Deliberations on Natural Products and Future Directions in the Pharmaceutical Industry

Kathrin Buntin, Peter Ertl, Dominic Hoepfner, Philipp Krastel, Edward J. Oakeley, Dominik Pistorius, Tim Schuhmann, Joanne Wong, and Frank Petersen§*

§SCS Senior Industrial Science Award 2020

Abstract: Natural Products (NPs) are molecular 'special equipment' that impart survival benefits on their producers in nature. Due to their evolved functions to modulate biology these privileged metabolites are substantially represented in the drug market and are continuing to contribute to the discovery of innovative medicines such as the recently approved semi-synthetic derivative of the bacterial alkaloid staurosporin in oncology indications. The innovation of low molecular weight compounds in modern drug discovery is built on rapid progress in chemical, molecular biological, pharmacological and data sciences, which together provide a rich understanding of disease-driving molecular interactions and how to modulate them. NPs investigated in these pharmaceutical research areas create new perspectives on their chemical and biological features and thereby new chances to advance medical research. New methods in analytical chemistry linked with searchable NP-databases solved the issue of reisolation and enabled targeted and efficient access to novel molecules from nature. Cheminformatics delivers high resolution descriptions of NPs and explores the substructures that systematically map NP-chemical space by sp³-enriched fragments. Whole genome sequencing has revealed the existence of collocated gene clusters that form larger functional entities together with proximate resistance factors thus avoiding self-inhibition of the encoded metabolites. The analysis of bacterial and fungal genes provides tantalizing glimpses of new compound-target pairs of therapeutic value. Furthermore, a dedicated investigation of structurally unique, selectively active NPs in chemical biology demonstrates their extraordinary power as shuttles between new biological target spaces of pharmaceutical relevance.

Keywords: Analytical chemistry · Chemical diversity · Drug discovery · Genetic engineering · Heterologous expression · Natural products · Natural products fragments · Target identification · Whole genome sequencing

Kathrin Buntin studied molecular biotechnology at the TU Munich and received her PhD at the Institute for Pharmaceutical Research, Saarland in 2010. Joining Novartis the same year, her scientific work focuses on the elucidation of biosynthetic pathways for the production of natural products, the genetic engineering of natural product producing microbes and physiological aspects of their cultivation. Together with her team, she contributed to the development of several drug candidates.

Peter Ertl is a Novartis Leading Scientist working in the field of chemical informatics. He is author of more than 120 publications covering all areas of cheminformatics and computational chemistry and several cheminformatics methods and algorithms accepted as industry standards. His major research interests include analysis and visualization of chemical space, bioisosteric design and interactive web tools supporting the work of medicinal chemists.

Dominic Hoepfner did his PhD in the lab of Prof. Philippsen, Biozentrum in Basel, Switzerland followed by a postdoc with Prof. Tabak, University of Utrecht, Netherlands, where he discovered a fundamental biogenesis pathway of peroxisomes. Joining the Novartis Institutes for BioMedical Research in 2005 he has been working in the field of chemical genetics and chemogenomics. Dominic and his team are leveraging the CRISPR/Cas9 gene-editing technology to identify novel disease targets and to deconvolute the mechanism of action of compounds. In addition to his role at Novartis, he teaches at the University of Basel and the ESBS Strasbourg.

Philipp Krastel studied organic chemistry at the University of Goettingen where he received his PhD in natural product research. In his thesis he investigated the metabolic profile of different endo-symbionts from beetles and insects. In 2002 he started as a research investigator in the natural products group at Novartis Basel. He has been the group leader of the natural product chemistry team since 2008, focusing on the analytical characterization and preparation of natural products being pursued in preclinical research, chemical development and production.

Correspondence: Dr. F. Petersen, E-mail: frank.petersen@novartis.com, Novartis Institutes for BioMedical Research, Novartis Pharma AG, CH-4002 Basel

Edward James Oakeley performed his PhD studies in biochemistry in the laboratory of Colin Lazarus at The University of Bristol (UK) studying auxin-binding proteins of *Zea mays* and *Fragaria x ananassa*. Postdoctoral research studying DNA methylation of higher plants and floral regeneration *in vitro* was performed in the laboratory of Jean-Pierre Jost at the Friedrich Miescher Institute (FMI, Switzerland). He has held lecturer positions in cell biology and bioinformatics at the University of Exeter (UK) and genomics group leader positions at the FMI and Novartis (Switzerland).

Dominik Pistorius received a PhD in pharmaceutical biotechnology from Saarland University in 2011. Under the supervision of Prof. Rolf Müller he investigated novel mechanisms of bacterial secondary metabolite biosynthetic pathways. In 2011 he joined the Natural Product group at Novartis in Basel. His research interests are oriented towards application and further development of modern dereplication methods for rapid identification of known natural products from complex mixtures and the optimization of production processes by variation of physiological parameters, feeding of rate-limiting precursors and genetic engineering. These methodologies are applied to advance drug discovery projects at various stages in close collaboration with the disease experts.

Tim Schuhmann studied organic chemistry at the University of Bayreuth. After his diploma he joined the group of Axel Zeeck (University of Goettingen) for his PhD on isolation, biosynthesis and mode-of-action investigations of natural products. In 2005 he started his scientific career in the biotech industry where he led various interdisciplinary research projects. In 2010 Tim took over a lab head position in preclinical research at Novartis where he worked in natural product chemistry which included the generation of a NP-inspired fragment library. His current role is in hit generation sciences where Tim follows his passion for combining different sources and technologies to identify and further develop tractable assay hits to advance therapeutic projects.
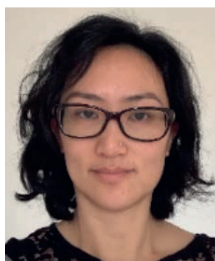
Joanne Wong received her PhD in biology, health and biotechnologies from the University of Toulouse (France) in 2007. After a postdoctoral fellowship spent at the Pasteur Institute (Paris) in the field of fungal genetics, she joined the Natural Product group at Novartis in 2011 as a lab head. Her current research interests address natural compound production and genetic engineering of biosynthetic pathways in eukaryotic systems (filamentous fungi, yeasts and plants).

Frank Petersen received his PhD with Hans Zähner at the University of Tübingen. He joined the Natural Product group at CIBA-Geigy in 1991 and took over the responsibility for Natural Product research at Novartis in 1999. He and his team elaborate concepts for the discovery of novel microbial natural products by linking HiRes-analytics with secondary metabolomics approaches. Together with further scientific expert teams Frank's group investigated novel natural products for the modulation of new target proteins. Since 2010 the research team has integrated Synthetic Biology concepts for the identification and engineering of new biosynthetic pathways. Frank initiated microbial bioprospection projects with institutes in Africa and Asia and received visiting professorships in China and South Korea.

1. Introduction

Natural products (NPs) interact with, and regulate, a plethora of physiological pathways in nature. In contrast to 'primary metabolites' which are essential for living cells, these 'secondary metabolites' are not survival-critical, however, they do represent chemical 'special equipment' for plants, microorganisms, and animals conveying survival advantages in their biological habitats.[1a–f] Due to their evolutionary relatedness, these 'fitness factors' can also modulate the activity of their target homologs and paralogs in higher animals.[2a–e] The biological functions of NPs make them a valuable source of chemical probes for discovering novel molecular interactions at the heart of diverse human diseases. Historically, numerous NPs have inspired therapeutic innovations in medicine. Only recently, the benzoylated derivative of staurosporin, developed by Novartis, received approval for the treatment of systemic mastocytosis and acute myeloid leukaemia, driven by the mutated genes Kit (D816V) and Flt3 (D835h/Y), respectively.[3] In 2019, this new drug received the Prix Gallien de Suisse. Although NPs are often perceived as being a predestinated source of antiproliferative or antibiotic molecules, in fact they play a key role in the discovery of many first-in-class-treatments in almost all therapeutic areas.[4a–c]

In drug discovery sciences, new concepts and technologies are constantly being investigated to maximize the chances of therapeutic innovation. Breakthroughs in synthetic chemistry are mirrored in the evolution of DNA-encoded or macrocyclic peptide libraries of breath-taking size. Advances in cheminformatics have enabled the screening of virtual compounds in algorithm-mediated screens. As we gain a richer understanding of disease mechanisms, biologics such as antibodies and recombinant peptides have changed the face of human therapy, while cell and gene therapies are currently writing a new chapter in transformative medical practice.

As a diverse set of chemical scaffolds, NPs are an important element of the drug discovery sciences contributing to the development of transformative medications for unmet medical needs (Fig. 1). However, to fully explore their potential in new pharmaceutical research directions, we have to go beyond the traditional path of identification and isolation of active metabolites from biological mixtures, which has been the prevailing route for decades.
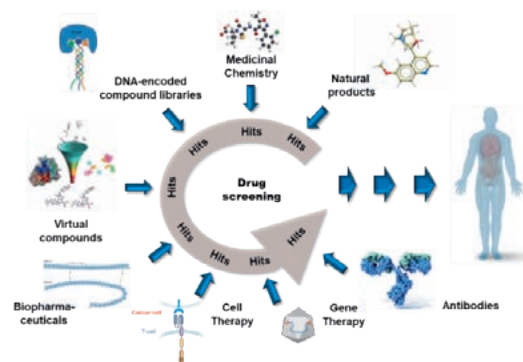

Fig. 1. Drug discovery in pharmaceutical research: sciences and technologies

Besides doing research *on* NPs our team is also exploring scientific questions by doing research *with* secondary metabolites. This involves improving access to new secondary metabolites by analytical methods or by genome sequencing, exploring the diversity of our in-house NP library, and finally, novel NPs and NP-like scaffolds provide potential therapeutic options when they can interact with new biological targets. In the following sections we give an insight into some of our research directions, which are of increasing importance in pharmaceutical NP sciences.

2. Accessing Natural Products Novelty

An efficient and stringent 'dereplication' process for the rapid identification of known secondary metabolites from complex biological matrices is a key requirement of NP research. In its original definition the term dereplication recapitulates the process of analysis of an active sample from preliminary screening by analytical methods and databases to identify known bioactive molecules and avoid the re-isolation of well-studied NPs.[5] Nowadays, the term has been detached from the bioactive sample and is used instead to explore the novelty of an analyte. There is a large and ever-growing number of both open and commercial NP databases and collections[6] (currently >120). Reconciling different structural classes, source organism metadata, experimental mass spectrometry (MS) or nuclear magnetic resonance (NMR) data can be extremely challenging for dereplication. Besides our internal and proprietary Novartis NP database, we make extensive use of the Chapman & Hall/CRC Press Dictionary of Natural Products (DNP) for dereplication purposes. For further enrichment and diversification of our collection, we follow a staggered approach. First, we dereplicate against our internal collection, if no hits are retrieved, we dereplicate against DNP. If we can identify the compound in DNP, we decide based on the compound class and the extent of research performed on the compound if a cultivation of the producer and subsequent purification of the compound for further biological profiling is worthwhile or not.

The technological advances in the chromatographic separation, spectrometric and spectroscopic methods and computational tools have made the dereplication process far more efficient and sensitive. The state-of-the-art analytical setup is a hyphenation of an ultra-high-performance liquid chromatography (UHPLC) system with a diode array detector (DAD) and an electrospray ionization high-resolution mass spectrometer (ESI-HRMS) greatly simplifies the dereplication of analytes from crude extracts. In this setting both UV/vis and MS detection can be highly biased concerning the signal intensities of a given analyte (small or large molar extinction coefficient and poor or strong ionization efficiency). To allow for the estimation of relative amounts of unknown analytes the additional coupling of a mass-flow sensitive detector such as a charged aerosol detector (CAD) or an evaporative light scattering detector (ELSD) is highly beneficial. The retention time (RT) of an analyte is an intrinsic chromatographic property and can be used as parameter for dereplication using in-house databases. The DAD records UV/vis spectra (*e.g.* 190–600 nm) of all analytes with at least a minimal chromophore after separation by liquid chromatography (LC). The UV/vis-maxima (λ_{max}) extracted from these spectra can be used for queries in external databases *e.g.* DNP. In the best-case scenario, the chromophore(s) of a compound can result in highly characteristic UV/vis spectra, which allow assignment to a certain compound family (*e.g.* polyenes, indolocarbazoles, phenazines, and actinomycins). However, a more powerful and reliable approach for dereplication is the computational matching of UV/vis spectra using in-house databases. The Chromeleon 7.3™ chromatography data system (ThermoFisher) offers a powerful and efficient UV/vis library search functionality. Thousands of reference UV/vis spectra can be queried within a few seconds. ESI-HRMS[(2)] data offer a rich and complementary set of information for the analyte. The accurate mass and isotopic

patterns obtained can either be used directly to perform database searches within a narrow range (*e.g.* 1–2 ppm), or they can be used to calculate the molecular formula which can be used for even more stringent database searches. At Novartis, we have systematically compiled relevant metadata such as RT, accurate mass and source organism in our in-house UV/vis library with a total of ~12,500 unique records from microbial sources (~ 4,700 actinomycetes, ~ 6,500 fungi, and ~1,300 from Gram-negative bacteria). This collection represents a one-stop solution for the unambiguous dereplication of analytes against our internal compound collection. However, such an approach is limited to in-house libraries that have been built from isolated and characterized pure compounds, as there is no extensive public repository for UV/vis spectra in a machine-readable format and the existing commercial software solutions for UV/vis spectral matching are mostly vendor specific. Additional structural information is contained in the HRMS[2] spectra of an analyte. The fragmentation spectra can either be used to differentiate isobaric compounds that cannot be discriminated based on their associated characteristics or it can be used to perform library searches against an MS[2] library. Large public repositories of MS[2] spectra like MassBank of North America[7] (MoNA; >200,000 records) and METLIN[8] (>1,000,000 records) are available, but these contain only a small fraction of NPs. The major drawback of classical MS[2] library searches is that only exact matches can be identified and no information concerning the structural class of an unidentified analyte can be made. This changed with the development of molecular networking in 2012.[9] The rationale behind this method is the observation that molecules with similar structural features share similar MS[2] spectra. The calculated spectral similarities are visualized as a molecular network with nodes, representing an MS[2] spectrum, and edges connecting those nodes above a user defined spectral similarity cut-off. This method even allows for the grouping of novel members of a previously characterized compound family without the need for an exact match to a reference spectrum. One cornerstone of this new methodology is that many researchers in the field have joined forces to make contributions to the Global Natural Products Social Networking (GNPS),[10] the most comprehensive collection of NP MS[2] spectra with currently ~10,000 unique records of annotated compounds. The concept of molecular networking has been further extended and refined by the integration of various tools into the GNPS platform.[11] Our experience with molecular networking has identified several limitations with its massive application in an industrial setting. First, the demand that proprietary datasets need to be uploaded onto the public servers of GNPS requires the setup of the computational services on internal servers to mirror these systems and the calculation of large networks requires significant computing time. Further, the public dataset is negatively impacted by inherent spectral differences caused by non-standardized fragmentation energies and the use of diverse types of instruments from different vendors. Finally, the accuracy of the network topology, and thereby the predictive power of the obtained grouping, is highly dependent on several user-defined parameters. These can be optimized for a certain compound class, but from our experience, it is not possible to define parameters that are optimal for datasets of thousands of spectra from extremely diverse compound classes. Nevertheless, the methodology has been a true game changer for many groups in academia that do not have the luxury of an extensive in-house library. In addition, there are computational tools like DEREPLICATOR+[12] that match *in silico* fragmentation of structures from databases with experimental MS[2] spectra.

The structure elucidation of complex NPs of higher molecular weight is performed mainly by interpretation of extensive 1D and 2D NMR experiments (^{1}H, ^{13}C, ^{1}H-^{1}H COSY, ^{1}H-^{13}C HSQC, ^{1}H-^{13}C HMBC, [1]H-^{1}H ROESY) and can be a laborious process. The technological progress in NMR instrumentation and probe minia-

turization over the past decade has led to a dramatic improvement in sensitivity and allows for the acquisition of high quality 2D heteronuclear correlation experiments from sample quantities below 1 mg. In the recent past, computational methods for computer-assisted structure elucidation (CASE) using these NMR datasets have matured[13] and commercial programs are available (*e.g.* ACD Labs Structure Elucidator and Mestrelab MNova Structure Elucidator) to support the analyst. The importance of HRMS[2] and HRMS[n] for *de novo* structure elucidation has likewise increased and there is an ever-growing number of computational tools available.[14] For specific compound classes (*e.g.* peptidic (sub)structures) the information obtained thereof, like the amino acid sequence, can be highly complementary to that obtained from NMR analysis. In addition, 'old-fashioned' technologies like UV/vis and infrared (IR) spectroscopy can offer crucial insights about functional groups that are elusive to NMR and MS. In summary, the analysts have recently been provided with a large and ever growing computational toolbox to support and speed up their work, but finally maximum efficiency in NP structure elucidation also requires teamwork, expert knowledge and experience.

Another technology that has positively affected our research is the development of systems for ultrafast sample cleanup based on solid phase extraction (SPE) coupled to a HRMS system, *e.g.* the RapidFire platform from Agilent. This platform permits the high-throughput analysis with cycle times <10 seconds per sample, thereby enabling the screening of many mutants, as typically obtained from random mutagenesis approaches that aim for the identification of mutants with increased productivity for a specific compound.

3. Assessing Natural Products Diversity

The progresses in analytical technologies have enabled our research group to isolate more than 700 NPs per year, all being new for Novartis. 10% of these are first-in-class-scaffolds. Surprisingly, this ratio has remained constant for more than 15 years now. Complemented by external compound sourcing partnerships we have created a library of 30'000 structurally elucidated natural compounds representing one of the largest NP sets in industry. This asset provides a highly competitive advantage for the multi-dimensional exploration of these molecules.

With the advances in computational chemistry, the description of NP specific features gains higher resolution and opens the door to rationally designing appropriate screening libraries. The *in silico* dissection of NPs into their substructures, being accessible by synthetic chemistry, represents an underexplored opportunity to interrogate biological pathways with sp[3]-enriched NP-like fragments. Some applications of these ideas are described in the following pages.

NPs are probably the most diverse class of molecules. Their structures range from small molecules consisting of only a few atoms, through medium-sized richly substituted aliphatic or aromatic ring systems, up to complex structures with numerous fused rings or large macrocycles. To illustrate this diversity 100 representative molecules from the Novartis collection are shown in Fig. 2.

To better visualize the chemical space covered by NPs and compare it with that of average drug-like molecules we displayed the chemical space of these structures as a 2-dimensional plot (Fig. 3).
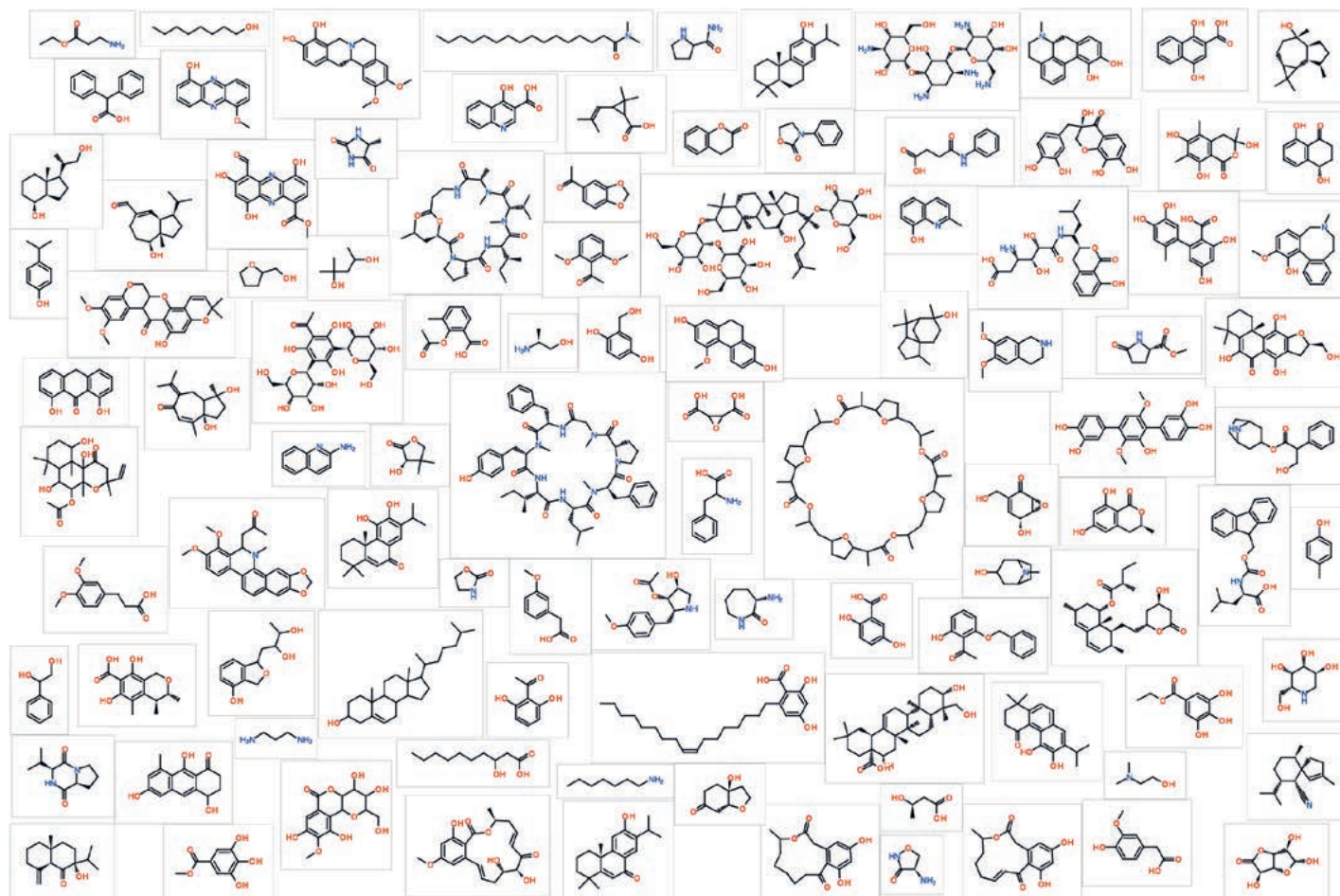

Fig. 2. Examples of NPs from the Novartis collection illustrate the high diversity of this class of molecules.
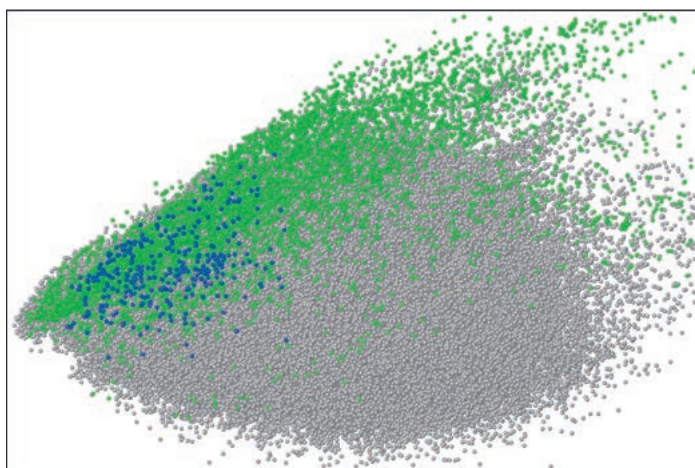
Fig. 3. Comparison of the chemical space covered by synthetic molecules from the Novartis screening deck (gray), NPs (green) and NP-inspired fragments (blue). The horizontal axis of the plot roughly represents the molecule size and the vertical axis complexity, or feature richness.

The molecules were characterized by simple substructure features called Scaffold Keys, developed in-house at Novartis[15] and used already with success in the analysis of chemical space for bioisosteric design.[16] The dimensionality of the set was reduced by principal component analysis (PCA). Of course, when reducing the dimensionality, some information is lost, but the PCA minimizes the loss of information for a given number of dimensions. The resulting plot is shown in Fig. 3. Novartis NPs (in green) are compared with synthetic molecules from the Novartis screening deck[17] (in grey) and a set of NP-inspired fragments (in blue) that have been obtained by fragmentation of NP molecules and are used in-house in fragment-based screening.[18] The horizontal axis of the plot represents the approximate molecule size (small molecules are on the left, large on the right) and the vertical axis complexity, or feature richness, with simple molecules at the bottom and more complex ones at the top of the graph. This plot clearly documents the unique characteristics of NPs that form a valuable addition to the 'standard' synthetic molecules contained

in the screening deck. A large portion of the Novartis NPs are novel. Approximately half of the in-house NP structures are found within the COCONUT[19] database of more than 420,000 published NP structures. All others are novel.

Numerous studies have already analyzed the difference in physicochemical properties between NPs and synthetic molecules, including also our earlier study published in CHIMIA.[20] Therefore, it is not necessary to cover this topic here. A more interesting and less explored area is the study of substructure features of NPs and how they differ from synthetic molecules. Our earlier study focused on the analysis of NP scaffolds not only found clear differences between scaffolds of NPs and synthetic molecules, but also could classify scaffolds into distinct groups based on the class of organism producing the particular NP.[21] Another very striking feature that distinguishes NPs from synthetic molecules are their unique functional groups (FGs). FGs are an important part of molecules forming specific interactions with target proteins *via* hydrogen bonds and other weaker interactions. FGs also determine the reactivity, bioavailability and metabolic stability of the parent molecules. To learn more about this interesting topic we analyzed FGs present in the Novartis NP collection and compared them with FGs present in synthetic molecules from the Novartis screening deck. The extraction of FGs has been performed using a method developed in-house and previously published.[22a,b] Comparison of the 50 most frequent FGs from each set (66 unique FGs in total) are shown in Fig. 4. The horizontal axis in the diagram represents FG frequency (the most common groups are on the left, less common on the right) whereas the vertical axis indicates the propensity of FGs for NPs (green area at the top) or synthetic molecules (blue area at the bottom). The figure nicely visualizes the differences between these two sets. FGs typical for NPs contain mostly oxygen atoms (hydroxy, ester, peroxide, polyglycol, epoxide rings), ethylene-derived groups and various α,β-unsaturated systems. In the synthetic molecules nitrogen-containing and chemically more easily accessible FGs are over-represented, for example amide, urea, sulfonamide, sulfone or imide functionalities and substituents like fluoro, cyano and nitro.

Another in-house developed cheminformatics methodology supporting NP research is an algorithm to calculate NP-likeness.[23a,b] This descriptor may be defined as the structure and
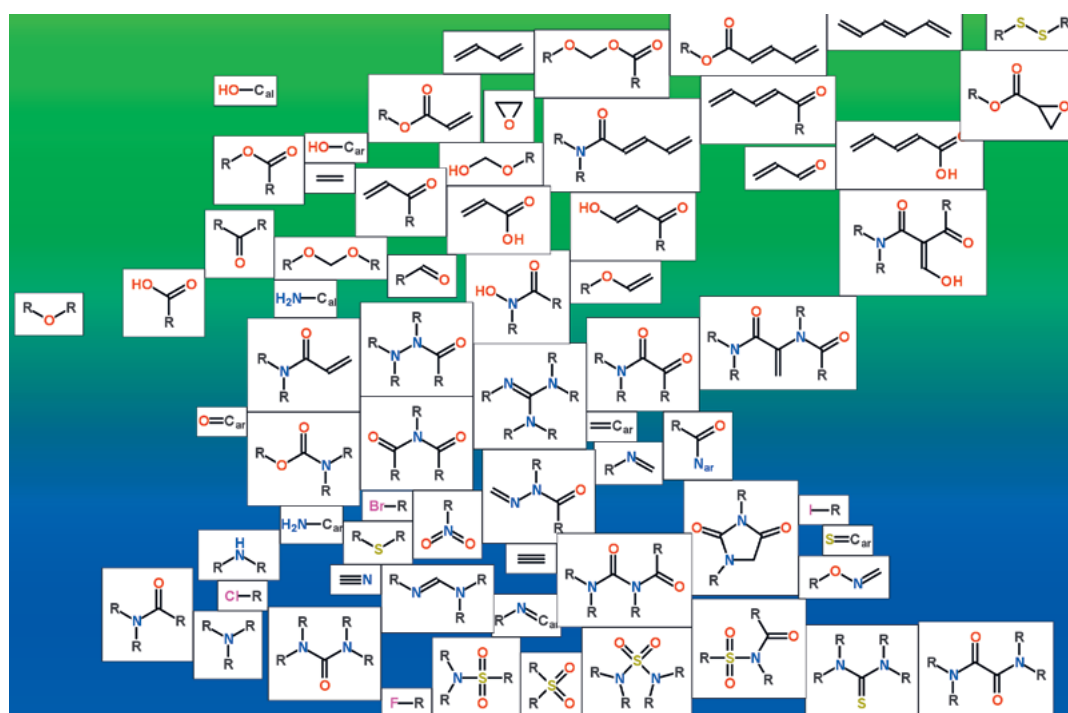

Fig. 4. Plot of common functional groups displaying their prevalence in NPs (green area) and synthetic molecules (blue area), position on the horizontal axis is proportional to the frequency of functional groups – the most common groups are on the left, less common on the right.

property similarity of analyzed molecules to the chemical space occupied by NPs. The NP-likeness is calculated as a Bayesian sum of fragment contributions, trained on large sets of NPs and synthetic molecules to best separate these two groups. Application possibilities of this descriptor in drug discovery are indeed numerous. The most obvious is virtual screening, whereby molecules, for example from commercial compound catalogues, are selected using not only the standard features like physicochemical properties or pharmacophores, but also based on their good NP-likeness. The same procedure may be applied also for the selection of fragments for fragment-based screening.[18] Another application area of NP-likeness is library design. The scaffolds and building blocks that are used for library construction may be selected to have good NP-likeness, leading to synthetic libraries that contain structural features of NPs. All these methods are actively applied in the ongoing drug discovery activities at Novartis.

In summary, the cheminformatics analysis of NPs in the Novartis compound collection demonstrates the high diversity and novelty of our molecules providing a competitive edge for Novartis research. Several new cheminformatics methods and algorithms have been developed in-house for the analysis of NPs to support our drug discovery, virtual screening, fragment-based screening and library design activities.

4. Natural Product Fragments

Fragment-based drug discovery (FBDD) has matured into a well-established and widely used technology that reliably delivers relevant chemical starting points for therapeutic drug projects, and which complements the outcome of more traditional hit finding strategies like high-throughput screening of large compound libraries.[24] This is achieved by screening small molecules using technologies suitable for detecting low-affinity interactions with the targets of interest. Resulting hits are advanced by applying structure-based design principles paralleled by computational

methods to yield potent leads with favorable molecular properties as demonstrated by the approval of four drugs and the initiation of dozens of clinical trials resulting from FBDD efforts.[25] To access a significantly large area of chemical space three-dimensionality has been identified as a key contributor to be considered for the design of screening libraries.[26]

NPs are inherently enriched for sp3 carbons and have by design a high propensity to interact with proteins as they are biosynthesized by enzymes with the task to specifically modulate biological processes – an attribute which has proven to be transferable to NP fragments.[27] For introducing the unique NP benefits to FBDD we decided to identify and include common NP-motifs of appropriate size and molecular properties in our fragment screening collections. We achieved this by acquisition and *de novo* synthesis of suitable molecules which included chemical degradation of larger NPs with known bioactivities as exemplified in Scheme 1.[18]

Screening of our NP-derived fragment set against multiple targets has provided several new chemical starting points for exploration within medchem teams and facilitated access to structural data supporting efficient rational ligand design. Similar library concepts have been applied by various other groups leading to the identification of potent p38a MAP kinase inhibitors,[28] fragments with antimalarial activity,[29] and – *via* a NP fragment-linking approach – also enabled the discovery of novel GLUT-1 and -3 inhibitors.[30] Another remarkable example illustrating the potential of the latter concept is the discovery of the potent and selective antimalarial compound cipargamin currently in clinical investigations.[31a,b] It is built up starting from the synthetic regioisomers of the naturally occurring *N*-methyltryptamine and 6-bromoisatin that undergo a Pictet-Spengler-type cyclization to form the unique spiroindolone motif of cipargamin (Scheme 2).[32]

In summary, NP-derived FBDD demonstrates a validated novel concept that has already proven its impact on several drug discovery projects. Further developments have shown the value of linking NP fragments to form unique pseudo-NPs. More efficient ^{19}F NMR- based screening strategies of larger NP-inspired fragment collections have been enabled by late-stage fluorination of these molecules.[33]

## 5. Molecular Genetics and Genome Sequencing

In recent years, we have seen a profound expansion in the way we discover new microbial NPs. Independent of cultivation, we now have the tools to get a 'biosynthetic NP inventory' by directly sequence genomes, and predict and classify biosynthetic clusters and their encoded NPs. Sequencing technologies either deliver data in short regions a few hundred bases in length (*e.g.* the HiSeq/NovaSeq platforms manufactured by Illumina Inc, San Diego, CA) or long regions of tens to hundreds of thousands of bases in length (*e.g.* the Sequel IIe platform manufactured by Pacific Biosciences, Menlo Park, CA; or the MinION/GridION platforms manufactured by Oxford Nanopore Technologies Ltd, Oxford, UK).[34a,b]

The first two complete bacterial genome sequences were published in 1995,[35] and today bacterial genome sequencing is affordable to a great number of labs, reflected by 345908 uploaded prokaryotic genome sequence entries at NCBI.[36] For a long time,
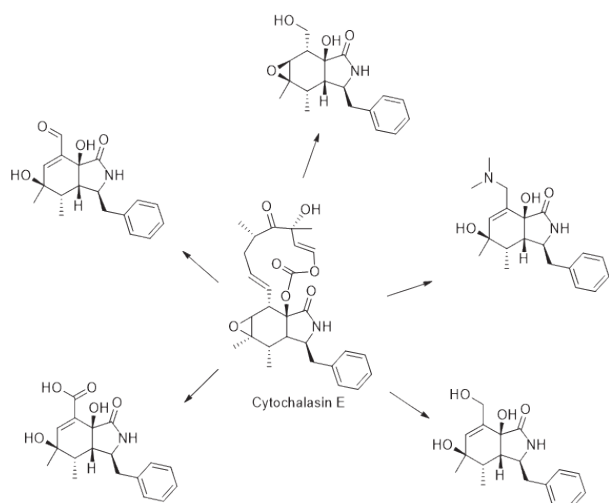
Scheme 1. Example for fragment generation by chemical cytochalasin E degradation.

Scheme 2. NP-like fragment merging leading to the antimalarial cipargamin.
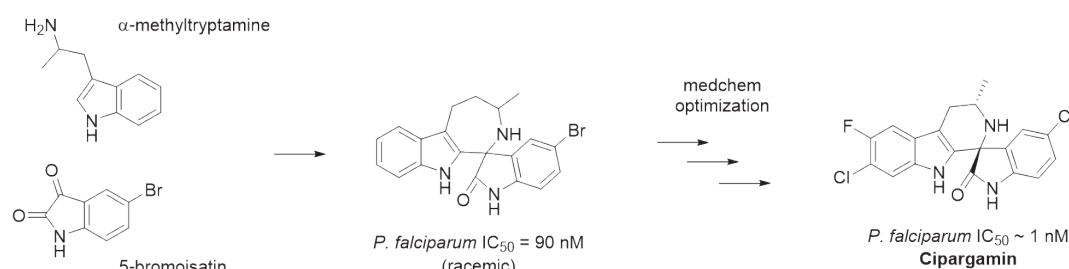
short read sequencing had the advantage in base quality with most reads between Q30-Q40,[37a,b] but the advent of HiFi sequencing has shifted the vast majority of long reads to an even higher quality level (Fig. 5). This allows accurate phasing and assembly of very closely related sequences. However, it is believed that only around 1% of all microbes are amenable to isolation and growth as pure isolates in the laboratory.[38a,b] An alternative approach is metagenomic sequencing where DNA is extracted from complex sources such as soil or the mammalian digestive tract and directly analyzed.[39] As the most abundant bacteria are not necessarily the easiest to culture, such sequencing efforts when combined with assembly using tools such as CANU[40] or HiFiASM[41] have the potential to unlock this microbial 'dark matter' because amongst the sequenced microbes, natural compound producers are present. The challenge is to analyze the generated data in a meaningful way.
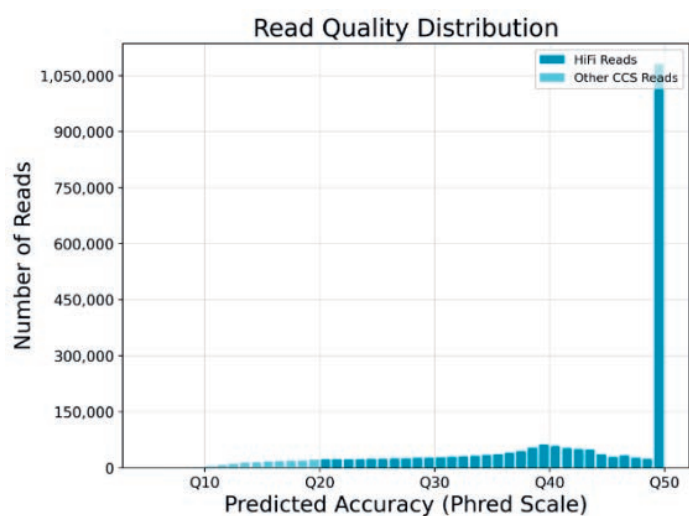

Fig. 5. Read quality bins on the Phred score (Q50 = one error in 100,000 bp) observed with HiFi sequencing on the Sequel IIe system (Pacific Biosciences). Source: *internal bacterial sequencing data, SMRT Link v10.1.*

In 2011, the first version of antiSMASH[42] was published, which can identify, annotate and analyze from currently 70 different types of secondary metabolite biosynthesis gene clusters (BGCs) based on bioinformatic algorithms. The rapid development of bioinformatic tools for natural product BGCs transformed the way in which novel secondary metabolites can be discovered today. *In silico* mining of sequenced microbial genomes makes it possible to identify potentially outstanding strains by the BGCs they harbor, before the cultivations even start in the lab. By mining public databases, novel taxonomic classes that have been neglected so far, might be identified as potential new NP sources. Comparison and collection of identified BGCs and their compounds in databases like MIBiG[43] allow a better analysis and understanding of NP biosynthesis. Repositories of BGC information are starting points for data science-driven approaches aiming at the discovery of pharmaceutically relevant compounds *via* at least two promising avenues, potentially impactful to prioritize uncharacterized BGCs (i) adjacent BGCs that encode synergistic compounds (ii) BGCs encoding a compound together with an insensitive isoform of its own target. In our lab, the discovery of a hypercluster producing rapamycin and actinoplanic acid in *Streptomyces rapamycinicus*,[44] and the colocalization of genes encoding cyclosporin biosynthesis and a mutated cyclophilin A variant (unpublished data) added respectively to a growing body of evidence in the literature.[45a,b] The availability of the complete

DNA as well as the protein sequence of a specific pathway also opens doors for tailored genetic engineering. Novel genetic engineering tools like CRISPR (Clustered Regularly Interspaced Short Palindromic Repeats) editing and silencing have been adapted to NP-producing bacteria[46] and allow targeted modifications of the genes resulting in *e.g.* novel analogs. Whole genome information of a producer strain facilitates titer increase of a specific secondary metabolite by upregulation of positive regulator genes, exchange of native promoters or deletion of negative regulators. In addition, deletion of BGCs from unwanted side products hampering the isolation of the target compound as well as engineering of metabolic fluxes to optimally support the biosynthesis with building blocks are targets for genetic modifications. To the surprise of the NP research community, it became obvious through the increasing number of genomes from NP-producing microbes that many BGCs are silent under standard laboratory conditions.[47] Consequently, many research groups including us initiated efforts to activate these 'silent' gene clusters by genetic manipulation or varying cultivation conditions. Besides working in the native producer strain, with decreasing DNA synthesis cost (0.09 US Dollar/bp), whole gene clusters can be printed synthetically and expressed in heterologous hosts with cleaner backgrounds and well-understood physiology. Complete synthesis of pathways might also allow mixing and matching of known pathways to generate novel compounds in the near future as researchers start to elucidate further details about the biosynthetic logic of certain pathways.[48]

6. Natural Products in Target Discovery

Identification of novel leads is accomplished by performing screens on purified proteins that have been validated as suitable targets or by using disease-relevant phenotypic readouts. The latter is currently seeing a revival as screens embracing biological complexity (by screening complex readouts in primary cells, cocultures, organoids or tissues) have delivered more first-in-class medicines than other approaches.[49] Unfortunately, this strategy requires known targets and mechanisms of action.

Unbiased target identification methods can be separated into two classes: biochemical and genetic assays (summarized in Table 1 and 2, in-house NP target identification examples summarized in Table 3, chemical structures in Fig. 6). Biochemical methods leverage the binding affinity of the compound to its target protein either by exposing it to a panel of pre-purified proteins in solution (SEC-TID),[50] immobilized proteins on arrays,[51a–c] or nonpurified proteins in complex cell lysates coupled with a readout measuring the interaction. Methodologies like the cellular thermal shift assay, CETSA[52a,b] target identification using drug affinity responsive target stability, DARTS[53] or limited proteolysis assay, LIP[54] can be conducted with the original lead molecule but protein arrays or affinity pull-down experiments require modifications of the chemical probe with an immobilization linker, fluorescent label or photo-crosslinking moiety.[51c,55] In particular, with complex molecules like NPs, chemical modifications can be a significant challenge as recently reviewed in detail by Chen *et al*.[56] But despite the requirement for elaborate chemistries, inhouse affinity pull-down experiments have supported the target identification of many natural compounds including cyclomarin A,[57a,b] nannocystin A,[58] OSW-1,[59] and haperforin G/harrpernoid D.[60]

Genetic methodologies do not require modification of the chemical probe and can provide resolution down to the binding pocket or individual, interacting amino acids. Beyond leveraging the natural expression profile of sensitive and insensitive cell lines that facilitated the target identification for englerin A,[61a,b] there are two general approaches: modulation of compound sensitivity of cells by genome-wide gene deletions or RNAi-based knock-down and generation of resistance-conferring residues by

Table 1. Compiled list of presented target identification approaches by biochemical methods.

	Biochemical				
Method	**SEC-TID** Size-Exclusion Chromatography for Target Identification	**CETSA** Cellular Thermal Shift Assay	**LIP** Limited Proteolysis **DARTS** Drug Affinity Responsive Target Stability	**Protein Array**	**Chemoproteomics** a) Affinity pull-down of target by immobilized compound b) Photo-affinity labelling of target by functionalized compound
Assay Principle	Compound is added to purified, arrayed proteins in solution followed by size-exclusion chromatography	Compound binding stabilizes target leading to decreased susceptibility of the target protein to unfolding at elevated temperatures	Compound binding stabilizes target leading to decreased susceptibility of the target protein to degradation by proteases	Compound is incubated with immobilized, purified proteins arrayed on a matrix. Specific binding is assayed by competition and washing	a) Compound is immobilized and exposed to cell lysates. Specific binding is assayed by competition and washing. b) Compound is modified with a photo-activatable cross-linker exposed to cell lysates and cross-linking induced
Modified / labeled compound required	No	No	No	Yes	Yes
Detection method	Mass spectrometry	Mass spectrometry	Mass spectrometry	Imaging	Mass spectrometry
Weakness / Blind spot	Covers only purified proteins present in array, false positive rate due to unspecific binding	Small difference in stabilization requires in-depth MS analysis of many samples from temperature gradient. Lowly expressed proteins are difficult to identify	Small difference in stabilization may require in-depth MS analysis of many samples from digestion time-course. Lowly expressed proteins are difficult to identify	Requires chemical functionalization of compound. Covers only purified proteins present in array. Linker of both, protein on array and compound linker may interfere with binding	Requires chemical functionalization of compound. Compound linker may interfere with binding. Lowly expressed proteins and membrane proteins are difficult to identify
References	[50]	[52]	[53,54]	[51]	[51c,55]

unbiased mutagenesis followed by isolation of resistant cells and genome-sequencing to find the underlying mutations. Sensitizing cells by using genome-wide deletion collections has first been implemented in diploid baker's yeast by introducing heterozygous deletions. The reduced gene-copy number leads to compound-induced haploinsufficiency in the strain carrying a deletion in the gene encoding the target.[62] The deletion strains are pooled, grown in the presence or absence of the test compound and the strain compositions of the pools compared. The simple growth requirements and rapid growth of yeast cells have enabled the profiling of large compound collections.[63a,b] For NPs, this assay has demonstrated the ability to produce high quality data. NPs are often potent and highly selective, and modulate evolutionarily conserved targets that are well represented in baker's yeast. One example that demonstrates this is cladosporin that was identified as possessing strong anti-plasmodial characteristics and thus antimalarial activity. Using the yeast system, it was found to target the ATP-binding pocket of lysyl-tRNA synthetases.[64] Despite the very high conservation of this pocket with the human enzyme, this natural compound shows ~1000 fold selectivity for the plasmodial enzyme. This demonstrates the power of naturally evolved compounds to confirm the tractability of targets for drug-discovery purposes that otherwise would likely be dismissed. As a consequence, this triggered a spur of activities to harness lysyl-tRNA synthetase as an anti-infective target.[65a,b] Other examples of highly conserved targets that were identified using the yeast

system are novolactone,[66] decatransin[67] or rocaglamide A[68] that were followed up because of their anti-cancer profile, and FR17145 that was a hit in a hepatitis C replicon assay.[69] Despite the diverse biological activity and origin, the efficacy targets could all be identified in the yeast assay.

For its relatively small genome of 13 Mb, lack of large intergenic regions and clonal growth on solid media, baker's yeast also provides a good starting point for the identification of resistance-conferring residues. Cells are mutagenized in a random fashion by exposure to UV irradiation or chemical mutagens, then plated on growth-inhibitory compound concentrations and surviving colonies picked for whole-genome sequencing and analysis. This approach has been used for decades to identify the targets of synthetic and natural compounds active on eukaryotic cells, notably in the identification of mTOR as the target of rapamycin.[70] Another classic example is cyclosporin A where yeast also played a role in deciphering the mechanism of action.[71a–c] Cyclosporin A has recently moved back into focus due to the discovered interactions of its target with viral polymerases, and thus potent antiviral activity that could also be confirmed against SARS-CoV-2.[72a,b] In the case of the natural compound argyrin A yeast mutagenesis was used as a validation for the target hypothesis generated by mutagenesis in bacteria[73] but it also served as an orthogonal target validation method for the chemogenomic profiling papers cited above.[64,66,69] Identification of resistance conferring residues allowed in many cases the binding pocket to be hypothesized and

Table 2. Compiled list of presented target identification approaches by genetics methods

	Genetic					
Method	**Expression profiling**	**Pooled HIP** Haplo-insufficiency profiling	**Pooled CRISPR profiling**	**Pooled RNAi sensitization profiling**	**Haploid cell profiling**	**Muta-Seq** Mutagenesis followed by sequence identification of resistance mediating mutations
Assay Principle	Linking compound sensitivity / resistance to gene expression profiles to deduce target information	Genome-wide, precise deletions of one gene allele in diploid cells compromise synthesis of the corresponding protein and sensitizes cells with deletion in the target-coding gene upon exposure of sub-lethal compound dose	Genome-wide gene-editing induces mutations, compromises protein synthesis and upon exposure to compound modulates response to compounds of cells with mutations in the target-coding gene (can be sensitivity, can be resistance)	Genome-wide knock-down of protein synthesis using RNA interference sensitizes cells with reduced target protein levels to sub-lethal compound dose	Insertional mutagenesis (transposon or retrovirus-based) to generate loss of function mutations in haploid cells (no compensatory second gene-copy) followed by screening for resistance	Random mutagenesis using irradiation or chemical mutagens to generate point mutations that lead to dominant resistance
Modified / labeled compound required	No	No	No	No	No	No
Detection method	Expression profiling / sensitivity profiling	Detection of deletions by microarray / sequencing	Determination of sgRNA abundance by sequencing	Determination of RNAi abundance by sequencing	Mapping of insertion sites by sequencing	Mutation detection by whole genome sequencing
Weakness / Blind spot	May require significant number of cell lines to delineate relevant group of correlating genes. Gene-expression may not be linked to compound sensitivity / resistance	Requires generation of precise, genome-wide haploid detection collection. Currently this is only possible in systems with efficient homologous recombination (fungi). Mostly restricted to live/death readouts	Functional gene redundancy can lead to false negative hits. Large assay volume / cell number required to keep good genome-wide coverage. Restricted to live/death or flow-cytometry compatible readouts	Functional gene redundancy can lead to false negative hits. Significant number of false positive hits due to off-target effects. Large assay volume / cell number required to keep good genome-wide coverage	Insertional bias of used system makes genome-wide coverage challenging and leads to high complexity. Mapping of insertion sites is not trivial. Restricted to haploid cells (mouse embryonic stem cells, human leukemia cancer cells) and mostly live/death readouts	Identification of resistance-conferring mutations requires good sequencing coverage (cost) and sequencing of good number of clones to differentiate between causative mutations and by-stander mutations. Complex analysis. Restricted to live/death readouts
References	[61]	[61,62]	[80,81]	[79]	[76,77]	[58,64,67,70,84]

subsequent *in silico* docking to be performed. In some cases, a co-crystal structure is now available, confirming the validity of the approach.[66,73,74]

Although selection of resistant cells and underlying mutations can be challenging in non-clonal growing organisms like yeast, the power of the approach has inspired researchers to successfully apply it to other species, *e.g.* plasmodia. In the case of the natural compound-derived antimalarial drug candidate spiroindolone NITD609 this allowed us to pinpoint the mechanism of action to a P-type ATPase.[31a,75]

But despite the high evolutionary conservation of yeast genes the simple fact that baker's yeast has ~6000 annotated genes, rather than the ~20,000 found in humans, explains in part why we need target identification approaches that operate directly in mammalian systems. The publication of protocols to produce haploid mouse embryonic stem cells opened the door for genome-wide forward and reverse genetics in mammalian cells. Using insertional mutagenesis, the essential components for a variety of natural probes including ricin toxin were discovered.[76a,b] Generation

of haploid leukemia-derived human cancer cells allowed similar experiments to be conducted in a human context. These have provided valuable insights into NP-derived microtubule poisons or viral entry factors.[77a,b]

Unravelling and harnessing of the CRISPR/Cas gene-editing mechanism and the availability of genome-wide sgRNA libraries has opened the door for genome-wide genetics in most cell lines or organisms of interest.[78a–c] Before CRISPR became available, genome-wide RNAi-derived methods were tested for target identification in human cells but obtained profiles that were complex and difficult to interpret.[79] The simplicity and precision of the CRISPR/Cas system produced much cleaner profiles and although gene-editing results in editing of both alleles over time, early time-points have allowed us to capture heterozygous states and recapitulate the compound-induced haploinsufficiency known from the yeast assay.[80] This assay was used in-house to successfully support the deconvolution of the mechanism of action of NPs like cavinafungin,[81] chivosazol F[82] and kendomycin. [83] As demonstrated with the cation chelator kendomycin, genetic

Table 3. Compiled list of in-house target identification projects

Compound	Target	Target identification technologies	Ref.
Argyrin B	mEFG	Resistance profiling	[73]
Cavinafungin	SEC11A	Mammalian chemogeno-mic profiling, resistance profiling	[81]
Chivosazol F	ACT1	Mammalian chemogeno-mic profiling, chemical proteomics	[82]
Cladosporin	P.f. LysRS	Yeast chemogenomic pro-filing, resistance profiling	[64]
Cyclomarin A	M.t. ClpC1	Affinity proteomics	[57a]
Cyclomarin A	P.f. Ap3Aase	Affinity proteomics	[57b]
Decatransin	SEC61A1	Yeast chemogenomic pro-filing, resistance profiling	[67]
Englerin A	TRPC4	Expression profiling	[61]
Harrpernoid D	Mitochon-drial proteins	Affinity proteomics	[60]
Kendomycin	Cation chelation	Mammalian chemoge-nomic profiling, yeast chemogenomic profiling	[83]
Nannocystin A	eEF1A	Affinity proteomics, che-mogenomic profiling	[58]
Novolactone	HSP70	Yeast chemogenomic pro-filing, resistance profiling	[66]
OSW-1	OSBP, OPR4L	Affinity proteomics	[59]
Rocaglamide A	eIF4A	Yeast chemogenomic pro-filing, resistance profiling	[68]
Spiroindolone NITD609	P.f. ATP4	Resistance profiling	[31a, 75]

profiling also has the advantage of still producing an informative dataset, even if the compound does not directly interact with a protein target.

Generation of resistance conferring residues and screening for resistant clones upon compound selection also works in mammalian cells[84] and was successfully used in-house to identify the target of natural compounds with anticancer properties like decatransin[67] or nannocystin.[58] But despite increased sequencing power, the size of the human genome and its epigenetic control mechanisms contribute to compound resistance. Thus, *de novo* target identification using this assay is still a complex undertaking. However, the availability of base-editor CRISPR systems that allow the focused mutagenesis of defined chromosomal regions may provide a powerful validation strategy once a target hypothesis has been generated by another method.[85]

In summary, target deconvolution has significantly benefitted from advances in mass spectrometry resolution and coverage, as well as from advances in genetic methodologies and sequencing capacity. But it is also clear that there is still no one size fits all solution, and each project requires the dedicated interplay of chemists, biologists, and data scientists.

7. Inclusion of Natural Products in Pharma Research

Industrial NP research depends on the scientific and technological innovations mainly elaborated by academic research groups. However, those teams are not only critical in progressing the basic research of NPs. They are of high importance in closing those scientific gaps which limit the application of NPs in pharmaceutical R&D. Therefore, academic and industrial NP research have also shared responsibility to move secondary metabolites closer to those scientific areas of high relevance in pharmaceutical sciences and medicinal chemistry. This requires an elevated interdisciplinary engagement and orthogonal perspective to our research mission.

The following research fields are among those in which medicinal chemistry and pharmaceutical biology are heavily investigated, to discover transformative therapies which address unmet medical needs: novel therapeutically relevant targets and new ways to regulate their activities;[86] new modalities for pathway
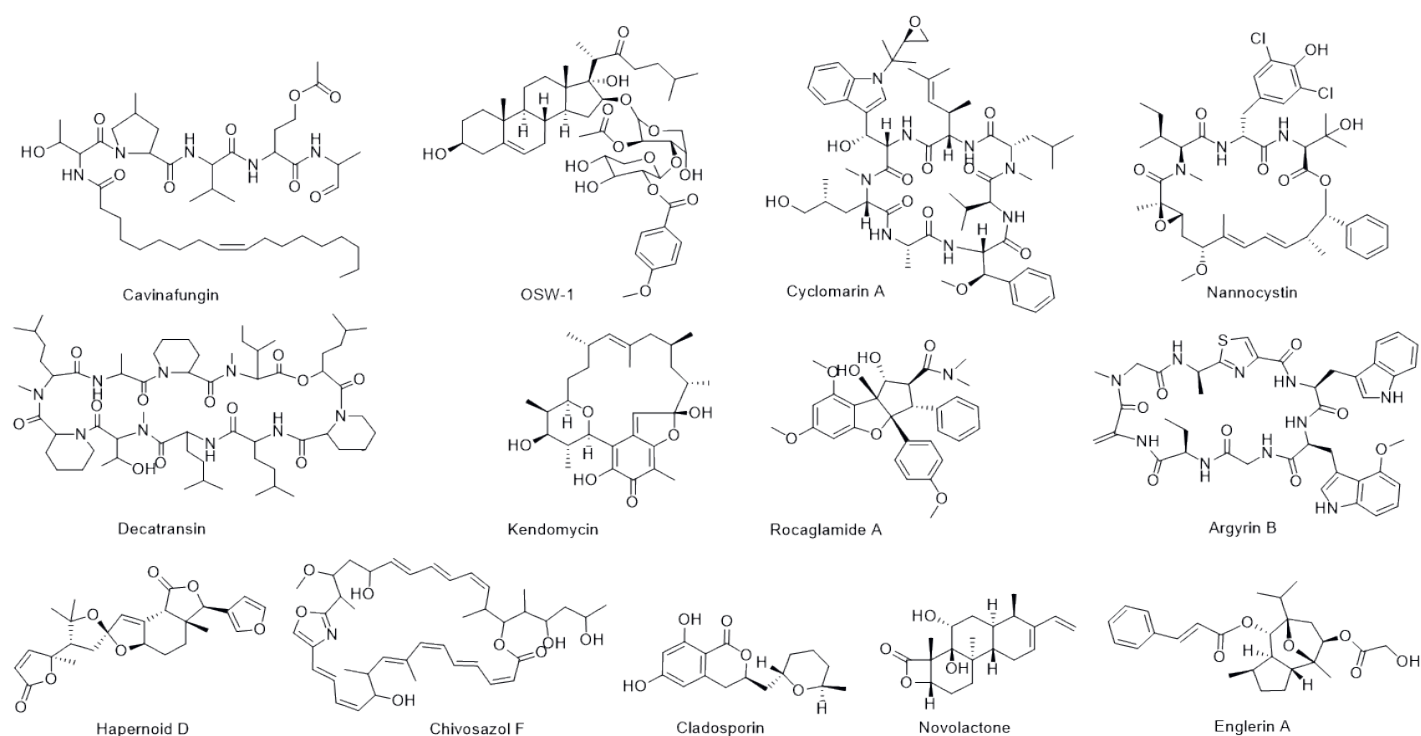

Fig. 6. NP structures of in-house target identification projects.

inhibitions;[87] novel biologically active matter as starting points for medicinal chemistry;[88] cellular uptake or tissue specific drug delivery.[89]

Strengthening the inclusion of modern NP research in these spotlights of pharmaceutical innovation is critical to exploit the full potential of these molecules in non-traditional directions. These areas of opportunity for academic and pharmaceutical research teams can be matched to the following scientific NP directions:

Novel NP-based Diversity[90a–f]

- Access to the 'dark matter' metabolome of new organism classes, microbial, symbiotic communities, and the eDNA to populate the uncharted NP-chemospace of biological relevance.
- Biodiversity-driven investigation of proven and underexploited microorganism groups combined with secondary metabolomics studies, HR-MS analytics, and database platforms as efficacious concepts for the description of new NPs.
- Emphasis on the discovery of new antibiotics with Gram-negative activity spectra by using genome/metagenome-driven concepts and by considering especially Gram-negative NP producers more closely. These molecules should have physico-chemical properties or biological characteristics which increase the likelihood of them being able to cross cell walls of both Gram-negative and infectious bacteria.

Computational and Systems Biology[91a–m]

- 'Big Data', using artificial intelligence analyses of TCM diagnoses and herbal treatments for proven efficacy in defined diseases to identify the active principles of therapeutic potential.
- Further development of appropriate CompBiol tools to improve genome mining, gene calling, comparative genomics, or structure prediction for the discovery of NP biosynthesis pathways in bacteria, fungi, or plants.
- Advancing predictive software systems which integrate transcriptome data with underlying metabolic fluxes to investigate time-dependent regulatory systems orchestrating the biotechnological production of microbial NPs to rationalize and accelerate titer optimizations.

Synthetic Biology[92a–m]

- Decipher the architectural plan of functional NRPS/ PKS to enable site-specific permutations of building blocks across macrocyclic NPs backbones for rational drug design.
- Optimize the success rate of NP cluster expression in a heterologous chassis suitable for process development on an industrial scale.
- Investigation and evolution of tailoring enzymes involved in the biosynthesis of chiral NPs and their specific precursors for building block synthesis in research and process chemistry.

Target Identification and Cheminformatics[93]

- Leverage structurally unique, specifically active secondary metabolites as shuttles to a novel, disease-relevant target space or to new modes of molecular interaction with known proteins.
- Identify modulators of therapeutically relevant targets by genome mining for co-localization signatures of mutated proteins as a resistance mechanism and of biosynthetic NPs pathways organized in hyperclusters with the potential to synergistically change biological responses in pro- and eukaryotes.
- Identify the physico-chemical or biological properties of macrocyclic NPs which convey cell penetration or tissue specificity.
- Design conjugates with cell-penetrating macrocyclic NPs and biochemically active/ high affinity synthetics/ binders either for cellular delivery of partner molecules lacking this feature, or for molecular glue concepts.
- Investigate NP substructures for fragment-based screening or construct new, sp³-enriched, chemical matter by fragment merging.

So far, NPs as drugs have contributed to many innovations in human therapy. Capitalizing on the explosion of new technological possibilities in cell biology, medicinal chemistry, molecular modeling, cheminformatics, molecular genetics, computational biology, AI (Artificial Intelligence) and 'Big Data' analysis, whole genome sequencing paired with DNA-synthesis, genetic design, enzyme evolution, chemical ecology, and metagenomics ensures that NPs will remain an innovative and inspiring scientific discipline in future pharmaceutical R&D.

[1] a: J. Gershenzon, N. Dudareva, *Nat. Chem. Biol.* **2007**, *3*, 408, https://doi.org/10.1038/nchembio.2007.5; b) J. M. Crawford, J. Clardy, *Chem. Commun. (Camb)* **2011**, *47*, 7559, https://doi.org/10.1039/c1cc11574j; c) C. Beemelmanns, H. Guo, M. Rischer, M. Poulsen, *Beilstein J. Org. Chem.* **2016**, *12*, 314, https://doi.org/10.3762/bjoc.12.34; d) M. Morita, E. W. Schmidt, *Nat. Prod. Rep.* **2018**, *35*, 357, https://doi.org/10.1039/c7np00053g; e) A. Deveau, G. Bonito, J. Uehling, M. Paoletti, M. Becker, S. Bindschedler, S. Hacquard, V. Herve, J. Labbe, O. A. Lastovetsky, S. Mieszkin, L. J. Millet, B. Vajna, P. Junier, P. Bonfante, B. P. Krom, S. Olsson, J. D. van Elsas, L. Y. Wick, *FEMS Microbiol. Rev.* **2018**, *42*, 335, https://doi.org/10.1093/femsre/fuy008; f) S. Khalid, N. P. Keller, *Environ. Microbiol.* **2021**, *23*, 1334, https://doi.org/10.1111/1462-2920.15410.

[2] a) P. Wang, J. Heitman, *Genome Biol.* **2005**, *6*, 226, https://doi.org/10.1186/gb-2005-6-7-226; b) C. A. Shertz, R. J. Bastidas, W. Li, J. Heitman, M. E. Cardenas, *BMC Genomics* **2010**, *11*, 510, https://doi.org/10.1186/1471-2164-11-510; c) J. S. Burg, P. J. Espenshade, *Prog. Lipid Res.* **2011**, *50*, 403, https://doi.org/10.1016/j.plipres.2011.07.002; d) S. Yasmin, L. Alcazar-Fuoli, M. Grundlinger, T. Puempel, T. Cairns, M. Blatzer, J. F. Lopez, J. O. Grimalt, E. Bignell, H. Haas, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E497, https://doi.org/10.1073/pnas.1106399108; e) Y. Li, H. Ren, Y. Zhao, J. Sun, Y. Fan, D. Jin, Y. Pei, *J. Invertebr. Pathol.* **2020**, *171*, 107334, https://doi.org/10.1016/j.jip.2020.107334.

[3] R. M. Stone, P. W. Manley, R. A. Larson, R. Capdeville, *Blood Adv.* **2018**, *2*, 444, https://doi.org/10.1182/bloodadvances.2017011080.

[4] a) R. Bade, H. F. Chan, J. Reynisson, *Eur. J. Med. Chem.* **2010**, *45*, 5646, https://doi.org/10.1016/j.ejmech.2010.09.018; b) D. J. Newman, G. M. Cragg, *J. Nat. Prod.* **2020**, *83*, 770, https://doi.org/10.1021/acs.jnatprod.9b01285; c) A. G. Atanasov, S. B. Zotchev, V. M. Dirsch, I. International Natural Product Sciences, C. T. Supuran, *Nat. Rev. Drug Discov.* **2021**, *20*, 200, https://doi.org/10.1038/s41573-020-00114-z.

[5] F. E. Koehn, G. T. Carter, *Nat. Rev. Drug Discov.* **2005**, *4*, 206, https://doi.org/10.1038/nrd1657.

[6] M. Sorokina, C. Steinbeck, *J. Cheminform.* **2020**, *12*, 20, https://doi.org/10.1186/s13321-020-00424-9.

[7] https://mona.fiehnlab.ucdavis.edu/.

[8] C. Guijas, J. R. Montenegro-Burke, X. Domingo-Almenara, A. Palermo, B. Warth, G. Hermann, G. Koellensperger, T. Huan, W. Uritboonthai, A. E. Aisporna, D. W. Wolan, M. E. Spilker, H. P. Benton, G. Siuzdak, *Anal. Chem.* **2018**, *90*, 3156, https://doi.org/10.1021/acs.analchem.7b04424.

[9]   J. Watrous, P. Roach, T. Alexandrov, B. S. Heath, J. Y. Yang, R. D. Kersten, M. van der Voort, K. Pogliano, H. Gross, J. M. Raaijmakers, B. S. Moore, J. Laskin, N. Bandeira, P. C. Dorrestein, *Proc. Natl. Acad. Sci. USA* **2012**, *109*, E1743, https://doi.org/10.1073/pnas.1203689109.

[10] M. Wang, J. J. Carver, V. V. Phelan, L. M. Sanchez, N. Garg, Y. Peng, D. D. Nguyen, J. Watrous, C. A. Kapono, T. Luzzatto-Knaan, C. Porto, A. Bouslimani, A. V. Melnik, M. J. Meehan, W. T. Liu, M. Crusemann, P. D. Boudreau, E. Esquenazi, M. Sandoval-Calderon, R. D. Kersten, L. A. Pace, R. A. Quinn, K. R. Duncan, C. C. Hsu, D. J. Floros, R. G. Gavilan, K. Kleigrewe, T. Northen, R. J. Dutton, D. Parrot, E. E. Carlson, B. Aigle, C. F. Michelsen, L. Jelsbak, C. Sohlenkamp, P. Pevzner, A. Edlund, J. McLean, J. Piel, B. T. Murphy, L. Gerwick, C. C. Liaw, Y. L. Yang, H. U. Humpf, M. Maansson, R. A. Keyzers, A. C. Sims, A. R. Johnson, A. M. Sidebottom, B. E. Sedio, A. Klitgaard, C. B. Larson, C. A. B. P, D. Torres-Mendoza, D. J. Gonzalez, D. B. Silva, L. M. Marques, D. P. Demarque, E. Pociute, E. C. O'Neill, E. Briand, E. J. N. Helfrich, E. A. Granatosky, E. Glukhov, F. Ryffel, H. Houson, H. Mohimani, J. J. Kharbush, Y. Zeng, J. A. Vorholt, K. L. Kurita, P. Charusanti, K. L. McPhail, K. F. Nielsen, L. Vuong, M. Elfeki, M. F. Traxler, N. Engene, N. Koyama, O. B. Vining, R. Baric, R. R. Silva, S. J. Mascuch, S. Tomasi, S. Jenkins, V. Macherla, T. Hoffman, V. Agarwal, P. G. Williams, J. Dai, R. Neupane, J. Gurr, A. M. C. Rodriguez, A. Lamsa, C. Zhang, K. Dorrestein, B. M. Duggan, J. Almaliti, P. M. Allard, P. Phapale, L. F. Nothias, T. Alexandrov, M. Litaudon, J. L. Wolfender, J. E. Kyle, T. O. Metz, T. Peryea, D. T. Nguyen, D. VanLeer, P. Shinn, A. Jadhav, R. Muller, K. M. Waters, W. Shi, X. Liu, L. Zhang, R. Knight, P. R. Jensen, B. O. Palsson, K. Pogliano, R. G. Linington, M. Gutierrez, N. P. Lopes, W. H. Gerwick, B. S. Moore, P. C. Dorrestein, N. Bandeira, *Nat. Biotechnol.* **2016**, *34*, 828, https://doi.org/10.1038/nbt.3597.

[11] A. E. Fox Ramos, L. Evanno, E. Poupon, P. Champy, M. A. Beniddir, *Nat. Prod. Rep.* **2019**, *36*, 960, https://doi.org/10.1039/c9np00006b.

[12] H. Mohimani, A. Gurevich, A. Shlemov, A. Mikheenko, A. Korobeynikov, L. Cao, E. Shcherbin, L. F. Nothias, P. C. Dorrestein, P. A. Pevzner, *Nat. Commun.* **2018**, *9*, 4035, https://doi.org/10.1038/s41467-018-06082-8.

[13] D. C. Burns, E. P. Mazzola, W. F. Reynolds, *Nat. Prod. Rep.* **2019**, *36*, 919, https://doi.org/10.1039/c9np00007k.

[14] M. A. Beniddir, K. B. Kang, G. Genta-Jouve, F. Huber, S. Rogers, J. J. J. van der Hooft, *Nat. Prod. Rep.* **2021**, https://doi.org/10.1039/D1NP00023C.

[15] P. Ertl, *J. Chem. Inf. Model.* **2014**, *54*, 1617, https://doi.org/10.1021/ci5001983.

[16] P. Ertl, *ChemRXiv* **2021**, 10.26434/chemrxiv.13525457.v2.

[17] A. Schuffenhauer, N. Schneider, S. Hintermann, D. Auld, J. Blank, S. Cotesta, C. Engeloch, N. Fechner, C. Gaul, J. Giovannoni, J. Jansen, J. Joslin, P. Krastel, E. Lounkine, J. Manchester, L. G. Monovich, A. P. Pelliccioli, M. Schwarze, M. D. Shultz, N. Stiefl, D. K. Baeschlin, *J. Med. Chem.* **2020**, *63*, 14425, https://doi.org/10.1021/acs.jmedchem.0c01332.

[18] H. Prescher, G. Koch, T. Schuhmann, P. Ertl, A. Bussenault, M. Glick, I. Dix, F. Petersen, D. E. Lizos, *Bioorg. Med. Chem.* **2017**, *25*, 921, https://doi.org/10.1016/j.bmc.2016.12.005.

[19] M. Sorokina, P. Merseburger, K. Rajan, M. A. Yirik, C. Steinbeck, *J. Cheminform.* **2021**, *13*, 2, https://doi.org/10.1186/s13321-020-00478-9.

[20] S. Wetzel, A. Schuffenhauer, S. Roggo, P. Ertl, H. Waldmann, *Chimia* **2007**, *61*, 355, https://doi.org/10.2533/chimia.2007.355.

[21] P. Ertl, T. Schuhmann, *Mol. Inform.* **2020**, *39*, e2000017, https://doi.org/10.1002/minf.202000017.

[22] a) P. Ertl, *J. Cheminform.* **2017**, *9*, 36, https://doi.org/10.1186/s13321-017-0225-z; b) P. Ertl, T. Schuhmann, *J Nat. Prod.* **2019**, *82*, 1258, https://doi.org/10.1021/acs.jnatprod.8b01022.

[23] a) P. Ertl, S. Roggo, A. Schuffenhauer, *J. Chem. Inf. Model.* **2008**, *48*, 68, https://doi.org/10.1021/ci700286x; b) K. V. Jayaseelan, P. Moreno, A. Truszkowski, P. Ertl, C. Steinbeck, *BMC Bioinform.* **2012**, *13*, 106, https://doi.org/10.1186/1471-2105-13-106.

[24] T. G. Davies, I. J. Tickle, *Top. Curr. Chem.* **2012**, *317*, 33, https://doi.org/10.1007/128_2011_179.

[25] W. Jahnke, D. A. Erlanson, I. J. P. de Esch, C. N. Johnson, P. N. Mortenson, Y. Ochi, T. Urushima, *J. Med. Chem.* **2020**, *63*, 15494, https://doi.org/10.1021/acs.jmedchem.0c01608.

[26] F. Lovering, J. Bikker, C. Humblet, *J. Med. Chem.* **2009**, *52*, 6752, https://doi.org/10.1021/jm901241e.

[27] E. A. Crane, K. Gademann, *Angew. Chem. Int. Ed.* **2016**, *55*, 3882, https://doi.org/10.1002/anie.201505863.

[28] B. Over, S. Wetzel, C. Grutter, Y. Nakai, S. Renner, D. Rauh, H. Waldmann, *Nat. Chem.* **2013**, *5*, 21, https://doi.org/10.1038/nchem.1506.

[29] H. Vu, L. Pedro, T. Mak, B. McCormick, J. Rowley, M. Liu, A. Di Capua, B. Williams-Noonan, N. B. Pham, R. Pouwer, B. Nguyen, K. T. Andrews, T. Skinner-Adams, J. Kim, W. G. J. Hol, R. Hui, G. J. Crowther, W. C. Van Voorhis, R. J. Quinn, *ACS Infect. Dis.* **2018**, *4*, 431, https://doi.org/10.1021/acsinfecdis.7b00197.

[30] J. Ceballos, M. Schwalfenberg, G. Karageorgis, E. S. Reckzeh, S. Sievers, C. Ostermann, A. Pahl, M. Sellstedt, J. Nowacki, M. A. Carnero Corrales, J. Wilke, L. Laraia, K. Tschapalda, M. Metz, D. A. Sehr, S. Brand, K.

Winklhofer, P. Janning, S. Ziegler, H. Waldmann, *Angew. Chem. Int. Ed.* **2019**, *58*, 17016, https://doi.org/10.1002/anie.201909518.

[31] a) N. J. Spillman, R. J. Allen, C. W. McNamara, B. K. Yeung, E. A. Winzeler, T. T. Diagana, K. Kirk, *Cell Host Microbe* **2013**, *13*, 227, https://doi.org/10.1016/j.chom.2012.12.006; b) N. J. White, S. Pukrittayakamee, A. P. Phyo, R. Rueangweerayut, F. Nosten, P. Jittamala, A. Jeeyapant, J. P. Jain, G. Lefevre, R. Li, B. Magnusson, T. T. Diagana, F. J. Leong, *N. Engl. J. Med.* **2014**, *371*, 403, https://doi.org/10.1056/NEJMoa1315860.

[32] B. K. Yeung, B. Zou, M. Rottmann, S. B. Lakshminarayana, S. H. Ang, S. Y. Leong, J. Tan, J. Wong, S. Keller-Maerki, C. Fischli, A. Goh, E. K. Schmitt, P. Krastel, E. Francotte, K. Kuhen, D. Plouffe, K. Henson, T. Wagner, E. A. Winzeler, F. Petersen, R. Brun, V. Dartois, T. T. Diagana, T. H. Keller, *J. Med. Chem.* **2010**, *53*, 5155, https://doi.org/10.1021/jm100410f.

[33] A. Vulpetti, U. Hommel, G. Landrum, R. Lewis, C. Dalvit, *J. Am. Chem. Soc.* **2009**, *131*, 12949, https://doi.org/10.1021/ja905207t.

[34] a) M. Gavrielatos, K. Kyriakidis, D. A. Spandidos, I. Michalopoulos, *Mol. Med. Rep.* **2021**, *23*, https://doi.org/10.3892/mmr.2021.11890; b) V. Murigneux, L. W. Roberts, B. M. Forde, M. D. Phan, N. T. K. Nhu, A. D. Irwin, P. N. A. Harris, D. L. Paterson, M. A. Schembri, D. M. Whiley, S. A. Beatson, *BMC Genomics* **2021**, *22*, 474, https://doi.org/10.1186/s12864-021-07767-z.

[35] a) R. D. Fleischmann, M. D. Adams, O. White, R. A. Clayton, E. F. Kirkness, A. R. Kerlavage, C. J. Bult, J. F. Tomb, B. A. Dougherty, J. M. Merrick, et al., *Science* **1995**, *269*, 496, https://doi.org/10.1126/science.7542800; b) C. M. Fraser, J. D. Gocayne, O. White, M. D. Adams, R. A. Clayton, R. D. Fleischmann, C. J. Bult, A. R. Kerlavage, G. Sutton, J. M. Kelley, R. D. Fritchman, J. F. Weidman, K. V. Small, M. Sandusky, J. Fuhrmann, D. Nguyen, T. R. Utterback, D. M. Saudek, C. A. Phillips, J. M. Merrick, J. F. Tomb, B. A. Dougherty, K. F. Bott, P. C. Hu, T. S. Lucier, S. N. Peterson, H. O. Smith, C. A. Hutchison, 3rd, J. C. Venter, *Science* **1995**, *270*, 397, https://doi.org/10.1126/science.270.5235.397.

[36] https://www.ncbi.nlm.nih.gov/genome/browse#!/ prokaryotes/.

[37] a) https://emea.illumina.com/content/dam/illumina-arketing/documents/ products/appnotes/novaseq-hiseq-q30-app-note-770-2017-010.pdf; b) https://www.illumina.com/content/dam/illumina-marketing/documents/ products/technotes/technote_understanding_quality_scores.pdf.

[38] a) D. K. Chaudhary, A. Khulan, J. Kim, *Sci. Rep.* **2019**, *9*, 6666, https://doi.org/10.1038/s41598-019-43182-x; b) S. R. Vartoukian, R. M. Palmer, W. G. Wade, *FEMS Microbiol. Lett.* **2010**, *309*, 1, https://doi.org/10.1111/j.1574-6968.2010.02000.x.

[39] P. D. Schloss, J. Handelsman, *Curr. Opin. Biotechnol.* **2003**, *14*, 303, https://doi.org/10.1016/s0958-1669(03)00067-3.

[40] S. Koren, B. P. Walenz, K. Berlin, J. R. Miller, N. H. Bergman, A. M. Phillippy, *Genome Res.* **2017**, *27*, 722, https://doi.org/10.1101/gr.215087.116.

[41] H. Cheng, G. T. Concepcion, X. Feng, H. Zhang, H. Li, *Nat. Meth.* **2021**, *18*, 170, https://doi.org/10.1038/s41592-020-01056-5.

[42] M. H. Medema, K. Blin, P. Cimermancic, V. de Jager, P. Zakrzewski, M. A. Fischbach, T. Weber, E. Takano, R. Breitling, *Nucleic Acids Res.* **2011**, *39*, W339, https://doi.org/10.1093/nar/gkr466.

[43] S. A. Kautsar, K. Blin, S. Shaw, J. C. Navarro-Munoz, B. R. Terlouw, J. J. J. van der Hooft, J. A. van Santen, V. Tracanna, H. G. Suarez Duran, V. Pascal Andreu, N. Selem-Mojica, M. Alanjary, S. L. Robinson, G. Lund, S. C. Epstein, A. C. Sisto, L. K. Charkoudian, J. Collemare, R. G. Linington, T. Weber, M. H. Medema, *Nucleic Acids Res.* **2020**, *48*, D454, https://doi.org/10.1093/nar/gkz882.

[44] P. Mrak, P. Krastel, P. Pivk Lukancic, J. Tao, D. Pistorius, C. M. Moore, *J. Biol. Chem.* **2018**, *293*, 19982, https://doi.org/10.1074/jbc.RA118.005314.

[45] a) M. Alanjary, M. H. Medema, *J. Biol. Chem.* **2018**, *293*, 19996, https://doi.org/10.1074/jbc.H118.006669; b) Y. Yan, N. Liu, Y. Tang, *Nat. Prod. Rep.* **2020**, *37*, 879, https://doi.org/10.1039/c9np00050j.

[46] F. Alberti, C. Corre, *Nat. Prod. Rep.* **2019**, *36*, 1237, https://doi.org/10.1039/c8np00081f.

[47] R. H. Baltz, *J. Ind. Microbiol. Biotechnol.* **2019**, *46*, 281, https://doi.org/10.1007/s10295-018-2115-4.

[48] K. A. J. Bozhuyuk, F. Fleischhacker, A. Linck, F. Wesche, A. Tietze, C. P. Niesert, H. B. Bode, *Nat. Chem.* **2018**, *10*, 275, https://doi.org/10.1038/nchem.2890.

[49] D. C. Swinney, J. Anthony, *Nat. Rev. Drug Discov.* **2011**, *10*, 507, https://doi.org/10.1038/nrd3480.

[50] M. Salcius, A. J. Bauer, Q. Hao, S. Li, A. Tutter, J. Raphael, W. Jahnke, J. M. Rondeau, E. Bourgier, J. Tallarico, G. A. Michaud, *J. Biomol. Screen.* **2014**, *19*, 917, https://doi.org/10.1177/1087057114522691.

[51] a) Y. Lee, E. K. Lee, Y. W. Cho, T. Matsui, I. C. Kang, T. S. Kim, M. H. Han, *Proteomics* **2003**, *3*, 2289, https://doi.org/10.1002/pmic.200300541; b) B. B. Haab, *Curr. Opin. Drug. Discov. Devel.* **2001**, *4*, 116; c) E. C. Hett, R. E. Kyne, Jr., A. Gopalsamy, M. A. Tones, H. Xu, G. L. Thio, E. Nolan, L. H. Jones, *ACS Comb. Sci.* **2016**, *18*, 611, https://doi.org/10.1021/acscombsci.6b00089.

[52] a) D. Martinez Molina, R. Jafari, M. Ignatushchenko, T. Seki, E. A. Larsson, C. Dan, L. Sreekumar, Y. Cao, P. Nordlund, *Science* **2013**, *341*, 84,

https://doi.org/10.1126/science.1233606; b) R. Jafari, H. Almqvist, H. Axelsson, M. Ignatushchenko, T. Lundback, P. Nordlund, D. Martinez Molina, *Nat. Protoc.* 2014, 9, 2100, https://doi.org/10.1038/nprot.2014.138.

[53] B. Lomenick, R. Hao, N. Jonai, R. M. Chin, M. Aghajan, S. Warburton, J. Wang, R. P. Wu, F. Gomez, J. A. Loo, J. A. Wohlschlegel, T. M. Vondriska, J. Pelletier, H. R. Herschman, J. Clardy, C. F. Clarke, J. Huang, *Proc. Natl. Acad. Sci. USA* 2009, 106, 21984, https://doi.org/10.1073/pnas.0910040106.

[54] M. Pepelnjak, N. de Souza, P. Picotti, *Trends Biochem. Sci.* 2020, 45, 919, https://doi.org/10.1016/j.tibs.2020.05.006.

[55] M. Schirle, M. Bantscheff, B. Kuster, *Chem. Biol.* 2012, 19, 72, https://doi.org/10.1016/j.chembiol.2012.01.002.

[56] X. Chen, Y. Wang, N. Ma, J. Tian, Y. Shao, B. Zhu, Y. K. Wong, Z. Liang, C. Zou, J. Wang, *Signal Transduct. Target Ther.* 2020, 5, 72, https://doi.org/10.1038/s41392-020-0186-y.

[57] a) E. K. Schmitt, M. Riwanto, V. Sambandamurthy, S. Roggo, C. Miault, C. Zwingelstein, P. Krastel, C. Noble, D. Beer, S. P. Rao, M. Au, P. Niyomrattanakit, V. Lim, J. Zheng, D. Jeffery, K. Pethe, L. R. Camacho, *Angew. Chem. Int. Ed.* 2011, 50, 5889, https://doi.org/10.1002/anie.201101740; b) N. Burstner, S. Roggo, N. Ostermann, J. Blank, C. Delmas, F. Freuler, B. Gerhartz, A. Hinniger, D. Hoepfner, B. Liechty, M. Mihalic, J. Murphy, D. Pistorius, M. Rottmann, J. R. Thomas, M. Schirle, E. K. Schmitt, *ChemBioChem* 2015, 16, 2433, https://doi.org/10.1002/cbic.201500472.

[58] P. Krastel, S. Roggo, M. Schirle, N. T. Ross, F. Perruccio, P. Aspesi, Jr., T. Aust, K. Buntin, D. Estoppey, B. Liechty, F. Mapa, K. Memmert, H. Miller, X. Pan, R. Riedl, C. Thibaut, J. Thomas, T. Wagner, E. Weber, X. Xie, E. K. Schmitt, D. Hoepfner, *Angew. Chem. Int. Ed.* 2015, 54, 10149, https://doi.org/10.1002/anie.201505069.

[59] A. W. Burgett, T. B. Poulsen, K. Wangkanont, D. R. Anderson, C. Kikuchi, K. Shimada, S. Okubo, K. C. Fortner, Y. Mimaki, M. Kuroda, J. P. Murphy, D. J. Schwalb, E. C. Petrella, I. Cornella-Taracido, M. Schirle, J. A. Tallarico, M. D. Shair, *Nat. Chem. Biol.* 2011, 7, 639, https://doi.org/10.1038/nchembio.625.

[60] H. Cho, Q. Shen, L. H. Zhang, M. Okumura, A. Kawakami, J. Ambrose, F. Sigoillot, H. R. Miller, S. Gleim, A. Cobos-Correa, Y. Wang, P. Piechon, G. Roma, F. Eggimann, C. Moore, P. Aspesi, Jr., F. A. Mapa, H. Burks, N. T. Ross, P. Krastel, M. Hild, T. J. Maimone, D. E. Fisher, D. K. Nomura, J. A. Tallarico, S. M. Canham, J. L. Jenkins, W. C. Forrester, *Cell Chem. Biol.* 2021, https://doi.org/10.1016/j.chembiol.2021.03.004.

[61] a) C. Carson, P. Raman, J. Tullai, L. Xu, M. Henault, E. Thomas, S. Yeola, J. Lao, M. McPate, J. M. Verkuyl, G. Marsh, J. Sarber, A. Amaral, S. Bailey, D. Lubicka, H. Pham, N. Miranda, J. Ding, H. M. Tang, H. Ju, P. Tranter, N. Ji, P. Krastel, R. K. Jain, A. M. Schumacher, J. J. Loureiro, E. George, G. Berellini, N. T. Ross, S. M. Bushell, G. Erdemli, J. M. Solomon, *PLoS one* 2015, 10, e0127498, https://doi.org/10.1371/journal.pone.0127498; b) J. Barretina, G. Caponigro, N. Stransky, K. Venkatesan, A. A. Margolin, S. Kim, C. J. Wilson, J. Lehar, G. V. Kryukov, D. Sonkin, A. Reddy, M. Liu, L. Murray, M. F. Berger, J. E. Monahan, P. Morais, J. Meltzer, A. Korejwa, J. Jane-Valbuena, F. A. Mapa, J. Thibault, E. Bric-Furlong, P. Raman, A. Shipway, I. H. Engels, J. Cheng, G. K. Yu, J. Yu, P. Aspesi, Jr., M. de Silva, K. Jagtap, M. D. Jones, L. Wang, C. Hatton, E. Palescandolo, S. Gupta, S. Mahan, C. Sougnez, R. C. Onofrio, T. Liefeld, L. MacConaill, W. Winckler, M. Reich, N. Li, J. P. Mesirov, S. B. Gabriel, G. Getz, K. Ardlie, V. Chan, V. E. Myer, B. L. Weber, J. Porter, M. Warmuth, P. Finan, J. L. Harris, M. Meyerson, T. R. Golub, M. P. Morrissey, W. R. Sellers, R. Schlegel, L. A. Garraway, *Nature* 2012, 483, 603, https://doi.org/10.1038/nature11003.

[62] G. Giaever, P. Flaherty, J. Kumm, M. Proctor, C. Nislow, D. F. Jaramillo, A. M. Chu, M. I. Jordan, A. P. Arkin, R. W. Davis, *Proc. Natl. Acad. Sci. USA* 2004, 101, 793, https://doi.org/10.1073/pnas.0307490100.

[63] a) A. Y. Lee, R. P. St Onge, M. J. Proctor, I. M. Wallace, A. H. Nile, P. A. Spagnuolo, Y. Jitkova, M. Gronda, Y. Wu, M. K. Kim, A. Cheung-Ong, N. P. Torres, E. D. Spear, M. K. Han, U. Schlecht, S. Suresh, G. Duby, L. E. Heisler, A. Surendra, E. Fung, M. L. Urbanus, M. Gebbia, E. Lissina, M. Miranda, J. H. Chiang, A. M. Aparicio, M. Zeghouf, R. W. Davis, J. Cherfils, M. Boutry, C. A. Kaiser, C. L. Cummins, W. S. Trimble, G. W. Brown, A. D. Schimmer, V. A. Bankaitis, C. Nislow, G. D. Bader, G. Giaever, *Science* 2014, 344, 208, https://doi.org/10.1126/science.1250217; b) I. Filipuzzi, S. Cotesta, F. Perruccio, B. Knapp, Y. Fu, C. Studer, V. Pries, R. Riedl, S. B. Helliwell, K. T. Petrovic, N. R. Movva, D. Sanglard, J. Tao, D. Hoepfner, *PLoS Genet.* 2016, 12, e1006374, https://doi.org/10.1371/journal.pgen.1006374.

[64] D. Hoepfner, C. W. McNamara, C. S. Lim, C. Studer, R. Riedl, T. Aust, S. L. McCormack, D. M. Plouffe, S. Meister, S. Schuierer, U. Plikat, N. Hartmann, F. Staedtler, S. Cotesta, E. K. Schmitt, F. Petersen, F. Supek, R. J. Glynne, J. A. Tallarico, J. A. Porter, M. C. Fishman, C. Bodenreider, T. T. Diagana, N. R. Movva, E. A. Winzeler, *Cell Host Microbe* 2012, 11, 654, https://doi.org/10.1016/j.chom.2012.04.015.

[65] a) X. Wang, D. E. Wedge, S. J. Cutler, *Nat. Prod. Commun.* 2016, 11, 1595; b) B. Baragana, B. Forte, R. Choi, S. Nakazawa Hewitt, J. A. Bueren-Calabuig, J. P. Pisco, C. Peet, D. M. Dranow, D. A. Robinson, C. Jansen, N. R. Norcross, S. Vinayak, M. Anderson, C. F. Brooks, C. A. Cooper, S. Damerow, M. Delves, K. Dowers, J. Duffy, T. E. Edwards, I. Hallyburton,

B. G. Horst, M. A. Hulverson, L. Ferguson, M. B. Jimenez-Diaz, R. S. Jumani, D. D. Lorimer, M. S. Love, S. Maher, H. Matthews, C. W. McNamara, P. Miller, S. O'Neill, K. K. Ojo, M. Osuna-Cabello, E. Pinto, J. Post, J. Riley, M. Rottmann, L. M. Sanz, P. Scullion, A. Sharma, S. M. Shepherd, Y. Shishikura, F. R. C. Simeons, E. E. Stebbins, L. Stojanovski, U. Straschil, F. K. Tamaki, J. Tamjar, L. S. Torrie, A. Vantaux, B. Witkowski, S. Wittlin, M. Yogavel, F. Zuccotto, I. Angulo-Barturen, R. Sinden, J. Baum, F. J. Gamo, P. Maser, D. E. Kyle, E. A. Winzeler, P. J. Myler, P. G. Wyatt, D. Floyd, D. Matthews, A. Sharma, B. Striepen, C. D. Huston, D. W. Gray, A. H. Fairlamb, A. V. Pisliakov, C. Walpole, K. D. Read, W. C. Van Voorhis, I. H. Gilbert, *Proc. Natl. Acad. Sci. USA* 2019, 116, 7015, https://doi.org/10.1073/pnas.1814685116.

[66] A. Q. Hassan, C. A. Kirby, W. Zhou, T. Schuhmann, R. Kityk, D. R. Kipp, J. Baird, J. Chen, Y. Chen, F. Chung, D. Hoepfner, N. R. Movva, R. Pagliarini, F. Petersen, C. Quinn, D. Quinn, R. Riedl, E. K. Schmitt, A. Schitter, T. Stams, C. Studer, P. D. Fortin, M. P. Mayer, H. Sadlish, *Chem. Biol.* 2015, 22, 87, https://doi.org/10.1016/j.chembiol.2014.11.007.

[67] T. Junne, J. Wong, C. Studer, T. Aust, B. W. Bauer, M. Beibel, B. Bhullar, R. Bruccoleri, J. Eichenberger, D. Estoppey, N. Hartmann, B. Knapp, P. Krastel, N. Melin, E. J. Oakeley, L. Oberer, R. Riedl, G. Roma, S. Schuierer, F. Petersen, J. A. Tallarico, T. A. Rapoport, M. Spiess, D. Hoepfner, *J. Cell Sci.* 2015, 128, 1217, https://doi.org/10.1242/jcs.165746.

[68] H. Sadlish, G. Galicia-Vazquez, C. G. Paris, T. Aust, B. Bhullar, L. Chang, S. B. Helliwell, D. Hoepfner, B. Knapp, R. Riedl, S. Roggo, S. Schuierer, C. Studer, J. A. Porco, Jr., J. Pelletier, N. R. Movva, *ACS Chem. Biol.* 2013, 8, 1519, https://doi.org/10.1021/cb400158t.

[69] S. B. Helliwell, S. Karkare, M. Bergdoll, A. Rahier, J. R. Leighton-Davis, C. Fioretto, T. Aust, I. Filipuzzi, M. Frederiksen, J. Gounarides, D. Hoepfner, A. Hofmann, P. E. Imbert, R. Jeker, R. Knochenmuss, P. Krastel, A. Margerit, K. Memmert, C. V. Miault, N. Rao Movva, A. Muller, H. U. Naegeli, L. Oberer, V. Prindle, R. Riedl, S. Schuierer, J. A. Sexton, J. Tao, T. Wagner, H. Yin, J. Zhang, S. Roggo, S. Reinker, C. N. Parker, *Nat. Commun.* 2015, 6, 8613, https://doi.org/10.1038/ncomms9613.

[70] J. Heitman, N. R. Movva, M. N. Hall, *Science* 1991, 253, 905, https://doi.org/10.1126/science.1715094

[71] a) B. Haendler, R. Keller, P. C. Hiestand, H. P. Kocher, G. Wegmann, N. R. Movva, *Gene* 1989, 83, 39, https://doi.org/10.1016/0378-1119(89)90401-0; b) R. E. Handschumacher, M. W. Harding, J. Rice, R. J. Drugge, D. W. Speicher, *Science* 1984, 226, 544, https://doi.org/10.1126/science.6238408; c) P. L. Koser, D. Sylvester, G. P. Livi, D. J. Bergsma, *Nucleic Acids Res.* 1990, 18, 1643, https://doi.org/10.1093/nar/18.6.1643.

[72] a) K. Watashi, N. Ishii, M. Hijikata, D. Inoue, T. Murata, Y. Miyanari, K. Shimotohno, *Mol. Cell* 2005, 19, 111, https://doi.org/10.1016/j.molcel.2005.05.014; b) A. Molyvdas, S. Matalon, *Eur. Respir. J.* 2020, 56, https://doi.org/10.1183/13993003.02484-2020.

[73] B. Nyfeler, D. Hoepfner, D. Palestrant, C. A. Kirby, L. Whitehead, R. Yu, G. Deng, R. E. Caughlan, A. L. Woods, A. K. Jones, S. W. Barnes, J. R. Walker, S. Gaulis, E. Hauy, S. M. Brachmann, P. Krastel, C. Studer, R. Riedl, D. Estoppey, T. Aust, N. R. Movva, Z. Wang, M. Salcius, G. A. Michaud, G. McAllister, L. O. Murphy, J. A. Tallarico, C. J. Wilson, C. R. Dean, *PLoS one* 2012, 7, e42657, https://doi.org/10.1371/journal.pone.0042657.

[74] a) P. Fang, H. Han, J. Wang, K. Chen, X. Chen, M. Guo, *Chem. Biol.* 2015, 22, 734, https://doi.org/10.1016/j.chembiol.2015.05.007; b) S. Iwasaki, W. Iwasaki, M. Takahashi, A. Sakamoto, C. Watanabe, Y. Shichino, S. N. Floor, K. Fujiwara, M. Mito, K. Dodo, M. Sodeoka, H. Imataka, T. Honma, K. Fukuzawa, T. Ito, N. T. Ingolia, *Mol. Cell* 2019, 73, 738, https://doi.org/10.1016/j.molcel.2018.11.026.

[75] a) H. Turner, *Future Med. Chem.* 2016, 8, 227, https://doi.org/10.4155/fmc.15.177; b) M. Rottmann, C. McNamara, B. K. Yeung, M. C. Lee, B. Zou, B. Russell, P. Seitz, D. M. Plouffe, N. V. Dharia, J. Tan, S. B. Cohen, K. R. Spencer, G. E. Gonzalez-Paez, S. B. Lakshminarayana, A. Goh, R. Suwanarusk, T. Jegla, E. K. Schmitt, H. P. Beck, R. Brun, F. Nosten, L. Renia, V. Dartois, T. H. Keller, D. A. Fidock, E. A. Winzeler, T. T. Diagana, *Science* 2010, 329, 1175, https://doi.org/10.1126/science.1193225.

[76] a) M. Leeb, A. Wutz, *Nature* 2011, 479, 131, https://doi.org/10.1038/nature10448; b) U. Elling, J. Taubenschmid, G. Wirnsberger, R. O'Malley, S. P. Demers, Q. Vanhaelen, A. I. Shukalyuk, G. Schmauss, D. Schramek, F. Schnuetgen, H. von Melchner, J. R. Ecker, W. L. Stanford, J. Zuber, A. Stark, J. M. Penninger, *Cell Stem Cell* 2011, 9, 563, https://doi.org/10.1016/j.stem.2011.10.012.

[77] a) J. E. Carette, C. P. Guimaraes, M. Varadarajan, A. S. Park, I. Wuethrich, A. Godarova, M. Kotecki, B. H. Cochran, E. Spooner, H. L. Ploegh, T. R. Brummelkamp, *Science* 2009, 326, 1231, https://doi.org/10.1126/science.1178955; b) N. M. Gerhards, V. A. Blomen, M. Mutlu, J. Nieuwenhuis, D. Howald, C. Guyader, J. Jonkers, T. R. Brummelkamp, S. Rottenberg, *Mol. Oncol.* 2018, 12, 953, https://doi.org/10.1002/1878-0261.12307.

[78] a) M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, E. Charpentier, *Science* 2012, 337, 816,

https://doi.org/10.1126/science.1225829; b) O. Shalem, N. E. Sanjana, E. Hartenian, X. Shi, D. A. Scott, T. S. Mikkelsen, D. Heckl, B. L. Ebert, D. E. Root, J. G. Doench, F. Zhang, *Science* **2014**, *343*, 84, https://doi.org/10.1126/science.1247005; c) T. Wang, J. J. Wei, D. M. Sabatini, E. S. Lander, *Science* **2014**, *343*, 80, https://doi.org/10.1126/science.1246981.

[79] S. Kittanakom, A. Arnoldo, K. R. Brown, I. Wallace, T. Kunavisarut, D. Torti, L. E. Heisler, A. Surendra, J. Moffat, G. Giaever, C. Nislow, *G3 (Bethesda)* **2013**, *3*, 1375, https://doi.org/10.1534/g3.113.006437.

[80] D. Estoppey, J. W. Hewett, C. T. Guy, E. Harrington, J. R. Thomas, M. Schirle, R. Cuttat, A. Waldt, B. Gerrits, Z. Yang, S. Schuierer, X. Pan, K. Xie, W. Carbone, J. Knehr, A. Lindeman, C. Russ, E. Frias, G. R. Hoffman, M. Varadarajan, N. Ramadan, J. S. Reece-Hoyes, Q. Wang, X. Chen, G. McAllister, G. Roma, T. Bouwmeester, D. Hoepfner, *Sci. Rep.* **2017**, *7*, 42728, https://doi.org/10.1038/srep42728.

[81] D. Estoppey, C. M. Lee, M. Janoschke, B. H. Lee, K. F. Wan, H. Dong, P. Mathys, I. Filipuzzi, T. Schuhmann, R. Riedl, T. Aust, O. Galuba, G. McAllister, C. Russ, M. Spiess, T. Bouwmeester, G. M. C. Bonamy, D. Hoepfner, *Cell Rep.* **2017**, *19*, 451, https://doi.org/10.1016/j.celrep.2017.03.071.

[82] I. Filipuzzi, J. R. Thomas, V. Pries, D. Estoppey, M. Salcius, C. Studer, M. Schirle, D. Hoepfner, *ACS Chem. Biol.* **2017**, *12*, 2264, https://doi.org/10.1021/acschembio.7b00385.

[83] D. Tranter, I. Filipuzzi, T. Lochmann, B. Knapp, J. Kellosalo, D. Estoppey, D. Pistorius, A. Meissner, V. O. Paavilainen, D. Hoepfner, *J. Nat. Prod.* **2020**, *83*, 965, https://doi.org/10.1021/acs.jnatprod.9b00826.

[84] S. A. Wacker, B. R. Houghtaling, O. Elemento, T. M. Kapoor, *Nat. Chem. Biol.* **2012**, *8*, 235, https://doi.org/10.1038/nchembio.779.

[85] A. V. Anzalone, L. W. Koblan, D. R. Liu, *Nat. Biotechnol.* **2020**, *38*, 824, https://doi.org/10.1038/s41587-020-0561-9.

[86] C. H. Emmerich, L. M. Gamboa, M. C. J. Hofmann, M. Bonin-Andresen, O. Arbach, P. Schendel, B. Gerlach, K. Hempel, A. Bespalov, U. Dirnagl, M. J. Parnham, *Nat. Rev. Drug Discov.* **2021**, *20*, 64, https://doi.org/10.1038/s41573-020-0087-3.

[87] M. J. Blanco, K. M. Gardinier, *ACS Med. Chem. Lett.* **2020**, *11*, 228, https://doi.org/10.1021/acsmedchemlett.9b00582.

[88] J. J. Vanden Eynde, A. A. Mangoni, J. Rautio, J. Leprince, Y. T. Azuma, A. T. Garcia-Sosa, C. Hulme, J. Jampilek, R. Karaman, W. Li, P. A. C. Gomes, D. Hadjipavlou-Litina, R. Capasso, A. Geronikaki, L. Cerchia, J. M. Sabatier, R. Ragno, T. Tuccinardi, A. Trabocchi, J. Y. Winum, F. J. Luque, K. Prokai-Tatrai, M. Spetea, M. Gutschow, I. Kosalec, C. Guillou, M. H. Vasconcelos, G. Kokotos, G. Rastelli, M. E. de Sousa, C. Manera, S. Gemma, S. Mangani, C. Siciliano, S. Galdiero, H. Liu, P. J. H. Scott, C. de Los Rios, L. A. Agrofoglio, S. Collina, R. C. Guedes, D. Munoz-Torrero, *Molecules* **2019**, *25*, https://doi.org/10.3390/molecules25010119.

[89] Z. Zhao, A. Ukidve, J. Kim, S. Mitragotri, *Cell* **2020**, *181*, 151, https://doi.org/10.1016/j.cell.2020.02.001.

[90] a) K. Alam, M. N. Abbasi, J. Hao, Y. Zhang, A. Li, *Molecules* **2021**, *26*, https://doi.org/10.3390/molecules26102977; b) S. Wiegand, M. Jogler, C. Boedeker, D. Pinto, J. Vollmers, E. Rivas-Marin, T. Kohn, S. H. Peeters, A. Heuer, P. Rast, S. Oberbeckmann, B. Bunk, O. Jeske, A. Meyerdierks, J. E. Storesund, N. Kallscheuer, S. Lucker, O. M. Lage, T. Pohl, B. J. Merkel, P. Hornburger, R. W. Muller, F. Brummer, M. Labrenz, A. M. Spormann, H. J. M. Op den Camp, J. Overmann, R. Amann, M. S. M. Jetten, T. Mascher, M. H. Medema, D. P. Devos, A. K. Kaster, L. Ovreas, M. Rohde, M. Y. Galperin, C. Jogler, *Nat. Microbiol.* **2020**, *5*, 126, https://doi.org/10.1038/s41564-019-0588-1; c) M. Rust, E. J. N. Helfrich, M. F. Freeman, P. Nanudorn, C. M. Field, C. Ruckert, T. Kundig, M. J. Page, V. L. Webb, J. Kalinowski, S. Sunagawa, J. Piel, *Proc. Natl. Acad. Sci. USA* **2020**, *117*, 9508, https://doi.org/10.1073/pnas.1919245117; d) B. M. Hover, S. H. Kim, M. Katz, Z. Charlop-Powers, J. G. Owen, M. A. Ternei, J. Maniko, A. B. Estrela, H. Molina, S. Park, D. S. Perlin, S. F. Brady, *Nat. Microbiol.* **2018**, *3*, 415, https://doi.org/10.1038/s41564-018-0110-1; e) V. Libis, N. Antonovsky, M. Zhang, Z. Shang, D. Montiel, J. Maniko, M. A. Ternei, P. Y. Calle, C. Lemetre, J. G. Owen, S. F. Brady, *Nat. Commun.* **2019**, *10*, 3848, https://doi.org/10.1038/s41467-019-11658-z; f) J. Masschelein, M. Jenner, G. L. Challis, *Nat. Prod. Rep.* **2017**, *34*, 712, https://doi.org/10.1039/c7np00010c.

[91] a) B. Ming T. Shuo, M. Mingsan, J. Jiaojiao X. Weiyun, *Int. J. Serv. Technol. Manag.* **2015**, *21*, 294, https://doi.org/10.1504/IJSTM.2015.07394; b) B. Liu, *Front. Med.* **2014**, *8*, 272, https://doi.org/10.1007/s11684-014-0364-9; c) Y. Tang, Z. Li, D. Yang, Y. Fang, S. Gao, S. Liang, T. Liu, *Chin. Med.* **2021**, *16*, 2, https://doi.org/10.1186/s13020-020-00409-8; d) Y. Guo, X. Ren, Y. X. Chen, T. J. Wang, *Chin. J. Integr. Med.* **2019**, *25*, 648, https://doi.org/10.1007/s11655-019-3169-5; e) M. Yang, J. L. Chen, L. W. Xu, G. Ji, *Evid. Based Complement Alternat. Med.* **2013**, *2013*, 731969, https://doi.org/10.1155/2013/731969; f) B. Behsaz, E. Bode, A. Gurevich, Y. N. Shi, F. Grundmann, D. Acharya, A. M. Caraballo-Rodriguez, A. Bouslimani, M. Panitchpakdi, A. Linck, C. Guan, J. Oh, P. C. Dorrestein, H. B. Bode, P. A. Pevzner, H. Mohimani, *Nat. Commun.* **2021**, *12*, 3225, https://doi.org/10.1038/s41467-021-23502-4; g) H. Ren, C. Shi, H. Zhao, *iScience* **2020**, *23*, 100795, https://doi.org/10.1016/j.isci.2019.100795; h) G. A. Vignolle, R. L. Mach, A. R. Mach-Aigner, C. Derntl, *BMC Genomics* **2020**, *21*, 258, https://doi.org/10.1186/s12864-020-6653-6; i) M. A. Skinnider, C. W. Johnston, M. Gunabalasingam, N. J. Merwin, A. M. Kieliszek, R. J. MacLellan, H. Li, M. R. M. Ranieri, A. L. H. Webster, M. P. T. Cao, A. Pfeifle, N. Spencer, Q. H. To, D. P. Wallace, C. A. Dejong, N. A. Magarvey, *Nat. Commun.* **2020**, *11*, 6058, https://doi.org/10.1038/s41467-020-19986-1; j) K. Scherlach, C. Hertweck, *Nat. Commun.* **2021**, *12*, 3864, https://doi.org/10.1038/s41467-021-24133-5; k) M. T. Robey, L. K. Caesar, M. T. Drott, N. P. Keller, N. L. Kelleher, *Proc. Natl. Acad. Sci. USA* **2021**, *118*, https://doi.org/10.1073/pnas.2020230118; l) F. Alberti, S. Kaleem, J. A. Weaver, *Biol. Open.* **2020**, *9*, https://doi.org/10.1242/bio.056010; m) H. U. Kim, P. Charusanti, S. Y. Lee, T. Weber, *Nat. Prod. Rep.* **2016**, *33*, 933, https://doi.org/10.1039/c6np00019c.

[92] a) K. A. J. Bozhueyueuk, J. Watzel, N. Abbood, H. B. Bode, *Angew. Chem. Int. Ed.* **2021**, https://doi.org/10.1002/anie.202102859; b) E. Kalkreuter, K. S. Bingham, A. M. Keeler, A. N. Lowell, J. J. Schmidt, D. H. Sherman, G. J. Williams, *Nat. Commun.* **2021**, *12*, 2193, https://doi.org/10.1038/s41467-021-22497-2; c) V. T. T. Pham, C. T. Nguyen, D. Dhakal, H. T. Nguyen, T.-S. Kim, J. K. Sohng, *Appl. Sci.* **2021**, *11*, 1851; d) D. Dhakal, M. Chen, H. Luesch, Y. Ding, *J. Ind. Microbiol. Biotechnol.* **2021**, *48*, https://doi.org/10.1093/jimb/kuab003; e) T. Nogawa, A. Terai, K. Amagai, J. Hashimoto, Y. Futamura, A. Okano, M. Fujie, N. Satoh, H. Ikeda, K. Shin-Ya, H. Osada, S. Takahashi, *J. Nat. Prod.* **2020**, *83*, 3598, https://doi.org/10.1021/acs.jnatprod.0c00755; f) J. Ke, Y. Yoshikuni, *Curr. Opin. Biotechnol.* **2020**, *62*, 88, https://doi.org/10.1016/j.copbio.2019.09.005; g) J. J. Zhang, X. Tang, B. S. Moore, *Nat. Prod. Rep.* **2019**, *36*, 1313, https://doi.org/10.1039/c9np00025a; h) M. M. Zhang, Y. Wang, E. L. Ang, H. Zhao, *Nat. Prod. Rep.* **2016**, *33*, 963, https://doi.org/10.1039/c6np00017g; i) S. L. Robinson, J. Piel, S. Sunagawa, *Nat. Prod. Rep.* **2021**, https://doi.org/10.1039/D1NP00006C; j) T. A. Scott, J. Piel, *Nat. Rev. Chem.* **2019**, *3*, 404, https://doi.org/10.1038/s41570-019-0107-1; k) Y. Tsunematsu, N. Maeda, M. Sato, K. Hara, H. Hashimoto, K. Watanabe, C. Hertweck, *J. Am. Chem. Soc.* **2021**, *143*, 206, https://doi.org/10.1021/jacs.0c08879; l) S. Galanie, D. Entwistle, J. Lalonde, *Nat. Prod. Rep.* **2020**, *37*, 1122, https://doi.org/10.1039/c9np00071b; m) H. Kries, S. E. O'Connor, *Curr. Opin. Chem. Biol.* **2016**, *31*, 22, https://doi.org/10.1016/j.cbpa.2015.12.006.

[93] a) S. E. Motika, P. J. Hergenrother, *Nat. Prod. Rep.* **2020**, *37*, 1395, https://doi.org/10.1039/d0np00059k; b) I. Kjaerbolling, T. Vesth, M. R. Andersen, *mSystems* **2019**, *4*, https://doi.org/10.1128/mSystems.00085-19; c) N. Liu, E. D. Abramyan, W. Cheng, B. Perlatti, C. J. B. Harvey, G. F. Bills, Y. Tang, *J. Am. Chem. Soc.* **2021**, *143*, 6043, https://doi.org/10.1021/jacs.1c01516; d) N. P. Keller, *Nat. Chem. Biol.* **2015**, *11*, 671, https://doi.org/10.1038/nchembio.1897; e) C. D. Payne, B. Franke, M. F. Fisher, F. Hajiaghaalipour, C. E. McAleese, A. Song, C. Eliasson, J. Zhang, A. S. Jayasena, G. Vadlamani, R. J. Clark, R. F. Minchin, J. S. Mylne, K. J. Rosengren, *Chem. Sci.* **2021**, *12*, 6670, https://doi.org/10.1039/d1sc00692d; f) L. K. Buckton, M. N. Rahimi, S. R. McAlpine, *Chemistry* **2021**, *27*, 1487, https://doi.org/10.1002/chem.201905385; g) P. G. Dougherty, A. Sahni, D. Pei, *Chem. Rev.* **2019**, *119*, 10241, https://doi.org/10.1021/acs.chemrev.9b00008; h) Z. Guo, S. Y. Hong, J. Wang, S. Rehan, W. Liu, H. Peng, M. Das, W. Li, S. Bhat, B. Peiffer, B. R. Ullman, C. M. Tse, Z. Tarmakova, C. Schiene-Fischer, G. Fischer, I. Coe, V. O. Paavilainen, Z. Sun, J. O. Liu, *Nat. Chem.* **2019**, *11*, 254, https://doi.org/10.1038/s41557-018-0187-4; i) C. J. Gerry, S. L. Schreiber, *Nat. Chem. Biol.* **2020**, *16*, 369, https://doi.org/10.1038/s41589-020-0469-1; j) S. L. Schreiber, *Cell* **2021**, *184*, 3, https://doi.org/10.1016/j.cell.2020.12.020.