

Zur Dokumentation chemischer Forschungsergebnisse

Von Dr. R. FUGMANN, Dr. WILHELM BRAUN und W. VAUPEL,

Farbwerke Hoechst AG., vormals Meister Lucius & Brüning, Frankfurt/Main-Höchst

Das ständige Anwachsen der naturwissenschaftlich-technischen Forschung läßt das Wiederfinden bekanntgemachter Forschungsergebnisse immer schwieriger werden. Die Dokumentation hat zum Ziel, diese Schwierigkeiten zu überwinden. Je nach Art und Umfang eines Sachgebietes wird man verschiedene Wege zu seiner Dokumentation einschlagen. Methoden und Möglichkeiten einer umfassenden Dokumentation der Chemie mit ihren Grenzgebieten werden dargelegt.

In den Farbwerken Hoechst begannen vor mehreren Jahren Forschungsarbeiten auf dem Gebiet der Dokumentation, diesem interessanten Grenzgebiet zwischen Natur- und Geisteswissenschaft. Die Arbeiten wurden von einem größeren Kreis von Fachleuten vorangetrieben. Sie führten zu einem Dokumentations-System, welches alle Belange der Chemie mit ihren Grenzgebieten zu berücksichtigen gestattet und welches bereits beträchtlichen Nutzen abwirft. Für dieses Dokumentations-System wurde die Kurzbezeichnung „GREMAS“ („Genealogisches Recherchieren durch Magnetbandspeicherung“) gewählt.

Wenn die Dokumentation eines chemischen Sachgebietes eingerichtet werden soll, so sind zuerst die Anforderungen und Anfragegepflogenheiten des künftigen Benutzerkreises festzustellen. Die Ansprüche des Chemikers an die Dokumentation seiner Literatur sind sehr vielseitig. Neben der rein präparativen organischen oder anorganischen Chemie mit ihren Molekelstrukturen und Reaktionen sind es die Eigenschaften der Stoffe, ihre Verwendungszwecke, verfahrenstechnische und apparative Details, theoretische Betrachtungen usw., die sämtlich registrierwürdig sind. Alle diese Angaben können bei der Literatursuche zu Suchbedingungen erhoben werden.

Welche Möglichkeiten für die Dokumentation dieser Daten z. Zt. zur Verfügung stehen, sei näher betrachtet.

Die spezifische und die allgemeine Anfrage

Bei allen Suchvorgängen kann man zwei Typen von Fragen unterscheiden, und dem ist bereits bei der Planung des Systems Rechnung zu tragen. Im einen Fall bezweckt man das Wiederfinden eines Sachverhaltes, für den man eine klare, erschöpfende Definition geben kann. Hierher gehören Autorennamen, individuelle chemische Verbindungen, Firmennamen als Patentinhaber usw. Der Fragesteller oder Dokumentar hat den Sachverhalt noch genau in Erinnerung. Es handelt sich hier um einen „Abruf“ von Sachverhalten, denen man früher einmal begegnet ist („question of recall“).

Einen anderen Zweck als diese spezifischen Fragen verfolgen die allgemein gehaltenen Fragen. Man sucht nicht nur individuelle Verbindungen, sondern Verbindungstypen, Reaktionen allgemeiner Art usw. Man kann oder will die Suche nicht auf bestimmte, noch erinnerliche Dinge beschränken. Man läßt sich gleichsam vom Speicher ein „An-

gebot“ über die Objekte machen, die mit dem behandelten Problem in einem mehr oder weniger engen Zusammenhang stehen. Aus diesem Angebot gedenkt man dann eine freizügige Auswahl zu treffen. Man rechnet damit, auf Literaturhinweise zu stoßen, nach denen man niemals spezifisch gesucht hätte („question of discovery“¹⁾).

Beide Fragetypen erfordern unterschiedliche Dokumentations-Methoden. Im ersten Falle nämlich läßt sich für den Gegenstand der Frage (etwa eine chemische Summenformel) im Registriersystem in stets eindeutiger und jederzeit reproduzierbarer Weise ein einziger Ort angeben, an welchem dieser Gegenstand einzuordnen und später wiederzufinden ist. Durch ein einziges, oft rein formales Einteilungsprinzip (z. B. alphabetische und numerische Reihenfolge der Summenformeln) ist die Ablagestelle eines solchen Objekts im Registriersystem eindeutig festgelegt. Daher bereitet das Recherchieren nach klar definierten Objekten in den Registerwerken und Steilkarteien keine Schwierigkeiten.

Während man somit bei hochspezifischen Fragestellungen mit den herkömmlichen Hilfsmitteln auskommt, erfordert hingegen das allgemeine Problem sowohl in der Theorie als auch in der Technik anspruchsvollere Methoden (siehe auch²⁻⁴⁾). Die allgemeinen Fragestellungen spielen in der Praxis die weitaus wichtigere Rolle. Denn eine spezifische gesuchte Verbindung ist auf jeden Fall in der allgemeinen Stoffgruppe mitenthalten, die durch die allgemeine Anfrage zusammengestellt wird.

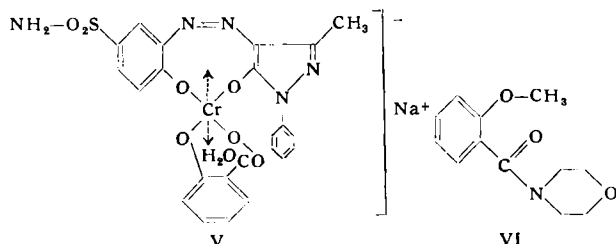
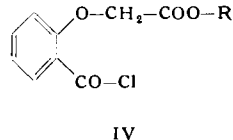
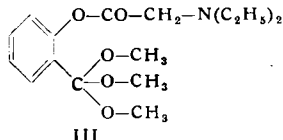
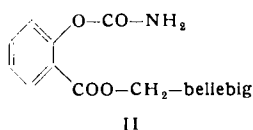
Zum andern aber gilt die allgemeine Fragestellung einem wichtigeren Zweck. Das Anhäufen von Einzeltatsachen wird letztlich immer dem Ziele dienen, eine umfassende Theorie zu entwickeln, um die weiteren Experimente planvoll auf bestimmte, besonders aussichtsreiche Richtungen zu konzentrieren. Je komplexer jedoch die Zusammenhänge sind, desto mehr Einzelergebnisse sind einordnend zu beurteilen, und desto schwieriger wird es, diesen Vorgang allein im Gedankenexperiment zu vollziehen. Die mechanisierte Gruppenbildung ist ein wirksames Hilfsmittel zu diesem Ziel, und hierin liegt ihre heuristische Bedeutung.

¹⁾ C. L. Bernier, Correlative Indexes VI. Amer. Documentation 17, 277 [1960].

²⁾ S. R. Ranganathan, Natural, Classificatory and Machine Languages, Annals of library Sci. 6, 65 [1959].

³⁾ S. R. Ranganathan, Postulational Approach to Faceted Classification, Annals of library Sci. 5, 35 [1958].

⁴⁾ J. W. Perry u. A. Kent, Tools for Machine Literature Searching. Interscience Publishers, Inc., New York, 1958, 14 ff.



oder nach Anwendungen oder nach Herstellungsverfahren unterteilen, wie das die moderne facettierte Klassifikation vorschreibt und nicht nach mehreren Gesichtspunkten gleichzeitig. Denn andernfalls werden die genannten Zusammenhänge zerrissen und lassen sich nicht mehr systematisch abfragen, weil die Begriffskategorien ungeordnet bleiben. In einem fortgeschrittenen Stadium machen sich bei einer solchen nicht „kategorie-reinen“ Klassifikation gewöhnlich noch andere Schwierigkeiten bemerkbar, die die Wirksamkeit des Systems stark beeinträchtigen können⁷⁾.

Auch die Auswahl der differenzierenden Charakteristika selbst unterliegt bestimmten Gesetzmäßigkeiten. Von diesen sei nur die Forderung der „Feststellbarkeit“ eines Charakteristikums erwähnt, „*Canon of ascertainability*“⁸⁾. Das Gesetz besagt, daß die Aufgliederung einer Gruppe von Objekten nur nach einem solchen Charakteristikum geschehen darf, das an den Objekten feststellbar ist. *Ranganathan* führt als Beispiel an, daß man eine Gruppe von Anwesenden wohl nach ihrem Geburtsjahr, nicht jedoch nach dem Datum ihres Todes klassifizieren kann. Die Dokumentation der Chemie läßt oft wesentliche Teile dieses Gesetzes außer Acht. Die differenzierenden Charakteristika unserer Formelregister, Art und Anzahl der Atome, lassen sich in Einspeicherung und Anfrage nur für einen Teil unserer chemischen Probleme angeben. Was die Anfrage betrifft, so sind diese Charakteristika sogar in der Mehrzahl der Fälle, nämlich bei allen allgemeinen Anfragen, nicht „feststellbar“.

Die natürlichen Einteilungsgesichtspunkte eines Sachgebietes herauszuarbeiten und praktisch zu erproben, ist somit die verantwortungsvollste Aufgabe desjenigen, der ein leistungsfähiges Dokumentations-System zu entwerfen hat. Glücklicherweise ist die Chemie ein Gebiet mit einer hohen naturgegebenen Systematik. Das spiegelt sich bereits in der weitgehend einheitlichen Stoffgliederung unserer Lehr- und Handbücher wider. Nach diesen Prinzipien formt sich auch das Stoffgruppenbild des Chemikers bei seiner Ausbildung. Der Dokumentar, der sich diese Prinzipien beim Aufbau seines Code zunutze macht, wird verbreitet auf Zustimmung stoßen.

Ermittlung von Partialstrukturen

Mit der vorweggenommenen Genusbildung beim Einspeichern nach einem klassifizierten Schlüsselssystem ist zwar eine wichtige Voraussetzung für die allgemeine Anfragemöglichkeit gegeben, doch bereitet selbst dann noch das sichere Wiederfinden einer jeden, im Verband einer größeren Molekel versteckten Partialstruktur große Schwierigkeiten. Will man nämlich das allgemeine Frageprinzip konsequent durchführen, dann müssen bei der Verschlüsselung einer auch nur mäßig komplizierten Formel in dieser viele Partialstrukturen berücksichtigt und aus den Molekelstrukturen herausgelesen werden. Ein Salicylsäure-Derivat kann gleichzeitig eine Pyrazol-Verbindung, ein Oxybenzol-sulfonsäureamid usw. sein und vieles andere mehr.

Die Hauptschwierigkeit für das sichere Wiederfinden von Partialstrukturen ist somit nicht so sehr technischer, als vielmehr begrifflicher Art. Es erhebt sich die Frage, wer die zahlreichen Partialstrukturen, welche als sinnvolle Suchbedingungen auftreten können, in einer Molekel alle erkennen und zur Registrierung vorschreiben soll. Hier sind zahlreiche Überlegungen notwendig. Es ist z. B. zwischen solchen Substituenten zu unterscheiden, die feste Bestandteile der Molekel sind und solchen, die nur vorkommen können, jedoch nicht vorkommen müssen („logische Summe“). Es muß weiterhin die Abwesenheit der fehlenden Substituenten konstatiert werden („logische Differenz“) und manches andere mehr. Die Aufgabe kann durch Menschenkraft allein kaum konsequent gelöst werden⁸⁾. Die bekannten Registerwerke bieten deshalb auch stets nur

eine Auswahl besonders wichtiger Partialstrukturen an. Aus technischen Gründen wird vielfach auch eine „Rangordnung“ eingehalten und die fragliche Molekel nur einmal, nämlich bei der ranghöchsten Partialstruktur aufgeführt („Prinzip der spätesten Stelle“). Als Folge davon sind die einzelnen Partialstrukturen über das ganze Register verstreut, je nachdem, mit welchen ranghöheren Merkmalen sie sonst noch in der Molekel zufällig vergesellschaftet sind.

Die modernen programmgesteuerten Rechenautomaten aber vermögen die genannten logischen Entscheidungen zu fällen und sehr viele sinnvolle Partialstrukturen zu erkennen. In unserem Dokumentations-System leistet die elektronische Großrechenanlage z. B. einen derartigen Beitrag zur Verschlüsselung und entlastet hierdurch entscheidend den Menschen. Wir können uns beim Verschlüsseln darauf beschränken, lediglich das Muster für die auszuführenden logischen Schritte anzugeben. Die Maschine greift diese Andeutungen auf und führt sie nach einem vorgegebenen Programm zu Ende⁸⁾. Erst durch einen solchen maschinellen Arbeitsbeitrag ist es möglich geworden, die für eine Großdokumentation der Chemie erforderliche Präzision in der Verschlüsselung zu erreichen und gleichzeitig die allgemeine Abfragbarkeit zu wahren.

Grundsätzlich die gleichen Überlegungen wie für chemische Strukturen gelten auch für chemische Reaktionen. Ebenso wenig wie in den Bezeichnungen „Maleinsäure“, „Muconsäure“, „Mesaconsäure“ usw. das gemeinsame Merkmal der olefinischen Dicarbonsäure in einer für Sortiermaschinen verwertbaren Form enthalten ist, so kann man auch aus den üblichen Bezeichnungsweisen für Reaktionen keine sortierbaren Charakteristika entnehmen. Die Bezeichnungen „Beckmannsche Umlagerung“, „Schmidtsche Reaktion“, „Natriumamid-Spaltung von Ketonen“ lassen äußerlich keinerlei Verwandtschaft erkennen, obwohl bei allen drei Reaktionen eine Keto-Gruppe in die Carbonsäureamid-Gruppe übergeführt wird. Zwar wird es in diesem einfachen Fall dem Chemiker keine Schwierigkeiten bereiten, die genannten Zusammenhänge zu überblicken. Anders ist es aber bereits bei Reaktionen, in denen beispielsweise ein Kohlenstoff-Substituent am Aromaten durch Halogen verdrängt wird oder bei denen α -Halogen-ketone in Acetylen-Verbindungen überführt werden. Hier leistet eine mechanisierte Gruppenbildung wertvolle Dienste. Ebenso wie bei analogen Fragestellungen nach Verbindungstypen müssen zu diesem Zwecke auch chemische Reaktionen in systematischer Weise, Kohlenstoffatom für Kohlenstoffatom, beschrieben werden.

Im Dokumentations-System *Gremas* kann jede organisch-chemische Reaktion durch die Beschreibung der beteiligten Kohlenstoff-Atome im Ausgangs- und Endzustand dargestellt werden. Hierdurch wird es möglich, auf maschinellem Wege alle Reaktionstypen zusammenzuführen, bei denen eine bestimmte Umwandlung entweder allein oder gleichzeitig im Verband mit beliebigen anderen Veränderungen stattfindet. Hierzu ist grundsätzlich kein neuer Code notwendig, denn es werden die bereits bei der Strukturbeschreibung benutzten Chiffren für die Kohlenstoff-Atome wiederverwendet. Lediglich ihre Anordnung wird verändert, um die reaktionschemischen Zusammenhänge wiederzugeben. Bei dieser Arbeitsweise ist der zusätzliche Zeitbedarf für die Verschlüsselung von chemischen Reaktionen denkbar gering.

Auf dem Gebiet der chemischen Reaktionen existieren zwar vorzügliche Handbücher. Der Nutzen dieser Werke wird jedoch durch die Nachteile einer jeden eindimensionalen Darstellungsweise in Buchform beeinträchtigt. Außerdem hängt die Qualität dieser Werke wiederum zum Teil vom Gedächtnis ihrer Autoren, bzw. von der Leistungsfähigkeit ihrer Registriermittel ab.

⁷⁾ A. J. Wells, *The Colon-Classification*, *Aslib Proc.* 2, 14 [1951].

⁸⁾ R. Fugmann, Über die Bedeutung der programmgesteuerten Rechenautomaten für die Dokumentation. *Nachr. f. Dokumentation* 12, 69 [1961].

Dokumentation der Grenzgebiete

Die Chemie dringt zunehmend in die verschiedensten Nachbargebiete ein. Für Chemikalien und Kunststoffe werden immer neue Anwendungsmöglichkeiten erschlossen. Die Technik der Synthese in Laboratorium und Betrieb und die Analysenverfahren verfeinern sich. Hieraus ergibt sich die Notwendigkeit, alle diese Daten ebenso sorgfältig festzuhalten wie die rein chemischen.

Wiederum gilt die Forderung der Gruppenbildung. Man stößt hier aber auf beträchtliche Schwierigkeiten. Die Einteilungsgrundsätze sind häufig umstritten und bei weitem nicht so klar wie in der reinen Strukturchemie. Ob man beispielsweise Nahrungs- und Genußmittel nach ihrem chemischen Aufbau, ihrer soziologischen Bedeutung, ihrer pflanzlichen oder tierischen Herkunft oder nach ihrer Haltbarkeit usw. einteilt, hängt von dem „Aspekt“ des Personenkreises ab, auf den die betreffende Dokumentation zugeschnitten ist. Da es meist mehrere gleichwertige Aspekte gibt, bleibt der Benutzerkreis einer vorgegebenen Systematik relativ klein. Dies wirkt sich stets nachteilig auf die Wirtschaftlichkeit einer solchen Dokumentation aus.

Die herkömmlichen Speichermittel, auf sich allein gestellt, haben sich auch den Anforderungen nicht gewachsen gezeigt, die an eine Dokumentation der Grenzgebiete der Chemie heute gestellt werden müssen. Entweder bieten sie nicht genügend Darstellungsmöglichkeiten für die unübersehbar zahlreichen Sachverhalte, oder die Sortierzeiten wachsen in bedrohlicher Weise mit der Größe des Speichers (Sortiermaschinen für Rand-, Schlitz- und Maschinenlochkarten). Oder aber sie sind wenig geeignet zur Gruppenbildung auf verschiedenen hierarchischen Niveaus und zur Ausführung anspruchsvollerer logischer Operationen (Sichtlochkarten). All diese Unzulänglichkeiten wirken sich auf die Lebensdauer der betreffenden Systeme aus. Man kann mit diesen Hilfsmitteln stets nur eine begrenzte Zeit erfolgreich arbeiten.

Wertbeständigkeit der geleisteten Arbeit

Die Leistungsgrenze einer Kartei tritt dann zu Tage, wenn das Herausholen von Informationen im Laufe der Zeit immer schwieriger geworden und zum Schluß nur noch in Ausnahmefällen möglich ist. In diesem Stadium ist die gesamte, in die Verschlüsselung und Einspeicherung bisher investierte Arbeit entwertet. Es ist daher wichtig, sich die Ursachen für eine solche fatale Entwicklung zu vergegenwärtigen. Einmal sind zumeist vom rein Mechanischen her Grenzen gezogen. Die Sortierzeiten werden immer länger, je mehr Material erfaßt worden ist. Zum anderen aber reicht mit dem Anwachsen des Speichers bald die Treffsicherheit der Selektion nicht mehr aus. Es wird auf eine Anfrage immer mehr „Ballast“ herausgegeben, d. h. unzutreffende oder nicht genügend genau zutreffende Literaturhinweise, und durch diese muß sich der Fachmann nach einer Selektion immer wieder aufs neue hindurcharbeiten, entweder der Dokumentar oder der Fragesteller selbst.

Dem kann ein größeres Dokumentationsvorhaben nur dadurch entgegen, daß man von Anbeginn der Verschlüsselungsarbeit eine sehr genaue und vollständige Beschreibung der wissenschaftlichen Sachverhalte betreibt. Denn diese Genauigkeit wird später beim Anwachsen des Speichers in zunehmendem Maße gebraucht, um die erforderliche Detaillierung bei der Selektion herbeizuführen⁹⁾.

Jede Sparsamkeit in diesem Punkte ist gefährlich, denn sie kann sich auf die Lebensdauer des Systems auswirken.

⁹⁾ B. C. Vickery, *Machines and Indexes*, Unesco Bull. Library 13, 252 [1959].

Auch muß die bei der Verschlüsselung einer Literaturstelle vermeintlich eingesparte Arbeit später, wenn die visuelle Sichtung der herausgegebenen Literaturhinweise einsetzt, vervielfacht nachgeholt werden. Das muß so oft geschehen, wie diese Literaturstelle jemals angesprochen wird, und zu einem Zeitpunkt, in welchem das Ergebnis dieser nachträglichen Überprüfung nicht mehr festgehalten werden kann.

Es bedeutet keine Lösung dieser Frage, wenn der Dokumentar dann beschließt, jeweils nur noch die letzten Jahrgänge in seine Recherchen einzubeziehen, um die aufzubringende Arbeit in Grenzen zu halten. Im Laufe der Zeit sinkt dann nämlich ein immer größerer Teil der geleisteten Verschlüsselungsarbeit in den Bereich des nicht mehr Nutzbaren ab. Je länger die Laufzeit eines solchen Systems ist, desto geringer wird der Nutzen, wenn man ihn mit der zum Verschlüsseln aufgebrauchten Arbeit vergleicht.

Es ist notwendig, einen gesunden Kompromiß zwischen Einspeicherungsarbeit und Recherchenarbeit zu finden. Dabei muß man berücksichtigen, daß der mit dem Recherchieren verbundene Arbeitsbetrag im Laufe der Zeit stark steigt: Die mit dem Durchsehen der selektierten Literaturhinweise verknüpfte Arbeit wächst mit wachsendem Speicher und ist um so größer, je geringer die Treffsicherheit ist. Dem kann nur dadurch begegnet werden, daß man durch erhöhte Genauigkeit bei der Verschlüsselung die Treffsicherheit der Selektion so hoch treibt, daß grundsätzlich überhaupt keine Sichtungsarbeit mehr anfällt.

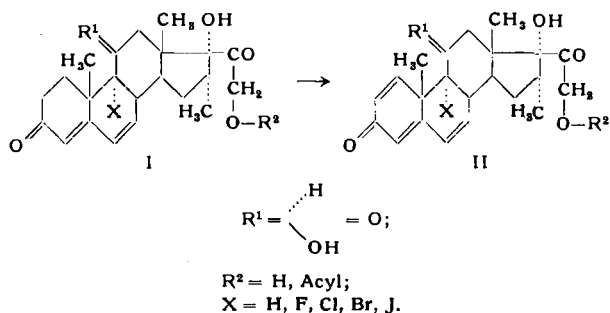
Wenn man versucht, einen solchen Kompromiß zu finden, so stößt man sehr rasch an die Leistungsgrenze der herkömmlichen maschinellen Hilfsmittel vor. Es ist mit ihnen nicht möglich, alle wichtigen Charakteristika aus den vielen fast unübersehbar gewordenen Arbeitsgebieten präzise, allgemein abfragbar und unverfälscht zu speichern.

In ähnlicher Weise aber, wie uns die Großrechenanlage bereits beim Ermitteln der Partialstrukturen behilflich war, so zeigt sich auch ihre Überlegenheit in der Speicherung sämtlicher registrierwürdiger Daten. Die Aufnahmefähigkeit des Magnetbandes ist praktisch unerschöpflich, sowohl was den Umfang eines Code anbetrifft, als auch bezüglich der Anzahl der Charakteristika, mit denen im Einzelfall eine Publikation beschrieben werden soll. Bei dieser Sachlage sind dem Experimentieren mit der erforderlichen Verschlüsselungsgenauigkeit keine Grenzen mehr gezogen. So versetzen uns diese Rechenanlagen erstmalig in die Lage, Dokumentations-Systeme von sehr hoher Lebenserwartung aufzubauen.

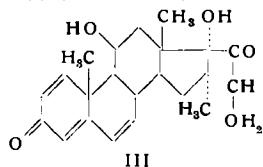
An einem Beispiel sei gezeigt, welche Verschlüsselungsgenauigkeit nach unseren Erfahrungen erforderlich ist, um auch bei sehr großen Sammlungen den Bestand der Arbeit sicherzustellen.

Aus dem Referat zu einem Patent aus der Chemie der Steroide wurden Formel I und II in den Magnetbandspeicher aufgenommen.

16-Methyl-11-hydroxy- bzw. -oxo-17 α , 21-dihydroxy- $\Delta^4,4,9$ -pregnatrien-3, 20-dion-Derivate mit der allg. Formel II und ihre Herstellung durch Einwirkung von Selenioxyd oder Mikroorganismen des Typus *Bacillus sphaericus* auf die entsprechenden $\Delta^4,9$ -Diene mit der allg. Formel I werden geschützt.



So erhält man z. B. durch Einwirkung von *Bacillus sphaericus* auf 16 α -Methyl-11 β , 17 α , 21-trihydroxy- $\Delta^{4,6}$ -pregnadien-3, 20-dion 16 α -Methyl-11 β , 17 α , 21-trihydroxy- $\Delta^{1,4,6}$ -pregnatrien-3, 20-dion (III). Die neuen Steroide zeigen z. B. entzündungshemmende Eigenschaften und werden bei Arthritis verwendet.



Die Formel II wurde so eingespeichert, daß sie beispielsweise auf folgende Fragestellungen anspricht:

3-Oxo- $\Delta^{1,4}$ -Steroide

17 α -21-Dihydroxy-20-oxo-Steroide

9 α -Fluor-16 α -methyl-Steroide

3,20-Dioxo-11 β , 17 α , 21-trihydroxy-16 α -methyl-Steroide

$\Delta^{1,4,6}$ -Steroide, die in 9-Stellung unsubstituiert sind.

Die Verschlüsselung spricht auf alle weiteren Fragestellungen an, bei denen die angegebenen, festen oder variablen Substituenten in beliebiger Kombination in vorgegebenen Positionen gefordert werden. Bei unerwünschten Substituenten erstreckt sich ein Verbot sinngemäß nur auf die festen Substituenten, nicht auf die variablen.

Weiterhin werden unter anderen die folgenden Fragestellungen bedient:

Gemischt cycloaliphatisch-aliphatische Ketone mit α -ständiger Hydroxy-Gruppe.

Dehydrierung von Cycloaliphaten allgemein.

Dehydrierung von Cycloaliphaten durch Übergang von zwei CH_2 -Gruppen in CH -Gruppen.

Reaktionen mit Selenioxyd.

Mikrobiochemische Reaktionen.

Dehydrierung mit *Bazillus sphäricus*.

Antiphlogistisch wirksame Stoffe.

Ein jeder Substituent am Steroid-Gerüst wird mit allen Details, die zur Selektion auch nach diesen Gesichtspunkten erforderlich sind, beschrieben, z. B. der Bernsteinsäurehalbesther-Rest, ein Trifluoracetyl-Derivat, ein Zucker-Rest usw.

Erst nachdem all diese Vorkehrungen getroffen sind, kann der Dokumentar die Formel II unbesorgt aus seinem Gedächtnis entlassen. Es ist sichergestellt, daß sie unter jedem einschlägigem Gesichtspunkt wiedergefunden werden kann.

Bis zur Fertigstellung eines systematischen anorganischen Code speichern wir die anorganischen Chemikalien noch in ihrer Formelschreibweise ein. Ähnlich wird ein Teil der biologischen Begriffe vorerst in Gestalt ihrer wissenschaftlichen Bezeichnungen festgehalten.

Zur Zeit sind ca. 80000 chemische Verbindungen und allgemeine Stoffgruppenformeln für die Magnetbandspeicherung verschlüsselt worden. Insgesamt wurden bisher ca. 4700 maschinelle Selektionen ausgeführt, von denen 2500 für die Herstellung von Sachregistern dienten.

Die Möglichkeit des eigenhändigen Recherchierens

Es liegt im Wesen der großen Dokumentations-Systeme, daß niemand mehr den Inhalt ihrer Speicher voll überblicken kann. Hierdurch gibt es bei den Selektionen immer wieder Überraschungen. So muß man beispielsweise nachträglich oft feststellen, daß man eine Frage allzu allgemein gestellt hat und hierdurch auch Sachgebiete angesprochen hat, die in dem vorgefaßten Zusammenhang nicht interessieren. Wenn man beispielsweise mit der Blickrichtung auf aneurin-ähnliche oder luziferin-ähnliche Stoffe als Suchbedingung „biologisch bedeutsame Stoffe mit dem Thiazol-Gerüst“ aufgegeben hat, so wird auch die gesamte, umfangreiche Penicillin-Literatur mit herausgegeben, obwohl in diesem Falle ein einziger Hinweis hierauf genügt hätte.

Man wird diese Erfahrung verwenden, um die Fragestellung zu wiederholen, diesmal mit einschränkenden Bedingungen, wodurch die unerwünschte Literatur zurückgehalten wird („feed back“). Die Wiederholung einer Frage in einer neuen Variante läßt sich auch durch noch so angestrengtes Nachdenken vor der Formulierung des Auftrages nicht immer vermeiden.

Der umgekehrte Fall ist ebenso häufig. Hier werden auf eine allzu spezifische Frage nur sehr wenige oder überhaupt keine Hinweise gegeben, so daß es notwendig ist, die Fragestellung in allgemeinerer Form zu wiederholen.

Solche Fehlschläge können bei eigenhändigem Recherchieren am eigenen Arbeitsplatz nicht vorkommen, da diese Erfahrungen nicht erst nach Beendigung des gesamten Vorganges, sondern bereits gleich zu Anfang gewonnen werden. Oft können die Suchbedingungen noch rechtzeitig geändert werden. Andererseits gewinnt man auch ganz unerwartete Ergebnisse, so daß man die Selektion mit einer anderen Formulierung fortsetzt.

Auch in mancherlei anderer Hinsicht hat die Handkartei am Arbeitsplatz unbestreitbare Vorzüge gegenüber einer zentralen Institution¹⁰⁾. Der größte Teil der Hemmnisse, die der Befragung einer Zentrale im Wege stehen und die Benützung einer Handkartei begünstigen, ist psychologischer Art. Aber auch diesen Fakten muß ein Dokumentationsvorhaben Rechnung tragen.

Wir haben daher nach Möglichkeiten gesucht, dem Praktiker das Einrichten, das laufende Fortführen und Handhaben seiner eigenen Kartei zu erleichtern. Vor allem muß der experimentierende Chemiker von routinemäßiger Verschlüsselungsarbeit entlastet werden. Nur der Teil der Codifizierung muß ihm überlassen bleiben, für den seine Spezialkenntnisse unentbehrlich sind. Es muß weiterhin vermieden werden, daß jeder Fachmann seine literarische Aktivität isoliert entfaltet. Denn dann steht jedem nur der Nutzen seiner eigenen Arbeit zur Verfügung und die Beiträge seiner Kollegen bleiben ihm verschlossen. Es müssen vielmehr für einen großen Kreis von Chemikern alle registrierwürdigen Daten nach einem weitgehend einheitlichen System verschlüsselt und zentral aufbewahrt werden, damit sie für jeden zugänglich sind. Es bleibt die Aufgabe der Zentrale, für die Verteilung der interessierenden Publikationen an die Partner zu sorgen.

Bei der Verwirklichung dieses Planes haben wir einige Ergebnisse erzielt. Unsere Chemiker haben heute folgende Möglichkeiten (siehe Abb. 1):

1. Der einfachste Fall ist eine Fragestellung zu einem aktuellen Thema. Die Frage kann hochspezifisch, aber auch sehr allgemein sein, wobei alle Übergänge möglich sind. Die Zentrale liefert lichtgepauste Referate über die entsprechenden Publikationen.

2. Jeder Informationswunsch kann als Dauerauftrag aufgegeben werden. Dann wird der Fragesteller laufend über alle Neuzugänge unterrichtet. Diese Variante erfreut sich für feste Arbeitsgebiete, auf denen auch die Patentlage überwacht werden soll, bei unseren Chemikern steigender Beliebtheit.

3. Bei größeren Arbeitsgebieten werden die Literatursammlungen am Arbeitsplatz des Chemikers bald wieder unübersichtlich. Für diese Fälle haben wir einen zweiten Code für die präparative organische Chemie ausgearbeitet, der dieses Gebiet mit nur 500 verschiedenen Merkmalen beschreibt. Wenn man Verbindungen nach diesem Kurz-Code verschlüsselt, so muß man allerdings bei der späteren Selektion eine wesentlich geringere Treffsicherheit in Kauf nehmen als bei dem Original-Code für die Magnetbandspeicherung. Sofern jedoch die Sammlung einen Umfang von etwa 10000 Karten nicht übersteigt, reicht die Genauigkeit einer solchen abgekürzten Verschlüsselung im allgemeinen aus. Die wichtigste

¹⁰⁾ C. L. Bernier, Information Resources, Western Reserve University, Cleveland, Ohio, 203; Interscience Publishers, Inc. New York, 1958.

Eigenart dieses gekürzten Code liegt darin, daß er mit einfachen Hilfsmitteln, wie Randlochkarten oder Schlitzlochkarten gehandhabt werden kann.

Dieser Kurz-Code wurde aus dem Original-Code entwickelt. Die Verwandtschaft dieser beiden Systeme geht so weit, daß der Elektronenrechner die Umwandlung einer Original-Chiffre in die gekürzte Chiffre nach einem vorgegebenen Programm selbst durchführen kann. Eine auf diesem maschinellen Wege entstandene Chiffre unterscheidet sich in nichts von einer solchen, die von Menschenhand auf direktem Wege bei unmittelbarer Verwendung des Kurz-Code entstanden ist.

Dies versetzt die Zentrale nun in die Lage, dem Fragesteller nicht nur Literaturstellen auszuliefern, sondern fertig umgeschlüsselte Literaturstellen, wobei die neue, gekürzte Chiffre von der Maschine ermittelt und in einer gedruckten Liste festgehalten worden ist. Eine technische Arbeitskraft der Zentrale oder des Chemikers kann diese gekürzten Chiffren auf gewöhnliche Lochkarten übertragen. Auf diesem Wege richten wir z.Zt. für verschiedene Arbeitsgebiete unserer Chemiker Arbeitskarteien ein, ohne daß der Chemiker Verschlüsselungsarbeit leisten muß.

4. Für manche Spezialgebiete und für größere Sammlungen benötigt man jedoch fast die unverminderte Aussagefähigkeit der Original-Chiffre auf dem Magnetband. Wenn es gelingt, die Aussagekraft dieser Formelbeschreibung beim Übergang auf die herkömmlichen Speichermittel zu bewahren, so bedeutet dies, daß man sich für das Recherchieren von der zentralen Rechenanlage unabhängig machen kann. Dies würde alle Vorzüge einer Dokumentation mit programmgesteuerten Rechenautomaten einem großen Kreis von Chemikern zugänglich machen und alle Nachteile beseitigen, die sonst unausweichlich mit einer jeden derartigen zentralen Institution verknüpft sind.

Dieser Übergang der Magnetbandchiffre auf kleinere Speichermittel ist tatsächlich durch einen „Transformations“-Prozess möglich⁴⁾. Zu diesem Zwecke bildet man sich ein großes Repertoire von sinnvollen chemischen Stichwörtern und beauftragt den Rechenautomaten in einer entsprechenden Zahl von „Modellrecherchen“, den gesamten Magnetbandspeicher zu durchsuchen. Jede dieser Modellrecherchen zielt dabei auf eines der vorgegebenen Stichwörter ab. Das Ergebnis dieser Selektionen wird von der Maschine in einer gedruckten Liste aufgezeichnet. In dieser sind unter einem jeden Stichwort die aufgefundenen Literaturstellen aufgeführt. Ein solcher Maschinendruck hat alle Eigenarten eines auf vollautomatischem Wege entstandenen Sachregisterwerkes. Man findet beispielsweise die Salicylsäure-Körper, Halogen-aminobenzole usw. jeweils unter ihren Stichwörtern.

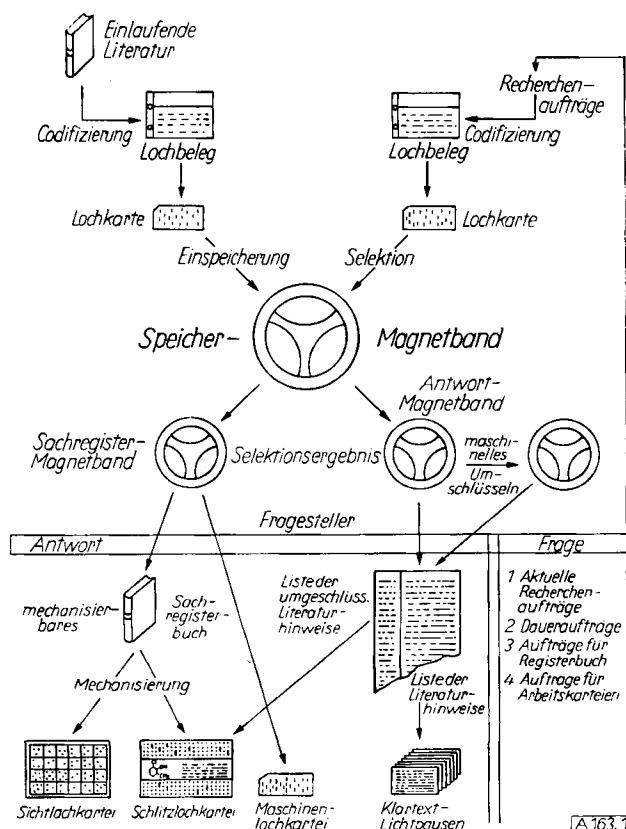


Abb. 1. Übersicht der Großdokumentation mit Magnetband-Speicher

Es gibt verschiedene Möglichkeiten, diese Sachregisterbücher zu mechanisieren, wenn in ungünstig gelagerten Fällen ihre visuelle Handhabung lästig erscheint. Denn diese Registerbücher sind so angelegt, daß sie der Anwendung der Indexmethoden zugänglich sind. Erprobt haben wir bisher die Mechanisierung mit Hilfe von Schlitzloch- und Sichtlochkarten. Weiterhin kann der Teilnehmer an einer solchen gemeinschaftlichen Dokumentation an Stelle des Sachregisters auch maschinelle gestanzte Lochkarten beziehen, um damit eine eigene Maschinenlochkartei mit einfachen Sortiermaschinen zu betreiben⁵⁾.

Somit können viele Partner bei der Erstellung einer in vollendeter Weise geordneten, umfangreichen Literatursammlung zusammenarbeiten. Die Kosten reduzieren sich, da sie sich auf die an diesem System teilnehmenden Partner verteilen. Jedem von ihnen steht jedoch der volle Nutzen der Gemeinschaftsarbeit zur Verfügung. Der Teilnehmer kann der Zentrale unmittelbar Aufträge für Recherchen übergeben. Er kann jedoch auch in regelmäßigen Abständen geeignete Informationsträger über sein Sachgebiet oder über die gesamte eingespeicherte präparative organische Chemie beziehen. Ihm bleibt es überlassen, ob er Maschinenlochkarten oder Sachregisterbücher auswählt, auf Grund derer er sich dann ohne jeden weiteren akademischen Arbeitsaufwand Schlitzloch- oder Sichtlochkarten anlegen kann, oder ob er diese Sachregisterbücher in der gewohnten Weise visuell handhaben will. Auf diesem Wege ist auch die Frage der Geheimhaltung des Recherchenprogramms gelöst.

Ausblick

Das Wiederverwerten aller vorliegenden Erfahrungen ist immer schwieriger geworden, je mehr Erfahrungen bekannt wurden, denn unsere herkömmlichen Registriermethoden für diese Erfahrungen basieren offenkundig oder in versteckter Weise auf dem begrenzten menschlichen Gedächtnis. Die Entwertung eines sich ständig vergrößernden Anteiles unserer Forschungsergebnisse durch banale Un auffindbarkeit wird mit naturgesetzlicher Konsequenz fortschreiten, wenn wir keine wirksamen Gegenmaßnahmen treffen. In einer nicht mehr fernen Zukunft wird die Wirtschaftlichkeit einer Institution der chemischen Forschung entscheidend davon abhängen, inwieweit diese Institution bereits vorhandene Forschungsergebnisse wiederzuverwerten vermag.

Gerade zur rechten Zeit haben zwei Zweige der Wissenschaft das Werkzeug geschaffen, das wir zur zuverlässigen und übersichtlichen Aufbewahrung der Ergebnisse unserer Forschung so dringend benötigen. Der Theorie der Klassifikation und Indexbildung verdanken wir einerseits die Richtlinien, nach denen die Begriffsbildung für Einspeicherung und Anfrage vollzogen werden muß. Andererseits ermöglichen uns die großen programmgesteuerten Rechenautomaten, die Forderungen der Theorie ohne Einschränkung zu verwirklichen.

Es kann gegen diese Hilfsmittel nicht der Einwand erhoben werden, daß ihre Anwendung zu kostspielig sei. Zweifellos ist die Einrichtung und Unterhaltung eines derartigen organisierten Dokumentations-Systems mit beträchtlichen Kosten verknüpft. Die reinen Kosten sind für ein Projekt der Wirtschaft jedoch bisher nur dann prohibitiv gewesen, wenn entweder billigere Wege zum gleichen Ziel bekannt waren, oder wenn darauf verzichtet werden konnte, das Ziel zu erreichen.

Zweifellos sind völlig neue Arbeitsgebiete der Gefahr einer Doppelbearbeitung wesentlich weniger ausgesetzt als ältere. Jedoch sind auch bei den modernen Arbeitsrichtungen stets gewisse Vorarbeiten nötig, die sich auf dem Gebiet der klassischen präparativen Chemie abspielen. Der Mangel an wirksamen Informationsmöglichkeiten verursacht be-

reits heute einen beträchtlichen Prozentsatz fehl disponierter Forschungsvorhaben. Um solche Fehldispositionen zu vermeiden, sind Aufwendungen erforderlich, die nur einige wenige Prozent derjenigen Beträge ausmachen, die man für die experimentelle Erarbeitung von Forschungsergebnissen unbedenklich auszugeben bereit ist. Also braucht man um die Wirtschaftlichkeit einer umfassenden Dokumentation keine Sorge zu haben. Gilt dies bereits heute, so erst recht für die Zukunft.

Allerdings muß ein größeres Dokumentationsvorhaben zunächst ein unrentables Anlaufstadium überwinden, ebenso wie eine jede andere Investition in der Wirtschaft. Denn bei vollen Kosten ist der Nutzen des Systems anfänglich klein, da die Speicher erst eine bestimmte Größe erreichen müssen, bevor sie anfangen, einen wesentlichen Ertrag abzuwerfen. Diese Phase kann sich über mehrere Jahre hinziehen.

Auf dem Wege zu einer umfassenden Dokumentation ist schließlich noch eine andere Schwierigkeit zu überwinden: Dem naturwissenschaftlich-technisch orientierten Fachmann leuchtet zwar die Notwendigkeit eines angemessenen

maschinellen Hilfsmittels rasch ein. Die Vorstellung jedoch, daß es für eine erfolgreiche Registrierung seines Wissensgebietes neben der profunden Kenntnis dieses Gebietes noch anderer, fachfremder Kenntnisse bedarf, ist noch wenig verbreitet. Dem Chemiker allerdings ist es bei der Vielseitigkeit seines Arbeitsfeldes nichts Ungewohntes, daß er einen Teil seines Gebietes, welches er bisher zunächst nach besten Kräften noch selbst bearbeitet hat, bald als separates Arbeitsfeld einem Experten überlassen muß. Diese Entwicklung bahnt sich heute für das literarische Gebiet an. Die Voraussetzungen hierfür und die in Aussicht stehenden Ergebnisse sind auf keinem Gebiet günstiger als in der Chemie.

Zu diesen sachlichen Voraussetzungen auch die personellen Grundlagen zu schaffen, hierin muß man heute eine dankbare Aufgabe für die chemische Lehre und Forschung der Zukunft erblicken.

Die Verfasser danken zahlreichen Kollegen und besonders Dr. A. Schwalbach für Ratschläge und Mitarbeit.

Eingegangen am 8. August 1961 [A 163]

Die Valenzeigenschaften des vierbindigen Kohlenstoff-Atoms in Molekülen der Symmetrie C_{3v}

Von Doz. Dr. W. ZEIL

Institut für Physikalische Chemie und Elektrochemie der T.H. Karlsruhe

Beim Übergang von Methyl- zu tert.-Butyl-Derivaten beobachtet man eine Vergrößerung des Bindungsabstandes am tertiären C-Atom. Desgleichen steigt die Polarität dieser Bindung bei den Halogeniden an. Es wird gezeigt, daß eine von Coulson früher gegebene Beziehung es gestattet, Valenzwinkel und Bindungsabstände sowie die Polarität der Bindung in Abhängigkeit vom Hybridisierungsparameter λ zu beschreiben. λ ist ein Maß für den prozentualen p-Charakter eines hybridisierten Orbitals.

Einleitung und Problemstellung

Die folgenden Betrachtungen nahmen ihren Ausgang an molekülphysikalischen Untersuchungen zum Problem der Hyperkonjugation¹⁾. Während man im allgemeinen für eine C-C-Einfachbindung, wie sie in gesättigten Kohlenwasserstoffen vorliegt, Abstände zwischen 1,534 und 1,543 Å beobachtete²⁾, fand man für die C-C-Einfachbindung neben einer Dreifachbindung, also z.B. im Methylacetylen³⁾ und Acetonitril⁴⁾, wesentlich verkürzte Abstände von etwa 1,46 Å. Diese Verkürzung des Abstandes wurde schon sehr früh von Pauling⁵⁾ und später von Coulson⁶⁾ zum Teil auf eine Änderung des kovalenten Radius und zum anderen Teil auf „Resonanzeffekte“ zurückgeführt. So ergeben z.B. die von Coulson⁷⁾ für den Kohlenstoff in den verschiedenen Hybridisierungszuständen angenommenen kovalenten Radien von

$$\begin{aligned} r_{C(sp^3)} & 0,771 \text{ Å} \\ r_{C(sp^2)} & 0,749 \text{ Å} \\ r_{C(sp)} & 0,735 \text{ Å} \end{aligned}$$

unter Benutzung der Regel von Pauling⁵⁾, die besagt, daß sich der Bindungsabstand durch einfache Addition der so-

genannten kovalenten Radien errechnen läßt, einen Abstand der C-C-Einfachbindung im Methylacetylen von 1,506 Å. Die Differenz des Wertes von 1,506 Å gegenüber dem gemessenen Wert von 1,459 Å wird von Coulson einem Hyperkonjugationseffekt zugeschrieben⁶⁾. Wichtig für das folgende ist, daß die Coulsonschen Radien aus C-H-Abständen unter Zugrundelegung eines Wertes von 0,322 Å für den Wasserstoff-Radius errechnet wurden.

Zur Klärung der Frage, ob die Abstandsverkürzung durch Hyperkonjugationseffekte bedingt ist, haben wir vor längerer Zeit mikrowellenspektroskopische Untersuchungen am Trichloracetonitril⁸⁾, Trimethylacetonitril¹⁾, tert.-Butylacetylen⁹⁾, tert.-Butylchloracetylen⁹⁾ sowie Phenylacetylen⁹⁾ begonnen. Unsere Untersuchungen wurden ergänzt durch unabhängig von uns ausgeführte Arbeiten aus dem National Bureau of Standards von Mann und Mitarb.¹⁰⁾. Die Untersuchungen von Mann erlauben es, mit ziemlicher Sicherheit die Länge der C-C-Einfachbindung neben der Dreifachbindung im Trimethylacetonitril und tert.-Butylacetylen festzulegen. In beiden Fällen werden Abstände von etwa 1,49 Å angegeben. Es wird also ein größerer Abstand als in den Methyl-Derivaten beobachtet.

Da unsere Untersuchungen schon in einem frühen Stadium zeigten, daß die Bindungsverhältnisse am tertiären C-Atom für eine Auswertung und Diskussion der Mikrowellenspektren von Bedeutung sind, haben wir auch die Mikrowellenspektren des tert.-

¹⁾ (a) W. Zeil u. J. F. Pfommer, Z. Elektrochem. 61, 938 [1957]; (b) H. Heel u. W. Zeil, ebenda 64, 962 [1960].

²⁾ A. Almenningen u. O. Bastiansen, Acta chem. scand. 9, 815 [1955]; H. C. Allen u. E. K. Plyler, J. chem. Physics 31, 1062 [1959]; G. E. Hansen u. D. M. Dennison, ebenda 20, 313 [1952].

³⁾ L. F. Thomas, E. I. Sherrer u. J. Sheridan, Trans. Faraday Soc. 51, 619 [1955].

⁴⁾ C. C. Costain, J. chem. Physics 29, 864 [1958].

⁵⁾ L. Pauling: The Nature of the Chemical Bond, Cornell University Press, Ithaca 1948.

⁶⁾ C. A. Coulson: Valence, Clarendon Press, Oxford 1952.

⁷⁾ C. A. Coulson, Contribution à l'étude de la structure moléculaire, Maison Desoer, Liège 1947.

⁸⁾ loc. cit. 1a), S. 939.

⁹⁾ W. Zeil, M. Winnewisser, H. K. Bodenseh u. H. Buchert, Z. Naturforsch. 15a, 1011 [1960].

¹⁰⁾ L. J. Nugent, D. E. Mann u. D. R. Lide Jr., N.B.S. Report 7162 [1961].