

АКАДЕМИЯ НАУК БЕЛОРУССКОЙ ССР  
ИНСТИТУТ МАТЕМАТИКИ  
ИНСТИТУТ БИООРГАНИЧЕСКОЙ ХИМИИ



**ПРОГРАММНОЕ  
ОБЕСПЕЧЕНИЕ  
ЭВМ**

ВЫПУСК 86

МИНСК 1989

АКАДЕМИЯ НАУК БЕЛОРУССКОЙ ССР  
ИНСТИТУТ МАТЕМАТИКИ  
ИНСТИТУТ БИООРГАНИЧЕСКОЙ ХИМИИ

ПРОГРАММНОЕ  
ОБЕСПЕЧЕНИЕ ЭВМ

ВЫПУСК 86

КОМПЬЮТЕРНЫЕ МЕТОДЫ БЕЛКОВОЙ ИНЖЕНЕРИИ

Под редакцией  
канд. физ.-мат. наук С. А. Шермана

МИНСК 1989

Рецензент д-р физ.-мат. наук Р.Г.Жбанков

Редакционный совет:

канд. физ.-мат. наук А.С.Метельский; А.Е.Данченко; канд. физ.-мат. наук Г.М.Адаменко; канд. физ.-мат. наук Н.С.Жаврид; канд. техн. наук Л.А.Золоторевич; Т.И.Кляус; Э.В.Ковалевич; канд. экон. наук М.Л.Петрович; канд. физ.-мат. наук А.В.Покатаев; Ю.С.Полоус; Л.А.Рута; канд. физ.-мат. наук А.А.Сенько; Н.Д.Соколова; канд. физ.-мат. наук Г.К.Столяров; канд. физ.-мат. наук В.И.Цагельский; Л.Т.Чупригина; канд. физ.-мат. наук В.А.Шкель; канд. физ.-мат. наук Н.В.Шкут

П 190210000-004 4-89  
М334-89

© Институт математики АН БССР,  
Институт биоорганической химии, 198

## СО Д Е Р Ж А Н И Е

Предисловие редактора . . . . .	4
Тимошук Т. А., Шерман С. А. Тезаурус для информационно-поисковой системы в области струк- турных исследований белков . . . . .	6
Леонова В.И., Цейтин В. М., Сломенец Т. И. НИМАН - система хранения аминокислотных последовательностей белков и пеп- тидов человека . . . . .	34
Шерман С. А., Дегтерева Л. Н., Ракова А. А. Расширение комплекса программ для работы с базой рентгеноструктурных данных белков . . . . .	53
Финкельштейн А. В. Пакет прикладных программ для расчета вторичной структуры белков и полипептидов . . . . .	66
Рубинштейн А. И., Поройков В. В. Программа предсказания вторичной структуры белков по аминокислотной последовательности - модифициро- ванный алгоритм Шераги . . . . .	77
Уржумцев А. Г., Дунин Б. Ю., Вернослова Е. А. FROG - комплекс программ для локального уточнения атомной структуры макромолекул . . . . .	86
Позигун Д. В., Кузьмин В. Е. Комплекс программ для конформационных расчетов биоорганических и бионеорганических соединений . . . . .	98
Ракова А. А. Структура дистрибутивной магнит- ной ленты и установка комплекса программ белковой инженерии . . . . .	123

УДК 548.73

А.Г.Уржумцев, В.Ю.Луни,  
Е.А.Верносова  
Пушино, Науч.-исслед. ЦИ  
АН СССР

## FROG- КОМПЛЕКС ПРОГРАММ ДЛЯ ЛОКАЛЬНОГО УТОЧНЕНИЯ АТОМНОЙ СТРУКТУРЫ МАКРОМОЛЕКУЛ

### I. Назначение и сфера применения

Комплекс *FROG* представляет собой пример разработки программного обеспечения для решения расчетных задач белковой инженерии и белковой кристаллографии, которое можно было бы легко адаптировать к новым постановкам задач, сохраняя при этом высокую эффективность работы.

#### I.1. Примеры задач, при решении которых может быть использован комплекс *FROG*

Таковыми примерами являются :

- уточнение структуры макромолекулы по данным рентгеновского эксперимента (с возможным использованием стереохимических ограничений);
- "энергетическое" уточнение структуры (теоретический конформационный анализ);
- уточнение структуры с использованием данных о фазах структурных факторов (по существу - уточнение структуры по совместным данным нескольких рентгеновских экспериментов с тяжелоатомными производными);
- уточнение ориентации и положения молекулы как целого по данным рентгеновского эксперимента.

После небольших добавлений комплекс *FROG* может быть использован для задач:

- "вписывания молекулы в электронную плотность";
- уточнения структуры с привлечением данных по ЯМР-спектроскопии и т.п.

## I.2. Основные особенности комплекса FROG

### I. Быстродействие.

Комплекс построен на базе последовательного применения алгоритма быстрого дифференцирования [1],[2]. При уточнении структуры гамма-кристаллина III $\beta$  при разрешении 3.0 Å (около 3000 атомов в независимой части ячейки, около 12000 рефлексов) по совокупности рентгеновских и стереохимических данных время цикла уточнения на ЭВМ I065 составляло около 30 минут CPU. С увеличением размеров белка требуемое процессорное время растет линейно.

### 2. Задание критериев уточнения (целевой функции) внешними файлами

Ядро комплекса составляет программа уточнения модели макромолекулы. При этом описание требований, предъявляемых к модели (т.е. того, чего мы хотим добиться от модели), вынесено во внешние файлы, формируемые независимо от основной программы. Это позволяет менять требования, предъявляемые к модели, не затрагивая основной программы. В частности, это позволяет легко вводить новые типы ограничений и уточнять модели различных классов макромолекул (белки, ДНК и т.д.).

### 3. Атомно-блочная модель макромолекулы

В описываемой версии комплекса модель уточняемого объекта задается в виде:

- а) совокупности свободных атомов, параметры которых могут меняться независимо друг от друга;
- б) набора жестких групп атомов, которые в процессе уточнения двигаются как твердое тело.

Количество и состав жестких групп может быть произвольным.

В частности:

- вся модель может состоять из свободных атомов (отсутствие жестких групп);
- жесткой группой может быть вся молекула, в этом случае 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 и т.д.

Взаимное расположение отдельных жестких групп может регулироваться введением соответствующих штрафных функций.

4. Возможность "жесткого" учета некристаллографической симметрии.

При наличии некристаллографической симметрии программе можно задать и независимо изменять модель только одной субъединицы. Остальные субъединицы будут при этом генерироваться в процессе работы при помощи заданных преобразований некристаллографической симметрии. Эта возможность позволяет также работать в произвольной пространственной группе.

5. Базовая версия комплекса FROG

В данной работе описана базовая версия комплекса FROG. По мнению авторов, дальнейшее развитие программ уточнения пойдет по линии сборки программы уточнения для решения конкретной задачи (или класса задач) по исследованию пространственной структуры. Поэтому возможны различные модификации базовой версии, адаптированные к условиям конкретной работы (повышение эффективности за счет учета специфики пространственной группы; учет особенностей используемой ЭВМ, в том числе создание вариантов комплекса для развитых персoкомпьютеров и т.п.).

Особенностями базовой версии являются:

- работа в пространственной группе P1 (остальные группы автоматически сводятся к P1, но при этом возможно замедление работы программы по сравнению с версией, учитывающей симметрию данной группы);
- атомно-блочная модель для описания структуры;
- ориентированность на работу с белками.

Комплекс написан на языке ЮТРАН. Объем комплекса - около 35000 операторов. Схема связи основных программ комплекса и файлов данных приведена на рис. I.

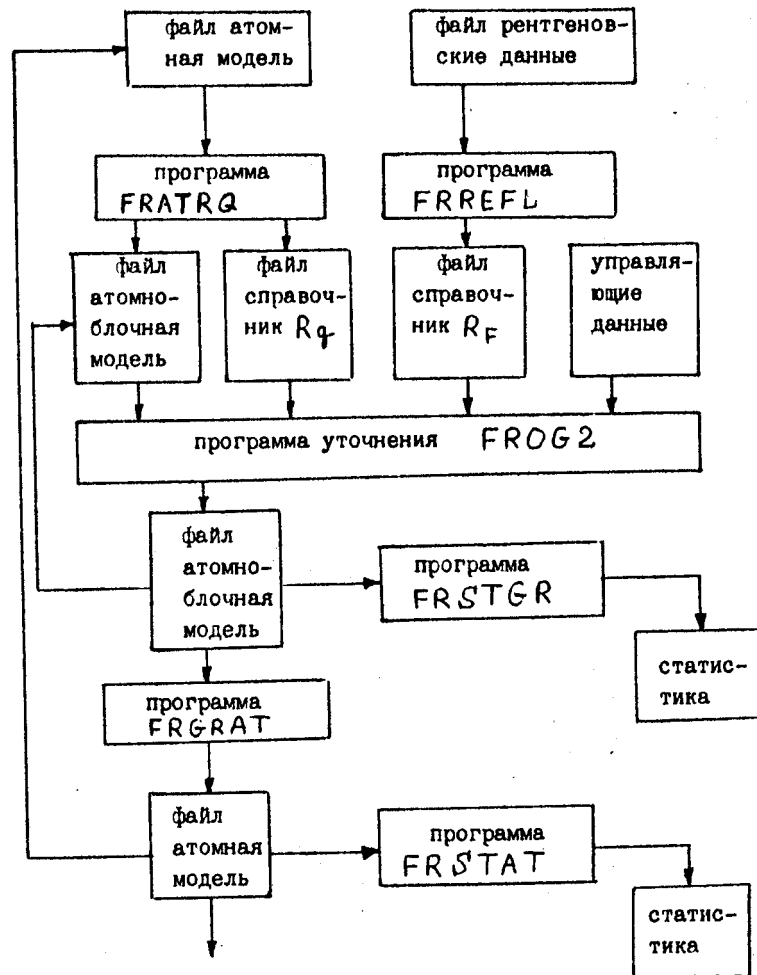


Рис. I. Схема функциональных связей программ и файлов комплекса



## 2. Программа уточнения FROG2

Ядро комплекса - программа FROG2 - предназначена для минимизации функции специального вида, зависящей от большого числа переменных. Ввиду большого числа переменных и сложной зависимости от них минимизируемой функции основную часть программы составляют подпрограммы, организующие расчет этой функции и ее градиента. Для минимизации применяется метод спуска с выбором направления спуска на базе антиградиента либо по методу сопряженных градиентов. Для поиска оптимальной точки на найденном направлении спуска применяется аппроксимация минимизируемой функции многочленом третьей степени по значениям функции и ее производной в концах отрезка аппроксимации.

В программе используется несколько уровней описания модели:

- $\chi$  - обобщенные параметры - это параметры жестких групп, на которые разбита модель (углы Эйлера и векторы трансляций, температурные факторы и т.п.);
- $q$  - параметры атомов, входящих в модель (декартовы координаты атомов, температурные факторы, коэффициенты заполнения);
- $p$  - распределение электронной плотности в ячейке кристалла (рассчитанное на некоторой сетке);
- $F$  - структурные факторы.

Минимизируемая функция имеет вид:

$$R = w_F R_F + w_p R_p + w_q R_q + w_\chi R_\chi,$$

где  $R_F$  выражается непосредственно через структурные факторы, рассчитанные по текущей модели;

$R_p$  - через значения электронной плотности;

$R_q$  - через параметры атомов;

$R_\chi$  - через обобщенные параметры (параметры жестких групп);

$w_F, w_p, w_q, w_\chi$  - задаваемые перед работой программы веса.

Каждой из функций  $R_F, R_p, R_q, R_\chi$  отвечает специальный файл входных данных, описывающий этот компонент минимизируемой функции. В свою очередь, каждый из таких компонентов минимизируемой функции может расщепляться на функции разного вида. Например:

$$R_q = w_{q,e} R_{q,e} + w_{q,d} R_{q,d} + w_{q,a} R_{q,a} + \dots$$

где  $R_{p,e}$  - энергия невалентных взаимодействий;  
 $R_{p,d}$  - штраф за отклонение в модели от стандартных длин связей;  
 $R_{p,a}$  - штраф за нарушение величин валентных углов;  
и т.п.

Примеры различных типов минимизируемых функций приведены в пп. 3.2.2. и 3.2.3.

Наконец, каждая из функций типа  $R_{p,e}, R_{p,d}, R_{p,a} \dots$  представляет собой сумму большого числа слагаемых, каждое из которых зависит от небольшого числа параметров (своего уровня описания модели). Например,  $R_{p,d} = \sum w_{i,j} (d_{i,j}^2 - (d_{i,j}^0)^2)$  где суммирование производится по всем парам валентно связанных атомов;

$d_{i,j}$  - расстояние между  $i$ -м и  $j$ -м атомами модели;  
 $d_{i,j}^0$  - стандартная длина связи для атомов такого типа;  
 $w_{i,j}$  - соответствующий вес.

Здесь отдельное слагаемое зависит от координат двух атомов.

### 3. Работа с комплексом FROG

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

#### 3.1. Исходные данные

##### I) Атомная модель

Этот файл содержит текущую информацию обо всех атомах, включенных в модель. Для каждого из атомов в нем присутствует запись, содержащая координаты атомов, температурный фактор, коэффициент заполнения и некоторую дополнительную информацию. В начале работы с новым объектом этот файл должен быть создан пользователем. В дальнейшем этот файл обновляется программами комплекса FROG в конце процедуры уточнения модели. Форматированный вид файла облегчает его ручную корректировку (добавление/удаление атомов)

при помощи текстовых редакторов.

2) Файл с "рентгеновскими" данными

Этот файл содержит информацию, полученную в процессе рентгеновского эксперимента, и необходим только в том случае, когда эта информация используется в процессе уточнения (так, при "энергетическом" уточнении этот файл может отсутствовать). В файле для каждого из отражений могут быть приведены значения модуля структурного фактора, стандартное отклонение для этого значения, коэффициенты  $A, B, C, D$  распределения вероятностей для фаз структурных факторов [4] и т.п.

3) Дополнительные виды исходной информации

При желании в работу по уточнению могут быть включены и другие виды исходной информации об исследуемом объекте, например, распределение электронной плотности (полученное каким-либо образом), в которую должна быть вписана модель. Другим примером являются данные по ЯМР-спектроскопии.

### 3.2. Входные данные программы уточнения

Для работы программы уточнения *FR062* необходимо задать:

- а) описание стартовой атомно-блочной модели, указав при этом, какие параметры модели разрешается менять в процессе уточнения;
- б) файлы, описывающие критерии, согласно которым будет проходить уточнение модели (то есть файлы, задающие минимизируемую программой *FR062* функцию);
- в) данные, управляющие процессом уточнения, например:
  - число циклов уточнения;
  - размеры элементарной ячейки;
  - веса, с которыми различные критерии включаются в суммарную минимизируемую функцию;
  - параметры, регулирующие ход уточнения.

В отличие от исходной информации (п.3.1.) и управляющих данных, которые хранятся в удобном для пользователя виде, подготовлена атомно-блочной модели и файлов, описывающих критерии уточнения (файлов-справочников), осуществляется специальными сервисными программами, входящими в комплекс *FR06*.

### 3.2.1. Атомно-блочная модель

Отличие атомно-блочной модели от исходной атомной модели в том, что в атомно-блочной модели указанные пользователем наборы атомов объединены в жесткие группы. Такая жесткая группа в процессе уточнения может перемещаться только как твердое тело, то есть взаимное расположение атомов, включенных в такую группу, не меняется в процессе уточнения. В данной версии программы существует ограничение - каждый атом может быть включен в состав только одной жесткой группы.

Программа FRATRQ, входящая в состав комплекса, позволяет преобразовать исходный файл атомов в атомно-блочную модель, объявив жесткими группами для любых (перечисленных пользователем) остатков:

- боковую цепь, начиная с  $C_{\beta}^k$  (здесь и далее верхний индекс обозначает номер остатка);
- пептидное звено  $(CO)^{k-1} N^k C_{\alpha}^k$ ;
- пептидное звено с прилегающей боковой цепью  $(CO)^{k-1} N^k C_{\alpha}^k R^k$ .

Программа FRATRQ позволяет также ввести 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, всю белковую глобулу и т.п.

Регулировка степеней свободы в модели молекулы (что можно менять в процессе уточнения) осуществляется в программе FR062 на двух уровнях флажками в управляющих данных для жестких групп (все координаты атомов группы, все температурные факторы и т.д.) и индивидуальными флажками для каждого из параметров в файле с атомной моделью. Флажок фиксации параметров в управляющих данных может иметь три значения:

- +1 - в этом случае все параметры такого типа фиксируются, невзирая на индивидуальные флажки фиксации;
- 1 - в этом случае все параметры такого типа могут меняться, невзирая на индивидуальные флажки фиксации;
- 0 - в этом случае вопрос о фиксации каждого параметра решается индивидуально, исходя из значения соответствующего флажка в файле с атомной моделью.

### 3.2.2. Файл - справочник $R_q$

Этот файл описывает критерии уточнения, выражающиеся непосредственно через декартовы координаты атомов. Два основных вида таких критериев, используемых в комплексе FROG - это:

- штрафы за нарушение стандартных стереохимических соотношений;
- энергия невалентных взаимодействий.

Программа FRATRQ, входящая в состав комплекса FROG, позволяет на основе исходного файла с атомной моделью создать файл - справочник  $R_q$ . При этом в зависимости от того, какие образованы жесткие группы, в этот файл заносятся лишь те штрафы, которые могут меняться при заданной разбивке на жесткие группы. Так, если жесткой группой объявлена вся молекула, то никакие стереохимические соотношения внутри молекулы в процессе уточнения измениться не могут. Поэтому в таком случае никаких штрафов в справочник  $R_q$  внесено не будет. Аналогичный учет разбивки на жесткие группы проводится при отборе пар атомов, энергию невалентного взаимодействия которых следует учитывать.

Виды штрафов за нарушение стереохимических соотношений приведены в таблице I.

Невалентные взаимодействия учитываются потенциалом

$$R_{q,E}(\bar{r}_1, \bar{r}_2) = \frac{A}{|\bar{r}_1 - \bar{r}_2|^2} + \frac{B}{|\bar{r}_1 - \bar{r}_2|^6}$$

с параметрами A и B, определенными в [3]. При этом отбираются пары атомов, находящихся друг от друга не далее чем на расстоянии  $r_{max}$ , задаваемом пользователем. При таком отборе анализируются все виды контактов атомов:

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

Относительные веса, с которыми входят в суммарный минимизируемый критерий штрафы различных типов, задаются пользователем в управляющих данных.

### 3.2.3. Файл - справочник $R_F$

Этот файл описывает критерии уточнения, выражающиеся непосредственно через модули и фазы структурных факторов, рассчитанных по текущей модели. Перечень таких видов критериев дан в таблице 2. Программа FRREFL, входящая в состав комплекса FROG, позволяет преобразовать исходный файл с "рентгеновскими" данными в файл-справочник  $R_F$ , введя указанные пользователем типы критериев из таблицы 2. Относительные веса, с которыми входят критерии в суммарную минимизируемую функцию, задаются пользователем. Отбор рефлексов, для которых соответствующие слагаемые включаются в сумму, осуществляется на двух этапах:

- на этапе подготовки справочника  $R_F$  отбираются рефлексы заданной зоны разрешения, со значением модуля структурного фактора, превосходящим заданный порог;
- на этапе работы программы уточнения FROG2 из файла-справочника  $R_F$  отбираются лишь рефлексы с  $t_1 \leq F^{calc} / F^{obs} \leq t_2$ , где  $t_1, t_2$  задаются пользователем в управляющих данных.

Шкальные коэффициенты, фигурирующие в критериях рассматриваемого типа, могут (по желанию пользователя) либо уточняться в ходе работы программы FROG2, либо оставаться неизменными.

В зависимости от желания пользователя в критериях таблицы 2 может использоваться одна из следующих весовых схем:

- а)  $w_j = 1/\delta_j^\alpha$ , где  $\alpha$  - заданный параметр (при  $\alpha=0$  все  $w=1$ );
- б)  $w_j = 1/\delta_j$ , где  $\delta_j$  - индивидуальное для данного рефлекса значение, взятое из исходного файла "рентгеновских" данных.

### 3.3. Обработка результатов работы программы уточнения FROG2

По окончании работы программа уточнения FROG2 создает файл с уточненной атомно-блочной моделью. Файл с соответствующей атомной моделью может быть получен при помощи программы FRGRAT, входящей в состав комплекса. В комплексе имеются также программы FRSTGR и FRSTAT, позволяющие получить информацию о точности выполнения стереохимических требований в имеющейся модели белка. Входным файлом для программы FRSTGR является файл с атомно-блочной моделью, а для FRSTAT - файл с атомной моделью.

Таблица I  
Виды штрафов за нарушение стандартных стереохимических соотношений

Штрафная функция	Примечание
$R_{\varphi, d}(\bar{r}_1, \bar{r}_2) = ( \bar{r}_1 - \bar{r}_2 ^2 - d_0^2)^2$	учитываются все пары валентно связанных атомов
$R_{\varphi, \alpha}(\bar{r}_1, \bar{r}_2, \bar{r}_3) = 1 - \cos(\alpha - \alpha_0)$	учитываются все тройки атомов, образующих валентный угол
$R_{\varphi, \omega}(\bar{r}_1, \bar{r}_2, \bar{r}_3, \bar{r}_4) = 1 - \cos m(\omega - \omega_0)$ здесь $\omega$ -угол между плоскостями, построенными по векторам $(\bar{r}_1, \bar{r}_2, \bar{r}_3)$ и $(\bar{r}_2, \bar{r}_3, \bar{r}_4)$	учитываются торсионные углы
$R_{\varphi, \rho}(\bar{r}_1, \bar{r}_2, \dots, \bar{r}_m) = \sum_{j=1}^m d^2(\bar{r}_j, L_{opt})$ здесь $L_{opt}$ -наилучшая плоскость для точек $\bar{r}_1, \bar{r}_2, \dots, \bar{r}_m$ ; $d(\bar{r}, L)$ -расстояние от точки $\bar{r}$ до плоскости $L$	учитываются плоские группы атомов в боковых цепях и пептидных звеньях
$R_{\varphi, h}(\bar{r}_1, \bar{r}_2, \bar{r}_3, \bar{r}_4) = (V - V_0)^2$ здесь $V$ -ориентированный объем тетраэдра, построенного на векторах $\bar{r}_1, \bar{r}_2, \bar{r}_3, \bar{r}_4$	учитываются четверки атомов $N^k, C_{\alpha}^k, C_{\beta}^k, C^k$

Таблица 2.  
Виды критериев, зависящих от структурных факторов

Критерий	Примечание
$\sum_j w_j  KF_j^{calc} - F_j^{obs} $	$K$ - шкальный коэффициент
$\sum_j w_j (KF_j^{calc} - F_j^{obs})^2$	$K$ - шкальный коэффициент
$\sum_j w_j (K(F_j^{calc})^2 - (F_j^{obs})^2)^2$	$K$ - шкальный коэффициент
$\sum_j (A_j \cos \rho_j^{calc} + B_j \sin \rho_j^{calc} + C_j \cos 2\rho_j^{calc} + D_j \sin 2\rho_j^{calc})$	$\rho^{calc}$ - фаза структурного фактора, рассчитанного по модели $A, B, C, D$ - коэффициенты Хендриксона и Латмана [4]

#### Литература

1. Ким К.В., Нестеров Д.Е., Черкасский Б.В. Оценка трудоемкости вычисления градиента // ДАН СССР. - 1984. - Т.275, №6. - С.1306-1309.
2. Lunin V.Yu., Urzhumtsev A.G. Program Construction for Macro - molecule Atomic Model Refinement Based on the Fast Fourier Transform and Fast Differentiation Algorithms // Acta Crystallogr. - 1985. - A41. - P.327-333.
3. Levitt M. Energy Refinement of Hen Egg-white Lysozym // J.Mol.Biol. - 1974. - V61.82. - P.393-420.
4. Hendrickson W.A., Lattman E.E. Representation of Phase Probability Distributions for Simplified Combination of Independent Phase Information // Acta Crystallogr. - 1970. - B26.-P.136-143.



Программное обеспечение ЭВМ  
Выпуск 86  
Компьютерные методы белковой инженерии

Редактор Л.Г.Радкевич  
Технический редактор А.А.Ракова

Подписано в печать 17.04.89. АТ 08749. Формат 60x84/16.  
Усл. печ. л. 7,7. Усл. кр.-отт. 8,05. Уч.-изд. л. 7,0.  
Тираж 300 экз. Заказ 148. Цена 35 к.

Институт математики АН БССР.  
220604, Минск, ул. Сурганова, II.  
Отпечатано на ротапринтере Института математики АН БССР.  
220604, Минск, ул. Сурганова, II.