Table 18.5.7.1. *Comparison of full-matrix $\sigma(r, B_{\rm avg})$ with the diffraction-component precision index (DPI)*

Protein	$(N_i/p)^{1/2}$	R	$d_{\rm min}$ (Å)	DPI $\sigma(r, B_{\rm avg})$ (Å)	Full-matrix $\sigma_{\rm diff}(r, B_{\rm avg})$ (Å)	Reference
Concanavalin A	0.148	0.128	0.94	0.034	0.033	(*a*)
Immunoglobulin	0.476	0.156	1.70	0.221	0.186	(*b*)

References: (*a*) Deacon *et al.* (1997); (*b*) Usón *et al.* (1999).

18.5.7. Examples of the diffraction-component precision index

18.5.7.1. *Full-matrix comparison with the diffraction-component precision index*

The DPI (18.5.6.9) with R was offered as a quick and rough guide for the diffraction-data-only error for an atom with $B = B_{\rm avg}$. The necessary data for the comparison with the two unrestrained full-matrix inversions of Section 18.5.5 are given in Table 18.5.7.1. For concanavalin A with $B_{\rm avg} = 14.8$ Å^2, the full-matrix quadratic (18.5.4.2*b*) gives 0.033 Å for a carbon atom and the DPI gives 0.034 Å for an unspecified atom. For the immunoglobulin with $B_{\rm avg} = 26.8$ Å^2, the full-matrix quadratic (18.5.4.2*a*) gives $\sigma_{\rm diff}(r) = 0.19$ Å for a carbon atom, while the DPI gives 0.22 Å.

For these two structures, the simple DPI formula compares surprisingly well with the unrestrained full-matrix calculations at $B_{\rm avg}$.

For the restrained full-matrix calculations on concanavalin A, the quadratic (18.5.4.2*c*) with $B = B_{\rm avg}$ gives $\sigma_{\rm res}(r) = 0.028$ Å for a carbon atom, which is only 15% smaller than the unrestrained 0.033 Å. This small decrease matches the discussion of $\sigma_{\rm res}(r)$ and $\sigma_{\rm diff}(r)$ in Section 18.5.4.1 following equation (18.5.4.1). But that discussion also indicates that for the immunoglobulin, the restrained $\sigma_{\rm res}(r, B_{\rm avg})$, which was not computed, will be proportionaly much lower than the unrestrained value of $\sigma_{\rm diff}(r, B_{\rm avg}) = 0.19$ Å, since the restraints are relatively more important in the immunoglobulin.

18.5.7.2. *Further examples of the DPI using R*

Table 18.5.7.2 shows a range of examples of the application of the DPI (18.5.6.9) using R to proteins of differing precision, starting with the smallest $d_{\rm min}$. In all the examples, N_i has been set equal to $n_{\rm atoms}$, the total number of atoms. The ninth and tenth columns show $\langle\Delta r\rangle$ values derived from Luzzati (1952) and Read (1986) plots described later in Section 18.5.8.

The first entry is for crambin at 0.83 Å resolution and 130 K (Stec *et al.*, 1995). Their results were obtained from an unrestrained full-matrix anisotropic refinement. Inversion of the full matrix gave s.u.'s $\sigma_{\rm diff}(x) = 0.0096$ Å for backbone atoms, 0.0168 Å for side-chain atoms and 0.0409 Å for solvent atoms, with an average for all atoms of 0.022 Å. The DPI $\sigma(r, B_{\rm avg}) = 0.021$ Å corresponds to $\sigma(x) = 0.012$ Å, which is satisfactorily intermediate between the full-matrix values for the backbone and side-chain atoms.

Sevcik *et al.* (1996) carried out restrained anisotropic full-matrix refinements on data from two slightly different crystals of ribonuclease Sa, with $d_{\rm min}$ of 1.15 and 1.20 Å. They inverted full-matrix blocks containing parameters of 20 residues to estimate coordinate errors. The overall r.m.s. coordinate error for protein atoms is given as 0.03 Å, and for all atoms (including waters and ligands) as 0.07 Å for MGMP and 0.05 Å for MSA. The DPI gives $\sigma(r, B_{\rm avg}) = 0.05$ Å for both structures.

The next entries concern the two lower-resolution (1.8 and 1.95 Å) studies of TGF-β2 (Daopin *et al.*, 1994). The DPI gives $\sigma(r) = 0.16$ Å for 1TGI and 0.24 Å for 1TGF. This indicates an r.m.s. position difference between the structures for atoms with $B_i = B_{\rm avg}$ of $(0.16^2 + 0.24^2)^{1/2} = 0.29$ Å. Daopin *et al.* reported the differences between the two determinations, omitting poor parts, as $\langle\Delta r\rangle_{\rm rms} = 0.15$ Å (main chain) and 0.29 Å (all atoms).

Human diferric lactoferrin (Haridas *et al.*, 1995) is an example of a large protein at the lower resolution of 2.2 Å, with a high value of $(N_i/p)^{1/2}$, leading to $\sigma(r, B_{\rm avg}) = 0.43$ Å.

Three crystal forms of thaumatin were studied by Ko *et al.* (1994). The orthorhombic and tetragonal forms diffracted to 1.75 Å, but the monoclinic $C2$ form diffracted only to 2.6 Å. The structures with 1552 protein atoms were successfully refined with restraints by *XPLOR* and *TNT*. For the monoclinic form, the number of parameters exceeds the number of diffraction observations, so (N_i/p) is negative and no estimate by (18.5.6.9) of the diffraction-data-only error is possible. The DPI (18.5.6.9) gives 0.17 and 0.16 Å for the orthorhombic and tetragonal forms, respectively.

18.5.7.3. *Examples of the DPI using $R_{\rm free}$*

As in the case of monoclinic thaumatin, for low-resolution structures the number of parameters may exceed the number of diffraction data. To circumvent this difficulty, it was proposed in Section 18.5.6.3 to replace $p = n_{\rm obs} - n_{\rm params}$ by $n_{\rm obs}$ and R by $R_{\rm free}$ in a revised formula (18.5.6.10) for the DPI. Table 18.5.7.3 shows examples for some structures for which both R and $R_{\rm free}$ were

Table 18.5.7.2. *Examples of diffraction-component precision indices (DPIs)*

Protein	N_i	$n_{\rm obs}$	$(N_i/p)^{1/2}$	$C^{-1/3}$	R	$d_{\rm min}$ (Å)	DPI $\sigma(r, B_{\rm avg})$ (Å)	Luzzati $\langle\Delta r\rangle$ (Å)	Read $\langle\Delta r\rangle$ (Å)	Reference
Crambin	447	23759	0.150	1.074	0.090	0.83	0.021	0.055		(*a*)
Ribonuclease MGMP	1958	62845	0.208	1.046	0.109	1.15	0.047		0.08	(*b*)
Ribonuclease MSA	1832	60670	0.204	1.016	0.106	1.20	0.045		0.05	(*b*)
TGF-β2 1TGI	948	∼14000	0.305	∼1.0	0.173	1.80	0.16	0.21	0.18	(*c*)
TGF-β2 1TFG	974	∼11000	0.370	∼1.0	0.188	1.95	0.24	0.23		(*c*)
Lactoferrin	5907	39113	0.618	1.036	0.179	2.20	0.43	0.25–0.30	0.35	(*d*)
Thaumatin C2	1552	4622	*	1.10	0.184	2.60	—	0.25		(*e*)

References: (*a*) Stec *et al.* (1995); (*b*) Sevcik *et al.* (1996); (*c*) Daopin *et al.* (1994); (*d*) Haridas *et al.* (1995); (*e*) Ko *et al.* (1994).
* (N_i/p) negative.

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

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

Protein	N_i	n_{obs}	$(N_i/p)^{1/2}$, $(N_i/n_{obs})^{1/2}$	$C^{-1/3}$	R, R_{free}	d_{min} (Å)	DPI $\sigma(r, B_{avg})$ (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 $(N_i/n_{obs})^{1/2}$, R_{free} and the DPI $\sigma(r, B_{avg})$ from (18.5.6.10).

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

Tickle *et al.* (1998*a*) have made full-matrix error estimates for isotropic restrained refinements of γB-crystallin with $d_{min} = 1.49$ Å and of βB2-crystallin with $d_{min} = 2.10$ Å. The DPI $\sigma(r, B_{avg})$ 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_{free} in (18.5.6.10). The full-matrix weighted averages of $\sigma_{res}(r)$ 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 $d_{min} = 2.50$ 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_{free}, $\sigma(r, B_{avg}) = 0.69$ Å.

The HyHEL-5–lysozyme complex (Cohen *et al.*, 1996) had $d_{min} = 2.65$ Å. Here the number of reflections is less than the number of parameters, but the R_{free} formula gives $\sigma(r, B_{avg}) = 0.69$ 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 $\sigma(r, B_{avg})$. 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, $\sigma^2_{diff}(x_i)$ is increased. So the DPI will be an underestimate of the diffraction component in low-resolution structures.

However, the true restrained variance $\sigma^2_{res}(x_i)$ in the new counterpart of (18.5.3.12) remains less than the diagonal diffraction result (18.5.3.11) $\sigma^2_{diff}(x_i) = 1/a$. 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 $\sigma_{res}(r) = 0.15$ Å, compared with the DPI 0.25 Å (on R) or 0.22 Å (on R_{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 $\sigma(r, B_{avg})$ is given by a formula of 'back-of-an-envelope' simplicity. B_{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 $\sigma_{diff}(r)$ 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_{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_{obs} had no errors, and that the