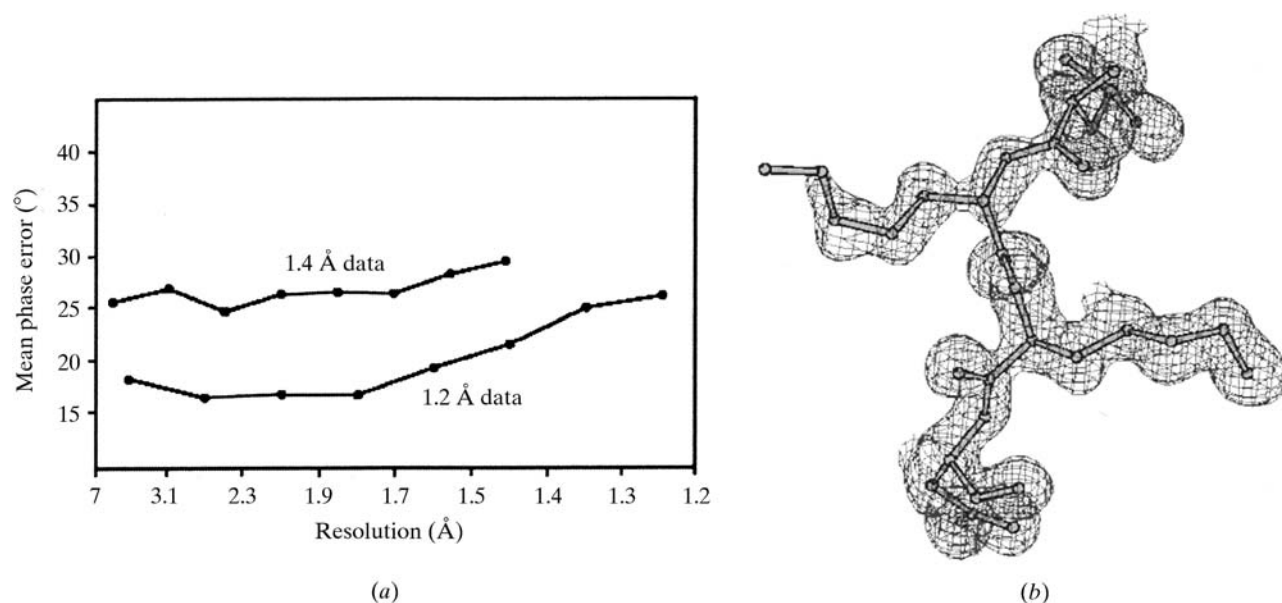


16. DIRECT METHODS

**Figure 16.1.1.2**

(a) Mean phase error as a function of resolution for the two independent *ab initio* SHELXD solutions of the previously unsolved protein hirustasin. Either the 1.2 Å or the 1.4 Å native data set led to solution of the structure. (b) Part of the hirustasin molecule from the 1.4 Å room-temperature data after one round of *B*-value refinement with fixed coordinates.

- (1) *Ab initio* structure solution with atomic resolution data: the whole crystal structure solution is required with most of the atomic sites sufficiently defined for least-squares refinement. Sheldrick's rule applies rigorously here with very few exceptions in the literature. There have been isolated successes at lower resolutions, but these mostly involved the presence of heavier atoms or data with truncated resolution from crystals that would have diffracted further.
- (2) Substructure solution: the determination of the positions of the heavy atoms only (often Se from selenomethionine, but also quite frequently heavy-atom salts, complexes and clusters). Sheldrick's rule is substantially relaxed and structure solutions at data resolutions of 5–6 Å are possible. This follows from the arguments of Morris and Bricogne: the distance beats of 1.1 Å are irrelevant to the substructure. By focusing on the heavy atoms, the complexity of the structure is much reduced, the distance between atoms is larger and the solution of the phase problem becomes easier.

The importance of the presence of several atoms heavier than oxygen for increasing the chance of obtaining a solution by the program *SnB* at resolutions less than 1.2 Å was noticed for truncated data from vancomycin and the 289-atom structure of conotoxin EpI (Weeks & Miller, 1999b). The results of SHELXD application to hirustasin, which contains ten sulfur and 457 carbon, nitrogen and oxygen atoms in the asymmetric unit, are consistent with this (Usón *et al.*, 1999). The 55-amino-acid protein hirustasin could be solved by SHELXD using either 1.2 Å low-temperature data or 1.4 Å room-temperature data. However, as shown in Fig. 16.1.1.2(a), the mean phase error (MPE) is significantly better for the 1.2 Å data over the whole resolution range. Although small-molecule interpretation based on peak positions worked well for the 1.2 Å solution (overall MPE = 18°), standard protein chain tracing was required for the 1.4 Å solution (overall MPE = 26°). As is clear from the corresponding electron-density map (Fig. 16.1.1.2b), SHELXD produced easily interpreted protein density even when bonded atoms are barely resolved from each other.

16.1.1.2. Data completeness

The relative effects of accuracy, completeness and resolution on *Shake-and-Bake* success rates using *SnB* for three large *P1* structures were studied by computing error-free data using the known atomic coordinates (Xu *et al.*, 2000). The results of these studies, presented in Table 16.1.1.1, show that experimental error contributed nothing of consequence to the low success rates for vancomycin and lysozyme. However, completing the vancomycin data up to the maximum measured resolution of 0.97 Å resulted in a substantial increase in success rate which was further improved to an astounding success rate of 80% when the data were expanded to 0.85 Å. As a result of problems with overloaded reflections, the experimental vancomycin data did not include any data at 10 Å resolution or lower. A total of 4000 reflections were phased in the process of solving this structure with the experimental data. Some of these data were then replaced with the largest error-free magnitudes chosen from the missing reflections at several different resolution limits. The results in Table 16.1.1.2 show a tenfold increase in success rate when only 200 of the largest missing magnitudes were supplied, and it made no difference whether these reflections had a maximum resolution of 2.8 Å or were chosen randomly from the whole 0.97 Å sphere. The moral of this story is that, *when collecting synchrotron data for direct methods, it pays to take a second pass using a shorter exposure time to fill in the low-resolution data.*

16.1.1.3. Summary

The basic theory underlying direct methods has been summarized in an excellent chapter (Giacovazzo, 2008) in *International Tables for Crystallography* Volume B (Chapter 2.2) to which the reader is referred for details. Suffice it for this chapter to say that classical direct methods attempt to reconstruct the missing phase information using native data alone by utilizing direct relationships between the crystallographic phases without any *a priori* structural information.

From a historical perspective, the first successful applications of direct methods to native data for structures that could legiti-