

Statistische Untersuchungen zur Strukturkomplementarität von Naturstoffen und synthetischen Substanzen

Thomas Henkel,* Roger M. Brunne, Hartwig Müller und Felix Reichel

Naturstoffe repräsentieren eine reiche Quelle an biologisch aktiven Verbindungen. Sie hatten bedeutenden Anteil an der Erforschung und Entwicklung neuer Arzneimittel sowie Pflanzenschutzprodukte, was sich aus der retrospektiven Analyse wichtiger Handelsprodukte zwanglos ableiten läßt.^[1, 2] Diese historische Betrachtung gibt jedoch keine Antworten auf die Frage, inwieweit sich Naturstoffe in ihren Struktureigenschaften von gut zugänglichen synthetischen Substanzen unterscheiden. Mit dem Einzug des High-Throughput-Screenings (HTS) sowie der kombinatorischen Chemie in den Wirkstofffindungsprozeß kommt dieser Frage jedoch zentrale Bedeutung zu, um die zukünftige Rolle von Naturstoffen in diesem Forschungsbereich zu bestimmen. Es gilt zu ermitteln, ob Naturstoffe einen strukturell einzigartigen Testsubstanzpool bilden, der durch vertretbare Syntheseanstrengungen nicht einfach ersetzbar ist. In diesem Zusammenhang ist auch die Bedeutung einzelner Naturstoffquellen von Interesse, um eine mögliche Fokussierung und Effizienzsteigerung bei der Substanzwahl zu diskutieren.

Im folgenden wird erstmals versucht, die Struktureigenschaften von Naturstoffen im Vergleich mit synthetischen Substanzen statistisch zu beschreiben, um Aussagen zur Komplementarität dieser Struktursammlungen abzuleiten; zu diesem Zweck wurden zwei Naturstoffdatenbanken (*DNP*, *BNPD*), in denen ein Großteil der publizierten Naturstoffe registriert sind, zwei chemische Datenbanken, die käufliche Chemikalien (*ACD*) bzw. einen repräsentativen Prüfsubstanzpool (*Synthetika*) umfassen, sowie eine Datenbank mit pharmazeutischen Produkten und Ent-

wicklungssubstanzen (*Drugs*) herangezogen (siehe methodischer Teil).

Zuerst wurden die Molekulargewichts- bzw. Heteroatomverteilungen aus den betrachteten Datenbanken ausgewertet (Abbildung 1–4). Naturstoffe weisen dabei durchschnittlich

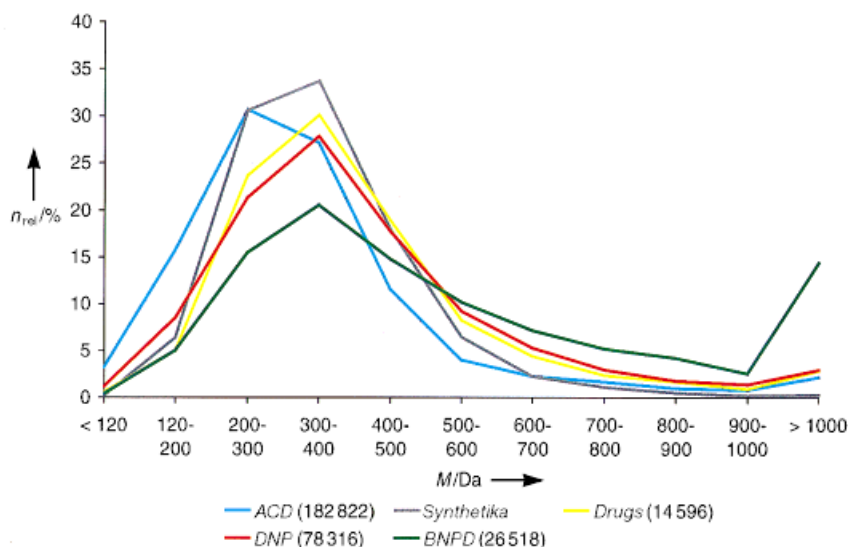


Abbildung 1. Verteilung n der Molekulargewichte in den fünf betrachteten Datenbanken. Die Zahlen in Klammern repräsentieren hier wie auch in allen folgenden Abbildungen die Zahl der berücksichtigten Substanzeinträge.

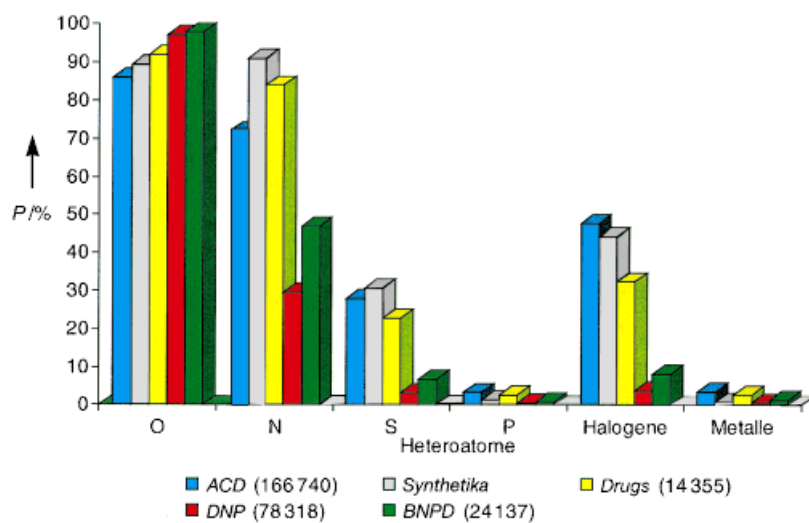


Abbildung 2. Populationen P der Heteroatome in den fünf betrachteten Datenbanken.

[*] Dr. T. Henkel, Dr. H. Müller
Bayer AG, Pharma Forschung
Life Science Center Natural Products (PH-R LSC-NP)
Aprather Weg, D-42096 Wuppertal
Fax: (+49) 202-368515
E-mail: THOMAS.HENKEL.TH@BAYER-AG.DE
Dr. R. M. Brunne
Bayer AG, Pharma Forschung, Structural Research (PH-R SR)
Dr. F. Reichel
Bayer AG, Zentrale Forschung Wirkstoffforschung (ZF-WFM)

Hintergrundinformationen zu diesem Beitrag sind im WWW unter <http://www.wiley-vch.de/home/angewandte/> zu finden oder können beim Autor angefordert werden.

höhere Molekulargewichte auf als synthetische Verbindungen; die Verteilung bei *Drugs* und *DNP* erscheint sehr ähnlich. Wie die Heteroatomverteilung zeigt, enthalten Naturstoffe weniger häufig Stickstoff-, Halogen- oder Schwefelatome, aber etwas häufiger Sauerstoffatome. Die Datenbank *Synthetika* weist mit ca. 90% den durchschnittlich höchsten Anteil an Stickstoffverbindungen auf.

In bezug auf topologische Parameter enthalten Naturstoffe (*DNP*) einen relativ höheren Anteil an Verbindungen mit sp^3 -hybridisierten Brückenkopfatomen. Auch ist die mittlere Zahl an Ringen sowie Chiralitätszentren pro Molekül größer

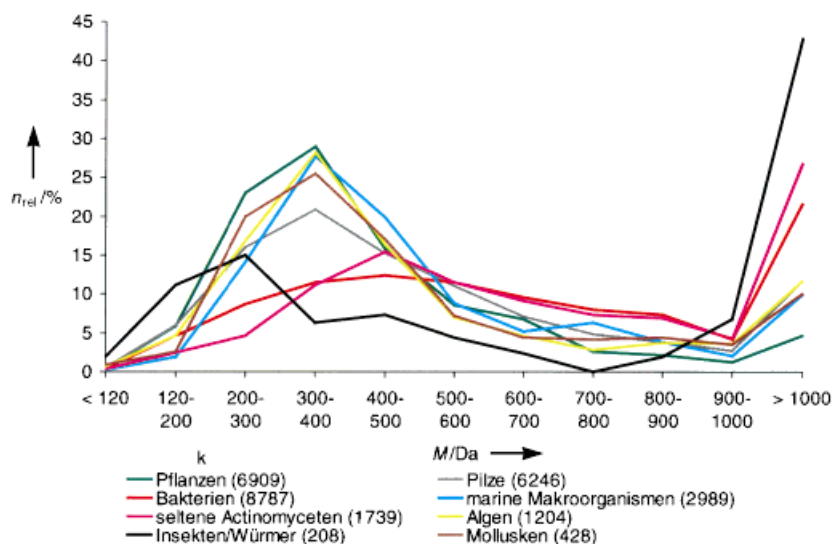


Abbildung 3. Molekulargewichtsverteilung von Naturstoffen aus unterschiedlichen Produzentenorganismen (Quelle: BNPD).

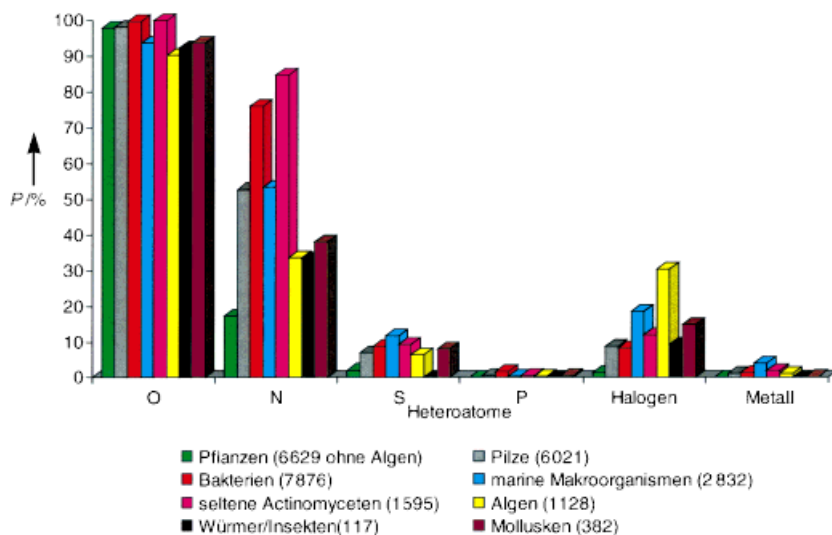


Abbildung 4. Populationen *P* der Heteroatome von Naturstoffen aus unterschiedlichen Produzentenorganismen (Quelle: BNPD).

(Tabelle 1). Damit lassen sich Naturstoffe statistisch als räumlich komplexere Strukturen beschreiben.

Hinsichtlich der durchschnittlichen Zahl verschiedener pharmakophorer Gruppen, z.B. Amidgruppen, je Verbindung weisen Naturstoffe (*DNP*, ca. 3.2 Pharmakophore) und synthetische Verbindungen (*Synthetika*, ca. 3.3) eine Mittelstellung zwischen den pharmazeutischen Wirkstoffen (*Drugs*, ca. 3.7) und den käuflichen Chemikalien (*ACD*, ca. 2.9) auf. Die Häufigkeit für bestimmte pharmakophore Gruppen weicht jedoch bei den Naturstoffen stark von der bei *Drugs* und *Synthetika* ab (Abbildung 5). Noch deutlicher unterscheiden sich die

Tabelle 1. Häufigkeiten ausgewählter Struktureigenschaften sowie durchschnittliche Zahl von Strukturmerkmalen pro Molekül bezogen auf alle Struktureinträge in drei exemplarischen Datenbanken.

| Eigenschaften | <i>Drugs</i> | <i>Synthetika</i> | <i>DNP</i> |
|--|--------------|-------------------|------------|
| Brückenkopf-atome mit drei Ringbindungen | 25 % | 9 % | 49 % |
| Brückenkopf-atome mit vier Ringbindungen | 4 % | 1.4 % | 13 % |
| C-C-Bindung, frei drehbar | 74 % | 58 % | 66 % |
| Ringe pro Molekül | 3.0 | 2.6 | 3.3 |
| Chiralitätszentren pro Molekül | 1.2 | 0.1 | 3.2 |
| frei drehbare Bindungen pro Molekül | 10.7 | 8.0 | 11.1 |

Datenbanken hinsichtlich der Häufigkeit von Kombinationen pharmakophorer Gruppen (Tabelle 2).

Über eine computergestützte Ähnlichkeitsanalyse wurde die bekannte hohe innere Redundanz der Naturstoffe – also die Häufigkeit, mit der zu jedem Strukturtyp mindestens ein strukturell verwandter aufzufinden ist – statistisch mit 85 % ermittelt. Schließt man diese Redundanz aus, so verringert sich der *DNP*-Pool von ca. 80000 Substanzen auf 11500 strukturell diverse

Tabelle 2. Häufigkeit von Kombinationen pharmakophorer Gruppen.

| pharmakophore Gruppen | <i>Drugs</i> [%] | <i>Synthetika</i> [%] | <i>DNP</i> [%] |
|-----------------------|------------------|-----------------------|----------------|
| Alkohol/Ether | 19 | 5 | 41 |
| Alkohol/Ester | 10 | 3 | 30 |
| Aren/Alkohol | 24 | 13 | 40 |
| Aren/Alkohol/Ether | 12 | 5 | 27 |
| Amin/Aren | 50 | 40 | 15 |
| Aren/Amid | 31 | 43 | 12 |
| Amin/Aren/Amid | 20 | 15 | 5 |

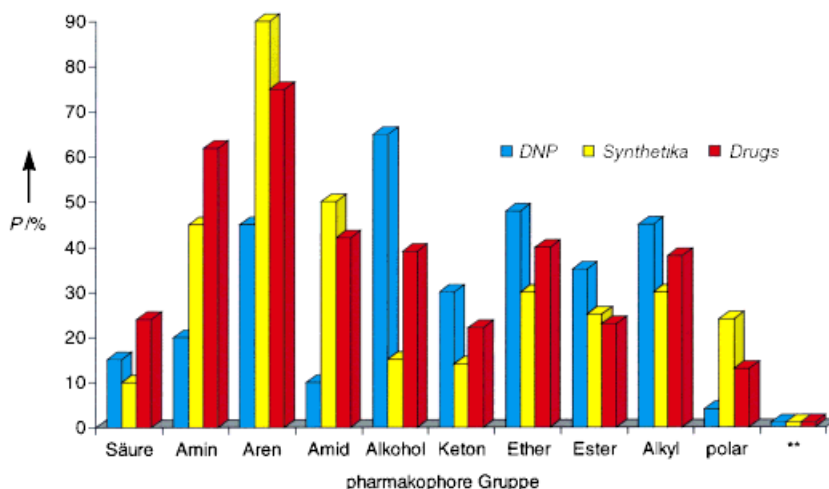


Abbildung 5. Populationen *P* von pharmakophoren Gruppen analysiert für die Datenbanken *DNP*, *Synthetika* und *Drugs*; ** = nicht klassifizierbar; polar = F, CN, NO₂.

Verbindungen. Vergleicht man hingegen die Strukturähnlichkeit der Verbindungen in *DNP* mit der der Substanzen in *Synthetika*, so sind ca. 40 % der Naturstoffe nicht durch synthetische Verbindungen repräsentiert.

Das Potential einzelner Naturstoffquellen wurde in bezug auf Struktureigenschaften zunächst durch Vergleich der Molekulargewichts- und Heteroatomverteilung der Verbindungen individueller Produzentenorganismen (aus *BNPD*) untersucht (Abbildung 3, 4). Hierbei ergeben sich signifikante Verteilungsunterschiede, z. B. die deutlich höheren Molekulargewichte von Naturstoffen, die aus seltenen Actinomyceten isoliert wurden, oder der niedrige durchschnittliche Stickstoffanteil in Pflanzeninhaltsstoffen. Analysiert man die Verteilung der in den Naturstoffdatenbanken angegebenen Strukturtypen (Abbildung 6), so findet sich ein hoher Anteil von Terpenen (ca. 35 %) und Alkaloiden (20 %) im *DNP*-Pool sowie von Peptiden (17 %) und Alicyclen (19 %) im *BNPD*-Pool. Analysiert man diese Verteilung hinsichtlich der Produzentenorganismen (Abbildung 7), so erhält man ebenfalls signifikante Populationsunterschiede. Im Vergleich zur Gesamtverteilung aller bioaktiven Naturstoffe in *BNPD* weisen solche von Bakterien beispielsweise einen überdurchschnittlich großen Anteil an Makrocyclen auf, der bei Pflanzen hingegen sehr gering ist. Offensichtlich kann ein bezüglich Strukturtypen repräsentativer Naturstoffpool nur unter Berücksichtigung aller Naturstoffquellen gebildet werden. Umgekehrt kann damit auch eine bedarfsgerechte individuelle Zusammenstellung bevorzugter Strukturtypen durchgeführt werden.

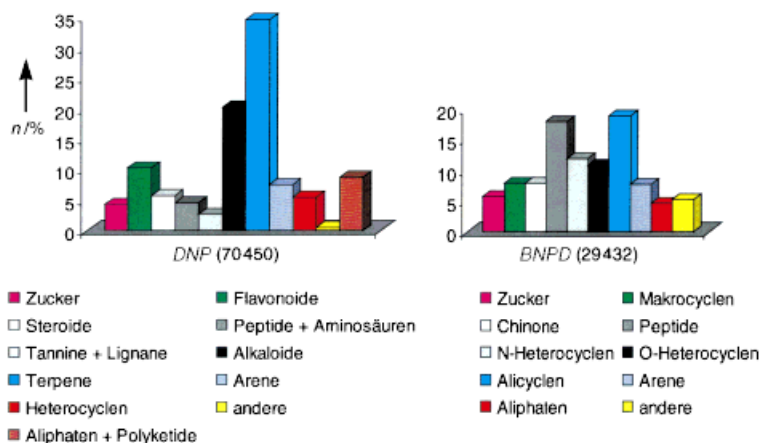


Abbildung 6. Verteilung der Strukturtypen unter den Naturstoffen der Datenbanken *DNP* und *BNPD*, entsprechend der internen Zuordnung in der jeweiligen Datenbank; beiden Datenbanken liegen unterschiedliche Klassifizierungen zugrunde, weshalb sie getrennt dargestellt sind.

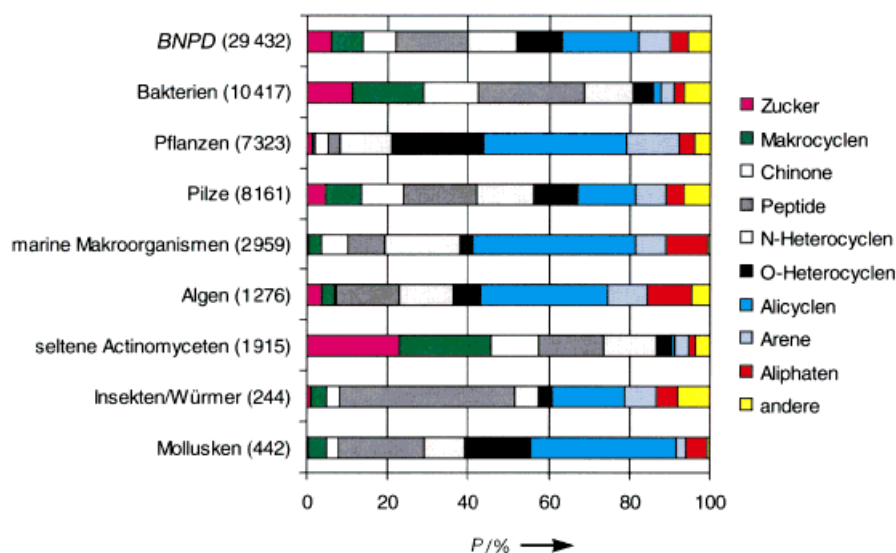


Abbildung 7. Populationen *P* der Strukturtypen von Naturstoffen aus unterschiedlichen Produzentenorganismen (Quelle: *BNPD*).

Aus der Zahl der bekannten Naturstoffe aus einzelnen Quellen (Abbildung 3, 4, 7) folgt, daß diese unterschiedlich produktiv sind und/oder unterschiedlich intensiv untersucht worden sind. Mehr als 10000 bioaktiven Bakterieninhaltsstoffen stehen weniger als 450 aus Mollusken gegenüber. Insgesamt verteilen sich die ca. 30000 bioaktiven Naturstoffe (*BNPD*) hinsichtlich ihrer Herkunft auf die vier großen Quellen Pflanzen, Bakterien, Pilze und Tiere zu 27, 33, 26 bzw. 13 %. Vor diesem Hintergrund wurde auch die Zahl neu beschriebener bioaktiver Naturstoffe (*BNPD*) nach Jahr und Quelle untersucht (Abbildung 8). Dabei wird deutlich, daß diese Zahl stetig gestiegen ist: von ca. 800 zwischen 1960–1965 auf mehr als 5000 zwischen 1990–1995. Diese Entwicklung betrifft alle untersuchten Quellen, besonders stark stiegen allerdings die Zahlen für Pilz- und marine Inhaltsstoffe an.

Alle dargelegten statistischen Auswertungen weisen eine deutliche Komplementarität der betrachteten Struktureigenschaften von Naturstoffen und synthetischer Verbindungen auf. Diese Verbindungen verfügen über typische Strukturelemente, die bisher nicht routinemäßig Ziele bei der Synthese für randomisierte Testsubstanzen waren. Die Strukturvarianz der Naturstoffverbindungen aus den unterschiedlichen natürlichen Quellen ist auffällig und sollte jeweils individuell entsprechend den Anforderungen bei der Wirkstoffsuche berücksichtigt werden. Das Potential an neuen Naturstoffen ist nicht erschöpft. Damit sind Naturstoffe nach wie vor eine bedeutende Quelle für die Wirkstofffindung. Das gelegentlich formulierte Vorurteil, Naturstoffe seien breit untersucht und folglich seien keine Neuerungen mehr zu erwarten, läßt sich damit eindeutig entkräften.

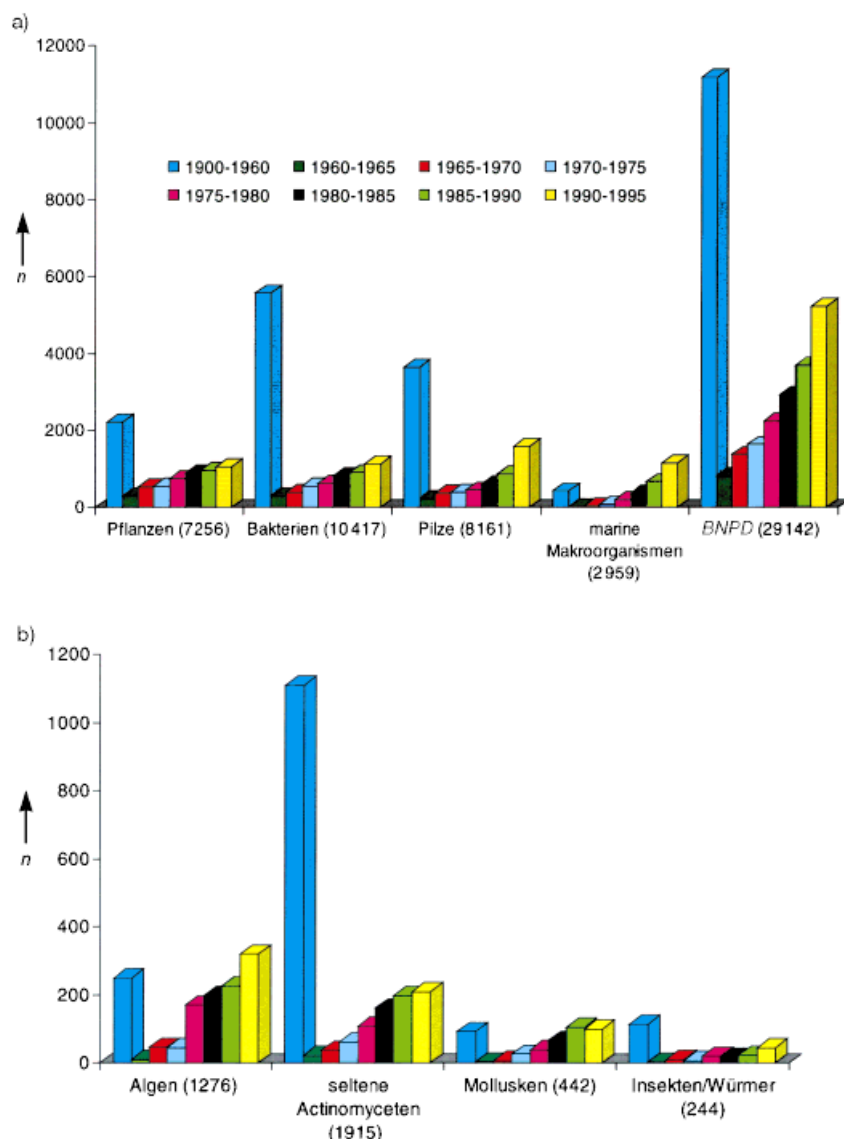


Abbildung 8. a, b) Zeitabhängigkeit von Neueinträgen in DNP, geordnet nach der biologischen Quelle.

Methodischer Teil

Als Datenbanken wurden zugrunde gelegt:

- **DNP**: Dictionary of Natural Products on CD-ROM (Chapman&Hall), 78318 Struktureinträge (Stand: Juni 1996)
- **BNPD**: Bioactive Natural Product Database, Szenzor Management Consulting Company, Budapest (Ungarn) (by Berdy), 29432 Einträge von Naturstoffen mit beschriebener biologischer Aktivität (Stand: Juli 1996)
- **Drugs**: pharmazeutische Produkte/Entwicklungssubstanzen, die in Pharmaprojects, RDFocus und dem bei der Bayer AG gepflegten Wirkstoffpool registriert sind, 14596 Einträge (Stand: Juni 1996)
- **ACD**: Available Chemicals Directory, Version 93.2, von Molecular Design Ltd. Information Systems Inc., San Leandro, CA (USA), 182822 Einträge
- **Synthetika**: repräsentativer Testsubstanzpool synthetischer Verbindungen der Bayer AG.

Aufgrund der Datenbankstrukturen ergaben sich einige Besonderheiten bei der Datenauswertung, die bei der Detailbetrachtung der ermittelten Auswertungen beachtet werden sollten: DNP umfaßt in der untersuchten

Version 78318 Verbindungen, von denen aber nur 70450 echte Naturstoffe sind. Je nach Betrachtung wurde wegen der Datenstruktur mit beiden Datensätzen gearbeitet. In DNP sind die Naturstoffquellen nicht auf einfache Weise identifiziert, so daß eine Auswertung der Produzentenorganismen auf BNPD begrenzt wurde. Da die Datensätze zum Teil unvollständig sind, sind die in den Abbildungen angegebenen Zahlen betrachteter Substanzen geringer als die Gesamtzahl in den Datenbanken. Zur besseren Übersichtlichkeit und Vereinfachung der Darstellung wurden in den Tabellen und Abbildungen zum Teil nur ausgewählte Strukturdatenbanken verglichen. Für die strukturbezogenen Auswertungen wurden die Datenbanken in MACCS^[3] und UNITY-Formate^[4] überführt und mit den dort verfügbaren Methoden analysiert. Für die Ähnlichkeitsanalysen der Molekülstrukturen wurde auf die 2D-Fingerprint-Deskriptoren in der Standarddefinition von UNITY^[4] zurückgegriffen. Die Ähnlichkeit/Unähnlichkeit zwischen Molekülpaares wurde nach Tanimoto^[5] berechnet und bezogen auf einen Grenzwert von 0.75 diskutiert. Die nach allgemeinem chemischen Verständnis definierten pharmakophoren Gruppen wurden mit bei der Bayer AG entwickelter Software analysiert. Nähere Angaben finden sich in den Hintergrundinformationen.

Eingegangen am 7. Oktober 1998 [Z12501]

International Edition:

Angew. Chem. Int. Ed. **1999**, 38, 643–647

Stichwörter: Molekulare Diversität • Naturstoffe • Statistische Analyse • Struktureigenschaften • Wirkstoff-Forschung

- [1] a) G. M. Gragg, D. J. Newmann, K. M. Snader, *J. Nat. Prod.* **1997**, 60, 52–60, zit. Lit.; b) J. P. Pachlatko, *Chimica* **1998**, 52, 29–47; c) Y.-Z. Shu, *J. Nat. Prod.* **1998**, 61, 1053–1071.
- [2] PharmaQuant, Ranking of Ethical Drug Products by Sales Value, Wood Mackenzie, **1998**.
- [3] MACCS-II, Version 2.3., Molecular Design Ltd. Information Systems, **1995**.
- [4] UNITY, Version 2.5., Tripos, Inc., St. Louis, MO (USA), **1996**.
- [5] P. Willet, *Similarity and Clustering in Chemical Information Systems*, Wiley, New York, **1987**.