

MLFSOM: simulating the diffraction experiment on an absolute scale

James M. Holton^{1,2}

¹Department of Biochemistry and Biophysics, University of California,
San Francisco, CA 94158

²Physical Biosciences Department, Lawrence Berkeley National Laboratory,
1 Cyclotron Road, Berkeley, CA 94720

Crystallography on the one-micron scale is going to require multi-crystal strategies. From first principles, it can be derived that diffracting an average of one photon into the 2.4 Å spots of a complete data set requires not less than one cubic micron of lysozyme crystal, and the practical reality of background and other sources of noise require that the total volume of scattering matter be larger than this. But what about non-lysozyme crystals? How small can a crystal of a given protein be before assembling a complete data set is theoretically impossible? How high can the “redundancy” be without adding “too much” read-out noise? What about a better detector? What about a perfect detector? Answering these questions requires that damage, noise and signal be placed on a common, absolute scale. To this end, a quantitative simulator of the entire diffraction experiment was created and called “MLFSOM” (MOSFLM in reverse). The input to the simulator is a protein data bank (PDB) file and parameters such as photon flux, beam size, crystal size and detector performance characteristics entered in conventional units such as photons/s and millimeters. MLFSOM was used to produce images in SMV format that were subsequently processed with ELVES. The general result of these trials was that one and only one of the many sources of noise in the diffraction experiment will dominate a given data set. For example, the noise introduced by background scattering limits the signal-to-noise ratio of faint, high-resolution spots, and detector read-out noise is only important in cases where the background is very low. Conversely, the read-out noise of a modern detector cannot have a significant impact on anomalous data, and the optimal strategy for MAD/SAD data collection was collecting a large number of very brief exposures, or “dose slicing”. A simple formula for calculating the required crystal size needed for a complete data set will be presented.