

Л. А. ГРИБОВ, В. А. ДЕМЕНТЬЕВ, М. Е. ЭЛЯШБЕРГ  
(Москва)

### О ВОЗМОЖНОСТИ АВТОМАТИЗАЦИИ ИССЛЕДОВАНИЯ СТРОЕНИЯ И СВОЙСТВ МОЛЕКУЛЫ ПО ЕЕ МОЛЕКУЛЯРНЫМ СПЕКТРАМ

**§ 1. Общая постановка задачи.** Последние два десятилетия характеризуются массовым внедрением физических методов исследования строения и свойств многоатомных молекул. Без преувеличения можно сказать, что, по-видимому, примерно половина всех решаемых вопросов, возникающих при детальном изучении строения многоатомной молекулы, решается с участием в той или иной степени таких физических методов исследований, как спектроскопия в ультрафиолетовой области, спектроскопия в инфракрасной области, спектроскопия комбинационного рассеяния, спектроскопия ядерного магнитного резонанса, электронного парамагнитного резонанса, масс-спектроскопия, и некоторых других.

Наилучшие результаты получаются при комплексном подходе, когда для решения одной и той же задачи одновременно применяется целый ряд физических и химических методов исследования. Такой подход сулит наибольший успех, но связан с обработкой большого количества экспериментальных данных.

Молекулярная спектроскопия, как и подавляющее большинство физических методов исследования, является косвенным методом. Само по себе наблюдение спектра ничего еще не может сказать о природе молекулы, если не установлена физическая связь между наблюдаемым спектром и ее строением.

Значительная часть работы, связанной с производством большого числа вычислений, может быть переложена на ЭВМ. В молекулярной спектроскопии подобного типа работа ведется уже в течение более 10 лет и разработан определенный набор алгоритмов и программ, осуществляющий большое количество различных операций. Постепенно эта работа систематизируется и создаются уже не отдельные, не связанные между собой программы, а ставится вопрос о создании комплексных программ, выполняющих вполне определенные сложные вычислительные операции.

Многие из алгоритмов решения тех или иных задач, встречающихся в теории молекулярных спектров многоатомных молекул, которые с успехом применялись в домашний период, оказываются неудобными при применении ЭВМ в самых разнообразных областях физики. Поэтому сейчас идет работа по преобразованию математической стороны физических теорий в вид, удобный для использования ЭВМ.

В создании алгоритмов таких программ значительный опыт накоплен именно в колебательной спектроскопии. Последние достижения в этой области дают основание надеяться, что удастся построить полные действующие комплексы, решающие достаточно сложные задачи.

До последнего времени развитие соответствующих алгоритмов шло в основном по линии автоматизации вычислительных операций, встречающихся в обычной теории колебательных спектров многоатомных молекул. Речь идет о вычислении частот и форм колебаний, смещений атомов из положения равновесия, определении значений силовых постоянных многоатомных молекул по их спектрам, вычислении интенсивности полос поглощения и т. д. Соответствующий математический аппарат был уже довольно давно разработан, а затем приведен к виду, пригодному для использования на ЭВМ.

В последние годы автоматизации подверглась еще одна важнейшая задача, а именно определение структуры молекулы по имеющемуся спектру и набору характеристических частот и обратная задача установления характеристических частот по формально заданным спектрам и известной структуре молекул. Такая задача отличается от привычного круга задач тем, что мы в этом случае имеем дело с так называемым «черным ящиком», т. е. мы можем знать, что какая-то структурная группа порождает вполне определенный набор полос, но мы можем не задумываться, почему это происходит и какая существует связь между этими полосами и особенностями данной структурной группы. Решение задач такого типа требует применения специального аппарата математики, в частности булевой алгебры.

Объединение всех этих подходов, по-видимому, может дать возможность построить схему, которая совершала бы следующие операции: отождествление молекул по спектру, построение грубой структурной формулы молекулы, уточнение этой структуры, сравнение получающихся при этом численных констант с типичными константами, характеризующими, например, индукционное влияние или сопряжение в молекуле, и т. д.

Поставим вопрос, из каких частей, из каких блоков должна состоять подобная идеализированная система и какие части ее уже отработаны. Перечислим основные этапы решения соответствующей задачи.

1. Предположим, что зарегистрирован ИК-спектр вещества. Будем исходить из предположения, что этот спектр относится к индивидуальному веществу, т. е. смесь, получающаяся, например, в результате реакции, уже подвергалась разделению с помощью, скажем, хроматографического метода. На первом этапе необходимо произвести идентификацию этого вещества. Может оказаться, что это вещество относится уже к классу изученных. В таком случае его распознавание может быть осуществлено путем формального сопоставления спектра вещества с атласными спектрами. Общее число спектров, накопленных на сегодняшний день в различных атласах, составляет около 100 тыс. Само собой разумеется, что такое сопоставление с помощью ручного перебора или некоторых отдельных примитивных устройств сортировки в ограниченный промежуток времени произведено быть не может. Необходимо, следовательно, применять для этих целей ЭВМ, закодировав специальным образом имеющийся атлас спектров и спектр исследуемого вещества, и производить соответствующее сопоставление с использованием специального алгоритма. Такого типа система создана, в частности, в Сибирском отделении Академии наук СССР [1].

2. В результате такого сопоставления могут возникнуть две ситуации: либо исследуемое вещество окажется тождественным одному из ранее

изученных (дальнейшие исследования можно прекратить), либо это вещество относится к классу не изученных ранее соединений (сопоставление с имеющимися спектрами не приведет к решению задачи). Однако может оказаться, что само это новое соединение состоит из отдельных относительно замкнутых группировок, которые уже ранее были изучены. Например, в нем может содержаться бензольное кольцо,  $\text{CH}_3$ -,  $\text{CH}_2$ -группы, винильные, кетонные группы и т. д. В таком случае спектр соответствующего вещества будет содержать ряд полос, относящихся к категории характеристических и встречающихся в других изученных уже ранее группировках. Если число таких характеристических полос будет достаточно большим, то тогда на основании корреляционных таблиц характеристических частот можно попытаться проделать соответствующий логический анализ и построить предполагаемую структуру молекулы. Общий подход к решению этой задачи описан в § 3 данной работы.

Необходимо заметить, что точность такого логического анализа в значительной степени зависит от того, насколько много в молекулярном спектре характеристических полос. Может оказаться, что почти все колебания данных молекул будут нехарактеристическими и тогда пользование корреляционными таблицами станет невозможным. Такой случай, однако, представляется все же довольно редким, по крайней мере в классе органических соединений.

Итак, предположим, что структура молекулы на основании корреляционной таблицы и соответствующего логического анализа установлена. Заметим, что такой анализ требует для своего проведения не только знания спектра и корреляционных таблиц, но и знания брутто-формулы и ряда дополнительных химических соображений, причем, как правило, устанавливается не одна, а несколько возможных структур, логически не противоречащих набору характеристических частот данной молекулы.

3. Но любая молекула наряду с характеристическими частотами, возникновение которых связано с существованием относительно замкнутых группировок в молекуле, будет иметь и целый ряд частот, связанных с относительными колебаниями этих группировок и в этом смысле характеристическими не являющихся.

При анализе на втором этапе такие частоты просто не используются, и соответствующая информация теряется. Однако можно в принципе найти все частоты колебаний, если произвести расчет колебаний, основываясь на предполагаемых структурах. Такого типа работа составляет третий этап решения задачи. На этом этапе предполагаемая структура молекулы рассчитывается средствами теории колебательных спектров молекул и расчетный спектр сопоставляется с экспериментальным. Таким образом, возникает обратная связь, которая представляется нам совершенно необходимой для полной идентификации вещества. Можно надеяться, что такой анализ во многих случаях позволит уточнить структуры молекул.

4. Приведенные расчеты и уточнения соответствующих структур дают возможность получить набор численных значений параметров, характеризующих свойства отдельных структурных элементов молекулы. Если говорить о колебательной спектроскопии, то к этому набору относятся силовые постоянные многоатомных молекул, дипольные моменты связей и производные от дипольных моментов связей по растяжениям связей и изменениям валентных углов (электрооптические параметры). Эта чрезвычайно ценная информация в дальнейшем может быть использована для сопоставления свойств данной молекулы с

типичными свойствами. Так, например, значения силовых постоянных связей существенно различаются в зависимости от того, с сопряженным или несопряженным типом соединений мы имели дело.

В таблице представлены некоторые значения силовых постоянных связей углерод—углерод для различных типов соединений [2]. Как мы видим, имеются заметные различия между насыщенным и ароматическим типами.

Силовые постоянные связей в соединениях различного типа

Соединение	Группа	$K_i$	$K_j$	$K_{ij}$	$K_{ik}$	Тип поля
Этан	C—C	7,0	—	—	—	Несопряженный
Пропан	C—C—C	7,0	7,0	0,1	—	
Пропилен	C=C—C	14,2	7,0	0,1	—	Ароматический
Бензол	C—C—C—C	11,4	11,4	1,0	—0,9	
Бутадион	C=C—C=C	12,0	10,4	0,4	—1,5	Мезомерный
Акрилонитрил	C=C—C≡	13,8	9,5	0,5	—0,1	Индуктивный

Такое сопоставление полученных значений силовых постоянных с типичными даст возможность не только установить структурную формулу молекулы, но и отнести ее по признаку свойств химических связей к некоторым определенным типам.

В значительной степени необходимые алгоритмы и программы специального назначения, решающие весь этот круг задач, уже известны. На рисунке представлена общая блок-схема программы, выполняющей все этапы исследования, о которых было написано выше. Такие программы мы дальше будем называть исследовательскими программами. Ныне действующие части программы обведены сплошными рамками. Те части программы, которые предстоит разработать, обведены штриховыми рамками.

Мы еще не имеем хорошо действующего алгоритма, который с учетом корреляционных таблиц, химической информации о возможных валентных состояниях атомов, информации о пространственном расположении отдельных частей молекулы, брутто-формулы и ряда других соображений автоматически строил бы структуру молекулы.

Если структура молекулы выстроена, то одной из очередных задач является установление типа симметрии молекулы и вычисление матрицы коэффициентов симметрии для последующего численного решения уравнений. После выполнения соответствующего вычисления необходимо провести сопоставление вычисленного и экспериментального спектров молекулы. Это сопоставление представляет собой сложную логическую задачу, причем еще не вполне ясны физические принципы проведения такого сопоставления. Соответствующий алгоритм также еще не разработан.

Дальнейшая задача заключается в уточнении структуры молекулы путем сближения расчетного и экспериментального спектров с помощью вариации параметров, в частности силовых постоянных. Эта задача уже разработана, и можно с уверенностью утверждать, что такая задача принципиальных трудностей не вызывает.

Таким образом, наряду с большим комплексом уже работающих программ, мы имеем еще и несколько этапов, в отношении которых существуют лишь только предварительные разработки.

К общему комплексу программ следует еще добавить программы, осуществляющие редукцию спектров к идеальному прибору. Спектры, вводимые в программу для последующей обработки, должны быть очи-

**Блок-схема программы  
для автоматического установления структуры и свойств молекулы по ее ИК-спектру**



шены от инструментальных погрешностей. На самом деле, любой прибор вносит соответствующие погрешности и необходимо на первом этапе решения иметь комплекс программ, осуществляющий перевод спектра в идеальный, а также перевод его в стандартную шкалу, например в молярную концентрацию. Соответствующие алгоритмы и программы разработаны и существуют в целом ряде вариантов.

**§ 2. Комплекс вычислительных программ.** Комплекс вычислительных программ существует в двух вариантах: для машин БЭСМ-3, БЭСМ-4, М-20 [3] и «Минск-22» [4—6]. Наиболее полный комплекс программ записан в кодах машины «Минск-22». Имеются следующие программы, связанные между собой и последовательно производящие целый ряд операций.

Первая программа [4] составляет колебательное уравнение. Для работы этой программы в машину вводится информация о геометрии молекулы в форме либо задания единичных векторов связей и их длин, либо инструкции по вычислению векторов связей на основе заданных валентных углов. Затем задаются значения масс атомов, тип и порядок следования естественных колебательных координат.

По этим данным производится формирование матрицы перехода от декартовых координат смещения атомов к естественным координатам. Эта матрица используется для автоматического формирования матрицы кинематических коэффициентов. Затем вводится матрица силовых постоянных в естественных координатах и матрица коэффициентов симметрии. Матрицы кинематических коэффициентов и силовых постоянных приводятся по симметрии и передаются далее для решения вековых уравнений. Максимальный порядок матриц равен 81, что соответствует 27 атомам в молекуле.

Второй программой, присоединенной к первой и решающей задачу вслед за решением задач по первой программе, является программа решения колебательных уравнений [4]. Производится последовательная диагонализация сначала матрицы кинематических коэффициентов, а затем матрицы силовых постоянных. На каждом этапе диагонализуется симметричная вещественная матрица и решение производится по методу вращения Якоби. Соответствующий алгоритм чрезвычайно удобен для ЭВМ и, что очень важно, нечувствителен к вырождениям.

В результате работы этой программы вычисляются частоты и формы нормальных колебаний молекулы в координатах симметрии, в естественных и декартовых координатах. Очень важно, что в программе предусмотрен автоматический поиск и исключение зависимых координат. Этот поиск основан на том свойстве матрицы кинематических коэффициентов, что между ее строками и столбцами имеется та же линейная зависимость, что и между координатами. При наличии зависимых координат матрицы кинематических коэффициентов будут особенные и иметь столько нулевых собственных чисел, сколько имеется зависимых координат. Элементы собственных векторов, относящиеся к нулевым собственным числам, дают коэффициенты в линейных комбинациях между зависимыми координатами. Матрица приведения кинематических коэффициентов к единичному виду превращается в прямоугольную матрицу отбрасыванием столбцов, отвечающих нулевым собственным числам. С помощью такой прямоугольной матрицы производится преобразование в новые координаты матрицы силовых постоянных. Фактически это означает переход к новым координатам, которые являются уже независимыми между собой и представляют собой линейные комбинации исходных естественных координат.

Практика показывает, что машина уверенно находит до 20 зависимых координат и отбрасывает их. Максимальный порядок решаемых уравнений составляет в нашем комплексе 42. Следовательно, молекулы в 27 атомах могут быть рассчитаны полностью только в том случае, если имеются хотя бы 1—2 элемента симметрии.

Следующая программа комплекса — программа решения обратных спектральных задач, т. е. уточнения значения матрицы силовых постоянных, если известны некоторые нулевые приближения матрицы и набор экспериментальных частот [5, 6].

Задача решается по методу наименьших квадратов, причем предусмотрено параметрическое задание в виде линейных комбинаций значений силовых постоянных многоатомных молекул. Это дает возможность учесть переносимость силовых постоянных из одной молекулы в другую и возможную функциональную связь между отдельными силовыми постоянными. Так, например, можно учесть тот известный факт, что с увеличением недиагональной силовой постоянной, как правило, симбатно растет и диагональная силовая постоянная.

Программа рассчитана таким образом, что она может производить уточнение силовых постоянных или параметров, от которых они зависят, одновременно для нескольких молекул разной симметрии, причем для каждой молекулы задача решается в своих координатах симметрии, в то время как исходная информация задается в естественных колебательных координатах. Поэтому запись исходной информации отличается большой простотой.

Программа решения обратной спектральной задачи отличается той особенностью, что дает возможность на каждом шаге итерации все время поддерживать первоначальное отнесение частот, что позволяет, в свою очередь, уверенно идти к минимуму. Одновременно можно варьировать до 20 параметров и сравнивать с экспериментом до 50 рассчитываемых частот.

На этом этапе решения задачи приходится сталкиваться с необходимостью сопоставления вычисленных и экспериментальных частот.

В зависимости от порядка следования вычисленных и экспериментальных частот будут получаться, вообще говоря, разные результаты. Это сопоставление представляет собой довольно трудную логическую задачу, и мы указывали, что пока оно выполняется исследователем на основании определенных логических соображений и интуитивного подхода и требует еще разработки специальных алгоритмов для автоматизации этого процесса.

После решений прямой и обратной спектральной задачи можно перейти к расчету интенсивностей полос поглощения в инфракрасных спектрах многоатомных молекул. Эта программа использует метод расчета, описанный в работе [7], и требует дополнительного введения электрооптических параметров.

Программа для расчета спектров комбинационного рассеяния нами не делалась на том основании, что существующая теория этого метода требует введения столь большого числа параметров, что вряд ли может быть широко использована.

**§ 3. Основы решения задачи о построении структуры молекул средствами математической логики.** Как уже указывалось, на первом этапе решения задачи необходимо провести логический анализ и установить структуру молекулы по набору характеристических частот. Такая задача решается средствами математической логики и вкратце сводится к следующему.

Молекула как совокупность структурных элементов и ее ИК-спектр представляют собой дискретные по природе объекты. Известно, что математическим аппаратом, применяемым для построения структурных в рассмотрении множества структурных элементов является логический спектр  $\omega_i$ . Элементам этих множеств как логическим переменным придается смысл элементарных высказываний:  $\omega_i$  ( $\bar{\omega}_i$ ) — спектр содержит (не содержит) частоту  $\omega_i$ ;  $A_i$  ( $\bar{A}_i$ ) — молекула содержит (не содержит) структурный элемент  $A_i$ .

В каждом конкретном случае подготовку данных начинают с того, что в экспериментальном спектре выделяются частоты, которые, по предположению, считаются характеристическими. Каждой из этих частот (от 10 до 20) ставится в соответствие набор конкурирующих функциональных групп (структурных элементов). Связь каждой частоты с группами-конкурентами представляется в виде импликаций — булевых функций, отражающих зависимость вида: если в спектре наблюдается частота  $\omega_i$ , то молекула содержит хотя бы одну функциональную группу из сопоставляемого набора. Свойство некоторых частот быть характеристическими для группы  $A_i$  также представляется соответствующей импликацией: если молекула содержит группу  $A_i$ , то в спектре наблюдается конъюнкция частот, характерных для этой группы.

Информация об экспериментальном спектре и спектр-структурные корреляции используются для составления логического уравнения специального вида. Решение уравнения сводится к вычислению такой булевой функции, которая содержит перечисление всех равновероятных комплексов структурных элементов, присутствие которых возможно в молекуле исследуемого соединения. Программа, разработанная для машины типа М-20, находит решение уравнения и автоматически выделяет те наборы фрагментов, которые удовлетворяют также и брутто-формуле. В соответствии с разрядностью машины оказывается возможным решение задач, для которых суммарное число рассматриваемых групп и частот не превышает 45. В типичном случае (15 групп и 15 частот) машинное время не превышает нескольких минут.

Далее необходимо из каждого допустимого набора фрагментов построить все молекулярные структуры, удовлетворяющие брутто-формуле и требованиям теории строения. Эта задача в принципе может быть решена в два этапа. На первом этапе путем решения системы целочисленных уравнений для данного комплекса структурных элементов находится в соответствии с брутто-формулой число функциональных групп каждого типа. Затем, используя алгоритм, основанный на методах математической логики и теории графов, «синтезируют» из фрагментов все возможные структуры.

Необходимо обратить внимание и на другие возможности, которые открывает применение математической логики в молекулярной спектроскопии. При использовании дискретной структурной модели системы молекула — спектр взаимосвязь строения молекул ряда соединений с их колебательными спектрами может быть представлена в виде системы логических уравнений. Решение этой системы позволяет установить все допустимые гипотезы о наборах частот, характерных для каждой из функциональных групп, а также отнести каждую из частот спектра



к колебаниям определенного структурного элемента или комбинации структурных элементов.

Программа решений системы логических уравнений, разработанная для машины типа М-20, дает возможность устанавливать спектр-структурные корреляции в случаях, когда число обсуждаемых групп и частот в отдельности не превышает 36.

Более подробно проблема применения математической логики в молекулярной спектроскопии рассмотрена в работах [8, 9].

Отметим, что в связи с общей задачей построения исследовательских программ приобретает чрезвычайно важное значение вопрос о построении высококачественных корреляционных таблиц, содержащих характеристические спектральные признаки. Имеющиеся корреляционные таблицы не могут быть признаны удовлетворительными. Для целей установления характеристических частот, по-видимому, могут быть применены как средства, описанные выше, основанные на методах математической логики, так и программы, основанные на методах распознавания образов. Такой опыт в настоящее время уже имеется.

Мы склонны думать, что предпочтение следует отдать методам, основанным на применении булевой алгебры, где соответствующая задача строго формулируется в терминах логических уравнений. В этом случае взаимосвязь между структурой вещества и спектром видна более наглядно, и результат может быть легко интерпретирован.

Успех разработки таких программ сулит значительные перспективы. Например, исследование структуры молекул по их рентгенограммам с использованием ЭВМ и специальных программ привело к значительному ускорению соответствующих расчетов. В результате был сделан целый ряд открытий, имеющих фундаментальное значение.

Есть все основания думать, что внедрение аналогичных программ в практику спектроскопических исследований также приведет к резкому возрастанию их уровня, повышению точности, скорости и в конечном счете к бурному росту соответствующей информации.

#### ЛИТЕРАТУРА

1. Ю. П. Дробышев, Р. С. Нигматуллин, В. И. Лобанов, И. К. Коробейничева, В. С. Бочкарев, В. А. Коптюг. Система опознавания химических соединений по их ИК-спектрам на базе ЦВМ БЭСМ-6. Семинар «Использование вычислительных машин в спектроскопии молекул». Тезисы докладов. Новосибирск, 1971.
2. Е. М. Попов, Г. А. Коган. О некоторых закономерностях силовых полей органических молекул.— ЖСХ, 1967, т. 8.
3. Л. А. Грибов, В. В. Жогина, С. Ф. Архипова. Использование электронных вычислительных машин в практике расчета колебательных спектров.— ЖПС, 1966, т. 5.
4. В. А. Дементьев, О. И. Кондратов, Л. А. Грибов, Л. И. Кашкан. Программа решения задач о колебаниях многоатомных молекул на электронной счетной машине «Минск-22».— Изв. ТСХА, 1970, вып. 2.
5. В. А. Дементьев, Л. А. Грибов. Программа решения обратной задачи теории колебательных спектров многоатомных молекул на ЭЦВМ «Минск-22».— Изв. ТСХА, 1971, вып. 2.
6. В. А. Дементьев, Л. А. Грибов. О решении обратной спектральной задачи теории частот колебаний многоатомных молекул с помощью метода наименьших квадратов.— ЖПС, 1971, т. 14.
7. Л. А. Грибов. Теория интенсивностей в инфракрасных спектрах многоатомных молекул. М., «Наука», 1963.
8. М. Е. Эляшберг, Л. А. Московкина. Структурно-групповой анализ по ИК-спектрам многоатомных молекул с применением ЭЦВМ.— ЖПС, 1968, т. 8.
9. Л. А. Грибов, М. Е. Эляшберг. Применение алгебры логики в спектроскопических исследованиях.— Изв. ТСХА, 1969, т. 3.

*Поступила в редакцию  
3 января 1972 г.*