

Computerunterstützte Zuordnung von ^{13}C -NMR-Spektren

Wolfgang Robien

Institut für Organische Chemie, Universität Wien, A-1090 Wien, Österreich

(Eingegangen 7. Oktober 1982. Angenommen 3. November 1982)

Computer-Assisted Assignment of Carbon-13 NMR Spectra

A computer program for the assignment of ^{13}C resonances to the respective carbons of a known structure is presented. The algorithm is based on the prediction of chemical shift ranges from a data base containing carbon-centered substructural environments and their corresponding chemical shifts. The method permits a stepwise solution of the assignment problem using chemical shift arguments up to a five-bond radius.

(Keywords: Computer program; C-13 Data base)

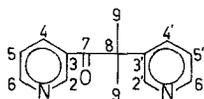
Einleitung

Obwohl der hohe Informationsgehalt der ^{13}C -NMR-Spektroskopie sehr früh erkannt wurde, war die routinemäßige Aufnahme dieses Kerns wegen seiner geringen Empfindlichkeit lange Zeit nicht möglich. Erst durch Anwendung der *Fourier*transformation und die Einführung der Hochfeldspektrometer wurden komplexe organische Moleküle in Mengen von wenigen Milligramm und darunter erfaßbar. Die Zuordnung der einzelnen Linien eines ^{13}C -NMR-Spektrums erfolgt einerseits durch Vergleich mit Literaturdaten von Molekülen mit ähnlichen Strukturelementen und andererseits durch Anwendung von speziellen Meßtechniken¹ wie selektiver Entkopplung, Shiftreagentien oder zweidimensionalen NMR-Experimenten^{2,3}. Diese Methoden erfordern relativ lange Meßzeiten und sind daher komplexen Problemstellungen vorbehalten; hingegen sind die chemische Verschiebung und die Signalmultiplizität Parameter, die dem Spektroskopiker leicht zugänglich

sind, weshalb versucht werden soll, daraus möglichst viel Information herzuleiten. Die Zuordnung von ^{13}C -NMR-Spektren wird sich daher primär auf diese Parameter und somit auf den Vergleich mit geeignetem Referenzmaterial⁴⁻¹¹ stützen. Hier erweist sich wegen der ungeheuren Fülle von Literaturdaten eine Systematisierung in Form einer spektroskopischen Datenbank¹²⁻¹⁵ als vorteilhaft. Bei der Auswertung von ^{13}C -NMR-Spektren ergeben sich zwei prinzipielle Problemstellungen: Einerseits die Ermittlung einer unbekanntens Struktur aus den spektralen Größen¹⁶⁻²⁴ und andererseits die Signalzuordnung zu einer bereits bekannten Struktur. Im folgenden soll ein einfacher Algorithmus hergeleitet werden, der die Zuordnung der Signale des ^{13}C -Spektrums zu einer vorgegebenen Struktur gestattet.

Methodik

Die chemische Verschiebung eines ^{13}C -Kerns wird hauptsächlich durch die Hybridisierung und die chemische Umgebung, also durch sterische und elektronische Wechselwirkungen bestimmt. Die Erfassung der Hybridisierung und der chemischen Umgebung kann mittels eines HOSE-Codes²⁵⁻²⁷ (Hierarchisch geordnete sphärische Darstellung der Umgebung) erfolgen. Dieser Code beschreibt die Einflüsse der Substituenten auf die chemische Verschiebung eines Kohlenstoffatoms in einer leicht verständlichen Symbolsprache. Mit diesem Code kann nun eine Datei erstellt werden, die Strukturelemente und die zugeordneten chemischen Verschiebungen enthält. Für eine vorgegebene Struktur kann mittels dieser gespeicherten Struktur-Shift-Korrelationen ein ^{13}C -Spektrum abgeschätzt werden²⁸. Diese Abschätzung kann für verschiedene Genauigkeiten, d. h. Anzahl der Bindungen, über die Substituenteneinflüsse berücksichtigt werden sollen, durchgeführt werden. In Tab. 1 sind die abgeschätzten Spektren für Metyrapone mit Genauigkeit 3 (γ -Substituenten) und 5 (ε -Substituenten) den gemessenen Werten gegenübergestellt. Mit diesen, aus dem vorhandenen Referenzmaterial berechneten Erwartungsbereichen der einzelnen Linien, kann nun die Zuordnung des experimentellen Spektrums durchgeführt werden.



Metyrapone

Tabelle 1. *Abgeschätzte und experimentelle chemische Verschiebungen von Metyrapone mit Genauigkeit 3 und 5.* [Mw = Mittelwert, Stdev = Standardabweichung, Min = minimaler Shiftwert, Max = maximaler Shiftwert, Anzahl = Zahl der verwendeten Referenzspektren; alle Shiftwerte in ppm gegen TMS]. Bei der hohen Genauigkeit 5 haben nur sehr wenige Verbindungen ausreichende strukturelle Ähnlichkeit mit Metyrapone, daher wird für einige C-Atome nur die Identität berücksichtigt (Anzahl = 1)

Berechnung mit Genauigkeit 3					Berechnung mit Genauigkeit 5					exp.
Mw	Stdev	Min	Max	Anzahl	Mw	Stdev	Min	Max	Anzahl	
150,1	0,5	149,1	150,7	6	150,1	0,6	149,1	150,7	5	150,7
131,4	0,4	131,1	132,1	5	131,4	0,4	131,1	131,6	2	131,6
136,1	0,4	135,5	136,6	7	136,3	0,2	136,0	136,6	5	136,6
122,2	6,4	110,4	140,3	35	123,1	0,2	123,0	123,5	5	123,0
150,6	2,4	146,3	155,1	21	152,7	0,6	152,1	153,8	6	152,2
147,8	—	147,8	147,8	1	147,8	—	147,8	147,8	1	147,8
139,9	—	139,9	139,9	1	139,9	—	139,9	139,9	1	139,9
128,3	7,1	123,3	133,5	2	133,5	—	133,5	133,5	1	133,5
123,1	5,0	113,8	130,8	7	123,8	—	123,8	123,8	1	123,8
150,6	2,4	146,3	155,1	21	148,6	—	148,6	148,6	1	148,6
199,8	1,1	198,5	201,3	5	201,3	—	201,3	201,3	1	201,3
50,4	0,2	50,2	50,5	2	50,5	—	50,5	50,5	1	50,5
26,8	0,4	26,2	27,3	5	27,3	—	27,3	27,3	1	27,3

Für eine Substanz mit n nicht äquivalenten Kohlenstoffatomen und n Linien wird eine Zuordnungsmatrix der Dimension $n \times n$ erstellt. Die Elemente a_{ij} dieser Matrix können nun die Werte 0 oder 1 annehmen. Wenn $a_{ij} = 1$ ist, bedeutet dies, daß dem Kohlenstoff C_i die Linie L_j zugeordnet ist. Zu Beginn der iterativen Spektrenzuordnung werden alle Elemente auf 1 gesetzt, das bedeutet, daß die Summe aller Matrixelemente n^2 ist. Eine eindeutige Zuordnung ist dann gegeben, wenn diese Summe auf den Wert n so reduziert werden kann, daß jedem C-Atom nur eine einzige Linie zugeordnet ist. Das erste Zuordnungskriterium ist die Signalmultiplizität. Es kann sowohl die Multiplizitätsinformation aus SFORD-Spektren oder J -modulierten Spektren verwendet werden, ebenso kann die Multiplizität vernachlässigt werden. Die weitere Zuordnung erfolgt mittels der Signalintensität. Für die Intensität sind nur ganzzahlige Werte erlaubt, daher ergibt sich für n Kohlenstoffatome immer eine $n \times n$ Matrix. Nach jeder Reduktion der Anzahl der auf 1 gesetzten Elemente wird die spektroskopische Relevanz der Matrix untersucht. Es wird überprüft, ob alle Linien verwendet werden und ob alle C-Atome zugehörige Linien besitzen. Weiters werden nach jeder Verminderung der auf 1 gesetzten Elemente eindeutig zugeordnete Linien oder Liniengruppen identifiziert und

damit weitere Zuordnungsmöglichkeiten ausgeschlossen. Daran anschließend beginnen die Iterationszyklen mit den abgeschätzten Erwartungsbereichen der einzelnen Linien. Im ersten Zyklus werden die Shiftbereiche unter Berücksichtigung der α -Substituenten verwendet, dann erfolgt die Einbeziehung der β -Substituenten, usw. Es werden insgesamt fünf Iterationszyklen durchgeführt, was einer Berücksichtigung der Shifteinflüsse der ϵ -Substituenten entspricht. Nach jeder Iteration wird die Matrix — wie bereits eingangs erwähnt — überprüft und gegebenenfalls weiter reduziert. Dieses iterative Verfahren bewirkt einerseits die Zuordnung auf Grund von bekannten Verschiebungsbereichen, jedoch auch für manche C-Atome eine Zuordnung nach dem Ausschließungsprinzip.

Der dargestellte Algorithmus entspricht der Vorgangsweise des Spektroskopikers, der zuerst Multiplizität und Intensität zur Zuordnung heranzieht und daran anschließend Shiftargumente beginnend bei einfachen funktionellen Gruppen einsetzt und nach jedem Zuordnungsschritt gesicherte Erkenntnisse aus den weiteren Überlegungen ausschließt. Die computerunterstützte Zuordnung kann jedoch Substituenteneinflüsse über einen Radius von fünf Bindungen berücksichtigen, was dem Spektroskopiker nicht mehr möglich ist.

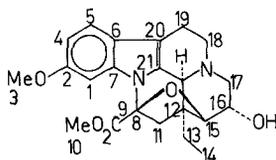
Für die Interpretation der Resultate wurden zwei Kenngrößen gewählt: Erstens wird die Datengüte der vorhandenen Referenzverbindungen bezüglich der speziellen Problemstellung beurteilt und zweitens die Übereinstimmung zwischen abgeschätztem Spektrum und den eingegebenen Linien berechnet. Die Datengüte kann Werte von 0% bis 100% annehmen und beschreibt über wie viele Bindungen Substituenteneinflüsse mit den vorhandenen Referenzverbindungen bei dieser Problemstellung berücksichtigt werden können. Der *R*-Wert beschreibt die Differenz von berechneten und experimentellen Werten, weiters werden C-Atome mit großer Abweichung von Berechnung und Messung markiert. Abschließend wird zur Überprüfung der Zuordnung durch den Spektroskopiker eine Liste relevanter Referenzspektren ausgegeben, was die Suche nach Vergleichsmaterial erspart.

Ergebnisse und Diskussion

Zuordnung von ^{13}C -NMR-Spektren

Mit insgesamt 2700 ^{13}C -Spektren wurde eine Datenbank mit der beschriebenen Struktur-Shift-Korrelationstabelle erstellt. Am folgenden Beispiel soll die Effizienz der computerunterstützten Zuordnung von ^{13}C -Spektren mittels des dargelegten Rechenschemas gezeigt werden:

Vincarodin besitzt ein hinreichend komplexes ^{13}C -Spektrum, das im folgenden zugeordnet wird und mit der bekannten Literaturzuordnung verglichen werden kann²⁹. Die Eingabeparameter sind lediglich die



Vincarodin

bekannte Struktur und die gemessenen chemischen Verschiebungen unter wahlweiser Berücksichtigung der Multiplizität. Vincarodin enthält 22 Kohlenstoffatome, daher enthält die Zuordnungsmatrix zu Beginn 484 Elemente mit dem Wert 1. Nach Verwendung der Multiplizitätsinformation kann dieser Wert auf 134 gesenkt werden. Die schrittweise Einbeziehung der Substituenten reduziert diese Anzahl weiter. Die endgültige Zuordnung ist nach Berücksichtigung der γ -Substituenten erreicht. Insgesamt ergeben sich für 22 Kohlenstoffatome 31 Zuordnungsmöglichkeiten. In Tab. 2 ist die Literaturzuordnung mit dem erhaltenen Resultat der Rechnung verglichen. Alle 22 Linien sind richtig zugeordnet, davon 15 eindeutig, obwohl die Daten-güte für diese Problemstellung nur 24,5% beträgt.

Für die aromatische Methoxygruppe wurde ein Wert von $54,5 \pm 1,7$ ppm aus 4 Vergleichsspektren mit einer Genauigkeit von fünf Bindungen abgeschätzt, für die Estermethylgruppe ein Wert von $53,7 \pm 0,7$ ppm aus 2 Referenzdaten. Auf Grund der teilweisen Überlappung der Erwartungsbereiche wurde auch keine eindeutige Zuordnung getroffen. Weder für C-8 noch C-21 konnten passende Vergleichsdaten gefunden werden, daher werden alle drei möglichen Singulett (137,8, 133,2 und 90,5 ppm) diesen C-Atomen zugeordnet. Eine einzige Struktur, die das Fragment von C-8 oder C-21 mit den α -Substituenten enthält, könnte weitere Zuordnungsmöglichkeiten ausschließen. Für die Probleme bei der Zuordnung von C-7, C-8 und C-21 ist ausschließlich die schlechte Datenqualität bei diesen Strukturelementen verantwortlich, trotzdem treten keine falschen Zuordnungen auf. Die Verwendung der Multiplizitätsinformation aus J -modulierten Spektren führt zu einer größeren Anzahl von vertauschbaren chemischen Verschiebungen; die Vernachlässigung der Multiplizität verschlechtert die Zuordnung nicht wesentlich, was die Effizienz der Verwendung von abgeschätzten Shiftbereichen und des Matrixalgorithmus demonstriert.

Tabelle 2. Vergleich der Literaturzuordnung²⁹ von Vincarodin mit der computerunterstützten Zuordnung. [a] Multiplizitätsinformation aus SFORD-Spektren; [b] Multiplizitätsinformation aus *J*-modulierten Spektren³⁰; [c] Ohne Berücksichtigung der Multiplizität

C Nr.	Zuordnung aus Lit. ²⁹ ppm	Computerunterstützte Zuordnung		
		[a] ppm	[b] ppm	[c] ppm
1	96,2	96,2	96,2	96,2
2	156,3	156,3	156,3	156,3
3	55,3	55,3; 52,9	52,9; 55,3; 56,5	52,9; 55,3; 56,5
4	109,6	109,6	109,6	109,6; 110,9
5	118,6	118,6	118,6	118,6
6	125,3	125,3	125,3	125,3
7	137,8	133,2; 137,8	133,2; 137,8	133,2; 137,8
8	90,5	90,5; 133,2; 137,8	90,5; 133,2; 137,8	90,5; 133,2; 137,8
9	168,6	168,6	168,6	168,6
10	52,9	52,9; 55,3	52,9; 55,3	52,9; 55,3
11	46,1; 45,0	45,0; 46,1	43,9; 45,0; 46,1	43,9; 45,0; 46,1
12	43,9	43,9	43,9; 45,0; 46,1	43,9; 45,0; 46,1
13	25,7	25,7	25,7	25,7
14	8,3	8,3	8,3	8,3
15	82,0	82,0	82,0	82,0
16	66,3	66,3	66,3	66,3
17	45,0; 46,1	45,0; 46,1	43,9; 45,0; 46,1	43,9; 45,0; 46,1
18	50,1	50,1	50,1	50,1; 52,9; 55,3; 56,5
19	18,4	18,4	18,4	18,4
20	110,9	110,9	110,9	109,6; 110,9
21	133,2	90,5; 133,2; 137,8	90,5; 133,2; 137,8	90,5; 133,2; 137,8
22	56,5	56,5	52,9; 55,3; 56,5	50,1; 52,9; 55,3; 56,5

Fehlerquellen

Mit dem vorgestellten Datensystem wurden 2700 Testrechnungen durchgeführt.

Mit diesem Datenmaterial gibt es insgesamt 422 891 Zuordnungsmöglichkeiten, die endgültige Zuordnung ist bei 45 521 Kombinationen gegeben. Auf Grund der Multiplizität, Intensität und der Shiftargumente konnten 362 030 mögliche Zuordnungen ausgeschlossen werden, was eine erreichte Anzahl von 60 681 Möglichkeiten ergibt. Dies bedeutet, daß im Mittel 1,34 Linien pro Kohlenstoffatom zugeordnet wurden, bzw. 65% der Linien eindeutig zugeordnet wurden. Die Fehlerquote beträgt dabei 0,35% bezogen auf die ausgeschiedenen Möglichkeiten und 2,81% bezüglich der endgültigen Zuordnung. Es

wurde bei 2700 Testbeispielen nur 1 von jeweils 48 Linien falsch zugeordnet, was die Effizienz der Methode demonstriert, obwohl bei der derzeitigen Programmversion die Stereochemie nicht berücksichtigt wird.

Zusammenfassung

Die Zuordnung von ^{13}C -NMR-Spektren kann durch den Einsatz von elektronischen Datenverarbeitungsanlagen wesentlich vereinfacht werden. Die Verwendung von abgeschätzten Verschiebungsbereichen unter Berücksichtigung der Substituenten über ein bis fünf Bindungen in Kombination mit einem Algorithmus, der die Zahl der Zuordnungsmöglichkeiten schrittweise verkleinert, hat sich als vorteilhaft erwiesen. Diese Strategie entspricht der Vorgangsweise des Spektroskopikers und führt von einer groben Zuordnung durch Multiplizität, Intensität und α -Substituenteneinflüssen bis zur Berücksichtigung der ϵ -Substituenten. Der Rechner kann und soll den Spektroskopiker nicht ersetzen, er bietet jedoch durch seine Schnelligkeit und Zuverlässigkeit eine wertvolle Hilfe. Selbstverständlich ist eine kritische Überprüfung der erhaltenen Resultate notwendig. Gerade durch aufgetretene Differenzen zwischen Spektrenabschätzung aus einem großen Reservoir an Referenzdaten und den gemessenen chemischen Verschiebungen können selektiv weitere Experimente geplant werden, die offene Problemstellungen lösen sollen.

Dank

Mein Dank gilt Dr. *E. Haslinger* für wertvolle Diskussionen, sowie Dr. *H. Kalchauer* und Hr. *J. Scharf* für die Hilfe bei der Datenerstellung.

Weiters danke ich dem Interfakultären Rechenzentrum der Universität Wien für die großzügige Vergabe von Rechenzeit.

Experimenteller Teil

Die Programme sind größtenteils in FORTRAN IV geschrieben, einige Unterprogramme jedoch in Maschinensprache. Die durchschnittliche Rechenzeit für die Zuordnung von komplexen organischen Molekülen (C_{20} — C_{30}) beträgt 3—8 s CPU-Zeit an einer CDC CYBER 170/720 Rechenanlage. Selbstverständlich sind noch zahlreiche andere Suchalgorithmen und Interpretationshilfen für ^{13}C -NMR-Spektren implementiert, wie Isomerensuche, Suche nach homologen Reihen oder nach Summenformelbereichen, Suche nach Linien und Liniengruppen, Suche nach Substanznamen oder Namensteilen und eine Teilstruktursuche*. Der Speicherplatzbedarf für das komplette Datensystem beträgt 12 K 60 Bit Worte.

* Die Datenbank ist allen Interessenten über das Interfakultäre Rechenzentrum der Universität Wien zugänglich.

Literatur

- ¹ Wehrli F. W., Wirthlin T., Interpretation of Carbon-13 NMR Spectra. London: Heyden. 1976.
- ² Freeman R., Morris G. A., Bull. Magn. Res. **1**, 5 (1979).
- ³ Kalchauer H., Haslinger E., Robien W., Monatsh. Chem. **113**, 805 (1982).
- ⁴ Pretsch E., Clerc T., Seibl J., Simon W., Tabellen zur Strukturaufklärung organischer Verbindungen mit spektroskopischen Methoden. Berlin-Heidelberg-New York: Springer. 1976.
- ⁵ Stothers J. B., Carbon-13 NMR Spectroscopy. New York: Academic Press. 1972.
- ⁶ Breitmaier E., Voelter W., ¹³C NMR Spectroscopy. Weinheim: Verlag Chemie. 1974.
- ⁷ Smith W. B., Annual Reports on NMR Spectroscopy **8**, 199 (1978).
- ⁸ Crabb T. A., Annual Reports on NMR Spectroscopy **8**, 1 (1978).
- ⁹ Blunt J. W., Stothers J. B., Org. Magn. Res. **9**, 439 (1977).
- ¹⁰ Wehrli F. W., Nishida T., Fortschr. Chem. Org. Naturst. **36**, 1 (1978).
- ¹¹ Bohlmann F., Zeisberg R., Klein E., Org. Magn. Res. **7**, 426 (1975).
- ¹² Bremser W., Z. Anal. Chem. **286**, 1 (1977).
- ¹³ Dubois J. E., Bonnet J. C., Anal. Chim. Acta **112**, 245 (1979).
- ¹⁴ Heller S. R., Milne G. W. A., Anal. Chim. Acta **122**, 117 (1980).
- ¹⁵ Milne G. W. A., Zupan J., Heller S. R., Miller J. A., Org. Magn. Res. **12**, 289 (1979).
- ¹⁶ Carhart R. E., Smith D. H., Brown H., Djerassi C., J. Amer. Chem. Soc. **97**, 5755 (1975).
- ¹⁷ Gray N. A. B., Crandell C. W., Nourse J. G., Smith D. H., Dageforde M. L., Djerassi C., J. Org. Chem. **46**, 703 (1981).
- ¹⁸ Carhart R. E., Smith D. H., Gray N. A. B., Nourse J. G., Djerassi C., J. Org. Chem. **46**, 1708 (1981).
- ¹⁹ Cheer C. J., Smith D. H., Djerassi C., Tursch B., Braekman J. C., Daloz D., Tetrahedron **32**, 1807 (1976).
- ²⁰ Smith D. H., Gray N. A. B., Nourse J. G., Crandell C. W., Anal. Chim. Acta **133**, 471 (1981).
- ²¹ Nourse J. G., Smith D. H., Carhart R. E., Djerassi C., J. Amer. Chem. Soc. **102**, 6289 (1980).
- ²² Abe H., Yamasaki T., Fujiwara I., Sasaki S., Anal. Chim. Acta **133**, 499 (1981).
- ²³ Shelley C. A., Munk M. E., Anal. Chim. Acta **133**, 507 (1981).
- ²⁴ Fujiwara I., Okuyama T., Yamasaki T., Abe H., Sasaki S., Anal. Chim. Acta **133**, 527 (1981).
- ²⁵ Jezl B. A., Dalrymple D. L., Anal. Chem. **47**, 203 (1975).
- ²⁶ Bremser W., Anal. Chim. Acta **103**, 355 (1978).
- ²⁷ Gray N. A. B., Nourse J. G., Crandell C. W., Smith D. H., Djerassi C., Org. Magn. Res. **15**, 375 (1981).
- ²⁸ Bremser W., Chem. Ztg. **104**, 53 (1980).
- ²⁹ Neuss N., Boaz H. E., Occolowitz J. L., Wenkert E., Schell F. M., Potier P., Kan C., Plat N. M., Plat M., Helv. Chim. Acta **56**, 2660 (1973).
- ³⁰ Cookson D. J., Smith B. E., Org. Magn. Res. **16**, 111 (1981).