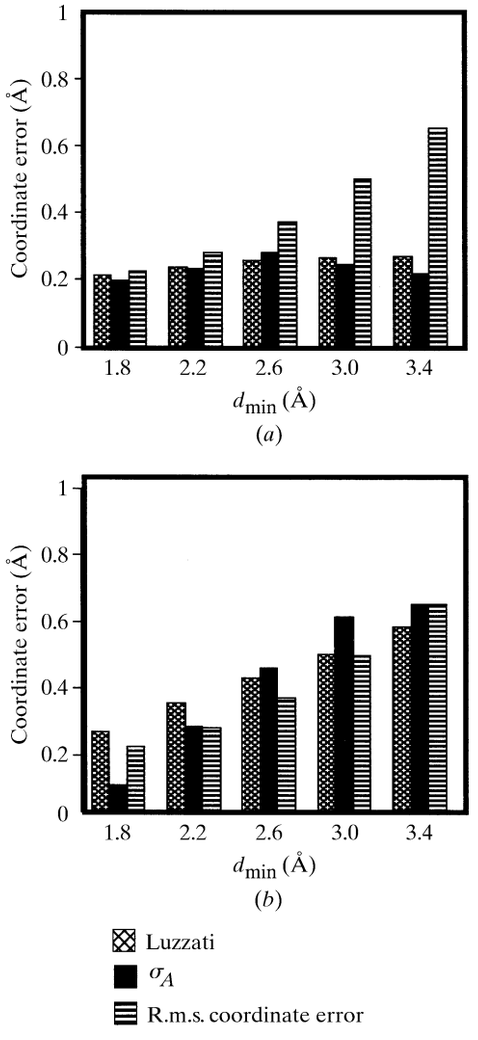

Fig. 18.2.2.1. Effect of resolution on coordinate-error estimates: accuracy as a function of resolution. Refinements were begun with the crystal structure of penicillopepsin (Hsu *et al.*, 1977) with water molecules omitted and with uniform temperature factors. The low-resolution limit was set to 6 Å. Inclusion of all low-resolution diffraction data does not change the conclusions (Adams *et al.*, 1997). The penicillopepsin diffraction data were artificially truncated to the specified high-resolution limit. Each refinement consisted of simulated annealing using a Cartesian-space slow-cooling protocol starting at 2000 K, overall *B*-factor refinement and individual restrained *B*-factor refinement. All refinements were carried out with 10% of the diffraction data randomly omitted for cross validation. (*a*) Coordinate-error estimates of the refined structures using the methods of Luzzati (1952) and Read (1986). All observed diffraction data were used, *i.e.* no cross validation was performed. The actual coordinate errors (r.m.s. differences to the original crystal structure) are shown for comparison. (*b*) Cross-validated coordinate-error estimates. The test set was used to compute the coordinate-error estimates (Kleywegt & Brünger, 1996).

interactions (Hendrickson, 1985). $E_{\text{X-ray}}$ is related to the difference between observed and calculated data, and $w_{\text{X-ray}}$ is a weight appropriately chosen to balance the gradients (with respect to atomic parameters) arising from the two terms.

18.2.3.1. *X-ray diffraction data versus model*

The traditional form of $E_{\text{X-ray}}$ consists of the crystallographic residual, E^{LSQ}, defined as the sum over the squared differences between the observed ($|\mathbf{F}_o|$) and calculated ($|\mathbf{F}_c|$) structure-factor amplitudes for a particular atomic model:

$$E_{\text{X-ray}} = E^{\text{LSQ}} = \sum_{hkl \in \text{working set}} (|\mathbf{F}_o| - k|\mathbf{F}_c|)^2, \qquad (18.2.3.2)$$

where *hkl* are the indices of the reciprocal-lattice points of the crystal and k is a relative scale factor.

Minimization of E^{LSQ} can produce improvement in the atomic model, but it can also accumulate systematic errors in the model by fitting noise in the diffraction data (Silva & Rossmann, 1985). The least-squares residual is a limiting case of the more general maximum-likelihood theory and is only justified if the model is nearly complete and error-free. These assumptions may be violated during the initial stages of refinement. Improved targets for macromolecular refinement have been obtained using the more general maximum-likelihood formulation (Bricogne, 1991; Pannu & Read, 1996; Adams *et al.*, 1997; Murshudov *et al.*, 1997). The goal of the maximum-likelihood method is to determine the likelihood of the model, given estimates of the model's errors and those of the measured intensities.

A starting point for the maximum-likelihood formulation of crystallographic refinement is the Sim (1959) distribution, *i.e.*, the Gaussian conditional probability distribution of the 'true' structure factors, $\mathbf{F}$, given a partial model with structure factors $\mathbf{F}_c$ and the model's error (Fig. 18.2.3.1) (Srinivasan, 1966; Read, 1986, 1990) (for simplicity we will only discuss the case of acentric reflections),

$$P_a(\mathbf{F};\, \mathbf{F}_c) = (1/\pi\varepsilon\sigma_\Delta^2)\exp[-(\mathbf{F} - D\mathbf{F}_c)^2/\varepsilon\sigma_\Delta^2], \qquad (18.2.3.3)$$

where σ_Δ is a parameter that incorporates the effect of the fraction of the asymmetric unit that is missing from the model and errors in the partial structure. Assuming a Wilson distribution of intensities, it can be shown that (Read, 1990)

$$\sigma_\Delta^2 = \langle|\mathbf{F}_o|^2\rangle - D^2\langle|\mathbf{F}_c|^2\rangle, \qquad (18.2.3.4)$$

where D is a factor that takes into account model error: it is unity in the limiting case of an error-free model and it is zero if no model is available (Luzzati, 1952; Read, 1986). For a complete and error-free model, σ_Δ therefore becomes zero, and the probability distribution, $P_a(\mathbf{F};\, \mathbf{F}_c)$, is infinitely sharp.

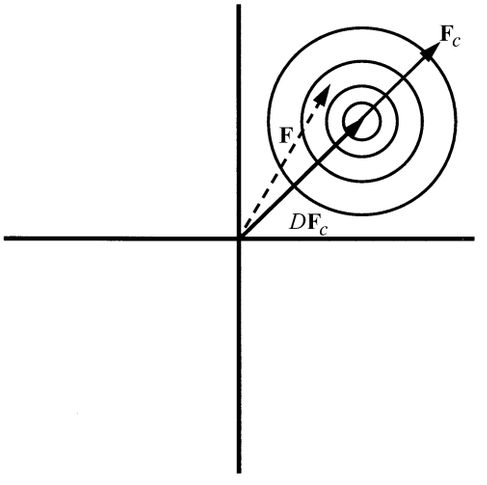

Fig. 18.2.3.1. The Gaussian probability distribution forms the basis of maximum-likelihood targets in crystallographic refinement. The conditional probability of the true structure factor, $\mathbf{F}$, given model structure factors, is a Gaussian in the complex plane [equation (18.2.3.3)]. The expected value of the probability distribution is $D\mathbf{F}_c$ with variance σ_Δ, where D and σ_Δ account for missing or incorrectly placed atoms in the model.

references