Computer Programs for Automatic Interpretation of Protein Difference Patterson and for Automatic Solving of Organic Structures

F PAVELČÍK[1] AND E J DODSON[2]

1 *Department of Inorganic Chemistry, Faculty of Natural Sciences, Comenius University, SK-842 15 Bratislava, Slovak Republic*
2 *Protein Structure Group, Department of Chemistry, York University, Heslington, YO1 5DD York, United Kingdom*

Key words: Patterson, organic structures

Abstract. Automatic methods have been applied to interpretation of macromolecular difference Patterson These methods combined with Fourier recycling can also be used for structure determination of organic structures Test calculations with programs XFPM and XFPL showed that these methods are useful alternative to traditional methods

Introduction

The computer program was born in 1986 (Pavelcik 1986) under the name XFPS as a program for automated Fourier, Patterson and superposition methods The program was later extended by second and third order symmetry minimum function, cross-vector function, by a full symmetry minimum translation function (Pavelcik 1988), procedures utilising Patterson peaks (Pavelcik *et al* 1992a) by an expert and graphics routines and released as the program for automatic structure determination (Pavelcik *et al* 1992b) Recently the program has been divided into a program for automatic structure determination and refinement, XFPA, and to a specialised program for interpretation of protein difference Patterson function, XFPM

Structure determination of organic structures is dominated by direct methods, which in 60-ties and early 70-ties replaced Patterson methods Recently (Pavelcik 1988, Pavelcik *et al* 1992a, Pavelcik 1994) powerful Patterson methods for automatic determination of heavy atom structures have been developed These methods are combined with Fourier recycling methods, which are very powerful tools for phase refinement (Pavelcik 1994, 1997) In this paper preliminary results of the fully automatic (Patterson based) structure determination methods combined with automatic isotropic refinement are described for organic structures These methods are of potential use for direct solving of small proteins measured at atomic resolution

The heavy atom replacement method remains the major tool to phase new macromolecular structures A difference Patterson shows peaks corresponding to the heavy atoms plus noise Several algorithms have been developed for automatic interpretation of the Patterson map e g HASSP, GROPAT or SHELXS The significant updating of the XFPS for analysis of difference Patterson function in the macromolecular crystallography represents a new tool for protein structure analysis

Methods

Four strategies for automatic structure determination of organic structures and for automatic interpretation of difference Patterson have been developed Two of strategies for structure determination are new, two are modification and extension of strategies A and D of Pavelcik (1994) The principal difference with (1994) paper is the change from a single-solution to an automatic multi-solution The phase problem is solved by superposition method The full symmetry atomic minimum superposition (Simpson *et al* 1965, Pavelcik 1986) in the crystal space group, or a minimum vector superposition (Buerger 1959) in the space group P1, are used Starting atoms for the atomic minimum superposition are generated either by symmetry minimum function (Simpson *et al* 1965, Pavelcik 1986, 1990) or full symmetry (SMF based) minimum translation function based on one Patterson vector (Pavelcik 1988) Shift vectors for vector superposition are either simple Patterson vectors, or several vectors resulting from the cross-vector superposition (Iljukhin *et al* 1981, Pavelcik 1994) Peaks from superposition map form input into Fourier recycling in crystal or P1 space groups Slightly modified computer implementation of Pavelcik (1994) is used Number of Fourier cycles depends on number of atoms in the asymmetric unit Structure solved in P1 space group is shifted to conventional cell origin (Pavelcik 1994) and further refined in the crystal space group Structure determination is supported by an automatic structure expansion and refinement (Pavelcik 1997), but only isotropic refinement was used for testing purposes

Computer program

For fully automatic Patterson interpretation of protein difference Patterson only following instructions are needed

CELL λ, a, b, c, α, β, γ, sp gr No
CHEM Z, Hg 3, N 100
SOLV n

The most relevant changes with respect to the program XFPS are

Input of macromolecular reflection data, these data can be prepared e g by CCP4 program mtz2various User oriented input Improved symmetry minimum function for atoms in special positions and second order symmetry minimum map Cross-vector superposition Minimum vector superposition and the SMF based search for the cell origin Modified and more powerful Patterson multi-solution based on fragment generators and cross-vector function with combined figure of merit Test for pseudo-symmetry in the atomic minimum superposition Fourier recycling Pseudo-Gaussian peak integrals for sorting electron-density peaks Five expert routines for Patterson multi-solution

The program has been written in FORTRAN77(90) and has been tested on an IBM compatible PC under MS DOS, Windows95 and LINUX, and on SGI under IRIX 6 2 User's instructions, program description and test structures are supplied as an ASCII text file MS WORD document and HTML files are under development The programs XFPA and XFPM are available on request from F P and XFPM will be distributed as the part of CCP4 program

Results and Discussion

Automatic Patterson interpretation of difference Patterson has been tested on structures RNase, Hg and Pt derivatives, SmIR, Se anomalous derivative, OPAL, U anomalous

Table 1. Crystal data for test structures and results of automatic solution and isotropic refinement SOLV means the strategy

Code	Sp group	Formula	Z	SOLV	R
FUNG	Pccn	$C_{17}H_{22}ClN_2O_4PS_2$	8	0	15 4
BETA	1C2/c	$C_{18}H_{29}ClN_2O_5$	8	0	14 2
APAPA	$P4_12_12$	$C_{30}H_{49}N_{15}O_{22}P_2$	8	1	15 8
MEXI	P 1	$C_{11}H_{18}ClON$	4	1	11 1
BL238	P 1	$C_{11}H_{15}NaN_2O$	16	2	11 6
DIAM	$P4_2/c$	$C_{14}H_{20}O$	8	1	14 5
DIOL	I-42d	$C_{10}H_{18}O_2$	16	1	14 4
BED	I4	$C_{26}H_{26}N_4O_4$	8	1	15 2
NO55	Fdd2	$C_{20}H_{24}N_4$	16	1	10 6
TURIO	$P6_322$	$C_{15}H_{24}O_2$	12	1	13 0
SKN1	$P3_1$	$C_7H_{16}ClNO_4$	3	0	9 3
CIME	Cc	$C_{10}H_{18}N$ OS	4	0	7 7
AMIDE	$Pbc2_1$	$C_{14}H_{18}N$ O_4	4	1	9 9
KARB	C2	$C_{22}H_{25}N_2O_4$	4	1	11 7
LOG	$P2_12_12_1$	$C_{17}H_{26}O_{10}$	4	1	9 1
Pn1A	$P2_1$	C_6 $H_{91}N_{18}O_{25}S_4$	1	3	17 3
DCHP	$P6_2$	$C_{18}H_{29}N$ O_{12} P	12	4	12 0
CYCLOD	$P2_1$	$C_{42}H_{92}O_{46}$	2	4	8 0
PROP3	$P2_1/n$	$C_{28}H_{30}O_4$	4	1	19 4

derivative, VALER, Hg derivative HEXS Hg derivative All test structures have been automatically interpreted

Tests for organic structures were restricted to equal atom'structures and to organic structures not containing atoms heavier than Cl Examples of test data and results are given in Table 1 The results show that methods developed can be used for routine structure determination of small and intermediate organic structures Presence of heavier atom like Na^+ or P simplifies the process of structure determination and in most cases heavy atom strategy can be used The cyclodextrine with 88 atoms (no heavy atoms) in the asymmetric unit of $P2_1$ space group is one of the most difficult structures solved

Acknowledgements. The project was supported by the COPERNICUS program of EU, and Grant VEGA No 1/4088/97 of MS SR Authors are grateful to Dr Z Dauter for test data

References

Buerger M J (1959) Vector Space New York Willey

Iljukhin V V , Kuzmin E A and Belov N V (1981) New Approaches to the Patterson Function Moscow Nauka (In Russian)

Pavelcik F (1986) XPFS, a program for automatic Fourier, Patterson and superposition methods J Appl Cryst 19 488-491

Pavelcik F (1988) Patterson Oriented Automatic Structure Determination Getting a Good Start Acta Cryst A**44**, 724-729

Pavelcik F (1994) Patterson Oriented Automatic Structure Determination Deconvolution Techniques in Space Group P1 Acta Cryst A**50**, 467 474

Pavelcik F (1997) XFPA an update to XFPS, A program for automatic Patterson solution and refinement Poster no P4 3-11 ECM-17 Lisboa

Pavelcik F , Kuchta L , and Sivy J (1992a) Patterson-Oriented Automatic Structure Determination Utilizing Patterson Peaks Acta Cryst A**48**, 791-796

Pavelcik F , Sivy J , Rizzoli C and Andreetti G D (1992b) XFPS a program for automatic solution of crystal structure by Patterson, Fourier and superposition methods Recent developments J Appl Cryst **25**, 328 329

Simpson P G , Dobrott R D and Lipscomb W N (1965) Symmetry minimum and high order superposition functions Acta Cryst **18**, 169-179

An Evolutionary Link between Sporulation and Prophage Induction Revealed by X-Ray Crystallography

RICHARD J LEWIS[1], JAMES A BRANNIGAN[1], WENDY A OFFEN[1], ISSAR SMITH[2] AND ANTHONY J WILKINSON[1]

1 *Department of Chemistry, University of York, York Y01 5DD, UK*
2 *Public Health Research Institute, New York, New York 10016, USA*

Key words: sporulation, SinI-SinR, phage induction, Bacillus subtilis

Spore formation is an extreme response of some bacteria to adversity In *Bacillus subtilis* the proteins of the *sin*, sporulation inhibition, region form a component of an elaborate molecular circuitry which regulates the commitment to sporulation SinR is a tetrameric repressor protein which binds to the promoters of genes essential for entry into sporulation and prevents their transcription This repression is overcome through the activity of SinI which disrupts the SinR tetramer through the formation of a SinI-SinR heterodimer The interactions governing this curious quaternary transition are revealed in the crystal structure of the SinI-SinR complex The most striking, and unexpected, finding is that the tertiary structure of the DNA binding domain of SinR is identical to that of the corresponding domains of the repressor proteins (CI and Cro) of bacteriophage 434 which regulate lysis/lysogeny This structural similarity greatly exceeds that between SinR and any bacterial protein or between the 434 repressor proteins and their homologues in the closely related bacteriophage λ This close evolutionary relationship implied by the structures of SinR and the 434 repressors provokes comparison both of their functions, and of the developmental pathways in which the respective proteins participate

Acknowledgements. To the BBSRC and Welcome Trust for funding, Copernicus for travel and E J Dodson, G G Dodson and I Barak for advice and encouragement