

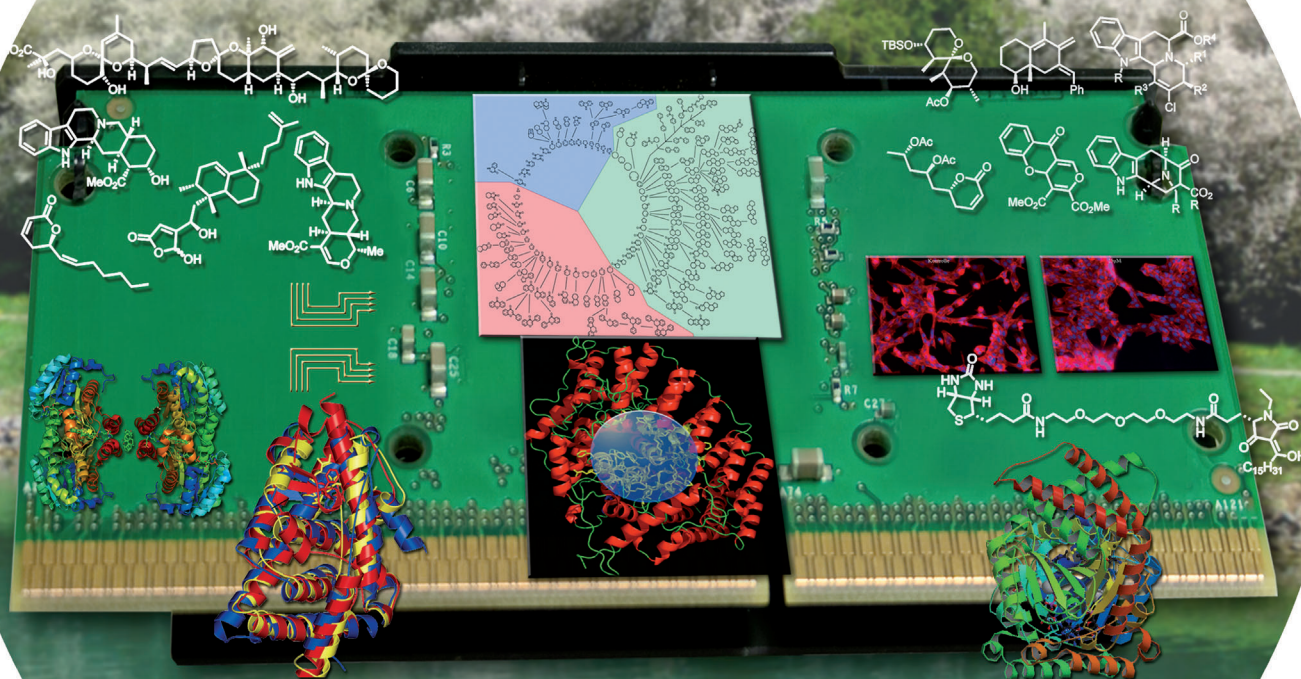
Biologie-orientierte Synthese (BIOS)

Stefan Wetzels, Robin S. Bon, Kamal Kumar und Herbert Waldmann*

Stichwörter:

Bioorganische Chemie ·
Chemieinformatik ·
Medizinalchemie ·
Naturstoffe ·
Syntheseplanung

Biologie-orientierte Synthese



Wirkstoffkandidaten
Chemische Sonden
Neue Zielstrukturen

Welche Substanzklassen eignen sich am besten als Werkzeuge für chemisch-biologische Forschung und als Inspiration für medizinisch-chemische Projekte? Der chemische Strukturraum ist riesig und kann durch Synthese von Verbindungen nicht vollständig exploriert werden. Daher werden Methoden benötigt, mit deren Hilfe die biologisch relevanten Anteile des chemischen Strukturraums identifiziert und kartiert werden können. Die mit diesen Methoden erzeugten Hypothesen inspirieren dann Syntheseprogramme, um die biologisch relevanten Teile des Strukturraumes mit realen Verbindungen zu füllen. Die Biologie-orientierte Synthese baut darauf auf, dass die Evolution von Proteinen und Naturstoffen sich auf bestimmte Strukturklassen beschränkt. Sie nutzt eine hierarchische Klassifikation bioaktiver Substanzen, die auf Substrukturbeziehungen und der Art der biologischen Aktivität beruht. Mit dieser Methode werden Gerüststrukturen biologisch aktiver Substanzklassen ausgewählt und als Startpunkte für die Synthese von Substanzkollektionen mit fokussierter Diversität genutzt. Scaffold Hunter, ein intuitiv zugängliches und hochgradig interaktives Computerprogramm, ermöglicht die Navigation im chemischen Strukturraum. In Bibliotheken niedermolekularer Substanzen, die nach dem BIOS-Konzept synthetisiert wurden, ist biologische Aktivität oft angereichert. Sie ermöglichen die Untersuchung komplexer biologischer Phänomene durch direkte Perturbation und können darüber hinaus auch als Inspiration in der Medikamentenentwicklung dienen.

„Die Natur macht nichts umsonst“
Aristoteles

1. Einleitung

Die Untersuchung biologischer Systeme mittels niedermolekularer Substanzen („small molecules“; Substanzen mit einem Molekulargewicht unter 800 g mol^{-1}) ist ein zentraler Bestandteil der chemischen Biologie. Bioaktive Moleküle können hervorragende Werkzeuge und molekulare Sonden für die Analyse komplexer biologischer Netzwerke und Systeme mit robuster und redundanter Funktionalität sein. Ihre Wirkung ist im Unterschied zu genetischen Methoden akut und nicht chronisch. Sie wirken schnell und reversibel, und ihre Anwendung ist steuer- und regelbar (durch Konzentrationsveränderung).^[1] Obwohl sich molekulare Sonden oft in ihren Eigenschaften von Wirkstoffen unterscheiden,^[2] sind erfolgreiche chemische Sonden eine wertvolle Inspirationsquelle für die Wirkstoffentwicklung. Im Lauf der vergangenen Jahrzehnte wurde eine Vielzahl niedermolekularer Substanzen entdeckt, die die Aktivität einer großen Zahl von Proteinen beeinflussen, und der Bedarf an qualitativ hochwertigen chemischen Sonden mit klar definierter Struktur, Wirkung, Selektivität, Wirkmechanismus und Verfügbarkeit wächst.^[3] Als ein großes Ziel der chemisch-biologischen Forschung wurde die Entwicklung eines synthetischen Moleküls für jedes Protein im humanen Genom vorgeschlagen,

Aus dem Inhalt

1. Einleitung	10991
2. Begrenzte Zahl von Strukturen und Diversität im Strukturraum der Naturstoffe und der Proteinbindungsstellen	10992
3. Biologie-orientierte Synthese (BIOS)	11001
4. Zusammenfassung und Ausblick: Woher kommen wir und wohin gehen wir?	11012
5. Resümee	11016

das die Aktivität genau dieses Proteins selektiv beeinflusst.^[4] Eine der größten Herausforderungen bei diesem Unterfangen ist die Identifikation geeigneter Substanzklassen zur Modulation einer bestimmten Proteinfunktion. Da aktuelle Schätzungen davon ausgehen, dass der Wirkstoff-ähnliche Strukturraum mehr als 10^{60} Moleküle umfasst, ist es unmöglich, alle möglichen Verbindungen zu synthetisieren.^[5] Tatsächlich reichen weder die Materie im

Universum noch die Zeit bis zum Ende des Universums aus, um alle diese Moleküle herzustellen. Die entscheidende Frage bei der Entwicklung niedermolekularer Substanzen für die chemisch-biologische Forschung, und in Analogie und Erweiterung auch für die Wirkstoffentwicklung, ist: Wie kann man die mit biologisch relevanten Substanzen angereicherten Teile des chemischen Strukturraums identifizieren? Oder anders formuliert: Wie kann man den biologisch relevanten Strukturraum identifizieren, kartieren und in ihm navigieren?^[5c]

Vergleichbar mit der beschränkten Zugänglichkeit biologisch relevanter Moleküle war die Natur sehr konservativ bei der Entwicklung des chemischen Strukturraums, der durch die Proteinbindungstaschen definiert wird. Ein Protein aus durchschnittlich 300 Aminosäurebausteinen, von denen jeder aus einer von 20 unterschiedlichen Aminosäuren besteht, kann auf mehr als 10^{390} einzigartige Möglichkeiten aufgebaut

[*] Dr. S. Wetzel, Dr. R. S. Bon, Dr. K. Kumar, Prof. Dr. H. Waldmann
Max-Planck-Institut für Molekulare Physiologie
Abteilung Chemische Biologie
Otto-Hahn-Straße 11, 44227 Dortmund (Deutschland)
und
Technische Universität Dortmund, Fakultät Chemie
Lehrbereich Chemische Biologie
Otto-Hahn-Straße 6, 44227 Dortmund (Deutschland)
E-Mail: herbert.waldmann@mpi-dortmund.mpg.de

werden.^[5a] Selbst die Genome der kompliziertesten Lebewesen kodieren jedoch nicht mehr als 10^4 – 10^5 Proteine, die oft innerhalb von Proteinfamilien hoch konservierte Subdomänen enthalten. Diese Konservierung von Strukturen hat eine beschränkte Zahl an möglichen Bindungstaschen für kleine Moleküle zur Folge, was wiederum die Entwicklung rationaler Methoden für die Entwicklung von Liganden und Inhibitoren beflügelt. Aktuelle hochmoderne Methoden basieren beispielsweise auf mechanistischen Überlegungen (Mechanismus-basierte Inhibitoren), evolutionären Argumenten (Sequenzhomologie), dreidimensionaler Proteinstruktur (strukturbasiertes Design)^[6] oder der Klassifikation von niedermolekularen Substanzen nach vorgegebenen molekularen Eigenschaften (Deskriptoren).^[7] Angesichts der beschränkten Diversität in der Evolution sowohl der synthetischen Moleküle als auch der Proteinbindungstaschen, entwickelten wir eine konzeptionell andere, strukturbasierte Methode, um biologisch relevante chemische Strukturräume zu analysieren und für die Entwicklung synthetischer Moleküle für chemisch-biologische und medizinisch-chemische Forschung zu nutzen. Wir bezeichnen diese Methode als Biologie-orientierte Synthese (BIOS). BIOS beruht auf der Analyse von Strukturen aus der Welt der Proteine sowie der niedermolekularen Substanzen in Kombination mit der begrenzten Zahl von Strukturen und Diversität in der Natur. In diesem Aufsatz entwickeln wir zunächst das gedankliche Gerüst, auf dem unsere Argumentation beruht. Danach beschreiben wir die Entwicklung der chemie- und bioinformatischen Methoden und Werkzeuge, die die Identifizierung, Analyse, Kartierung und Navigation in biologisch relevanten chemischen Strukturräumen ermöglichen. Anschließend zeigen wir Beispiele

für die Anwendung dieser Methoden in der Entwicklung und Synthese von Substanzkollektionen. Schließlich demonstrieren wir, wie chemische Sonden gemäß der Logik von BIOS entwickelt werden können und wie ihre Anwendung zu neuen Einsichten in biologische Phänomene geführt hat.

2. Begrenzte Zahl von Strukturen und Diversität im Strukturräum der Naturstoffe und der Proteinbindungsstellen

Natürliche Sekundärmetabolite niedermolekularer Substanzen definieren einen Teil des biologisch relevanten chemischen Strukturrums, der für die Entdeckung bioaktiver Moleküle von Bedeutung ist. Naturstoffe waren und sind eine reichhaltige und umfangreiche Inspirationsquelle für die Wirkstoffentwicklung.^[8] Naturstoff-abgeleitete und Naturstoff-inspirierte Substanzbibliotheken haben höhere Trefferquoten in biochemischen und biologischen Tests gezeigt (zur Definition von Naturstoff-abgeleiteten und Naturstoff-inspirierten Substanzen, siehe Abbildung 1),^[9] und höhere molekulare Komplexität und eine hohe Zahl an Stereozentren sind zwei molekulare Eigenschaften, die Naturstoffe von typischen, vorzugsweise auf der Basis von synthetischer Zugänglichkeit entworfenen kombinatorischen Bibliotheken unterscheiden. Es konnte gezeigt werden, dass ein direkter Zusammenhang zwischen diesen molekularen Eigenschaften und der Erfolgswahrscheinlichkeit in der Entwicklung von Substanzen vom Wirkstoffvorläufer zum Medikament in der Klinik besteht.^[10] In ihrer Entstehung binden Naturstoffe an mehrere Proteine. Zum einen verläuft die Biosynthese von



Robin S. Bon promovierte 2007 bei Prof. Romano Orru an der Freien Universität Amsterdam in Organischer Chemie. Danach absolvierte er ein Postdoktorat in der Abteilung von Prof. Herbert Waldmann am Max-Planck-Institut für molekulare Physiologie in Dortmund (Alexander von Humboldt-Stipendiat). Seit November 2009 arbeitet er als Nachwuchswissenschaftler an der Universität Leeds. Gegenstand seiner Forschung ist die Entwicklung von biologisch aktiven Molekülen und Werkzeugen für biochemische Assays und die In-vivo-Bildgebung.



Stefan Wetzel studierte Chemie an den Universitäten Regensburg und Heidelberg und promovierte in der Abteilung von Prof. Waldmann am Max-Planck-Institut für molekulare Physiologie über computergestützte Methoden zur Entwicklung fokussierter, biologisch relevanter Substanzbibliotheken. Zurzeit absolviert er ein Postdoktorat bei Novartis auf den Gebieten der quantitativen Biologie und Computer-Systembiologie.

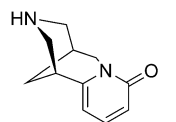


Kamal Kumar promovierte an der G.N.D. Universität in Amritsar, Indien, bei Prof. M. P. S. Ishar. Nach einem Postdoktorat bei Prof. M. Beller in Rostock (Alexander von Humboldt-Stipendiat) wechselte er 2002 in die Gruppe von Prof. H. Waldmann in der Abteilung für Chemische Biologie am Max-Planck-Institut für molekulare Physiologie. Seit Mai 2006 leitet er eine Gruppe in dieser Abteilung. Seine Forschungsinteressen gelten der Entwicklung neuer Synthesemethoden für Naturstoff-basierte Substanzbibliotheken, Kaskadenreaktionen, komplexitätserzeugenden Anellierungsreaktionen und der Erforschung biologischer Funktionen mit niedermolekularen Substanzen.

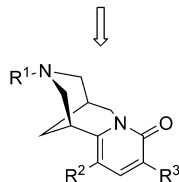


Herbert Waldmann wurde am 11. Juni 1957 in Neuwied geboren. Er promovierte 1985 in organischer Chemie bei Prof. H. Kunz an der Universität Mainz. Danach folgte ein Postdoktorat bei Prof. G. Whitesides an der Harvard University. Es folgten Berufungen auf Professuren an den Universitäten Bonn (1991), Karlsruhe (1993) und Dortmund (1991) sowie als Direktor am MPI für molekulare Physiologie. Seine Forschungsinteressen gelten dem Gebiet der chemischen Biologie, der Entwicklung bioaktiver niedermolekularer Verbindungen, Proteinsonden und der Mikroarray-Technologie.

Naturstoff-abgeleitete Bibliotheken



Cytisin



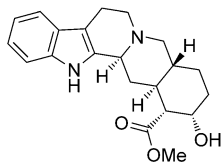
Gerüst ist identisch zum Naturstoffgerüst.

Gerüst wird typischerweise durch Abbau des Naturstoffs erhalten (auch die Synthese ist möglich).

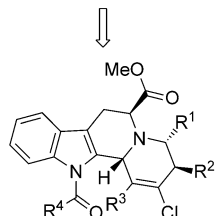
Substitutionsmuster ist weitgehend durch die Reaktivität und Gerüststruktur des Naturstoffs festgelegt.

Kein Variieren der Stereochemie.

Naturstoff-inspirierte Bibliotheken



Yohimbin



Gerüst ist eng verwandt mit dem Naturstoffgerüst.

Gerüst wird typischerweise de novo synthetisiert.

Substituenten werden während der Synthese eingeführt.

Substitutionsmuster kann von dem des Naturstoffs abweichen.

Stereochemie kann variiert werden.

Abbildung 1. Naturstoff-abgeleitete und Naturstoff-inspirierte Substanzbibliotheken.

Naturstoffen typischerweise über mehrfache, einander nachgeschaltete Bindung von Zwischenstufen an verschiedene Enzyme. Zum anderen zeigen viele Naturstoffe eine Vielzahl biologischer Wirkungen, sowohl innerhalb eines Organismus als auch bei mehreren Spezies. Gemeinsam legen diese Befunde nahe, dass Naturstoffe höchstwahrscheinlich an verschiedene Zielproteine binden und deren Funktion beeinflussen. Das trifft besonders auf eng verwandte Analoga zu, die eine gesamte Naturstoffklasse definieren. Tatsächlich sind die Gerüststrukturen von Naturstoffen in der Natur hochgradig konserviert (siehe auch Abschnitt 2.1). Viele Naturstoffe, die über ein gemeinsames Gerüst, aber verschiedene Substitutionsmuster verfügen, zeigen verschiedene Profile biologischer Aktivität. Daher definieren Naturstoffgerüststrukturen evolutionär ausgewählte „privilegierte Strukturen“.^[11] Diese verleihen der gesamten Substanzklasse die Fähigkeit zur Interaktion mit und Bindung an verschiedene Zielproteine und kodieren somit die strukturellen Eigenschaften, die für eine Bindung benötigt werden. Auf Grund dieser Eigenschaften definieren Naturstoffgerüststrukturen diejenigen biologisch relevanten Teile des riesigen chemischen Strukturraums, die in der Evolution besetzt wurden. Gleichzeitig sind sie die Antwort der Natur auf die Frage nach der Kartierung und Navigation im chemischen Strukturraum. Wichtig ist jedoch, dass dies nicht die einzige Antwort auf diese Frage darstellt – wie die Entwicklung von Wirkstoffen, die nicht auf Naturstoffen basieren, durch die pharmazeutische Industrie seit mehr als einem Jahrhundert zeigt.

Die hochselektive Erkennung von Naturstoffen und ihren Vorstufen durch die Biosynthesemaschinerie, spezifische Rezeptoren und Zielproteine setzt starke molekulare Wechselwirkungen zwischen den Naturstoffen und den Ligandbindungstaschen der Proteine voraus. Daher muss die Proteinstruktur exakt auf die Strukturmerkmale der Naturstoffe abgestimmt sein. Die dreidimensionale Struktur von Proteinen wird durch die Anordnung von Sekundärstrukturelementen, wie α -Helices und β -Faltblättern, im Rückgrat des Proteins bestimmt. Dadurch entstehen die charakteristischen Faltungstypen der einzelnen Proteindomänen, die dann zusammen die gesamte Proteinstruktur bilden. Kleinere Faltungsbereiche innerhalb von Domänen, so genannte Subfaltungen, legen die Größe und Form der Ligandbindungstasche sowie die Anordnung der katalytischen und Ligand-erkennenden Aminosäuren fest. Die jeweilige Struktur und die chemischen Eigenschaften der Aminosäurebausteine hingegen, vor allem die Seitenketten, bestimmen die Art der Liganden, die gebunden werden können. Die Struktur der Proteinfaltung ist stärker konserviert als die Aminosäuresequenz, und Proteindomänen mit niedriger Sequenzähnlichkeit können sehr ähnliche Faltungen einnehmen. Die Gesamtzahl der Faltungstypen in der Natur wird auf 1000–8000 geschätzt und fällt für die großen Proteinfamilien noch niedriger aus.^[12]

In der Evolution sowohl der Naturstoffgerüststrukturen als auch der Proteinfaltungen ist die Natur konservativ, aber divers in den Aminosäureseitenketten der Proteine und den Substituenten der Naturstoffe. Auf dieser Erkenntnis fußend haben wir eine mögliche Analogie zwischen den Naturstoffgerüststrukturen und den Subfaltungen der Ligandbindungstaschen sowie integrierten Ankerpunkten für die Bindung vorgeschlagen und untersucht. Wir vermuteten, dass hoch konservierte Naturstoffgerüststrukturen auf die ebenfalls hoch konservierten Subfaltungen der Ligandenbindungstaschen abgestimmt sind. Die Wechselwirkung der diversen Naturstoff-Substituenten mit den ebenso diversen Aminosäureseitenketten in der Ligandbindungstasche führt zu einer selektiven und starken Bindung. In diesem Fall bestimmt das Naturstoffgerüst die räumliche Anordnung der Substituenten und daher auch, wie der Ligand in Ligandbindungstaschen mit komplementärer Form und Größe, also die komplementären Subfaltungen, passt (siehe Abbildung 2). Eine Bindung kommt jedoch nur dann zu Stande, wenn die Eigenschaften und Größe der Naturstoff-Substituenten und der Aminosäureseitenketten in der Ligandbindungstasche ebenfalls gut zueinander passen. Nach diesem Konzept binden Naturstoffe (und möglicherweise andere Klassen von niedermolekularen Substanzen) mit ähnlichen Gerüststrukturen höchstwahrscheinlich an Proteine mit ähnlichen Subfaltungen der Ligandbindungstasche. Daher könnte die Identifikation der strukturellen Ähnlichkeit zwischen Naturstoffgerüststrukturen und Subfaltungen in Proteinen die Entwicklung Naturstoff-inspirierter Verbindungsbibliotheken lenken. Im Idealfall enthalten solche Substanzbibliotheken, die auf Naturstoffgerüststrukturen basieren und mit ausreichend diversen Substituenten versehen sind, Liganden für mehrere Proteindomänen mit ähnlichen Subfaltungen.

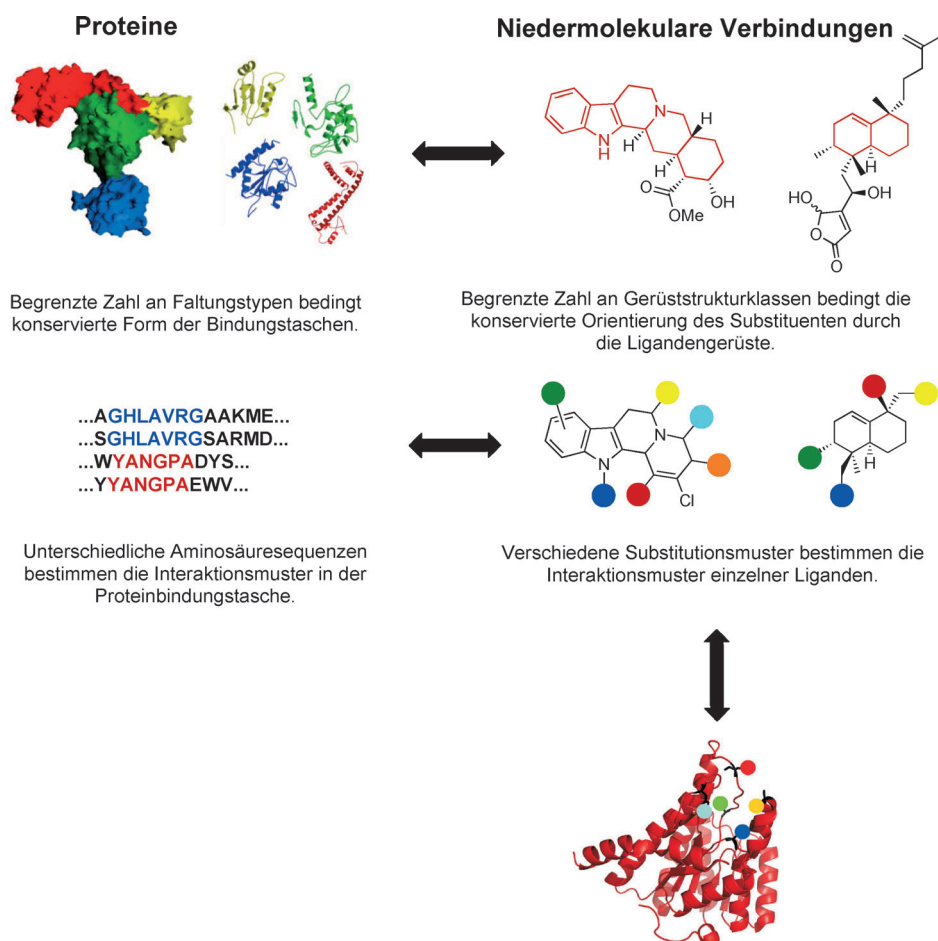


Abbildung 2. Gerüstsubstituentenanalogie zwischen niedermolekularen Substanzen und Proteinen. Das Gerüst der niedermolekularen Substanz legt die räumliche Orientierung der Substituenten fest, während die Subfaltung die Aminosäureseitenketten räumlich ausrichtet. Eine Bindung resultiert, wenn kompatible Substituenten in ihrer räumlichen Ausrichtung aufeinander treffen, sodass sie miteinander wechselwirken können.

Diese Argumentation rückt die Struktur der Moleküle und der Ligand-erkennenden Proteinkernstrukturen in das Rampenlicht der Auffindung bioaktiver Substanzen. Diese Chemie-zentrierte Methode zur Substanzentwicklung ist konzeptionell alternativ zu anderen validen Methoden, die beispielsweise auf mechanistischen und evolutionären Überlegungen beruhen oder auf die Maximierung chemischer Diversität abzielen (siehe oben).

Von besonderer Wichtigkeit ist, dass die strukturelle Diversität der Ligandbindungstaschen, die durch die verschiedenen Aminosäuren bei gleicher Subfaltung entsteht, die Entwicklung Naturstoff-inspirierter Substanzbibliotheken zwingend erfordert. Eine solche Substanzbibliothek wird nur dann Liganden für mehrere Proteine beinhalten, wenn die Diversität der Substituenten der jeweiligen Naturstoffgerüststruktur auf die Diversität der Aminosäureseitenketten abgestimmt ist, die möglicherweise in den strukturell ähnlichen Subfaltungen auftreten.

Zur Prüfung und Entwicklung dieses Konzepts und der Möglichkeit, neue biologisch prävalidierte Startpunkte im chemischen Strukturraum für die Entwicklung synthetischer

Substanzbibliotheken zu finden, haben wir neue chemie- und bioinformatische Methoden entwickelt. Diese ermöglichen die Identifikation, Analyse, Kartierung und Navigation in den Strukturräumen der biologisch relevanten Moleküle sowie der Proteinbindungstaschen. Diese Methoden wurden weiterhin genutzt, um die Entwicklung von Substanzbibliotheken zu lenken und die Bioaktivitätsprofile bestimmter Substanzklassen vorherzusagen.

2.1. Strukturelle Klassifikation der Naturstoffe (SCONP)

Frühe Untersuchungen der Eigenschaften von Naturstoffen zielten darauf ab,^[13] die Unterschiede zwischen Naturstoffen und typischen Verbindungen aus medizinisch-chemischen Programmen zu verstehen sowie die molekularen Eigenschaften zu entschlüsseln, die für die biologische Relevanz der Naturstoffe verantwortlich sind. Später wurde das Konzept der Gerüststrukturbäume („scaffold trees“) eingeführt und auf das „Dictionary of Natural Products“ (DNP) angewendet, woraus die erste strukturelle Klassifikation der Naturstoffe (SCONP) hervorging.^[14]

Diese Klassifikation kartierte effektiv den chemischen Strukturraum der Naturstoffe, wie er im DNP, der umfassendsten Datenbank von Naturstoffen, enthalten ist (Abbildung 3).^[15]

Zur Reduktion der hohen strukturellen Diversität der Naturstoffe auf ein beherrschbares Maß wurden Gerüststrukturen, definiert als alle Ringe des jeweiligen Moleküls sowie die sie verbindenden aliphatischen Ketten und ringständige Doppelbindungen, anstelle von Molekülen klassifiziert und hierarchisch angeordnet. Dabei wird für jede Gerüststruktur ein Ast erstellt, der durch iterativen Abbau der Gerüststruktur um jeweils einen Ring pro Schritt entsteht. Dieser Abbau wird durch einen Regelsatz gesteuert. Das resultierende kleinere Gerüst wird jeweils als „Elter“ bezeichnet, das größere Gerüst als „Kind“. Ringe werden so lange wie möglich wiederholt entfernt (in der Regel bis nur noch ein einzelner Ring übrig ist), wodurch eine Sequenz von Gerüststrukturen entsteht. In dieser Abfolge, die auch als „Ast“ des Gerüststrukturbaums bezeichnet wird, stellt jedes Eltergerüst eine Substruktur des entsprechenden Kindgerüsts dar. Anders ausgedrückt erwächst ein Kindgerüst aus dem

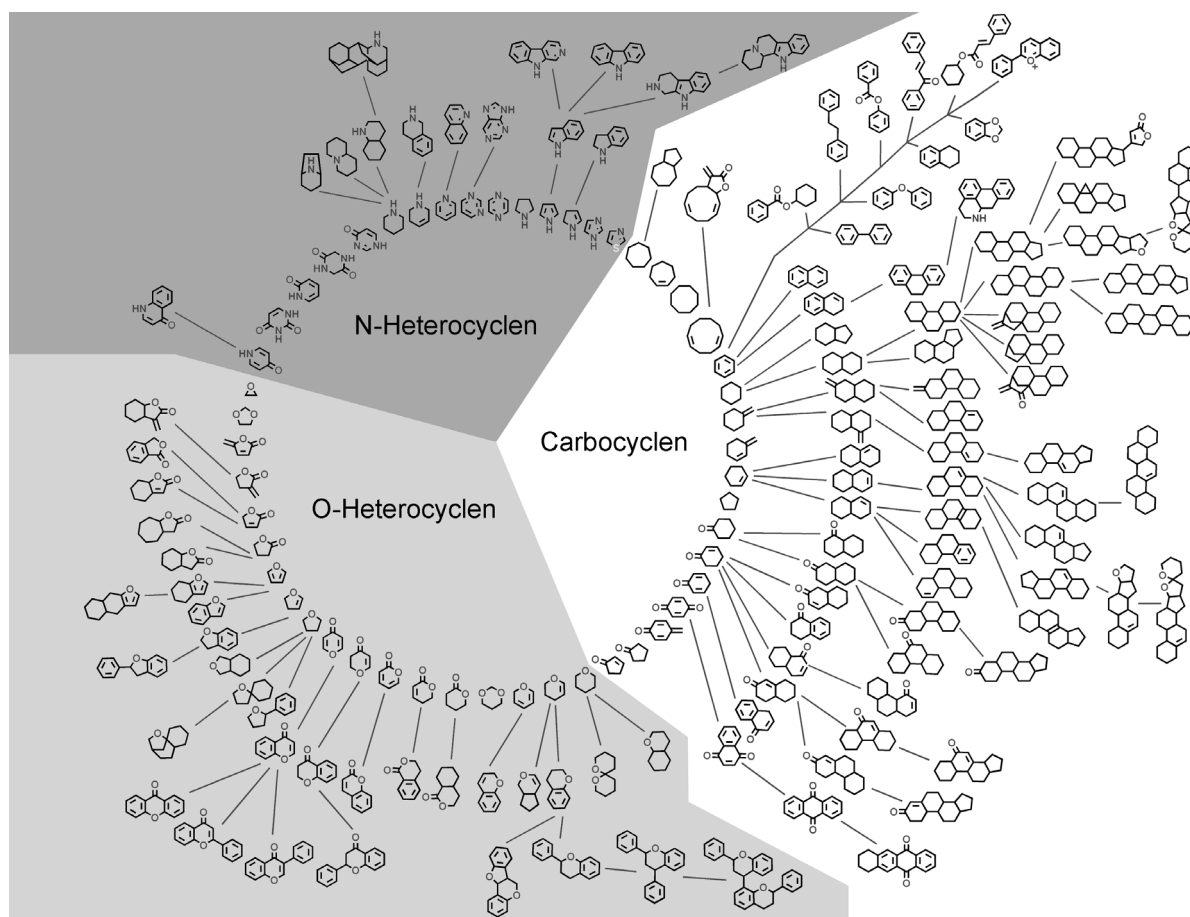


Abbildung 3. Der Gerüststrukturbaum der Naturstoffe.

Eltergerüst. Im letzten Schritt werden alle Äste zusammengefasst und bilden dann den endgültigen Gerüststrukturbaum. Der „Naturstoffgerüstbaum“ diente als Orientierungshilfe und zur Inspiration bei der Planung und Synthese mehrerer Naturstoff-inspirierter Substanzkollektionen sowie bei der Gewinnung neuer biologischer Erkenntnisse. Jedoch sind nicht nur die Naturstoffe biologisch relevant, sondern auch zahlreiche andere Substanzklassen, die nicht aus der Natur stammen, insbesondere viele Wirkstoffe und Agrochemikalien. Aus diesem Grund war es notwendig, das Konzept über die Naturstoffe hinaus zu erweitern. Die Integration weiterer Substanzklassen bedingte Änderungen am ursprünglichen Regelsatz, der für die Erstellung der strukturellen Klassifikation der Naturstoffe genutzt wurde, und in dem beispielsweise eine Regel verlangte, dass jedes Eltergerüst tatsächlich die Gerüststruktur mindestens eines Naturstoffs sein musste.

Zu diesem Zweck wurde ein neuer Satz von 13 Regeln entwickelt,^[2b,16] der weitere Eigenschaften und dirigierende Argumente einführte. So stellen diese Regeln beispielsweise sicher, dass jedes Kindergerüst mit genau einem Eltergerüst verbunden ist. Obwohl durch diesen Schritt die Information über alternative Äste größtenteils verlorengeht, ist genau diese Reduktion der Schlüsselschritt, um ein visuell verständliches Baumdiagramm mit deutlich erkennbarer Logik statt eines komplexen Netzwerks zu erhalten. Solche Ver-

einfachungen ermöglichen die „Skalierung der menschlichen Wahrnehmung“,^[17] das heißt, sie erlauben dem menschlichen Verstand, mit großen Datenmengen umzugehen und diese zu verarbeiten. Die Auswahl an Regeln und deren Reihenfolge im Regelsatz basiert unter anderem auf synthetischem und medizinischchemischem Wissen und ist daher teilweise subjektiv. Diese Vorgehensweise ermöglicht eine sehr flexible, aber dennoch intuitiv verständliche Klassifikation, die praktisch jedes Molekül aufnehmen und mit anderen über Substrukturbeziehungen verbinden kann.

Die anfängliche Konzentration auf den von der Natur genutzten chemischen Strukturraum in der ersten Version des Naturstoffgerüstbaumes führte zu vielen Lücken im Gerüststrukturbaum. Die fehlenden Strukturen waren entweder im Laufe der Evolution nicht synthetisiert oder bis dato nicht entdeckt worden. Die Erstellung und Überlagerung von Gerüststrukturbaumen verschiedener Molekülbibliotheken wurde durch diese Lücken sehr schwierig, wenn nicht sogar unmöglich. Die verbesserte Version des Regelsatzes erlaubt und generiert „virtuelle Gerüststrukturen“, um den Baum zu vervollständigen. Diese Gerüststrukturen repräsentieren keine Moleküle des zu analysierenden Datensatzes. Sie entstehen durch den iterativen Abbau der Gerüststrukturen und existieren nur *in silico*. Diese virtuellen Gerüststrukturen füllen die Lücken und stellen offensichtliche, vielversprechende Ansatzpunkte für chemische und biologische For-

schung dar. Der Begriff der „Brachiation“ (deutsch: „Schwinghängeln“) wurde der Anthropologie entliehen, in der er die Bewegung von Gibbons in den Bäumen ihres Lebensraumes beschreibt. Im Zusammenhang mit Gerüststrukturbäumen wird dieser Begriff verwendet, um die Bewegung von großen, komplexeren zu kleineren, strukturell weniger komplexen Gerüststrukturen entlang der Äste zu beschreiben. Es wird vorausgesetzt, dass die Art der biologischen Aktivität während der Brachiation erhalten bleibt, wenn auch in leicht veränderter Form, z.B. in Bezug auf Potenz (siehe Abbildung 4). Insbesondere erfolgt die Brachiation durch den Naturstoffgerüstbaum bei diesem Ansatz entlang der Pfade biologischer Relevanz. Dadurch unterscheidet sie sich wesentlich von strukturellen Vereinfachungen, die allein auf chemischen Kriterien basieren, z.B. Synthetisierbarkeit der kleineren Gerüststrukturen oder retrosynthetischen Überlegungen. Brachiation basiert auf der Hypothese, dass kleinere Gerüststrukturen Eigenschaften mit größeren Gerüststrukturen, deren Teil sie sind, gemeinsam haben – dem Konzept, auf dem sich die Fragment-basierte Wirkstoffentwicklung gründet.

Die Brachiation regte dazu an, den streng Chemie-basierten Aufbau des Naturstoffgerüstbaumes um Biologie-geleitete Strukturbäume zu ergänzen und zu erweitern.^[18] So scheiterten beispielsweise die Versuche, Morphin in den ersten Naturstoffgerüstbaum einzufügen (siehe Abbildung 22, Abschnitt 4). Die Ergänzung des Gerüststrukturbaumes um Gerüststrukturen von Substanzen, die keine Naturstoffe sind, aber ähnliche biologische Aktivität wie Morphin zeigen, hätte es ermöglicht, die Lücken zu schließen. Biologie-geleitete Strukturgerüstbäume bieten eine alternative Perspektive auf den chemischen Strukturraum, indem sie biologische Aktivität als Leitkriterium zum Aufbau der Äste von Gerüststrukturen heranziehen. In diesem Sinne stellen sie auf der Brachiation basierende Gerüststrukturbäume dar, also Sequenzen von Gerüststrukturen mit sehr ähnlicher aber leicht veränderter biologischer Aktivität. Bei der Biologie-gesteuerten Navigation im chemischen Strukturraum werden in jedem Abbauschritt alle möglichen Elter-Kind-Paare für das jeweilige Kindgerüst gebildet. Der Aufbau von Ästen erfolgt dann nach gleicher biologischer Aktivität, z.B. In-vitro-Aktivität gegen ein bestimmtes Protein. Sofern mehrere

Elter-Kind-Paare ähnliche biologische Aktivität zeigen, werden entsprechend mehrere Äste erstellt. Im letzten Schritt wird dann der längste Ast mit den wenigsten Lücken ausgewählt. Die Kombination dieser Äste zum endgültigen Gerüststrukturbaum erfolgt analog zu den Chemie-basierten Gerüststrukturbäumen.

Der Wert der Gerüststrukturbäume beruht wesentlich auf zusätzlicher verknüpfter Information aus verschiedenen Quellen, beispielsweise dem Ursprung, der Häufigkeit, der mittleren biologischen Aktivität der Gerüststrukturen oder der Information über Zielproteine. Die Notwendigkeit, hochgradig annotierte Gerüststrukturbäume darzustellen und dynamisch mit ihnen zu interagieren, um sie intuitiv nutzbar zu machen, führte zur Entwicklung des JAVA-basierten Computerprogramms „Scaffold Hunter“.^[19] Scaffold Hunter ermöglicht die automatische Darstellung, das Filtern von und Navigieren in Gerüststrukturbäumen auf intuitive Weise. Es bietet Eigenschafts- und Struktur-basierte Filter, eine große Auswahl von Methoden zur farblichen Hervorhebung sowie eine Vielzahl an individuell anpassbaren Einstellungen. Die interaktive Navigation in den Gerüststrukturbäumen schließt Vergrößern und Verkleinern, das Schwenken des Sichtfeldes sowie die automatische Erstellung von Teilbäumen basierend auf ausgewählten Gerüststrukturen ein. Ein zweites Programm namens „ScaffoldTreeGenerator“ ermöglicht die Erstellung von Gerüststrukturdatenbanken aus SD-Dateien,^[20] einem Dateiformat, das von vielen weit verbreiteten Strukturzeichenprogrammen, wie unter anderem ChemDraw^[21] und ISIS Draw,^[22] exportiert werden kann. Des Weiteren erlaubt es auch den Import von zusätzlichen Daten, z.B. Bioaktivitätsdaten. Gemeinsam ermöglichen es die beiden Programme Chemikern und Biologen, die oft keine Experten in Chemieinformatik sind, Gerüststrukturdatenbanken aus praktisch jedem Datensatz von chemischen Strukturen und damit verknüpften Daten aufzubauen und zu analysieren. Beide Programme sind frei verfügbar unter scaffoldhunter.sourceforge.net.

Die Anwendung von Scaffold Hunter sowie der Chemie- und Biologie-basierten Gerüststrukturbäume zur Entdeckung bioaktiver Substanzen hängt von dem zu analysierenden Datensatz und der Problemstellung ab. Chemie-basierte Gerüststrukturbäume können für jede verfügbare Zahl von

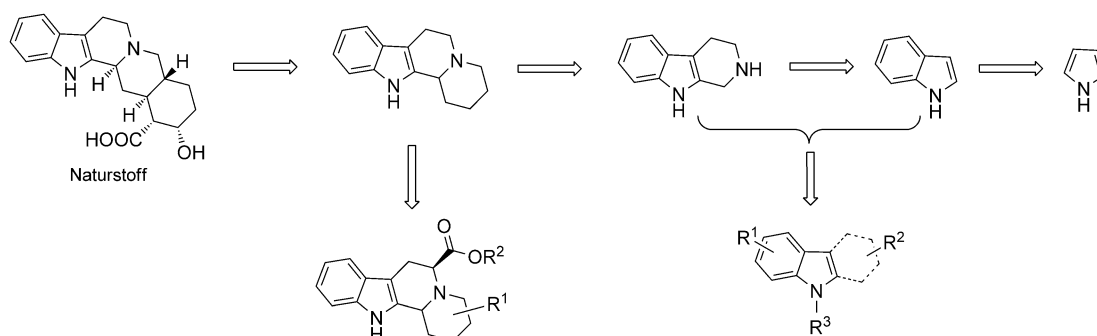


Abbildung 4. „Schwinghängeln“ im Ast eines Gerüststrukturbaums verdeutlicht an einem Beispiel aus den N-Heterocyclen des Naturstoffgerüstbaumes. In diesem Fall führt das Schwinghängeln vom pentacyclischen Gerüst des Yohimbins, eines Inhibitors der Phosphatase Cdc25A, zu tetra-, tri- und bicyclischen Gerüststrukturen. Ausgehend von diesen Gerüststrukturen wurden Substanzbibliotheken synthetisiert, in denen mehrere Cdc25A-Inhibitoren identifiziert wurden (Details siehe Abschnitt 3.2).

Molekülen und unabhängig von den übrigen verfügbaren Daten generiert werden. Sie können genutzt werden, um verschiedene Datensätze zusammenzufügen und um Syntheseprojekte zu inspirieren. Die Bioaktivitäts-basierten Gerüststrukturbäume können große Mengen von Bioaktivitätsdaten integrieren und erlauben Vorhersagen von Bioaktivitäten. In vielen Fällen werden sich beide Methoden gegenseitig ergänzen und sollten parallel angewendet werden, sofern es die Daten erlauben.

Eine besonders vielversprechende und offensichtliche Anwendung ist die Nutzung virtueller Gerüststrukturen in Chemie-basierten Gerüststrukturbäumen. Da virtuelle Gerüststrukturen Eigenschaften mit den benachbarten Gerüststrukturen teilen, sollten sie gute Ausgangspunkte für die Planung von mit biologischer Aktivität angereicherten Substanzbibliotheken darstellen. Ähnliches sollte auch für die Biologie-basierten Gerüststrukturbäume gelten. Gerüststrukturen, die Lücken in den Ästen darstellen, also Substanzklassen ohne Annotation für das betrachtete Protein, können gute Startpunkte für die Entwicklung von Substanzkollektionen mit einer bestimmten, erwarteten Bioaktivität sein.^[19]

Um die Frage zu beantworten, ob virtuelle Gerüststrukturen, welche die Lücken in Chemie-basierten Gerüststrukturbäumen besetzen, vielversprechende Startpunkte für die organische Synthese darstellen, wurde ein Gerüststrukturbaum aus 765 135 Strukturen der PubChem-Datenbank erstellt, die mindestens einen Ring enthalten und für die Daten aus biologischen oder biochemischen Tests verfügbar waren.^[19] Die in PubChem angegebenen Zielproteine wurden dann mit einer Liste aus der WOMBAT-Datenbank verglichen. WOMBAT^[23] ist eine Datenbank mit Substanzen sowie deren Bioaktivitätsdaten aus der wissenschaftlichen Literatur. Es wurden vielversprechende virtuelle Gerüststrukturen gesucht, die neben Gerüststrukturen, die gegen das jeweilige Zielprotein Aktivität aufweisen, liegen. Dabei wurden nur Proteine berücksichtigt, für die Daten in beiden Datenbanken verfügbar waren. In WOMBAT wurde dann nach Verbindungen gesucht, welche die virtuelle Gerüststruktur aus PubChem enthalten und gegen dasselbe Protein aktiv sind, also die Lücken im PubChem-Datensatz schließen.

Das Potenzial der Methode zur Vorhersage von vielversprechenden Gerüststrukturen wurde anhand des Pyruvatkinase-Datensatzes in PubChem untersucht.^[24] Vier Gerüststrukturen wurden für die Erstellung einer kleinen Substanzkollektion ausgewählt. Diese wurde dann biochemisch auf Inhibition und Aktivierung von Pyruvatkinase untersucht. Neun Substanzen zeigten einen AC_{50} -Wert $< 10 \mu\text{M}$ in den Tests. Virtuelle Gerüste aus Ästen mit inhibitorischer Aktivität ergaben sechs Inhibitoren, und virtuelle Gerüste aus Ästen mit aktivatorischer Aktivität führten zu drei Aktivatoren. Eine Suche in Chemical Abstracts ergab, dass keine dieser Substanzen vorher in irgendeiner Form mit Kinaseinhibition in Verbindung gebracht worden war (Abbildung 5).^[25]

Wie vorstehend erwähnt, können Biologie-basierte Gerüststrukturbäume auf ähnliche Weise verwendet werden. Eine entsprechende Analyse der WOMBAT-Datenbank mit Bioaktivitäts-basierten Gerüststrukturbäumen zeigte, dass

Schwinghangeln für 1/3 aller Zielproteine möglich ist – in vielen Fällen über 3 bis 9 Schritte. Unter diesen Zielproteinen finden sich Vertreter aller großen Klassen von Proteinen, die von Wirkstoffen adressiert werden, d.h. Kinasen, Opioidrezeptoren, G-Protein-gekoppelte Rezeptoren und Enzyme (Abbildung 6). Es stellte sich heraus, dass Brachiation wesentlich weiter verbreitet ist als erwartet und eher die Regel als eine Ausnahme darstellt.

Die Möglichkeit, Lücken in Ästen für die Substanzentwicklung zu nutzen, wurde anhand von zwei Sequenzen für 5-Lipoxygenase (5-LOX) und Östrogenrezeptor alpha ($ER\alpha$) biochemisch untersucht. Der Ast für 5-LOX enthielt Verbindungen mit 1 bis 7 Ringen, wobei auf der Ebene der Tricyclen eine Lücke in der Annotation auftrat (Abbildung 7). Der $ER\alpha$ -Ast reichte vom Gerüst mit sechs Ringen bis zum Monocyclen und wies eine Lücke auf der Bicyclen-Ebene auf (Abbildung 8). Von vier Verbindungen, die ausgehend vom tricyclischen 5-LOX-Gerüst entworfen wurden, zeigten zwei einstellig mikromolare IC_{50} -Werte in einem zellbasierten Testsystem. Basierend auf dem bicyclischen Gerüst für $ER\alpha$ wurden acht Verbindungen entworfen. Konzentrationsabhängige Messungen mit einem fluoreszenzbasierten Testsystem identifizierten einen Inhibitor mit einem IC_{50} -Wert von $20 \mu\text{M}$ gegen $ER\alpha$ und $4.6 \mu\text{M}$ gegen $ER\beta$. Obwohl die Wirksamkeit auf den ersten Blick nicht allzu stark erscheint, ergibt eine nähere Analyse, dass der nicht-optimierte Inhibitor nur ca. 100-fach weniger potent ist, als das natürliche Substrat Estradiol.

Diese Ergebnisse zeigen, dass Scaffold Hunter die Identifikation von Gerüststrukturen ermöglicht, die Lücken füllen, und zwar sowohl in Chemie-basierten als auch in Biologie-basierten Gerüststrukturbäumen. Diese Strukturen stellen vielversprechende Startpunkte zum Entwurf von fokussierten Kollektionen niedermolekularer Substanzen dar, in denen biologische Relevanz für das jeweils betrachtete Protein angereichert ist.

Brachiation kann eine praktikable Strategie sein, um strukturell einfachere Analoga zu identifizieren, besonders bei der Entwicklung von Naturstoff-inspirierten Substanzbibliotheken. In vielen Fällen sind jedoch kaum Informationen über das Bioaktivitätsprofil der Naturstoffe verfügbar. Daher sind Methoden zur Vorhersage von Bioaktivität von großem Wert. Wie bereits oben erwähnt, ist Brachiation ein weitverbreitetes Phänomen, und Gerüststrukturklassen, die Lücken in (nicht) annotierten Gerüststrukturbäumen besetzen, können ähnliche biologische Aktivität wie ihre benachbarten Gerüststrukturen aufweisen. Die beiden Beobachtungen führten zu der Hypothese, dass biologische Annotation von annotierten auf nicht-annotierte Substanzbibliotheken übertragen werden könnte, indem die Gerüststrukturbäume der beiden Datensätze vereint werden (Abbildung 9). Neben der direkten Annotation durch in beiden Datensätzen enthaltenen Gerüststrukturen sollte sich die Annotation entlang der Äste des Gerüststrukturbaumes übertragen lassen, analog zur Brachiation. Daher kann eine viel umfassendere Annotation erreicht werden, da auch Gerüststrukturen, die ausschließlich im annotierten Datensatz enthalten sind, ihre Annotation an benachbarte Gerüststrukturen im selben Ast übertragen können.

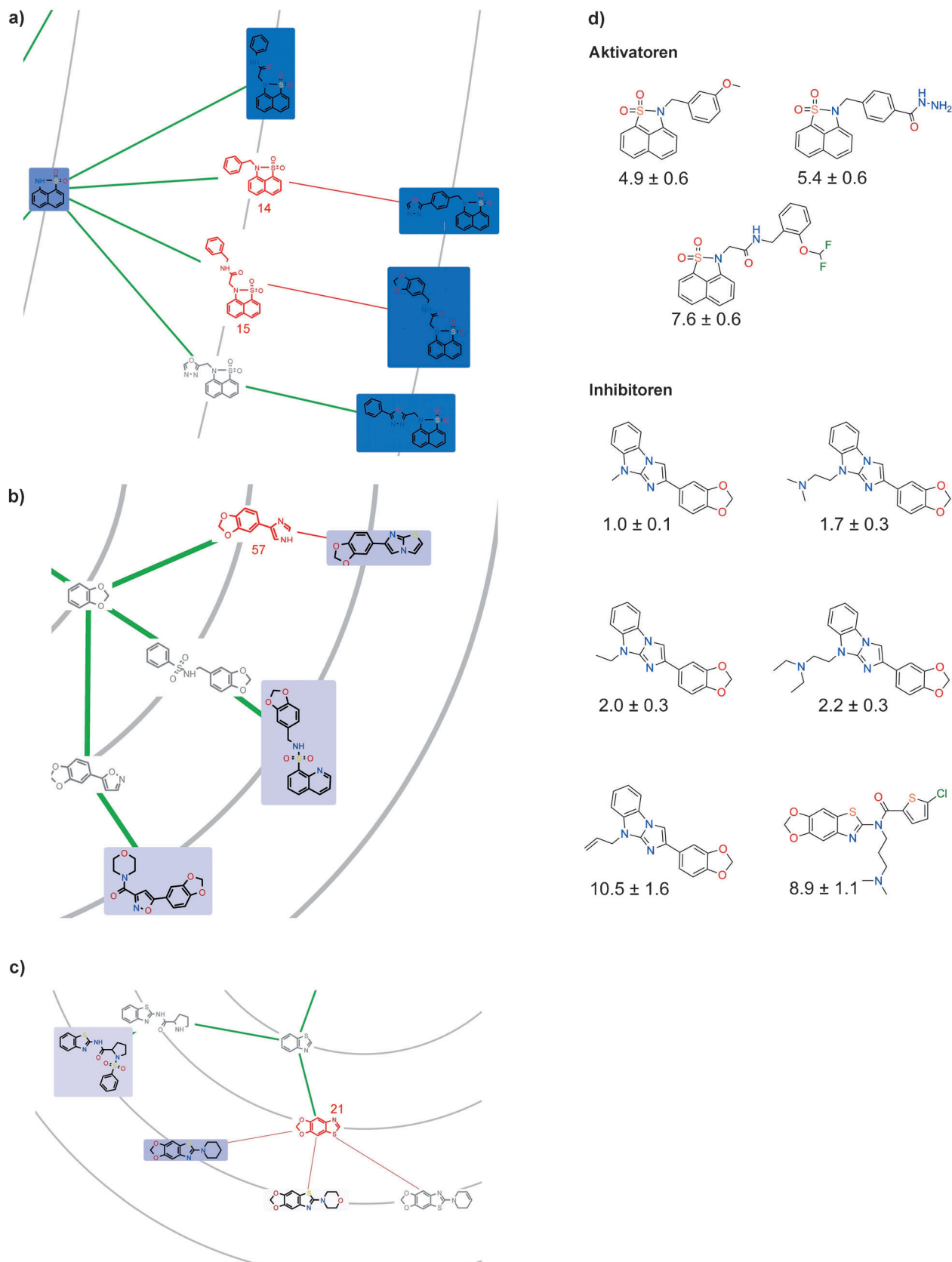


Abbildung 5. Ausgewählte Äste des Gerüststrukturbaumes basierend auf dem Pyruvatkinase-Datensatz und den Testergebnissen. Die vier virtuellen Gerüststrukturen sind rot dargestellt, zusammen mit der Zahl an Verbindungen für das jeweilige Gerüst. a) Dieser Ast enthält mehrere Gerüststrukturen, die gute Aktivatoren der Pyruvatkinase repräsentieren. b, c) Äste, die Inhibitoren der Pyruvatkinase repräsentieren. Zusätzliche virtuelle Gerüststrukturen sind grau dargestellt. Die Blautöne repräsentieren die mittleren $\log(\text{AC}_{50})$ -Werte aus dem Datensatz (dunkleres Blau bedeutet höhere Aktivität). Die Bilder wurden aus Scaffold Hunter exportiert. d) Inhibitoren und Aktivatoren der Pyruvatkinase mit einem IC_{50} -Wert $\leq 10 \mu\text{M}$ aus dem Test der Pyruvatkinase (Daten verfügbar in PubChem als Assay ID 2941). Abdruck nach *Nat. Chem. Biol.* **2008**, 5, 581–583.

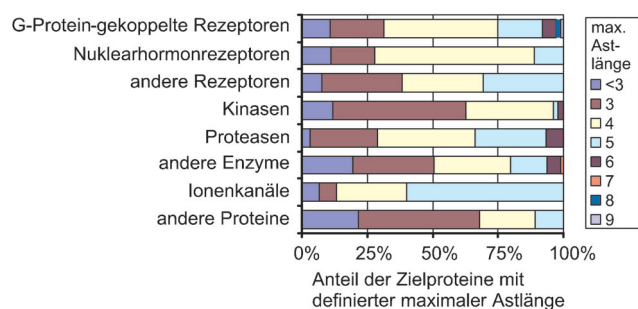


Abbildung 6. Die Länge einer Brachiations-Sequenz, das heißt die Zahl an Ringen, die unter Beibehaltung ähnlicher Bioaktivität entfernt werden können, für die wichtigsten Klassen von durch Wirkstoffe adressierten Proteinen. Die Verteilung der Länge der längsten Äste pro Protein über die Proteinklassen hinweg zeigt, dass für die meisten Proteinklassen mehr als die Hälfte der Äste eine Länge von 4 oder mehr aufweist.

Diese Hypothese wurde durch Annotation der Naturstoffstrukturen des γ -Pyron-Astes des DNP mit Bioaktivitätsinformation aus der WOMBAT-Datenbank geprüft. Dafür wurden die entsprechenden Gerüststrukturbäume aus dem DNP und WOMBAT verschmolzen.^[26]

Aus WOMBAT wurden mehrere Gerüststrukturen identifiziert, bei denen dieselbe Aktivität für mehr als zwei der fünf Gerüste in einem Ast bekannt war. Eine Substanzkollektion aus 500 γ -Pyronen, die fünf Gerüststrukturen auf drei Hierarchieebenen abdeckten, wurde zusammengestellt. Diese Bibliothek wurde auf Inhibition der in WOMBAT beschriebenen Proteine Monoaminoxidase A und B, der STAT-Proteine STAT1, STAT3 und STAT5b (STAT = signal transducers and activators of transcription) sowie der sauren Sphingomyelinase getestet. Bemerkenswerterweise wurden Inhibitoren für alle Proteine entdeckt (Abbildung 10), die in

einigen Fällen selektiv (auf Iso-Enzymebene) und ähnlich zu Strukturen waren, welche unabhängig in ergebnisoffenen Testreihen gefunden wurden.^[27]

Diese Ergebnisse belegen, dass Gerüststrukturbäume gut geeignet sind, um neue Gerüststrukturen für die Entwicklung von Substanzbibliotheken zu identifizieren, indem man die Lücken in bekannten Datensätzen füllt. Des Weiteren eröffnen sie die Möglichkeit, durch die Verschmelzung von Gerüststrukturbäumen Bioaktivitätsannotationen von Substanzklassen ohne Annotation vorherzusagen. Die prospektive Annotation ist besonders wichtig für die Vorhersage der biologischen Funktion von Naturstoffklassen. Allerdings ist die Methode auf die Gerüststrukturen beschränkt und kann nicht notwendigerweise die Annotation von biologischer Aktivität für einzelne Naturstoffe leisten.

2.2 Proteinstrukturähnlichkeitsclustering (PSSC)

Bindende molekulare Wechselwirkungen bedingen komplementäre Strukturen in Ligand und Protein, wobei meist die Substituenten des Liganden und die Seitenketten der im Protein enthaltenen Aminosäuren direkt beteiligt sind. Hohe Sequenzähnlichkeit bei Proteinen führt meist zu sehr ähnlichen Strukturen und somit auch zur Bindung ähnlicher Liganden. Die Definition der Komplementarität im PSSC-Konzept geht über die Sequenzähnlichkeit hinaus und integriert Strukturen mit geringer Sequenz-, aber hoher struktureller Ähnlichkeit. Während Sequenzähnlichkeit durch Analyse der Sequenzhomologie bestimmt werden kann, werden für strukturelle Ähnlichkeit strukturbasierte Methoden benötigt.

Die dreidimensionale Anordnung der Interaktionspunkte im Raum wird durch das Strukturgerüst bestimmt, d. h. durch die molekulare Gerüststruktur oder die Anordnung des

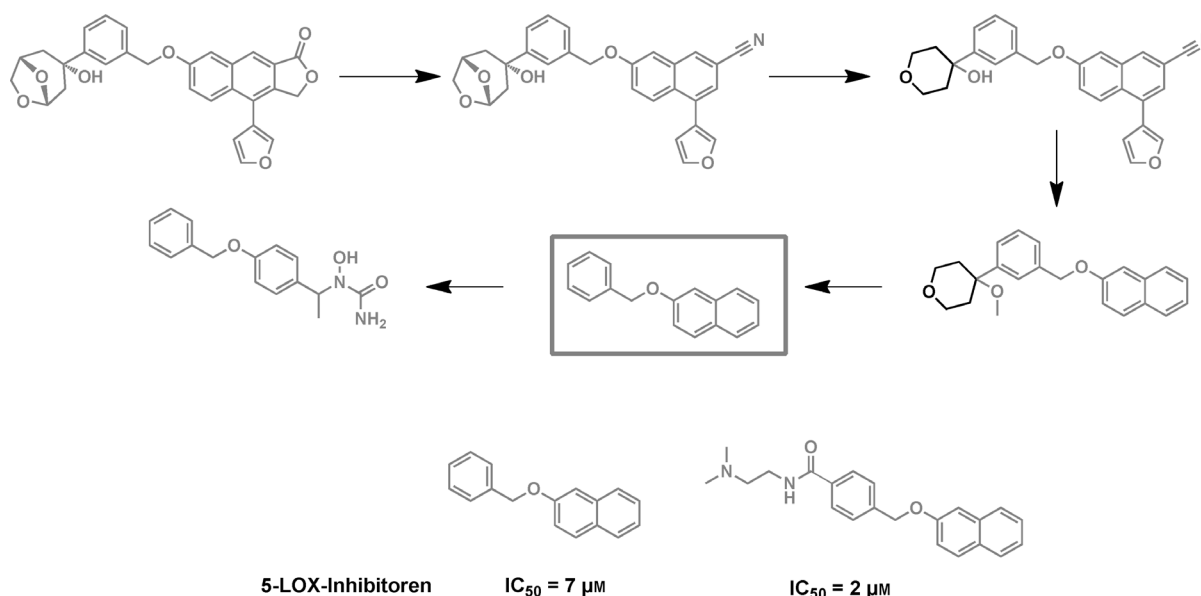
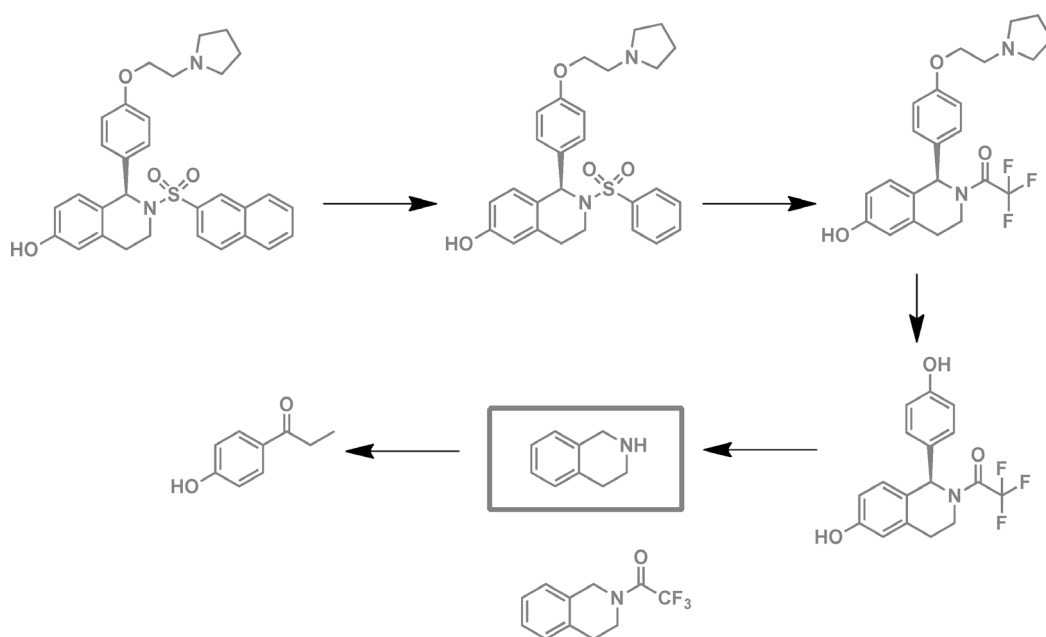


Abbildung 7. Ast eines Biologie-basierten Gerüststrukturbäums für 5-LOX-Inhibitoren. Für die Gerüststruktur im Kasten ist keine Aktivitätsinformation für 5-LOX bekannt. Sie diente als Vorlage für eine kleine Kollektion von Substanzen, die zwei mikromolare Inhibitoren enthielt.




Inhibition von ER α / β : IC₅₀ = 20 μ M und 4.6 μ M

Abbildung 8. Ast des Gerüststrukturbauemes für Inhibitoren des Östrogenrezeptors α/β (ER α/β). Es waren keine Verbindungen bekannt, welche die Gerüststruktur im Kasten enthielten und die Aktivität von ER α/β beeinflussen. Daher diente dieses Gerüst als Vorlage für eine kleine Kollektion an Verbindungen, die einen Inhibitor enthielt.

Proteinrückgrats (= Subfaltungstyp) in den Ligand-erkennenden Bindungsstellen des Proteins (Abbildung 2). Folglich sollte für eine Bindung die Komplementarität auf der abstrakteren Ebene der Gerüststrukturen ebenfalls Voraussetzung sein. Liganden mit ähnlichen Gerüststrukturen werden daher von Proteinen mit ähnlichem Subfaltungstyp der Bindungstasche gebunden. Diese Hypothese bildet die Grundlage des Proteinstrukturähnlichkeitsclusterings (PSSC), bei dem Proteine nach der strukturellen Ähnlichkeit des Subfaltungstyps ihrer Bindungstasche gruppiert werden (Abbildung 2). Die resultierenden Gruppen können dann genutzt werden, um beispielsweise vielversprechende Substanzklassen für Proteine oder alternative Zielproteine für eine bestimmte Substanzklasse zu identifizieren.

Ein erstes Beispiel für eine mögliche Anwendung des PSSC basiert auf Literaturdaten.^[28] Der Farnesoid-X-Rezeptor (FXR), ein nukleärer Hormonrezeptor, der eine Schlüsselrolle in Stoffwechselkrankheiten einnimmt, wurde gemeinsam mit dem Östrogenrezeptor β (ER β)^[29] und dem peroxisomalen Proliferations-aktivierten Rezeptor γ (PPAR γ)^[30] gruppiert. Obwohl die Sequenzähnlichkeit der drei Proteine weniger als 20 % beträgt, ist der Subfaltungstyp der Bindungstasche hochgradig konserviert (Abbildung 11). Eine anschließende Literatursuche führte zu dem Naturstoff Genistein, einem bekannten Inhibitor von ER β und PPAR γ . Der Wirkstoff Troglitazon beruht auf derselben Gerüststruktur und ist ein bekannter Regulator von PPAR γ . Bemerkenswerterweise wären auch einige aktive Moleküle aus einem Screening einer Substanzbibliothek mit 10000 Benzopyranen gegen FXR durch PSSC vorhergesagt worden. Zusätzlich ergab die Benzopyranbibliothek auch Liganden


 für die anderen Proteine des Clusters. Dieser Befund unterstreicht die Anwendung von PSSC bei der Entwicklung von Substanzbibliotheken.

Zunächst wurden Suchen nach struktureller Ähnlichkeit in der FSSP-Onlinedatenbank genutzt, um die Proteinstrukturähnlichkeitscluster zu definieren. Die FSSP-Datenbank basiert auf dem Vergleich der Faltungstypen eines nicht-redundanten Teils der Protein Data Bank (PDB) mithilfe des DALI-Algorithmus. Die resultierende Trefferliste wurde dann nach Ähnlichkeit und interessanten Treffern analysiert, z.B. Überlagerungen

mit großer Struktur-, aber niedriger Sequenzähnlichkeit. In einem abschließenden Validierungsschritt wurden die Ligand-erkennenden Kernstrukturen der Proteine des Clusters manuell extrahiert. Dabei handelt es sich um einen kugelförmigen Ausschnitt der Proteinstrukturen, dessen Zentrum die Bindungsstelle des jeweiligen Proteins bildet. Anhand der Überlagerung dieser Kernstrukturen wurde die strukturelle Ähnlichkeit visuell verifiziert. Allerdings führte diese Methode zu vielen falsch-positiven Überlagerungen, bei denen Teile des Proteins, die nicht in der Nähe der Bindungstasche liegen, sehr gut überlagern und so eine irreführend hohe Ähnlichkeit ergeben. In einer verbesserten Methode wurde daher zunächst die Ligand-erkennende Kernstruktur,^[31] d.h. die Subfaltung um die Bindungstasche herum, automatisch extrahiert. Anschließend wurden in der FSSP-Datenbank alle Strukturen gesucht, die strukturell ähnlich zu diesem „Ligand-erkennenden Kern“ sind. Dieses Verfahren stellt sicher, dass die Ähnlichkeit wirklich in dem Bereich um die Bindungstasche besteht und verringerte die Zahl an falsch-positiven Ergebnissen drastisch, wodurch fokussierte Folgeuntersuchungen möglich wurden (Abbildung 12).

Die erste Version der PSSC-Methode wurde verwendet, um neue Strukturtypen von Inhibitoren zu identifizieren. Dazu wurde die Gerüststruktur eines Naturstoffs und Inhibitors der Phosphatase Cdc25A genutzt, um neue Inhibitoren für die anderen Proteine der Gruppe zu finden, z.B. für Acetylcholinesterase (AChE) oder die 11 β -Hydroxysteroid-Dehydrogenase 1 (HSD1, Abbildung 13). Eine Substanzbibliothek von Hydroxybutenoliden, einem für die Inhibition von Phosphatasen wichtigen Strukturmotiv, ergab neue

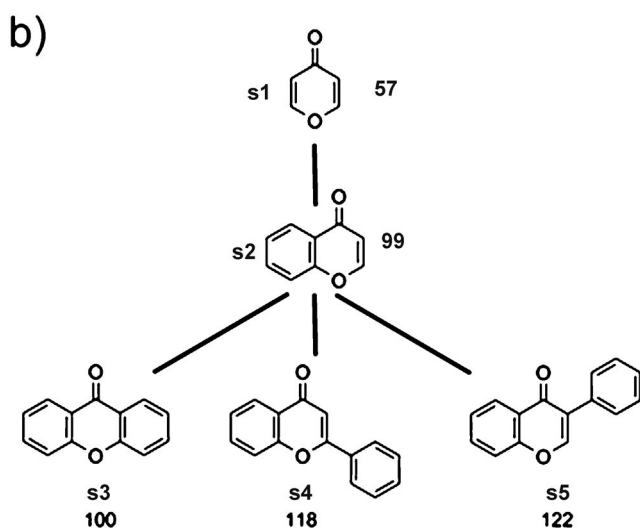
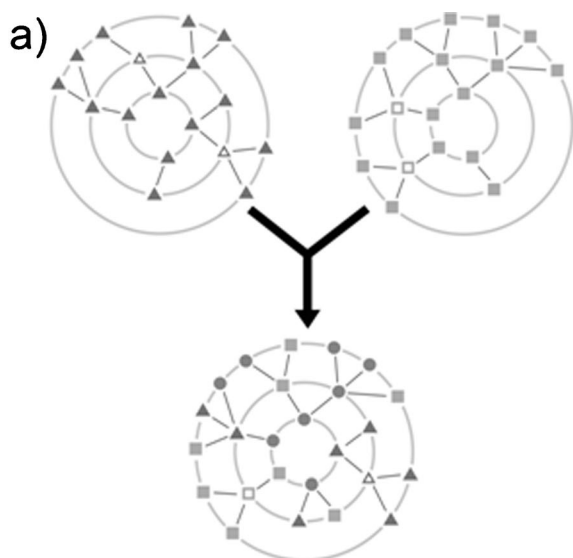


Abbildung 9. a) die Verschmelzung zweier Gerüststrukturbäume (Dreiecke und Quadrate) erzeugt einen neuen Baum. In diesem neuen Gerüststrukturbaum repräsentieren die Knoten Moleküle, die entweder aus beiden Bäumen (gefüllte Kreise), aus nur einem Baum (gefüllte Dreiecke/Quadrate) oder aus keinem Baum (umrissene Dreiecke/Quadrate) stammen. Annotationen, beispielsweise für Zielproteine, können direkt übertragen werden, sofern ein Knoten Moleküle aus beiden Bäumen repräsentiert, oder indirekt durch Brachiation. b) Eine γ -Pyrone-Bibliothek mit 500 Substanzen wurde aufgebaut, die fünf verschiedene Gerüststrukturen im Baum abdeckt.

Inhibitoren für AChE- und 11 β -HSDs (für eine detailliertere Diskussion dieser Ergebnisse siehe Abschnitt 3.2).^[32]

In einem neueren Beispiel wurde ein Proteinstrukturähnlichkeitscluster um gastrische Lipase gesucht, einem Enzym das durch Substanzen mit dem β -Lacton-Motiv moduliert wird, einschließlich des kommerziellen Wirkstoffs Tetrahydrolipstatin (Orlistat). Die Ähnlichkeitssuche mit der Ligand-erkennenden Kernstruktur der gastrischen Lipase ergab als strukturell ähnliches Protein unter anderem Acylproteinthioesterase 1 (APT1). Eine auf Tetrahydrolipstatin basierende Substanzbibliothek wurde synthetisiert und bio-

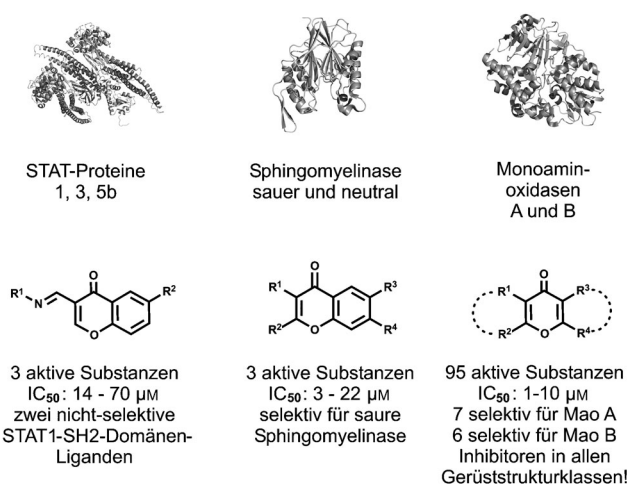


Abbildung 10. Proteinklassen, die in WOMBAT als Ziele des γ -Pyrone-Astes annotiert sind, und für die Inhibitoren gefunden wurden. Interessanterweise wurden aktive Substanzen für alle Proteine gefunden, die für nicht-optimierte Substanzen eine bemerkenswerte Selektivität und Wirksamkeit aufwiesen.

chemisch gegen APT1 getestet. Interessanterweise konnten mehrere Inhibitoren von APT1 mit IC₅₀-Werten im mikro- und nanomolaren Bereich identifiziert werden (für eine detailliertere Diskussion der Ergebnisse siehe Abschnitt 3.2).^[33]

Diese beiden Beispiele für die erfolgreiche Anwendung des PSSC legen nahe, dass die Methode in der Tat einen gangbaren Weg zur Identifikation neuer Protein-Ligand-Paare bietet. Für Schlussfolgerungen zur generellen Anwendbarkeit des PSSC oder über den Anwendungsbereich der Methode ist es in Anbetracht der geringen Zahl an momentan abgeschlossenen Studien allerdings zu früh.^[34]

3. Biologie-orientierte Synthese (BIOS)

Die strukturelle Klassifikation der Naturstoffe (SCONP) einschließlich ihrer Erweiterung auf Strukturklassen, die keine Naturstoffe sind, und PSSC sind zwei komplementäre Herangehensweisen, um biologisch relevante Strukturklassen im riesigen chemischen Strukturraum zu finden. Allein oder gemeinsam synergistisch angewendet bilden sie die Grundlage des von uns entwickelten Prinzips der Biologie-orientierten Synthese (BIOS) (Abbildung 14).^[5c,9b,35]

BIOS verlangt nach biologischer Relevanz als primäres Kriterium für die Auswahl von Substanzklassen und Gerüststrukturen. Diese inspirieren die Synthese von Substanzbibliotheken, die mit biologischer Aktivität angereichert sind. BIOS-basierte Substanzbibliotheken sind typischerweise nicht groß und müssen es auch nicht sein. Nach unserer Erfahrung ergibt das Screening solcher Bibliotheken Trefferquoten von 0.2–1.5 %, woraus sich eine Bibliotheksgröße von 200–500 Verbindungen ergibt, um Substanzen für die weitere Entwicklung zu erhalten. Die Synthese kann die Anwendung anspruchsvoller chemischer Methoden und mehrstufiger Sequenzen erfordern, besonders wenn Naturstoff-inspirierte Bibliotheken synthetisiert werden sollen. Diese Investition in

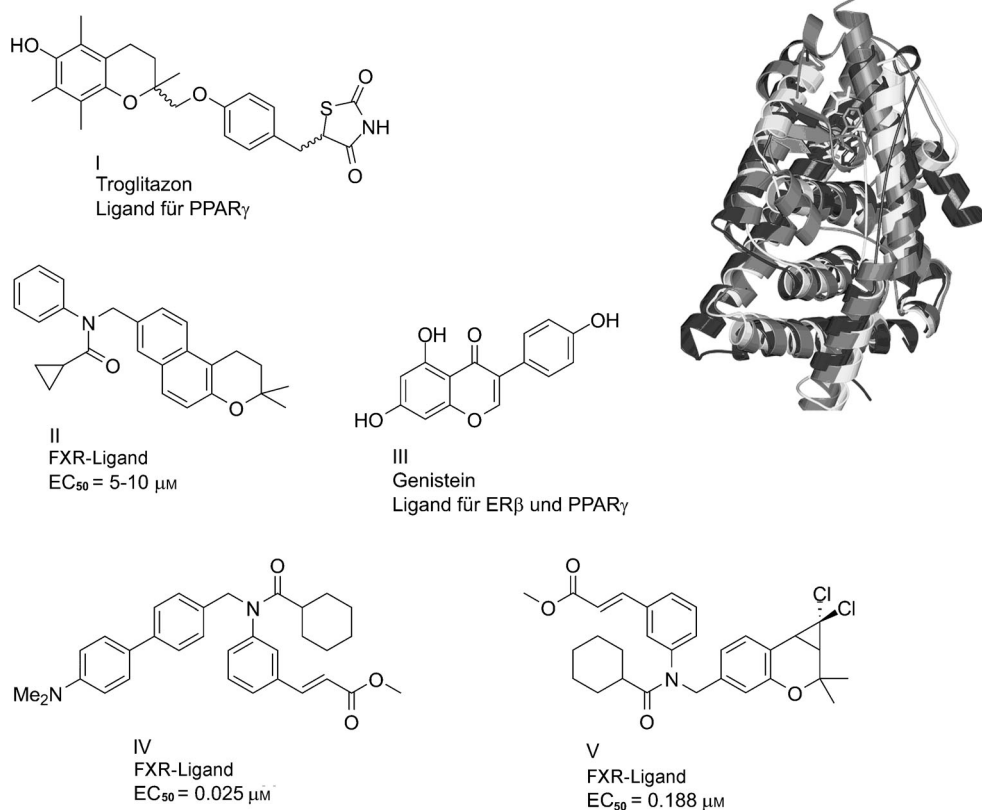


Abbildung 11. Proteinstrukturähnlichkeitscluster von ER β im Komplex mit Genistein (III, dunkelgrau), PPAR γ mit Rosiglitazon (grau) und Farnesoid-X-Rezeptor (hellgrau). Die Überlagerung der Ligand-erkennenden Kernstrukturen verdeutlicht die strukturelle Ähnlichkeit der Strukturen von Benzopyran-basierten Liganden für ER β , PPAR γ und FXR.



Abbildung 12. Die verbesserte PSSC-Methode. Die Ligand-erkennende Kernstruktur wird vor der Ähnlichkeitssuche extrahiert und statt der Gesamtstruktur für die Suche verwendet. Dadurch wird die Suche nach struktureller Ähnlichkeit auf den relevanten Bereich um die Bindungstasche fokussiert und so die Quote an falsch-positiven Ergebnissen deutlich gesenkt.

die chemische Entwicklung gleicht sich jedoch sehr gut durch die kleinere benötigte Bibliotheksgröße aus. BIOS führt zu relevanteren Substanzklassen, verlangt aber unter Umständen nach anspruchsvollerer Chemie.

Bedingt durch die Reduktion der strukturellen Komplexität im Vergleich zu den als Inspiration genutzten Naturstoffen können die initialen Treffer aus dem Screening von BIOS-Bibliotheken unselektiv sein und mehrere Proteine mit ähnlichen Ligand-erkennenden Kernstrukturen binden. Gleichzeitig können sie eine begrenzte Potenz aufweisen, z. B. IC₅₀-Werte im mikromolaren Bereich. Bei solchen Konzentrationen muss außerdem mögliche unspezifische Bindung in den Screens in Betracht gezogen werden. Dies erfordert eine gründliche Entwicklung der Screening-Bedingungen und Folgeexperimente einschließlich angemessener Kontroll-

experimente. Dabei handelt es sich allerdings um generelle Probleme, die häufig beim Screening und der Entwicklung von so genannten „tool compounds“ (Substanzen, die zur Erforschung biologischer Systeme eingesetzt werden) und frühen Wirkstoffkandidaten auftreten. Sie bedingen die weitere Entwicklung der initialen Treffer, um potente und selektive „tool compounds“ zu generieren.^[2a] Diese Notwendigkeit ist in der medizinisch-chemischen und chemisch-biologischen Forschung jedoch stets gegeben.

BIOS wurde ursprünglich auf der Basis der Strukturen von Naturstoffen entwickelt. Das Schlüsselkriterium für BIOS ist jedoch die biologische Relevanz, nicht das Vorkommen in der Natur. Daher schließt BIOS die Naturstoffe mit ein, ist aber nicht auf diese beschränkt, sondern umfasst auch weite Teile des übrigen chemischen Strukturraums. Insbesondere umfasst diese

Erweiterung auch die zahlreichen Verbindungsklassen, die aus mehr als 100 Jahren pharmazeutischer Entwicklung hervorgegangen sind und nicht zu den Naturstoffen gehören.

Sowohl SCONP- als auch PSSC-basierte Substanzbibliotheken beruhen auf struktur-basierten Argumenten. Bei den aus SCONP abgeleiteten Substanzbibliotheken lenken die Gerüststrukturen bioaktiver Moleküle die Entwicklung und Synthese, wobei sich die biologische Relevanz vom biosynthetischen

Ursprung und der biologischen Aktivität herleitet. Gemeinsame Proteinstrukturen bilden die Basis für PSSC-basierte Bibliotheken und belegen deren biologische Relevanz. Beide Methoden für sich alleine genügen, um BIOS zu inspirieren und können als Hypothesengenerator dienen.

Darüber hinaus können SCONP und PSSC gemeinsam auf synergistische Weise angewendet werden, wobei sie sich gegenseitig bestärken. Die Entwicklung von 11 β HSD1- und APT1-Inhibitoren, die oben kurz erwähnt wurden und unten detaillierter diskutiert werden, sind überzeugende Beispiele für die Leistungsfähigkeit der Methode.

Die synergistische Anwendung wird oft durch die fehlenden Daten für Bioaktivität erschwert, besonders bei Naturstoffen, sowie durch fehlende Proteinkristallstrukturen,

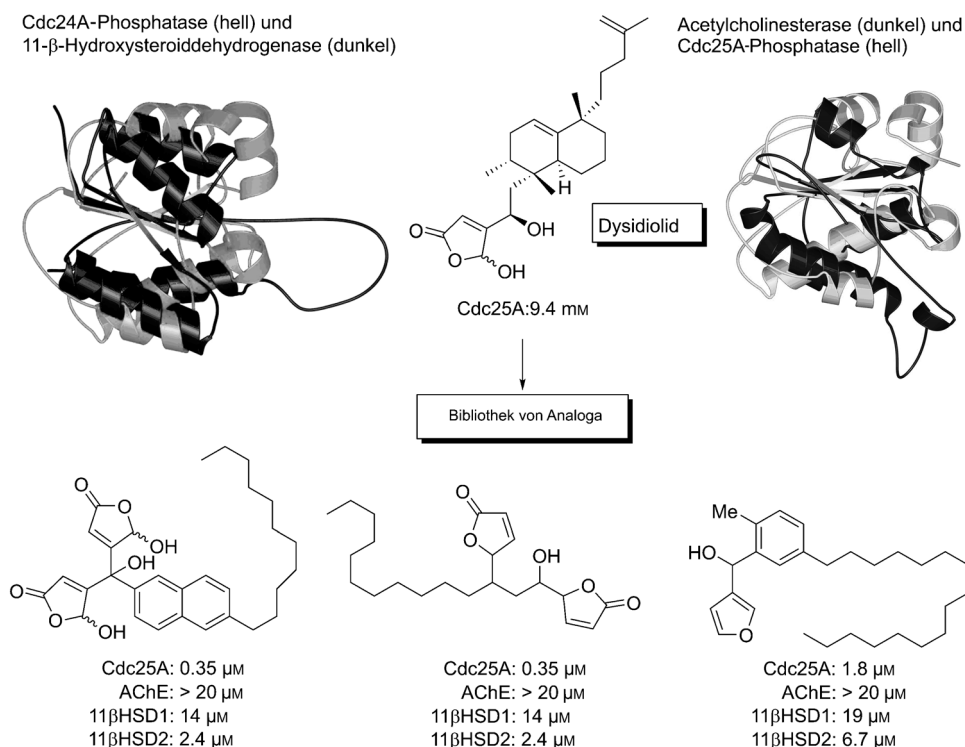


Abbildung 13. Cluster der Proteine HSD, Cdc25A und AChE sowie der zugehörigen aktiven Substanzen aus einer Dysidiolid-inspirierten Substanzbibliothek.

insbesondere mit gebundenen Liganden oder Inhibitoren. Das Fehlen von Proteinkristallstrukturen fällt besonders ins Gewicht, wenn biosynthetische Argumente zur Unterstüt-

zung der BIOS-Methode genutzt werden sollen. Prinzipiell sollte die strukturelle Ähnlichkeit zwischen Proteinen der Naturstoffbiosynthese und anderen Proteinen auf mögliche Zielproteine für diese Naturstoffe hinweisen.^[36] So zeigte ein Vergleich der Proteinfaltungstopologie (PFT) der Enzyme Chalcon-synthase (CHS), Chalconiso-merase (CHI) und Anthocy-anidinsynthase (AND), die Schlüsselschritte in der Bio-synthese von natürlich vor-kommenden Chalconen und Flavanoiden katalysieren, Ähnlichkeit mit dem katalyti-schen Zentrum der Phospho-inositid-3-kinase (PI3K). Obwohl CHS, CHI, ANS und PI3K ihrem Faltungstyp nach als sehr verschieden gelten, fanden sich ähnliche Anord-nungen von verschiedenen Sekundärstrukturen, d. h. PFTs. In der Tat waren Chalcone unter den ersten gefunden Klassen von Kinase-inhibitoren.^[37] Dieser Ansatz stellt eine Verbindung zwischen den biosynthetischen Enzymen eines Naturstoffs und seinen

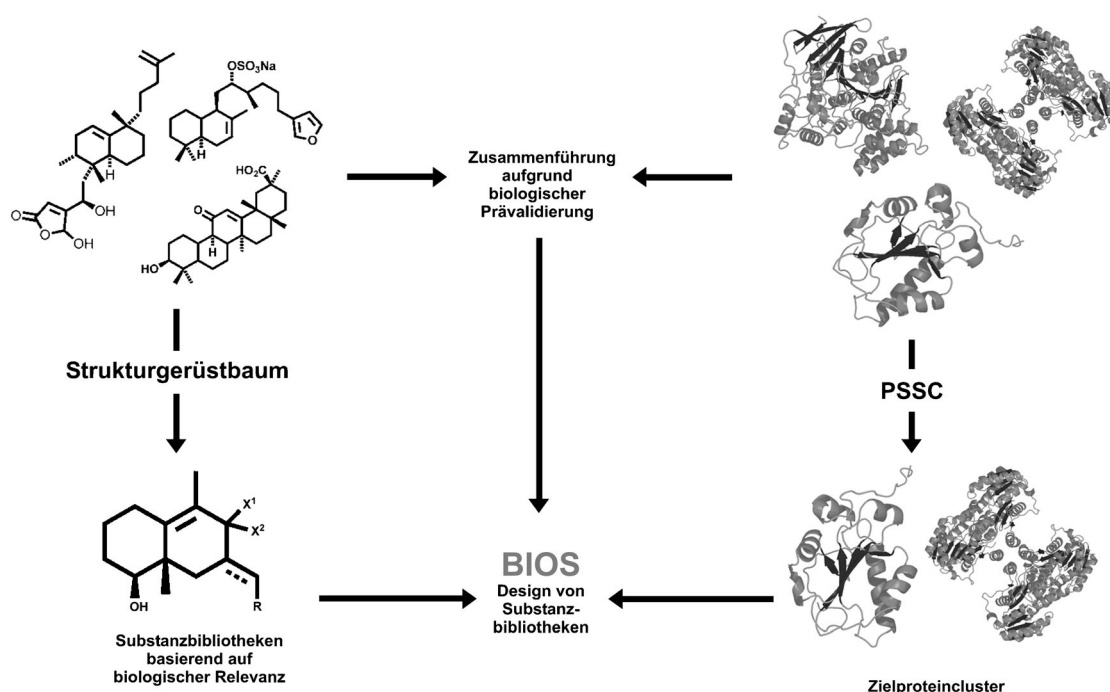


Abbildung 14. BIOS verbindet den chemischen und biologischen Raum, d. h. Proteinstrukturähnlichkeitscluster und Substanzbibliotheken, durch biologische Vorvalidierung. Dabei geht es weit über Naturstoffe hinaus und schließt alle Substanzklassen mit bekannter biologischer Relevanz mit ein.

möglichen Zielproteinen her. Das lückenhafte Wissen über viele biosynthetische Enzyme und ihre Strukturen begrenzt allerdings die Anwendbarkeit dieser Methode.

3.1. BIOS in der Entwicklung von Substanzbibliotheken

Eine Nutzung von BIOS für medizinisch-chemische und chemisch-biologische Forschung verlangt nach der Synthese von Substanzbibliotheken basierend auf biologisch relevanten Strukturgerüsten. Naturstoffe stellen eine große Quelle von bioaktiven Molekülen dar. Der begrenzte Zugang zu diesen Molekülen aus natürlichen Quellen oder durch synthetische und semi-synthetische Verfahren schränkt die weitere Nutzung in den Biowissenschaften aber oft ein. Es entsteht daher ein Bedarf an Synthesen komplexer naturstoffähnlicher Moleküle in ausreichender Menge und Zahl, der nach der Entwicklung neuer, mit den Randbedingungen der Bibliothekssynthese kompatibler Strategien und Methoden verlangt. Ein synergistischer Ansatz zur Nutzung der Leistungsfähigkeit aktueller organischer Synthesen und der Technologie der kombinatorischen und parallelen Synthese wird benötigt, um fokussierte Bibliotheken auf der Basis von Kerngerüststrukturen von Naturstoffen und anderer biologisch relevanter Strukturtypen zu synthetisieren. Besonders begehrt und schwierig sind chemische Reaktionen und Reaktionssequenzen (im Bezug auf hohe Gesamtausbeuten und weniger einzelner Reaktionsschritte), die leicht verfügbare Substrate nutzen und komplexe Naturstoff-basierte Strukturen aufbauen. Dieser Herausforderung wurde beispielsweise durch die kürzlich entwickelten Mehrkomponentenreaktionen, Kaskaden- und Dominoreaktionen, multikatalytische Eintopfreaktionen und asymmetrische Festphasensynthesen zur Herstellung Naturstoff-ähnlicher Moleküle begegnet.^[38] Angesichts der verschiedenen Ringsysteme und Kerngerüststrukturen in Naturstoffen ist die Wahl eines bevorzugten Ringsystems als Ziel für die Bibliothekssynthese oft alles andere als trivial. Statistische Analysen der Gerüststrukturen der Naturstoffe im DNP zeigten, dass mehr als die Hälfte aller Naturstoffe in der Datenbank mit einem Molekulargewicht unter 1000 g mol^{-1} zwei, drei oder vier Ringe enthalten. Systeme mit zwei bis vier Ringen sind also gute Startpunkte für die Entwicklung naturstoffbasierter Substanzbibliotheken (Abbildung 15). Bei der Synthese sollte bedacht werden, dass bei Naturstoff-abgeleiteten Substanzbibliotheken die Gerüststruktur identisch mit der des jeweiligen Naturstoffs ist, wohingegen in Naturstoff-inspirierten Bibliotheken die Gerüststruktur möglicherweise nicht identisch, aber nahe verwandt zur Struktur des zugrundeliegenden Naturstoffs ist. Die Gerüststruktur wird typischerweise durch De-novo-Synthese aufgebaut, wobei das Einführen von Substituenten sowie die Variation der Substitutionsmuster und der Stereochemie ermöglicht werden (siehe Abbildung 1). Im Folgenden fassen wir ausgewählte Synthesen Naturstoff-inspirierter Substanzkollektionen zusammen. Für einen aktuellen Überblick über das Gebiet empfehlen wir dem Leser umfassendere Übersichtsartikel.^[39]

Eine Naturstoff-inspirierte Synthese wurde entwickelt, um biologisch aktive Analoga des natürlich vorkommenden

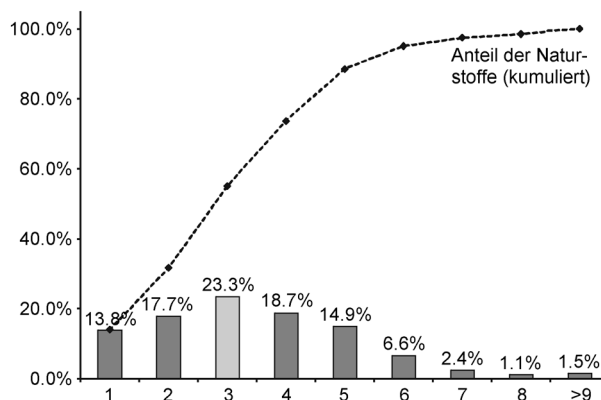
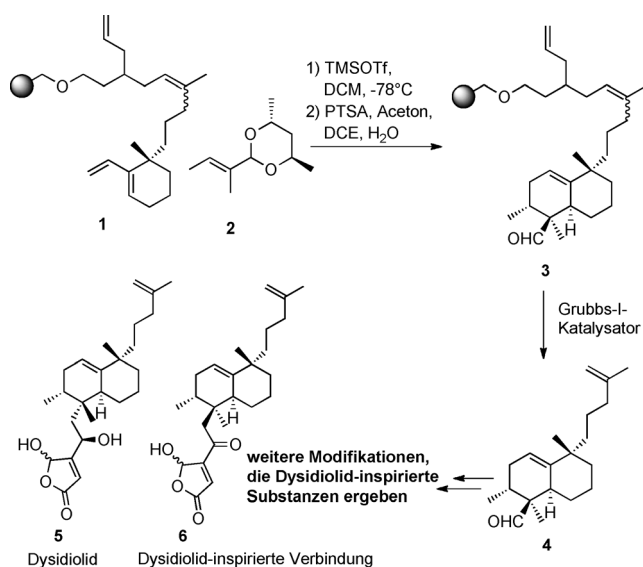


Abbildung 15. Vorkommen von Gerüststrukturen mit verschiedenen Zahlen an Ringen in Naturstoffen. 20.8% aller Naturstoffe enthalten drei Ringe und markieren das Maximum der Verteilung. Die Gerüststrukturen mit zwei und vier Ringen liegen innerhalb einer Standardabweichung, und 52.8% aller Naturstoffe enthalten Gerüststrukturen mit zwei, drei oder vier Ringen.

Phosphataseinhibitors Dysidiolids (**5**; Schema 1) zu identifizieren. Dafür wurde das chirale Dienophil **2** eingesetzt, um den stereodirigierenden Einfluss des an die feste Phase gebundenen chiralen Diens **1** zu verstärken.^[40] Das bicyclische Gerüst **3** wurde durch eine Diels-Alder-Reaktion des Diens **1** mit dem von Tigraldehyd abgeleiteten Acetal **2** aufgebaut, wobei ein *endo/exo*-Verhältnis von 91:1 und eine Selektivität von 95:5 zugunsten des gewünschten *endo*-Isomers beobachtet wurden. Das Cycloaddukt **4** wurde durch Ringschlussmetathese (RCM) freigesetzt. Weitere Modifikationen des Cycloaddukts **4** ergaben Dysidiolid-Analoga. Die biologische Untersuchung der fokussierten, kleinen Substanz-



Schema 1. Festphasensynthese von Dysidiolid-inspirierten Verbindungen. TMS = Trimethylsilyl, DCM = Dichlormethan, PTSA = *p*-Toluolsulfonsäure, DCE = Dichlorethan.

bibliothek ergab mehrere Phosphataseinhibitoren mit cytotoxischer Aktivität gegen verschiedene Krebszelllinien, wobei das Dysidiolid-ähnliche Molekül **6** als potentester Inhibitor der Phosphatase Cdc25A mit einem IC_{50} -Wert von $0.8 \mu\text{M}$ identifiziert wurde.

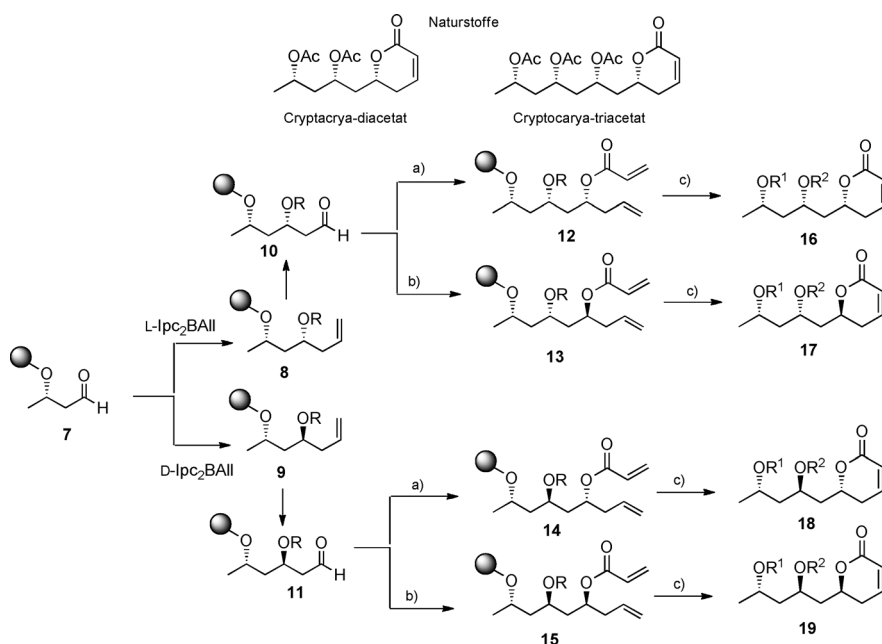
Die Synthese einer Substanzbibliothek an der festen Phase verlangt oft nach einer Anpassung bekannter Synthesemethoden an ein Format zur Bibliothekssynthese. Entwicklungen bei der asymmetrischen Festphasensynthese haben beispielsweise die Herstellung Naturstoff-inspirierter Substanzbibliotheken ermöglicht.^[38b] Ein eindrucksvolles Beispiel ist die enantioselektive Carbonylallylierung – eine der bedeutendsten Umwandlungen funktioneller Gruppen – für die stereoselektive Festphasensynthese einer Bibliothek Naturstoff-inspirierter δ -Lactone (Schema 2).^[41] Die Synthe-

Das all-*syn*-Isomer aus *Cryptocarya diacetat* wurde in einer Ausbeute von 11 % mit einfacher Flash-Chromatographie isoliert. Alle möglichen acht Stereoisomere des δ -Lactons wurden mit dieser Reaktionssequenz hergestellt, wobei die Allylierungsreaktion stereokomplementär verwendet wurde. Die Anpassung einer asymmetrischen Reaktion an die feste Phase ist oft nicht einfach. Dennoch zeigt das Beispiel des Lactons, dass die heutigen Synthesemethoden im Prinzip die Synthese aller Stereoisomere eines Naturstoffs erlauben.

Ein weiteres Beispiel der asymmetrischen Festphasensynthese bilden stereokontrollierte Aldolreaktionen, mit deren Hilfe eine Naturstoff-inspirierte Substanzbibliothek von Spiroacetalen aufgebaut wurde. Substanzen mit Spiroacetalen kommen in der Natur häufig vor und sind für ihre diverse biologische Aktivität bekannt. Insbesondere das

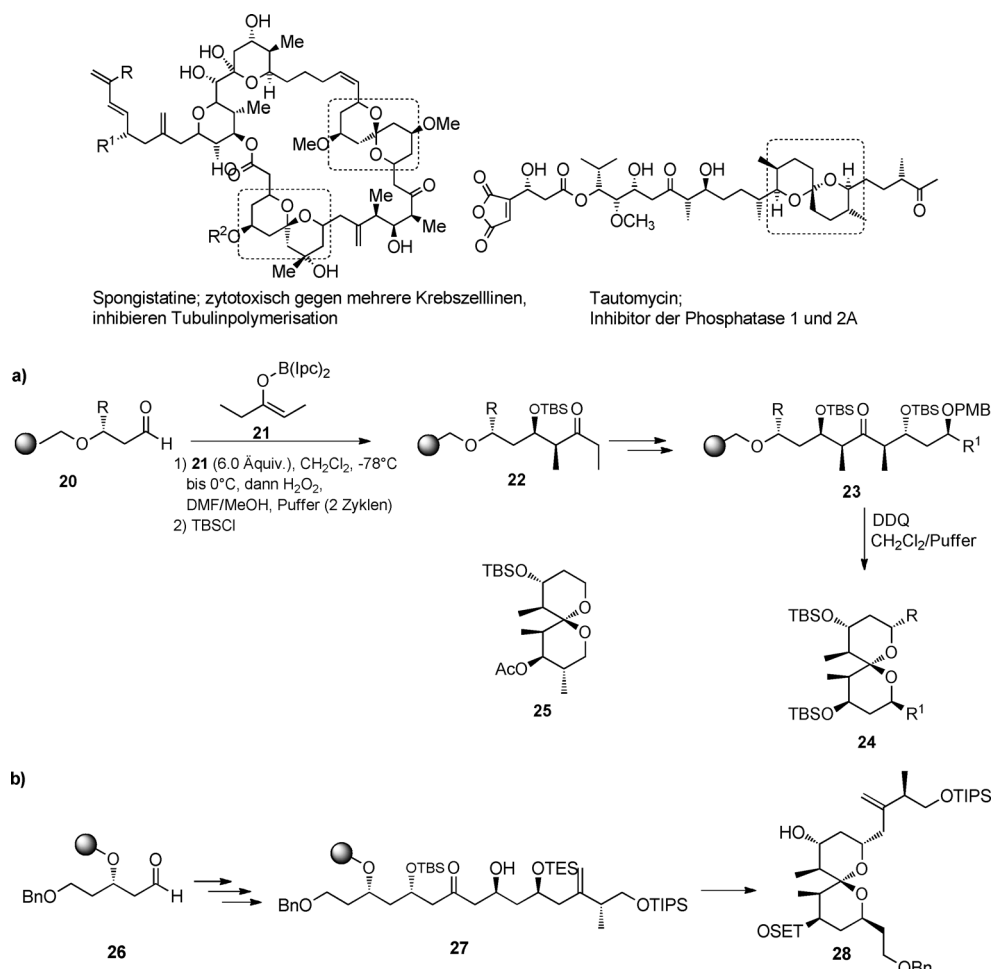
rigide Spiro[5.5]ketal-Ringsystem findet sich als Fragment in vielen komplexen Naturstoffen mit einer breiten Palette an biologischer Aktivität.^[42] Die außergewöhnlich potenten Spongistatine, welche die Tubulin-Polymerisation inhibieren, enthalten ein Spiroacetal-Fragment in ihrem macrocyclischen Gerüst, ebenso wie der Proteinphosphatase-Inhibitor Tautomycin.^[43] Zur Untersuchung der biologischen Funktionen des Spiroacetal-motivs, einschließlich der mit dem zugrunde liegenden Naturstoff vergleichbaren Bioaktivität, wurde eine Festphasensynthese von Spiroacetalen mit der asymmetrischen Aldolreaktion als Schlüsselschritt entwickelt. Um zu Spiro[5.5]ketalen zu gelangen, wurde eine Aldolreaktion des festphasengebundenen Aldehyds **20** mit dem vorab gebildeten *Z*-Borenolat **21** durchgeführt, die das enantiomerenangereicherte Aldodaddukt **22** ergab. Im Unterschied zur Synthese in flüssiger Phase benötigte die Reaktion an fester Phase zwei Wiederholungen mit sechs Äquivalenten des chiralen Re-

gens **21** bis zur vollständigen Umsetzung des Aldehyds. Eine weitere *anti*-selektive Aldolreaktion an der festen Phase mit dem *E*-Borenolat ergab die geschützten Bis- β -hydroxyketone **23** als fortgeschrittene Vorstufen der finalen Spiroacetale **24**. Gleichzeitige Entschützung der PMB-Gruppe und Acetalbildung wurden durch oxidative Spaltung mit DDQ erreicht, wobei das Spiroketal **24** freigesetzt wurde. Das Verhältnis der Diastereomere des Produkts zeigt, dass die zweite Aldolreaktion in den Fällen mit gleicher Stereochemie („matched case“) lediglich ein Diastereomer des Spiroacetals **24** ergibt, während in den Fällen mit unterschiedlicher Stereochemie („mismatched case“) eine geringere Stereoselektivität beobachtet wurde. Für das Spiroacetal **25** (Schema 3) aus der Substanzbibliothek wurde die Inhibitor der Phosphatasen VHR und PTP1b mit IC_{50} -Werten von $6 \mu\text{M}$ und $39 \mu\text{M}$ nachgewiesen.



Schema 2. Synthese stereoisomerer δ -Lactone durch asymmetrische Allylierung von Aldehyden an fester Phase als Schlüsselschritt. a) 1. L-Ipc₂Ball, 2. Acrylierung; b) 1. D-Ipc₂Ball, 2. Acrylierung; c) 1. Grubbs-Katalysator der 2. Generation, 2. Abspaltung von der festen Phase.

seroute zu den Naturstoff-Analoga beinhaltet mehrere stereokomplementäre Allylierungsreaktionen am polymeren Träger gefolgt von einer Ringschlussmetathese. Daher wurden vor der Bibliothekssynthese die Reaktionsbedingungen für die Allylierung immobilisierter Aldehyde mittels B-allyl(diisopinocampheyl)boran (Ipc₂Ball) in hoher Ausbeute und Enantioselektivität optimiert. Die Allylierung des polymergebundenen Aldehyds **7** mittels L-Ipc₂Ball ergab polymergebundenes **8** in einem *syn:anti*-Verhältnis von 85:15. Vorsichtige Ozonolyse der Doppelbindung ergab Aldehyd **10**. Dieser wurde einer zweiten Allylierung mit L-Ipc₂Ball unterworfen, und der entstandene sekundäre Alkohol in den Acrylsäureester **12** umgewandelt. Ringschlussmetathese mit dem Grubbs-II-Katalysator ergab das α,β -ungesättigte Lacton **16**, das vom polymeren Träger abgespalten (TFA, Trifluoressigsäure) und acetyliert wurde.



Schema 3. Synthese Naturstoff-inspirierter Spiroacetale. TBS = *tert*-Butyldimethylsilyl, DDQ = 2,3-Dichlor-5,6-dicyan-1,4-benzochinon, TIPS = Triisopropylsilyl, PMB = *p*-Methoxybenzyl, TES = Triethylsilyl, Bn = Benzyl.

Darüber hinaus st\u00f6rte die Verbindung **25** den geordneten Aufbau des Mikrotubulinnetzwerks in einer humanen Krebszelllinie.^[44]

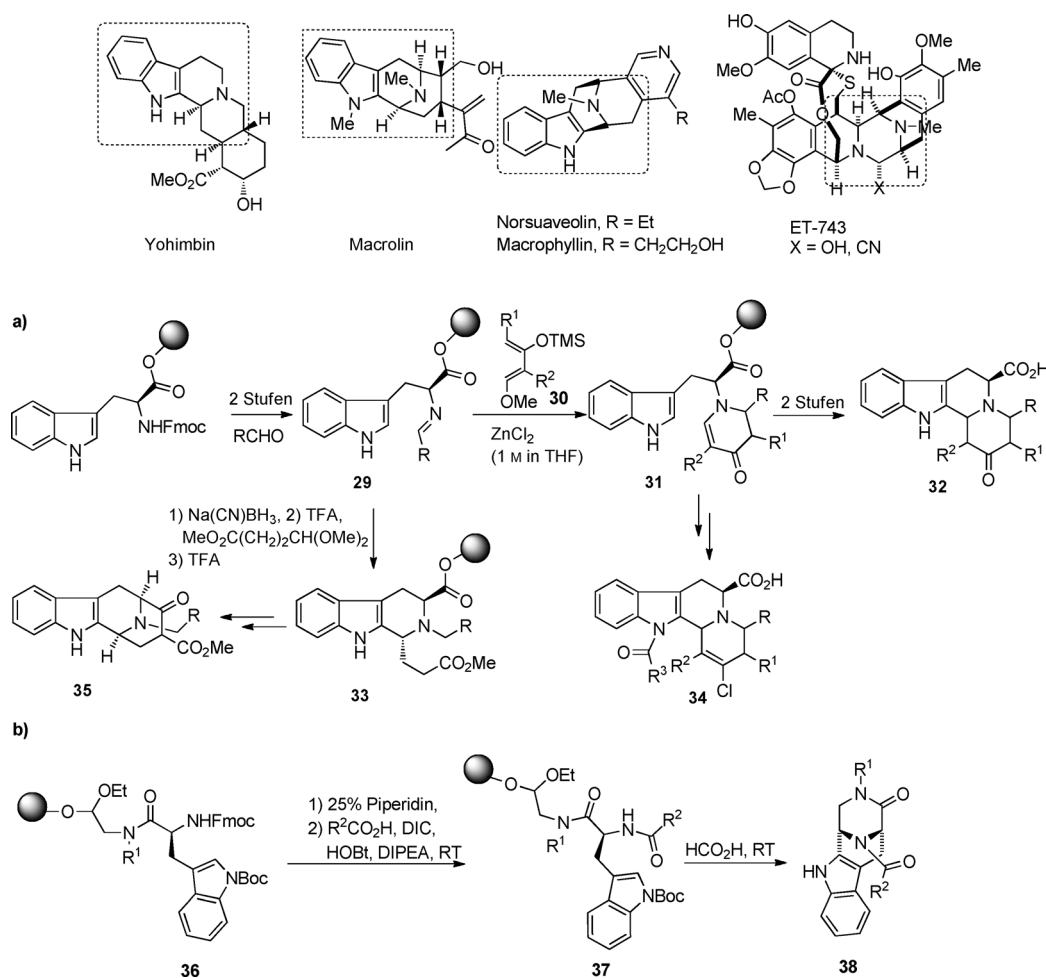
Auf \u00e4hnliche Weise wurde ein Fragment des Naturstoffs Spongistatin, der die Spiroketalstruktur **28** enth\u00e4lt, an der festen Phase hergestellt.^[45] Dazu wurde der immobilisierte β -Hydroxyaldehyd **26** zwei aufeinander folgenden stereoselektiven Aldolreaktionen unterworfen, die das Bis- β -hydroxyketon **27** ergaben. Aus dem gesch\u00fctzten Polyol **27** ging das Spiroacetal **28** durch Abspaltung von der festen Phase und In-situ-Cyclisierung hervor.

Diese Beispiele verdeutlichen, wie BIOS zur Identifikation strukturell einfacherer Startpunkte f\u00fcr die Entwicklung von Substanzbibliotheken beitr\u00e4gt und gleichzeitig zu neuen Inhibitor Klassen f\u00fchrt. In einem weiteren Versuch, den Strukturraum der Naturstoffe mit BIOS zu erkunden, wurden Indolalkaloidger\u00fcste als Zielstrukturen ausgew\u00e4hlt. Die Wahl basiert auf den strukturell komplexen Alkaloiden Yohimbin und Ajamalicin, die beide die Proteinphosphatase Cdc25A inhibieren. Analyse mit SCONP sowie Brachiation (= Bewegung entlang eines Astes) entlang der nat\u00fcrlichen Pr\u00e4validierungslinie f\u00fchrten zu tetracyclischen Indolo[2,3-*a*]chinolizidinen (Schema 4). Eine Festphasensynthese von Indolo-

[2,3-*a*]chinolizidinen ergab 450 Verbindungen in sechs bis acht Reaktionsschritten. Der Syntheseentwurf basiert auf der Lewis-S\u00e4ure-vermittelten Mannich-Michael-Reaktion zwischen immobilisierten D- oder L-Tryptophanaminen **29** und elektronenreichen Silyloxydienen **30**. Die Enaminonprodukte **31** wurden anschlie\u00dfend durch Behandlung mit S\u00e4ure oder Phosgen cyclisiert, wodurch tetracyclische Ketone und Vinylchloride entstanden. Diese wurden weiter derivatisiert sowie basen- oder s\u00e4urekatalysiert vom polymeren Tr\u00e4ger abgespalten und ergaben die Indolochinolizidine **32** und **34** (Schema 4) in hohen Gesamtausbeuten.^[46] Die Indolochinolizidin-Bibliothek enth\u00e4lt zwei Cdc25A-Inhibitoren mit IC₅₀-Werten vergleichbar dem des Naturstoffs. Die Tryptophanamine **29** wurden ebenfalls genutzt, um eine Macrolin-inspirierte Substanzbibliothek^[47] zu synthetisieren, bestehend aus tetracyclischen Indol-Derivaten **35** mit einem gemeinsamen Cycloocta-

[*b*]Indol-Ger\u00fcst. Die Reduktion des Imins **29** gefolgt von einer Pictet-Spengler-Reaktion mit Methyl-4,4-dimethoxybutyrat f\u00fchrte zu den 1,3-*trans*- β -Carbolinen **33**. Die zur Herstellung des tetracyclischen Ger\u00fcsts notwendige 1,3-*cis*-Umlagerung wurde durch Abspaltung von der festen Phase und regioselektive Epimerisierung unter basischen Bedingungen ausgel\u00f6st. Das dabei entstehende *cis*-Isomer durchl\u00e4uft eine Dieckmann-Cyclisierung zum β -Ketoester **35**. Die so hergestellte Macrolin-inspirierte Substanzbibliothek mit ca. 100 Verbindungen ergab potente Inhibitoren der mycobakteriellen Phosphatase MptpB.

Eine andere Naturstoff-inspirierte Synthese zielte auf das Diaza-verbr\u00fcckte tetracyclische Indolger\u00fcst ab, das Bestandteil vieler Alkaloide ist (Schema 4).^[48] Das marine Alkaloid Yondelis (ET-743, Schema 4), das eine Diaza-verbr\u00fcckte Ger\u00fcststruktur enth\u00e4lt, erhielt von der FDA 2005 den Status als einzig bekannter Wirkstoff („orphan drug status“) f\u00fcr die Behandlung von Ovarialkrebs in den USA. Obwohl viele Naturstoffe Diaza-verbr\u00fcckte Ringsysteme enthalten, schr\u00e4nkten das seltene Vorkommen in der Natur sowie die strukturelle Komplexit\u00e4t die Entwicklung als Krebstherapeutikum ein. Um Zugang zu Substanzklassen mit Diaza-verbr\u00fcckten Ringsystemen zu gewinnen, die m\u00f6gli-

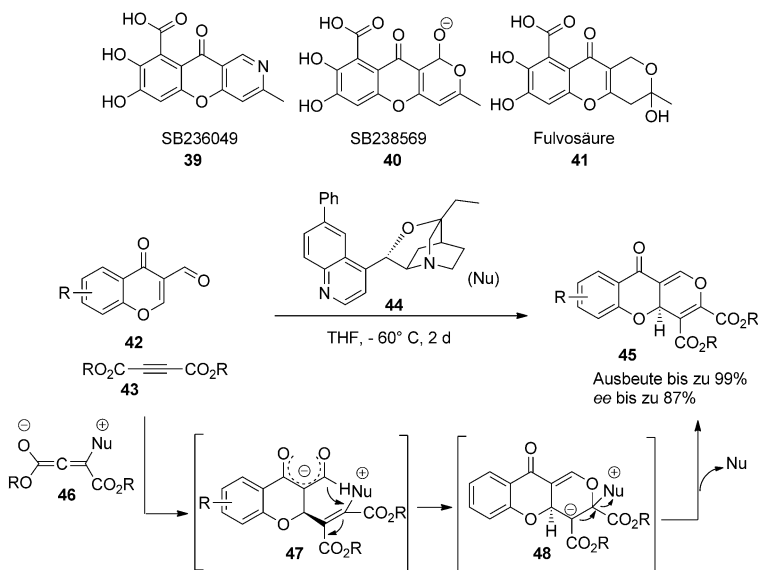


Scheme 4. Synthese polycyclischer Alkaloid-inspirierter Substanzbibliotheken. Fmoc = 9-Fluorenylmethoxycarbonyl, DIC = Diisopropylcarbodiimid, DIPEA = Diisopropylethylamin, Boc = *tert*-Butyloxycarbonyl, HOBT = 1-Hydroxybenzotriazol.

cherweise breite biologische Aktivität zeigen (Schema 4b),^[49] wurde das immobilisierte Tryptophanacetal **36** entschützt und acetyliert, sodass die Cyclisierungsvorstufe **37** entsteht. Die engültige regio- und diastereoselektive Cyclisierung wurde in reiner Ameisensäure durchgeführt, wobei simultan die Abspaltung von der festen Phase und die Pictet-Spengler-Cyclisierung über in situ generierte Iminiumionen erfolgte. Die Diaza-verbrückten Moleküle **38** wurden als einzelne Diastereome in hohen Ausbeuten und mit hoher Reinheit erhalten. Aus Fmoc-geschütztem Tryptophan und Fmoc-geschütztem (o-diTBS)DOPA als Substrat wurde eine Bibliothek mit 384 Substanzen mit 3.9-Diazabicyclo-

[3.3.1]non-6-enon-Gerüsten hergestellt, die mit Indol und Dihydroxybenzol verbrückt sind und an den verbrückenden Stickstoffatomen diversifiziert wurden.

Die Untersuchung von tricyclischen Benzopyranen, in denen der Benzopyranring mit weiteren Heterocyclen (**39** und **40**) verbrückt ist, zeigte inhibitorische Aktivität gegen Metallo-β-lactamasen und folglich potenzielle antibakterielle Aktivität gegen wirkstoffresistente Bakterienstämme.^[50] Inspiriert durch diese Naturstoffe wurde eine neue [4+2]-Anellierungsroute zum tricyclischen Kerngerüst der Benzopyrane entwickelt, um eine fokussierte Bibliothek tricyclischer Benzopyrane (**45**, Schema 5) herzustellen.^[51] Die Anellierung zweier elektronenarmer Systeme, in diesem Fall des Oxadiens **42** und des Acetylen-carboxylats **43**, wurde durch nucleophile Katalyse mit tertiären Phosphinen oder Aminen möglich. Das Zwitterion **46**, das durch die Addition eines Organokatalysators **44** an die Alkine **43** entstand, durchlief eine Reaktionssequenz aus Michael-Addition/Michael-Addition/Eliminierung, die zur gewünschten Zielstruktur führt. Cinchona-abgeleitete β-Isocupre-



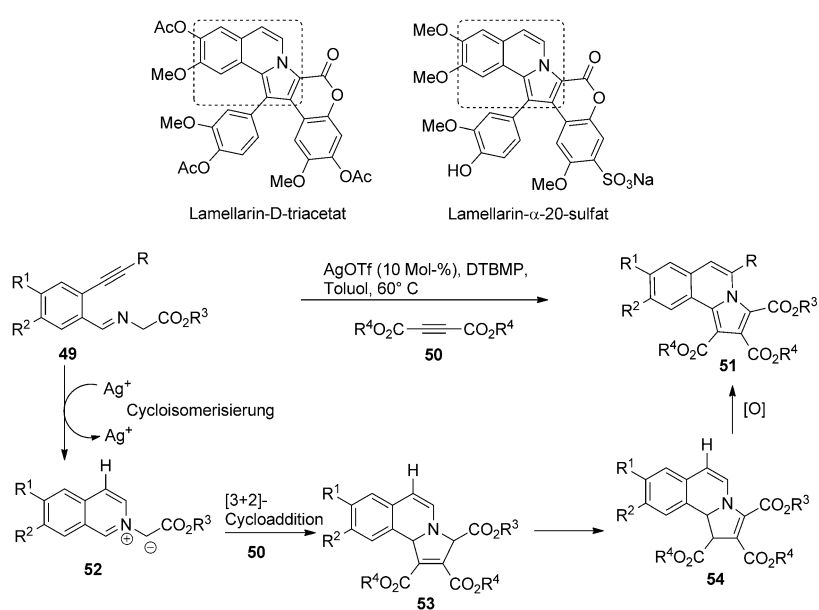
Scheme 5. Synthese Naturstoff-inspirierter tricyclischer Benzopyrane.

dine als Katalysatoren eröffneten eine enantioselektive Route zu (*S*)-**45**.

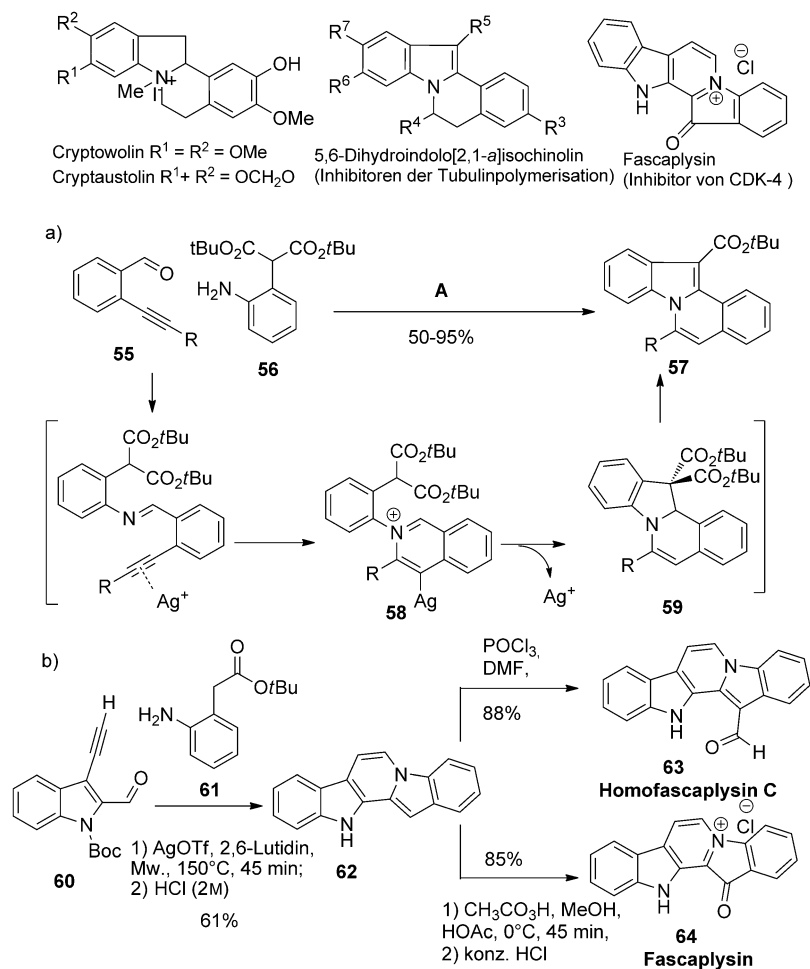
Die Synthese Naturstoff-inspirierter Substanzbibliotheken setzt oft mehrstufige Synthesesequenzen voraus, die Naturstoff-ähnliche Komplexität aufbauen. Diese Voraussetzung erschwert in vielen Fällen die Synthese von mittelgroßen oder großen Substanzbibliotheken und verlangt nach der Entwicklung von Reaktionen, die Komplexität erzeugen und schnell und effizient komplexe Naturstoff-basierte Gerüststrukturen aufbauen.^[52] Kaskaden- oder Dominoreaktionen können effiziente Lösungen zur Meisterung dieser Herausforderung sein. So nutzt beispielsweise eine effiziente Synthese von Pyrroloisochinolonen eine Dominoreaktion.^[53] Die Pyrroloisochinolone sind mit den Lamellarin-Alkaloiden verwandt (Schema 6), einer Familie von marinen Naturstoffen, die ein hochsubstituiertes Pyrroloisochinolongerüst enthalten und unter anderem humane Topoisomerase I und HIV-1-Integrase inhibieren.^[54] Die Ag^I-katalysierte Cycloisomerisierung von Aklinyln-benzylidenglycinaten **49** zu einem Azomethinylid **52** gefolgt von einer dipolaren Cycloaddition an das Acetylendicarboxylat ergab die Zwischenstufe **53**. Isomerisierung und nachfolgende oxidative Aromatisierung ergab das Pyrroloisochinolon **51** in einer effizienten Eintopfsynthese.

Auf ähnliche Weise wurde kürzlich eine Kaskadenreaktion, die Silber-katalysierte Cycloisomerisierung von Acetylenaldehyden als Schlüsseltransformation nutzt, bei der Synthese von diversen Alkaloidringsystemen eingesetzt.^[55] Indoloisochinolone, eine medizinisch bedeutende Substanzklasse mit antikanzerogenen Eigenschaften, wurden einfach und effizient mittels solcher Kaskadenmethoden hergestellt. Dazu werden das aus einem Acetylenbenzaldehyd hergestellte Imin und ein Anilin mit einem entsprechenden Nucleophil (**56**) in der entscheidenden Silber-katalysierten Cycloisomerisierung unter Mikrowellenbedingungen zu den Isochinoliniumkationen **58** umgesetzt. Der nucleophile Angriff des entsprechenden Nucleophils auf das Iminiumkation führt zur Zwischenstufe **59**, die nach decarboxylierender Aromatisierung die Zielstruktur Indolo-[2,1-*a*]isochinolon (**57**; Schema 7a) in guten Ausbeuten ergibt.

Nach dieser Methode wurden die marinen Naturstoffe Homofascaplysin C und der CDK-5-Inhibitor Fascaplysin synthetisiert (Schema 7b). Ausgehend von Boc-geschütztem 3-Ethynylindol-2-carbaldehyd **60** als ge-



Schema 6. Kaskadensynthese Lamellarin-inspirierter Moleküle. DTBMP = 2,6-Di(*tert*-butyl)-4-methylpyridin.



Schema 7. Kaskadensynthese Alkaloid-basierter Substanzkollektionen.

meisamer Vorstufe und dem Anilin **61** führt die Mikrowellen-unterstützte Silber-katalysierte Kaskadencyclisierung zum pentacyclischen Gerüst **62**. Die Formylierung von **62** mit POCl_3 führt zu reinem Homofascaplysin C, wohingegen die Oxidation der pentacyclischen Kernstruktur **62** mit Peressigsäure und nachfolgende Behandlung mit Säure effizient Fascaplysin ergaben.

Ausgewählte Gerüststrukturen weiterer Substanzbibliotheken, die von Naturstoffen abgeleitet wurden und Inhibitoren oder molekulare Sonden für biologische Anwendungen hervorgebracht haben, sind in Abbildung 16 zusammenge-

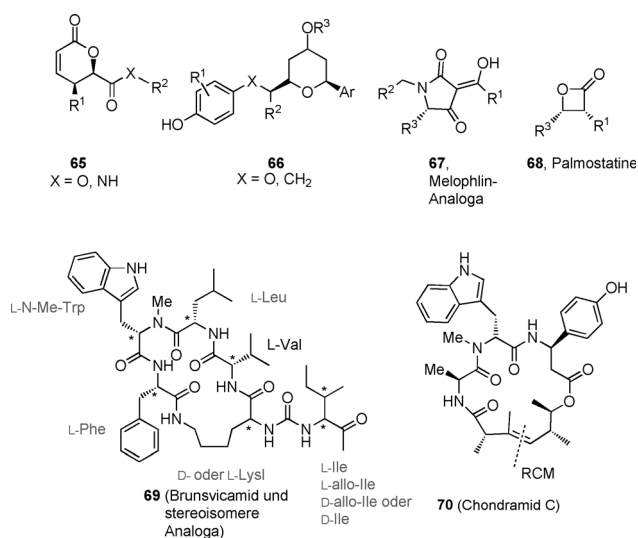


Abbildung 16. Auf der BIOS-Methode basierende Substanzbibliotheken.

fasst. Der Weg zu enantiomerenreinen Naturstoff-abgeleiteten α,β -ungesättigten Lactonen führt über eine Hetero-Diels-Alder-Reaktion Sauerstoff-substituierter Diene mit einem Glyoxylat in Gegenwart eines chiralen Titankatalysators und ergibt die gewünschten Dehydrolactone mit hohen Enantiomeren- und Diastereomerenüberschüssen. Die biologische Prüfung der Substanzen in zellulären Assays identifizierte neue Modulatoren des Zellzyklus und Inhibitoren des Eindringens von Viren in Zellen.^[56] In einem weiteren Projekt wurde die Hetero-Diels-Alder-Reaktion eines festphasengebundenen Aldehyds mit einem Danishefsky-Dien in Gegenwart eines chiralen Katalysators genutzt, um Lactone in hoher Ausbeute und mit großem Enantiomerenüberschuss herzustellen. Die Lactone wurden darüber hinaus an der festen Phase modifiziert, um eine Naturstoff-inspirierte Substanzbibliothek basierend auf dem Tetrahydropyran-Gerüst zu erhalten (Abbildung 16, **66**).^[57]

Melophlin A und B sind Naturstoffe aus der Klasse der Tetramsäuren, welche die Morphologie von HRas-transformierten NIH3T3-Fibroblasten bei einer Konzentration von $5 \mu\text{g mL}^{-1}$ rückgängig machen.^[58] Zur Identifizierung ihres Zielproteins sowie ihrer biologischen Funktion im Ras-Signaltransduktionsnetzwerk wurde eine Melophlin-abgeleitete Substanzbibliothek synthetisiert (**67**, Abbildung 16). Die biologische Untersuchung sowie anschließende chemische

Proteomanalysen enthüllten, dass Melophlin A in Zellen unerwartet an Dynamine bindet und so die Signalweiterleitung durch das Ras-Netzwerk indirekt beeinflusst, indem es die Endocytose von MEK, einem Zielprotein von Ras, verhindert.^[59]

β -Lactone kommen in verschiedenen Naturstoffen vor und wurden als Gerüststruktur für die Synthese der Palmostatine (**68**, Abbildung 16) verwendet. Palmostatin B wurde als Inhibitor der Acylproteinthioesterase 1 (APT1)^[33] entwickelt und erfolgreich genutzt, um die generelle Rolle dieser Thioesterase bei der Regulierung der Lokalisierung, des intrazellulären Transports und der Signalweiterleitung der S-palmitoylierten H- und N-Ras-Proteine aufzuklären (siehe Abschnitt 3.2).^[60]

Cyclopeptidische Gerüststrukturen sind häufig in Naturstoffen anzutreffen. Die Brunsvicamide sind modifizierte Cyclopeptide aus Cyanobakterien, die durch die ϵ -Aminogruppe einer D-Lysineinheit cyclisiert und mit Harnstoffeinheiten funktionalisiert wurden. Sie zeigen potente Inhibition von Carboxypeptidasen. Eine Bibliothek modifizierter Brunsvicamide wurde synthetisiert, indem die Aminosäurebausteine und die Stereochemie in einer kombinierten Fest- und Flüssigphasensynthese variiert wurden.^[61] Die kleine Bibliothek wurde biochemisch auf Inhibition der Carboxypeptidase A untersucht. Die Ergebnisse zeigen die Bedeutung der verschiedenen Aminosäuren und besonders die hohe Relevanz der Stereochemie des Lysins für die inhibitorische Aktivität. Darüber hinaus wurde eine Synthese für von Chondramid C abgeleitete Cyclopeptide entwickelt und zum Aufbau einer Bibliothek möglicher Modulatoren der Aktin-filament-Polymerisation eingesetzt.^[62] Der entscheidende Makrocyclisierungsschritt wurde mittels einer Ruthenium-katalysierten Ringschlussmetathese (RCM) durchgeführt, die im Verlauf der Bibliothekssynthese erkennbare Trends in der Reaktivität und der E/Z-Selektivität offenbarte. Der wachstumshemmende Effekt der synthetisierten Verbindungen wurde gemessen, und Struktur-Aktivitäts-Beziehungen wurden ermittelt, die in guter Übereinstimmung mit den relevanten Daten zu den Naturstoffen zu sein scheinen. Eine Anzahl weiterer potenter, nicht in der Natur vorkommende und vereinfachter Analoga wurden für weitere Studien zum Wirkmechanismus ausgewählt, insbesondere zur Untersuchung des Verhältnisses zwischen der Cytotoxizität der Verbindungen und ihrer Eigenschaft, das Actin-Gerüst zu zerstören.

Zusätzlich zu diesen illustrierenden Beispielen sind verschiedene Arbeiten publiziert worden, welche die erfolgreiche Synthese von Naturstoff-abgeleiteten Substanzbibliotheken beschreiben (für umfassende Übersichten siehe Lit. [38,39], für eine Übersicht mit ausgewählten Beispielen siehe Lit. [9b]). Auf der Grundlage dieser gemeinsamen weltweiten Anstrengung kann mit Sicherheit geschlossen werden, dass die verfügbaren Synthesemethoden die zuverlässige und schnelle Synthese Naturstoff-abgeleiteter Substanzbibliotheken ermöglichen.

3.2. Anwendung von BIOS in der Entwicklung von Liganden und Inhibitoren sowie in der chemischen Biologie

3.2.1. Die Anwendung von PSSC bei der Entwicklung eines APT1-Inhibitors

Die H- und N-Ras-Proteine sind S-palmitoylierte membrangebundene GTPasen, die eine wichtige Rolle bei der Weiterleitung von Wachstumsfaktorsignalen über die Plasmamembran spielen. Mutationen in Ras-Proteinen treten in ca. 30 % aller Krebserkrankungen auf. Die reversible Abspaltung und Anknüpfung von Palmitinsäure von und an Cysteine am C-Terminus (Depalmitoylierung und Palmitoylierung) der H- und N-Ras-Isoformen reguliert deren Membranverankerung und spezifische Lokalisierung. Die Dynamik des Palmitat-Umsatzes ist essentiell für die Signalweiterleitung über H- und N-Ras und zur Aufrechterhaltung des Ras-Transportkreislaufs zwischen der Plasmamembran und dem Golgi-Apparat (Abbildung 17).^[63] Acylproteinthioester-

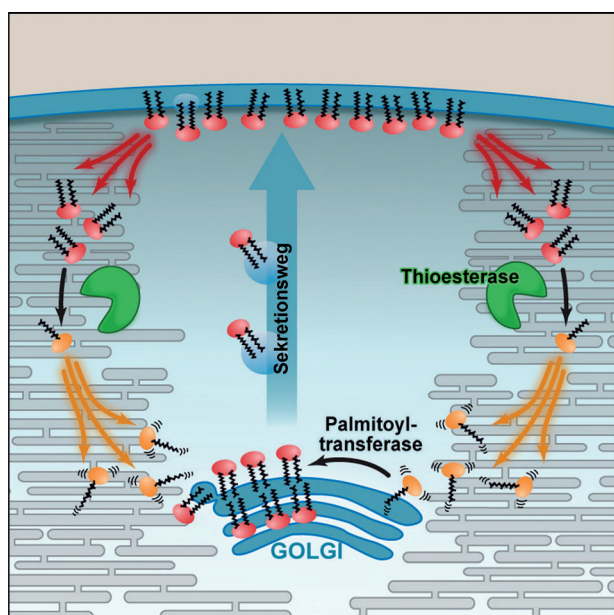


Abbildung 17. Der dynamische Ras-Zyklus. Abdruck nach *Cell* 2010, 141, 458–471.

rase 1 (APT1) war das einzige bekannte Enzym, das H- und N-Ras depalmitoyliert, ihre mögliche Rolle im Ras-Zyklus war aber unklar. Da der Transport von Ras zwischen den verschiedenen zellulären Bereichen innerhalb von Sekunden bis Minuten abläuft, ist ein chemisch-genetischer Ansatz, der eine schnelle APT1-Inhibition ermöglicht, besonders zur Aufklärung der Rolle von APT1 im dynamischen Ras-Zyklus geeignet. Da ein für diesen Zweck geeigneter Inhibitor nicht verfügbar war, wurde PSSC zur Entwicklung eines APT1-Inhibitors eingesetzt.

Die Suche nach Subfaltungen mit Ähnlichkeit zur Ligandbindungstasche von APT1 im Rahmen der PSSC-Analyse ergab eine hohe strukturelle Ähnlichkeit mit der gastrischen Lipase aus Hunden, trotz einer sehr geringen Sequenzähnlichkeit (kleiner 25 %).^[33] Überlagerung der Struk-

turen der beiden aktiven Zentren zeigt eine sehr ähnliche räumliche Anordnung der katalytischen Aminosäuren (Abbildung 18). Dies impliziert, dass Substanzen, die ähnlich zu

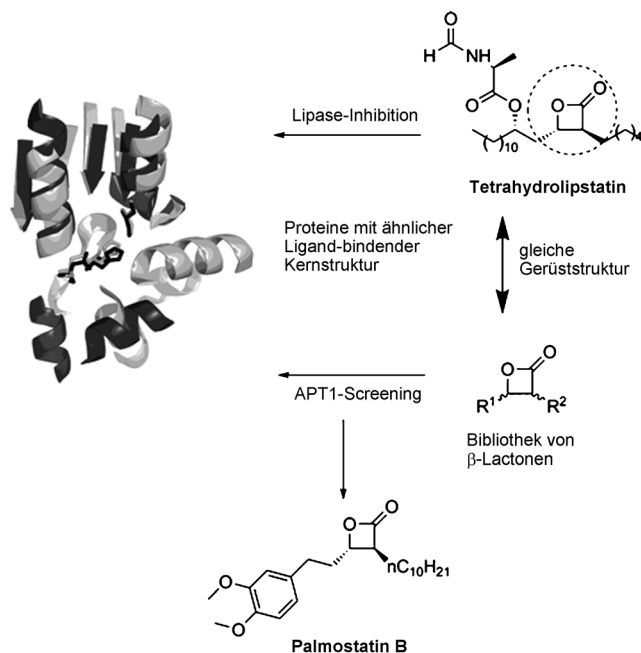


Abbildung 18. Proteinstrukturähnlichkeitscluster von APT1 (dunkelgrau) und der gastrischen Lipase (hellgrau) sowie die Logik für Synthese und Test einer β -Lactonbibliothek, die den APT1-Inhibitor Palmostatin B ergab.

Lipaseinhibitoren sind, möglicherweise gute APT1-Inhibitoren sein könnten. Der Naturstoff-abgeleitete, auf dem Markt befindliche Wirkstoff Tetrahydrolipstatin (Orlistat) enthält ein β -Lacton, das durch das Enzym angegriffen wird und es durch die Bildung eines Acyl-Enzym-Intermediats inhibiert. Ausgehend von dieser Analyse wurde eine Bibliothek von β -Lactonen synthetisiert. Die potenteste Substanz, Palmostatin B, wurde im Detail charakterisiert (Abbildung 18). Palmostatin B inhibiert APT1 kompetitiv mit einem IC_{50} -Wert von 670 nM durch reversible Acylierung des nucleophilen Serins der katalytischen Triade des Enzyms. Der resultierende APT1-Komplex von Palmostatin B wird langsam hydrolysiert, und die Substanz selbst hat eine Halbwertszeit von 58 h in wässriger Lösung. Palmostatin B ist ausreichend löslich und zellpermeabel, um eine nützliches Hilfsmittel bei der Untersuchung der Funktion von APT1 im Ras-Acylierungs-/Deacylierungszyklus darzustellen.

Folglich wurde Palmostatin B in einer Reihe biochemischer Experimente und Untersuchungen in lebenden Zellen eingesetzt, einschließlich zeitaufgelöster Fluoreszenzmikroskopiestudien, die bewiesen, dass die Verbindung in Zellen mit APT1 interagiert, selektiv für APT1 gegenüber anderen intrazellulären Hydrolasen ist und die Depalmitoylierung von H- und N-Ras in Zellen inhibiert. Palmostatin B stört den zellulären Acylierungskreislauf auf der Ebene der Depalmitoylierung und führt dadurch zum Verlust der präzisen Lokalisierung von palmitoyliertem Ras an der Membran im

stationären Zustand durch entropiegetriebene Verteilung des Proteins über die zellulären Membranen (Abbildung 17). Sehr unerwartet schwächt es dadurch das H-/N-Ras-Signal und induziert die mögliche Reversion von H-Ras-transformierten MDCK-F3-Zellen zum nicht-transformierten Phänotyp.

Die Rolle von APT1 als eine wesentliche Thioesterase im Acylierungszyklus wurde so eindeutig belegt, und APT1 wurde damit zu einem naheliegenden Zielprotein zur Bekämpfung von Krebs. Palmotatin B ist möglicherweise ein wertvoller Startpunkt für die Entwicklung von Modulatoren für die gestörte Signaltransduktion durch palmitoylierte Ras-Proteine.

3.2.2. Anwendung von SCONP zur Identifizierung neuer Phosphatase-Inhibitoren

Proteinphosphatasen sind entscheidende Regulatoren einer Vielzahl biologischer Prozesse und Zielproteine in Wirkstoffentwicklungsprogrammen, z.B. für Diabetes und die Krebstherapie.^[64] Die Inhibition von Phosphatasen in Zellen und in vivo erwies sich allerdings als schwierig, und daher ist der Bedarf an neuen Phosphatase-Inhibitoren groß.

In einem Versuch, solche neuen Verbindungen zu finden, wurde eine Bibliothek von 354 Naturstoffen auf Inhibition mehrerer Phosphatasen getestet.^[65] Überraschenderweise wurde das pentacyclische Alkaloid Yohimbin als Inhibitor ($IC_{50} = 22.3 \mu M$) der dual-spezifischen Phosphatase Cdc25A identifiziert, die als Zielprotein in der Krebstherapie angesehen wird. Da die Synthese einer Substanzbibliothek ausgehend vom pentacyclischen Gerüst des Yohimbins eine sehr große Herausforderung darstellen würde, wurde eine SCONP- und Brachiationsanalyse für diesen Naturstoff durchgeführt. Diese Analyse führte über tetra- und tricyclische zu bicyclischen Naturstoffgerüststrukturen, welche dann die Synthese von Substanzbibliotheken inspirierten (Abbildung 4 und 19).

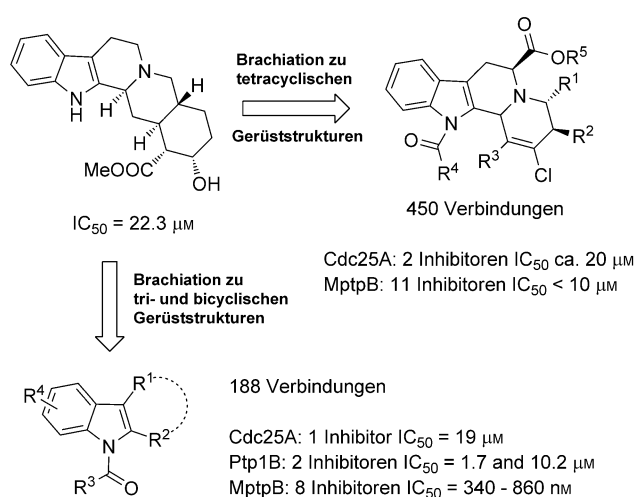


Abbildung 19. Brachiation durch den Indol-Ast des Naturstoff-Gerüststrukturaumes und Entwicklung neuer Naturstoff-abgeleiteter Phosphatase-Inhibitorklassen.

Eine Bibliothek von 450 tetracyclischen Indolochinolizinen wurde wie im Schema 4 gezeigt synthetisiert. Zusätzlich wurde eine weitere Bibliothek von 188 tri- und bicyclischen Indolderivaten mittels Fischer-Indol-Synthese und einer Capture-and-Release-Strategie an Harzkügelchen synthetisiert. Die biochemische Analyse der Substanzkollektion auf Cdc25A-Inhibition identifizierte zwei tetracyclische und eine tricyclische Verbindung mit IC_{50} -Werten, die dem des Naturstoffs vergleichbar sind. Weiteres Screening gegen andere Phosphatasen ergab neue Inhibitoren der Proteintyrosinphosphatase 1B (PTP1B), einem wichtigen Zielprotein in der Diabetesforschung, sowie nanomolare Inhibitoren der mycobakteriellen Proteintyrosinphosphatase B (MptpB), die ein vielversprechendes Zielprotein zur Entdeckung neuer Antituberkulosewirkstoffe ist.

Diese Ergebnisse belegen die erfolgreiche Brachiation durch den Ast der N-Heterocyclen des SCONP-Baumes. Sie verdeutlichen, dass der BIOS-Ansatz eine wesentliche Reduktion der molekularen Komplexität unter Erhaltung der Art der Bioaktivität erlaubt und eine Möglichkeit zur Entdeckung neuer, einfach zugänglicher von komplexen Naturstoffstrukturen abgeleiteter Inhibitorklassen bietet.

3.2.3. PSSC und SCONP kombiniert: Decaline als selektive 11β HSD1-Inhibitoren

Glucocorticoide sind Steroidhormone, die den Stoffwechsel von Glucose regulieren. Hohe Spiegel von Glucocorticoiden führen möglicherweise zur Entwicklung des Metabolischen Syndroms.^[66] Das aktive Glucocorticoid Cortisol wird von 11β -Hydroxysteroiddehydrogenase (11β HSD1) durch katalytische Reduktion aus Cortison hergestellt. 11β HSD1 wird hauptsächlich in der Leber, dem Fettgewebe und dem Gehirn exprimiert. In den Nieren katalysiert 11β HSD1 die Inaktivierung von Cortisol durch Oxidation zu Cortison und schützt so den Körper vor Cortisol-induziertem Bluthochdruck. In Mäusen führt der globale genetische Knock-out von 11β HSD1 zu erhöhter Insulinsensitivität und Resistenz gegen ernährungsbedingte Fettleibigkeit, Hypoglykämie und Dyslipidämie (Verschiebung der Lipidzusammensetzung im Blut). Diese Ergebnisse legen nahe, dass 11β HSD1-Inhibitoren bei der Behandlung von Diabetes Typ 2 und dem Metabolischen Syndrom sowie zur Vorbeugung von Arteriosklerose von Nutzen sein könnten.^[67]

Mehrere Pharma- und Biotechfirmen haben nicht-steroidale 11β HSD1-Inhibitoren mit positiver Wirkung in Tiermodellen der Arteriosklerose und Diabetes Typ 2 entwickelt, und die Suche nach Isoenzym-selektiven 11β HSD1-Inhibitoren geht weiter. Die synergistische Kombination von PSSC und SCONP führte zu neuen Typen von selektiven 11β HSD1-Inhibitoren mit zellulärer Aktivität.^[14]

Mittels PSSC wurden 11β HSD1 und 11β HSD2 einem Cluster mit der dual-spezifischen Phosphatase Cdc25A und der Acetylcholinesterase (AChE, Abbildung 13) zugeordnet. Obwohl Cdc25A und AChE mechanistisch nicht mit den 11β HSDs verwandt sind und ihre Sequenzähnlichkeit sehr niedrig ($< 10\%$) ist, zeigen die Subfaltungstypen des katalytischen Zentrums und die Positionen der katalytischen Aminosäuren eine gute Überlagerung (Abbildung 20B).

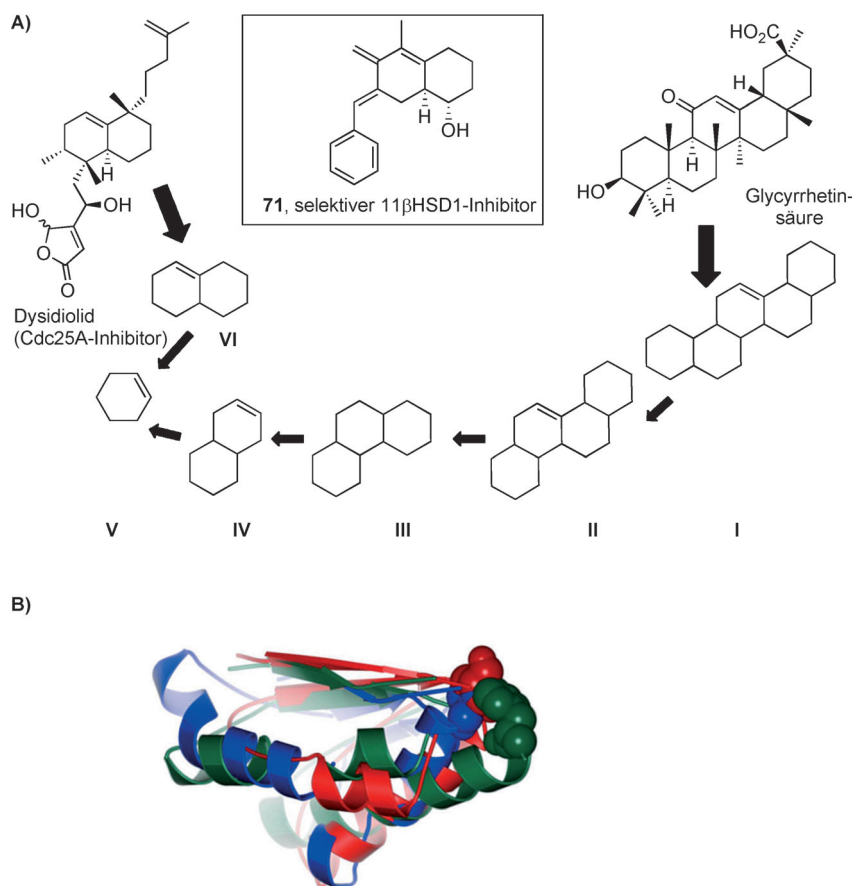


Abbildung 20. A) SCONP-Analyse von Glycyrrhetinsäure und Dysidiol sowie die Überlegungen, die zur Identifizierung des selektiven 11βHSD1-Inhibitors **71** führten. B) Überlagerung der katalytischen Zentren von Cdc25A (rot), 11βHSD1 (grün) und AChE (blau). Die katalytischen Aminosäuren, Cys-430 (Cdc25A), Tyr-183 (11βHSD1) und Ser-200 (AChE), sind als raumfüllende Darstellung gezeigt.

Zum Zeitpunkt der Analyse war die Struktur von 11βHSD1 nicht bekannt, sodass ein Homologiemodell für die PSSC-Analyse entwickelt und benutzt wurde. Später wurde die Proteinstruktur bestimmt und publiziert.^[68] Der Vergleich bestätigte die Qualität des Homologiemodells und zeigte, dass prinzipiell hochaufgelöste Kristallstrukturen von Proteinen für eine erste, hypothesengenerierende PSSC-Analyse nicht unbedingt notwendig sind.

Anschließend wurden der in der Natur vorkommende Cdc25A-Inhibitor Dysidiol und der 11βHSD1-Ligand Glycyrrhetinsäure (GS) mithilfe des SCONP-Baumes analysiert. Der schrittweise Abbau des pentacyclischen Gerüsts von GS führte zur bicyclischen 3,4-Dehydrodecalin-Gerüststruktur **IV**, während die SCONP-Analyse von Dysidiol das bicyclische Elterngerüst 1,2-Dehydrodecalin **VI** ergab (Abbildung 2A). Da **VI** als alternatives Teilgerüst von GS angesehen werden kann, wurde eine Naturstoff-abgeleitete, auf der Gerüststruktur **VI** basierende Bibliothek von 483 Decalinen synthetisiert.^[69] Neben mehreren niedrig-mikromolaren AChE-Inhibitoren enthielt die Bibliothek auch drei stark Isoenzym-spezifische nanomolare 11βHSD1-Inhibitoren. Für den selektiven 11βHSD1-Inhibitor **71** wurde nachfolgend gezeigt, dass er die Cortisol-vermittelte Translokation des Glucocorticoidrezeptors in den Zellkern von HEK-293-

Zellen in niedrig-mikromolarer Konzentration inhibiert. Dies zeigt, dass diese neue Substanz 11βHSD1 auch in Zellen inhibiert.

4. Zusammenfassung und Ausblick: Woher kommen wir und wohin gehen wir?

Bioaktive niedermolekulare Substanzen bieten einzigartige und bisher nicht verfügbare Möglichkeiten für die Untersuchung komplexer biologischer Phänomene durch schnelle, temporäre, konditionelle, selektive und regelbare Beeinflussung biologischer Systeme, ohne diese zu verändern. Der Schlüssel zur Entdeckung bioaktiver Substanzen liegt weniger in der Adressierung des gesamten chemischen Strukturraums als vielmehr in der Entwicklung und Anwendung von Methoden, die es ermöglichen, den biologisch relevanten chemischen Strukturraum zu identifizieren, zu kartieren und darin zu navigieren. Letztlich müssen solche entsprechend einzusetzenden Methoden die vorausblickende Erforschung des chemischen Strukturraumes sowie die Vorhersage von Bioaktivität für bestimmte Substanzklassen ermöglichen.

Um dieses Ziel zu erreichen, haben wir einerseits eine strukturelle Klassifikation der Naturstoffe (SCONP) eingeführt. Die zugrunde liegenden Gerüststrukturen der

Naturstoffe stellen evolutionär ausgewählte Strukturen dar, welche die notwendigen Eigenschaften kodieren, um an Proteine zu binden. Ihre Gerüststrukturen definieren biologisch relevante und prävalidierte Teile des chemischen Strukturraums, die von der Natur im Laufe der Evolution erschlossen wurden. Folglich ist zu erwarten, dass auf der Grundlage von Naturstoffen entworfene Substanzbibliotheken reich an biochemischer und biologischer Aktivität sind. Die baumartige Anordnung der Naturstoffgerüststrukturen ist dabei ein effizientes Werkzeug zur Erzeugung von Ideen und Hypothesen für die Entwicklung und Synthese von Substanzbibliotheken.

Des Weiteren haben wir das Proteinstrukturähnlichkeitsclustering (PSSC) als analoge Methode zur Erzeugung von Hypothesen eingeführt. PSSC beruht darauf, dass Proteinstrukturen in der Evolution konserviert sind und nutzt die strukturelle Ähnlichkeit von Bindungstaschen zur Identifizierung neuer Ligandentypen für die jeweils betrachteten Proteine. Beide Ansätze, PSSC und SCONP, deuten darauf hin, dass die Natur Konserviertheit und Diversität synergistisch nutzt: in der Evolution von Naturstoffen, der diese Naturstoffe biosynthetisierenden Proteine sowie der Proteine, an welche die Naturstoffe binden, um ihre biologische Funktion zu erfüllen. Auf der Ebene der Gerüststrukturen

war die Natur sowohl in der Welt der niedermolekularen Substanzen als auch in der Welt der Proteine konservativ. In beiden Fällen wurde die Konserviertheit durch Diversität ergänzt, und zwar durch Diversität der Substituenten in den Molekülen und deren Anknüpfungspunkten sowie durch Diversität der Seitenketten der Aminosäuren in den Liganden-bindenden Kernstrukturen. Die gegenseitige Abstimmung der Gerüststruktur sowie der Struktur der Substituenten und deren Position innerhalb der Moleküle und Proteine ermöglichen den Entwurf und die Identifikation biologisch relevanter Molekülklassen.

Sowohl PSSC als auch SCONP inspirieren die Auswahl von Gerüststrukturen für Substanzbibliotheken basierend auf Relevanz für und Prävalidierung durch die Natur. Wir bezeichnen Synthesen, die auf diesen Kriterien beruhen, als Biologie-orientierte Synthesen (BIOS). In BIOS können entweder SCONP- oder PSSC-Analysen sowohl einzeln als auch synergistisch genutzt werden, um Hypothesen und Ideen als Leitlinien der Synthese von Substanzbibliotheken zu entwickeln. Innerhalb von BIOS wird fokussierte Diversität um einen biologisch relevanten Startpunkt im riesigen chemischen Strukturraum generiert. BIOS baut zunächst auf der durch die Natur in der Evolution geschaffenen Diversität auf und zielt darauf ab, Gebiete nachgewiesener Relevanz durch Naturstoff-inspirierte und Naturstoff-abgeleitete Substanzbibliotheken lokal zu erweitern. Jedoch stellen auch nicht in der Natur vorkommende Gerüststrukturtypen mit nachgewiesener biologischer Relevanz valide Startpunkte für BIOS dar. BIOS ist also nicht auf Naturstoff-Gerüststrukturklassen beschränkt. Es verlangt nach biologischer Relevanz als Leitkriterium und nicht nach Vorkommen in der Natur. Geleitet von der biologischen Relevanz kann BIOS neue Möglichkeiten für die Entdeckung bisher unbekannter Proteinliganden- und Inhibitoriklassen bei relativ hohen Erfolgsquoten und vergleichsweise kleinen Bibliotheksgrößen bieten. Durch Brachiation entlang der Äste der Gerüststrukturbaume kann BIOS auch als Hypothesengenerator dienen, um zu strukturell einfacheren Gerüststrukturen zu gelangen, die eine gleiche Art der Bioaktivität, oft mit abgestufter Potenz und Selektivität aufweisen.

BIOS bietet hauptsächlich eine strukturbasierte, chemozentrische Perspektive auf das Problem der Identifizierung neuer bioaktiver Substanzen und Strukturtypen. Die grundlegende Idee zur Entwicklung von SCONP, PSSC und BIOS wurde in der zweiten Hälfte der 90er Jahre geboren und entwickelt, d. h. zu einer Zeit, als die erste Welle der kombinatorischen Chemie und des Hochdurchsatz-Screenings durch die industrielle und akademische Forschung gerollt war und sehr große Substanzbibliotheken synthetisiert wurden – hauptsächlich basierend auf der Grundlage chemischer Machbarkeit und kommerzieller Verfügbarkeit von Ausgangsverbindungen. Zu dieser Zeit begann sich abzuzeichnen, dass das Hochdurchsatz-Screening solcher Substanzbibliotheken sehr niedrige Trefferquoten lieferte. Dazu im Gegensatz standen etwa zwei Größenordnungen höhere Trefferquoten der historisch gewachsenen Substanzkollektionen der pharmazeutischen Industrie und der reinen Naturstoffbibliotheken.^[9c] Selbst in vollem Angesicht dieses Widerspruchs – für den es keine offensichtliche Erklärung

gab – führen die meisten pharmazeutischen Unternehmen trotzdem fort, die Naturstoffe aus ihren Screeningbibliotheken zu entfernen. Naturstoffe schienen strukturell zu komplex, um sie weiter zu verfolgen und zu synthetisieren, zu groß und in der Natur oft nicht in für die weitere Entwicklung ausreichender Menge verfügbar. Die hauptsächlich durch Technologie getriebene Entwicklung der Hochdurchsatz-Screeningmethoden schien geeignet, die Nachfrage nach einer zunehmenden Zahl von aktiven Substanzen, Leitstrukturen, Wirkstoffen und auch chemischen Sonden für biologische Untersuchungen zu befriedigen. Dennoch zeigte sich schnell, dass eine höhere Zahl von Screens, Substanzbibliotheken und Datenpunkten hierzu nicht ausreichen würden, sondern dass qualitativ hochwertige Substanzbibliotheken benötigt werden, die zusätzliche Kriterien erfüllen, wie beispielsweise biologische Relevanz, Wirkstoffartigkeit („drug-likeness“), strukturelle Komplexität und Diversität. Im Bewusstsein dieser Tatsachen und Entwicklungen, insbesondere der herausragenden Erfolge von Naturstoffen und der widersprüchlichen gleichzeitigen Entscheidung, Naturstoffe aus der industriellen Wirkstoffsuche zu verbannen, begannen wir uns zu fragen, ob es eine Logik geben könnte, mit der die strukturelle Komplexität der Naturstoffe reduziert werden könnte, und zwar unter Erhaltung von Bioaktivität. Könnte eine grundlegende Logik entwickelt werden, um die strukturelle Komplexität der Naturstoffe, ihr Verhältnis zueinander und auch die strukturelle Diversität der Bindungsstellen der Zielproteine systematisch zu analysieren? Könnte eine solche Logik die Synthese von Substanzbibliotheken inspirieren? Würde es chemisch möglich sein, Substanzbibliotheken von Strukturen, annähernd so komplex wie Naturstoffe, mit den dafür benötigten Methoden, wie beispielsweise der Festphasensynthese, herzustellen? Und falls dem so wäre, könnten diese Bibliotheken auch an die Erfolge der Naturstoffe in biochemischen und biologischen Screens anknüpfen? Würden sie also die Qualitätskriterien erfüllen und relativ hohe Trefferquoten bei vergleichbar kleiner Bibliotheksgröße liefern und so den Bedarf für den Einsatz von Hochdurchsatztechnologien verringern?

Im Falle des Erfolges könnten eine solche Logik und ein solcher Ansatz die Wiedereinführung von Naturstoffen in der Wirkstoffsuche und Entwicklung von Wirkstoffkandidaten anregen und vorantreiben, sowohl in der Medizinalchemie als auch in der chemisch biologischen Forschung – dann allerdings mit zuverlässiger Beherrschung der molekularen Komplexität und verfügbaren Syntheserouten mit Entwicklungspotenzial.

Als Antwort auf diese und ähnliche Fragen wurden SCONP, PSSC und BIOS entwickelt, als leitende grundlegende Logik, um biologisch relevante Klassen von Gerüststrukturen zu identifizieren, zu analysieren und hierarchisch zu ordnen, um Syntheseprojekte anzuregen und sogar die Art der Bioaktivität für Klassen von Gerüststrukturen vorherzusagen. Die durch BIOS gewonnenen Ergebnisse zeigen, dass eine Reduktion der strukturellen Komplexität von Naturstoffen und anderen Substanzen bei Erhaltung der Art der Bioaktivität in der Tat möglich ist und dass diese Beobachtung für alle aktuellen großen Klassen von Zielproteinen von Wirkstoffen gilt.

Wir betonen, dass Brachiation nach der Logik von BIOS sich von Versuchen unterscheidet, Naturstoffstrukturen rein nach chemischen Argumenten zu vereinfachen, beispielsweise nach höherer Effizienz der Synthese oder retrosynthetischen Überlegungen. In BIOS folgt Brachiation den Pfaden der biologischen Relevanz, die beispielsweise durch das Vorkommen der kleineren Gerüststruktur in der Natur oder verfügbare Bioaktivitätsdaten definiert werden. Die Chemie zur Synthese von Substanzbibliotheken mit kleineren Gerüststrukturen muss entsprechend ausgewählt werden. Daher basiert die Wahl der Syntheseeziele in BIOS auf biologischen Überlegungen und Auswahlkriterien.

Auf BIOS basierende Substanzbibliotheken sind klein, fokussiert und zeigen hohe Trefferquoten in Screens. Unser eigener Beitrag zur Synthese von Naturstoff-inspirierten und -abgeleiteten Substanzkollektionen sowie inzwischen viele ausgezeichnete Arbeiten verschiedener anderer Gruppen weltweit^[9a,38,39] lassen uns wohlbegründet schlussfolgern, dass die Methoden der organischen Synthese genügend entwickelt und leistungsfähig sind, um zuverlässigen und flexiblen Zugang zu solchen Substanzbibliotheken bei vertretbarem Aufwand zu gewähren, und dies sowohl im akademischen wie auch im industriellen Umfeld.

Der chemische Aufwand für die Synthese solcher Bibliotheken mag hoch sein und mehr Zeit in der Entwicklung benötigen, wird aber zu biologisch relevanten Substanzbibliotheken führen. In gewisser Hinsicht verlangt BIOS nach einer intensiveren Investition in den chemischen Teil der Entwicklung bioaktiver Moleküle, die sich durch bessere Moleküle für biologische Forschung bezahlt macht.

BIOS kehrt aber auch die Begründung und Inspiration für die Etablierung solche Syntheseprojekte um, die beispielsweise von der Methodenentwicklung oder der Suche nach einer Totalsynthese für einen bestimmten Naturstoff geleitet werden. Bioaktivität und Relevanz der Gerüststruktur eines bestimmten Naturstoffs bestimmen das Syntheseeziel, wobei die chemischen Synthesestrategien und -methoden an die daraus resultierenden Anforderungen angepasst werden müssen.

In rein synthetischen Untersuchungen gibt oft die Methode vor, welcher Naturstoff synthetisiert wird. Totalsynthesen führen zur Synthese eines Naturstoffs in allen strukturellen Einzelheiten, während BIOS die Naturstoffstruktur auf die Gerüststruktur reduziert, wobei die verschiedenen Substituenten und die Stereochemie variieren können

und sollen. Die präzise Herstellung der leitenden Naturstoffe in allen Einzelheiten wird gewürdigt und begrüßt, ist aber nicht zwingend notwendig.

Als die Ideen, die zur Entwicklung von SCONP, PSSC und BIOS führten, erste Form annahmen, war, wie bereits erwähnt, die pharmazeutische Industrie dabei, die Naturstoffe aus den Screeningbibliotheken zu verbannen und die Naturstoff-Forschungsabteilungen abzuschaffen und auszugliedern. Bis auf wenige Ausnahmen befinden sich heute alle wesentlichen Naturstoffbibliotheken überwiegend in der Hand kleinerer, spezialisierter Unternehmen, wie der InterMed Discovery GmbH und der Analyticon GmbH. Naturstoffe wurden als zu groß, zu komplex und nicht synthetisierbar angesehen, um in den zeitlichen Ablauf der Wirkstoffentwicklungsprojekte und Pipelines der Pharmafirmen zu passen. In dieser Begründung und dem nachfolgenden Prozess mögen jedoch Meinungen des öfteren die Fakten dominiert haben. So hat die statistische Analyse des SCONP-Baumes gezeigt, dass mehr als die Hälfte aller Naturstoffe Gerüststrukturen mit zwei, drei oder vier Ringen aufweisen und dass ihr Van-der-Waals-Volumen am unteren Ende der Größenverteilung von Proteinbindungstaschen liegt. Folglich hat also die Mehrheit der Naturstoffe genau die richtige Größe (!), um als Startpunkt für die Hit- und Lead-Suche und Entwicklungsprogramme zu dienen, einschließlich der Einführung weiterer Substituenten (Abbildung 21). Die SCONP-Analyse kann prinzipiell auch die Anknüpfungspunkte von Substituenten an den Gerüststrukturen aufzeigen und so das Design zusätzlich inspirieren.

Darüber hinaus kann der iterative Abbau von Gerüststrukturen so durchgeführt werden, dass die kleineren Gerüststrukturen „Fragment-ähnlich“ im Sinne der Fragment-

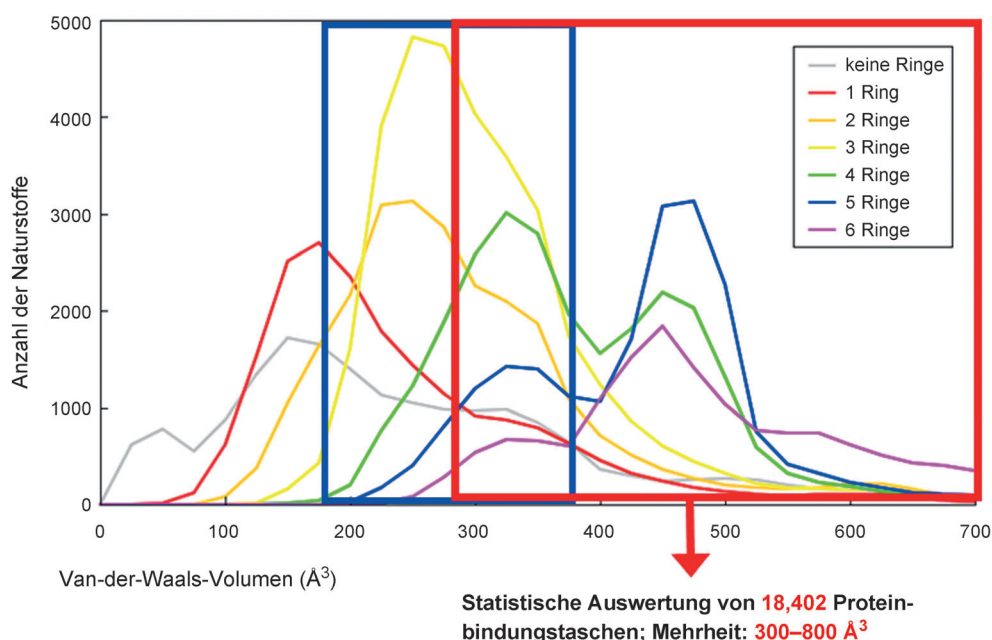


Abbildung 21. Vergleich der Van-der-Waals-Volumina der Naturstoffgerüststrukturen mit unterschiedlichen Ringzahlen mit den Volumina von Proteinbindungstaschen.^[70] Das Volumen der Naturstoffgerüste mit 2–4 Ringen liegt am unteren Ende der Größenverteilung der Volumina von Proteinbindungstaschen. Dies legt nahe, dass diese Gerüststrukturen nicht zu groß für die weitere Entwicklung sind.

basierten Wirkstoffentwicklung sind.^[71] Diese Analyse liefert in gewisser Hinsicht die „Fragmente der Natur“ und wird die Synthese von Naturstoff-inspirierten Substanzbibliotheken und deren Design weiter beflügeln. Schließlich hat die Synthese Naturstoff-inspirierter Substanzkollektionen, die oben zusammengefasst wurde, gezeigt, dass Substanzbibliotheken mit Naturstoff-ähnlicher Komplexität in der Tat erreichbar sind und dass die aktuell verfügbaren Synthesemethoden leistungsfähig genug für die Synthese solcher Bibliotheken im akademischen und industriellen Umfeld sind.

Zusammengefasst legen diese Ergebnisse und die Schlussfolgerungen daraus nahe, dass es sich für die pharmazeutische Industrie als klug erweisen könnte, das Naturstoff-basierte Design von Substanzbibliotheken, deren Synthese und Screening wieder in ihre Forschungs- und Entwicklungsprogramme aufzunehmen. Das muss nicht notwendigerweise auf die früher praktizierte Weise geschehen, bei der man sich auf einzelne Naturstoffe fokussierte. Naturstoff-inspirierte und -abgeleitete Substanzbibliotheken und Naturstoff-abgeleitetes Fragment-basiertes Design sollten die wohlverstandenen Notwendigkeiten und Randbedingungen erfüllen, die in einem industriellen Umfeld oft akzeptiert werden müssen. Die Logik von BIOS zeigt auch, dass die Verwendung von Naturstoffen alleine als Inspiration zur Identifizierung, Kartierung und Erforschung des biologisch relevanten chemischen Strukturraums nicht ausreicht und „Lücken“ im chemischen Strukturraum entstehen. Stattdessen ist es notwendig, die Analyse auf so viele bioaktive – und damit biologisch relevante – Substanzklassen wie möglich auszudehnen, unabhängig davon ob es sich um Naturstoffe handelt. Idealerweise würde man alle bekannten bioaktiven Substanzen einbeziehen. Am besten verdeutlicht eines der

erfolgreichsten Beispiele für strukturelle Vereinfachung bei Erhaltung vergleichbarer Bioaktivität diese Notwendigkeit. Morphin kann nicht in den SCONP-Baum integriert werden, da eine geeignete tetracyclische Gerüststruktur in der Natur bisher nicht gefunden wurde. Nicht in der Natur vorkommende, tetracyclische Morphinderivate wurden jedoch zu marktfähigen Wirkstoffen entwickelt (Abbildung 22).

Bei der Entwicklung von BIOS wurden ursprünglich Naturstoffe als prävalidierte Leitbeispiele benutzt; BIOS ist aber nicht auf diese beschränkt. Naturstoffe spiegeln die von der Natur in der Evolution entwickelte Lösung zur Identifizierung biologisch relevanter Bereiche im chemischen Strukturraum wider. Selbstverständlich gibt und wird es andere Lösungen geben, die nicht durch die Natur genutzt wurden.

Um die Abdeckung des chemischen Strukturraums zu erweitern, wird es notwendig sein, Zugang zu deutlich größeren Datensätzen zu erhalten, die Struktur und Bioaktivität verknüpfen, als in der unserer Analyse zugrunde liegenden WOMBAT-Datenbank verfügbar. In der jüngsten Vergangenheit wurde die öffentlich verfügbare ChEMBL-Datenbank im Internet gestartet,^[72] die eine Fülle von Bioaktivitätsdaten aus der wissenschaftlichen Literatur enthält. Darüber hinaus stellt auch PubChem große Datensätze für Analysen zur Verfügung. Die Information in diesen Datenbanken sollte gemeinsam mit dem DNP und WOMBAT eine deutlich verbesserte Analyse des biologisch relevanten chemischen Strukturraums erlauben.

Ein weiter Schritt in der Entwicklung solcher Ressourcen könnte die Anwendung von vollautomatisiertem Text-Mining auf die gesamte Literatur sein, einschließlich der Beziehungen zwischen der chemischen Struktur und der Bioaktivität der jeweiligen Moleküle. Die größten Quellen von qualitativ

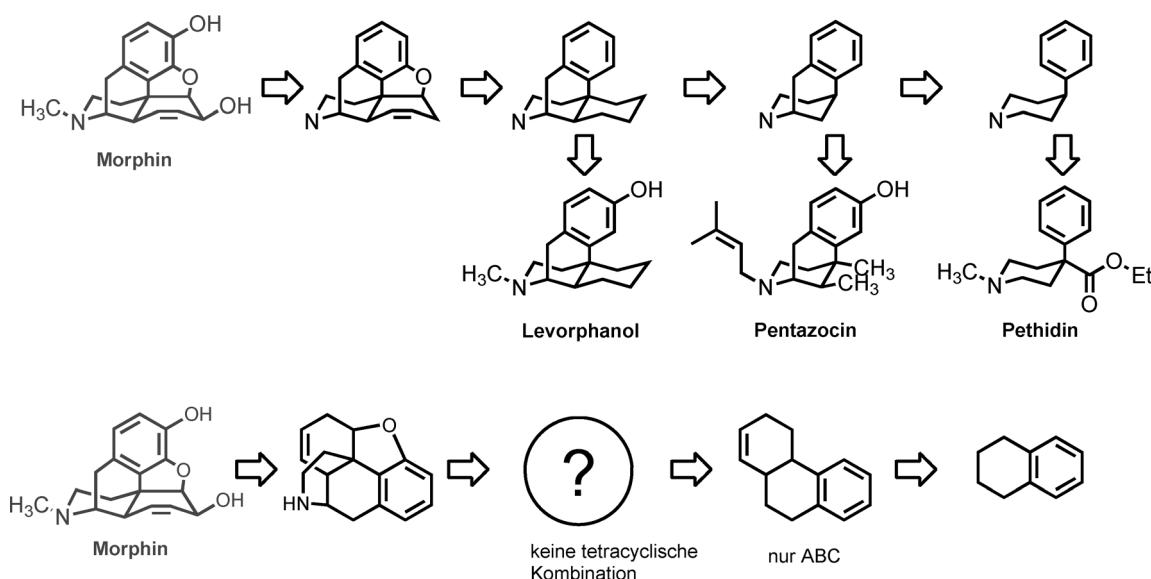


Abbildung 22. Das Dilemma beim Versuch, Morphin und seine verwandten Strukturen mit kleineren Gerüststrukturen in den Naturstoffbaum einzufügen. Morphin wurde in der pharmazeutischen Entwicklung strukturell vereinfacht, woraus kommerzialisierte Wirkstoffe mit drei und vier Ringen hervorgingen. Dennoch zeigt die Analyse von Morphin mittels SCONP, dass keine vom Morphin abgeleitete tetracyclische Gerüststruktur in der Natur vorkommt und dass nur tricyclische Gerüststrukturen bekannt sind. Somit entsteht eine Lücke im chemischen Strukturraum der Naturstoffe, die im Bioaktivitätsraum nicht existiert. Dieses Dilemma zeigt, dass die SCONP-Analyse durch die Einbeziehung von bioaktiven Substanzen, die nicht in der Natur vorkommen, ergänzt werden muss, um zu einer besseren Analyse des biologisch relevanten Strukturraums zu kommen.

hochwertigen Daten sind jedoch ausschließlich innerhalb der großen Pharmafirmen verfügbar, die über Jahrzehnte hinweg Millionen von Verbindungen in Hunderten von biochemischen und biologischen Screens getestet haben. Könnte man Zugang zu diesen Daten bekommen und könnten diese dann einer Analyse entsprechend des BIOS-Ansatzes unterzogen werden, so sollte diese Analyse eine Vielzahl neuer Forschungsprojekte ins Leben rufen. Diese neuen Projekte würden die chemisch-biologische und die medizinisch-chemische Forschung vorantreiben und möglicherweise zur schnelleren Identifizierung und Entwicklung neuer und besserer Wirkstoffe beitragen. Wir erwarten nicht, dass dies in nächster Zukunft geschehen wird. Eher wird die akademische Forschung durch die Analyse von Datenbanken wie ChEMBL und PubChem inspiriert werden. Aber wir fragen uns auch, ob die pharmazeutischen Firmen wirklich wissen, wo die Lücken in ihren Substanzbibliotheken, Datenbanken und Patenten liegen.

5. Resümee

Synthetische niedermolekulare Substanzen bieten einzigartige Möglichkeiten, komplexe biologische Systeme direkt zu beeinflussen und zu untersuchen. Für die Entdeckung solcher Moleküle werden Methoden zur Eingrenzung und Kartierung des biologisch relevanten Strukturraums benötigt, und um in diesem Raum zu navigieren und ihn zu befüllen. Die Biologie-orientierte Synthese (BIOS) adressiert diese Problem einerseits durch eine Chemie-zentrierte Analyse der Ligand-bindenden Kernstrukturen, die in den Faltungstypen von Proteindomänen eingebettet sind, sowie andererseits den Gerüststrukturen aus der Evolution hervorgegangener Naturstoffklassen und weiteren, nicht in der Natur vorkommenden bioaktiven Substanzklassen. Für solche Analysen wurden das Proteinstrukturähnlichkeitsclustering (PSSC) sowie die strukturelle Klassifikation von Naturstoffen (SCONP) und deren Erweiterung auf bioaktive synthetische Verbindungen entwickelt. Sowohl einzeln als auch gemeinsam angewendet dienen diese Methoden der Erzeugung von Hypothesen, um biologisch relevante Klassen synthetischer niedermolekularer Substanzen zu identifizieren. Die so identifizierten Gerüststrukturen bilden die Basis für die Synthese kleiner und mittelgroßer Substanzbibliotheken, beispielsweise inspiriert durch die Struktur eines Naturstoffs. Diese Substanzkollektionen bevölkern durch ihre fokussierte Diversität den Raum um einen biologisch relevanten Startpunkt im riesigen chemischen Strukturraum. Scaffold Hunter, ein intuitiv zugängliches und hochgradig interaktives Computerprogramm, ermöglicht die Analyse biologisch relevanter Teile des chemischen Strukturraums, indem es die Gerüststrukturen hierarchisch anordnet – entweder nach deren chemischer Struktur oder basierend auf der biologischen Aktivität.

Naturstoff-basierte Substanzkollektionen mit fokussierter Diversität können mithilfe verschiedener Methoden effizient synthetisiert werden, z.B. mit mehrstufigen Verfahren in Lösung oder an der festen Phase, Domino- und Kaskadenreaktionen oder auch Mehrkomponentenreaktionen, die von

Polymer-immobilisierten Abfangreagentien und neuen Trenntechniken profitieren. Nach der Logik von BIOS hergestellte, Naturstoff-abgeleitete Substanzkollektionen erwiesen sich als mit biologischer Aktivität angereichert und ergaben typischerweise Trefferquoten von ca. 0.2–1.5 % aktiver Substanzen in biochemischen und zellbasierten Tests.

Die erfolgreiche Entwicklung des BIOS-Ansatzes ebnet den Weg für die Nutzung der biologischen Prävalidierung der Naturstoffstrukturen durch die Evolution in der chemisch-biologischen und medizinisch-chemischen Forschung. Dadurch überwindet BIOS die Hürde der synthetischen Zugänglichkeit oder Verfügbarkeit von Naturstoffen und inspiriert die zukünftige Nutzung von Naturstoffen und davon abgeleiteten Substanzkollektionen in der Wirkstoffentwicklung.

Die Entwicklung und experimentelle Validierung des BIOS-Konzepts spiegelt die Arbeit zahlreicher früherer und heutiger Mitglieder unserer Arbeitsgruppe über mehr als ein Jahrzehnt wider, denen wir mehr als dankbar sind. Ihre Namen finden sich in den Publikationen, die aus unserer Gruppe hervorgegangen sind und die in diesem Aufsatz zitiert wurden. Sie verfügten und verfügen über die notwendige Furchtlosigkeit, bedeutende Fragen zu stellen und wirklich schwierige und grundlegende Probleme der chemisch-biologischen und medizinisch-chemischen Forschung anzugehen. Kennzeichnend sind ihre intellektuellen und experimentellen Fähigkeiten und Fertigkeiten, die es ihnen ermöglichten, sich den Herausforderungen zu stellen, die ein multidisziplinärer Ansatz mit sich bringt, der die Methoden und Kulturen der Chemie, Biologie und Informatik integriert. Wir danken auch unseren Kooperationspartnern aus den verschiedenen Projekten, deren Namen in den Autorenlisten unserer gemeinsamen Veröffentlichungen zu finden sind. Ohne ihre fortwährenden Beiträge und die vertrauensvolle Zusammenarbeit hätten viele Projekte nicht erfolgreich in die Tat umgesetzt werden können. Unsere Forschung wurde unterstützt durch die Max-Planck-Gesellschaft, die Deutsche Forschungsgemeinschaft, das Bundesministerium für Bildung und Forschung, die Alexander von Humboldt-Stiftung, die Volkswagen-Stiftung, die Europäische Union (Finanzierung durch den Europäischen Forschungsrat im Rahmen des siebten Rahmenprogramms der Europäischen Union (FP7/2007-2013)/ERC Grant agreement no.268309), das Land Nordrhein-Westfalen, den Fonds der Chemischen Industrie, die Novartis AG, Bayer CropScience, die BASF AG und die Analyticon GmbH.

Eingegangen am 8. November 2010

- [1] E. Zamir, P. I. H. Bastiaens, *Nat. Chem. Biol.* **2008**, *4*, 643–647.
- [2] a) S. V. Frye, *Nat. Chem. Biol.* **2010**, *6*, 159–161; b) S. Wetzel, A. Schuffenhauer, S. Roggo, P. Ertl, H. Waldmann, *Chimia* **2007**, *61*, 355–360; c) K. Grabowski, K. H. Baringhaus, G. Schneider, *Nat. Prod. Rep.* **2008**, *25*, 892–904.
- [3] http://www.nature.com/nchembio/chemical_probes.html.
- [4] S. L. Schreiber, *Nat. Chem. Biol.* **2005**, *1*, 64–66.
- [5] a) C. M. Dobson, *Nature* **2004**, *432*, 824–828; b) R. S. Bohacek, C. McMartin, W. C. Guida, *Med. Res. Rev.* **1996**, *16*, 3–50; c) R. S. Bon, H. Waldmann, *Acc. Chem. Res.* **2010**, *43*, 1103–1114.

- [6] a) G. Klebe, *Drug Discovery Today* **2006**, *11*, 580–594; b) K. J. Simmons, I. Chopra, C. W. G. Fishwick, *Nat. Rev. Microbiol.* **2010**, *8*, 501–510.
- [7] a) M. Rupp, T. Schroeter, R. Steri, H. Zettl, E. Proschak, K. Hansen, O. Rau, O. Schwarz, L. Müller-Kuhrt, M. Schubert-Zsilavecz, K. R. Müller, G. Schneider, *ChemMedChem* **2010**, *5*, 191–194; b) P. Willett, *Drug Discovery Today* **2006**, *11*, 1046–1053; c) J. Rosén, J. Gottfries, S. Muresan, A. Backlund, T. I. Oprea, *J. Med. Chem.* **2009**, *52*, 1953–1962.
- [8] D. J. Newman, G. M. Cragg, *J. Nat. Prod.* **2007**, *70*, 461–477.
- [9] a) K. Kumar, H. Waldmann, *Angew. Chem.* **2009**, *121*, 3272–3290; *Angew. Chem. Int. Ed.* **2009**, *48*, 3224–3242; b) M. Kaiser, S. Wetzel, K. Kumar, H. Waldmann, *Cell. Mol. Life Sci.* **2008**, *65*, 1186–1201; c) R. Breinbauer, I. R. Vetter, H. Waldmann, *Angew. Chem.* **2002**, *114*, 3002–3015; *Angew. Chem. Int. Ed.* **2002**, *41*, 2878–2890; d) S. C. K. Sukuru, J. L. Jenkins, R. E. J. Beckwith, J. Scheiber, A. Bender, D. Mikhailov, J. W. Davies, M. Glick, *J. Biomol. Screening* **2009**, *14*, 690–699; e) D. H. Drewry, R. Macarron, *Curr. Opin. Chem. Biol.* **2010**, *14*, 289–298.
- [10] F. Lovering, J. Bikker, C. Humblet, *J. Med. Chem.* **2009**, *52*, 6752–6756.
- [11] B. E. Evans, K. E. Rittle, M. G. Bock, R. M. Dipardo, R. M. Freidinger, W. L. Whitter, G. F. Lundell, D. F. Veber, P. S. Anderson, R. S. L. Chang, V. J. Lotti, D. J. Cerino, T. B. Chen, P. J. Kling, K. A. Kunkel, J. P. Springer, J. Hirshfield, *J. Med. Chem.* **1988**, *31*, 2235–2246.
- [12] R. I. Sadreyev, N. V. Grishin, *BMC Struct. Biol.* **2006**, *6*, 6.
- [13] a) T. Henkel, R. M. Brunne, H. Müller, F. Reichel, *Angew. Chem.* **1999**, *111*, 688–691; *Angew. Chem. Int. Ed.* **1999**, *38*, 643–647; b) M. L. Lee, G. Schneider, *J. Comb. Chem.* **2001**, *3*, 284–289; c) M. Feher, J. M. Schmidt, *J. Chem. Inf. Comput. Sci.* **2003**, *43*, 218–227.
- [14] M. A. Koch, A. Schuffenhauer, M. Scheck, S. Wetzel, M. Casalta, A. Odermatt, P. Ertl, H. Waldmann, *Proc. Natl. Acad. Sci. USA* **2005**, *102*, 17272–17277.
- [15] *Dictionary of Natural Products*, Chapman & Hall/CRC Informa, London, **2005**.
- [16] A. Schuffenhauer, P. Ertl, S. Roggo, S. Wetzel, M. A. Koch, H. Waldmann, *J. Chem. Inf. Model.* **2007**, *47*, 47–58.
- [17] K. O. Elliston, Talk at the 18th [BC]2 Basel Computational Biology Conference, **2010**.
- [18] S. Renner, W. A. L. van Otterlo, M. D. Seoane, S. Möcklinghoff, B. Hofmann, S. Wetzel, A. Schuffenhauer, P. Ertl, T. I. Oprea, D. Steinhilber, L. Brunsfeld, D. Rauh, H. Waldmann, *Nat. Chem. Biol.* **2009**, *5*, 585–592.
- [19] S. Wetzel, K. Klein, S. Renner, D. Rauh, T. I. Oprea, P. Mutzel, H. Waldmann, *Nat. Chem. Biol.* **2009**, *5*, 696–696.
- [20] *CTfile formats*, Symyx Technologies, San Ramon, CA, USA, **2007**; <http://www.symyx.com/downloads/public/ctfile/ctfile.pdf>.
- [21] Cambridge Soft Corporation, **2006**.
- [22] ISIS Draw, Symyx Technologies, San Ramon, CA, USA.
- [23] a) M. Olah, M. Mracec, L. Ostopovici, R. Rad, A. Bora, N. Hadaruga, I. Olah, M. Banda, Z. Simon, T. I. Oprea in *WOMBAT: World of Molecular Bioactivity*, Vol. 23, Wiley-VCH, Weinheim, **2005**; b) M. Olah, R. Rad, L. Ostopovici, A. Bora, N. Hadaruga, D. Hadaruga, R. Moldovan, A. Fulias, M. Mracec, T. I. Oprea, *Chemical Biology: From Small Molecules to Systems Biology and Drug Design*, Vol. 1–3 (Hrsg.: S. L. Schreiber, T. M. Kapoor, G. Wess), Wiley-VCH, Weinheim, **2007**, S. 760–786.
- [24] J. Inglese, D. S. Auld, A. Jadhav, R. L. Johnson, A. Simeonov, A. Yasgar, W. Zheng, C. P. Austin, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 11473–11478.
- [25] a) W. Schulz, *Chem. Eng. News* **1996**, *74*(25), 43–44; b) Chemical Abstract Service (CAS), Columbus, Ohio, USA, **2007**.
- [26] S. Wetzel, W. Wilk, S. Chammaa, B. Sperl, A. G. Roth, A. Yektaoglu, S. Renner, T. Berg, C. Arenz, A. Giannis, T. I. Oprea, D. Rauh, M. Kaiser, H. Waldmann, *Angew. Chem.* **2010**, *122*, 3748–3752; *Angew. Chem. Int. Ed.* **2010**, *49*, 3666–3670.
- [27] T. Berg, *ChemBioChem* **2008**, *9*, 2039–2044.
- [28] a) K. C. Nicolaou, R. M. Evans, A. J. Roecker, R. Hughes, M. Downes, J. A. Pfefferkorn, *Org. Biomol. Chem.* **2003**, *1*, 908–920; b) K. C. Nicolaou, J. A. Pfefferkorn, H. J. Mitchell, A. J. Roecker, S. Barluenga, G. Q. Cao, R. L. Affleck, J. E. Lillig, *J. Am. Chem. Soc.* **2000**, *122*, 9954–9967.
- [29] V. C. Jordan, *Cancer Cell* **2004**, *5*, 207–213.
- [30] C. A. de La Lastra, S. Sanchez-Fidalgo, I. Villegas, V. Motilva, *Curr. Pharm. Des.* **2004**, *10*, 3505–3524.
- [31] S. Wetzel, Dissertation, Universität Dortmund, **2009**; available at <http://eldorado.tu-dortmund.de/handle/03/26470>.
- [32] M. A. Koch, L. O. Wittenberg, S. Basu, D. A. Jeyaraj, E. Gourzoulidou, K. Reinecke, A. Odermatt, H. Waldmann, *Proc. Natl. Acad. Sci. USA* **2004**, *101*, 16721–16726.
- [33] F. J. Dekker, O. Rocks, N. Vartak, S. Menninger, C. Hedberg, R. Balamurugan, S. Wetzel, S. Renner, M. Gerauer, B. Schölermann, M. Rusch, J. W. Kramer, D. Rauh, G. W. Coates, L. Brunsfeld, P. I. H. Bastiaens, H. Waldmann, *Nat. Chem. Biol.* **2010**, *6*, 449–456.
- [34] a) F. J. Dekker, M. A. Koch, H. Waldmann, *Curr. Opin. Chem. Biol.* **2005**, *9*, 232–239; b) M. A. Koch, H. Waldmann, *Drug Discovery Today* **2005**, *10*, 471–483.
- [35] a) H. Waldmann, *Drugs Future* **2009**, *34*, 24–25; b) W. Wilk, T. J. Zimmermann, M. Kaiser, H. Waldmann, *Biol. Chem.* **2009**, *391*, 491–497.
- [36] B. M. McArdle, R. J. Quinn, *ChemBioChem* **2007**, *8*, 788–798.
- [37] B. M. McArdle, M. R. Campitelli, R. J. Quinn, *J. Nat. Prod.* **2006**, *69*, 14–17.
- [38] a) J. P. Nandy, M. Prakesch, S. Khadem, P. T. Reddy, U. Sharma, P. Arya, *Chem. Rev.* **2009**, *109*, 1999–2060; b) T. Lessmann, H. Waldmann, *Chem. Commun.* **2006**, 3380–3389.
- [39] a) L. A. Wessjohann, *Curr. Opin. Chem. Biol.* **2000**, *4*, 303–309; b) J. W. H. Li, J. C. Vederas, *Science* **2009**, *325*, 161–165; c) M. S. Butler, *Nat. Prod. Rep.* **2005**, *22*, 162–195; d) J. Y. Ortholand, A. Ganesan, *Curr. Opin. Chem. Biol.* **2004**, *8*, 271–280.
- [40] a) D. Brohm, S. Metzger, A. Bhargava, O. Müller, F. Lieb, H. Waldmann, *Angew. Chem.* **2002**, *114*, 319–323; *Angew. Chem. Int. Ed.* **2002**, *41*, 307–311; b) D. Brohm, N. Philippe, S. Metzger, A. Bhargava, O. Müller, F. Lieb, H. Waldmann, *J. Am. Chem. Soc.* **2002**, *124*, 13171–13178.
- [41] a) V. Mamane, A. B. Garcia, J. D. Umarye, T. Lessmann, S. Sommer, H. Waldmann, *Tetrahedron* **2007**, *63*, 5754–5767; b) A. B. Garcia, T. Lessmann, J. D. Umarye, V. Mamane, S. Sommer, H. Waldmann, *Chem. Commun.* **2006**, 3868–3870; c) J. D. Umarye, T. Lessmann, A. B. Garcia, V. Mamane, S. Sommer, H. Waldmann, *Chem. Eur. J.* **2007**, *13*, 3305–3319.
- [42] F. Perron, K. F. Albizati, *Chem. Rev.* **1989**, *89*, 1617–1661.
- [43] D. A. Evans, M. J. Dart, J. L. Duffy, D. L. Rieger, *J. Am. Chem. Soc.* **1995**, *117*, 9073–9074.
- [44] a) O. Barun, S. Sommer, H. Waldmann, *Angew. Chem.* **2004**, *116*, 3258–3261; *Angew. Chem. Int. Ed.* **2004**, *43*, 3195–3199; b) O. Barun, K. Kumar, S. Sommer, A. Langerak, T. U. Mayer, O. Müller, H. Waldmann, *Eur. J. Org. Chem.* **2005**, 4773–4788.
- [45] I. Paterson, D. Gottschling, D. Menche, *Chem. Commun.* **2005**, 3568–3570.
- [46] a) A. Nören-Müller, I. Reis-Corrêa, H. Prinz, C. Rosenbaum, K. Saxena, H. J. Schwalbe, D. Vestweber, G. Cagna, S. Schunk, O. Schwarz, H. Schiewe, H. Waldmann, *Proc. Natl. Acad. Sci. USA* **2006**, *103*, 10606–10611; b) I. R. Corrêa, Jr., A. Nören-Müller, H. D. Ambrosi, S. Jakupovic, K. Saxena, H. Schwalbe, M. Kaiser, H. Waldmann, *Chem. Asian J.* **2007**, *2*, 1109–1126.
- [47] A. Nören-Müller, W. Wilk, K. Saxena, H. Schwalbe, M. Kaiser, H. Waldmann, *Angew. Chem.* **2008**, *120*, 6061–6066; *Angew. Chem. Int. Ed.* **2008**, *47*, 5973–5977.

- [48] a) A. G. Myers, D. W. Kung, *J. Am. Chem. Soc.* **1999**, *121*, 10828–10829; b) J. F. González, E. de La Cuesta, C. Avendano, *Tetrahedron* **2004**, *60*, 6319–6326; c) C. Chan, R. Heid, S. P. Zheng, J. S. Guo, B. S. Zhou, T. Furuuchi, S. J. Danishefsky, *J. Am. Chem. Soc.* **2005**, *127*, 4596–4598.
- [49] S. C. Lee, S. B. Park, *J. Comb. Chem.* **2006**, *8*, 50–57.
- [50] D. J. Payne, J. A. Hueso-Rodriguez, H. Boyd, N. O. Concha, C. A. Janson, M. Gilpin, J. H. Bateson, C. Cheever, N. L. Nikonovich, S. Pearson, S. Rittenhouse, D. Tew, E. Diez, P. Perez, J. de La Fuente, M. Rees, A. Rivera-Sagredo, *Antimicrob. Agents Chemother.* **2002**, *46*, 1880–1886.
- [51] H. Waldmann, V. Khedkar, H. Duckert, M. Schumann, I. M. Oppel, K. Kumar, *Angew. Chem.* **2008**, *120*, 6975–6978; *Angew. Chem. Int. Ed.* **2008**, *47*, 6869–6872.
- [52] a) D. G. Rivera, O. E. Vercillo, L. A. Wessjohann, *Org. Biomol. Chem.* **2008**, *6*, 1787–1795; b) L. F. Tietze, M. E. Lieb, *Curr. Opin. Chem. Biol.* **1998**, *2*, 363–371; c) A. Ulaczyk-Lesanko, D. G. Hall, *Curr. Opin. Chem. Biol.* **2005**, *9*, 266–276.
- [53] S. Su, J. A. Porco, *J. Am. Chem. Soc.* **2007**, *129*, 7744–7745.
- [54] a) C. Bailly, M. Facompre, C. Tardy, C. Mahieu, C. Perez, I. Manzanares, C. Cuevas, *Clin. Cancer Res.* **2003**, *9*, 6112s–6113s; b) E. Marco, W. Laine, C. Tardy, A. Lansiaux, M. Iwao, F. Ishibashi, C. Bailly, F. Gago, *J. Med. Chem.* **2005**, *48*, 3796–3807; c) J. Kluza, M. A. Gallego, A. Loyens, J. C. Beauvillain, J. M. F. Sousa-Faro, C. Cuevas, P. Marchetti, C. Bailly, *Cancer Res.* **2006**, *66*, 3177–3187; d) M. V. R. Reddy, M. R. Rao, D. Rhodes, M. S. T. Hansen, K. Rubins, F. D. Bushman, Y. Venkateswarlu, D. J. Faulkner, *J. Med. Chem.* **1999**, *42*, 1901–1907; e) A. Aubry, X. S. Pan, L. M. Fisher, V. Jarlier, E. Cambau, *Antimicrob. Agents Chemother.* **2004**, *48*, 1281–1288.
- [55] H. Waldmann, L. Eberhardt, K. Wittstein, K. Kumar, *Chem. Commun.* **2010**, *46*, 4622–4624.
- [56] T. Lessmann, M. G. Leuenberger, S. Menninger, M. Lopez-Canet, O. Muller, S. Hümmer, J. Bormann, K. Korn, E. Fava, M. Zerial, T. U. Mayer, H. Waldmann, *Chem. Biol.* **2007**, *14*, 443–451.
- [57] M. A. Sanz, T. Voigt, H. Waldmann, *Adv. Synth. Catal.* **2006**, *348*, 1511–1515.
- [58] S. Aoki, K. Higuchi, Y. Ye, R. Satari, M. Kobayashi, *Tetrahedron* **2000**, *56*, 1833–1836.
- [59] T. Knoth, K. Warburg, C. Katzka, A. Rai, A. Wolf, A. Brockmeyer, P. Janning, T. F. Reubold, S. Eschenburg, D. J. Manstein, K. Hubel, M. Kaiser, H. Waldmann, *Angew. Chem.* **2009**, *121*, 7376–7381; *Angew. Chem. Int. Ed.* **2009**, *48*, 7240–7245.
- [60] O. Rocks, M. Gerauer, N. Vartak, S. Koch, Z. P. Huang, M. Pechlivanis, J. Kuhlmann, L. Brunsveld, A. Chandra, B. Ellinger, H. Waldmann, P. I. H. Bastiaens, *Cell* **2010**, *141*, 458–471.
- [61] a) T. Walther, S. Renner, H. Waldmann, H. D. Arndt, *Chem-BioChem* **2009**, *10*, 1153–1162; b) T. Walther, H. D. Arndt, H. Waldmann, *Org. Lett.* **2008**, *10*, 3199–3202.
- [62] a) H. Waldmann, T. S. Hu, S. Renner, S. Menninger, R. Tannert, T. Oda, H. D. Arndt, *Angew. Chem.* **2008**, *120*, 6573–6577; *Angew. Chem. Int. Ed.* **2008**, *47*, 6473–6477; b) R. Tannert, L. G. Milroy, B. Ellinger, T. S. Hu, H. D. Arndt, H. Waldmann, *J. Am. Chem. Soc.* **2010**, *132*, 3063–3077.
- [63] O. Rocks, A. Peyker, M. Kahms, P. J. Verveer, C. Koerner, M. Lumbierres, J. Kuhlmann, H. Waldmann, A. Wittinghofer, P. I. H. Bastiaens, *Science* **2005**, *307*, 1746–1752.
- [64] a) L. Bialy, H. Waldmann, *Angew. Chem.* **2005**, *117*, 3880–3906; *Angew. Chem. Int. Ed.* **2005**, *44*, 3814–3839; b) V. V. Vintonyak, A. P. Antonchick, D. Rauh, H. Waldmann, *Curr. Opin. Chem. Biol.* **2009**, *13*, 272–283.
- [65] C. Rosenbaum, P. Baumhof, R. Mazitschek, O. Muller, A. Giannis, H. Waldmann, *Angew. Chem.* **2004**, *116*, 226–230; *Angew. Chem. Int. Ed.* **2004**, *43*, 224–228.
- [66] a) M. Wang, *Nutr. Metab.* **2005**, *2*, 3; b) G. Arnaldi, A. Angeli, A. B. Atkinson, X. Bertagna, F. Cavagnini, G. P. Chrousos, G. A. Fava, J. W. Findling, R. C. Gaillard, A. B. Grossman, B. Kola, A. Lacroix, T. Mancini, F. Mantero, J. Newell-Price, L. K. Nieman, N. Sonino, M. L. Vance, A. Giustina, M. Boscaro, *J. Clin. Endocrinol. Metab.* **2003**, *88*, 5593–5602.
- [67] a) Y. Kotelevtsev, M. C. Holmes, A. Burchell, P. M. Houston, D. Schmoll, P. Jamieson, R. Best, R. Brown, C. R. W. Edwards, J. R. Seckl, J. J. Mullins, *Proc. Natl. Acad. Sci. USA* **1997**, *94*, 14924–14929; b) N. M. Morton, J. M. Paterson, H. Masuzaki, M. C. Holmes, B. Staels, C. Fievet, B. R. Walker, J. S. Flier, J. J. Mullins, J. R. Seckl, *Diabetes* **2004**, *53*, 931–938.
- [68] D. J. Hosfield, Y. Q. Wu, R. J. Skene, M. Hilgers, A. Jennings, G. P. Snell, K. Aertgeerts, *J. Biol. Chem.* **2005**, *280*, 4639–4648.
- [69] M. Scheck, M. A. Koch, H. Waldmann, *Tetrahedron* **2008**, *64*, 4792–4802.
- [70] H. Gohlke, G. Klebe, *Angew. Chem.* **2002**, *114*, 2764–2798; *Angew. Chem. Int. Ed.* **2002**, *41*, 2644–2676.
- [71] a) P. J. Hajduk, *Nat. Chem. Biol.* **2006**, *2*, 658–659; b) P. J. Hajduk, *J. Med. Chem.* **2006**, *49*, 6972–6976.
- [72] <http://www.ebi.ac.uk/chembl/index.php>.