

Описание программы AVERAGE

(Это описание справедливо для версии от 27.05.2001 и позже)

Содержание.

- 1. Назначение.**
- 2. Используемые понятия.**
- 3. Алгоритм работы программы.**
- 4. Входной файл управляющих данных:**
 - 4.1. Имя шага задания.**
 - 4.2. Входной файл наборов фаз в формате .OUT.**
 - 4.3. Номер пространственной группы.**
 - 4.4. Параметры элементарной ячейки.**
 - 4.5. Информация об эталонных фазах.**
 - 4.6. Информация об использовании экспериментальных модулей.**
 - 4.7. Диапазон разрешения для расчета расстояния между наборами фаз.**
 - 4.8. Возможность переворачивать синтез при расчете расстояния между наборами фаз.**
 - 4.9. Информация для FOM-статистики.**
- 5. Входной файл FAM.ITS.**
- 6. Выходной файл структурных факторов**
- 7. Выходной файл - протокол работы.**
- 8. Имеющиеся ограничения.**
- 9. Определение используемых понятий.**
- 10. Литература**

1. Назначение.

Программа AVERAGE предназначена для усреднения множества наборов фаз. Эти наборы фаз («варианты») могут быть предварительно получены программами GENMEM, FAMREF, RING либо другим способом и должны находиться в файле в формате .OUT.

2. Используемые понятия.

Набор фаз (вариант)

Эталонные фазы

Коэффициент корреляции и расстояние, отвечающие двум наборам фаз.

Выравнивание наборов фаз.

(Знание этих понятий необходимо для понимания дальнейшего текста. Краткое определение используемых понятий приведено в конце описания. Для более подробного знакомства с этими понятиями рекомендуется обратиться к используемой литературе.)

3. Алгоритм работы программы.

Перед усреднением фазовых наборов, содержащихся в файле, их необходимо сдвинуть к единому началу (при условии, что используемая пространственная группа допускает несколько альтернативных сдвигов начала координат). С этой целью сначала в программе находится центр облака, которое образуют в конфигурационном пространстве всех фазовых наборов содержащиеся в файле наборы фаз, а именно такой набор фаз, сумма квадратов расстояний от которого до всех других минимальна. Такой набор объявляется центральным. Затем все остальные наборы сдвигаются допустимыми сдвигами начала координат так, чтобы соответствие с центральным была максимальным. Сдвинутые фазы усредняются с получением усредненных фаз Ph_ave и показателей достоверности FOM. Затем программа производит усреднение в несколько других условиях с получением модулей T-calc и фаз Ph_calc (в частном случае эти результаты могут совпадать). Результаты обоих усреднений заносятся в выходной файл структурных факторов. В файл сообщений записывается некоторая статистика, отражающая процесс усреднения,

Несколько более строго все сказанное выглядит следующим образом:

Во входном файле содержится информация об экспериментальных модулях Fobs (для каждого рефлекса), фазах Ph (для каждого рефлекса каждого набора), а также еще некоторая информация F (для каждого рефлекса каждого набора), которая может быть либо весом, приписываемым фазе, либо индивидуальным модулем.

Расстояние между наборами фаз вычисляется как расстояние между синтетами Фурье. И при этом могут быть использованы как экспериментальные модули, одинаковые для одного рефлекса всех наборов, так и индивидуальные значения модулей.

Для каждого из наборов фаз, содержащихся в файле (за исключением, может быть, первого – в зависимости от заданных управляющих данных) вычисляется сумма расстояний до всех других наборов в заданном диапазоне по разрешению. Затем находится такой, для которого эта сумма является минимальной. Этот вариант объявляется центральным.

После этого каждый набор фаз посредством допустимых сдвигов начала координат сдвигается к центральному для достижения максимального соответствия.

Затем происходит усреднение сдвинутых наборов фаз с использованием экспериментальных значений модулей для всех рефлексов. При усреднении каждый рефлекс получает как значение фазы Ph_ave, так и показатель достоверности FOM.

Затем происходит усреднение фаз с учетом индивидуальных модулей. В результате такого усреднения получают значения рассчитанных модулей T-calc и Ph_calc.

После этого рассчитанные величины выводятся в выходной файл структурных факторов. В каждой записи такого файла содержится информация об одном рефлексе:

H K L Fobs Test_flag FOM Ph_ave T_calc Ph_calc Ph_ex

если точные фазы присутствовали во входном файле наборов фаз, либо

H K L Fobs Test_flag FOM Ph_ave T_calc Ph_calc

в противном случае.

В протокол заносится некоторая статистическая информация о фазах, содержащихся в файле, и корреляция усредненных фаз с эталонными, рассчитанная разными способами.

Замечание. В частном случае, когда все индивидуальные модули равны единице, величины FOM и T_calc должны совпадать между собой. Также в этом случае должны совпадать величины Ph_ave и Ph_calc.

4. Входной файл управляющих данных.

Общие замечания. Строки в файле управляющих данных, которые начинаются знаком “!”, воспринимаются как комментарии. Это позволяет для начинающего пользователя взять некоторый ГОТОВЫЙ ПОЛНЫЙ файл управляющих данных и адаптировать его к своим нуждам, попросту закомментарив ненужные строки. Все комментарии при этом очень полезно оставить в файле, поскольку они существенно облегчают изменение параметров и уменьшают число ошибок при их изменении.

Числовые данные вводятся по свободному формату.

Буква “Y” или “N” в ответах типа “ДА/НЕТ” может быть как прописной, так и строчной.

Буквы “R” или “D” в информации об единицах измерения углов могут быть как прописными, так и строчными.

Пример управляющих данных. Подробное описание каждого параметра содержится ниже.

```
!  
!   Script for the program AVERAGE (version 1.0)  
!  
! Code of step (up to 4 symbols)  
a8  
!  
! The path for the input file containing variants for averaging  
a7_gmem.out  
!  
! Space group number (in accordance with FAM.ITS dictionary)  
19  
!  
! Unit cell parameters (in Å and degrees)  
34.900 40.300 42.200 90.00 90.00 90.00  
!  
! The key (Y/N) for the presence of reference phases in the input  
! file and the [optional] key (D/R) for phase values unit  
! (degree/radian) in the input and output files  
! e.g. Y D or N R or Y etc.  
Y  
!  
! The next data define the mode to calculate the map correlation  
! coefficient:  
!  
! Whether Fobs will be used for all variants (Y) or individual  
! magnitudes will be taken from the input file for every  
! variant (N)  
Y
```

! Resolution limits (Dmin, Dmax) for the maps calculation and
! [optional] the step (in Å) for the search of the optimal
! alignment (may be important in the case of monoclinic or
! triclinic groups, e.g.)

12. 9999. 2.

! Whether the map flip is considered as a permitted map
! transformation(Y/N) in the alignment

N

! Number of resolution zones to calculate FOM-statistics

5

! Resolution limits (Dmin, Dmax) for every zone

16. 9999.

13. 9999.

12. 9999.

8. 9999.

4. 9999.

!

! the end of the script

4.1. Имя шага задания.

(! Code of step (up to 4 symbols))

Это не более чем 3-символьная метка, которая будет сопровождать все выходные файлы программы. Во время работы шага s1 будут созданы выходные файлы

s1_ave.uf - файл структурных факторов в формате .UF;

s1_ave.mes - протокол работы программы;

Рекомендуется часть символов имени сделать цифрами и увеличивать их в процессе работы.

4.2. Входной файл наборов фаз в формате .OUT.

(! The path for the input file containing variants for averaging)

Имя файла (не более 72 символов), где находятся наборы фаз, предназначенные для усреднения.

Структура файла следующая:

первая запись (формат A72) - символьная информация;

вторая запись (формат A72) - символьная информация;

третья запись (формат I4) - целое число NREF - количество рефлексов, для которых далее задана информация;

далее следуют NREF записей (формат 3I4, 2G12.4), одна на рефлекс, в каждой из которых содержатся индексы рефлексов H, K, L, экспериментальное значение модуля Fobs и тест-флаг Tstflg, принимающий значение 1 для рабочих рефлексов и 0 для свободных рефлексов. Программа AVERAGE эту информацию не использует;

далее следуют записи формата ((6G12.5)) или ((6G12.6)), по одной на каждый набор фаз. В одной такой записи содержится $2 \cdot NREF + 1$ вещественное число. В начале записи содержится одно число - значение некоторого критерия, которое программой AVERAGE используется только для сбора статистики. Затем следуют NREF вещественных чисел - значения фаз для каждого рефлекса $\Phi(i,j)$, где i - номер рефлекса, а j - номер варианта. Затем следуют еще NREF вещественных чисел $F(i,j)$. В зависимости от того, какой программой получен файл, это могут быть либо веса, приписываемые фазам каждого рефлекса, либо значения рассчитанных модулей (например, при восстановлении модулей неизмеренных рефлексов). Поэтому при работе с таким файлом необходим признак - будем ли мы при вычислении расстояния между наборами фаз использовать экспериментальные модули, то есть

одинаковые значения модулей для всех наборов фаз, либо свои индивидуальные значения модулей для каждого набора.

Фазы в файле могут быть заданы в градусах либо в радианах. Если входные фазы были заданы в градусах, то тогда используется формат ((6G12.5)). Если входные фазы были заданы в радианах, то используется формат ((6G12.6)).

Первая запись такого типа может содержать некоторые особые фазы, которые мы будем называть эталонные. Это могут быть как фазы, рассчитанные по модели (если для изучаемого объекта может быть построена какая-либо модель), так и фазы, полученные некоторым другим методом, с которыми нам будет интересно сравнить наше решение. Они могут быть очень далеки от правильного решения. Но в любом случае такие фазы не должны включаться в усреднение. Параметр Y/N, находящийся в файле управляющих данных после информации об элементарной ячейке, сообщает, является ли первая запись такого типа записью с эталонными фазами. Если этот параметр задан как "Y", то фазы в первой записи интерпретируются как эталонные и в усреднение не включаются.

4.3. Номер пространственной группы.

(! Space group number (in accordance with FAM.ITS dictionary))

Целое число, соответствующее номеру пространственной группы в Международных кристаллографических таблицах. Информация о такой группе должна содержаться в файле FAM.ITS, который поставляется вместе с программой. Этот файл может быть пополнен для пространственных групп, информация по которым еще не внесена авторами.

4.4. Параметры элементарной ячейки.

(! Unit cell parameters (in Å and degrees))

6 вещественных чисел - длины ребер ячейки и углы. Длины задаются в ангстремах, углы - в градусах.

4.5. Информация об эталонных фазах.

(! The key (Y/N) for the presence of reference phases in the input

! file and the [optional] key (D/R) for phase values unit

! (degree/radian) in the input and output files

! e.g. Y D or N R or Y etc.)

При создании файла в формате .OUT в качестве первого набора могли быть записаны некоторые особые фазы – возможно, точные, если структура уже известна. В таком случае эти фазы не следует включать в усреднение. В этой строке управляющих данных содержится ответ на вопрос, является ли первый набор фаз особым, не подлежащим усреднению. Такие фазы мы будем называть эталонными.

После этого по крайней мере через один пробел может следовать информация о том, в каких единицах заданы углы – в градусах (D/d) или радианах (R/r). Алгоритм использования этой информации следующий:

В любом случае, независимо от того, задана ли информация о единицах измерения углов, программа делает попытку проинтерпретировать единицу измерения фаз самостоятельно. Для этого вычисляется среднее значение суммы абсолютных величин первого набора фаз. Если это среднее по абсолютной величине не превосходит 5, то фазы интерпретируются в радианах; если среднее оказывается не меньше 50, то фазы будут интерпретируются в углах. В противном случае программа считает, что углы не могут быть проинтерпретированы.

Возможны 4 случая сочетания наличия входной информации и успешности интерпретации.

1. Входной информации нет, самостоятельная интерпретация безуспешна

Программа заканчивает работу с аварийным сообщением.

2. Входной информации нет, самостоятельная интерпретация успешна.

Используются единицы, полученные в результате самостоятельной интерпретации.

3. Входная информация есть, самостоятельная интерпретация безуспешна.

Используется входная информация.

4. Входная информация есть, самостоятельная интерпретация успешна.

И тогда возможны 2 варианта:

4а. Единицы измерения, полученные независимым образом, совпадают.

Они и используются.

4б. Единицы измерения находятся в конфликте.

Используются единицы, заданные во входной информации.

Выдается предупреждающее сообщение.

4.6. Информация об использовании экспериментальных модулей.

(! Whether Fobs will be used for all variants (Y) or individual

! magnitudes will be taken from the input file for every

! variant (N))

При создании файла наборов фаз всегда задается информация об экспериментальных модулях. Кроме того, в некоторых случаях каждый набор фаз снабжается своим индивидуальным набором модулей. Эта строка управляющих данных содержит информацию о том, будет ли при подсчете корреляции синтезов использоваться набор экспериментальных модулей, один для всех наборов фаз, либо синтезы будут рассчитываться со своими индивидуальными модулями. Буква Y означает, что будет использован экспериментальный набор модулей.

4.7. Диапазон разрешения для расчета расстояния между наборами фаз.

(! Resolution limits (Dmin, Dmax) for the maps calculation and

! [optional] the step (in A) for the search of the optimal

! alignment (may be important in the case of monoclinic or

! triclinic groups, e.g.)

2 или 3 вещественных числа, из которых первые 2 являются обязательными. Они задают минимальное и максимальное разрешение, определяющие зону в обратном пространстве, в которой будет подсчитываться корреляция сгенерированных фаз с эталонным набором (если он есть) либо с первым набором в файле (если эталонные фазы отсутствуют). Эти 2 числа могут быть заданы в любом порядке. При расчете корреляции сравниваемые наборы фаз предварительно выравниваются допустимыми для данной пространственной группы изменениями начала координат и/или энантиomorфа (информация о допустимых сдвигах и изменении энантиomorфа задается в файле FAM.ITS). Если для одной из осей возможен любой выбор начала координат (например, сдвиг вдоль оси Y для пространственной группы P21), то третье число в строке управляющих данных задает шаг, с которым будут проверяться сдвиги при выравнивании вдоль этой оси. Это третье число может отсутствовать, и тогда программа делает попытку назначить длину шага как Dmin/4.

4.8 Возможность переворачивать синтез при расчете расстояния между наборами фаз.

(! Whether the map flip is considered as a permitted map

! transformation(Y/N) in the alignment)

Этот параметр определяет, будет ли при выравнивании фаз рассматриваться не только “прямой”, но и “перевернутый” синтез. Если число допустимых сдвигов равно, например, 4, указана возможность смены энантиomorфа, а возможность “переворачивать” синтез не указана, то в качестве корреляции пробного синтеза с точным будет выбрано максимальное из 8 чисел, каждое из которых соответствует корреляции при некотором допустимом сдвиге и любом выборе энантиomorфа. А если возможность “переворачивать” синтез указана, то будет выбрано максимальное из 16 чисел, образующих 8 пары; в каждой паре одно число

является корреляцией “прямых” синтезов, а другое - корреляцией “перевернутого” пробного и “прямого” точного синтеза.

4.9. Информация для FOM-статистики.

! Number of resolution zones to calculate FOM-statistics

5

! Resolution limits (Dmin, Dmax) for every zone

16. 9999.

13. 9999.

12. 9999.

8. 9999.

4. 9999.

!

Некоторой характеристикой качества усредненных фаз является статистика их показателей достоверности. В процессе усреднения каждый рефлекс получает показатель достоверности FOM. В программе подсчитывается их среднее значение в заданных зонах по разрешению. Высокое среднее значение свидетельствует о том, что фазы находятся недалеко друг от друга (в смысле расстояния между фазами в конфигурационном пространстве), и, следовательно, можно надеяться, что результат их усреднения будет находиться недалеко от правильного решения.

Этот раздел управляющих данных содержит целое число – количество зон NZON. Если это число не равно нулю, то далее следуют NZON строк, каждая из которых содержит 2 вещественных числа Dmin и Dmax - границы зон по разрешению для FOM-статистики (они могут следовать в любом порядке).

5. Входной файл FAM.ITS.

Входной файл FAM.ITS содержит информацию о некоторых пространственных группах в виде, доступном для программы. Ниже приведена информация для группы P212121. Если файл не содержит информацию о группе, необходимой для вашей работы, вы можете сами пополнить его по приведенному здесь образцу либо обратиться к разработчикам.

NEWGROUP P212121

заголовок, отмечающий начало нового блока информации

19 (the group number)

номер пространственной группы (по этому номеру этот блок находится программами, использующими этот файл)

4 (number of symmetries)

количество элементов симметрии в пространственной группе

1 0 0 0 1 0 0 0 1 0 0 0

-1 0 0 0 -1 0 0 0 1.5 0.5

1 0 0 0 -1 0 0 0 -1 .5 .5 0

-1 0 0 0 1 0 0 0 -1 0.5 .5

коэффициенты уравнений симметрии в следующем порядке

$\Gamma_{11}, \Gamma_{21}, \Gamma_{31}, \Gamma_{12}, \Gamma_{22}, \Gamma_{32}, \Gamma_{31}, \Gamma_{32}, \Gamma_{33}, t_1, t_2, t_3$

3 (number of centrosymmetric zones)

число зон, содержащих центр симметрии

0 0 1 .5 0 0

0 1 0 0 0.5

1 0 0 0.5 0

каждая из центросимметричных зон должна быть определена

шестью параметрами: $m_1, m_2, m_3, a_1, a_2, a_3$:

рефлекс hkl принадлежит этой зоне тогда и только тогда, когда $m_1 \cdot h + m_2 \cdot k + m_3 \cdot l = 0$;

при этом допустимые значения фазы для этого рефлекса:

$$\alpha = (a_1 \cdot h + a_2 \cdot k + a_3 \cdot l) \cdot \pi \quad \text{либо} \quad \alpha + \pi.$$

0 (number of axes with any shift of origin)

определяет число осей (0, 1 или 3), таких, что любой сдвиг вдоль них приводит к эквивалентному выбору начала координат

8 (number of the possible origin choices)

количество вариантов дискретного выбора эквивалентных начал координат; если допустимы произвольные сдвиги начала вдоль какой-то из осей, то эти сдвиги добавляются к каждому из дискретных вариантов выбора начала;

0 0 0

.5 0 0

0 .5 0

.5 .5 0

0 0 .5

.5 0 .5

0 .5 .5

.5 .5 .5

соответствующие начала координат

1 (possibility of enantiomorph choice, 1 - if possible)

указывает, совпадает ли энантиомер данной группы с ней самой

6. Выходной файл структурных факторов.

Выходным файлом программы является файл структурных факторов с именем

<имя_шага>_ave.uf

Этот форматированный файл организован следующим образом (так называемый формат UF):

- первая запись (72 символа) - первый заголовок файла; эта информация формируется следующим образом: первые 22 символа – текст «Output data of AVERAGE», затем следуют параметры элементарной ячейки и номер пространственной группы;

- вторая запись (72 символа) - второй заголовок файла; он будет таким:

H K L Fobs Test_flag FOM Ph_ave T_calc Ph_calc

если эталонные фазы не присутствовали во входном файле, и таким:

H K L Fobs Test_flag FOM Ph_ave T_calc Ph_calc Ph_ex

если эталонные фазы присутствовали во входном файле;

- третья запись (целое число, которое читается по формату I4) - LREC, размер каждой из последующих записей (количество чисел в одной записи; не должно превышать 100; как синоним понятия “количество чисел” мы будем также использовать выражение “количество колонок”);

- далее до конца файла следуют одинаковые записи, состоящие из LREC чисел (которые называются также колонками), отвечающих одному рефлексу; из этих чисел первые три - целые (это индексы H, K, L), а остальные – вещественные, содержащие разнообразную информацию о данном структурном факторе. Такие записи вычитываются по формату

(3I4, 5G12.6) или (3I4, 5G12.6)

если LREC не больше 8, и по формату

(3I4, 5G12.5/(6G12.5)) или (3I4, 5G12.5/(6G12.5))

в противном случае.

В этом файле все записи содержат информацию одинакового типа в одной и той же позиции, в соответствии со вторым заголовком (экспериментальные модули в 4-ой позиции, усредненные фазы в 7-ой, их показатели достоверности в 6-ой и так далее).

Замечание. Формат G12.6 используется в том случае, если углы во входном файле наборов фаз (а значит, и в выходном файле) задаются в радианах. Если же углы задаются в градусах, то вместо этого используется формат G12.5.

7. Выходной файл - протокол работы.

Протокол работы имеет имя <имя_шага>_ave.mes.

Вот пример протокола работы программы AVERAGE:

```

                                AVERAGE                                Version 1.0    27.05.2001
Input file name: a7_gmem.out
Input file titles are:
protg   34.9 40.3 42.2 90. 90. 90.   P212121
h k l d Fobs F(mod) Phi(mod)
Space group number: 19
Unit cell:   34.90  40.30  42.20    90. 90. 90.
Whether reference phase values are present in the input files: Y
No RADIAN/DEGREE information is defined in the script
Maps alignment mode:
  Whether Fobs will be used for all maps: Y
  resolution limits:   12.00 - 9999.00
  minimal trial step (in A) for the origin shift:    2.00
  whether the map flip is a permitted transformation: N
Number of resolution zones for FOM-statistics:  5
  the corresponding resolution limits:
    16.00- 9999.00
    13.00- 9999.00
    12.00- 9999.00
     8.00- 9999.00
     4.00- 9999.00

***** Input file statistics *****

Phases are interpreted in radians
  20 phase variants was found in the input file
The search criterion values (qmin, qmax, ave, rms):
    0.1801      0.8705      0.5748      0.1907

Map correlation with respect to the reference phases:
  cmin,cmax,ave,rms:   0.2098  0.6915  0.3880  0.1223

    0    0    0    0    0    0    0    0    0    0
    0    0    0    0    0    0    0    0    0    0
    0    0    0    0    1    3    5    6    2    0
    0    0    1    2    0    0    0    0    0    0

The radius of gyration:
rmin,rmax,ave,rms:   1.00    1.13    1.05    0.04
    0    0    0    0    0    0    0    0    0    0
    0    0    0    0    0    0    0    0    0    1
   11    6    2    0    0    0    0    0    0    0
    0    0    0    0    0    0    0    0    0    0

***** Averaging *****

The program has defined the variant   18 as a middle one
```

Map correlation of input variants with respect to the middle phases:

cmin,cmax,ave,rms: 0.2487 1.0000 0.5047 0.1613

0	0	0	0	0	0	0	0	0	0
0	0	0	0	0	0	0	0	0	0
0	0	0	0	1	0	2	1	5	3
1	4	1	0	1	0	0	0	0	1

Map correlation with respect to reference phases for:

(F_obs,P_ave)-map: 0.2429

(F_obs,FOM,P_ave)-map: 0.4825

(F_obs,P_calc)-map: 0.2429

(T_calc,P_calc)-map: 0.3194

FOM-statistics:

resolution	ref	<FOM>
16.-9999.	15	0.3426
13.-9999.	22	0.3407
12.-9999.	28	0.3364
8.-9999.	85	0.2459
4.-9999.	580	0.2005

В протокол заносится, прежде всего, вся информация из входного файла управляющих данных.

Если не была задана информация о том, в радианах или градусах приведены реперные фазы, программа сообщает о своих результатах интерпретации фаз (в приведенном ниже протоколе они проинтерпретированы в радианах; это говорит о том, что и выходные фазы будут выражены в радианах).

Далее в протоколе содержатся :

1. Некоторая статистика входного критерия (минимальное, максимальное, среднее и среднеквадратичное отклонение). Входной критерий – это первое вещественное число в записях с набором фаз во входном файле.
2. Статистика корреляции входных фаз с эталонными (минимальное, максимальное, среднее, среднеквадратичное отклонение и гистограмма значений корреляции).
3. Характеристика «облака» входных фазовых наборов (так называемый радиус инерции).
4. Номер варианта, который является центральным для этого облака.
5. Статистика корреляции по отношению к «центральному» набору фаз (минимальное, максимальное, среднее, среднеквадратичное отклонение и гистограмма значений).
6. 4 величины корреляции эталонного синтеза (эталонные фазы, экспериментальные модули) с различными синтезами: (F_obs,P_ave), (F_obs,FOM,P_ave), (F_obs,P_calc), (T_calc,P_calc). Наиболее интересной характеристикой из этих четырех представляется вторая – корреляция эталонного синтеза со взвешенным усредненным.

В случае, когда все индивидуальные модули равны единице, первая и третья величины должны совпадать.

7. Средние значения показателей достоверности, рассчитанных в заданных зонах по разрешению.

8. Имеющиеся ограничения

В программе приняты некоторые ограничения, изменить которые можно, перетранслировав программу. Это

- максимально возможное число рефлексов во входном файле - 5000;
- максимально возможное число симметрий в пространственной группе - 48;
- максимально возможное число centrosymmetric зон в файле FAM.ITS - 20;

- максимально возможное число сдвигов начала координат в файле FAM.ITS - 8;
- максимально возможный размер массива, используемый при подсчете корреляции с использованием “выравнивания” карт для моноклинных или триклинных групп - 10000. Размер этого массива вычисляется как $a_{\text{cell}}/d_{\text{step}}$ для случая моноклинной группы при выделенной оси **a** (или, соответственно, $b_{\text{cell}}/d_{\text{step}}$ для случая моноклинной группы при выделенной оси **b**) и $a_{\text{cell}}*b_{\text{cell}}*c_{\text{cell}}/d_{\text{step}}^3$ для случая триклинной группы. Если ваши данные таковы, что это ограничение оказалось нарушенным, рекомендуется увеличить задаваемую величину шага d_{step} ;
- максимально возможное число вариантов – 1000;
- максимально возможное число зон для статистики показателей достоверности (FOM) – 20;
- максимально возможное число бинов для гистограмм – 40.

9. Определение используемых понятий.

Набор фаз (вариант).

Для входного файла структурных факторов, содержащего NREF рефлексов, это NREF чисел, каждое из которых представляет собой фазу очередного рефлекса (в радианах или градусах, в зависимости от используемого признака, описанного в 4.5). В качестве синонима к словосочетанию “набор фаз” могут быть использованы слова “фазовый вариант”.

Эталонные фазы.

Некоторый набор фаз, по отношению к которым рассчитывается корреляция фаз. Это могут быть как фазы, рассчитанные по известной модели, так и полученные некоторым независимым способом. Эталонные фазы используются не для определения фаз текущего шага, а лишь для статистики распределения корреляции и аналогичной информации. Фактически они могут быть очень далеки от правильного решения.

Коэффициент корреляции и расстояние, отвечающие двум наборам фаз.

В качестве основного показателя, характеризующего степень близости двух наборов фаз $\{\varphi_1(\mathbf{h})\}$ и $\{\varphi_2(\mathbf{h})\}$, в программе используется коэффициент корреляции карт распределения электронной плотности, рассчитанных с этими фазами:

$$C = \frac{\int \rho_1(\mathbf{r})\rho_2(\mathbf{r})dV_r}{\sqrt{\int \rho_1(\mathbf{r})^2 dV_r} \sqrt{\int \rho_2(\mathbf{r})^2 dV_r}} = \frac{\sum_{\mathbf{h}} F_1(\mathbf{h})F_2(\mathbf{h})\cos(\varphi_1(\mathbf{h}) - \varphi_2(\mathbf{h}))}{\sqrt{\sum_{\mathbf{h}} F_1(\mathbf{h})^2} \sqrt{\sum_{\mathbf{h}} F_2(\mathbf{h})^2}}$$

Этот коэффициент корреляции зависит не только от значений фаз, но и от значений модулей структурных факторов, с которыми рассчитываются эти карты. Программа допускает две возможности при его расчете:

- использовать всегда один и тот же набор экспериментальных модулей;
- для каждого набора фаз брать соответствующий ему индивидуальный набор модулей структурных факторов.

Иногда удобно вместо коэффициента корреляции карт анализировать «расстояние» между ними, определяемое как

$$D = \sqrt{\int \left(\frac{\rho_1(\mathbf{r})}{\sqrt{\int \rho_1(\mathbf{r})^2 dV_r}} - \frac{\rho_2(\mathbf{r})}{\sqrt{\int \rho_2(\mathbf{r})^2 dV_r}} \right)^2 dV_r} = \sqrt{2(1 - C)}$$

Два набора фаз тем ближе, чем меньше это расстояние.

Следует также отметить, что значение коэффициента корреляции зависит от того, по какому набору рефлексов (по какой зоне разрешения) определяется эта корреляция.

Выравнивание наборов фаз.

Два набора фаз, будучи формально совсем разными, могут приводить к почти одинаковым изображениям на картах, отличающихся только выбором начала координат. Поэтому, прежде чем оценивать близость карт, их необходимо максимально хорошо «выровнять» сдвигом одной карты относительно другой. При работе в группе P1 должны анализироваться всевозможные сдвиги. При наличии у объекта других групп симметрии при выравнивании рассматриваются только «допустимые» для этой группы сдвиги карты (допустимые выборы начала координат), то есть такие, которые сохраняют вид уравнений симметрии карты. Набор допустимых сдвигов, вообще говоря, свой для каждой пространственной группы.

Если при переходе к энантиоморфной структуре пространственная группа симметрий не изменяется, то при выравнивании наборов фаз в качестве допустимого преобразования второй карты рассматривается также возможность ее замены энантиоморфной ей картой.

При работе с макромолекулярными объектами при низком разрешении иногда целесообразно добавить к числу допустимых преобразований второй карты ее «переворачивание»: $\rho(\mathbf{r}) \rightarrow -\rho(\mathbf{r})$ (модули структурных факторов у таких карт совпадают, а фазы различаются на π).

Оптимальным считается такое преобразование второй карты (допустимый сдвиг, плюс, может быть, переход к энантиоморфу, плюс, может быть, переворачивание карты), при котором ее корреляция с первой картой становится максимальной. Коэффициент корреляции (или расстояние) между выровненными картами и рассматривается в программе как адекватная оценка степени близости двух фазовых наборов. Все значения, выдаваемые программой, относятся к выровненным фазовым наборам.

10. Литература.

1. Lunin, V.Yu. & Woolfson, M.M. (1993) "The mean Phase Error and the Map Correlation Coefficient". Acta Cryst. **D49**, 530-533.
2. Lunin, V.Yu. & Lunina, N.L. (1996) "The Map Correlation Coefficient for Optimally Superposed Maps". Acta Cryst. **A52**, 365-368.
3. Lunin V.Y., Lunina N.L., Petrova T.E., Skovoroda T.P., Urzhumtsev A.G. & Podjarny A.D. (2000) "Low-resolution ab initio phasing: problems and advances". Acta Cryst. **D56**, 1223-1232.