International Tables for Crystallography (2006). Vol. F, Section 18.5.8.1, pp. 412–413.

Table 18.5.7.3. *Comparison of DPIs using R and* $R_{\rm free}$

The second row for each protein contains values appropriate to the DPI equation (18.5.6.10) using $R_{\rm free}$.

Protein	N_i	$n_{\rm obs}$	$(N_i/p)^{1/2}$, $(N_i/n_{\rm obs})^{1/2}$	$C^{-1/3}$	R, $R_{\rm free}$	$d_{\rm min}$ (Å)	DPI $\sigma(r, B_{\rm avg})$ (Å)	Luzzati $\langle \Delta r \rangle$ (Å)	Read $\langle \Delta r \rangle$ (Å)	Reference
Concanavalin A	2130	116712	0.148 0.135	1.099	0.128 0.148	0.94	0.034 0.036	0.06		(*a*)
γB-Crystallin	1708	26151	0.297 0.256	1.032	0.180 0.204	1.49	0.14 0.14	0.16	0.12	(*b*)
βB2-Crystallin	1558	18583	0.356 0.290	~1.032	0.184 0.200	2.10	0.25 0.22	0.21	0.17	(*b*)
Ribonuclease A with RI	4416	18859	1.922 0.484	1.145	0.194 0.286	2.50	1.85 0.69	0.32	0.57	(*c*)
Fab HyHEL-5 with HEWL	4333	11754	* 0.607	1.111	0.196 0.288	2.65	— 0.69	0.30		(*d*)

References: (*a*) Deacon *et al.* (1997); (*b*) Tickle *et al.* (1998*a*); (*c*) Kobe & Deisenhofer (1995); (*d*) Cohen *et al.* (1996).
* (N_i/p) negative.

available. The second row for each protein shows the alternative values for $(N_i/n_{\rm obs})^{1/2}$, $R_{\rm free}$ and the DPI $\sigma(r, B_{\rm avg})$ from (18.5.6.10).

For the structures with $d_{\rm min} \leq 2.0$ Å, the DPI is much the same whether it is based on R or $R_{\rm free}$.

Tickle *et al.* (1998*a*) have made full-matrix error estimates for isotropic restrained refinements of γB-crystallin with $d_{\rm min} = 1.49$ Å and of βB2-crystallin with $d_{\rm min} = 2.10$ Å. The DPI $\sigma(r, B_{\rm avg})$ calculated for the two structures is 0.14 and 0.25 Å with R in (18.5.6.9), and 0.14 and 0.22 Å with $R_{\rm free}$ in (18.5.6.10). The full-matrix weighted averages of $\sigma_{\rm res}(r)$ for all protein atoms were 0.10 and 0.15 Å, for only main-chain atoms 0.05 and 0.08 Å, for side-chain atoms 0.14 and 0.20 Å, and for water oxygens 0.27 and 0.35 Å. Again, the DPI gives reasonable overall indices for the quality of the structures.

For the complex of bovine ribonuclease A and porcine ribonuclease inhibitor (Kobe & Deisenhofer, 1995) with $d_{\rm min} = 2.50$ Å, the number of reflections is only just larger than the number of parameters, so that $(N_i/p)^{1/2} = 1.922$ is very large, and the DPI with R gives an unrealistic 1.85 Å. With $R_{\rm free}$, $\sigma(r, B_{\rm avg}) = 0.69$ Å.

The HyHEL-5–lysozyme complex (Cohen *et al.*, 1996) had $d_{\rm min} = 2.65$ Å. Here the number of reflections is less than the number of parameters, but the $R_{\rm free}$ formula gives $\sigma(r, B_{\rm avg}) = 0.69$ Å.

18.5.7.4. *Comments on the diffraction-component precision index*

The DPI (18.5.6.9) or (18.5.6.10) provides a very simple formula for $\sigma(r, B_{\rm avg})$. It is based on a very rough approximation to a diagonal element of the diffraction-data-only matrix. Using a diagonal element is a reasonable approximation for atomic resolution structures, but for low-resolution structures there will be significant off-diagonal terms between overlapping atoms. The effect can be simulated in the two-atom protein model of Section 18.5.3.2 by introducing positive off-diagonal elements into the diffraction-data matrix (18.5.3.3). As expected, $\sigma^2_{\rm diff}(x_i)$ is increased. So the DPI will be an underestimate of the diffraction component in low-resolution structures.

However, the true restrained variance $\sigma^2_{\rm res}(x_i)$ in the new counterpart of (18.5.3.12) remains less than the diagonal diffraction result (18.5.3.11) $\sigma^2_{\rm diff}(x_i) = 1/a$. Thus for low-resolution structures, the DPI should be an overestimate of the true precision given by a restrained full-matrix calculation (where the restraints act to hold the overlapping atoms apart). This is confirmed by the results for the 2.1 Å study of βB2-crystallin (Tickle *et al.*, 1998*a*) discussed in Section 18.5.7.3 and Table 18.5.7.3. The restrained full-matrix average for all protein atoms was $\sigma_{\rm res}(r) = 0.15$ Å, compared with the DPI 0.25 Å (on R) or 0.22 Å (on $R_{\rm free}$). The ratio between the unrestrained DPI and the restrained full-matrix average is consistent with a view of a low-resolution protein as a chain of effectively rigid peptide groups. The ratio no doubt gets much worse for resolutions of 3 Å and above.

The DPI estimate of $\sigma(r, B_{\rm avg})$ is given by a formula of 'back-of-an-envelope' simplicity. $B_{\rm avg}$ is taken to be the average B for fully occupied sites, but the weights implicit in the averaging are not well defined in the derivation of the DPI. Thus the DPI should perhaps be regarded as simply offering an estimate of a typical $\sigma_{\rm diff}(r)$ for a carbon or nitrogen atom with a mid-range B. From the evidence of the tables in this section, except at low resolution, it seems to give a useful overall indication of protein precision, even in restrained refinements.

The DPI evidently provides a method for the comparative ranking of different structure determinations. In this regard it is a complement to the general use of $d_{\rm min}$ as a quick indicator of possible structural quality.

Note that (18.5.6.3) and (18.5.6.4) offer scope for making individual error estimates for atoms of different B and Z.

18.5.8. Luzzati plots

18.5.8.1. *Luzzati's theory*

Luzzati (1952) provided a theory for estimating, at any stage of a refinement, the average positional shifts which would be needed in an idealized refinement to reach $R = 0$. He did not provide a theory for estimating positional errors at the end of a normal refinement.

(1) His theory assumed that the $F_{\rm obs}$ had no errors, and that the

Table 18.5.8.1. $R = \langle|\Delta F|\rangle/\langle|F|\rangle$ *as a function of* $s\langle\Delta r\rangle$ *in the Luzzati model for three-dimensional noncentrosymmetric structures* ($s = 2\sin\theta/\lambda$)

$s\langle\Delta r\rangle$	R	$s\langle\Delta r\rangle$	R
0.00	0.000	0.10	0.237
0.01	0.025	0.12	0.281
0.02	0.050	0.14	0.319
0.03	0.074	0.16	0.353
0.04	0.098	0.18	0.385
0.05	0.122	0.20	0.414
0.06	0.145	0.25	0.474
0.07	0.168	0.30	0.518
0.08	0.191	0.35	0.548
0.09	0.214	∞	0.586

$F_{\rm calc}$ model (scattering factors, thermal parameters *etc.*) was perfect, apart from coordinate errors.

(2) The Gaussian probability distribution for these coordinate errors was assumed to be the *same for all atoms*, independent of Z or B.

(3) The atoms were not required to be identical, and the position errors were not required to be small.

Luzzati gave families of curves for R *versus* $2\sin\theta/\lambda$ for varying average positional errors $\langle\Delta r\rangle$ for both centrosymmetric and noncentrosymmetric structures. The curves do not depend on the number N of atoms in the cell. They all rise from $R = 0$ at $2\sin\theta/\lambda = 0$ to the Wilson (1950) values 0.828 and 0.586 for random structures at high $2\sin\theta/\lambda$. Table 18.5.8.1 gives $R = \langle|\Delta F|\rangle/\langle|F|\rangle$ as a function of $s\langle\Delta r\rangle$ for three-dimensional noncentrosymmetric structures.

In a footnote (p. 807), Luzzati suggested that at the end of a normal refinement (with R nonzero due to experimental and model errors, *etc.*), the curves would indicate an upper limit for $\langle\Delta r\rangle$. He noted that typical small-molecule $\sigma(r)$'s of 0.01–0.02 Å, if used as $\langle\Delta r\rangle$ in the plots, would give much smaller R's than are found at the end of a refinement.

As examples, the Luzzati plots for the two structures of TGF-$\beta 2$ are shown in Fig. 18.5.8.1. Daopin *et al.* (1994) inferred average $\langle\Delta r\rangle$'s around 0.21 Å for 1TGI and 0.23 Å for 1TGF.

Of the three Luzzati assumptions summarized above, the most attractive is the third, which does not require the atoms to be identical nor the position errors to be small. For proteins, there are very obvious difficulties with assumption (2). Errors do depend very strongly on Z and B. In the high-angle data shells, atoms with large B's contribute neither to ΔF nor to $|F|$, and so have no effect on R in these shells. In their important paper on protein accuracy, Chambers & Stroud (1979) said 'the [Luzzati] estimate derived from reflections in this range applies mainly to [the] best determined atoms.'

Thus a Luzzati plot seems to allow a cautious upper-limit statement about the precision of the best parts of a structure, but it gives little indication for the poor parts.

One reason for the past popularity of Luzzati plots has been that the R values for the middle and outer shells of a structure often roughly follow a Luzzati curve. Evidently, the effective average $\langle\Delta r\rangle$ for the structure must be decreasing as $2\sin\theta/\lambda$ increases, since atoms of high B are ceasing to contribute, whereas the proportionate experimental errors must be increasing. This also suggests that the upper limit for $\langle\Delta r\rangle$ for the low-B atoms could be estimated from the lowest Luzzati theoretical curve touched by the experimental R plot. Thus in Fig. 18.5.8.1 the upper limits for the low-B atoms could be taken as 0.18 and 0.21 Å, rather than the 0.21 and 0.23 Å chosen by Daopin *et al.*

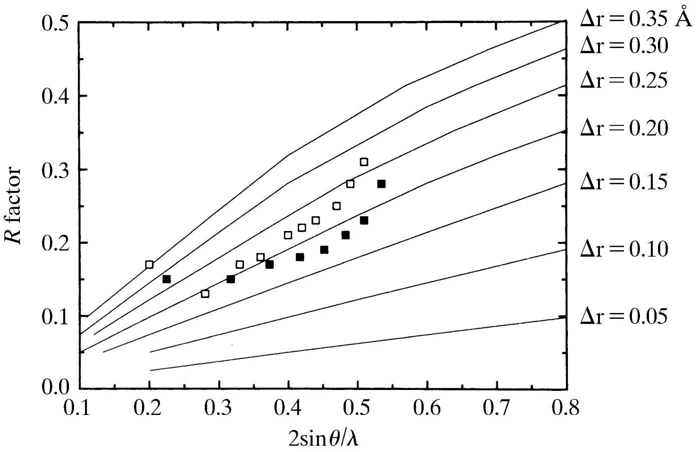

Fig. 18.5.8.1. Luzzati plots showing the refined R factor as a function of resolution for 1TGI (solid squares) and 1TGF (open squares) (Daopin *et al.*, 1994).

From the introduction of $R_{\rm free}$ by Brünger (1992) and the discussion of $R_{\rm free}$ by Tickle *et al.* (1998*b*), it can be seen that Luzzati plots should be based on a residual more akin to $R_{\rm free}$ than R in order to avoid bias from the fitting of data.

The mean positional error $\langle\Delta r\rangle$ of atoms can also be estimated from the σ_A plots of Read (1986, 1990). This method arose from Read's analysis of improved Fourier coefficients for maps using phases from partial structures with errors. It is preferable in several respects to the Luzzati method, but like the Luzzati method it assumes that the coordinate distribution is the same for all atoms. Luzzati and/or Read estimates of $\langle\Delta r\rangle$ are available for some of the structures in Tables 18.5.7.2 and 18.5.7.3. Often, the two estimates are not greatly different.

18.5.8.2. *Statistical reinterpretation of Luzzati plots*

Luzzati plots are fundamentally different from other statistical estimates of error. The Luzzati theory applies to an idealized incomplete refinement and estimates the average shifts needed to reach $R = 0$. In the least-squares method, the equations for shifts are quite different from the equations for estimating variances in a converged refinement. However, Luzzati-style plots of R *versus* $2\sin\theta/\lambda$ can be reinterpreted to give statistically based estimates of $\sigma(x)$.

During Cruickshank's (1960) derivation of the approximate equation (18.5.6.2) for $\sigma(x)$ in diagonal least squares, he reached an intermediate equation

$$\sigma^2(x) = N_i \bigg/ \left[4\sum_{\rm obs}(s^2/R^2)\right]. \qquad (18.5.8.1)$$

He then assumed R to be independent of s $(= 2\sin\theta/\lambda)$ and took R outside the summation to reach (18.5.6.2) above.

Luzzati (1952) calculated the acentric residual R as a function of $\langle\Delta r\rangle$, the average radial error of the atomic positions. His analysis shows that R is a linear function of s and $\langle\Delta r\rangle$ for a substantial range of $s\langle\Delta r\rangle$, with

$$R(s,\ \langle\Delta r\rangle) = (2\pi)^{1/2}s\langle\Delta r\rangle. \qquad (18.5.8.2)$$

The theoretical Luzzati plots of R are nearly linear for small-to-medium $s = 2\sin\theta/\lambda$ (see Fig. 18.5.8.1). If we substitute this R in the least-squares estimate (18.5.8.1) and use the three-dimensional-Gaussian relation $\sigma(r) = 1.085\langle\Delta r\rangle$, some manipulation (Cruickshank, 1999) along the lines of Section 18.5.6 eventually yields a statistically based formula,