

## ГЛАВА 2

### FROG - комплекс программ для уточнения атомной структуры макромолекул.

Комплекс программ FROG /19-23/ предназначен для решения задач рентгеновской кристаллографии, конформационного анализа, расчетных задач белковой инженерии. Он позволяет изменить в модели исследуемого объекта значения координат атомов, их температурных факторов и коэффициентов заполнения таким образом, чтобы добиться:

- лучшего соответствия данным рентгеновского эксперимента (минимизировать R-фактор);
- более точного соответствия стандартным стереохимическим соотношениям (длины связей, валентные углы, плоские группы, хиральность и т.д.);
- меньшего значения энергии невалентных взаимодействий (в том числе межмолекулярных);
- лучшего вписывания молекулы в электронную плотность;
- выполнения любых других требований, выражаемых формально через координаты атомов модели, значения соответствующей модели функции распределения электронной плотности или ее коэффициентов Фурье.

По желанию пользователя в работе может использоваться любое из перечисленных требований или произвольная их комбинация.

Комплекс FROG представляет собой пример разработки програм-много обеспечения, которое можно легко адаптировать к новым поста-новкам задач, сохраняя при этом высокую эффективность работы

#### **2.1. Основные особенности комплекса FROG.**

##### **1. Быстродействие.**

Комплекс построен на основе последовательного применения алгоритма быстрого дифференцирования /105-108/ и быстрого преобразования Фурье/75-77,109/. Размер используемой оперативной

памяти и время работы программы зависят от размеров уточняемого объекта и разрешения, при котором идет работа. При уточнении структуры молекулы белка (около 3000 атомов в независимой части элементарной ячейки) на разрешении 2.5 Å (12000 рефлексов) по совокупности рентгеновских и стереохимических данных время цикла уточнения составляет около 30 минут процессора на ЭВМ ЕС 1055М с быстродействием 300 000 оп./сек.

На IBM PC 486/487 цикл уточнения молекулы белка размером около 5000 атомов в независимой части ячейки на разрешении 1.7 Å (55 000 рефлексов) с учетом стереохимии составляет 5 минут. С увеличением размеров белка требуемое процессорное время для уточнения растет линейно.

## **2. Атомно-блочная модель макромолекулы.**

Модель уточняемого объекта в комплексе FROG представляет собой конгломерат из свободных атомов и жестких групп. Параметры свободных атомов могут меняться независимо друг от друга; жесткая группа в процессе уточнения двигается как твердое тело, а взаимное положение атомов внутри группы остается неизменным. Такую модель далее будем называть атомно-блочной.

Количество и состав жестких групп может быть произвольным:

- вся модель может состоять из свободных атомов;
- жесткой группой может быть вся молекула, тогда в процессе уточнения будут меняться только ориентация и положение центра молекулы как твердого тела;
- можно объявить жесткими группами отдельные домены, спиральные участки, отдельные пептидные звенья, боковые цепи и т.д.

## **3. Задание критериев уточнения внешними файлами.**

Ядро комплекса составляет программа уточнения модели макромолекулы. При этом описание требований, предъявляемых к модели (критериев уточнения), вынесено во внешние файлы, формируемые независимо от основной программы. Это позволяет менять требования,

предъявляемые к модели, без корректировки текста программы. В частности, это позволяет легко вводить новые ограничения и уточнять модели разных классов макромолекул (белков, ДНК, различных комплексов и т.д.) со своими собственными стереохимическими ограничениями. Подготовка этих файлов осуществляется специальными вспомогательными программами комплекса, описанными ниже (пп. 2.4-2.6). Если возможности этих программ оказываются недостаточными, пользователь может создать свои программы подготовки входных файлов с требуемыми критериями уточнения.

#### **4. Применение R-free фактора.**

Имеется возможность при уточнении использовать R-free фактор в качестве критерия оценки качества модели (см. п.1.5.). Для этого во входном файле структурных факторов выбирается случайным образом подмножество рефлексов  $T$  (в объеме 10% от полного набора, например) и им приписывается признак “free”. В дальнейшем уточнение проводится по оставшейся части рефлексов, а по рефлексам с признаком “free” производится только расчет R-фактора для контроля за процессом уточнения.

#### **5. Возможность “жесткого” учета некристаллографической симметрии.**

Некристаллографическая (н/к) симметрия может учитываться при уточнении двумя способами:

- “жесткое” введение н/к симметрии. В этом случае программе можно задать и независимо изменять параметры атомов лишь одной субъединицы, остальные субъединицы при этом генерируются при помощи заданных преобразований н/к симметрии. Эта возможность позволяет также работать в произвольной пространственной группе.
- “мягкое” введение н/к симметрии. Параметры атомов всех субъединиц в этом случае изменяются независимо, но в минимизируемый критерий включается штраф, величина которого тем меньше, чем точнее выполняются соответствующие условия симметрии.

## **6. Уточнение положения молекулы и параметров некристаллографической симметрии.**

Объявляя жесткой группой всю молекулу, можно решать задачу уточнения положения молекулы внутри элементарной ячейки и уточнения параметров н/к симметрии. При этом можно использовать различные критерии уточнения: энергетический критерий; соответствие рассчитанных по модели модулей структурных факторов эксперименту; вписывание молекулы в "экспериментальную" электронную плотность и т.п.

## **7. Проверка геометрии модели, полученной при уточнении.**

В комплексе FROG имеются программы сбора статистики (п.2.7.), позволяющие получить информацию о нарушении стереохимических и энергетических соотношений в имеющейся модели молекулы белка. Программы сообщают по каждому из типов контролируемых параметров не только величины средних ошибок, но и максимальные отклонения от стандартных значений. При необходимости более тщательного изучения ошибок в геометрии модели можно распечатать всю интересующую информацию об отклонениях, превосходящих заданные пользователем пороговые величины. Кроме того, рассчитываются гистограммы распределения отклонений от заданных значений по всем критериям, включенными в уточнение.

## **8. Адаптируемость к размеру доступной памяти.**

Программы комплекса легко адаптируются к любому размеру доступной памяти. Имеется две версии программы уточнения: файловая и бесфайловая. Файловая версия в процессе расчетов использует вывод промежуточных результатов во внешние рабочие файлы, освобождая оперативную память для последующих действий. Это позволяет работать на любой доступной памяти, но увеличивает время работы. При этом единственным ограничением является то, что в буферный массив должно помещаться по крайней мере два сечения рассчитанной электронной плотности. В бесфайловой версии все

необходимые данные хранятся в памяти, за счет чего ускоряется процесс уточнения, но эта версия рассчитана на большие объемы оперативной памяти.

Количество атомов в модели, размеры элементарной ячейки, разрешение набора дифракционных данных не ограничиваются. Программы работают с любой пространственной группой симметрии.

Комплекс написан на языке ФОРТРАН, его объем - около 35000 операторов. Схема связи основных программ комплекса и файлов данных приведена на рис.2.

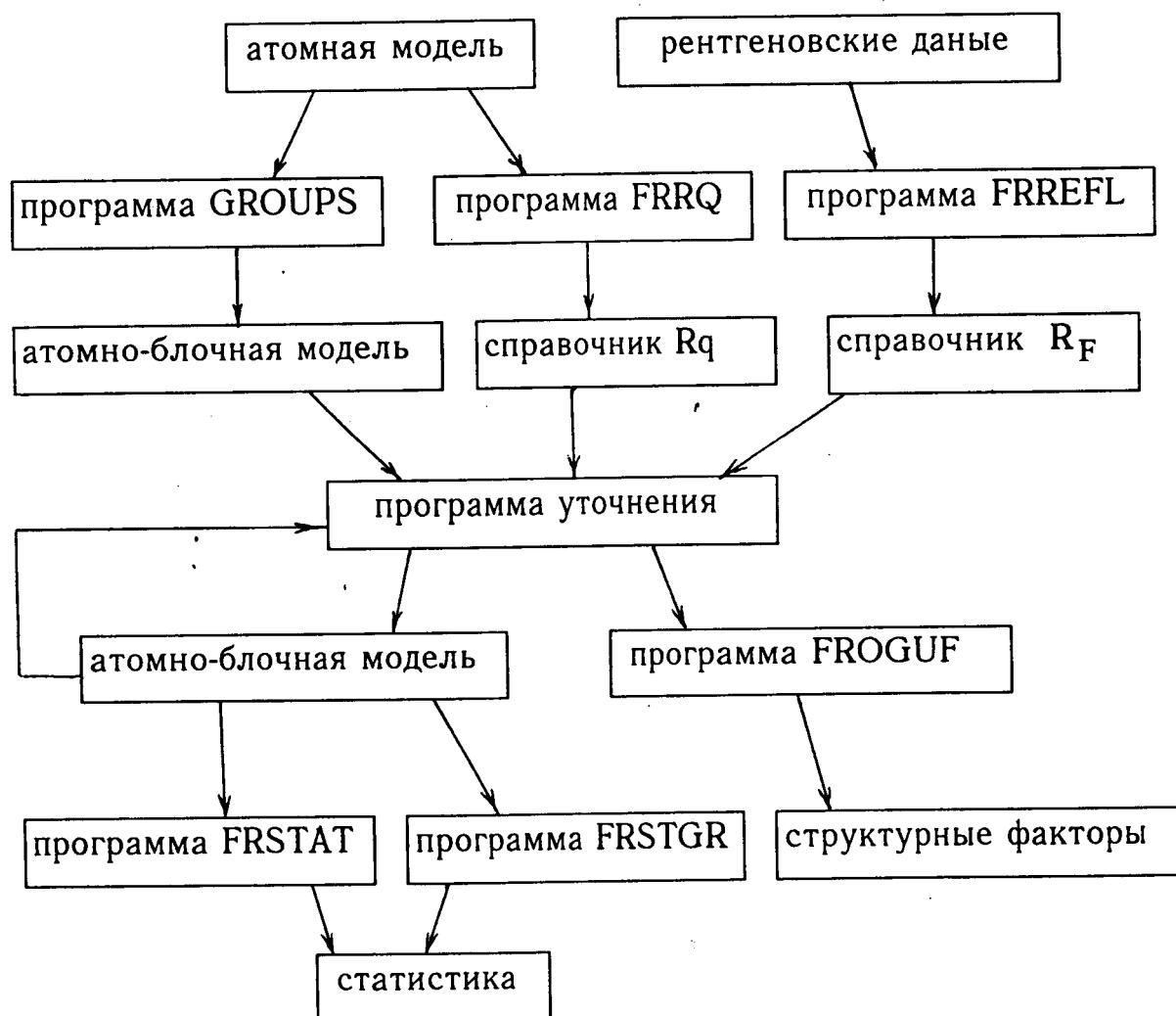


Рис.2. Схема функциональных связей программ и файлов комплекса FROG.

## **2.2. Работа с комплексом FROG.**

Ниже представлены основные сведения, поясняющие ход работы с комплексом FROG. Подробное описание форматов файлов, режимов работы и управляющих данных всех программ имеется в документации, поставляемой вместе с комплексом.

### **1. Исходные данные для уточнения.**

В начале работы над новым объектом создаются исходный файл с атомами модели в формате PDB с координатами, изотропными температурными факторами и коэффициентами заполнения, и файл структурных факторов (в формате UF, п.3.4.2.), в котором для каждого отражения может быть приведено значение модуля структурного фактора, стандартное отклонение  $\sigma$  для этого значения, коэффициенты A,B,C,D распределения вероятностей для фаз структурных факторов /110/, весовые множители и т.п. Эти файлы имеют достаточно простую организацию, облегчающую их редактирование и просмотр.

Для запуска программы уточнения исходные файлы преобразуются во внутренний формат комплекса FROG с добавлением информации о режиме уточнения. На этом этапе могут создаваться файлы:

#### ***файл со стартовой атомно-блочной моделью***

Этот файл присутствует всегда; в нем хранятся параметры модели и информация о том, как атомы объединены в жесткие группы и какие параметры разрешается менять в процессе уточнения.

#### ***файл с рентгеновскими данными - справочник $R_f$***

Этот файл должен присутствовать, если в процессе уточнения рассчитанные по текущей модели структурные факторы сравниваются с экспериментально определенными значениями (так, при энергетическом уточнении этот файл может отсутствовать).

#### ***файл со стереохимическими данными - справочник $R_q$***

В файле содержится информация о том, какие стереохимические и энергетические соотношения требуется контролировать в процессе

уточнения. Если никаких ограничений на модель не накладывается, этот файл может отсутствовать.

Подготовка атомно-блочной модели и файлов, описывающих критерии уточнения (справочников  $R_f$  и  $R_q$ ), осуществляется специальными сервисными программами, входящими в комплекс FROG (пп. 2.4-2.6).

## 2. Критерии уточнения.

С математической точки зрения программа уточнения должна изменять значения параметров модели таким образом, чтобы минимизировать значение функционала, являющегося количественным представлением всех требований, предъявляемых к модели. Программа уточнения комплекса FROG фиксирует общую структуру критерия, но не его конкретную форму. Конкретный вид функционала описывается специальными входными файлами-справочниками, которые позволяют его легко варьировать, не внося изменений в текст программы уточнения. Функционал может содержать любые требования к модели, которые выражаются в терминах параметров атомно-блочной модели, атомных параметров, распределения электронной плотности или структурных факторов.

В общем виде минимизируемый критерий  $R$  можно записать как

$$R = w_F R_F + \sum_{j=1}^6 w_{q,j} R_{q,j}, \quad (29)$$

где  $R_F$  описывает соответствие модели рентгеновскому эксперименту;

$R_{q,j}$  описывает точность выполнения заданных стереохимических соотношений;

$w_F, w_{q,j}$  - весовые константы, задаваемые в управляющих данных..

Критерий  $R_F$  может иметь следующий вид:

$$R_F = \sum_s w_s (|F_{obs}(s)| - k |F_{calc}(s)|)^2 \quad (30)$$

$$R_F = \sum_s w_s | |F_{obs}(s)| - k |F_{calc}(s)| | \quad (31)$$

$$R_F = \sum_s w_s (|F_{obs}(s)|^2 - k|F_{calc}(s)|^2)^2 \quad (32)$$

$$R_F = \sum_s (A_s \cos(\phi_s) + B_s \sin(\phi_s) + C_s \cos(2\phi_s) + D_s \sin(2\phi_s)) , \quad (33)$$

где  $A_s, B_s, C_s, D_s$  - заданные коэффициенты функции распределения вероятности фаз  $\phi_s$ .

Критерии  $R_{q,j}$  задаются на языке параметров атомов модели и описывают соответствие модели стереохимическим соотношениям:

$$R_{q,1} = \sum_{i,j} \left( \frac{A_{ij}}{|r_i - r_j|^{12}} + \frac{B_{ij}}{|r_i - r_j|^6} \right) \quad (34)$$

энергия невалентных взаимодействий по невалентно связанным парам атомов, находящихся в пределах заданного расстояния друг от друга;

$$R_{q,2} = \sum_{i,j} (|r_i - r_j|^2 - d_0^2)^2 \quad \text{по всем валентным связям} \quad (35)$$

$$R_{q,3} = \sum_k (1 - \cos(\alpha_k - \alpha_0)) \quad \text{по всем валентным углам} \quad (36)$$

$$R_{q,4} = \sum_k (1 - \cos m(\theta_k - \theta_0)) \quad \text{по всем торсионным углам} \quad (37)$$

$$R_{q,5} = \sum_k d^2(r_{j,k}, L_{opt}) \quad \begin{aligned} &\text{отклонения атомов плоских групп от} \\ &\text{соответствующих плоскостей.} \end{aligned} \quad (38)$$

$$R_{q,6} = \sum_k (V_k - V_0)^2 \quad \begin{aligned} &\text{по четверкам атомов, образующих} \\ &\text{хиральные объемы.} \end{aligned} \quad (39)$$

### **3. Возможные действия после работы программы уточнения.**

По завершении работы программы возможны следующие действия:  
а). Продолжение уточнения прежней атомно-блочной модели.

Если принято решение продолжить уточнение прежней модели, не меняя количества и состава введенных жестких групп, то полученный файл с уточненной моделью можно использовать как входной файл для следующего запуска программы уточнения, не меняя при этом входные файлы-справочники.

## б). Изменение количества и состава жестких групп.

Если принято решение пересмотреть состав жестких групп либо провести ручную правку модели, то атомно-блочная модель переводится в атомную модель, производится разбивка на новые группы и пересчитывается справочник Rq.

## в). Расчет структурных факторов.

При желании можно задать режим расчета структурных факторов по уточненной модели.

## г). Проверка стереохимических параметров модели.

После любого этапа уточнения можно осуществить проверку корректности модели с точки зрения ее стереохимии и неразрешенных контактов (программы FRSTAT и FRSTGR в п.2.7.).

## **2.3. Основные характеристики программы FROG.**

Автоматическое уточнение атомной модели макромолекулы обычно заключается в минимизации некоторого функционала  $R(\chi)$ : чем меньше величина  $R(\chi)$ , тем лучше соответствуют рассчитанные характеристики модели экспериментальным величинам. Основные проблемы уточнения возникают из-за большого размера пространства варьируемых параметров  $\chi$ , трудоемкости вычисления таких характеристик модели, как, например, структурные факторы, и немонотонности поведения  $R(\chi)$ . Все это приводит к локальному характеру минимизации.

### **1. Общая структура программы.**

Ввиду большого числа переменных и сложной зависимости от них минимизируемой функции основную часть программы уточнения комплекса FROG составляют подпрограммы, организующие расчет этой функции и ее градиента. Для минимизации применяется метод спуска с выбором направления по антиградиенту либо метод сопряженных градиентов. Для поиска оптимальной точки на найденном направлении спуска применяется аппроксимация

минимизируемой функции многочленом третьей степени по значениям функции и ее производной на концах отрезка аппроксимации.

## 2. Параметры, описывающие модель.

В программе используется несколько уровней описания модели:

$\chi$  - обобщенные параметры жестких групп, на которые разбита модель (углы Эйлера, вектор трансляции, групповой температурный фактор, групповой коэффициент заполнения);

$q$  - параметры атомов, входящих в модель (декартовы координаты, изотропные температурные факторы, коэффициенты заполнения);

$\rho$  - распределение электронной плотности в ячейке кристалла, рассчитанное по атомной модели на некоторой сетке  $r$ ;

$F$  - структурные факторы, соответствующие рентгеновскому рассеянию распределением электронов  $\rho(r)$ .

Связь между этими уровнями выглядит как  $\chi \Rightarrow q \Rightarrow \rho \Rightarrow F$  (40)

Здесь стрелки показывают направление, в котором вычисления не вызывают особых затруднений. Обратный путь обычно проблематичен. Например, преобразование  $q \Leftarrow \rho$  представляет собой процесс интерпретации карт электронной плотности.

Отметим, что некоторые программы уточнения вычисляют структурные факторы прямо по атомным параметрам  $q \Rightarrow F$  с помощью аналитических формул. Однако двухшаговое преобразование  $q \Rightarrow \rho \Rightarrow F$ , реализованное в FROG, для макромолекул более эффективно и позволяет накладывать такие ограничения на модель, которые выражаются в терминах распределения электронной плотности. Поэтому распределение плотности выделено в отдельный уровень.

## 3. Структура функционала.

Требования, которые накладываются на модель, имеют общую особенность: каждое из них является локальным, т.е. связывает всего несколько параметров.

Например, условие  $|F_{obs}(s)| - |F_{calc}(s)| )^2 \Rightarrow \min$  (41)

ограничивает величину единственного структурного фактора  $F_{\text{calc}}(s)$ .

Другой пример: условие  $(|\mathbf{r}_j - \mathbf{r}_k| - d_{jk}^0)^2 \Rightarrow \min$  (42)

связывает координаты только двух конкретных атомов. Каждое условие применяется в терминах одного из перечисленных в (40) уровней переменных. Поэтому общий минимизируемый функционал  $R$  представляется в виде суммы большого числа условий (ограничений), выраженных в терминах  $\chi$ ,  $q$ ,  $\rho$  или  $F$ , и подразделяется на 4 компоненты:

$$R = R\chi + Rq + Rp + RF, \quad (43)$$

где  $R\chi$  выражается непосредственно через обобщенные параметры

жестких групп текущей атомно-блочной модели;

$Rq$  - через параметры атомной модели;

$Rp$  - через значения электронной плотности;

$RF$  - через структурные факторы, рассчитанные по текущей модели.

Каждый из 4 компонентов может расщепляться на функции разного вида. Например,  $Rq$  может включать в себя ограничения на длины связей, валентные углы, планарность, невалентные контакты и т.д. (все они выражаются непосредственно через атомные координаты). Наконец, каждая из таких конкретных функций (например, энергия невалентных взаимодействий) представляет собой сумму большого числа однотипных слагаемых, зависящих от небольшого числа переменных.

Во входных данных функциям  $R\chi$ ,  $Rq$ ,  $Rp$ ,  $RF$  отвечают соответствующие файлы-справочники. Например, в справочнике  $Rq$  для каждого элемента критерия, контролирующего отклонение длин валентных связей от стандартных значений, задается весовой коэффициент, номера атомов, образующих валентную связь, и стандартная величина связи  $d_{jk}^0$ . Аналогично в справочнике  $RF$  для каждого слагаемого рентгеновского критерия (41) задается весовой

коэффициент, индексы рефлекса  $h, k, l$  и экспериментальное значение модуля структурного фактора  $|F_{\text{obs}}(h, k, l)|$ .

Использование файлов-справочников имеет несколько преимуществ перед обычным заданием функционала непосредственно внутри программы. Это дает возможность уточнения не только белковых структур, но и моделей других типов макромолекул, например, молекул ДНК, без изменения программы; возможность включения в процесс уточнения нестандартных ограничений; использования различных весовых схем; индивидуальной работы с каждым элементом функционала; включения рефлекса в несколько критериев одновременно; независимость критериев, что позволяет легко их изменять и добавлять новые критерии (для этого нужно изменить или добавить небольшую подпрограмму расчета конкретного элемента критерия).

#### 4. Расчет функционала и градиента.

Иерархия уровней описания модели (40) и возможность представления общего функционала  $R$  в виде компонентов (43), зависящих только от переменных  $\chi, q, \rho$  и  $F$ , позволяет применить следующую оптимальную схему расчета функции  $R$  и градиента (рис.3):

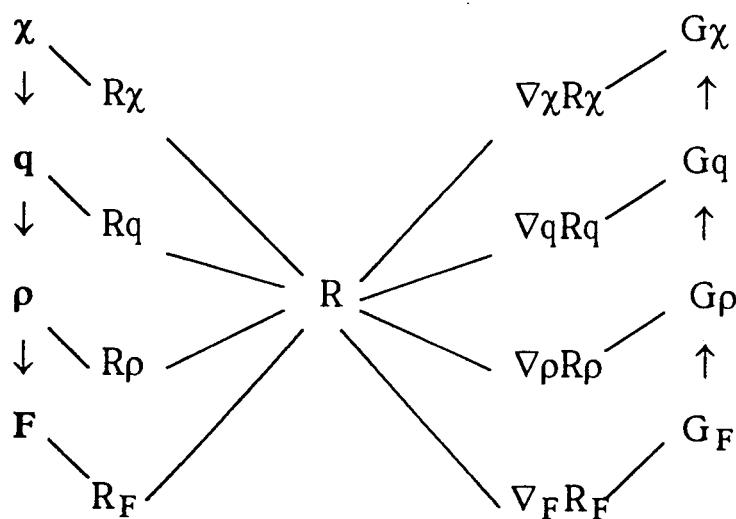


Рис.3. Схема расчета функционала и его градиента.

Здесь  $\nabla_x R_x$  - градиент функции  $R_x$  по переменной  $x$ ,

$$G_F = \nabla_F R_F,$$

$$G_\rho = \nabla_\rho (R_\rho + R_F), \quad (44)$$

$$G_q = \nabla_q (R_q + R_\rho + R_F),$$

$$G_\chi = \nabla_\chi (R_\chi + R_q + R_\rho + R_F)$$

Суть алгоритма быстрого дифференцирования заключается в том, что процесс вычисления минимизируемого функционала сначала представляется в виде цепочки наиболее простых замен переменных. Градиент затем вычисляется как обратная цепочка переходов от градиента по одним переменным к градиенту по предшествующим (в расчете значения функционала) переменным. Здесь каждое преобразование  $x \rightarrow y$  с предыдущего уровня параметров  $x$  к последующему  $y$  влечет за собой преобразование  $\nabla_y f \rightarrow \nabla_x f$  в схеме вычисления градиента.

Как было показано в работе /108/, для преобразования  $x \rightarrow y$  обратное преобразование  $\nabla_x f \rightarrow \nabla_y f (\frac{dy}{dx})$  можно вычислить за ту же цену, что и прямое преобразование. Из этого факта следует, что вычисление всех  $n$  компонент градиента любой функции  $R(x_1, \dots, x_n)$  можно осуществить за то же время  $T_v$ , что и вычисление самого критерия  $T_R$ , т.е.  $T_v \approx T_R$ , а не  $T_v \approx nT_R$ .

#### 2.4. Подготовка атомно-блочной модели: программа GROUPS.

Программа уточнения может работать как с атомными моделями, представляющими собой набор атомов, для которых задаются координаты, изотропные температурные факторы и коэффициенты заполнения, так и с жесткими группами. Жесткая группа включает в себя координаты всех входящих в нее атомов, а ее текущее положение задается углами Эйлера и вектором трансляции. Остальными параметрами жесткой группы являются групповой температурный фактор и коэффициент заполнения.

Для жесткой группы, состоящей из  $N$  атомов, таким образом, можно уточнять всего 8 параметров (углы Эйлера, вектор трансляции, групповой температурный фактор и коэффициент заполнения) вместо  $5N$  параметров модели со свободными атомами. При этом не накладывается никаких ограничений на количество и состав жестких групп. В частности, все атомы могут уточняться как индивидуальные. И наоборот, всю молекулу можно уточнять как жесткое тело. Возможны и все промежуточные варианты. Специальная программа комплекса GROUPS позволяет легко конструировать на основе исходного файла координат атомов файл с атомно-блочной моделью. Этот файл, кроме параметров атомов, содержит информацию о том, каким образом атомы объединяются в жесткие группы. Программа GROUPS предоставляет пользователю некоторые стандартные способы введения таких групп.

Жесткими группами для любых (указанных пользователем) номеров остатков можно объявить:

- боковую цепь, начиная с  $C_{\beta}^k$ -атома,
- пептидное звено  $(CO)^{k-1}-N^k-C_{\alpha}^k$ ,
- пептидное звено с прилегающей боковой цепью  $(CO)^{k-1}-N^k-C_{\alpha}^k-R^k$ .

Верхний индекс обозначает номер остатка. Можно ввести также более крупные жесткие группы атомов, объявляя такой группой произвольный набор пептидных звеньев (с прилегающими боковыми группами или без них). Например, можно объявить жесткой группой  $\alpha$ -спираль,  $\beta$ -слой, главную цепь, аминокислотные остатки внутри белковой глобулы, всю молекулу и т.п. Жесткая группа в процессе уточнения может перемещаться только как твердое тело, т.е. взаимное расположение атомов, входящих в такую группу, не меняется в процессе уточнения.

## **Способы организации жестких групп.**

Возможные способы фиксации жестких групп приведены в следующей таблице:

типы жестких групп	номера остатков
боковые цепи (начиная с Сβ-атомов отдельных остатков)	задаются номера всех остатков, для которых боковые цепи объявляются самостоятельными жесткими группами
пептидные звенья $(CO)^{k-1}\text{-N}^k\text{-C}_\alpha^k$ отдельных остатков	задаются номера всех остатков, для которых пептидные звенья объявляются самостоятельными жесткими группами
целиком аминокислотные остатки	задаются номера всех остатков, которые объявляются самостоятельными жесткими группами
единая группа из пептидных звеньев нескольких остатков (без боковых цепей)	задаются номера всех остатков, участвующих в образовании данной жесткой группы
единая группа из нескольких остатков	задаются номера всех остатков, участвующих в образовании данной жесткой группы

## **Способы задания состава жестких групп.**

*Основной вариант* - необходимые номера остатков разделяются одним или несколькими пробелами:

2 4    6 7  8 9    10 15

*Сокращенная запись* - вместо нескольких подряд идущих номеров могут быть заданы границы участка в виде n1-n2:

2 4    6 -10  15

при этом допустимы пробелы до и после тире.

*Перенос на следующую карту:* если не все требуемые номера остатков умещаются на одной карте, то ставится символ продолжения \* и оставшиеся номера переносятся на следующую карту:

2 4 6 -10

\* 15

### **Способы задания температурных факторов для атомов жестких групп.**

Во входном файле каждый атом имеет индивидуальный температурный параметр  $B_0(j)$ . В атомно-блочной модели для каждой жесткой группы вводится общий параметр  $B_{gr}$ , который может уточняться, а для каждого атома группы определяется индивидуальный температурный параметр  $B_{ind}(j)$ , не меняющийся в процессе уточнения. Текущее значение температурного фактора каждого атома в процессе уточнения определяется как  $B(j) = B_{gr} + B_{ind}(j)$ .

В зависимости от управляющих данных величины  $B_{ind}(j)$  и  $B_{gr}$  определяются следующим образом:

a)  $B_{gr} = \min(B_0(j))$ ,  $B_{ind}(j) = B_0(j) - B_{gr}$  (45)

b)  $B_{gr} = \langle B_0(j) \rangle$ ,  $B_{ind}(j) = 0$  (46)

### **Способы задания коэффициентов заполнения для атомов жестких групп.**

Во входном файле каждый атом имеет индивидуальный коэффициент заполнения  $T_0(j)$ . В атомно-блочной модели для каждой жесткой группы вводится общий коэффициент заполнения  $T_{gr}$ , который может уточняться, а для каждого атома группы определен индивидуальный  $T_{ind}(j)$ , не меняющийся в процессе уточнения. Текущее значение коэффициента заполнения  $T(j)$  каждого атома в процессе уточнения определяется как  $T_{gr} \times T_{ind}(j)$ . В зависимости от управляющих данных величины  $T_{ind}(j)$  и  $T_{gr}$  определяются следующим образом:

a)  $T_{gr} = \max(T_0(j))$ ,  $T_{ind}(j) = T_0(j) / T_{gr}$  (47)

b)  $T_{gr} = \langle T_0(j) \rangle$ ,  $T_{ind}(j) = 1$  (48)

Программа GROUPS содержит около 2000 операторов ФОРТРАНа.

## **2.5. Подготовка справочника Rq стереохимических и энергетических критериев: программа FRRQ.**

Программа FRRQ позволяет ввести в процесс уточнения штрафы за нарушение стандартных стереохимических требований: заданных величин длин валентных связей, валентных и двугранных углов, выход из плоскости атомов, принадлежащих плоским группам, нарушение ориентированного объема заданной четверки атомов (хиральность) и т.п. Кроме того, программа может заносить в справочник Rq перечень пар атомов, между которыми существуют невалентные контакты, для расчета при уточнении модели энергии невалентных взаимодействий.

При этом могут учитываться все виды взаимодействий:

- контакты внутри молекулы,
- контакты атомов молекулы со всеми окружающими ее в кристалле атомами других молекул (в том числе связанных с исходной некристаллографической симметрией),
- всевозможные невалентные контакты.

В справочник Rq заносятся сведения только о тех стереохимических требованиях и невалентных взаимодействиях, которые не фиксируются введенными жесткими группами. Например, если жесткой группой объявлена вся молекула, то при движении ее как твердого тела никаких изменений во взаимном расположении атомов молекулы произойти не может. Поэтому в справочник Rq в этом случае может быть занесена только информация о контакте данной молекулы с соседними. Основные типы штрафных функций, используемых в программе FRRQ, перечислены в формулах (34)-(39). Значения стандартных стереохимических параметров вводятся из внешнего файла и согласованы с последними данными и словарями, используемыми в современных программах уточнения X-PLOR, TNT, PROLSQ, EREF /36,111,112/.

### **Алгоритм отбора пар атомов с невалентными контактами.**

По атомной модели требуется составить справочник попарных контактов для последующего расчета в программе уточнения энергетического функционала (34), причем необходимо учесть взаимодействия атомов не только внутри исходной молекулы, но и с атомами соседних молекул. Пара невалентно связанных атомов заносится в справочник, если атомы находятся на расстоянии  $d_{ij} < d$  друг от друга, где  $d$  - задаваемая пользователем величина, при которой учитываются невалентные взаимодействия. Для отобранный пары атомов в справочник заносятся их номера и константы функционала  $R_{q,1}$  ( $A_{ij}$  и  $B_{ij}$ ).

Работу программы можно разделить на 3 этапа.

1). Для учета взаимодействий между соседними молекулами формируется полный набор, содержащий атомы исходной и всех размноженных по симметриям молекул.

Для этого по входному файлу атомных координат  $\{x_j, y_j, z_j\}$  определяется параллелепипед, ограничивающий исходную молекулу:  $X_{min} \leq x_j \leq X_{max}$ ,  $Y_{min} \leq y_j \leq Y_{max}$ ,  $Z_{min} \leq z_j \leq Z_{max}$ . Затем вычисляется интересующая нас область  $\mathfrak{X}$  такая, что :

$$\begin{aligned} X_{min}-d &\leq x \leq X_{max}+d, \\ Y_{min}-d &\leq y \leq Y_{max}+d, \\ Z_{min}-d &\leq z \leq Z_{max}+d, \end{aligned} \tag{49}$$

где  $d$  - максимальное расстояние, на котором учитываются взаимодействия атомов.

Теперь с помощью заданных уравнений симметрии (кристаллографических и некристаллографических) и использования трансляций по периоду формируется полный набор атомов и из него отбираются только попавшие в интересующую нас область  $\mathfrak{X}$ . Среди этих атомов затем будут выбираться невалентно связанные пары.

2). Чтобы избежать проверки по всей области  $\mathcal{R}$  на взаимодействие каждого атома с каждым, параллелепипед  $\mathcal{R}$  разбивается на клетки размером  $d$ , и для атомов полного набора вычисляется номер клетки, в которую он попадает. После проделанных операций у каждого атома из полного набора появляется следующая информация:

Nat, Nкл, Poi, NS, TX, TY, TZ, x,y,z,

где Nat - номер атома во входном файле;

Nкл - номер клетки, в которую попал атом;

Poi = 1 или 0 (атом  $\in$  исходной или соседней молекуле);

NS - номер уравнения симметрии;

TX,TY,TZ - трансляции вдоль осей, приводящие атом в область  $\mathcal{R}$ ;

x,y,z - координаты атомов после азмножения по симметрии.

Полученный набор атомов упорядочивается по возрастанию номеров клеток, и теперь можно проверять пары атомов на возможное взаимодействие только внутри каждой клетки или между соседними клетками. Если требуется составить справочник только для взаимодействий атомов внутри исходной молекулы, то в буфер отбираются пары с  $Poi=1$  для обоих атомов. Если нужно учесть контакты только между соседними молекулами, то выводится пара атомов, у одного из которых  $Poi=0$ , а у другого  $Poi=1$ . При отборе всех возможных контактов в буфер заносятся пары с любым сочетанием  $Poi$ , кроме ситуации с одновременными значениями  $Poi=0$ .

Далее для каждой отобранный таким образом пары вычисляются расстояния  $d_{ij}$  между атомами, и при  $d_{ij} < d$  пара выводится в рабочий файл с соответствующими константами энергетического функционала  $A_{ij}$  и  $B_{ij}$ .

3). Осталось из отобранных контактирующих атомов удалить валентно связанные пары. Для этого по входной модели и файлу, содержащему описание стандартной стереохимии, составляется список валентно связанных атомов, и за один просмотр этого списка и рабочего файла,

полученного на предыдущем этапе, осуществляется удаление валентных пар.

Программа FRRQ содержит около 4300 операторов ФОРТРАНа. В ней используется один рабочий массив X(N), поэтому ее легко настроить на любой размер доступной оперативной памяти.

### **Замечание.**

В файле со стандартной стереохимией и параметрами невалентных взаимодействий для каждого типа остатка имеется информация о том, какие атомы образуют валентные связи, валентные и двугранные углы, плоские группы и с какими индивидуальными весами их надо включать в уточнение. Например, для аминокислотного остатка LYS соответствующий блок записей выглядит следующим образом:

LYS 22 N C O CA CB CG CD CE NZ HN HA HB1 HB2 HG1 HG2 HD1 HD2

HE1 HE2 HZ1 HZ2 HZ3

BON 22 N CA 1.458 1639.889 C CA 1.525 1342.404 C O 1.231 1480.

C N+ 1.329 3020.408 CA CB 1.53 1480. N HN 0.98 405.

CB CG 1.52 657.778 CG CD 1.52 657.778 CD CE 1.52 657.778

.....

ANG 39 N CA C 111.2 247.886 O C CA 120.8 672.465 O C N+ 123.759.15

CA C N+ 116.2 485.856 C N+ CA+ 121.7 599.823 N CA CB 110.5 672.465

C CA CB 110.1 538.345 C N+ HN+ 120.30. CA N HN 120.35.

.....

PLA 1 5 C O CA N+ CA+ 100.

CHI 1 CA N C CB 2.6 10.

DIH 10 CA- C- N CA 5.180.1 C- N CA C 1.60.4 N CA C N+ 1.30.4

N CA CB CG 1.60.6 CA CB CG CD 1.60.6 CB CG CD CE 1.60.6

CG CD CE NZ 1.60.6 CD CE NZ HZ1 1.60.6 CD CE NZ HZ2 1.60.6

## **2.6. Подготовка справочника Rf рентгеновских критериев:**

### **программа FRREFL.**

Справочник Rf описывает критерии уточнения, выражающиеся непосредственно через модули и фазы структурных факторов, рассчитанных по текущей модели (30)-(33). Программа FRREFL,

входящая в состав комплекса FROG, позволяет преобразовать исходный файл с рентгеновскими данными в справочник Rf, введя указанные пользователем типы критериев. Относительные веса, с которыми входят критерии в суммарную минимизирующую функцию, задаются пользователем. Отбор рефлексов, для которых соответствующие слагаемые включаются в сумму, осуществляется на двух этапах:

- 1) на этапе подготовки справочника Rf отбираются рефлексы заданной зоны разрешения со значениями модулей структурных факторов, превосходящих заданный порог;
- 2) на этапе работы программы уточнения из справочника Rf отбираются лишь рефлексы с  $t_1 \leq F_{\text{calc}}/F_{\text{obs}} \leq t_2$ , где  $t_1$  и  $t_2$  задаются в управляемых данных.

В уточняемых критериях (30)-(33) может использоваться одна из следующих весовых схем:

$$1) w_s = |s|^{-\alpha}, \quad (50)$$

где  $\alpha$  - заданный параметр (при  $\alpha=0$  все веса  $w=1$ );

$$2) w_s = 1/\sigma_s, \quad (51)$$

где  $\sigma_s$  - индивидуальное для данного рефлекса значение, взятое из исходного файла с рентгеновскими данными.

Чтобы при уточнении использовать R-free фактор в качестве критерия оценки качества модели (см. п.1.5.), во входном файле структурных факторов выбирается случайным образом подмножество рефлексов (например, 10% от полного набора,) и они заносятся в справочник Rf со специальным признаком "free". В дальнейшем уточнение проводится по оставшейся части рефлексов, а по рефлексам с признаком "free" производится только расчет R-фактора для контроля за процессом уточнения.

Программа FRREFL содержит около 740 операторов ФОРТРАНа.

## **2.7. Обработка результатов уточнения: программы FRSTAT и FRSTGR.**

В комплексе FROG имеются программы FRSTAT и FRSTGR сбора статистики, позволяющие получить информацию о нарушении стереохимических соотношений в имеющейся модели молекулы белка. Программы сообщают по каждому из типов контролируемых параметров стандартное значение этого параметра, число контролируемых соотношений, среднее и максимальное отклонение от стандартного значения. Кроме того, рассчитываются гистограммы распределения отклонений от заданных значений.

В программе FRSTAT распечатываются следующие величины:

- а) отклонения длин валентных связей, валентных и двугранных углов от стандартных стереохимических величин;
- б) отклонения атомов, входящих в плоские группы, от плоскостей;
- с) нарушение хиральности.

В программе FRSTGR, кроме перечисленной информации, еще выдаются сообщения о невалентных взаимодействиях. Различия этих программ состоят в том, что FRSTAT собирает статистику по всем атомам модели и для нее не требуется справочник Rq. FRSTGR собирает статистику только по тем стереохимическим соотношениям, по которым проводилось уточнение и которые отражены в справочнике Rq.

Программа FRSTAT содержит около 3300 операторов, а FRSTGR - 1600 операторов ФОРТРАНа.

## **2.8. Версия комплекса FROG для персокомпьютеров.**

Комплекс программ FROG несколько лет использовался в его первоначальном варианте на ЕС-1055М и на IBM PC 486/487 компьютерах в работах с реальными объектами (см. гл.4). В 1994 г. комплекс был расширен некоторыми вспомогательными программами и частично обновлен /113/. В частности, был полностью обновлен файл стандартной стереохимической информации на основе работ /36,111,114/. Для облегчения управления процессом уточнения и

задания управляющих данных Ивановым М.Е. /115/ была создана программная “оболочка” на IBM PC (рис.4), позволяющая пользователю войти в среду комплекса FROG и в интерактивном графическом режиме проводить все необходимые действия: подготовить и задать управляющие данные и файлы, запустить программы на счет, получить листинги и результаты уточнения, т.е. организовать любую необходимую последовательность действий в удобном пользователю режиме работы, не выходя из среды комплекса FROG.

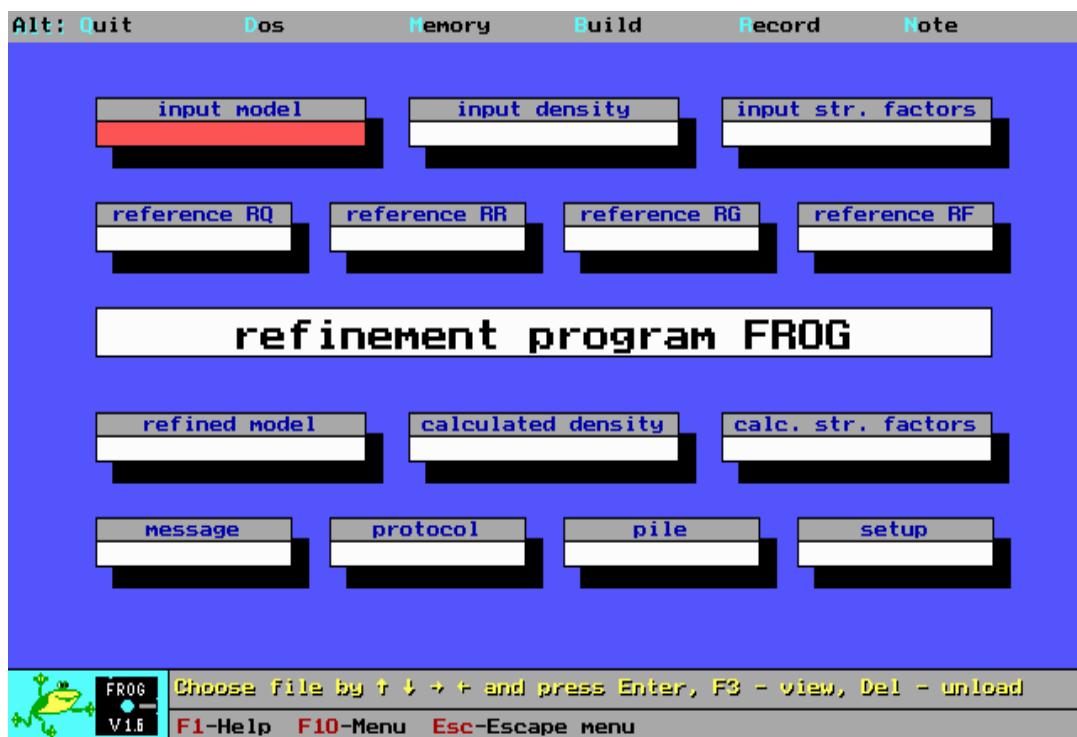


Рис.4. Комплекс программ FROG на IBM PC.

## ГЛАВА 3

### Серия графических программ на IBM PC для задач рентгеноструктурного анализа.

На всех этапах расшифровки структуры методами рентгеноструктурного анализа используются компьютерные вычисления и программы самого широкого профиля. В процессе развития и усовершенствования программного обеспечения для задач рентгеновской кристаллографии появляются все более сложные многофункциональные комплексы программ, рассчитанные на мощные дорогие суперкомпьютеры и графические станции для анализа и интерпретации результатов. В то же время постоянно растущие возможности персональных компьютеров (как по быстродействию, так и по памяти) позволяют выполнять большую часть исследований структур, используя гораздо более дешевое и легко доступное оборудование. Отсюда возникает потребность в разработке и развитии кристаллографических программ, использующих особенности РС.

В данной главе описывается серия прикладных программ для задач рентгеноструктурных исследований, разработанных в Институте математических проблем биологии РАН (Пущино) в лаборатории кристаллографии макромолекул /116-121/.

Программы могут работать на машинах IBM PC/AT, PS/2 или совместимых с ними, в системе MS-DOS версии 3.3+, с EGA, VGA графическими адапторами, сами определяют размер доступной оперативной памяти.

Все программы этой серии имеют единообразный графический интерактивный интерфейс и взаимодействуют друг с другом посредством файлов фиксированных форматов (см.п.3.3.). При этом имеются средства контроля за используемыми файлами; в любое время можно обратиться к подсказке "Help".

Набор возможных операций в каждой программе достаточно разнообразен для решения поставленных задач, но это не делает их

громоздкими и запутанными при их освоении и эксплуатации, что, к сожалению, характерно для некоторых больших программных комплексов.

В описываемую серию входят программы FFT, FAN, CAN и FANS, позволяющие рассчитать синтезы Фурье и Паттерсона любого типа; вывести синтезы на экран монитора в режиме моно или стерео и провести их визуальное исследование; сравнить несколько синтезов, накладывая один на другой; одновременно с синтезом вывести на экран атомную модель исследуемого объекта и провести с ней некоторые манипуляции (вращение, трансляцию) и т.д.

### **3.1. Программа FFT - расчет дискретного трехмерного преобразования Фурье.**

Программа FFT (Fast Fourier Transform) предназначена для осуществления дискретного трехмерного преобразования Фурье на IBM PC - совместимых компьютерах и ориентирована на задачи кристаллографии. Программа позволяет строить синтезы Фурье и Паттерсона, разностные синтезы Фурье и Паттерсона и комбинированные синтезы любого вида; имеет возможность создавать выводной синтез в одном из нескольких форматов. Программа имеет удобный для пользователя графический интерактивный режим ввода параметров, необходимых для расчета синтеза. Имеется возможность расчета синтеза в любой пространственной группе, в одной из 6 ориентаций кристаллографических осей. При расчете происходит контроль за полнотой входного набора структурных факторов.

#### **1. Дискретное трехмерное преобразование Фурье.**

На различных этапах работы по расшифровке структуры методом рентгеноструктурного анализа возникает потребность вычисления трехмерного преобразования Фурье при расчете функции распределения электронной плотности  $\rho(r)$  исследуемого объекта.

Функция  $\rho(\mathbf{r})$  - периодическая, т.е. существует три некомпланарных вектора  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$  таких, что  $\rho(\mathbf{r}) = \rho(\mathbf{r+a}) = \rho(\mathbf{r+b}) = \rho(\mathbf{r+c})$ . Если ввести декартовую систему координат таким образом, чтобы оси были параллельны векторам  $\mathbf{a}$ ,  $\mathbf{b}$ ,  $\mathbf{c}$ , а длины этих векторов  $A$ ,  $B$ ,  $C$  принять за единицы вдоль соответствующих осей, то от функции  $\rho(\mathbf{r})$  в абсолютной системе координат можно перейти к функции  $\rho(x,y,z)$  в относительной системе координат, где элементарной ячейкой вместо параллелепипеда со сторонами  $A$ ,  $B$ ,  $C$  будет единичный куб  $0 \leq x \leq 1$ ,  $0 \leq y \leq 1$ ,  $0 \leq z \leq 1$ . Значения функции в такой элементарной ячейке полностью определяют значения функции на всем пространстве в силу периодичности  $\rho(\mathbf{r})$ :

$$\rho(x,y,z) = \rho(x+1,y,z) = \rho(x,y+1,z) = \rho(x,y,z+1) \quad (52)$$

Итак, требуется вычислить функцию

$$\rho(x,y,z) = \sum_{hkl} F_{hkl} e^{-2\pi i(hx+ky+lz)} \quad (53)$$

где  $F_{hkl}$  - заданные числа (в общем случае комплексные). Величины  $F_{hkl}$  могут быть структурными факторами, интенсивностями отражений (при расчете функции Паттерсона), а в общем случае - произвольными коэффициентами Фурье. Хотя функция  $\rho(x,y,z)$  определена на всем пространстве, на практике ее вычисляют на равномерной трехмерной сетке, т.е. разбивают ребра элементарной ячейки, идущие вдоль осей  $x$ ,  $y$ ,  $z$ , на  $NX$ ,  $NY$ ,  $NZ$  частей соответственно, и вычисляют величины  $\rho_{jx,jy,jz} = \rho\left(\frac{jx}{NX}, \frac{jy}{NY}, \frac{jz}{NZ}\right)$  для  $jx=0,1,\dots,NX-1$ ;  $jy=0,1,\dots,NY-1$ ;  $jz=0,1,\dots,NZ-1$  по формуле (54):

$$\rho_{jx,jy,jz} = \sum_{h=0}^{NX-1} \sum_{k=0}^{NY-1} \sum_{l=0}^{NZ-1} F_{hkl} e^{-2\pi i (h \frac{jx}{NX} + k \frac{jy}{NY} + l \frac{jz}{NZ})} \quad (54)$$

Вычисления осуществляются на базе алгоритма Кули-Тьюки /75/ быстрого преобразования Фурье, и используется вычислительное ядро комплекса программ, разработанных Л.Тен Эйком /76,77/ и поставленных в ИМПБ РАН Луниным В.Ю. для ЕС ЭВМ /109/. Подробное описание алгоритма быстрого преобразования Фурье и организации вычислений дается в работах /75-77,109/.

При вычислениях размеры массивов коэффициентов Фурье  $F_{nkl}$  и результирующего синтеза  $r(x,y,z)$  могут оказаться столь велики, что не смогут разместиться целиком в оперативной памяти. Программа сама определяет область допустимой для работы оперативной памяти и рассчитывает синтез порциями, используя рабочий файл для хранения промежуточных результатов. Такой алгоритм работы позволяет осуществлять расчет практически любого по размеру синтеза, единственным условием является размещение в памяти целиком по крайней мере одного сечения.

## **2. Схема работы программы FFT.**

После запуска программа переходит в графический интерактивный режим работы, позволяя пользователю задавать и корректировать все параметры и контролировать процесс расчета синтеза. В любой момент можно прервать работу программы (клавиша "Esc") или получить подсказку Help (клавиша F1).

Следующая схема отражает основные шаги, выполняемые при работе с программой:



### **3. Начало работы с программой.**

Запуск программы возможен двумя способами:

- a) FFT ;
- б) FFT <имя файла параметров> (например, FFT EXAMPLE.PAR).

В первом случае стартовые значения всех параметров для расчета синтеза (размеры элементарной ячейки; границы, в которых будет рассчитан синтез; ориентация синтеза; пространственная группа; имена входного и выходного файлов и т.п.) берутся из стандартного файла FFT.PAR. Далее в ходе работы эти параметры могут быть изменены пользователем.

Во втором случае стартовые значения всех параметров берутся из указанного при запуске файла и также могут быть далее изменены

пользователем в ходе работы. Файл параметров может автоматически создаваться при завершении работы программы (см.п. 3.1.6).

После запуска программы пользователь задает имя входного файла, содержащего структурные факторы, и выводного файла с рассчитанным синтезом. Если программа не нашла файла с введенным именем, имя файла набрано неверно или при создании файла была допущена ошибка в формате, программа выдает предупреждение об ошибке, и тогда можно набрать новое имя файла либо прекратить работу ("Esc").

Если указанный входной файл обнаружен и имеет требуемый формат, на экране появляется информация из заголовков файла и количество позиций в записях по каждому рефлексу. Можно начать работу с введенным файлом либо ввести другое имя файла.

Входной файл структурных факторов имеет простой формат (п.3.3.2.) и содержит всю информацию, необходимую для расчета синтеза: индексы рефлексов  $h,k,l$ ; модули, фазы и т.д.

#### **4. Ввод параметров для расчета синтеза.**

Это основная сцена программы в интерактивном режиме. Существует три группы параметров, которые должны быть заданы для дальнейшего расчета синтеза.

1. Тип синтеза и правила для отбора коэффициентов Фурье.
2. Параметры элементарной ячейки и пространственная группа.
3. Параметры, шкалирование и формат выводного синтеза.

С помощью переключателя **Tab** ( $\leftrightarrow$ ) можно переходить от одной группы данных к другой.

**В первой группе** параметров можно выбрать один из 5 вариантов расчета синтеза:

1. Синтез Фурье с коэффициентами

$$W_s \times (K \times F_s) \times {}^{i\varphi} s \quad (55)$$

2. Синтез Паттерсона с коэффициентами

$$W_s \times (K \times F_s)^2 \quad (56)$$

### 3. Разностный синтез Фурье с коэффициентами

$$W_s \times (K1 \times F1_s - K2 \times F2_s) \times e^{i\varphi_s} \quad (57)$$

### 4. Разностный синтез Паттерсона с коэффициентами

$$W_s \times (K1 \times F1_s - K2 \times F2_s)^2 \quad (58)$$

### 5. Синтез в общем виде

$$\sum_{j=1}^N W_s^j \times (K_j \times F_s^j) \times e^{i\varphi_s^j} \quad (59)$$

Здесь  $s$  - трехмерный индекс (миллеровские индексы рефлекса);

$W_s$ ,  $F_s$ ,  $\varphi_s$  - определены для каждого рефлекса во входном файле;

$K$ ,  $K1$ ,  $K2$  - определяются пользователем в процессе работы.

При расчете первых четырех типов синтезов задаются границы разрешения  $D_{MIN}$  и  $D_{MAX}$  (в Å), при котором будет рассчитан синтез; номера позиций во входном файле, из которых берутся значения модулей  $F_s$ , фаз  $\varphi_s$  и весов  $W_s$ . Здесь же определяются пользователем шкальные коэффициенты  $K$ ,  $K1$ ,  $K2$  (рис.5).

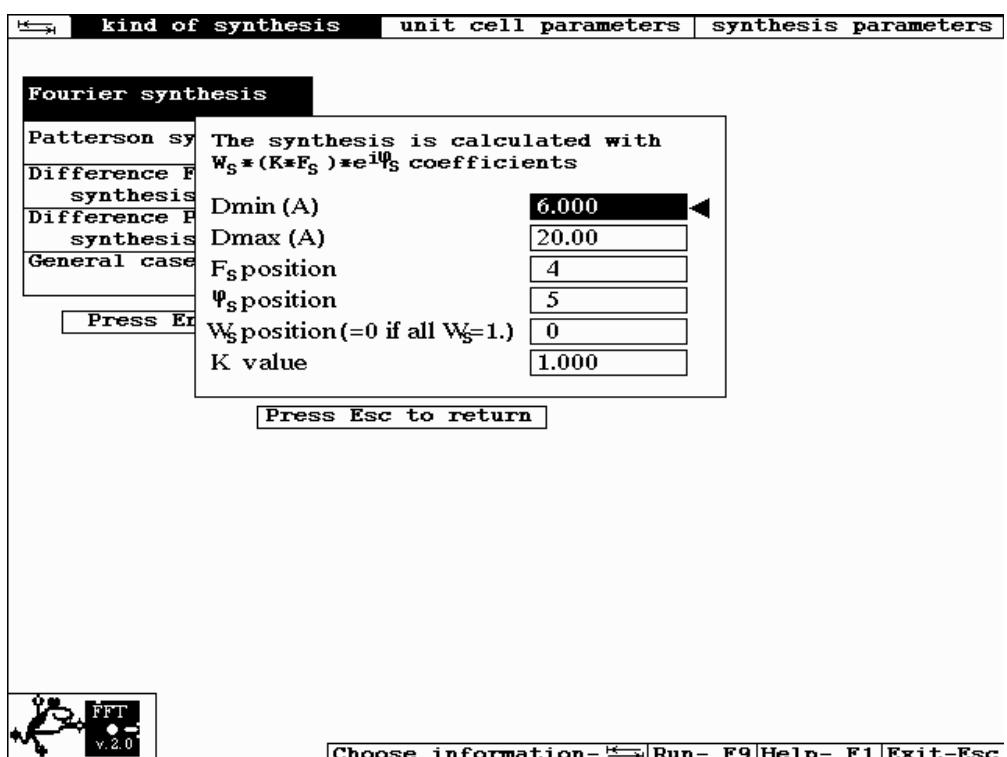


Рис.5. Задание правила формирования коэффициентов Фурье.

При расчете синтеза общего вида правило, по которому вычисляются коэффициенты для синтеза, может быть разным для разных зон по разрешению. Пользователь задает количество зон разрешения NZON ( $\leq 5$ ). Для каждой зоны задается разрешение DMIN, DMAX и число слагаемых N в сумме (59). Для каждого слагаемого вводятся номера позиций, содержащих  $F_s$ ,  $\varphi_s$ ,  $W_s$  и шкальный коэффициент K.

**Во второй группе** данных задаются параметры элементарной ячейки A, B, C,  $\alpha$ ,  $\beta$ ,  $\gamma$  (в Å и градусах) и пространственная группа кристалла (рис.6). Если пользователь собирается работать в пространственной группе, которая отсутствует в общем списке, он может сам задать уравнения симметрии, подведя курсор к окошку "Other" и нажав клавишу "Enter".

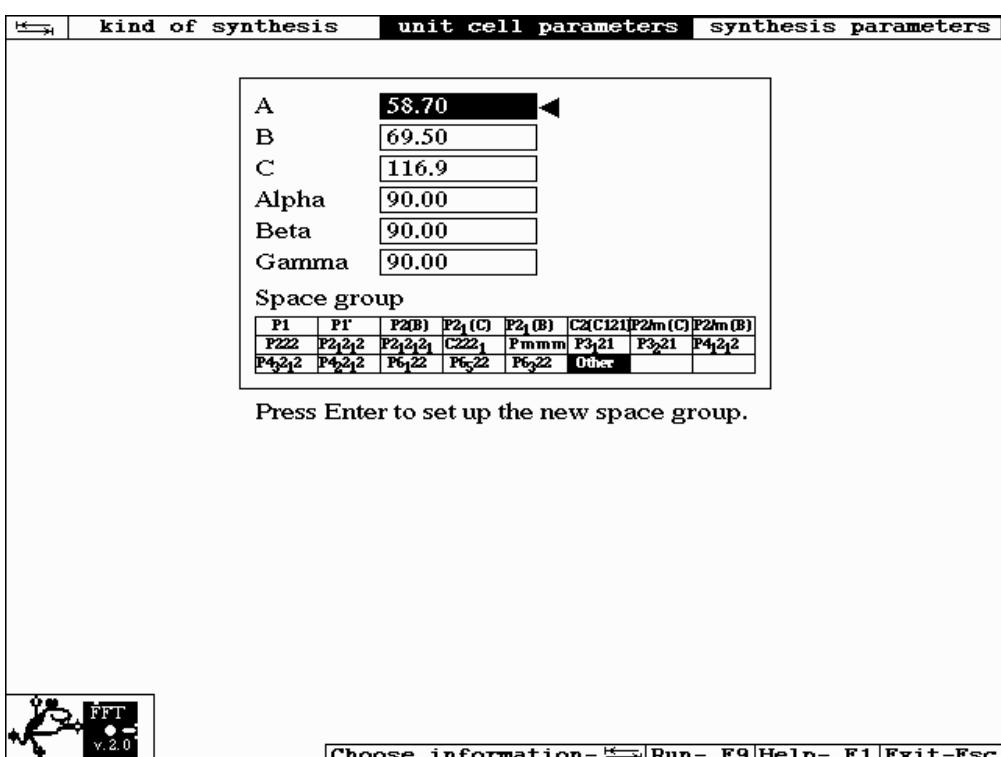


Рис.6. Задание параметров ячейки и пространственной группы.

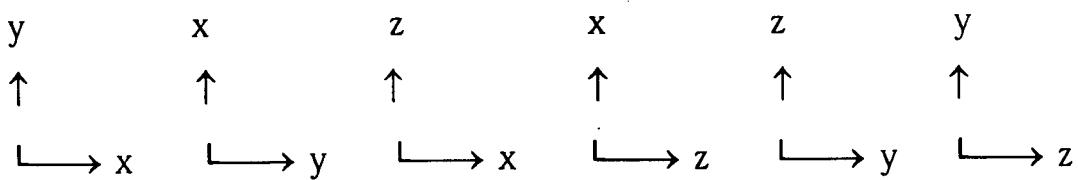
При этом, оказавшись в режиме задания произвольной пространственной группы, нужно ввести количество уравнений

симметрии и задать матрицы симметрии и вектора трансляции. Коэффициенты матриц могут быть равны 0, 1, -1, а элементы векторов - 0, 1/2, 1/3, 1/4, 1/6, 2/3, 2/4, 2/6, 3/4, 3/6, 4/6, 5/6 (рис.7).

		kind of synthesis	unit cell parameters	synthesis parameters
Input the necessary coefficients for the symmetry transforms.				
<b>8</b> symmetry relations will be used for new reflections generating				
1	X' = <input type="text" value="1"/> *X + <input type="text" value="0"/> *Y + <input type="text" value="0"/> *Z + <input type="text" value="0"/> Y' = <input type="text" value="0"/> *X + <input type="text" value="1"/> *Y + <input type="text" value="0"/> *Z + <input type="text" value="0"/> Z' = <input type="text" value="0"/> *X + <input type="text" value="0"/> *Y + <input type="text" value="1"/> *Z + <input type="text" value="0"/>	4 X' = <input type="text" value="0"/> *X + <input type="text" value="1"/> *Y + <input type="text" value="0"/> *Z + <input type="text" value="1/2"/> Y' = <input type="text" value="-1"/> *X + <input type="text" value="0"/> *Y + <input type="text" value="0"/> *Z + <input type="text" value="1/2"/> Z' = <input type="text" value="0"/> *X + <input type="text" value="0"/> *Y + <input type="text" value="1"/> *Z + <input type="text" value="3/4"/>		
2	X' = <input type="text" value="-1"/> *X + <input type="text" value="0"/> *Y + <input type="text" value="0"/> *Z + <input type="text" value="0"/> Y' = <input type="text" value="0"/> *X + <input type="text" value="-1"/> *Y + <input type="text" value="0"/> *Z + <input type="text" value="0"/> Z' = <input type="text" value="0"/> *X + <input type="text" value="0"/> *Y + <input type="text" value="1"/> *Z + <input type="text" value="1/2"/>	5 X' = <input type="text" value="0"/> *X + <input type="text" value="1"/> *Y + <input type="text" value="0"/> *Z + <input type="text" value="0"/> Y' = <input type="text" value="1"/> *X + <input type="text" value="0"/> *Y + <input type="text" value="0"/> *Z + <input type="text" value="0"/> Z' = <input type="text" value="0"/> *X + <input type="text" value="0"/> *Y + <input type="text" value="-1"/> *Z + <input type="text" value="0"/>		
3	X' = <input type="text" value="0"/> *X + <input type="text" value="-1"/> *Y + <input type="text" value="0"/> *Z + <input type="text" value="1/2"/> Y' = <input type="text" value="1"/> *X + <input type="text" value="0"/> *Y + <input type="text" value="0"/> *Z + <input type="text" value="1/2"/> Z' = <input type="text" value="0"/> *X + <input type="text" value="0"/> *Y + <input type="text" value="1"/> *Z + <input type="text" value="1/4"/>	6 X' = <input type="text" value="0"/> *X + <input type="text" value="-1"/> *Y + <input type="text" value="0"/> *Z + <input type="text" value="0"/> Y' = <input type="text" value="1"/> *X + <input type="text" value="0"/> *Y + <input type="text" value="0"/> *Z + <input type="text" value="0"/> Z' = <input type="text" value="0"/> *X + <input type="text" value="0"/> *Y + <input type="text" value="-1"/> *Z + <input type="text" value="1/2"/>		
 <span>PgDn - next page</span> <span>Help - F1   Exit - Esc</span>				

Рис.7. Задание произвольных матриц симметрии.

**В третьей группе** данных вводится заголовок для выводного синтеза, разбиение элементарной ячейки NX, NY, NZ; границы для расчета синтеза MINX, MAXX, MINY, MAXY, MINZ, MAXZ и ориентация (рис.8). Кроме того, выбирается тип шкалирования и формат выводного синтеза. Можно задать одну из 6 ориентаций выводного синтеза:



Ориентация определяет порядок индексов, в котором выводится трехмерный массив, содержащий рассчитанный синтез. Наиболее быстро меняется индекс, соответствующий оси абсцисс, затем оси ординат. В выводной файл синтез заносится сечениями по третьей оси.

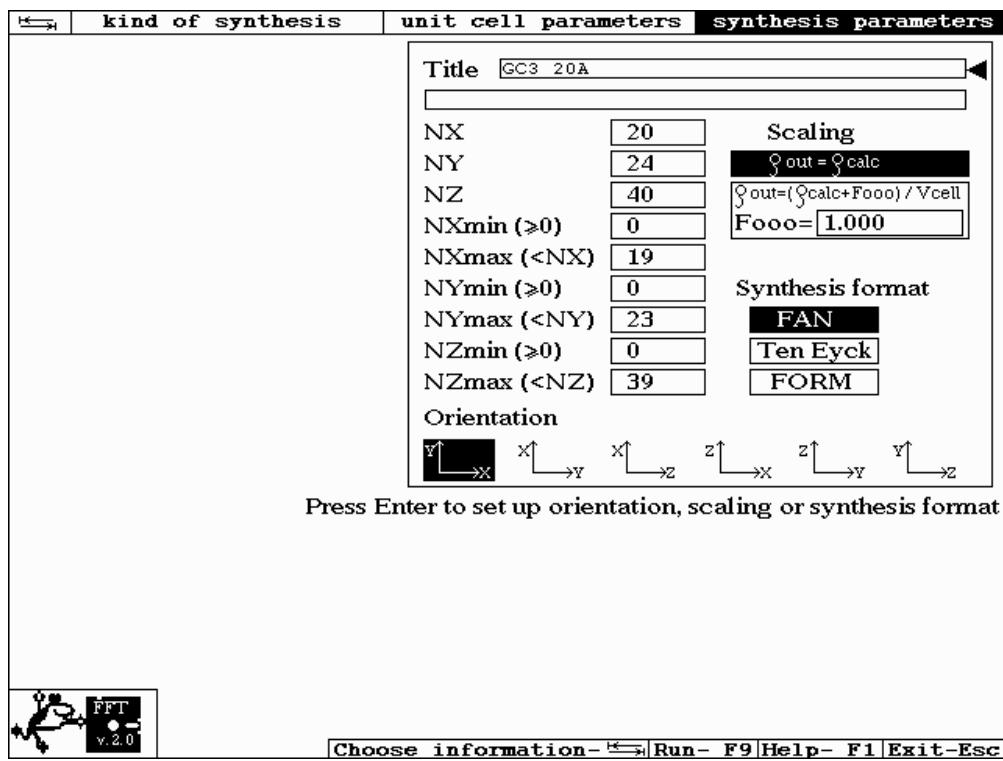


Рис.8. Задание разбиения синтеза, его ориентации и формата.

### Выбор способа шкалирования.

Синтез может быть рассчитан в одном из двух вариантов:

а) без специального шкалирования, т.е.  $\rho_{out} = \rho_{calc}$

б) со шкалированием  $\rho_{out} = (F_{000} + \rho_{calc}) / V_{cell}$ ,

где  $V_{cell}$  - объем элементарной ячейки

$F_{000}$  - шкальный коэффициент.

В этом случае программа потребует ввести параметр  $F_{000}$ .

### Выбор формата выводного синтеза.

Синтез может быть получен в одном из трех форматов: **FAN**, **Ten Eyck** или **FORM** (см.п.3.3.).

- 1) Синтез в специальном формате **FAN** можно непосредственно использовать для дальнейшего визуального просмотра и анализа, подавая его на вход программе FAN, CAN или FANS.
- 2) Файл с синтезом в формате **Ten Eyck** - бесформатный.
- 3) Форматизированный файл в формате **FORM**, удобный для работы в других операционных системах. Его можно использовать как входной

для работы с программой FAN или CAN, но он занимает гораздо больше места на диске и поэтому неудобен в работе.

### **Переход к расчету синтеза.**

После установки всех необходимых параметров нажатием клавиши F9 можно запустить расчет синтеза. При этом на экране появятся все характерные параметры, выбранные для расчета синтеза. Может оказаться, что число точек разбиения элементарной ячейки было выбрано неудачно, тогда программа автоматически корректирует эти величины и выдает предупреждение. При выборе чисел разбиения существует несколько ограничений:

a)  $NX \geq 2 \times h_{\max} + 2; 0 \leq MINX \leq MAXX \leq NX-1;$

$$NY \geq 2 \times k_{\max} + 2; 0 \leq MINY \leq MAXY \leq NY-1;$$

$$NZ \geq 2 \times l_{\max} + 2; 0 \leq MINZ \leq MAXZ \leq NZ-1;$$

- б)  $NX, NY, NZ$  - четные, должны разлагаться в произведение не более чем 14 простых сомножителей и среди них не должно быть множителей  $\geq 19$ .

Если в итоге выбранные параметры удовлетворяют пользователю, он может продолжить счет. В противном случае он может вернуться в предыдущую сцену выбора параметров и изменить их.

### **5. Контроль за полнотой набора входных данных.**

В процессе расчета синтеза осуществляется контроль за полнотой входного набора данных. На экран выводятся три сечения обратного пространства ( $h=0; k=0; l=0$ ), на которых отражено наличие соответствующих рефлексов (рис.9). Кроме того, количество рефлексов, имеющихся во входном файле, сравнивается с теоретически возможным количеством рефлексов в полном наборе соответствующего разрешения. При подсчете учитывается общее количество рефлексов после размножения по уравнениям кристаллографической симметрии.

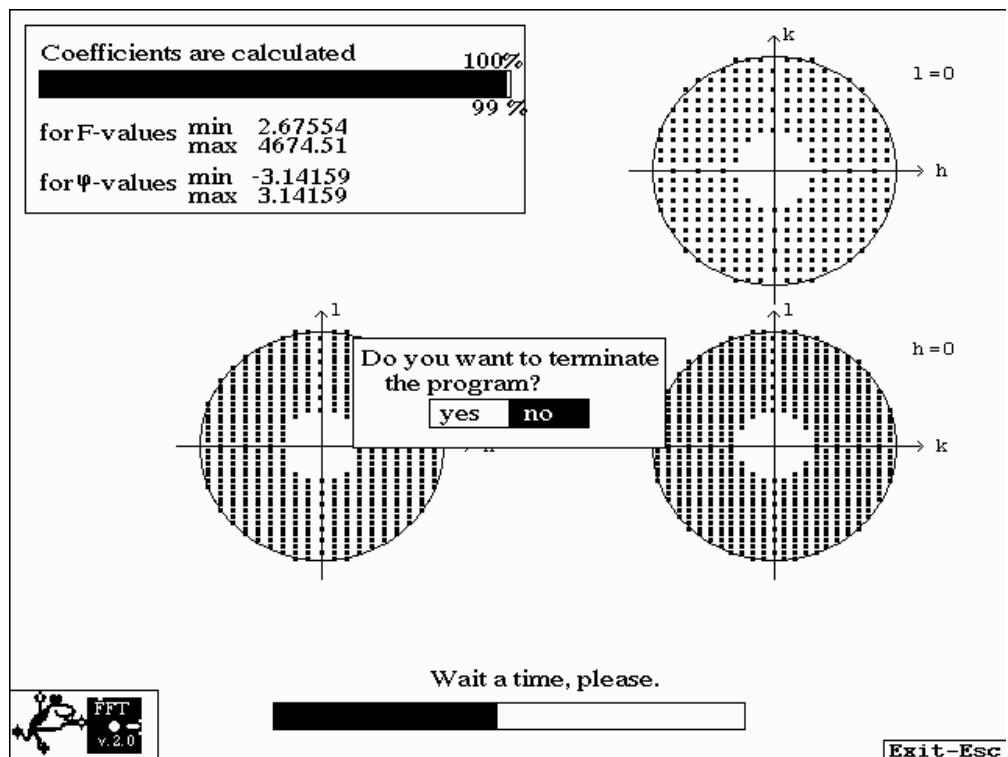


Рис.9. Контроль за полнотой набора данных.

## 6. Создание файла параметров по окончании работы.

В процессе работы программа выдает запрос о том, следует ли сохранить текущие параметры (имена входного и выводного файлов, тип синтеза, способ выбора коэффициентов Фурье, параметры элементарной ячейки, пространственная группа, разбиение и ориентация синтеза и т.д.) в специальном файле. При дальнейших запусках программы FFT можно подавать этот файл на вход (см.п. 3.1.2.), облегчив себе дальнейшую работу по установке параметров для расчета синтеза.

### 3.2. Программы FAN, CAN и FANS - визуальный анализ

#### синтезов электронной плотности и атомной модели.

Информация о многих физических свойствах трехмерных объектов, выражаемых посредством скалярных полей, часто представляется в виде набора значений функции, вычисленных на сетке в трехмерном пространстве. В задачах кристаллографии, в частности, исследуемой функцией является функция распределения электронной плотности

$\rho(r)$ , рассчитанная на трехмерной равномерной сетке (см. п.3.1.1), которая подается во входном файле программы как массив значений. Функция выводится на экран монитора в виде изолиний (линий, соединяющих точки с равными значениями функции), отвечающих ее двумерным сечениям.

Предполагается, что  $\rho(r)$  может быть периодической по трем направлениям в пространстве, и тогда значения функции, заданные в единичной элементарной ячейке (в относительной системе координат, см.п.3.1.1.), определяют область значений  $\rho(r)$  на всем пространстве.

Итак, исследуемая функция  $\rho(r)$  задана в виде массива чисел

$$\rho_{ijk} = \rho\left(\frac{i}{NX}, \frac{j}{NY}, \frac{k}{NZ}\right),$$

где NX, NY, NZ - разбиение элементарной ячейки,

$$0 \leq \min x \leq i \leq \max x < NX;$$

$$0 \leq \min y \leq j \leq \max y < NY;$$

$$0 \leq \min z \leq k \leq \max z < NZ.$$

Заметим, что  $\rho(r)$  может быть задана не во всей элементарной ячейке, а лишь в ее части, определяемой границами  $\min x, \max x, \min y, \max y, \min z, \max z$ .

Серия программ FAN, CAN, FANS предназначена для визуального анализа функций трех переменных (распределения электронной плотности) на IBM PC-совместимых компьютерах.

**Программа FAN (scalar Field ANalysis)** позволяет :

- получить на экране дисплея картину линий уровня в одном из двумерных сечений;
- получить суммарную картину линий уровня в нескольких сечениях;
- легко манипулировать с набором сечений, отображаемых на экране (добавлять и удалять отдельные сечения);

- оперативно изменять масштаб изображения и сдвигать "окно" в пространстве (при большом увеличении не вся интересующая нас область может уместиться на экране);
- легко варьировать набор и цвет изображаемых линий уровня;
- определять относительные координаты любой точки экрана;
- в любое время запросить подсказку ("Help").

**Программа FANS (scalar Field Analysis Stereo)** позволяет:

- выполнять все функции программы FAN в стереорежиме;
- выводить на экран стереоизображение скелетной модели в том же масштабе, что и изображение синтеза;
- сдвигать и вращать изображение модели как жесткого тела относительно изображения анализируемой функции;

**Программа CAN (Comparative ANalysis)** позволяет получить на экране дисплея по отдельности или одновременно изображение трех объектов:

- двух функций распределения электронной плотности, зависящих от трех пространственных переменных (далее будем называть одну из этих функций основной, а вторую - дополнительной);
- атомной модели белковой молекулы.

Атомная модель задается набором координат атомов и изображается в виде проекции на экран скелетной модели, в которой атомы изображаются шариками, а химические связи - отрезками.

Программа CAN дает возможность:

a) для основной анализируемой функции

- получить на экране дисплея картину линий уровня в одном из двухмерных сечений;
- получить наложенную картину линий уровня в нескольких сечениях;
- легко манипулировать с набором сечений, отображаемых на экране (добавлять и удалять отдельные сечения);
- легко варьировать набор и цвет изображаемых линий уровня;

- оперативно изменять масштаб изображения и сдвигать "окно" в пространстве;
- выводить функцию сечениями по любой из трех координатных осей.

#### **б) для дополнительной функции**

- осуществлять те же операции, что и для основной функции;
- сдвигать изображение дополнительной функции относительно изображения основной и отражать его зеркально в координатных плоскостях (переходить к энантиоморфу);
- менять в процессе работы дополнительную функцию, считывая ее из указанного пользователем файла.

#### **в) для атомной модели**

- выводить на экран изображение скелетной модели в том же масштабе, что и изображение анализируемых функций;
- сдвигать и вращать изображение модели как жесткого тела относительно изображения анализируемой функции;
- выводить помимо изображения атомной модели молекулы белка изображения молекул, связанных с исходной заданными преобразованиями симметрии.

Программы ориентированы на задачи кристаллографии. Они обеспечивают автоматическое доопределение заданной функции по периодичности на все пространство и возможность работы с косоугольными элементарными ячейками, что позволяет изучать также кристаллографическое окружение исследуемого объекта. В любой момент существует возможность получить подсказку "Help".

### **1. Запуск программ.**

Запуск программ возможен в двух режимах:

- имя программы
- имя программы <имя файла параметров>

В первом случае стартовые значения некоторых внутренних параметров (имена входных файлов, масштаба и координаты центра изображения, значения уровней изолиний и т.п.) устанавливаются

стандартным образом. Далее в ходе работы они могут быть изменены пользователем.

Во втором случае стартовые значения внутренних параметров берутся из заданного файла и также могут быть далее изменены пользователем в ходе работы. Файл параметров может автоматически создаваться при завершении работы программы (см. п.3.2.10).

На различных стадиях работы пользователь может видеть на экране информацию разных типов и иметь в распоряжении разный набор возможных действий. Эти ситуации, определяемые характером отображенной на экране информации и набором возможных действий, называются далее "сценами". Ниже дается описание каждой из сцен.

## **2. Выбор входного файла с синтезом электронной плотности.**

В эту сцену пользователь попадает сразу после запуска программы. Для выбора файла, содержащего значения анализируемой функции, следует набрать на клавиатуре имя файла (оно продублируется на экране) и нажать клавишу "Enter". Для коррекции ошибок в процессе ввода имени может быть использована клавиша "Bsp". Неправильное задание имени файла или недопустимый формат указанного файла может привести к отказу от работы, этом случае можно набрать новое имя файла либо прекратить работу ("Esc").

### **Нормальный ввод**

Если указанный файл обнаружен и имеет требуемый формат, на экране появляется информация об объекте и области, в которой рассчитана исследуемая функция. Предполагается, что она рассчитана на равномерной сетке в параллелепипеде (вообще говоря, косоугольном) со сторонами A1, A2, A3 и противолежащими этим сторонам плоскими углами  $\alpha_1$ ,  $\alpha_2$ ,  $\alpha_3$ . Функция может быть периодически продолжена из этого параллелепипеда ("элементарной ячейки") на все пространство.

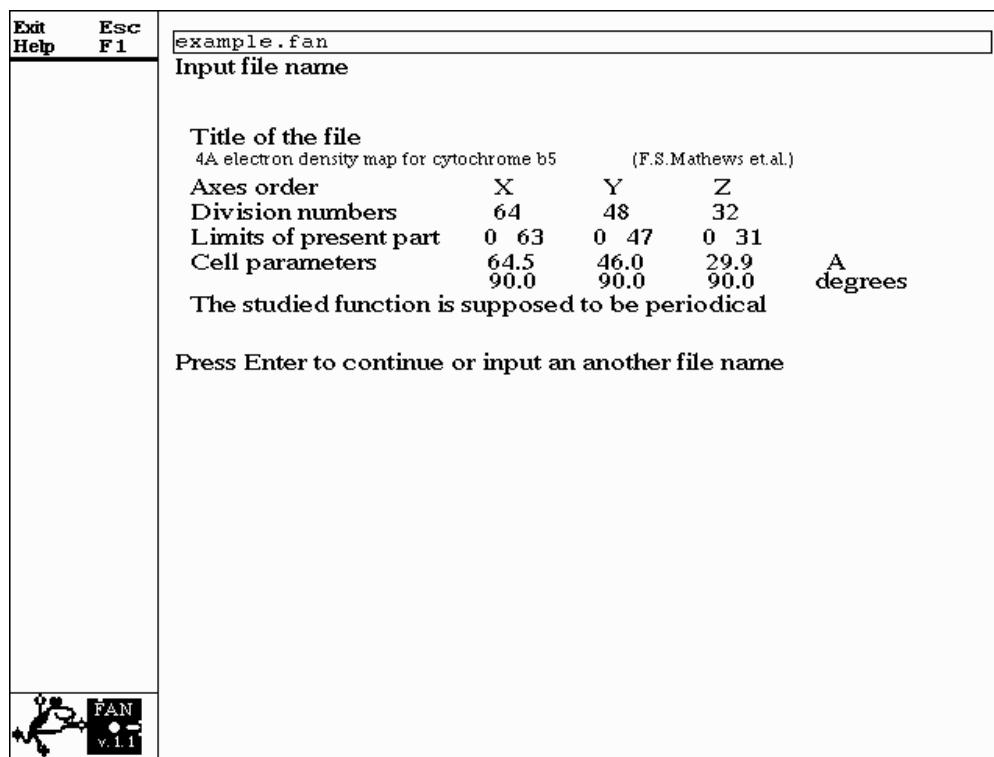


Рис.10. Информация о выводимом на экран синтезе.

На экран выводится следующая информация (рис.10):

- заголовок, идентифицирующий изучаемую функцию;
- величины сторон A<sub>1</sub>, A<sub>2</sub>, A<sub>3</sub>, название единицы их измерения и величины углов  $\alpha_1$ ,  $\alpha_2$ ,  $\alpha_3$  (в градусах);
- названия осей A<sub>1</sub>, A<sub>2</sub>, A<sub>3</sub>, привычные пользователю (например, a, b, c или X, Y, Z);
- числа делений N<sub>1</sub>, N<sub>2</sub>, N<sub>3</sub> вдоль сторон элементарной ячейки при выборе сетки.

На экран могут быть выведено изображение объекта в сечениях, параллельных одной из трех координатных плоскостей. Пользователю предоставляется возможность выбрать один из 6 требуемых вариантов ориентации изображения.

### **Формат входного файла**

Входной файл с синтезом содержит значения изучаемой функции, вычисленные в узлах сетки, и некоторую информацию об области пространства, в которой она задана (параметры "элементарной" ячейки

и т.п.). Кроме того, файл имеет заголовок, идентифицирующий изучаемый объект. Работа может происходить с тремя типами входного файла (**FORM**, **Binary** или **FAN**). Файлы с синтезами типа **FORM** и **FAN** можно рассчитать программой FFT (п.3.1).

Файл типа **FORM** - форматизованный файл последовательного доступа, имеет простую структуру, но занимает много места на диске и поэтому неудобен для работы. Входной файл в таком формате преобразуется в процессе работы в рабочий файл типа **FAN**, который можно запомнить по окончании работы программы и затем использовать как входной при ее следующих запусках.

Файл типа **Binary** - бесформатный файл последовательного доступа, занимает примерно в 2.5 раза меньше места на диске, чем файл типа **FORM**. Он более удобен для работы, но могут возникнуть трудности при его создании, т.к. не все операционные системы поддерживают формат **binary**. В процессе работы такой файл также преобразуется в файл типа **FAN**, который можно запомнить и затем использовать как входной.

Файл типа **FAN** наиболее удобен для работы. Это файл прямого доступа в специально упакованном формате, занимает примерно в 10 раз меньше места на диске, чем соответствующий файл типа **FORM**. Такой файл либо можно рассчитать программой FFT, либо нужно стартовать FAN, CAN или FANS с файлом типа **FORM** или **Binary** и запомнить рабочий файл, созданный после работы программы.

Форматы синтезов описаны в п.3.4.

### 3. Основная сцена.

Проиллюстрируем схему возможных действий программ визуализации и анализа синтезов и модели на примере программы CAN. Почти вся работа программы происходит в основной сцене. Большая часть экрана в ней занята изображением анализируемых функций и модели.

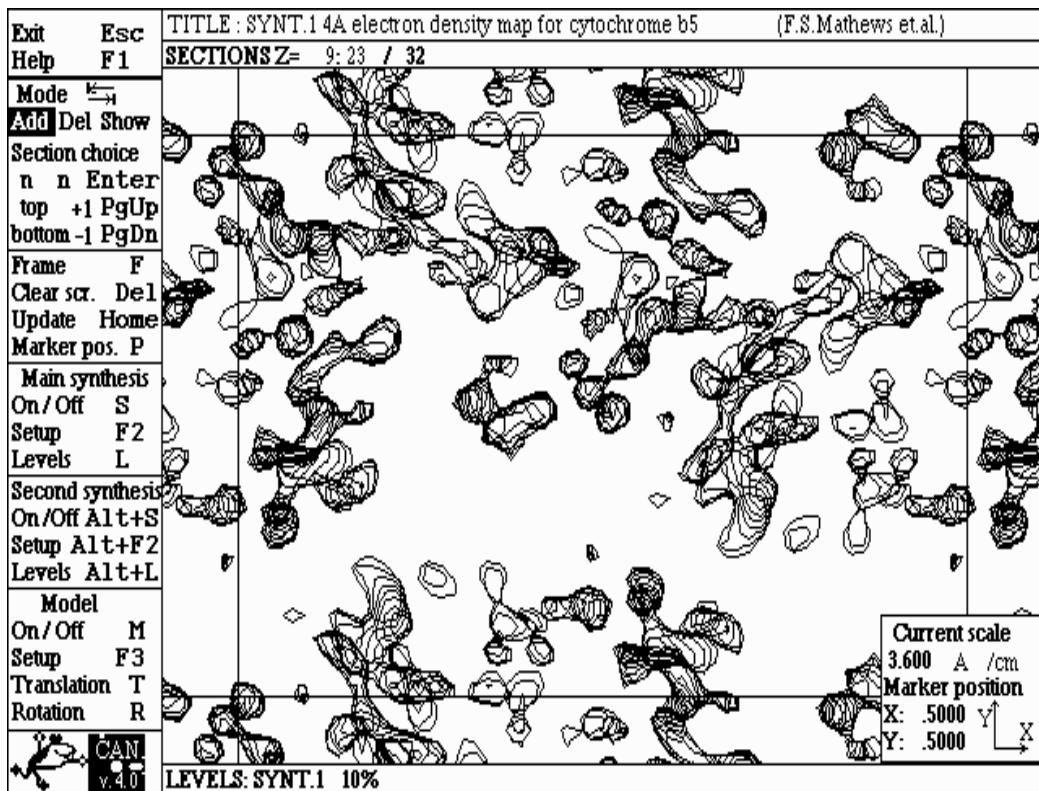


Рис.11. Изображение функции электронной плотности (9-23 сечения).

### Информационные сообщения

В верхней части экрана выводятся заголовки, взятые из файлов с основной и дополнительной функцией. Ниже находится строка с информацией о номерах сечений, изображенных в данный момент на экране. Например, запись SECTION Y = 5 10 20:25 / 48 означает, что сетка, на которой рассчитаны значения основной функции, содержит 48 узлов вдоль оси Y, и на экране представлены сечения по оси Y, отвечающие узлам с номерами 5, 10 и всеми номерами от 20 до 25.

В нижней части экрана показаны относительные значения уровней (см.п.3.3.5.), отображенных на экране (рис.11). В центре экрана расположен маркер, который можно двигать с помощью клавиш со стрелками и снимать относительные координаты интересующей точки.

### Формирование набора сечений

Формирование набора отображенных на экране сечений может происходить в трех режимах: **Add**, **Del** или **Show**. Переключение режимов производится клавишей "Tab".

**Add** - используется для добавления к изображению, имеющемуся на экране, линий уровня нового сечения.

**Del** - используется для удаления с экрана линий уровня одного из отображенных на экране сечений.

**Show** - используется для отображения на экране линий уровня лишь одного (выбранного пользователем) сечения; при выборе нового сечения предыдущее стирается с экрана.

На экран выводятся линии уровня в сечениях, параллельных указанной координатной плоскости. Во всех трех режимах для выбора сечения можно либо набрать его номер, либо вывести необходимую группу сечений, манипулируя клавишами "PgUp" и "PgDn". В табл.3 приведены возможные варианты вывода сечений на экран.

Табл.3

режим	выбор сечения при нажатии	
	PgUp	PgDn
Add	сечение с номером на NS	сечение с номером на NS
Show	больше, чем максимальный из номеров уже изображенных сечений	меньше, чем минимальный из номеров уже изображенных сечений
Del	сечение с максимальным номером из изображенных на экране	сечение с минимальным номером из изображенных на экране

Шаг "NS" для просмотра сечений можно задать в сцене "Setup".

## **Общее управление изображением**

**Esc (Exit)** - выход в сцену окончания работы (п.3.3.10.);

**F1 (Help)** - вызов подсказки;

**Tab (Mode)** - выбор режима работы **Add**, **Del** или **Show**;

**n** |

**PgUp** } - выбор сечения для рисования;

**PgDn** |

**F (Frame)** - изменение масштаба изображения и сдвиг "окна";

**Del (Clear screen)** - очистка экрана;

**Home (Update)** - обновление экрана (рекомендуется после работы в режиме **Del**);

**P (Marker position)** - вывод на экран информации о координатах маркера и масштабе изображения. При повторном нажатии клавиши "P" эта информация исчезает.

## **Управление изображением основной функции**

**S (On/Off)** - включение и выключение изображения основной функции;

**F2 (Setup)** - установка параметров изображения основной функции (п.3.3.6);

**L (Levels)** - изменение значений, количества и цвета линий уровней (п.3.3.5).

## **Управление изображением дополнительной функции**

**Alt+S (On/Off)** - включение и выключение изображения дополнительной функции;

**Alt+F2 (Setup)** - установка параметров изображения дополнительной функции (п.3.3.7);

**Alt+L (Levels)** - изменение значений, количества и цвета линий уровней дополнительной функции (п.3.3.5).

## **Управление изображением модели**

**M (On/Off)** - включение и отключение изображения модели;

**F3** (Setup) - установка параметров изображения модели (п.3.3.8);

**T** (Translation) - выход в режим сдвига модели (п.3.3.9);

**R** (Rotation) - выход в режим вращения модели (п.3.3.9).

#### **4. Изменение масштаба изображения и сдвиг окна.**

Сцена изменения масштаба (Frame) вызывается нажатием клавиши "F". В левое информационное поле выводятся значения масштаба, координаты центра и границы окна, существовавшие в основной сцене, и текущие значения этих параметров внутри данной сцены (при входе в сцену они совпадают). Масштаб показывает, сколько единиц измерения, обозначенных во входном файле, укладывается в 1 см экрана. Границы окна - относительные координаты границ в системе коор-динат элементарной ячейки (в этой системе оси идут вдоль ребер ячейки; A1, A2, A3 - единицы масштаба вдоль осей).

При входе в сцену в центре окна появляется маркер, который можно перемещать клавишами со стрелками. Положение маркера задает центр нового окна, которое появится после возвращения в основную сцену. При движении маркера в левом информационном поле меняются текущие координаты маркера (центра окна) и относительные координаты границ окна.

Текущее значение масштаба можно изменено следующим образом:

- вводом числового значения желаемого масштаба;
- нажатием "+" ("") для увеличения (уменьшения) текущего значения.

При нажатии клавиши "Del" происходит восстановление исходных параметров изображения (масштаба, границ окна и координат центра картинки), существовавших на момент входа в сцену "Frame".

При уменьшении значения текущего масштаба движущиеся границы в окне показывают область, которая займет все окно, если будет взято текущее значение масштаба (рис.12-13).

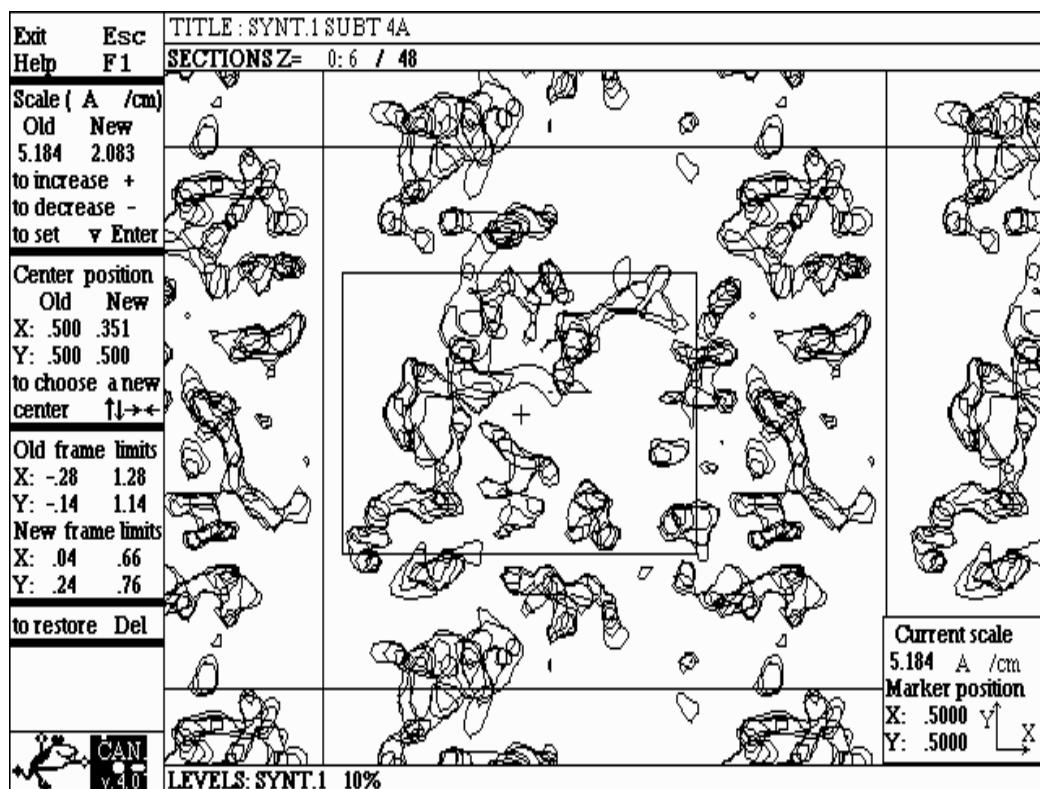


Рис.12. Выбор части изображения, которое нужно вывести на экран.

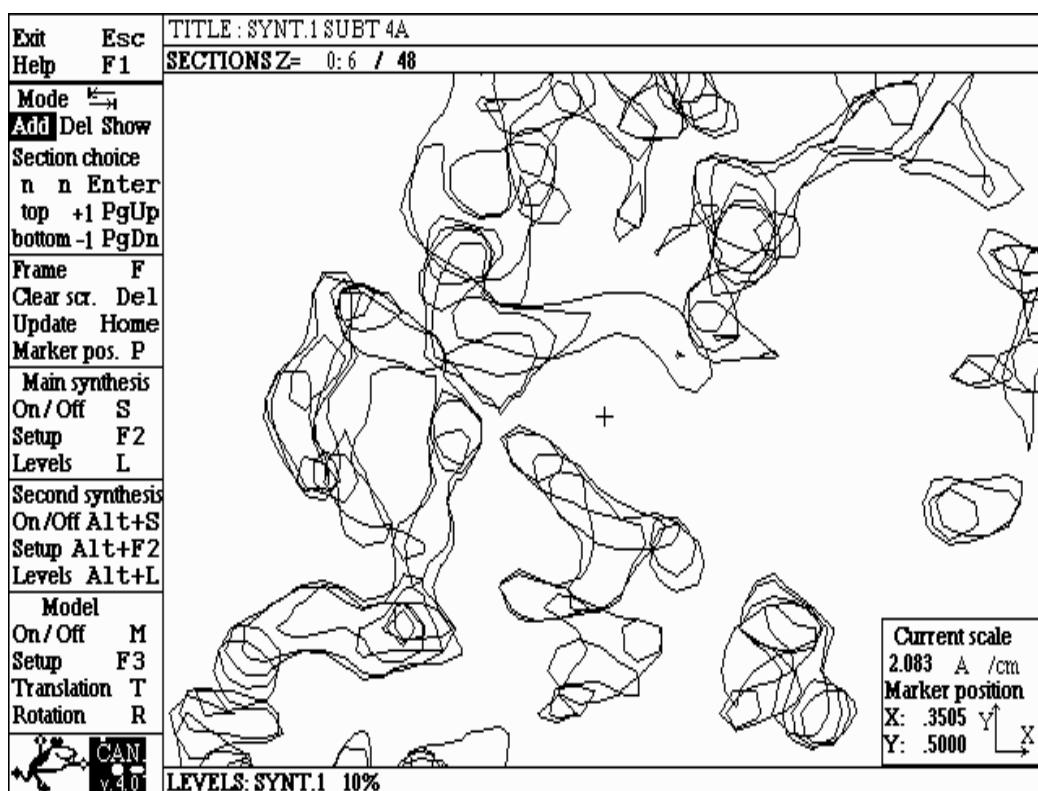


Рис.13. Изображение объекта в увеличенном масштабе и в новых границах.

## **5. Выбор значений уровней.**

Сцена выбора значений отображаемых уровней ("Levels") вызывается нажатием клавиши "L" для основной функции или "Alt+L" для дополнительной.

С каждой поверхностью уровня функции трех переменных  $\rho(r)$  связаны две величины:

- 1) абсолютное значение уровня  $\rho_0$ , которому соответствует поверхность  $\rho(r) = \rho_0$ ;
- 2) относительный объем (по отношению к полному объему  $V_{cell}$  ячейки) части элементарной ячейки  $V_p$ , ограниченной этой поверхностью:

$$V = (V_p / V_{cell}) \times 100\%,$$

где  $V_p$  - объем области в ячейке, в точках которой  $\rho(r) \geq \rho_0$ .

Для задания уровня можно использовать любую из этих двух характеристик. При этом другая будет рассчитана программой автоматически.

В режиме выбора значений уровней можно:

- изменить количество изображаемых изолиний;
- изменить значения уровней;
- изменить цвет изолиний.

### **Активные и пассивные уровни**

При входе в сцену на экране появляется таблица из трех колонок (рис.14). Первая колонка - значения возможных уровней в абсолютной шкале, вторая - процент объема ячейки, выделяемый этим уровнем, третья - код цвета, которым изображается соответствующая изолиния.

Часть уровней, присутствующих в таблице, изображена красным цветом - это "активные" уровни. На экране изображаются изолинии, отвечающие только активным уровням. Прочие ("пассивные") уровни хранятся в таблице, но при рисовании не используются.

Для того, чтобы перевести пассивный уровень в активный (или наоборот), следует подвести курсор к желаемому значению и нажать “Insert”.

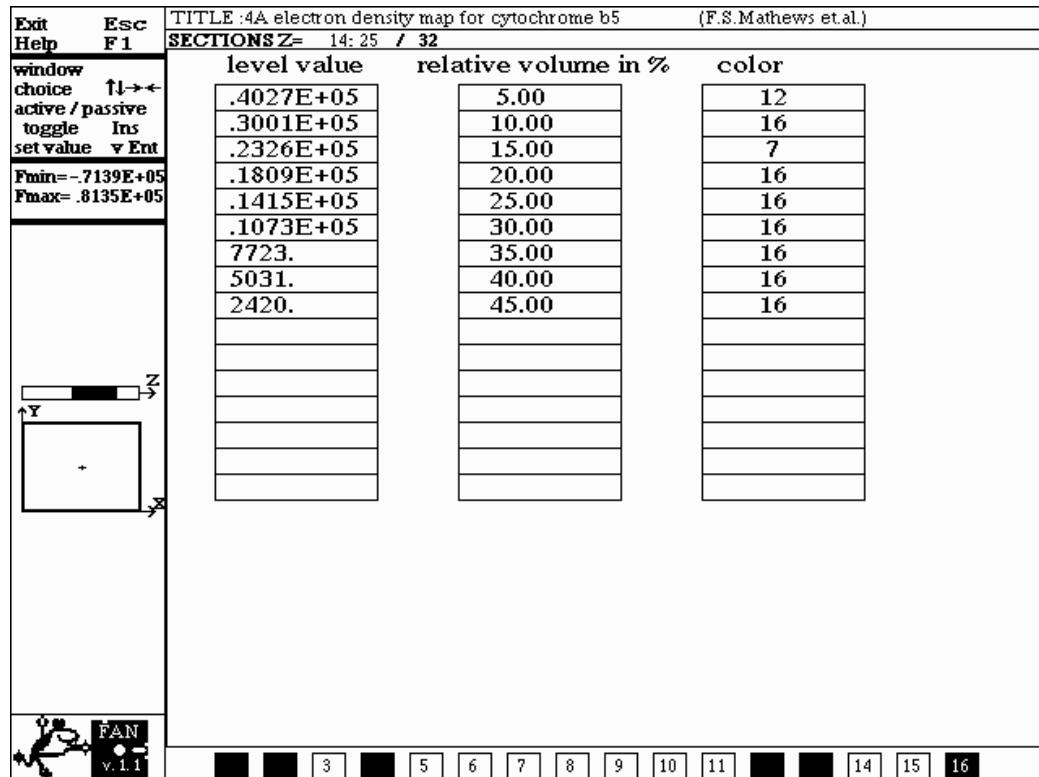


Рис.14. Выбор количества, значений уровней и цвета изолиний.

## Изменение цвета линии уровня

Для изменения цвета следует подвести стрелками курсор в нужную позицию в третьей колонке, ввести новый код цвета (от 1 до 16) и нажать "Enter".

## Запись нового значения в таблицу

Для введения нового значения уровня в таблицу следует подвести стрелками курсор к одной из позиций в таблице (неважно, занятой числом или свободной), ввести желаемое значение и нажать "Enter".

## **6. Установка внутренних параметров работы программы.**

В сцене установки общих параметров программы можно изменить следующие параметры (рис.15):

X

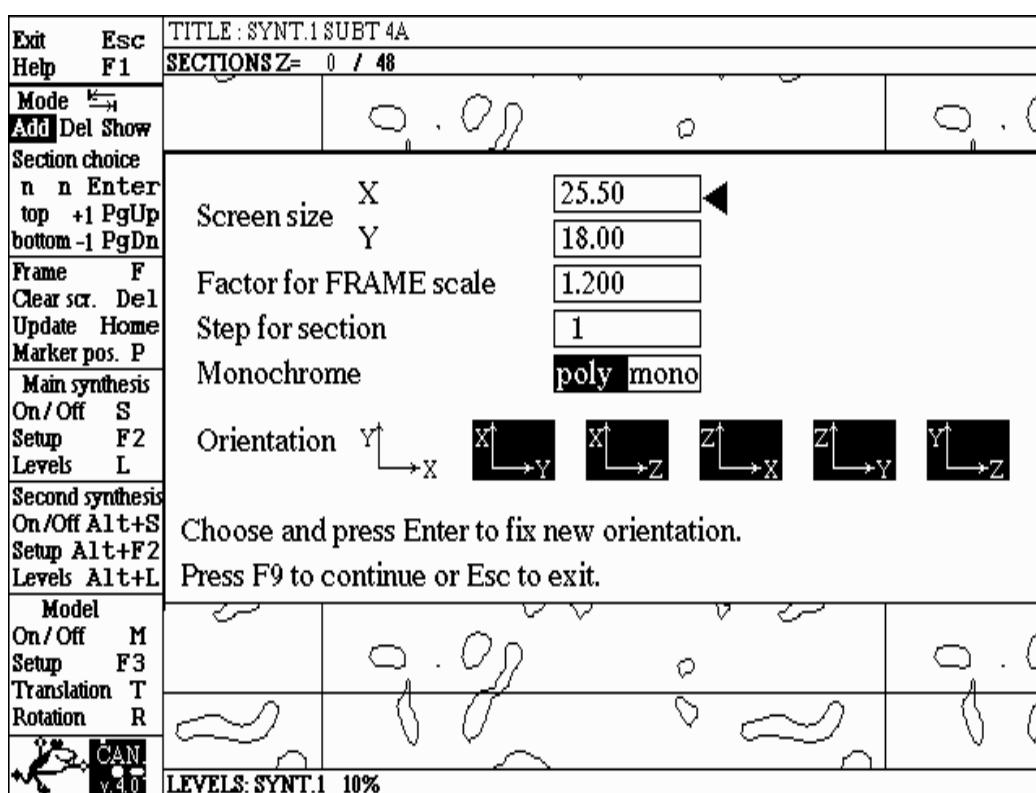
**Screen size** } - размеры окна монитора по горизонтальной и вертикальной осям (в см.);  
Y

**Factor for frame scale** - коэффициент, использующийся в сцене "Frame" при изменении масштаба изображения;

**Step for section** - шаг, с которым в основной сцене при нажатии клавиш "PgDn" или "PgUp" берутся очередные сечения;

**Monochrome** - установка режима работы в цветном или монохромном варианте;

**Orientation** - выбор новой ориентации синтеза.



## 7. Установка параметров работы второго синтеза.

В этой сцене ("Alt+S") можно задать или изменить имя файла с дополнительной функцией. Кроме того, можно задать сдвиг и (или) инверсию изображения дополнительной функции относительно

основной. Когда все необходимые изменения внесены, нажимается клавиша "F9" и на экран выводится информация о втором синтезе (рис.16). После выхода в основную сцену на экране будут изображены наложенные сечения двух синтезов. Их можно задать разным цветом и просматривать, например, по одному сечению или небольшими порциями.

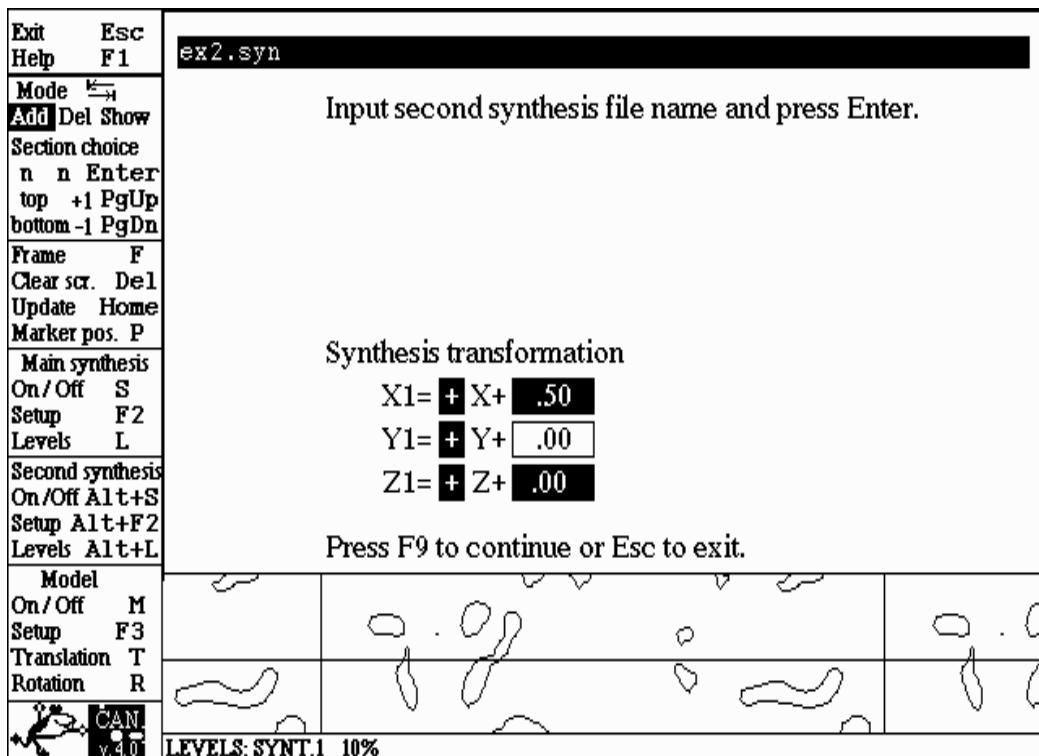


Рис.16. Сцена ввода в работу второго синтеза электронной плотности.

## 8. Установка и изменение параметров модели.

В этой сцене можно ввести или изменить имя файла с моделью, которую требуется вывести на экран. Атомная модель изображается в виде проекции на экран скелетной модели, в которой атомы изображаются шариками, а химические связи - отрезками. Все связанные по симметрии молекулы изображаются разным цветом, поэтому ситуацию с плохой упаковкой молекул в кристалле, когда атомы соседних молекул "налезают" друг на друга, легко определить визуально. Кроме того, можно задать следующие параметры:

**Factor for translation step** - при нажатии клавиши +(-) шаг трансляции, используемый в сцене “Translation” (п.3.3.9.), умножается (делится) на этот коэффициент.

**Factor for rotation step** - при нажатии клавиши +(-) шаг вращения, используемый в сцене “Rotation” (п.3.3.9.), умножается (делится) на этот коэффициент.

**First residue number** - номер первого остатка, который подписывается при изображении модели.

**Step for residues numbering** - шаг, с которым меняются номера подписываемых остатков.

**Model colour** - номер цветовой палитры для изображения модели.

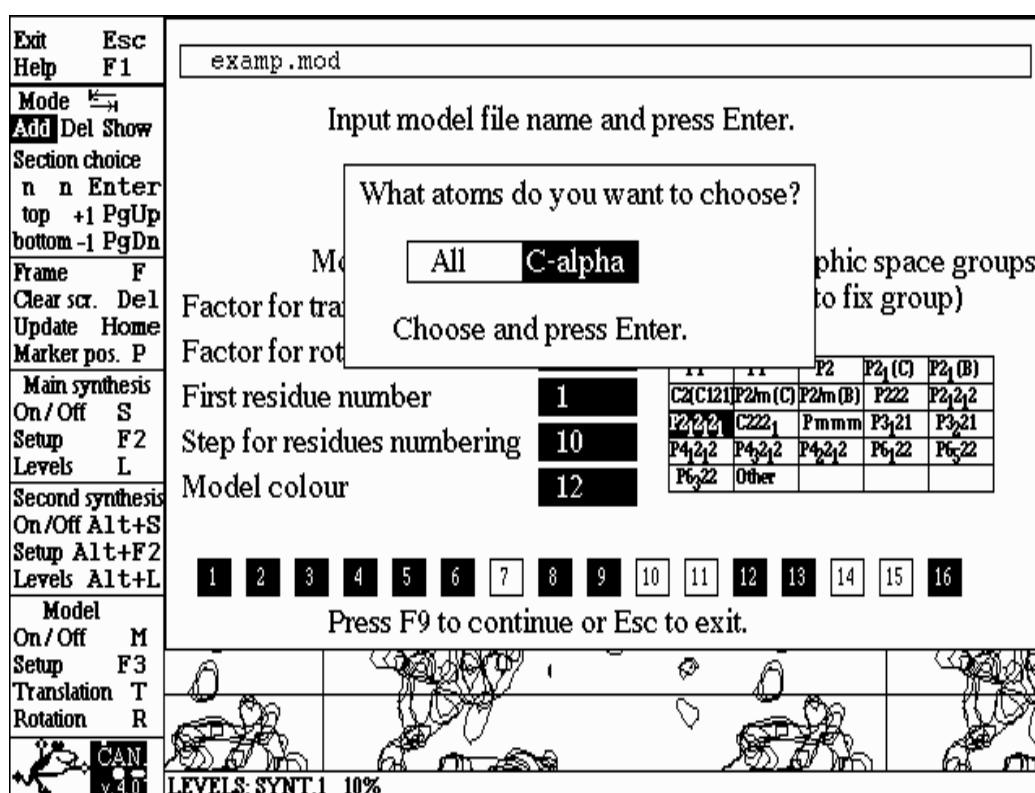


Рис.17. Задание параметров для рисования атомной модели.

После установки нужных параметров пользователю предоставляется возможность ввести в работу либо всю атомную модель, либо только последовательность С<sub>α</sub> атомов (рис.17).

## 9. Модификация модели (сдвиг и вращение).

Выведенную на экран модель (рис.18) можно двигать и вращать как жесткое тело.

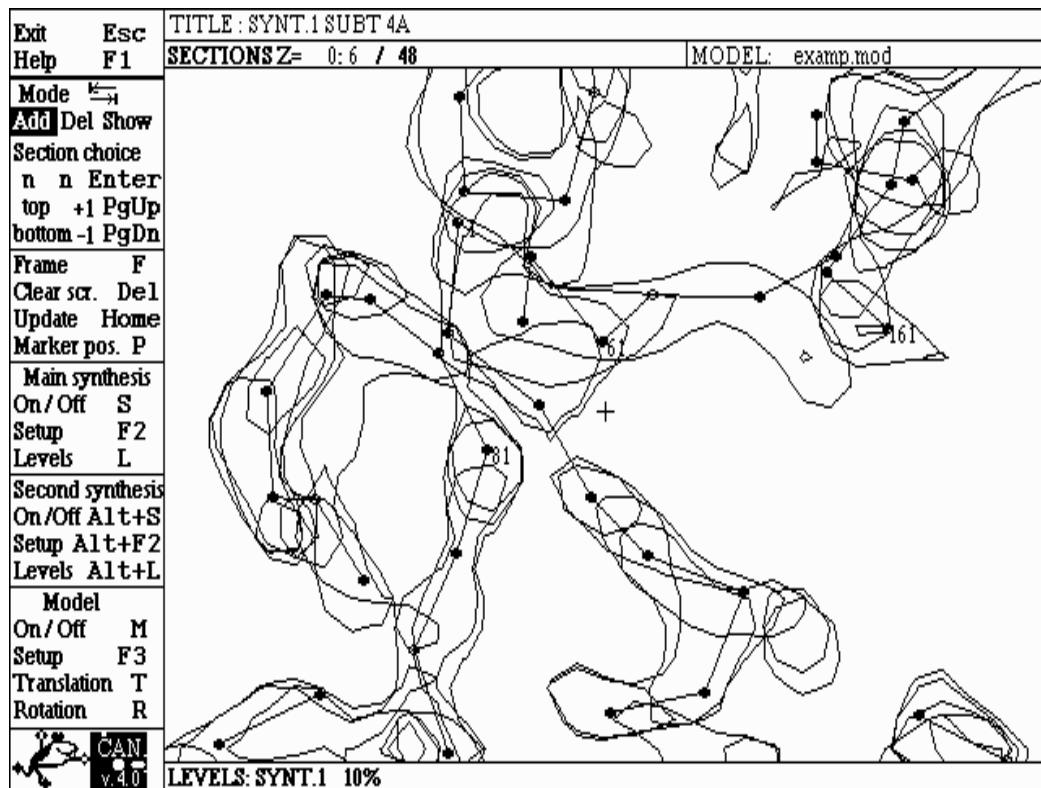


Рис.18. Изображение синтеза электронной плотности с наложенной атомной моделью.

В режиме модификации модели возможны следующие действия:

**Translation** - переход в режим трансляции (клавиша "T").

**Rotation** - переход в режим вращения (клавиша "R"). Текущий режим работы задан красным цветом.

**Update** - восстановление исходного положения модели (клавиша "Home").

**Save** - запоминание текущего положения модели (клавиша "S").

**Get** - восстановление положения модели, которое запомнили ранее (клавиша "G")

**New step** - увеличение (уменьшение) шага трансляции или вращения (клавиша + или - ).

**Shift** - трансляция или вращение модели вокруг осей координат (клавиши →, ←, ↑, ↓, PgUp, PgDn)

В отдельном окне изображена таблица с информацией о векторе трансляции и углах поворота модели. В ней выводятся величины вектора сдвига текущей модели по сравнению с исходной (в Å) и углы вращения модели (в градусах).

## **10. Завершение работы программы.**

Для завершения работы следует, находясь в основной сцене, нажать клавишу "Esc". При этом, если по ходу работы программы входной файл с синтезом электронной плотности претерпевал изменения (менялась ориентация координатных осей либо формат входного файла отличался от внутреннего формата программы), то можно сохранить этот рабочий файл и в дальнейшем его использовать при новых запусках. Аналогично можно сохранить модифицированную модель, если с ней происходила какая-либо работа, и текущие значения внутренних параметров (имена входных файлов, масштаб, положение окна, значения уровней, номера выведенных на экран сечений и т.п.) в специальном файле параметров.

При выборе ответа "No" или нажатии клавиши "Esc" работа прекращается, при выборе ответа "Yes" программа просит ввести имена соответствующих файлов, после чего прекращает работу.

### **3.3. Форматы используемых файлов.**

Программа FFT использует в качестве входного файл структурных факторов в формате UF, а на выходе создает файл, содержащий синтез электронной плотности в одном из трех форматов: FORM, Ten Eyck или FAN.

Программы FAN, CAN и FANS в качестве входного используют файл с синтезом электронной плотности в форматах FORM, FAN или Binary, а CAN и FANS еще файл с атомной моделью в формате PDB.

## **1. Файлы с синтезами электронной плотности.**

### **Формат файла типа FORM.**

1 запись: CODE, S1, S2, S3, PAR (A4, 3A1, A3)

character\*4 CODE, character\*1 S1,S2,S3, character\*3 PAR

CODE - ключевое слово, содержит символы \$\$GA для периодических функций или \$UGA для непериодических,

S1, S2, S3 - символы для обозначения координатных осей,

PAR - обозначение единицы измерения сторон элементарной ячейки.

2 запись: TITLE (A80) содержит 80 символов произвольного текста для обозначения исследуемой функции.

3 записи: A1, A2, A3, AL, BE, GA (6G10.4)

содержит 6 действительных чисел по 10 байтов каждое, параметры элементарной ячейки. Величины A1, A2, A3 измеряются в PAR единицах, углы AL, BE, GA в градусах.

4 запись: N1, N2, N3, MIN1, MAX1, MIN2, MAX2, MIN3, MAX3 (9I5)

содержит 9 пятибайтовых целых чисел - разбиение элементарной ячейки и границы выбранной части функции.

Следующие (MAX3 - MIN3 + 1) записей

(( SEC(J1,J2), J1=MIN1,MAX1 ), J2=MIN2,MAX2 ) (8G10.4))

содержат (MAX1-MIN1+1)\*(MAX2-MIN2+1) действительных чисел каждая - значения функции.

### **Формат файла типа Binary.**

1 запись - CODE, S1, S2, S3, PAR, TITLE, A1, A2, A3, AL, BE, GA,  
N1, N2, N3, MIN1, MAX1, MIN2, MAX2, MIN3, MAX3

следующие N3 записей -

((SEC(J1,J2,J3),J1=MIN1,MAX1),J2=MIN2,MAX2),J3=MIN3, MAX3)

Величины

SEC(J1,J2,J3), A1, A2, A3, AL, BE, GA - 4-байтовые действительные числа;  
N1,N2,N3,MIN1,MAX1,MIN2,MAX2,MIN3,MAX3 - 4-байтовые целые числа;  
S1,S2,S3 - 1-байтовые, PAR - 3-байтовые, TITLE - 80-байтовые символьные переменные.

Первые 4 байта ( CODE ) должны содержать \$B4 - для периодической функции, \$UB4 - для непериодической функции.

### **Формат файла типа Тип Еуcк**

Это бесформатный файл, имеет следующий вид:

1 запись - TITLE (18 слов);

затем следуют по две записи на каждое сечение результата:

JZ, MINX, MAXX, MINY, MAXY

((R(I, J), I=M1, M2), J=L1, L2)

Здесь R - массив результата; MINX,MAXX,MINY,MAXY - граничные значения индексов выводимой функции;

JZ - номер выводимого сечения;

M1=MINX+1, M2=MAXX+1, L1=MINY+1, L2=MAXY+1

В конце файла выводится запись (-1,-1,-1,-1,-1).

### **2. Файл структурных факторов в формате UF.**

Файл структурных факторов в записи по каждому рефлексу может содержать любую доступную информацию: экспериментальное значение модуля структурного фактора, рассчитанные по модели модуль и фазу,  $\sigma$ , весовой множитель, коэффициенты А, В, С, D распределения вероятностей фаз Хендрикsona-Латтмана /122/ и т.д. Обязательными величинами являются индексы рефлексов h,k,l.

Файл форматизованный, имеет следующий формат:

TITLE1 (18A4) - произвольный текст

TITLE2 (18A4) - произвольный текст

LRECL (I4) - количество чисел в последующих записях

IH, IK, IL, (R(I), I=1, LRECL-3) - для каждого рефлекса

(3I4, 5G12.6) для LRECL ≤ 8

(3I4, 5G12.6, / (6G12.6)) для LRECL > 8