restrained and unrestrained $\sigma(r)$. The reason is obvious. Most atoms are linked by 1–2 bond restraints to two or three other atoms. Even a carbonyl oxygen atom linked to its carbon atom by a 0.02 Å restraint is also subject to 0.04 Å 1–3 restraints to chain C_α and N atoms. Consequently, for a high-B atom, when the restraints are applied it is coupled to several other atoms in a group, and its $\sigma_{\rm res}(r)$ is lower, compared with the diffraction-data-only $\sigma_{\rm diff}(r)$, by a greater amount than would be expected from the two-atom model.

18.5.4.2. *Unrestrained inversion for an immunoglobulin*

Sheldrick has provided the results of the unrestrained lower-resolution refinement of a single-chain immunoglobulin mutant (T39K) with 218 amino-acid residues, with data to 1.70 Å refined isotropically (Usón *et al.*, 1999). Fig. 18.5.4.3 shows $\sigma_{\rm diff}(r)$ *versus* $B_{\rm eq}$ for the fully occupied protein atoms. Superposed on the data points are least-squares quadratic fits. In a first very rough approximation for $\sigma_{\rm diff}(x_i)$ suggested later by equation (18.5.6.3), the dependence on atom type is controlled by $1/Z_i$, the reciprocal of the atomic number. Sheldrick found that a $1/Z_i$ dependence produced too little difference between C, N and O. The proportionalities between the quadratics for $\sigma(r)$ in Figs. 18.5.4.1 and 18.5.4.3 are based on the reciprocals of the scattering factors at $\sin\theta/\lambda = 0.3$ Å^{-1}, symbolized by $Z_i^{\#}$. For C, N and O, these are 2.494, 3.219 and 4.089, respectively. For potential use in later work, the least-squares fits to the $\sigma(r_i)Z_i^{\#}$ in Å are recorded here as

$$0.11892 + 0.00891B + 0.0001462B^2, \qquad (18.5.4.2a)$$

$$0.01826 + 0.001043B + 0.0002230B^2 \text{ and} \qquad (18.5.4.2b)$$

$$0.00115 + 0.004414B + 0.0000214B^2 \qquad (18.5.4.2c)$$

for the immunoglobulin (unrestrained), concanavalin A (unrestrained) and concanavalin A (restrained), respectively.

As might be expected from the lower resolution, the lowest $\sigma_{\rm diff}(r)$'s in the immunoglobulin are about six times the lowest $\sigma_{\rm diff}(r)$'s in concanavalin. But at $B = 50$ Å^2, the immunoglobulin curve for carbon gives $\sigma_{\rm diff}(r) = 0.37$ Å, which is only 50% larger than the concanavalin value of 0.25 Å.

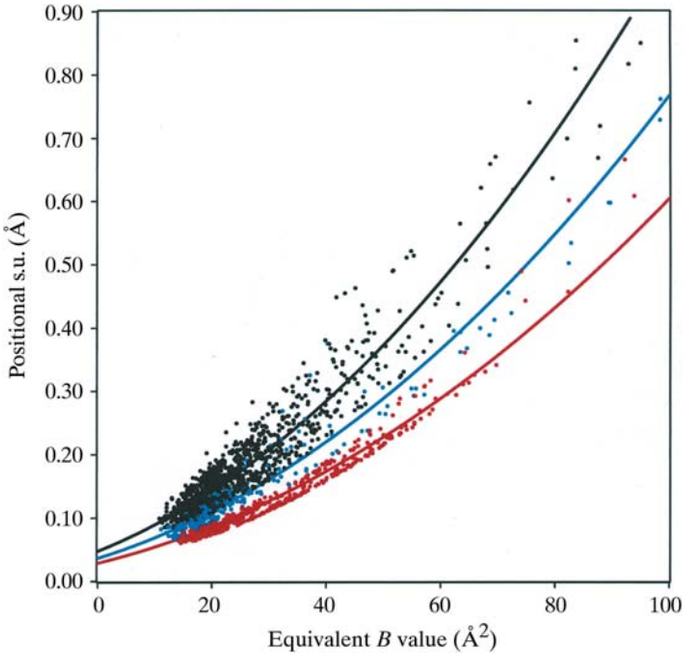

Fig. 18.5.4.3. Plot of $\sigma_{\rm diff}(r)$ *versus* $B_{\rm eq}$ from an unrestrained full matrix for immunoglobulin mutant (T39K) with 1.70 Å data. Carbon black, nitrogen blue, oxygen red.

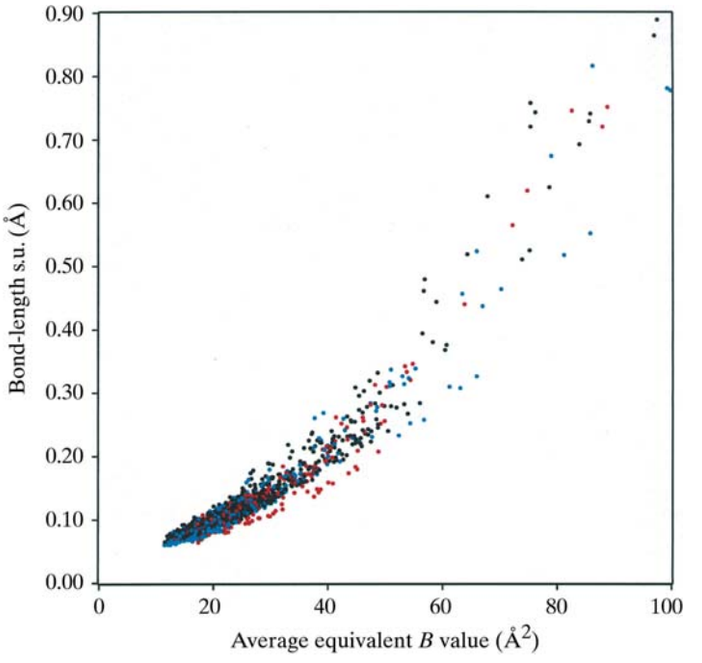

Fig. 18.5.4.4. Plot of $\sigma_{\rm diff}(l)$ *versus* average $B_{\rm eq}$ from an unrestrained full matrix for immunoglobulin mutant (T39K) with 1.70 Å data. C—O black, C—N blue, C—O red.

Fig. 18.5.4.4 shows $\sigma_{\rm diff}(l)$ *versus* $B_{\rm eq}$ for the immunoglobulin. Note that the lowest immunoglobulin unrestrained $\sigma_{\rm diff}(l)$ is about 0.06 Å, which is three times the 0.02 Å $\sigma_{\rm geom}(l)$ bond restraint.

18.5.4.3. *Comments on restrained refinement*

Geometric restraint dictionaries typically use bond-length weights based on $\sigma_{\rm geom}(l)$ of around 0.02 or 0.03 Å. Tables 18.5.7.1–18.5.7.3 show that even 1.5 Å studies have diffraction-only errors $\sigma_{\rm diff}(x, B_{\rm avg})$ of 0.08 Å and upwards. Only for resolutions of 1.0 Å or so are the diffraction-only errors comparable with the dictionary weights. Of course, the dictionary offers no values for many of the configurational parameters of the protein structure, including the centroid and molecular orientation.

18.5.4.4. *Full-matrix estimates of precision*

The opening contention of this chapter in Section 18.5.1.1 is that the variances and covariances of the structural parameters of proteins can be found from the inverse of the least-squares normal matrix. But there is a caveat, chiefly that explicit account would not be taken of disorder of the solvent or of parts of the protein. Corrections by Babinet's principle of complementarity or by mask bulk solvent models are only first-order approximations. The consequences of such disorder problems, which make the variation of calculated structure factors nonlinear over the range of interest, may in future be better handled by maximum-likelihood methods (*e.g.* Read, 1990; Bricogne, 1993*a*; Bricogne & Irwin, 1996; Murshudov *et al.*, 1997). Pannu & Read (1996) have shown how the maximum-likelihood method can be cast computationally into a form akin to least-squares calculations. Full-matrix precision estimates along the lines of the present chapter are probably somewhat low.

It should also be noted that full-matrix estimates of coordinate precision are most reliably derived from matrices involving both

references