Strukturermittlung ist der Knackpunkt

Kleine Moleküle und Molekülfragmente als Hinweisgeber – Bioinformatik als treibende Kraft

Franziska Hufsky und Rolf Kickuth

Zur Zeit machen die Ergebnisse zur Erforschung der Desoxyribonukleinsäure, ihrer Variationen und verwandten Moleküle, ihrer 3D-Struktur und ihrer Interaktion mit anderen Substanzen Furore. Das zeigen beispielsweise die Artikel über Epigenetik in dieser CLB, aber auch die Entwicklung von RNA-Impfstoffen. Tatsächlich sind es aber oft kleine Moleküle oder Molekülteile, die das Interesse der Chemiker, Biologen und Mediziner auf sich ziehen. Auch bei der Epigenetik geht es z. B. um Methylierungen, Acetylierungen etc. Oft hört man in diesen Tagen auch Hinweise auf exponentielles Wachstum, und man ist versucht, den Wissenszuwachs über Molekülstrukturen mit solch einem Modell zu beschreiben. Das trifft jedoch nicht zu. Dieser Artikel beschreibt, wie man sich mit kleinen Molekülen der Kenntnis über große Moleküle nähern kann, und wie sich die Kenntnis von Molekülstrukturen über die Zeit entwickelt hat. Wen wundert es. Eine große Rolle spielt dabei jetzt die Bioinformatik.

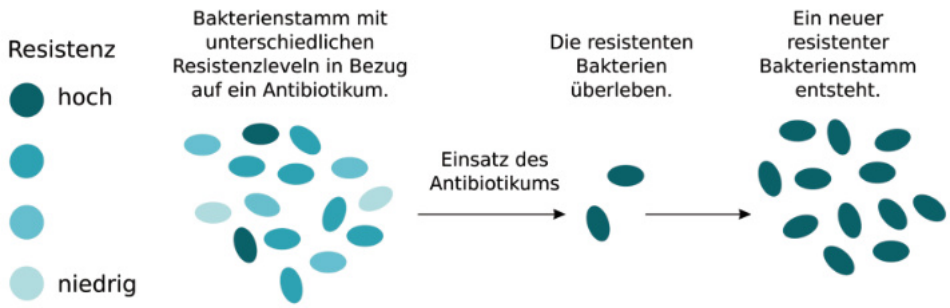
1. Kleine Moleküle im Fokus

Kleine Moleküle — insbesondere Naturstoffe — sind die Grundlage der Entwicklung neuer Medikamente, z.B. neuer Antibiotika. Multiresistente Bakterien, gegen die kein uns bekanntes Antibiotikum mehr hilft, werden immer mehr gefürchtet. Um diese Bakterien bekämpfen zu können, braucht man vollkommen neue Antibiotika, deren Entdeckung zunehmend schwieriger wird. Mittels bioinformatischer Methoden lässt sich die Struktur neuer potentieller Wirkstoffe wesentlich schneller entschlüsseln.

1.1 Gefahr durch multiresistente Bakterien

Seit der Entdeckung der antibiotischen Wirksamkeit von Penicillin durch Alexander Fleming (1929) hat man viele durch Bakterien ausgelöste Infektionskrankheiten, wie z.B. Cholera oder Syphilis, durch Antibiotika unter Kontrolle bekommen. Leider haben aber viele Bakterien im Laufe der Jahre Resistenzen gegen ein oder mehrere Antibiotika entwickelt (Abbildung 1). Dies liegt unter anderem am häufig verantwortungslosen Umgang mit Antibiotika, z.B. beim präventiven Einsatz in der Tierhaltung, als Medikament gegen Viren oder bei vorzeitigem Abbruch der Einnahme. Um multiresistente Bakterien bekämpfen zu können, braucht man vollkommen neue Antibiotika, die jedoch immer seltener auf den Markt kommen.

Abbildung 1: Gefahr durch multiresistente Bakterien: Schema der Entwicklung eines resistenten Bakterienstamms.



1.2 Die Kraft der Natur nutzen

Zur Herstellung von Medikamenten verwenden Menschen seit jeher Naturstoffe. Zur Entdeckung neuer Antibiotika macht man sich Gifte zu Nutze, die von Bakterien und Pilzen zur Abwehr gegen andere Bakterienstämme produziert werden. Um diese Wirkstoffe nicht umständlich aus Pflanzen oder Bakterien gewinnen zu müssen, muss man die molekulare Struktur der Wirkstoffe entschlüsseln. Strukturaufklärung von kleinen Molekülen ist ein aufwendiger und kostenintensiver Prozess, der

Die Autorin / der Autor:
Dr. Franziska Hufsky ist Bioinformatikerin. Sie studierte und promovierte an der Friedrich-Schiller-Universität Jena. Seit 2017 ist sie dort wissenschaftliche Koordinatorin des European Virus Bioinformatics Center (Foto: Anne Günther, Universität Jena).
Rolf Kickuth (68) ist Chemiker und Verleger der CLB.

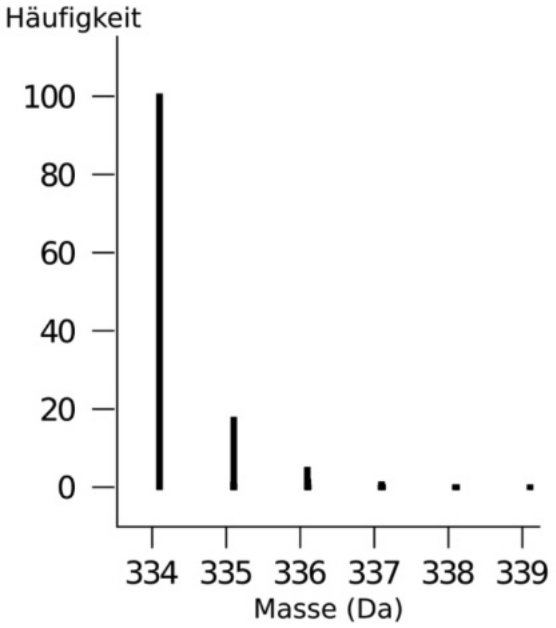


Abbildung 2: Isotopenmuster von Penicillin G ($C_{16}H_{18}N_2O_4S$), dem vom Schimmelpilz *Penicillium notatum* produzierten, natürlichen Antibiotikum..

mittels bioinformatischer Methoden wesentlich beschleunigt werden kann.

1.3 Waage statt Lupe zur Molekülerkennung

Mit klassischen Lichtmikroskopen lassen sich Strukturen dieser Größen- (oder eher Kleinen-)ordnung nicht mehr erkennen. Stattdessen werden die Moleküle gewogen. Solche winzigen Massen lassen sich mit der Massenspektrometrie ermitteln. Kennt man die exakte Masse, kann man daraus die Summenformel des Moleküls berechnen. Die Summenformel gibt bekannterweise die Zusammensetzung der Elemente im Molekül an, also wie viele Wasserstoff-, Kohlenstoffatome etc. im Molekül vorkommen.

Auch wenn sich die Messtechniken in den letzten Jahren extrem verbessert haben, kann man die Masse leider nicht ganz exakt bestimmen. Beim Berechnen der Summenformel muss man daher einen bestimmten Fehlertoleranzbereich beachten. Durch diesen Fehlertoleranzbereich ergeben sich plötzlich extrem viele Summenformeln, die die ungenaue Masse erklären könnten.

Um die Summenformel dennoch bestimmen zu können, braucht man mehr Information. Deswegen schaut man sich das Isotopenmuster des Moleküls an. Die Elemente auf unserer Erde kommen nämlich in unterschiedlicher Form vor. Zwei Atome des gleichen Elements können unterschiedliche Anzahl an Neutronen enthalten, und haben somit ein unterschiedliches Gewicht (siehe auch in CLB 1/2-2016 den Artikel von Marlene Görger: Stabilisotopenanalysen am Menschen – Was die Isotopie unseres Körpergewebes über uns verrät– und was nicht; S. 28-40).

So gibt es z.B. neben dem „normalen“ Wasserstoff (kein Neutron) auch schweren Wasserstoff (ein Neutron) und überschweren Wasserstoff (zwei Neutronen), der jedoch radioaktiv und somit instabil ist. Normaler Wasserstoff kommt auf unserem Planeten zu 99,985 % vor, schwerer nur zu 0,015 %. Durch die unterschiedliche Häufigkeit der Isotopen entsteht im Massenspektrum ein Muster – das Isotopenmuster (Abbildung 2), das sich für unterschiedliche Moleküle ändert. So kann man zwei Moleküle mit sehr ähnlicher Masse aber unterschiedlicher Summenformel voneinander unterscheiden. Mittels bioinformatischer Methoden, lässt sich das Isotopenmuster aller möglichen Summenformeln, die die gemessene Masse erklären würden, simulieren und mit dem gemessenen Isotopenmuster vergleichen, um die wahre Summenformel zu bestimmen.

1.4 Summenformel vs. Strukturformel

Die Summenformel beschreibt leider noch nicht die Struktur des Moleküls, d.h. die Anordnung der einzelnen Atome. Diese Anordnung bestimmt die chemischen Eigenschaften des Moleküls und somit auch dessen Wirksamkeit als Medikament. Für eine einzelne Summenformel gibt es wiederum Unmengen verschiedener atomarer Anordnungen und somit Unmengen möglicher Strukturformeln (Abbildung 3).

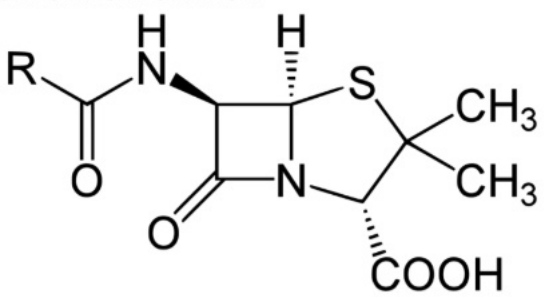
Um mit dem Massenspektrometer mehr Information als nur die Gesamtmasse des Moleküls zu ermitteln, wird das Molekül in Fragmente zerlegt, indem man es z.B. mit anderen Molekülen zusammenstoßen lässt. Die Massen der einzelnen Fragmente werden dann als ein für das Molekül typisches Spektrum erfasst. Experten werten solche Spektren aus, indem sie Diagramme zeichnen, welche die einzelnen Schritte, wie das Molekül bei der Kollision zerbrochen ist, per Hand nachvollzie-

Abbildung 3: Summenformel (oben) und Strukturformel (unten) der Kernstruktur der Penicilline, eine Gruppe von antibiotischen Wirkstoffen aus der Natur, die von Schimmelpilzen gebildet werden. Die Seitenkette „R“ ist variabel.

Summenformel:

$R\text{-}C_9H_{11}N_2O_4S$

Strukturformel:



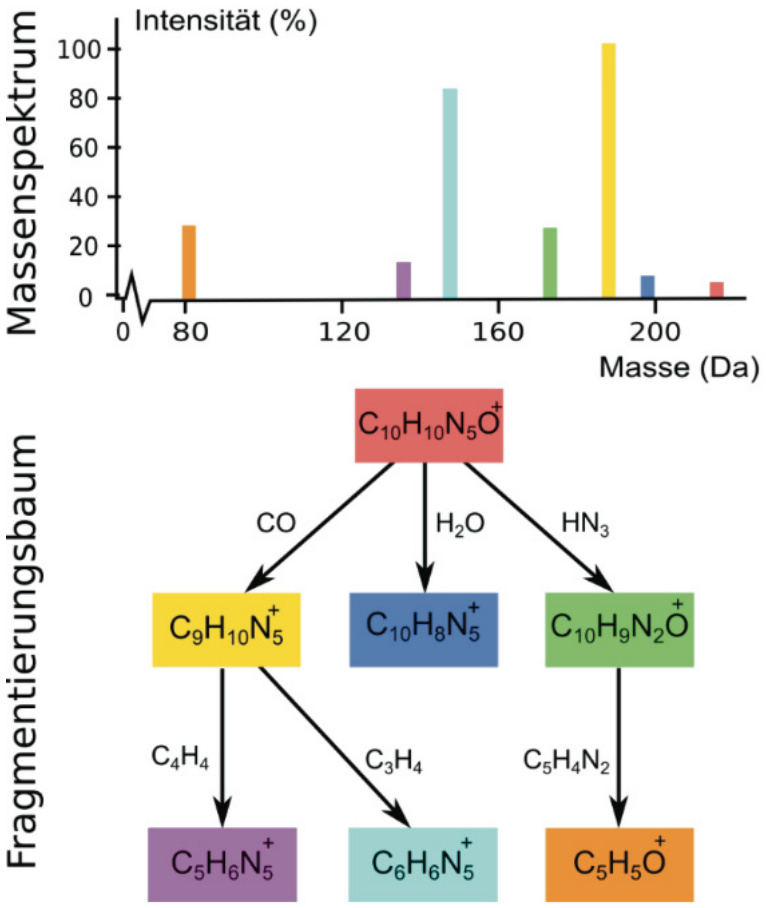


Abbildung 4: Jedes Fragment, das im Massenspektrum als Signal gemessen wurde, wird im Fragmentierungsbaum durch eine Summenformel erklärt.

hen. Eine solche manuelle Interpretation ist jedoch umständlich, zeitaufwendig und erfordert Expertenwissen sowie Kenntnis über die Molekülstruktur. Interessant für die Entdeckung neuer Medikamente sind aber genau solche Moleküle, deren Struktur bisher unbekannt ist.

1.5 Spezielle Bäume erklären die Chemie

Um die Analyse von Spektren unbekannter Moleküle in Zukunft zu erleichtern und wesentlich zu beschleunigen, hat die Arbeitsgruppe um Prof. Sebastian Böcker, Lehrstuhl für Bioinformatik der Uni Jena, das Konzept der Fragmentierungsbäume entworfen. Wenn in der Informatik von Bäumen die Rede ist, handelt es sich nicht um einen Baum im biologischen Sinn. Gemeint ist vielmehr das Konzept, das auch Stammbäumen zugrunde liegt.

Ein Fragmentierungsbaum ist eine Art Diagramm, welches den Fragmentierungsprozess des Moleküls veranschaulicht – ähnlich der manuellen Interpretation durch einen Experten, nur automatisiert, schnell und ohne Vorwissen (Abbildung 4). Startend mit der Summenformel des Moleküls beschreibt ein Fragmentierungsbaum nicht nur die Summenformeln der einzelnen Fragmente, sondern auch, wie diese Schritt für Schritt aus dem Molekül herausgebrochen wurden. Ein solcher Baum lässt sich in weniger als einer Sekunde berechnen.

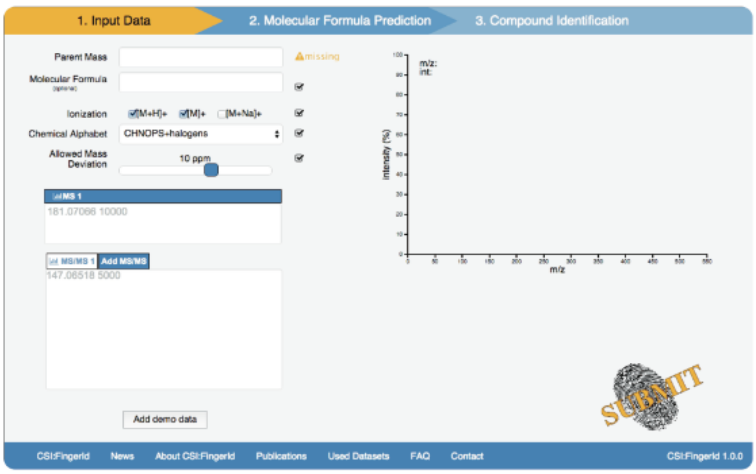


Abbildung 5: Bildschirmfoto der Strukturformel-Suchmaschine CSI:FingerID.

1.6 Ein bisschen CSI

Dieser Baum dient nun wiederum — zusammen mit dem gemessenen Fragmentierungsspektrum — als Grundlage für eine sehr starke Strukturformel-Suchmaschine: CSI:FingerID (Abbildung 5). Gibt man ihr das Fragmentierungsspektrum eines unbekannten Moleküls, sucht sie damit vollautomatisch in einer Strukturdatenbank, ohne die Spektren der darin enthaltenen Moleküle zu benötigen. Das ist wichtig, denn Spektren sind für neue, unbekannte Wirkstoffe noch nicht gemessen und referenziert wurden. CSI:FingerID beruht auf einem Machine Learning Ansatz. Aus dem gemessenen Spektrum und dem daraus berechneten Fragmentierungsbaum werden Eigenschaften abgelesen, z.B. die Masse des Moleküls und seiner Fragmente oder aber die Ähnlichkeit des Baums zu Bäumen anderer Moleküle. Mittels dieser Eigenschaften lässt sich durch das maschinell erlernte Wissen ein struktureller Fingerabdruck des Moleküls berechnen. Das klingt schon alles ein bisschen nach CSI — ist es auch, nämlich *Compound Structure Identification*. Mit diesem strukturellen Fingerabdruck kann man dann in der Strukturdatenbank nach dem Molekül suchen, das am wahrscheinlichsten diesen Fingerabdruck erzeugt hat.

Mittlerweile gibt es CANOPUS (Akronym für „(class assignment and ontology prediction using mass spectrometry)“, ein Rechenwerkzeug für die systematische Annotation von Verbindungsklassen. Auch hier die treibende Kraft: Aufklärungen bei „Metabolomics“, der analytischen Erfassung von niedermolekularen Stoffwechselverbindungen in einem biologischen System (s. Literatur „Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra“).

Mit nicht zielgerichteter Tandem-Massenspektrometrie (mehrere Analysatoren hintereinander) können Tausende von Molekülen in einer biologi-

schen Probe nachweisen. Die Annotation von Strukturmolekülen ist jedoch auf Strukturen beschränkt, die in Bibliotheken oder Datenbanken vorhanden sind, was die Analyse und Interpretation experimenteller Daten einschränkt. CANOPUS verwendet ein tiefes neuronales Netzwerk, um 2497 Verbindungsklassen aus Fragmentierungsspektren vorherzusagen, einschließlich aller biologisch relevanten Klassen. Das neue Bioinformatiker-Werkzeug zielt explizit auf Verbindungen ab, für die weder spektrale noch strukturelle Referenzdaten verfügbar sind, und sagt Klassen voraus, denen Tandem-Massenspektrometrie-Trainingsdaten fehlen.

1.7 Das System lernt selbst

CANOPUS funktioniert in zwei Phasen: Zunächst erzeugt das Verfahren, aus dem mittels Massenspektrometrie gemessenen Fragmentierungsspektrum, einen molekularen Fingerabdruck. Dieser beinhaltet Informationen über die Struktureigenschaften des gemessenen Moleküls. In einem zweiten Schritt ordnet das System den Metaboliten mithilfe des Fingerabdrucks einer bestimmten Stoffklasse zu, ohne diesen dafür identifizieren zu müssen (Abbildung 6).

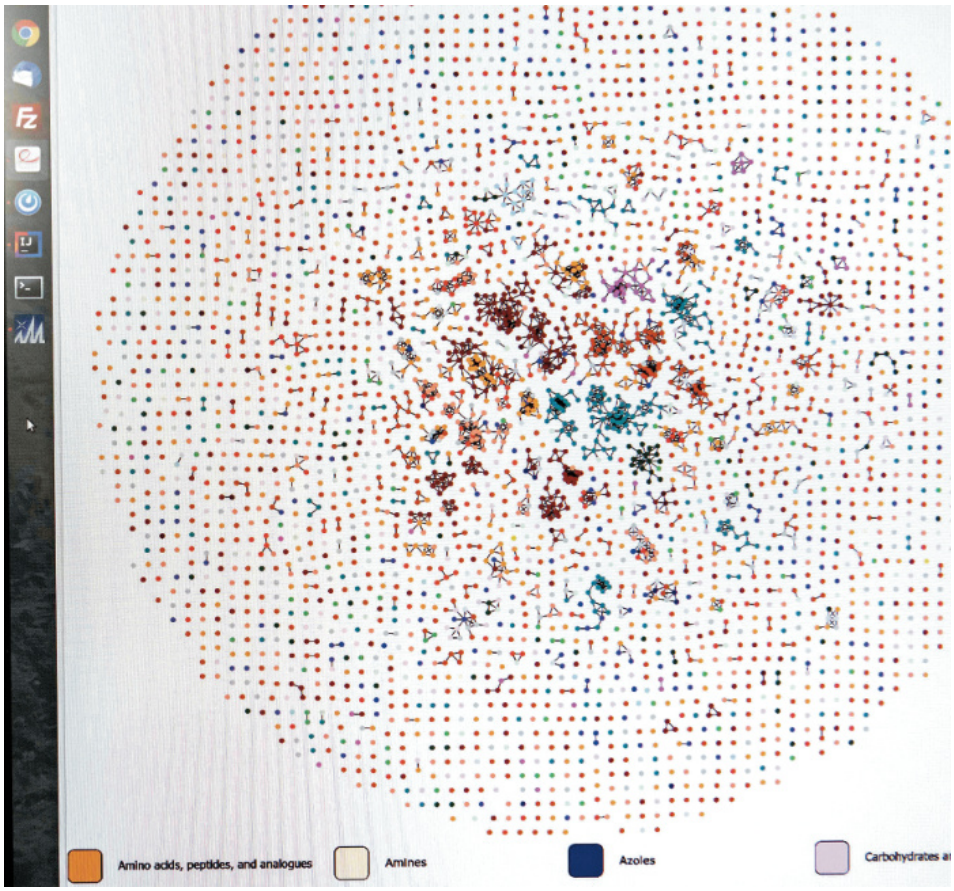


Abbildung 6: Bildschirmfoto von der Visualisierung eines mit CANOPUS genessenen Datensatzes (Abb.: Uni Jena/Jens Meyer).

Bemerkenswert an dem System: Maschinelle Lernen benötigt in der Regel große Datenmengen, um trainiert zu werden. Das neue zweistufige Verfahren hingegen ermöglicht es, im ersten Schritt auf einer vergleichsweise kleinen Datenmenge von zehntausenden Fragmentierungsspektren zu trainieren, um dann im zweiten Schritt aus Millionen von Strukturen die charakteristischen Struktureigenschaften zu bestimmen, die für eine Stoffklasse signifikant sind. Das System detektiert also diese Struktureigenschaften bei einem unbekannten Molekül innerhalb einer Probe und ordnet es dann einer bestimmten Stoffklasse zu. Allein diese Information reicht bereits aus, um viele wichtige Fragestellungen zu beantworten.

Mit ihrer Methode haben die Jenaer Bioinformatiker beispielsweise die Darmflora von Mäusen verglichen, bei denen eine Versuchsgruppe mit Antibiotika behandelt worden war. Die Untersuchungen zeigen, welche Metabolite die Maus und ihre Darmflora produzieren. Solche Forschungsergebnisse können wichtige Erkenntnisse über das menschliche Verdauungs- und Stoffwechselsystem ermöglichen.

Mit der neue Methode erweitern die Jenaer Bioinformatiker die Möglichkeiten der Suchmaschine für molekulare Strukturen „CSI:FingerID“, die sie der internationalen Forschungsgemeinschaft seit rund fünf Jahren zur Verfügung stellen .Es gibt noch weitere Informatik-Werkzeuge dieser Art, die sich auch teilweise ergänzen und sich in Kombination verwenden lassen.

1.8 Pharmazie, Forensik, Ernährung, Umwelt...

Neben der Pharmazie, spielt die Identifizierung kleiner Moleküle noch in vielen weiteren Forschungsfeldern eine große Rolle. In der Forensik z.B. testet man auf kleine Moleküle, um Abbauprodukte von Drogen, Giften etc. zu identifizieren. Ähnliches gilt für die Ernährungs- oder Umweltforschung. Aber auch bei der Aufklärung des Stoffwechsels von kaum erforschten Organismen stößt man ständig auf unbekannte Moleküle, die es zu entschlüsseln gilt.

2. Die Geschichte der Erforschung des chemischen Universums

Die Aufklärung chemischer Strukturen wird immer mehr von komplizierten Maschinen und von Computern unterstützt. Man könnte annehmen: Die Zahl aufgeklärter Strukturen nimmt exponentiell zu. Nun:

Vor einiger Zeit bin ich auf eine interessante Veröffentlichung über die Geschichte der Chemie gestoßen, genauer gesagt, über die Entdeckung neuer chemischer Strukturen (siehe Literatur: Exploration of the chemical space and its three historical regimes). Spannend finde ich das, weil Strukturdatenbanken die grundlegende Ressource meines Doktorarbeitsthemas waren, und weil neue chemische Moleküle immer auch das Potential für die Entwicklung neuer Medikamente bieten.

Um die wissenschaftlichen Entdeckung der Chemie aus mehr als 200 Jahren zu analysieren, haben die Autoren natürlich nicht tausende Veröffentlichungen gewälzt, sondern die Geschichte der Chemie mit einem vollständig datengetriebenen Ansatz aufgerollt. Soll heißen, sie haben sich die größte Reaktionsdatenbank geschnappt, die bis ins 19. Jahrhundert zurückreicht und in der sich etwa 42 Millionen Reaktionen mit über 20 Millionen Substanzen befinden, und darauf eine Menge Statistiken berechnet.

Das „chemische Universum“ umfasst alle möglichen chemischen Strukturen, von kleinen Molekülen wie Wasser, bis zu riesigen Molekülen wie Proteinen. Dazu zählen sowohl Substanzen, die auf natürliche Weise vorkommen, aber auch Substanzen, die chemisch synthetisiert werden müssen, das heißt durch eine Reaktion von zwei oder mehreren Verbindungen hergestellt werden. Jedes Jahr dehnt sich das „chemische Universum“ um 4,4 Prozent aus, das heißt, jedes Jahr werden 4,4 Prozent neue chemische Strukturen entdeckt; davon etwa die Hälfte durch Synthese. Und das relativ konstant seit 1800. „Relativ“, weil dieses Wachstum natürlich auch Schwankungen unterlag, unter anderem durch gesellschaftliche, aber auch wissenschaftliche Ereignisse.

Abbildung 7: Blick in das Labor des chemischen Instituts der Universität Leipzig im Jahre 1906 (Foto: Hermann-Walter (1838-1909))

2.1 Die drei Zeitalter der Strukturchemie

Wissenschaftlich lassen sich in den Daten drei verschiedene Perioden erkennen, deren Übergänge sich überraschenderweise ziemlich scharf abtrennen lassen: das protoorganische, das organische und das organometallische „Zeitalter“. Während der Übergänge zwischen den Zeitaltern waren die Wachstumsraten leicht gestört, aber die Chemie kehrte schnell zu ihrem Wachstumstrend von 4,4 Prozent zurück.

Im *protoorganischen Zeitalter* (1800–1860) und vor allem in den ersten Jahren, schwankte die jährliche Wachstumsrate noch recht stark. In diesem Zeitalter dominierten anfangs Kohlenstoff- und Wasserstoff-Verbindungen, später Verbindungen auf Kohlenstoff-, Wasserstoff-, Stickstoff-, Sauerstoff- und Halogenbasis. In dieser Zeit wurde der chemische Raum noch hauptsächlich durch Extraktion von Substanzen aus tierischen und pflanzlichen Produkten erweitert; Synthese fand wenn, dann basierend auf typisch anorganischen Verbindungen statt.

Im zweiten Zeitalter, dem *organischen Zeitalter* (1861–1980), wurden neue Substanzen bereits mit einer viel größeren Regelmäßigkeit entdeckt (Abbildung 7). Diese Regelmäßigkeit lässt sich wissenschaftlich auch mit der Einführung der Valenz- und Strukturtheorie in Verbindung bringen, die die Forschung in der organischen Chemie um 1860 veränderte. Ab da gingen die Chemiker in ihrer Arbeit wesentlich planmäßiger vor und die Neuentdeckungen waren dadurch weniger vom Zufall bestimmt als in der Epoche zuvor.

Im *organometallischen Zeitalter* (1981 bis heute) erlebten die Metallverbindungen ein Revival, sowohl als Ausgangsstoffe als auch als Endprodukte chemischer Reaktionen. Zehn Prozent der neuen Verbindungen basierten auf Platinmetallen. Siliciumverbindungen, bisher kaum beachtet in der Geschichte der Chemie, wurden zu Stars. Noch stärker als in der vorherigen Epoche wurde regelmäßig eine Verbindung nach der anderen synthetisiert. Ab etwa 1995 gewannen vor allem bioorganische Verbindungen an Interesse.

2.2 Das „Abgrasen“ des chemischen Universum

Interessanterweise wurden nicht nur neue Strukturen entdeckt, sondern auch neue Stoffzusammensetzungen, also die Menge an Elementen, aus denen die Moleküle bestehen (unabhängig von deren Verhältnis, vgl. *Summenformel*). Unangefochten auf Platz 1 stehen seit 1890 die Substanzen, die sich rein aus Kohlen-

stoff, Wasserstoff, Stickstoff und Sauerstoff zusammensetzen. Alles in allem wurde das „chemische Universum“ also bisher eher ohne rechten Plan abgegrast. Nur wenige der Stoffzusammensetzungen sind überhaupt ausführlich untersucht.

2.3 Einfluss gesellschaftlicher Ereignisse: Chemie in Zeiten der Weltkriege

Wie zu erwarten, hatten die beiden Weltkriege einen drastischen Einfluss auf die Wissenschaft und führten zum vorübergehenden Einbruch in der Anzahl der neu entdeckten chemischen Substanzen. Die Auswirkungen des Ersten Weltkriegs waren besonders verheerend, insbesondere weil sich die chemische Industrie und Forschung in der Zeit vor dem Ersten Weltkrieg auf Deutschland konzentrierte. Im Nachhinein führte der Krieg zu einem raschen Aufstieg der Chemie in anderen Ländern, insbesondere den USA.

Vielleicht auch deshalb waren die Auswirkungen des Zweiten Weltkrieges auf die Anzahl chemischer Neuentdeckungen weniger stark zu spüren. Nach beiden Kriegen erholte sich die chemische Forschung (so wie auch andere Forschungsgebiete) von diesen Rückschlägen relativ schnell und kehrte zu ihrer Wachstumskurve von etwa 4,4 Prozent pro Jahr zurück.

Die Kriege führten aber auch zu einer Verlagerung der chemischen Forschung. Während des Ersten Weltkriegs nahm unter anderem die Bedeutung von Arsen-, Antimon- und Bismut-Verbindungen zu, während die von Aluminium, Gallium, Indium und Thallium abnahm. Das Interesse an Arsen-Verbindungen erklärt sich vermutlich durch die verschiedenen im Ersten Weltkrieg entwickelten Arsen-Kampfstoffe. Während des Zweiten Weltkrieges wurden Stickstoffverbindungen und Alkalimetalle zunehmend uninteressant, während Schwefel, Bor, Phosphor und Silicium an Interesse gewannen. Phosphorverbindungen wurden vor allem relevant, als ihr Rolle in alltäglichen Anwendungen und als neuartige Insektizide und andere industrielle Materialien bekannt wurde.

2.4 Die Lieblinge der Chemiker

Interessanterweise, waren und sind die Chemiker recht konservativ bei der Wahl ihrer Ausgangsstoffe. Die meisten Substrate werden nur ein einziges Mal als Ausgangsstoff verwendet, die „Lieblingsstoffe“ jedoch immer wieder.

Ein möglicher Grund ist sicher die leichte Verfügbarkeit dieser Substanzen. In der Tat umfassen die meisten Reaktionen typischerweise zwei Ausgangsstoffe: eine weniger bekannte Substanz und einen „Klassiker“ aus dem Synthesewerkzeugkasten. Zu Beginn des 19. Jahrhunderts waren vor allem starke Säuren und Basen beliebt, später vor allem organische Substanzen.

Einer der Topstars ist Essigsäureanhydrid, das 1852 synthetisiert wurde und seit 1880 besonders gern für Acetylierungsreaktionen verwendet wird. Eine Acetylierung ist der Austausch von einem Wasserstoffatom durch eine Acetylgruppe (C_2H_3O), wobei entsprechende Verbindungen entstehen (zum Beispiel Aceton, Heroin oder Himbeerketon, die Hauptgeruchskomponente von Himbeeren).

2.5 Wie geht’s weiter?

Alles in allem wirft die Studie die Frage auf, warum die Chemie trotz gesellschaftlicher und wissenschaftlicher Störfaktoren eine so stabile Wachstumsrate von 4,4 Prozent beibehält?

In dem vorangegangenen Kapitel über „Kleine Moleküle im Fokus“ wurde primär über solche Moleküle berichtet, die im Metabolom eine Rolle spielen, in dem Reaktionsnetzwerk des Stoffwechsels von Zellen und Organismen. Beim Auffinden molekularer Strukturen hat sich in jüngerer Zeit aber auch ein Schwerpunkt auf das Proteom ausgebildet, also auf die Eiweißstoffe, die eine Zelle, ein Lebewesen ausmachen. Eiweiße sind sehr komplexe Moleküle, und für ihre Wirkung ist entscheidend, welche dreidimensionale Struktur sie einnehmen. Auf diesem Gebiet gibt es erstaunliche Fortschritte zu vermelden; von daher sei ihnen ein kurzer Abschnitt gewidmet.

2.5.1 Proteinstrukturdatenbanken im Auftrieb

Großen Auftrieb hatten in der jüngeren Vergangenheit Proteinstrukturdatenbanken. Man findet verschiedene Proteinstrukturdatenbanken wie zum Beispiel die Database of Macromolecular Movements, JenaLib, Proteopedia oder PDBWiki. Die bekannteste ist jedoch die Protein Data Bank (PDB). Erste Arbeiten dazu zur Aufstellung geeigneter Datenformate und anschließender graphischen Aufarbeitung gab es ab 1969 an der Texas A & M University, unterstützt vom Brookhaven National Laboratory. Die PDB wurde dann im Oktober 1971 in der Zeitschrift Nature New Biology als Joint Venture zwischen dem Cambridge Crystallographic Data Centre (Großbritannien) und dem Brookhaven National Laboratory (USA) angekündigt. Das ursprüngliche Format wurde noch durch die Breite der Computer-Lochkarten auf 80 Zeichen pro Zeile beschränkt; 1996 führte man schrittweise ein neues, leistungsfähigeres Format ein (macromolecular Crystallographic Information file format, abgekürzt mmCIF).

Zum 1. April 2020 umfasste die PDB 150 423 Proteine. Der weitaus größte Teil der Strukturen – 135 170 – wurde mit der Röntgenstrukturanalyse ermittelt. An zweiter Stelle kommen NMR-spektroskopische Methoden; sie ermöglichten bei 11 337 Proteinen die Strukturermittlung. Kryo-elektronenmikroskopisch fand man ab etwa 2013 3475 Strukturen; die restlichen 441 Strukturen wurden mit anderen bzw. hybriden Methoden ermittelt. Historisch gesehen ist die Anzahl der Strukturen in der PDB mit 100 registrierten Strukturen im Jahr 1982,

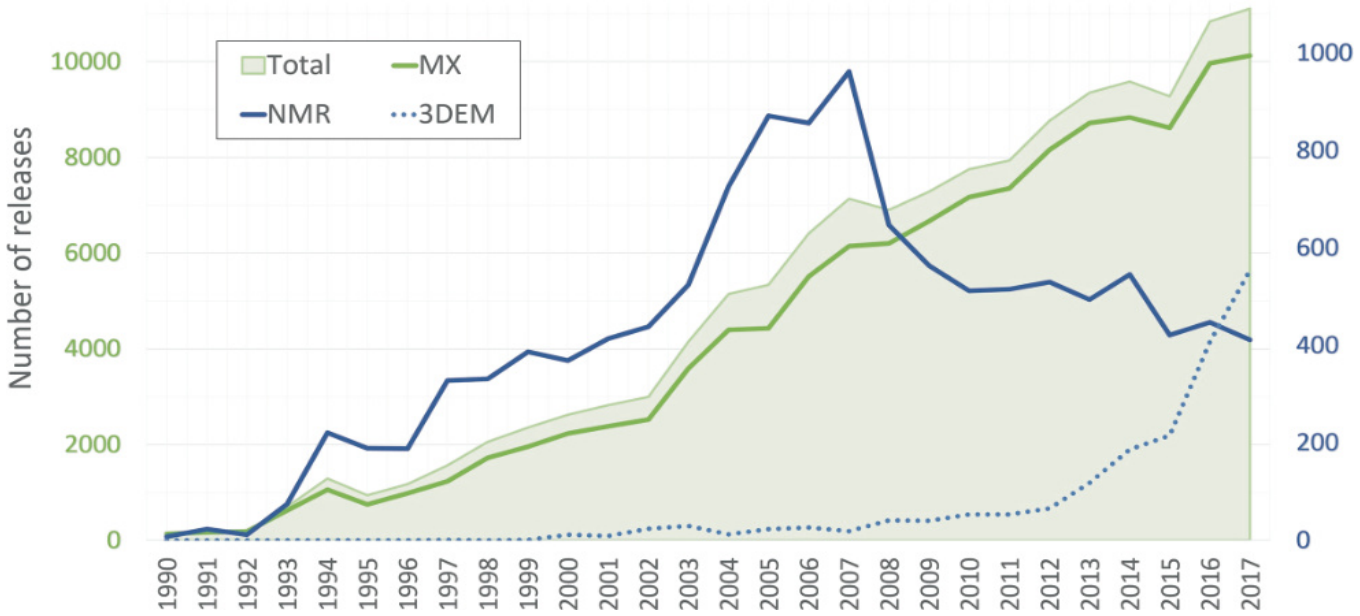


Abbildung 8: Anzahl der jährlich veröffentlichten PDB-Strukturen. Alle PDB Core Archive-Strukturen sind hellgrün schattiert, und MX-Strukturen (Makromolekulare Kristallografie) sind mit einer durchgezogenen grünen Linie dargestellt, die in Bezug auf die grüne Primärachse (links) aufgetragen ist. NMR-Strukturen (blaue durchgezogene Linie) und 3DEM-Strukturen (3D-Elektronenmikroskopie; blaue gestrichelte Linie) sind in Bezug auf die blaue Sekundärachse (rechts) aufgetragen (Abb.: wwPDB-Konsortium; CC 4.0 by).

1000 Strukturen im Jahr 1993, 10 000 im Jahr 1999 und 100 000 im Jahr 2014 annähernd exponentiell gewachsen (Abbildung 8). Seit 2007 befindet sich die Sammlung neuer Proteinstrukturen eher auf einem hohen Plateau.

2.5.2 Klärt AlphaFold alle Proteinstrukturen auf?

Ende November 2020 war es Google mit „AlphaFold 2“ gelungen, CASP14 zu gewinnen, noch viel deutlicher als es bereits Version 1 im Jahre 2018 schaffte (siehe den Artikel in CLB 11/12-2020: Ein komplexes Problem quasi gelöst – Die Faltung von Proteinen…“ S. 552-575). CASP ist ein Wettbewerb, in dem sich Proteinfaltungs-Simulationen messen können. CASP steht für „Critical Assessment of Techniques for Protein Structure Prediction“, also für die kritische Überprüfung von Techniken zur Vorhersage von Proteinstrukturen. Und AlphaFold ist ein von DeepMind entwickeltes Programm für künstliche Intelligenz. DeepMind dürfte aus einem anderen Zusammenhang her bekannt sein: Die jetzige Google-Tochter entwickelte mit AlphaGo (2015) und AlphaZero (2017) künstliche Intelligenzen, die auch den besten Go-Spielern der Welt das Fürchten lehren. Mit solch einem frühen Erfolg hatte niemand gerechnet, ist doch die Komplexität des Go-Spiels um Klassen größer als die des Schachspiels.

In diesem Juli hat Google nun erstmals Details zu diesem Werkzeug veröffentlicht und sogar den Quellcode freigegeben. Um die Datenbank einzurichten und zu betreiben, arbeitet DeepMind mit dem European Molecular Biology Laboratory (EMBL) zusammen, dessen Zentrale in Heidelberg ist. Zudem wurde bekannt, dass DeepMind mit Hilfe seiner KI die Formen fast aller Proteine im menschlichen Körper vorhersagen kann – und noch viel mehr: Auch die Form hunderttausender anderer Proteine, die in 20 der am häufigsten untersuchten Organismen vorkommen (u.a. Hefe, Fruchtfliege und Maus), ließe sich vorhersagen. Derzeit liegen laut DeepMind 350 000 neu vorhergesagten Strukturen vor. DeepMind gab an, dass es in den nächsten Monaten die Strukturen für mehr als 100 Millionen Proteine errechnen und veröffentlichen wird. Das ist so ziemlich alles dieser Art, was derzeit überhaupt bekannt ist… Man darf gespannt sein!

Literatur / Quellen

Kai Dührkop, Huibin Shen, Marvin Meusel, Juho Rousu and Sebastian Böcker: Searching molecular structure databases with tandem mass spectra using CSI:FingerID; Proc Natl Acad Sci USA, 112(41):12580-12585, 2015.

Florian Rasche, Aleš Svatoš, Ravi Kumar Maddula, Christoph Böttcher and Sebastian Böcker: Computing fragmentation trees from tandem mass spectrometry data; Anal Chem, 83(4):1243-1251, 2011.

Sebastian Böcker, Matthias Letzel, Zsuzsanna Lipták and Anton Pervukhin: SIRIUS: Decomposing isotope patterns for metabolite identification; Bioinformatics, 25(2):218-224, 2009.

Dührkop, K., Nothias, LF., Fleischauer, M. et al. Systematic classification of unknown metabolites using high-resolution fragmentation mass spectra. Nat Biotechnol 39, 462–471 (2021)

EJ Llanos, W Leal, DH Luu, J Jost, PF Stadler, G Restrepo: Exploration of the chemical space and its three historical regimes; Proc Natl Acad Sci USA, pii: 201816039, 2019

Der neueste Stand – aktuelle Informationen zu früheren CLB-Artikeln

Erste ambulante CAR-T-Zell-Therapie

Zu CLB 3/4-2020, 120-128: CAR-T-Zellen enttarnen Krebs

Bereits in der vorangegangenen CLB 5/6-2021 berichteten wir zum neuesten Stand bei CAR-T-Zellen (CAR: Chimeric Antigen Receptor). Das Gebiet macht schnelle Fortschritte, sodass hier nochmals die Thematik angesprochen wird.

Bislang war für diese Therapie ein mehrwöchiger stationärer Aufenthalt erforderlich. Nun möchte man in München an der Medizinischen Klinik III diese Therapie im ambulanten Setting etablieren. Es kommen derzeit jedoch nur bestimmte Krebspatienten dafür in Frage. Diese werden im Vorfeld nach festgelegten medizinischen Kriterien untersucht und für die ambulante CAR-T-Zell-Therapie im Rahmen einer klinischen Studie ausgewählt. Ambulant bedeutet, dass die Patienten noch am selben Tag der Gabe der gentechnisch veränderten Zellen nach Hause gehen können. Ein Spezialteam der Immunotaskforce der Medizinischen Klinik III kümmert sich um die ambulante (Nach)-Betreuung und Überwachung. Das Studiensetting sieht bestimmte Vorgaben dafür vor: In den ersten Wochen kommt der Patient jeden Tag in die Klinik. Wenn er stabil ist und die Therapie ohne Nebenwirkungen verträgt, greift ab der 3. Woche zusätzlich zu den dann seltener werdenden Besuchen in der Klinik ein Telefonmonitoring. Die Studienleiterin Marion Subklewe relativiert jedoch noch die Erwartungen: „Die ambulante CAR-T-Zell-Therapie ist ein wichtiger Schritt, diese neue Therapieform noch mehr Patienten zur Verfügung zu stellen. Bislang ist die ambulante Therapie aber noch Studienpatienten vorbehalten. Bis die ambulante Gabe zum Standardverfahren geworden ist, wird es aber sicherlich noch einige Zeit dauern."

Weltweit erstmals CAR-T-Zellen erfolgreich gegen Autoimmunerkrankung eingesetzt

Zu CLB 3/4-2020, 120-128: CAR-T-Zellen enttarnen Krebs

Mit Gelenkschmerzen und einem roten Gesichtsausschlag fing es an: Die damals 16-jährige Thu-Thao V. hatte bereits mehrere ärztliche Untersuchungen in drei Städten hinter sich, als sie im Februar 2017 am Universitätsklinikum Erlangen die Diagnose erhielt: Systemischer *Lupus erythematodes* (SLE). Bei der lebensbedrohlichen Autoimmunerkrankung, die vor allem junge Frauen betrifft, greift das Immunsystem eigene Körperzellen in verschiedenen Organsystemen an. Nachdem auch unterschiedliche immununterdrückende Therapien die Symptome der jungen Frau nicht nachhaltig verbessern konnten, wurde Thu-Thao V. von Forschenden des Deutschen Zentrums Immuntherapie (DZI) des Uni-Klinikums Erlangen im März 2021 ein Präparat mit CAR-T-Zellen verabreicht. Knapp ein halbes Jahr nach der Zelltherapie ist nun gewiss: Die Gelenkschmerzen sind verschwunden, der Organismus der 20-Jährigen hat sich komplett erholt, kann sogar wieder normal Sport machen.

Beim SLE spielen Teile des Immunsystems verrückt und bilden Antikörper gegen die eigene Erbsubstanz, was unweigerlich zu schweren Entzündungsreaktionen in den Organen führt. In der schlimmsten Phase der Erkrankung musste die Fachoberschülerin jeden Tag knapp 20 Tabletten einnehmen, damit ihr Körper die Strapazen ihres fehlgeleiteten Immunsystems kompensieren konnte. Zu den Gelenkschmerzen kamen auch Wassereinlagerungen durch meine Niereninsuffizienz, starkes Herzklopfen und Haarausfall. Nach einem akuten Schub waren die Beschwerden besonders schlimm. Alle Therapien, die darauf abzielten, das fehlgesteuerte Immunsystem der jungen Patientin zu unterdrücken, scheiterten.

Im Falle der jungen SLE-Patientin wurde wählte man daher eine Therapie mit chimären Antigenrezeptoren (CAR). Immunzellen, also T-Zellen der Patientin wurden im Labor mithilfe eines gentechnischen Verfahrens mit dem CAR ausgestattet. Dieser erkennt spezielle Antigene auf der Oberfläche der Zielzellen und zerstört diese. Den CAR-T-Zellen wurde die Fähigkeit beigebracht, diejenigen Immunzellen (B-Zellen) unschädlich zu machen, die Antikörper gegen körpereigene Zellen bilden.

Im März 2021 erhielt Thu-Thao V. als weltweit erste Patientin mit SLE CAR-T-Zellen. Ergebnis: Die Mediziner waren sehr überrascht, wie schnell sich der Zustand der Patientin unmittelbar nach der Zellinfusion besserte. Die CAR-T-Zellen haben ihre Aufgabe ausgezeichnet erledigt und haben die krankheitsvermittelnden B-Zellen rasch zerstört. Zusammen mit den Antikörpern gegen die eigene Erbsubstanz verschwanden auch alle Krankheitssymptome des SLE. Thu-Thao V. konnte alle immununterdrückenden Medikamente inklusive Kortison absetzen. Die Patientin ist nun seit knapp einem halben Jahr vollkommen beschwerdefrei, bisher gibt es keine Anzeichen für ein erneutes Auftreten der Erkrankung. Die Mediziner planen nun eine klinische Studie mit CAR-T-Zellen bei Patientinnen und Patienten mit Autoimmunerkrankungen.

Veröffentlichung: G. Krönke et al.: CD19-Targeted CAR T Cells in Refractory Systemic Lupus Erythematosus; N Engl J Med; 385:567-569, 5 August 2021.