

АВТОМАТИЗАЦИЯ ХИМИЧЕСКИХ ИССЛЕДОВАНИЙ

УДК 539.194 : 681.3.06

Л. А. ГРИБОВ, В. А. ДЕМЕНТЬЕВ, А. И. ТИЩЕНКО,
М. Е. ЭЛЯШБЕРГ, Э. З. ЯКУПОВ

(Москва)

О ВОЗМОЖНОСТИ СТРУКТУРНО-ГРУППОВОГО АНАЛИЗА ВЕЩЕСТВА ПО ЕГО МОЛЕКУЛЯРНЫМ СПЕКТРАМ С ПРИМЕНЕНИЕМ МАШИННЫХ КОРРЕЛЯЦИОННЫХ ТАБЛИЦ

Опыт использования математической логики в спектрохимических исследованиях [1, 2] показал, что задача извлечения структурной информации из молекулярных спектров может быть разбита на два этапа. На первом этапе исследователь, пользуясь своим опытом и априорными сведениями о веществе, выделяет в спектре предположительно характеристические признаки и с помощью корреляционных таблиц составляет систему логических уравнений, которые связывают спектральные признаки со структурными группировками. На втором этапе электронно-вычислительная машина решает систему логических уравнений, выдавая все возможные сочетания структурных группировок, которыми может быть объяснен наблюдаемый спектр.

Неоднозначность решения указанной задачи связана с пересечением множеств характеристических признаков различных структурных элементов и с тем, что характеристические признаки спектра не чувствительны к изомерии, так что существует некоторое множество структур, соответствующих данному набору признаков. Множество структур можно сократить до единственной структуры, если использовать характеристические признаки, содержащиеся в нескольких молекулярных спектрах, полученных независимыми физическими методами. С математической точки зрения это соответствует поиску структуры исследуемого вещества в пересечении множеств решений, получаемых из отдельных молекулярных спектров.

В настоящей работе предпринята попытка объединить два метода установления структуры вещества: инфракрасную спектроскопию и масс-спектрометрию — и автоматизировать на основе применения ЭВМ процесс структурно-группового анализа вещества по его ИК- и масс-спектрам.

При этом автоматизации подвергнут и первый этап анализа, а именно составление логических уравнений с использованием машинных корреляционных таблиц. С этой целью для ЭВМ «Минск»-22 написана на языке ФОРТРАН программа, которая допускает как отдельную независимую обработку информации по ИК- и масс-спектрам вещества, так и совместную обработку этой информации. В целом данная программа входит в цикл программ по автоматизации спектрохимических исследований. Задача создания такого цикла программ была поставлена в [3].

Опишем схему работы программы структурно-группового анализа вещества по его ИК- и масс-спектрам в случае, когда в машину вводятся оба этих спектра (рис. 1).

ИК-спектр вещества вводится в машину в виде списка наблюдаемых полос поглощения, частоты которых выражены в см^{-1} . Особым образом отмечаются широкие полосы (задается начало и конец интервала интенсивного поглощения) и частоты «свободных» групп ОН, NH, СО, измеренные в разбавленных растворах (снабжаются специальным

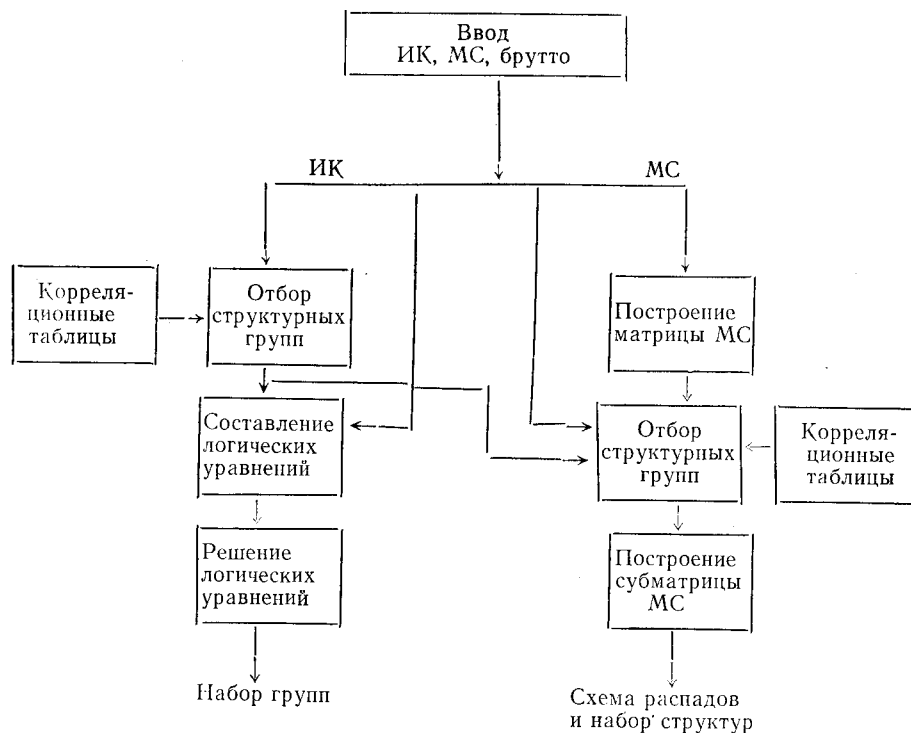


Рис. 1.

дополнительным признаком). В машину вводятся все частоты, наблюдаемые в ИК-спектре. Если некоторые области спектра по условиям эксперимента зарегистрированы с большими искажениями или не записаны вовсе, то такие не используемые при анализе «темные области» вводятся в машину в виде интервалов. Исследователь сообщает машине номера трех полос спектра, которые он полагает характеристическими. Именно для этих полос будут составляться и решаться логические уравнения.

Масс-спектр низкого разрешения вводится в машину в виде масс наблюдаемых пиков в порядке убывания масс. Первый пик должен быть молекулярным.

Обязательно вводится брутто-формула вещества в виде количества атомов С, Н, N, О.

Брутто-формула кодируется условным числом, в котором для указания количества атомов С и Н отводятся по два разряда, а для N и О — по одному.

В качестве примера приведем данные о соединении $\text{C}_9\text{H}_{10}\text{O}_2$ в том виде, как они были введены в машину.

Брутто-формула: 91002.

ИК-спектр: 3052, 2970, 2900, 1750, 1615, 1595, 1505, 1460, 1440, 1383, 1370, 1228, 1082, 1028, 1005, 967, 922, 902, 837, 824, 750, 738, 698.

Номера частот для составления логических уравнений: 1, 2, 4, 6, 8, 11, 12, 14, 21, 23.

Масс-спектр: 150, 108, 107, 91, 90, 79, 43.

Работа программы начинается с обработки информации, содержащейся в ИК-спектре. Пользуясь корреляционными таблицами, хранящимися на магнитной ленте, машина отыскивает такие структурные группы, которые не противоречат брутто-формуле вещества и у которых все характеристические полосы проявились в заданном ИК-спектре.

Корреляционные таблицы на магнитной ленте записаны в виде карточек. Образец карточки для расшифровки ИК-спектра (группа Ag—COOC) показан на рис. 2. В карточке кодируется брутто-формула группы, номер группы и интервалы частот, в которых лежат все характеристические полосы данной группы. Кроме того, особым образом кодируются широкие характеристические полосы и те узкие полосы групп ОН,

$\begin{array}{c} \text{O} \\ \parallel \\ \text{Ag}-\text{C}-\text{O}-\text{C}- \\ \\ \text{H} \end{array}$			80002
1100—1150	1250—1310	1480—1525	
1575—1625	1715—1730		

Рис. 2.

НН, СО, которые наблюдаются только в случае мономеров. Карточки составлены на основе известных корреляционных таблиц [4, 5].

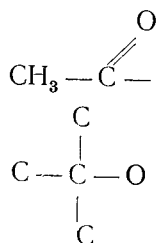
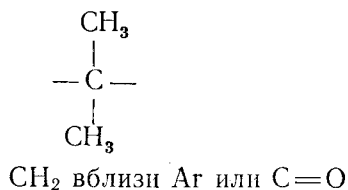
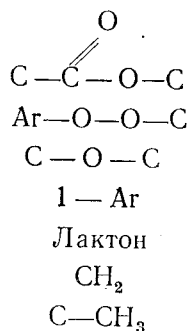
Программа вызывает в оперативную память изображения карточек и, просматривая заданный спектр, проверяет, все ли характеристические полосы данной группы наблюдаются в спектре. Если некоторые характеристические полосы попадают в «темные» места спектра, сравнение по таким полосам не производится.

Группы, получившие в таком смысле полное подтверждение в спектре и не противоречащие брутто-формуле вещества, заносятся в список возможных в данном веществе групп. Кроме того, номер каждой группы заносится в списки номеров групп при тех частотах спектра, которые исследователь отметил как характеристические. Допустим, частота спектра ω_i участвовала в подтверждении возможных для данного вещества групп A_3, A_5, A_{11} . Тогда в машине возникнет список, который в дальнейшем будет воспринят как импликация: $\omega_i \rightarrow A_3 + A_5 + A_{11}$. Совокупность таких импликаций и представляет собой автоматически построенную систему логических уравнений для определения структурно-группового состава вещества по его ИК-спектру.

Решение системы логических уравнений ищется по схеме, рассмотренной ранее [1, 2]. Отметим только, что в программе предусмотрено вмешательство человека в процесс составления и решения логических уравнений. Если исследователь считает это необходимым, он имеет возможность дополнить систему уравнений, составленную машиной, любыми логическими уравнениями относительно возможных групп, записанными в виде дизъюнктивных нормальных форм. Это позволяет использовать дополнительную априорную информацию о групповом составе вещества и сократить множество решений, получаемых из ИК-спектра и брутто-формулы вещества.

Решение системы логических уравнений машина выдает в виде совокупностей номеров тех групп, которые могут одновременно присутствовать в веществе. Как указывалось выше, характеристические полосы различных групп могут перекрываться в спектре, поэтому машина выдает несколько наборов групп. Некоторые из решений могут быть отброшены после проверки по брутто-формуле вещества.

Приведем промежуточный результат работы программы для рассмотренного выше примера. Просмотрев предложенный ИК-спектр, машина отобрала следующие структурные группировки, удовлетворяющие спектру и брутто-формуле вещества:



Решение логических уравнений мы не приводим, так как оно не дает единственного набора структурных группировок, из которых можно было бы построить искомую структуру соединения $\text{C}_9\text{H}_{10}\text{O}_2$. Таким образом, здесь мы сталкиваемся с ситуацией, когда из данных ИК-спектроскопии можно автоматически извлечь некоторую структурную информацию, однако эта информация еще недостаточна для однозначного построения структуры. Далее мы покажем, что дает в этом случае привлечение данных, содержащихся в масс-спектре вещества.

Заключив обработку ИК-спектра, программа приступает к работе с масс-спектром. В начале составляется треугольная матрица, которую можно назвать матрицей масс-спектра данного вещества. Эта матрица отражает все возможные пути фрагментаций молекулы с заданной брутто-формулой. Диагональные элементы матрицы масс-спектра содержат брутто-формулы, которые могут соответствовать наблюдаемым в спектре массам, а недиагональные элементы — брутто-формулы разностей масс всех наблюдаемых пиков.

Матрица масс-спектра строится следующим образом. Для каждого пика масс-спектра, кроме молекулярного (его брутто-формула известна), программа генерирует все возможные брутто-формулы, а также брутто-формулы для разностей каждого пика со всеми предыдущими. Из всех брутто-формул отбираются и заносятся в матрицу только те формулы, которые удовлетворяют условиям:

$$1 \leq I_n \leq 2I_c + I_n + 2 \quad (\text{для пиков [6]});$$

$$0 \leq I_n \leq 2I_c + I_n + 2 \quad (\text{для разностей}).$$

Так же отбрасываются брутто-формулы разностей, которые совпадают со списком запрещенных брутто-формул осколков (CH_2 , N и т. д.).

Нулевой недиагональный элемент говорит о том, что нельзя из соответствующего тяжелого иона получить соответствующий легкий ион. Если какая-то строка недиагональных элементов целиком равна нулю, то масс-спектр соответствует смеси веществ. Матрица, построенная машиной для масс-спектра соединения $\text{C}_9\text{H}_{10}\text{O}_2$, показана в таблице.

При вышеописанном способе составления матрица отражает все возможные процессы, происходящие с молекулой под электронным ударом. Диагональные элементы матрицы — это ионы, образующиеся как в результате простой диссоциации молекулы на ион и радикал, так и вследствие различных перегруппировочных процессов.

91002																					
20201	70801																				
30600	0	60402																			
20301	100	0	70701																		
30700	0	100	0	60302																	
20302	101	0	1	0	70700																
30701	0	101	0	1	0	60301															
20402	201	0	101	0	100	0	70600														
30801	0	201	0	101	0	100	0	60201													
50000	0	0	0	0	0	0	0	41002													
30302	10101	0	10001	0	0	0	0	0	60700												
40701	20500	10101	20400	10001	0	0	0	0	0	50301											
60302	40101	0	40001	0	0	0	0	0	10302	0	30700										
70701	50500	40101	50400	40001	0	0	0	0	0	0	0	20301									

В масс-спектре низкого разрешения отсутствует непосредственная информация об элементном составе ионов. Поэтому каждому значению массового числа диагональных элементов матрицы может соответствовать несколько ионов с различными брутто-формулами, не противоречащими брутто-формуле исходного соединения. В примере, приведенном в таблице,— это ионы с m/e 108 (70801 и 60402), 107 (70701 и 60302), 91 (70700 и 60301), 90 (70600, 60201 и 41002) и т. д. Это обстоятельство приводит к тому, что наряду с истинными путями фрагментации молекулярного иона матрица содержит также и ложные пути.

Сделать правильный выбор в общем случае затруднительно без привлечения множества структурных фрагментов, полученных независимым способом (из анализа ИК-спектра), однако при анализе путей фрагментации молекулярного иона возможно использование вероятностных характеристик для значений брутто-формул разностей массовых чисел (вертикальные элементы матрицы). Все эти оценки производятся при анализе матрицы человеком. В результате матрица масс-спектра может быть значительно упрощена и из нее выделены наиболее вероятные пути фрагментации молекулярного иона, которые приведут к наиболее вероятному набору масс-спектрометрических фрагментов.

В нашем примере образование иона с m/e 90 возможно несколькими путями.

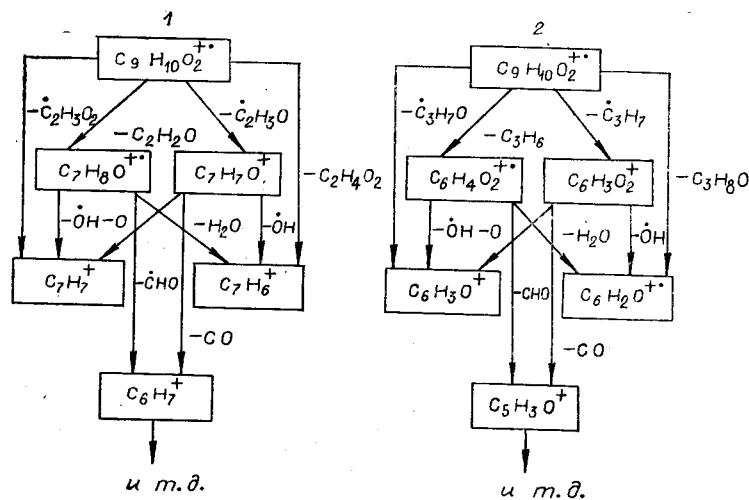
1. Отщеплением от молекулярного иона группы атомов $C_2H_4O_2$ (20402) с образованием иона $C_7H_6^{+}$ (70600). [Этот же ион может образоваться при потере ионом $C_7H_8O^{+}$ (70801) молекулы воды (201) или ионом $C_7H_7O^{+}$ (70701) группы атомов OH (101).]

2. Отщеплением от молекулярного иона группы атомов C_3H_8O (30801) с образованием иона $C_6H_2O^{+}$ (60201). [Этот же ион может образоваться при потере ионом $C_6H_4O_2^{+}$ (60402) молекулы воды (201) или ионом $C_6H_3O_2^{+}$ (60302) группы атомов OH (101).]

3. Отщеплением от молекулярного иона группы атомов C_5 (50000) с образованием иона $C_4H_{10}O_2^{+}$ (41002).

Очевидно, что наиболее вероятен первый путь — отщепление от молекулярного иона группы атомов $C_2H_4O_2$; второй путь (отщепление C_3H_8O) менее вероятен, но не может быть полностью исключен, тогда как третий путь (отщепление C_5) практически невероятен.

В результате такого рассмотрения мы имеем право отбросить брутто-формулу иона с m/e 90 $C_4H_{10}O_2$ и ее дальнейшие превращения исключить из матрицы. Формула C_6H_2O не может быть исключена и нам приходится рассматривать 2 независимых пути фрагментации молекулярного иона, приведенные ниже:



Как видно из схемы, каждый путь приводит к своему набору фрагментов. При построении структурной формулы исходного соединения на основании этих наборов фрагментов мы получим минимум 2 результата. Для правильного выбора мы должны найти подтверждения строения фрагментов на основании сравнения со строением отдельных групп атомов, полученным при анализе спектров ИК. Эта операция делается машиной следующим образом.

В машине на магнитной ленте хранятся корреляционные таблицы, составленные на основе [7] в виде карточек. Образец масс-спектроскопической карточки для подтверждения масс-спектрометрических фрагментов представлен на рис. 3. Для каждого массового числа кодируются все наиболее часто встречающиеся брутто-формулы. Для каждой брутто-формулы указываются вероятные структуры, из которых данный осколок может быть отщеплен при электронном ударе. Для каждой структуры указываются номера входящих в нее структурных группировок, которые дают характерные полосы поглощения в ИК-спектрах.

Программа вызывает в оперативную память изображение тех карточек, брутто-формулы которых имеются в диагональных элементах матрицы масс-спектра. Отбираются те брутто-формулы и структуры, которые полностью подтверждаются номерами групп, отобранных при анализе ИК-спектра. Для каждой подтвержденной брутто-формулы пика строится субматрица данного пика, представляющая собой полную схему распада, приводящую к образованию пика именно с данной брутто-формулой и структурой. Субматрица выдается на печать. Такая субматрица может быть изучена и на ее основе исследователь уже легко может сделать заключение о возможных структурных соотношениях между данным и более тяжелыми продуктами распада. В приведенном примере полностью подтвердилось строение следующих фрагментов:

	Строение	m/e	Брутто-формула
Альтернативное строение	$\text{CH}_2 - \text{C} - [\text{X}]$	107	$\text{C}_7 \text{H}_7 \text{O} (70701)$
	$\text{CH}_2 - [\text{X}]$	91	$\text{C}_7 \text{H}_7 (70700)$
	$[\text{X}]$ CH_3	91	$\text{C}_7 \text{H}_7 (70700)$
	$[\text{X}_1]$ $\text{C} - [\text{X}_2]$ $[\text{X}_3]$	91	$\text{C}_7 \text{H}_7 (70700)$
	$\text{CH}_3 - \text{C} - [\text{X}]$ \parallel O	43	$\text{C}_2 \text{H}_3 \text{O} (20301)$
	$\text{C}_3 \text{H}_7 - \text{O} - \overset{\text{O}}{\parallel} \text{C} - [\text{X}]$	43	$\text{C}_3 \text{H}_7 (30700)$
	$\text{C}_3 \text{H}_7 - \text{O} - [\text{X}]$	43	$\text{C}_3 \text{H}_7 (30700)$
	$\text{C}_3 \text{H}_7 - (\text{C}_n \text{H}_{2n+1})$	43	$\text{C}_3 \text{H}_7 (30700)$

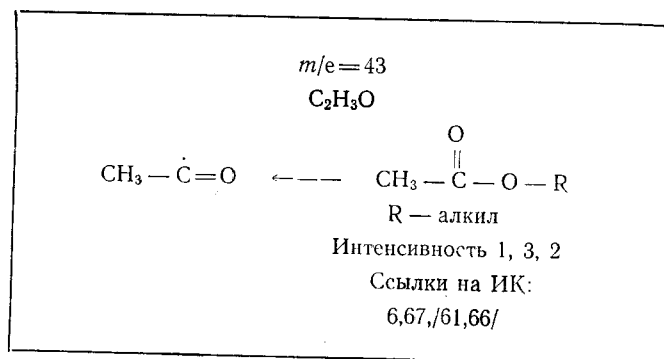
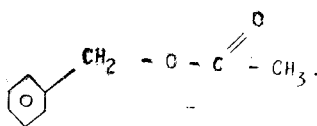


Рис. 3.

По пути 2 ни один из приведенных на схеме ионов не был подтвержден. Таким образом, первый путь фрагментации молекулярного иона является истинным и набор фрагментов, приведенный выше, следует использовать для конструирования структурной формулы. Поскольку масс-спектроскопические фрагменты сравнительно крупнее, чем группировки, ответственные за появление характеристических полос в ИК-спектре, легче построить структурную формулу соединения, пользуясь масс-спектрометрическими фрагментами. Видно, что брутто-формуле нашего соединения удовлетворяет только объединение фрагментов C_7H_7O и C_2H_3O , что дает искомую структуру



Таким образом, лишь совместная обработка ИК-спектра и масс-спектра в данном примере привела к однозначному установлению структуры вещества.

ЛИТЕРАТУРА

1. L. A. Gribov, M. E. Elyashberg, L. A. Moskovkina. Solution of Spectral Problems by Methods of Symbolic Logic.— J. Mol. Struct., 1971, v. 9, p. 357.
2. М. Е. Эляшберг, Л. А. Московкина, Л. А. Грибов. О применении математической логики для установления структуры сложных молекул по колебательным спектрам.— Журнал прикладной спектроскопии, 1971, т. 15, с. 843.
3. Л. А. Грибов, В. А. Деметьев, М. Е. Эляшберг. О возможности автоматизации исследования строения и свойств молекулы по ее молекулярным спектрам.— Автометрия, 1972, № 4.
4. К. Наканиси. Инфракрасные спектры и строение органических соединений. М., Изд-во иностран. лит., 1961.
5. Беллами. Инфракрасные спектры сложных молекул. М., Изд-во иностран. лит., 1963.
6. Дж. Бейнон. Масс-спектрометрия и ее применение в органической химии. М., «Мир», 1964.
7. F. W. McLafferty. Mass Spectral Correlations. Washington, 1963.

Поступила в редакцию 28 ноября 1973 г.