiplexed Fluorescent Dyes. *Nat. Protoc.* **2016**, *11*, 1757.
- (53) Ziegler, S.; Sievers, S.; Waldmann, H. Morphological Profiling of Small Molecules. *Cell Chem. Biol.* **2021**, *28*, 300.
- (54) Dunn, Z. D.; Wever, W. J.; Economou, N. J.; Bowers, A. A.; Li, B. Enzymatic Basis of "Hybridity" in Thiomarinol Biosynthesis. *Angew. Chem., Int. Ed.* **2015**, *54*, 5137.
- (55) Shiozawa, H.; Kagasaki, T.; Takahashi, S.; Kinoshita, T.; Haruyama, H.; Domon, H.; Utsuib, Y.; Kodama, K. Thiomarinol, a New Hybrid Antimicrobial Antibiotic Produced by a Marine Bacterium Fermentation, Isolation, Structure, and Antimicrobial Activity. *J. Antibiot.* **1993**, *46*, 1834.
- (56) Martino, E.; Casamassima, G.; Castiglione, S.; Cellupica, E.; Pantalone, S.; Papagni, F.; Rui, M.; Siciliano, A. M.; Collina, S. Vinca Alkaloids and Analogues as Anti-Cancer Agents: Looking Back, Peering Ahead. *Bioorg. Med. Chem. Lett.* **2018**, *28*, 2816.
- (57) Zhu, J.; Wang, M.; Wen, W.; Yu, R. Biosynthesis and Regulation of Terpenoid Indole Alkaloids in *Catharanthus roseus*. *Pharmacogn. Rev.* **2015**, *9*, 24.
- (58) Prakash, V.; Timasheff, S. N. Mechanism of Interaction of Vinca Alkaloids with Tubulin: Catharanthine and Vindoline. *Biochemistry* **1991**, *30*, 873.
- (59) Chi, S.; Xie, W.; Zhang, J.; Xu, S. Theoretical Insight into the Structural Mechanism for the Binding of Vinblastine with Tubulin. *J. Biomol. Struct. Dyn.* **2015**, *33*, 2234.
- (60) Suzuki, K.; Yasui, Y. *Hybrid Natural Products*; Hanessian, S., Ed.; Wiley-VCH: Weinheim, 2014.
- (61) Jiang, M.; Wu, Z.; Liu, L.; Chen, S. The Chemistry and Biology of Fungal Meroterpenoids (2009–2019). *Org. Biomol. Chem.* **2021**, *19*, 1644.
- (62) Schor, R.; Schotte, C.; Wibberg, D.; Kalinowski, J.; Cox, R. J. Three Previously Unrecognised Classes of Biosynthetic Enzymes Revealed during the Production of Xenovulene A. *Nat. Commun.* **2018**, *9*, 1963.
- (63) Yee, D. A.; Kakule, T. B.; Cheng, W.; Chen, M.; Chong, C. T. Y.; Hai, Y.; Hang, L. F.; Hung, Y. S.; Liu, N.; Ohashi, M.; Okorafor, I. C.; Song, Y.; Tang, M.; Zhang, Z.; Tang, Y. Genome Mining of Alkaloidal Terpenoids from a Hybrid Terpene and Nonribosomal Peptide Biosynthetic Pathway. *J. Am. Chem. Soc.* **2020**, *142*, 710.
- (64) Miyanaga, A.; Kudo, F.; Eguchi, T. Protein-Protein Interactions in Polyketide Synthase-Nonribosomal Peptide Synthetase Hybrid Assembly Lines. *Nat. Prod. Rep.* **2018**, *35*, 1185.
- (65) Tietze, L. Secologanin, a Biogenetic Key Compound—Synthesis and Biogenesis of the Iridoid and Secoiridoid Glycosides. *Angew. Chem., Int. Ed.* **1983**, *22*, 828.
- (66) Burhop, A.; Bag, S.; Grigalunas, M.; Woitalla, S.; Bodenbinder, P.; Brieger, L.; Strohmman, C.; Pahl, A.; Sievers, S.; Waldmann, H. Synthesis of Indofulvin Pseudo-Natural Products Yields a New Autophagy Inhibitor Chemotype. *Adv. Sci.* **2021**, *8*, 2102042.
- (67) Salabei, J. K.; Gibb, A. A.; Hill, B. G. Comprehensive Measurement of Respiratory Activity in Permeabilized Cells Using Extracellular Flux Analysis. *Nat. Protoc.* **2014**, *9*, 421.
- (68) Liu, J.; Flegel, J.; Otte, F.; Pahl, A.; Sievers, S.; Strohmman, C.; Waldmann, H. Combination of Pseudo-Natural Product Design and Formal Natural Product Ring Distortion Yields Stereochemically and Biologically Diverse Pseudo-Sesquiterpenoid Alkaloids. *Angew. Chem., Int. Ed.* **2021**, *60*, 21384.
- (69) Depeursinge, A.; Racoceanu, D.; Iavindrasana, J.; Cohen, G.; Platon, A.; Poletti, P.-A.; Muller, H. Purporphamine Induces Osteogenesis by Activation of the Hedgehog Signaling Pathway. *Chem. Biol.* **2004**, *11*, 1229.
- (70) Gally, J.-M.; Pahl, A.; Czodrowski, P.; Waldmann, H. Pseudo-Natural Products Occur Frequently in Biologically Relevant Compounds. *J. Chem. Inf. Model.* **2021**, *61*, 5458.

- (71) Ertl, P.; Roggo, S.; Schuffenhauer, A. Natural Product-Likeness Score and Its Application for Prioritization of Compound Libraries. *J. Chem. Inf. Model.* **2008**, *48*, 68.
- (72) Böttcher, T. An Additive Definition of Molecular Complexity. *J. Chem. Inf. Model.* **2016**, *56*, 462.
- (73) Demoret, R. M.; Baker, M. A.; Ohtawa, M.; Chen, S.; Lam, C. C.; Khom, S.; Roberto, M.; Forli, S.; Houk, K. N.; Shenvi, R. A. Synthetic, Mechanistic, and Biological Interrogation of Ginkgo Biloba Chemical Space En Route to (–)-Bilobalide. *J. Am. Chem. Soc.* **2020**, *142*, 18599.
- (74) Ji, Q.; Lirag, R. C.; Miljanić, O. Š. Kinetically Controlled Phenomena in Dynamic Combinatorial Libraries. *Chem. Soc. Rev.* **2014**, *43*, 1873.
- (75) Eigen, M.; Schuster, P. The Hypercycle. A Principle of Natural Self-Organization. Part A: Emergence of the Hypercycle. *Naturwiss.* **1977**, *64*, 541.
- (76) Granier, S.; Manglik, A.; Kruse, A. C.; Kobilka, T. S.; Thian, F. S.; Weis, W. I.; Kobilka, B. K. Structure of the  $\delta$ -Opioid Receptor Bound to Naltrindole. *Nature* **2012**, *485*, 400.
- (77) Klein, J.; Heal, J. R.; Hamilton, W. D. O.; Boussemghoune, T.; Tange, T. Ø.; Delegrange, F.; Jaeschke, G.; Hatsch, A.; Heim, J. Yeast Synthetic Biology Platform Generates Novel Chemical Structures as Scaffolds for Drug Discovery. *ACS Synth. Biol.* **2014**, *3*, 314.
- (78) Ramström, O.; Lehn, J. M. Drug Discovery by Dynamic Combinatorial Libraries. *Nat. Rev. Drug Discovery* **2002**, *1*, 26.
- (79) Zhou, Y.; Li, C.; Peng, J.; Xie, L.; Meng, L.; Li, Q.; Zhang, J.; Li, X. D.; Li, X.; Huang, X.; Li, X. DNA-Encoded Dynamic Chemical Library and Its Applications in Ligand Discovery. *J. Am. Chem. Soc.* **2018**, *140*, 15859.
- (80) Zhou, Y.; Shen, W.; Peng, J.; Deng, Y.; Li, X. Identification of Isoform/Domain-Selective Fragments from the Selection of DNA-Encoded Dynamic Library. *Bioorg. Med. Chem.* **2021**, *45*, 116328.
- (81) Huang, Y.; Li, X. Recent Advances on the Selection Methods of DNA-Encoded Libraries. *ChemBioChem.* **2021**, *22*, 2384.
- (82) Wu, G.; Yu, G.; Yu, Y.; Yang, S.; Duan, Z.; Wang, W.; Liu, Y.; Yu, R.; Li, J.; Zhu, T.; Gu, Q.; Li, D. Chemoreactive-Inspired Discovery of Influenza A Virus Dual Inhibitor to Block Hemagglutinin-Mediated Adsorption and Membrane Fusion. *J. Med. Chem.* **2020**, *63*, 6924.
- (83) Whitmarsh-Everiss, T.; Olsen, A. H.; Laraia, L. Identification of Inhibitors of Cholesterol Transport Proteins Through the Synthesis of a Diverse, Sterol-Inspired Compound Collection. *Angew. Chem., Int. Ed.* **2021**, *60*, 26755.
- (84) Cunha, M. R.; Bhardwaj, R.; Carrel, A. L.; Lindinger, S.; Romanin, C.; Parise-Filho, R.; Hediger, M. A.; Reymond, J. L. Natural Product Inspired Optimization of a Selective TRPV6 Calcium Channel Inhibitor. *RSC Med. Chem.* **2020**, *11*, 1032.
- (85) Zhu, H.; Cai, Y.; Ma, S.; Futamura, Y.; Li, J.; Zhong, W.; Zhang, X.; Osada, H.; Zou, H. Privileged Biorenewable Secologanin-Based Diversity-Oriented Synthesis for Pseudo-Natural Alkaloids: Uncovering Novel Neuroprotective and Antimalarial Frameworks. *ChemSusChem* **2021**, *14*, 5320.
- (86) Atanasov, A. G.; Zotchev, S. B.; Dirsch, V. M.; Orhan, I. E.; Banach, M.; Rollinger, J. M.; Barreca, D.; Weckwerth, W.; Bauer, R.; Bayer, E. A.; Majeed, M.; Bishayee, A.; Bochkov, V.; Bonn, G. K.; Braid, N.; Bucar, F.; Cifuentes, A.; D'Onofrio, G.; Bodkin, M.; Diederich, M.; Dinkova-Kostova, A. T.; Efferth, T.; El Bairi, K.; Arkells, N.; Fan, T. P.; Fiebich, B. L.; Freissmuth, M.; Georgiev, M. I.; Gibbons, S.; Godfrey, K. M.; Gruber, C. W.; Heer, J.; Huber, L. A.; Ibanez, E.; Kijjoo, A.; Kiss, A. K.; Lu, A.; Macias, F. A.; Miller, M. J. S.; Mocan, A.; Müller, R.; Nicoletti, F.; Perry, G.; Pittalà, V.; Rastrelli, L.; Ristow, M.; Russo, G. L.; Silva, A. S.; Schuster, D.; Sheridan, H.; Skalicka-Woźniak, K.; Skaltsounis, L.; Sobarzo-Sánchez, E.; Bredt, D. S.; Stuppner, H.; Sureda, A.; Tzvetkov, N. T.; Vacca, R. A.; Aggarwal, B. B.; Battino, M.; Giampieri, F.; Wink, M.; Wolfender, J. L.; Xiao, J.; Yeung, A. W. K.; Lizard, G.; Popp, M. A.; Heinrich, M.; Berindan-Neagoie, I.; Stadler, M.; Daglia, M.; Verpoorte, R.; Supuran, C. T. Natural Products in Drug Discovery: Advances and Opportunities. *Nat. Rev. Drug Discovery* **2021**, *20*, 200.
- (87) Feher, M.; Schmidt, J. M. Property Distributions: Differences between Drugs, Natural Products, and Molecules from Combinatorial Chemistry. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 218.
- (88) Chen, Y.; Garcia De Lomana, M.; Friedrich, N. O.; Kirchmair, J. Characterization of the Chemical Space of Known and Readily Obtainable Natural Products. *J. Chem. Inf. Model.* **2018**, *58*, 1518.
- (89) Wetzel, S.; Schuffenhauer, A.; Roggo, S.; Ertl, P.; Waldmann, H. Cheminformatic Analysis of Natural Products and Their Chemical Space. *Chimia (Aarau).* **2007**, *61*, 355.
- (90) Lovering, F. Escape from Flatland 2: Complexity and Promiscuity. *Med. Chem. Comm.* **2013**, *4*, 515.
- (91) Sauer, W. H. B.; Schwarz, M. K. Molecular Shape Diversity of Combinatorial Libraries: A Prerequisite for Broad Bioactivity. *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 987.
- (92) Lovering, F.; Bikker, J.; Humblet, C. Escape from Flatland: Increasing Saturation as an Approach to Improving Clinical Success. *J. Med. Chem.* **2009**, *52*, 6752.
- (93) Taylor, R. D.; Maccoss, M.; Lawson, A. D. G. Rings in Drugs. *J. Med. Chem.* **2014**, *57*, 5845.
- (94) Ertl, P. Magic Rings: Navigation in the Ring Chemical Space Guided by the Bioactive Rings. *J. Chem. Inf. Model.* **2021**. DOI: 10.1021/acs.jcim.1c00761.