

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

3. Crystallographic Programs

Sachiko Okada*

Faculty of Engineering, Science University of Tokyo, 1-3 Kagurazaka, Shinjuku-ku, Tokyo 161, Japan

Kenji Okada

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

Received February 24, 1994*

The DIRECT-SEARCHER automatic system (DS*SYSTEM)^{1,2} is designed to carry out the calculations for the structure analysis of organic compounds running on mainframe and personal computers (PC). This system consists of heavy-atom methods (PSL + SEARCHER), direct methods (DIRECTER, MULTAN series), and other crystallographic programs. In the first³ and the second paper,⁴ the automation of heavy-atom methods and direct methods for crystallographic phasing are described. In this paper details of the other improved crystallographic programs are described, respectively. DS*SYSTEM is suitable for use by organic chemists as well as professional crystallographers.

INTRODUCTION

When organic compounds include sulfur or a heavier atom, heavy-atom methods are usually used to determine the molecular structure, and direct methods are used when compounds consist only of light atoms. DS*SYSTEM, illustrated in Figure 1, is designed to carry out the analytical part rather than the experimental part and consists of the heavy-atom methods, direct methods, and other crystallographic 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 assumptions or human intervention. DIRECTER is also based on the symbolic-addition method using the tangent formula with the multisolution process to determine the atomic coordinates of all atoms automatically without any chemical assumptions or human intervention.

In this paper we describe other crystallographic programs. The symbol * means that the main program produces the plot file, and the symbol † means an introduced program.

DESCRIPTION

AFCR Program. The output data of AFC (automatic four-circle controlled) manufactured by Rigaku Co. or Mac Science Co. include many experimental values (the standard reflections, flags, and I_{b1} , I_{b2} , T_1 , T_2 , ...) which are not used in the structure analysis. AFCR edits the raw data to prepare the necessary data for crystallographic computing. The equivalent reflections and extinction rule are treated depending upon space group. The observed structure factor $F_o(h)$ is adjusted by the ratio of standard reflection of each block with decay corrections. If the reflection is counted twice or more, the value is selectable by user option (averaged, previous, or last). $\sigma(I_o(h))$ is written as follows:

$$\sigma(I_o(h)) = \left(I_p + \left(\frac{T_p}{2T_B} \right)^2 (B_1 + B_2) \right)^{1/2} \quad (1)$$

where I_p is the peak intensity, scan count; T_p is the peak measurement time, scan time; T_B is one background mea-

Experimental Part

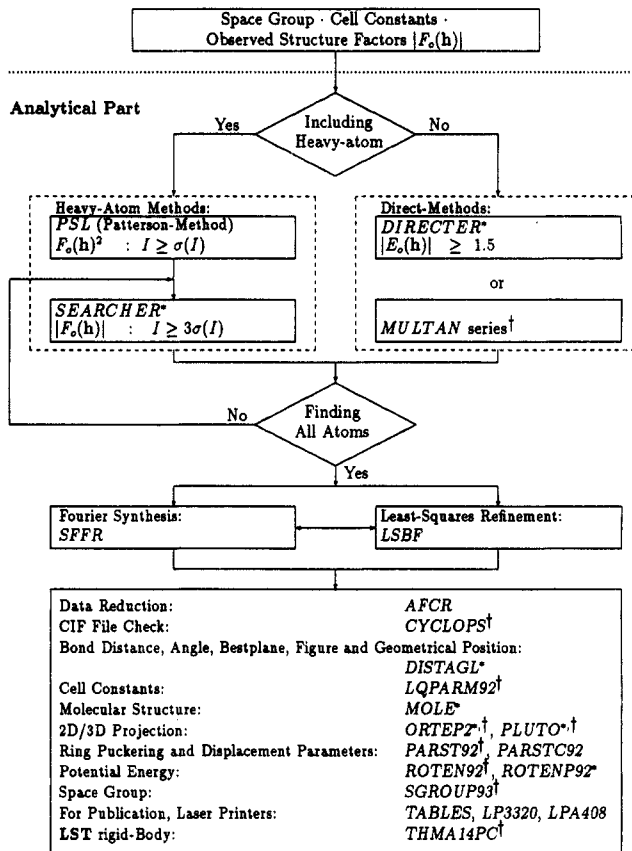


Figure 1. Simplified layout of DS*SYSTEM: (*) produced plot file; (†) introduced software.

surement time; B_1 is the lower background peak, count; and B_2 is the upper background peak, count.

The output is the reflection data h , k , l , $F_o(h)$, $\sigma(F_o(h))$, and $\sigma(I_o(h))$, which are sorted with user options h , k , or l .

CYCLOPS Program†. This program reads the crystallographic information file (CIF) and checks CIF texts and data, comparing with the dictionary CIFDIC.C91. Messages are printed out to the standard output device. CYCLOPS† was prepared by the International Union of Crystallography (IUCr) in 1992.⁵

* Abstract published in *Advance ACS Abstracts*, May 15, 1994.

DISTAGL Program*. DISTAGL* handles the following calculations through a keyword.

DIST calculates atomic bond distance d_{ij} and angle θ_{123} with standard deviation δ_d defined as follows:

$$d_{ij} = (\Delta G \Delta)^{1/2} \quad (2)$$

$$\cos \theta_{123} = \frac{\Delta_{12} G \Delta_{32}}{(\Delta_{12} G \Delta_{12})^{1/2} (\Delta_{32} G \Delta_{32})^{1/2}} \quad (3)$$

$$\delta_d = \left\{ \sum_{i=1}^N \left(\frac{\partial f}{\partial p_i} \right)^2 \sigma_i^2 \right\}^{1/2} \quad (4)$$

where Δ is the difference row vector between the fractional coordinates; G is the metric tensor, $\equiv \mathbf{a} \mathbf{a}^T$ (\mathbf{a} is a row vector of cell constants; \mathbf{a}^T is a transposed column vector of \mathbf{a}); N is the number of variables; f is the function of computation (distance or angle); p_i is the parameter (cell constants and atomic coordinates); and σ_i is the standard deviation of the parameter.

BOND calculates intra/interatomic bond distances and angles within a molecule and between molecules with standard deviations, defined by eqs 2–4.

BESTPLANE finds the principal plane for specified atoms. The method of calculation used is an iterative technique⁶ that minimizes the weighted sum of squares of residuals M .

$$M = \sum_{i=1}^N w_i (x_i | \underline{n})^2 \quad (5)$$

where \underline{n} is the unit normal row vector; w_i is the weight for atom i ; x_i is the coordinate row vector of atom i ; x_i^T is the transposed column vector of x_i ; and N is the number of atoms defining the plane.

Dihedral angles ω_{12} are calculated by the normals \underline{n}_1 and \underline{n}_2 to the planes.

$$\cos \omega_{12} = \frac{\underline{n}_1 G^* \underline{n}_2}{(\underline{n}_1 G^* \underline{n}_1)^{1/2} (\underline{n}_2 G^* \underline{n}_2)^{1/2}} \quad (6)$$

where $G^* = G^{-1}$.

TORSION calculates the torsion angle defined by four atoms with the right-hand rule.⁷

FIGURE draws a two-dimensional projection (a^* , b^* , c^*) of the molecules in a unit cell with the scale of 1 Å = 0.5–2.54 cm and prints bond distances and all equivalent positions of atoms.

GENERAL leads the orthogonal matrix O and generates several atom positions geometrically:

$$O = \begin{pmatrix} a & b \cos \gamma & c \cos \beta \\ 0 & b \sin \gamma & -c \sin \beta \cos \alpha^* \\ 0 & 0 & c \sin \beta \sin \alpha^* \end{pmatrix} \quad (7)$$

HEXGNL generates three atoms in hexagonal, **LINEAR** generates one atom in linear, **OCTHED** generates four atoms in octahedral, **SQRPLN** generates two atoms in square planar, **TECHAN** generates two atoms in tetrahedral chain, **TETERM** generates three atoms in tetrahedral terminal, and **TRIGON** generates one atom in trigonal conformation.

LP3320 and LPA408 Programs. These programs prepare a print file or a plot file from the output files of each program. These programs edit print files with user selections of a carriage control, number of copies, and number of characters per line for a laser printer LP3320-SP4 (Ricoh) or A40x (Canon).

LQPARM92 Program†. LQPARM92† reads the reflection data h, k, l, θ , or approximate cell constants with θ and refines the cell constants considering crystal type by the least-squares method. LQPARM92† is rewritten from the original version of LQPARM.⁸

LSBF Program. LSBF refines atomic parameters by the least-squares method with a block-diagonal matrix approximation or full matrix. It is used to adjust the values of the scale factor, the overall temperature factor, and atomic parameters (site occupancy, positional and temperature factors) in order to minimize the difference in structure factors $F_o(\mathbf{h})$ and $F_c(\mathbf{h})$. The algorithm is based on ORFLS,⁹ and the refinement is based on $F(\mathbf{h})$. We made the following improvements from ORFLS so that the program can solve the normal equation by a block-diagonal matrix approximation or full matrix, handle the anomalous dispersion, treat several dumping factors through the user's option, calculate intra/interbond distance and angle, and unify the input/output format and the weight for DS*SYSTEM.

LSBF stores the square matrix M , which multiplies the rows p by the columns p and solves the weighted normal equation $M \cdot p = \Delta$. The matrix inversion of the block-diagonal approximation is the combinations of a single 2×2 for a scale factor and an overall temperature factor, and 5×5 or 10×10 atom blocks for each atomic parameter (a, x, y, z, B , or β_{ij}). The other coefficients are ignored. On the other hand the full-matrix inversion solves M one time, the same as the ORFLS. The standard deviation σ_p of each parameter p is expressed as follows:

$$\sigma_p = \left(\frac{R_w}{m_{pp}(N-P)} \right)^{1/2} \quad (8)$$

where m_{pp} is the p th diagonal elements of M^{-1} ; N is the total number of reflection data, $n = 1, 2, \dots, N$; and P is the total number of parameters.

It is able to select the dumping factor, up to six items, and multiply it independently by the parameters such as coordinates or temperature factors of hydrogen or non-hydrogen atoms. LSBF continues the execution and indicates a warning message if a temperature factor becomes negative.

LSBF provides the weighting schemes of eight kinds such as 1.0, $1/|F_o(\mathbf{h})|$, etc. If a chemist wants to apply the special function to the refinement, it is easy to add a subprogram WTSUB with Fortran77 language. LSBF provides chi-squared S and chi-squared sigma σ_S measures of refinement added on R -factor and weighted R -factor wR .

At the last cycle of the refinement, LSBF calculates intra/interbond distance and angle for the confirmation and examination of the result.

MOLE Program*. MOLE* draws any lines and circles changed with the line width and line type (solid or dashed lines) and draws ASCII characters with any angle and arbitrary size. This program uses the drawing of chemical structures and chemical reactions.

MULTAN Series†. This well-known series is made up of the following programs: **NORMAL78** computes normalized structure amplitudes $|E_o(\mathbf{h})|$ from observed structure factor $|F_o(\mathbf{h})|$. **MULTAN78** makes Σ_2 relationships, finds an origin set, finds the starting reflections by application of Σ_1 , and determines phases using the weighted tangent formula. **EXFFT78** computes an E -map by the first Fourier transform. **PEAKSER78** finds the coordinates of the highest peaks from the E -map and looks for groups of peaks which form a molecular fragment. **MULTAN series†** is rewritten from the original version of **MULTAN78**.¹⁰

ORTEP2 Program*†. This program draws thermal motion probability ellipsoids at atomic sites and produces stereoscopic pairs of molecules. Overlap is eliminated automatically. ORTEP2*† is rewritten from the version of ORTEP-II.¹¹

PARST92 Program†. This program can calculate the following items through a keyword: Niggli's reduced cell, principal axes of a thermal ellipsoid, bond length (uncorrected and corrected for a thermal motion), the ring-puckering parameters (Q_T , q , ϕ , θ),¹² the displacement asymmetry parameters (DAP, D_s , D_2),¹³ of rings, spherical polar coordinates for stereographic projection, possible hydrogen bonds, comparison of the coordinates of subsets of atoms. PARST92† is a modified version of PARST.¹⁴

PARSTC92 Program. PARSTC92 reads the output file of PARST92† and punches out CIF data. This file is submittable to IUCr as a structural paper with a little amendment.

PLUTO78 Program*†. This program is rewritten from the package of PLUTO¹⁵ for the production of crystal and molecular illustrations and is a general plot package for the production of crystal and molecular illustrations.

ROTE92 Program†. ROTEN92† calculates the difference molecular potential energy after rotations of one or two molecular fragments about a given direction. ROTEN92† uses a function of the type $E_{ij} = -A_{ij}r_{ij}^{-6} + B_{ij} \exp(-C_{ij}r_{ij})$. The atomic charges are considered according to the methods of Smith¹⁶ or Gasteiger and Marsili.¹⁷ ROTEN92† is a modified version from ROTENER¹⁸ with ATOMCHAR.¹⁹

ROTE92 Program*. ROTENP92* reads the output files of ROTEN92 and draws calculated difference potential-energy profiles ΔE and contour lines for the rotation of two molecular fragments.

SFFR Program. SFFR calculates structure factors and various Fourier syntheses with any grids and prints out the Fourier maps with three-dimensional peak positions. It is very convenient to find the correct hydrogen atom positions.

In the two-dimensional electron density distributions, the coordinates and peak height of the Fourier peak are calculated using a suitable second-order polynomial by Lagrangian interpolation from nine values. 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 is improved from a predetermination to an automatic adjustment.

SGROUP93 Program†. This program reads the output file that includes h , k , l , $F_o(h)$, and $\sigma(F_o(h))$ and prints out a list of the detected absence followed by a list of possible space groups in decreasing order of space group number. SGROUP93† is the implementation version of SGROUP.²⁰

TABLES Program. TABLES converts actual anisotropic temperature factors β_{ij} to U_{ij} and calculates equivalent isotropic thermal parameters B_{eq} and U_{eq} from β_{ij} as follows:²¹

$$B_{eq} = (4/3) \left\{ \sum_i \sum_j \beta_{ij} a_i a_j \right\} \quad (9)$$

$$U_{eq} = (1/6\pi^2) \left\{ \sum_i \sum_j \beta_{ij} a_i a_j \right\} \quad (10)$$

TABLES calculates the following items: the bond distance, angle, and torsion angle; molecular weight M_r ; calculated density D_x ; absorption coefficient μ . Distances and angles are tabled with their standard deviations in parentheses. Necessary data to submit IUCr are punched out automatically in CIF format.²²

THMA14PC Program†. This program is for a thermal motion analysis of the anisotropic displacement parameters

(ADPs), is carried out in terms of the LST rigid-body approximation,^{23,24} and considers the correlations in the internal and overall motions.²⁵ THMA14PC† is rewritten from the 1992 version of THMA14.²⁶

CONCLUSION

The crystallographic programs are used in the process of the structure determination of organic compounds. There are many purposes such as the initial stage (AFCR, LQPARM92,† SGROUP93†), the confirmations of hydrogen atoms (DISTAGL*, SFFR), the least-squares refinement (LSBF), the discussion of the final stage (DISTAGL*, ROTEN92,† THMA14PC†), and the publication (CYCLOPS,† ORTEP2,*† PLUTO78,*† TABLES). In case several atom groups have large thermal motions, thermal motion analysis in LST rigid-body approximation of the THMA14PC program† and potential energy calculations of the ROTEN92 program are used to confirm the static or dynamic disorder for atoms and molecules. The examples of structure analysis using DS*SYSTEM are reported in refs 27–30.

DS*SYSTEM is the computer system researched and developed so that noncrystallographers can determine the crystal structures of their own organic compounds by themselves. It is very important to unify the computer system especially in input and output. Also, graphical outputs are very useful to understand the results immediately.

Program Availability. These programs are written in standard Fortran77 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 51 000 source code lines. Documentation and the latest executable files of some programs are available from K.O.

ACKNOWLEDGMENT

The authors gratefully acknowledge Dr. H. Koyama of Kinki University, who encouraged and helped us to accomplish this work.

REFERENCES AND NOTES

- Okada, K.; Okada, S.; Koyama, H. 32-bit Personal Computer Version of DIRECT-SEARCHER System. *Acta Crystallogr., Sect. A* **1990**, *A47*, C70.
- Okada, K.; Koyama, H. 32 bit Personal Computer Version of the DIRECT-SEARCHER System. *J. Appl. Crystallogr.* **1991**, *24*, 1067–1070.
- Okada, K. Application of Automated Structure Analysis to Some Organic Compounds Using PCs: Part I. Automated Patterson Techniques. *J. Chem. Inf. Comput. Sci.* **1993**, *33*, 871–878.
- Okada, K. Application of Automated Structure Analysis to Some Organic Compounds Using PCs: Part II. Automation of the Symbolic-Addition Method. *J. Chem. Inf. Comput. Sci.*, in press.
- Hall, S. R. CIF Applications. III. CYCLOPS: for validating CIF Data Names. *J. Appl. Crystallogr.* **1993**, *26*, 480–481.
- Schomaker, V.; Waser, W.; Marsh, E. R.; Bergman, G. To Fit a Plane or a Line to a Set of Points by Least Squares. *Acta Crystallogr.* **1959**, *12*, 600–604.
- Klyne, W.; Prelog, V. Description of Steric Relationships Across Single Bonds. *Experientia* **1960**, *XVI*, 521–523.
- Nardelli, M. *LQPARM: A Fortran Routine for Refining the Lattice Parameters by the Least-Squares Procedure*; University of Parma: Parma, Italy, 1992.

- (9) Busing, W. R.; Martin, K. O.; Lavy, H. A. *ORFLS: A Fortran Crystallographic Least-Squares Program*, ORNL-TM305; Oak Ridge National Laboratory: Oak Ridge, TN, August 1962.
- (10) Main, P.; Hull, S. E.; Lessinger, L.; Germain, G.; Declercq, J.-P.; Woolfson, M. M. *MULTAN78: A System of Computer Programs for the Automatic Solution of Crystal Structures from X-ray Diffraction Data*; Univ. of York: York, England, 1978.
- (11) Johnson, C. K. *ORTEP-II: A Fortran Thermal-Ellipsoid Plot Program for Crystal Structure Illustrations*; Report ORNL-5138 (3rd revision of ORNL-3794) revised; Oak Ridge National Laboratory: Oak Ridge, TN, 1976.
- (12) Cremer, D.; Pople, J. A. A General Definition of Ring Puckering Coordinates. *J. Am. Chem. Soc.* **1975**, *97*, 1354–1358.
- (13) Nardelli, M.; Ring Asymmetry Parameters from Out-of-Plane Atomic Displacements. *Acta Crystallogr., Sect. C* **1983**, *C39*, 1141–1142.
- (14) Nardelli, M. PARST: A System of Fortran Routines for Calculating Molecular Structure Parameters from Results of Crystal Structure Analysis. *Comput. Chem.* **1983**, *7*, 95–98.
- (15) Motherwell, W. D. S.; Clege, W. *PLUTO: Program for Plotting Molecular and Crystal Structures*; University of Cambridge: England, 1978.
- (16) Smith, D. W. A New Method of Estimating Atomic Charges by Electronegativity Equilibration. *J. Chem. Educ.* **1990**, *67*, 559–562.
- (17) Gasteiger, G.; Marsili, M. Iterative Partial Equalization of Orbital Electronegativity—A Rapid Access to Atomic Charges. *Tetrahedron* **1980**, *36*, 3219–3228.
- (18) Nardelli, M. *ROTENER: A Fortran Routine for Calculating Non-Bonded Potential Energy*; University of Parma: Parma, Italy, 1992.
- (19) Nardelli, M. *ATOMCHAR: A Computer Routine for Calculating the Atomic Charges in a Molecule*; University of Parma: Parma, Italy, 1992.
- (20) Yerkess, J. A Program for Determinations of Possible Space Groups from Structure-Factor Listings. *J. Appl. Crystallogr.* **1990**, *23*, 142–143.
- (21) Fisher, R. X.; Tillmanns, E. The Equivalent Isotropic Displacement Factor. *Acta Crystallogr. Sect. C* **1988**, *C44*, 775–776.
- (22) Hall, S. R.; Allen, F. H.; Brown, I. D. In the Crystallographic Information File (CIF), A New Standard Archive File for Crystallography; *Acta Crystallogr. Sect. A* **1991**, *A47*, 655–685.
- (23) Schomaker, V.; Trueblood, K. N. On the Rigid-Body Motion of Molecules in Crystals. *Acta Crystallogr., Sect. B* **1968**, *B24*, 63–76.
- (24) Trueblood, K. N. Analyses of Molecular Motion with Allowance for Intermolecular Torsion. *Acta Crystallogr., Sect. A* **1978**, *A34*, 950–954.
- (25) Dunitz, J. D.; White, D. N. J. Non-Rigid-Body Thermal-Motion Analysis. *Acta Crystallogr., Sect. A* **1973**, *A29*, 93–94.
- (26) Trueblood, K. N. *THMA14: The Computer Program for Thermal Motion Analysis including Internal Torsion*; University of California: Los Angeles, CA, 1992.
- (27) Okada, K.; Koyama, H. Structure of the Germacrane Furanosessquiterpenoid Linderane. *Acta Crystallogr., Sect. C* **1993**, *C49*, 341–343.
- (28) 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.
- (29) Okada, K.; Koyama, H. Crystal Structure and Conformational Analysis of 3 β ,17 β -diacetoxy-8,9-seco-5 α -androstane-8,9,11-trione. *Bull. Chem. Soc. Jpn.* **1994**, *67*, 189–195.
- (30) Okada, K.; Takekuma, S. Crystal Structure and Conformational Analysis of 7,13-Abietadien-18-oic Acid. *Bull. Chem. Soc. Jpn.* **1994**, *67*, 807–815.