for any crystallographic orbit that refers to a comprehensive complex of that lattice complex if, in addition, it may be described by the same coordinate triplets as an orbit of the regarded lattice complex.

The authors thank the unknown referee of this paper for his helpful and encouraging comments.

References

BARTH, H.-U. (1980). *Über umfassende Komplexe kubischer Gitterkomplexe.* Diplomarbeit, Univ. Münster.

BURZLAFF, H. & ZIMMERMANN, H. (1974). *Z. Kristallogr.* **139**, 252-269.

ENGEL, P. (1983). *Z. Kristallogr.* **163**, 243-249.

ENGEL, P., MATSUMOTO, T., STEINMANN, G. & WONDRATSCHEK, H. (1984). *Z. Kristallogr.* Suppl. No. 1.

FISCHER, W., BURZLAFF, H., HELLNER, E. & DONNAY, J. D. H. (1973). *Space Groups and Lattice Complexes. Natl Bur. Stand. (US) Monogr.* No. 134. Washington: National Bureau of Standards.

FISCHER, W. & KOCH, E. (1974). *Z. Kristallogr.* **139**, 268-278.

FISCHER, W. & KOCH, E. (1978). *Z. Kristallogr.* **147**, 255-273.

FISCHER, W. & KOCH, E. (1983). *Acta Cryst.* A**39**, 907-915.

HERMANN, C. (1935). *Gitterkomplexe.* In *Internationale Tabellen zur Bestimmung von Kristallstrukturen*, Vol. I. Berlin: Bornträger.

International Tables for Crystallography (1983). Vol. A. Dordrecht, Boston: D. Reidel.

Internationale Tabellen zur Bestimmung von Kristallstrukturen (1935). Vol. I. Berlin: Bornträger.

KOCH, E. (1974). *Z. Kristallogr.* **140**, 75-86.

KOCH, E. & FISCHER, W. (1975). *Acta Cryst.* A**31**, 88-95.

KOCH, E. & FISCHER, W. (1978). *Z. Kristallogr.* **147**, 21-38.

MATSUMOTO, T. & WONDRATSCHEK, H. (1979). *Z. Kristallogr.* **150**, 181-198.

STEINMANN, G. (1984). *Kristallographische Orbits im dreidimensionalen Raum.* Dissertation, Univ. Karlsruhe.

WONDRATSCHEK, H. (1976). *Z. Kristallogr.* **143**, 460-470.

WONDRATSCHEK, H. (1980). *Commun. Math. Chem.* **9**, 121-125.

Acta Cryst. (1985). A**41**, 426-433

Restrained Structure-Factor Least-Squares Refinement of Protein Structures Using a Vector Processing Computer

BY ILYAS HANEEF, DAVID S. MOSS,* MICHAEL J. STANFORD AND NIVEDITA BORKAKOTI

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

(*Received 2 August* 1983; *accepted* 20 *March* 1985)

Abstract

A least-squares refinement program *RESTRAIN* has been developed, which is capable of refining macromolecular structures using structure amplitudes, phases from isomorphous replacement or anomalous scattering and pseudo-energy restraints. In addition to positional parameters and isotropic temperature factors, anisotropic mean-square displacements may be refined either as individual atomic **U** tensors or as **TLS** tensors applied to groups of atoms. Anharmonic effects may be handled by coupling together occupancies to enable the electron density of an atomic group to be distributed over more than one subsite. A novel way of restraining groups of atoms to be planar has been developed that does not require dummy atoms and does not restrain the plane to lie in its current orientation.

Introduction

Techniques for the refinement of macromolecular structures from diffraction data using geometrical restraints are now well established. Before 1976 most refinements of protein structures were undertaken using difference Fourier methods. Several techniques for automating this approach were developed (Diamond, 1971; Freer, Alden, Carter & Kraut, 1975) and real-space refinement has recently been applied to the protein component of a virus (Jones & Liljas, 1984).

Reciprocal-space least-squares refinement techniques followed later and imposed geometrical restraints on the positional parameters in terms of bond lengths, bond angles and non-bonded interactions. Systems that have minimized functions that contain both structure amplitude and restraint terms (Konnert, 1976; Sussman, Holbrook, Church & Kim, 1977; Moss & Morffew, 1982) have been widely used in the refinement of protein and RNA structures (see, for example, Borkakoti, Palmer, Haneef & Moss, 1983; Sielecki, Hendrickson, Broughton, Delbaere, Bryer & James, 1979; Girling, Houston, Schmidt & Amma, 1980). Other systems, which impose the restraints in a separate least-squares or energy-minimization step outside the structure-amplitude refinement (Agarwal, 1978; Jack & Levitt, 1978), have also been successfully employed (Baker, 1980).

* To whom all correspondence should be addressed.

0108-7673/85/050426-08$01.50

A major problem affecting the least-squares refinement of macromolecules is the heavy demand on computer time. Several ways of reducing this problem may be adopted. Firstly, the use of restraints enables the refinement to proceed satisfactorily with a normal matrix, which is diagonal or block-diagonal with respect to the structure-factor derivatives. Secondly, advantage may be taken of the increasing availability of vector/array processors. These are found both as a part of the central processor of a scalar machine such as the Cray-1, Cyber 205 or Norsk 570 or as a peripheral unit such as the floating-point array processors that may be attached to a general-purpose machine such as a VAX 11/780 (Furey, Wang & Sax, 1982). Thirdly, the structure factors and their derivatives may be supplied by a fast Fourier transform algorithm. Programs using this approach have been written by Agarwal (1978) and Jack & Levitt (1978).

The desire to relate structure to function in macromolecular studies often requires that the maximum amount of dynamic information should be derived from a structure refinement. The Bragg reflections from a crystal are determined by the time- and lattice-averaged structure, but information about the correlation of atomic motions is lost in the thermal or disorder diffuse scattering. Any *a priori* assumptions that can be made about these correlations can be employed to restrain atomic displacement parameters and enable more accurate models of these displacements to be used with macromolecular diffraction data of limited resolution. Konnert & Hendrickson (1980) have described a method of anisotropic refinement that is based on the small magnitude of the relative displacements of bonded atoms in the bond direction. The purpose of the present paper is to describe a program *RESTRAIN*, which takes advantage of the vector hardware of the Cray-1 computer and offers various facilities for the modelling of anisotropic and anharmonic atomic displacements.

The function minimized

The refinement of a macromolecular structure is usually characterized by a poor observation-to-parameter ratio. This leads to higher thermal or disorder diffuse scattering and weaker Bragg reflections, which only extend to a limited resolution. The number of structure amplitudes from a protein crystal becomes equal to the number of positional parameters when a resolution between 3·2 and 2·5 Å is attained, depending on the solvent content of the crystal.

A satisfactory refinement of a macromolecule must therefore call upon sources of information other than structure amplitude data otherwise unacceptably large random errors will occur in the refined parameters. Phases derived from isomorphous replacement of anomalous scattering can be employed to restrain the model but due to the large errors usually present in such phase estimates, this strategy may not contribute much to the later stages of refinement. Target values for stereochemical or pseudo-energy restraints are known with higher precision and contribute most significantly to the function minimized. In *RESTRAIN* this function is

$$\begin{aligned} M = &\sum W_f(|F_o| - G|F_c|)^2 + \sum W_p(\varphi_o - \varphi_c)^2 \\ &+ \sum W_d(d_t - d_c)^2 + \sum W_b(b_o - b_{\min})^2 \\ &+ \sum W_v|\mathbf{V}| \quad (\text{for } b_o < b_{\min}), \end{aligned}$$

where W_f = structure-amplitude weighting coefficients, $|F_o|$ = observed structure amplitudes, G = scale factor, $|F_c|$ = calculated structure amplitudes, W_p = phase weighting coefficients, φ_o = observed phases, φ_c = calculated phases, W_d = restrained distance weighting coefficients, d_t = target interatomic distances, d_c = calculated interatomic distances, b_o = observed distance between two non-bonded atoms, $b_{\min}$ = minimum distance allowed for such atoms, W_b = weighting coefficients for non-bonded interactions, $|\mathbf{V}|$ = determinant of the product-moment matrix of a planar group of atoms, W_v = weighting coefficients for planarity restraints.

M may be written as a function of three terms:

$$M = A + B + C, \tag{1}$$

where

$$A = \sum W_f(|F_o| - G|F_c|)^2$$

is the term conventionally found in crystallographic least-squares refinement procedures. The weight W_f given to the individual squared terms is calculated from one of a choice of two formulae. One is a modification of a formula due to Cruickshank (1961):

$$W_f = a(\sin\varphi/\lambda)^b/(c + |F_o| + d|F_o|^2);$$

and the second utilizes the standard deviations $\sigma(|F_o|)$ that are derived from the intensity measurements:

$$W_f = a(\sin\varphi/\lambda)^b/[\sigma^2(|F_o|) + c|F_o|^2].$$

The values of a, b, c and d may be supplied by the program user.

The term B in (1) is given by

$$B = \sum W_p(\varphi_o - \varphi_c)^2,$$

where φ_o is the estimate of the phase from isomorphous and anomalous data, φ_c is the phase calculated from the model. Phase data are weighted by the term

$$W_p = am(180 - |\varphi_o - \varphi_c|)^b,$$

where a and b are coefficients that may be supplied by the user and m is the figure of merit. The phase difference occurring in W_p allows for the cyclic nature of the phase data, which implies that a calculated phase that is 180° from its observed value cannot contribute to the refinement. Hence centric data make no contribution. Correct weighting of term B relative to other terms should enable both A and B to decrease during refinement. Most of the calculations required for phases and their derivatives are also needed in the structure-amplitude refinement and thus the computer time increases by about 15% when each reflection contributes phase information for inclusion in B. The relevant algebraic expressions are given in Appendix I.*

The third term in (1) represents pseudo-potential energy terms:

$$C=\sum W_d(d_t-d_c)^2+\sum W_b(b_o-b_{\min})^2$$
$$+\sum W_v|\mathbf{V}| \quad (\text{for } b_o<b_{\min}).$$

The interatomic distance terms correspond to the central force-field approximation in vibrational spectroscopy but the force constants are crudely approximated by assuming the same value for all bonded distances, another value for atoms separated by two bonds and a third value for pairs separated by three bonds. The second term in C prevents undesirably close contacts occurring during refinement between atoms separated by more than three bonds. The chirality of specified tetrahedra of atoms may be restrained by giving equal weights to the restraints along the tetrahedral edges. Restraint specifications are supplied in a dictionary file of residues but supplementary restraints may be given in the steering data.

RESTRAIN calculates values for weighting coefficients that would be required to produce root-mean-square deviations from target distances comparable with the dispersion of these values found in small-molecule structures. Use of these weights is suitable in the final cycles of refinement while softer restraints may assist convergence at earlier stages. Application of harder restraints may severely reduce the rate of convergence.

When a libration tensor $\mathbf{L}$ (see below) is refined the relevant interatomic distances (d_c) are calculated using the expression

$$d_c=d_o\{1+\tfrac{1}{2}[\operatorname{tr}(\mathbf{L})-\mathbf{n}'\mathbf{L}\mathbf{n}]\},$$

which includes a libration correction. The uncorrected distance is d_o, tr denotes the trace operation, $\mathbf{n}$ is a column matrix denoting a unit vector along the interatomic direction and $\mathbf{n}'$ denotes its transpose. A root-mean-square libration of 7° about a direction perpendicular to a bond gives a correction of 0.01 Å to the calculated distance.

The last term in C is a planarity restraint. For planar or pseudo-planar groups such as the phenyl or imidazole rings of peptide side chains, the central force field is inadequate for maintaining the geometry imposed by π-electron delocalization. In spectroscopic calculations force fields have be introduced with a high proportion of off-diagonal terms in order to deal with such situations (Califano, 1976). Planar restraints in geometric least squares have been applied by positioning a dummy atom at some distance from the plane (Dodson, Isaacs & Rollett, 1976) or by minimizing the current least-squares best plane (Hendrickson & Konnert, 1980). Both these techniques not only restrain the atoms to be planar but also restrain them to the current least-squares plane. In a refinement where the orientation of the plane is implicitly refined a method is desirable that does not dampen changes in orientation produced by other terms in M. In Appendix II it is shown that the necessary and sufficient condition for a set of atoms to be planar is that the determinant of the matrix $\mathbf{V}$ be zero, where $\mathbf{V}$ is the product-moment matrix

$$\begin{pmatrix}\sum X_iX_i & \sum X_iY_i & \sum X_iZ_i\\ \sum Y_iX_i & \sum Y_iY_i & \sum Y_iZ_i\\ \sum Z_iX_i & \sum Z_iY_i & \sum Z_iZ_i\end{pmatrix}.$$

The summations in this matrix are over all the atoms in the plane and the coordinates are Cartesian with respect to the centroid of the planar group as origin. This determinant is not a quadratic form and also includes many off-diagonal terms. The pseudo-force constant W_v associated with this determinant may be chosen to yield root-mean-square deviations from the refined least-squares plane of less than 0·02 Å. Fig. 1 illustrates the use of planarity restraints applied in this manner to a tyrosine residue in the refinement of ribonuclease A (Borkakoti, Moss & Palmer, 1982).

Refinement parameters

The function M may be minimized with respect to some or all of the following parameters:

(a) overall scale factor (G);

(b) overall atomic displacement parameter (U);

(c) atomic coordinates (x_j, y_j and z_j);

(d) individual isotropic atomic displacement parameters (U_j);

(e) individual atomic anisotropic displacement parameters (U_j^{mn});

(f) rigid-body displacement parameters, including a translation tensor ($\mathbf{T}_i$), a libration tensor ($\mathbf{L}_i$) and a screw rotation tensor ($\mathbf{S}_i$) of the ith rigid group (Ibers & Hamilton, 1974);

(g) atomic occupancy factors (n_j).

The structure-factor expression employed is

$$F(hkl)=G\sum_{j=1} n_jf_jR_jS_j,$$

* Appendices I and III have been deposited with the British Library Lending Division as Supplementary Publication No. SUP 42182 (6 pp.). Copies may be obtained through The Executive Secretary, International Union of Crystallography, 5 Abbey Square, Chester CH1 2HU, England.

where

$$R_j = \exp\,[2\pi i(hx_j + ky_j + lz_j)]$$

$$S_j = \exp\,[-8\pi^2 U \sin^2\theta/\lambda^2]$$

$$\text{or}\quad \exp\,[-8\pi^2 U_j \sin^2\theta/\lambda^2]$$

$$\text{or}\quad \exp\,[-2\pi^2(h^2U_j^{11} + k^2U_j^{22} + l^2U_j^{33} + 2klU_j^{23} + 2lhU_j^{31} + 2hkU_j^{12})].$$

In the above formulation, reciprocal-lattice coordinates (h, k, l), atomic coordinates (x, y, z), and the **U** tensor components are all with respect to an orthonormal basis. The atomic form factor is denoted by f.

Individual atomic **U** tensors may be refined or rigid-body anisotropic refinement of atomic groups may be undertaken by refining the rigid-body displacement tensors **T**, **L** and **S**, in which case the **U** tensor components of the jth atom in the ith group are given by the equation

$$\mathbf{U}_j = \mathbf{T}_i + \mathbf{A}_j\mathbf{L}_i\mathbf{A}_j^T + \mathbf{A}_j\mathbf{S}_i + \mathbf{S}_i^T\mathbf{A}_j^T,$$

where

$$\mathbf{A}_j = \begin{pmatrix} 0 & z_j & -y_j \\ -z_j & 0 & x_j \\ y_j & -x_j & 0 \end{pmatrix}.$$

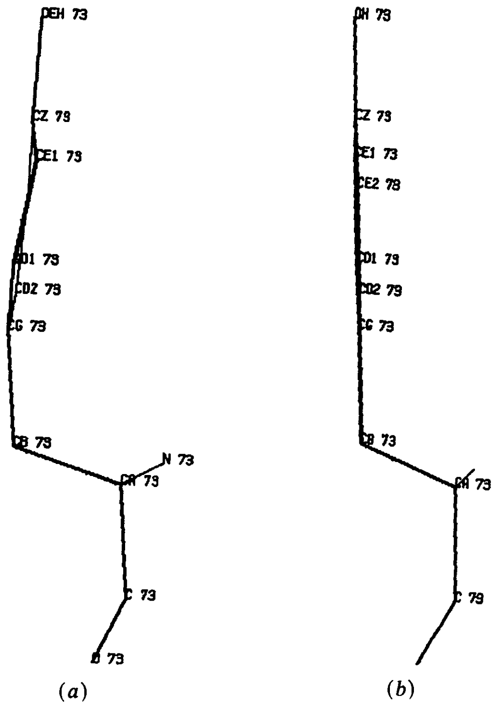

Fig. 1. Effect of applying planarity restraints on the side group Tyr 73 of bovine pancreatic ribonuclease A. (*a*) Planarity restrained by distance restraints only. (*b*) Planarity restraints implemented as described in the text.

Several subsites cannot be adequately treated by anisotropic refinement, which can only model unimodal distributions of atomic positions. *RESTRAIN* has facilities for coupling together occupancies so that their sum is unity. Fig. 2 shows two sites of histidine 119 in bovine pancreatic ribonuclease that were refined in this way.

Formation of normal equations

RESTRAIN sets up and solves an approximation to least-squares normal equations. A sketch of the relevant theory of function minimization is given in Appendix I.* The equations may be written as

$$\mathbf{N}(\mathbf{p})\delta\mathbf{p} = [\mathbf{K}(\mathbf{p}) + \lambda\mathbf{I}]\delta\mathbf{p} = -\mathbf{g}(\mathbf{p}), \quad \text{where } \lambda \geq 0.$$

The matrix **K** is an approximation to a normal matrix where contributions to off-diagonal terms are included for the energy restraints and 3 × 3 blocks are used for the contributions from the positional parameters of the atoms. All other off-diagonal terms are taken as zero. The correlation between the overall scale factor and the displacement parameters is taken into account in the way described by Rollett (1965). The column vector **p** describes the current parameter set and $\delta\mathbf{p}$ represents the parameter increments obtained by solving the normal equations. The vector $g(\mathbf{p})$ is $\nabla\, M(\mathbf{p})$, the gradient vector of M at **p**. **I** is a unit matrix and λ is a fraction of the mean diagonal element of **K**

$$\lambda = m\langle K_{ii}(\mathbf{p})\rangle_i.$$

The parameter m is chosen by the user but can also be modified by the program. This parameter is useful in situations where the normal equations are ill-conditioned. This may often happen in macromolecular refinement. Firstly, if a refinement is being undertaken with little or no diffraction data as when geometric regularization is carried out then M may only possess a weak minimum, *i.e.* the minimum is not a zero-dimensional subspace of parameter space. Secondly, if atoms are present in the refinement with high displacement parameters or low occupancies then small diagonal elements will appear in the normal matrix. This commonly occurs with solvent molecules. The parameter m is also useful when a refinement is undertaken with atomic positions with relatively large errors and M may depart significantly from quadratic curvature in the neighbourhood of **p**. In these cases values of m in the range $0 < m < 0{\cdot}5$ produce equations that are well conditioned and give a solution vector biased towards the direction of steepest descent. In a recent refinement of a B-DNA hexamer using interproton distances from NMR spectroscopy and no diffraction data (Clore, Gronenborn,

* See deposition footnote.

Table 1. *Comparison of computer processing times for various stages of refinement with and without vectorization*

Timings on the Cray-1S for the general structure-factor subroutine in *RESTRAIN* with and without vectorization. Timings are also given for the same routine after various modifications to vectorize the inner DO loops. Data refer to avian pancreatic polypeptide (aPP), which has 36 residues (293 non-hydrogen atoms + 38 water molecules), space group *C*2 and 17027 reflections (Glover *et al.*, 1983).

	Atomic displacements	Vectorization on/off	Look-up function	Internal function	Time (s)
(i)	Isotropic	off		sin/cos/exp	106
(ii)	Isotropic	on	sin/cos/exp		32·6
(iii)	Isotropic	on	sin/cos	exp	30·9
(iv)	Isotropic	on		sin/cos/exp	22·6
(v)	Anisotropic	on		sin/cos/exp	24·4

The most significant improvements occur owing to the use of the Cray-1S trigonometric functions to vectorize the inner DO loops [compare (iii) with (iv)]. Other changes, such as using the Cray-1S *SSUM* and Vector/Merge functions do not make significant differences in this routine. The final timing for this routine after optimization is 19·5 s for the isotropic calculation.

Central processor times spent on different parts of a least-squares refinement calculation on the Cray-1S computer with and without using the vector facilities. Data are taken from the refinement of avian pancreatic polypeptide (Glover *et al.*, 1983).

V = on: vectorization switched on; V = off: vectorization switched off.

Function	Time (s)		Time per unit calculation
	V = on	V = off	for V = on
Structure-factor calculation	22·56	105·89	2×10^{-6} s atom^{-1} reflection^{-1} (general equivalent position)$^{-1}$
Assemble normal equations	6·12	35·43	1×10^{-6} s atom^{-1} reflection^{-1}
Geometry restraints	0·81	0·81	2×10^{-3} s atom^{-1}
Solution of normal equations by Gauss–Seidel method	0·71	0·78	2×10^{-3} s atom^{-1}

Table 2. *Analysis of the Gauss–Seidel solution of the normal equations; a cycle of glucagon refinement*

$\delta\mathbf{p}\,.\,\mathbf{g}/(\|\delta\mathbf{p}\|\,\|\mathbf{g}\|) = 0{\cdot}824$; $\rho = 4{\cdot}46$; i = iteration number; $\delta\mathbf{p}^i$ = solution vector at iteration i; $\mathbf{q}^i$(mean) = mean of elements of $(\delta\mathbf{p}^i - \delta\mathbf{p}^{i-1})$; $\mathbf{q}^i$(max.) = maximum element of $(\delta\mathbf{p}^i - \delta\mathbf{p}^{i-1})$; $\mathbf{g}$ = gradient vector; ρ is defined in Appendix III.

i	$\mathbf{q}^i$(mean)	$\mathbf{q}^i$(max.)	$\|\delta\mathbf{p}^i - \delta\mathbf{p}^{i-1}\|/\|\delta\mathbf{p}^i\|$
4	0·0034	0·0080	0·3143
5	0·0016	0·0058	0·1401
6	0·0008	0·0044	0·0755
7	0·0005	0·0037	0·0459
8	0·0003	0·0030	0·0303
9	0·0002	0·0034	0·0211
10	0·0001	0·0027	0·0152

Moss & Tickle, 1985) values of m greater than 2 were successfully used to combat severe ill conditioning. In order to assess the condition of the normal matrix and the relation between the directions of the solution vector and the gradient vector, *RESTRAIN* computes a statistic ρ, which measures the condition of the matrix (see Appendix III)* and also the scalar

$$\delta\mathbf{p}\,.\,\mathbf{g}(\mathbf{p})/\|\delta\mathbf{p}\|\,\|\mathbf{g}(\mathbf{p})\|.$$

If this quantity is close to zero the solution vector is approximately orthogonal to the steepest-descent vector and progress of the refinement may be improved by increasing m. The use of normal equations modified by the term $\lambda\mathbf{I}$ is sometimes known as the Levenberg–Marquardt method and its use in pathological cases has been widely discussed (see for example Wolfe, 1978).

Solution of normal equations

In order to solve the modified normal equations

$$\mathbf{N}(\mathbf{p})\,\delta\mathbf{p} = -\mathbf{g}(\mathbf{p})$$

the Gauss–Seidel method is employed. A brief description of the method is given in Appendix III. This method has been used for solving the large linear systems of equations that arise from the solution of partial differential equations. It is a stationary iterative method, which is extremely simple to program and requires only that one normal equation be held in central memory at any given time. As the matrix $\mathbf{N}(\mathbf{p})$ is sparse, each row is held in an array where only non-zero elements are stored. A second array contains the column numbers of the corresponding elements in the first array. These arrays are held on a scratch file. In the Gauss–Seidel method the rows of $\mathbf{N}(\mathbf{p})$ are repeatedly multiplied by a column vector whose elements gradually approximate to the solution of the equations. Because of the use of the auxiliary array to index the rows, the DO loops contain implied subscripted subscripts and vectorization is inhibited. Nevertheless, the solution of the equations typically occupies less than 15% of the total central processor time of the restrained least-squares calculation (see Table 1). Table 2 shows the progress of the Gauss–Seidel iterations during a cycle of refinement of glucagon, which is a polypeptide of 29 residues and crystallizes in space group $P2_13$ (Sasaki, Dockerill, Adamiak, Tickle & Blundell, 1975).

The precision required in the solution of the normal equations need only be sufficient to ensure that errors

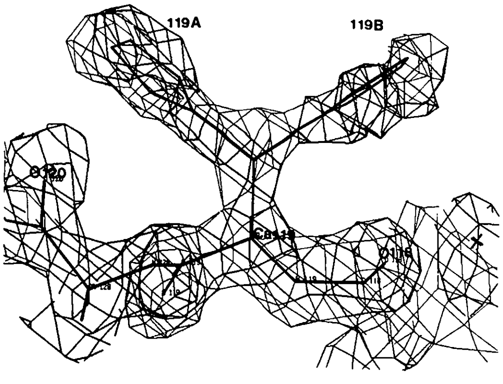

Fig. 2. Electron density $(2|F_o| - |F_c|)$ in the region of His 119 in bovine pancreatic ribonuclease A. Occupancies of sites A and B are 0·8 and 0·2 respectively. Contour level is 0·6 e Å^{-3}.

* Appendix III has been deposited. See deposition footnote.

in the solution vector are small compared with Taylor-series truncation errors inherent in the Gauss–Newton theory described in Appendix I. *RESTRAIN* terminates the iterations when

$$\|\delta\mathbf{p}^i - \delta\mathbf{p}^{i-1}\| / \|\delta\mathbf{p}^i\| < s,$$

where superscripts denote iteration numbers and s may be chosen by the user but defaults to 0·02. Values of s as high as 0·2 can still enable satisfactory progress to be made in the earlier stages of refinement.

Owing to the Taylor-series approximation referred to above the block-diagonal approximation used for structure-factor derivatives, the solution vector $\delta\mathbf{p}$ obtained from the Gauss–Seidel iterations does not in general determine the minimum of $M(\mathbf{p})$. A better approximation may be economically determined by accepting the direction of $\delta\mathbf{p}$ and determining the magnitude of the vector by recalculating the function M for three parameter vectors $\mathbf{p}$, $\mathbf{p}+\alpha\delta\mathbf{p}$ and $\mathbf{p}+2\alpha\delta\mathbf{p}$, where α is a partial shift factor supplied by the user and typically is in the range 0·2 to 0·4. The optimum shift vector is found by fitting a quadratic curve through the three function values and using the shift factor corresponding to the minimum point. In order to shorten computing time the calculation of M for shift-factor determination is carried out by sampling the terms in the summation. Typically only one tenth of the terms are re-evaluated. At 1·5 Å resolution optimum shift factors are typically in the range 0·8 to 1·0 but at lower resolution (2·5 Å) values in the range 0·4 to 0·8 are encountered owing to the more serious neglect of off-diagonal terms in the normal matrix.

Vector processing

A vector processor differs from a conventional serial processor such as an IBM 3081 or a VAX 11/750 by its ability to process a set of operands with one instruction. The Fortran statements

```
   DO 10 I=1,1000
10 A(I)=SC*B(I)
```

constitute a loop, which on a serial processor requires the execution of branch on-condition instructions in addition to the floating-point multiplications. On a Cray-1 computer, segments of the arrays 64 elements long can be processed by one instruction. In general, vector processors are well suited to algorithms that call for identical operations to be carried out on the elements of an array. An essential condition for vectorization is usually that the addresses of successive array elements accessed must increase in arithmetic progression and the Cray-1 computer currently imposes the restriction that only the innermost DO loops of the program may be vectorized by the Fortran compiler.

The computation of structure factors and their derivatives is by far the most expensive part of the calculation on a scalar machine. Fortunately a traditional algorithm for these calculations (Rollett, 1965) is very amenable to vectorization. For each reflection all loops over the number of atoms make use of the vector registers of the Cray-1. The scattering factors are calculated from a four-Gaussian approximation and it is interesting to note that this takes approximately the same processor time as consulting a look-up table, which cannot be vectorized. This is illustrated in Table 1 where timings for the various stages of the refinement process are given with and without vectorization. Fortunately, as this table shows, the greatest gains in speed with vectorization are in the structure-factor routine.

The vectorization is optimized by:

(*a*) using few loops with long code blocks in preference to many short code loops;

(*b*) putting long loops inside short loops since only innermost loops are vectorized on the Cray-1;

(*c*) removing any code (*e.g.* IF statements, I/O requests) from loops where it would inhibit vectorization;

(*d*) storing data that occur with irregular address increments into temporary arrays with regular address increments;

(*e*) replacing look-up tables for any functions by built-in Cray-1 functions that are able to use the vector registers.

The latter changes increase both the speed and precision because the use of look-up tables involves irregular addressing of arrays, and more approximate function evaluation. Gains in speed of over 20% were achieved by replacing look-up tables with internal Cray-1 trigonometric functions in the structure-factor calculation.

A second way of speeding up the structure-factor calculations involves the use of the product forms of the geometric structure-factor formulae. These expressions are space-group specific and involve the factorization of the temperature-factor expression and hence are only suited to isotropic refinement. However, the use of such a code on the Cray-1 usually speeds up the calculation by less than a factor of two. We have observed larger gains in speed using space-group-specific subroutines on scalar machines.

The scope for vectorization in the energy part of the calculation and in the solution of normal equations is more limited because these processes inherently involve much irregular addressing.

Discussion

It is interesting to compare the refinement method implemented in *RESTRAIN* with that used by other workers. Structure-amplitude terms supplemented by distance restraints have been used in several other computer programs (Sussman *et al.*, 1977; Hendrickson & Konnert, 1980) for refinement of macromolecules. The use of isomorphous phases in reciprocal-space refinement is related to the real-space refinement method proposed by Diamond (1971). In the latter method the difference between the observed electron density (ρ_o) and that calculated from trial atomic positions (ρ_c) is minimized. This least-squares problem may be reformulated in reciprocal space:

$$\int (\rho_0 - \rho_c)^2 \mathrm{d}V = (2/V) \sum |F_o - F_c|^2.$$

This sum of squares may be expressed in terms of the real and imaginary components of the structure factors:

$$\sum |F_o - F_c|^2 = \sum (A_o - A_c)^2 + \sum (B_o - B_c)^2.$$

It will be noted that real-space refinement gives equal weight to each squared term and also assumes zero correlation between the errors in the A and B parts of the structure factor. The formulation in *RESTRAIN* allows different squared terms but assumes zero correlation between the errors associated with amplitudes and phases. Ignoring correlation in both cases can only be justified by computational convenience and the difficulty in assessing the magnitude of the effect.

The large errors typically associated with isomorphous phases particularly at higher Bragg angles limit their use to the earlier stages of structure refinement. Our own experience of the use of phases between resolutions of 2·6 and 4·0 Å have shown that successful refinement is highly dependent on correct phase weighting, which is more difficult to achieve than correct structure-amplitude weighting owing to the cyclic nature of the phase data. Constrained refinement of atomic positions can correct larger positional errors and phases may contribute more to such a refinement at low resolution where phases tend to have higher figures of merit.

Refinement of free anisotropic displacement parameters can only be undertaken when diffraction data up to a resolution approaching 1 Å are available. Yet the observed displacements may be highly anisotropic (Glover, Haneef, Pitts, Wood, Moss, Tickle & Blundell, 1983). In such cases the use of the rigid-body **TLS** option provides a particularly suitable method of refining side chains containing ring systems and other rigid groups. The mean-square amplitudes of vibration of intramolecular distances in a benzene ring are about 0·003 Å^2 (Kimuro & Kubo, 1960). This is at least an order of magnitude smaller than the mean-square displacements observed in protein structures. The rigid-body approximation is therefore a valid approximation for such rings and requires only a modest increase in the number of parameters refined. A **TLS** refinement of the eight atoms of a tyrosine side chain requires 20 **TLS** parameters whereas a free anisotropic refinement requires 48 mean-square displacement parameters. Pawley (1970) has discussed the refinement of **TLS** parameters in small-molecule structures where the rigid-body displacements are much smaller than in macromolecules. It is interesting to consider the range of applicability of the **TLS** approximation in macromolecular refinement. Rigid-group vibrations and some internal modes such as the B_{2g} ring-puckering vibrations of the carbon atoms of a benzene ring can be represented exactly by **TLS** tensors. The planar-ring side groups of residues such as histidine, phenylalanine, tyrosine, tryptophan and the bases of nucleic acids are clearly candidates for anisotropic rigid-body treatment. Tetrahedral moieties such as occur in isoleucine, leucine, threonine and valine as well as smaller planar groups such as carboxyl, amide or guanidinium can also be refined in this way but the economy in parameters is smaller than with larger groups. Interpretation of results must take into account the fact that **TLS** components contain a linear dependency where a group consists of five or fewer coplanar atoms (Ibers & Hamilton, 1974).

The magnitude of librational disorder present in protein structures has a significant effect on calculated interatomic distances and this is particularly relevant in geometrically restrained refinements. Our recent **TLS** refinements of avian pancreatic polypeptide and ribonuclease (unpublished results) at resolutions of 0·98 and 1·45 Å respectively have shown distance corrections of up to 0·02 Å in the well ordered side-chain rings and up to 0·06 Å in an active-site sulphate ion. Libration effects in main chains are less significant. However, the extensive disorder often shown by side chains of residues such as lysine and arginine on the surface of protein molecules could call into question the validity of applying distance restraints to such groups unless the anharmonic and anisotropic disorder can be correctly modelled. Unfortunately the diffraction information needed for such modelling is in the diffuse scattering rather than in the Bragg reflections. It is hoped that molecular-dynamics simulations may provide a basis for improved least-squares models (Kuriyan, Karplus, Levy & Petsko, 1984).

The use of the Gauss–Seidel method for solving the linear equations arising from the minimization process has also been employed by Hoard & Nordman (1979) in structure refinement. Nordman, however, updates the structure factors during the Gauss–Seidel iterations. The method of conjugate gradients has been employed for solving systems of linear equations with large sparse matrices (Hendrickson & Konnert, 1980). The time requirements of these two techniques are similar (Ralston, 1965) and as the solution of the equations occupies less than 20% of the computer time for a refinement cycle, it is difficult to see a clear advantage in either method.

The use of the Cray-1 vector processor allows a significant increase in speed owing to the highly vectorizable classical algorithms for the computation of structure factors and their derivatives. Fortunately it is just these algorithms that consume most of the processor time when only scalar processing is employed.

APPENDIX II

The condition for a set of atoms to lie in a plane

Consider a set of atoms with position vectors with respect to their centroid represented by column matrices $\mathbf{x}_i$. Let $\mathbf{n}$ be a column matrix representing a unit vector perpendicular to the plane and let primed symbols represent row matrices. Then for all i

$$\mathbf{n}'\mathbf{x}_i = 0.$$

Hence

$$\begin{aligned} 0 &= \sum (\mathbf{n}'\mathbf{x}_i)^2 \\ &= \mathbf{n}' \sum (\mathbf{x}_i\mathbf{x}_i^T)\mathbf{n} \\ &= \mathbf{n}'\mathbf{V}\mathbf{n}, \end{aligned}$$

where the summations are taken over the atoms in the plane and $\mathbf{V}$ is the product-moment matrix. Thus, $\mathbf{n}'\mathbf{V}\mathbf{n}$ can be expressed as a sum of squares that will equal zero when the atoms are coplanar. The matrix is therefore positive semidefinite and its determinant is zero.

To show that a zero determinant of the matrix of product moments is a sufficient condition for planarity, the steps in the above argument may be reversed.

The first and second derivatives of $|V|$ required in the Newton method are computed by a forward difference technique. If $|V|(x, y, z)$ is the determinant when a given atom in the plane is at (x, y, z) then first and second derivatives are calculated from expressions such as

$$\partial|V|/\partial x = [|V|(x+\delta x, y, z) - |V|(x, y, z)]/\delta x$$

and

$$\begin{aligned} \partial^2|V|/\partial x\partial y = [&|V|(x+\delta x, y+\delta y, z) - |V|(x+\delta x, y, z) \\ &- |V|(x, y+\delta y, z) + |V|(x, y, z)]/\delta x\,\delta y. \end{aligned}$$

The increments δx and δy are 0·01 Å. It should be noted that each recalculation of $|V|$ requires a re-evaluation of the centroid of the planar group.

References

Agarwal, R. C. (1978). *Acta Cryst.* **A34**, 791-809.

Baker, E. M. (1980). *J. Mol. Biol.* **141**, 441-484.

Borkakoti, N., Moss, D. S. & Palmer, R. A. (1982). *Acta Cryst.* **B38**, 2210-2217.

Borkakoti, N., Palmer, R. A., Haneef, I. & Moss, D. S. (1983). *J. Mol. Biol.* **169**, 743-755.

Califano, S. (1976). *Vibrational States.* New York: Wiley.

Clore, G. M., Gronenborn, A. M., Moss, D. S. & Tickle, I. J. (1985). *J. Mol. Biol.* In the press.

Cruickshank, D. W. J. (1961). *Computing Methods and the Phase Problem in X-ray Crystal Analysis*, edited by R. Pepinski, J. M. Robertson & J. C, Speakman. London: Pergamon Press.

Diamond, R. (1971). *Acta Cryst.* **A27**, 436-452.

Dodson, E. J., Isaacs, N. W. & Rollett, J. S. (1976). *Acta Cryst.* **A32**, 311-315.

Freer, S. T., Alden, R. A., Carter, C. W. Jr & Kraut, J. (1975). *J. Biol. Chem.* **250**, 46-54.

Furey, W. Jr, Wang, B. C. & Sax, M. (1982). *J. Appl. Cryst.* **15**, 160-166.

Girling, R. L., Houston, T. E., Schmidt, W. C. & Amma, E. L. (1980). *Acta Cryst.* **A36**, 43-50.

Glover, I., Haneef, I., Pitts, J., Wood, S., Moss, D., Tickle, I. & Blundell, T. (1983). *Biopolymers*, **22**, 293-304.

Hendrickson, W. A. & Konnert, J. H. (1980). *Computing in Crystallography*, ch. 13, pp. 13.01-13.25. Bangalore: Indian Academy of Sciences.

Hoard, L. G. & Nordman, C. E. (1979). *Acta Cryst.* **A35**, 1010-1015.

Ibers, J. A. & Hamilton, W. C. (1974). *International Tables for X-ray Crystallography*, Vol. IV, pp. 320-322. Birmingham: Kynoch Press. (Present distributor D. Reidel, Dordrecht.)

Jack, A. & Levitt, M. (1978). *Acta Cryst.* **A34**, 931-935.

Jones, T. A. & Liljas, L. (1984). *Acta Cryst.* **A40**, 50-57.

Kimuro, K. & Kubo, M. (1960). *J. Chem. Phys.* **32**, 1776-1780.

Konnert, J. H. (1976). *Acta Cryst.* **A32**, 614-617.

Konnert, J. H. & Hendrickson, W. A. (1980). *Acta Cryst.* **A36**, 344-350.

Kuriyan, J., Karplus, M., Levy, R. M. & Petsko, G. A. (1984). Abstracts of 8th International Biophysics Congress, Bristol, England, p. 71.

Moss, D. S. & Morfew, A. J. (1982). *Comput. Chem.* **6**(1), 1-3.

Pawley, G. S. (1970). *Acta Cryst.* **A26**, 289-292.

Ralston, A. (1965). *A First Course in Numerical Analysis*, pp. 431-445. New York: McGraw-Hill.

Rollett, J. S. (1965). *Computing Methods in Crystallography*, pp. 38-56. Oxford: Pergamon Press.

Sasaki, K., Dockerill, S., Adamiak, D. A., Tickle, I. J. & Blundell, T. L. (1975). *Nature (London)*, **257**, 751-757.

Sielecki, A. R., Hendrickson, W. A., Broughton, C. G., Delbaere, L. T. J., Bryer, G. D. & James, M. N. G. (1979). *J. Mol. Biol.* **134**, 781-804.

Sussman, J. L., Holbrook, S. R., Church, G. M. & Kim, S. H. (1977). *Acta Cryst.* **A33**, 800-804.

Wolfe, M. A. (1978). *Numerical Methods for Unconstrained Optimisation*, pp. 225-254. London: Van Nostrand-Reinhold.