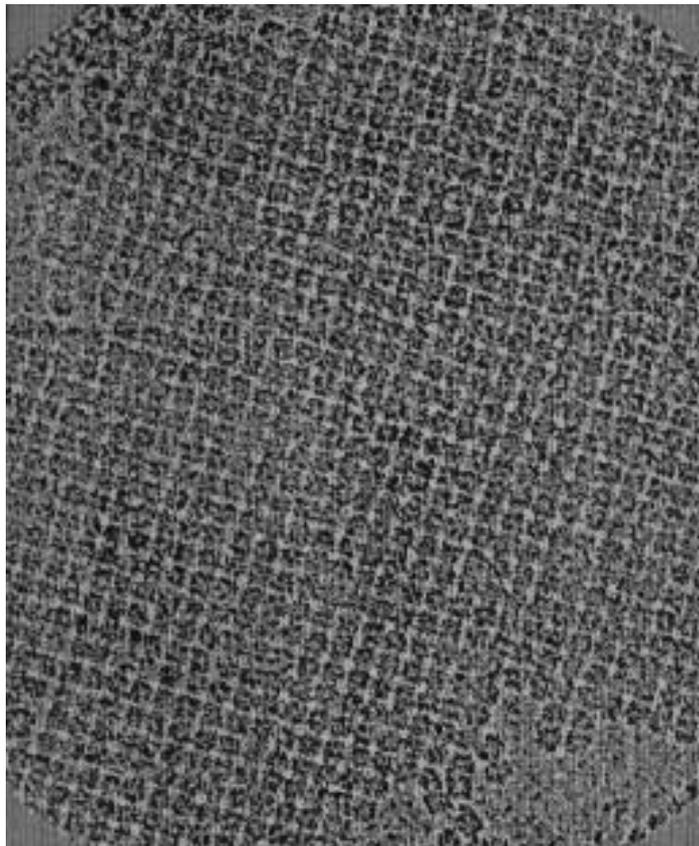


2-D CRYSTALS IN IMAGIC-5

HANDS-ON



In the following you will find a list of the IMAGIC commands to analyse 2-D crystals in IMAGIC-5.

All rights reserved

© Image Science Software GmbH
Berlin, Germany

Version: 14-May-2003

2-D crystals in IMAGIC-5

Please note that there are many other commands that you probably can use during your analysis. Do not hesitate to "go your own way".

Title page:

Negatively stained 2-D crystal of bacterial D-Ribulose-1,5-Bisphosphate Carboxylase / Oxygenase (RuBisCO) after band-pass filtering (with friendly permission of Dr. Andreas Holzenburg, University of Leeds, UK)

2-D crystals in IMAGIC-5

Don't hesitate to punch "?" to understand the meaning of a certain **IMAGIC-5** question). **IMAGIC-5** will give you help.

(1) STARTING THE IMAGIC-5 SYSTEM

Call **IMAGIC**:

```
imagic
```

or

```
i
```

(or any other command name you have defined in your `.cshrc` or in your `login.com`).

(2a) SCAN MICROGRAPHS & CREATE IMAGIC IMAGES

Scan the desired electron micrographs with your scanner and copy them to your workstation (see corresponding manuals).

Use **IMAGIC** command **IMPORT-EXPORT** to convert the images into **IMAGIC** format. Let us assume that the scanned micrograph is stored in TIFF format in the file "`my_crystal.tif`":

```
IMAGIC-COMMAND: IMPORT-EXP
** EM2EM welcomes you **
```

2-D crystals in IMAGIC-5

```
IMAGIC-COMMAND: IMPORT-EXP

** EM2EM welcomes you **

Data format of the input to be converted : TIFF
Export to which data format [?]          : IMAGIC
Input = set of 2D sections of a 3D volume: NO
Import a SET of input image files [NO]   : NO
Input image file [?]                     : my_crystal
                                           You select

Output image file (NO EXT.) [?]          : my_crystal
                                           You select

Pixel spacing in Angstroem (X,Y,Z)      : 5.0
                                           You select
```

For further information on how to use the **IMPORT-EXPORT** command please see the "**em2em**" manual.

(2b) CHECK THE IMAGES

Use IMAGIC command **DISPLAY** to check the images.

You can also calculate a power spectrum of your crystalline image to check its quality. Call **FFT2D** and use option **FORWARD** (see chapter (6)).

Then **DISPLAY** the Fourier transform. **DISPLAY** will automatically show you the (logarithm of the) power spectrum of the transform.

(3) IMAGE SIZE

One will normally extract from the larger raw image a smaller area containing the crystal of interest. To find the co-ordinates of the smaller area, we use the COOS option together with the NO STORE option in the DISPLAY command:

```
IMAGIC-COMMAND: DISP
** DISPLAY welcomes you **
Input image file, image loc#s           : my_crystal
                                           you select
           ...
Change parameter settings
(ERASE,SCALE,LOC,START,SCAN,BYE...)? [NO] : COOS
Should the values be stored in a PLT file?: NO STORAGE
           ...
Change parameter settings
(ERASE,SCALE,LOC,START,SCAN,BYE...)? [NO]: NO
```

One normally needs to know the top-left X-Y co-ordinates as well as the bottom-right co-ordinates. To find the corresponding co-ordinates move the cursor to the top left edge and press the middle button of the mouse. DISPLAY will print the co-ordinate values on the screen. Continue with the lower right edge. If you want to stop select the same point twice.

Continue with the next crystalline area or leave DISPLAY by typing BYE or EXIT.

2-D crystals in IMAGIC-5

This way we know how large the extracted image should be and where it starts. The actual extraction is performed by the command **CUT-IMAGES** (option **GENERAL**):

```
IMAGIC-COMMAND: CUT-IM

** SUBFILE welcomes you **

Input file, image loc#s      : my_crystal           you
Output file, image loc#s    : crys_patch,1         select

Mode of operation:
  APERIODIC,  CENTER,      GENERAL,  PAD_GENERAL,
  LARGE_PAD,  SMALL_PAD,   PERIODIC,  SQUARE,
  CHECKERS
Give mode of operation      : GENERAL
Output file dimensions X,Y  : 100,100
Top left X, Y coordinates   : 23,18
in the old image

you select
according to the
values found with
DISPLAY/COOS
```

Another way to extract the crystalline area is to take the coordinates of the centre of the desired region.

These coordinates are determined with command **DISPLAY** (option **COOS**) and stored in a file (**COOS** option: **YES STORE**):

2-D crystals in IMAGIC-5

```
IMAGIC-COMMAND: DISP

** DISPLAY welcomes you **
Input image file, image loc#s           : my_crystal
                                           you select
    ...

Change parameter settings
(ERASE,SCALE,LOC,START,SCAN,SAVE,BYE...)? : COOS
Should the values be stored in a PLT file? : YES
Output PLT file [my_crystal]             : my_crystal
    ...

Change parameter settings
(ERASE,SCALE,LOC,START,SCAN,SAVE,BYE...)? : NO
```

In the example the central coordinates are stored in file [my_crystal.plt](#).

You can again use command **CUT-IMAGE**, now with option **APERIODIC** to extract the desired crystalline patch(es):

```
IMAGIC-COMMAND: CUT-IM

** SUBFILE welcomes you **

Input file, image loc#s           : my_crystal
Output file, image loc#s         : crys_patch,1
                                           you select

Mode of operation:
  APERIODIC,  CENTER,      GENERAL,   PAD_GENERAL,
  LARGE_PAD,  SMALL_PAD,  PERIODIC,  SQUARE,
  CHECKERS
Give mode of operation           : APERIODIC
Output file dimensions X,Y       : 100,100
                                           you select
                                           according to the
                                           values found with
                                           DISPLAY/COOS
```

2-D crystals in IMAGIC-5

```
Coordinates (plt) file      : my_crystal.plt
```

**output from
DISPLAY
(option COOS)**

NOTE:

The image dimensions should be chosen such that they are compatible with the FFT algorithm (preferably factors of 2, 3 and 5). To find out interactively which sizes are compatible, type a ? when the **CUT-IMAGES** command prompts you for the image size.

Finally, make the image size square. Although this is not strictly necessary, this will facilitate your thinking. If you have a rectangular region of interest, you can impose an ellipse on the square image later on using **PRETREAT-IMAGE** (see later).

(4) BAND-PASS FILTER

This is an operation that can be performed at different stages of the analysis. However, it shall be introduced here since it is a form of image preparation and can be logically done here. This is an optional operation. Proceed to section (5) if desired.

A band-pass-filter can be used to suppress the very low and the very high spatial frequencies. The very high spatial frequencies represent noise that will affect the analysis, and so we would like to ameliorate its effects quite early on. The very low spatial frequencies represent things like average densities and density ramps that depend on illumination conditions and have nothing to do with the molecular structure in which we are interested. For crystal analysis, all the low frequency information up to the first order reflections can be left out without affecting the information we are interested in. For correlation averaging it even should be left out.

2-D crystals in IMAGIC-5

The **(INCORE-)BAND-PASS-FILTER** command Fourier transforms (FFT) the input image and imposes a mask on it that attenuates low and high frequencies, but lets intermediate ones through. The inverse Fourier transform is also computed automatically. The parameters requested are the low frequency cut-off and the high frequency cut-off. These depend on the situation at hand, e.g., how much noise is present in your images and how finely they have been digitised. However, the default values are reasonable. Also, one is requested for the percentage remaining transmission of low frequency. For example, a value of 0.02 means that the low frequency Fourier components are NOT set to zero but instead are multiplied by 0.02. Thus, although they are suppressed now they can be re-boosted to their original values using the command **INVERSE-BAND-PASS-FILTER**. You cannot retrieve the high frequency components, however.

```
IMAGIC-COMMAND: INC-BAND-PASS

** INCBAND welcomes you **

Give input file, loc#s           : crys_patch           you
Give output file, loc#s          : crys_filt       select

Inverse band-pass filter         : NO
The image will be band-pass filtered
Please specify:
Low frequency cut off           : 0.05            have to be adapted
Remaining low-freq. transm.     : 0.005          (see text above),
High frequency cut off          : 0.02           to get started the
                                   default values can
                                   help you

ASQ filter the images too?      : NO
```

If you want to calculate your frequency cut-off in terms of spatial frequency, then here is an example. Say that the image is scanned such that each pixel is of size x . Then in the Fourier transform, the edge of the image corresponds to the maximum spatial frequency $1 / (2 * x)$. Remember that the centre of the transform is zero spatial frequency. Thus a fraction f (between 0 and 1) corresponds to a spatial frequency of

$$\frac{f}{2 * x}$$

Note:

If your images are too large (i.e., do not fit in core) please use command **BAND-PASS-FILTER**.

(5) PREPARATION OF THE CRYSTAL IMAGE

Prior to any analysis one will normally prepare the image so to avoid some obvious artefacts and to bring the statistical properties of the image such as average density and variance to standard values.

(5a) IMPOSE-PLT-CONTOUR

If the crystal does not quite fill the image one may mask the surrounding background away by contouring the crystal area. The coordinates of the contour are determined with command **DISPLAY** using the **COOS** option. The mask is actually imposed later using the command **IMPOSE-PLT-CONTOUR**. The image statistics inside the contour are normalised meaning that the average density is set to zero and the standard deviation of the image to 10:

```
IMAGIC-COMMAND: IMPOSE-PLT

** CONTPLT welcomes you **

Input file, image loc#s           : crys_filt
Contour file name                  : crys_filt      you
Output file, image loc#s          : crys_norm      select
Normalise densities inside contour : YES
```

(5b) PRETREAT-IMAGE

If the crystal fills the larger part of the image one may standardise the statistics in the whole image or within a circular mask in the image, in which situations one uses the command **PRETREAT-IMAGE**.

```
IMAGIC-COMMAND: PRETREAT-IMAGE

** PRETREAT welcomes you **

Input file, image loc#s           : crys_filt       you
Output file, image loc#s         : crys_norm     select
Mode of pretreatment:
  MASKONLY  NORMVARIANCE  NVNOMASK  ZEROFLOAT
Please specify option              : NORMVAR

Pretreatment mask to be used
  CIRCLE  RECTANGLE        : WHOLE

Give the desired new sigma        : 10.0
```

NOTE:

In all important operations (FFT) during the analysis, images are treated as being cyclical. It is therefore very important to avoid having the image information extend to the edge of the image. This can be avoided by using a mask (as above) or by padding an image with an edge of zeroes (**IMAGIC** command: **CUT-IMAGES** with options **SMALLPAD** or **GENPAD**).

(5c) MULTIPLY-BY-CONSTANT

Normally images are analysed in which protein (in general: the density of the particles to be analysed, correlated etc.) appears light on the display, whereas negative stain, for example, appears dark. After floating and normalisation of densities above, this means that higher (positive) density values should represent protein, i.e., desired structural information. This may not be the case with your images. If not, then use the command **MULTIPLY-BY-CONSTANT**, option **MULTIPLY** to multiply your image by -1.0 at this stage. Otherwise, just carry on.

2-D crystals in IMAGIC-5

```
IMAGIC-COMMAND: MULTIPLY-BY-CONST

** ARITHM welcomes you **

Input file, image loc#s           : crys_norm
Output file, image loc#s         : crys_norm
Choose arithmetic operation       : MULT
Give CONSTANT to be multipl. to  : -1.0
```

(6) THE FOURIER TRANSFORM AND THE POWER SPECTRUM

For periodic objects like a 2D-crystal of biological macromolecules it is worth looking at the Fourier transform since the periodic information is concentrated on a regular "reciprocal lattice". All the non-periodic information lies outside the reciprocal lattice peaks and that allows us to, for example, place a mask over the transform that lets the peaks through and stops everything else (**FOURIER-PEAK-FILTERING**).

The Fourier transform from the input image is calculated by using the command **FFT2D** with option **FORWARD**:

```
IMAGIC-COMMAND: FFT2D

** FFT2D welcomes you **

Input file, image loc#s           : crys_norm   You
Output file, image loc#s         : crys_fft     select
Forward or Reverse Fourier transform : FORWARD
```

2-D crystals in IMAGIC-5

We can then look at the Fourier transformed image by using the DISPLAY command. If not otherwise specified **DISPLAY** will by default display the logarithm of the squares (i.e. the log the power spectrum) whenever complex (Fourier transformed) data are to be displayed. The default **GREYVALUES** are 0.0 and 5.0, respectively. It may be worth playing a bit with these values to find the optimum representation of the power spectrum. For example, 1.5 to 4.0 :

```
IMAGIC-COMMAND: DISP

** DISPLAY welcomes you **

Input image file, image loc#s           : crys_fft
                                           you select,
                                           Fourier transform
                                           of your crystal
                                           ...

Change parameter settings
(ERASE,SCALE,LOC,START,SCAN,BYE...)? [NO] : GREY

Please give option to get the greyvalues
(AUTOMATIC,INTERACTIVE,WINDOW)[AUTO]     : INTERACTIVE
Give black, white levels (0,0 = automatic): 1.5,4
                                           ...

Change parameter settings
(ERASE,SCALE,LOC,START,SCAN,BYE...)? [NO]: NO
```

NOTE:

The Fourier transform is always transposed (X-Y exchanged) as compared to the input image. For publications, etc., the transposing can be undone by using the command **TRANSPPOSE-IMAGE**.

(7) DETERMINING THE RECIPROCAL LATTICE VECTORS

The first step is to get an approximate idea about the positions of the diffraction peaks. For this purpose we **DISPLAY** the Fourier transformed image and then use the **COOS** option (reads the cursor position in terms of the image coordinates) to position reciprocal lattice peaks.

```
IMAGIC-COMMAND: DISP

** DISPLAY welcomes you **

Input file name, image loc#s           : crys_fft
                                         Fourier transform
                                         of your crystal
                                         ...

Change parameter settings
(ERASE,SCALE,LOC,START,SCAN,BYE...)? [NO] : COOS
Should the values be stored in a PLT file?: NO
                                         ...

Change parameter settings
(ERASE,SCALE,LOC,START,SCAN,SAVE,BYE...)? [NO] : NO
```

You have to index the diffraction pattern, that is, to determine what the unit cell is. See handbooks on diffraction methods and look for the 17 possible plane patterns.

Then we use the command **FOURIER-PEAK-SEARCH** to get the precise positions of the diffraction peaks. In this command, a least-squares fit is performed on a 3x3 environment of the peak to find refined peak coordinates. A list of the peaks is stored in a "plt" file which you should send to a printer.

2-D crystals in IMAGIC-5

```
IMAGIC-COMMAND: FOURIER-P-S

** ALLPEAK welcomes you **

Input file, image loc#s           : crys_fft
                                   you select,
                                   Fourier transform
                                   of your crystal

Output filename for peak coordinates : crys_peaks.plt
                                   you select,
                                   output "plt" file
                                   with coordinates

Linear size of search area (3-5-7-9-11) : 5      you select

Minimal distance allowed between peaks : 100     you select,
                                                unit cell is
                                                minimum

Minimum distance from the edge of image?: 50     you select,
                                                ½ unit cell
                                                is minimum

Mode of operation
  NORMAL, DEVIATION                : normal
Number of peaks wanted maximally   : 300
                                   depends on your
                                   crystal, be conservative
                                   at the beginning
```

Instead of using option NORMAL you may prefer option DEVIATION in which peaks are only selected if the peak height is higher than a specified standard deviation added to the background (type ? to get help).

Look for the peaks that fall on the lattice and calculate the refined reciprocal lattice coordinates from these refined peak positions. Note: because of the Hermitian symmetry the diffraction peaks come in pairs (Friedel pairs).

2-D crystals in IMAGIC-5

You can then check the positions of the peaks with command **DISPLAY** and option **PLOT**.

Finally, use the command **COMPUTE-RECIPROCAL-VECTORS** to convert the peak coordinates from the peak list into true reciprocal lattice vectors. Really, though, you only just need a pocket calculator.

(8) FILTERING THE PEAKS IN THE FOURIER TRANSFORM

Using the command **FOURIER-PEAK-FILTER** we can focus on the periodic information present in the image and disregard everything else. If the crystal is large and ideally well ordered, the peaks in the Fourier transforms would be only one pixel wide and would have a perfectly defined phase. However, in practice things look different; peaks may be smeared out over an area and we thus are obliged to use holes containing more than a single pixel in Fourier space.

The hole size parameter is a very important one. It determines the effective area over which one averages in real-space. The smaller the hole, the more unit cells are averaged into each point of the resulting filtered image. The (linear) size of the averaged area in real-space may be roughly estimated by the formula:

$$\text{LINEAR AVERAGING SIZE} = \frac{\text{LINEAR IMAGE SIZE}}{2 * \text{HOLE RADIUS}}$$

NOTE: This **FOURIER-PEAK-FILTER** command does only the masking. You thus must provide it with an input Fourier transform (command **FFT2D**, option **FORWARD**), and you will have to reverse Fourier transform (command **FFT2D**, option **REVERSE**) the masked transform to obtain an image you can look at.

2-D crystals in IMAGIC-5

```

IMAGIC-COMMAND: FOURIER-P-FILT

** FOUFILT welcomes you **

Input file, image loc#s ?           : crys_fft
                                     you select,
                                     Fourier transform
                                     of your crystal

Output file, loc#s?                 : crys_opt_filt
                                     you select

Mode to get the mask characteristics.
Input from USER or FILE            : USER

Reciprocal lattice vector 1 : X1 , Y1 : 12.3,11.6 you
Reciprocal lattice vector 2 : X2 , Y2 : 8.2,6.3  have
                                     to know

Give radius of HOLES in mask (pixels) [2]: 4      you select

Low pass filter radius [0.667]       : 0.8      you select

Leave out the (0,0) peak [YES]        : YES      you select,
                                     usually: YES

```

During the masking of the Fourier transform, the program will determine in detail how coherent the phases in each hole of the masked image are. A quality factor Q is calculated which is defined to be:

$$Q = \frac{\text{VECTOR SUM OF PEAK AMPLITUDES}}{\text{SUM OF MODULI OF PEAK AMPLITUDES}}$$

A significance threshold for Q is also defined and printed. Note that, unless the crystal is ideal and centred, it may be very difficult to interpret the peak phases directly in terms of the plane group of the crystal.

NOTE: This kind of processing (**FOURIER-PEAK-FILTER**) is not very subtle since it is based on the implicit assumption that the input crystal is a perfect crystal with a repeating motif placed on an ideal undistorted lattice. **CORRELATION-AVERAGING** is much more subtle and can replace/ supersede this Fourier space manipulation procedure almost always see chapter 10).

(9) THE "QUASI-OPTICALLY" FILTERED IMAGE

The filtered image has some interesting interpretations. In each point this image contains the local average of an environment of the input (unfiltered) image. Thus, if the holes are not chosen too small, the filtered image will still contain all the long range distortions present in the original and can thus be used as a first judgement of quality.

Also, bad areas of the crystal may be recognised visually in the filtered image.

Yet, quasi-optical filtering does not allow us to correct for the distortions and thus the possibilities are fundamentally limited. The number of unit cells averaged is limited to the characteristic "coherency" area of the crystal.

(10) CORRELATION AVERAGING

Correlation averaging (CA) allows us to correct for long range distortions of the crystal and thus allows us to average a much larger number of unit cells than with the filtering technique.

To apply correlation averaging we need a reference image since we are essentially looking for "matches" of the reference motif ("matched filtering") in the crystal image. The choice of the reference is important and some attention must be devoted to create good reference images.

(10a) LOOK FOR A REFERENCE IMAGE

(i) Extracting a patch from the filtered image

Since the quasi-optically filtered image already contains considerably less noise than the original it is a good idea to extract an area from that image to be used as a reference patch for the correlation image.

The extraction can be done in the following way: **DISPLAY** your quasi-optically filtered image. If the image is displayed use option **COOS**. Move the cursor to the middle of a good area and press the middle button of the mouse. You are asked whether the coordinates should be stored, give **YES** and specify the name of the output ".plt" file (usually the same name as the input file). To stop press the middle button and give **STOP**. Leave **DISPLAY** by typing **BYE** or **EXIT**.

To extract the small reference patch use command **CUT-IMAGE** with option **APERIODIC**. Input file is the quasi-optically filtered image and the corresponding "plt" file created in **IMAGIC**:

2-D crystals in IMAGIC-5

```
IMAGIC-COMMAND: CUT-IM

** SUBFILE welcomes you **

Input file, image loc#s      : crys_opt_filt      you
Output file, image loc#s    : crys_ref           select

Mode of operation:
  APERIODIC,  CENTER,      GENERAL,   PAD_GENERAL,
  LARGE_PAD,  SMALL_PAD,   PERIODIC,  SQUARE,
  CHECKERS
Give mode of operation      : APERIODIC
Output file dimensions X,Y  : 100,100           you select,
                                                                    type ? to
                                                                    get "good" image
                                                                    sizes. The value
                                                                    depends on the
                                                                    size of the unit
                                                                    cell of the crystal

Coordinates (plt) file      : my_micrograph.plt
                                                                    output from
                                                                    DISPLAY
                                                                    (option COOS)
```

(ii) Using a patch from the original image as reference

We can also start with a patch of the original crystal as a reference. For that purpose we simply extract a good sub-area of the full crystal instead of using the quasi-optically filtered image. The same commands shown in (i) can be used.

(10b) CENTRING THE "MOTIF"

Once we have selected a reference patch we should try to centre the motif in the (still small) reference image. Call command **CENTER-IMAGE** and use options **SELF** or **MASS**. Type **?** to get help):

```
IMAGIC-COMMAND: CENTER-IM

** ALIMASS welcomes you **

Input file, loc. numbers      : crys_ref      you
Output file, loc. numbers    : crys_ref_cent select
Options for centering:
MASS MASK ABSOLUTE SELF CENTER
Please specify option         : mass          type ? to
                                                get help

Radius of circle for mass calcn : 0.85      you
Density threshold for mass calcn : -999.    select,
                                                -999. means
                                                no threshold
```

If the first reference is a patch from the original noisy crystal, it may be difficult to centre this noisy reference and one may wait until after the first iteration(s) of the correlation averaging procedure with the centring of the reference image. If the reference image is not centred, that leads to correlation peaks which will be systematically shifted with respect to the unit cell of the crystal. This may become a problem if the reference patch is of the same order of size as the unit cell. In this case the reference image size should be larger than the unit cell.

(10c) PADDING WITH ZEROES

For the cross-correlation function calculation, the image and reference image have to be of exactly the same size. For this reason the small reference image has to be padded with zeroes up to the size of the crystal image.

To achieve this, use command **PAD-IMAGE** with option **CENTER**.

```
IMAGIC-COMMAND: PAD-IM

** SUBFILE welcomes you **

Input file, image loc#s      : crys_ref_cent      you
Output file, image loc#s    : crys_ref_ca        select

Mode of operation:
  APERIODIC,  CENTER,      GENERAL,    PAD_GENERAL,
  LARGE_PAD,  SMALL_PAD,   PERIODIC,   SQUARE,
  CHECKERS
Give mode of operation      : CENTER
Output file dimensions X,Y  : 1024,1024          size of the
                                                                    crystal image
```

(10d) CROSS-CORRELATION

To start the correlation averaging procedure you must start the **CORRELATION-AVERAGING** command. This command will perform all the operations mentioned below up to the level of iterating the whole procedure (chapters (10e) and (11)).

The cross correlation function (CCF) or the mutual correlation function (MCF) is calculated by (conjugate) multiplying the Fourier transforms of the crystal image and the reference image with each other. The result is then Fourier transformed back to real space and is the CCF (or MCF) image containing a peak at each position where the original matches the reference.

The correlation peak heights are determined by how well an area of the crystal matches the reference motif. However, factors such as the local amount of embedding material (stain) in a given area or other low frequency disturbances which are not (necessarily) structure related may lead to misleading correlation peak values. It is therefore a good practice to suppress the low frequency components during the CCF calculation. This can be done by either **BAND-PASS-FILTERING** the crystal input image (see chapter (4) or by multiply the CCF in Fourier space by a filtering function (command **ARITHMETIC** and option **BANDPASS**).

The peaks in the CCF (or MCF) are searched and written into a (sorted) table of the so-many highest peaks that were found in the CCF. The peak searching algorithm performs a least-squares-fit of a paraboloid over a 3x3 environment of the peaks and thus finds peaks coordinates within a fraction of the inter pixel distance. The algorithm is the same as that mentioned earlier to find the peaks within the power spectrum of the crystal image.

(10e) EXTRACTING THE SMALL PATCHES

Using the CCF (or MCF) peak list, **CORRELATION-AVERAGING** then extracts the environments of the correlation maxima from the input crystal image and puts these as a sequence of small images in a standard **IMAGIC** image file. Please note that (at the moment) the sub-images are only extracted at integer co-ordinate positions.

The whole correlation averaging procedure is calculated with **IMAGIC** command **CORRELATION-AVERAGING**:

2-D crystals in IMAGIC-5

```
IMAGIC-COMMAND : co-aver
** COR_AVG welcomes you **
Input crystal (patches) : crys_norm
                        your crystal
FULL analysis or NEW peaks only : FULL
Input REFERENCE patch : crys_ref_ca
                        your reference
Mode of operation
  CCF or MCF : CCF
                        at the beginning
                        we prefer CCF
Give name for CCF file [crysccfs] : crys_ccf
                        output CCF file,
                        you select
# of PEAKS exp. in data set : 500
                        be conservative at the
                        beginning. Check the
                        extracted patches later
                        to correct this number
                        in further iteration(s)
Distance from the edge of file [32] : 50 you
Minimal distance between peaks [10.] : 100 select,
                        depends on
                        the size of the
                        reference motif
Name for extracted patches : crys_sml you select
Size of these subfiles [64,64] : 200,200
                        you select
                        depends on the size
                        of the reference motif
A scratch file is necessary : junk_ref_ft you
2nd scratch file needed : junk_xxx select
```

(10f) SUMMING THE SMALL PATCHES

Then all the images are summed into a total sum which can be inspected and may be used to re-iterate the process. Use command **INCORE-SUM** with option TOTSUM:

```
IMAGIC-COMMAND: INC-SUM

** SUMMER welcomes you **

Mode of summing:
  CONDITIONAL_SUM, SOME_SUM, TOTAL_SUM
Please specify option           : TOTSUM
Input file, NO loc#s          : crys_sml  you select
Output file, image loc#s      : crys_sum  output from
                               :          COR-AVER
Standard deviation images     : none
```

Note that since the peak-list is sorted by correlation values, the first of the extracted patches may be better than the last ones. It may thus be wise to restrict the analysis to the first so-and-so-many extracted images.

If desired, you can use the **INCORE-SUM** command with the **SOMESUM** option to sum only the first 100 or so of your extracted images, and not all of them.

(11) ITERATING THE CORRELATION AVERAGING PROCEDURE

Each time correlation averaging procedure is repeated, a better average is obtained with less noise. Since it has less noise, we may make the reference patch smaller and that, in turn, leads to a better correction of the distortions of the crystal.

2-D crystals in IMAGIC-5

Do not forget, however, to centre the motif since with a small, uncentred reference patch you may miss the molecules in the CCF.

In a later stage of the analysis the finer details will become more and more important. MCF can then be used instead of CCF.

(12) MSA CLASSIFICATION OF THE EXTRACTED PATCHES

The analysis up to this point has considered that the crystal might have been distorted but we have always assumed that the unit cells are equal. This may NOT be true. Differences in preparation, rotational shifts of the molecules within the unit cell, small molecular differences etc. cannot be detected with the procedure which we have used until now.

To find differences between the patches we first use multivariate statistical analysis (MSA) (representation of the extracted patches in a "hyper space" and data reduction) and automatic classification.

(12a) MSA MASK

First a MSA mask is needed which is a binary mask to define the area of interest within the image frame that should be considered during multivariate statistical analysis and classification.

For this, **DISPLAY** the resulting average image of the previous correlation averaging procedure (use a large **SCALE** factor). Then use options **COOS** and **YES OUTPUT TO PLT FILE** (store), and mark the boundary (not very close) around the interesting area.

2-D crystals in IMAGIC-5

Having created this contour "plt" file, call **MSA-MASK**:

```
IMAGIC-COMMAND: MSA-MASK

** MSAMASK welcomes you **

Input file, image loc#s                : crys_sum
                                         you select,
                                         average of the
                                         extracted patches

Input contour coos (plt) file [cry_sum] : totsum_X.plt
                                         contour "plt"
                                         file created
                                         with DISPLAY
                                         and option COOS

Give output file, image loc#s          : msamask
                                         you select
```

(12b) MULTIVARIATE STATISTICAL ANALYSIS (MSA)

Then perform multivariate statistical analysis. Input images are the patches extracted from the previous correlation procedure. The results are mainly stored in the header file of the input images, in a registration file (**msa.lis**, for example) and in the (output) eigenimages (**eigenim**, for example).

2-D crystals in IMAGIC-5

Call **MSA-RUN**:

```
IMAGIC-COMMAND: MSA-RUN

** MSA welcomes you **

Choose mode of operation:
  FRESH_MSA  REFINE
Please specify option           : fresh

MSA distances:
  EUCLIDIAN  CHISQUARE  MODULATION
Please specify option           : modulation
use modulation distance

Input aligned "images"           : crys_sml
you select, extracted patches of correlation averaging

Are there any inactive images?   : no
if wanted some patches can be excluded

Input MSA mask file [none]       : msamask
the MSA mask created, you can give "none" if no mask is wanted

Eigenimages output file         : eigenim   output
Pixel coordinates output file   : pixcoos   you
Eigenpixel vectors output file  : eigenpix  select
Give number of iterations (<65) [24] : 24       you select
Give number of eigenimages (<70) [24] : 69       you select
Overcorrection factor (0< ocf< 0.9) : 0.8      you select
Rootname for results files, no ext. : msa      you select
no extension please !
```

(12c) CLASSIFICATION

The next step is the classification within the hyper-space created by [MSA-RUN](#). Call [MSA-CLASSIFY](#):

```
IMAGIC-COMMAND: MSA-CLASSIFY

** CLASSIFY welcomes you **

Input to be classified:
  IMAGES  PIXEL-VECTORS  SEQUENCES  FABOSA-CONTOURS
Please specify option          : images
Input images file (treated by MSA)      : crys_sml

                                     extracted crystal patches
                                     treated by MSA-RUN

Percentage of images to be ignored [0]  : 0
Active factors for classification? [68] : 68

                                     the number must be less/equal the
                                     one used in MSA-RUN

Tricky normalisation of factorial coos  : no  always NO !
Weight the factorial coordinates [NO]    : no  always NO !
Plot the classification trees [YES]      : no  usually NO !
Other than default output information    : no
Classification (MSA) maps stored [NO]    : no
What NUMBER of classes do you wish      : 500 you select!
                                           large number

Name of output results files            : classes
                                           you select!

Mass threshold parameter [0]            : 0
```

[MSA-CLASSIFY](#) asks for the number of classes into which the data set should be divided. Try to start with a huge number, use [MSA-SUM](#) (12d) and look how the class-sum images look like. If too many classes show the same view you can reduce the number of classes (call [MSA-CLASSIFY](#) (12c) and [MSA-SUM](#) (12d)). [MSA-RUN](#) (12b) is not necessary).

(12d) CLASS-SUM IMAGES

Then calculate the class-sum images. Call **MSA-SUM-IMAGES**:

```
IMAGIC-COMMAND: MSA-SUM

** CLASSUM welcomes you **

Input images to be summed           : crys_sml
                                     extracted crystal patches
                                     treated by MSA and CLASSIFY

Rootname of input classification files : classes
                                     output from MSA-CLASSIFY

Output class averages [classsums]    : classsums
                                     you select, output patch averages

Downweight small classes? [YES]     : yes
Produce summing statistics           : no
Fraction of worst class members to ignore: 5   you select
```

Now look at the class-sum images, if necessary re-do the classification (steps [12c](#) - [12d](#)) demanding less or more number of classes.

(13) PUBLICATION

Commands which you may need for polishing your results are, for example: **BLOW-UP** to bring up the image to a larger format (by interpolation), **CONTOUR-IMAGE** to create contour lines (but be careful not to interpret too many contour lines and hence less "objectivity"), **AVERAGE-NFOLD-ROTATIONAL** to impose in-plane symmetries, **INFORMATION-IN-SEQUENCE** to calculate various resolution criteria... See manual(s) to get information about all available **IMAGIC** commands or use command **MENU *** .

Then write a nice paper for Nature, Science, Structure, JMB, etc. Or you may feel that the most appropriate thing to do at this point is command **DELETE-IMAGE** and go have a beer (the latter option is not yet implemented in **IMAGIC**).

Content

(1) STARTING THE IMAGIC-5 SYSTEM	3
(2A) SCAN THE MICROGRAPHS AND CREATE IMAGIC IMAGES	3
(2B) CHECK THE IMAGES	4
(3) IMAGE SIZE	5
(4) BAND-PASS FILTER	8
(5) PREPARATION OF THE CRYSTAL IMAGE	10
(5A) IMPOSE-PLT-CONTOUR	10
(5B) PRETREAT-IMAGE	11
(5C) MULTIPLY-BY-CONSTANT	11
(6) THE FOURIER TRANSFORM AND THE POWER SPECTRUM	12
(7) DETERMINING THE RECIPROCAL LATTICE VECTORS	14
(8) FILTERING THE PEAKS IN THE FOURIER TRANSFORM	16
(9) THE "QUASI-OPTICALLY" FILTERED IMAGE	18

2-D crystals in IMAGIC-5

(10) CORRELATION AVERAGING	18
(10A) LOOK FOR A REFERENCE IMAGE	19
(10B) CENTRING THE "MOTIF"	21
(10C) PADDING WITH ZEROES	22
(10D) CROSS-CORRELATION	22
(10E) EXTRACTING THE SMALL PATCHES	23
(10F) SUMMING THE SMALL PATCHES	25
(11) ITERATING THE CORRELATION AVERAGING PROCEDURE	25
(12) MSA CLASSIFICATION OF THE EXTRACTED PATCHES	26
(12A) MSA MASK	26
(12B) MULTIVARIATE STATISTICAL ANALYSIS (MSA)	27
(12C) CLASSIFICATION	29
(12D) CLASS-SUM IMAGES	30
(13) PUBLICATION	31

CONTENT