

В. А. БАРХАШ, Ю. П. ДРОБЫШЕВ, С. П. СОКОЛОВ, Л. Ф. СЕКЕРИНА,
член-корреспондент АН СССР В. А. КОПТЮГ

ОПЫТ ПОСТРОЕНИЯ МАШИННОЙ СИСТЕМЫ ОПРЕДЕЛЕНИЯ СТРУКТУРНЫХ ФРАГМЕНТОВ НЕИЗВЕСТНЫХ СОЕДИНЕНИЙ ПО СПЕКТРАМ Я.М.Р.

Широкое внедрение в химические исследования методов спектроскопии молекул привело к накоплению столь больших массивов спектральной информации, что ее эффективное использование становится возможным лишь с помощью вычислительной техники. Весьма перспективно, в частности, применение ЭЦВМ для логической обработки накопленной информации при решении задач, связанных с установлением строения химических соединений. Комплекс машинных систем, создаваемых для этой цели, должен включать: 1) системы опознавания химических соединений по их спектральным характеристикам; 2) системы извлечения из различных видов спектров информации о структурных фрагментах молекулы исследуемого вещества; 3) систему «сборки» возможных структур из выявленных структурных фрагментов с учетом полного элементного состава исследуемого соединения, определяемого классическими методами химического анализа или методом масс-спектрометрии высокого разрешения.

К настоящему времени в ряде стран уже созданы и эксплуатируются системы первого типа, использующие высокую характеристичность и-к. спектров (¹⁻³). Возможность построения машинных систем второго типа обсуждалась применительно к инфракрасной спектроскопии (^{4, 5}) и особенно широко для масс-спектрометрии (^{6, 7}). Сообщалось также о создании системы определения хромофорных группировок органических соединений по их электронным спектрам поглощения (⁸). Попытки же машинного извлечения структурной информации из спектров п.м.р. относятся пока к ограниченным группам соединений (⁹⁻¹¹)*. В этой работе мы предлагаем сравнительно простой и достаточно общий принцип построения машинной системы определения структурных фрагментов неизвестных соединений по спектрам п.м.р.

При извлечении структурной информации из спектров п.м.р. широко используются имеющиеся в литературе сведения о зависимости химических сдвигов сигналов водородсодержащих групп от их химического окружения. Существует, однако, большое число факторов, искажающих простые спектро-структурные корреляции (эффекты анизотропии магнитной восприимчивости, влияние электрических диполей удаленных полярных групп, затрудненность вращения вокруг некоторых типов ординарных связей и т. д.). Можно, конечно, пытаться непрерывно расширять корреляционные таблицы, но этот путь представляется нам мало перспективным при построении машинной системы. Целесообразнее использовать для этой цели те спектро-структурные корреляции, которые содержатся в неявном виде в больших коллекциях спектров п.м.р., полагая, что в та-

* Имеется большое число работ по машинному расчету вида спектров п.м.р. различных типов спиновых систем. Соответствующие программы, так же как и вычислительные программы для других видов спектроскопии, могут успешно использоваться на заключительном этапе для уточнения деталей строения молекулы исследуемого вещества.

ких коллекциях практически всегда найдутся спектры соединений, содержащих хотя бы часть крупных структурных блоков исследуемого вещества. Отыскание таких модельных соединений по спектру п.м.р. исследуемого вещества позволило бы идентифицировать структурные фрагменты.

Если анализ мультиплетности сигналов (с использованием, в случае необходимости, двойного ядерного магнитного резонанса) позволяет выделить в спектре п.м.р. исследуемого соединения наборы сигналов, принадлежащих водородсодержащим группам, образующим связанные структурные блоки, то поиск модельных соединений можно вести по каждому из этих наборов сигналов или по их совокупности.

Рассмотрим в качестве примера спектр п.м.р. изобутилового эфира бензойной кислоты $C_6H_5COO-CH_2CH(CH_3)_2$ (спектр № 75 по каталогу ⁽¹²⁾). В этом спектре присутствуют пять сигналов: дублет при 1,00, мультиплет при 2,08, дублет при 4,10 и мультиплеты с центрами при 7,48 и 8,06 м.д. (δ -шкала); соотношение интенсивностей сигналов 6 : 1 : 2 : 3 : 2. Анализируя этот спектр, можно прийти к выводу, что три первых сигнала принадлежат фрагменту $(CH_3)_2CHCH_2-$, а два последних — монозамещенному бензольному кольцу. Поиск модельных соединений по этим двум наборам сигналов позволил бы решить вопрос о характере групп, с которыми связаны эти фрагменты (в данном примере — о характере перемычки между ними). Однако в общем случае выделение в спектре п.м.р. наборов сигналов, принадлежащих достаточно крупным структурным блокам, и интерпретация этих наборов не является столь простым, как в приведенном примере. Поэтому целесообразно рассмотреть все возможные комбинации из n имеющихся в спектре сигналов по m сигналов, принимая последовательно m равным: n , $n-1$, $n-2$ и т. д. При $m = n$ фактически решается задача опознания исследуемого соединения по его спектру п.м.р., а при $m < n$ ведется поиск модельных соединений, содержащих те же структурные фрагменты, что и исследуемое вещество.

Весьма существенным является вопрос о том, какие параметры спектров п.м.р. должны быть представлены в машинном каталоге системы для того, чтобы был обеспечен эффективный поиск модельных соединений. По нашему мнению, при составлении такого каталога можно ограничиться указанием химических сдвигов и относительных интенсивностей сигналов без описания их мультиплетности (см. также ⁽¹³⁾), поскольку последнее во многих случаях представляет большие трудности.

Для проверки изложенных принципов построения машинной системы определения структурных фрагментов неизвестных соединений по их спектрам п.м.р. нами была создана модель этой системы на базе ЦВМ БЭСМ-6, обеспечивающая возможность обработки различных вариантов поиска по спектру уже упоминавшегося изобутилового эфира бензойной кислоты ⁽¹⁾. С этой целью из каталога спектров п.м.р. лаборатории Садтлера ⁽¹²⁾, охватывающего 12 000 соединений, с помощью одного из указателей были отобраны спектры, в которых присутствуют сигналы, попадающие хотя бы в один из следующих интервалов: 0,95—1,05, 2,03—2,13, 4,05—4,15, 7,43—7,53, и 8,01—8,11 м.д. (всего около 1300 спектров). Сведения об этих спектрах были записаны на магнитной ленте в виде пар чисел — значения химического сдвига сигнала, попавшего в один из указанных интервалов, и номера соединения по каталогу — с упорядочением этих пар по первому числу.

Составленные программы обеспечивают возможность работы в двух режимах. В первом режиме (режим опознания) поиск в машинном каталоге ведется по полному набору сигналов исследуемого соединения ($m = n$), причем приложение сигналов указывается с учетом возможных экспериментальных погрешностей ($\pm \Delta$). Если в машинном каталоге отсутствует спектр п.м.р., удовлетворяющий запросу, может быть осуществлен переход ко второму режиму работы. При этом m последовательно уменьшается на единицу и для каждого значения m находятся все воз-

возможные комбинации сигналов, а по ним — спектры п.м.р., содержащие сигналы в указываемых совокупностях спектральных интервалов. Выдаваемые машиной ответы содержат номера отобранных спектров. Последующий анализ отобранных спектров по критерию интенсивности сигналов и выявление структурных фрагментов общих для соединений, соответствующих этим спектрам, проводились на данном этапе исследования оператором.

При машинной обработке запросов по спектру упоминавшегося выше изобутилового эфира бензойной кислоты (1) мы получили следующие результаты. При $m = n = 5$ и $\Delta = \pm 0,02$ или $\pm 0,05$ м.д. в ответах содержалось только соединение (1). По пяти комбинациям сигналов при $m = 4$ и $\Delta = \pm 0,02$ м.д. также было отобрано только соединение (1), а при $\Delta = \pm 0,05$ м.д. еще четыре соединения общей формулы $ArCOOCH_2CH(CH_3)_2$, где Ar — замещенное бензольное кольцо. Комбинациям сигналов при $m = 3$ и $\Delta = \pm 0,02$ м.д. удовлетворяли соединения (1), четыре соединения типа $ArCOOCH_2CH(CH_3)_2$ и три соединения с иной структурой, причем последние исключались при использовании сведений об относительной интенсивности сигналов. При $m = 3$ и $\Delta = \pm 0,05$ м.д. в ответах были представлены соединения (1), шестнадцать соединений типа $ArCOOCH_2CH(CH_3)_2$ и шесть соединений иной структуры, спектры которых, как и в предыдущем случае, не удовлетворяли критерию относительной интенсивности сигналов.

Приведенные результаты показывают, что используя данные о химических сдвигах сигналов и их относительной интенсивности, можно из большого массива спектров п.м.р. отобрать спектры структурно-подобных соединений и определить структурные фрагменты, присутствующие в исследуемом соединении. Если бы, например, спектр соединения (1) не был представлен в каталоге (¹²), полученные в описанном выше эксперименте результаты позволили бы заключить, что оно является изобутиловым эфиром ароматической карбоновой кислоты. Характер ароматического остатка легко определить, проведя поиск по двум сигналам (7,48 и 8,06 м.д.), остающимся после вычитания из спектра п.м.р. сигналов идентифицированного фрагмента — $COOCH_2(CH_3)_2$. Заметим одновременно, что поиск по всем возможным комбинациям сигналов при $m = 2$ не целесообразен, так как число выдаваемых при этом спектров велико, а получаемая информация ограничена из-за малой характеристичности таких комбинаций.

Изложенный принцип можно использовать и для других видов спектроскопии я.м.р., например, для извлечения структурной информации из спектров углеродного магнитного резонанса.

Новосибирский институт органической химии
Вычислительный центр Сибирского отделения
Академии наук СССР
Новосибирск

Поступило
19 IV 1972

ЦИТИРОВАННАЯ ЛИТЕРАТУРА

- ¹ D. H. Anderson, G. L. Covert, *Anal. Chem.*, **39**, 1288 (1967). ² L. H. Cross, J. Haw, D. J. Shields, *Molecular Spectroscopy*, London, 1968, p. 189.
³ Ю. П. Дробышев, Р. С. Нигматуллин и др., *Вестн. АН СССР*, № 8, 75 (1970). ⁴ Л. Н. Дроздов-Тихомиров, *Оптика и спектроскопия*, **27**, 156 (1969).
⁵ Л. А. Грибов, М. Е. Эляшберг, *Изв. ТСХА*, в. 3, 182 (1969). ⁶ В. В. Разников, В. Л. Тальрозе, *Избр. тр. II Всесоюз. конф. по масс-спектрометрии*, 1971. ⁷ P. C. Jurs, V. R. Kowalski et al., *Anal. Chem.*, **42**, 1387 (1970). ⁸ Ю. П. Дробышев, Р. С. Нигматуллин и др., *Тез. докл. семинара Использование вычислительных машин в спектроскопии молекул*, Новосибирск, 1971, стр. 11. ⁹ S. Sasaki, H. Abe et al., *Anal. Chem.*, **40**, 2220 (1968). ¹⁰ Пат. ФРГ, 1943 828, 1970.
¹¹ А. В. Кессених, Л. Н. Дроздов-Тихомиров, А. А. Белорусский, *Изв. АН СССР, сер. хим.*, 1972, 95. ¹² *Sadtler NMR Spectra*, Philadelphia, 1971.
¹³ F. Erni, J. T. Clerc, *Chimia*, **24**, 388 (1970).