and a ΔF map which results in a map better showing the changes due to errors. Some investigators prefer using further amplified ΔF contributions by using a $(3F_o - 2F_c)$ map or higher-order terms.

The contribution of the disordered solvent continuum has been discussed previously. Macromolecular crystals also contain significant quantities of discrete or partially discrete solvent molecules (*i.e.* water). Care needs to be taken in adding solvent to a model. Errors in models generate peaks in Fourier maps that can be interpreted as solvent peaks. Hence, adding solvent peaks too early in the refinement process may, in fact, lead to model errors. Automatic water-adding programs are becoming more common; examples include *SHELXL*98 and *ARP*/*wARP* (Lamzin & Wilson, 1997). These programs check if the waters are with in reasonable bonding distances of hydrogen-bonding atoms. There is a distribution of solvent molecules ranging from ones with low *B* factors at unit occupancy to ones with very large *B* factors. Various criteria are used to decide on a cutoff in the discrete solvent contribution. A rule of thumb for ambient-temperature data sets is frequently about one solvent molecule per residue in a protein molecule. As more data are being collected at cryogenic temperatures, this ratio is tending to go up. Noise is being fitted if too many peaks in a ΔF map are being assigned as solvent molecules. This can also contribute to reducing *R* factors on incorrect models. Solvent sites may not be fully occupied. Because of the large *B* factors and limited range of the diffraction data, the *B* factors and occupancy are highly correlated. Refinement of occupancy does not usually contribute either to improving a model or to reduction of *R* factors in structures with up to 2.0 Å resolution data. Beyond 1.5 Å data, it may be possible to refine solvent water occupancies and *B* factors. At even higher resolution, some programs, such as *SHELXL*98, provide anisotropic refinement methods which may further improve the solvent model while reducing *R* factors including $R_{\rm free}$.

18.1.9.3. *R and* $R_{\rm free}$

Cross validation is a powerful tool for avoiding over-interpretation of the data by a too elaborate model. The introduction of cross validation to crystallography (Brünger, 1992) has been responsible for significant improvement in the quality of structure determinations. A subset of the reflections, chosen randomly, is segregated and not used in the refinement. If the model is correct and the only errors are statistical, these reflections should have an *R* factor close to that of the reflections used in the refinement. Changes to the model should affect both *R* and $R_{\rm free}$ similarly. Kleywegt & Jones (1997) have pointed out that it is necessary to treat the selection of free reflections very carefully in the presence of noncrystallographic symmetry.

18.1.10. Conclusion

It is always important to bear in mind that macromolecular crystal structures are models intended to explain a particular set of observations. Statistical measures can determine how well the model explains the observations, but cannot say whether the model is true or not. The distinction between precision and accuracy must always be kept in mind. The objective should not be simply to obtain the best fit of a model to the data, but, in addition, to find all of the ways in which a model does *not* fit the data and correct them. Until the day when all crystals diffract to atomic resolution, the primary objective of refinement of the models will be to determine just how well the structures are or are not determined.

references