

Große Moleküle aus der virtuellen Backstube – eine Lücke in der Strukturforschung wird geschlossen**

Peter Luger*

Daß präparativ arbeitende Chemiker ihre neuen Verbindungen in heißen Öfen zusammenkochen oder auch backen, ist tägliche Praxis. Daß aber vor allem große Moleküle in den Gigabyte-Welten moderner Computer geschüttelt und gebacken, manchmal auch nur halbgebacken werden, ist neu. Sehr erfolgreich wird dadurch eine häufig beklagte Lücke in der ansonsten sehr wirkungsvollen röntgenographischen Strukturforschung geschlossen. Als vor 6 Jahren die japanische Arbeitsgruppe um Kondo^[1] die Kristallstruktur des tiefblauen Blütenfarbstoffes Commelinin aufklärte, erregte dieses Ergebnis ein solches Aufsehen, daß *Nature* ihr die Titelseite ihres ersten August-Heftes widmete. Ferner wurde die Gruppe mit dem Preis des besten Posterbeitrags zum 16. Internationalen Kristallographiekongreß in Peking 1993 ausgezeichnet.^[2] Da sich die „scientific community“ daran gewöhnt hatte, daß Kristallstrukturen alltäglich in großer Zahl in die wissenschaftliche Literatur eingingen, war eine derartig Aufsehen erregende Wirkung eigentlich nicht zu verstehen. Ein Grund für die große Resonanz war sicherlich ästhetischer Art: Das wunderschöne Aussehen der Commelininstruktur (Abb. 1 oben) steht dem der zugehörigen Blüte kaum nach.

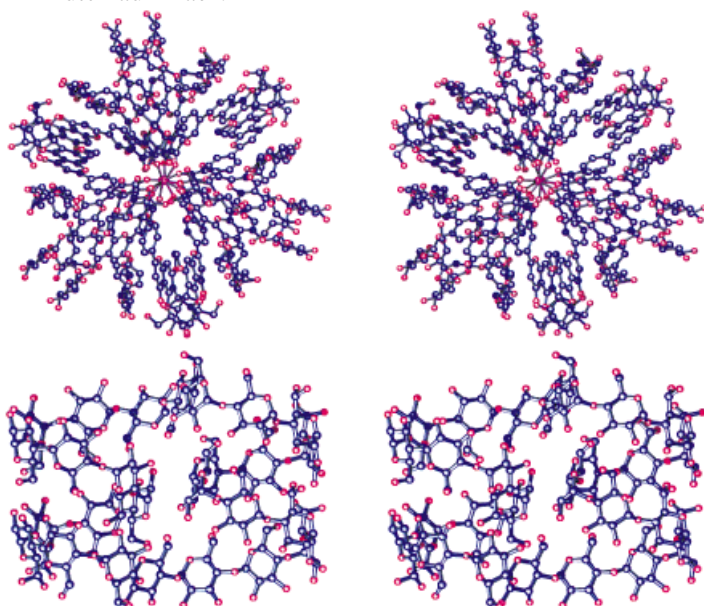


Abb. 1. Oben Struktur des Blütenfarbstoffes Commelinin,^[2] die Anfang der 90er Jahre mit klassischen Direkten Methoden gelöst wurde; Unten mit SHELXD gelöste Struktur einer aus 26 Glucose-Einheiten bestehenden Cycloamylose,^[22] einer der größten atomar aufgelösten Leichtatomstrukturen; beide Stereobildpaare sind mit SCHAKAL^[30] gezeichnet.

[*] Prof. Dr. P. Luger
Institut für Kristallographie der Freien Universität
Takustraße 6, D-14195 Berlin
Fax: (+49) 30-838-3464
E-mail: luger@chemie.fu-berlin.de.

[**] Diese Arbeit wurde von der Deutschen Forschungsgemeinschaft (SFB 312, Gerichtete Membranprozesse) und vom Fonds der Chemischen Industrie gefördert.

Aber nicht nur der schönen Bilder wegen erregte die Struktur großes Aufsehen. Sie enthält 253 Atome (ohne H-Atome) in der asymmetrischen Einheit und gehörte damit zu jener Gruppe von mittelgroßen Strukturen, deren Lösung für die Kristallographen auch Anfang der 90er Jahre noch eine Herausforderung war. Zwar nimmt seit etwa 30 Jahren die Zahl der im Cambridge Crystallographic Data Centre (CCDC)^[3] registrierten Kristallstrukturanalysen fast exponentiell zu, derzeit um jährlich mehr als 10000 Strukturen. Auch in der Proteinkristallographie hat eine dem Trend nach vergleichbare Entwicklung stattgefunden, wenn auch zahlenmäßig in geringerem Umfang. Immerhin wächst inzwischen die Proteindatenbank in Brookhaven^[4] auch um etwa 1000 Einträge pro Jahr (Abb. 2). Dennoch begünstigten diese

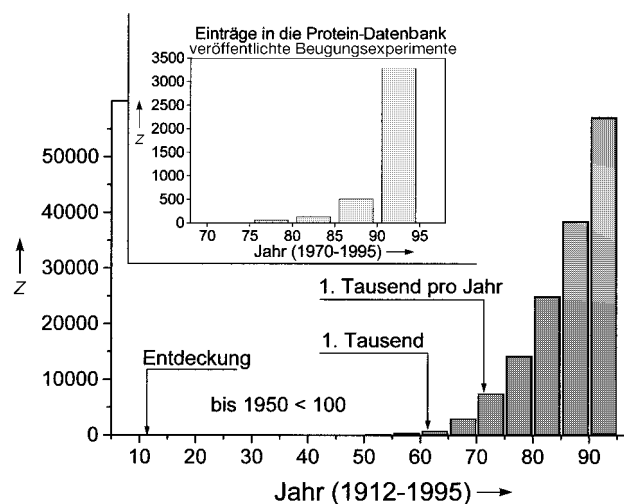


Abb. 2. Zahl Z der Einträge neuer Kristallstrukturen im Cambridge Data File des Cambridge Crystallographic Data Centre (CCDC) in Fünfjahres-Intervallen. Einschub: Entsprechende Einträge in die Proteindatenbank in Brookhaven.

Entwicklungen nur die kleinen Strukturen, wobei die Schallgrenze zunächst bei 100, später bei etwa 200 Atomen lag, und die großen Strukturen mit Molekulargewichten im Kilodalton-Bereich. Strukturen mit einigen 100 Atomen in der asymmetrischen Einheit blieben nach wie vor problematisch. So findet man z. B. unter den reinen organischen Strukturen weniger als 300 Einträge mit mehr als 100 C-Atomen in der Summenformel und nur ganze 16 Einträge im Cambridge Data File mit mehr als 200 C-Atomen.

In einem „Research-News“-Artikel in *Science*, der im April 1992 erschien,^[5] wurde die Situation mit der einer kinderreichen Familie verglichen, bei der häufig den größten und den kleinsten Kindern die meiste Zuwendung gilt, während die mittleren eher wenig beachtet nebenher mitlaufen. Nun sind Kristallographen keine Rabeneltern, sondern die Gründe für ihre Achtlosigkeit waren schwerwiegend. Ein Problem

ist die Kristallisierbarkeit, die mit zunehmender Molekülgröße zunächst stark abnimmt; erst sehr große Moleküle, die eine weniger flexible, mehr globuläre Gestalt haben, neigen dann wieder eher zum Kristallisieren als die Moleküle mittlerer Größe. Hinzu kommt, daß in die Kristallisation biologischer Makromoleküle erhebliche Anstrengungen investiert worden sind, sogar extraterrestrische Versuche wurden einbezogen.^[6]

Ein weiteres Problem aber ist, daß bei Molekülen mittlerer Größe die sonst omnipotenten „Direkten Methoden“ zur Lösung des kristallographischen Phasenproblems nicht mehr und die Methoden zur Lösung von Proteinstrukturen noch nicht funktionieren. Daher stand für mittelgroße Strukturen kaum eine brauchbare Bestimmungsmethode zur Verfügung. Konnten sie in Ausnahmefällen gelöst werden, erregten sie, wie das Commelinin-Beispiel zeigt, großes Aufsehen.

Wie die unlängst in der *Angewandten Chemie* erschienene Arbeit von Sheldrick et al.^[7] über zwei Actinomycin-Strukturen (Actinomycin D und Z₃ mit 314 bzw. 307 Atomen in der asymmetrischen Einheit) dargelegt hat, kann diese Lücke jetzt geschlossen werden, wenn atomar (d.h. bis ca. 1 Å) aufgelöste Intensitätsdatensätze vorliegen. Dazu mußten allerdings völlig neue Strategien zur Strukturbestimmung entwickelt werden, und ein Initiator dieser Entwicklung war wieder einmal Herbert Hauptman, einer der „Väter“ der Direkten Methoden, dem zusammen mit Jerome Karle für ihre Pionierarbeiten auf diesem Gebiet 1985 der Nobel-Preis für Chemie verliehen wurde. 1988 hat Hauptman eine Phasen-Minimums-Funktion $R(\varphi)$ eingeführt,^[8] bei der die bereits in den klassischen Direkten Methoden dominierenden Drei- und Vierphasen-Strukturinvarianten, die Triplets und Quartetts $\varphi_{hk} = \varphi_h + \varphi_k + \varphi_{-h-k}$ bzw. $\varphi_{lmn} = \varphi_l + \varphi_m + \varphi_n + \varphi_{-l-m-n}$ (h, k, l, \dots sind reziproke Gittervektoren von Bragg-Reflexen) eine zentrale Rolle spielen. Für die Triplett- und Quartett-Beziehungen (genauer für $\cos \varphi_{hk}$ und $\cos \varphi_{lmn}$) lassen sich Erwartungswerte ε_T und ε_Q berechnen, mit denen $R(\varphi)$ schließlich definiert ist durch Gleichung (1).

$$R(\varphi) = \frac{\sum_{h,k} A_{hk} [\cos \varphi_{hk} - \varepsilon_T]^2 + \sum_{l,m,n} |B_{lmn}| [\cos \varphi_{lmn} - \varepsilon_Q]^2}{\sum_{h,k} A_{hk} + \sum_{l,m,n} |B_{lmn}|} \quad (1)$$

Nach dem ebenfalls von Hauptman eingeführten Minimumsprinzip^[9] sollte die Minimierung von $R(\varphi)$ zur richtigen Struktur führen,^[10] was nach der Bedeutung der ε_T und ε_Q unmittelbar einleuchtet.

Zur praktischen Realisierung wurde der sogenannte Shake-and-bake-Algorithmus eingeführt.^[11–13] Dieser unterscheidet sich von den klassischen Direkten Methoden (die allein im reziproken Raum arbeiten) hauptsächlich dadurch, daß er in einem cyclischen Prozeß ständig zwischen reziprokem und direktem Raum wechselt. Auf diese Weise wird automatisch die Bedingung eingebracht, daß die richtige Struktur nicht nur überall eine positive Elektronendichte, sondern auch hohe lokale Elektronendichtemaxima haben soll, denen Atompositionen zugeordnet werden können. Damit wird der atomare Aufbau einer chemischen Struktur berücksichtigt.

Shake-and-bake läuft nun in mehreren Schritten ab, die abwechselnd im direkten und im reziproken Raum stattfinden^[14] (Abb. 3). Nachdem eine Teststruktur generiert wurde – in der Regel über einen Zufallsgenerator – liefern die nach

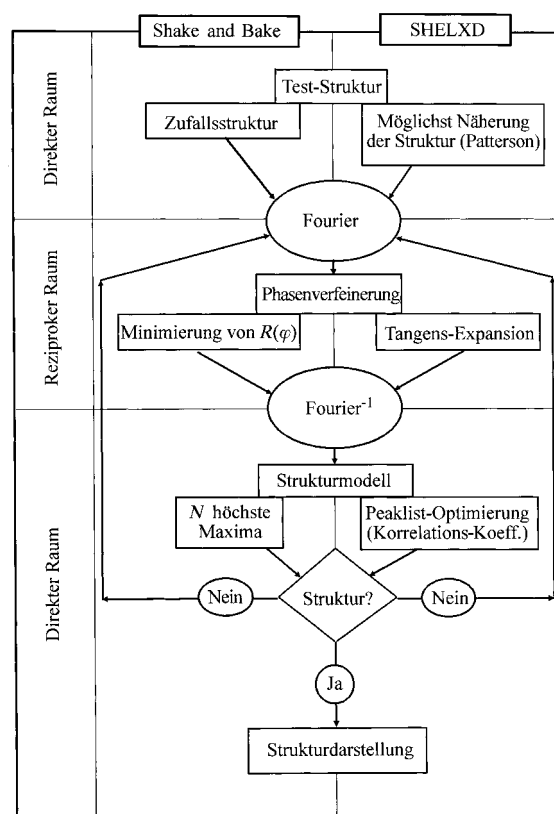


Abb. 3. Schematische Darstellung des Ablaufs von Shake-and-bake (links) und von SHELXD (rechts).

der Fourier-Transformation in den reziproken Raum erhaltenen Strukturfaktoren die Anfangsphasen. Danach findet der „Shake“-Schritt statt: Durch Phasenverfeinerung in einem iterativen Prozeß wird die Minimierung von $R(\varphi)$ angestrebt. Nach einem solchen Minimierungscyclus wird eine neue Struktur „gebacken“. Mit den nun hoffentlich verbesserten Phasen wird wieder im direkten Raum eine neue Elektronendichteverteilung generiert. Daraus werden die N höchsten Maxima ausgesucht. Diese werden erneut Fourier-transformiert, so daß neue Anfangsphasen in den nächsten „Shake“-Schritt eingebracht werden.

Dieses Verfahren muß in der Regel einige hundert Shake-and-bake-Cyclen durchlaufen, bevor am Verlauf von $R(\varphi)$ erkannt werden kann, ob es zur richtigen Struktur führt. Mit einer Teststruktur allein führt Shake-and-bake trotz des Aufwandes selten zum Ziel, so daß in der Praxis einige tausend Anläufe mit verschiedenen, zufällig gewählten Teststrukturen berechnet werden müssen.

Angeregt durch Shake-and-bake hat George Sheldrick mit SHELXD ein Programm speziell zur Lösung von Strukturen mit 200–2000 Atomen entwickelt.^[15,16] Sein Verfahren unterscheidet sich an einigen Stellen von Shake-and-bake (Abb. 3). Sheldrick versucht, auf die zufälligen Ausgangsstrukturen zu verzichten. Sind z.B. Schweratome vorhanden, werden teilweise sehr raffinierte Patterson-Interpretationen^[17,18] eingebracht. In diesem Ansatz sind die Anfangsphasen zuverlässiger als in einer völlig zufällig gewählten Struktur. Es schließt sich eine konventionelle Tangens-Verfeinerung an, und eine inverse Fourier-Transformation führt dann zu einem neuen

Strukturmodell. Nun werden aber die N höchsten Maxima nicht kritiklos für das weitere Verfahren übernommen, sondern es findet eine „Peaklist“-Optimierung statt.^[19] Ausgehend vom niedrigsten Peak wird fortlaufend entschieden, ob der Verbleib oder die Wegnahme dieses Peaks einen von Fujinaga und Read^[20] eingeführten Korrelationskoeffizienten verbessert oder nicht. Erst die bereinigte, optimierte „Peaklist“ wird in den nächsten Cyclus eingegeben.

Durch diese „Peaklist“-Optimierung (und die mögliche Patterson-Interpretation am Anfang) arbeitet SHELXD mit größerem Gewicht im direkten Raum als Shake-and-bake. Eine Zeit lang bezeichnete Sheldrick selbst sein Verfahren als „half-baked“,^[19] aber angesichts der inzwischen erzielten Erfolge kann man es heute durchaus „fully-baked“ nennen. So sind mit SHELXD außer den bereits erwähnten Actinomycin-Strukturen eine Reihe weiterer bislang für unlösbar gehaltener Strukturen aufgeklärt worden. Beispiele sind Vancomycin,^[21] ein wichtiges Lantibiotikum mit 313 unabhängigen Nicht-H-Atomen in der asymmetrischen Einheit, und eine Cycloamylose aus 26 Glucose-Einheiten mit annähernd 650 Atomen (2 Moleküle pro asymmetrische Einheit),^[22] die eine geraume Zeit als die größte atomar aufgelöste Leichtatomstruktur galt (Abb. 1 unten). Übertroffen wird sie von dem vermutlichen derzeitigen „Weltrekordhalter“, der Struktur des Lantibiotikums Mersacidin,^[23] das mit 850 unabhängigen Atomen der Größe eines Proteins mit 120 Aminosäuren entspricht. Etwas außer Konkurrenz, weil schon vorher bekannt, läuft die trikline Struktur des 1001 Atome enthaltenden Hühnereiweiß-Lysozyms (Abb. 4), die in der Arbeitsgruppe von Sheldrick neu gelöst wurde und bei

der damit erstmals die 1000-Atome-Grenze überschritten wurde!^[24]

Angesichts der Tatsache, daß für eine Strukturbestimmung einige tausend Anläufe und für einen Anlauf mehrere hundert Cyclen zu rechnen sind, ergeben sich exorbitante Rechenzeiten, die in VAX-Jahren angegeben werden können.^[19] Da noch in den 80er Jahren die VAX-Rechner zur üblichen Laborausstattung gehörten, wären weder Shake-and-bake noch SHELXD einsetzbar gewesen, selbst wenn die Algorithmen schon existiert hätten. Da mit heutigen Prozessoren die Rechengeschwindigkeit um mindestens zwei Größenordnungen besser ist, verringern sich VAX-Jahre auf annehmbare Tage oder Stunden.

Da die neuen Algorithmen, wie bereits erwähnt, atomar aufgelöste Datensätze benötigen, hätten sie kaum Erfolgsaussichten, wenn nicht parallel in jüngster Zeit einige bemerkenswerte experimentelle Entwicklungen stattgefunden hätten. Neben Fortschritten bei den Kristallisationstechniken^[25] sind hier vor allem wesentlich verbesserte Möglichkeiten der Intensitätsdatensammlung zu nennen. Durch den Einsatz von Synchrotronstrahlung, möglichst bei tiefen Temperaturen, läßt sich das Verhältnis von Reflexintensität zum Untergrund besonders bei hohen Beugungsordnungen erheblich steigern. Kombiniert man das detektorseitig mit flächenhafter Detektion („Flächenzähler“) wie etwa den neuen CCD-Systemen (z.B. dem bereits im April 1994 auf den Markt gekommenen SMART-Diffraktometer^[26]), so trägt dies nicht nur zur weiteren Steigerung der Auflösung bei, sondern es kann auch die Meßzeit gegenüber den konventionellen seriellen Diffraktometern um Größenordnungen

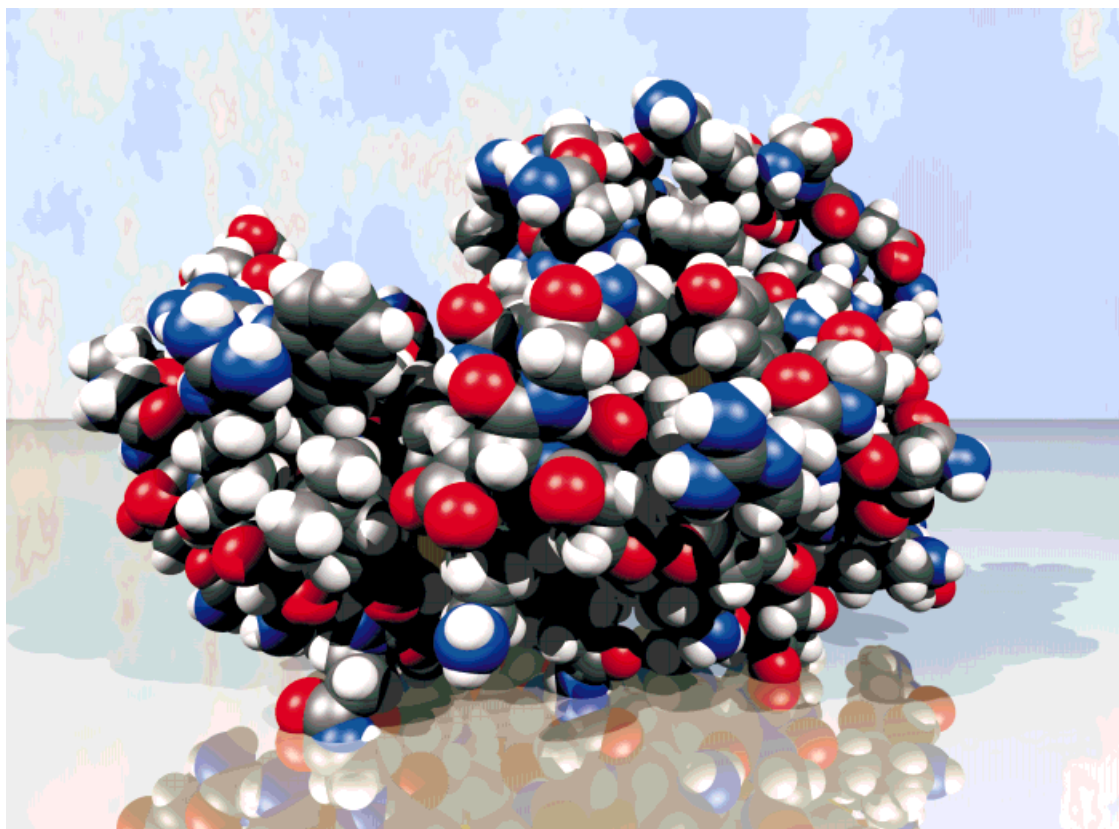


Abb. 4. 1001 unabhängige Atome in der Struktur der trikline Form von Hühnereiweiß-Lysozym.^[24]

verkürzt werden,^[27] was bei den umfangreichen Datensätzen der großen Strukturen besonders zum Tragen kommt. Damit ist – wie schon in den 60er Jahren – erneut ein Innovations-schub in der Kristallstrukturforschung mehreren gleichzeitig erfolgten Entwicklungen in Theorie, Experiment und Computertechnologie zu verdanken. Die Lücke, die sich bisher für die mittelgroßen Moleküle von einigen 100 Atomen auftat, schließt sich also noch rechtzeitig zur Jahrtausendwende.

Problemstellungen bietet die moderne Chemie reichlich. Allein in der Klasse der antibiotisch oder cytostatisch wirksamen Naturstoffe, deren zunehmende Bedeutung unter anderem durch den jüngst erschienenen Highlight-Artikel von Lindel^[28] belegt ist, werden rasch Strukturgrößen von einigen 100 Atomen erreicht, die bisher nur in Ausnahmefällen,^[29] nunmehr aber mit großer Aussicht auf Erfolg röntgenographisch aufgeklärt werden können.

Stichwörter: Computerchemie • Direkte Methoden • Röntgenbeugung • Röntgenstrukturanalyse

- [1] T. Kondo, K. Yoshida, A. Nakagawa, T. Kawai, H. Tamura, T. Goto, *Nature* **1992**, 358, 515–518.
- [2] A. Nakagawa, K. Yoshida, T. Kondo, T. Kawai, T. Goto, *Acta Crystallogr. Sect. A* **1993**, 49, 201.
- [3] F. H. Allen, O. Kennard, *Chem. Des. Autom. News* **1993**, 8, 31–37.
- [4] F. C. Bernstein, T. F. Koetzle, G. J. B. Williams, E. F. Meyer, Jr., M. D. Brice, J. R. Rodgers, O. Kennard, T. Shimanouchi, M. Tasumi, *J. Mol. Biol.* **1977**, 112, 535–542.
- [5] A. S. Moffat, *Science* **1992**, 256, 309–310.
- [6] F. Rosenberger, *Z. Kristallogr.* **1998**, Suppl. 15, 18.
- [7] M. Schäfer, G. M. Sheldrick, I. Bahner, H. Lackner, *Angew. Chem.* **1998**, 110, 2482–2485; *Angew. Chem. Int. Ed.* **1998**, 37, 2381–2384.
- [8] H. A. Hauptman, *Proceedings of the ACA Meeting*, Philadelphia, USA, **1988**, Abstract R4.
- [9] H. A. Hauptman in *Crystallogr. Computing 5* (Hrsg.: D. Moras, A. D. Podjarny, J. C. Thierry), University Press, Oxford, **1991**, S. 324–332.
- [10] C. M. Weeks, G. T. DeTitta, R. Miller, H. A. Hauptman, *Acta Crystallogr. Sect. D* **1993**, 49, 179–181.
- [11] R. Miller, G. T. DeTitta, R. Jones, D. A. Langs, C. M. Weeks, H. A. Hauptman, *Science* **1993**, 259, 1430–1433.
- [12] G. T. DeTitta, C. M. Weeks, P. Thuman, R. Miller, H. A. Hauptman, *Acta Crystallogr. Sect. A* **1994**, 50, 203–210.
- [13] C. M. Weeks, G. T. DeTitta, H. A. Hauptman, P. Thuman, R. Miller, *Acta Crystallogr. Sect. A* **1994**, 50, 210–220.
- [14] R. Miller, S. M. Gallo, H. G. Khalak, C. M. Weeks, *J. Appl. Crystallogr.* **1994**, 27, 613–621.
- [15] G. M. Sheldrick, *SHELXD-CCP4 Workshop*, York, **1997**, 147–157.
- [16] G. M. Sheldrick in *Direct Methods for Solving Macromol. Structures* (Hrsg. S. Fortier), Kluwer Academic Publishers, Niederlande, **1998**, S. 401–411.
- [17] G. M. Sheldrick in *Crystallogr. Computing 6* (Hrsg. H. D. Flack, L. Parkanyi, L. Simon), University Press, Oxford, **1993**, S. 100–110.
- [18] G. M. Sheldrick in *Direct Methods for Solving Macromol. Structures* (Hrsg. S. Fortier), Kluwer Academic Publishers, Niederlande, **1998**, S. 131–141.
- [19] G. M. Sheldrick, R. O. Gould, *Acta Crystallogr. Sect. B* **1995**, 51, 423–431.
- [20] M. Fujinaga, R. J. Read, *J. Appl. Crystallogr.* **1987**, 20, 517–521.
- [21] M. Schäfer, Th. R. Schneider, G. M. Sheldrick, *Structure* **1996**, 4, 1509–1515.
- [22] K. Geßler, I. Uson, T. Takaha, N. Krauß, S. Smith, G. M. Sheldrick, W. Saenger, *6. Jahrestagung Deutsche Gesellsch. f. Kristallogr.*, Karlsruhe, **1998**, Abstract.
- [23] J. Kärcher, P. Lubini, E. Pohl, T. R. Schneider, G. M. Sheldrick, *6. Jahrestagung Deutsche Gesellsch. f. Kristallogr.*, Karlsruhe, **1998**, Abstract.
- [24] G. M. Sheldrick, persönliche Mitteilung, **1998**.
- [25] J. Hulliger, *Angew. Chem.* **1994**, 106, 151–171; *Angew. Chem. Int. Ed. Engl.* **1994**, 33, 143–162.
- [26] SMART, CCD-based Structure Determination System, Bruker-AXS Inc., Madison, WI, USA, **1994**.
- [27] T. Koritsanzsky, R. Flaig, D. Zobel, H.-G. Krane, W. Morgenroth, P. Luger, *Science* **1998**, 279, 356–358.
- [28] T. Lindel, *Angew. Chem.* **1998**, 110, 806–808; *Angew. Chem. Int. Ed.* **1998**, 37, 774–776.
- [29] D. A. Langs, *Science* **1988**, 241, 188–191.
- [30] E. Keller, *SCHAKAL88. User Manual*, Universität Freiburg, **1988**.
- [31] Zu den Abbildungen 1 und 4 haben die Arbeitsgruppen von A. Nakagawa (Sapporo, Japan), G. M. Sheldrick (Göttingen) und W. Saenger (Berlin) dankenswerterweise Koordinaten und Bildmaterial beigetragen.