32 bit personal computer version of the DIRECT-SEARCHER system. By Kenji Okada, Research and Development Center, Ricoh Co. Ltd, Kouhoku-ku, Yokohama 223, Japan and Hirozo Koyama, Research Institute for Science & Technology, Kinki University, 3-4-1, Kowakae, Higashi-Osaka 577, Japan

(Received 10 April 1991; accepted 18 June 1991)

Abstract
A new version of DIRECT-SEARCHER (DIRECT-SEARCHER automatic system) for personal computers (PC) consisting of several crystallographic programs has been developed for automatic structure analysis of organic compounds as an innovation of the DS*SYSTEM [Koyama & Okada (1975). Acta Cryst. A31, S18]. Each program is improved and rewritten in Fortran77 for the 32 bit PC. The PC version of DS*SYSTEM is used by organic chemists rather than professional crystallographers.

Introduction
Since Patterson presented the Patterson function in 1935, a great number of crystal structures have been solved using his technique. After Karle & Karle solved structures with the direct method using the symbolic addition procedure in 1964, direct methods become the major crystallographic tool, especially for organic compounds. However, many kinds of computer are used for crystallographic computing such as supercomputer, mainframe, superminicomputer, engineering workstation (EWS) and personal computer (PC). There are only three crystallographic packages for all kinds of computer from supercomputer to PC, for example, the XTAL3.0 system (Flack et al., 1990) for laptop PC to Cray, an interactive program system NRC/VAX (Gabe, Le Page, Charland, Lee & White, 1989) for IBM, VAX and IBM PC-XT/AT and the SHELX-90 program system (Sheldrick, 1990) for Cray to PC. The program systems reported for PC are the following: an integrated package CRYSRULER (Rizzoli, Sangemaro, Calestani & Andreetti, 1987) for IBM-PC; APPLE-CRYST83 (Kennard, 1984) for AppleI-Plus; microcomputer programs for automated crystal structure analysis (Tanaka, Yamauchi, Katsube & Ashida, 1989) for the NEC-9801 PC in Japan.

We developed DS*SYSTEM for automatic structure analysis on a mainframe CDC6600/CDC6000 (Koyama & Okada, 1969, 1970a, b). DS*SYSTEM consists of several crystallographic programs. It is intended to be used by organic chemists rather than professional crystallographers. We have improved the Cray-I version (Koyama & Okada, 1975, 1978) by adding vector facilities and optimizations. Recently we focused on the computational speeds, storage capacities, ease of use and costs of PC. In this paper we describe the details of the PC version of DS*-SYSTEM which has been implemented on a 32 bit PC with several modifications, adding new programs and bag fix (Okada, Okada & Koyama, 1990).

Structure of the DS*SYSTEM
DS*SYSTEM is designed to carry out the calculations necessary to solve the crystal structures of organic compounds. This system consists of the following three groups which make up the analysis part as shown in Fig. 1: heavy-atom analysis (PSL, SEARCHER), direct methods (DIRECTER, MULTAN78) and others.

1. Heavy-atom analysis
A heavy-atom position is determined automatically by the PSL program which has the following five steps: (i) The Harker sets and non-Harker sets in the Patterson map are calculated. (ii) The possible atomic coordinates are stored together with the height of the peak in memory by solving the Patterson matrix. (iii) The structure factors are computed for all possible atoms. (iv) The R factors are

![Fig. 1. A simplified layout of the DIRECT*SYSTEM automatic system. † Introduced software.](image-url)
stored in increasing order and the R factors are compared to identify the same atom. At this stage, the heavy atom is positioned at the lower R factors. (v) The heavy-atom coordinates are refined by full-matrix least squares and R factors are sorted once more in increasing order. Finally, the heavy atom is selected out at the top of the R factors. The original version of PSL was developed by Koyama & Okada (1974) for the CDC 6600 (131K word) and modified to be applicable to the Cray-1 (500K word) by Okada (1980).

Program SEARCHER calculates the structure factors using the defined heavy atom and finds a set of atomic positions to be included in the next calculation of structure factors (part A). Based on the temperature factor using block-matrix least squares, SEARCHER selects a set of atomic sites to be provided for the next cycle of SEARCHER (part B). In part A the following are carried out: (i) Calculation of the structure factors, the scale factor k and the R factor. (ii) Comparison between the number of input atoms for the structure-factor calculation (NSF) and the maximum NSF (NLS). If NSF < NLS, part A is to be continued otherwise part B is started. (iii) Calculation of three-dimensional Fourier sections in an asymmetric unit at the predetermined intervals. (iv) Calculation of distances between the input atoms and the peaks. (In the first step of each calculation, this part passes through.) (v) Selection of the atomic sites based on the ‘reasonable’ interatomic distances (1-1-1.7 Å) between neighboring light-atom peaks. (vi) Selection of the atoms (NLH) to be included in the next Fourier calculations.

The second part (B) selects out a set of atomic sites from the atoms found in part A and these atoms are used for the next cycle of SEARCHER. (i) If NSF ≥ NLS, calculate the least squares with the block-diagonal method. (ii) Determine which atom to eliminate under the condition B > 10-0 Å² for the next least-squares calculations. (iii) Select the atomic sites based only on the temperature factor. These processes are carried out automatically.

SEARCHER draws the two-dimensional projection diagrams of molecules in a unit cell and prints the result of the atomic coordinates, the structure-factor tables, the final R factor and the bond distances. The atomic coordinates are written on the punch file which is able to be used as the input for the next calculation. The original version of SEARCHER was developed by Koyama & Okada (1970b) for the CDC3600 (90K word), modified for the CDC6600 in 1977 with addition of a full-matrix option and modified for the Cray-1 by Okada (1981).

2. Direct methods

Program DIRECTER based on the symbolic-addition method has been reconstructed to solve the crystal structure without any human intervention among the several programs such as TANGFORM (Brenner & Gum, 1968), ASIG (Brenner & Gum, 1968), SNOOPY (Motherwell, 1971), TANTWO (Drew, Larson & Motherwell, 1971), SEARCHER (Koyama & Okada, 1970b) and MULTAN71 (Main, Woolfson & Germain, 1971).

The output data of DIRECTER are projection diagrams of the molecules in the unit cell, atomic coordinates and bond distances. The atomic coordinates are also written on the punch file.

After loading the data, the following six logical steps are carried out in sequence. (i) ORIGIN: define three kinds of origin set with phased symbols. (ii) SNOOPY: choose additional symbols to generate up to 95% of the reflections from the largest 50 |E(h)| values. (iii) TANGENT: refine the phase and select the best ten run sets from each starting set. The criteria are determined from R_{Kale}, Q_{Drew}, a and t. The values of R_{Kale} (30-10%) and Q_{Drew} (45-10%) are good indicators of the correct phase set. A phase is accepted if it satisfies the following criteria: a > 2-5, t > 0-25, |E(h)|E(k)|E(h - k)| ≥ 10.0 and Δρ_{shuf} < π/2 (from cycle to cycle). (iv) SELECTION: select the best five run sets from 30 runs. (v) E-MAP: calculate Fourier synthesis by E(h), look for the peaks, store the coordinates of each peak with the height and calculate the structure factors. (vi) FIGURE: draw two-dimensional projections of the molecules in a unit cell.

The initial version of DIRECT (Koyama & Okada, 1978) was developed on a CDC6600 in 1975. DIRECT80 was transformed to a Cray-1 in 1980 from DIRECT and is improved on the PC as DIRECTER.

MULTAN78 which is rewritten from the original MULTAN78 (Main, 1978) is used for the automatic multisolution procedures.

3. Others

We also developed the following eight programs to carry out the calculations required to solve the structure of crystals. The symbol* means that the main program produces a plotter file of a two- or three-dimensional projection. (i) program AFCR is designed for the direct data which are obtained by the AFC (automatic four-circle controlled) diffractometer manufactured by Rigaku Co. and prepares the necessary data for structure analysis. The direct data include many experimental values which are not used in the structure analysis. (ii) DISTAGL* handles the following calculations through a keyword: calculates atomic distances and angles with standard deviations (DIST), the best planes with given atoms (BESTPL) and bond distances and angles within a molecule and between molecules (BOND); plots two-dimensional projections of a unit cell (FIGURE); generates one general atom (GENRAL), three hexagonal atoms (HEXGNL), one linear atom (LINEAR), four octahedral atoms (OCTHED), two square-planar atoms (SQRPLN), two tetrahedral atoms (TECHAN), three tetrahedral terminal atoms (TERTERM) and one trigonal atom (TRIGON). (iii) LP3320 prepares a print file or a plot file from the output files of each program. LP3320 edits a carriage control, number of copies and characters per line for the print file and paper size for the plot file for the Ricoh LP3320-SP4 (240 d.p.i.) laser printer. (iv) LSBF refines the atomic parameters by the least-squares method with block-diagonal or full matrix. The refinement is based on structure factor F and can be obtained by minimizing the sum of the weighted squares of the residual. It is used to adjust the scale factors, the overall temperature factor and the atomic parameters (site occupancy, coordinates and temperature factors). The algorithm of LSBF is based on ORFLS (Busing, Martin & Levy, 1962) and we made the following improvements so that the program can: (a) solve the normal equations by block-diagonal matrix approximation.
or full matrix; (b) handle the anomalous dispersion; (c) treat the dumping factor through a user's option; (d) calculate inter/intrabond distances and angles; (e) unify the input/output formats and the weights for DS*SYSTEM. (v) ORTEP2* is the same as ORTEPII (Johnson, 1976) for crystal structure illustrations. (vi) PLUTO78* is the same as PLUTO, written by Motherwell (1978), for the production of crystal and molecular illustrations. (vii) SFFR calculates the structure factors with any weights and various Fourier syntheses with any grids and prints out the Fourier maps with three-dimensional peak positions. (viii) TABLES prepares the tables for a publication such as bond distances and angles and equivalent isotropic thermal parameters $B_{eq}$ using the actual atomic parameters.

The PC version of DS*SYSTEM is designed primarily to operate in an interactive environment and is easy to use. The main four merits are as follows: (i) After each program is started, it asks for the input/output file assignments from the user on the PC screen. The PC screen is also used to display the stage of data processing. This is different from supercomputer or mainframe computer which operates batch mode processing. (ii) The data preparation is easier using the text editor than for mainframe computers. Then the output files of the programs such as PSL, SEARCHER and DIRECTER have standardized data architecture and text readable format. The user can create, collect and correct the input data easily from keyboard or output data from various programs on the editor to the next calculation. (iii) The PC version keeps the same capabilities and holds the same input/output data structures as the mainframe DS*SYSTEM. The user can select the platform according to the size. (iv) Only a laser printer is used for output of the PC version instead of two output modes such as a printer and a pen plotter. The mainframe DS*SYSTEM uses a high-speed line printer and online/offline colored pen plotter. We have recently developed program LP3320 and library GR-LIB as plotting programs to use the same laser printer.

### System requirements

The PC version of DS*SYSTEM runs under the following hardware configurations and software requirements.

1. **Hardware and software**

   The standard configuration of the PC is formed from the CPU of Intel 80X86 with 640 kbyte memory equipped with auxiliary memories on floppy disks. To be most useful the PC should have an additional co-processor (Intel 80X87) and the faster hard disks, and the newer and faster processor achieves the shortest times. As the CRT is used only for conversation with the PC and informs the user of the progress of the calculation, it is not necessary to have a color CRT. The laser printer is a very convenient device as printer and plotter. The Ricoh LP3320-SP4 laser printer we used is capable of printing at 240 d.p.i. and has A4/A3 paper cassettes. This device also has a PR201H emulator for a printer, an HP-GL emulator for a plotter and a TK401X emulator for hard copy. With regard to the software, we use MS-DOS (Version 3.X) operating system. The PC is able to use 520 kbyte as the user memory because there are no device drivers. DS*SYSTEM is written by the full set of Fortran77 and no assembler language is used. We use an MS-Fortran (Version 4.1X) compiler with the huge model option of emulation mode. For the plotter we have developed the Calcomp/HP-GL conversion library GRLIB to maintain the compatibility with the supercomputer and mainframes. The screen editor is the most powerful and least time-consuming tool for editing the data and controlling statements.

2. **Restrictions**

   The limitation of DS*SYSTEM is unified by using common libraries. Number of reflections $\leq 4800$, $\Sigma Z$ relationships $\leq 10,000$, number of atoms $\leq 200$ and number of grids in one section $\leq 200 \times 200$.

### Practical tests

DS*SYSTEM has been developed on several computers such as the Cray-1 supercomputer, the mainframe CDC, 16 bit PC to 32 bit PC. We use the NEC-9801 PC which is the most common in Japan and the Ricoh LP3320-SP4 laser printer. More than 50 organic compounds have been solved by PSL + SEARCHER for an automatic heavy-atom analysis which derives from a heavy-atom position the positions of all the atoms without any chemical assumptions, and more than 21 compounds have been determined by DIRECTER without any human intervention.

We summarize the results of computational time and the number of atoms obtained by PSL + SEARCHER on seven organic compounds in Table 1. Another seven organic compounds solved by DIRECTER are shown in Table 2 compared with MULTAN78.

### Summary and concluding remarks

Structure analysis with the heavy-atom method takes 2 min by Cray-1 supercomputer and about 2 h by a 32 bit PC (20 MHz) using PSL + SEARCHER. MULTAN78 can solve the structures within 15 min, whereas DIRECTER takes from 1 h to 1 d with the PC. Comparing the computational speed, the PC has the capability of one sixtieth of a Cray-1, one third of a CDC6600 and 1.5 times a CDC3600. The CPU speed of PCs will be quicker in the near future and it is a matter of time before the capability of the PC exceeds the old mainframe computer.

If part of the molecular geometry is known or some fragments are obtained by DIRECTER, SEARCHER is a powerful tool to find the remaining atoms from the viewpoint of saving time and cost. SEARCHER finds the positions of all atoms and draws a two-dimensional projection diagram of molecules in a unit cell automatically.

There are many merits of using a PC such as computational time, ease of operation to compute several program steps and cost efficiency. Non-crystallographers can determine the crystal structures of their own organic compounds using the PC version of DS*SYSTEM.

The crystallographic program system installed on both supercomputers and PC will become popular for researchers who can select which computer to use according to the scale of the compound.
Table 1. Comparison of computer speeds with heavy-atom analysis programs

<table>
<thead>
<tr>
<th>Compound</th>
<th>Space group</th>
<th>DIRECTER (PSL)</th>
<th>MULTAN78</th>
<th>DIRECTER (SEARCHER)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>806 8036 8038</td>
<td>806 8036 8038</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td>(8MHz) (16MHz)</td>
<td>(20MHz)</td>
<td>(8MHz) (16MHz) (20MHz)</td>
</tr>
<tr>
<td>( \text{Cs}_2 \text{H}_6 \text{Br}_2 \text{N}_2 \text{O}_5 )</td>
<td>( P_2_1/c )</td>
<td>5:11s 8min34s 5min03s 4min10s</td>
<td>9min33s 1h29min24s</td>
<td>45:1s 3h35min30s 1h13min03s 57min23s</td>
</tr>
<tr>
<td>( \text{Cs}_2 \text{H}_6 \text{O}_8 )</td>
<td>( P_2_2_2_2 )</td>
<td>5:15s 8min34s 5min03s 4min10s</td>
<td>9min33s 1h29min24s</td>
<td>45:1s 3h35min30s 1h13min03s 57min23s</td>
</tr>
<tr>
<td>( \text{Cs}_2 \text{H}_6 \text{Br}_2 \text{O}_7 )</td>
<td>( P_2_2_2_2 )</td>
<td>8:49s 34min34s 20min55s 16min40s</td>
<td>14min36s 20min42s</td>
<td>56:4s 2h43min47s 8h23min28s 2h45min47s</td>
</tr>
<tr>
<td>( \text{Cs}_2 \text{H}_6 \text{Br}_2 \text{O}_5 )</td>
<td>( P_1 )</td>
<td>11:96s 1h07min38s 25min17s 20min09s</td>
<td>22min01s 91:5s</td>
<td>38min14s 2h13min17s 48min38s 23:23</td>
</tr>
<tr>
<td>( \text{Cs}_2 \text{H}_6 \text{Br}_2 \text{O}_5 )</td>
<td>( P_1 )</td>
<td>7:52s 17min57s 8min59s 7min14s</td>
<td>14min36s 20min42s</td>
<td>56:4s 2h43min47s 8h23min28s 2h45min47s</td>
</tr>
<tr>
<td>( \text{Cs}_2 \text{H}_6 \text{Br}_2 \text{O}_5 )</td>
<td>( P_1 )</td>
<td>23:50s 52min20s 32min28s 25min38s</td>
<td>19min44s 94:4s</td>
<td>38min14s 2h13min17s 48min38s 23:23</td>
</tr>
<tr>
<td>( \text{Cs}_2 \text{H}_6 \text{Br}_2 \text{O}_5 )</td>
<td>( P_1 )</td>
<td>3:60s 9min04s 4min11s 3min22s</td>
<td>19min44s 94:4s</td>
<td>38min14s 2h13min17s 48min38s 23:23</td>
</tr>
</tbody>
</table>

The computational times of PSL, SEARCHER and DIRECTER could be made shorter by applying a fast Fourier transform (FFT) in the Fourier summations.

The DS*SYSTEM described in this paper is available in executable form from the authors.

## References


