

Ю. П. ДРОБЫШЕВ, В. А. КОПТЮГ
(Новосибирск)

**КОМПЛЕКСНАЯ МАШИННАЯ СИСТЕМА
ДЛЯ РЕШЕНИЯ СТРУКТУРНЫХ ЗАДАЧ
МЕТОДАМИ МОЛЕКУЛЯРНОЙ СПЕКТРОСКОПИИ**

Методы, используемые химиками при решении структурных задач, непрерывно совершенствуются. Хотя химические превращения продолжают оставаться важным инструментом получения информации о строении молекулы исследуемого вещества, в настоящее время на первый план выдвинулись физические методы, причем наиболее широкое распространение получили различные методы спектроскопии молекул — колебательная и электронная спектроскопия, масс-спектрометрия, ядерный магнитный резонанс и др. При решении структурных задач нельзя отдать предпочтение какому-либо одному из упомянутых выше спектральных методов. Они дополняют друг друга, и максимальный эффект достигается при их комбинированном использовании. Нередко анализ набора спектров позволяет сделать однозначный вывод о строении молекулы исследуемого соединения без или с ограниченным привлечением значительно более трудоемких химических способов расшифровки структуры молекул. В результате комбинированного использования различных методов молекулярной спектроскопии усилия, затрачиваемые на установление строения достаточно сложной молекулы, составляют лишь небольшую долю (видимо, не более двух процентов [1]) от тех затрат труда, которые требовались для этого всего два-три десятилетия назад.

Широкое внедрение методов молекулярной спектроскопии в химию привело к накоплению больших массивов спектральных характеристик молекул. Например, число представленных в опубликованных коллекциях инфракрасных спектров поглощения превышает уже 120 тыс. [2], а число систематизированных электронных спектров поглощения и спектров ядерного магнитного резонанса измеряется десятками тысяч. Эффективное использование накопленной спектральной информации открывает новые возможности резкого увеличения производительности труда исследователей при решении структурных задач. Для реализации этих возможностей в Сибирском отделении АН СССР создается комплексная машинная система обработки информации по различным видам спектроскопии молекул на базе ЦВМ БЭСМ-6*.

* Работа ведется сотрудниками Новосибирского института органической химии (В. А. Коптюг, И. К. Коробейничева, В. С. Бочкарев, М. И. Горфинкель, В. А. Бархаш, С. А. Нехорошев) и Вычислительного центра Сибирского отделения АН СССР (Ю. П. Дробышев, Р. С. Нигматуллин, В. И. Лобанов, С. П. Соколов, Л. Е. Ильгина, В. И. Смирнов).

Несмотря на успехи вычислительной техники, полная замена человека электронной вычислительной машиной при решении широкого круга структурных задач вряд ли реальна, и поэтому мы отказались (по крайней мере, на первом этапе) от попытки полной автоматизации, так как это неизбежно привело бы к сужению круга решаемых задач (ср., например, [3]). Более гибкой представляется система «человек — машина», в которой возможно тесное взаимодействие человека и ЭВМ в процессе решения структурной задачи. Принципы построения и отдельные составные части такой системы описываются ниже.

Основным информационным фондом всей системы является созданная НИОХ и ГПНТБ СО АН СССР специализированная библиотека спектров, в которой достаточно полно представлены картотеки, каталоги, атласы и другая справочная литература по молекулярной спектроскопии. Соответствующим образом закодированная спектральная информация, а также некоторые дополнительные сведения о веществе (брутто-формула, молекулярный вес, в некоторых случаях сведения о структурных фрагментах молекулы и т. д.) записываются на магнитную ленту и образуют серию машинных каталогов (по инфракрасной спектроскопии, электронной спектроскопии и т. д.). Эти каталоги в совокупности со специальными программами, позволяющими осуществлять выборку из машинного каталога соединений, удовлетворяющих задаваемому набору признаков, представляют собой первую составную часть комплексной системы для молекулярной спектроскопии. Поясним роль этой части системы при решении структурных задач.

Когда исследователь приступает к установлению строения соединения*, полученного в результате химических превращений или выделенного из смесей искусственного или естественного происхождения, существует большая вероятность того, что это соединение окажется известным, поскольку счет охарактеризованных химических соединений ведется уже на миллионы. Но как узнать, известно или неизвестно вещество, попавшее в руки исследователя? Видимо, следует попытаться опознать его по каким-либо характерным признакам. В качестве таких признаков целесообразно выбрать (ср. [4]) инфракрасные спектры поглощения**. При наличии машинного каталога инфракрасных спектров задача просмотра десятков тысяч спектров становится вполне реальной. Подробно два варианта действующей машинной системы опознания химических соединений по их инфракрасным спектрам поглощения с общим объемом машинного каталога порядка 40 тыс. спектров описаны в [6, 7]. Здесь уместно остановиться лишь на одной особенности этой системы.

При проведении опознания соединений по их инфракрасным спектрам поглощения возникает трудность, связанная с тем обстоятельством, что условия записи спектров не стандартизованы. Поскольку спектры одного и того же соединения, записанные в разных условиях (фазовое состояние, растворитель, концентрация), могут весьма существенно отличаться друг от друга, нецелесообразно искать в машинном каталоге спектр, совершенно точно совпадающий с представленным для опознания. Положение несколько упрощается благодаря тому, что исследователь может заранее сказать, в каких областях спектра могут наблюдаться искажения при изменении условий регистрации, и поэтому на степень совпадения спектральных кривых в этих областях можно на-

* Существующие методы разделения веществ (например, различные виды хроматографии) в большинстве случаев обеспечивают возможность работы с индивидуальными соединениями.

** В принципе опознание соединения может проводиться и по другим видам спектров (см., например, [5]).

ложить при поиске менее жесткие требования, чем в остальных спектральных интервалах. В упомянутых выше вариантах системы опознания это достигается введением в поисковый запрос двух типов спектральных признаков — обязательных и желательных. Машина отбирает в машинном каталоге соединения, удовлетворяющие набору обязательных признаков, и затем упорядочивает их по степени совпадения желательных признаков. Время работы центрального процессора ЭВМ БЭСМ-6 составляет при этом всего 5—6 с в расчете на 10 тыс. просмотренных спектров. В машинном ответе на запрос, составленный с целью опознания соединения, указываются номера отобранных соединений по каталогам основного фонда. Совпадение спектра исследуемого вещества с одним из отобранных машиной спектров, как правило, приводит к опознанию вещества и, следовательно, к решению задачи. В противном случае вводится в действие вторая часть системы, предназначенная для расшифровки структуры исследуемого вещества. Для этого необходимо ответить на вопрос, из какого набора атомов построена молекула, т. е. определить ее брутто-формулу ($C_nH_lN_mO_k \dots$). Это может быть сделано разными путями, но наиболее привлекательным является метод масс-спектрометрии высокого разрешения, который позволяет «взвесить» молекулу с очень большой точностью. Одна из программ описываемой комплексной машинной системы позволяет, учитывая нецелочисленность атомных весов элементов, находить по точному значению молекулярного веса возможные брутто-формулы исследуемого соединения [8].

Располагая сведениями о наборе атомов, из которых построена молекула исследуемого соединения, еще нельзя приступать к построению ее возможной структуры, так как даже сравнительно ограниченному набору атомов соответствует огромное число возможных структурных формул, например, брутто-формула $C_{10}H_{22}O$ отражает около тысячи возможных структур. «Для веществ, содержащих более 20 атомов в скелете, решение задачи перебора всех структурных изомеров быстро становится бессмысленным не только с помощью современных ЭВМ, но и тех, появление которых ожидается в будущем» [9]. Сказанное делает очевидным необходимость предварительного получения информации об укрупненных структурных фрагментах молекулы, т. е. о «блоках атомов».

Богатым источником такой информации являются опять-таки молекулярные спектры. Используемые химиками корреляции «спектр — структура» имеют преимущественно эмпирический характер и включают в себя наиболее резко бросающиеся в глаза связи спектральных и структурных характеристик. Довольно очевидно, что привлечение вычислительной техники для выявления более широких корреляционных связей и для проведения анализа спектров с целью извлечения структурной информации таит в себе очень большие возможности.

Построение машинных систем извлечения структурной информации из спектров молекул, естественно, должно проводиться с учетом физической сущности каждого из спектральных методов. Рассмотрим это на примере электронной спектроскопии.

Поглощение молекулой излучения в ультрафиолетовой (200—400 нм) и видимой (400—760 нм) областях спектра соответствует возбуждению наименее прочно связанных электронов молекулы. Такими электронами являются электроны кратных связей (π -электроны) и электроны, входящие в состав неподеленных атомных пар n -электроны). Сочетание нескольких кратных связей или кратных связей и атомов с неподеленными электронными парами типа $A=B-C-D \equiv E \dots$, допускающее возможность перекрывания орбиталей π - и n -электронного спектра погло-

щения молекулы, носит название хромофорной системы молекулы. Соединения с одинаковыми хромофорами имеют очень близкие электронные спектры поглощения. Поэтому если мы найдем в имеющихся фондах электронный спектр, совпадающий со спектром исследуемого соединения, то тем самым получим указание на возможное строение большого блока молекулы. Подробно система определения хромофорного фрагмента молекулы, использующая машинный каталог электронных спектров поглощения, описана в [10].

Машинная система извлечения структурной информации из спектров протонного магнитного резонанса строится на ином принципе. Она основана на идентификации водородсодержащих фрагментов молекулы путем сопоставления различных комбинаций сигналов в протонном спектре исследуемого соединения со спектрами соединений, содержащихся в соответствующем машинном каталоге.

Проведение подобного машинного анализа набора спектров исследуемого соединения позволяет получить информацию о присутствии в молекуле этого соединения ряда структурных блоков. Располагая этой информацией и (если имеются) данными химического исследования, можно попытаться приступить к «сборке» молекулы. На этом этапе задача формулируется следующим образом — имеется набор структурных фрагментов, каждый из которых обладает определенным числом свободных валентностей, и набор одиночных атомов (находится по разности между брутто-формулой и суммарным составом структурных фрагментов), которые необходимо объединить так, чтобы все указанные части входили в состав одной связанной структуры и чтобы не осталось неиспользованных валентностей. Совокупность программ для решения этой задачи и представляет собой следующую часть комплексной машинной системы для молекулярной спектроскопии.

Решая указанную задачу машинным способом*, исследователь получает набор **всех** возможных структурных формул, отвечающих имеющейся сумме спектральной и химической информации. Если возможной оказывается только одна структурная формула, являющаяся плоским отображением молекулы, исследователь приступает далее к выяснению деталей ее пространственного строения. При выдаче машиной ограниченного набора структурных формул (не более 20—30) целесообразно подвергнуть их дополнительному анализу на соответствие имеющейся совокупности данных об исследуемом веществе.

На этом этапе должна вводиться в действие еще одна составная часть комплексной машинной системы, работа над которой только начинается. В основе ее будет лежать машинный каталог, содержащий описание структур соединений (см., например, [11]), спектры которых представлены в упоминавшихся выше спектральных машинных каталогах. Эта часть системы позволит уточнить соответствие предполагаемых структурных формул исследуемого соединения его молекулярным спектрам. Рассмотрев выданные машиной на предыдущем этапе структуры, исследователь отбрасывает те из них, которые не удовлетворяют химическим данным, а в остальных выделяет те особенности строения, которые отличают их друг от друга. Если для выявленных отличительных особенностей структур (например, наличие определенным образом замещенного бензольного кольца; наличие комбинации атомов с магнит-

* В настоящее время разработан алгоритм построения возможных структур из набора фрагментов, который реализован в соответствующей программе для ЭВМ «Наири» в НИОХ СО АН СССР (М. И. Горфинкель и С. А. Нехорошев), а также написана программа на языке «Альфа» для ЭВМ БЭСМ-6.

ными ядрами, образующей спиновую систему определенного типа, и т. д.) спектральные проявления хорошо известны (содержатся в опубликованных корреляционных таблицах), исследователь может сам провести дополнительную проверку соответствия оставшегося набора структур спектрам исследуемого соединения и еще более сузить круг возможных структур. Если, однако, спектральные проявления отличительных особенностей структур неизвестны (нельзя составить корреляционные таблицы, предусматривающие все возможные случаи), он обращается к машинной системе с заданием просмотреть машинный каталог структур и отобрать в нем соединения, содержащие заданный блок структуры. По номерам этих соединений машина отбирает далее в соответствующем спектральном машинном каталоге их спектры и выявляет характерные для отобранной группы соединений спектральные признаки. Таким образом, на рассматриваемом этапе машинная система используется для выявления ранее неизвестных спектро-структурных корреляций. На этом же этапе целесообразно использование различных вычислительных программ, позволяющих рассчитывать ожидаемый для данной структуры спектр или определенную часть спектра (см., например, [12—14]).

Если после проведенного уточнения соответствия возможных структур набору молекулярных спектров исследуемого соединения и его химическим свойствам остается несколько структур, а также в тех случаях, когда на этапе «сборки» было получено слишком большое число возможных структур, должен быть сделан вывод о недостаточности имеющейся структурной информации и предприняты шаги для получения дополнительной информации, например, химическим путем. После получения дополнительной информации цикл может быть повторен, начиная с этапа «сборки».

Важно подчеркнуть, что рассматриваемая комплексная система для молекулярной спектроскопии, с одной стороны, позволит использовать достаточно унифицированный подход к решению структурных задач, а с другой — будет обладать необходимой гибкостью, поскольку позволит исследователю вносить коррективы на всех этапах решения задачи.

ЛИТЕРАТУРА

1. Chemistry, a New Look. Benjamin Inc., p. 44, 1966.
2. K. Whetzel. Infrared Spectroscopy.— Chem. Eng. News, 1968, v. 44, № 32.
3. Verfahren und Einrichtung zur Structuranalyse organischer Verbindungen; Пат. ФРГ 1 943 828 (1968);— Chem. Eng. News, 1970, Bd. 46, № 53.
4. D. H. Anderson, G. L. Covert. Computer Search System for Retrieval of Infrared Data.— *Analyt. Chem.*, 1967, v. 39
5. В. В. Разников, В. Л. Тальрозе. Некоторые вопросы автоматической идентификации индивидуальных органических соединений по их масс-спектрам.— В сб. «Избранные труды I Всесоюзной конференции по масс-спектрометрии». Л., 1969.
6. Ю. П. Дробышев, Р. С. Нигматуллин, В. И. Лобанов, И. К. Коробейничева, В. С. Бочкарев, В. А. Коптюг. Использование ЭВМ для опознавания химических соединений по спектральным характеристикам, сообщение I.— *Вестник АН СССР*, 1970, 75.
7. Те же авторы. Система опознавания химических соединений по их ИК-спектрам на базе ЦВМ БЭСМ-6.— Тезисы докладов семинара «Использование вычислительных машин в спектроскопии молекул». Новосибирск, 1971.
8. М. И. Горфинкель, С. А. Нехорошев. Алгоритм и программа для подбора возможных эмпирических формул по заданному точному значению молекулярного веса соединения.— Тезисы докладов семинара «Использование вычислительных машин в спектроскопии молекул». Новосибирск, 1971.
9. В. В. Разников, В. Л. Тальрозе. Автоматическое построение полного набора структурных формул изомеров соединения определенного элементарного состава и молекулярного веса.— *ЖСХ*, 1970, т. 11.

10. В. С. Бочкарев, Ю. П. Дробышев, В. А. Коптюг, И. К. Коробейничева, В. И. Лобанов, Р. С. Нигматуллин. Машинная информационно-поисковая система для электронной спектроскопии.— *Автометрия*, 1972, № 4.
11. Г. Э. Владуц, Э. Д. Стоцкий, В. К. Финн. О некоторых системах записи структурных формул органической химии.— Доклады на конференции по обработке информации, машинному переводу и автоматическому чтению текста, вып. 4. М., ВИНТИ, 1961.
12. Л. А. Грибов. Состояние и перспективы автоматизации исследований многоатомных молекул по их колебательным спектрам. — Тезисы докладов семинара «Использование вычислительных машин в спектроскопии молекул». Новосибирск, 1971.
13. В. А. Деметьев. Комплекс программ для расчета колебательных спектров многоатомных молекул на ЭЦВМ «Минск-22». — Тезисы докладов семинара «Использование вычислительных машин в спектроскопии молекул». Новосибирск, 1971.
14. А. А. Фомичев. Алгол-60 — программы для анализа спектров ЯМР высокого разрешения.— Тезисы докладов семинара «Использование вычислительных машин в спектроскопии молекул». Новосибирск, 1971.

*Поступила в редакцию
11 января 1972 г.*