

## 17 Mathematik in der Chemie: Die Suche nach Synthesewegen für neue Medikamente

Nicht nur Zuschauer der Serie *Breaking Bad* sondern auch viele Wissenschaftler halten chemische Synthese für eine Art Kunstform. Gute Chemische Synthesewege zu finden, ist eine große Herausforderung bei der Suche nach neuen Wirkstoffen in Medizin oder Pflanzenschutz. Dass Mathematik und künstliche Intelligenz hier weiterhelfen können, überrascht vielleicht.

Dr. Georg Mogk, Principal Expert Applied Mathematics, Bayer AG, Leverkusen und Priv. Doz. Dr. Thomas Mrziglod, Head of Applied Mathematics, Bayer AG, Leverkusen geben ein Beispiel zur Rolle der Mathematik in der Chemie.

Mathematik haftet der Ruf an, sehr abstrakt und vielleicht auch etwas weltfremd zu sein. Das dem nicht so ist, wollen wir im Folgenden am Beispiel der Chemie zeigen. Hier kann Mathematik dabei helfen, denkbare Moleküle (z.B. als Wirkstoffe von Medikamenten) in die Realität zu holen. Das Bild des Chemikers, der in seinem Labor herumwerkelt und dabei sensationelle neue Wirkstoffe findet, gehört immer mehr der Vergangenheit an. Tatsächlich werden potenzielle Wirkstoffe am Computer auf eine bestimmte Wirkung hin designt. Hierbei werden Algorithmen verwendet und dabei oft hunderttausende Substanzen generiert, die eine bestimmte Wirkung erzielen könnten. Für die vielversprechendsten Kandidaten muss dies dann natürlich noch im Labor überprüft werden. Dazu muss jeder Wirkstoffkandidat in ausreichenden Mengen synthetisiert werden.



Abbildung 17: Synthese Roboter bei Bayer

Die Vorhersage, ob sich ein Molekül tatsächlich synthetisieren lässt, ist allerdings sehr schwierig. Experten schätzen, dass auf ein synthetisierbares Molekül mindestens eine Milliarde Moleküle kommen, die zwar existieren könnten, aber mit dem heutigen Stand der Chemie nicht synthetisiert werden können. Dies führt uns zu zwei Fragen: *kann man vorhersagen, ob ein am Rechner designtes Molekül synthetisierbar ist?* Und wenn ja, *wie kann es synthetisiert werden?* Das Konzept, mit dem man diese Fragen beantworten kann, heißt Retro-Synthese und wurde ab den späten 1960er Jahren von dem amerikanischen Chemiker Elias James Corey entwickelt.

Bei der Retro-Synthese geht man gedanklich jede Bindung eines Moleküls durch, scheidet sie durch und fragt sich, durch welches Reaktionsschema die Bindung entstehen könnte. Weiter fragt man, wie die Ausgangsprodukte (auch Edukte genannt) aussehen müssten, damit bei einer Vorwärts-Synthese genau die gesuchte Bindung entsteht. Rekursiv wendet man nun das gleiche Verfahren auf die gerade ermittelten Edukte an, bis man schließlich bei kaufbaren Molekülen oder in einer Sackgasse landet. Das Ergebnis ist dann der sogenannte Synthesebaum, der alle denkbaren Synthesewege enthält. Der Chemiker muss dann *nur* aus dem Synthesebaum die beste Syntheseroute auswählen. Für das Konzept der Retro-Synthese erhielt Corey 1990 den Nobelpreis für Chemie.

So einfach das Konzept sich anhört, so hat es doch seine Tücken, die sich in zwei Kernfragen zusammenfassen lassen:

- Wo und wie kann man überhaupt ein Molekül schneiden?
- Die Synthesebäume können sehr groß werden. Gibt es einen Syntheseweg, der ausschließlich bei kaufbaren Bausteinen landet?

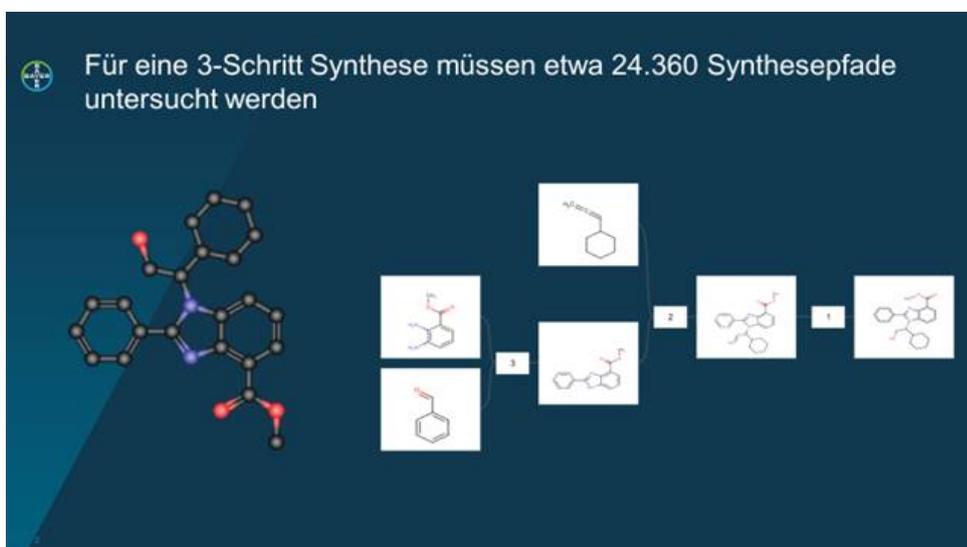


Abbildung 18: Drei-stufige Syntheseroute

Hier kommt nun Mathematik ins Spiel. Da für beide Fragen extrem viele Möglichkeiten bestehen, bieten sich Computer an, um diese anzugehen. Regelbasierte Ansätze für die erste Frage stoßen auf Grund des explorierenden chemischen Wissens schnell an ihre Grenzen. Ein modernerer Ansatz ist es, den Computer die Schneideregeln selbst lernen zu lassen. Ein Update des Systems bei Weiterentwicklung der Chemie ist dann sehr einfach: Man lässt den Computer auf den neuen Daten lernen. Innerhalb der Bayer AG verfolgen wir diesen Ansatz. Dafür haben wir 19 Mio. veröffentlichte Reaktionen der letzten 250 Jahre und unsere eigenen Laborjournale genutzt, um ein Künstliches Neuronales Netz zu trainieren.

Bevor der Computer lernen kann, sind aber einige Hindernisse zu überwinden: Die chemische Nomenklatur für Moleküle und Reaktionen ist für Menschen sehr gut zugänglich, sie muss aber erst in eine für Computer geeignete Darstellung von Molekülen überführt werden. Auch bei den zu lernenden Reaktionen gibt es Herausforderungen: In Publikationen fokussieren sich Forscher auf das Wesentliche. „Unwichtige“ Nebenprodukte, wie Wasser oder CO<sub>2</sub> werden oft nicht dokumentiert. Oder in den Reaktionen spielen doch plötzlich Atome aus dem Lösungsmittel eine Rolle, die in den Reaktionsgleichungen aber nicht auftauchen. Kurz: in vielen publizierten Reaktionen stimmen die Massenbilanzen nicht. Dies bereitet Rechnern beim richtigen Verstehen von Reaktionen große Probleme und man muss mit intelligenten mathematischen Algorithmen aushelfen (Stichwort *Atom-Mapping*).

Eine weitere Herausforderung liegt in dem Publikationsverhalten, da üblicherweise nur funktionierende Reaktionen veröffentlicht werden. Wenn wir Menschen lernen, lernen wir aber auch aus dem, was nicht funktioniert. Das gilt auch für Computer: „Der Mensch lernt aus eigenen Fehlern, Computer aus den Fehlern anderer“. Dabei geht es vor allem darum, *sinnvolle* nicht funktionierende Reaktionen zu finden. Das sind solche, bei denen einem Experten nicht sofort klar ist, dass die Reaktion nicht funktioniert. Solche Reaktionen finden wir beispielsweise in den Laborjournalen von Bayer. Bayer ist über 150 Jahre alt und die chemische Forschung bei Bayer hat in den letzten Jahren etwa 12 Mio. elektronische Laborjournaleninträge hervorgebracht. Diese müssen allerdings ausgelesen und interpretiert werden. Für die automatische Interpretation der Laborjournaleninträge verwenden wir ebenfalls ein neuronales Netz. Das Neuronale Netz liest alle 12 Mio. Laborbucheinträge und bewertet jeden einzelnen, ob die dort beschriebene Reaktion funktioniert hat oder nicht.

Nach all diesen Vorbereitungen haben wir einen maschinen-lesbaren Datensatz aus funktionierenden und nicht-funktionierenden Reaktionen. Mit diesen Daten trainieren wir ein Neuronales Netz für eine sogenannte Ein-Schritt-Retrosynthese.

Damit hätten wir eine moderne Antwort auf die erste der beiden obigen Fragen, *wo schneide ich?* Kommen wir nun zu der zweiten Frage: *Wie finden wir einen Syntheseweg, der bei bekannten oder kaufbaren Ausgangsstoffen landet?* Bei der Beantwortung dieser Frage haben wir mit der sogenannten *kombinatorischen Explosion* zu kämpfen. Auch wenn wir uns hier mit der Synthese von sogenannten *small molecules* beschäftigen, können diese bis zu 100 Verbindungen zwischen Schweratomen haben. Nehmen wir beispielsweise ein Molekül mit 100 Verbindungen, welches wir in 10 Schritten syn-

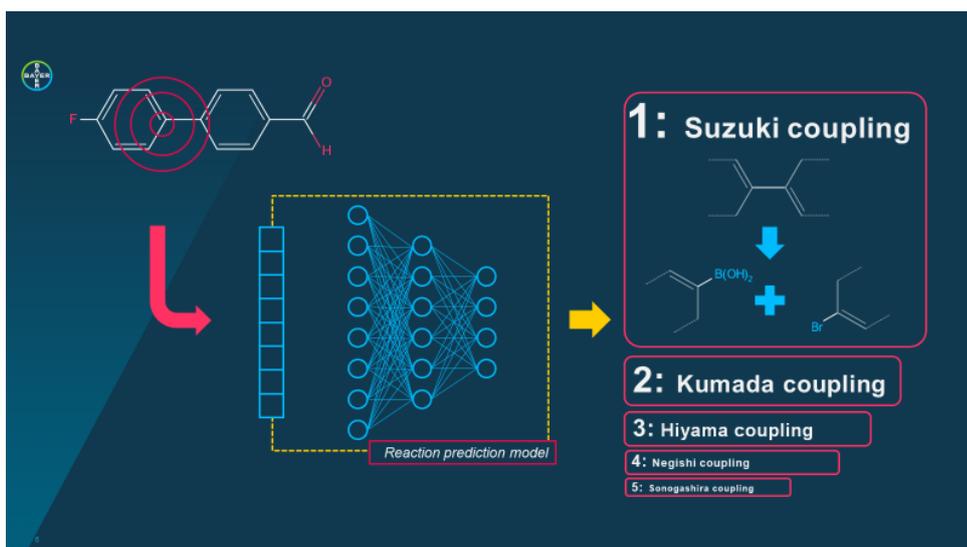


Abbildung 19: Ein neuronales Netz zur Vorhersage geeigneter Reaktionsmechanismen

thetisieren wollen. Dann müssten wir  $5 \cdot 10^{21}$  mögliche Syntheserouten untersuchen. Zum Vergleich das Universum existiert erst seit ca.  $7,8 \cdot 10^{18}$  Sekunden. Aber reichen 10 Stufen?

Um dieser enormen Vielfalt Herr zu werden, wenden wir Algorithmen an, die für Strategiespiele, wie Schach oder GO, entwickelt wurden. Wie bei Strategiespielen ist es im Allgemeinen nicht möglich, den gesamten Baum zu erkunden. Daher werden effiziente Algorithmen benötigt, um diese Art von Bäumen zu erkunden und möglichst schnell gute Lösungen zu liefern.

Eines der schönen Dinge der Mathematik ist, dass sie einen immer wieder überrascht, wo man Ergebnisse aus anderen Teildisziplinen der Mathematik anwenden kann. Hier verwenden wir Ergebnisse der Spieltheorie für die chemische Synthese.

An dem Beispiel der Retro-Synthese sieht man gut, wie die Arbeit eines modernen Mathematikers aussieht: Sie ist interdisziplinär und teamorientiert. Nur durch die Zusammenarbeit verschiedener Disziplinen können Projekte wie dieses erfolgreich umgesetzt werden. In dem Projektteam finden sich neben Mathematikern und Chemikern auch Chem-Informatiker und Computer Linguisten. Die Probleme müssen in Mathematik übersetzt werden, werden dann mit Mathematik gelöst, und die Lösung muss zurück in die Fachsprache der Fragesteller übertragen werden.