

## EMICOFV.DOC

EMICOFV is a "brute force" (translates, "sledgehammer") icosahedral view-refining program. It calculates the view parameters (THETA,PHI,OMEGA) for a particle assumed to have icosahedral symmetry. The entire icosahedral asymmetric unit is scanned in 1 degree steps in THETA, PHI, and OMEGA. The asymmetric unit of an icosahedron is a triangular region (one third of an icosahedral face) with its corners at a three-fold axis (THETA,PHI=69.09,0.0) and two adjacent five-fold axes (90.0,-31.717474 and 90.0,31.717474). The THETA,PHI,OMEGA coordinate system is that defined by Klug and Finch (J. Mol. Biol. 31:1-12 [1968]).

\*\*\*\*\*

\* BACKGROUND \*

\*\*\*\*\*

EMICOFV is adapted from the original MRC routine FINDVIEW. It has been extensively modified by S. Fuller (Heidelberg) and T. Baker (Purdue). The original MRC "phase residual" or "vector residual" calculation was modified to compensate for the common lines degeneracies that led to improper relative weighting of the residuals for different orientations. The degeneracy of the common lines is compensated in 2 ways:

1. The common lines subroutine, ICO\_COMLIN, now uses a flag to identify truly degenerate common lines (IFLAG(NVIEW)).
2. For a given pair of common lines, the radius at which they are RADMIN apart (DEFAULT = 1.0 TPU) is calculated, and residuals are determined only for those common points for which their separation is greater than RADMIN.

The program calculates the chi\_square for each orientation and the probability of that chi\_square based on:

1. An estimate of the variance obtained from the matrix of residuals for a fixed THETA,PHI and varying OMEGA.
2. The number of degrees of freedom, which is determined from the number of actual points for which residuals were evaluated and the number of samples on the complete line.

The value of the unweighted residual is then multiplied by the probability value. Thus, low residuals that are highly probable remain virtually unchanged whereas those with the same value that are very improbable are increased.

\*\*\*\*\*

\* PROGRAM INPUT \*

\*\*\*\*\*

1. FILE NAME (A)  
1B. IMG\_NUM1,IMG\_NUM2,IMG\_NUM\_STEP (3I)
2. ICO\_IDIM (I)
3. MINR,MAXR,NSAMPL,FMIN,RADMIN,ILIST (3I,2F,I)
4. ID,DUMT,DUMP,DUMO,FFT\_ORIGX,FFT\_ORIGY (I,5F)

1. FILE NAME (A)

-----

Name of file that contains the IMAGE data. If the file is in BYTE-packed format (see [TSB.DOC]EMPROGS.DOC), then the following additional input is necessary:

1B. IMG\_NUM1,IMG\_NUM2,IMG\_NUM\_STEP (3I)

-----

IMG\_NUM1 and IMG\_NUM2 identify the first and last IMAGES to be analyzed. By setting IMG\_NUM\_STEP > 1 (DEFAULT = 1), IMAGES in the data file will be left out of the analysis. For example, if IMG\_NUM1,IMG\_NUM2,IMG\_NUM\_STEP = 1,20,3, then only IMAGES # 1, 4, 7, 10, 13, 16, and 19 in the BYTE-packed file will be used in EMICOFV.

2. ICO\_IDIM (I)

-----

ICO\_IDIM is the size of the FFT. The only allowed values for ICO\_IDIM are 128, 256, 512 or 1024.

3. MINR,MAXR,NSAMPL,FMIN,RADMIN,ILIST (3I,2F,I)

-----

MINR and MAXR are the inner and outer radii (in INTEGER TPU steps) of the band of data thought to be correlated icosahedrally. These parameters should be carefully chosen as outlined below.

MINR and MAXR specify the lower and upper resolution limits of the data to be used in the refinement calculations. These values are expressed in INTEGER transform pixel units (TPU). The DEFAULTS for these parameters are 1 and ((ICO\_IDIM/2)-2).

As an example, for a particle image of bovine papilloma virus (BPV) recorded at 36,000X, scanned at 25 microns, and Fourier transformed in a 128 by 128 pixel array, the following relationships can be derived:

$$\begin{aligned} \text{Each pixel} &= \text{scan raster in nm} / \text{magnification} \\ &= (25000 \text{ nm}/36000) \\ &= 0.694 \text{ nm} \end{aligned}$$

Each TPU = 1.0 / (pixel size \* ICO\_IDIM)  
= magnification / (scan raster in nm \* ICO\_IDIM)  
= (36000/(25000 nm\*128))  
= 0.0113 nm<sup>-1</sup>  
= 1/88.9 nm

Judicious choices of MINR and MAXR significantly improve the success of the orientation search computed by EMICOFV.

MINR and MAXR are typically estimated by inspection of a few representative particle Fourier transforms (displayed with EMFFTDSP) and the same values are then used for all particle images analyzed by EMICOFV. Typically, the value for MAXR is set slightly lower than the value corresponding to the maximum expected resolution of the data. This helps reduce uncertainties introduced by the presence of high resolution noise. For BPV images, the maximum resolution was determined to be about 36 TPU (= 3.5 pixel resolution = 2.5 nm resolution). For EMICOFV, MAXR was set = 26 (4.9 pixel resolution; 3.4 nm resolution).

NOTE: MAXR must be < (ICO\_IDIM/2)-1, the limit of the Fourier transform data. Since micrographs typically should be scanned at a raster step size (pixel resolution) AT LEAST TWICE AS FINE as the expected resolution, MAXR should never exceed (ICO\_IDIM/2)-1. If, for example, the data are digitized at a raster step size three times finer than the expected resolution, MAXR should not exceed ICO\_IDIM/3.

The choice of MINR also has a significant effect on the success of the orientation search. Setting MINR too low has the effect of overemphasizing very low resolution features, which generally carry little if any useful information concerning the orientations of the icosahedral symmetry axes of the particle. MINR is estimated by noting where the particle transform no longer appears circularly symmetric. In the BPV example, MINR was determined to be 6 TPU (21.3 pixel resolution; 14.8 nm resolution) which is a resolution corresponding to about one fourth the particle diameter. Thus, for BPV, the spherically symmetric part of the transform extends to about 14.8 nm resolution and this part of the transform carries no useful information about the orientation of the icosahedral symmetry axes.

NSAMPL is the number of radial sample steps to be taken within the band. NSAMPL is almost always set equal to the DEFAULT value (MAXR-MINR+1). At this value the sample step size equals one TPU.

FMIN is the fractional minimum amplitude (relative to the mean) of data used in the refinement. Data points with smaller amplitudes are ignored. For FMIN > 0.0 the average value of F in each radial zone is

calculated and these values are fit, via a Wilson plot, to a straight line to establish the expected average value of F which is then used as a cutoff. For  $F_{MIN} < 0.0$  the average value of F for the entire required area is calculated and this cutoff is applied to all the F's. The DEFAULT for  $F_{MIN}$  is 1.0, a value that is useful in most instances. Use  $F_{MIN} > 1.0$  or  $< 1.0$  WITH CAUTION.

RADMIN is the distance in reciprocal space beyond which data points are considered to be independent. This is usually equal to the reciprocal of twice the particle diameter ( $1/2 * \text{diameter}$ ) expressed in  $REAL * 4$  TPU. The DEFAULT for RADMIN is 1.0.

RADMIN can be estimated in the following way:

$$RADMIN = ICO\_IDIM / (2 * ICO\_DIAM)$$

where  $ICO\_IDIM$  is the FFT dimension (see 2.) and  $ICO\_DIAM$  is the diameter of the particle (box) in pixels.

For example, for BPV images, the particle box size was 91 by 91 (NCOL by NROW), and the size of the FFT used was 128 by 128. Thus,  $RADMIN = (128 / (2 * 91)) = 0.7$ .

NOTE: It is always a bit safer to set RADMIN slightly larger than the value determined by the above calculation. Thus, in the BPV example, RADMIN was set = 1.0 instead of 0.7.

ILIST controls the amount of output written to EMICOFV.TMP.

- ILIST = 0 only prints out the probability weighted phase residual.
- = 1 also performs a 5 cycle origin refinement for each of the 10 best particle orientations found (essentially using the EMICOORG algorithm). Use of this option increases total program CPU usage by about 10%.
- = 2 also prints out the original MRC phase residual (with a modified version of the common lines subroutine).

#### 4. ID,DUMT,DUMP,DUMO,FFT\_ORIGX,FFT\_ORIGY (I,5F)

-----

ID,DUMT,DUMP,DUMO are dummy variables, not used by EMICOFV. They are present only for convenience as they are part of the standard EMICO.DAT input data format.  $FFT\_ORIGX, FFT\_ORIGY$  are the pixel coordinates of the particle center (the point 0.0,0.0 corresponding to the lower left corner of the boxed particle image). Initial estimates of  $FFT\_ORIGX$  and  $FFT\_ORIGY$  are usually determined with EMCORORG. There is one line of this input data for each particle. A separate EMICOFV.TMP file is produced for each particle. Only one EMICOGRA.DAT file is produced

for each run of EMICOFV.

```
*****  
* RUNNING THE PROGRAM *  
*****
```

The program is typically executed as a BATCH job and output in the \*.LOG, EMICOFV.TMP, and EMICOGRAD.TMP files should be examined very carefully to determine the most probable THETA,PHI,OMEGA orientation angle.

PLEASE NOTE: Output from this program requires careful analysis, since the correct orientation angle QUITE OFTEN turns out NOT to be the orientation at which the BEST residual value is found due to the low signal-to-noise in most EM data (especially of frozen-hydrated specimens). EMICOFV gives the most trustworthy results ONLY if the particle phase origins (FFT\_ORIGX,FFT\_ORIGY) are fairly accurately known. This is not always easy to do, especially for particle image data that is noisy (the case for most images of frozen-hydrated specimens). In addition, the results of EMICOFV are highly sensitive to the values chosen for MINR, MAXR, and FMIN (see 3.).

```
*****  
* FINAL NOTES *  
*****
```

1. A VACCELLERATOR version of this program, [TSB.AEXE]A\_EMICOFV.EXE, is available and works the same as EMICOFV, but much faster, and thus it is preferred.
2. For those wishing to do limited searches (finer sampling of THETA, PHI, or OMEGA over restricted regions of the asymmetric unit) EMICOFV2 must be used. Note that this version of the program is not documented, requires FFT data input, and other parameters are input differently than in EMICOFV. To recompile and link this routine you will need subroutines (ICOFV\_\*) that may be found in [TSB.NEWFV]SDFLIB.SUBS.
3. See [TSB.FOR]EMICOFV.BCH for an example BATCH command procedure.

```
=====
```

The FORTRAN code for EMICOFV is in [TSB.FOR]EMICOFV.FOR.  
This documentation is in [TSB.DOC]EMICOFV.DOC

2-Oct-91

```
-----
```

```
*****
MAIN      *
(EMICOFV.FOR) *
*****
```

FLOW CHART FOR EMICOFV PROGRAM

```

*
*-- PIRADDEG
*
*-- ICO_ROTMAT
*
*-- IMG_OPEN ----- | -- STRING_UPPER
*                    | -- IMG_PACK_BIMGCOM_CLEAR
*                    | -- FILE_CHECK
*-- IMG_PACK_GETRANGE
*
*-- IMG_PACK_FIND - IMG_PACK_NRECS
*
*-- IMG_FILL
*
*-- IMG_PACK_FILL - IMG_PACK_NRECS
*
*-- IMG_FFT_FILL - FFT_CLEAR
*
*-- FFT_2D - FOURT - L6TOL9
*
*-- FFT_TO_ATBT - PIRADDEG
*
*-- ICO_POLART ----- | -- PIRADDEG
*                    | -- ICO_INTERP
*
*-- ZMIN - LINFIT      | -- ICO_COMLIN_A
*                    | -- ICO_CHECKANG
*-- ICO_BRUTE - ICO_BRUTE_RSYM -- | -- PCHISQ - CHIPROB - NORMAL
*                    | -- ICO_BESTLIST
*                    | -- ZMINC - LINFIT          | -- PIRADDEG
*                    | -- ICOFV_ORG ----- | -- ICO_COMLIN
*                    | -- ICO_BRUTE_RESREAD   | -- ICO_INTERP

```