Acta Cryst. (1993). D**49**, 100–106

Programs for Phasing by Entropy Maximization as Implemented in *Xtal*3.2: a Crystallographic Software System

BY JAMES M. STEWART

University of Maryland, College Park, MD 20742, USA

DOUGLAS M. COLLINS

The Naval Research Laboratory, Washington, DC, USA

KEITH D. WATENPAUGH

The Upjohn Co., Kalamazoo, MI, USA

EDWARD PRINCE

National Institutes for Science and Technology, Gaithersburg, MD, USA

AND SYDNEY R. HALL

University of Western Australia, Nedlands, Australia

(*Received* 30 *June* 1992; *accepted* 18 *August* 1992)

Abstract

*Xtal*3.2, a crystallographic software package, is an international development project involving about 40 researchers over a full spectrum of crystallographic interests. This development has been supported by many national and international agencies and commercial institutions since the first version in 1983. The 1992 release, *Xtal*3.2, contains software for 95 different calculations. These range from the processing of raw diffraction data to interactive molecular graphics, atomic charge estimation, electronic publication preparation, and the structure solution and refinement of small and large molecules. Tests of the *Xtal* programs for phase determination and phase refinement by the application of 'maximum entropy' are presented.

1. Introduction

*Xtal*3.2 (Hall & Stewart 1992) is a collection of programs designed to perform the calculations necessary to solve, refine, interpret and publish structures determined by diffraction techniques. It consists of a 'nucleus' of programs which manage program invocation, input/output of data and results, formation, maintenance and manipulation of data files, and monitoring of perceived errors. The library of crystallographic and 'housekeeping' programs runs under the control of the nucleus programs. Each program is modular and works against defined data files, both binary and ASCII. The structure of these files is defined in such a way that each calculation may extract the information needed and insert the results which are computed into update files for use by any other program in the system. The preparation of the ASCII input files which invoke the programs is documented in book form (Hall & Stewart, 1992). The ASCII data files which result from a structure analysis are in the CIF format described by Hall, Allen & Brown (1991). The CIF format files are suitable for communication to and from other systems, to some journals, and as archive files.

All the programs are written in FORTRAN; however, the symbolic representation is stored in a form which is first treated by a preprocessor, *RFPP* (Hall, 1985), a FORTRAN77 variant of *RATMAC* (Munn, Stewart, Norden & Pagoaga, 1980), which is itself a FORTRAN program. The macro feature of the preprocessor is a convenient way to deal automatically with the differences in machines, FORTRAN compilers and operating system characteristics. In addition to the preprocessor there are other 'software tools' for preparing, applying, and distributing updates and enhancements to the crystallographic programs. Between major releases updates are handled by means of an automatic e-mail reply facility 'sendme' from the University of Western Australia.

Detailed *Xtal* documentation is published as two manuals. These give details of the file structure, system control commands, terminology used and descriptions of each of the crystallographic calculations. Some examples of input are also included. The programs are supplied with some 'decks' of test input and output. The documentation includes information on implementation of the programs and use of the 'tools' programs.

*Xtal*3.2 may be installed on any computer or operating system which has a FORTRAN compiler. The macros supplied allow for the setting of local computer characteristics, *e.g.* available memory. The standard distribution is

0907-4449/93/010100-07$06.00

from the University of Western Australia on a single 0.5″ magnetic tape with implementation and application documentation. A customized version is also available for PCs on disk. The versatility and portability of the *Xtal* package works well in the increasingly common networked computing and workstation environments.

In the following section, the current programs of *Xtal*3.2 are listed in alphabetical order of their calling mnemonic with a brief description of the purpose of each program. Later sections give special emphasis to the programs designed to aid in macromolecular calculations. It should be noted that some of the 'service' routines, such as file creation and reflection sort/merge serve both the small-molecule and macromolecule programs. Therefore a straightforward categorization of the programs is not possible. Finally, there is a section devoted to the experimental programs for the utilization of maximum entropy for phase determination and extension.

2. Crystallographic and service programs in *Xtal*3.2

In what follows 'BDF' designates a 'binary data file'.

ABSCAL	Absorption correction of intensities using φ spin data
ABSORB	Absorption correction using Gaussian/analytical/spherical algorithms
ADDATM	Load, update and/or edit atomic parameters
ADDMUL	Add Xengen area detector intensity data to a BDF
ADDPAT	Add intensity data for a powder pattern to a BDF
ADDREF	Add reflection-intensity or *F* data to a BDF
APMASK	Apply a density solvent-flattening mask prepared by *MKMASK*
ATABLE	Generate a table of atomic parameters in publication format
BAYEST	Develop Bayesian statistics of intensity data
BFOURR	Set up Fourier coefficients for protein data for *FOURR*
BONDAT	Generate idealized bonded atoms from geometric considerations
BONDLA	Generate contact and bond lengths and angles from atomic parameters
CEDAR	Refine atomic parameters based on diffraction, energy and dynamics
CHARGE	Calculate atomic charges
CIFIO	Generate and read CIF archive ASCII files from BDF
CONTRS	Prepare contour plots from Fourier maps
CONVOL	Direct-space convolution of two functions by a reciprocal-space multiplication
CREDUC	Determination of cell reduction and twin laws
CRITIQ	Cull multiple intensity measurements of outliers
CRYLSQ	Crystallographic least-squares refinement of atomic parameters
DIFDAT	Read and translate diffractometer tapes to extract intensity data
EDTBDF	Edit the contents of a BDF
FC	Calculate structure factors by summation over atoms
FINDKB	Find linear and exponential scaling between parent and derivative protein
FODIFF	Prepare a Fourier map from the difference between two Fourier maps
FOGEN	Generate any volume of a Fourier map from an asymmetric portion
FOGNU	Generate any volume of a Fourier map from a full cell map
FOMERG	Generate a Fourier map by merging Fourier maps of partial structures
FOSTAT	Prepare statistics on Fourier-map densities
FOURR	Compute a reciprocal cell-to-direct cell Fourier transform; Beevers–Lipson and fast Fourier transform algorithms
GENEV	Evaluation of normalized structure factors, *E*; also scale and temperature parameters
GENMAP	Generate an electron-density map from atom parameters
GENSIN	Generate structure invariants for direct methods
GENTAN	Tangent-formula generation of phases
LATCON	Unit-cell parameters from 2θ values
LISTFC	List reflection structure factors for publication
LSABS	Lausanne-Gaussian or analytical absorption with crystal dimension estimation
LSLS	Lausanne least-squares refinement of atomic parameters from raw intensities
LSQPL	Least-squares planes and lines with respect to atomic coordinates
LSRES	Editor for setting restraints in Lausanne least-squares program
MAKBRK	Make a file from *FOURR* output suitable for use by *FRODO* (Jones, 1985; external program)
MAPLST	Dump a Fourier map as ASCII numbers for use by external programs
MEDENS	Form a constrained exponential electron-density distribution from an electron-density map
MEFFIT	Modify a *MEDENS* map in a maximum-entropy step to phase and fit observed structure moduli
MEPHAS	Explore phase assignments by maximum-entropy fitting of structure moduli by the method of Prince (1989)
MERGDS	Merge data sets; *e.g.* parent and isomorphous in preparation for multiple isomorphous replacement run
MERGOB	Merge equivalent reflection intensities from two multiple observation BDFs
MERUN	Set up control lines for *MEDENS*/*MEFFIT* maximum-entropy refinement run
MESTAR	Clear and initialize reflection phases in a BDF; *e.g.* for *MEPHAS* run
MIND	Output atom sites for *MindTool* (external program)
MIR	Determination of reflection phases for a protein from multiple isomorphous replacement data
MKMASK	Make solvent-density flattening mask for *APMASK*
MODEL	Search electron-density peak sites for connected sets
MODHKL	Modify reflection data on BDF
MOGIN	Generate input file from BDF for use by *MOGLI* (Evans and Sutherland Computer Corporation, 1985)
MULIST	Listings of multiply observed reflection intensity data
NEWCEL	Transform the unit cell in a BDF to a new or different unit cell
NICNAK	Raw intensity data-tape processor for Nicolet–Siemens diffractometer
ORTEP	Oak Ridge thermal ellipsoid plotting program
PARTN	Hirshfield partitioning of a pseudo-atom-fragment electron-density map
PATSEE	Rotational and translational search for atomic coordinates from a Patterson function
PEAKIN	Placing of sets of idealized sites for use in *MODEL*
PEKPIK	Search a Fourier map for coordinates of ranked positive or negative densities (peaks or holes)
PERFAC	Estimate a monochromator perfection factor
PHACMP	Compare the phases of reflections phased by two different methods
PHONYD	Generate idealized controlled error F_{obs} from F_{calc}
PIG	Portable interactive graphics
PLOT	Plotter interface using local hardware; serves *ORTEP*, *CONTRS*, *etc.*
PLOTX	Plotter interface using local software
PRECED	Preliminary data loading for *CEDAR* refinement by energy and dynamics
PROATM	Load protein atomic parameters and descriptors from Brookhaven data bank format
PROLSQ	Konnert–Hendrickson restrained structure-factor least squares

PROTIN	Form restraints for proteins for Konnert-Hendrickson *PROLSQ*
RCALC	Calculate structure-factor *R* factors between BDF items
REFCAL	Process raw reflection intensity diffractometer data
REFM90	Read and write SCFS-90 (Brown, 1988) file format
REGFE	Crystallographic functions and errors from least squares
REGWT	Regina analysis and modification of least-squares weights
REMSET	Remove a complete data set from BDF
REVIEW	Review of structure invariants produced by *GENSIN* after phases are solved
RFOURR	Fourier transform from direct space to reciprocal space
RIGBOD	Formation of rigid groups for use in least-squares refinement, *CRYLSQ*
RMAP	Produce *R* factors as a function of coordinates of a translated fragment
RSCAN	Produce reflection *R* factors as a function of $\sin\theta/\lambda$, *F, hkl*
*SCALE*1	Scale non-intersecting data sets
SHELIN	Read and execute *SHELX* input line
SIMPEL	Symbolic addition phasing procedure
SIMWGT	Calculate Sim weights of F_{obs} for combination with weighted phases from other sources (*e.g.* Henderson-Lattman coefficients)
SKLOUT	Generate input line file for *SCHAKAL* (Keller, 1988; external program)
SLANT	Produce general section Fourier maps
SORTRF	Sort reflections in BDF on *h, k, l* in any chosen order
STARTX	*Ab initio* binary data file builder
VUBDF	View the contents of the BDF or any selected records thereof
VUFILE	View contents of any ASCII line-output file
XTINCT	Isotropic Zachariasen extinction coefficient from correlation of intensity differences

3. *Xtal*3.2 programs for macromolecular structures

*Xtal*3.2 requires a user to invoke crystallographic routines based on an order which is dictated by the problem at hand. The narrative which follows should give the reader a sense of the programs available in the order in which they would often be used. The programs *STARTX, ADDMUL, ADDREF, SORTRF, MERGDS* and *REMSET* are used to form and fill data files with the unit-cell information, the intensity data for the parent and isomorphous derivatives, including sorting, merging and scaling of multiple data sets from heavy-atom derivatives. Once known, atom parameters may be loaded by means of *PROATM* or, for the heavy-atom parameters prior to a multiple isomorphous replacement run, by *ADDATM*. Programs for carrying out initial and extended phasing include *MIR, SIMWGT, PATSEE, MESTAR, MERUN* and *MEPHAS*. The programs *BFOURR, FOURR, RFOURR, CONVOL, MKMASK, APMASK, FOSTAT, FOMERG* and *MAKBRK* allow for the preparation and manipulation of Fourier transforms. *FOURR* is a general reciprocal space to direct space Fourier transform program. *RFOURR* does the reverse transformations. *BFOURR* is used to form the special coefficients such as Bijvoet differences, $|F_+ - F_-|$, useful in determining macromolecular structures. While *MKMASK* and *APMASK* are used for flattening or other simple modifications of the electron-density maps. *MAKBRK* is an example of an interfacing program in *Xtal*. It produces a file which is to be read by the external *FRODO* program. Once the trial structure is established by Fourier fitting methods the atom parameters may be refined using *PRECED* and *CEDAR* or *PROTIN* and *PROLSQ*. These programs are specifically coded to deal with the refinement of macromolecular structures. *RCALC* will give *R* factors as a function of various quantities, *e.g.* $\sin\theta/\lambda$. There are also a number of programs which serve for housekeeping or program checking tools. *PHACMP* will allow the comparison of phases determined by two different methods; *EDTBDF, MODHKL, VUBDF* allow editing and examination of binary files. *CIFIO* will produce an ASCII file from a binary data file so that a local line editor may be used to edit data files. *PHONYD* generates ideal *F* data with a specified amount of pseudo-random error. It can be useful for making test runs before trying a solution or refinement method with 'real' data.

4. Maximum-entropy implementation in *Xtal*

There are two distinct algorithmic approaches to the use of maximum entropy in the *Xtal* system. One is due to Collins (Collins, 1982; Collins & Mahar, 1983) and the other to Prince (Prince, Sjölin & Alenljung, 1988; Prince, 1989). Collins' scheme is embodied in the programs *MERUN, MEDENS* and *MEFFIT*. While *MEPHAS* is an *Xtal* adaptation of the programs of Prince. In both procedures it is necessary to have a '*prior*' electron-density map. This map may be generated using, for example, multiple isomorphous replacement phases or by inserting only phases for the origin- and enantiomorph-defining reflections by means of *MESTAR*. The purpose of *MESTAR* is to allow the insertion of structure-factor phases for any chosen reflections and the resetting of the remainder to 'unknown'. This *prior* electron-density map is then used to compute a constrained exponential electron-density distribution by means of *MEDENS*. *MEDENS* is modeled on subroutine *MAXENT* by Prince. The map prepared by use of *MEDENS* is the *prior* used in further calculations. The output function from *MEDENS* is a function of maximum entropy and will be used for the extrapolation, interpolation, determination and smoothing of reflection phases. All negative areas will have been scaled above zero and all large positive regions will have been sharpened. In *MEDENS* the input electron density is scanned to establish the maximum, minimum and average electron density. The process is very sensitive to average electron density, and will not work if the F_{000} term is omitted from the original electron-density calculation, or is far from the true value. The process becomes ill conditioned and will fail as the mean of the electron density approaches zero from the positive. This may present a challenge in the case of macromolecules where the contribution of the solvent portion is initially not well known. It is also important that the grid of the input electron density be 'fine' enough to provide the resolution that will allow the phase extension

in the subsequent *RFOURR* runs. The output electron density with unit sharpening is an exponential representation which satisfies two constraints:

(1) The mean of the electron density remains constant.

(2) The mean square of the electron density remains constant.

This is accomplished by calculating the new electron density, ρ_{new}, from the old electron density, ρ_{old}, by

$$\rho_{new} = \exp[ZB + (1 - Z)A],$$

where

$$Z = [\rho_{old} - \min(\rho_{old})]/[\max(\rho_{old}) - \min(\rho_{old})]$$

The scale factors A and B are obtained by a Newton-Raphson iteration.

5. *Xtal*3.2 program *MEFFIT* for phase and moduli refinement

MEFFIT combines a positive-definite map of constrained exponential electron density formed by *MEDENS* and a difference map prepared by *FOURR* using phases calculated by *RFOURR* from the *MEDENS* output map. The effect of this process is to produce a new map such that the maximum-entropy phases and the observed structure moduli will be produced upon Fourier transformation of the new map. The program may be run by use of the *Xtal* program *MERUN* which sets up an input stream to drive all the programs needed in the refinement process.

The focus of program *MEFFIT* is adjustment of a positive-definite density map toward agreement with observed structure-factor magnitudes. The designed application is sequential improvement of an imperfectly phased set of structure factors through manipulation of its corresponding noisy or low-resolution density function by *MEDENS*, *MEFFIT* and necessary Fourier routines. On a grid suitably fine for the desired final resolution, an initial electron-density function is converted by *MEDENS* into a maximum-entropy density, which is a positive-definite exponential function. This map is used as input to *RFOURR* to calculate structure factors (including phases), which, it is expected, are subsequently used in *FOURR* to form a companion difference electron-density map. *MEFFIT* generates a new positive-definite exponential density in a maximum-entropy adjustment of the *prior* exponential density. The new density is computed by pointwise multiplication of the *prior* by

$$\rho_{new} = \rho_{old}\exp(\text{constant} \times \Delta\rho)$$

The ρ_{new} is then scaled to restore the original mean value. As a stand alone process, this combination of a positive-definite density and the difference density, $\Delta\rho$, constitutes one maximum-entropy step in the adjustment of a positive-definite density to match more closely the experimentally observed moduli. Watenpaugh has tested phase extension using this algorithm by setting up a test using the refined structure of phospholipase A_2 (K49 variant) (space group $P4_32_12_1$, $a = b = 71.53$, $c = 57.59$ Å, $R = 14.7\%$) (Holland *et al.*, 1990). Table 1 shows the average phase error for acentric reflections and the count of the number of agreements for centric reflections when the phases of the final structure factors are compared to the phases recovered from the maximum-entropy process. The phase extension power of the method seems to become useful when extending phases from the 3.0–2.5 Å region to 2.0 Å or possibly higher resolution. Starting at lower resolution produces quite flat electron-density maps that can be easily satisfied with small changes in the starting phases resulting in little phase extension. These low-resolution maps have very little negative density and are, therefore, not greatly modified by the maximum-entropy criterion. At higher resolution preventing negative density is a very powerful element in the phase extension process. After the desired number of iterations have been run, the last step of which is *RFOURR* to compute structure factors, the final F_{obs} Fourier map may be calculated. This gives the electron density in its standard formulation but with structure-factor phasing corresponding to the final exponential density. This whole process has the structure: *FOURR*; *MEDENS*; *RFOURR*; (*FOURR*; *MEFFIT*; *RFOURR*); *FOURR*, in which the parentheses show the inner loop of program usage. By this process an electron-density map for which the conventional R value is zero may be produced, *i.e.* every $|F_{obs}|$ matches every $|F_{calc}|$.

6. *Xtal*3.2 program *MEPHAS* for phase determination

MEPHAS is the name given to the *Xtal* program version of the method of Prince. In this process a constrained exponential input electron-density map, the *prior*, based on a smaller number of starting reflections, is used as a basis to establish, by maximum-entropy criteria, the phases of the chosen reflections. This process is used to extend the number of phased reflections. At each stage in the process of establishing the maximum-entropy phases for the chosen reflections the constrained exponential electron-density function is held everywhere positive, and the magnitudes of the F_{calc} implied by the electron density is refined by a dual function procedure which forces a match in the magnitudes of the F_{obs} and F_{calc} of the chosen subset of reflections. The *prior* map may have been prepared by the use of: (i) as few reflections as the number of phases allowed to define the origin and enantiomorph of the space group being treated; (ii) the reflections in (i) plus others determined from other information; (iii) the reflections in (i) plus those determined by a previous run of *MEPHAS* itself. In the description that follows two terms are used: trial and iteration. A trial refers to the use of one set of phases applied to a subset of reflections to be phased. An iteration refers to an iteration of the application of Prince's dual function to bring the moduli to conformance.

Table 1. *Results of maximum-entropy phase extension by MEFFIT*

[Start/End] shows starting map resolution (Å)/extended map resolution (Å). ⟨Acentric⟩ shows average phase error (°)/total number of acentric reflections phased. ⟨Centric⟩ shows number concordant/total number of centric reflections phased. Test case: phospholipase A_2 (K49 variant) (space group $P4_32_12_1$, $a = b = 71.53$, $c = 57.59$ Å, $R = 14.7\%$).

	Shells of d spacing (Å) which show statistics of phase extension							
	4.0	3.2	2.8	2.6	2.4	2.2	2.1	2.0
[Start/End]								
[2.0/2.0]								
⟨Acentric⟩	7.3/955	5.2/1046	5.8/1063	6.1/1074	6.4/1085	6.1/1097	6.8/1102	7.3/916
⟨Centric⟩	412/421	261/262	226/228	196/197	184/184	170/170	160/160	116/117
[2.5/2.0]								
⟨Acentric⟩	7.0/957	6.0/1048	6.9/1063	7.5/1074	41.0/1053	50.7/1060	53.7/1077	55.1/892
⟨Centric⟩	429/434	265/265	228/229	196/197	117/163	118/150	107/144	73/111
[3.0/2.4]								
⟨Acentric⟩	5.6/957	5.4/1048	32.8/994	61.5/973	59.6/259			
⟨Centric⟩	430/435	264/265	167/197	122/165	26/38			
[3.0/2.0]								
⟨Acentric⟩	7.4/957	7.1/1048	36.8/1030	67.0/1037	68.1/1051	66.7/1063	67.7/1070	65.6/896
⟨Centric⟩	429/435	264/265	176/212	125/178	91/168	96/160	91/147	63/105
[4.0/2.8]								
⟨Acentric⟩	5.4/957	62.8/829	74.2/380					
⟨Centric⟩	431/435	111/193	39/76					
[6.0*/3.2]								
⟨Acentric⟩	38.5/431	89.3/54						
⟨Centric⟩	225/267	13/33						

* In the 6 Å test 2.4/261 acentric and 186/186 centric phased reflections were used to begin the test.

The *MEPHAS* program is designed to run in four distinct modes:

(1) This is called the '*ME*sign' mode. Sixteen centric reflections are tested. From 49 to 289 separate trials are carried out until the phases corresponding to the maximum entropy of the constrained exponential electron density have been found. In each trial dual-function iterations are performed until a good fit of the 16 reflection moduli is achieved. The first 32 trials consist of assigning 16 patterns of centric phases and their supplements to the chosen subset of reflections. The difference in entropy associated with the maps from each pair of trials is used in Yates' algorithm (Yates, 1937) to predict the maximum-entropy phase set. A check is made to see if any of the 32 trials show a greater entropy than the Yates result. The phases from one or the other source are then tested by taking the supplement of each phase reflection by reflection. If the entropy rises, the phase is accepted and the next reflection tested until all have been 'checked'. This checking will be confined to checking each of the 16 reflections once, or optionally, repeatedly until no changes are detected. That is a possible maximum of 256 trials. The final phases and their F_{calc} are output without changing any other reflections previously treated. In addition the final constrained exponential density, as modified, is saved to become the next *prior*.

(2) This is called the '*ME*osc' mode. A specified number of centric reflections are oscillated by 180° in phase. At each trial, when the *F*-fitting iterations are completed, if the entropy has risen the current value of the reflection phase is saved and left unchanged as the trials continue. This is tantamount to running the last part of the *ME*sign procedure.

(3) This is called the '*ME*cyc' mode. A small number of general reflections, usually one, may be specified along with a phase 'step' to be applied. These reflections will then be tested against the maximum-entropy criterion at each step from 0 to 360° less the step. Once the maximum-entropy phase is established for each reflection in turn that value is used while the next reflection is tested.

(4) This is called the '*ME*fit' mode. A number of general reflections may be specified for moduli fitting. In this mode the initial phases will be those obtained from the input BDF. Usually a previous run of *RFOURR* using the last previous constrained exponential electron-density map will be required. In this case the map will be refined in one trial of as many iterations as are required to bring the moduli into conformance. This is the many-reflection moduli-fitting mode of operation. Note that, unlike the Collins approach except for special cases, input phases are held fixed in the logarithm of density while the constrained exponential map is modified to fit exactly the input $|F_{obs}|$. The phases derived from the modified map will then show a departure from the phases of the *prior* electron-density map.

7. Preliminary tests of *Xtal* program *MEPHAS*

For the purpose of testing the behavior of the phasing algorithm a test case using idealized data for bovine pancreatic phospholipase A_2 (space group $P2_12_12_1$, $a = 47.07$, $b = 64.45$, $c = 57.59$ Å, $R = 14.7\%$) (Dijkstra, Kalk, Hol & Drenth, 1981) was prepared. In these tests no examination of the maps in the manner described by Sjölin, Prince, Svensson & Gilliland (1991) was carried out. Four of the largest centric reflections suitable for defining the

origin and enantiomorph were assigned the F_{calc} phases and the sixteen largest *F*-centric reflections were treated by *MEPHAS*. In each of the 49 trials the entropy of the map is very similar, ranging from −0.102472 to −0.091110. The trial with the refined structure F_{calc} phases forced into it showed an entropy of −0.102346. It is next to the lowest. The 16 reflections in the maximum-entropy trial yielded seven which were in agreement with the F_{calc} phases.

In a second *a priori* test the four starting reflections used were those determined by the direct-methods program *GENTAN* while the 16 reflections to be determined were again those of highest 'unphased' *F* values. In this test the entropy over the 49 trials ranged from −0.084946 to −0.075564. The trial with the refined-structure F_{calc} phases introduced showed an entropy of −0.084726, again one above the minimum-entropy trial. The 16 reflections from the maximum-entropy trial showed eight in agreement with the F_{calc} phases.

For contrast with the first two tests, a *prior* map with all 13295 reflections with their F_{calc} phases was formed and converted to an exponential map. Then the same high *F* reflections which were used in the first test were treated by *MEPHAS*. In this test the entropy over the 49 trials ranged from −0.582014 to −0.452263. The trial with the refined-structure F_{calc} phases introduced showed an entropy of −0.582014. This was the minimum-entropy trial. The next highest entropy was −0.565033. Of the 16 reflections from the maximum-entropy trial only two were in agreement with the F_{calc} phases. Table 2 presents a summary of this and the first test showing the number of phases of the sixteen tested which agree with the F_{calc} phases, as a function of decreasing entropy.

These results contrast with those of Sjölin *et al.* (1991) on recombinant bovine chymosin. In that study maximum entropy was an effective figure of merit, but in these studies it was not. Ultimately the important question is: 'Can an automated procedure find a set of phases that lead to an *interpretable* map?' The answer to this question will require continuing modification of the programs and further simulation studies.

8. Summary and concluding remarks

In trials of these programs it has been found that both the Collins method and the Prince method cause the formation of an electron-density map which is everywhere positive and which includes the observed moduli for all the reflections in the *prior*. In fact the application of maximum entropy to the modification of the density such that the reflection moduli are expressed is outstanding. However, it has become clear that the entropy alone is not always a sufficient criterion for selecting the correct phases for a structure. When the number of reflections is low, the algorithm can powerfully fit the moduli without allowing the density to be non-positive regardless of the phases chosen. The three tests for *MEPHAS* shown here suggest that the entropy of the map with the correct phases is not necessarily the maximum-entropy map. This observation indicates that an additional criterion will be required if this method is to succeed as an *a priori* phasing scheme.

Table 2. *Examples of maximum-entropy phase trials by MEPHAS*

The left trial is for a prior map with 20 contributing reflections. The right trial is for a prior map with 13295 contributing reflections. In both trials the 16 centric reflections which were tested are those of highest *F* magnitude. Test case: bovine pancreatic phospholipase A_2 (space group $P2_12_12_1$, a = 47.07, b = 64.45, c = 57.59 Å, R = 14.7%).

Trials which show entropy and number of 'correct' centric phases			
Entropy	Phases in agreement	Entropy	Phases in agreement
−0.091110	7	−0.452263	2
−0.091111	6	−0.452312	3
−0.091112	7	−0.453501	1
−0.091116	5	−0.455762	2
−0.091117	5	−0.455887	2
−0.091160	7	−0.456051	2
−0.091357	7	−0.457769	3
−0.091360	8	−0.458128	0
−0.091707	7	−0.458134	0
−0.091740	7	−0.458254	2
−0.092402	5	−0.458684	3
−0.092584	8	−0.460324	2
−0.092611	7	−0.462914	2
−0.092648	7	−0.464381	2
−0.092856	5	−0.465170	2
−0.093088	5	−0.467918	2
−0.093263	8	−0.469093	2
−0.093276	7	−0.475868	8
−0.094018	8	−0.487321	2
−0.094761	0	−0.491157	8
−0.094959	8	−0.495285	8
−0.095355	8	−0.495840	8
−0.095617	8	−0.496521	8
−0.096082	7	−0.500127	8
−0.096534	8	−0.500587	8
−0.096688	8	−0.506093	8
−0.096774	8	−0.512417	8
−0.097042	5	−0.512438	8
−0.097179	8	−0.515169	8
−0.097231	8	−0.516545	8
−0.097410	8	−0.519113	8
−0.097498	8	−0.519557	8
−0.097729	8	−0.520387	8
−0.097764	8	−0.521138	8
−0.097772	8	−0.521164	8
−0.098088	8	−0.522756	8
−0.098189	8	−0.522836	8
−0.098878	8	−0.523666	8
−0.098928	8	−0.525135	8
−0.099077	8	−0.526614	8
−0.099077	8	−0.527371	8
−0.099106	8	−0.530395	8
−0.099112	8	−0.534232	8
−0.099616	8	−0.535994	8
−0.099685	8	−0.536338	8
−0.100367	8	−0.537877	8
−0.102198	8	−0.540039	8
−0.102346	16	−0.565033	8
−0.102472	8	−0.582014	16

Other methods for making use of information in addition to the magnitudes of the moduli have been proposed. Bricogne (1988) has put forward a log-likelihood figure of merit for ranking phase sets. Lunin (1988) has shown that the use of electron-density histograms can remove the uncertainty in the early stages of phasing. Zhang & Main (1990) have described a program *SQUASH* which, starting from a low-resolution map, uses histograms, solvent flattening, Sayre's equation and density modification based

on a least-squares procedure to improve macromolecular models. Because of the open flexible structure of the *Xtal* system, these approaches and others like them will be incorporated into this package. This will be done either to enhance the effectiveness of present programs or to add new methods to the library as it exists.

The contribution of JMS to this work has been supported under contracts N00014-88-K-0323 and N00014-92-J-1556 of the Office of Naval Research.

References

BRICOGNE, G. (1988). *Acta Cryst.* A**44**, 517-545.
BROWN, I. D. (1988). *Acta Cryst.* A**44**, 232.
COLLINS, D. M. (1982). *Nature (London)*, **298**, 49-51.
COLLINS, D. M. & MAHAR, M. C. (1983). *Acta Cryst.* A**39**, 252-256.
DIJKSTRA, B. W., KALK, K. H., HOL, W. G. & DRENTH, J. (1981). *J. Mol. Biol.* **147**, 97-123.
Evans and Sutherland Computer Corporation (1985). *MOGLI*. Evans and Sutherland Computer Corporation, Salt Lake City, USA.
HALL, S. R. (1985). *RFPP. Ratmac to Fortran77 Preprocessor Manual.* Univ. of Western Australia, Australia.
HALL, S. R., ALLEN, F. R. & BROWN, I. D. (1991). *Acta Cryst.* A**47**, 655-685.
HALL, S. R. & STEWART, J. M. (1992). Editors. *Xtal3.2 Reference Manual.* Univs. of Western Australia, Australia, and Maryland, USA. In preparation.
HOLLAND, D. R., CLANCY, L. C., MUCHMORE, S. W., RYDEL, T. J., EINSPAHR, H. M., FINZEL, B. C., HEINRICKSON, R. L. & WATENPAUGH, K. D. (1990). *J. Biol. Chem.* **265**, 17649-17656.
JONES, T. A. (1985). *Methods Enzymol.* **115**, 144-150.
KELLER, E. (1988). *SCHAKAL. A FORTRAN Program for the Graphical Representation of Molecular and Crystallographic Models.* Kristallographisches Institut der Univ. Frieburg, Germany.
LUNIN, V. Y. (1988). *Acta Cryst.* A**44**, 144-150.
MUNN, R. J., STEWART, J. M., NORDEN, A. P. & PAGOAGA, M. K. (1980). *Ratmac Primer.* Tech. Rep. LBL-11847. NRCC, Canada, and Univ. of Maryland, USA.
PRINCE, E. (1989). *Acta Cryst.* A**45**, 200-203.
PRINCE, E., SJÖLIN, L. & ALENLJUNG, R. (1988). *Acta Cryst.* A**44**, 216-222.
SJÖLIN, L., PRINCE, E., SVENSSON, L. A. & GILLILAND, G. L. (1991). *Acta Cryst.* A**47**, 216-223
YATES, F. (1937). *The Design and Analysis of Factorial Experiments.* London: Imperial Bureau of Soil Science.
ZHANG, K. Y. J. & MAIN, P. (1990). *Acta Cryst.* A**46**, 377-381