The diffuse spot around $(h,0,0)$ and $(0,k,0)$ will of course be similar; the coefficients of l_1 and l_2 in (12) acquire the same form as that of the coefficient of l_3 on appropriate transformation of axes.

In order to observe these diffuse spots, single crystals of cubic ZnS doped with, say, Se must be grown. Se is expected to replace S and form distortion centres of T_d symmetry. The difficulties in these experiments are the faintness of the diffuse scattering due to low strength of the substitutional distortion centres, and the presence of thermal diffuse scattering. However, by working at low temperatures, it may be possible to observe the anisotropic effect in the reciprocal plane (normal to **H**) through some strong axial reflexion.

References

COCHRAN, W. (1956). *Acta Cryst.* **9**, 259–262.

COCHRAN, W. & KARTHA, G. (1956). *Acta Cryst.* **9**, 941–943.

DAWSON, B. (1967*a*). *Proc. R. Soc. London Ser. A*, **298**, 264–288.

DAWSON, B. (1967*b*). *Proc. R. Soc. London Ser. A*, **298**, 379–394.

DAWSON, B. (1967*c*). *Proc. R. Soc. London Ser. A*, **298**, 395–401.

EKSTEIN, H. (1945). *Phys. Rev.* **68**, 120–124.

JAHNKE, E. & EMDE, F. (1945). *Tables of Functions with Formulae and Curves*, p. 112. New York: Dover Publications.

MERISALO, M. & JARVINEN, M. (1977). *Philos. Mag.* **B37**, 233–240.

POWELL, J. L. & CRASEMAN, B. (1961). *Quantum Mechanics*, p. 247. London: Addison-Wesley.

PRATHER, J. L. (1961). *Atomic Energy Levels in Crystals, Natl Bur. Stand. (US) Monogr.* **19**, 3–6.

VON DER LAGE, F. C. & BETHE, H. A. (1947). *Phys. Rev.* **71**, 612–622.

Acta Cryst. (1980). A**36**, 43–50

Macromolecular Structure Refinement by Restrained Least-Squares and Interactive Graphics as Applied to Sickling Deer Type III Hemoglobin

BY R. L. GIRLING, T. E. HOUSTON, W. C. SCHMIDT JR AND E. L. AMMA

Department of Chemistry, University of South Carolina, Columbia, SC 29208, *USA*

(*Received* 27 *February* 1979; *accepted* 25 *July* 1979)

Abstract

A restrained least-squares (RLS) computer program and two interactive graphics (IG) systems have been used in combination to refine the structure of deer type III hemoglobin. By alternating applications of RLS with examinations and corrections of the atomic model superposed on electron density maps (IG), the residual has been reduced from ~0·42 to ~0·25 and the sites of dubious fit between model and map reduced to ~6% of the residues or ~3% of the atoms. It was possible to fit routinely ~4600 atoms to X-ray intensity data sets ranging from less than 6000 (9·0–4·0 Å resolution) to ~21 500 points (6·0–1·98 Å resolution) employing RLS, which uses interatomic distances to retain structural integrity. Convergence was rapid and many shifts greater than 1 Å were recorded. An in-house graphics display allowed the placement of atoms not in the original atomic model and GRIP, a fast-response interactive graphics system, was used to correct any gross conformational misfit of the atomic model to the electron density maps. The man hours needed to run both GRIP and RLS is less than previously reported real-space methods. The strategy of how RLS and IG can be best applied and how the molecular structure changed during refinement are discussed.

Introduction

A key factor in an X-ray protein structural refinement is the improvement in the interpretability of the electron density maps so that structural features are more readily perceived. The use of high-resolution X-ray data to improve the atomic coordinates of a macromolecule whose basic structure is 'solved' at low resolution has followed diverse paths (Blundell & Johnson, 1976). Until recently, the refinement methods reported were: real-space fitting by visual fit in a Richards Box (Richards, 1968) or its present equivalent, the electronic optical comparator (Collins, Cotton, Hazen, Meyer & Morimoto, 1975); automated real-space fitting (*e.g.* Diamond, 1966, 1974; Huber, Kukla, Bode, Schwager, Bartels, Deisenhofer & Steigemann, 1974; Freer, Alden, Carter & Kraut, 1975; Ladner, Heidner & Perutz, 1977; Moews & Kretsinger, 1975);

0567-7394/80/010043-08$01.00

conventional reciprocal-space fitting (Watenpaugh, Sieker, Herriott & Jensen, 1973); and energy minimization procedures (Levitt, 1974; Levitt & Lifson, 1969; Rasse, Warme & Scheraga, 1974; Hermans & McQueen, 1974; Hingerty, Brown & Jack, 1978). In the past, real-space refinement has been the chosen method because energy minimization techniques have failed to fit the X-ray data to the same precision and conventional reciprocal-space refinements had far greater computing requirements than the real-space methods. Two recent developments have made a reciprocal-space refinement viable and have substantially reduced the number of man hours needed to implement a refinement: (1) a restrained least-squares (RLS) procedure which uses known structural parameters as observational equations to supplement the X-ray data (Konnert, 1976; Schmidt, Girling & Amma, 1977; Hendrickson & Konnert, 1979; Anderson, Stenkamp & Steitz, 1978; Sussman, Holbrook, Warrant, Church & Kim, 1978; Sussman, Holbrook, Church & Kim, 1977); (2) the availability of high-resolution* interactive graphics systems (IG) for rapid examination of electron density maps (GRIP, 1975) and, if necessary, for modification of refined atomic positions to conform to these maps. RLS was chosen here because it biases the solution towards chemically reasonable geometry, it converges rapidly without high-resolution data, and it can be implemented easily by those familiar with conventional least-squares procedures. RLS differs from conventional least-squares refinement by : (1) using molecular geometry as if it were observational data, (2) solving the normal equations using the conjugate-gradient method, and (3) saving only selected elements of the derivative matrix, thereby reducing computer memory requirements to 1% of that required by conventional full-matrix least squares. This report analyzes the results of applying a combination of RLS and IG to refine an initial model of sickling deer type III hemoglobin [Hb(DIII)] obtained by a molecular-replacement solution (Schmidt, Girling, Houston, Sproul, Amma & Huisman, 1977) from an initial R† of ~0·42 (4372 atoms/5589 reflections) to a final R of ~0·25 (4556 atoms/21 469 reflections). The RLS procedure has not previously been applied to a molecule of this size (Konnert, 1976; Anderson, Stenkamp & Steitz, 1978) nor to a molecular replacement solution unrefined by other methods (Sussman *et al.*, 1978; Anderson, Stenkamp & Steitz, 1978). The degree of 'hands off' RLS refinement is determined and the supplemental value of IG to the RLS is shown. The combination of these techniques may have wide applicability and other workers may profit from our experience in terms of the power, limitations and most effective usage in macromolecular problems. (For a detailed report, see Girling, Schmidt, Houston & Amma, 1978.)

* Greater than 4 display points per mm.

† $R = \sum ||F_o| - |F_c||_I / \sum |F_o|_I$.

Experimental and refinement parameters

The source, purification, crystallization conditions, unit-cell data and details of the structure solution of deer hemoglobin β-chain type III [Hb(DIII)] can be found in Schmidt, Girling, Houston, Sproul, Amma & Huisman (1977).

In RLS, molecular geometry is incorporated as observed data so that the sums of differences between the n observed and calculated structure amplitudes (F_o, F_c) and between the m 'ideal' geometric terms, d_o, (Hendrickson, 1977) and the corresponding terms (d_c) in the observed structure are minimized. The equations to be minimized take the form:

$$\zeta = \sum_{i=1}^{n} W_i(|F_o|_i - |F_c|_i)^2 + \sum_{l=1}^{m} W_l(d_{o,l}^2 - d_{c,l}^2)^2,* \quad (1)$$

where W_l is the weight given the idealization of the lth geometric real-space term (types 1–4, below) and W_i is the weight of the ith X-ray intensity. The specific geometric terms (restrictions) as applied to Hb(DIII) fall into four types (see Table 1 in Girling, Schmidt, Houston & Amma, 1978, for more details). Excluding hydrogen atoms, there was a restriction for: type 1, every covalently bonded atom pair (1,2 distance); type 2, every bond angle (1,3 distance); type 3, each peptide link specifying the alpha carbon to carbonyl oxygen [C$\alpha\cdots$O] (1,4 distance); type 4, the planarity of the peptide link, the planarity of the appropriate part of the side chains in the phenylalanine, tryptophan, arginine, asparagine, aspartic acid, tyrosine, glutamine, glutamic acid and histidine residues and the planarity of the four

* In later versions of the program $(d_{o,l}^2 - d_{c,l}^2) \rightarrow (d_{o,l} - d_{c,l})$.

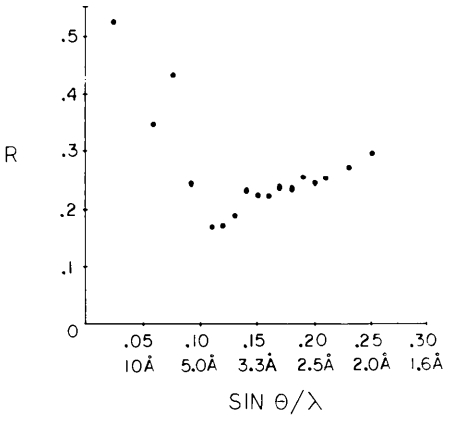

Fig. 1. The residual, $\sum ||F_o| - |F_c|| / \sum |F_o|$, of various shells of intensity data (solvent not included). The additional scale in Å is $d = \lambda/2 \sin\theta$.

Table 1. *Summary of the parameters associated with the refinement of* Hb(DIII)

NAT is the number of atoms in the refinement giving 3 × NAT + 2 variables associated with the atomic coordinates, RT is the total number of restraints between bonded atoms (σ = 0·062), non-bonded atoms (σ = 0·125) and atoms in a plane (σ = 0·1), where the weight associated with bonded or non-bonded restraints is $[1/2(\sigma)$ (ideal distance)$]^2$ and that associated with plane restraints is $1/\sigma^2$; NPT is the number of X-ray intensity data points in the resolution range, RES (in Å). W_i is the weight for the X-ray intensity data; R_B is the conventional residual $\sum||F_o| - |F_c||/\sum|F_o|$ calculated prior to the least-squares fit for the substep; R_{EST} is the residual for an ~3% representative sample of the data after the shifts and damping factors for that substep have been applied to the atomic coordinates; $\langle\Delta\varphi_1\rangle$ is the average phase change $\langle|\varphi_n - \varphi_{n-1}|\rangle$ in that substep for a group of reflections near 5 Å resolution; $\langle S\rangle$ is the average shift for each cycle in Å after the damping factor has been applied; $\langle S_T\rangle$ is the average accumulated total shift in Å from the beginning of that step of the substep being examined; D_B is the bonded r.m.s. deviation from the ideal distance after the shifts for that cycle have been applied; T is the CPU time in h on an IBM 370-168 computer, where the maximum core was 1500 K, the PDP 11-40 (P), or the time spent viewing and refining the molecule on the GRIP system (G), where the appropriate letter in parentheses designates the operating system. A more complete listing of refinement parameters are to be found in Girling *et al.* (1978).

Step no.	Refinement type	Substep (n)	NAT	RT	NPT	RES	W_i	R_B	R_{EST}	$\langle\Delta\varphi_1\rangle$	$\langle S\rangle$	$\langle S_T\rangle$	D_B	T
1	Restrained least squares	CYCLE 1	4372	11804	5589	4·0–9·0	1/169	0·419	0·404	17	0·24	0·24	0·041	2·7
		CYCLE 2	4110	11265	5589	4·0–9·0	1/169	0·387	0·362	11	0·19	0·35	0·079	2·7
		CYCLE 3	4110	11265	5589	4·0–9·0	1/121	0·357	0·350	10	0·18	0·44	0·106	2·7
		CYCLE 4	4110	11265	5589	4·0–9·0	1/100	0·337	0·342	5	0·08	0·45	0·105	2·7
		CYCLE 5	4110	11265	9524	3·0–5·0	1/81	0·400	0·380	5	0·21	0·51	0·091	4·5
		CYCLE 6	4110	11265	9524	3·0–5·0	1/72	0·358	0·356	5	0·15	0·56	0·062	4·5
		CYCLE 7	4110	11265	9524	3·0–5·0	1/58	0·336	0·342	4	0·13	0·61	0·045	4·5
		CYCLE 8	4110	11265	9524	3·0–5·0	1/49	0·321	0·332*	4	0·12	0·64	0·049	4·5
		CYCLE 9	4110	11265	11647	2·8–5·0	1/49	0·321	0·320	4	0·13	0·68	0·053	5·4
		CYCLE 10	4110	11265	11647	2·8–5·0	1/27	0·306	0·307	4	0·12	0·72	0·039	5·4
		CYCLE 11	4110	11265	13934	2·6–5·0	1/21	0·306	0·292	4	0·10	0·73	0·042	6·3
		CYCLE 12	4110	11265	13934	2·6–5·0	1/17	0·290	0·290	3	0·10	0·77	0·036	6·3
		CYCLE 13	4110	11265	13934	2·6–5·0	1/14	0·275	0·265	3	0·09	0·79	0·028	6·3
		CYCLE 14	4110	11265	13934	2·6–5·0	1/14	0·265	0·255	2	0·07	0·81	0·037	6·3
		CYCLE 15	4110	11265	19930	2·0–5·0	1/25	0·289	0·288	3	0·11	0·81	0·040	8·9
		CYCLE 16	4110	11265	19930	2·0–5·0	1/17	0·284	0·279		0·09	0·83	0·029	8·9
2	Addition of 415 atoms from molecular graphics		⟨3613⟩	–	13929	2·6–5·0	–	⟨0·32⟩	–		–	–	–	22·0 200·0 (P)
3	Restrained least squares	CYCLE 17	4525	12314	8419	3·3–6·0	1/17	0·340	0·290	15	0·26	0·26	0·430	4·2
		CYCLE 18	4525	12314	8419	3·3–6·0	1/17	0·278	0·262	9	0·17	0·30	0·270	4·2
		CYCLE 19	4525	12314	8419	3·3–6·0	1/16	0·254	0·248	5	0·12	0·35	0·220	4·2
		CYCLE 20	4525	12314	21469	2·0–6·0	1/16	0·310	0·280	3	0·17	0·36	0·136	10·3
		CYCLE 21	4552	12380	21469	2·0–6·0	1/12	0·286	0·263	6	0·12	0·40	0·055	10·5
4	Rebuild structure on GRIP system		⟨3613⟩	–	13929	2·6–5·0		⟨0·32⟩	–		–	–	–	45·0 ~35·0 (G)
5	Restrained least squares	CYCLE 22	4552	12380	21469	2·0–6·0	1/12	0·367	0·315	21	0·29	0·29	0·125	10·5
		CYCLE 23	4552	12380	21469	2·0–6·0	1/12	0·307	0·287	9	0·17	0·32	0·078	10·5
		CYCLE 24	4552	12380	21469	2·0–6·0	1/12	0·287	0·273	5	0·12	0·39	0·046	10·5
		CYCLE 25	4552	12380	21469	2·0–6·0	1/9	0·275	0·256	3	0·11	0·42	0·040	10·5
6	Rebuild structure on GRIP system		4552	–	21469	2·0–6·0	–	~0·260						5·0 ~128·0 (G)

* Estimated value.

pyrrole rings and the two carboxyl groups of each heme group. In the heme group, the iron atoms were considered bound to the four nitrogen atoms of the porphyrin resulting in four bonding (type 1) and eight non-bonding Fe–N–C (type 2) constraints.

Each reflection was assigned the same weight, W_i, but it is more correct to use weights determined by counting statistics or scattering angle (Fig. 1). Under the above conditions all atoms may move nearly independently if $W_i \gg W_l$ but their movements are increasingly restrained as W_i decreases for a given set of W_l's. Konnert (1976) suggests that the ratio of W_i to W_l be set by approximating $(d_o^2 - d_c^2)$ with $2d\Delta d$ and then equating the two right-hand terms of (1). In this way, W_i and W_l can be related to the variances of F and d, factors routinely estimated in X-ray structure determinations.*

The second part of the refinement (IG) utilized the PDP-11/40 with a VT11 graphics processor and a 0·43 m display or the GRIP system (1975) to inspect visually and to adjust the hemoglobin model structure after RLS 'convergence'. Atoms not in the model were located by estimating coordinates from the PDP-11 stereo display of the model superposed on an electron density map (step 2, Table 1). The PDP-11 display system was inadequate for major changes in the backbone, such as when more than two or three residues

* The W_i/W_l ratio will differ in more recent versions of the program. See footnote to equation (1).

had to be shifted, and subsequent IG examinations (steps 4, 6, Table 1) and molecular manipulations utilized the GRIP system, where up to ten residues could be displayed and fitted to the electron density map.

For the first two inspections (steps 2, 4) eight modified* electron density maps were examined. Such a modified electron density map was used in order to remove bias from the part of the map being examined. The final inspection was made with maps using $(|2F_o| - |F_c|) \exp(i\varphi_c)$ and $(|F_o| - |F_c|) \exp(i\varphi_c)$ coefficients in which F_c and φ_c were calculated from the model. One direction was contoured with the former coefficients and the other two with the latter set in order that changes in three-dimensional atomic positioning would fit our interpretation of both maps.† Only positive density at one level in each map was displayed so that incorrectly positioned atoms would not be within the contours. This procedure tends to eliminate fitting false details in difference maps. Contour levels varied between 0·2 and 3·0 e/Å^3.

Results and discussion‡

The refinement plan envisioned cycles of RLS until the residual indicated convergence, followed by an IG inspection of the refined atomic coordinates superimposed on electron density maps (phased by the refined coordinates) to assess the validity of and, if necessary, to change atomic positioning or to include additional atoms in RLS. If the IG changes were substantial, a repetition of RLS followed by visual inspection was carried out until the correlation between the refined atomic positions and the electron density was satisfactory. The results of the plan can be conveniently divided into six steps (Table 1): (1) 16 cycles (substeps) of RLS with starting atomic coordinates from the molecular-replacement solution (Schmidt, Girling, Houston, Sproul, Amma & Huisman, 1977); (2) locating additional atoms by means of IG; (3) five cycles of RLS, primarily to refine the new atomic coordinates; (4) rebuilding portions of the molecule by IG where atomic positions do not coincide with the electron density; (5) four cycles of RLS with the atoms of step 4; and (6) final IG examination and adjustment.

(*a*) *Restrained least-squares refinement*

(1) *Parameter initialization and in-stream changes for the restrained least-squares refinement* (*steps* 1, 3, 5, *Table* 1). The parameters in RLS which are different from conventional least-squares refinement are: the resolution limits of the X-ray intensity data set, the number of iterations for the conjugate-gradient solution, a damping factor applied to the atomic shifts to ensure a minimal residual, the weights associated with the various restrictions and a criterion for terminating the refinement.

Selecting the initial X-ray data set (step 1, Table 1) was based on the premise that lower-resolution sets give valid shifts over a larger region of real space than higher-resolution data (Konnert, 1976), that very-low-resolution data are affected by the solvent and that computer time can be saved by the use of small data sets. The data set may be small because the RLS does not depend upon maintaining a conventional ratio of X-ray intensity observations to variables. The conventional ratio does not reflect the geometric restraint terms, which also act as observables and are nearly as numerous as the variables themselves. As the refinement in step 1 progressed, the initial data set was expanded to include all 'observed' data greater than 5 Å resolution. In later steps (3, 5), more high-resolution data were included earlier because most of the atoms had been verified in the electron density.

The number of iterations to be used in the conjugate-gradient method of finding parameter shifts was determined by analysis of the parameter shift *vs* the number of iterations and set at 200. The analysis found that the shifts reached a plateau after ~200–300 iterations but by ~1500 iterations the solution had been rendered meaningless by round-off error.

The damping factor, which when applied to the shifts from the conjugate-gradient solution minimized the residual, was experimentally determined by four separate calculations with a representative sample (3%) of the intensity data. Damping factors of 0·25, 0·50, 0·75 and 1·0 were applied to the shifts and the best value was calculated by linear interpolation when the residual differences were significant. In addition, atoms deviating more than 1 Å from ideal distances after the first cycle of step 1 were removed from the coordinate list because movements of this magnitude seemed at the time unreasonable in view of the restrictions imposed on the molecular geometry. No atoms were removed from the refinement after this point and any required correction was applied by IG.

The weighting scheme used in the refinement can be related to statistical considerations which give the reciprocal of the standard deviation squared as the best weight (Jeffrey & Shiono, 1959). The estimated standard deviation of F_o was taken as $\sim\langle \Delta F \rangle$ of the previous structure-factor calculation. Estimated standard deviations for the various distance types (1–4) were set at 0·062, 0·125, 0·125 and 0·1 Å (Hendrickson, 1977), respectively, throughout the refinement. Since the W_i's remained constant through the refine-

* Each map was computed with phases in which ~1/10 of the atoms (differing atoms in each case) was omitted giving phases $\varphi_C - \varphi_{AO}$; where φ_C is the phase calculated with all atoms and φ_{AO} is the phase contribution of atoms omitted. Only that part of the map where the atoms had been deleted was used for fitting.

† This is a standard feature of GRIP; *i.e.* up to three different types of maps can be stored and displayed simultaneously.

‡ For more detail see Girling *et al.* (1978).

ment and $\langle \Delta F \rangle$ decreased from step to step (thereby increasing W_i), the intensity data took on greater weight as refinement progressed.

Estimated standard deviations for the atomic coordinates were not readily available and convergence for all steps was arbitrarily assumed if the residual was not likely to decrease by more than 0·01 in the next cycle or RLS refinement.

(2) *Performance of the restrained least-squares (RLS).* We have evaluated the results of the RLS from trends in the calculated real-space and reciprocal-space parameters, some of which are given in Table 1, and by the fit of the RLS atomic coordinates to the Fourier maps (see Fourier techniques). Upon an examination of the calculated parameters one can observe: (1) the pattern of convergence or divergence from the intensity data in reciprocal space and the restrictional data in real space; (2) the nature of the shifts in the atomic coordinates; and (3) the power or limitations of the method with respect to atomic placement.

The RLS convergence pattern for the initial intensity data set consists of large changes; in the residual and phases in reciprocal space, in the shifts in real space and an increase in the deviation from interatomic-distance idealization for the first few cycles. Later cycles show decreasing changes in the first three variables and a reassertion of the geometric idealization (see Table 1 and Fig. 2). Adding an equal number of higher-resolution intensity data points to the refinement will cause a repeat of the original pattern starting near the original R with the following exceptions: the shifts are generally smaller due to addition of higher-resolution data and the rate of change in phase angle decreases ($\langle \Delta\varphi_1 \rangle$ in Table 1) for intensity data previously refined. The ideality of the interatomic distances (*e.g.* D_B in Table 1) is not well maintained in early cycles because of the large shifts demanded by the reciprocal-space (ΔF) term in the minimization. In steps 3 and 5 the poor starting geometry of some side chains of the structure masks any trend towards an increase in D_B as was found in step 1.

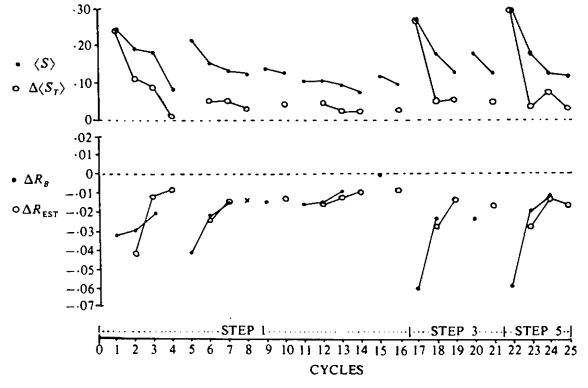

Fig. 2. Changes in different parameters associated with the progress of the RLS (see Table 1 for definition). All points connected by solid lines are associated with identical F_o intensity data sets and the same number and kinds of atoms. Some points are not included due to the large amounts of computer time required for their calculation.

The 'convergence' of the RLS refinement was determined by a change in R of less than ~0·01. The average shift at this point was ~5% of the nominal resolution of the intensity data. The final R is dependent upon the ratio W_i/W_l and easing the geometric restraint (increasing W_i/W_l) will significantly reduce the residual. The value used for W_i at the end of the refinement was approximately midway between $(1/\Delta F)^2$ and the average variance due to the counting statistics. An additional increase in the value of W_i may be justified, but a better estimate of the variance of F_o and the variance of d_c is needed to set the most realistic limits on the W_i/W_l ratio.

Since one of the purported strengths of RLS is the ability to shift atomic positions relatively large distances, it is appropriate to examine the shift magnitudes for each part of the refinement, locate those atoms which moved substantially during the refinement and measure the relative oscillation in their shifts. The pattern for average shifts (see $\langle S \rangle$ and $\langle S_T \rangle$ in Table 1) is that of large changes in the early cycles of a given step particularly with low-resolution data, followed by smaller changes in the later cycles with higher-resolution data. The shift magnitudes do not decrease as sharply during the progress of the refinement for a given step as in full-matrix least-squares and atoms moving relatively large distances (>0·5 Å) from their starting position still change significantly in the later cycles of the step (Girling, Schmidt, Houston & Amma, 1978). In the last two RLS steps (3 and 5), where the structure had previously been refined, the total average shift is about one half that of the initial step. The total $\langle S_T \rangle$ of ~0·4 Å for both steps may indicate the approximate precision of the coordinates, due in part to the absence of more high-resolution data. In contrast to collective parameters such as the average shifts, an examination (Girling, Schmidt, Houston & Amma, 1978) of the individual atoms moving >0·5 Å allows identification of specific locations in the structure which are rapidly changing. For instance, the 100 backbone atoms near the termini demonstrate a higher mobility than the remainder of the atoms, side-chain atoms show substantially greater freedom of movement than the backbone atoms. In this case large shifts were observed in the helices as well as in the turns. In the latter stages of the refinement (steps 3 and 5) there were no adjacent residues whose backbone atoms had average shifts >1 Å.

A crude measure of the change in the direction or oscillation of the parameter shifts during a given step was made by comparing the sum of the $\langle S \rangle$'s for the first n cycles in a step to the $\langle S_T \rangle$ of the nth cycle. The requirement for the ideal case of no oscillation (uni-

directional shifts) is $\sum \langle S \rangle / \sum \langle S_T \rangle = 1$ and as the ratio increases above this ideal value, oscillation increases. The $\sum_n \langle S \rangle / \langle S_T \rangle$ ratios are 2·5, 2·3 and 1·6 for steps 1, 3 and 5, respectively, indicating a tendency towards decreased oscillation with the high-resolution data.

The limitations of the RLS used here were apparent from the large number of inter- and intra-molecular non-bonded contacts less than 2·5 Å. A new version considers this problem (Hendrickson, 1978). The most serious obstacle to automated refinement *via* RLS is the detection of incorrectly placed atoms which do not violate the geometrical constraints. Moving relatively few (~5%) atoms during a reconstruction affects the progress of the structure refinement, as measured by R, as much as or more than the final two or three cycles of RLS in any given step (Table 1). It is not yet clear to what extent the final cycles of RLS serve to improve the Fourier maps derived from their results.

The power of RLS lies in its ability to improve the protein atomic positions without the over-determination normally associated with small-molecule X-ray structural refinement. A ratio of 1·3 structure amplitudes per atom is sufficient because the 'observational' restrictions provide a much more powerful means of retaining structural integrity than an equal number of X-ray data points, assuming the initial structure is nearly correct. This concept is further supported by our recent results on hemoglobin C (Paslay, Houston, Girling, Amma & Huisman, 1979). A surprising but useful result is that RLS was able to shift several misplaced atoms in two side chains up to 20 Å towards their correct positions without destroying molecular integrity.

(3) *Suggested changes for greater efficiency.* In retrospect, Table 1 suggests several alterations in either the initial parameters or their subsequent values to achieve the highest rate of convergence consistent with retention of molecular geometry and efficient use of computer time. Specifically, the convergence criterion and selection of the intensity data set may be liberalized as there is no discernable tendency of the molecular geometry to become ill-behaved during refinement.

To examine the convergence and the structural changes, plots were made of the change per cycle of the residuals, R_B and R_{EST}, of the average phase angle near 5 Å, $\langle \varphi_1 \rangle$, of the average difference between the observed and calculated structure factors, $\langle ||F_o| - |F_c|| \rangle$, and of the average shift of the atomic coordinates, $\langle S \rangle$ (see Table 1 caption for definitions). Also considered was the difference in the average total shift, $\langle S_T \rangle$, between the beginning of the step to cycle n and cycle $n-1$ (for plots of $\langle ||F_o| - |F_c|| \rangle$ and the average phase angle *vs* n near 3 Å, see Girling, Schmidt, Houston & Amma, 1978). These plots (Fig. 2) for step 1 showed that the significant changes in the structure were completed in the early cycles for a given X-ray intensity data set, suggesting the arbitrary convergence criteria of $\Delta R_B \sim 0{\cdot}01$ was too low. Raising the ΔR_B to 0·02 eliminated 'fine tuning' cycles such as 4, where the average shifts, $\langle S \rangle$, were less than half of the next cycle and cycles 7–12, where only small changes were registered in the plots of $\Delta\langle S \rangle$, $\Delta\varphi$, *etc.* Deletion of cycles 4 and 7–12 could have saved as much as 15 h CPU time, (Table 1). Steps 3 and 5 generally meet the $\Delta R_B \simeq 0{\cdot}02$ criterion except for the final cycles.

Given $\Delta R \simeq 0{\cdot}02$ as a convergence criterion, the selection of the intensity data set can be improved by doubling the number of reflections from one converged data set to the next set at higher resolution. The effect of differing sizes of intensity data sets is illustrated by steps 3 and 5. Step 3 has three cycles with low-resolution data followed by two at high resolution with a total reduction in R of ~0·08 in ~33 h CPU time* as opposed to step 5 in which four cycles with high-resolution data reduced R by ~0·09 in 42 h CPU time.* The most efficient method seems to be that used in step 3.

Care should be taken when applying these results to data where the starting coordinates are not as well determined as hemoglobin was in this case. The restraints serve to retain a generally correct original model as opposed to improving initial coordinates from a heavy-metal isomorphous-replacement phased Fourier map. Some investigators (Fermi, 1975; Deisenhofer & Steigemann, 1975) reported average shifts of up to 50° from the MIR starting phase angles during refinement as opposed to the 37° observed for Hb(DIII). It remains to be determined whether differences in the starting models implied by these phase-shift differences substantially affect the convergence of RLS.

(*b*) *Fourier techniques*

This analysis was restricted to two exhaustive studies rather than numerous attempts to define the fit of small portions of the molecule during the course of the refinement. The first study was divided into two parts. The first (step 2, Table 1) utilized in-house graphics and the second part (step 4, Table 1) used the GRIP system.

(1) *Initial Fourier structure evaluation and refinement* (*steps* 2, 4). In step 2, the first part of the initial study, the residues in which one or more atoms had not been previously located were superposed on the appropriate electron density map. The 442 missing atoms from 153 residues were located but the placement of 27 of them was at a low confidence level. No attempt was made to idealize the geometry of the additional atoms and their coordinates were simply estimated from the graphics display. The new atomic positions were displayed and some blatantly incorrect geometries were improved before step 3.

* This has now decreased in our local environment by a factor of 12. A factor of six arises from an improved algorithm and the additional factor of two because of the change from an IBM 370-168 to an IBM 3033.

For the second part of the first study, step 4, where the atoms located in step 2 had been refined, the new coordinates (step 3) and the eight Fourier maps used in step 2 were displayed on the GRIP system. Large portions of the molecule could now be manipulated with relative ease. Gross corrections (>1 Å) were made to 249 residues, but only 46 of these were so large that RLS could not make the appropriate change. The atoms not examined on GRIP were displayed and changed on the PDP-11 CRT (due to time limitations, the first half of the α_1 chain and the heme groups). The most significant changes in configuration involved long sidechains or rebuilding the chain termini, but there were a few sections (two or more adjacent residues), primarily at the chain termini, which did not fit particularly well, even after exhaustive trial and error fittings.

The quality of the electron density maps used in both steps was variable, but they yielded enough well-defined density to place ~95% of the main chain and ~60% of the side-chain atoms to within 1 Å. When superimposed on the maps the favorable aspects of the molecular structure were: the retention of geometry in spite of background noise in the electron density; less than 10% of the atoms could be shifted to a better fit in the density without a major change in configuration; and the ease of detecting ill-fitting areas in the structure. Major difficulties encountered in the fit were: the variance in the contour levels required to outline a molecular position; the absence of positive electron density surrounding the main chain (~3%) and the side chains (~10%), particularly at protein–solvent interfaces and at the ends of the protein chains; background levels which would allow alternate positions for side chains; a change in the handedness of approximately 10% of the residues (now restricted in a new version of RLS by Hendrickson, 1978); and atoms in well defined density having non-bonded contacts at variance with recognized geometry. [The new RLS is also substantially improved in this respect and we are also using the energy-minimization procedure of Hermans, Ferro, McQueen & Wei (1976) to alleviate unreasonably close contacts.] Although the parts of the eight modified F_o maps which we used may be relatively unbiased, they had a generally high noise level due to series termination errors. The noise level may have resulted in some spurious 'improvements' where only small changes were made to the atomic positioning, but the large changes (46 residues) significantly improved the compatability of the atomic positions with the map. The large number of residues adjusted (~40%) do not adequately reflect the generally correct nature of the structure.

(2) *Final Fourier study* (*step* 6). The final phase angles from step 5 were used with (ΔF) and $(2F_o - F_c)$ in order to calculate two electron density maps. The structural changes made by superimposing all four chains and heme groups on the two maps were either a residue which could be clearly shifted into a better position or short contacts which dictated small adjustments. In all, 52 residues were repositioned because they were clearly incorrect and 41 required minor adjustments because of short contacts. Particular attention in this case was paid to the fit of the atoms involved in intermolecular contacts less than 5 Å. By generating the symmetry-related atoms, the re-fitting was satisfactorily completed except for the $\beta_2(1–8) - \beta_2'(1–8)$ interaction about the twofold rotation axis. There were only a few other parts of the polypeptide chain whose positioning was unsatisfactory. The heme groups were well defined and peaks in the difference map suggest some occupancy of the sixth coordination site of the iron as would be expected from the cyanomet form of the hemoglobin. The planarity of the heme group was approximately maintained and the iron atoms showed a tendency to be nearly in the heme planes although the latter observation may be an artifact of the restraints.

The restraints on the polypeptide appeared to have achieved the desired effect except for an occasional atom which showed a tendency to deviate from its designated plane.

Finally, the overall fit was considerably better than the previous Fourier refinement. Only 51, or less than 11%, of the residues required serious retrofitting and less than 6% still needed some attention.

The electron density maps were much improved relative to the modified (F_o) maps in the previous Fourier study. The $(2F_o - F_c)$ map showed few breaks in the main chain density and most side chains (~85%) fitted readily into positive density. In contrast, the (ΔF) map ranged from relatively simple (zero) to very complicated (noisy). The background level seemed to correlate with position, the interior of the molecule being much freer of difference density than the exterior. Non-helical parts of the molecule exposed to solvent were likely to be surrounded by a hollow column of positive density. Proper adjustment of the thermal parameters probably could help to alleviate this feature.

Conclusions

RLS refinement coupled with IG offers an attractive alternative to real-space refinements. Real-space methods have a clear advantage in their relatively modest computing times (Freer, Alden, Carter & Kraut, 1975) whereas RLS–IG is more automated and converges faster. The most advantageous method is dependent upon the cost of computing *vis-à-vis* the availability of manpower, but RLS is becoming more attractive as computing power increases and its cost decreases. The success of RLS was diluted in the present refinement by the extensive use of electron density maps.

Recent experience suggests that conventional $2F_o - F_c$ and $F_o - F_c$ maps can be used efficiently in reconfiguring incorrectly placed atoms. Where these maps are ill-defined, a substantial improvement may often be obtained by deleting those atoms from the model structure which do not fit the map, refining the structure by RLS and calculating a difference map. This method minimizes bias and series termination errors.

Given the improved RLS program and present experience, we predict that an initial RLS will satisfactorily fit more than two thirds of the residues in a structure such as Hb(DIII). An exhaustive rebuilding with IG such as GRIP should reduce the structural ambiguities to less than 10% of the residues and less than 5% of the atomic positions. Further RLS can then be used to re-idealize the geometry and reduce the residual for final difference maps. Given IG and a large computer, the crystallographer should be able to complete the high-resolution refinement of a 50000 Dalton structure in one man year. A somewhat disappointing feature, however, is that RLS at this time is not a 'hands off' refinement. Inclusion of intermolecular constraints and variable atom temperature factors may improve this situation, but IG is likely to remain an important factor in such refinements for some time.

We wish to express our appreciation to J. H. Konnert and W. A. Hendrickson for the restrained least-squares program and their advice concerning its use. We also wish to thank the University of South Carolina Computer Services Division and, in particular, L. Zarpentine for program scheduling. The major portion of the interactive graphics results were derived with the aid of the GRIP-75 molecular graphics system built by the University of North Carolina Department of Computer Science with the help of many users. Major system contributions were made by E. G. Britton, J. S. Lipscomb and M. E. Pique, under the direction of W. V. Wright, J. E. McQueen Jr and J. Hermans. F. P. Brooks Jr is Principal Investigator. The system has been developed under NIH Biotechnology Research Resource grant No. RR00898, NSF grant No. GJ-34697, AEC contract No. AT-(40-1)-3817, and the IBM Corporation. Day-to-day administrative assistance on GRIP was rendered by J. Leonarz and real-time programming handled by W. V. Wright, M. Pique, J. Lipscomb, W. Siddell and D. Voss. The PDP-11 stereo display was derived from a program kindly supplied by B. Wishner and Dr W. E. Love. S. B. Belk and E. Reeves provided local software support and advice on interfacing the various computer programs. This work was supported by NIH grant HL-15158.

References

ANDERSON, C., STENKAMP, R. & STEITZ, T. (1978). *J. Mol. Biol.* **123**, 15–33.

BLUNDELL, T. L. & JOHNSON, L. N. (1976). *Protein Crystallography*, pp. 420–442. New York: Academic Press.

COLLINS, D. M., COTTON, F. A., HAZEN, E. E. JR, MEYER, E. F. JR & MORIMOTO, C. N. (1975). *Science*, **190**, 1047–1053.

DEISENHOFER, J. & STEIGEMANN, W. (1975). *Acta Cryst.* B**31**, 238–250.

DIAMOND, R. (1966). *Acta Cryst.* **21**, 253–266.

DIAMOND, R. (1974). *J. Mol. Biol.* **82**, 371–391.

FERMI, G. (1975). *J. Mol. Biol.* **97**, 237–256.

FREER, S. T., ALDEN, R. A., CARTER, C. W. JR & KRAUT, J. (1975). *J. Biol. Chem.* **250**, 46–54.

GIRLING, R. L., SCHMIDT, W. C. JR, HOUSTON, T. E. & AMMA, E. L. (1978). Tech. Rep. No. A01 XRAY REF. PPOT, Department of Chemistry, University of South Carolina, Columbia, SC 29208, USA.

GRIP (1975). GRIP-75 molecular graphics system built at the University of North Carolina. See acknowledgements.

HENDRICKSON, W. A. (1977). Personal communication. *Lists Compiled from the Residues of Diamond* (1974) *and Small Molecule X-ray Structures.*

HENDRICKSON, W. A. (1978). *Restrained Least-Squares Program Package.* August 1978.

HENDRICKSON, W. A. & KONNERT, J. H. (1979). *Madras Symposium on Biomolecular Structure, Conformation, Function and Evolution.* In the press.

HERMANS, J. JR, FERRO, D., MCQUEEN, J. JR & WEI, S. C. (1976). *Environmental Effects on Molecular Structure and Properties*, edited by B. PULLMAN, pp. 459–483. Dordrecht, Holland: D. Reidel.

HERMANS, J. JR & MCQUEEN, J. E. JR (1974). *Acta Cryst.* A**30**, 730–739.

HINGERTY, B., BROWN, R. & JACK, A. (1978). *J. Mol. Biol.* **124**, 523–534.

HUBER, R., KUKLA, D., BODE, W., SCHWAGER, P., BARTELS, K., DEISENHOFER, J. & STEIGEMANN, W. (1974). *J. Mol. Biol.* **89**, 73–101.

JEFFREY, G. A. & SHIONO, R. (1959). *Acta Cryst.* **12**, 819.

KONNERT, J. H. (1976). *Acta Cryst.* A**32**, 614–617.

LADNER, R. C., HEIDNER, E. J. & PERUTZ, M. F. (1977). *J. Mol. Biol.* **114**, 385–414.

LEVITT, M. (1974). *J. Mol. Biol.* **82**, 393–420.

LEVITT, M. & LIFSON, S. (1969). *J. Mol. Biol.* **46**, 269–279.

MOEWS, P. & KRETSINGER, R. (1975). *J. Mol. Biol.* **91**, 201–228.

PASLAY, R. C., HOUSTON, T. E., GIRLING, R. L., AMMA, E. L. & HUISMAN, T. H. J. (1979). *Am. Crystallogr. Assoc. Winter Meeting, Honolulu.* Abstract K7.

RASSE, D., WARME, P. & SCHERAGA, H. (1974). *Proc. Natl Acad. Sci.* **71**, 3736–3740.

RICHARDS, F. M. (1968). *J. Mol. Biol.* **37**, 225–230.

SCHMIDT, W. C. JR, GIRLING, R. L. & AMMA, E. L. (1977). *Acta Cryst.* B**33**, 3618–3620.

SCHMIDT, W. C. JR, GIRLING, R. L., HOUSTON, T. E., SPROUL, G. D., AMMA, E. L. & HUISMAN, T. H. J. (1977). *Acta Cryst.* B**33**, 335–343.

SUSSMAN, J. L., HOLBROOK, S. R., CHURCH, G. M. & KIM, S. H. (1977). *Acta Cryst.* A**33**, 800–804.

SUSSMAN, J. L., HOLBROOK, S. R., WARRANT, W., CHURCH, G. M. & KIM, S. H. (1978). *J. Mol. Biol.* **123**, 607–630.

WATENPAUGH, K., SIEKER, L. C., HERRIOTT, J. & JENSEN, L. H. (1973). *Acta Cryst.* B**29**, 943–956.