

На основе описанной процедуры был разработан машинный алгоритм и одна из его возможных реализаций на языке АЛЬФА. Длина рабочей программы ~ 2000 команд.

Численный эксперимент проводился на модельном материале. Мощность исходного множества функций GN_c была равна 10, а размерность вспомогательного базиса (соответствующая среднеквадратичной ошибке аппроксимации $\epsilon=0,005$) 30. В качестве вспомогательного базиса использовалась ортонормированная система тригонометрических функций. На БЭСМ-6 оптимальный базис был найден за 1 мин.

ЛИТЕРАТУРА

1. F. V. Rупe, E. F. Mecl usky. On Essay on Primeimplicant Tables Fundamental Product Table of Combinations or Truthtable.— J. Soc. JAM, 1961, v. 9, № 4.
2. Т. Л. М а й с т р о в а. Линейное программирование и задачи минимизации нормальных форм булевских функций.— В кн.: Проблемы передачи информации. Вып. 12. М., Изд-во АН СССР, 1963.
3. В. Г. Н о в о с е л о в. Нахождение кратчайших покрытий.— В кн.: Труды СФТИ. Вып. 48. Томск, 1966.
4. В. Д. К а з а к о в. Структурная теория релейных устройств.— В кн.: Нахождение минимальных нормальных форм логической функции методом ограниченного перебора. М., Изд-во АН СССР, 1963.
5. А. Д. З а к р е в с к и й. О сокращении переборов при решении некоторых задач синтеза дискретных автоматов.— ИВУЗ. Сер. радиофизика, 1969, т. 7, № 1.
6. А. Д. З а к р е в с к и й. Оптимизация покрытий множеств.— В кн.: Логический язык для представления алгоритмов синтеза релейных устройств. М., «Наука», 1966.
7. А. М. З а е з д н ы й, Г. В. Э й д у к я в и ч ю с. Сокращенное представление сигналов с помощью систем ортогональных функций.— Радиотехника, 1963, № 11.

Поступила в редакцию 10 января 1974 г.

УДК 518.5 : 541.63+539.193

Б. С. ЖОРОВ

(Ленинград)

МОДЕЛИРОВАНИЕ НА ЭВМ ПРОСТРАНСТВЕННОЙ СТРУКТУРЫ ОРГАНИЧЕСКИХ СОЕДИНЕНИЙ

При решении многих задач теоретической химии и молекулярной биологии возникает потребность в определении устойчивых пространственных структур органических соединений. Установление энергетически выгодных конформаций молекул часто является необходимым этапом расчета молекулярных спектров. Важнейший, но весьма трудоемкий экспериментальный метод исследования конформаций молекул — рентгеноструктурный анализ — позволяет определить пространственную структуру молекул, находящихся только в кристаллическом состоянии. В тех же случаях, когда необходимо иметь представление о всех потенциальных конформационных возможностях соединения, достаточно полную информацию могут дать только расчетные методы.

В настоящее время методы теоретического конформационного анализа успешно применяются для изучения пространственной структуры различных химических соединений. Однако дело осложняется тем обстоятельством, что весьма дорогие программы для ЭВМ, реализующие эти методы, часто оказываются либо приспособленными к узкому классу соединений, либо требуют для расчета конкретной молекулы многочис-

ленных исходных данных, подготовка которых становится одним из самых трудоемких этапов работы.

Для более интенсивного внедрения расчетных методов в конформационные исследования необходимо создание комплексов программ, ориентированных на широкий круг соединений, и в то же время доступных для специалистов, не знакомых с вычислительной техникой и программированием. Ниже излагаются важнейшие принципы построения такой системы. На их основе были созданы конформационные программы для ЭВМ «Минск-22», «Наири-3» и «Минск-32».

Суть метода теоретического конформационного анализа состоит в исследовании функции

$$E = f(\bar{q}),$$

где E — конформационная энергия молекулы; \bar{q} — вектор обобщенных координат, однозначно определяющих геометрию молекулы (длины связей, валентные углы, углы внутреннего вращения). Значения обобщенных координат, при которых функция E имеет минимумы, характеризуют устойчивые конформации молекулы.

В приближении атом-атом потенциалов [1] конформационная энергия молекулы представляется в виде суммы взаимодействий, часть которых является функциями межатомных расстояний. Поэтому необходимым элементом конформационных расчетов является вычисление координат атомов по заданному набору обобщенных координат. Несмотря на то что проблема эта принципиально решена [2—6], именно с нею связаны основные трудности построения универсальной конформационной программы.

Для вычисления и минимизации конформационной энергии молекулы произвольной структуры требуется хранить в памяти ЭВМ около тридцати массивов, описывающих математическую модель молекулы. Все эти массивы могут быть подготовлены специальной сервисной программой на основании исходных данных, которые можно разбить на три группы: а) характеристики, присущие одним и тем же типам атомов и связей в любой молекуле; б) минимальное описание структуры конкретной молекулы и особенности расчетной модели; в) характеристики, присущие только исследуемой молекуле. Поскольку наибольший удельный вес в общем объеме исходной информации имеют характеристики первой группы, отказ от необходимости их ввода для каждой конкретной молекулы значительно сокращает этап подготовки к расчетам. Поэтому целесообразно собрать эти данные для широкого класса молекул, заложить их в программы и менять только по мере их уточнения и расширения.

Важнейший принцип построения универсальной конформационной программы основан на той же идее, которая используется в известных наборах механических атомных моделей, позволяющих из «атомов» ограниченного числа типов собирать модели разнообразных молекул. Математическую модель молекулы (совокупность массивов для программ исследования конформаций) также можно собирать из ограниченного набора хранящихся в сервисной программе описаний атомов. Основным элементом такого описания является векторная модель атома (ВМА), представляющая собой набор единичных векторов (лучей), число которых равно числу одиночных связей, образуемых данным атомом, а углы между лучами равны валентным углам. Каждый луч имеет свой номер, причем вся векторная модель ориентирована относительно собственной локальной системы координат таким образом, чтобы первый луч был направлен противоположно оси x_A , а ось z_A совпадала с направлением векторного произведения лучей 1 и 2 ($\vec{L}_2 \times \vec{L}_1$). На рис. 1 приведена ВМА тетраэдрического типа. В памяти машины векторные модели представлены в виде матриц размером $N \times 3$, где N — число лучей данной

ВМА. Значения ковалентных радиусов, соответствующих каждому лучу ВМА всех типов, образуют отдельный нормативный массив сервисной программы.

Помимо геометрических параметров атому каждого типа сопоставляется набор энергетических характеристик и типовых признаков, из которых сервисная программа формирует массивы, используемые при расчете энергии молекулы.

Описание молекул. Структуру конкретной молекулы можно задать, если указать порядок соединения векторных моделей атомов, образующих эту молекулу. Такое описание представляет собой совокупность строк, каждая из которых характеризует одну связь в молекуле. Атом, стоящий в описании связи первым, называется базовым, а второй — присоединяемым. Строка, как правило, содержит номер, тип и номер луча базового атома; номер, тип и номер луча присоединяемого атома; признак, указывающий на возможность внутреннего вращения относительно данной связи, и значение двугранного угла, если вращение запрещено. В качестве примера на рис. 2 приведено описание молекулы *l*-аланина. Процесс задания структуры в известной мере напоминает сборку модели молекулы из отдельных атомов: запись одной строки равносильна присоединению еще одного атома к уже сформированной части. Локальная система координат векторной модели первого атома является глобальной системой координат всей молекулы. Если после задания структуры остаются свободными лучи каких-либо атомов, то описание дополняется автоматически за счет присоединения к этим лучам атомов водорода. Для построения модели бимолекулярного комплекса описание второй молекулы помещается вслед за первой, причем нумерация атомов не прерывается.

После задания структурной формулы вводится следующая информация, характеризующая особенности расчетной модели молекулы и значения некоторых потенциалов: а) способы отсчета двугранных углов, задаваемые номерами четверок атомов, дисконформации которых соответствуют нулевым значениям углов; б) номера атомов, валентные углы которых могут варьироваться при поиске устойчивых конформаций; в) типы и потенциалы возможных водородных связей; г) параметры торсионных потенциалов; д) заряды на атомах. Поскольку значения зарядов задаются, как правило, на основании квантовохимических расчетов, их получение также может быть автоматизировано. Например, в комплексе для ЭВМ «Минск-22» заряды рассчитываются по методу Дель Ре [7].

На основании перечисленных исходных данных и собственных нормативных массивов сервисная программа формирует массивы (целочисленные — в упакованном виде), которые используются непосредственно конформационными программами.

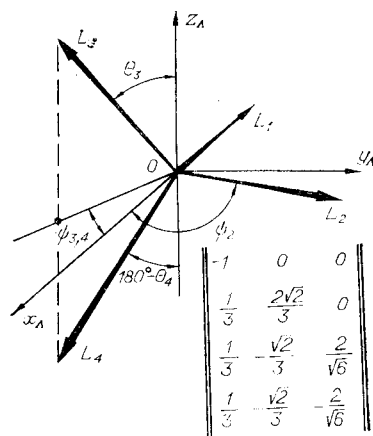


Рис. 1. Векторная модель атомов тетраэдрического типа. Элементы строк матрицы есть проекции лучей на оси локальной системы координат. Направления лучей однозначно задаются пятью переменными.

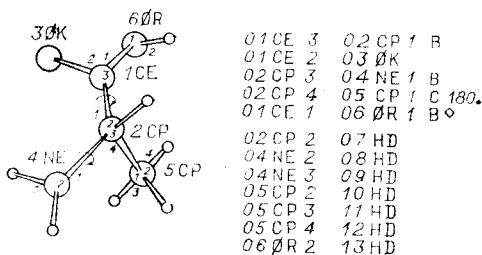


Рис. 2. Описание структуры молекулы *l*-аланина.

Вычисление декартовых координат атомов (сборка молекул). Входной информацией для этой программы является массив векторных моделей атомов (МВМА); адреса векторных моделей атомов в массиве МВМА (АВМА); размеры векторных моделей; массивы номеров атомов и лучей, характеризующие порядок соединения векторных моделей; длины связей; текущие значения варьируемых геометрических параметров; массив номеров связей, к которым относятся варьируемые геометрические параметры.

Перед сборкой молекулы ориентация векторных моделей атомов в локальных системах координат (массив МВМА) в общем случае отличается от той, которую эти ВМА имеют в нормативном массиве сервисной программы. У каждой ВМА есть 2 особых луча: луч \vec{L}_p^i , с помощью которого векторная модель j -го атома присоединяется к уже сформированной части молекулы, образуя k -ю связь; и луч \vec{L}_q^i , задающий вместе с лучом \vec{L}_p^i одну из плоскостей двугранного угла внутреннего вращения вокруг k -й связи. Номера этих лучей определяются при задании структуры молекулы и способов отсчета двугранных углов и засылаются сервисной программой в особые массивы. В массиве МВМА векторная модель j -го атома ориентирована таким образом, чтобы луч \vec{L}_p^i был направлен противоположно оси x_A . Ориентация луча \vec{L}_q^i зависит от того, допускается или нет вращение вокруг k -й связи. В первом случае луч \vec{L}_q^i фиксируется на плоскости $x_A O y_A$; во втором — проекция луча M_k^i на плоскость $y_A O z_A$ образует с положительным направлением оси y_A угол φ_k , равный фиксированному значению двугранного угла внутреннего вращения вокруг k -й связи. Если структура молекулы описана так, что для некоторых атомов $p \neq 1$ и (или) $q \neq 2$, и (или) $\varphi \neq 0$, то сервисная программа ориентирует их векторные модели в локальных системах координат надлежащим образом. За счет этого все постоянные преобразования векторных моделей выносятся за пределы программы сборки молекул.

Сборке молекул предшествует коррекция векторных моделей атомов с упругими валентными углами. Сами валентные углы не являются обобщенными координатами, но могут быть легко получены из векторной модели. Направления же лучей векторной модели, например тетраэдрического атома, однозначно определяются пятью переменными (см. рис. 1). Первоначально значения этих переменных, соответствующие валентным углам выбранной векторной модели, определяются сервисной программой. В процессе минимизации энергии эти переменные наряду с углами внутреннего вращения рассматриваются уже как обобщенные координаты.

Программа сборки молекулы просматривает последовательно характеристики связей в том порядке, как они заданы в описании молекулы. При этом определяются координаты присоединяемых атомов, а их векторные модели ориентируются в глобальной системе координат и засылаются в массив ориентированной векторной модели атома (ОВМА), структура которого аналогична массиву МВМА. Рассмотрим операции, выполняемые при сборке одной связи. Пусть k -я связь образуется лучами \vec{L}_R^i базового атома i и \vec{L}_p^j присоединяемого атома j (рис. 3). Пусть, далее, известны координаты \vec{Q}_i атома i и направления всех его лучей в глобальной системе координат (матрица A_i по

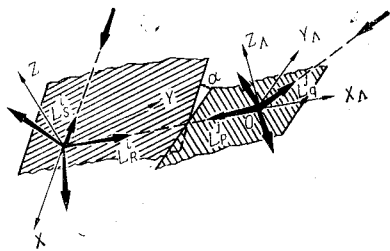


Рис. 3. Взаимное расположение векторных моделей атомов в связи.

адресу АВМА [i] в массиве ОБМА), а также длина связи d_k и угол внутреннего вращения α_k относительно этой связи. Тогда координаты присоединяемого атома в глобальной системе координат определяется следующим образом:

$$\bar{Q}_j = \bar{Q}_i + d_k \bar{L}_R^i.$$

Повернем векторную модель M_j присоединяемого атома j вокруг луча \bar{L}_R^i на угол α по часовой стрелке (если смотреть со стороны оси x_A):

$$M'_j = M_j \begin{vmatrix} 1 & 0 & 0 \\ 0 & \cos \alpha & -\sin \alpha \\ 0 & \sin \alpha & \cos \alpha \end{vmatrix}.$$

Сориентируем теперь локальную систему координат таким образом, чтобы ось x_A совпала с лучом L_R^i , а ось z_A — с векторным произведением $\bar{L}_R^i \times \bar{L}_S^i$, и определим проекции лучей векторной модели присоединяемого атома на оси координат глобальной системы:

$$A_j = M'_j \begin{vmatrix} L_{Rx}^i L_{Ry}^i L_{Rz}^i \\ I_x I_y I_z \\ K_x K_y K_z \end{vmatrix},$$

где A_j — матрица векторной модели присоединяемого атома в глобальной системе координат; $L_{Rx}^i, L_{Ry}^i, L_{Rz}^i$ — компоненты луча L базового атома i (в глобальной системе координат); K_x, K_y, K_z — компоненты единичного вектора \bar{K} , совпадающего по направлению с векторным произведением $\bar{L}_R^i \times \bar{L}_S^i$; I_x, I_y, I_z — компоненты вектора $\bar{I} = \bar{K} \times \bar{L}_R^i$. Для связи, в которой участвует водород (или подобный атом), достаточно определить лишь координаты присоединяемого атома.

При вычислении координат атомов бимолекулярного комплекса алгоритм полностью сохраняется, но в качестве первого шага задаются координаты базового атома первой связи свободной молекулы, а его векторная модель ориентируется в глобальной системе координат. Значения шести переменных, необходимых для выполнения этих операций, хранятся в массиве обобщенных координат.

Вычисление и минимизация энергии. В приближении атом-атом потенциалов вычисление конформационной энергии при известных координатах атомов не представляет затруднений, поскольку основные энергетические составляющие являются простыми функциями межатомных расстояний [8]. Для достижения большей гибкости модели вычисление отдельных составляющих энергии (невалентные и электростатические взаимодействия, торсионные потенциалы, водородные связи, деформация валентных углов, свободная энергия гидратации) осуществляется отдельными блоками, которые могут выполняться или не выполняться в зависимости от состояния управляющих параметров. Для расчета невалентных взаимодействий используется несколько наборов потенциалов, предложенных разными авторами [8].

При вычислении градиента энергии в пространстве обобщенных координат принимается во внимание тот факт, что частная производная от энергии по i -й обобщенной координате равна по модулю обобщенной силе, сопряженной с этой обобщенной координатой. Для углов внутреннего вращения, валентных углов и эйлеровых углов ориентации свободной молекулы обобщенные силы есть моменты относительно соответст-

вующих осей вращения, обусловленные действием внутримолекулярных сил. Поскольку сила взаимодействия двух атомов определяется как производная от энергии этого взаимодействия по расстоянию, то складывая векторы этих сил при соответствующих атомах, можно одновременно с расчетом энергии всех парных взаимодействий определить и силы воздействия на каждый атом со стороны остальных атомов системы. Тогда момент относительно i -й оси вращения может быть найден путем суммирования вкладов от сил, приложенных к тем атомам системы, положение которых зависит от i -й обобщенной координаты. Для определения номеров соответствующих атомов используется формируемая сервисной программой булева матрица, элемент которой $B_{ji}=1$, если положение j -го атома зависит от i -й обобщенной координаты, и $B_{ji}=0$ — в противном случае. Эта же матрица используется для исключения из расчета тех межатомных взаимодействий, которые не зависят ни от одной из обобщенных координат. Для такой пары атомов соответствующие строки матрицы B идентичны, а поскольку строка занимает одну ячейку памяти (если число переменных меньше, чем длина разрядной сетки машины), весь анализ сводится к сравнению двух элементов массива. Вычисление градиента с помощью обобщенных сил позволяет многократно экономить машинное время по сравнению с методом конечных разностей, поскольку для нахождения частных производных требуется порядка 10% от времени расчета энергии, и величина эта слабо зависит от числа переменных. Аналогичные результаты дает метод, основанный на аналитическом представлении градиента [9].

Поиск минимумов энергии на первом этапе осуществляется при помощи градиентного спуска с автоматическим выбором шага [10], а в области малых значений градиента — методом Давидона [11]. На ЭВМ «Минск-32» поиск одной устойчивой конформации молекулы из 30 атомов с 20 степенями свободы занимает около трех минут.

Управление моделью и представление результатов. Настройка на определенный режим работы осуществляется специальной программой, которая вначале задает стандартные значения управляющим параметрам, обеспечивая тем самым некоторый оптимальный режим. В случае необходимости в произвольной последовательности вводятся идентификаторы и новые значения управляющих параметров, что дает возможность исключать отдельные типы взаимодействий, изменять наборы потенциалов, выбирать форму представления выходной информации, задавать критерии сходимости и т. п. Программа настройки позволяет также корректировать некоторые элементы и участки массивов, сфор-

мированных сервисной программой, что позволяет изменять жесткие геометрические параметры и другие характеристики. Предусмотрена возможность фиксировать любые координаты при поиске минимумов. В целом настроенная программа в комплексе с блочной организацией конформационных программ позволяет быстро (в системе для ЭВМ «Наири-3» в режиме диалога человек — машина) изменять структуру математической модели, что дает возможность оперативно исследовать влияние различных факторов на конформационные свойства молекул.

Для найденных устойчивых конформаций предусмотрен вывод изображения трехмерной структуры молекул. На АЦПУ или пишущей машинке печатаются номера атомов, расположенные на плоскости листа таким образом, что если провести связи,

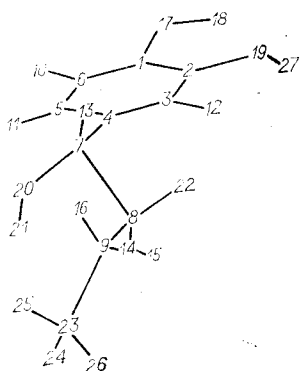


Рис. 4. Изображение трехмерной структуры молекулы адреналина, полученное на выводной пишущей машинке ЭВМ «Наири-3».

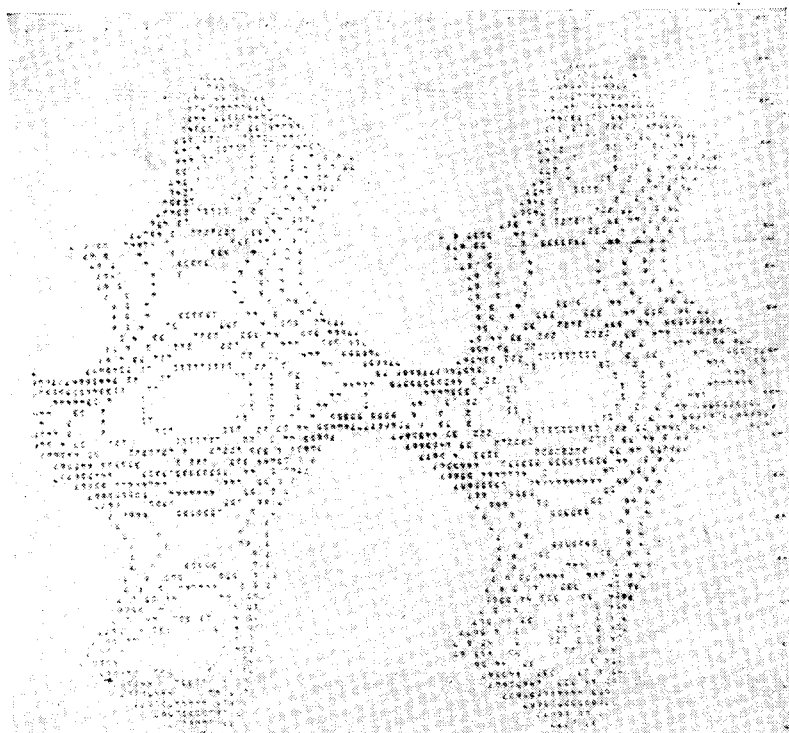


Рис. 5. Распечатка на АЦПУ конформационной карты. Цифры указывают величину уровня (ккал/моль) относительно глобального минимума.

получится аксонометрическая проекция молекулы (рис. 4). При расчете конформационных карт изображение энергетической поверхности выводится на АЦПУ в виде совокупности линий уровня (рис. 5).

ЛИТЕРАТУРА

1. А. И. Китайгородский. Молекулярные кристаллы. М., «Наука», 1971.
2. H. E. Eyring. The Resultant Electric Moment of Complex Molecules.— *Phys. Rev.*, 1932, v. 39, p. 746.
3. D. A. Grant, P. J. Flory. The Configuration of Random Polypeptide Chain. 11. Theory.— *J. Chem. Soc.*, 1965, v. 87, p. 2791.
4. C. R. Eddy. Configuration of the Spatial Locations of Atoms of a Chain Molecules Undergoing Intermolecular Rotations.— *J. Chem. Phys.*, 1963, v. 38, p. 1032.
5. T. Ooi, R. A. Scott, G. Vandercooi, H. A. Scheraga. Conformational Analysis of Macromolecules. IV.— *J. Chem. Phys.*, 1967, v. 46, p. 4410.
6. H. B. Thompson. Calculation of Cartesian Coordinates and their Derivatives from Internal Molecular Coordinates.— *J. Chem. Phys.*, 1967, v. 47, p. 3407.
7. J. Del Re. A Simple HO-LCAO Method for the Calculation of Charge Distribution in Saturated Organic Molecules.— *J. Chem. Soc.*, 1958, № 11.
8. В. Г. Дашевский. Конформации олиго- и полипептидов.— В кн.: Итоги науки. Сер. химия. (Высокомолекулярные соединения). М., ВИНТИ, 1970.
9. С. Г. Галактионов. Описание пространственной структуры сложных молекул.— В кн.: Конформационные расчеты сложных молекул. Минск, Изд. Ин-та теплообмена, 1970.
10. И. М. Гельфанд, Е. Б. Вул, С. Л. Гинсбург, Ю. Г. Федоров. Метод оврагов в задачах рентгеноструктурного анализа. М., «Наука», 1966.
11. R. Fletcher, M. Powell. A Rapidly Convergent Descent Method for Minimization.— *Computer J.*, 1963, v. 6, p. 163.

Поступила в редакцию 22 декабря 1973 г.;
окончательный вариант — 16 апреля 1974 г.