Acta Cryst. (1992). A**48**, 42-45

Some Different Strategies of Least-Squares Refinement of a Molecule

BY GILLIAN W. HARRIS* AND DAVID S. MOSS

Department of Crystallography, Birkbeck College, Malet Street, London WC1E 7HX, England

(*Received* 4 *October* 1988; *accepted* 15 *July* 1991)

Abstract

Refinements of a macromolecule (ribonuclease-A) based on structure amplitudes, $|F|$, and structure amplitude squares, $|F|^2$, were carried out and the results compared. Although the conventional R values are higher for the $|F|^2$ refinement, positional parameters from both types of refinement were not significantly different. However, the mean-square displacements from $|F|^2$ refinements were systematically higher than for those using $|F|$. Various resolution windows and weighting schemes were employed during the work. Electron density maps were examined for $|F|^2$ refinements and were very similar to those using $|F|$ in spite of a conventional R factor of 0·29 using all 1·4 Å data. While $|F|^2$ refinements may be formally more correct than $|F|$ refinements, there is little evidence that $|F|^2$ refinement is superior provided that a reasonable weighting strategy is adopted.

Introduction

Since the method of least-squares refinement was introduced into X-ray crystallography by Hughes (1941), crystal structure refinements have almost always used structure amplitudes $|F|$ as the 'observed' quantities. The most commonly used agreement index or R factor has also been based on structure amplitudes (Booth, 1947) rather than intensities ($|F|^2$).

In kinematic scattering theory the intensity is proportional to the square of the structure amplitude (Rees, 1977). The use of structure amplitude squares ($|F|^2$) in place of structure amplitudes ($|F|$) in refinements has been advocated on the grounds that refinements based on $|F|$ can give false minima since the first derivative of the function being minimized is not continuous (Rollett, McKinlay & Haigh, 1976). Most arguments, however, have been based on statistical theory. A statistical bias towards a too low $|F|$ is introduced if $|F|$ is estimated as the square root of intensity (Ibers & Hamilton, 1964; Rees, 1977; Wilson, 1979). Wilson (1973, 1976*b*) has shown that only parameters estimated by minimizing an unweighted residual based on intensities are free from statistical bias. Hence there is a consensus of opinion, at least in theory, that refinements should be based on intensities rather than on structure amplitudes.

However, in practice the majority of refinements continue to be based on structure amplitudes. It might be wrong, but it appears to work! In some instances, small-molecule refinements have been carried out using both intensities and structure amplitudes (Freer & Kraut, 1965; Seiler, Schweizer & Dunitz, 1984) to see what effect this has on the structure. Seiler, Schweizer & Dunitz (1984) also investigated the effect of refining with and without zero reflections (*i.e.* those reflections for which the measured background count exceeds the peak count). However, most differences, especially in the atomic coordinates, were found to be insignificant. Any investigation which focuses on the function to be minimized in crystal structure refinement must also consider the question of weighting. In refinements, zero reflections are often given zero weight, but Hirshfeld & Rabinovich (1973) have shown that in a least-squares refinement, preferably based on $|F|^2$, the arbitrary exclusion of unobserved reflections biases the input data and may cause systematic error in the refined parameters. Wilson (1976*a*) has discussed bias due to the use of non-unit weights in least-squares refinements. The effect of omitting zero reflections and the use of unit weights was also investigated.

Current work on diffuse X-ray scattering of macromolecular crystals at our laboratory involves the measurement of integrated X-ray intensities between reciprocal-lattice positions and its use in refinement along with Bragg diffraction data. Since there appeared to be no reason for applying a square-root transformation to the diffuse intensity data, it was decided to investigate the effect of using intensities in the refinement of the enzyme ribonuclease-A using, initially, the Bragg reflections only. In this paper we report these $|F|^2$ refinements and compare some results with refinements based on structure amplitudes.

Seiler, Schweizer & Dunitz (1984) showed that, for small-molecule crystal structures where the model is overdetermined, refinements based on $|F|$ or $|F|^2$, with or without exclusion of weak reflections, gave very similar atomic parameters. In contrast to the work reported here, these authors had diffraction data available to beyond 0·5 Å resolution.

* Present address: Department of Protein Engineering, AFRC Institute of Food Research, Reading Laboratory, Shinfield, Reading RG2 9AT, England.

0108-7673/92/010042-04$03.00

However, two problems are encountered in macromolecular refinement which do not regularly occur in small-molecule refinements. The observation-to-parameter ratio is poor. Thus, at 2 Å resolution the ratio of observation to positional parameters may be less than three. A further problem is the extensive static and dynamic disorder usually associated with macromolecules which is often poorly modelled in the conventional structure-factor formula. With these considerations in mind we decided to investigate the effect of using $|F|^2$ rather than $|F|$ in macromolecular refinement.

The problem of weighting diffraction data is more acute in macromolecular refinements where the structure-factor model is usually poorer than in crystal structures of small molecules. Consequently, standard deviations from counting statistics are a poor basis for weighting (Hendrickson, 1980). The work reported in this paper has used unit weights and also a weighting scheme. The appropriateness of weights has been assessed by analysis of variance, which ensures that undue weight is not being thrust onto any subset of the data.

Method

Ribonuclease-A was chosen for this study as the structure has been well refined by crystallographic methods independently in two laboratories (Borkakoti, Moss & Palmer, 1982; Wlodawer & Sjölin, 1983). A comparison between the two structures (Wlodawer, Borkakoti, Moss & Howlin, 1986) found the errors of the main-chain atomic coordinates of the two structures to be about 0.08 Å. Such an estimate of the precision of the model is not readily available from a single macromolecular refinement because the normal matrix is not easily inverted to give standard deviations in the conventional way.

The refinements of ribonuclease-A used the 1.45 Å X-ray data of Borkakoti, Moss, Stanford & Palmer (1984). Unless otherwise stated, all measured reflections were used in the refinements and all the statistics are for all reflections. Over 15% of the reflections were weak ($<2\sigma$) and at 1.45 Å resolutions about 35% of them were weak in the outermost shell. The structure refinements based on structure amplitudes ($|F|$) were carried out using the least-squares-refinement program *RESTRAIN* (Haneef, Moss, Stanford & Borkakoti, 1985). The necessary program modifications were made to *RESTRAIN* for the refinements based on structure amplitudes squared ($|F|^2$). The starting coordinates used were those of Borkakoti, Moss, Stanford & Palmer (1984). Refinements were carried out with three different high-resolution cut-offs (1.4, 1.6, 1.8 Å) to investigate the effect of varying the resolution window. A low-resolution cut-off of 8 Å was applied in all work so as to reduce the effect of scattering from the disordered water in the crystals. All the results quoted are after 20 cycles of refinement.

Table 1. *R values for the $|F|$ and $|F|^2$ refinements of ribonuclease-A*

(*A*) $\|F\|$ refinements					
High-resolution cut-off (Å)	1.4[e]	1.4[f]	1.6[g]	1.8[h]	1.4[f,i]
R[a]	0.194	0.223	0.189	0.163	0.221
RDASH[b]	0.230	0.261	0.227	0.200	0.204
(*B*) $\|F\|^2$ refinements					
High-resolution cut-off (Å)	1.4[e]	1.4[f]	1.6[g]	1.8[h]	1.4[f,j]
R[a]	0.244	0.290	0.253	0.216	0.263
$R2$[c]	0.367	0.383	0.377	0.352	0.290
$RW2$[d]	0.276	0.374	0.292	0.222	0.094

Notes: (*a*) $R = \sum |F_o| - G|F_c| / \sum |F_o|$.
(*b*) RDASH $= [\sum w(|F_o| - G|F_c|)^2 / \sum w|F_o|^2]^{1/2}$.
(*c*) $R2 = \sum |F_o^2 - G'F_c^2| / \sum F_o^2$.
(*d*) $RW2 = \sum w(F_o^2 - G'F_c^2)^2 / \sum wF_o^4$.
(*e*) For 17 565 non-zero reflections.
(*f*) For all 19 098 reflections.
(*g*) 14 207 reflections.
(*h*) 10 931 reflections.
(*i*) Using constant weights, $w = 0.39 \times 10^{-4}$.
(*j*) Using $w = 0.39 \times 10^{-4}/(|F| + 1.0)$.

The R factors employed in the refinements are given in Table 1. The R factors for the $|F|^2$ refinements were chosen from a consideration of the literature on the subject (Lenstra, 1982; Wilson, 1973, 1976*a*, *b*, 1979) and the desire to retain as close an analogy as possible to the $|F|$ refinement.

The purpose of weighting the data in protein refinement is not only to allow for experimental errors but also to allow for errors in the structure-factor model which may be considerable due to significant anisotropic and anharmonic disorder. Two strategies were used for weighting the experimental data. In the first strategy, weights were assigned so that average values of the weighted residuals were approximately constant when analysed in terms of resolution shells or in terms of intensity bins. These weights will be called 'constant-variance weights'. This approach ensures that the course of the refinement is not dominated by any subset of the reflection data. This method was implemented by calculating weights using the following expressions [based on a modification of the formula due to Cruickshank (1965) for $|F|$] for both the $|F|$ and $|F|^2$ refinements:

$|F|$ refinements,

$$w = 0.02/(250 + |F_0|);$$

$|F|^2$ refinements,

$$w = 10^{-5}/(15000 + |F_0|^2 + 0.0001|F_0|^4).$$

The constants in the above expressions were determined by adjustment until the desired constant variance of the weighted residual was attained.

A second weighting strategy which was carried out on the 1.4 Å data was to use constant weights ($w = 0.39 \times 10^{-4}$) for an $|F|$ refinement. Such a strategy is

Table 2. *Coordinate comparison of the $|F|$ and $|F|^2$ refinements of ribonuclease-A*

The table gives r.m.s. deviation and mean deviation in brackets.

High-resolution cut-off (Å)	R.m.s. deviation (mean) (Å) Main-chain only	All atoms
1.4[a]	0.085 (0.076)	0.130 (0.107)
1.4[b]	0.068 (0.061)	0.088 (0.075)
1.4[b,c]	0.094 (0.085)	0.114 (0.098)
1.6	0.080 (0.073)	0.123 (0.102)
1.8	0.073 (0.065)	0.108 (0.087)

Notes: (*a*) For 17 565 non-zero reflections.
(*b*) For all 19 098 reflections.
(*c*) Using constant weights ($w = 0.39 \times 10^{-4}$) for $|F|$ refinements and $w = 0.39 \times 10^{-4}/(|F| + 1.0)$ for $|F|^2$ refinements.

not possible with $|F|^2$ refinement because it throws too much weight on the strong reflections. Instead weights calculated by $w = 0.39 \times 10^{-4}/(|F| + 1.0)$ were employed. This corresponds approximately to the assumption that the variance of $|F|$ is constant.

A coordinate comparison of the structures from $|F|$ and $|F|^2$ refinements were carried out using the program *EULER* (Moss, Howlin & Haneef, 1985). The results are given in Table 2.

Results

(*a*) *R factors*

The final *R* factors for the $|F|$ and $|F|^2$ refinements are given in Table 1. The conventional *R* factors, based on $|F|$, are 0.04 to 0.07 higher for the $|F|^2$ refinements than for the corresponding $|F|$ analyses. For the $|F|^2$ refinements, *R* factors based on $|F|^2$ were also calculated and these were 0.03 to 0.12 higher than the conventional *R*-factor values. The expected feature that *R* factors are lowered by limiting the resolution or omitting zero reflections applies to both types of refinement.

(*b*) *Coordinates*

The statistics from a comparison of the coordinates of the $|F|$ and $|F|^2$ refinements at three resolutions are given in Table 2. For the main-chain atoms only, in all three cases the root-mean square (r.m.s.) deviation between the two structures is less than 0.08 Å. When side-chain atoms are also included, the r.m.s. deviation between the structures does not exceed 0.13 Å. It is notable that excluding the zero reflections or using constant weights both produce less-consistent results in the comparison between the two types of refinement.

The effect of omitting zero reflections was investigated for both the $|F|$ and $|F|^2$ refinements at 1.4 Å using constant-variance weighting. The statistics of this coordinate comparison are given in Table 3, where it can be seen that the $|F|^2$ refinement is more sensitive to the omissions but all the r.m.s. deviations are less than 0.12 Å. Very similar results were obtained when the effect of weighting strategy was investigated and the results are shown in Table 4.

Table 3. *Coordinate comparison of refinements of ribonuclease-A at* 1.4 Å *with and without the* 1533 *zero reflections*

	R.m.s. deviation (mean) (Å) Main-chain only	All atoms
$\|F\|$ refinement	0.055 (0.050)	0.071 (0.060)
$\|F\|^2$ refinement	0.091 (0.082)	0.113 (0.099)

Table 4. *Coordinate comparison of refinements of ribonuclease-A at* 1.4 Å, *using different weighting schemes*

Constant-variance weights *versus* constant weights for $|F|$ refinements and constant-variance weights *versus* $w = 0.39 \times 10^{-4}/(|F| + 1.0)$ for $|F|^2$ refinements. All 19 098 reflections were used.

	R.m.s. deviation (mean) (Å)	
$\|F\|$ refinement	0.042 (0.037)	0.064 (0.050)
$\|F\|^2$ refinement	0.110 (0.097)	0.143 (0.121)

Table 5. *Average values* (Å^2) *of the isotropic mean-square displacement amplitude (m.s.d.a.) for the $|F|$ and $|F|^2$ refinements of ribonuclease-A*

M.s.d.a. $= \sum u_j^2 / N$, where u_j^2 is the mean-square displacement of the *j*th atom and *N* is the number of atoms.
M.s.d.a. values for starting coordinates: main chain 0.169; side chain 0.276; all protein atoms 0.221.

Atom group (number of atoms)	High-resolution cut-off (Å)				
(*A*) $\|F\|$ refinement					
	1.4[a]	1.4[b]	1.6	1.8	1.4[b,c]
Main chain (496)	0.159	0.168	0.165	0.160	0.165
Side chain (455)	0.271	0.277	0.276	0.281	0.278
All protein (951)	0.213	0.220	0.218	0.218	0.219
Overall U[e]	0.139	0.160	0.154	0.138	0.154
(*B*) $\|F\|^2$ refinement					
	1.4[a]	1.4[b]	1.6	1.8	1.4[b,d]
Main chain (496)	0.171	0.198	0.182	0.163	0.205
Side chain (455)	0.281	0.297	0.284	0.272	0.306
All protein (951)	0.244	0.246	0.231	0.215	0.254
Overall U[e]	0.166	0.290	0.206	0.123	0.244

Notes: (*a*) For 17 565 non-zero reflections.
(*b*) For all 19 098 reflections.
(*c*) Constant weights used $w = 0.39 \times 10^{-4}$.
(*d*) Weights used $w = 0.39 \times 10^{-4}/(|F| + 1.0)$.
(*e*) Overall atomic displacement parameter.

(*c*) *Mean-square displacements*

Table 5 shows the effect on overall mean-square atomic displacement amplitudes (U) of using $|F|^2$ rather than $|F|$ in least-squares refinement. The overall U values in the $|F|^2$ refinements are generally higher and very sensitive to the weighting strategy, resolution cut-off and treatment of weak reflections. For both the $|F|^2$ and $|F|$ refinements, the general trend is that lowering the high-resolution cut-off lowers the overall U value.

The average value of the individual isotropic atomic mean-square displacement amplitude (m.s.d.a.) of the protein atoms was investigated. The m.s.d.a. values are higher for the $|F|^2$ than the $|F|$ refinements at the same resolution and the m.s.d.a. values are greater the lower the high-resolution cut-off for both the $|F|$ and $|F|^2$ refinements. At 1.4 Å resolution the m.s.d.a. values are slightly lower for the $|F|$ and $|F|^2$ refinements omitting the 1533 zero reflections. The same is true for the overall U values. (See Table 5.)

(*d*) *Weighting*

The use of $w = 0.39 \times 10^{-4}/(|F| + 1.0)$ for the $|F|^2$ refinement gives lower R values, the weighted R, $RW2$, being significantly lower than constant-variance weights (see Table 1*B*). The individual isotropic mean-square displacement amplitudes obtained are similar (Table 5*B*). However, the $|F|^2$ refinements are more sensitive to weighting strategy. This is reflected in the larger r.m.s. deviation of 0.143 Å between the two structures for all atoms (Table 4).

(*e*) *Electron density maps*

Electron density maps $(2F_o - F_c)$ were calculated from structure factors F_c for the $|F|$ and $|F|^2$ refinements at 1.4 Å and examined on an interactive graphics system (Evans & Sutherland Picture System 2). Differences observed between the electron density maps from the $|F|^2$ refinement and from the $|F|$ refinement were usually small and did not lead us to reinterpret the structure, despite the higher R value.

Concluding remarks

The use of $|F|^2$ instead of $|F|$ in refinements underlines the fact that the conventional R factor is only of limited value in assessing the quality of a refinement. It is surprising to see clean electron density maps at a contour level of 0.5 eA^{-3} with an R factor of 0.29 using all reflections at 1.4 Å resolution.

The coordinate differences between $|F|$ and $|F|^2$ refinements are generally not significant when error estimation is based on comparison of two independently refined structures (Wlodawer *et al.*, 1986). However, $|F|^2$ refinements do produce small but systematic differences in m.s.d.a.'s.

We conclude that, while $|F|^2$ refinements may be formally more correct than $|F|$ refinements, there is little evidence that $|F|^2$ refinement is superior provided that a reasonable weighting strategy is adopted. Table 5, in fact, shows that $|F|$ refinements produce more homogeneous U values than the $|F|^2$ refinements. $|F|^2$ refinements will certainly not be a popular option among crystallographers keen on publishing low R factors!

One of us (GWH) would like to acknowledge the Science and Engineering Research Council for support during this work.

References

BOOTH, A. D. (1947). *Nature (London)*, **160**, 196.
BORKAKOTI, N., MOSS, D. S. & PALMER, R. A. (1982). *Acta Cryst.* B**38**, 2210-2217.
BORKAKOTI, N., MOSS, D. S., STANFORD, M. J. & PALMER, R. A. (1984). *J. Crystallogr. Spectrosc. Res.* **14**, 467-494.
CRUICKSHANK, D. W. J. (1965). In *Computing Methods in Crystallography*, edited by J. S. ROLLETT, pp. 112-116. Oxford: Pergamon Press.
FREER, S. T. & KRAUT, J. (1965). *Acta Cryst.* **19**, 992-1002.
HANEEF, I., MOSS, D. S., STANFORD, M. J. & BORKAKOTI, N. (1985). *Acta Cryst.* A**41**, 426-433.
HENDRICKSON, W. A. (1980). In *Refinement of Protein Structures*, edited by P. A. MACHIN, J. W. CAMPBELL & M. ELDER, pp. 1-8. Warrington: SERC Daresbury Laboratory.
HIRSCHFELD, F. L. & RABINOVICH, D. (1973). *Acta Cryst.* A**29**, 510-513.
HUGHES, E. W. (1941). *J. Am. Chem. Soc.* **63**, 1737-1752.
IBERS, J. A. & HAMILTON, W. C. (1964). *Acta Cryst.* **17**, 781-782.
LENSTRA, A. T. H. (1982). In *Crystallographic Statistics*, edited by S. RAMASESHAN, M. F. RICHARDSON & A. J. C. WILSON, pp. 195-224. Bangalore: Indian Academy of Science.
MOSS, D. S., HOWLIN, B. & HANEEF, I. (1985). *EULER*. Unpublished program. Birkbeck College, London, England.
REES, B. (1977). *Isr. J. Chem.* **16**, 154-158.
ROLLETT, J. S., MCKINLAY, T. G. & HAIGH, N. P. H. (1976). In *Crystallographic Computing Techniques*, edited by F. R. AHMED, pp. 413-419. Copenhagen: Munksgaard.
SEILER, P., SCHWEIZER, W. B. & DUNITZ, J. D. (1984). *Acta Cryst.* B**40**, 319-327.
WILSON, A. J. C. (1973). *Acta Cryst.* B**29**, 1488-1490.
WILSON, A. J. C. (1976*a*). *Acta Cryst.* A**32**, 781-783.
WILSON, A. J. C. (1976*b*). *Acta Cryst.* A**32**, 994-996.
WILSON, A. J. C. (1979). *Acta Cryst.* A**35**, 122-130.
WLODAWER, A., BORKAKOTI, N., MOSS, D. S. & HOWLIN, B. (1986). *Acta Cryst.* B**42**, 379-387.
WLODAWER, A. & SJÖLIN, L. (1983). *Biochemistry*, **22**, 2720-2728.