

Abb. 2. Molekülstrukturen von **2** und **6** (abgebildet) im Kristall. Ausgewählte Bindungslängen [Å] und -winkel [°]; Werte für **6** in Klammer: M'-N(1) 2.066(6) [1.855(2)], M'-N(2) 2.082(7) [1.885(5)], N(1)-C(1) 1.449(11) [1.444(9)]; N(1)-M'-N(2) 101.8(3) [104.7(2)], M'-N(2)-Si(2) 115.3(4) [116.8(3)], M'-N(1)-C(1) 124.8(5) [126.1(4)].

ne auf der NMR-Zeitskala schnelle Rotation um die C-N-Bindungen denkbar. Die M'-N-Abstände und N-M'-N-Winkel sind vergleichbar mit denen in kristallinem $\text{Sn}(\text{NR}_2)_2$ [2.096(1), 2.088(6) Å; 104.7(2)°]^[2] und denen in kristallinem $\text{Ge}(\text{NR}_2)_2$ [1.878(5), 1.873(5) Å; 107.1(2)°]^[1]. In **2** und **6** ist die M'-N(Aryl)-Bindung deutlich kürzer als die M'-N(R_2)-Bindung; letztere ist geringfügig kürzer als die entsprechenden Bindungen in $\text{M}'(\text{NR}_2)_2$.

Die neuen verbrückenden Liganden 1,4- $\text{C}_6\text{H}_4(\text{NR})_2$ könnten in der Koordinationschemie nützlich sein, insbesondere bei der Fragestellung, ob die stickstoffgebundenen metallhaltigen Einheiten miteinander „kommunizieren“ können, was bei aromatischen Verbindungen durchaus denkbar wäre.

Experimentelles

2: 1,4-Bis(trimethylsilylamino)benzol [12] (8.0 g, 31 mmol) in 70 mL Hexan wird bei 25 °C mit 2 Äquiv. $n\text{BuLi}$ (in n -Hexan, 41 mL, 62 mmol) behandelt. Die Mischung wird 12 h gerührt, die flüchtigen Bestandteile im Vakuum entfernt, wobei 1,4- $\text{C}_6\text{H}_4[\text{N}(\text{R})\text{Li}]_2$ als gelber Feststoff zurückbleibt (8.1 g, 98%). Dieser (1.3 g, 4.92 mmol) wird bei 25 °C mit $[\text{Sn}(\mu\text{-Cl})(\text{NR}_2)_2]$ [8] (3.09 g, 4.92 mmol) in 40 mL n -Hexan umgesetzt, die Mischung 2 h gerührt und filtriert. Flüchtige Komponenten werden im Vakuum entfernt, der Rückstand mit 10 mL Diethylether extrahiert. Nach 2 d werden bei -30 °C orange Kristalle von **2** isoliert (2.78 g, 70%); korrekte Elementaranalyse.

1 und **6:** Wie für **2** beschrieben, jedoch mit $\text{trans-1,4-C}_6\text{H}_4[\text{N}(\text{R})\text{Li}]_2$ und $[\text{Sn}(\mu\text{-Cl})(\text{NR}_2)_2]$ (**1**) oder mit 1,4- $\text{C}_6\text{H}_4[\text{N}(\text{R})\text{Li}]_2$ und $[\text{GeCl}(\text{NR}_2)(\text{tmada})]$ (**6**) hergestellt. **6** entsteht in Form farbloser Kristalle (2.86 g, 86%) durch Mischen von $[\text{GeCl}_2(\text{dioxan})]$ (2.00 g, 8.64 mmol) mit LiNR_2 (1.45 g, 8.64 mmol) und $[(\text{Me}_2\text{NCH}_2)_2(\text{tmada})]$ (1.00 g, 17.28 mmol) in 30 mL Et_2O (16 h, 25 °C) nach Entfernen der flüchtigen Komponenten im Vakuum und nach Extraktion mit Toluol/Hexan (1:1).

4: Zu **2** (1.5 g, 19.00 mmol) und $\text{Sn}(\text{OAr})_2$ (2.1 g, 37.0 mmol) werden bei 25 °C 40 mL Toluol gegeben. Die hellbraune Reaktionsmischung wird 16 h gerührt, dann filtriert. Das Filtrat wird im Vakuum eingedunstet und gekühlt (-5 °C), wobei kristallines **4** (0.86 g, 50%) erhalten wird; korrekte Elementaranalyse.

Eingegangen am 13. Januar 1994 [Z 6614]

- [1] R. W. Chorley, P. B. Hitchcock, M. F. Lappert, W.-P. Leung, P. P. Power, M. M. Olmstead, *Inorg. Chim. Acta* **1992**, 198–200, 203.
- [2] T. Fjeldberg, H. Hope, M. F. Lappert, P. P. Power, A. J. Thorne, *J. Chem. Soc. Chem. Commun.* **1983**, 639.
- [3] B. Çetinkaya, I. Gümrükçü, M. F. Lappert, J. L. Atwood, R. D. Rogers, M. J. Zaworotko, *J. Am. Chem. Soc.* **1980**, 102, 2088.
- [4] P. B. Hitchcock, M. F. Lappert, B. J. Samways, E. L. Weinberg, *J. Chem. Soc. Chem. Commun.* **1983**, 1492.
- [5] T. Fjeldberg, A. Haaland, B. E. R. Schilling, M. F. Lappert, A. J. Thorne, *J. Chem. Soc. Dalton Trans.* **1986**, 1551.

- [6] D. E. Goldberg, P. B. Hitchcock, M. F. Lappert, K. M. Thomas, A. J. Thorne, T. Fjeldberg, A. Haaland, B. E. R. Schilling, *J. Chem. Soc. Dalton Trans.* **1986**, 2387.
- [7] a) H. Braunschweig, R. W. Chorley, P. B. Hitchcock, M. F. Lappert, G. A. Lawless, *J. Chem. Soc. Chem. Commun.* **1992**, 1311; b) P. Jutzi, A. Becker, H. G. Stammer, B. Neumann, *Organometallics* **1991**, 10, 1647.
- [8] R. W. Chorley, P. B. Hitchcock, B. S. Jolly, M. F. Lappert, G. A. Lawless, *J. Chem. Soc. Chem. Commun.* **1991**, 1302.
- [9] Kristalldaten von **1**: $M = 814.7$, triklin $P\bar{1}$, $a = 8.448(3)$, $b = 10.273(4)$, $c = 12.160(5)$ Å, $\alpha = 100.63(3)$, $\beta = 106.73(3)$, $\gamma = 92.53(3)^\circ$, $V = 988.0$ Å³, $Z = 1$, $\rho_{\text{ber.}} = 1.37$ g cm⁻³, $F(000) = 420$, $\mu = 14.7$ cm⁻¹, $T = 173$ K. 3241 unabhängige Reflexe, $R = 0.027$, $R_w = 0.040$, $S = 1.6$, 163 Parameter. **2**: $M = 808.7$, monoklin, $C2/c$ (Nr. 15), $a = 10.891(2)$, $b = 14.229(3)$, $c = 25.747(5)$ Å, $\beta = 94.08(1)^\circ$, $V = 3979.7$ Å³, $Z = 4$, $\rho_{\text{ber.}} = 1.35$ g cm⁻³, $F(000) = 1656$, $\mu = 14.6$ cm⁻¹, $T = 293$ K. 1924 unabhängige Reflexe, $R = 0.058$, $R_w = 0.074$, $S = 2.1$, 163 Parameter. **6**: $M = 716.5$, monoklin, $C2/c$ (Nr. 15), $a = 10.698(5)$, $b = 14.317(8)$, $c = 25.592(7)$, $\beta = 93.39(3)^\circ$, $V = 3912.7$ Å³, $Z = 4$, $\rho_{\text{ber.}} = 1.22$ g cm⁻³, $F(000) = 1512$, $\mu = 17.1$ cm⁻¹, $T = 293$ K. 3597 unabhängige Reflexe, $R = 0.056$, $R_w = 0.065$, $S = 1.7$, 163 Parameter. Meßbereich jeweils $2 < \theta < 25^\circ$, monochromatisierte MoK_α -Strahlung ($\lambda = 0.71069$ Å), Enraf-Nonius-CAD4-Diffraktometer, Strukturlösung mit Schweratommethoden (SHELXS 86 [10]) mit $I \geq 2\sigma I$. Absorptionskorrektur mit DIFABS [11]. Weitere Einzelheiten zu den Kristallstrukturuntersuchungen können beim Direktor des Cambridge Crystallographic Data Centre, 12 Union Road, GB-Cambridge CB2 1EZ, unter Angabe des vollständigen Literaturzitats angefordert werden.
- [10] G. M. Sheldrick in *Crystallographic Computing 3* (Hrsg.: G. M. Sheldrick, C. Krüger, R. Goddard), Oxford, **1985**, S. 175.
- [11] N. Walker, D. Stuart, *Acta Crystallogr. Sect. A* **1983**, 39, 158.
- [12] L. Birkoff, H. P. Kuhltal, A. Ritter, *Chem. Ber.* **1969**, 93, 2810.

Die Bestimmung größter deckungsgleicher Teilstrukturen mit einem genetischen Algorithmus: Anwendung in der Syntheseplanung und zur strukturellen Analyse biologischer Aktivität

Markus Wagener und Johann Gasteiger*

In der Chemie müssen häufig die Strukturen zweier Verbindungen verglichen werden: bei der Analyse biologischer Aktivität, der Suche nach geeigneten Ausgangsmaterialien für die Synthese einer Zielverbindung, der Aufklärung von Biosynthesewegen sowie der Bestimmung der direkt an einem Reaktionsgeschehen beteiligten Atome und Bindungen. Eine wesentliche Hilfe ist dabei die Bestimmung der größten Teilstruktur, die den zu vergleichenden Molekülen gemeinsam ist (maximum common substructure, MCSS). Dementsprechend spielt die MCSS-Suche eine wichtige Rolle in der computerunterstützten Verarbeitung chemischer Information.

Dies spiegelt sich auch in der großen Zahl von Arbeiten wider, in denen Lösungen des MCSS-Problems vorgeschlagen werden^[1]. Alle Verfahren beruhen letztlich darauf, die Menge der möglichen Lösungen zu durchmustern, wobei unterschiedliche Heuristiken zur Begrenzung der Suche eingesetzt werden. Wir zeigen hier, daß das MCSS-Problem alternativ mit einem genetischen Algorithmus gelöst werden kann. Dadurch läßt sich die zeitaufwendige systematische Suche vermeiden, was besonders bei größeren Strukturen von Vorteil ist.

Genetische Algorithmen (GA) sind Optimierungsverfahren, die Mechanismen imitieren, mit denen sich die Natur an eine sich ändernde Umgebung anpaßt^[2]. Für ein gegebenes Problem

[*] Prof. Dr. J. Gasteiger, Dr. M. Wagener
Computer-Chemie-Centrum der Universität Erlangen-Nürnberg
Nägelbachstraße 25, D-91052 Erlangen
Telefax: Int. + 9131/85-6566

wird zu Beginn der Optimierung eine Anzahl zufällig bestimmter Lösungen („Population“) erzeugt, die alle Annäherungen an das gesuchte Optimum darstellen, sich aber in ihrer Güte („Fitness“) unterscheiden. Auf diese Population werden iterativ eine Reihe genetischer Operatoren angewendet, die häufig ein Vorbild in der natürlichen Genetik haben (Abb. 1). Der Selektionsoperator setzt die Population einer jeden neuen Generation bevorzugt aus solchen Lösungsversuchen (Individuen) der vorangegangenen Generation zusammen, die eine hohe Fitness haben. Dadurch wird der Anteil schlechter Lösungen, also solcher mit unterdurchschnittlicher Fitness, in der Population zunehmend geringer. Die restlichen Operatoren bilden ausgehend von der vorhandenen Population neuartige Individuen mit potentiell besseren Eigenschaften.

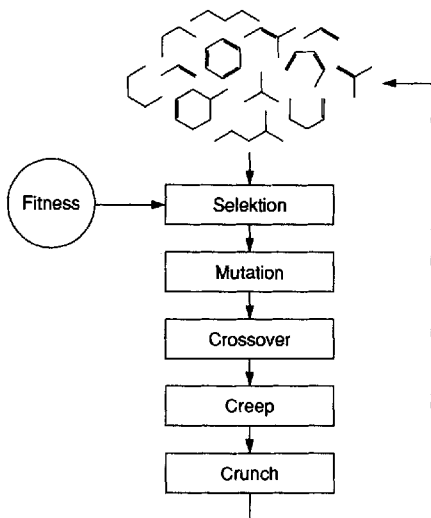


Abb. 1. Ein Überblick über den verwendeten genetischen Algorithmus. Die Rechtecke versinnbildlichen die genetischen Operatoren, und die Population ist durch Strukturfragmente angedeutet.

Genetische Algorithmen sind robust, d.h. auf eine Vielfalt von Problemen anwendbar^[2]. In der Chemie wurden sie allerdings bisher – außer für chemometrische Verfahren^[3] und zur Bestimmung der minimalen chemischen Distanz^[4] – noch wenig genutzt.

Bei der Suche nach der größten gemeinsamen Substruktur zweier Verbindungen mit einem genetischen Algorithmus werden Bindungen der betrachteten Moleküle einander zugeordnet. Diese Zuordnung entspricht einem Individuum der GA-Population. Sie kann einer gemeinsamen Substruktur entsprechen, muß es aber nicht. Die Fitness F einer derartigen Zuordnung wird im vorliegenden Verfahren gemäß (a) berechnet. N

$$F = N - V - (T_1 + T_2 - 2) \quad (a)$$

steht dabei für die Zahl der Bindungen, die in beiden Strukturen Bestandteil der Zuordnung sind. V gibt an, wie häufig es vorkommt, daß zwei benachbarte Bindungen der einen Struktur auf zwei Bindungen der anderen Struktur abgebildet werden, die nicht benachbart sind; je größer V also ist, um so schlechter passen die beiden Teilstrukturen aufeinander. T_1 und T_2 stehen für die Zahl der nicht miteinander verbundenen Teile in den beiden Strukturen. Die Konstante 2 wurde eingeführt, damit F die Zahl der aufeinander abgebildeten Bindungen wiedergibt, wenn eine zusammenhängende und passende Teilstruktur gefunden worden ist. Der obere Teil von Abbildung 2 zeigt den

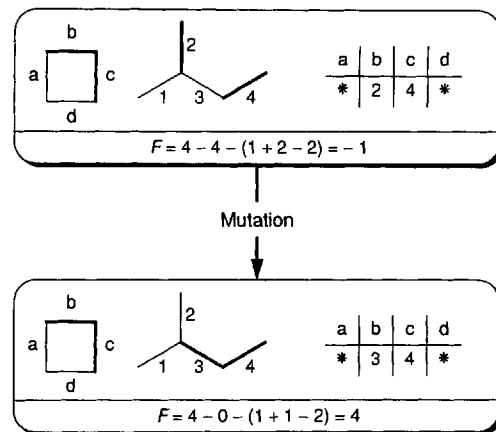


Abb. 2. Der Mechanismus des Mutationsoperators: Die Zuordnung der Bindung b im Cyclobutan zur Bindung 2 im Methylbutan wird in eine Zuordnung zur Bindung 3 umgewandelt. Dies erhöht den Fitness-Wert von -1 auf $+4$.

Vergleich von Cyclobutan mit Methylbutan. Die fett gezeichneten Bindungen sind einander zugeordnet. Dabei ist also, wie auch rechts von den beiden Strukturen in Form einer Tabelle zu sehen ist, Bindung b auf Bindung 2 und Bindung c auf Bindung 4 abgebildet. Den übrigen Cyclobutanbindungen sind keine Methylbutanbindungen zugeordnet, was durch einen Stern in der Tabelle angegeben wird. Die Fitness von -1 setzt sich hier wie folgt zusammen: Vier Bindungen werden insgesamt aufeinander abgebildet ($N = 4$). Im Cyclobutan sind die Bindungen benachbart, im Methylbutan nicht, so daß alle vier Bindungen die Nachbarschaftsbeziehungen verletzen ($V = 4$). Die Substruktur im Cyclobutan besteht aus einem einzigen Teil ($T_1 = 1$), wohingegen die im Methylbutan zwei Fragmente umfaßt ($T_2 = 2$).

Die Population aus solchen Individuen, also aufeinander abgebildeten Strukturpaaren, wird mit Hilfe von fünf genetischen Operatoren in eine neue Population übergeführt: Selektion, Mutation, Crossover, Creep und Crunch^[5].

Zur Selektion wird ein Verfahren verwendet^[6], das stochastische Fehler gegenüber denen anderer Selektionsverfahren reduziert. Mutation ändert willkürlich die Abbildung eines Bindungspaares. In Abbildung 2 wird der Bindung b nach der Mutation nicht mehr die Bindung 2 des Methylbutans zugeordnet, sondern die Bindung 3. Dadurch steigt der Wert der Fitness von -1 auf $+4$. Crossover tauscht Teillösungen in Form von Teilstrukturzuordnungen zwischen zwei Individuen aus. Abbildung 3 zeigt, wie dabei die Zuordnung der Bindung c zur Bin-

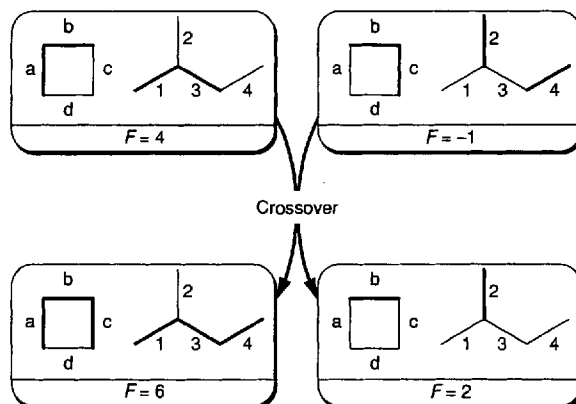


Abb. 3. Der Mechanismus des Crossover-Operators: Die Zuordnung der Bindung c zur Bindung 4 wird vom zweiten Strukturpaar (Individuum) auf das erste übertragen.

dung 4 vom rechten auf das linke Individuum übertragen und so die gemeinsame Substruktur vergrößert wird.

Damit die MCSS-Suche möglichst effizient durchgeführt werden kann, wurden zwei weitere, speziell auf das Problem zugeschnittene Operatoren entwickelt, Creep und Crunch. Sie nutzen im Gegensatz zu den klassischen Operatoren Mutation und Crossover Informationen über die Konstitution der betrachteten Strukturen, um gezielt zu Individuen mit höherer Fitness zu gelangen. So vergrößert der Creep-Operator eine Zuordnung der Bindungen um ein Bindungspaar, sofern dies möglich ist, ohne Nachbarschaftsbeziehungen innerhalb der beiden Strukturen zu verletzen (vgl. Summand V der Fitnessfunktion (a)). Dies ist besonders hilfreich, wenn nur noch ein Bindungspaar zur größten gemeinsamen Substruktur fehlt, da der letzte Schritt zum globalen Maximum vom Creep-Operator leicht vorgenommen werden kann.

Der Crunch-Operator verringert im Gegensatz dazu die Zahl der Bindungen, die einander zugeordnet werden. Ist eine Teilstruktur nicht zusammenhängend, so werden alle Bindungen, die am kleinsten Fragment beteiligt sind, vom Crunch-Operator vollständig aus der Zuordnungsliste entfernt. Dadurch können lokale Maxima, die häufig von kleinen, nicht zusammenhängenden Fragmenten geprägt sind, sicher wieder verlassen werden.

Zu starke Konkurrenz zwischen den Individuen einer Population kann, vor allem zu Beginn der Optimierung, dazu führen, daß nur ein lokales Maximum gefunden wird. Dies wird häufig auch als vorzeitige Konvergenz bezeichnet. Daher haben wir die Population in Teilpopulationen, Nischen, aufgeteilt. Konkurrenz tritt dann verstärkt innerhalb einer Nische auf, wird aber zwischen den Nischen reduziert.

Genetische Algorithmen sind zufallsgesteuerte Verfahren und führen somit nicht zwingend zum globalen Optimum. In den folgenden drei Beispielen haben wir daher jedes Strukturzuordnungsproblem 50mal mit einer jeweils anderen Initialisierung des Zufallszahlengenerators behandelt. Aus den Ergebnissen dieser 50 verschiedenen Läufe läßt sich abschätzen, wie sicher der genetische Algorithmus das globale Optimum erreicht. Bei all diesen Suchen wurde die Ordnung (Einfach-, Doppel-, Dreifachbindung) einander zugeordneter Bindungen nicht berücksichtigt, damit möglichst große Strukturfragmente gefunden werden können. Eine einzelne Berechnung erstreckte sich über 400 Generationen des genetischen Algorithmus. Die Populationen umfaßten jeweils 100 Individuen, wobei die Größe eines einzelnen Individuums durch die Zahl der Bindungen in der größeren der beiden Strukturen bestimmt wird. Jedes der folgenden Beispiele illustriert ein anderes Anwendungsgebiet unseres Verfahrens.

Die strukturellen Voraussetzungen biologischer Aktivität lassen sich mit unserer Methode analysieren. So zeigt Abbildung 4 die MCSS der beiden starken Analgetica Morphin 1 und Methadon 2, die in 48 von 50 Programmläufen gefunden wurde. Die Rechenzeit für einen Programmlauf betrug etwa 20 Sekunden auf einer Sparc-Station 10/20. Endgültige Schlüsse über die biologische Aktivität dürfen aber erst nach einem Vergleich der räumlichen Strukturen gezogen werden. Daher arbeiten wir ge-

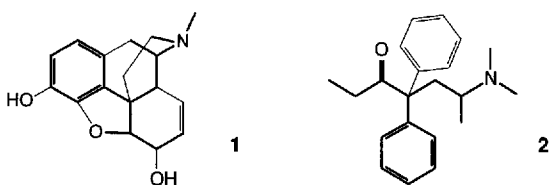


Abb. 4. Die größte gemeinsame Substruktur der beiden Analgetica Morphin 1 und Methadon 2 ist durch Fettdruck hervorgehoben.

genwärtig an der Entwicklung eines Verfahrens zur Bestimmung der größten gemeinsamen dreidimensionalen Teilstruktur durch einen genetischen Algorithmus.

Auch zur Analyse von Biosynthesewegen kann das Verfahren eingesetzt werden. Geissoschizin 4 ist ein wichtiges Zwischenprodukt bei der Biosynthese der Monoterpen-Indolalkaloide^[7]. Abbildung 5 zeigt die größte gemeinsame Substruktur von Strychnin 3 und Geissoschizin 4, die in 50 Optimierungsexperimenten 31mal gefunden wurde (Fitness: 48.0; der mittlere Wert der Fitness in allen Läufen betrug 44.0). Alle Einzelheiten der Umwandlung von Geissoschizin in die Strychnosalkaloide sind noch nicht geklärt; die hier gefundene gemeinsame Substruktur kann daher wichtige Hinweise für weitere Untersuchungen geben.

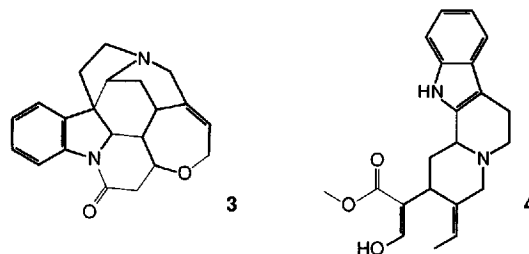


Abb. 5. Die größte gemeinsame Substruktur von Strychnin 3 und Geissoschizin 4.

Die Anwendung des Verfahrens bei der Suche nach geeigneten Ausgangsmaterialien zur Synthese einer gewünschten Verbindung wird am Beispiel der Strychninsynthese von Woodward et al.^[8] illustriert. Abbildung 6 zeigt die größte gemeinsame

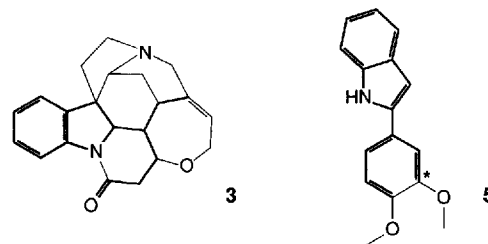


Abb. 6. Die größte gemeinsame Substruktur von Strychnin 3 und 2-Veratrylindol 5, dem Ausgangsstoff der Synthese von 3.

Substruktur von Strychnin 3 und 2-Veratrylindol 5, das als Ausgangspunkt der Synthese gedient hat. Eine Substruktur dieser Größe (Fitness: 34.0) wurde bei 50 Programmläufen 35mal gefunden; der durchschnittliche Fitness-Wert betrug 32.1. Da die gemeinsame Substruktur fast das ganze Gerüst von 5 umfaßt, eignet sich diese Verbindung hervorragend als Ausgangsmaterial. Diese hohe Ähnlichkeit wurde auch von Woodward et al. genutzt; bis auf das in Abbildung 6 mit einem Stern gekennzeichnete Veratrylindolatom werden alle Atome der MCSS entsprechend ihrer Zuordnung zum Strychninaufbau verwendet.

Um geeignete Ausgangsmaterialien für eine Synthese zu finden, kann man auch nach deckungsgleichen Teilstrukturen innerhalb ein und derselben Molekülstruktur suchen. Diese können einen Hinweis darauf geben, daß statt zweier unterschiedlicher Vorstufen nur eine einzige hergestellt werden muß, die zweimal im Lauf der Synthese eingesetzt werden kann. Je größer dabei die Struktur der mehrmals verwendeten Vorstufe ist, um so leistungsfähiger ist diese Synthesestrategie. Auch die

Natur macht häufig von diesem Prinzip Gebrauch; zu den bekanntesten Beispielen zählen die aus Isopreneinheiten aufgebauten Terpene.

Der beschriebene genetische Algorithmus konnte so erweitert werden, daß damit die größte Teilstruktur, die zweimal in einem Molekül enthalten ist, gefunden wird. Der erweiterte Algorithmus übernimmt sowohl die Fitnessfunktion als auch die genetischen Operatoren Selektion, Mutation, Crossover, Creep und Crunch. Die Repräsentation der betrachteten Substruktur allerdings mußte so erweitert werden, daß nun zwei Substrukturen, die in einer einzigen Verbindung vorliegen, verarbeitet werden können. Dabei muß natürlich darauf geachtet werden, daß weder ein Atom noch eine Bindung beiden Teilstrukturen gemeinsam ist.

Bei der Suche nach zwei identischen, möglichst großen Synthesebausteinen in Badion A 6, einem Hutfarbstoff des Maronenröhrlings (*Xerocomus badius*)^[9], mit dem genetischen Algorithmus wird das in Abbildung 7 dargestellte Ergebnis erhalten. Die zweimal vorhandene Teilstruktur ist fettgedruckt; sie entspricht der ebenfalls im Maronenröhrling enthaltenen Pulvinsäure 7, aus der Badion A wahrscheinlich durch oxidative Dimerisierung gebildet wird^[10]. In 50 Läufen des genetischen Algorithmus wurde die in Abbildung 7 dargestellte Teilstruktur mit einer Fitness von 54.0 27mal gefunden. Die durchschnittliche Fitness in allen Läufen betrug 48.1.

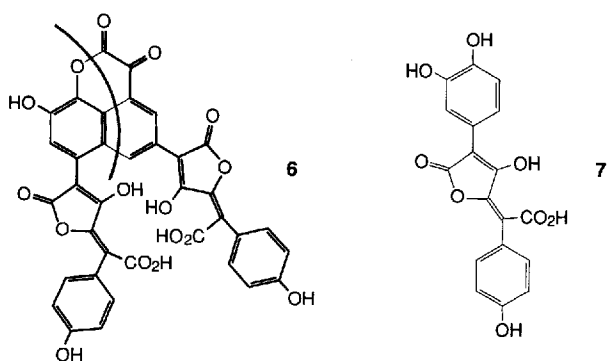


Abb. 7. Badion A 6 und die potentielle Ausgangsverbindung der Biosynthese, die Pulvinsäure 7. Die fett gezeichneten, mit einem Kreisbogen voneinander abgesetzten Teilstrukturen von 6 sind vom genetischen Algorithmus gefunden worden.

Um zu untersuchen, wie häufig dieses Syntheseprinzip verwendet wird und welche Strukturtypen und welche Reaktionen dabei eingesetzt werden, wurde eine systematische Suche in einer Reaktionsdatenbank durchgeführt. Aus der ChemInform-RX-Reaktionsdatenbank^[11] wurden 600 Reaktionen des Jahrgangs 1991 zufällig ausgewählt. Die Analyse ergab, daß bei 30 Reaktionen ein Edukt zweimal in das Reaktionsprodukt eingebaut worden ist. Eine dieser Reaktionen, die Kondensation eines Arens mit Formaldehyd, ist in Abbildung 8 dargestellt. Dieses Beispiel zeigt, daß mit dem Verfahren auch Reaktionen analysiert werden können, bei denen das Molekül außer den beiden deckungsgleichen Teilstrukturen auch noch Teile von

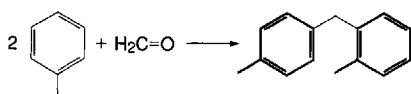


Abb. 8. Eine Reaktion, die zweimal dasselbe Edukt verwendet. Sie steht stellvertretend für die 30 derartigen Reaktionen, die bei einer systematischen Suche in einem Teil der ChemInform-RX-Reaktionsdatenbank gefunden wurden.

anderen Ausgangsstoffen enthält. Da das Reaktionsprodukt keine Symmetrie aufweist, ist die Betrachtung der Symmetrie allein nicht ausreichend als Synthesestrategie. Erst die Suche nach deckungsgleichen Teilstrukturen macht es möglich, derartige Synthesen zu planen.

Eingegangen am 15. Oktober, veränderte Fassung am 27. Dezember 1993 [Z 6421/6422]

- [1] G. Levi, *Calcolo* **1972**, 9, 1–12; M. M. Cone, R. Venkataraghavan, F. W. McLafferty, *J. Am. Chem. Soc.* **1977**, 99, 7668–7671; T. H. Varkony, Y. Shi-loach, D. H. Smith, *J. Chem. Inf. Comput. Sci.* **1979**, 19, 105–111; M. Bersohn, K. Mackay, *ibid.* **1979**, 19, 137–141; J. J. McGregor, P. Willett, *ibid.* **1981**, 21, 137–140; J. J. McGregor, *Software – Practice and Experience* **1982**, 12, 23–34; M. Bersohn, *J. Chem. Soc. Perkin Trans 1* **1982**, 631–637; H. Scsibran, K. Varmuza, *Fresenius J. Anal. Chem.* **1992**, 334, 220–222; L. Chen, W. Robien, *J. Chem. Inf. Comput. Sci.* **1992**, 32, 501–506.
- [2] D. E. Goldberg, *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, Reading, MA, **1989**.
- [3] D. Wienke, C. Lucasius, G. Kateman, *Anal. Chim. Acta* **1992**, 265, 211–225; M. Bos, H. T. Weber, *ibid.* **1991**, 247, 97–105; T. Li, C. Lucasius, G. Kateman, *ibid.* **1992**, 268, 123–134; C. Lucasius, G. Kateman, *TrAC Trends Anal. Chem. Pers. Ed.* **1991**, 10, 254–261.
- [4] E. Fontain, *Anal. Chim. Acta* **1992**, 265, 227–232; *J. Chem. Inf. Comput. Sci.* **1992**, 32, 748–752.
- [5] Einzelheiten der Optimierung des Verfahrens sollen in *J. Comput. Chem.* veröffentlicht werden.
- [6] J. E. Baker in *Genetic Algorithms and their Applications: Proceedings of the Second International Conference on Genetic Algorithms* (Hrsg.: J. J. Grefenstette), Lawrence Erlbaum, Hillsdale, NJ, **1987**, S. 14.
- [7] D. R. Dalton, *The Alkaloids*, Marcel Dekker, New York, **1979**.
- [8] R. B. Woodward, M. P. Cava, W. D. Ollis, A. Hunger, H. U. Daeniker, K. Schenker, *Tetrahedron* **1963**, 19, 247–288.
- [9] B. Steffan, W. Steglich, *Angew. Chem.* **1984**, 96, 435–437; *Angew. Chem. Int. Ed. Engl.* **1984**, 23, 445–447.
- [10] W. Steglich, H.-T. Huppertz, B. Steffan, *Angew. Chem.* **1985**, 97, 716–717; *Angew. Chem. Int. Ed. Engl.* **1985**, 24, 711–712.
- [11] A. Parlow, C. Weiske, J. Gasteiger, *J. Chem. Inf. Comput. Sci.* **1990**, 30, 400–402; J. Gasteiger, C. Weiske, *Nachr. Chem. Tech. Lab.* **1992**, 40, 1114–1120; die ChemInform-RX-Reaktionsdatenbank wird vom FIZ CHEMIE, Berlin, und von der Bayer AG, Leverkusen, mit Daten gefüllt.

Synthese und Charakterisierung von Thiolatokomplexen mit zweifach koordiniertem Eisen(II) **

Jeffrey J. Ellison, Karin Ruhlandt-Senge und Philip P. Power *

Komplexe mit zweifach koordinierten Übergangsmetallen, die eine unvollständig besetzte d-Schale haben, sind sehr selten und ausschließlich auf solche mit sperrigen Alkyl- oder Amido-liganden beschränkt^[1]. Für Liganden, die über ein Element der dritten Periode wie Schwefel oder Phosphor an das Metallzentrum gebunden sind, ist eine zweifache Koordination in kondensierter Phase unbekannt, und dreifach koordinierte Verbindungen sind sehr selten^[2]. Übergangsmetallkomplexe mit niedrigen Koordinationszahlen (2 oder 3) sind seit kurzem von großem Interesse, was hauptsächlich auf die Veröffentlichung der Kristallstrukturanalyse des Fe-Mo-Cofaktors der Nitrogenase von *Azotobacter vinelandii* zurückgeführt werden kann. In diesem

[*] Prof. P. P. Power, J. J. Ellison, Dr. K. Ruhlandt-Senge
Department of Chemistry
University of California
Davis, CA, 95616 (USA)
Telefax: Int. + 916/752-8995

[**] Diese Arbeit wurde von der National Science Foundation und dem Petroleum Research Fund gefördert.