‘dm’: An Automated Procedure for Phase Improvement by Density Modification.

Dr. K. D. Cowtan, University of York, Heslington, York, YO1 5DD.

Density Modification: The Problem.

The techniques of protein structure solution by X−ray diffraction methods are now well established and are the method of choice for obtaining structural information about many types of biological macromolecule. In many cases however a structure solution by X−ray methods is still far from routine, and in particular the step of obtaining phases for the measured x−ray intensities can present problems. Experimental techniques, such as the measurement of intensities from single or multiple isomorphous derivatives or the differences due to anomalous scattering give estimates to some of the phase information, which may or may not be enough to deduce the structure of the molecule.

Examination of approximate electron density maps phased by such experimental methods shows that they fall into two groups:

1. Those maps in which there are visible structural features, and which it is possible at least to attempt an interpretation. Once a significant portion of the map has been interpreted it is often possible to use this information to improve the rest of the map.

2. Those maps which do not appear to contain structural motifs, from which it is not possible to attempt an interpretation.

The distinction may seen obvious, but in fact we could imagine a third case; when a map is not good enough to show the correct structure, then it could clearly show an incorrect structure. The fact that an experienced worker can distinguish an interpretable map from an uninterpretable map suggests that it may be possible in some manner to quantify the property of map quality. If we the modify the map in such a manner to increase this quantity then it may be possible to convert a uninterpretable map into an interpretable one.

The Method

How are good electron density maps distinguished from poor ones? Even without knowing the structure of a particular protein there are some features which can be expected to appear in the correct map, such as:

1. Flat solvent. The solvent in which the protein is crystallized will fill the gaps caused by imperfect packing of the molecules. This solvent is usually disordered from cell to cell, and so in the diffraction structure the electron density will be constant to a first approximation. This is applied to an estimated map through the process of solvent flattening, or ‘Wanging’[7].

2. Predictable distribution of density values in the protein. Most proteins have fairly similar proportions of atomic types, distributed throughout the cell according to known constraints of atomic spacing. As a result the distribution of density values in the protein region is fairly similar from protein to protein. This information is applied through the process of histogram matching[8].

3. Connectivity of electron density. Proteins consist of linearly connected chains of peptide units. A good map will show this linear connectivity, and it is this feature which provides the starting point for building a model of the structure. The connectivity of the map can be enhanced by a process of iterative skeletonisation, as developed in the ‘PRISM’ package[1,4,5].

4. Known properties of the structure. Some features of the structure may be known from analysis of the x−ray data, or from biochemical sources. These include the presence of non−crystallographic symmetries relating different parts of the map, or relationships with other molecules, from which a portion of the structure can be deduced. This information can be applied through non−crystallographic symmetry[2] averaging and molecular replacement respectively.

5. Atomicity. In the high resolution limit, the electron density map for a structure will show

resolved atomic features. Atomicity can be enhanced by modifying the map to satisfy Sayre's equation[3,8].

Note that all of these features can be distinguished in a map without knowing the desired structure. Techniques have been developed by which a poor density map can be altered in a sensible manner in order to bring it into line with each of these constraints, thes are set out in more detail in the references given above.

The phase improvement calculation is cyclic and involves modification of the phases in order to obtain best agreement with the known constraints on the electron density and the observed structure factor data. This process is shown diagramatically in Figure 1.

Results of Current Density Modification Methods

The test data featured here is from the structure O^6–methylguanine–DNA methyltransferase[6], which is involved in DNA repair. The diffraction data was measured to 2.1 Angstrom resolution, with MIR phases available to 3.5 Angstrom. The structure was originally solved by averaging of two crystal forms and a variety of density modification techniques.

The map quality of the initial map and improved maps for the different density modification techniques is show in Table 1. The correlation coefficient is calculated from the agreement between elements in the approximated map and a map calculated from the known coordinates. Phase errors are unweighted and are listed for both the initial (MIR) phases and for the extrapolated (high resolution) phases.

The map quality is compared for the initial MIR map, and maps generated using solvent flattening (Wanging), solvent flattening and histogram matching alone, and in combination with Sayre's equation and skeletonisation. The skeletonisation method is similar to the PRISM method[1]. The connectivity of the map can be inferred from the location of the critical points in the electron density (maxima, minima, and saddle points of two kinds)[4,5]. These point can rapidly located by the method of Swanson[7].

A poor section from the MIR map, and the corresponding histogram matched density is shown in figure 2.

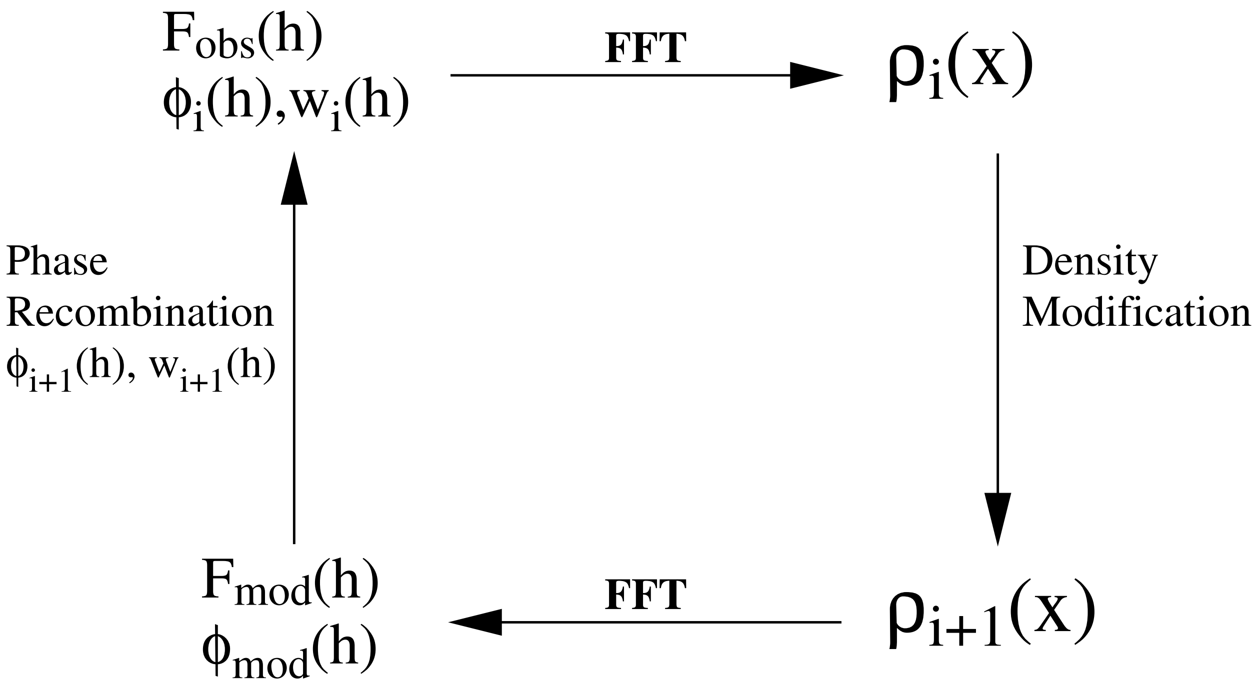

Figure 1: The Density modification process

Density Map	Map Correlation	MeanPhase Error	
		(Initial −3.0A)	(Extended 3.0−2.4A)
MIR	0.539	66.96	90.68
Solv.Flat.	0.599	60.94	88.79
Hist.Match+Solv.Flat.	0.632	59.91	82.98
Sayre+Hist.Match.+Solv.Flat.	0.653	60.19	83.08
Hist.Match+Solv.Flat.+Skeleton	0.677	58.17	81.08

Table 1: Phase Extension Results

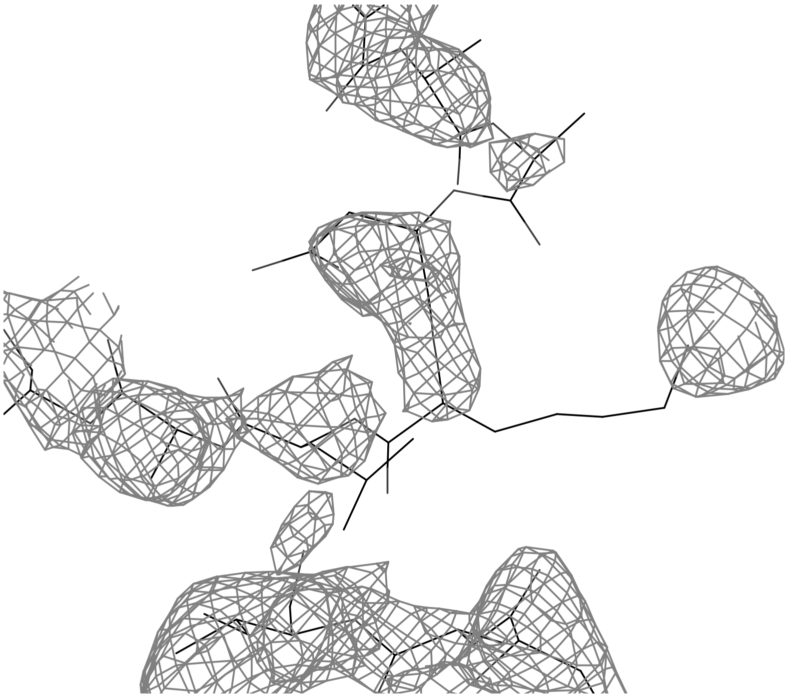

Figure 2(a): MIR map (residue 138−139, contour 1.5 σ)

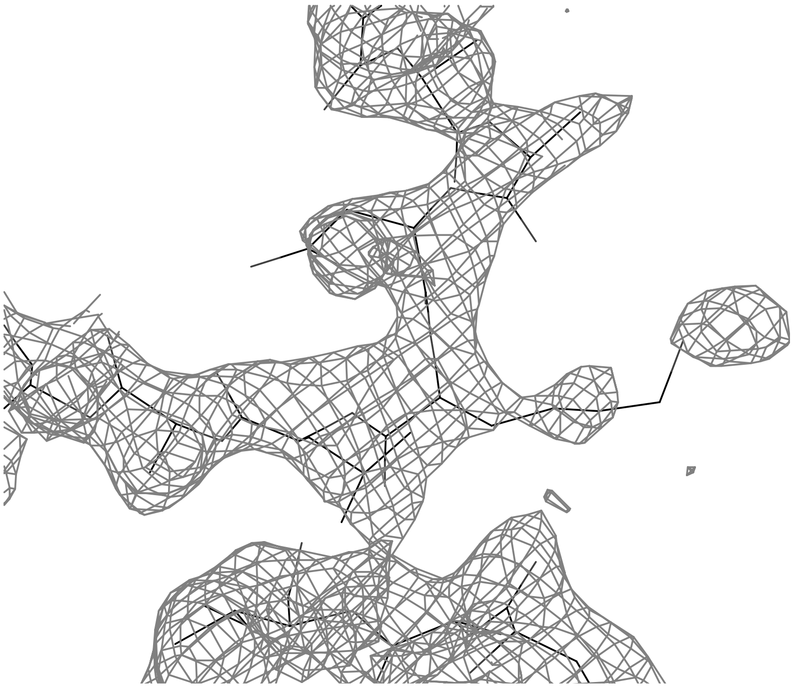

Figure 2(b): Histogram matched and Solvent flattened map

Using Density Modification in CCP4

A new program, 'dm', provides an automatic procedure for using the techniques described above, and has recently been incorporated into the CCP4 library. A single run of the 'dm' program will read an initial reflection file, apply multiple cycles of density modification, and produce an updated reflection file.

The initial reflection file must contain at least the observed structure factor magnitudes and their deviations, and the estimated phases and their weights. In the case of an MIR dataset the phases and weights will come direct from the MIR phasing (typically from the program 'mlphare'). It is also possible to use MAD data, although typically the phase extension scheme will be different since there is usually some phase information at all resolutions. This sort of calculation has proven particularly effective.

Density modification can also be applied in molecular replacement calculation. In this case the correctly oriented model is used to generate a new set of structure factor magnitudes and phases. The agreement between the observed and model magnitudes is then used to generate weights for the model phases (typically using the program 'sigmaa', PARTIAL option). The calculated phase and resultant weight are then input into the density modification calculation.

The use of sensible defaults means that the program can be operated very simply. A typical command file for a basic density modification calculation is shown in figure 3. In the case of a non−crystallographic symmetry averaging calculation, it is necessary to generate a mask covering one non−crystallographic subunit (usually one molecule), and in addition the non−crystallographic symmetry matrices must be included in the command file.

The most common application of the program involves use of the histogram matching and solvent flattening methods. These are very quick (~1 minute for 10 cycles with a unit cell of $2x10^5$A on an SGI workstation) and in the majority of cases lead to a good map improvement. In the case of very poor starting maps, the histogram matching can lead to a very broken map, in which case it may help to apply a temperature factor to smooth the map, or else repeat the calculation with solvent flattening alone. Averaging is always useful if the symmetry elements and mask can be determined. Sayre's equation is usually only helpful when the starting map is very good, and is very slow.

Other features available include input of a user−defined solvent mask, output of the internally calculated solvent mask, averaging of multiple non−crystallographic symmetry related domains with different matrices, and automatic refinement of non− crystallographic symmetry matrices.

The output of the program can be examined using the 'xloggraph' jiffy. The most useful indicator of whether the method has been successful is the 'Free−R factor'. This is analogous to the Free−R factor used in coordinate refinement, in that it is calculated using a set of reflections which are excluded from the initial map calculation. However, once phase relationships between structure factors are introduced it becomes impossible to completely isolate a set of reflection since the absence of a set of reflections leads to a systematically error in the phasing of the rest of the reflections. Thus it is necessary to change the Free−R set from cycle to cycle, or if time is not a factor to run each cycle twice with and without a Free−R set.

The density modification Free−R is a property of both the density modification technique and the initial data set, so it can not be used to compare different density modification techniques. During a typical calculation the density modification Free−R will drop from ~0.60 to ~0.45.

In the case of averaging calculations, the correlation between non−crystallographic symmetry related regions provides a good indication of whether the mask and matrices have been correctly determined. In later cycles this agreement will have increased through the averaging of the regions, and so becomes less informative.

```
#
dm
    hklin gmt_mir.mtz
    hklout gmt_dm.mtz
  << 'MY-DATA'
NCYC 10
SOLC 0.35
MODE SOLV HIST
SCHEME RES FROM 3.0
LABI FP=FP SIGFP=SIGFP PHIO=PHIB FOMO=FOM
LABO PHIDM=PHISF FOMDM=FOMSF
END
'MY-DATA'
#
```

Input reflection file

Ouput reflection file

10 cycles of phase improvement

Solvent Content

Density modification techniques to use

Phase extension in RESolution steps starting from 3.0 Angstrom

Input column labels

Output column labels

Figure 3: A Typical 'dm' Command File

References:

1. Baker D., Bystroff C., Fletterick, R. J., Agard, D. A. (1994). Acta Cryst. D49, 429–439.
2. Bricogne, G. (1974). Acta Cryst. A30, 395–405.
3. Cowtan K. D. & Main, P. (1993). Acta Cryst. D49, 148–157.
4. Gross E. (1994). correspondence.
5. Johnson C. (1994). correspondence.
6. Moore, M. H., Gulbis, J. M., Dodson, E. J., Demple, B. & Moody, P. C. E. (1994). EMBO Journal 13, 1495–1501.
7. Swanson, S. (1994). Acta Cryst. D50, 695–708.
8. Wang, B. C. (1985). Methods Enzymol. 115, 90–111.
9. Wigley, D.B., Roper, D.I. and Cooper R.A. (1989) J. Mol. Biol., 210, 881–882.
10. Zhang, K. Y. J. & Main P. (1990). Acta Cryst. A46, 377–381.