Acta Cryst. (1994). D**50**, 760–763

The *CCP*4 Suite: Programs for Protein Crystallography

BY COLLABORATIVE COMPUTATIONAL PROJECT, NUMBER 4*

*SERC Daresbury Laboratory, Warrington WA*4 4*AD, England*

(*Received* 8 *December* 1993; *accepted* 21 *March* 1994)

Abstract

The *CCP*4 (Collaborative Computational Project, number 4) program suite is a collection of programs and associated data and subroutine libraries which can be used for macromolecular structure determination by X-ray crystallography. The suite is designed to be flexible, allowing users a number of methods of achieving their aims and so there may be more than one program to cover each function. The programs are written mainly in standard Fortran77. They are from a wide variety of sources but are connected by standard data file formats. The package has been ported to all the major platforms under both Unix and VMS. The suite is distributed by anonymous ftp from Daresbury Laboratory and is widely used throughout the world.

Introduction

CCP4 (Collaborative Computational Project, number 4) was established in 1979 and involved a group of protein crystallographers who have pooled their programming effort and expertise to create a suite of computer programs for the processing and analysis of protein crystallographic data. The project started in the UK, supported by the Science and Engineering Research Council (SERC), but is now integrated with the European Science Foundation (ESF) Network of the European Association of the Crystallography of Biological Macromolecules (EACBM).

The *CCP*4 program suite is a set of separate programs which communicate *via* standard data file formats. The advantage of this organization is that it is easy to add new programs or to modify existing ones without interfering with the other parts of the suite and it is possible to combine programs in new ways, linked by the Unix shell or VMS command language. Most of the computer code is written in standard Fortran77 and conversion of new programs to follow the *CCP*4 standard is relatively staightforward providing that they are written in standard Fortran77.

* Contact person S. Bailey (CCP4 Secretary), SERC Daresbury Laboratory, Warrington WA4 4AD, England.

Table 1. *Examples of protein structures recently determined using CCP*4 *programs*

Protein	Reference
δ-Endotoxin	Li *et al.* (1991)
U1 small nuclear ribonucleoprotein-A	Nagai *et al.* (1990)
F1 ATPase	Abrahams *et al.* (1993)
Galactose oxidase	Ito *et al.* (1991)
Mouse major urinary protein	Bocskei *et al.* (1992)
Chicken annexin V	Bewley *et al.* (1993)
Mouse DHFR	Groom *et al.* (1991)
Tyrosine phenol-lyase	Antson *et al.* (1992)
Rubredoxin	Dauter *et al.* (1992)
Aldose reductase	Tête-Favier *et al.* (1993)
NII-Domain from duck ovotransferrin	Lindley *et al.* (1993)
Concanavalin A	Naismith *et al.* (1993)
Trypanothione reductase	Hunter *et al.* (1992)
Ribulose 1,5-biphosphate carboxylase/oxygenase	Newman *et al.* (1993)
E. coli triosephosphate isomerase	Noble *et al.* (1993)
Carboxypeptidase A/L-phenyl lactate complex	Teplyakov *et al.* (1993)
Rat CD4, domains 3 and 4	Brady *et al.* (1993)
Chimeric Fab′ fragment	Brady *et al.* (1992)
Human H ferritin	Lawson *et al.* (1991)
Glutamate dehydrogenase	Baker *et al.* (1992)

Briefly, the suite contains programs for the reduction and analysis of intensity data, structure solution by isomorphous replacement and molecular replacement, least-squares refinement and analysis of the structure. There are programs for displaying electron-density maps and plotting molecules. The suite has been used for the determination and refinement of a large number of protein structures, of which some recent examples are listed in Table 1.

The program suite

The program package comprises several different elements which are briefly described.

Programs

The programs have been collected from a variety of sources and cover most aspects of macromolecular crystallography, with alternatives for some processes. They are distributed by CCP4 thanks to the generosity of the program authors. Table 2 lists the programs which are currently distributed as part of the suite.

Table 2. *CCP4 programs and their functions*

Program	Function	Program reference (if program published)
ABSCALE	Film and image-plate scaling	
ABSURD	Initial processing of intensity files from *MADNES*	
ACT	Analysis of coordinates	
AGROVATA/ROTAVATA	Scaling intensities and averaging repeated measurements	
ALMN	Crowther's rotation function using *FFT*	Crowther (1972)
AMORE	Molecular replacement	Navaza (1990)
CAD	Collecting assorted reflection data	
COMPLETE	Statistics on completeness of the data set	
CONTACT	Calculates contacts in protein structure	
COORDCONV	Interconverts various coordinate formats	
DISTANG	Distances and angles in protein molecule	
ECALC	Calculates normalized structure amplitudes	
ENVELOPE	Generates molecular envelope for solvent flattening	Wang (1985), Leslie (1987)
EXTEND	Extends map to specified volume of the unit cell	
F2MTZ	Converts a formatted reflection file to MTZ file	
FFT, FFTBIG	Crystallographic fast Fourier transformation	Ten Eyck (1973)
FHSCAL	Scaling derivative to native by Kraut's method	
FLATMAP	Flattens solvent region of electron-density map	Wang (1985), Leslie (1987)
GENSYM	Generate sites by symmetry	
HEAVY	Heavy-atom refinement and phasing program	Terwilliger & Eisenberg (1983, 1987)
HKLPLOT	Plots a pseudo-precession picture from a reflection file	
ICOEFL	Scaling of multiple F_{calc}s to F_{obs}	
LCF2MTZ	Converts an LCF reflection file to MTZ one	
LSQKAB	Optimize fit of atomic coordinates from two data sets	Kabsch (1978)
MAPSIG	Print statistics on signal/noise for translation-function map	
MAPTONA4	Converts binary map file to na4 format and reverse	
MLPHARE	Phase calculation and refinement	Otwinowski (1991)
MTZ2VARIOUS	Makes ASCII reflection files from MTZ file	
MTZDUMP	Displays the contents of the MTZ file	
MTZTONA4, NA4TOMTZ	Transforms MTZ file to transferable ASCII file and reverse	
MTZUTILS	Reflection data file utility program	
OVERLAPMAP	Calculate the overlap of two maps	
PDBSET	Various useful manipulations on coordinate files	
PEAKMAX	Search for peaks in the electron-density map	
PLTDEV	Convert plot84 metafiles to other graphics formats	
NPO	Plot molecules and electron-density maps	
POLARRFN	Fast rotation function in polar coordinates	
POSTREF	Post-refinement of film data	Winkler *et al.* (1979)
PRMAP	Prints sections of electron-density maps	
PROCHECK	Protein stereochemistry checking	Laskowski *et al.* (1993)
PROLSQ	Structure refinement	Hendrickson (1985)
PROTIN	Prepares file with restraints for *PROLSQ*	Hendrickson (1985)
REINDEX	Reindex MTZ reflections file	
ROTAPREP	Produces multirecord MTZ file	
ROTMAT	Interconverts *CCP4/MERLOT/X-PLOR* rotation angles	
RSEARCH	*R*-factor search	
RSPS	Heavy-atom positions from derivative difference Patterson maps	Knight (1989)
RSTATS	Scales two data sets	
SCALEIT	Various derivative and native scaling	
SFALL	Structure-factor calculations using FFT	
SIGMAA	Phase combination	Read (1986)
SORTMTZ	Sorting MTZ file	
SURFACE	Accessible surface area	
TFFC	Translation function	Tickle (1992)
TRACER	Reduced cell calculations	
TRUNCATE	Converts intensities to amplitudes	French & Wilson (1978)
UNDUPL	Remove duplicates from *MADNES* data, after *ABSURD*	
UNIQUE	Generate unique reflection data set	
VECREF	Vector space refinement of the heavy atoms	Tickle (1991)
VECTORS	Generates Patterson vectors from atomic coordinates	
VOLUME	Polyhedral volume around selected atoms	
WATERSORT	Assign waters to nearest protein atoms	
WATERTIDY	Rearranges water position	
WATPEAK	Selects peaks	
WILSON	Produces Wilson plot	

Library

CCP4 maintains a library of subroutines for performing the basic crystallographic and programming operations. There are routines for reading and writing the standard format files for reflections, atomic coordinates, maps and plotting, and routines for parsing the keyworded input. The library also contains forward and reverse fast Fourier transform routines (Ten Eyck, 1973), routines for handling crystallographic symmetry, some mathematical

functions and a small number of machine-specific routines.

Documentation

The documentation is in several sections. The *CCP*4 manual describes installation and gives an overview of the suite with information on how to use the programs to perform various tasks. There is specific documentation for the individual programs, which is available in the form of 'man' pages on Unix systems, and there is documentation for the subroutine libraries. The suite also contains example files for many crystallography procedures and related test data.

Using the programs

In order to use the programs certain environment variables need to be set to appropriate values. A setup script is supplied to do this. It requires initial modification to adapt it to an individual system and should be executed by *CCP*4 users when they log in. The programs are run interactively or in batch mode using shell scripts with Unix systems and command files with VMS systems. Example scripts and command files are distributed which can be modified by the user. The control data for most programs is keyworded and is of the form of records with a leading keyword, usually followed by arguments which might be numbers or strings or keyword/value pairs. Log files are produced which contain information on the progress of the program and which may be printed. Tables within the log files are formatted so that statistics can be extracted and plotted as graphs using the facility *Xloggraph* which is supplied with the suite.

File formats

There are four types of file format for reflection data, map data, coordinate data and graphics metafiles. The coordinate data files are ASCII but the map and reflection data files are binary. The binary format is compact and can be read fast without loss of numerical accuracy relative to a text format. To ease the transfer of binary files, the file headers contain information about the computer on which they were created. This information is used by the subroutines which read and write reflection files to allow files created by *CCP*4 programs to be used by *CCP*4 programs on most types of computer in a manner which is transparent to the user.

The reflection file format is known as MTZ and it uses fixed length records with, in general, four bytes for each data item, with a minimum of three and a maximum of 200 data items per record. The files comprise two classes of record; header records and reflection data records. The header records contain information such as titles, cell parameters and symmetry operators. The reflection data are stored notionally as columns of real numbers, each referred to by an alphanumeric label. The first three columns are usually the Miller indices of the reflection, with labels *H*, *K* and *L*. The programs allow the user to set up a correspondence between the labels in the data file and names of the data columns expected by the program. *Standard* MTZ files have one record per reflection. However, during the initial data processing, unmerged data are stored in *multirecord* files in which each record represents one observation.

The standard format adopted for coordinate data is that used in the Brookhaven Protein Data Bank (Abola, Bernstein, Bryant, Koetzle & Wang, 1987; Bernstein *et al.*, 1977). The programs of the suite will handle either complete files or files containing only a subset of the types of records which may be present in a complete file, usually the ATOM and HETATM records. There is an intention to move to the CIF format when this becomes the standard format for storage of macromolecular coordinate data.

Maps are stored in a randomly accessible binary file as a three-dimensional array preceded by a header which contains all the necessary information about the map including dimensions of the array, cell parameters, symmetry information and maximum, minimum and mean density values. *CCP*4 currently uses a graphics metafile format called PLOT84 and routines are supplied for its implementation. Programs are distributed to convert PLOT84 into other graphics formats including PostScript.

Installation

The program suite has been implemented on a large number of hardware platforms under the Unix and VMS operating systems. Installation is straightforward and full instructions are given in the *CCP*4 manual. Installation under Unix utilizes a script to configure to the appropriate hardware and operating system variant prior to building. Command files are provided to carry out the installation under VMS. A small number of the library routines and one of the programs are written in C.

Distribution

The program suite is licensed free to academic institutes by Internet ftp or on a variety of media for a small handling/media charge. The programs may be obtained by Internet ftp from anonymous@ccp4.dl.ac.uk:pub/ccp4. Separate arrangements are made for commercial organizations who

should contact CCP4 directly. For further details about *CCP*4 or to obtain the programs please contact the CCP4 Secretary at Daresbury Laboratory (email: ccp4@dl.ac.uk).

A large number of people have contributed to CCP4 over the years and we thank them for their time and effort. CCP4 is supported by the SERC and the ESF Network of the EACBM.

References

ABOLA, E., BERNSTEIN, F. C., BRYANT, S. H., KOETZLE, T. F. & WANG, J. (1987). *Crystallographic Databases – Information Content, Software Systems, Scientific Applications*, edited by F. H. ALLEN, G. BERGERHOFF & R. SIEVERS, pp. 107–132. Bonn/Cambridge/Chester: Data Commission of the International Union of Crystallography.

ABRAHAMS, J. P., LUTTER, R., TODD, R. J., VAN RAAIJ, M. J., LESLIE, A. G. W. & WALKER, J. E. (1993). *EMBO J.* **12**, 1775–1780.

ANTSON, A. A., STROKOPYTOV, B. V., MURSHUDOV, G. N., ISUPOV, M. N., HARUTYUNYAN, E. H., DEMIDKINA, T. V., VASSYLYEV, D. G., DAUTER, Z., TERRY, H. & WILSON, K. (1992). *FEBS Lett.* **302**, 256–260.

BAKER, P. J., BRITTON, K. L., ENGEL, P. C., FARRANTS, G. W., LILLEY, K. S., RICE, D. W. & STILLMAN, T. J. (1992). *Proteins Struct. Funct. Genet.* **12**, 75–86.

BERNSTEIN, F. C., KOETZLE, T. F., WILLIAMS, G. J. B., MEYER, E. F., BRICE, M. D., ROGERS, J. R., KENNARD, O., SHIMANOUCHI, T. & TASUMI, M. (1977). *J. Mol. Biol.* **112**, 535–542.

BEWLEY, M. C., BOUSTEAD, C. M., WALKER, J. H. & WALLER, D. A. (1993). *Biochemistry*, **32**, 3923–3929.

BOCSKEI, Z., GROOM, C. R., FLOWER, D. R., WRIGHT, C. E., PHILLIPS, S. E. V., CAVAGGIONI, A., FINDLAY, J. B. C. & NORTH, A. C. T. (1992). *Nature (London)*, **360**, 186–188.

BRADY, R. L., DODSON, E. J., DODSON, G. G., LANGE, G., DAVIS, S. J., WILLIAMS, A. F. & BARCLAY, A. N. (1993). *Science*, **260**, 979–983.

BRADY, R. L., EDWARDS, D. J., HUBBARD, R. E., JIANG, J.-S., LANGE, G., ROBERTS, S. M., TODD, R. J., ADAIR, J. S., EMTAGE, J. S., KING, D. J. & LOW, D. C. (1992). *J. Mol. Biol.* **227**, 253–264.

CROWTHER, R. A. (1972). *The Molecular Replacement Method*, edited by M. G. ROSSMANN, pp. 173–178. New York: Gordon and Breach.

DAUTER, Z., SIEKER, L. C. & WILSON, K. S. (1992). *Acta Cryst.* B**48**, 42–59.

FRENCH, G. S. & WILSON, K. S. (1978). *Acta Cryst.* A**34**, 517–525.

GROOM, C. R., THILLET, J., NORTH, A. C. T., PICTET, R. & GEDDES, A. J. (1991). *J. Biol. Chem.* **266**, 19890–19893.

HENDRICKSON, W. A. (1985). *Methods Enzymol.* **115**, 252–270.

HUNTER, W. N., BAILEY, S., HABASH, J., HARROP, S. J., HELLIWELL, J. R., ABOAGYE-KWARTENG, T., SMITH, K. & FAIRLAMB, A. H. (1992). *J. Mol. Biol.* **227**, 322–333.

ITO, N., PHILLIPS, S. E. V., STEVENS, C., OGEL, Z. B., MCPHERSON, M. J., KEEN, J. N., YADAV, K. D. S. & KNOWLES, P. F. (1991). *Nature (London)*, **350**, 87–90.

KABSCH, W. (1978). *Acta Cryst.* A**32**, 922–923.

KNIGHT, S. (1989). PhD thesis, Swedish Univ. of Agricultural Sciences, Uppsala, Sweden.

LASKOWSKI, R. A., MACARTHUR, M. W., MOSS, D. S. & THORNTON, J. M. (1993). *J. Appl. Cryst.* **26**, 283–291.

LAWSON, D. M., ARTYMIUK, P. J., YEWDALL, S. J., SMITH, J. M. A., LIVINGSTONE, J. C., TREFFRY, A., LUZZAGO, A., LEVI, S., AROSIO, P., CESARENI, G., THOMAS, C. D., SHAW, W. V. & HARRISON, P. M. (1991). *Nature (London)* **349**, 541–544.

LESLIE, A. G. W. (1987). *Acta Cryst.* A**34**, 134–136.

LI, J., CARROLL, J. & ELLAR, D. J. (1991). *Nature (London)*, **353**, 815–821.

LINDLEY, P. F., BAJAJ, M., EVANS, R. W., GARRATT, R. C., HASNAIN, S. S., JHOTI, H., KUSER, P., NEU, M., PATEL, K., SARRA, R., STRANGE, R. & WALTON, A. (1993). *Acta Cryst.* D**49**, 292–304.

NAGAI, K., ONBRIDGE, C., JESSEN, T. H., LI, J. & EVANS, P. R. (1990). *Nature (London)*, **348**, 515–520.

NAISMITH, J. H., HABASH, J., HARROP, S., HELLIWELL, J. R., HUNTER, W. N., WAN, T. C. M. & WEISGERBER, S. (1993). *Acta Cryst.* D**49**, 561–571.

NAVAZA, J. (1990). *Acta Cryst.* A**46**, 619–620.

NEWMAN, J., BRÄNDÉN, C.-I. & JONES, T. A. (1993). *Acta Cryst.* D**49**, 548–560.

NOBLE, M. E. M., ZEELEN, J. PH., WIERENGA, R. K., MAINFROID, V., GORAJ, K., GOHIMONT, A. C. & MARTIAL, J. A. (1993). *Acta Cryst.* D**49**, 403–417.

OTWINOWSKI, W. (1991). *Isomorphous Replacement and Anomalous Scattering, Proceedings of the CCP*4 *Study Weekend*, edited by W. WOLF, P. R. EVANS & A. G. W. LESLIE, pp. 80–86. Warrington: SERC Daresbury Laboratory.

READ, R. J. (1986). *Acta Cryst.* A**42**, 140–149.

TEN EYCK, L. (1973). *Acta Cryst.* A**33**, 486–492.

TEPLYAKOV, A., WILSON, K. S., ORIOLI, P. & MANGARI, S. (1993). *Acta Cryst.* D**49**, 534–540.

TERWILLIGER, T. C. & EISENBERG, D. (1983). *Acta Cryst.* A**39**, 813–817.

TERWILLIGER, T. C. & EISENBERG, D. (1987). *Acta Cryst.* A**43**, 6–13.

TÊTE-FAVIER, F., RONDEAU, J.-M., PODJARNY, A. & MORAS, D. (1993). *Acta Cryst.* D**49**, 246–256.

TICKLE, I. J. (1991). *Isomorphous Replacement and Anomalous Scattering, Proceedings of the CCP*4 *Study Weekend*, edited by W. WOLF, P. R. EVANS & A. G. W. LESLIE. pp. 87–95. Warrington: SERC Daresbury Laboratory.

TICKLE, I. J. (1992). *Molecular Replacement, Proceedings of the CCP*4 *Study Weekend*, edited by E. J. DODSON, S. GOVER & W. WOLF, pp. 20–32. Warrington: SERC Daresbury Laboratory.

WANG, B. C. (1985). *Methods Enzymol.* **115**, 90–112.

WINKLER, F. K., SCHUTT, C. E. & HARRISON, S. C. (1979). *Acta Cryst.* A**35**, 901–911.