Acta Cryst. (1997). D**53**, 41–48

Weighting Macromolecular Diffraction Data

G. David Smith

Hauptman–Woodward Medical Research Institute, 73 High Street, Buffalo, NY 14203 USA, and Roswell Park Cancer Institute, Elm and Carlton Streets, Buffalo, NY 14263 USA. E-mail: smith@hwi.buffalo.edu

(*Received 15 April 1996; accepted 17 September 1996*)

Abstract

The results from refinements with six different weighting schemes in *PROFFT*, *SHELXL*93 and *X-PLOR* show that experimental weights, and especially unit weights, tend to overweight the low-angle data at the expense of the higher resolution data. This suggests that the use of unit weights are inappropriate except during initial cycles of refinements. Two alternative weighting schemes, one-line and two-line empirical weights, are shown to produce a relatively even distribution of the weighted r.m.s. error [r.m.s. Err = $(\sum w\Delta F^2/n)^{1/2}$] as a function of $\sin\theta/\lambda$ and result in linear $\delta(R)$ normal probability plots. Free R values and free weighted r.m.s. errors from the refinement of a scorpion toxin structure show that the alternative weighting schemes are superior to unit weights, and the r.m.s. deviations of bond distances and angles from the target values and G factors, as calculated by *PROCHECK*, confirm the superiority of the empirical weighting scheme. It was also observed that changing the value of WA, the contribution of the X-ray gradient to the total function minimized, produced little variation in R free but was directly proportional to the r.m.s. deviation of the bond distances from the target values.

1. Introduction

The refinement of a small-molecule crystal structure, using a full-matrix least-squares procedure with three positional parameters and six anisotropic thermal parameters from non-H atoms, is usually relatively straightforward given that data to atomic resolution or better are available. Towards the end of the refinement, all H atoms are frequently located and refined, and residuals of less than 0.04 are not uncommon. Results such as these in part result from the fact that all atoms which contribute to the scattering are included in the model. If reasonable care has been taken in the measurement of the structure factors, most systematic errors, with the exception of perhaps absorption, have been minimized. The function which is frequently minimized is $\sum w(F_o - F_c)^2$, and assuming that all known sources of error including not only counting statistics but also 'the instability constant' have been propagated into the estimated error of each structure factor (Blessing, 1987), the diffraction data are typically weighted by the inverse of the variance of the structure factors $[w = 1/\sigma^2(F_o)]$. Under these circumstances, a typical small-molecule structure (Howell, Pangborn, Marshall, Zabrocki & Smith, 1995) will exhibit a relatively small variation in the residuals, ΔF, and the goodness of fit $\{\text{GOF} = [\sum w\Delta F^2/(n-m)]^{1/2}$, where n is the number of data and m is the number of parameters$\}$ as a function of $\sin\theta/\lambda$. A more sensitive way to evaluate errors in the data or to validate a weighting scheme is through a $\delta(R)$ plot (Abrahams & Keve, 1971; Howell & Smith, 1992). In this technique, $\delta(\text{Real}) = (F_o - F_c)/\sigma(F_o)$ is plotted against δ(expected), assuming a normal distribution of errors. Although a $\delta(R)$ plot should be linear with a slope of unity and an intercept of zero, the slope is usually greater than unity since the error of each observation is typically underestimated.

Nearly all macromolecular refinements suffer from problems not encountered in small-molecule refinements. Atomic resolution data are seldom available, H-atom contributions are frequently ignored, and there are only a few examples in the Protein Data Bank (Bernstein *et al.*, 1977) for which all atoms in the unit cell have been included in the model. The introduction of restraints, such as employed in *PROLSQ* (Hendrickson & Konnert, 1980), *SHELXL*93 (Sheldrick, 1993), or *X-PLOR* (Brünger, Kuriyan & Karplus, 1987), permits a protein structure to be refined to a geometrically reasonable solution even in the absence of atomic resolution data. However, the inability to model the entire contents of the unit cell, particularly the bulk solvent, results in large discrepancies in ΔF in the lower resolution data, and these discrepancies are not necessarily because of errors in the geometry of the model but rather because of the incompleteness of the model.

Numerous weighting schemes are available in the various restrained least-squares refinement programs used to refine macromolecular structures. While some of these schemes may be sensible under a given set of circumstances, a unit weighting scheme, which is the default in *X-PLOR*, will downweight the contribution of the higher resolution data and is inappropriate in the

ISSN 0907-4449

latter stages of refinement. Reported here are several schemes that are shown to be superior to either unit weights or experimental weights.

2. Experimental

2.1. *Insulin dimer*

Test calculations were carried out on an insulin dimer, complexed with *p*-hydroxybenzamide (Smith, Ciszak & Pangborn, 1996) that had been refined to convergence. The final model consisted of 807 protein atoms, a total of 136 water molecules as both ordered and disordered atoms (accounting for approximately half of the water in the crystal), two *p*-hydroxybenzamide molecules, two zinc ions and one chloride ion; the positions of protein protons were calculated on the basis of idealized geometry and refined. All H atoms were subsequently excluded from the above model, and a series of refinements were performed using *SHELXL*93 (Sheldrick, 1993). *PROFFT* (Hendrickson & Konnert, 1980; Finzel, 1987) and *X-PLOR* (Brünger *et al.*, 1987), refining both positional and thermal parameters and with a variety of weighting schemes. Data between the resolution limits of 8.0 and 1.4 Å and with $F_o \geq 2\sigma(F_o)$ were included in the refinements. Since it has been suggested that the optimal value of WA should be two to three times smaller than that required to balance the energy and X-ray gradients (Brünger, 1997), one-third of the value of WA from the 'check' protocol was used for the *X-PLOR* simulated-annealing refinements. In the original version of *PROLSQ* (Hendrickson & Konnert, 1980), an empirical weighting scheme as a function of $\sin\theta/\lambda$ is provided [σ(applied) $= A + B(\sin\theta/\lambda - 1/6)$]. The expression for σ(applied) is a simple straight line, referred to in this paper as a one-line weighting scheme. While this scheme is sufficient for data at moderate resolution, a single function cannot adequately describe the distribution of ΔF for higher resolution data. *PROFFT* was therefore modified to incorporate a two-line empirical weighting scheme (Fig. 1) in which two straight-line segments for two different ranges of $\sin\theta/\lambda$ are used to model the distribution of ΔF. The weighting schemes employed, the r.m.s. deviations of the bond distances from that of the target values, the residuals, and the weighted r.m.s. errors [r.m.s. Err $= (\sum w\Delta F^2/n)^{1/2}$] from the refinements are given in Table 1. The one-line weighting scheme was also tested using conjugate-gradient refinement in *X-PLOR*.

2.2. *Scorpion toxin*

In a second series of test calculations, simulated-annealing and thermal-parameter refinement using unit and two-line weights were performed with *X-PLOR* on a partially refined structure of a 64-residue scorpion toxin (Housset, Habersetzer-Rochat, Astier & Fontecilla-Camps, 1994) This structure had originally been reported at 1.3 Å resolution, but was recently solved *ab initio* at 0.964 Å resolution in our laboratories using the *SnB* program (DeTitta, Weeks, Thuman, Miller &

Table 1. *Refinements and overall statistics for the insulin dimer*

Code	Method and weighting scheme	Residual	R.m.s. Err*	R.m.s.d. bond (Å)†
U	*PROFFT* refinement Unit weights $w = 1.0$	0.153	1.332	0.014
S	*PROFFT* refinement Experimental σ's $w = 1/\sigma^2(F_o)$	0.159	1.235	0.014
2L	*PROFFT* refinement Two-line weighting scheme $\sin\theta/\lambda \leq 0.284$: $w = 1/[A + B(\sin\theta/\lambda = 1/6)]^2$ $\sin\theta/\lambda > 0.284$: $w = 1/[C + D(\sin\theta/\lambda = 1/6)]^2$	0.154	1.538	0.015
F2	*SHELXL*93 refinement (F^2)‡ Default weights $w = 1.0/[\sigma^2(F_o^2) + (aP)^2 + bP]$ where $P = (F_o^2 + 2F_c^2)/3$	0.158 (0.273)	1.365 (1.122)	0.018
CG	*X-PLOR* conjugate-gradient refinement Default weights $w = 1.0$	0.177	3.285	0.007
SA	*X-PLOR* simulated-annealing refinement Default weights $w = 1.0$	0.183	3.371	0.006

* R.m.s. Err $= (\sum w\Delta F^2/n)^{1/2}$. † R.m.s.d. bond = r.m.s. deviation of bond distances from target values. ‡ Statistics in parentheses are compiled on the basis of F^2.

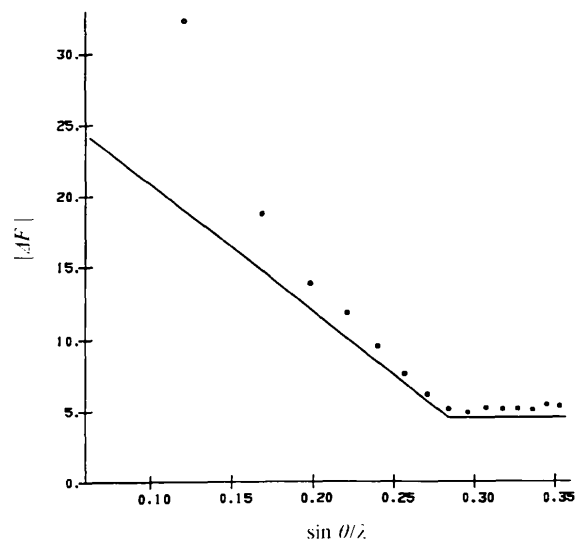


Fig. 1. A plot of $|\Delta F|$ for the insulin dimer in 15 equal volume shells *versus* $\sin\theta/\lambda$ in the resolution range 8.0–1.4 Å. The two-line weighting scheme is also illustrated: the first line is applicable between 0.0625 and 0.284 Å^{-1} in $\sin\theta/\lambda$ and the constants A and B were 14.91 and −88.63, respectively; the constants C and D for the second line were 4.5 and 0.0 and were applied to data between 0.284 and 0.357 Å^{-1}.

Table 2. *Residuals, weighted r.m.s. errors and r.m.s. deviations of bonds and angles from target values for the 1.0 Å scorpion toxin refinements*

	Unit weights	Two-line weights
Residual (working data)	0.202	0.223
Residual (test data)	0.222	0.241
ΔR (test − working data)	0.020	0.018
R.m.s. Err* (working data)	0.284	0.250
R.m.s. Err (test data)	0.326	0.272
Δ r.m.s. Err (test − working data)	0.042	0.022
R.m.s.d. bonds (Å)	0.0067	0.0054
R.m.s.d. angles (°)	1.399	1.348

*R.m.s. Err $= (\sum w\Delta F^2/n)^{1/2}$.

Hauptman, 1994). The data used in the refinements were restricted to those reflections with $F_o \geq 2\sigma(F_o)$ and a resolution between the limits of 8.0 and 1.0 Å. The model consisted of 487 protein atoms, 71 water molecules, and alternate conformations of two cysteine residues, for a total of 570 atoms. The value of WA from the check protocol was used without modification. For these two test cases, a free R value (Brünger, 1992*b*) was calculated using the same test reflections (approximately 10%) for both refinements. Residuals, free R values, and free and r.m.s. weighted errors are listed in Table 2 along with the differences between the figure of merit and their 'free' counterpart.

As most protein crystals do not diffract to a resolution of 1.0 Å or better, additional refinements were performed using data truncated to 2.0 Å, a resolution more typical for an average protein. A minimal model was used for these refinements, excluding water molecules and alternate conformations of the two cysteine residues, and the value of WA, the constant required to balance the X-ray gradient to that of the energy gradient, was multiplied by a constant which varied from 0.2 to 1.5 in increments of 0.1. One-line weights were substituted for the two-line scheme since the distribution of ΔF as a function of $\sin\theta/\lambda$ approximates a straight line at 2 Å resolution.

3. Results

3.1. *Insulin dimer*

The overall residuals, the r.m.s. weighted errors, and the r.m.s. deviations of bond distances from the target values are listed for each of the six refinements in Table 1 and, with the exception of the two *X-PLOR* refinements, do not show a wide variation. However, these overall figures of merit are quite misleading. Average values of the r.m.s. weighted error plotted against $\sin\theta/\lambda$ in equal volume shells and $\delta(R)$ plots for each of the six refinements are illustrated in Fig. 2 and 3, respectively. Fig. 2 clearly shows that several of the refinements do not result in an even distribution of the r.m.s. weighted error as a function of $\sin\theta/\lambda$ and this is particularly true for the two *X-PLOR* refinements. However, a relatively even distribution is observed for the default weighting scheme in *SHELXL*93 and for the two-line weighting scheme in *PROFFT*. These results are reflected in the $\delta(R)$ plots shown in Fig. 3. For the construction of these and subsequent plots, $\delta(\text{Real}) = (F_o - F_c)/\sigma(\text{applied})$ where $\sigma(\text{applied})$ was the reciprocal square root of the weight actually used in the refinement, $\sigma(\text{applied}) = 1/w^{1/2}$. While the *SHELXL*93 and two-line *PROFFT* refinements produce linear plots, the other four plots are decidedly non-linear, suggesting that the weighting schemes are poor, particularly for the *X-PLOR* refinements. It should be noted that the r.m.s. deviations of the bond distances from the target values from the *X-PLOR* refinements are less than half that of the other four refinements, reflecting the reduced value of WA. The r.m.s. deviations of the bond distances from the target values for the *PROFFT* and *SHELXL*93 refinements vary from 0.014 to 0.018 Å, suggesting that the ratio of the restraints for the geometrical to the X-ray terms are comparable for these four refinements.

In the manual for *X-PLOR*, Version 3.1, §12.5.4 (Brünger, 1992*a*) a procedure is described for incorporating a one-line empirical weighting scheme [$\sigma(\text{applied}) = A + B(\sin\theta/\lambda - 1/6)$; $w = 1/\sigma^2(\text{applied})$] as originally described by Hendrickson & Konnert (1980). Appropriate values for A and B were chosen to construct a straight line with the same slope and intercept as the plot of mean ΔF *versus* $\sin\theta/\lambda$ and a conjugate-gradient refinement was performed using *X-PLOR*. The r.m.s. weighted error and the $\delta(R)$ plot, illustrated in Figs. 4 and 5, show marked improvement over that obtained from a unit weighting scheme (Fig. 3, CG) and are comparable to those of the two-line

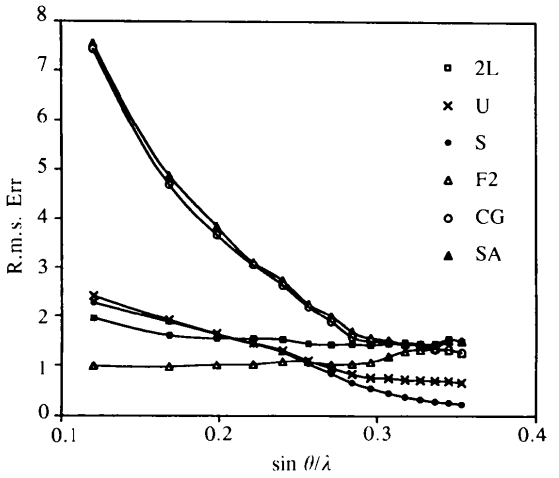


Fig. 2. A plot of the weighted r.m.s. error (r.m.s. Err) for the six insulin dimer refinements. Codes for the refinements are given in Table 1.

scheme in *PROFFT* (Fig. 3, 2L) and the default weighting scheme in *SHELXL*93 (Fig. 3, F2).

3.2. *Scorpion toxin*

A comparison of the residuals and r.m.s. weighted errors for these two refinements might at first suggest that the unit-weighted refinement is marginally better. However, plots of these figures of merit (Figs. 6 and 7) as a function of $\sin\theta/\lambda$ again show that this is not in fact the case. Fig. 6 shows that at low resolution the residual obtained with the two-line weights increases significantly while at higher resolution, the two-line weighted residual is smaller than that of the unit-weighted residual. Fig. 7 clearly shows why this is the case. The distribution of the r.m.s. weighted error is relatively constant for the two-line weighted refinement, indicating that all data are making an even contribution to the function minimized. In contrast, the r.m.s. weighted error for the unit-weighted refinement falls off uniformly from approximately 0.73 at low resolution to 0.14 at higher resolution, because of the fact that the low-resolution data is dominating the refinement. Thus, the observed distribution of the residuals should not be unexpected. It is also of interest to compare the sums of the function minimized $[\sum w(F_o - F_c)^2]$ as a function of resolution which again explains the distribution of the residuals. In the resolution ranges 8.0–2.44 Å and 1.02–1.00 Å (the first and last of 15 equal volume shells in $\sin\theta/\lambda$), the sums $\sum w(F_o - F_c)^2$ for the unit-weighted refinement differ by a factor of 38. In contrast, a factor of only 1.8 is calculated for the two-line weighted refinement. The effect of downweighting the higher resolution data also affects the phases, as mean and r.m.s. phase differences between the two refinement protocols were calculated to be 11.8 and 24.1°; not surprisingly, the largest differences are observed at the highest and lowest resolution limits.

Illustrated in Fig. 8 are the $\delta(R)$ plots for the two refinements. The slope and intercept of the plot from the unit-weighted refinement is calculated to be 0.269 and 0.015, respectively, and the plot is non-linear suggesting that the weights are not optimal. While the slope and intercept of 0.251 and 0.022, respectively, for the two-line refinement are comparable to the values from the unit-weighted refinement, the linearity of this plot indicates that the weighting scheme reflects the distribution of errors expected from a normal distribution better.

It should also be noted (Table 2) that the difference between the residuals for the working data and the test data (free R) for the unit-weighted refinement is 0.020 while the difference obtained from the two-line weights is 0.018. Although the difference in these two values is

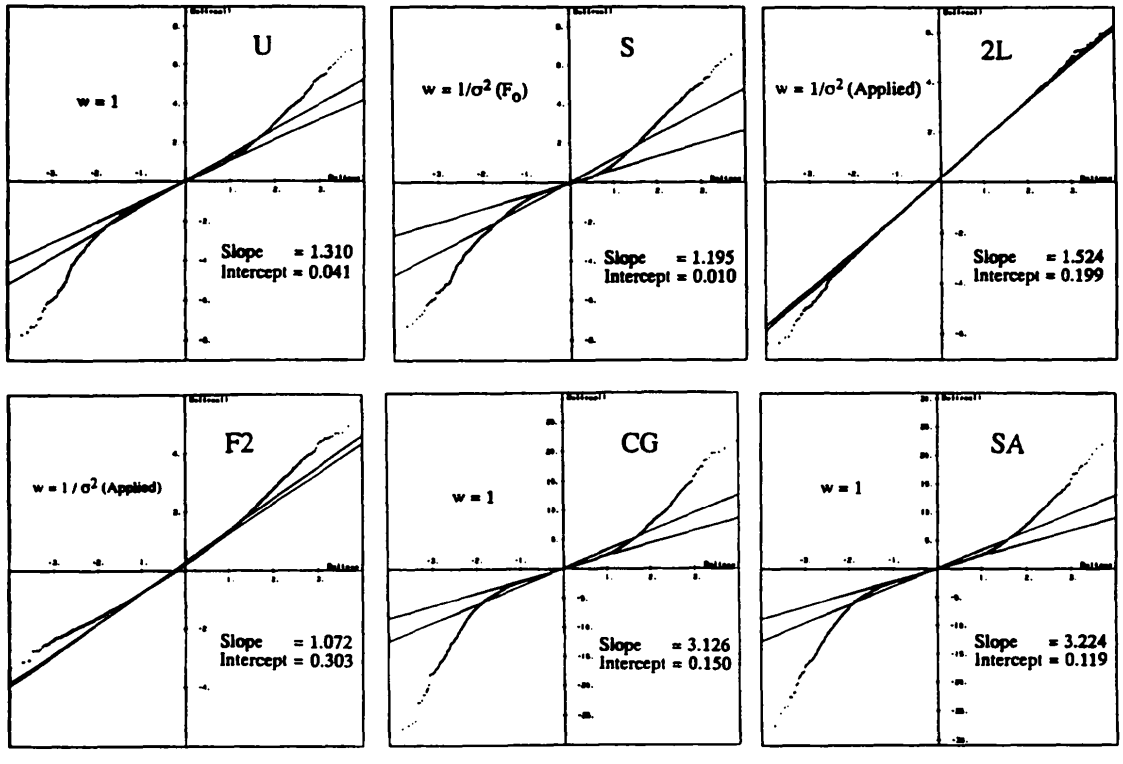


Fig. 3. $\delta(R)$ plots for the six insulin dimer refinements. Codes for the refinements are given in Table 1. σ(applied) for the 2L plot is: $\sin\theta/\lambda \le 0.284$, $\sigma(\text{applied}) = 14.91 - 88.63\ (\sin\theta/\lambda - 1/6)$; $\sin\theta/\lambda > 0.284$, $\sigma(\text{applied}) = 4.5$. σ(applied), for the $F2$ plot: $\sigma(\text{applied}) = 1.0/[\sigma^2(F_o^2) + (0.383p)^2 + 371.96p]^{1/2}$ where $p = (F_o^2 + 2F_c^2)/3.0$.

small, these differences do suggest that the two-line weighting scheme is superior to unit weights. In the same way that a free R is defined, it is also possible to define a free weighted r.m.s. error [r.m.s. Err $=(\sum w\Delta F^2/n)^{1/2}$ for the test data only]. In Table 2, the weighted r.m.s. error for the test data increases by 0.042 as compared to the working data for unit weights, but an increase of only 0.022 is noted for the two-line weights, a difference that does appear to be significant.

The scorpion toxin data, truncated to 2.0 Å and refining the minimal model with unit and one-line weights, behaved in a similar fashion to that of the complete set of data. Residuals, free R values, the weighted r.m.s. error and the free weighted r.m.s. error, obtained by applying the value of WA as calculated by the check protocol, are given in Table 3. The differences in the residuals and free R are probably insignificant, and somewhat larger differences are noted for the free and r.m.s. weighted errors. Average residuals and free R values, in 15 equal volume shells in $\sin\theta/\lambda$, behave in a similar fashion as noted in Fig. 6. The sums of $\sum w(F_o - F_c)$ for the first and last ranges in 15 equal shells of $\sin\theta/\lambda$ differ by a factor of 5.3 for unit weights but only 1.2 for the one-line weights. While the $\delta(R)$ plot from the unit-weighted refinement was not as sigmoidal as was observed for the

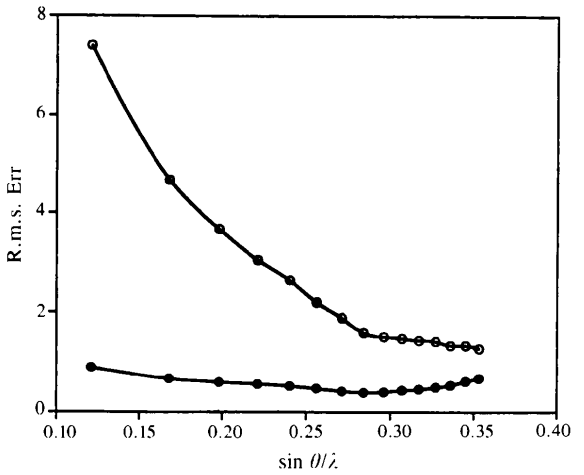


Fig. 4. Plot of the weighted r.m.s. error in 15 equal volume shells of $\sin\theta/\lambda$ for a unit-weighted (○) and one-line weighted (●) conjugate-gradient refinement in *X-PLOR* for the insulin dimer.

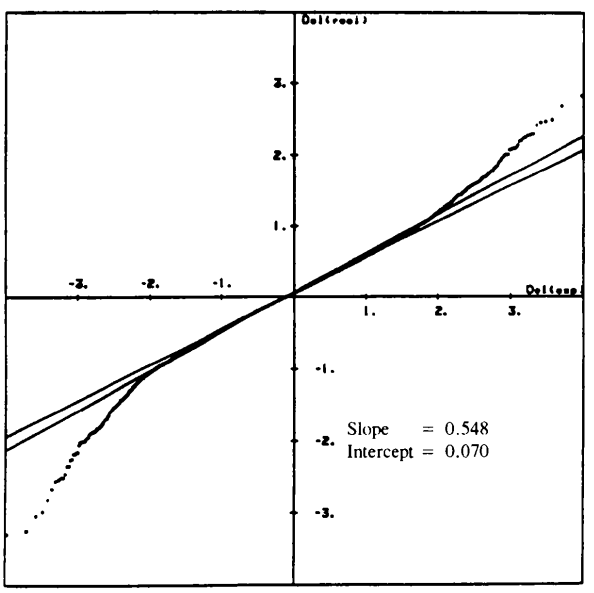


Fig. 5. $\delta(R)$ plot for the one-line weighted conjugate-gradient refinement in *X-PLOR* for the insulin dimer.

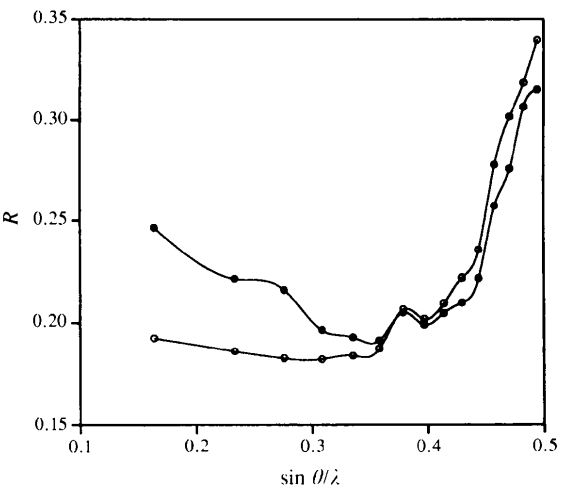


Fig. 6. A comparison of the residual in 15 equal volume shells of $\sin\theta/\lambda$ for the scorpion toxin simulated-annealing refinement at 1.0 Å resolution using unit weights (○) and two-line weights (●) in *X-PLOR*.

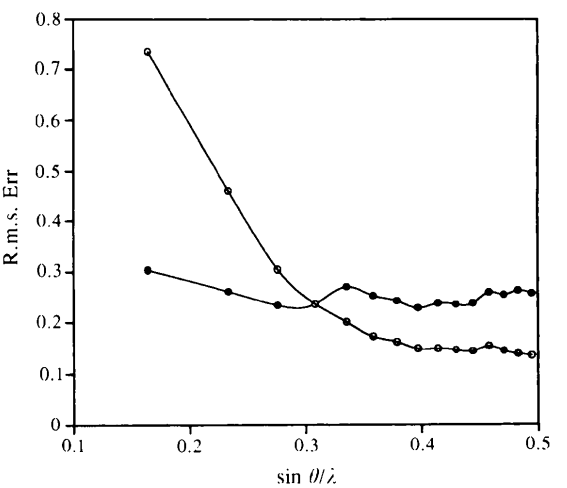


Fig. 7. A plot of the weighted r.m.s. error in 15 equal volume shells of $\sin\theta/\lambda$ for the scorpion toxin simulated-annealing refinement at 1.0 Å resolution using unit weights (○) and two-line weights (●) in *X-PLOR*.

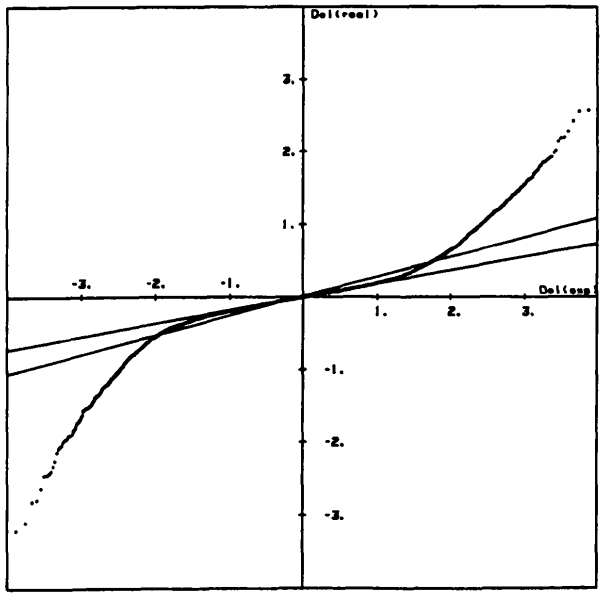
1.0 Å refinement, it was not as linear as the $\delta(R)$ plot from the one-line weighted refinement.

Fig. 9 illustrates the effect of varying WA for both the unit and one-line weighted refinements. With the exception of the smallest value of WA, both the residual and free R value exhibit minimal variations, in contrast to the behavior noted by Brünger (1997). While the insensitivity of the residuals to the contribution of X-ray gradient may be peculiar to this structure, it is also possible that the penicillopepsin crystal structure

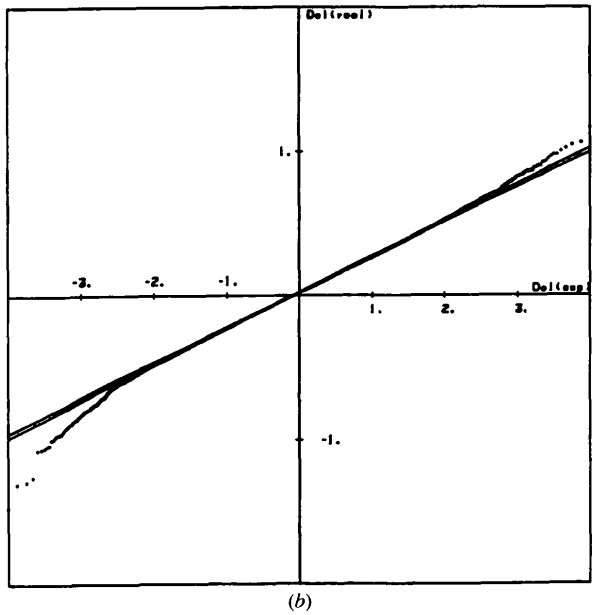
Table 3. *Residuals and weighted r.m.s. errors from the scorpion toxin refinements at 2.0Å resolution*

	Unit weights	One-line weights
Residual (working data)	0.243	0.243
Residual (test data)	0.279	0.278
ΔR (test − working data)	0.036	0.035
R.m.s. Err* (working data)	1.018	0.836
R.m.s. Err (test data)	1.201	1.013
Δ r.m.s. Err (test − working data)	0.183	0.177

* R.m.s. Err = $(\sum w\Delta F^2/n)^{1/2}$.



(*a*)



(*b*)

Fig. 8. $\delta(R)$ plots for the (*a*) unit weighted and (*b*) two-line weighted simulated-annealing refinement at 1.0 Å resolution in *X-PLOR* for the scorpion toxin structure.

(Brünger, 1997) is particularly sensitive to the applied value of WA. However, a plot of the r.m.s. deviation of the bond distances from their target values as a function of the fraction of applied WA shown in Fig. 10, clearly shows a direct relationship, and for every value of WA, the r.m.s. deviations of the bond distances are smaller for the one-line weighting scheme.

Finally, the statistics discussed above strongly suggest that the one- or two-line weighting scheme is superior to unit weights, but the question arises as to whether the resulting model is superior. The stereochemistry of the main- and side-chain atoms obtained from both 1.0 Å refinements was evaluated with *PROCHECK* (Laskowski, MacArthur, Moss & Thornton, 1993), and the resulting G factors are

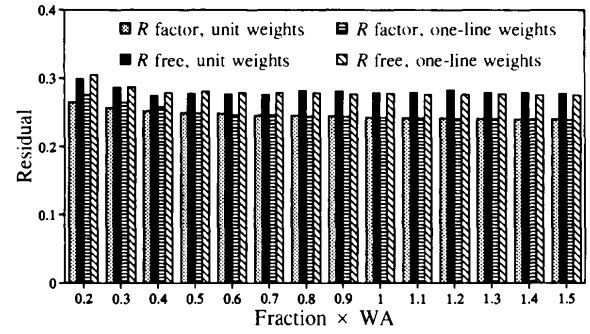


Fig. 9. Residual and free R value obtained by applying a fraction of WA in the simulated-annealing procedure for unit and one-line weights for the scorpion toxin data at 2.0 Å resolution.

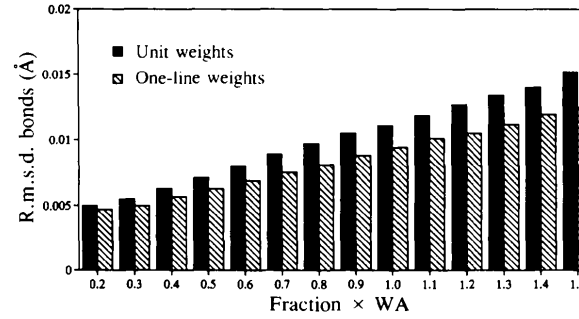


Fig. 10. The effect of varying WA on the r.m.s. devaition of bond distances from their target values for unit and one-line weights for the scorpion toxin data at 2.0 Å resolution.

Table 4. *G factors for the scorpion toxin structures obtained from the unit and two-line weighted refinements at 1.0Å resolution*

	Unit weights	Two-line weights
Dihedral angles		
φ–ψ distribution	−0.21	−0.16
χ^1–χ^2 distribution	0.20	0.24
χ^1 only	−0.19	−0.08
χ^3 and χ^4	0.13	0.47
ω	0.42	0.44
Average	0.11	0.18
Main-chain covalent forces		
Main-chain bond lengths (Å)	0.63	0.66
Main-chain bond angles (°)	0.35	0.38
Average	0.47	0.50
Overall average	0.26	0.31

tabulated in Table 4. In every category, the *G* factors are larger and, therefore, better for the structure refined with two-line weights as compared to unit weights. The r.m.s. deviations of bond distances and angles from the target values are listed in Table 2 for the refinement at 1.0 Å resolution and r.m.s. deviations of bond distances are illustrated in Fig. 10 for the 2 Å refinement. Without exception, the one- and two-line weighting schemes result in the smaller deviations as compared with that of the unit weighting scheme. The differences in the *G* factors and the bond distances and angles clearly show that the model has been improved, validating the use of a one- or two-line empirical weighting scheme.

4. Discussion

The behavior of the residuals and the weighted r.m.s. error at low resolution can be explained on the basis of an incomplete model. The large values of the average ΔF do not result from errors in positional and thermal parameters, but rather from the incompleteness of the model because of the omission of bulk solvent, disordered solvent and protein atoms, and H atoms, all of which make contributions primarily to the low-angle data. Since ΔF is much larger for the low-angle data as compared with the higher resolution data, the use of a unit weighting scheme in the refinement maximizes the effect of these largest differences and at the same time minimizes the effect of the higher angle data on the refinement. For example, the lowest resolution shell of the 15 equal volume shells accounts for only 7.4% of the data, but 48% of the total $\sum w(F_o - F_c)^2$ when unit weights are applied. In contrast, this shell is only 11% of the total $\sum w(F_o - F_c)^2$ when the two-line weights are applied.

A simple plot of the average ΔF *versus* $\sin\theta/\lambda$ in equal volume shells can be used as a guide to choose values for A and B for a one-line weighting scheme. If one considers σ(applied) to be equivalent to ΔF, then it is only necessary to choose A and B such that the resulting line models the distribution of the mean ΔF as a function of $\sin\theta/\lambda$. If the data warrant the use of a two-line scheme, as indicated by a break in the distribution of ΔF as shown in Fig. 1, then following the choice of $\sin\theta/\lambda$ to define the limits of the two lines, C and D are chosen in a manner similar to that of A and B.

The two examples given above are both small proteins that diffract to high resolution. However, these weighting schemes are just as applicable to larger proteins at lower resolution. A one-line scheme was used in the simulated-annealing refinement using *X-PLOR* at 2.0 Å resolution for duck δ crystallin, which contains over 1800 residues in the asymmetric unit. A comparison of these results with those obtained from a unit-weighted refinement was similar to the results from the insulin dimer and the scorpion toxin refinements, namely a linear as opposed to sigmoidal $\delta(R)$ plot and an even distribution of the weighted r.m.s. error as a function of resolution (Howell & Turner, 1995).

Crystallographers expend considerable effort in growing macromolecular crystals and in measuring data in their laboratories or at a synchrotron source in order to obtain high-quality data to the highest possible resolution. However, the choice of an inappropriate weighting scheme in the refinement protocol reduces the effect the high-angle data has upon the refinement, and thus, to some extent, negates the effort expended in obtaining this data. Care should also be taken to routinely examine the distribution of the weighted r.m.s. error as a function of $\sin\theta/\lambda$ to insure that all data are contributing equally to the function minimized. While the free R value and the free weighted r.m.s. error may be able to provide additional information regarding the validity of the weighting scheme, a linear $\delta(R)$ plot is indicative of a good refinement.

The author wishes to thank Drs B. M. Burkhart and E. Ciszak for their assistance in performing the refinements, to Drs P. Lynne Howell and Robert H. Blessing for many helpful discussions, to Drs J. C. Fontecilla-Camps and D. Housset for providing the scorpion toxin data, and to the referees for helpful suggestions. This research was supported by NIH grant GM46733.

References

Abrahams, S. C. & Keve, E. T. (1971). *Acta Cryst.* A**27**, 157–165.

Bernstein, F. C., Koetzle, T. F., Williams, G. J. B., Meyer E. F. Jr, Brice, M. D., Rodgers, J. R., Kennard, O., Shimanouchi, T. & Tasumi, M. (1977). *J. Mol. Biol.* **112**, 535–542.

Blessing, R. H. (1987). *Crystallogr. Rev.* **1**, 3–58.

Brünger, A. T. (1992*a*) *X-PLOR Version 3.1: A system for X-ray Crystallography and NMR*. Yale University Press, New Haven, CT, USA.

Brünger, A. T. (1992*b*). *Nature (London)*, **355**. 472–474.

Brünger, A. T. (1997). *Methods Enzymol.* In the press.

Brünger, A. T., Kuriyan, J. & Karplus, K. (1987). *Science*, **235**, 458–460.

DeTitta, G. T., Weeks, C. M., Thuman, P., Miller, R. & Hauptman, H. A. (1994). *Acta Cryst.* A**50**, 203–210.

Finzel, B. C. (1987). *J. Appl. Cryst.* **20**, 53–55.

Hendrickson, W. A. & Konnert, J. H. (1980). In *Computing in Crystallography*, edited by R. Diamond, S. Ramaseshan & K. Venkatesan, pp. 13.01–13.25. Bangalore: Indian Academy of Sciences.

Housset, D., Habersetzer-Rochat, C., Astier, J.-P. & Fontecilla-Camps, J. C. (1994). *J. Mol. Biol.* **238**, 88–103.

Howell, P.L., Pangborn, W. A., Marshall, G. R., Zabrocki, J. & Smith, G. D. (1995). *Acta Cryst.* C**51**, 2575–2579.

Howell, P. L. & Smith, G. D. (1992). *J. Appl. Cryst.* **25**, 81–86.

Howell, P. L. & Turner, M. A. (1995). Personal communication.

Laskowski, R. A., MacArthur, M. W., Moss, D. S. & Thornton, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

Sheldrick, G. M. (1993). *SHELXL93. A Program for the Refinement of Crystal Structures*. University of Göttingen, Germany.

Smith, G. D., Ciszak, E. & Pangborn, W. (1996). *Protein Sci.* **5**, 1502–1511.