

Application of Automated Structure Analysis to Some Organic Compounds Using PCs.

1. Automated Patterson Techniques

Kenji Okada

Research and Development Center, Ricoh Company Ltd., Kouhoku-ku, Yokohama 223, Japan

Received June 28, 1993*

When organic compounds include sulfur or a heavier atom, heavy-atom methods are usually used to determine the molecular structure. We describe an improved PC version of the PSL+SEARCHER programs. PSL+SEARCHER is based on the classical Patterson method which solves the Patterson function, selects out one heavy-atom position and finds all remaining atom positions automatically without any chemical assumption or human intervention. The improved features present are described in detail, and a few selected examples calculated on PCs demonstrate the power of PSL+SEARCHER programs. These programs comprise the DS*SYSTEM^{1,2} and are suitable for use by organic chemists as well as professional crystallographers.

INTRODUCTION

In structure determination of organic compounds which include a heavy atom, the atomic coordinates can be obtained by solving a Patterson function. Several methods have been proposed during more than 50 years, and relevant computer programs have been developed. When a part of the molecular geometry is known, the following programs have proved valuable: combination with Patterson-search method and direct methods,^{3,4} automated real space Patterson rotation search PATSEEP,⁵ vector search and feedback methods for macromolecules,⁶ intermolecular Patterson vectors,⁷ PATMET using partial-fragment rotation function,⁸ ROTSEARCH using rotation search,⁹ and automatic Patterson deconvolution by superposition method XFPS89 based on the symmetry minimum function.¹⁰ When there is no information about known structure fragments, the following techniques have been used: use of overlapping cross vectors for the determination of the position of two atoms by the superposition method,¹¹ IMPAS using generalized Harker vectors,⁷ and an automated procedure to search the difference Patterson function HASSP.¹²

The first version of SEARCHER for automatic structure analysis was developed on the mainframe CDC3600/CDC6600,^{13,14} and to that was added PSL¹⁵ which solved Patterson function automatically. These programs have been improved upon the Cray-1 version by adding vector facilities and optimization,¹⁶ rewritten in Fortran 77 and implemented on 32-bit personal computers (PCs) with several modifications and a bug fix.^{1,2} The PC 1993 version of PSL+SEARCHER has been improved to change the algorithms of peak summation, achieve a reduction of calculation time, and better conform to user requirements. Structure analysis of more than 50 organic compounds has been carried out by PSL+SEARCHER programs. Once several atom positions composing molecular fragment are determined, SEARCHER can be carried out to find all atom positions subsequently.

DESCRIPTION

Theoretical Background. The diffraction of X-ray by crystals occurs by the vibration of electrons which consist of the crystals for the incident upon X-ray beam. The angles θ of incidence and reflection depend on the repetitive nature of

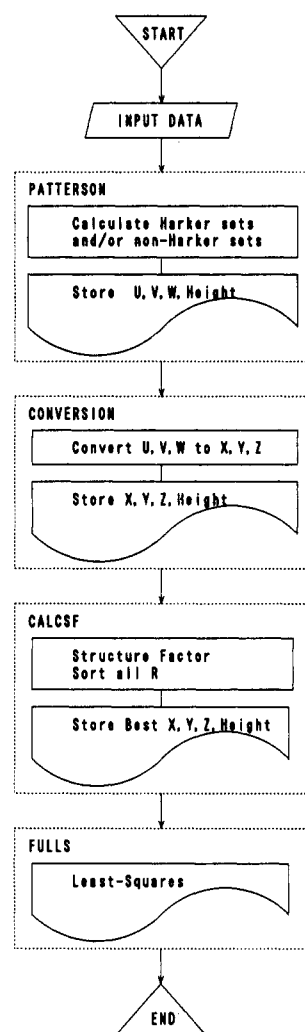


Figure 1. Simplified PSL program layout of calculation. PSL calculates Patterson functions, solves Patterson matrices, calculates structure factor, and refines a heavy-atom position by full-matrix least-squares method automatically.

the lattice and the positions \mathbf{r} of the atom. This is Bragg's law expressed as eq 1 by the scalar products of wavelength λ and

$$\frac{2 \sin \theta}{\lambda} = \frac{1}{d(\mathbf{h})} = \mathbf{h} \cdot \mathbf{r} \quad (1)$$

the vector \mathbf{h} with components h, k, l in reciprocal space, where

* Abstract published in *Advance ACS Abstracts*, September 15, 1993.

Table I. Molecular Properties of the Structures Solved by Using PSL + SEARCHER^a

	I	II	III	IV	V	VI	VII
A. crystal data							
formula	C ₁₉ H ₂₀ O ₂ N ₃ SBr	C ₁₉ H ₃₀ O ₂ S	C ₂₆ H ₃₇ O ₇ Br	C ₁₉ H ₂₂ O ₃ S	C ₂₃ H ₂₀ O ₃ S	C ₁₄ H ₁₇ N ₃ O ₃ S	C ₂₇ H ₄₃ NOHI
mol wt	462.4	322.5	565.5	330.44	411.49	307.30	525.55
space group	P2 ₁ /c	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁	P1	Pna2 ₁	C2/c	P2 ₁
Z	4	4	4	2	4	8	2
a (Å)	12.46(3)	22.806(1)	15.746(1)	9.127(1)	7.979(1)	23.186(1)	13.722(1)
b (Å)	13.510(3)	6.276(1)	13.622(1)	11.955(1)	27.459(1)	8.654(1)	8.229(1)
c (Å)	12.006(3)	11.978(1)	12.203(1)	8.783(1)	9.330(1)	16.664(1)	11.765(1)
α (deg)	90.0	90.0	90.0	93.08(1)	90.0	90.0	90.0
β (deg)	93.45(1)	90.0	90.0	117.70(1)	90.0	116.60(2)	100.27(2)
γ (deg)	90.0	90.0	90.0	85.49(1)	90.0	90.0	90.0
V (Å ³)	2017.4(2)	17144(4)	2617.4(4)	845.7(1)	2044.1(4)	2989.9(8)	1307.2(2)
D _m (g·cm ⁻³)	1.519	1.334	1.428	1.282			1.324
D _x (g·cm ⁻³)	1.522	1.250	1.435	1.298	1.327	1.365	1.335
B. problem	molecular conformation	←	←	structure unknown	←	←	←
C. refinement	full-matrix	←	block-matrix	full-matrix	←	←	←
temp factor	all-aniso	←	←	←	←	←	←
R	0.053	0.057	0.068	0.058	0.080	0.083	0.065
D. experiment							
λ	Mo Kα	←	←	←	←	←	←
scan range	θ ≤ 27.5°	←	←	←	←	←	←
scan technique	ω/2θ	←	←	←	←	←	←
filter	Zr	←	←	←	←	←	←
obs reflcn							
≥ 0σ	2117	2065	2725	3072	1855	2691	2523
≥ 1σ	1900	2061	2191	3038	1813	2581	2362
≥ 3σ	1898	1743	1692	2468	1214	1970	1734
E. ref	X-83	X-111	X-114	X-144	X-149	X-153	X-158

^a I, *p*-bromophenyl-*N*-carbamoylthiamine anhydride; II, 17β-hydroxy-5α-androstan-2α,3α-episulfoxide; III, 3-*o*-(*p*-bromobenzoyl)sarcostine; IV, 17β-acetoxy-18-methyl-1-thia-4-norestra-3,9(10),11-trien-2-one; V, 3-benzoyl-1a,2-dimethoxycarbonyl-1-methylthio-1a,7b-dihydrocyclopropa[*a*]naphthalene; VI, 6-ethyl-4-methylthio-5-oxo-5,6-dihydrohydrido[2,3-*d*]byridazine-3-carboxylate; VII, shinonomenine hydriodide.

Table II. Computational Results of the Structures Solved by PSL + SEARCHER

compd	space group	PSL							SEARCHER							
		Harker sets no. of peaks		possible atomic coordinates no. of coordinates			R of LS groups ^a	heavy atom R	NSF	NLH	NLS	part A ^b cycle R	part B ^c cycle R	atoms		
X-83	P2 ₁ /c	[0,v, ¹ / ₂] 4	[u, ¹ / ₂ ,w] 22		(y) 2	(x,z) 12		24	0.460–0.576 4	Br lowest	1	5	38	28 8	28 5	28/28
X-111	P2 ₁ 2 ₁ 2 ₁	[¹ / ₂ ,v,w] 10	[u, ¹ / ₂ ,w] 60	[u,v, ¹ / ₂] 50	(y,z) 12	(x,z) 18	(x,y) 24	29	0.517–0.556 5	S lowest	1	5	30	0.241 30 8	0.147 22 5	22/22
X-114	P2 ₁ 2 ₁ 2 ₁	[¹ / ₂ ,v,w] 55	[u, ¹ / ₂ ,w] 47	[u,v, ¹ / ₂] 55	(y,z) 30	(x,z) 18	(x,y) 50	69	0.444–0.571 9	Br lowest	1	5	45	0.348 45 10	0.163 36 6	36/36
X-144	P $\bar{1}$	[u,v,w] 79			(x,y,z) 46			38	0.561–0.609 38	S lowest	1	5	31	0.263 31 6	0.182 23 6	23/23
X-149	Pna2 ₁	[¹ / ₂ ,v,0] 11	[u, ¹ / ₂ , ¹ / ₂] 5	[u,v, ¹ / ₂] 49	(x) 6	(y) 10	(x,y) 60	33	0.563–0.637 12	S lowest	1	5	39	0.387 39 9	0.154 26 14	26/28
X-153	C2/c	[u,0,w] 73	[0,v, ¹ / ₂] 5		(y) 9	(x,z) 62		223	0.583–0.630 20	S 9th	1	5	28	0.381 28 7	0.307 21 6	21/21
X-158	P2 ₁	[u, ¹ / ₂ ,w] 49			(x,z) 46			20	0.294–0.523 9	I lowest	1	5	40	0.390 40 8	0.178 30 7	30/30
														0.213	0.164	

^a Groups: number of groups after LS. ^b part A: number of obtained atoms at part A. ^c part B: number of obtained atoms at part B.

θ = Bragg angle and $d(\mathbf{h})$ = the interplanar spacing in the crystal lattice. The intensity $I_0(\mathbf{h})$ passes through the reflecting angle is related experimentally to the square of the structure factor amplitude $F_0(\mathbf{h})$ and expressed as follows:

$$I_0(\mathbf{h}) \approx |F_0(\mathbf{h})|^2 \quad (2)$$

The structure factor $F_0(\mathbf{h})$ is the integral over its volume of the product of the electron density $\rho(\mathbf{r})$ and the dot product

$\mathbf{h} \cdot \mathbf{r}$, and is expressed as follows:

$$F_0(\mathbf{h}) = \int \rho(\mathbf{r}) \exp(2\pi i \mathbf{h} \cdot \mathbf{r}) d\mathbf{v} \quad (3)$$

The electron density $\rho(\mathbf{x})$ is expressed by eq 4 as the Fourier transform of the structure factor $F_0(\mathbf{h})$ defined by eq 3, where

$$\rho(\mathbf{x}) = \int F_0(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}) d\mathbf{v}^* \quad (4)$$

\mathbf{x} = the coordinates in real space and $d\mathbf{v}^*$ = any volume element

in reciprocal space defined \mathbf{h} . In reciprocal space the structure factor $F_0(\mathbf{h})$ is not continuous and has only the value at the reciprocal vector \mathbf{h} . The electron density $\rho(\mathbf{x})$ is rewritten in terms of Fourier summation of \mathbf{h} as follows:

$$\rho(\mathbf{x}) = \frac{1}{V} \sum_{\mathbf{h}} |F_0(\mathbf{h})| \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}) \quad (5)$$

where V = the unit cell volume. Namely, the structure factor expressed in eq 3 is calculated by using the atomic position in the unit cell and the phase angles. But there is no way to determine the phase angle in general methods. Two main methods are employed: heavy-atom analysis and direct methods. In crystallography the measurement data consist only of the structure magnitudes and do not have their phases.

The data read to solve the structure of organic compound include cell constants, space group, number of chemical unit, kind of atom and number of atoms in unit cell, and measured intensity data $I_0(\mathbf{h})$ defined by eq 2.

If the organic compound includes sulfur or a heavier atom, heavy-atom methods are used to solve the Patterson function. The Patterson function can be expressed as follows:

$$P(\mathbf{u}) = \frac{1}{V} \sum_{\mathbf{h}} |F_0(\mathbf{h})|^2 \exp(2\pi i \mathbf{h} \cdot \mathbf{u}) \quad (6)$$

where \mathbf{u} = the coordinate vector in Patterson space (u, v, w) and V = the volume of the unit cell. The heavy-atom position can be obtained by using the methods such as Harker/non-Harker sets, minimum function, space replacement method, and anomalous dispersion method. The calculation of structure factor $F_c(\mathbf{h})$ expressed by eq 7 is performed using heavy

$$F_c(\mathbf{h}) = \sum_{j=1}^N f_j(\mathbf{h}) T_j(\mathbf{h}) \exp(2\pi i \mathbf{h} \cdot \mathbf{r}_j) \quad (7)$$

atom coordinates \mathbf{x} (x, y, z), where N = the number of atoms in a unit cell, $j = 1, 2, \dots, N$, $f_j(\mathbf{h})$ = the atomic scattering factor of the j th atom, and $T_j(\mathbf{h})$ = the temperature factor of the j th atom. Once approximate phases are determined, we are able to calculate the electron density $\rho(\mathbf{x})$ expressed as follows:

$$\rho(\mathbf{x}) = \frac{1}{V} \sum_{\mathbf{h}} F_c(\mathbf{h}) \exp(-2\pi i \mathbf{h} \cdot \mathbf{x}) \quad (8)$$

From the electron density $\rho(\mathbf{x})$ map, more correct coordinates are calculated by considering peak heights, atomic distances, and angles. Thus we continue the calculation of $F_c(\mathbf{h})$ and $\rho(\mathbf{x})$ until the molecular structure is apparent. Next we refine the atomic coordinates by using least-squares method and minimize the discrepancy R_w expressed by eq 9, where $w(\mathbf{h})$ = the weight and k = the scale factor to $F_c(\mathbf{h})$, $\sum |F_0(\mathbf{h})| = k \sum |F_c(\mathbf{h})|$, to get more accurate positions and to find out ghost atoms.

$$R_w = \sum_{\mathbf{h}} w(\mathbf{h}) (|F_0(\mathbf{h})| - k|F_c(\mathbf{h})|)^2 \quad (9)$$

PSL+SEARCHER. In the case of heavy-atom methods, two programs are used: PSL solves the Patterson function automatically, and SEARCHER finds all remaining atom positions. For PSL and SEARCHER programs, we improved on the input and output file assignments by adopting interactive facilities and, on the optimization of DO-loop sequences in Fourier summation, structure factor calculation and least-squares refinement.

PSL: Solve Patterson Function Automatically. The flow chart of PSL (Patterson, structure factor and least-squares methods) is shown in Figure 1. Input data are as follows: the cell constants, the names of real and Patterson space, the symbol of the heavy atom, the reflection data [$h, k, l, F_0(\mathbf{h}), \sigma(I_0(\mathbf{h}))$], and Harker sets (sections and lines) to calculate Patterson functions. To use more accurate data, the standard deviation $\sigma(I_0(\mathbf{h}))$ of each reflection should be satisfied $\geq 1\sigma$.

(a) Calculate Patterson Matrix. In real space the equivalent position \mathbf{x}_n of the coordinate \mathbf{X} is written as follows:

$$\mathbf{x}_n = \mathbf{R}_n \cdot \mathbf{X} + \mathbf{r}_n \quad (10)$$

where \mathbf{x}_n = the column vector of symmetry n , $(x_n, y_n, z_n)^T$, \mathbf{R}_n = the rotation matrix of symmetry n , $(\mathbf{R}_{ij}, 3 \times 3)$, \mathbf{X} = the coordinate vector $(x, y, z)^T$, and \mathbf{r}_n = the translation vector of symmetry n , $(r_1, r_2, r_3)^T$.

And the Patterson column vector \mathbf{u}_n of symmetry n is defined as follows in Patterson space:

$$\mathbf{u}_n = \mathbf{P}_n \cdot \mathbf{X} + \mathbf{p}_n \quad (11)$$

where \mathbf{u}_n = the Patterson column vector of symmetry n , $(u_n, v_n, w_n)^T$, \mathbf{P}_n = the Patterson rotation matrix of symmetry n , $(=\mathbf{R}_i - \mathbf{R}_j, 3 \times 3)$, and \mathbf{p}_n = the Patterson translational vector of symmetry n , $(=r_i - r_j)$. These Patterson matrix \mathbf{P}_n and vector \mathbf{p}_n depend upon each space group and are determined by the difference vector (intermolecular vector) of two equivalent positions i and j of the heavy atom. We first calculate the Patterson function for all Harker sets. For example, the calculation ranges of Harker section $[u, 1/2, w]$ are from 0.0 to 1.0 for u and w in fractional coordinates. The calculation step should be less than 0.25 Å for Δu and Δw in crystal unit for peaks that may overlap and weak vectors that may be missing. To separate the Patterson peak clearly, $P(\mathbf{u})$ is often multiplied by the coefficient $\exp(B \sin^2 \theta / \lambda^2)$. The map is known as a sharpened Patterson. Other functions can also be used, for example E values. The PSL stores the local maximum peaks as the positions $\mathbf{u}_n(u, v, w)$ with peak height (H).

(b) Convert u, v, w to x, y, z . To obtain a solution for \mathbf{X} of eq 11 all peaks (u, v, w, H) of Harker vectors are converted to coordinates (x, y, z) as follows:

$$\mathbf{X} = \frac{\mathbf{u}_n - \mathbf{p}_n}{\mathbf{P}_n} \quad (12)$$

For example, the Harker section $[u, 1/2, w]$ in space group $P2_1$ means atomic coordinates $(2x, 1/2, 2z)$. All Harker vectors are converted to $(x), (y), (z), (x, y), (x, z), (y, z)$, or (x, y, z) coordinates with peak height by solving eq 12. PSL makes the combinations of these coordinates, removes values which are the same, and stores these combinations as the possible atom (x, y, z, H) . Some space groups have the coordinate free (as y coordinate in $P2_1$); then PSL sets that coordinate to 0.123. At this stage PSL removes the centers of symmetry such as $(0,0,0), (1/2, 1/2, 1/2), \dots, (1,1,1)$ and equivalent positions.

(c) Calculate Structure Factor and Refine Atom Positions (IS). PSL calculates the structure factors with isotropic temperature factor (B) for all possible atom positions and gets the reliability factor R (R factor) expressed as follows:

$$R = \frac{\sum |k|F_0(\mathbf{h})| - |F_c(\mathbf{h})||}{\sum |F_0(\mathbf{h})|} \quad (13)$$

where k = the scale factor to $F_0(\mathbf{h})$, $k \sum |F_0(\mathbf{h})| = \sum |F_c(\mathbf{h})|$. Once more PSL collects up the equivalent atom positions,

Table III. Comparison of the Computer Speed Using PSL + SEARCHER

compd	formula space group	PSL ^a			SEARCHER ^a				
		Cray-1	80386 (20 MHz)	i486 + ODP (25 MHz)	CDC3600	CDC6600	Cray-1	80386 (20 MHz)	i486 + ODP (25 MHz)
X-83	C ₁₉ H ₂₀ O ₂ N ₅ SBr	5.11 _s	4 _m 10 _s	45 _s	1 _k 29 _m 24 _s	9 _m 33 _s	45.8 _s	57 _m 23 _s	10 _m 00 _s
	P2 ₁ /c	0.144	5.556	1.000	8.940	0.655	0.076	5.738	1.000
X-111	C ₁₉ H ₃₀ O ₂ S	8.95 _s	6 _m 12 _s	1 _m 00 _s		22 _m 03 _s	59.3 _s	1 _k 29 _m 39 _s	15 _m 35 _s
	P2 ₁ 2 ₁ 2 ₁	0.149	6.200	1.000		1.415	0.063	5.753	1.000
X-114	C ₂₈ H ₃₇ O ₇ Br	8.49 _s	16 _m 40 _s	2 _m 33 _s		21 _m 01 _s	91.5	1 _k 40 _m 00 _s	17 _m 10 _s
	P2 ₁ 2 ₁ 2 ₁	0.055	6.536	1.000		1.224	0.089	5.825	1.000
X-144	C ₁₉ H ₂₂ O ₃ S	11.96 _s	20 _m 09 _s	3 _m 15 _s		14 _m 36 _s	36.4 _s	38 _m 14 _s	12 _m 42 _s
	P1	0.061	6.200	1.000		1.150	0.074	3.010	1.000
X-149	C ₂₃ H ₂₀ O ₅ S	7.52 _s	7 _m 14 _s	1 _m 10 _s		20 _m 42 _s	107.4 _s	2 _k 09 _m 59 _s	22 _m 33 _s
	Pna2 ₁	0.107	6.200	1.000		0.918	0.079	5.764	1.000
X-153	C ₁₄ H ₁₇ N ₃ O ₃ S	23.30 _s	25 _m 58 _s	4 _m 22 _s		22 _m 18 _s	53.3 _s	1 _k 33 _m 17 _s	16 _m 36 _s
	C2/c	0.089	5.947	1.000		1.343	0.054	5.619	1.000
X-158	C ₂₇ H ₄₃ NOHI	3.60 _s	3 _m 22 _s	32 _s		19 _m 44 _s	94.4 _s	47 _m 23 _s	18 _m 05 _s
	P2 ₁	0.113	6.313	1.000		1.091	0.087	2.620	1.000
	ratio	0.098	6.136	1.000	8.940	1.114	0.075	4.904	1.000

^a Upper value: actual CPU time. ^b Lower value: ratio.

stores independent atom positions, and prints the possible atom positions with *R*-factor in increasing order. Each parameter (*x*, *y*, *z*, *B*) of these independent possible atoms is refined to minimize *R_w* of eq 9 with the least-squares method by full matrix using the isotropic temperature factor and 3–5 iteration cycles. And PSL sorts the final *R* factor in increasing order, so a heavy atom is located in the lower part of the *R* factor.

SEARCHER: Find All Atom Positions Automatically. The flow chart of the SEARCHER (SEARCH all atom positions) program was shown in the previous paper.¹³ SEARCHER consists of part A (read input data, calculate structure factor and Fourier synthesis, and scan Fourier peaks) and part B (select the possible atom positions by least-squares method and draw the molecules of two-dimensional projection with final atomic coordinates). SEARCHER reads the data in three sets. The first contains the cell constants, the name of the space group, the kinds of atoms, the heavy-atom position (*x*, *y*, *z*), and the reflection data satisfying its standard deviation greater than 3σ(*I*₀(*h*)). The temperature factor for a heavy atom usually uses a standard value 4.5 Å² for Ag, I, Br atoms and 2.0 Å² for Cl, S atoms, respectively. An amount of 1.3 times the number of one molecule is usually used as the total number of atoms (NLS) due to the ghost atoms included during the selection process. The second set of data includes the indicator of Fourier axes, calculation range of Fourier synthesis, minimum value of the ρ map (RHOMIN), scale factor for ρ(*x*), and the additional number of atoms (NLH) for the structure factor calculation of the next cycle. Usually NLH employs 5 as the odd number. The third set contains the indicator of restrictions of the space group, the number of cycles, the damping factors, and the maximum temperature factor (BCUT) to remove the ghost atoms from the atom list for the least-squares method.

(a) Calculation of the Structure Factor. In the initial stage, the number of input atoms (NSF) is always one, *viz.*, a heavy atom only. The calculation of the structure factor uses the controlled reflection data by the combination of NSF and sin θ/λ. The sin θ/λ of each reflection is usually less than 0.4 or 0.45 Å⁻¹ in the first step. This is the reason why NSF is small in the initial stage, since we want to use highly reliable reflections of low values of sin θ/λ. SEARCHER calculates the structure factor, and if NSF is less than NLS, part A is continued; otherwise (NSF ≥ NLS) part B is started.

(b) Fourier Synthesis. The calculation of three-dimensional Fourier synthesis is performed in an asymmetric unit at the predetermined intervals (Δ*x*, Δ*y*, Δ*z*) that should be less than

0.1–0.25 Å to avoid missing peaks. The selection of the first axis in Fourier synthesis is important. This axis should satisfy one of the following conditions: the shortest cell constant or the angle of the other two axes is nearly equal to 90°. The scale factor for ρ(*x*) (SFOUR) usually uses 100.0 for the convenient peak selection.

(c) Peak Selection. The coordinates (*x*, *y*, *z*) and peak height (*H*) of the center of the Fourier peak are calculated using a suitable second-order polynomial by Lagrangian interpolation from twenty-seven values in three neighboring sections of three-dimensional electron density distributions. When the distance of each peak between the neighboring Fourier sections is less than SAMAX (0.5 Å), those peak pairs will be the same peak. The limitation in peak selection was improved from a predetermination to an automatic adjustment. Compared with the peak heights and RHOMIN (input), the number of peaks is adjusted by revising the initial RHOMIN. SEARCHER removes the equivalent peaks, summarizes the same peaks, sorts the peak heights in descending order, and stores the individual peaks. And, SEARCHER checks the distance between input atoms and whether peaks are reasonable and/or able to be additively. The selection of the atomic sites is based on the "reasonable" interatomic distances (1.1–1.7 Å) between neighboring low atomic number peaks. In the case of the first calculation of part A, that is, NSF = 1, the reasonable distance is replaced as the range of the ghost peak. SEARCHER then skips these peaks and adds the next higher peaks as new additional atom positions until NLH atoms are revealed. This usually means a heavy atom does not have a bond connection. Then SEARCHER stores the reasonable atom positions to expand other remaining atom positions. The new additional atoms are treated as a carbon atom with a temperature factor (*B*) equals to 3.5 Å². Then NSF is added to NSF + NLH for the next Fourier calculations.

(d) Part A. After new NSF atoms are determined, SEARCHER calculates structure factors once more. In the following cycles, additional atoms (NLH) must satisfy the reasonable distance condition. At this stage if NSF is less than NLS, part A is continued; otherwise (NSF ≥ NLS or NSF_{new} ≤ NSF_{old}) part B is started.

(e) Least-Squares Method. Since the NLS atoms selected in part A may include several ghost atoms, the SEARCHER minimizes *R_w* by refining the parameters (*x*, *y*, *z*, *B*, scale factor, overall temperature factor) with the least-squares method by block-diagonal matrix until ghost atoms are removed. The selection of the atomic sites in part B is based only on the

• Example Input data of PSL

```

-----1-----2-----3-----4-----5-----6-----7-----8
* X-83(ALPHA) C19.H20.02.N5.S.BR P21/C *
12.460 13.510 12.006 0.0 -0.060177 0.0
P21/C BR P2/M 000
      1
34 (4(3I3,F8.4,I1))
PAT 1 1
1 50 0 0 2 100 0 100 3 50 25 25
PAT 1 1
2 50 25 25 3 100 0 100 1 100 0 100
SF 1 4.5
LS 1 4.5
END
-----1-----2-----3-----4-----5-----6-----7-----8

```

• Example Input data of SEARCHER

```

-----1-----2-----3-----4-----5-----6-----7-----8
* X-83(ALPHA) C19.H20.02.N5.S.BR P21/C *
12.460 13.510 12.006 0.0 -0.060177 0.0
P21/C BR C
      3 1 0.40 38 0.60
34 1(4(3I3,F8.4,I1))
1 38 1
BR19 1 0.1132 0.1469 0.2001 1
4.5
      30 C-- 2
2 100 0 50 3 100 0 50 1 100 0 100
1 2.0 1 2 3 1.5
-----1-----2-----3-----4-----5-----6-----7-----8

```

• Example Reflection data : $h, k, l, F_o(h), \sigma(I_o(h))$

```

-----1-----2-----3-----4-----5-----6-----7-----8
0 0 2 47.45973 0 0 4 19.11983 0 0 6 40.35063 0 0 8 29.86753
0 0 10 14.03453 0 0 12 17.94223 0 1 1 31.70183 0 1 2 52.41983
0 1 3 81.06263 0 1 4 73.73073 0 1 5 6.24023 0 1 6 34.43743
0 1 7 36.80253 0 1 10 10.67143 0 1 11 15.44563 0 2 0 50.76813
0 2 1 32.27683 0 2 2 57.51923 0 2 3 74.29393 0 2 4 10.10883
. . .
. . .
. . .
13 2 3 8.91923 13 2 -1 16.79043 13 2 -3 10.85323 13 3 3 9.97643
13 4 0 11.00873 13 4 2 11.07683 13 4 -1 8.22263 13 4 -3 7.10833
13 4 -4 8.09563 13 4 -5 6.78163 13 5 2 8.46403 13 5 4 12.35123
13 5 -4 11.33893 13 6 0 11.31073 14 0 0 10.06363 14 0 -2 7.70283
14 1 -2 9.58073 14 2 1 9.81593 14 2 3 12.00903 14 3 -3 7.30383
99999999999999999999
-----1-----2-----3-----4-----5-----6-----7-----8

```

Figure 2. Input data of PSL and SEARCHER programs. PSL reads the title, the cell constants, the names of real space and Patterson space, the symbol of the heavy atom, reflection data, and Harker sets to calculate Patterson functions. SEARCHER reads the title, the cell constants, the name of space group, reflection data, a heavy-atom position, and an asymmetric unit for Fourier synthesis. The standard deviation of each reflection should be satisfied $\geq \sigma(I_o(h))$.

atomic temperature factor. SEARCHER determines whether atoms are to be eliminated ($B > BCUT$) for the next least-squares calculations, removes these atoms, and refines these parameters until two cycles produce no elimination. We usually use the BCUT which is 10.0 \AA^2 for Br, Ag, I and 8.0 \AA^2 for Cl, S.

(f) 2D Projection. The outputs of SEARCHER are two-dimensional projection diagrams of molecules in the unit cell with the scale of $1 \text{ \AA} = 2.54 \text{ cm}$ and print the results of the atomic coordinates.

(g) Part B. The atomic coordinates (x, y, z) of all equivalent positions are punched out and used as input data to the next cycles.

APPLICATIONS AND RESULTS

We summarize the results of molecular properties of seven organic compounds solved by PSL+SEARCHER in Table I. Table II shows a summary of the results in the Harker sets, possible atomic coordinates, the order of heavy atom by PSL,

```

* X-83(Alpha) C19.H20.O2.N5.S.Br P21/C *
DATE = 1993- 6- 1, TIME = (11:16:18), PROGRAM = PSL

A      = 12.46000 (A),      A*      = .080403 (A)
B      = 13.51000 (A),      B*      = .074019 (A)
C      = 12.00600 (A),      C*      = .083443 (A)
COS(A) = .00000,      ALPHA = 90.000 (Deg),      COS(A*) = .000000,      ALPHA* = 90.000 (Deg)
COS(B) = -.06018,      BETA  = 93.450 (Deg),      COS(B*) = .060177,      BETA*  = 86.550 (Deg)
COS(G) = .00000,      GAMMA = 90.000 (Deg),      COS(G*) = .000000,      GAMMA* = 90.000 (Deg)
SIN(A) = 1.000000,      SIN(A*) = 1.000000
SIN(B) = .998188,      SIN(B*) = .998188
SIN(G) = 1.000000,      SIN(G*) = 1.000000
VOLUME = 2017.36 (A**3),      VOLUME* = .0004957 (A**3)

SPACE GROUP NCENT NEQV NEQR
P21/C      1      2      4

CENTRO SYMMETRIC

SYMMETRY OPERATION
      TRANSFORMED X      TRANSFORMED Y      TRANSFORMED Z
1..  .000  1.  0.  0.      .000  0.  1.  0.      .000  0.  0.  1.
2..  .000  1.  0.  0.      .500  0. -1.  0.      .500  0.  0.  1.
3..  .000 -1.  0.  0.      .000  0. -1.  0.      .000  0.  0. -1.
4..  .000 -1.  0.  0.      .500  0.  1.  0.      .500  0.  0. -1.

PATTERSON SPACE GROUP = P2/M
PATTERSON EQUIVALENT POSITIONS
      U      V      W
.0000  1.  0.  0.      .0000  0.  1.  0.      .0000  0.  0.  1.
.0000  1.  0.  0.      .0000  0. -1.  0.      .0000  0.  0.  1.

PATTERSON INDIVIDUAL MATRIX
CODE      U      V      W
1..  3  3      .000  0.  0.  0.      .500  0. -2.  0.      .500  0.  0.  0.
2..  8  8      .000  2.  0.  0.      .000  0.  2.  0.      .000  0.  0.  2.
3..  6  6      .000  2.  0.  0.      .500  0.  0.  0.      .500  0.  0. -2.
4..  6  6      .000  2.  0.  0.      .500  0.  0.  0.      .500  0.  0.  2.
5..  8  8      .000  2.  0.  0.      .000  0. -2.  0.      .000  0.  0.  2.
6..  3  3      .000  0.  0.  0.      .500  0.  2.  0.      .500  0.  0.  0.

R-FACTOR BEST(LS) 24
BEST      R      NO      ATOM      X      Y      Z      B      SCALE      HEIGHT
1      .4599      4      BR 19      .1132      .1469      .2001      4.153      1.5955      378.1
2      .4599      2      BR 20      .1132      .8531      .2001      4.153      1.5955      378.1
3      .4599      3      BR 23      .3868      .1469      .7999      4.153      1.5955      378.1
4      .4599      1      BR 24      .3868      .8531      .7999      4.153      1.5955      378.1
5      .5071      7      BR 9      .0855      .1463      .9351      3.886      1.6555      378.1
6      .5071      8      BR 10     .0855      .8537      .9351      3.886      1.6555      378.1
7      .5071      6      BR 15     .4146      .1463      .0648      3.885      1.6558      378.1
8      .5071      5      BR 16     .4146      .8537      .0648      3.885      1.6558      378.1
9      .5536      10     BR 3      .3078      .1467      .8410      3.703      1.7294      378.1
10     .5536      12     BR 7      .1922      .1467      .1590      3.703      1.7294      378.1
11     .5536      11     BR 4      .3078      .8533      .8410      3.703      1.7294      378.1
12     .5536      9      BR 8      .1922      .8533      .1590      3.703      1.7294      378.1
13     .5651      15     BR 11     .0782      .1466      .0626      3.444      1.8142      378.1
14     .5651      13     BR 14     .4218      .8534      .9374      3.444      1.8142      378.1
15     .5651      14     BR 13     .4218      .1466      .9374      3.444      1.8142      378.1
16     .5651      16     BR 12     .0782      .8534      .0626      3.444      1.8142      378.1
17     .5723      19     BR 22     .3863      .8530      .2035      3.773      1.7775      378.1
18     .5723      18     BR 17     .1137      .1470      .7965      3.773      1.7775      378.1
19     .5723      20     BR 21     .3863      .1470      .2035      3.773      1.7775      378.1
20     .5723      17     BR 18     .1137      .8530      .7965      3.773      1.7775      378.1
21     .5755      24     BR 1      .3095      .1466      .1632      3.510      1.8502      378.1
22     .5755      21     BR 5      .1905      .1466      .8368      3.510      1.8502      378.1
23     .5755      23     BR 6      .1905      .8534      .8368      3.510      1.8502      378.1
24     .5755      22     BR 2      .3095      .8534      .1632      3.510      1.8502      378.1

TOTAL CPU TIME = 292 seconds ( = 0 H 4 M 52 S)

*****END *****
END OF THIS JOB

```

Figure 3. Output of PSL program. PSL prints out input data, intermediate peaks $[u, v, w, H]$, possible atomic coordinates and height $[(x, y, z, H)]$, and refined possible atom positions.

and the number of obtained atoms at parts A and B by SEARCHER. The computational times from a comparison of large computers with recent PCs are summarized in Table III.

X-83. $C_{19}H_{20}O_2N_5SBr$, $P2_1/c$, $Z = 4$. PSL found 4 peaks at Harker line $[0, v, 1/2]$, 22 peaks at Harker section $[u, 1/2, w]$, and 4 peaks at Harker line $[0, v, 1/2]$ with input data as shown in Figure 2. By solving the Patterson matrix, 24 possible atomic coordinates were obtained from 2 values of (y) and 12 pairs of (x, z) . R factors were from 0.477 to 0.632. Each R factor included the same 4 atom positions. Figure 3 shows

the output of PSL. These R factors were refined to 6 groups from 0.460 to 0.576 by the least-squares method. As shown in Figure 3, the lowest R factor 0.460 had 4 atomic coordinates, and these coordinates were the same atom positions related by the origin arbitrarily. Then we employed SEARCHER with NSF = 1, NLH = 5, and NLS = 38, and (0.3868, 0.1469, 0.7999) as Br atomic coordinates. SEARCHER computed 8 cycles as part A and found 28 atoms with $R = 0.147$ after 5 cycles at part B. These 28 atom positions are illustrated in the b^* -axis projection in Figure 4 and indicated all atoms of the compound.

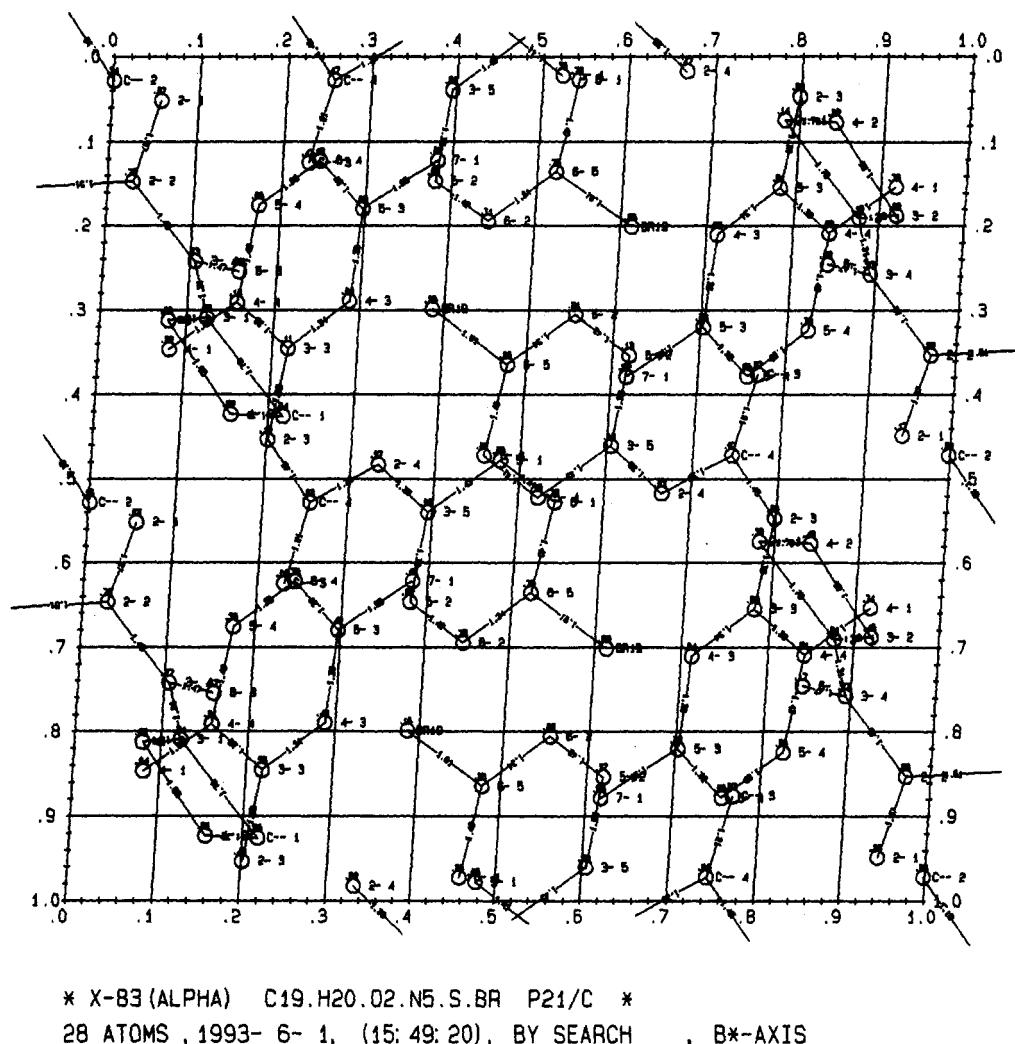


Figure 4. Output of SEARCHER program. SEARCHER draws the molecules of two-dimensional projection with final atomic coordinates.

X-144. $C_{19}H_{22}O_3S$, $P\bar{1}$, $Z = 2$. PSL found 79 peaks at the non-Harker section $[u, v, w]$. A set of 38 possible atomic coordinates was obtained from 46 pairs of (x, y, z) . We selected the atomic coordinates of lowest R factor as the Br atom. SEARCHER found all atoms of the compound. The structure of X-144 was solved¹⁷ initially by DIRECTER.¹⁸

X-149. $C_{23}H_{20}O_5S$, $Pna2_1$, $Z = 4$. PSL found 11, 5, and 49 peaks at Harker lines $[1/2, v, 0]$, $[u, 1/2, 1/2]$ and Harker section $[u, v, 1/2]$ with 200 divides equally along the b axis. A set of 66 possible atomic coordinates was obtained from 6 values of (x) , 10 values of (y) , and 60 pairs of (x, y) . These atoms reduced to 33 atoms by removing equivalent positions, and R factors were refined to 12 groups from 0.563 to 0.637. The lowest R factor had 2 atomic coordinates; we selected the lowest one as S atom considering z axis arbitrarily. SEARCHER found 39 atoms after 9 cycles at part A and refined 26 atoms with $R = 0.307$ after 14 cycles at part B. Another 2 atoms were obtained from the second cycle of SEARCHER.

CONCLUSION

PSL finds many heavy-atom positions with arbitrary origin selection. Users can then easily select one heavy-atom from one group. Heavy atoms at PSL usually appear in the lowest group of R factors. However, sometime atoms such as Cl or S are positioned in the lower part because the contributions of such atoms do not have large magnitudes for $F_0(h)$. To find the correct position of these heavy-atoms, the user must

select the atom position of PSL in order and use SEARCHER until the correct position is obtained. If the compounds do not include a heavy atom but some fragments are obtained from other procedures such as direct-methods, SEARCHER can be a powerful tool to find the remaining atom positions. SEARCHER also saves time and is cost efficient because all atom positions are found automatically and the two-dimensional projection diagram of molecules in a unit cell is drawn. Nonexpert crystallographers can determine the crystal structures of their own organic compounds by using PSL+SEARCHER programs.

The PC 1993 version of PSL+SEARCHER programs has changed the algorithms of peak summarization by adopting an automatic adjustment in Fourier summation, achieved a reduction of calculation time to exchange the sequences in DO-loops, and conformed to several user requirements such as the interactive file assignments and the additional options in the reflection data printing.

When we use PSL+SEARCHER, a structure analysis which took about 30 min running on the latest PC (i486 + overdrive processor, 25 MHz) took 2 min on the supercomputer Cray-1. Comparing the calculation speed of the computers, a PC has the capability of $1/10$ th of Cray-1, 1.1 times that of CDC6600, and 9 times that of CDC3600. Though the PC 1993 version takes more time rather than the supercomputer, the CPU speed of PCs will be faster in the near future, and it seems only a matter of time until the capability of a PC will exceed that of the mainframe computers of the past. This

version is very easy to operate, very handy to calculate anywhere, and relatively inexpensive.

Program Availability. These programs are written in standard Fortran 77 and have been fully tested under MS-DOS 3.x. Only standard options were applied. The limitation is unified by using DS*SYSTEM common libraries: the number of reflections ≤ 4800 , the number of atoms ≤ 200 , and the number of grids in one section $\leq 200 \times 200$. These programs need 520 kbyte of memory and 2 Mbyte of disk space. A laser printer with HP-GL emulation is needed for drawing. Each program consists of a main segment and DS*SYSTEM libraries totaling about 29 000 source code lines. Documentation and the latest executable files are available from the author.

ACKNOWLEDGMENT

The author gratefully acknowledges Dr. H. Koyama of Kinki University who encouraged and helped K.O. to accomplish this work.

REFERENCES AND NOTES

- (1) Okada, K.; Okada, S.; Koyama, H. 32-bit Personal Computer Version of DIRECT-SEARCHER System. *Acta Crystallogr., Sect. A* **1990**, *A47*, C70.
- (2) Okada, K.; Koyama, H. 32 bit Personal Computer Version of the DIRECT-SEARCHER System. *J. Appl. Crystallogr.* **1991**, *24*, 1067–1070.
- (3) Egret, E. Patterson Search—an Alternative to Direct Methods. *Acta Crystallogr., Sect. A* **1983**, *A39*, 936–940.
- (4) Hai-Fu, F.; Jin-Zi, Q.; Jia-Xing, Y.; Chao-De, Z.; Quan, H. The Program SAPI and its Application. II. The Combination of Patterson Superposition and Direct Methods. *Acta Crystallogr., Sect. A* **1988**, *A44*, 691–692.
- (5) Egret, E.; Sheldrick, G. M. Search for a Fragment of Known Geometry by Integrated Patterson and Direct Methods. *Acta Crystallogr., Sect. A* **1985**, *A41*, 262–268.
- (6) Rossmann, M. G.; Arnord, E.; Vriend, G. Comparison of Vector Search and Feedback Methods for Finding Heavy-Atom Sites in Isomorphous Derivatives. *Acta Crystallogr., Sect. A* **1986**, *A42*, 325–334.
- (7) Luger, P.; Fuchs, J. IMPAS—A Simple Structure Determination Procedure Based on Intermolecular Patterson Vectors. *Acta Crystallogr., Sect. A* **1986**, *A42*, 380–386.
- (8) Wilson, C. C. Increasing the Size of Search Fragments for Use in Patterson Method Calculations—the Partial-Fragment Rotation Function. *Acta Crystallogr., Sect. A* **1988**, *A44*, 478–481.
- (9) Rius, J.; Miravittles, C. An Automated Full-Symmetry Patterson Search Method. *J. Appl. Crystallogr.* **1987**, *20*, 261–264.
- (10) Pavelik, F.; Rizzole, C.; Andreetti, G. D. A Program for Automatic Solution of Crystal Structure by Patterson, Fourier and Superposition Methods. Recent Developments. *J. Appl. Crystallogr.* **1992**, *25*, 328–329.
- (11) Jacobson, R. A.; Beckman, D. E. A Modified Patterson Superposition Technique Using Multiple Vectors. *Acta Crystallogr., Sect. A* **1979**, *A35*, 339–340.
- (12) Terwilliger, T. C.; Kim, S.; Eisenberg, D. Generalized Method of Determining Heavy-Atom Positions Using the Difference Patterson Function. *Acta Crystallogr., Sect. A* **1987**, *A43*, 1–5.
- (13) Koyama, H.; Okada, K. Automatic Heavy-Atom Analyses of Some Organic Compounds. *Acta Crystallogr., Sect. B* **1970**, *B26*, 444–447.
- (14) Koyama, H.; Okada, K. SEARCHER: Automatic Heavy-Atom Analysis for the Determination of the Crystal Structure; Osaka, Japan, 1970.
- (15) Koyama, H.; Okada, K. Automation of the Crystal Structure Analyses. The DIRECT-SEARCHER Program System. *Acta Crystallogr., Sect. A* **1975**, *A31*, S18.
- (16) Koyama, H.; Okada, K. Automatic Determination of Crystal Structure, I. Heavy-atom Method, II. Direct-Method. *Acta Crystallogr., Sect. A* **1978**, *A34*, S40.
- (17) Okada, K.; Koyama, H. Structure of 17 β -Acetoxy-18-methyl-1-thia-A-norestra-3,9(10),11-trien-2-one. *Acta Crystallogr., Sect. C* **1993**, *C49*, 1123–1125.
- (18) Okada, K. To be published.