

Berichte aus
Arbeitskreisen
der
DGK

Nr. 9



**VII. Workshop Powder Diffraction
Structure Determination
and Refinement
from Powder Diffraction Data**

Robert E. Dinnebier (ed.)

Deutsche Gesellschaft für Kristallographie
2000

Acknowledgement

The VII international workshop on powder diffraction, October 4. – 8, 2000 at the Laboratory of Crystallography of the University of Bayreuth (Germany) was organized by the Powder Diffraction Group of the German Society of Crystallography DGK. We are grateful to the

Universität Bayreuth



for supplying technical assistance, lecture halls and PC pools free of charge.

We are further grateful for financial support from

- **Deutsche Gesellschaft für Kristallographie (DGK)**



- **Deutsche Mineralogische Gesellschaft (DMG)**



- **International Union of Crystallography (IUCr)**



- **International Centre for Diffraction Data (ICDD)**



- **Clariant GmbH, Frankfurt**



- **Stoe & Cie, Darmstadt**



- **Bruker AXS, Karlsruhe**



- **Cambridge Crystallographic Data Centre (CCDC), Cambridge**

- **Molecular Simulation Inc. (MSI), Cambridge**



- **Böhringer Ingelheim, Biberach a.d. Riss**



- **Crystal Impact, Bonn**



- **BASF AG, Ludwigshafen**



- **Philips Analytical, Almelo**



PHILIPS

Due to the generous support, it was possible to keep registration costs for this workshop low and to award 15 travel grants to students to cover all local costs at Bayreuth as well as part of their travel costs.

Preface

The complexity of crystal structures which can be solved and/or refined from powder diffraction data has been increasing steadily throughout the last few years. In particular further developments of “traditional” algorithms for structure determination in reciprocal space as well as the application of global optimization algorithms in direct space have reached a level which for the first time allows powder diffraction to be used as an alternative technique to single crystal diffraction for inorganic and small molecule structures. As a consequence, the powder diffraction method is gaining a lot of interest, particularly in the chemical and pharmaceutical industry.

Despite this success, we can not yet speak of it being a routine method. Therefore, because this field of research is developing rapidly, so there is a need for suitable training.

How is the workshop organized ? As for the previous 6 workshops of this series, there are lectures, tasks and practical exercises but the training concept is different. In view of the short time available, the participant will decide by himself, which programs he wants to try out and which of the many offered exercises he wants to perform. All tutorials and most of the software can be found on the distributed workshop CD.

The course is intended for beginners as well as advanced users of powder diffraction methods in materials, geological, physical, chemical and pharmaceutical sciences. The course will provide an overview of the state of the art methods for structure determination and refinement from powder data. Examples at different levels of complexity will be provided including: data reduction from powder diffraction patterns taken at ambient and non-ambient conditions, peak fitting and intensity extraction (LeBail- and Pawley methods), traditional structure determination methods for powders (direct methods, Patterson search method), global optimization methods for structure determination in real space, combination of structure refinement and energy minimization, the Rietveld method (including the use of soft constraints and rigid bodies) and a comparison of Fourier analysis versus maximum entropy.

Due to the large number of contributions, it was necessary to publish the Rietveld tutorial for this workshop in a separate issue of this series (*Berichte aus Arbeitskreisen der DGK Nr. 8*).

All relevant information of this workshop is available online at <http://www.uni-bayreuth.de/departments/crystal/workshop2000/>.

CONTENTS

Workshop Schedule.....	1
Freely Available Software to Assist in Solving Structures from Powder Diffraction.....	7
<i>Lachlan M.D. Cranswick</i>	
EXPO2000: a New Package for <i>Ab-initio</i> Structure Solution from Powder Data.....	49
<i>Angela Altomare, Carmelo Giacovazzo , Antonietta Guagliardi, Anna Grazia Giuseppina Moliterni & Rosanna Rizzi</i>	
The Solution of Molecular Structures by Patterson Search Methods Using Powder Diffraction Intensity Data.....	75
<i>Jordi Rius</i>	
Monte Carlo Methods	85
<i>Maryjane Tremayne & Colin Seaton</i>	
Crystal Structure Determination of Tetracycline Hydrochloride with Powder Solve	97
<i>Marcus A. Neumann, Frank J. J. Leusen, G. Engel, C. Conesa-Moratilla & S. Wilke</i>	
DASH Tutorial 1.....	121
<i>Kenneth Shankland & William I. F. David</i>	
DASH Tutorial 2.....	141
<i>Harriott Nowell & Kenneth Shankland</i>	
Combined Method for "Ab Initio" Structure Solution from Powder Diffraction Data: The ENDEAVOUR, Software.....	149
<i>Holger Putz</i>	
Structure Determination from Powder Data with TOPAS.....	169
<i>Arnt Kern</i>	
Energy Minimization Techniques.....	181
<i>Martin U. Schmidt</i>	
Maximum Entropy Method Applied to Crystallographic Problems.....	191
<i>Martin Schneider & Sander van Smaalen</i>	
BRASS: The Bremen Rietveld Analysis and Structure Suite	201
<i>Reinhard X. Fischer, Thomas Messner & Dethart Kassner</i>	
High Pressure Powder Diffraction.....	209
<i>Ross J. Angel</i>	

Application of X-Ray Powder Diffraction in Pharmaceutical Sciences	229
<i>Peter Sieger</i>	
Structure Determination from X-Ray Powder Data of Notoriously Difficult Materials.....	239
<i>Hermann Gies</i>	
The Importance of the Inorganic Crystal Structure Database ICSD for the Application of Rietveld-Methods	253
<i>Rudolf Allmann</i>	

Workshop Schedule

Events	Building	Room
Registration Desk	Naturwissenschaften II	Foyer
Industrial Exhibition + Posters	Naturwissenschaften II	Foyer
Lectures	Naturwissenschaften II	H19
Exercises	BIX	24
	GEO	S24b
	Geisteswissenschaften II	U.17.2
Coffee/Tea Breaks	Naturwissenschaften II	Foyer
Lunch (Wednesday till Friday)	Mensa	
Pizza (Saturday)	Naturwissenschaften II	Foyer
Philips Mixer (Wednesday)	Bayerisches Geoinstitut (BGI)	2 nd floor
Conference Dinner (Friday)	Speisegaststätte "Röhrensee" Pottensteinerstrasse 5 Tel: 0921-64593	

Wednesday, October 4, 2000
Direct Space methods (Global optimization)

When ?	What ?	Who ?
8.00-10.00	<i>Registration</i>	
9.30	Opening and Welcome	R. Dinnebier, S. van Smaalen
9.45	Simulated Annealing (DASH program)	W.I.F.David
10.30	Global Optimization Algorithm	K. Shankland
<i>11.15</i>	<i>Coffee/Tea</i>	
11.45	Monte Carlo Methods	M. Tremayne
<i>12.30</i>	<i>Lunch Break (Mensa)</i>	
13.30	Full-profile Simulated Annealing Method for Solving Complex Molecular Structures from Powder Diffraction Data	P. Bruce, Y. Andreev
14.15	Powder Solve in Action	F. Leusen
<i>14.45</i>	<i>Coffee/Tea</i>	
15.15-18.45	Exercises Part I	All
<i>19.00</i>	<i>Mixer (sponsored by Philips Analytical)</i>	

Thursday, October 5, 2000
Direct-, Patterson methods/ Energy minimization

Time ?	What ?	Who ?
9.00	Advances into the EXPO Package	C. Giacobazzo
10.00	Patterson Search Methods	J. Rius
<i>11.00</i>	<i>Coffee/Tea</i>	
11.30	Energy Minimization Techniques	M .U. Schmidt
<i>12.30</i>	<i>Lunch Break (Mensa)</i>	
13.30	Simultaneous Optimization of Pattern Difference and Potential Energy: Structure Solution Using Endeavor	H. Putz
14.00	DASH Demonstration	S. Maginn
<i>14.30</i>	<i>Coffee/Tea</i>	
15.00-18.45	Exercises Part II	All

Friday, October 6, 2000
Rietveld refinement

Time ?	What ?	Who ?
9.00	New Rietveld Program JANA	S. van Smaalen
9.45	Presentation of a new Rietveld Program	R. X. Fischer
<i>10.30</i>	<i>Coffee/Tea</i>	
11.00	Case Study of a Complicated Rietveld Refinement using GSAS (Preparation for Exercise)	R. Dinnebier
<i>12.30</i>	<i>Lunch (Mensa)</i>	
13.30	Rietveld Refinement at High Pressure	R. Angel
14.30	Structure Determination from Powder and Single Crystal Data by Simulated Annealing.	A. Kern
<i>15.00</i>	<i>Coffee/Tea</i>	
15.30-19.00	Exercises part III	All
<i>19.30</i>	<i>Informal Get Together and Dinner</i>	

Saturday, October 7, 2000
Special strategies/ applications/ data bases

Time ?	What ?	Who ?
9.00	The ICSD Database	R. Allmann
9.30	The ICDD Database	J. Faber
10.00	The CCDC Database	J. Cole
<i>10.30</i>	<i>Coffee/Tea</i>	
11.00	Freely Available Software Tools to Aid in Structure Solution from Powder Diffraction	L. Cranswick
11.45	Structure Determination of Complicated Framework Structures from Powder Data	H. Gies
12.30	The Maximum Entropy Method	M. Schneider
<i>13.00</i>	<i>Pizza</i>	
14.00	Application of Powder Diffraction in Pharmaceutical Sciences	P. Sieger
<i>14.30</i>	<i>Coffee/Tea</i>	
15.00-19.00	Exercises / Demonstrations part IV	All

Sunday, October 8, 2000
Exercises

Time ?	What ?	Who ?
9.00-12.00	Exercises Part V	all
12.00	Closing and "Good-bye"	

Freely Available Software to Assist in Solving Structures from Powder Diffraction

Lachlan M.D. Cranswick

*CCP14 (Collaborative Computation Project Number 14 for Single Crystal
and Powder Diffraction)*

*CLRC Daresbury Synchrotron Laboratory, Warrington, Cheshire, WA4
4AD, UK*

WWW: <http://www.ccp14.ac.uk>

E-mail: L.Cranswick@dl.ac.uk

Solving a structure from powder diffraction data normally requires starting from raw data conversion through to displaying a publication quality structure model. To help make this type of endeavour a going concern, a large amount of crystallographic software is presently freely available via the Internet to academics and students (commercial users should contact the program authors to determine licence conditions). This ranges from data conversion, peak profiling, indexing, structure solution, crystal structure validation up to photo-realistic rendering of the final refined solution.

The following notes provide web links to most of the presently known and freely available software; and also tries to give a feel for what some of these programs look like. In the past few years there has been a trend for to put GUIs (Graphical User Interfaces) and advanced graphics onto freely available software to make them more powerful and friendly to use. These notes are as current as possible for August 2000.

Single crystal programs are also cited as presently they can be very useful in assisting in structure solution from powder diffraction. Users are advised that they can greatly enhance and improve the effectively of their crystallographic resources by investigating available and newly developed programs to build a software toolkit optimised to the phases systems and structures. The difference between an intractable or unsolvable structure and one that is relatively routine could purely be the difference in which software is applied to the problem. Many of the cited packages are under continual development; thus this document may quickly go out of with new functionality added to programs. Readers are encouraged to take nothing for granted and investigate for themselves the following mentioned programs in creating a robust tool set for the analysis of their crystallographic problems.

With regard to the web references provided in the following text. The Internet is presently very quite dynamic and unstable. Web and ftp sites that exist one month might disappear or move the next. Thus the reader should not consider these World Wide Web links to be written in stone. If a web site provided in the below text has ceased to exist, the keywords of the program name may allow convenient finding of this or related software by using the program title in a web search engine. Scientists can be most at risk at not knowing what they do not know. Thus much of this information is provided in helping keywords that have been found to work with World Wide Web search engines are also provided. Most of the below stated programs are also mirrored (with tutorials and hints for many of the programs) at the CCP14 website at <http://www.ccp14.ac.uk>.

(If links are broken, refer to the CCP14 Mirrors area:

<http://www.ccp14.ac.uk/mirror/>

or the IUCr's Sincris software list at:

<http://www.iucr.org/sincris-top/logiciel/>)

Agenda

- Has the structure been solved already?
- Powder X-ray Diffraction Search Match and Phase Identification
- Crystal Structure Databases
- Powder Data Conversion
- Structure Data Conversion and Transformation via Graphical User Interfaces
- Powder Pattern Viewing and Processing
- Peak Finding and Peak Profiling
- Powder Indexing
- Spacegroup Assignment
- Unit Cell Refinement
- Full Profile Fitting (Pawley, LeBail)
- Texture analysis software
- Single Crystal Suites
- Structure Solution Specifically for Powder Diffraction
- Structure Solution Using Single Crystal Software
- 2D to 3D molecular Model Generation
- Single Crystal Refinement programs for assisting in building up the structure
- Rietveld structure refinement
- Rietveld structure refinement software that is Fourier Map/Q Peak Friendly
- Hydrogen Placement Using Single Crystal Software
- Free Standing Powder and Single Crystal Fourier Map software
- Structure Validation including Platon/Addsym
- Crystallographic Structure Visualisation: during structure solution and refinement
- Powder Diffraction Pattern Calculation
- Photo realistic rendering of crystal structures and Fourier contour maps
- Miscellaneous Links
- Setting up dual and multiboot PC computers containing Windows and UNIX
- IUCr and ICSU sponsored Crystallographic Nexus CD-ROMs for academics and students isolated from the internet.
- Useful Crystallographic Websites
- On-line crystallographic web tutorials
- Web Search Engines

Phase ID and Search Match from Powder Diffraction Data

Search Match Databases

1. ICDD (commercial at this time costing US\$6,200 for a starting copy):
<http://www.icdd.com>
2. Pauling Database (expected release 2007?)
3. Consider making your own up from synthesised phases and calculated powder patterns.

Search Match Software (as of August 2000) :

Summary list of search-match software at:

<http://www.ccp14.ac.uk/solution/search-match.htm>

All search-match software should be considered commercial unless mentioned otherwise

AXES for DOS:	ftp://ftp.physic.ut.ee/pub/pc/axes/
Bede ZDS for Windows	http://www.bede.com/
Bruker "DIFFRACplus" for Windows	http://www.bruker-axs.com/
Crystallographica Search-Match" for Windows	http://www.crystallographica.co.uk
Diffraction Technology "Traces" for Windows	http://www.ozemail.com.au/~difftech/products/traces.htm
"DRXWin" for MS-Windows	http://icmuv.uv.es/drxwin/
Radicon "LookPDF" for Windows	http://www.radicon.xraysite.com/
Macdiff (freeware):	http://www.geol.uni-erlangen.de/macsoftware/macdiff/MacDiff.html
MacPDF for Mac	http://world.std.com/~crose/MacPDFWebSite/MacPDF_V3.html
MDI "Jade" for Windows	http://www.materialsdata.com/products.htm
MicroPDSM	?????
Portable Logic for UNIX (freeware)	http://www.ncnr.nist.gov/programs/crystallography/software/logic.html
PADS New Edition	E-mail: RMSKempten@aol.com
Philips Search-Match for MS-Windows	http://www-eu.analytical.philips.com/products/xrd/
Socabim	????
Siefert "RayfleX" Software for MS-Windows:	http://www.roentgenseifert.com/seif4.9.htm
Mark Raven - CSIRO, "XPLOT" for Windows:	http://www.clw.csiro.au/services/mineral/xplot.htm
"XPowder" for Windows	http://www.ugr.es/~jdmartin/
"ZDS" for DOS and Windows	http://krystal.karlov.mff.cuni.cz/xray/zds/zdscore.htm

Crystal Structure Databases (as of August 2000)

Database	Address	Specialisation	Global Free Access?
ICSD	CD-ROM: http://www.fiz-karlsruhe.de/icsd_.html WEB: http://barns.ill.fr/dif/icsd/	Inorganic and Minerals	No
CCDC/Camb ridge	http://www.ccdc.cam.ac.uk/	Organic and Organometallic	No
Crysmet	http://www.tothcanada.com	Metals and Alloys	No
MINCRYST	http://database.iem.ac.ru/mincryst/ http://database.iem.ac.ru/mincryst/s_full.php3	Minerals	Yes
Mineral Web	http://www.man.ac.uk/Geology/MineralWeb/Mineral_Web.html	Minerals	Yes
American Mineralogist	http://www.geo.arizona.edu/xtal-cgi/test/	Minerals	Yes
Lama Modulated Structures	http://www.cryst.ehu.es/icsdb/	Modulated Structures	Yes

(Academics and students may already have free access to many of these via national facilities.) e.g., EPSRC funded CDS (Chemical Database Service) for UK Students and Academics:

<http://cds3.dl.ac.uk/cds/cds.html>

Some single crystal suites have a point and click interface into the Cambridge database allowing connectivity checks, cell searches to be performed in an easy to do manner. Platon/System S is one of these style of programs which links into the CSD Quest program. While Crystals for Windows does dynamic bond-length and angle comparison of data from the Cambridge database.

Platon/System S/Pluton (by Ton Spek):

UNIX:

<http://www.cryst.chem.uu.nl/platon/>

<ftp://xraysoft.chem.uu.nl/pub/>

Windows (ported by Louis Farrugia):

<http://www.chem.gla.ac.uk/~louis/software/>

Crystals (D. Watkin, R. Cooper, et al):

<http://www.xtl.ox.ac.uk/> <ftp://darkstar.xtl.ox.ac.uk/pub/crys32/>

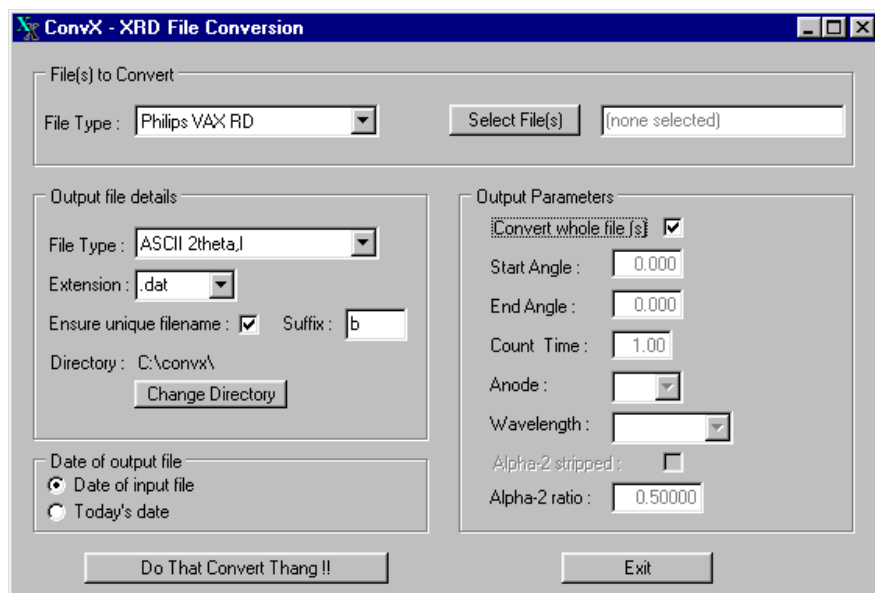
Powder Data Conversion

Summary list at: <http://www.ccp14.ac.uk/solution/powderdataconv/>

Tutorial on techniques to convert X-ray powder diffraction (Debye-Scherrer) to a synthetic digital diffractogram.	http://www.indiana.edu/~xl10rd/XRDFilmTutorial.htm		
ConvX Windows Mass Data Conversion	http://www.ceramics.irl.cri.nz/Convert.htm	Philips VAX RD, ASCII X-Y, Siemens/Bruker/DiffracPlus (RAW), Philips Binary (RD, SD), Sietronics (CPI), GSAS (DAT), DBWS Based (LHPM, RIET7, Fullprof) (DAT), ScanPI (INT)	Philips VAX RD, ASCII X-Y, Siemens/Bruker/DiffracPlus (RAW), Philips Binary (RD, SD) Sietronics (CPI), GSAS (DAT), DBWS Based (LHPM, RIET7, Fullprof) (DAT), ScanPI (INT)
DLConvert	http://www.ccp14.ac.uk/projects/dl-conv/index.htm	Daresbury Laboratory Beamlines: 9.1 angular dispersive, 2.3 angular dispersive, 16.4 Dispersive; Binary MCA Ortec CHN, Argonne Energy Dispersive Data; XT Data	XFIT Dat, CPI, GSAS, XY, linear interpolated data
Powder v2.00 for Windows	http://www.chem.t.u-tokyo.ac.jp/appchem/labs/kitazawa/dragoe/html/software.html	DBWS, GSAS CW, GSAS CW ESD, LHPM, Philips RD/SD binary, Philips UDF, MXP18 Binary, RIET7, Scintag, Siemens ASCII, Sietronics CPI, WPPF/Profit, Y free ascii, XY free ascii, free ascii. Line; X, XY, XYZ	DBWS, GSAS CW, GSAS CW ESD, LHPM, Philips RD/SD binary, Philips UDF, MXP18 Binary, RIET7, Scintag, Siemens ASCII, Sietronics CPI, WPPF/Profit, Y free ascii, XY free ascii, XYZ free ascii. Line; X, XY, XYZ

PowderX for Windows	http://www.ccp14.ac.uk/tutorial/powderx/index.htm	Mac Science ASCII, BD90 (Raw), X-Y, Rigaku (DAT), Sietronics (CPI), TsingHua Rigaku (USR) Siemens/Bruker ASCII (UXD), Siemens/Bruker Binary (RAW), Philips ASCII (UDF), Philips Binary (RD) Mac Science Binary, RIET7 (DAT), ORTEC Maestro (CHN)	ALLHKL (POW), Sietronics (CPI), FOURYA/ XFIT/ Koalariet (XDD), Fullprof (DAT), GSAS (DAT), Rietan (INT), Simpro (DUI), X-Y (XRD), DBWS (DAT), LHPM (DAT)
Winfit for Windows	http://www.geol.uni-erlangen.de/html/software/soft.html	Geol. Dept. Erlangen (DFA), Siemens/Bruker Diffrac V 2.1 (1 range) (RAW), (TRU), ASCII X-Y, ICDD Format (PD3), ZDS (ZDS), Software of F Nieto (CRI), Philips (UDF), Philips Binary (RD), STOE (RAW), JADE (MDI), MacDiff of Rainer Petschick (DIF), Converted RAW File (Bish, Eberl,..) (ASC), XDA Rietveld (XDA)	Siemens/Bruker Diffrac V 2.1 (1 range) (RAW), Philips Binary (RD), ASCII X-Y, XDA Rietveld (XDA)

ConvX Powder Diffraction Data Interconverter Interface

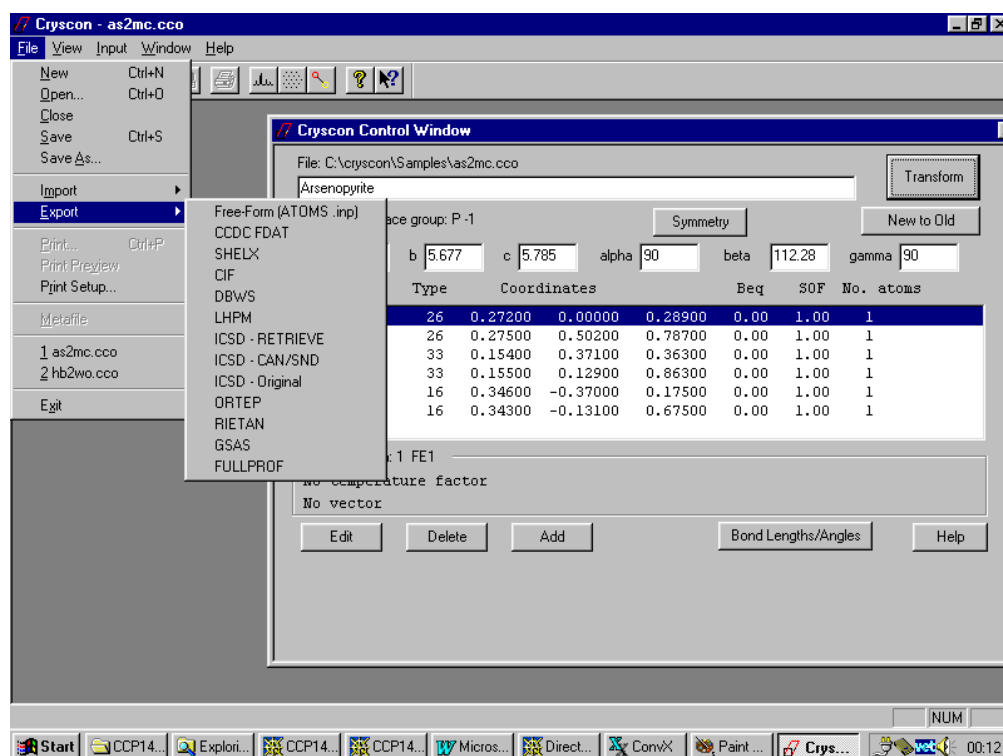


Structure Data Conversion and Transformation via Graphical User

Name	Web	Conv ersion ?	Trans form- ation	HKL data	Formats In	Formats Out
Cryscon	http://www.shapesoftware.com	Yes	Yes	Yes	Free Format, CCDC FDAT, Shelx, CIF, DBWS, LHPM, ICSD, ORTEP, XTLVIEW, ATOMS, RIETAN, GSAS, Am Mineral, Fullprof, Vibratz	Free Formation, CCDC FDAT, Shelx, CIF, DBWS, LHPM, ICSD, ORTEP, Rietan, GSAS, Fullprof
GUI WinORTEP	http://www.chem.gla.ac.uk/~loUIS/ortep3/	Yes	No	No	Shelx, CIF, GX, SPF/Platon, ORTEP, CSD/CCDC FDAT, CSSR XR, Crystals, GSAS, Sybol MOL/MOL2, MDL MOL, XYZ file, Brookhaven PDB, Rietica-LHPM, Fullprof	ORTEP, Shelx, XYZ

WinGX	http://www.chem.gla.ac.uk/~loUIS/wingx/	Yes	Yes	Yes	CSSR, Shelx, Cif, CSD/CCDC FDAT, GX	Shelx, CIF, GX, SPF/Platon, CACAO
Powder Cell	http://www.bam.de/a_v/v_1/powder/e_cell.html	Yes	Yes (Int Tab)	No	Powder Cell (CEL), Shelx, ICSD	Powder Cell, BGMN, Shelx, Opal
ICSD Web	http://barns.ill.fr/dif/icsd/	Yes	No	No	Inorganic Database	ICSD, Shelx, Fullprof, GSAS Macro, Powder Cell, Lazy Pulvarix

Structure transformation and inter-conversion using the shareware Cryscon program

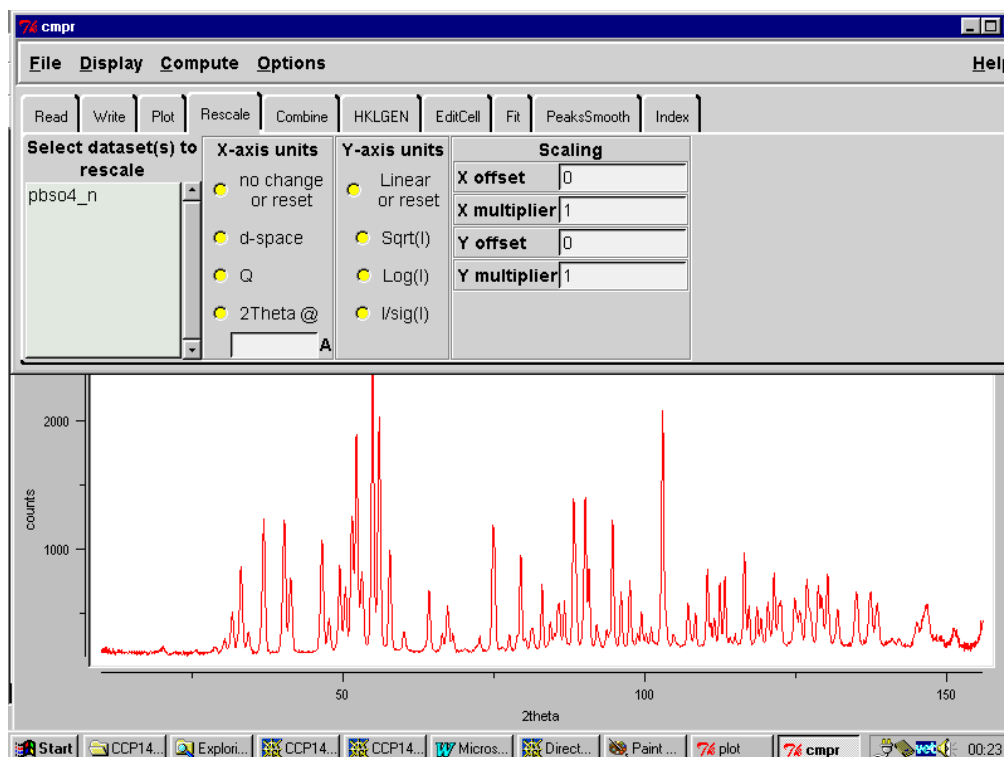


Powder Diffraction Pattern Viewing and Processing

Program	Address	Windows	UNIX	Data Formats
CMPR	http://www.ncnr.nist.gov/programs/crystallography/software/cmpr/cmpr.html	Yes	Yes	BT-1, DBWS, Fullprof, LHPM, GSAS, BNL X7A, XY Data
Powder Cell	http://www.bam.de/a_v/v_1/powder/e_cell.html	Yes	No	Siemens RAW, Philips UDF, XY, CPI, Riet7, APX63
Powder v 2.00	http://www.ccp14.ac.uk/ccp/web-mirrors/ndragoe/appchem/labs/kitazawa/dragoe/html/software.html	Yes	No	DBWS, GSAS CW, GSAS CW ESD, LHPM, Philips RD/SD binary, Philips UDF, MXP18 Binary, RIET7, Scintag, Siemens ASCII, Sietronics CPI, WPPF/Profit, Y free ascii, XY free ascii, free ascii. Line; X, XY, XYZ
Powder X	http://www.ccp14.ac.uk/ccp/web-mirrors/powderx/Powder/	Yes	No	Mac Science ASCII, BD90 (Raw), X-Y, Rigaku (DAT), Sietronics (CPI), TsingHua Rigaku (USR) Siemens/Bruker ASCII (UXD), Siemens/Bruker Binary (RAW), Philips ASCII (UDF), Philips Binary (RD) Mac Science Binary, RIET7 (DAT), ORTEC Maestro (CHN)
Winfit	http://www.geol.uni-erlangen.de/html/software/winsoft.html	Yes	No	Geol. Dept. Erlangen (DFA), Siemens/Bruker Diffrac V 2.1 (1 range) (RAW), (TRU), ASCII X-Y, ICDD Format (PD3), ZDS (ZDS), Software of F Nieto (CRI), Philips (UDF), Philips Binary (RD), STOE (RAW), JADE (MDI), MacDiff of Rainer Petschick (DIF), Converted RAW File (Bish, Eberl,..) (ASC), XDA Rietveld (XDA)

Winplotr	http://www-llb.cea.fr/winplotr/winplotr.htm ftp://bali.saclay.cea.fr/pub/divers/winplotr/	Yes	No	Fullprof, GSAS, XRFIT, HRMPD G42, 6TI, G41/G61, RX, DMC/PSI, XY
XFIT/Koalariet	http://www.ccp14.ac.uk/tutorial/xfit-95/xfit.htm	Yes	No	Riet7 DAT, CPI, XDD, XDA, XY, SCN

The CMPR Powder Diffraction Toolkit for UNIX and Windows

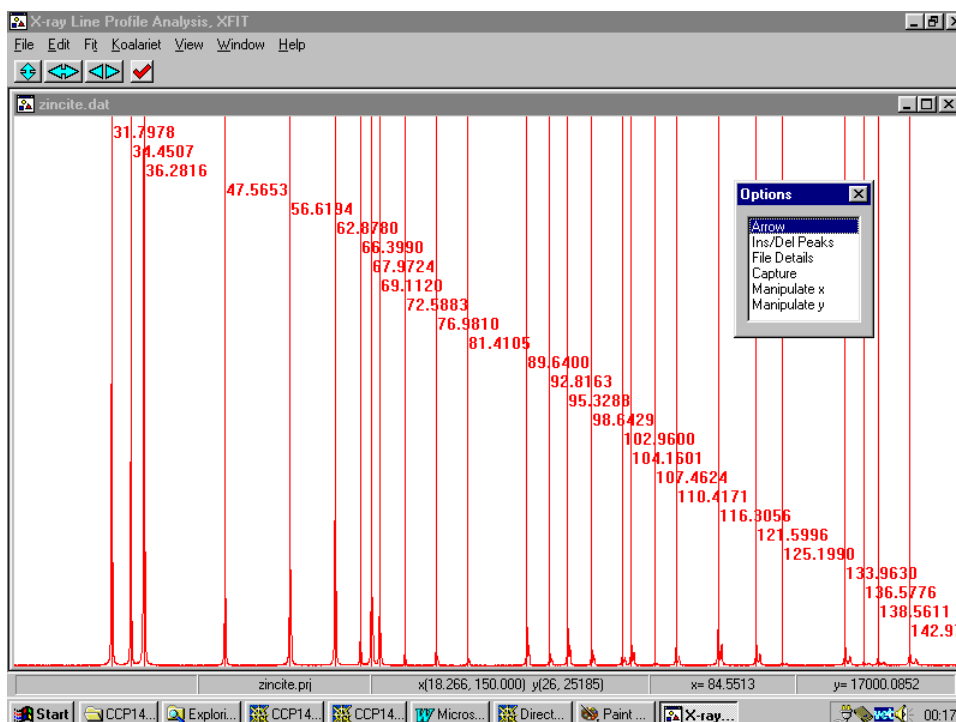
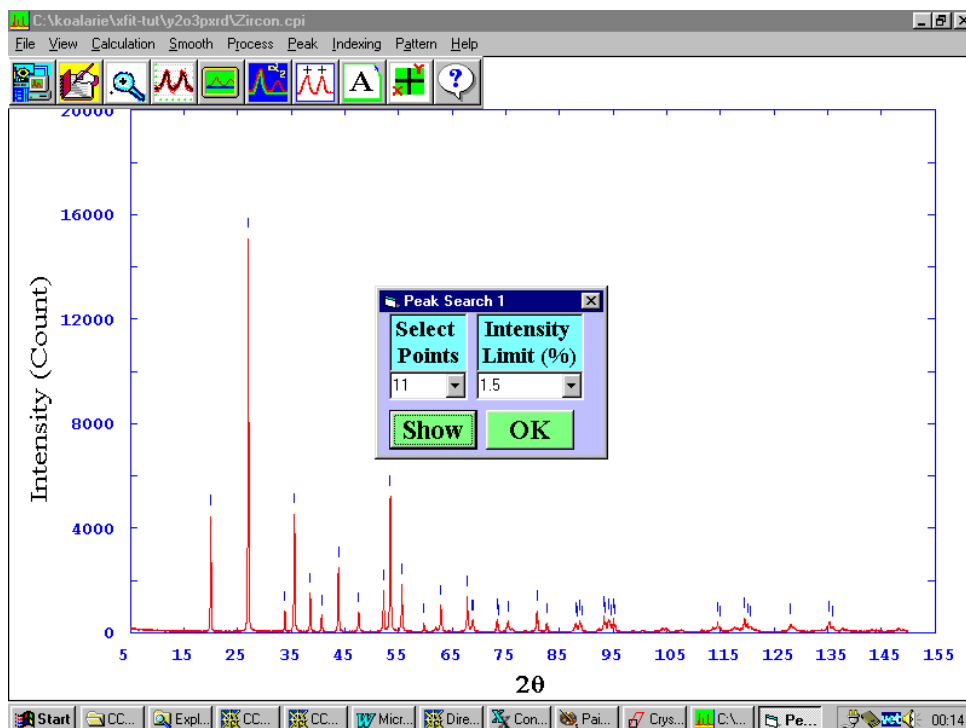


Peak Finding and Peak Profiling (As of August 2000)

Program	Address	Peak Finding	Peak Profiling	Data Formats
CMPR	http://www.ncnr.nist.gov/programs/crystallography/software/cmpr/cmpr.html	No	Yes	BT-1, DBWS, Fullprof, LHPM, GSAS, BNL X7A, XY Data
GSAS RAWPLOT	ftp://ftp.lanl.gov/public/gsas/	No	Yes	GSAS Format
Powder v 2.00	http://www.ccp14.ac.uk/ccp/web-mirrors/n_dragoe/appchem/labs/kitazawa/dragoe/html/software.html	Yes	No	DBWS, GSAS CW, GSAS CW ESD, LHPM, Philips RD/SD binary, Philips UDF, MXP18 Binary, RIET7, Scintag,

				Siemens ASCII, Sietronics CPI, WPPF/Profit, Y free ascii, XY free ascii, free ascii. Line; X, XY, XYZ
Powder X	http://www.ccp14.ac.uk/ccp/web-mirrors/powderx/Powder/	Yes	No	Mac Science ASCII, BD90 (Raw), X-Y, Rigaku (DAT), Sietronics (CPI), TsingHua Rigaku (USR) Siemens/Bruker ASCII (UXD), Siemens/Bruker Binary (RAW), Philips ASCII (UDF), Philips Binary (RD) Mac Science Binary, RIET7 (DAT), ORTEC Maestro (CHN)
Winfit	http://www.geol.uni-erlangen.de/html/software/winsoft.html	No	Yes	Geol. Dept. Erlangen (DFA), Siemens/Bruker Diffrac V 2.1 (1 range) (RAW), (TRU), ASCII X-Y, ICDD Format (PD3), ZDS (ZDS), Software of F Nieto (CRI), Philips (UDF), Philips Binary (RD), STOE (RAW), JADE (MDI), MacDiff of Rainer Petschick (DIF), Converted RAW File (Bish, Eberl,..) (ASC), XDA Rietveld (XDA)
Winplotr	http://www-llb.cea.fr/winplotr/winplotr.htm ftp://bali.saclay.cea.fr/pub/divers/winplotr/	Yes	Yes	Fullprof, GSAS, XRFIT, HRMPD G42, 6TI, G41/G61, RX, DMC/PSI, XY
XFIT/Koalariet	http://www.ccp14.ac.uk/tutorial/xfit-95/xfit.htm	No	Yes	Riet7 DAT, CPI, XDD, XDA, XY, SCN

a) Peak finding in PowderX for Windows and b) peak profiling in XFIT for Windows



Powder Indexing

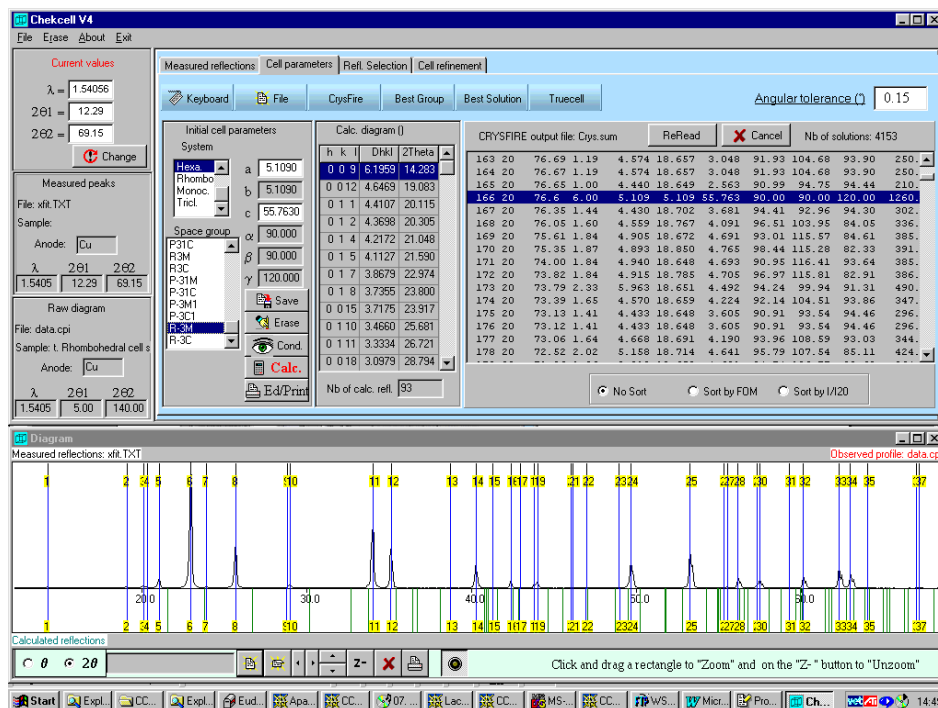
Programs such as Ito, Dicvol and Treor have multiple downloads via the Internet. Arnel Le Bail's site is given as the cited download site.

Indexing Programs	Download	Within Suites
Dicvol	http://sdpd.univ-lemans.fr/ftp/dicvol91.zip	Crysfire, Powder v 2.0, Winplotr
Eflech/Index	http://www.bgm.de/related.html	
Fjzn	http://www.ccp14.ac.uk/tutorial/crys/	Crysfire
Ito	http://sdpd.univ-lemans.fr/ftp/ito13.zip	Crysfire, Powder v 2.0
Kohl	http://www.ccp14.ac.uk/tutorial/crys/	Crysfire
Lzon	http://www.ccp14.ac.uk/tutorial/crys/	Crysfire
Losh	http://www.ccp14.ac.uk/tutorial/crys/	Crysfire
Supercel	http://www-llb.cea.fr/winplotr/winplotr.htm ftp://bali.saclay.cea.fr/pub/divers/winplotr/	Winplotr
Taup/Powder	ftp://hprib.lps.u-psud.fr/pub/powder/ http://www.ccp14.ac.uk/ccp/ccp14/ftp-mirror/taupin-indexing/pub/powder/	Crysfire
Treor	http://sdpd.univ-lemans.fr/ftp/treor90.zip	Crysfire, Powder v 2.0, Powder X, Winplotr
Xray Scan	http://phyhp.phy.ncku.edu.tw/~hjs/hjseng.html http://www.ccp14.ac.uk/ccp/web-mirrors/xrays-can-indexing/~hjs/hjseng.html	

Suites that link into Indexing Programs

Program	Address	Indexing Programs Linked (August 2000)
CMPR	http://www.ncnr.nist.gov/programs/crystallography/software/cmpr/cmpr.html	Manual Indexing via interactive slide bars
Crysfire	http://www.ccp14.ac.uk/tutorial/crys/	Ito, Dicvol, Treor, Losh, Lzon, Kohl, Taup/Powder
Powder X	http://www.ccp14.ac.uk/ccp/web-mirrors/powderx/Powder/	Treor, Manual Indexing
Powder V 2.00	http://www.ccp14.ac.uk/ccp/web-mirrors/ndragoe/appchem/labs/kitazawa/dragoe/html/software.html	Ito, Dicvol, Treor
Winplotr	http://www-llb.cea.fr/winplotr/winplotr.htm ftp://bali.saclay.cea.fr/pub/divers/winplotr/	Dicvol, Treor, Supercell
Chekcell	http://www.ccp14.ac.uk/tutorial/lmgp/	Crysfire suite, Manual Indexing

Example of Chekcell for Graphically Finding the Best Cell/Spacegroup Combination.



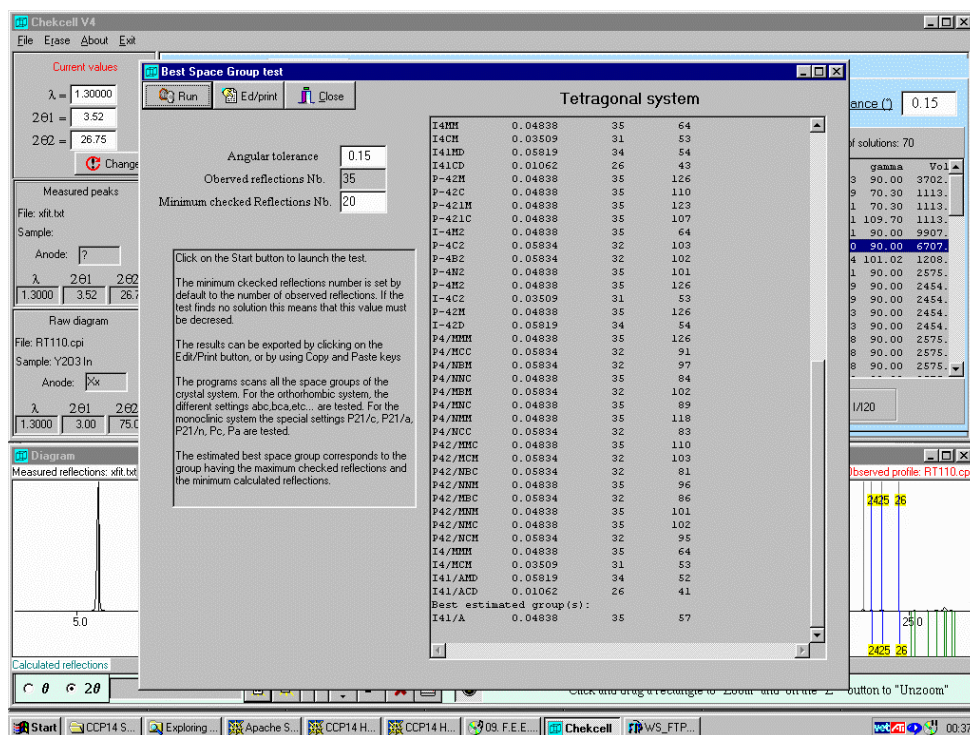
Spacegroup Assignment

Non Whole Profile Fitting (based on HKLs)

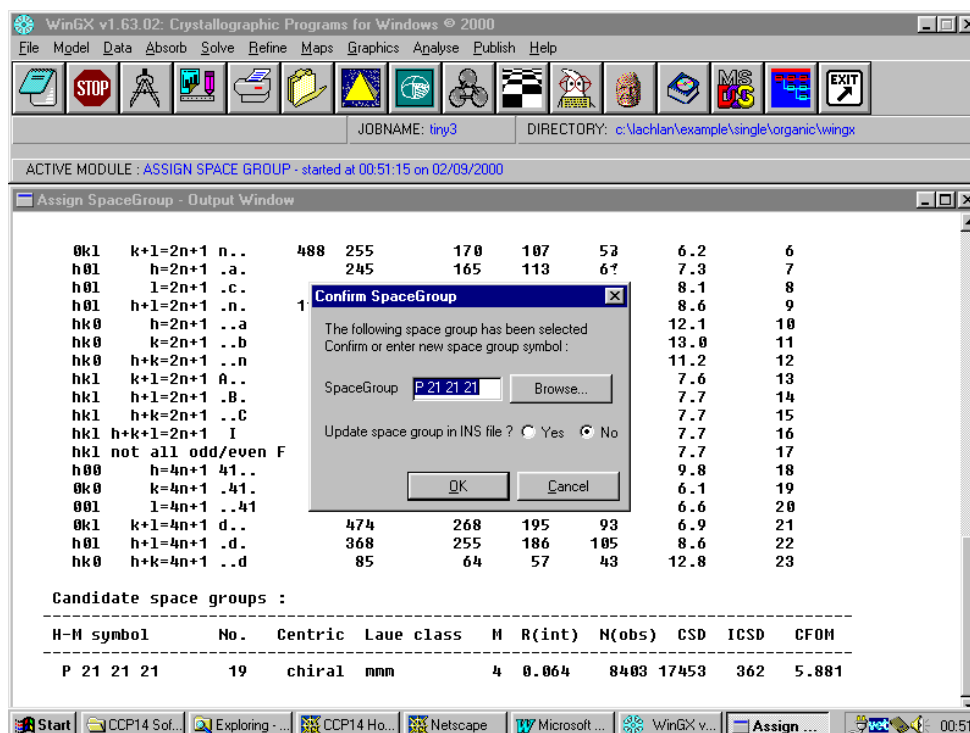
Program	Address	Method
Chekcell	http://www.ccp14.ac.uk/tutorial/lm/gp/	Graphically and automatic determination using peak list file and powder data
Platon	http://www.cryst.chem.uu.nl/platon/	HKL File
Absen (within ORTEX and WinGX)	http://www.nuigalway.ie/cryst/software.htm http://www.chem.gla.ac.uk/~louis/wingx/	HKL File
International Tables	http://www.iucr.org/iucr-top/it/	Manually from HKLs

Whole Profile Le Bail/Pawley methods (being guided by R factors).
Refer to Le Bail and Pawley fitting programs.

Example of running Chekcell for automatic determining optimum cell and spacegroup combinations.



Example of running Absen spacegroup assignment software within WinGX

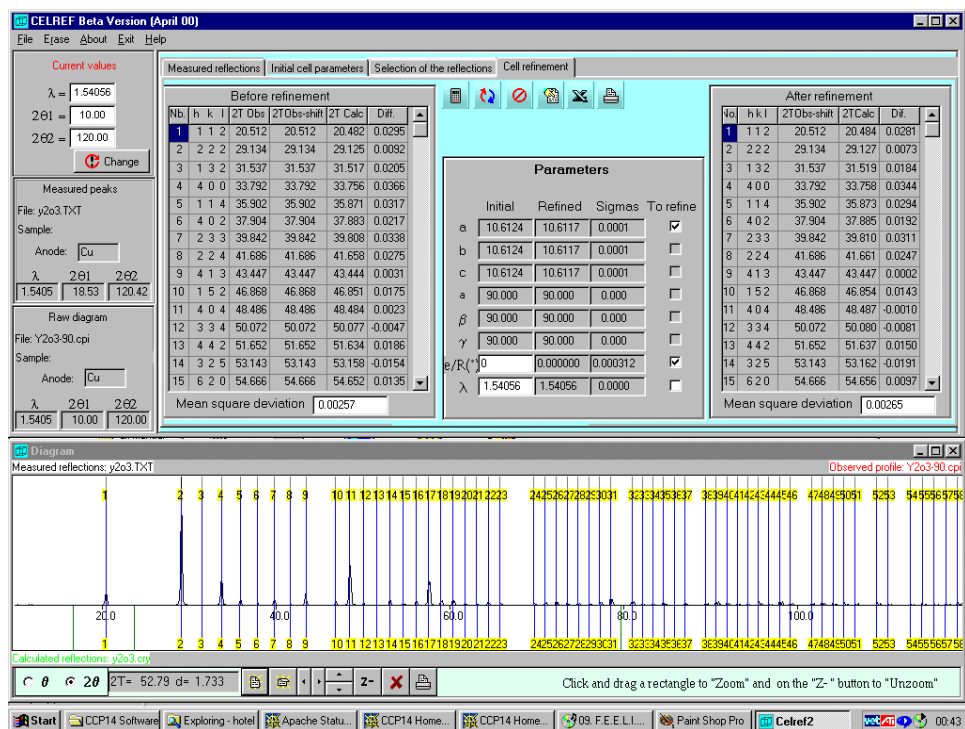


Unit Cell Refinement

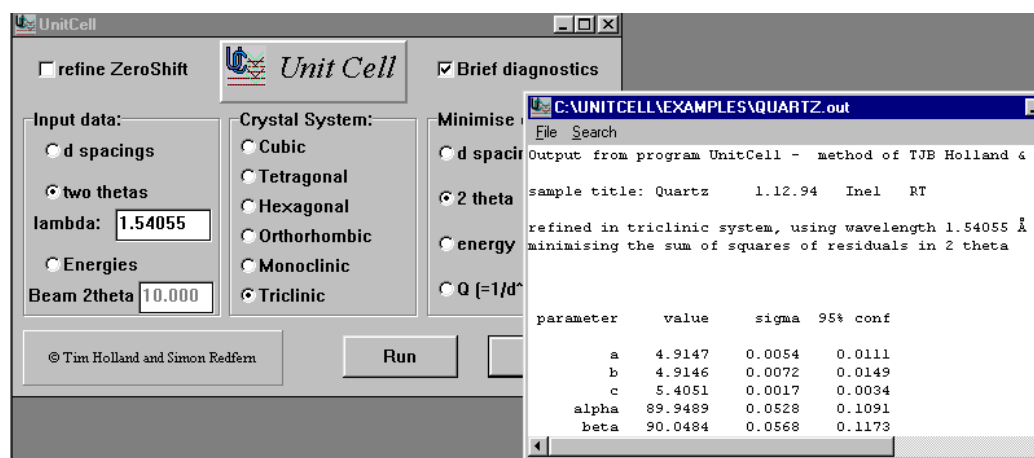
Summary list at: <http://www.ccp14.ac.uk/solution/unitcellrefine/>

Program	Address	Can over-lay Raw diffraction data	Can use Internal Standard
Celref	http://www.ccp14.ac.uk/tutorial/lmgp/	Yes	No
Eracel	http://sdpd.univ-lemans.fr/ftp/eracel.zip	No	No
LAPOD	http://www.ccp14.ac.uk/ccp/web-mirrors/lapod-langford/	No	No
LAPODS	http://www.ccp14.ac.uk/solution/ccp/web-mirrors/powderx/lapod/	No	No
Powder v2.00	http://www.ccp14.ac.uk/ccp/web-mirrors/n_dragoe/appchem/labs/kitazawa/dragoe/html/software.html	No	Yes
Refcel	ftp://img.cryst.bbk.ac.uk/pdpl/ http://www.ccp14.ac.uk/ccp/ccp14/ftp-mirror/profil/PDPL/	No	No
UNITCELL	http://www.esc.cam.ac.uk/astaff/holland/	No	No
Wincell	http://fazel.rajabalee.free.fr/delphi_an.htm	No	No
XLAT	ftp://jgiqc.llnl.gov/	No	Yes

Example of running Celref for Windows



Example of running UNITCELL for Windows (Mac and UNIX versions available)



Full Profile Fitting (Pawley, LeBail)

Summary lists at:

Le Bail <http://www.ccp14.ac.uk/solution/lebaill/>

Pawley <http://www.ccp14.ac.uk/solution/pawley/>

Overlap by Armel Le bail can take Le Bail extracted intensity files in Shelx, Sirpow and EXPO format converting them to F, Fsq and Intensity files in Shelx format <http://sdpd.univ-lemans.fr/ftp/overlap.zip>

Le Bail Extraction Programs

Program	Address	Known to recycle Intensities
ARIT	http://sdpd.univ-lemans.fr/arit.html	
BGMN	http://www.bgm.de	
EXPO	http://www.ba.cnr.it/IRMEC/SirWare.html	
EXTRACT (part of XRS-82)	http://www.kristall.ethz.ch/LFK/software/xrs/	
Fullprof	ftp://bali.saclay.cea.fr/pub/divers/	Yes
GSAS	ftp://ftp.lanl.gov/public/gsas/	Yes
LHPM-Rietica	ftp://ftp.ansto.gov.au/pub/physics/neutron/rietveld/Rietica_LHPM95/	Yes
Powder Cell	http://www.bam.de/a_v/v_1/powder/e_cell.html	Yes
Mprofil	http://www.dl.ac.uk/CCP/CCP14/ccp/ccp14/ccp14-by-program/mprep5-mprofil5/	
RIETAN 2000	http://www.nirim.go.jp/~izumi/rietan/angle_dispersive/an	

	gle_dispersive.html	
WinMprof	http://pecdc.univ-lemans.fr/WinMProf/WinMProf.htm	Yes
XND	ftp://old-labs.polycnrs-gre.fr/pub/xnd/ ftp://labs.polycnrs-gre.fr/pub/xnd/	

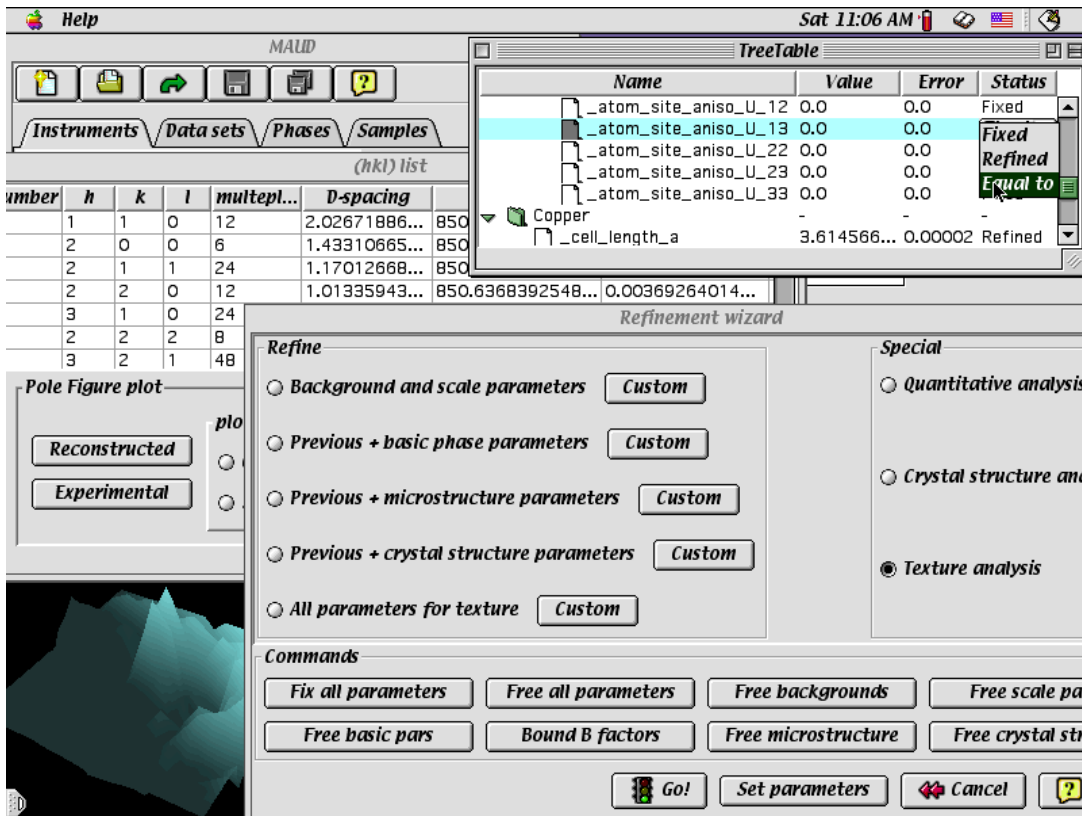
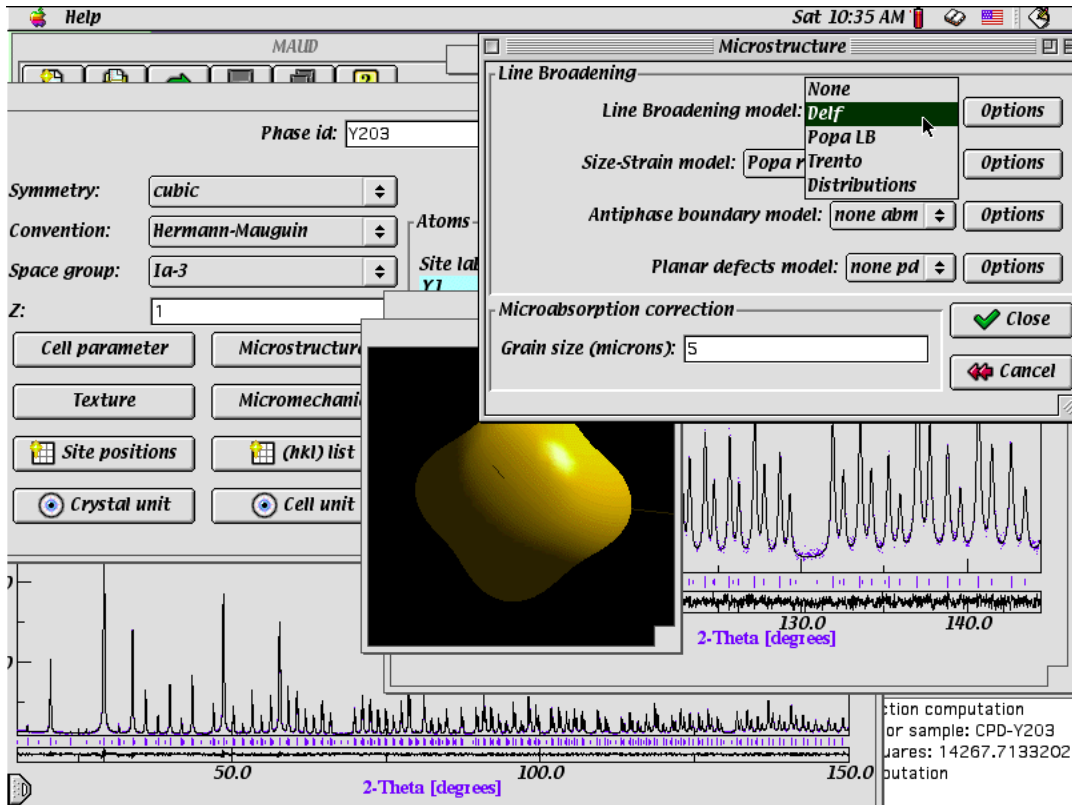
Pawley Extraction Programs

Program	Address	Known to recycle Intensities
Simpro	Http://www.uni-tuebingen.de/uni/pki/simref/simpro.html	
WPPF	Ftp://ftp.ccp14.dl.ac.uk/wdpp/WinNT/Table12b/Wppf/ ftp://ftp.ccp14.dl.ac.uk/wdpp/Win95/Table12b/Wppf/ ftp://ftp.ccp14.dl.ac.uk/wdpp/VaxVMS/Table12b/Wppf/ ftp://ftp.ccp14.dl.ac.uk/wdpp/MS-DOS/Table12b/Wppf/ ftp://ftp.ccp14.dl.ac.uk/wdpp/AlphaVMS/Table12b/Wppf/	

Texture analysis software

Program	Address
BEARTEX	http://www.seismo.berkeley.edu/~wenk/beartex.htm
GSAS	ftp://ftp.lanl.gov/public/gsas/
MAUD for Java (GPL'd)	http://www.ing.unitn.it/~luttero/
POFINT	http://pecdc.univ-lemans.fr/pofint/pofint.htm
popLA	http://www.mst.lanl.gov/cms/poplaapp.html
Symmet	http://www.ccp14.ac.uk/ccp/web-mirrors/chalk_river_pole_figure/

Examples of the MAUD for Java Graphical User Interface



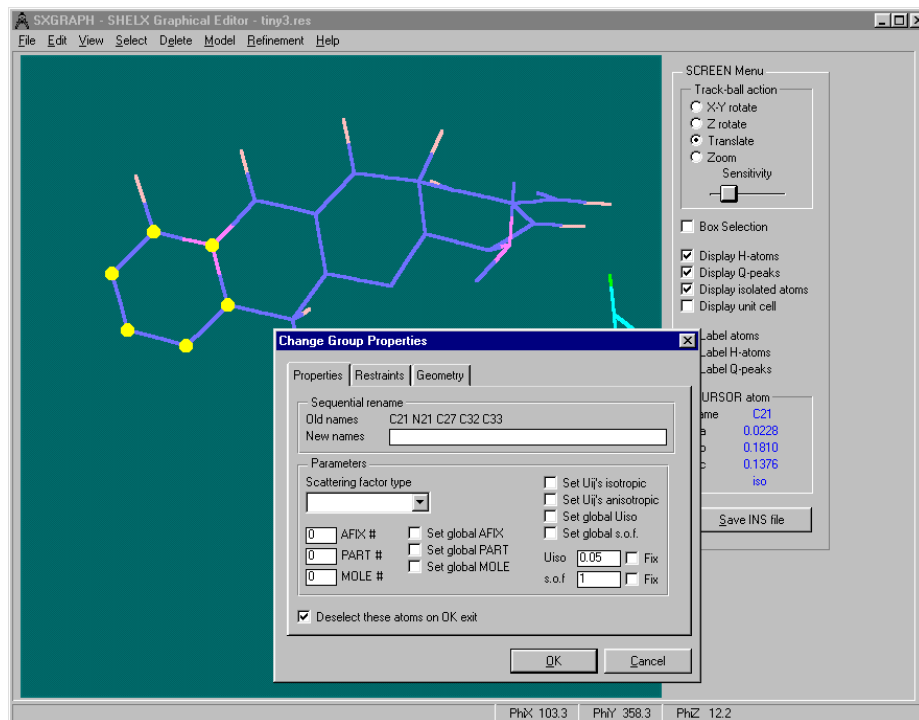
Single Crystal Suites

Program	Address	Linked Solution Software	Refinement
Crystals	http://www.xtl.ox.ac.uk/	Sir, Shelxs	Crystals
NRCVAX	Contact Peter White: pwhite@pyrite.chem.unc.edu	Solver	NRCVAX
ORTEX	http://www.nuigalway.ie/cryst/software.htm	Shelxs	Shelxl 97
Platon / System S	http://www.cryst.chem.uu.nl/platon/	Sir, Shelxs, Crunch, Dirdif	Shelxl 97
WinGX	http://www.chem.gla.ac.uk/~louis/wingx/	Sir, Shelxs, Dirdif, Patsee	Shelxl 97
Xtal (GPL'd)	http://xtal.crystal.uwa.edu.au/	Crisp, Patsee	Xtal (3 different programs)

Examples of the WinGX graphical user interface



The SXGRAPH Graphical User Interface over Shelxl97 provided with WinGX.



Structure Solution Software Specifically for Powder Diffraction

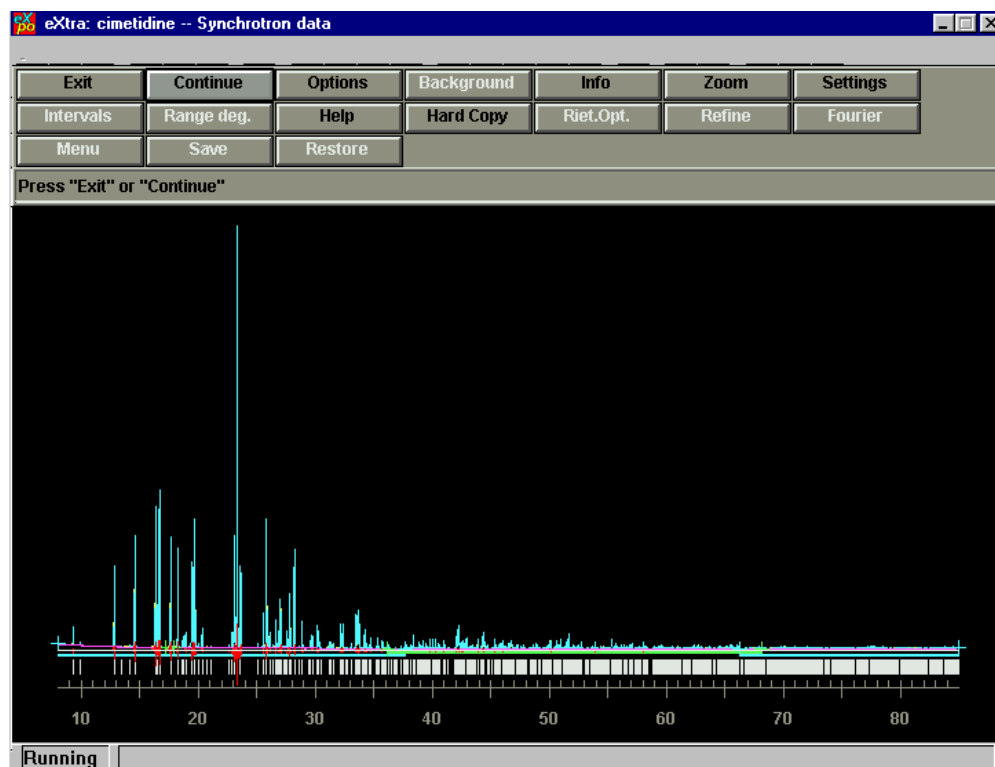
Freely Available Software

Program	Address	Method
EXPO 2000	http://www.ba.cnr.it/IRMEC/SirWare_main.html	Direct Methods and Real Space
ESPOIR (GPL'd)	http://sdpd.univ-lemans.fr/sdpd/espoir/	Real Space
Focus	http://www.kristall.ethz.ch/LFK/software/	Chemical Information, Zeolites
ZEFSA II (GPL'd)	http://www.mwdeem.chemeng.ucla.edu/zefsaII/	Real Space, Zeolites

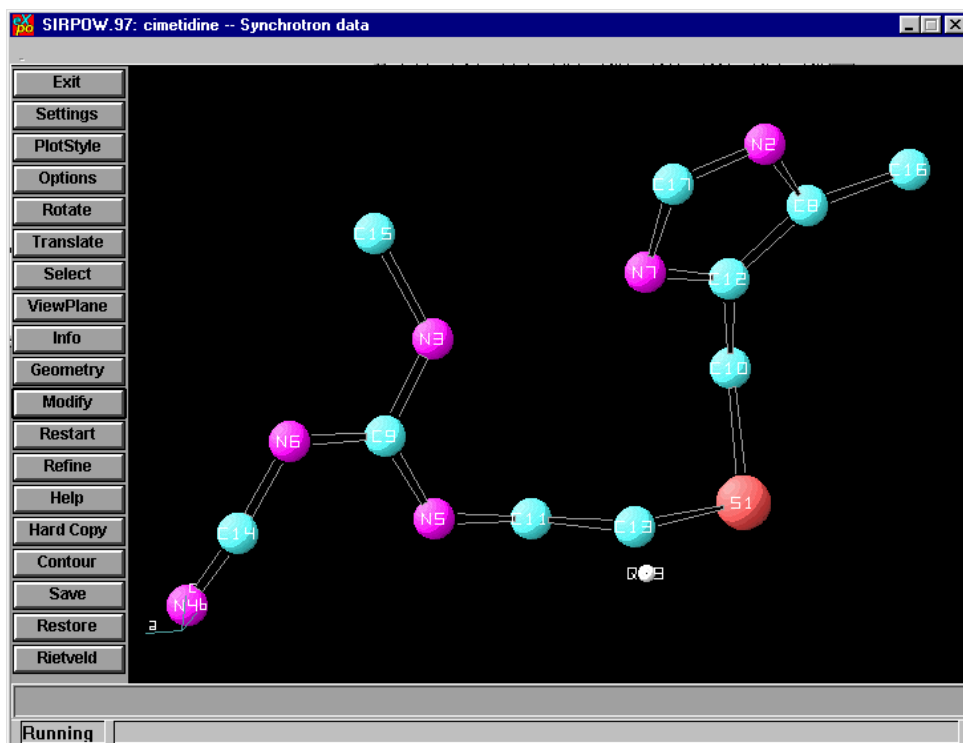
Commercially Available Software

Program	Address	Method
Cerius 2	http://www.msi.com/life/products/cerius2/	Real Space
WinCSD/CSD	http://imr.chem.binghamton.edu/zavalij/CSD.html	
DIFFRACplus TOPAS	http://www.bruker-axs.com/production/products/xrd/software/topas/	Real Space
Endeavour	http://www.crystalimpact.com/endeavour/	Real Space
Dash mentioned here(?)	http://www.esrf.fr/info/science/highlights/1999/chemistry/powder.html	Real Space

Example of automatic Le Bail fitting within EXPO



Example of EXPO completing the Fourier cycling to give a solve structure



Structure Solution Using Single Crystal Software

Program	Address	Method
CAOS	http://www.mlib.cnr.it/isc/caos/	Patterson
Crisp	http://xtal.crystal.uwa.edu.au/	Direct Methods
Crunch	http://chemb0b.leidenuniv.nl:80/~rag//	Direct Methods
Dirdif	http://www-xtal.sci.kun.nl/xtal/documents/software/dirdif.html	Patterson, Fragment
Patsee	http://www.org.chemie.uni-frankfurt.de/egert/html/patsee.html	Fragment
Shake'n'Bake (SnB)	http://www.hwi.buffalo.edu/SnB/	SnB
Shelxs 86/97/d	http://shelx.uni-ac.gwdg.de/SHELX/index.html	Direct Methods, Patterson, SnB style
Sir92/97/2000	http://www.ba.cnr.it/IRMEC/SirWare_main.html	Direct Methods, SnB style
Solver	Contact Peter White (pwhite@pyrite.chem.unc.edu)	Direct Methods
XFPA	Contact Frantisek Pavelcik (pavelcik@fns.uniba.sk)	Patterson

Screen images of a) Dirdif for Windows and b) Sir97 solving structures

The image shows two software windows. The top window is titled "DIRDIF-99.2 for Windows © 2000" and displays a terminal-style output for a crystallographic job. The output includes the user name "Lachlan M D Cranswick", job number "RUN 18", and details of the DDMAIN, FOUR, and PATTY programs. The bottom window is titled "SIR97: XYZN file = SHELXL INS file, from DIRDIF output for EG1SCR" and shows a 3D molecular model of a protein-ligand complex. A menu is overlaid on the model, showing options like "data", "normal", "invariants", "phase", "fourier", "menu", and "end". The "phase" option is highlighted with a blue arrow. The status bar at the bottom of the SIR97 window indicates "Running".

```
DIRDIF-99.2 for Windows © 2000
Files Auto NoRecycle Interactive Utilities Help

Lachlan M D Cranswick [ personal copy for CCP14 demo' only ]

Automatic run
Run number assigned for this job: RUN 18
Input from control data file CRYSIN for compound EG1SCR
Input from control data file CRYSIN for compound EG1SCR

===== Execute program DDMAIN ===== for compound: EG1SCR
Number of reflections (merged) output: 1874

===== Execute program FOUR ===== for compound: EG1SCR
Approximate nr of intermediate Fourier-transform passes needed: 2
Intermediate transform, pass 2
Final transform, pass 2
Final transform, pass 4

===== Execute program PATTY ===== for compound: EG1SCR
Calculate symmetry map
----- and interpret its peaks
Search for 2 atoms ready
Search for 3 atoms ready
Number of ATOM sets output: 7

===== Execute program DDMAIN ===== for compound: EG1SCR

CCODE: eg1scrat WORK FOLDER : c:\lachlan\example\single\inorg\wingx\ PROCESS: system
STATUS: Program DDMAIN now running .....
```

SIR97: XYZN file = SHELXL INS file, from DIRDIF output for EG1SCR

At 23:28: 8 SIR97 runs on: eg1scrat

- data
- normal B = 1.549 There are pseudo effects (82%)
- invariants
- phase Trial 3 / 128 CFom = 0.273 (max = 0.407)
- fourier
- menu
- end

Assign phases to reflexions

Running

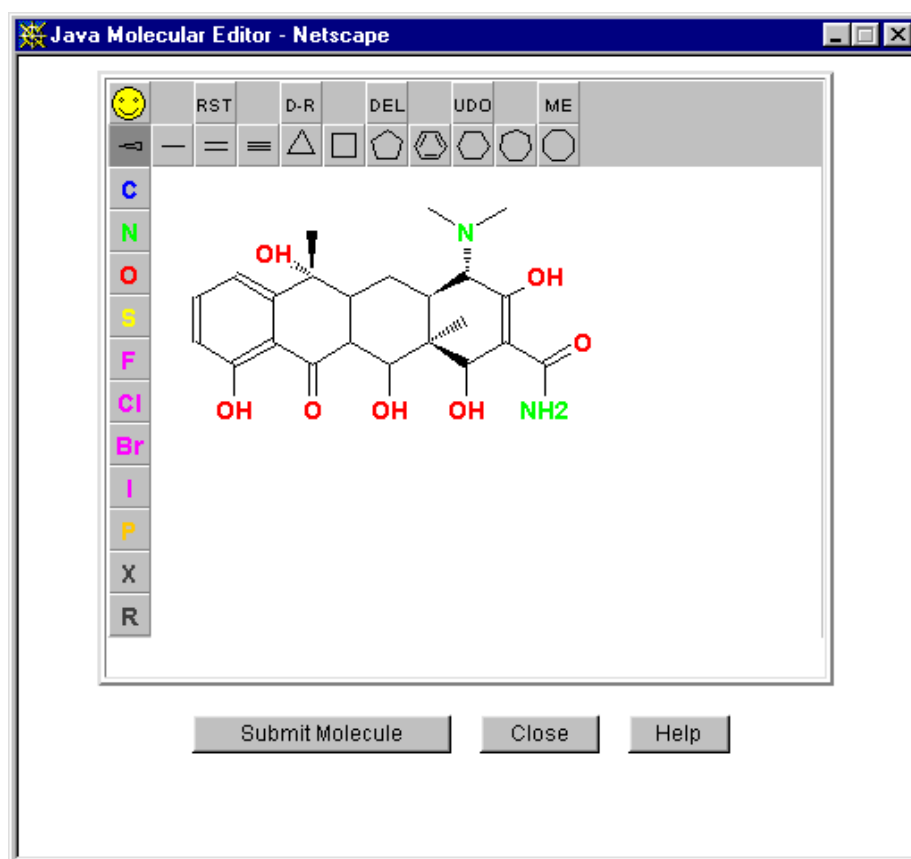
2D to 3D molecular Model Generation

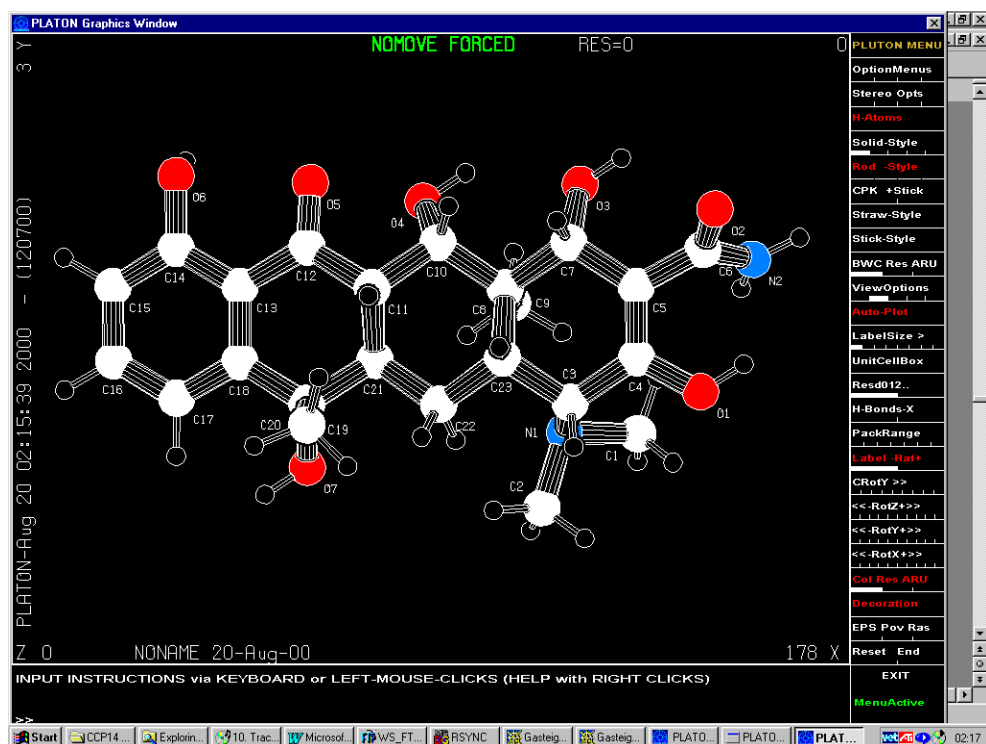
These can be helpful for fragment based search methods for real space and reciprocal space methods. Some 2D to 3D model builders are better than others. Three freely available ones are listed. Corina via a web interface is the easiest to start with.

Summary List at: http://www.ccp14.ac.uk/solution/2d_3d_model_builders/

Program	Address	Caveates
CORINA	http://www2.ccc.uni-erlangen.de/software/corina/free_struct.html	Web Based
Momo	http://www.org.chemie.uni-frankfurt.de/egert/html/momo.html	From the makers of Patsee
Dirdif	http://www-xtal.sci.kun.nl/xtal/documents/software/dirdif.html	Can generate 3D models with hydrogens from 2D models

Screen Images showing a) Generating a 2D Fragment in Corina via a Java Applet and b) viewing the resulting 3D optimised fragment in PDB format in Platon

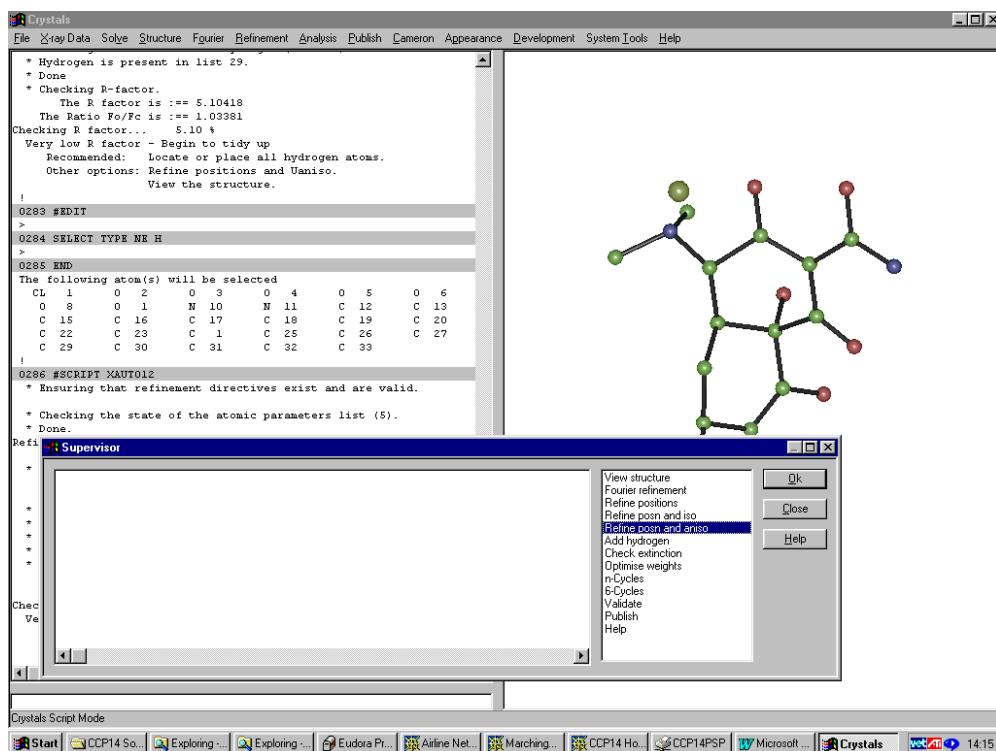




Single Crystal Refinement programs for assisting in building up the structure

Program	Address
CAOS	http://www.isc.mlib.cnr.it/caos/
Crystals	http://www.xtl.ox.ac.uk/
NRCVAX	Contact Peter White (pwhite@pyrite.chem.unc.edu)
Shelxl	http://shelx.uni-ac.gwdg.de/SHELX/index.html
Xtal (GPL'd)	http://xtal.crystal.uwa.edu.au/
WinGX, GUI CalcOH, GUI Xhydex	http://www.chem.gla.ac.uk/~louis/wingx/

Screen Image of Crystals for Windows

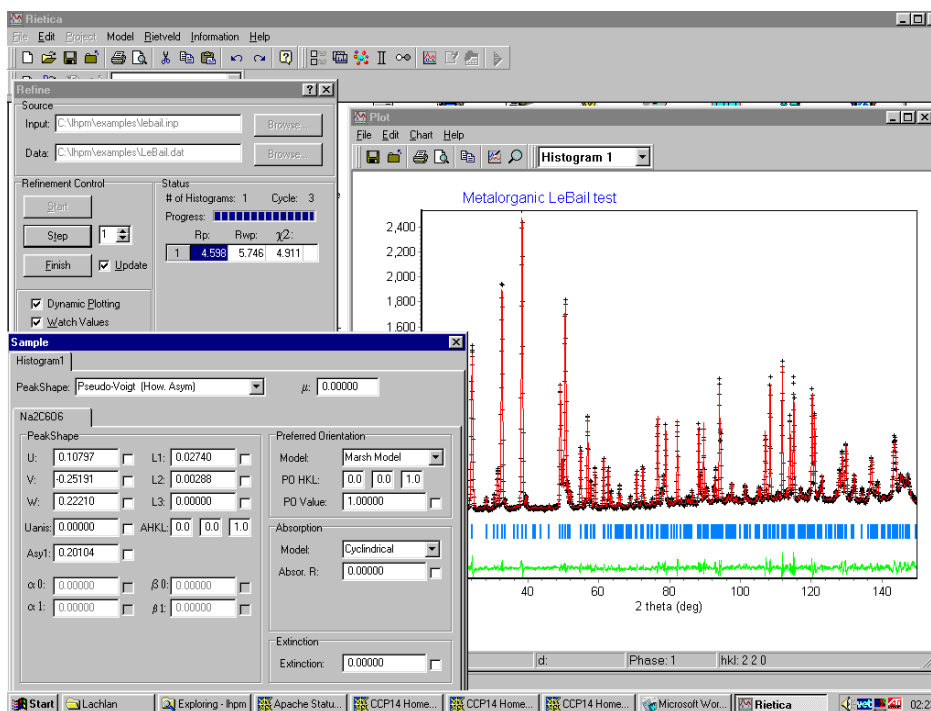
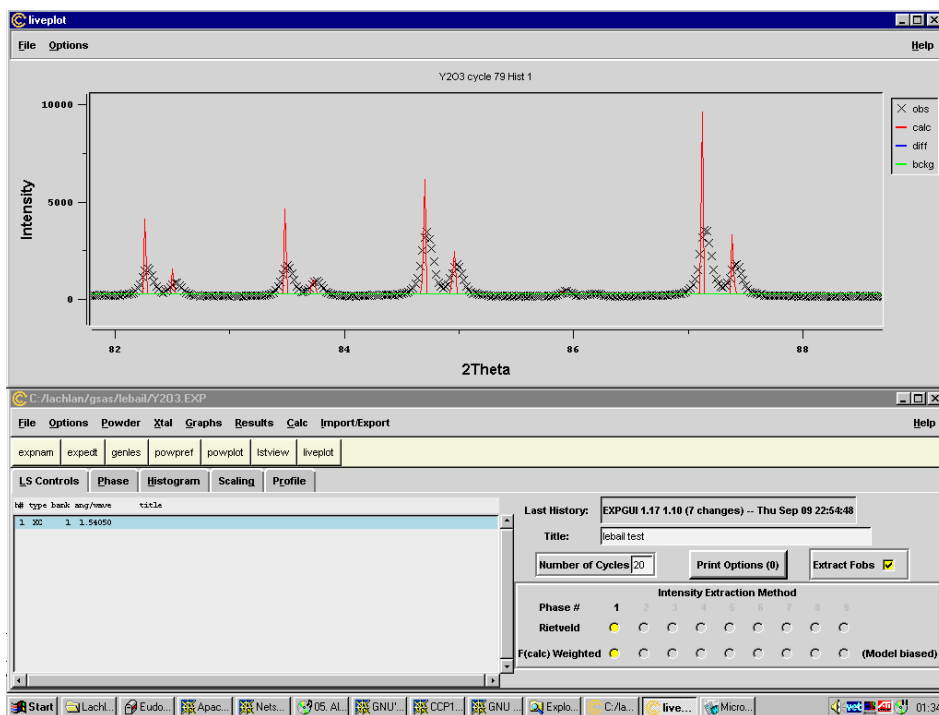


Rietveld structure refinement

Program	Address	Advertised Specialisations
ARITVE	http://sdpd.univ-lemans.fr/aritve.html	Glasses
BGMN	http://www.bgmn.de	Fundamental Parameters
DBWS	Contact Ray Young (r.young@physics.gatech.edu)	
DEBVIN	ftp://ftp.cc.uniud.it/DEBVIN/	Polymers
EXPO	http://www.ba.cnr.it/IRMEC/SirWare_main.html	
Fullprof	ftp://charybde.saclay.cea.fr/pub/divers/	General, Incommensurate, Magnetic, Combined Neutron/X-ray/Single Crystal, TOF, Fourier Map Facility via GFOUR
GSAS	ftp://ftp.lanl.gov/public/gsas/ EXP GUI Interface: http://www.ncnr.nist.gov/programs/crystallography/software/expgui/expgui_intro .	General, Magnetic, Combined Neutron/X-ray/Single Crystal, Texture, TOF, Fourier

	html	Map facility
Koalariet	http://www.ccp14.ac.uk/tutorial/xfit-95/xfit.htm	Fundamental Parameters
LHPM/Rietica	ftp://ftp.ansto.gov.au/pub/physics/neutron/rietveld/Rietica_LHPM95/	Full GUI Interface. Has in built Bond Valence software.
MAUD for Java (GPL'd)	http://www.ing.unitn.it/~luttero/	Materials Analysis, Texture
PREMOS/REMOS	http://www.nirim.go.jp/~yamamoto/	Modulated Structures
ProDD	http://www.cus.cam.ac.uk/~jpw22/	TOF
Profil	ftp://img.cryst.bbk.ac.uk/pdpl/	
Riet7/SR5	ftp://ftp.minerals.csiro.au/pub/xtallography/ http://www.ccp14.ac.uk/ccp/ccp14/ftp-mirror/csirominerals-anon-ftp/pub/xtallography/	Variable Count Time
RIETAN-2000 (GPL'd)	http://www.nirim.go.jp/~izumi/rietan/angle_dispersive/angle_dispersive.html	TOF, maximum-entropy method (MEM)
Rietquan	http://www.ing.unitn.it/~luttero/	Quantitative Analysis
Simref	http://www.uni-tuebingen.de/uni/pki/	Incommensurate, High temperature multiple dataset refinement
WinMprof	http://pecdc.univ-lemans.fr/WinMProf/WinMProf.htm	Le Bail fitting, In built Fourier peak finding for building up structures.
XND	ftp://old-labs.polycnrs-gre.fr/pub/xnd/ ftp://labs.polycnrs-gre.fr/pub/xnd/	Incommensurate, High temperature multiple datasets refinement
XRS-82/DLS-76	http://www.kristall.ethz.ch/LFK/software/	Zeolites

Screen Image of a) part of the EXPGUI Interface of GSAS and b) Rietica-LHPM Graphical User Interface

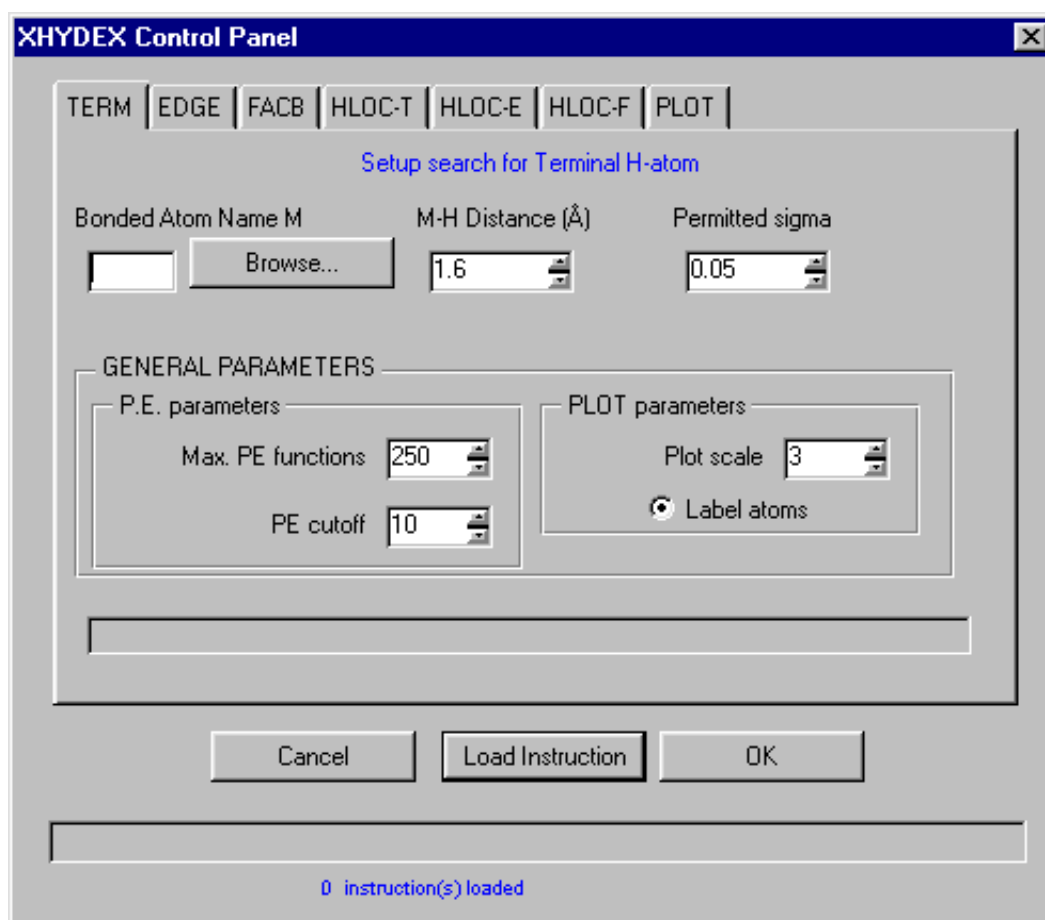


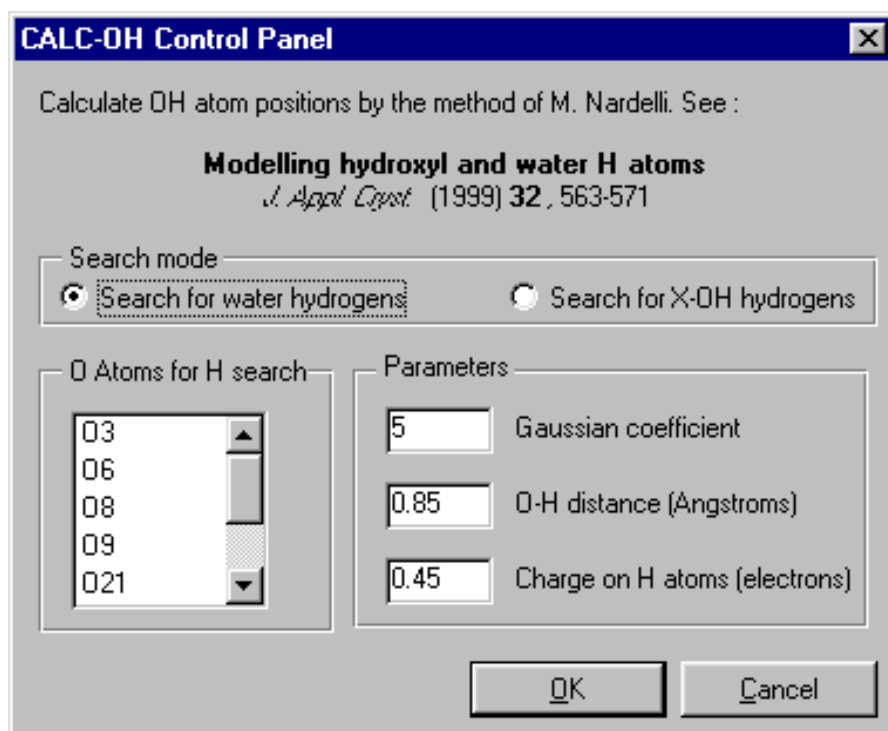
Hydrogen Placement Using Single Crystal Software

Important to try and place hydrogens if they are present in the structure during final stages of refinement.

Program	Address
CAOS	http://www.isc.mlib.cnr.it/caos/
Crystals	http://www.xtl.ox.ac.uk/
Dirdif	http://www-xtal.sci.kun.nl/xtal/documents/software/dirdif.html
Shelxl	http://shelx.uni-ac.gwdg.de/SHELX/index.html
Xtal (GPL'd)	http://xtal.crystal.uwa.edu.au/

Screen Images of a) GUI XHYDEX and b) GUI CALC-OH Provided within the WinGX Single Crystal Suite for finding of Hydrogens



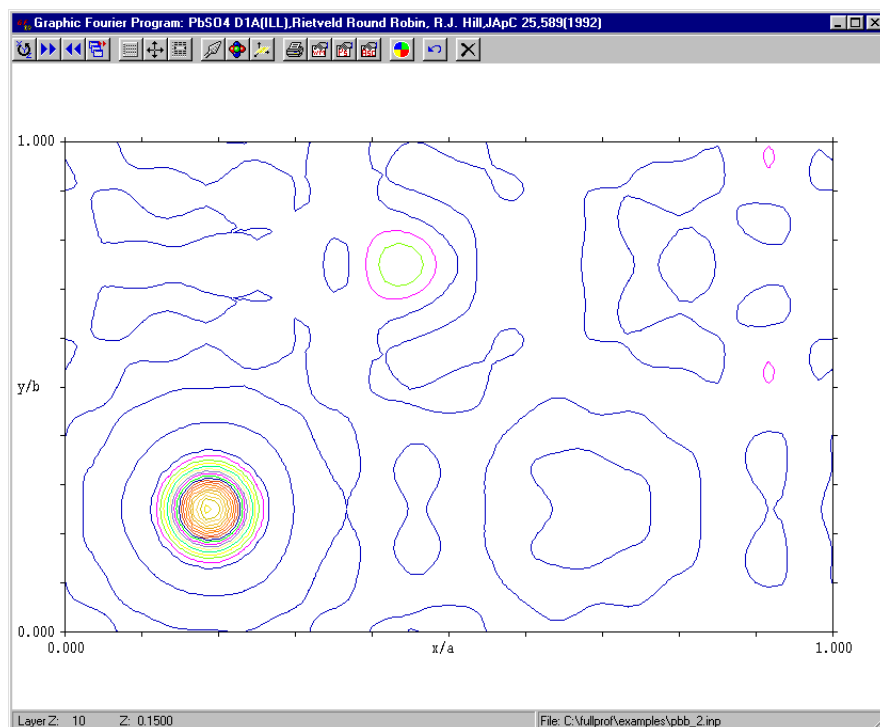


Rietveld structure refinement software that is Fourier Map/Q peak Friendly

Summary list at: http://www.ccp14.ac.uk/solution/rietveld_fourier_maps/

Program	Address	Advertised Specialisat.
BGMN	http://www.bgm.de	Outputs Shelx FCF files that can be read by various single crystal suites
EXPO	http://www.ba.cnr.it/IRMEC/SirWare_main.html	Has in built Fourier Map capability
Fullprof	ftp://charybde.saclay.cea.fr/pub/divers/	Links into the GFOUR program
GSAS	http://shelx.uni-ac.gwdg.de/SHELX/index.html EXP GUI Interface: http://www.ncnr.nist.gov/programs/crystallography/software/expgui/expgui_intro.html	In built Fourier including the ability to output VRML based Fourier Maps.
LHPM/Rietica	ftp://ftp.ansto.gov.au/pub/physics/neutron/rietveld/Rietica_LHPM95/	In built Fourier Map display
RIETAN-2000 (GPL'd)	http://www.nirim.go.jp/~izumi/rietan/angle_dispersive/angle_dispersive.html	Maximum-entropy method (MEM) for generating Fourier Maps
WinMprof	http://pecdc.univ-lemans.fr/WinMProf/WinMProf.htm	In built Fourier peak finding for building up structures.

Screen Image of GFOUR for Windows which interlinks with Fullprof Rietveld

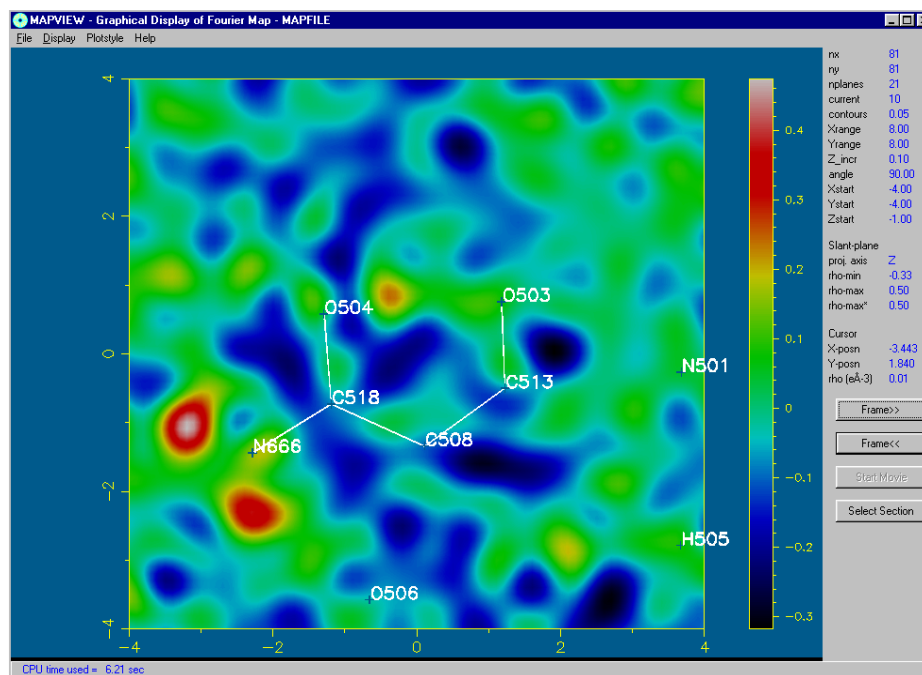


Free Standing Powder and Single Crystal Fourier Map software

Summary list at: http://www.ccp14.ac.uk/solution/rietveld_fourier_maps/

Program	Address	Caveates
GFOUR for Windows	ftp://charybde.saclay.cea.fr/pub/divers/progs_pc/fourier/	Wants Fullprof/GFOUR happy files to generate Maps
Platon for UNIX/Windows	http://www.cryst.chem.uu.nl/platon/	Wants Shelx format FCF format files to generate Maps
WinGX for Windows. Also has Map viewing via Contour and Mapview	http://www.chem.gla.ac.uk/~louis/software/	Wants Shelx format FCF files to generate the Fourier Map; or a WinGX format MAP file
Marching Cubes 3D Fourier Map viewer	http://mysak.umbr.cas.cz/~husakm/Public/MarchingCubeELD/MarchingCubeELD.htm	Views WinGX, Crystals or Project XD map files
FOUE (GPL'd)	http://www.ccp14.ac.uk/ccp/web-mirrors/s cott-belmonte-software/	Converts GSAS Binary Fourier Map files into WinGX, Crystals, Project XD and ASCII formats

Screen Image of the Mapview Fourier Map Viewing Software Provided with the WinGX Single Crystal Suite



Structure Validation (including Platon/Addsym)

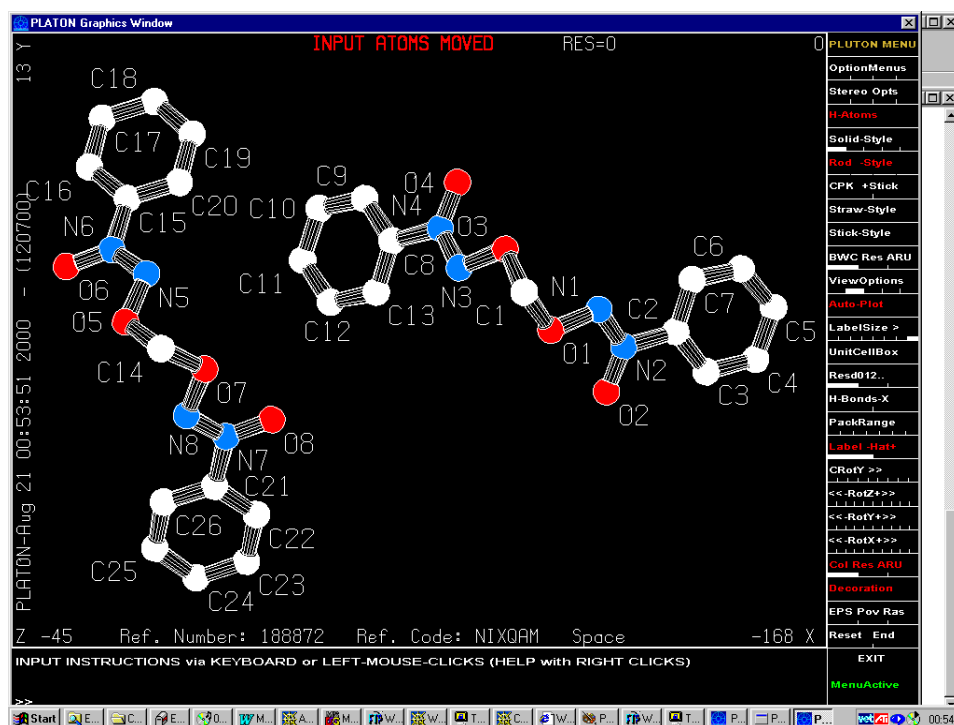
Most of these programs are made for single crystal applications but are very applicable to powder diffraction problems.

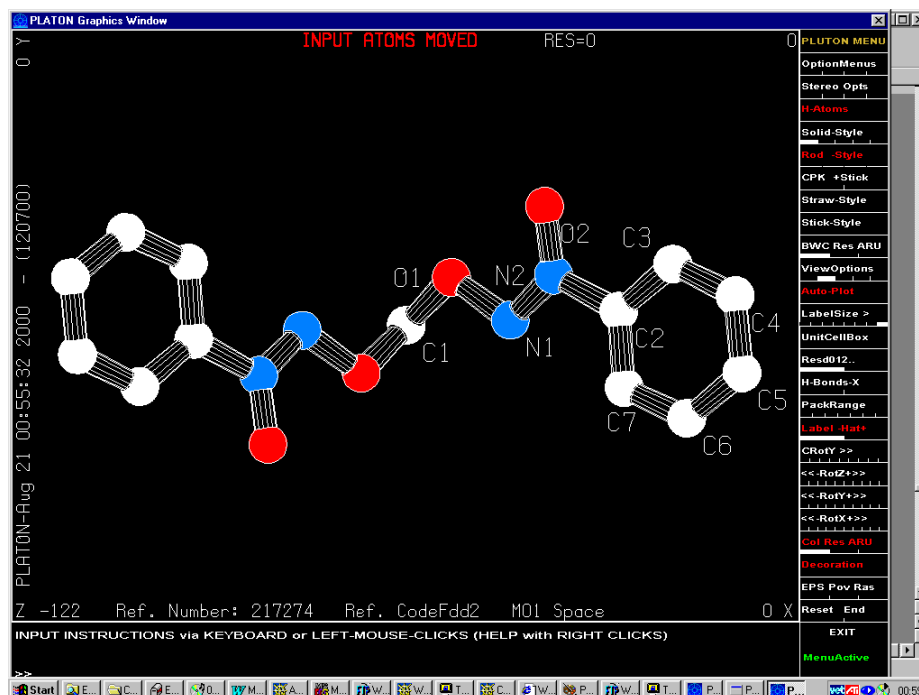
Program	Address	Caveates
Valist for Windows	ftp://ftp.ill.fr/pub/dif/valist/	Bond Valence Checking. Can read common file formats
Valence	http://www.ccp14.ac.uk/ccp/web-mirrors/valence/	Bond Valence Checking
Platon for UNIX/Windows	http://www.cryst.chem.uu.nl/platon/	Includes Addsym. Mandatory software to check for Missing Symmetry as well as other common problems!
WinGX for Windows	http://www.chem.gla.ac.uk/~louis/software/	Links to a variety of validation packages including Platon
ORTEX	http://www.nuigalway.ie/cryst/software.htm	GUI based Void finding
Crystals	http://www.xtl.ox.ac.uk/	Bond length and angle checking with the Cambridge database

Example of Addsym finding better symmetry based on "Short Communication: "P1 or P-1? Corrigendum", Acta Cryst B56 (2000) 744 from Richard E. Marsh"

1. Original Structure was published in P1 Triclinic in 1997
2. Reinterpreted by March as C2 Monoclinic in 1999
3. Found by a click of a button with Addsym to be Fdd2 Orthorhombic

Screen Image of a) original P1 Triclinic structure and b) resulting Fdd2 structure in Platon after analysis with Addsym



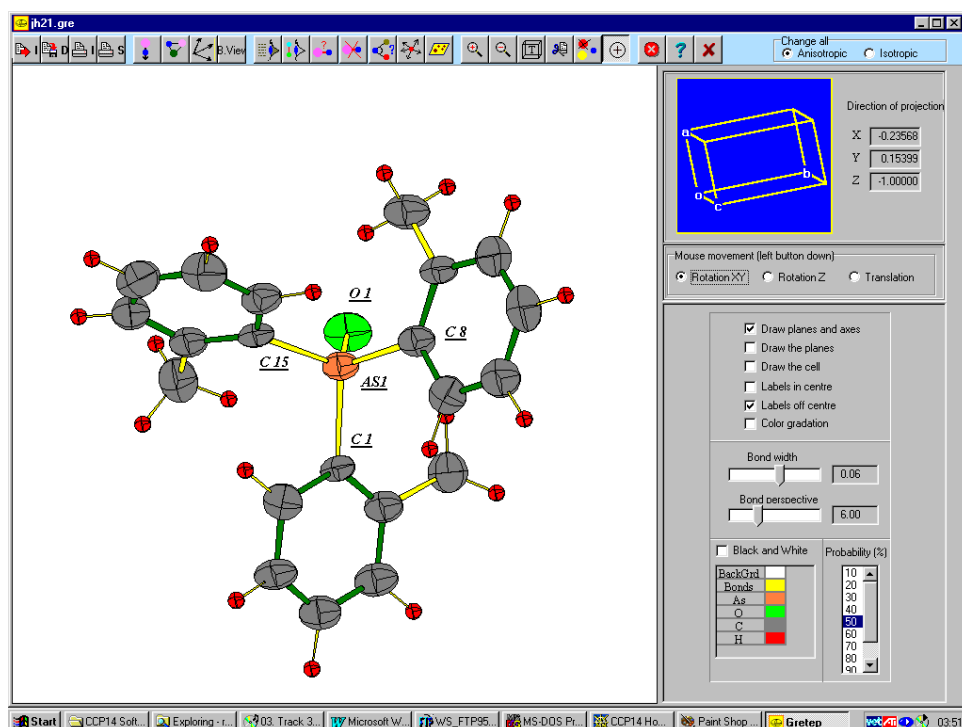


Crystallographic Structure Visualisation: during structure solution and refinement

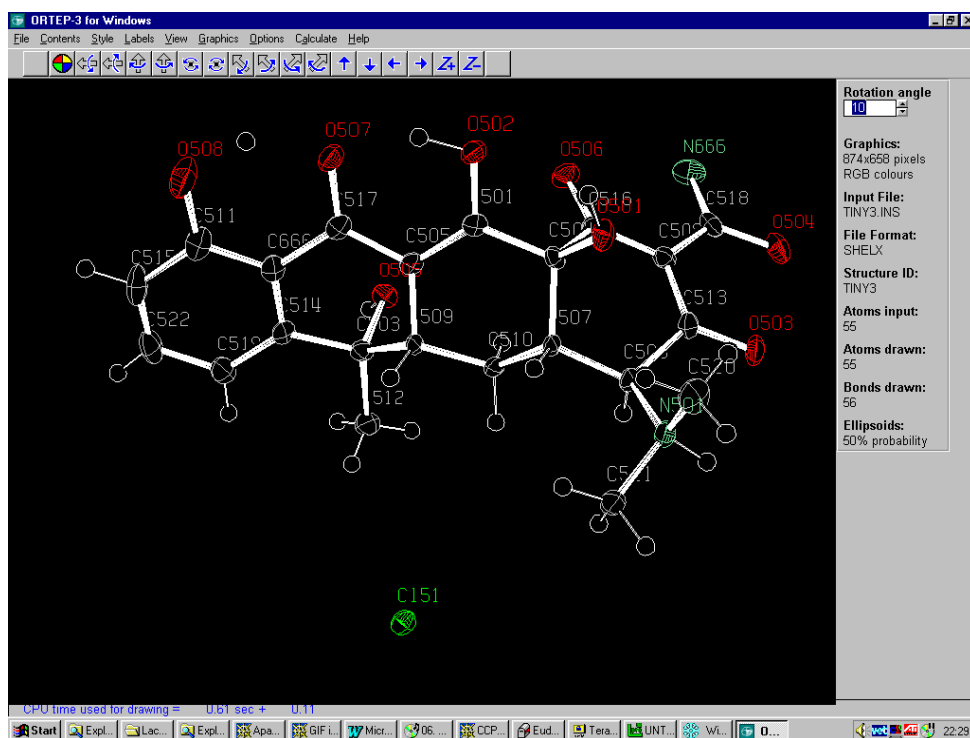
Most of these programs are made for single crystal applications but are very applicable to powder diffraction problems. They also can generate information on bond lengths, angles and local co-ordination at the click of a button.

Program	Address	Compatible File Formats
Platon for UNIX/Windows	http://www.cryst.chem.uu.nl/platon/	Cif, Platon SPF, Shelx INS / RES, PDB
GUI WinORTEP / GUI WinStruplo/ WinGX for Windows	http://www.chem.gla.ac.uk/~louis/software/	Shelx, CIF, GX, SPF/Platon, ORTEP, CSD/CCDC FDAT, CSSR XR, Crystals, GSAS, Sybol MOL/MOL2, MDL MOL, XYZ file, Brookhaven PDB, Rietica-LHPM, Fullprof
ORTEX	http://www.nuigalway.ie/cryst/software.htm	Shelx INS/RES
Crystals, Cameron	http://www.xtl.ox.ac.uk/	Shelx INS/RES, Crystals
Gretep	http://www.ccp14.ac.uk/tutorial/lmgp/ http://www.ccp14.ac.uk/ccp/web-mirrors/lmgp-laugier-bochu/	Shelx INS/RES, Gretep, Powder Cell, Lazy Pulvarix

Structure Viewing in Gretep (can handle inorganics and alloys as well):



Structure Viewing in GUI WinORTEP (reads many file formats including LHPM, Fullprof and GSAS files)



Powder Diffraction Pattern Calculation

Pretty much every Rietveld program can calculate powder patterns but specialised programs may be of benefit.

Program	Address	Compatible File Formats and major features
GUI Powder Cell for Windows	http://www.bam.de/a_v/v_1/powder/e_cell.html	Powder Cell (CEL), Shelx, ICSD. Structure transformation options.
Lazy Pulvarix for Windows within WinGX	http://www.chem.gla.ac.uk/~louis/software/	Formats importable into GUI WinOPTeP, then saved as Shelx. Shelx, CIF, GX, SPF/Platon, ORTEP, CSD/CCDC FDAT, CSSR XR, Crystals, GSAS, Sybol MOL/MOL2, MDL MOL, XYZ file, Brookhaven PDB, Rietica-LHPM, Fullprof
Powdis and Powutl / ORTEX Suite	http://www.nuigalway.ie/cryst/software.htm	Shelx INS/RES, CIF
GUI Poudrix for Windows	http://www.ccp14.ac.uk/tutorial/lmgp/ http://www.ccp14.ac.uk/ccp/web-mirrors/lmgp-laugier-bochu/	Shelx INS/RES, Poudrix, Powder Cell, Lazy Pulvarix. Calculates anomalous dispersion correctly for synchrotron wavelengths.
Lazy Pulvarix from the author, Prof. Erwin Parthe	E-mail to erwin.parthe@chiam.unige.ch	Lazy Pulvarix
Lazy Pulvarix via the ICSD web	http://barns.ill.fr/dif/icsd/	

Examples of the a) Powder Cell Interface and b) Poudrix GUI Interface

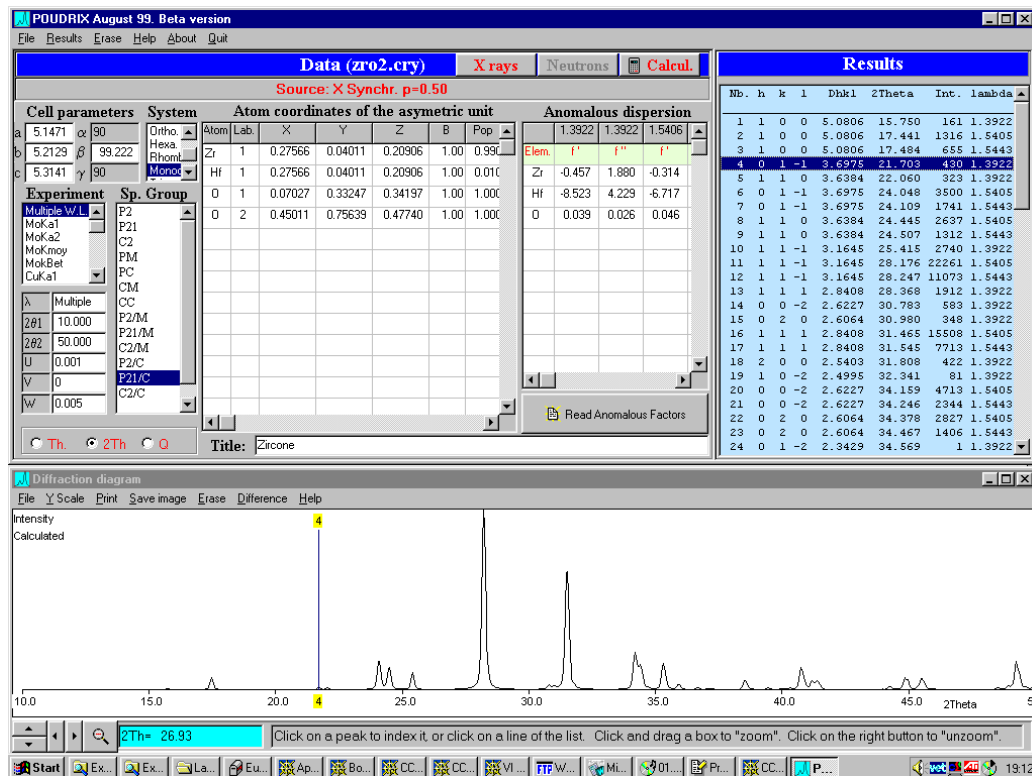
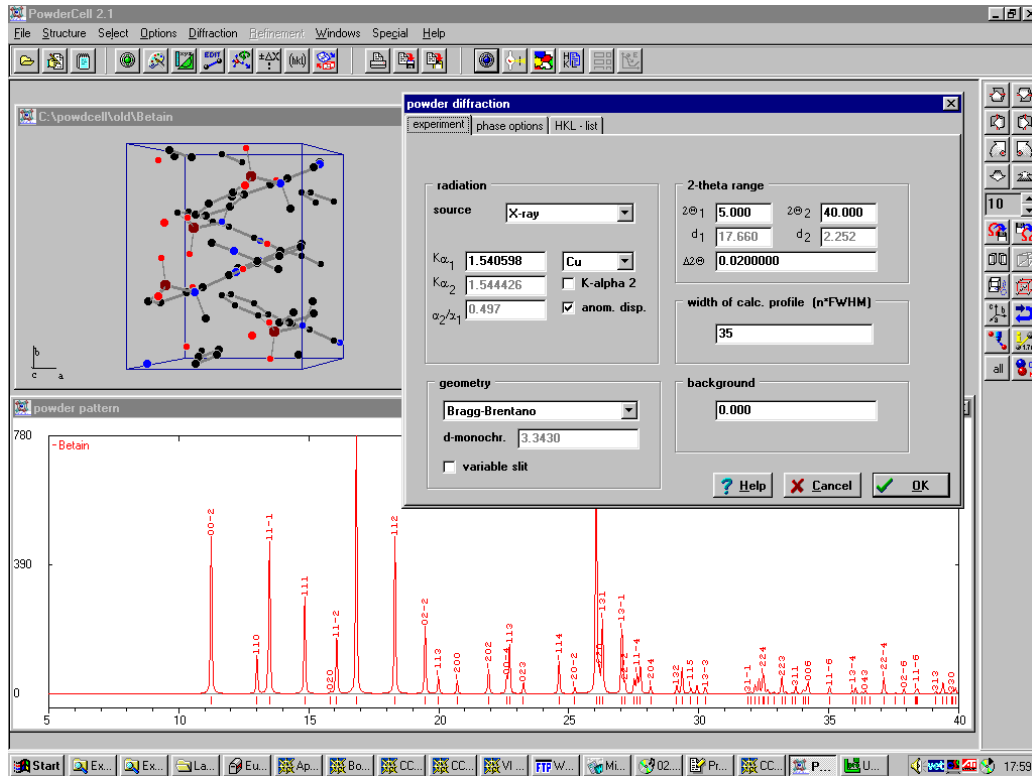
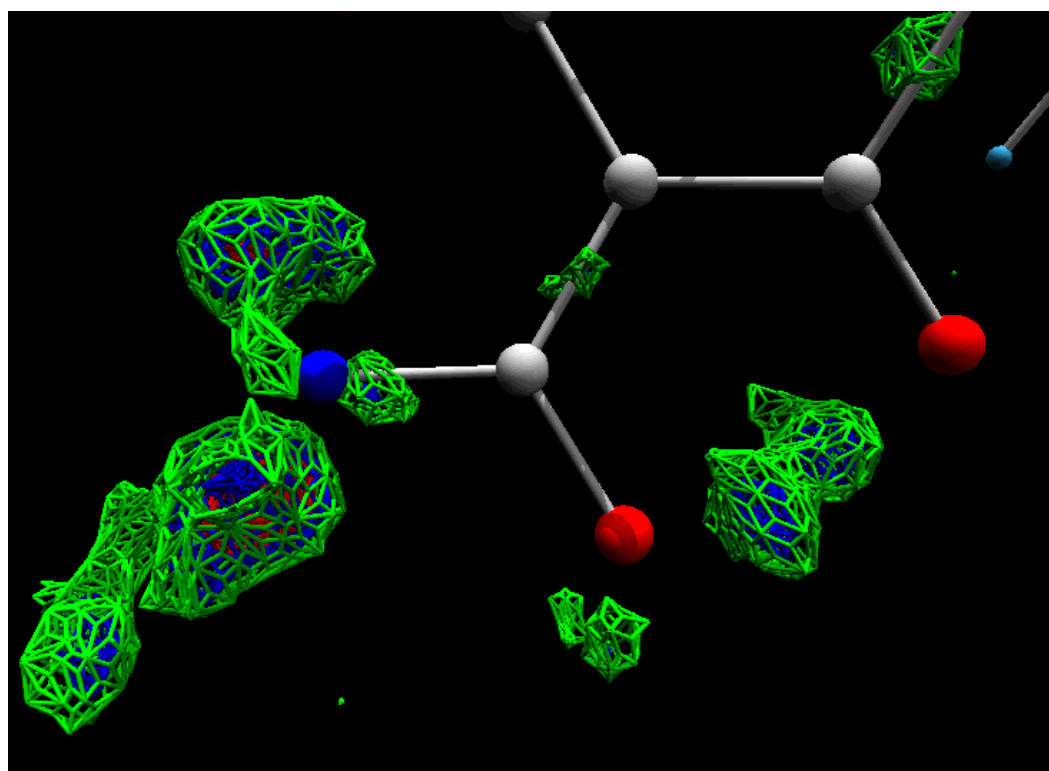
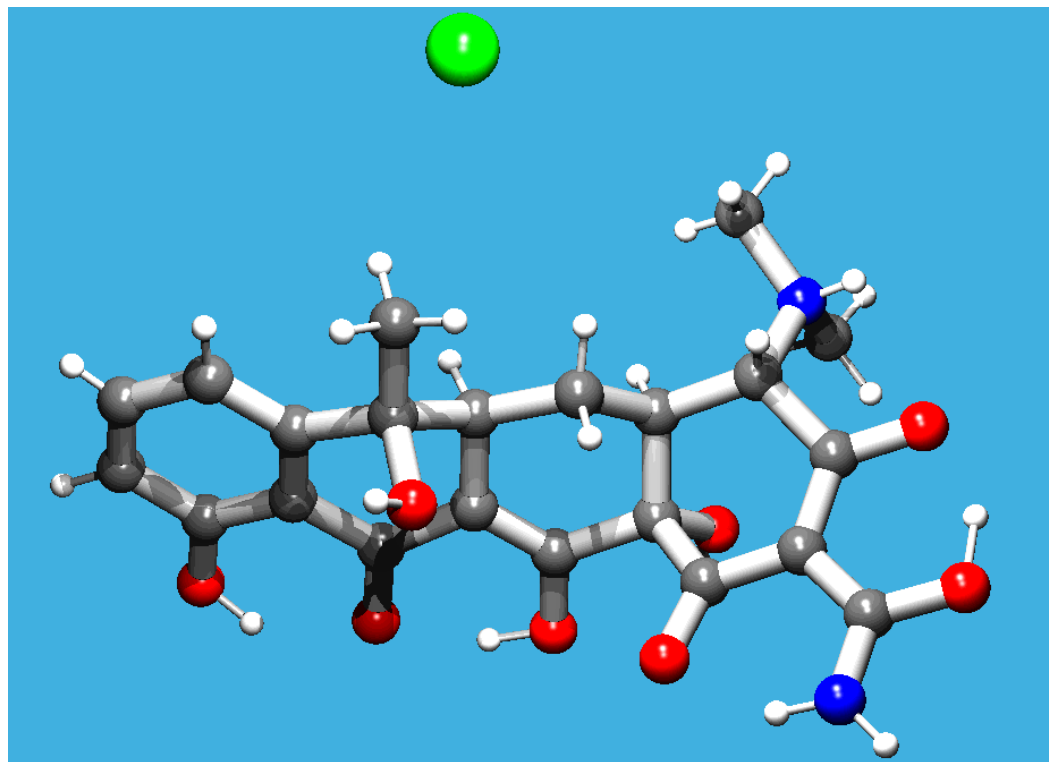


Photo realistic rendering of crystal structures and Fourier contour maps

This is trivial to perform and is available via a wide variety of software. For most programs you will need the freely available Povray software (<http://www.povray.org>)

Program	Address	Rendering engine	Compatible File Formats
GUI Powder Cell for Windows	http://www.bam.de/a_v/v_1/powder/e_cell.html	Povray	Powder Cell (CEL), Shelx, ICSD..
GUI WinORTEP, GUI WinStruplo, WinGX	http://www.chem.gla.ac.uk/~louis/software/	Raster3D and Povray	Formats importable into GUI WinORTEP and GUI WinStruplo. Shelx, CIF, GX, SPF/Platon, ORTEP, CSD/CCDC FDAT, CSSR XR, Crystals, GSAS, Sybol MOL/MOL2, MDL MOL, XYZ file, Brookhaven PDB, Rietica-LHPM, Fullprof. Anisotropic atoms.
ORTEX	http://www.nuigalway.ie/cryst/software.htm	Raster3D	Shelx INS/RES, CIF. Rendering of Movies. Anisotropic atoms.
Platon	http://www.cryst.chem.uu.nl/platon/	Raster3D and Povray	Shelx INS/RES, Poudrix, Powder Cell, Lazy Pulvarix. Calculates anomalous dispersion correctly for synchrotron wavelengths. Anisotropic atoms.
Xtal	http://xtal.crystal.uwa.edu.au/		Lazy Pulvarix
Marching Cubes	http://mysak.umbr.cas.cz/~husakm/Public/MarchingCubeELD/MarchingCubeELD.htm	Povray	Fourier Maps: Crystals, WinGX, Project XD. Refer also to FOUE for converting GSAS Fourier Maps into Marching Cubes format

Photorealistic Rendering a) Structures using Platon via Povray and b) Fourier Maps using Marching Cubes 3D via Povray



Miscellaneous Links

Site	Address
Crystallographic Nexus CD-ROM for crystallographers isolated from the internet	http://www.unige.ch/crystal/stxnews/nexus/index.htm
FreeBSD and Windows Multiboot Tutorial	http://www.ccp14.ac.uk/solution/bsdunix/
Linux and Windows Multiboot Tutorial	http://www.ccp14.ac.uk/solution/linux/linwin95a.html
Linux for Crystallography	http://www.ccp14.ac.uk/solution/linux/
IUCr Website and mirrors Including Crystallography World Wide and Sincris Crystallographic Software Library	http://www.iucr.org/iucr-top/ http://www.us.iucr.org/iucr-top/ http://www.ch.iucr.org/iucr-top/ http://www.se.iucr.org/iucr-top/ http://www.ru.iucr.org/iucr-top/ http://www.fr.iucr.org/iucr-top/ http://www.il.iucr.org/iucr-top/ http://www.au.iucr.org/iucr-top/ http://www.jp.iucr.org/iucr-top/ http://www.za.iucr.org/iucr-top/
Crystallography World Wide Educational Resources (and Mirrors)	http://www.iucr.org/cww-top/edu.index.html http://www.us.iucr.org/cww-top/edu.index.html http://www.ch.iucr.org/cww-top/edu.index.html http://www.se.iucr.org/cww-top/edu.index.html http://www.ru.iucr.org/cww-top/edu.index.html http://www.fr.iucr.org/cww-top/edu.index.html http://www.il.iucr.org/cww-top/edu.index.html http://www.au.iucr.org/cww-top/edu.index.html http://www.jp.iucr.org/cww-top/edu.index.html http://www.za.iucr.org/cww-top/edu.index.html
Advanced Certificate in Powder Diffraction on the Web: Birkbeck College, University of London	http://pd.cryst.bbk.ac.uk/pd/
Internet course: Learn how to succeed in a Structure Determination by Powder Diffractometry (SDPD) with the help of experts by the University of Maine (France)	http://sdpd.univ-lemans.fr/course/
QTA Internet Course Learn how to Quantitatively Determine the Crystallographic Texture of materials using x-ray and neutron diffraction by D. Chateigner and A. Jouanneaux	http://pecdc.univ-lemans.fr/qta/
Strategies in Structure Determination from Powder Data by A. Le Bail	http://sdpd.univ-lemans.fr/iniref/tutorial/indexa.html

Web Search Engines

Web Search Engine	Address
Google	http://www.google.com
Altavista	http://www.altavista.com
Links2Go Crystallography	http://www.links2go.com/topic/Crystallography
Inference Find	http://www.infind.com/
EZResult	http://www.easyresults.com/
Hotbot Web Search	http://www.hotbot.com
Multicrawl	http://www.multicrawl.com/
Meta Crawler	http://www.metacrawler.com/
All The Web	http://www.alltheweb.com/

EXPO2000: a New Package for *Ab-initio* Structure Solution from Powder Data

Angela Altomare, Carmelo Giacovazzo, Antonietta Guagliardi, Anna Grazia Giuseppina Moliterni & Rosanna Rizzi

*IRMEC c/o Dipartimento Geomineralogico, Università di Bari,
Campus Universitario, Via Orabona 4, 70125 Bari, Italy*

Abstract

New techniques have been integrated into EXPO (Altomare et al., 1999), a package devoted to extracting the integrated intensities from an indexed powder diffraction pattern and to solving crystal structures *via* Direct Methods. An efficient peak-search algorithm and an improved version of the program TREOR90 (Werner et al., 1985) have been introduced in order to offer new tools for a more powerful powder pattern indexing (Altomare et al., 2000a). A procedure aiming at completing a partial crystal structure *via* the prior information on the coordination polyhedra has been also implemented (Altomare et al., 2000b). The updated version of EXPO, EXPO2000, is indeed able to produce and complete a structural model directly from the experimental unindexed powder diagram.

Introduction

The main steps in a crystal structure determination process from powder data are: a) indexing of the diffraction pattern; b) space group determination; c) crystal structure solution; d) crystal structure refinement.

The success in the step a) is a preliminary condition for any other study. Due to the larger size unit cells studied by modern powder methods peak overlap is often severe. The possible simultaneous presence of origin shifts, impurity phases, inaccuracy in the experimental measures, etc., makes the number of failures of the indexing process still non negligible. Luckily the most used indexing programs, like ITO (Visser, 1969), TREOR90 (Werner et al., 1985), DICVOL91 (Boultif & Louër, 1991), exploit the experimental data *via* different approaches so that one program can succeed if the others fail. One of them, TREOR90, a trial-and-error program that searches for solutions in the index space, is particularly suitable to accept modifications of the procedure without losing its simple basic indexing approach. This suggested us to associate new code to TREOR90 in EXPO2000 in order to face the indexing process as

follows: i) an effective peak search procedure is defined; ii) the coded crystallographic decisions of TREOR90 are modified to increase its effectiveness and to reduce the user intervention. Such a new version will be called N-TREOR (Altomare et al., 2000a) from now on; iii) the cell suggested by N-TREOR is automatically refined *via* PIRUM, a least-squares d-refinement program (Werner, 1969; Eriksson & Werner, 1995); iv) all the procedure is made friendly *via* a graphic interface.

While steps b) and d) did not recently benefit by important advances, the step c) became more straightforward owing to improvements in Direct Methods techniques (Altomare et al., 1999) and to the introduction of new search procedures based on the prior knowledge of the molecular geometry (Harris et al., 1994; Andreev et al. 1997; Kariuki et al., 1997; Shankland et al., 1997, Altomare et al., 2000b). The last techniques operate as follows: structural fragments are rotated and translated through the unit cell under simulated annealing and genetic algorithm techniques. Internal coordinates can take the molecular flexibility into account. Criteria based on the agreement between the experimental and calculated diffraction patterns evaluate the feasibility of each trial model. About fifty structures have been solved *via* the application of these techniques.

On one hand Direct Methods only need information on unit cell content, cell parameters and space group but require high quality for the decomposition process of the experimental pattern in single intensities. On the contrary, Monte Carlo Methods are less sensible to the quality of the experimental data, but require a larger prior information on the molecular geometry.

A third approach has been recently proposed (Altomare et al., 2000b). Direct Methods are first applied to locate heavy atoms then the complete structural model is constructed *via* Monte Carlo techniques on the basis of the connectivity of the heavy-atoms provided by the Direct Methods solution. The approach (POLPO procedure) combines the advantages of Direct Methods with the computing power exploited by Monte Carlo techniques.

Both the new indexing procedure and POLPO have been introduced in EXPO2000, a revised version of EXPO (Altomare et. al., 2000a-b). EXPO2000 is therefore a program that may supply a complete structural model directly from the unindexed experimental powder diagram.

Indexing in EXPO2000

The peak search procedure

An accurate and precise peak location is not a trivial task. The most applied methods are: i) the peak profile fit (Huang & Parrish, 1975), ii) the derivative method (Sonneveld & Visser, 1975; Huang, 1988). The noise-sensitivity of the derivative method often generates spurious peaks: indeed smoothing of experimental data is advisable before its application (Savitzky & Golay, 1964) but even so spurious peaks can still survive. The procedure implemented in EXPO2000 exploits both profile fitting and derivative methods *via* the following sequential steps:

- a) use of a gaussian profile filter (from now on "filter technique");
- b) derivative technique applied to a cubic polynomial function;
- c) cluster profile fitting process.

The filter technique consists of:

- 1) search for the maximum count value in the pattern $[y_{\text{obs}}(i), 2\theta_i]$. Let $2\theta_{\text{max}}$ be its 2θ value;
- 2) calculation of a gaussian profile $\{y_{\text{calc}}(i)\}$ centred at $2\theta_{\text{max}}$, after preliminary estimate of the FWHM and asymmetry parameters;
- 3) search for the maximum count value in the difference pattern $[y_{\text{obs}}(i), 2\theta_i]$ so obtained: $y_{\text{obs,new}}(i) = y_{\text{obs,old}}(i) - y_{\text{calc}}(i)$. If $y_{\text{obs}}(i) < 0$, $y_{\text{obs}}(i)$ is set to zero.

Steps 1) - 3) are repeated until N peaks with an observed count higher than a threshold THR value, are selected.

If the profile fit (especially for strongly asymmetric observed peaks) is not satisfactory or a strong noise effect occurs, the search in the difference pattern $\{y_{\text{obs}}\}$ can generate artefact peaks (see, f. e., Figs.1-2 for two of the test structures quoted in Table 1).

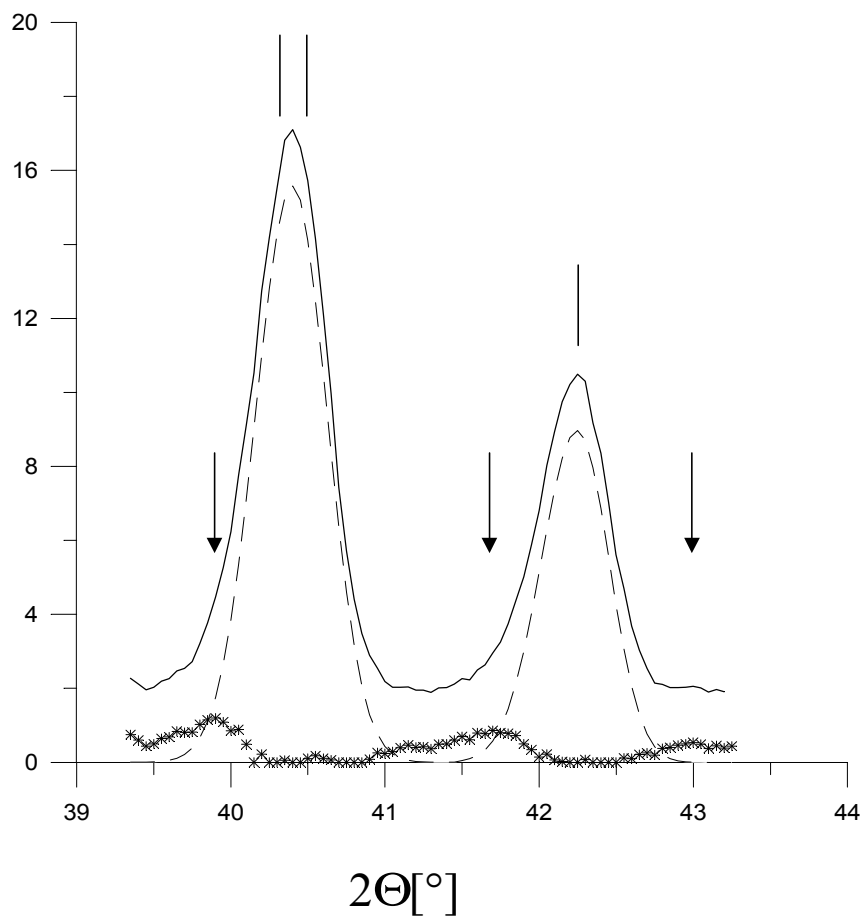


Fig. 1: *BACO: experimental profile (full line), calculated profile (dashed line) and difference profile (star line). The vertical arrows indicate false peak positions as a result of the filter method application. The vertical bars indicate the reflection positions calculated from the published cell parameters.*

Table 1: Code name and crystallochemical data for the test structures. X: data collected by home diffractometer; N: Neutron data; S: Synchrotron data.

Structure Code	Space group	Unit cell content
AGPZ (X) ¹	Pbca	Ag ₈ N ₁₆ C ₂₄ H ₂₄
BACO (N) ²	C2/m	Ba ₄ C ₈ O ₂₀ D ₈
BAMO (X) ³	P2 ₁	Ba ₄ Mo ₁₂ O ₄₀
BENZ (S) ⁴	P2 ₁ /a	C ₂₄ F ₁₂
CF3BR (N) ⁵	P2 ₁ /a	C ₄ Br ₄ F ₁₂
CFCL (N) ⁶	Fdd2	C ₈ F ₁₆ Cl ₁₆
CIME (S) ⁷	P2 ₁ /n	S ₄ C ₄₀ N ₂₄ H ₆₄
CROX (X) ⁸	P $\bar{1}$	Cr ₈ O ₂₁
CUPZ (X) ⁹	Pbca	Cu ₈ N ₁₆ C ₂₄ H ₂₄
DADA (X) ¹⁰	P2 ₁ 2 ₁ 2 ₁	Ti ₈ K ₄ Si ₁₂ O ₄₀
EMT (S) ¹¹	P6 ₃ /mmc	(Si,Al) ₉₆ Na ₂₈ O ₂₀₄
GAPO (S) ¹²	Pbca	Ga ₃₂ P ₃₂ O ₁₂₈ F ₈ C ₅₆
KUO (S) ¹³	P4/ncc	U ₄ P ₄ K ₄ O ₃₆ D ₂₄
LAMO (S) ¹⁴	P2 ₁ /a	La ₄ Mo ₂₀ O ₃₂
LASI (N) ¹⁵	P2 ₁ /c	La ₈ Si ₈ O ₂₈
LASO (X) ¹⁶	C 2/c	Mo ₂₀ O ₆₄ C ₄₈ N ₈ H ₂₄
LEV (X) ¹⁷	R3m	[Si ₅₄ O ₁₀₈] ₃ C ₈ NH ₁₆
MCM (S) ¹⁸	P6/mmm	Si ₇₂ O ₁₄₄
MES (X) ¹⁹	P2 ₁ /c	C ₂₄ N ₄ O ₂₀ S ₄ H ₅₂
METYL (S) ²⁰	I222	Na ₁₆ C ₁₆ H ₄₈
NAGE (S) ²¹	R3m	Ge ₂₁ Na ₉ N ₃ O ₆₆ D ₄₈
NBPO (S) ²²	C2/c	Nb ₂₀ O ₁₂₀ P ₂₈
NIZR (S) ²³	P2 ₁ /n	Ni ₄ Zr ₈ P ₄ O ₁₆
PBS (S) ²⁴	Pbca	Pb ₈ S ₁₆ O ₂₄
SAPO (X) ²⁵	Pmmn	Si ₃₂ O ₆₄ N ₂ C ₄₈
SBPO (S) ²⁶	P2 ₁ /n	Sb ₈ P ₁₄ O ₄₈
SGT (S) ²⁷	I4 ₁ /amd	Si ₆₄ O ₁₂₈ C ₁₀₄
SULPH (N) ²⁸	Pbcm	S ₈ D ₁₆
UTM-1(S) ²⁹	C2/m	Si ₄₄ O ₈₈
VFI (S) ³⁰	P6 ₃	Al ₁₈ P ₁₈ O ₁₁₄
YONO (S) ³¹	P2 ₁	Y ₈ O ₂₆ N ₂ H ₁₈

1) Masciocchi et al. (1994); 2) Christensen (1992); 3) Werner et al. (1997); 4) Williams et al. (1992); 5) Jouanneux et al. (1992); 6) Cockcroft & Fitch (1991); 7) Cernik et al. (1991); 8) Norby et al. (1991); 9) Masciocchi et al. (1994); 10) Dadachov & Le Bail (1997); 11) Baerlocher et al. (1994); 12) Meden et al. (1997); 13) Fitch & Cole (1991); 14) Hibble et al. (1988); 15) Christensen (1994); 16) Lasocha & Schenk (1997); 17) McCusker (1993); 18) Cambor et al. (1998); 19) Christensen et al. (1993); 20) Weiss et al. (1990); Roberts & Fitch (1991); 22) Zah-Letho et al. (1992); 23) Jouanneux et al. (1991a); 24) Christensen et al. (1991); 25) Estermann et al. (1992); 26) Jouanneux et al. (1991b); 27) McCusker (1988); 28) Cockcroft & Fitch (1990); 29) Plévert et al. (1999); 30) McCusker et al. (1991); 31) Christensen et al. (1992)

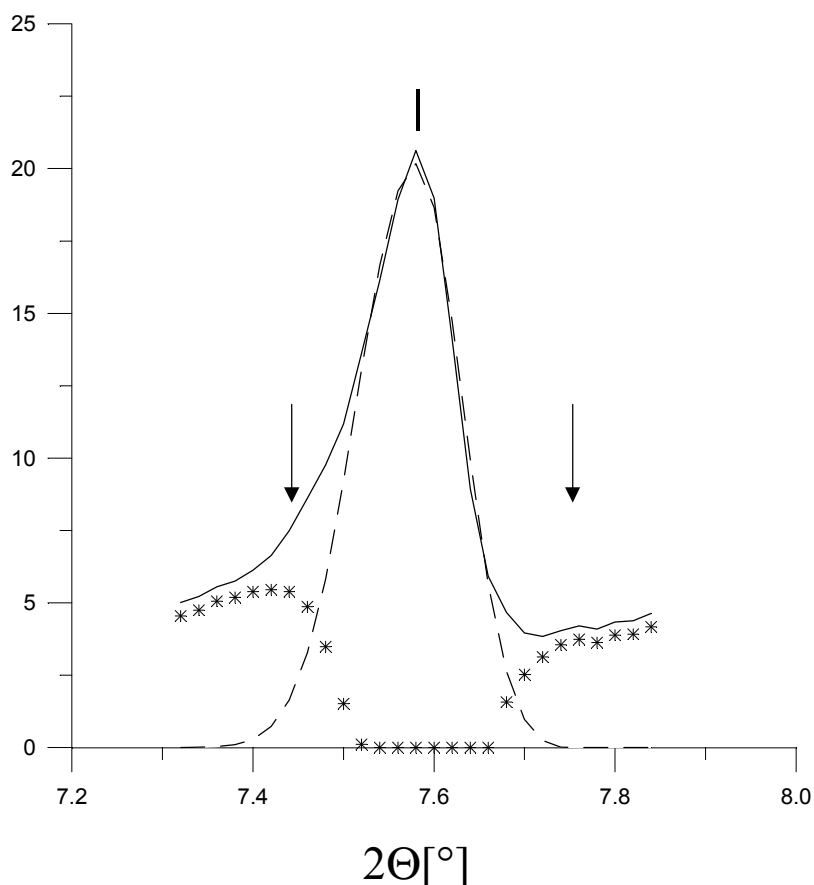


Fig. 2: SAPO: experimental profile (full line), calculated profile (dashed line) and difference profile (star line). The vertical arrows indicate false peak positions as a result of the filter method application. The vertical bars indicate the reflection positions calculated from the published cell parameters.

Most of the artifact peaks are recognized and discarded by letting the filter technique be followed by a derivative method applied to a cubic polynomial, i. e., the best fit curve for few experimental points belonging to neighborhoods of the candidate peaks.

Via a graphic interface the user can set new search conditions or change the peak search results, by adding more peaks and/or deleting the undesired ones.

The last step of the peak search procedure, the cluster profile fitting process, aims at improving the peak location via a preliminary refinement taking into account peak overlap. For each selected cluster of peaks the procedure alternates integrated intensity estimates via the Le Bail method (Le Bail et al., 1988) with least squares cycles refining all the profile parameters, the scale factors, background coefficients and Bragg 2θ positions.

The indexing procedure

At the end of the peak search procedure the refined d values are automatically supplied to N-TREOR. Unlike in TREOR90, N-TREOR is able to automatically take several decisions when a default run fails or when a plausible solution is found:

- 1) if satisfactory results are not found by a default run, the unit cell search is repeated with wider error limits;
- 2) if no plausible solution is still found, the maximum (h,k,l) Miller indices of the orthorhombic and monoclinic base lines are increased, and the error limits are narrowed to 50 per cent of the default values. That avoids the generation of a great number of big, erroneous unit cells and the elimination of correct solutions before their refinement;
- 3) if the maximum observed d value is larger than 10 Å, N-TREOR sets the maximum cell volume and the maximum axis value to be investigated to 4000 Å³ and 35 Å respectively (their default values in TREOR90 are 2000 Å³ and 25 Å respectively);
- 4) the dominant-zone tests are improved. If more than the first 6 peaks can be indexed by a two-dimensional rectangular cell, but no three-dimensional solution is found, N-TREOR indicates the first non-indexed peak as a possible impurity line that may be deleted by a second run;
- 5) the error tolerance $\Delta = |\sin^2(\theta_{\text{obs}}) - \sin^2(\theta_{\text{calc}})| < \varepsilon$ is wavelength dependent, where θ_{obs} and θ_{calc} are observed and calculated (from the trial cell) Bragg θ -values respectively, ε is a threshold value. In TREOR90 the default value of ε is optimised for Cu-K α radiation data (ε_{Cu}); in N-TREOR the ε value is a function of the used wavelength and it is related to ε_{Cu} according to the relation $\varepsilon = \left(\frac{\lambda}{\lambda_{\text{Cu}}} \right)^2 \cdot \varepsilon_{\text{Cu}}$, where λ is the experimental neutron or X-ray wavelength, λ_{Cu} is the Cu-K α wavelength;
- 6) monoclinic solutions are checked for possible rhombohedral symmetry. If the symmetry is likely to be rhombohedral, N-TREOR calculates the possible hexagonal axes;
- 7) a possible zero shift in the 2θ -Bragg angles (see below) is automatically found and applied to calculations.
The effectiveness of the implemented procedure is confirmed by the tests presented in the next chapter. The procedure may be schematically described as follows (see flow diagram in Fig. 3):
 - i) N-TREOR is first run by using the estimated d values. If a monoclinic or triclinic cell is found, N-TREOR calculates first the reduced cell and

then the conventional cell *via* the program REDUCT (Westdahl & Werner, 1995). More candidate cells can be suggested: in order to select the most probable one, a modified de Wolff (de Wolff, 1968) figure of merit M'_{20} is used, with $M'_{20} = (7 - N_{\text{par}}) \cdot M_{20}$, (1) where N_{par} is the number of cell parameters to be determined (6 for triclinic system, 4 for monoclinic,...), and M_{20} is the original de Wolff figure of merit. *Via* the heuristic relation (1) N-TREOR prefers a higher symmetry solution.

- j) at the end of the first run (say the standard run), N-TREOR systematically applies origin shifts in both directions along the 2θ axis, in step of $\Delta_{2\theta}$ ($\Delta_{2\theta}$ is two times the experimental 2θ step). For each shift a new N-TREOR run is performed: the origin shift corresponding to the maximum M'_{20} value $[(M'_{20})_{\text{max}}]$ is selected as the shift leading to the correct 2θ set. The automatic 2θ correction is stopped when a cell with $M'_{20} < (M'_{20})_{\text{max}}$ is found.

If a cell has been found with figures of merit $M'_{20} (\equiv M_{20}) \geq 20$ for the triclinic crystal system, or $M'_{20} \geq 30$ ($M_{20} \geq 10$) for monoclinic or higher symmetry crystal systems, then the unit cell parameters are refined by PIRUM. Originally this was an interactive program that has been suitably modified and implemented in EXPO2000 in order to automatically refine the unit cell parameters without any user intervention. The refinement step is followed by a statistical study of index parity of the calculated reflections to detect the presence of doubled axes or of additional lattice points (A, B, C, I, R or F centred cell), and to change the reflection conditions for new refinement cycles. The application of PIRUM includes two supplementary shifts of $\pm \Delta_{2\theta}/4$ to the selected origin shift, and some refinement cycles are again performed aiming at obtain a higher M'_{20} value.

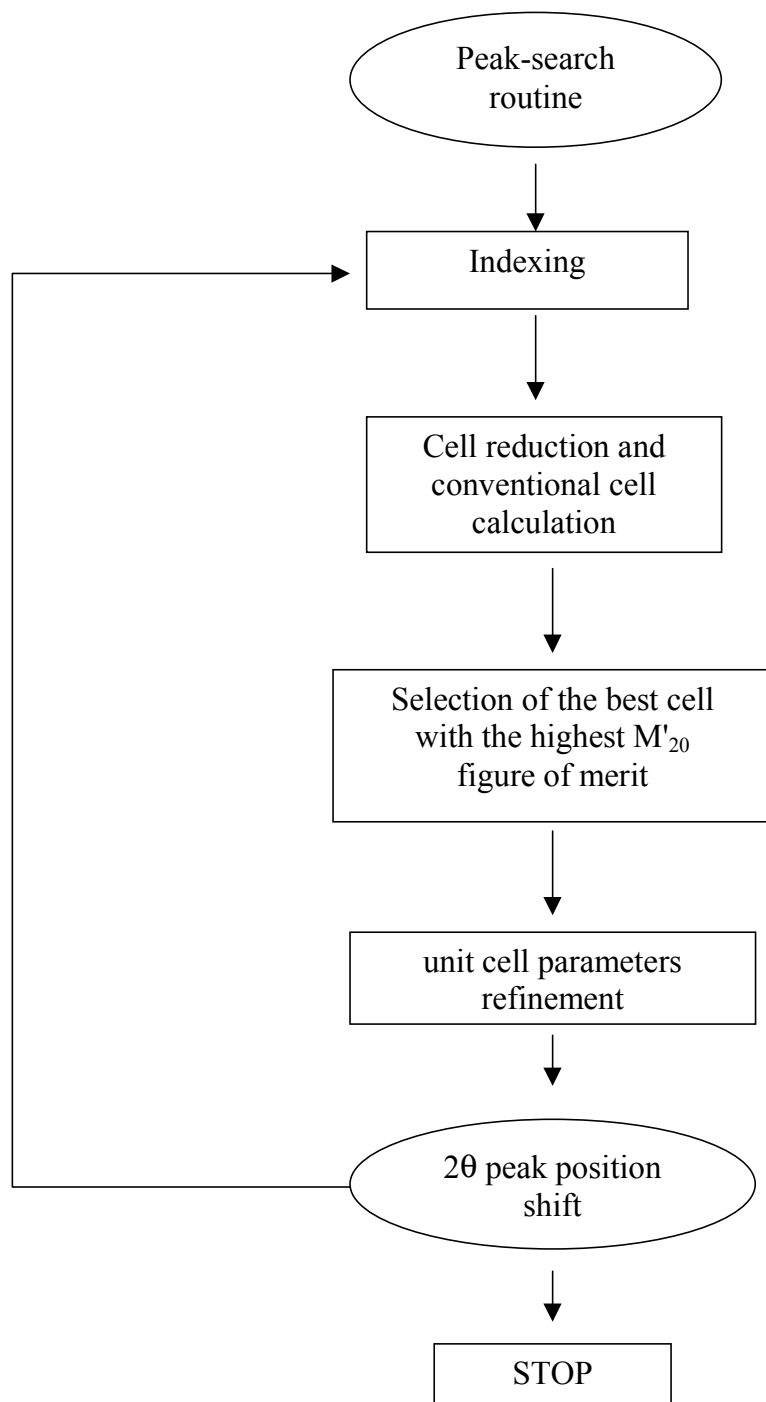


Fig. 3: Flow diagram of the indexing process in EXPO2000.

Applications

The method described in §2.2 has been applied to the set of test structures quoted in Table 1. The results of the application of TREOR90 and N-TREOR are listed in Table 2, where Y and N indicate that the correct unit cell has been found in a default mode and has not been found in a default mode, respectively.

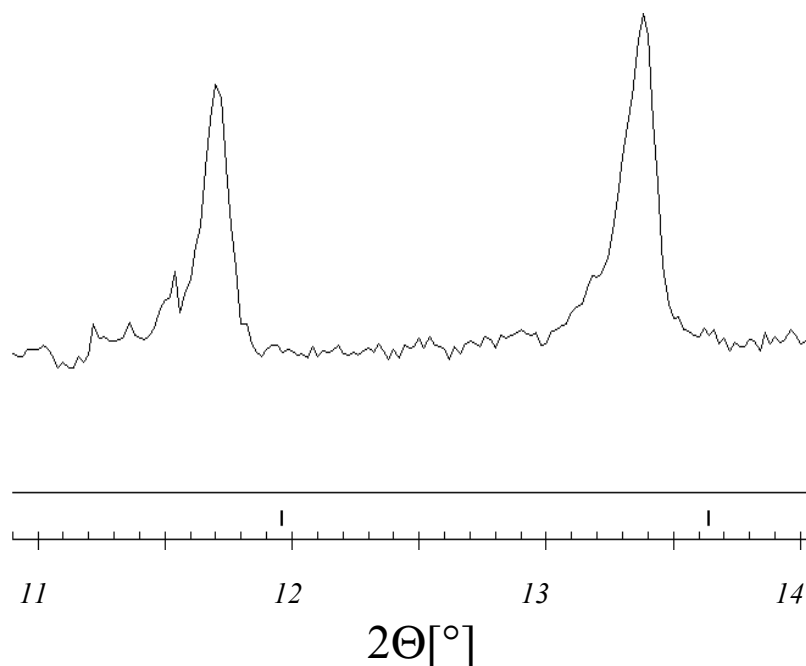


Fig.4: *MES: Zoom of the experimental pattern. The vertical bars at the bottom indicate the reflection positions calculated from the published unit cell parameters.*

TREOR90 is able to correctly index most of the structures in the default mode except AGPZ, BACO, CF3BR, EMT, LASO, MES, NBPO, SULPH and UTM-1. N-TREOR correctly indexes AGPZ, BACO, CF3BR, MES and SULPH only when a 2θ origin shift correction is applied by the program. The cases of MES, LASO and NBPO may be illustrative examples. For MES, the experimental diffraction pattern is characterised by a strong mismatch between the corrected and observed 2θ Bragg values (see Fig. 4). If positive origin shifts or no shift are applied N-TREOR did not suggest a monoclinic unit cell. Only applying negative shifts larger (in absolute value) than 0.14, monoclinic cells are found with increasing M_{20} as well as M'_{20} values up to a 2θ shift of 0.25 (for higher values of the shift M_{20} and M'_{20} decrease, see Fig. 5). This trend suggests a larger confidence in the reliability of the selected monoclinic unit cell. For the test structures whose experimental pattern is not affected by 2θ origin shift, M_{20} and M'_{20} follow again a similar trend but their maximum value is attained for vanishing origin shifts.

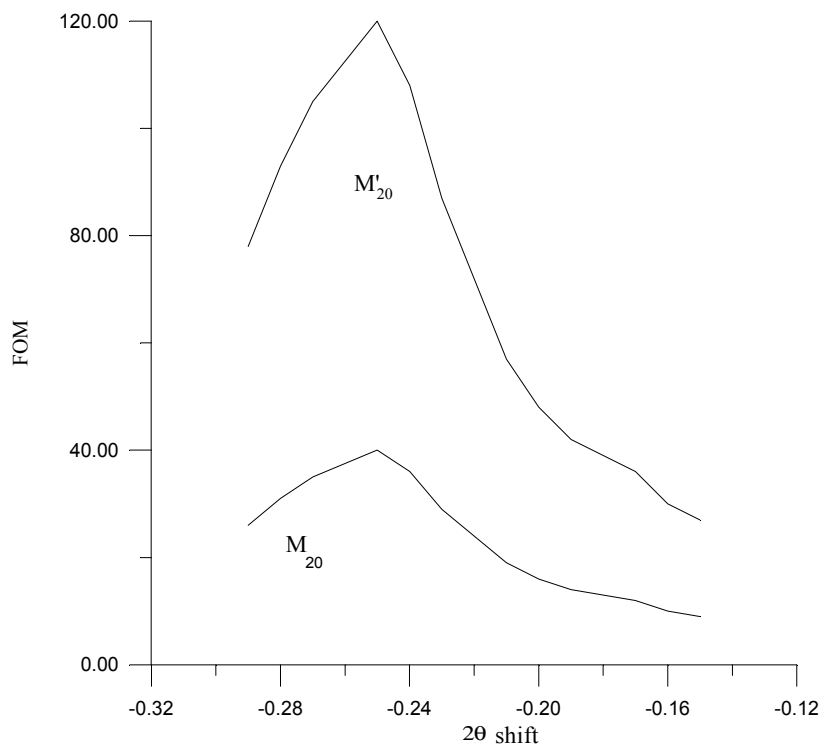


Fig.5: *MES: M_{20} and M'_{20} versus the origin shift as a result of the automatic indexing procedure implemented in EXPO2000.*

LASO and NBPO are also interesting cases. Both these structures have a long cell axis ($a = 28.691 \text{ \AA}$ and 29.8661 \AA respectively). TREOR90 does not suggest a plausible unit cell for LASO and finds a triclinic one for NBPO. The tools described in §2.2 (see, in particular, the procedure in point 2) make N-TREOR able to find the correct cell parameters. The strategy described in point 3 of the previous chapter, enables N-TREOR (unlike TREOR90) to correctly index in default mode, UTM-1 and EMT which have a long cell axis ($b=30.6991$, and $c=28.3459$ respectively). In many cases (e.g., for structures indexed by TREOR90, like CUPZ, LEV, METYL, NAGE, NIZR, PBS, SAPO, YONO in Table 3) the new features of N-TREOR enable the user to obtain unit cells with considerably higher figures of merit, reinforcing the user confidence in this program. Particularly significant cases are CUPZ, METYL, NIZR, PBS and YONO, for which the low-level M_{20} figure of merit obtained by TREOR90 is replaced by a higher figure obtained after PIRUM refinement.

Table 2: For each test structure the published cell parameters and the result of the application of TREOR90 and N-TREOR (both in their default modes) are respectively given. Y denotes success, N denotes failure.

Structure Code	Cell parameters	TREOR90	N-TREOR
AGPZ (X)	a=6.53 b=20.06 c=6.47	N	Y
BACO (N)	a=10.06 b=7.93 c=6.85 $\beta=122.35$	N	Y
BAMO (X)	a=14.69 b=7.57 c=6.96 $\beta=100.38$	Y	Y
BENZ (S)	a=9.52 b=7.43 c=7.54 $\beta=95.60$	Y	Y
CF3BR (N)	a=8.14 b=5.85 c=7.96 $\beta=111.72$	N	Y
CFCL (N)	a=10.17 b=14.96 c=5.10	Y	Y
CIME (S)	a=10.70 b=18.83 c=6.83 $\beta=111.31$	Y	Y
CROX (X)	a=5.43 b=6.56 c=12.12 $\alpha=106.37$ $\beta=95.74$ $\gamma=77.95$	Y	Y
CUPZ (X)	a=6.67 b=19.83 c=6.08	Y	Y
DADA (X)	a=7.14 b=9.91 c=12.94	Y	Y
EMT (S)	a=17.39 c=28.35	N	Y
GAPO (S)	a=14.68 b=15.43 c=16.97	Y	Y
KUO (S)	a=6.99 c=17.78	Y	Y
LAMO (S)	a=9.91 b=9.09 c=7.57 $\beta=109.05$	Y	Y
LASI (N)	a=5.41 b=8.80 c=14.29 $\beta=112.74$	Y	Y
LASO (X)	a=28.69 b=5.69 c=14.37 $\beta=113.22$	N	Y
LEV (X)	a=13.06 c=22.61	Y	Y
MCM (S)	a=14.21 c=29.94	Y	Y
MES (X)	a=8.59 b=9.93 c=11.11 $\beta=93.75$	N	Y
METYL (S)	a=6.81 b=18.86 c=6.65	Y	Y
NAGE (S)	a=10.87 c=13.74	Y	Y
NBPO (S)	a=29.87 b=8.72 c=8.79 $\beta=91.77$	N	Y
NIZR (S)	a=12.39 b=8.93 c=8.84 $\beta=90.55$	Y	Y
PBS (S)	a=7.18 b=6.92 c=16.11	Y	Y
SAPO (X)	a=21.94 b=13.69 c=7.12	Y	Y
SBPO (S)	a=11.94 b=8.73 c=8.32 $\beta=91.12$	Y	Y
SGT (S)	a=10.24 c=34.38	Y	Y
SULPH (N)	a=4.08 b=13.38 c=6.72	N	Y
UTM-1 (S)	a=9.48 b=30.70 c=7.31	N	Y
VFI (S)	a=18.98 c=8.10	Y	Y
YONO (S)	a=9.38 b=16.39 c=3.63 $\beta=101.07$	Y	Y

The integration of N-TREOR with PIRUM in EXPO2000 can be particularly useful when more plausible and not related cells with similar M_{20} and/or M'_{20} figures of merits are found by N-TREOR. Also the results of the intensity extraction process performed (in a subsequent step) by EXPO2000 using the various plausible unit cells can help in choosing the correct one: the

most probable cell is usually marked by the lowest profile fitting residual (Altomare et al., 2000b).

```
%structure mes
%job MES - X-ray data by courtesy of A. N.
%init
%data
range 5.0 88.0 0.02
pattern mes.pow
wave 1.5406
peak
%ntreor
%end
```

Fig. 6: MES: indexing input file for EXPO2000.

The EXPO2000 input file to be used in order to perform a powder pattern indexing is shown in Fig. 6. The directive 'peak' of the command '%data' has to be provided to run the peak-search procedure. The minimum and maximum 2θ values and the experimental step size are set by the directive 'range'. The peak-search procedure can be skipped providing EXPO2000 by a set of d values via an external file. The command '%ntreor', able to manage all the TREOR90 keywords with the same syntax, automatically performs the powder pattern indexing.

The POLPO method

General considerations

Often crystal structures with few heavy atoms and many light atoms (i.e. minerals and inorganic structures) are not completely determined by Direct Methods if powder diffraction data are used. Heavy atoms are usually easily found while light atoms are missed or roughly located. The study of the cation-cation distances and possible prior crystallochemical information can suggest the nature of the coordination polyhedra (see, f. e., the MCM zeolite in Table 4 and Fig. 7, with 13 oxygen atoms in the framework and 8 silicon atoms showing a tetrahedral coordination). The main aim of POLPO is to suitably exploit this information to integrate Direct Methods results and recover the complete structure *via* Monte Carlo techniques. Most of the candidate models are rejected because violating the prior information available and/or the stated crystallochemical rules. If more chemically consistent configurations are retained (the *feasible configurations*), they should be variants (i.e. with slight differences in bond distances and angles) of the same structural model.

Table 3: For the test structures indexed by TREOR90 the M_{20} values of the cells found by TREOR90 $[(M_{20})_T]$ and by N-TREOR as refined by PIRUM $[(M_{20})_{NT}]$ are given.

Structure Code	$(M_{20})_T$	$(M_{20})_{NT}$
BAMO (X)	39	39
BENZ (S)	80	81
CFCL (N)	12	13
CIME (S)	125	125
CROX (X)	45	47
CUPZ (X)	14	25
DADA (X)	70	70
GAPO (S)	34	38
KUO (S)	110	110
LAMO (S)	56	57
LASI (N)	15	15
LEV (X)	26	59
MCM (S)	52	55
METYL (S)	10	45
NAGE (S)	149	311
NIZR (S)	14	56
PBS (S)	11	56
SAPO (X)	25	30
SBPO (S)	29	29
SGT (S)	27	27
YONO (S)	11	78

Whatever the connectivity is, rich or poor, criteria based on the agreement between the whole experimental diffraction pattern and the calculated one are used to select the correct configuration among the feasible ones supplied by Monte Carlo techniques. The most probable configurations will be refined and/or completed by Rietveld refinement. The procedure is not time consuming since the wrong structural model are usually immediately refused, being not compatible with chemistry and/or crystallochemical constraints.

Table 4: *Shortest Si-Si distances for MCM: published data.*

Si(1) - Si(2) = 3.03	Si(5) - Si(7) = 3.02
- Si(2) = 3.03	- Si(7) = 3.02
- Si(2) = 3.03	- Si(7) = 3.02
- Si(1) = 3.16	- Si(4) = 3.24
Si(2) - Si(1) = 3.03	Si(6) - Si(3) = 3.16
- Si(3) = 2.99	- Si(6) = 3.14
- Si(3) = 2.99	- Si(7) = 3.17
- Si(4) = 3.08	- Si(7) = 3.17
Si(3) - Si(2) = 2.99	Si(7) - Si(5) = 3.02
- Si(3) = 3.11	- Si(6) = 3.17
- Si(6) = 3.16	- Si(6) = 3.17
Si(4) - Si(2) = 3.08	- Si(8) = 3.11
- Si(2) = 3.08	Si(8) - Si(7) = 3.11
- Si(2) = 3.08	- Si(8) = 2.96
- Si(5) = 3.24	- Si(8) = 3.13
	- Si(8) = 3.13

Two main sources of errors can interfere in POLPO performances: a) a strong misplacement of the heavy atoms, causing wrong indications on the cation connectivity (e.g., the assumption of false anion bridges between polyhedra, or, viceversa, the omission of really existing bridges). Indeed, incorrect assumptions generate false constraints or neglect some true ones; b) a distortion of polyhedra in the real structure. When no prior information about specific geometrical irregularity is available, POLPO usually builds regular polyhedra. This behaviour could generate a unreliable structural models. However, some distortions in the generated polyhedra are automatically introduced by POLPO because, in order to establish a satisfactory chemical connectivity, it tries to locate the anions introducing a tolerance on the assumed average bond angles and distances.

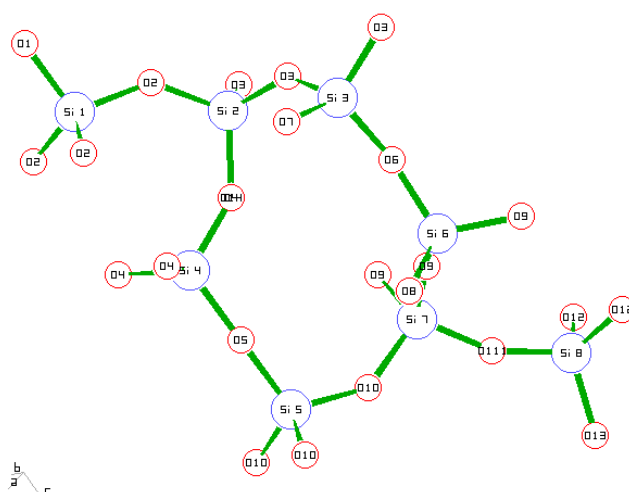


Fig. 7: MCM: Coordination polyhedra (published data) of the eight symmetry independent Si cations.

When the crystal structures may be described in terms of polyhedra, POLPO presents several advantages over the classic direct-space procedures mentioned in the introduction. These methods, in fact: i) cannot be used *ab initio* if no prior structural model may be formulated; ii) have to vary a large number of parameters to attain the necessary flexibility when several polyhedra are in the asymmetric unit.

On the contrary, POLPO has been designed to progressively build the (a priori unknown) structural model, to verify its crystallochemical consistence at each step of the procedure and to generate several variants of the model until one or more feasible configurations are defined.

1.1.1 The procedure

The POLPO procedure aims at fixing the coordinates of n_a anions (indicated by $A(j)$, $j=1, \dots, n_a$ from now on) starting from a set of n_k heavy atoms (indicated by $K(j)$, $j=1, \dots, n_k$ from now on) correctly (even if approximately) located by the Direct Methods section in EXPO2000. The accuracy of the atomic coordinates may be rough: therefore the values are improved by an automatic Rietveld refinement process, which uses, without varying them, the profile parameters estimated in the full pattern decomposition stage in EXPO2000. The next step performed by POLPO is the analysis of the interatomic distances between heavy atoms in order to detect the cation connectivity. The expected coordination polyhedra associated to the cations in the asymmetric unit, the average bond distances (indicated by $B(j)$, $j=1, \dots, n_k$ from now on) and the tolerance on bond distances and angles are supplied by the user *via* directives. POLPO starts the process by positioning anions bonded to cations with rich connectivity so as to exploit geometrical constraints useful for the right location

of the anions. Let us suppose that the anion A(1) is bonded to K(1) and K(2): then it is expected to lie on the circle intersection of two spheres, having radii B(1) and B(2) and centres in K(1) and K(2). A random point belonging to the circle is chosen as a trial position of A(1): then POLPO locates the coordination polyhedron of K(1) by a random rotation of it around the K(1)-A(1) axis, up to when its position satisfies the expected connectivity about K(1). Obviously a high connectivity strongly limits the free rotation around the K(1)-A(1) axis: accordingly wrong polyhedra orientations are immediately rejected. If the positions of all the anions belonging to the K(1) polyhedron are consistent with the expected connectivity they are retained as feasible atomic positions. POLPO checks the site symmetry of each feasible anion: if it is very closed to a symmetry element, it is moved on the symmetry element. Atoms symmetry equivalent to the feasible ones, say the A(1,q), q=1,...m, are calculated, located and their bonds established. A similar procedure is used to orient the polyhedra in the asymmetric unit. The process stops when all the anions have been located.

The profile residual $R_{\text{prof}} = \sum |y_i(\text{obs.}) - y_i(\text{calc.})| / \sum y_i(\text{obs.})$ is calculated for each feasible configuration, where $\{y_i\}$ are the profile counts. The smallest values of R_{prof} should correspond to the most reliable models. Each of them is refined *via* the Rietveld method in order to improve the quality of the atomic positions and discard the uncorrect ones.

1.1.2 Applications

The procedure has been applied to eight crystal structures: CROX, NIZR, SAPO, MCM, SGT, EMT, UTM-1, and VFI (see Table 1). We will describe the most relevant applicative aspects and the final results for four of them (MCM, SGT, UTM-1 and CROX). An analogous outcome is obtained for the other test structures; results are summarised for brevity.

MCM - The published crystal structure is depicted in Fig. 7: the shortest Si-Si distances (published data) are shown in Table 4. The cation-cation connectivity in Table 4 suggests (as expected for a zeolite) that the Si polyhedra are tetrahedra. The user directives fixing the connectivity conditions are:

```
tetr Si(1) 1.60 0.2 0.2
tetr Si(2) 1.60 0.2 0.2
tetr Si(3) 1.60 0.2 0.2
tetr Si(4) 1.60 0.2 0.2
tetr Si(5) 1.60 0.2 0.2
tetr Si(6) 1.60 0.2 0.2
tetr Si(7) 1.60 0.2 0.2
```

tetr Si(8) 1.60 0.2 0.2

We note: a) 1.60Å is the expected average Si-O distance; b) x=0.2, y=0.2 are the tolerance parameters for Si-O distances and O-Si-O angles respectively. I.e., 1.60(1±x)Å defines the range for the allowed bond distances and 109.47(1±y)° defines the range for the allowed bond angles for the tetrahedral coordination. Analogously, 90(1±y)° should define the range for the allowed bond angles for the octahedral coordination.

POLPO found 3 feasible solutions (cpu time 257 sec. using a Compaq Personal Workstation 500au SPECfp95: 19.5) all of them being variants of the same solution. The configuration corresponding to the best R_{prof} value (0.23) is assumed as the most informative. At the end of the procedure, 13 O atoms are assigned by POLPO. The average distance between our and published atomic positions is $\langle d \rangle = 0.23\text{Å}$. 20 cycles of automatic Rietveld refinement dropped down R_{prof} to 0.16 with $\langle d \rangle = 0.18\text{Å}$.

SGT - The published crystal structure is depicted in Fig. 8: the minimal Si-Si distances (published data) are shown in Table 5. Also in this case the Si atoms are tetrahedrally coordinate. Accordingly, the following directives are given by the user:

tetr Si(1) 1.60 0.2 0.2

tetr Si(2) 1.60 0.2 0.2

tetr Si(3) 1.60 0.2 0.2

tetr Si(4) 1.60 0.2 0.2

Table 5: Shortest Si-Si distances for SGT: published data.

Si(1) - Si(1) = 3.16	Si(3) - Si(1) = 3.16
- Si(1) = 3.16	- Si(2) = 3.11
- Si(3) = 3.16	- Si(2) = 3.11
- Si(4) = 2.98	- Si(3) = 3.15
Si(2) - Si(3) = 3.11	Si(4) - Si(1) = 2.98
- Si(3) = 3.11	- Si(2) = 3.09
- Si(4) = 3.09	- Si(2) = 3.09
- Si(4) = 3.09	- Si(4) = 3.09

POLPO found 6 feasible solutions (cpu time 98 sec.), all the structure models are variants of the same model. That with the lowest value of R_{prof} (0.49) is assumed as the most informative. At the end of the procedure, 7 O atoms are assigned by POLPO, with average distance (with respect to the published O positions) equal to 0.45Å. 20 cycles of automatic Rietveld refinement dropped down R_{prof} to 0.28 with $\langle d \rangle = 0.21\text{Å}$.

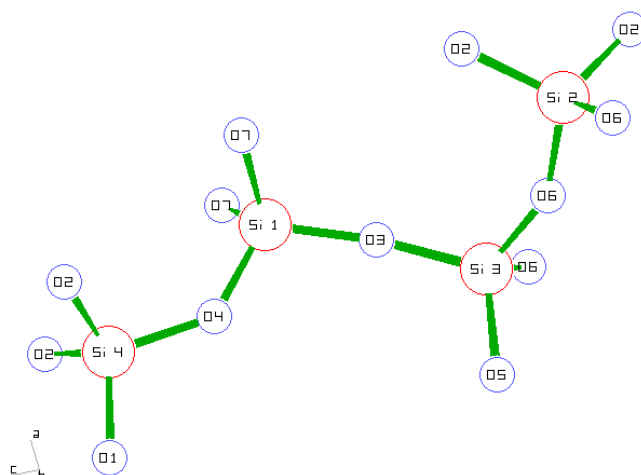


Fig. 8: SGT: Coordination polyhedra (published data) of the four symmetry independent Si cations.

UTM-1 - The published crystal structure is depicted in Fig. 9: the cation-cation connectivity is shown in Table 6 and suggests a tetrahedral coordination for Si atoms. Accordingly, the following directives are given:

tetr Si(1) 1.60 0.2 0.2

tetr Si(2) 1.60 0.2 0.2

tetr Si(3) 1.60 0.2 0.2

tetr Si(4) 1.60 0.2 0.2

tetr Si(5) 1.60 0.2 0.2

tetr Si(6) 1.60 0.2 0.2

Table 6 - Shortest cation–cation distances for UTM-1: published data.

Si(1) - Si(1) = 3.0	Si(4) - Si(2) = 3.13
- Si(1) = 3.08	- Si(3) = 3.09
- Si(3) = 3.08	- Si(5) = 3.15
- Si(5) = 3.14	- Si(6) = 3.02
Si(2) - Si(2) = 3.12	Si(5) - Si(1) = 3.14
- Si(4) = 3.13	- Si(4) = 3.15
- Si(6) = 3.04	- Si(5) = 3.06
- Si(6) = 3.10	- Si(6) = 3.17
Si(3) - Si(1) = 3.08	Si(6) - Si(2) = 3.04
- Si(1) = 3.08	- Si(2) = 3.10
- Si(4) = 3.09	- Si(4) = 3.02
- Si(4) = 3.09	- Si(5) = 3.17

POLPO found 9 feasible solutions (cpu time 94 sec.), both being variants of the same solution. That with the lowest R_{prof} value (0.23) is assumed as the

most informative. At the end of the procedure, 13 O atoms are located by POLPO. For them the average distance (with respect to the published O positions) is equal to 0.34Å. 20 cycles of automatic Rietveld refinement dropped down R_{prof} to 0.11 with $\langle d \rangle = 0.20\text{\AA}$.

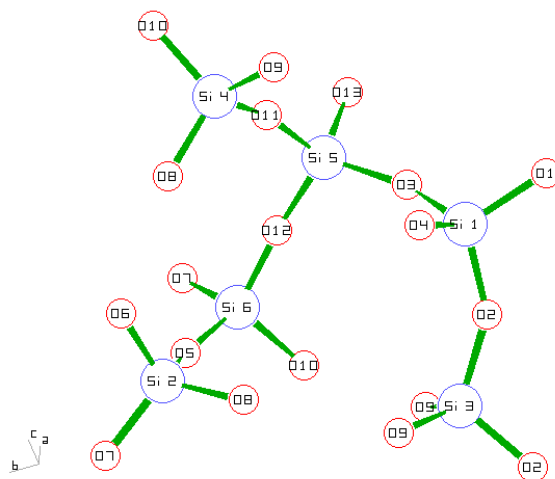


Fig. 9: *UTM-1: Coordination polyhedra (published data) of the six symmetry independent Si cations.*

CROX - The published crystal structure is depicted in Fig. 10: the shortest Cr–Cr distances (published data) are shown in Table 7. The cation-cation connectivity in Table 7 suggests that the Cr(1) polyhedron is an octahedron, and that Cr(2)–Cr(4) atoms coordinate tetrahedrally. Accordingly, the following directives are given by the user:

octa Cr(1) 1.92 0.2 0.2

tetr Cr(2) 1.75 0.2 0.2

tetr Cr(3) 1.75 0.2 0.2

tetr Cr(4) 1.75 0.2 0.2

Table 7 - Shortest Cr-Cr distances for CROX: published data.

Cr(1) - Cr(1) = 2.98	Cr(3) - Cr(4) = 3.22
- Cr(2) = 3.28	- Cr(1) = 3.52
- Cr(2) = 3.30	- Cr(1) = 3.53
- Cr(2) = 3.47	Cr(4) - Cr(3) = 3.22
- Cr(3) = 3.52	- Cr(4) = 3.35
- Cr(3) = 3.53	
- Cr(2) = 3.58	
Cr(2) - Cr(1) = 3.28	
- Cr(1) = 3.30	
- Cr(1) = 3.47	
- Cr(1) = 3.58	

POLPO found 5 feasible solutions (cpu time 33 sec) all of them being variants of the same solution. That with the lowest R_{prof} value (0.22) is assumed as the most informative. At the end of the procedure 11 O atoms are assigned by POLPO, with average distance (with respect to the published O positions) $\langle d \rangle = 0.24 \text{ \AA}$. 20 cycles of automatic Rietveld refinement dropped down R_{prof} to 0.16 with $\langle d \rangle = 0.28 \text{ \AA}$.

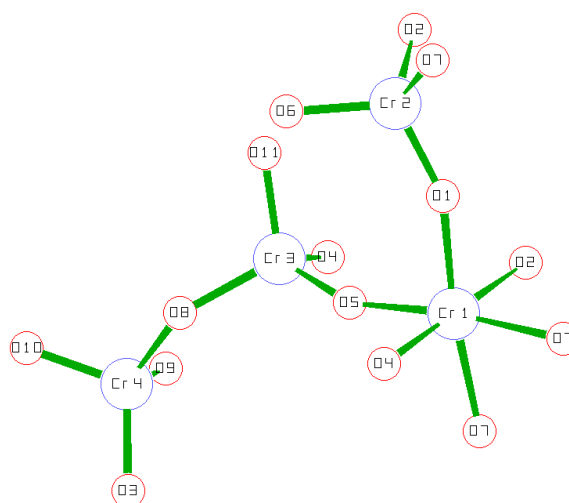


Fig. 10: CROX: Coordination polyhedra (published data) of the four symmetry independent Cr cations.

NIZR - POLPO found 2 feasible solutions (cpu time 34 sec.), both being variants of the same solution. That with the lowest R_{prof} value (0.22) is assumed as the most informative. At the end of the procedure 13 O atoms are located by POLPO instead of the expected 12 O atoms. For them the average distance (with respect to the published O positions) is equal to 0.23 \AA . 20 cycles of automatic Rietveld refinement dropped down R_{prof} to 0.15 with $\langle d \rangle = 0.16 \text{ \AA}$. The 13 atom was recognised to be wrong owing to the quite high vibrational parameter.

SAPO - POLPO found 6 feasible solutions (cpu time 38 sec), all of them being variants of the same solution. That with the lowest value of R_{prof} (0.29) is assumed as the most informative. At the end of the procedure 10 O atoms are assigned by POLPO, with average distance (with respect to the published O positions) equal to 0.27Å. 20 cycles of automatic Rietveld refinement dropped down R_{prof} to 0.14 with $\langle d \rangle = 0.19\text{Å}$.

EMT- POLPO found 2 feasible solutions (cpu time 103 sec), all of them being variants of the same solution. That with the lowest value of R_{prof} (0.12) is assumed as the most informative. At the end of the procedure 12 framework O atoms are assigned by POLPO, with average distance (with respect to the published O positions) equal to 0.23Å. After 20 cycles of automatic Rietveld refinement the R_{prof} is 0.14 with $\langle d \rangle = 0.23\text{Å}$.

VFI - POLPO found 8 feasible solutions (cpu time 81 sec), all of them being variants of the same solution. That with the lowest value of R_{prof} (0.25) is assumed as the most informative. At the end of the procedure 14 framework O atoms are assigned by POLPO, with average distance (with respect to the published O positions) equal to 0.37Å. 20 cycles of automatic Rietveld refinement dropped down R_{prof} to 0.11 with $\langle d \rangle = 0.30\text{Å}$.

Conclusions

A new indexing technique and a new procedure aiming at completing via Monte Carlo method a partial model supplied by Direct Methods have been described. Both the procedures (N-TREOR and POLPO respectively) have been implemented in EXPO2000, the heir of the EXPO package. N-TREOR presents some new features, absent in TREOR90, making the powder pattern indexing process more efficient. The crystal structure completion techniques requires only the prior information on the coordination polyhedra and, for each polyhedron, on the average cation–anion distance. The procedure can only handle atoms with octahedral and tetrahedral coordination, and assumes that all the heavy atoms have been correctly (even if approximately) located. It can be particularly useful when the completion of the structure is not straightforward because of the moderate quality of the experimental pattern and/or of the low fractional heavy atom scattering power. The method has been successfully applied to nine test structures.

References

- Altomare, A., Burla, M. C., Camalli, M., Carrozzini, B., Cascarano, G., Giacobazzo, C., Guagliardi, A., Moliterni, A. G. G., Polidori, G., Rizzi, R., (1999). *J. Appl. Cryst.* **32**, 339-340
- Altomare, A., Giacobazzo, C., Guagliardi, A., Moliterni, A. G. G., Rizzi, R. & Werner, P.-E., (2000a). *J. Appl. Cryst.* **33**, in press
- Altomare, A., Giacobazzo, C., Guagliardi, A., Moliterni, A. G. G. & Rizzi, R., (2000b). *J. Appl. Cryst.*, Submitted.
- Andreev, Y. G., MacGlashan, G. S. & Bruce, P. G. (1997). *Phys. Rev. B* **55**, 12011-12017.
- Boultif, A. & Louër, D. (1991). *J. Appl. Cryst.* **24**, 987-993.
- De Wolff, P. M. (1968). *J. Appl. Cryst.* **1**, 108-113.
- Eriksson, L. & Werner, P.-E. (1995). World Directory of Powder Diffraction Programs, Release 3 (1995) by IUCr. Ed. Gorter, S. & Smith, D.
- Harris, K. D. M., Tremaine, M., Lightfoot, P. & Bruce, P. G. (1994). *J. Am. Chem. Soc.* **116**, 3543,3547
- Kariuki, B. M., Serrano-Gonzales, H., Johnston, R. L. & Harris, K. D. M. (1997). *Chem. Phys. Lett.* **280**, 189-195.
- Huang, T. C. (1988). *Aust. J. Phys.* **41**, 201-212.
- Huang, T. C. & Parrish, W. (1975). *Appl. Phys. Lett.* **27**, 123-124.
- Le Bail, A., Duroy, H. & Fourquet, J.L. (1988). *Math. Res. Bull.* **23**,447-452.
- Savitzky, A. & Golay, M. J. E. (1964). *Anal. Chem.* **36**, No. 8, 1627-1639.
- Shankland, K., David, W. I. F. & Csoka, T. (1997). *Z. Kristallogr.*, **212**, 550-552
- Sonneveld, E. J. & Visser, J. W. (1975). *J. Appl. Cryst.* **8**, 1-7.
- Visser, J. W. (1969). *J. Appl. Cryst.* **2**, 89-95.
- Werner, P.-E. (1969). *Ark. kemi.* **31**, 513-516.
- Werner, P.-E., Eriksson, L. & Westdahl, M. (1985). *J. Appl. Cryst.* **18**, 367-370.
- Westdahl, M. & Werner, P.-E. (1995). World Directory of Powder Diffraction Programs, Release 3 (1995) by IUCr. Ed. Gorter, S. & Smith, D.
- Baerlocher, C., McCusker, L. B. & Chiappetta, R. (1994). *Microporous Materials* **4**, 269-280.
- Camblor, M. A., Corma, A., Díaz-Cabañas, M.-J. & Baerlocher, C. (1998). *J. Phys. Chem. B* **102**, 44-51.

- Cernik, R. J., Cheetham, A. K., Prout, C. K., Watkin, D. J., Wilkinson, A. P. & Willis, B. T. M. (1991). *J. Appl. Cryst.* **24**, 222-226.
- Christensen, A. N. (1992). *Acta Chem. Scandin.* **46**, 240-243.
- Christensen, A. N. (1994). *Z. Kristallog.* **209**, 7-13.
- Christensen, A. N., Hazell, R. G., Hewat, A. W. & O'Reilly, K. P. J. (1991). *Acta Chem. Scandin.* **45**, 469-473.
- Christensen, A. N., Nielsen, M., O'Reilly, K. P. J. & Wroblewski, T. (1992). *Acta Chem. Scandin.* **46**, 224-230.
- Christensen, A. N., Hazell, R. G., Lehmann, M. S. & Nielsen, M. (1993). *Acta Chem. Scand.* **47**, 753-756.
- Cockcroft, J. K. & Fitch, A. N. (1990). *Z. Kristallog.* **193**, 1-19.
- Cockcroft, J. K. & Fitch, A. N. (1991). *Z. Kristallog.* **197**, 121-130.
- Dadachov, M.S. & Le Bail, A. (1997). *Eur. J. Solid State Inorg. Chem.* **34**, 381-390.
- Estermann, M. A., McCusker, L. B. & Baerlocher, C. (1992). *J. Appl. Cryst.* **25**, 539-543.
- Fitch, A. N. & Cole, M. (1991). *Mat. Res. Bull.* **26**, 407-414. Hibble, S. J., Cheetham, A. K., Bogle, A. R. L., Wakerley, H. R. & Cox, D. E. (1988). *J. Am. Chem. Soc.* **110**, 3295-3296.
- Jouanneaux, A., Verbaere, A., Piffard, Y., Fitch, A. N. & Kinoshita, M. (1991a). *Eur. J. Solid State Inorg. Chem.* **28**, 683-699.
- Jouanneaux, A., Verbaere, A., Guyomard, D., Piffard, Y., Oyetola, S. & Fitch, A. N. (1991b). *Eur. J. Solid State Inorg. Chem.* **28**, 755-765.
- Jouanneaux, A., Fitch, A. N. & Cockcroft, J. K. (1992). *Molec. Phys.* **71**, No.1, 45-50.
- Lasocha, W. & Schenk, H. (1997). *J. Appl. Cryst.* **30**, 909-914.
- Masciocchi, N., Moret, M., Cairati, P., Sironi, A., Ardizzoia, G. A. & La Monica, G. (1994). *J. Am. Chem. Soc.* **116**, No. 17, 7668-7676.
- McCusker, L. (1988). *J. Appl. Cryst.* **21**, 305-310.
- McCusker, L. B. (1993). *Mat. Sci. Forum* **133-136**, 423-434.
- Meden, A., Grosse-Kunstleve, R. W., Baerlocher, C. & McCusker, L. B. (1997). *Z. Kristallog.* **212**, 801-807.
- Norby, P., Christensen, A. N., Fjellvåg, H., Lehmann, M. S. & Nielsen, M. (1991). *J. Solid State Chem.* **94**, 281-293.

- Plévert, J., Yamamoto, K., Chiari, G. & Tatsumi, T. (1999). *J. Phys. Chem. B* **103**, 8647-9649.
- Roberts, M. A. & Fitch, A. N. (1991). *J. Phys. Chem. Solids* **52**, No. 10, 1209-1218.
- Weiss, E., Corbelin, S., Cockcroft, J. K. & Fitch, A. N. (1990). *Chem. Ber.* **123**, 1629-1634.
- Werner, P.-E., Moustiakimov, M., Marinder, B.-O. & Knight, K. S. (1997). *Z. Kristallog.* **212**, 665-670.
- Williams, J. H., Cockcroft, J. K. & Fitch, A. N. (1992). *Angew. Chem. Int. Ed. Engl.* **31**, No.12, 1655-1657.
- Zah-Letho, J. J., Jouanneaux, A., Fitch, A. N., Verbaere, A. & Tournoux, M. (1992). *Eur. J. Solid State Inorg. Chem.* **29**, 1309-1320.
- McCusker, L.B., Baerlocher, Ch. (1991). *Zeolites* **11**, 308-313.

The Solution of Molecular Structures by Patterson Search Methods Using Powder Diffraction Intensity Data

Jordi Rius

*Institut de Ciència de Materials de Barcelona (CSIC)
Campus de la UAB, 08193-Bellaterra,
Catalunya (Spain)
E-mail: Jordi.Rius@icmab.es*

Introduction.

Until recently, the solution of crystal structures of relatively large molecular compounds required in most cases single-crystal intensity data ranging to atomic resolution. In 1988, Rius & Miravittles showed that Patterson search methods could be successfully applied to intensity data extracted from powder patterns, if large enough search model were available. Since then a number of crystal structures have been solved by this method. Nowadays, traditional Patterson search methods coexist with other direct-space approaches that control the model modification using efficient algorithms like the Monte Carlo, the simulated annealing, the genetic algorithm or error correcting codes among others.

Patterson search methods exploit the molecular geometry known from previous experiments or from force field calculations to place a similar molecular model in a crystal with unknown structure. In general, this process involves the determination of six parameters: three angular ones describing the rotation Ω to be applied to the input model, and three more corresponding to the coordinates of the shift \mathbf{t} necessary to position the already oriented model. Consequently, a six-dimensional search is needed. This search demands a lot of computing effort. Fortunately, this six dimensional search can be divided into two three dimensional ones if the Patterson function is explored with the rotation and the translation functions. The rotation function yields the more probably correct rotations, and once the model orientation has been found, the translation function provides the possible locations.

To understand how Patterson search methods work, first the principal properties of the Patterson function will be introduced followed by the description of the rotation and translation functions. To complete this

introduction, the direct methods sum function will be briefly discussed as a special type of Patterson search method.

Finally, some examples from the literature describing different situations where Patterson search methods were applied successfully to powder intensity data, will be presented.

The Patterson function and the molecular structure factor.

The value of the Patterson function at point \mathbf{u} is defined by the integral

$$P(\mathbf{u}) = \int_V \rho(\mathbf{r}) \cdot \rho(\mathbf{r} + \mathbf{u}) \cdot d\mathbf{r}$$

$$\rho(\mathbf{r}) = \sum_i Z_i \cdot \exp \left(-(\mathbf{r} - \mathbf{r}_i)^2 / \sigma^2 \right)$$

where the integration extends over the whole unit cell volume V . The physical meaning can be best understood if the electron density ρ at point \mathbf{r} is expressed as the sum of gaussian-like distributions centered at each atomic position \mathbf{r}_i , since then P can be rewritten as the double summation

$$P(\mathbf{u}) = \sum_j \sum_k Z_j Z_k \cdot \exp \left(-((\mathbf{r}_k - \mathbf{r}_j) - \mathbf{u})^2 / 2\sigma^2 \right)$$

It can be easily seen that the exponential factor will be the unity only for those points satisfying the condition $\mathbf{u} = \mathbf{r}_k - \mathbf{r}_j$. Each one of these points corresponds to the center of the peak between atoms j and k . If the unit cell contains N atoms, there will be N^2 peaks in the Patterson function unit cell from which N coalesce at $\mathbf{u} = 0$ to give a strong origin peak. In addition, the widths of the interatomic peaks are much larger than the atomic peaks. For molecular compounds where the height Z of all atoms are similar, there will be severe peak overlap in the Patterson function so that it will be impossible to identify individual peaks. This is in contrast with compounds containing a reduced number of heavy atoms. Since the height is equal to the product $Z_j \cdot Z_k$, interatomic peaks between heavy atoms will dominate over the rest of peaks.

For practical purposes it is convenient to decompose the Patterson function in two subsets of interatomic peaks:

- *self-Patterson*. This subset only includes the interatomic peaks between atoms j and k belonging to the same molecule. The vector $\mathbf{r}_k - \mathbf{r}_j$ only depends on the orientation of the molecule and its length is normally short. Consequently, these peaks will be concentrated around the origin of the Patterson function.

- *cross-Patterson*. This subset includes the peaks with j and k belonging to different molecules. Here, the vectors $\mathbf{r}_k - \mathbf{r}_j$ depend not only on the orientations of the molecules but also on their respective locations in the unit cell.

The Patterson function is normally computed in the form of a Fourier series with the intensities $I(\mathbf{H}) = |F(\mathbf{H})|^2$ as coefficients. To identify the parts of the intensity which contribute to the self- and to the cross-Pattersons it is advantageous to express the structure factor $F(\mathbf{H})$ as a function of the molecular structure factors $S_j(\mathbf{H})$,

$$F(\mathbf{H}) = \sum_{\mathbf{H}} S_j(\mathbf{H}) \exp(i2\pi\mathbf{H}\mathbf{r}_j), \quad (1)$$

where \mathbf{r}_j is the local origin of the j molecule with respect to the origin of the unit cell ($S_j(\mathbf{H})$ is computed with the atoms in molecule j referred to this local origin). If there is only one molecule in the asymmetrical unit, the number of molecules will be the order of the space group and the different \mathbf{r}_j will be related by the space group symmetry operations. Hence, the product $|F(\mathbf{H})|^2 = F(\mathbf{H})^* \cdot F(\mathbf{H})$ yields

$$I(\mathbf{H}) \propto \sum_j |S_j(\mathbf{H})|^2 + \sum_j \sum_{\mathbf{k} \neq j} S_j(\mathbf{H})^* S_{\mathbf{k}}(\mathbf{H}) \exp(-i2\pi\mathbf{H}(\mathbf{r}_j - \mathbf{r}_{\mathbf{k}})) \quad (2)$$

The first summation are the Fourier coefficients giving rise to the self-Patterson and the double summation those corresponding to the cross-Patterson. Since both \mathbf{r}_j and $\mathbf{r}_{\mathbf{k}}$ follow the space group symmetry, the difference vector $\mathbf{r}_j - \mathbf{r}_{\mathbf{k}}$ can be replaced by the expression

$$\mathbf{r}_j - \mathbf{r}_{\mathbf{k}} = (\mathbf{R}_j - \mathbf{R}_{\mathbf{k}}) \cdot \mathbf{r} + (\mathbf{t}_j - \mathbf{t}_{\mathbf{k}}) \quad (3)$$

in order to express the intensity as a function of the local origin vector \mathbf{r} of a single molecule (\mathbf{R}_j and \mathbf{t}_j are the rotational and translational components of the j space group symmetry operation).

The sum function.

There exist different functions to measure the coincidence or the discrepancy between the observed and the calculated values of the intensity data. Two of the most important ones are the Residual value (R) and the Correlation coefficient (C). Their expressions in terms of a collectivity Φ of arbitrary variables are

$$R(\Phi) = \sum_{\mathbf{H}} I_o(\mathbf{H})^2 + \sum_{\mathbf{H}} I_c(\mathbf{H})^2 - \sum_{\mathbf{H}} I_o(\mathbf{H}) \cdot I_c(\mathbf{H})$$

and

$$C(\Phi) = \sum_{\mathbf{H}} I_o(\mathbf{H}) \cdot I_c(\mathbf{H}) / (\sum_{\mathbf{H}} I_o(\mathbf{H})^2)^{1/2} \cdot (\sum_{\mathbf{H}} I_c(\mathbf{H})^2)^{1/2}$$

which must be ideally zero and the unity, respectively. Notice that both functions have the summation $\sum_{\mathbf{H}} I_o(\mathbf{H}) \cdot I_c(\mathbf{H})$ in common which must be a large positive quantity for the correct calculated values. This summation is called sum function and since it has no squared terms, it is specially well suited for Fourier analysis. Effectively, since the intensities are the Fourier coefficients of the Patterson function, the physical meaning of the sum function can be easily

interpreted as the integral of the product of observed and calculated Patterson functions extended over the whole volume of the unit cell

$$S(\Phi) = \int_V P_o(\mathbf{u}) \cdot P_c(\mathbf{u}, \Phi) \cdot d\mathbf{u}$$

In contrast to the R function, the sum function is less sensitive to scaling errors because it only contains products. Strictly speaking, all structure solution methods based either on the R or on the sum function of Patterson type functions should be considered as Patterson search methods. In the practice, however, this name is reserved for those based on the sum function only. This includes not only the rotation and translation functions but also the 'direct methods' sum function which as will be seen later only differs in the selection of the variables.

The rotation function.

Since the internal geometry of the molecular model is known, the corresponding set of intramolecular interatomic vectors can be readily computed. The calculated Patterson function is obtained by weighting each vector \mathbf{r}_{jk} with the product $Z_j Z_k$ (Z_j = atomic number of atom j). The rotation Ω of the vector set can be given either in Eulerian angles or in the axial rotation system. Although the former have considerable advantages in macromolecular crystallography because they allow the introduction of the Fast Fourier Transform algorithm in the computation of the rotation function, this advantage disappears in the case of powder diffraction because the models are rather small. Consequently, the simpler axial rotation system is normally used. It consists on two spherical polar angles (longitude ϕ and co-latitude ψ) which define the spin axis and a third angle χ specifying the rotation around this axis. To cover all possible model orientations, the following angular interval must be sampled:

$$0 \leq \phi \leq 360^\circ \quad 0 \leq \psi \leq 90^\circ \quad 0 \leq \chi \leq 360^\circ$$

If the rotation function (Rossmann & Blow, 1962) is expressed in terms of the Fourier coefficients of the observed and calculated Patterson functions

$$R(\Omega) = \sum_{\mathbf{H}} \mathbf{I}(\mathbf{H}) \cdot \sum_j \sum_k Z_j Z_k \cos(2\pi \mathbf{H} \Omega \mathbf{r}_{jk})$$

then it reduces to

$$R(\Omega) = \sum_j \sum_k Z_j Z_k P(\Omega \mathbf{r}_{jk})$$

if the \mathbf{H} summation is replaced by the value of the Patterson function at point $\Omega \mathbf{r}_{jk}$. Consequently, the values of the Patterson function must be stored using a small sampling grid. The expression of $R(\Omega)$ in terms of the interatomic vectors also allows to select in a very comfortable way the region U to be explored around the origin.

In the case of anisometric vector sets, it may occur that the vector set happens to coincide, for an arbitrary orientation Ω , with the origin of a neighbouring Patterson cell, thus given a wrong solution. To avoid this problem, the origin peak of the Patterson function is normally removed using the coefficients $I(\mathbf{H}) - \langle I(\mathbf{H}) \rangle$.

Normally, the general strategy followed is to perform a coarse search followed by a finer one. The coarse search consists on a χ -scan at each node of a (ϕ, ψ) lattice (Beurskens et al., 1984). In ROTSEARCH (Rius & Miravittles, 1987), 406 points with a maximum spherical angular error of 3.5° are employed. The maximum vector length is 6 \AA and $\Delta\chi=5^\circ$. The values of R are ranked according to $(R(\Omega) - \langle R \rangle) / \sigma$. After removal of the symmetry-equivalent top-ranked solutions, the surviving solutions are refined by recomputing the R -function using a finer grid (maximum vector length $\approx 9 \text{ \AA}$).

Very often, besides the value of the rotation function, the value of the correlation function for that particular orientation is also calculated. The reason for this is its higher discriminating power for small models.

In most cases, the rotation function only explores the orientation of the model. However, it can also be expressed in terms of additional variables like the torsion angles. If the two parts of the model which are connected by the torsion angle are of similar weight, one can refine the torsion angle without positioning the molecule first. Notice that if the torsion angle is varied gradually and the value of the rotation function remains unchanged, this very probably indicates a wrong solution.

The translation function.

According to Crowther & Blow (1967), Harada et al. (1981) and Rius et al. (1986) the translation function is defined by

$$T(\mathbf{r}) = \int_V P_{cross}(\mathbf{u}) \cdot P_{cross}(\mathbf{u}, \mathbf{r}) \cdot d\mathbf{u}$$

wherein \mathbf{r} is the position vector of the model to be placed. Since $T(\mathbf{r})$ is a sum function, it can be expressed as the sum of the product of the Fourier coefficients of the observed and calculated cross-Patterson functions. In view of the above definition of $T(\mathbf{r})$ and of expressions (2) and (3) it follows,

$$T(\mathbf{r}) = \sum_{\mathbf{H}} \{ I(\mathbf{H}) - \sum_j |S_j(\mathbf{H})|^2 \} \cdot \{ \sum_j \sum_{\mathbf{k} \neq \mathbf{j}} S_j(\mathbf{H})^* S_{\mathbf{k}}(\mathbf{H}) \exp(i2\pi \mathbf{H}(\mathbf{t}_{\mathbf{k}} - \mathbf{t}_{\mathbf{j}})) \times \exp(-i2\pi \mathbf{H}(\mathbf{R}_{\mathbf{j}} - \mathbf{R}_{\mathbf{k}}) \cdot \mathbf{r}) \}$$

Since the coordinates of the vector $\mathbf{H}(\mathbf{R}_{\mathbf{j}} - \mathbf{R}_{\mathbf{k}})$ are always integers, $T(\mathbf{r})$ can be calculated using the Fast Fourier Transform algorithm.

The translation function can be generalized to molecular crystals with more than one symmetry-independent molecule in the unit cell in order to take advantage of the molecular orientations found in the rotation search. If the symmetry-independent molecules are similar, a single rotation search furnishes all correct orientations. Once the first molecular fragment is positioned, another symmetry-independent oriented molecular fragment can be placed with respect to it.

If F_p denotes the known part of the structure, the observed and calculated Fourier coefficients in the translation function are

$$I(H) - |F_p(H)|^2 - \sum_j |S_j(H)|^2$$

and

$$\sum_j \sum_{k \neq j} S_j(H)^* S_k(H) \exp(i2\pi H(t_k - t_j)) \exp(-i2\pi H(R_j - R_k) \cdot r) + \sum_j F_p(H) S_j(H)^* \exp(-i2\pi H t_j) \exp(-i2\pi H R_j \cdot r)$$

Notice that the last summation causes the convolution of the inverted search fragment with the modified α function (Rius & Miravittles, 1988).

If the search model has no internal symmetry higher than the identity, the asymmetrical unit of T coincides with the unit cell of the normalizer of the space group (International Tables of Crystallography, 1987). However, as soon as one fragment has been located, the asymmetrical unit of the T function becomes the whole unit cell (except for non-primitive lattices).

Unlike rotation function searches which are safer, translation function searches are less reliable. To be sure that the correct solution is not overlooked, a rather large number of maxima of T must be checked.

The direct methods sum function

The sum function can also be used to explore other Patterson-type synthesis. In the case of direct methods the explored Patterson-type function is the modulus synthesis which is calculated with the moduli of the structure factors as coefficients. In direct methods, the most important preliminary information about the compound which is exploited is the atomicity, i.e. the knowledge that the electron density is concentrated forming maxima. Consequently, the experimental resolution of the intensity data must be high enough to produce separate maxima in the corresponding electron density Fourier synthesis. The variables which are refined in direct methods, are the set Φ of phases associated with the large structure factors. The expression of the direct methods sum function is then

$$S(\Phi) = \int_V P_o'(\mathbf{u}) \cdot P_c(\mathbf{u}, \Phi) \cdot d\mathbf{u},$$

where P_o' represents the observed modulus function with removed origin peak (Rius, 1993). In the practice, the coefficients used to compute P_o' are corrected for the form factor decay as well as for the atomic thermal vibration. The subtraction of the origin peak, i.e. the introduction of the coefficients $|F(H)| - \langle |F(H)| \rangle$ causes the weak structure factors to play an active role in the maximization of $S(\Phi)$. The set of refined phases which maximizes the direct methods sum function is normally considered as the most probably correct solution and is used to compute the final electron density map.

Nowadays, active research fields in this area are the application of the direct methods sum function to data from patterns with systematic overlap (Rius et al., 1999), and also the simultaneous solution of the crystal structure and the accidental peak overlap. This is achieved by considering the partitioning coefficients as additional variables in the sum function.

Practical examples

The rotation and translation functions are applied to powder diffraction data of molecular crystals when no single crystals can be grown. In general, the corresponding powder patterns have line broadening effects due to the particle size, and very often, the particles are anisometric e.g. needles or plates. To avoid preferred orientation effects it is better to measure in transmission mode with the sample included in a glass capillary. For in-house diffractometers and purely organic compounds, the useful information in the patterns ends in most cases at 2θ values in the range 50-60°.

The possibility of applying Patterson search methods depends on the availability of suitable search models and on the number of symmetry-independent molecules in the crystal. This is the principal reason why these methods have found limited application in the study of drug substances. However, they are fast and reliable when the molecules have rings. Three examples will illustrate their usefulness:

1. The first example describes the structure solution of a precipitate which decomposes when trying to grow larger crystals (Cirujeda et al., 1995). The crystal structure of the 2-(3,4-dihydroxyphenyl)- α -nitronyl nitroxide radical was determined to establish the hydrogen bond network and in this way to understand the competing ferromagnetic and antiferromagnetic interactions observed in this molecular solid. Notice that Patterson search methods provides valuable information about the position of the molecules and about the torsion angles, so that they can be used quite effectively to find hydrogen bonds. The structure belongs to space group $P2_1/a$ and was solved by introducing the intensities of 50 low resolution reflections into ROTSEARCH (Rius et al., 1987). The selected model molecule was derived

from the previously reported coordinates of a related radical. The best solution supplied by the program was refined with SHELX76 to a R value of 0.34. The correct position of the missing hydroxyl group (i.e. not included in the model) from the two possible positions was discriminated by performing two refinements (obtained R values were 0.28 and 0.37). The atomic arrangement with the lower R value was the correct one. This indicates that at this stage the intensities are very sensitive to the inclusion of electron density but not to the accurate positions of the atoms.

The Rietveld refinement of the structure was performed with the rigid body approach. The final estimated standard deviations of the rotational and positional parameters depend on the maximum 2θ value of the intensities introduced in the Rietveld refinement (0.6° for the Euler angles and 0.025\AA for the position of the molecule; $2\theta_{\text{max}} = 45^\circ$ for $\text{CuK}\alpha_1$ radiation).

2. The system of 2-bromonaphthalene has been selected to describe the solution of the crystal structure of a phase directly formed from a metastable one (Sañé et al. 1996). Above 332K this compound is liquid. When it cools, a metastable form called BrII ($P2_1/n$, $Z=4$) appears at 298K which transforms within months into the stable one BrIII ($P2_1/a$, $Z=2$). This transformation causes the destruction of the single crystals of BrII. The powder pattern of BrIII was measured with the sample included in a glass capillary and could be indexed unambiguously. However, all attempts to interpret the Patterson function to find the Br atoms failed. The reason was the low resolution of the extracted intensity data which caused different Br-Br Patterson peaks to appear as a single one in the map. Finally, application of the direct methods sum function as programmed in XLENS (Rius, 1999) to 77 extracted integrated intensities (20 large + 16 weak) furnished the positions of the two symmetry-independent Br atoms. These two atoms were refined with SHELX76 to an R value of 0.35. The rest of the molecules were located by rotating the fragments around the found Br positions and calculating the corresponding R values. The best solution ($R=0.138$) was the correct one as confirmed by the subsequent rigid body Rietveld refinement (mean standard deviations of Euler angles and molecular position 0.35° and 0.014\AA respectively). The crystal structure of phase BrII is formed by layers of Br-naphthalene molecules. Within each layer, the molecules are arranged forming two types of parallel rows i.e. with all the Br atoms within each row either up or down. During the transformation into phase BrIII, part of the molecules rotate to produce identical rows i.e. with the Br atoms alternatively placed up and down within each row.
3. This third selected example corresponds to the structure solution of the sublimated form of tris(3,5-dimethylpyrazol-1-yl)methane., a propeller-like molecule. The material has a 'cotton-wool' aspect. All attempts to obtain large crystals from the liquid failed since this resulted in a different phase

(Ochando et al., 1997). The phase is trigonal symmetry with cell parameters $a = 16.152 \text{ \AA}$, $c = 5.353$, space group $P3$, $C_{16}H_{22}N_6$ and $Z=3$. The powder pattern was measured with the sample in a glass capillary. The structure was solved taking a related compound as search model and the final Rietveld refinement converged to $R_{wp}=0.077$. This study is one of the first examples of Patterson search structure determinations from an hemihedral space group using powder data. The complexity of the structural determination is increased by the presence of three molecules in the asymmetric unit. This study showed that two of the molecules are similar but that the third one has the rings rotated in the opposite direction.

The support of the Direcció Gral. de Enseñanza Superior (DGES) of the Spanish Government (project PB98-0483) and of the Direcció Gral. de la Recerca of the Catalan Government (Grant SGR-00460, 1995) are gratefully acknowledged.

References

- BEURSKENS,P.T., BEURSKENS,G., STRUMPEL,M. & NORDMAN,C. (1984). Proceedings of the Patterson Symposium, Philadelphia, edited by J.P.Glusker.
- CIRUJEDA,J., OCHANDO,L.E., AMIGÓ,J.M., ROVIRA,C., RIUS,J. & VECIANA,J. (1995). *Angew.Chem.Int.Ed.Engl.* **34**, 55-57.
- CROWTHER, R.A. & BLOW,D.M. (1967) *Acta Cryst.* **23**, 544-548.
- HARADA,Y., LIFCHITZ,A., BERTHOU,J. & JOLLES,P. (1981) *Acta Cryst.* **A37**, 398-406.
- OCHANDO,L.E., RIUS,J., LOÜER,D., CLARAMUNT,R.M., LOPEZ,C., ELGUERO,J. & AMIGÓ,J.M. (1997) *Acta Cryst.* **B53**, 939-944.
- RIUS,J. & MIRAVITLLES,C. (1986) *Acta Cryst.* **A42**, 402-404.
- RIUS,J. & MIRAVITLLES,C. (1987) *J.Appl.Cryst.* **20**, 261-264.
- RIUS,J. & MIRAVITLLES,C. (1988) *Acta Cryst.* **A44**, 79-80.
- RIUS,J. & MIRAVITLLES,C. (1988) *J.Appl.Cryst.* **21**, 224-227.
- RIUS,J. (1993) *Acta Cryst.* **A49**, 406-409.
- RIUS,J. (1999) *Powder Diffraction* **14** (4) 267-273.
- RIUS,J., MIRAVITLLES,C., GIES,H. & AMIGÓ,J.M. (1999) *J.Appl.Cryst.* **32**, 89-97.
- ROSSMANN,M.G. & BLOW,D.M. (1962) *Acta Cryst.* **15**, 24-31.
- SAÑÉ,J., RIUS,J., LOÜER,D. & CHANH,N.B. (1996) *J.Phys.Chem.Solids* **57**, 1251-1257.

Monte Carlo Methods

Maryjane Tremayne & Colin Seaton

*School of Chemistry
University of Birmingham,
Edgbaston, Birmingham. B15 2TT. UK.
E-mail: m.tremayne@bham.ac.uk*

Introduction

The term ‘Monte Carlo’ refers to a group of methods in which physical or mathematical problems are simulated by using random numbers. Unlike deterministic algorithms where the use of the same input twice will generate two identical execution patterns and results, the Monte Carlo approach, as a type of probabilistic algorithm, includes steps that depend not only on the input but also on the results of some random events.

Monte Carlo methods have been employed in numerous fields of science [1-6]. In molecular modelling and structural chemistry, the Monte Carlo algorithm is used to generate configurations of a system by making random changes to the positions, orientations and conformations of the species under consideration. In this way, a collection of configurations is accumulated, and it is essential to formulate a procedure that determines whether each new configuration will be accepted within the group. The most commonly used strategy is the Metropolis importance sampling approach [7]. The crucial feature of this Metropolis scheme is that it biases the generation of configurations towards those that make the most significant contribution to a solution. The Metropolis Monte Carlo method has become so widely adopted in these fields that it is usually referred to simply as ‘the Monte Carlo method’.

Over the last few years, the Monte Carlo technique has been developed and applied to the determination of crystal structures from powder diffraction data [8,9]. This approach differs considerably from traditional methods of structure solution, in that it operates in direct space (rather than reciprocal or Patterson space). The direct-space strategy involves generation of trial crystal structures independently of the experimental powder diffraction data, with the suitability of each trial structure assessed by direct comparison between the powder diffraction pattern calculated for the trial structure and the experimental powder diffraction pattern. This comparison is quantified using an appropriate figure-of-merit such as R_{wp} or chi-squared, and the best structure solution used as the starting model for subsequent structure refinement.

Monte Carlo Methodology

In the Monte Carlo approach to structure solution from powder diffraction data, a series of structures (denoted $x(j)$ for $j=1, \dots, N$) are generated for consideration as potential structure solutions. Each structure generated during the calculation is not produced “from scratch” but derived from the previous structure forming a Markov chain. Hence the outcome of each Monte Carlo move (in which structure $x(j+1)$ is generated from structure $x(j)$) depends only upon the preceding structure and not on previous structure trials.

The first structure $x(1)$ is generally chosen as a random position of the structural model in the unit cell. The procedure for each Monte Carlo move then comprises the following steps:

- i) Starting from structure $x(j)$, the structural model is displaced by a random amount to generate a trial structure $x(\text{trial})$. The exact form of this displacement depends on the system of interest, and may be constrained depending on symmetry and other structural considerations. In general this will involve one or more of the following: (a) translation in a random direction by variation of $[x,y,z]$; (b) rotation around orthogonal axes by variation of $[\theta,\phi,\psi]$; (c) variation of selected internal degrees of freedom, e.g. torsion angles ($\tau_1, \tau_2, \dots, \tau_N$) to introduce conformational flexibility.
- ii) The agreement between the powder diffraction pattern calculated for the trial structure and the experimental powder diffraction pattern is assessed, using a criterion such as the profile R-factor $R_{\text{wp}}(x(\text{trial}))$.
- iii) The trial structure is then accepted or rejected by considering the difference Z between R_{wp} for the structure $x(\text{trial})$ and R_{wp} for the structure $x(j)$ [i.e. $Z=R_{\text{wp}}(x(\text{trial}))-R_{\text{wp}}(x(j))$] and invoking the Metropolis importance sampling algorithm. If $Z \leq 0$, the trial structure is accepted automatically as the new structure [i.e. $x(j+1) = x(\text{trial})$]. If $Z > 0$, the trial structure is accepted only with probability $\exp(-Z/S)$ and rejected with probability $[1-\exp(-Z/S)]$, where S is an appropriate scaling factor of Z . If the trial structure is rejected, structure $x(j+1)$ is taken to be the same as $x(j)$.

These steps are summarized in figure 1. The above procedure is repeated to generate a large number of structures, with each structure derived from the previous one through small random displacements. After a sufficiently wide range of parameter space has been explored, the best structure solution (corresponding to lowest R_{wp}) is used as the starting model for structure refinement calculations.

Implementation

The direct-space strategy of structure solution aims to find the crystal structure that has

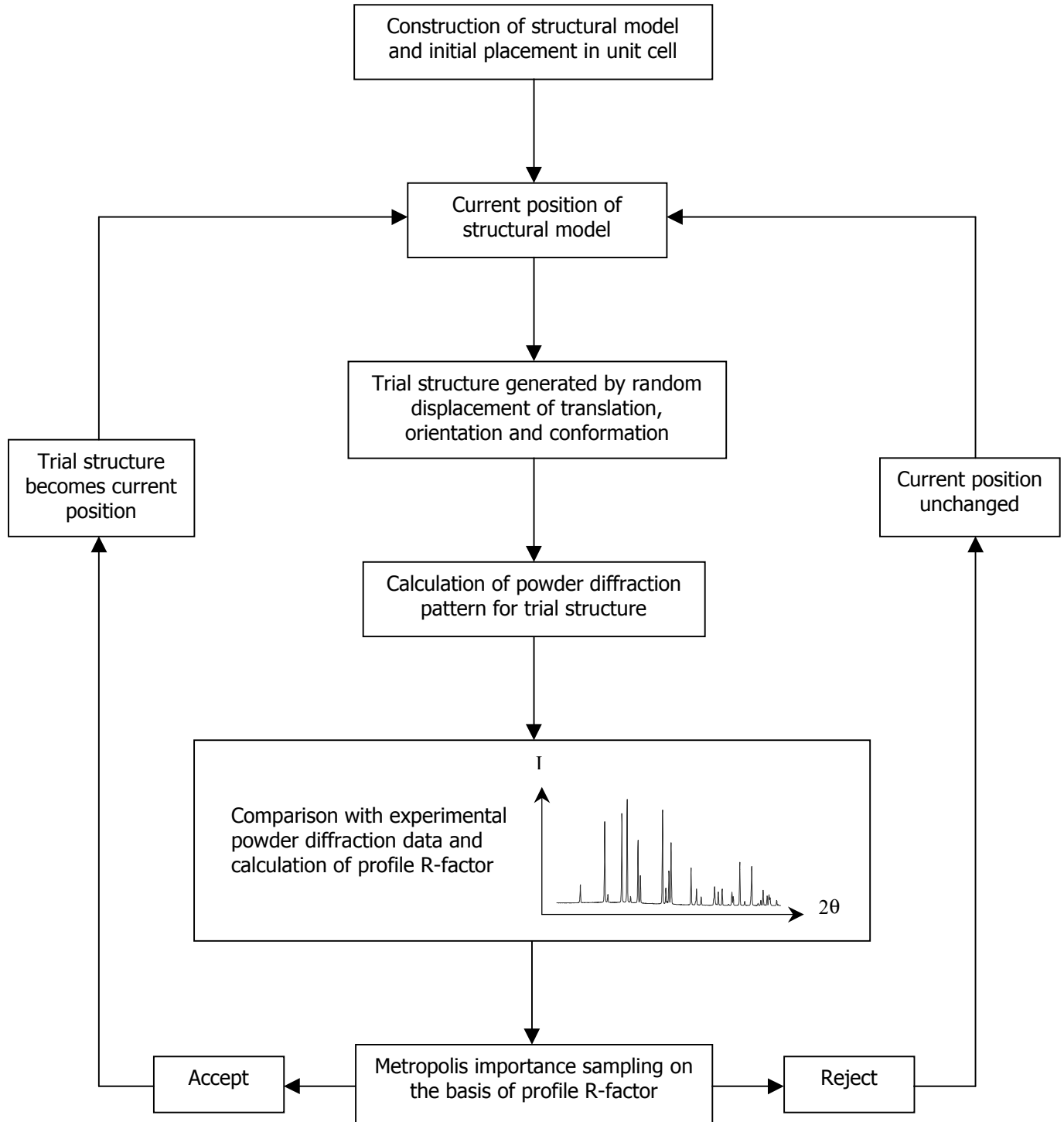


Figure 1: A summary of the steps involved in a Monte Carlo move.

the lowest possible R-factor in a way that is equivalent to exploring an R-factor hypersurface to find the global minimum. The form of a hypersurface is defined by the set of variables $[x, y, z, \theta, \phi, \psi, \tau_1, \tau_2, \dots, \tau_N]$ that are used to represent the position of the structure in the unit cell. A simple representation of an R-factor hypersurface is given in Figure 2. This is a two-dimensional section through the R_{wp} hypersurface for lithium zirconate. The structural model used to calculate this surface contained only the zirconium atom (the dominant X-ray scatterer) so that the structure is defined by three variables, the position $[x, y, z]$ of the zirconium atom. The section shown, $R_{wp}(y, z)$, was obtained by fixing the x-coordinate of the zirconium atom at the value in the known crystal structure. The deep minimum in R_{wp} corresponds to the correct structure solution, but the plot clearly shows the existence of other significant local minima.

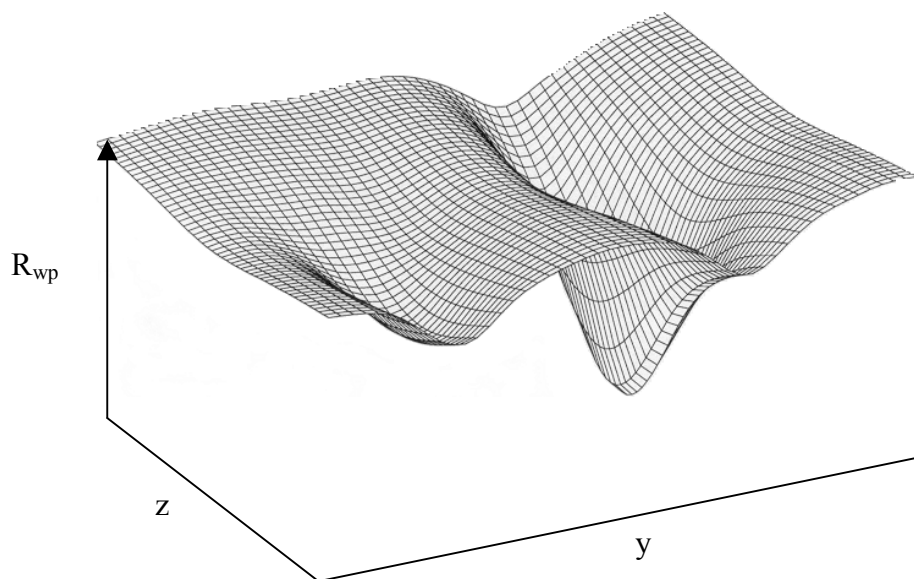


Figure 2: A section of the R_{wp} surface for lithium zirconate.

The Monte Carlo algorithm acts on a single point – the current structure – on the hypersurface. In each Monte Carlo move, a new point is selected from the neighbourhood of the current point. Local search methods such as hill-climbing and descent techniques use an iterative improvement technique on this point. However, these methods only provide locally optimum R_{wp} values that depend on the selection of the starting point. By using the Metropolis sampling algorithm in the Monte Carlo technique, we ensure that our point explores the hypersurface giving emphasis to regions with low R_{wp} , but with the ability to escape from local minima.

The efficiency of the Monte Carlo approach can be maximized either by modification of the hypersurface being searched (by appropriate choice of the structural model and range of data) or by optimization of the parameters that

control the algorithm itself. To achieve this we need a greater understanding of how the Monte Carlo algorithm is implemented.

As described earlier, a new structure is accepted if its R_{wp} is lower than the previous structure i.e. a downhill move into the nearest minima. If the R_{wp} is higher, the new structure is accepted with a probability $\exp(-Z/S)$ so that the smaller the uphill move, the greater the probability that the move will be accepted. This is achieved by comparing the factor $\exp(-Z/S)$ to a random number between 0 and 1. The random number generator at the heart of every Monte Carlo simulation program is accessed not only to generate new configurations but also to decide whether a given move should be accepted or not. If $\exp(-Z/S)$ is greater than the random number then the new structure is accepted. If it is smaller then the new structure is rejected. Thus if the R_{wp} of the trial structure is very close to that of the current structure, then $\exp(-Z/S)$ will be very close to 1, and so is more likely to be accepted. If Z is very large however, then $\exp(-Z/S)$ will be close to 0 and the move is unlikely to be accepted.

To ensure maximum efficiency in the propagation of the Monte Carlo algorithm, two adjustable parameters are used so that the optimum proportion (approximately 40%[10]) of trial moves is accepted. These are:

- i) The scale factor S , which operates in a manner analogous to temperature in applications of Monte Carlo techniques in energy simulations, and may be fixed or varied in a controlled manner during the calculation. Clearly the higher the value of S , the greater the probability that trial structures with $Z > 0$ will be accepted. The fundamental difference between Monte Carlo and simulated annealing techniques concerns the way in which the parameter S is used to control the sampling algorithm. Unlike the Monte Carlo method in which S is fixed or varied manually, in simulated annealing, S is decreased systematically according to an annealing schedule or temperature reduction procedure. These approaches only represent minimization of R_{wp} if $S=0$.
- ii) The maximum displacement d_{max} , which governs the size of the structural displacement in each Monte Carlo move. Each trial structure is generated by displacement of one or more parameters of the set $[x, y, z, \theta, \phi, \psi, \tau_1, \tau_2, \dots, \tau_N]$ by an amount ξd_{max} where ξ is a random number in the range -1 to 1 . This allows moves in both positive and negative directions. If d_{max} is too small then many moves will be accepted but the structures will be very similar and the hypersurface will be explored very slowly. Too large a displacement and the trial structures are effectively being chosen randomly from a large area of the hypersurface, resulting in a highly inefficient search in which nearby minima can be completely missed.

Both these parameters can be adjusted while the program is running so that the desired acceptance ratio is achieved.

A typical plot of R_{wp} for trial structures generated in a Monte Carlo structure solution calculation versus the corresponding Monte Carlo move number is shown in Figure 3. This clearly illustrates that after the location of each minimum (move numbers 2300 and 3974), the Monte Carlo algorithm does not converge but continues to search more of the parameter space.

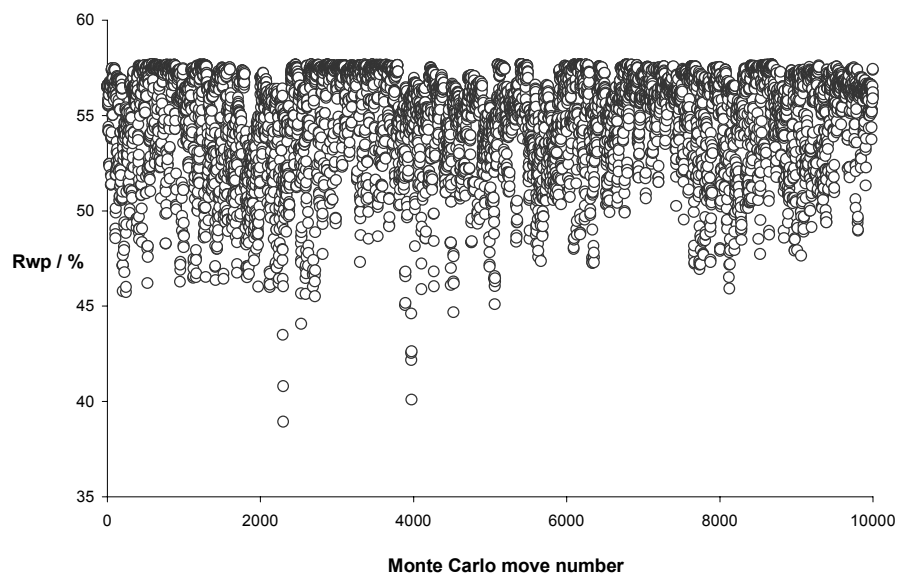


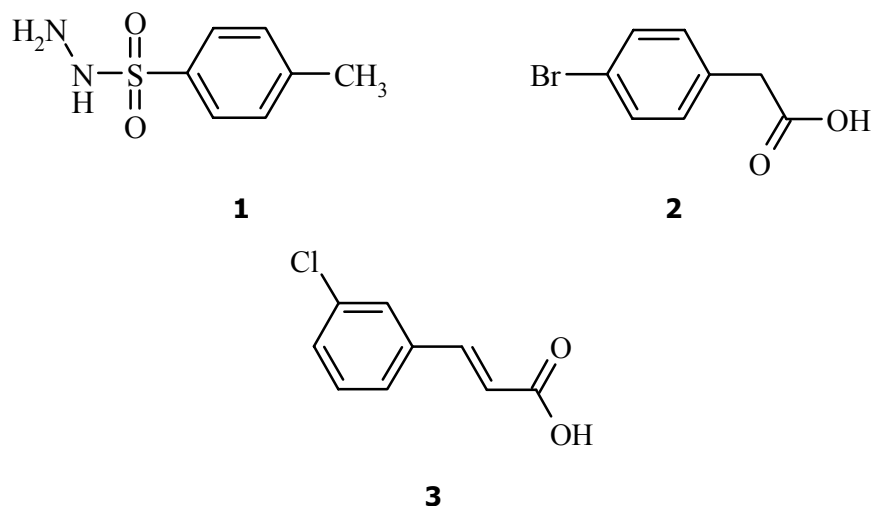
Figure 3: R_{wp} vs move number for a typical Monte Carlo calculation

Applications of the Monte Carlo method

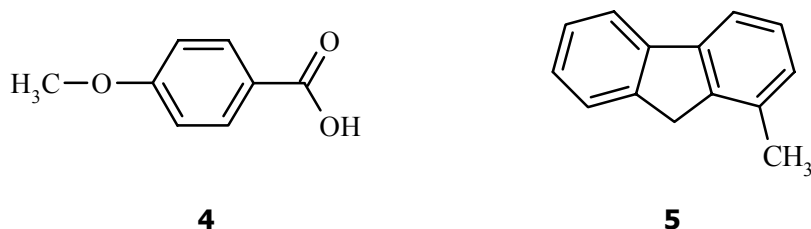
The Monte Carlo method has been used to solve the structures of molecular materials taken from a range of fields throughout supramolecular chemistry and materials science. Both single and multi-component systems described by structural models with varying degrees of intramolecular flexibility have been successfully solved from both laboratory and synchrotron X-ray and neutron powder diffraction data.

The first applications of the method involved only the simplest case of translation of the structural fragment through the unit cell (valid for a single atom problem) and rotation of the structural fragment around a fixed point in space (with no translation) [8]. This approach involving two separate Monte Carlo calculations was used to solve systems such as *p*-toluenesulfonylhydrazine (**1**) and *p*-bromophenylacetic acid (**2**) that contain a dominant scatterer (S and Br respectively). Initial location of the dominant scatterer in the unit cell enabled subsequent rotation of the remainder of the molecule around the fixed position of the dominant scatterer found in the first calculation.

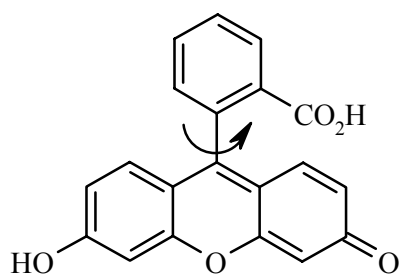
A similar approach was used to solve the γ phase of 3-chloro-*trans*-cinnamic acid (**3**) in which the carboxylic group was maintained at a fixed distance from the crystallographic centre of symmetry, with the centre of symmetry lying in the molecular plane, and the molecule then rotated about this point [11].



The Monte Carlo method was then generalized to consider the simultaneous translation and rotation of structural models. Initially, studies were confined to systems that could be described as rigid bodies, such as *p*-methoxybenzoic acid (**4**) [12] and 1-methylfluorene (**5**) [13].

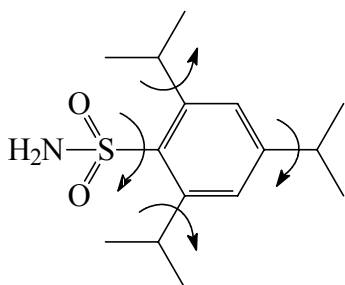


Further developments extended the technique by the introduction of intramolecular degrees of freedom enabling the study of structures that require conformational flexibility in the structure solution calculation. The structure determination of the red phase of fluorescein (**6**) (which finds applications as a yellow dye with intense green fluorescence) not only illustrated this advance in methodology but represents the resolution of a long-standing problem [14]. A curved arrow indicates the internal rotation allowing variation of the angle between the planar rigid benzoic acid and hydroxyxanthenone groups.

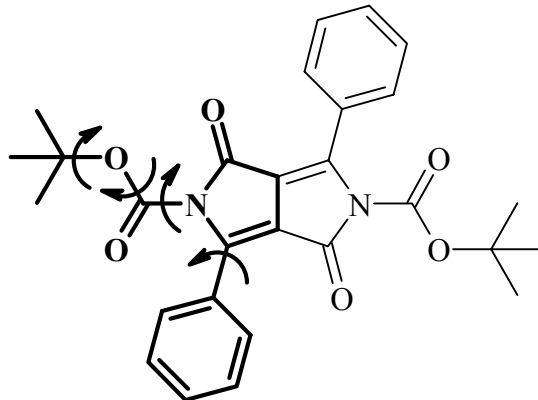


6

Structures requiring a greater degree of conformational flexibility have also been solved using the Monte Carlo method through the variation of multiple torsion angles. The structure determination of 2,4,6-tri-isopropylbenzene sulfonamide (**7**) [15] has added to a study of several related sulfonamino compounds also by powder diffraction techniques. Earlier attempts to solve this structure by traditional techniques were unsuccessful. Another example concerns the structure determination of 1,4-diketo-2,5-di-*t*-butoxycarbonyl-3,6-diphenylpyrrolo[3,4-*c*] pyrrole (DPP-Boc (**8**)), an important derivative of the commercial red pigment DPP. Routine powder diffraction studies of this system revealed the existence of a new polymorph and the Monte Carlo method was used in the subsequent structure determination [16]. The structural models used in the structure solution of these systems are shown below.



7



8

In both cases, the required amount of intramolecular flexibility was achieved through the variation of four torsion angles, although the movement of the molecule itself differs between these systems.

In the first instance, the complete sulfonamide molecule (excluding the hydrogen atoms of the methyl groups) was used in the Monte Carlo calculation. Internal rotation of the isopropyl and sulfonamino groups relative to the aryl ring together with translation and rotation of the molecule within the unit cell resulted in a structure solution calculation defined by 10 variables $[x, y, z, \theta, \phi, \psi, \tau_1, \tau_2, \tau_3, \tau_4]$.

The structural fragment used in the solution of DPP-Boc comprised half the molecule (shown in bold) excluding all hydrogen atoms, with the molecular inversion centre constrained to reside on a crystallographic inversion centre (space group P-1). The Monte Carlo calculation thus involved free rotation of the whole molecule around this pivot point and variation of the torsion angles within the *t*-butoxycarbonyl group.

In the application of direct-space methods, the presence of more than one component in the asymmetric unit increases the complexity of the problem, both in terms of the number of degrees of freedom needed for structure solution and to an extent the effect on R-factor discrimination. In the structure solution of sodium chloroacetate, both the sodium cation and the chloroacetate anion were included in the structural model and were handled as independent structural units [17]. From knowledge of this crystal structure, we were able to rationalize the solid state polymerization reaction of sodium chloroacetate to produce the biodegradable non-toxic polymer polyglycolide.

Although salts and hydrates are multi-component structures, one of the units is often described as a single atom and only requires translation. Systems containing more than one crystallographically distinct molecule in the asymmetric unit require translation, rotation and conformational flexibility if applicable. The situation is made more complex still if the structure comprises entirely different, completely independent molecular components, as the correct location of each molecule in the unit cell is unique. An example of this is the structure solution of an organic cocrystal, 1,2,3-trihydroxybenzene–hexamethylenetetramine (1/1) using the Monte Carlo method [18]. Although the two molecules were treated as rigid units in the structure solution, the number of degrees of freedom required for random movement is increased to 12 in this case [$\{x_1, y_1, z_1, \theta_1, \phi_1, \psi_1\}$ and $\{x_2, y_2, z_2, \theta_2, \phi_2, \psi_2\}$]. Trial crystal structures were generated by completely independent translation and rotation of the two molecular components within the unit cell. Subsequent rationalization of the intermolecular forces in this system revealed a three-dimensional network of both weak and strong hydrogen bonds. Figure 4 shows one of the puckered molecular ribbons in this structure, formed by alternating 1,2,3-trihydroxybenzene and hexamethylenetetramine (HMTA) molecules in which HMTA acts as an acceptor of three hydrogen bonds. This behaviour differs from the majority of related systems and hence the exact hydrogen-bonded relationship between the two components could not be predicted in advance.

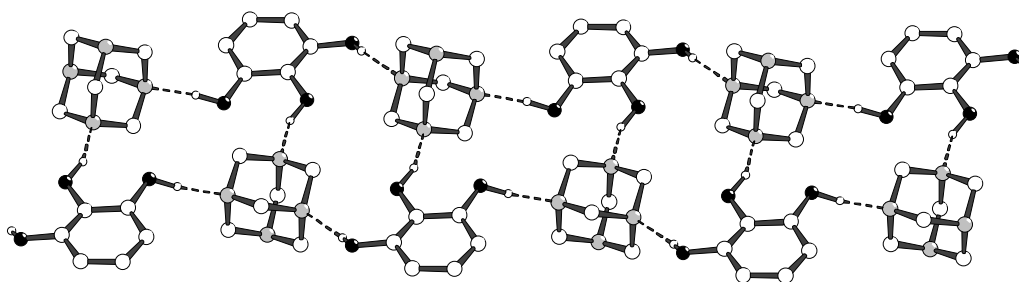
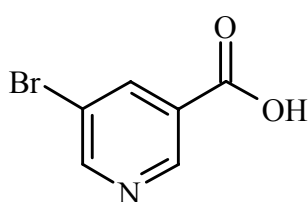


Figure 4: Crystal structure of 1,2,3-trihydroxybenzene–hexamethylenetetramine (1/1) with the intermolecular O-H...N hydrogen bonds indicated by dashed lines.



9

A further Monte Carlo study has shown that direct-space structure solution techniques may be more robust than traditional techniques when the diffraction data are distorted by preferred orientation. The use of a complete structural model in the Monte Carlo method has enabled the structure determination of 5-bromonicotinic acid (**9**) from experimental powder diffraction data significantly affected by preferred orientation [19].

Data was initially collected on a conventional laboratory diffractometer with the sample packed in a flat disc. This data was used for structure solution, but after significant distortion of the structure at the refinement stage, a second data set was collected in capillary mode. Comparison of the two powder patterns shows the high level of preferred orientation present (Figure 4). Once the direction of preferred orientation had been identified, corrections were made in refinement by variation of a preferred orientation parameter.

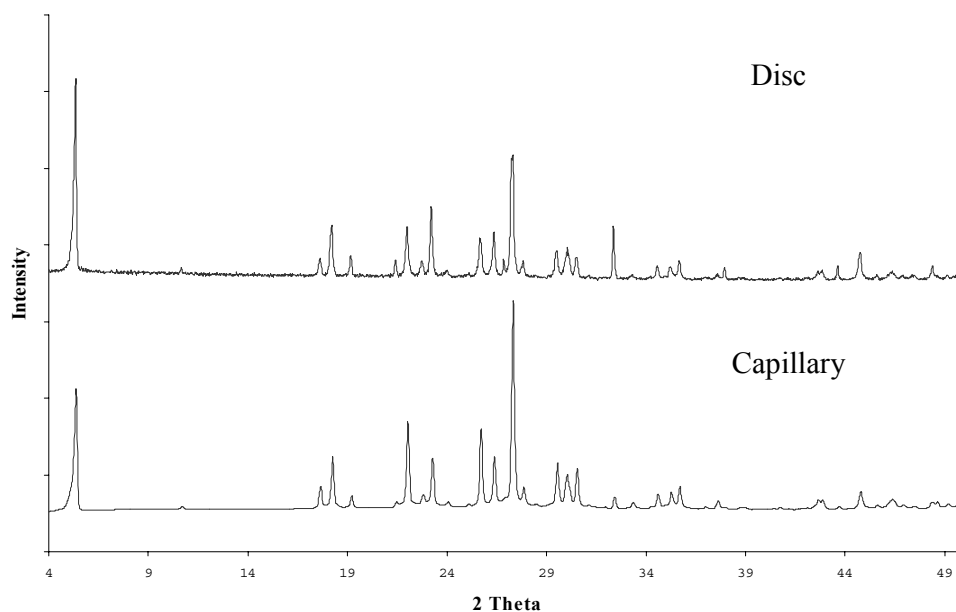


Figure 5: Powder diffraction patterns for 5-bromonicotinic acid collected in disc and capillary geometries.

It is clear from the refinement that the preferred orientation had a much greater effect on the disc data, but despite this, the Monte Carlo method successfully located the correct structure solution. The resulting structure contains molecules linked across an inversion centre forming the common dimeric carboxylic acid hydrogen bonded motif, an arrangement that is unusual compared to other nicotinic acid derivatives, but similar to that of 5-bromobenzoic acid.

Concluding Remarks

The structure determination of molecular solids from powder diffraction data is a rapidly expanding field, mainly due to the development of direct-space structure solution methods such as the Monte Carlo technique. The Monte Carlo algorithm provides an efficient method of global optimization that has been used to study a range of systems from small planar molecules to those that require conformational flexibility or contain multiple components. As the method is developed further and applied to problems of increasing complexity it will surely continue to make an invaluable contribution to the structural study and understanding of important solid state materials.

MT would like to thank the Royal Society for funding this work through the award of a University Research Fellowship, and the University of Birmingham and SmithKline Beecham Pharmaceuticals for a studentship for CS.

References

- [1] B. H. Kaye, 'A random walk through fractal dimensions', **1989**, VCH, New York
- [2] M. P. Allen & D. J. Tildesley, 'Computer Simulation of Liquids', **1987**, OUP, Oxford
- [3] P. H. Leslie & D. Chitty, *Bimetrika*, **1951**, 38, 269
- [4] G. H. Weiss, *Am. Sci.*, **1983**, 71, 65
- [5] A. R. Leach, 'Molecular Modelling, Principles and Applications', **1996**, Addison Wesley Longman Ltd.
- [6] R. L. McGreevy & L. Pusztai, *Mol. Simul.*, **1988**, 1, 359
- [7] N. Metropolis, A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller & E. Teller, *J. Chem. Phys.* **1953**, 21, 1087
- [8] K. D. M. Harris, M. Tremayne, P. Lightfoot & P. G. Bruce, *J. Am. Chem. Soc.* **1994**, 116, 3543
- [9] K. D. M. Harris & M. Tremayne, *Chem. Mat.*, **1996**, 8, 2554
- [10] M. Rao, C. Pangali & B Berne, *J. Mol. Phys.* **1979**, 37, 1773.
- [11] B. M. Kariuki, D. M. S. Zin, M. Tremayne & K. D. M. Harris, *Chem. Mat.* **1996**, 8, 565
- [12] M. Tremayne, B. M. Kariuki & K. D. M. Harris, *J. Appl. Cryst.* **1996**, 29, 211
- [13] M. Tremayne, B. M. Kariuki & K. D. M. Harris, *J. Mat. Chem.* **1996**, 6, 1601
- [14] M. Tremayne, B. M. Kariuki & K. D. M. Harris, *Angew. Chem. Int. Ed. Engl.* **1997**, 36, 770
- [15] M. Tremayne, E. J. MacLean, C. C. Tang & C. Glidewell, *Acta Cryst.* **1999**, B55, 1068
- [16] E. J. MacLean, M. Tremayne, B. M. Kariuki, K. D. M. Harris, A. F. M. Iqbal & Z. Hao, *J. Chem. Soc., Perkin Trans. 2*, **2000**, 1513
- [17] L. Elizabe, B. M. Kariuki, K. D. M. Harris, M. Tremayne, M. Epple & J. M. Thomas, *J. Phys. Chem. B*, **1997**, 101, 8827
- [18] M. Tremayne & C. Glidewell, submitted for publication
- [19] C. B. Aakeroy, A. M. Beatty, D. M. Rowe & M. Tremayne, in preparation

Crystal Structure Determination of Tetracycline Hydrochloride with Powder Solve

M. A. Neumann, F. J. J. Leusen, G. Engel, C. Conesa-Moratilla, S. Wilke

*MSI, 230/250 The Quorum,
Barnwell Road,
Cambridge CB5 8RE,
England*

Abstract

The crystal structure of tetracycline hydrochloride is solved and refined with the software package Powder Solve. Prior to structure solution, a conformational analysis of the carbon ring system of the molecular ion is carried out using molecular dynamics and semi-empirical calculations. Molecular mechanics is used to assist the final structure refinement.

Introduction

Powder Solve [1] is a software package for crystal structure determination from powder diffraction data. In this tutorial, Powder Solve is used to solve the crystal structure of tetracycline hydrochloride, a compound that recently served in a blind test to assess the capability of various methods for structure solution from powder diffraction data [2]. The original structure determination with Powder Solve was sent in as a post-deadline submission. Powder Solve is fully integrated in MSI's Cerius² modeling environment, which gives access to a wide range of modeling techniques, including force-field based molecular dynamics and mechanics, semi-empirical methods and *ab initio* calculations. Reliable potential energy calculations significantly enhance the potential of modern methods for structure solution from powder data. In the present case, potential energy calculations are used to determine the conformation of the molecular ion prior to structure solution and to improve the quality of the crystal structure in the final structure refinement.

Structure determination with Powder Solve is basically of four-step process, including indexing, profile fitting, structure solution, and structure refinement. In the first step, the crystal class and approximate lattice parameters are derived from the peak positions in the powder diffraction pattern, using either TREOR90 [3] or DICVOL91 [4]. Then, accurate lattice constants and various parameters related to the experimental setup and the texture of the sample are determined by profile fitting with the program POWDER FIT [1]. The refined parameters include the zero-point shift of the diffractogram, background parameters and profile parameters. All peak intensities are treated as independent parameters at this stage and are not related to the unit cell content. Performing profile fitting for a variety of relevant space groups, it is possible to eliminate those space groups that are incompatible with the experimental powder diffraction pattern due to systematic extinctions.

In the third step, the atomic arrangement in the asymmetric unit is determined with the program POWDER SOLVE [1]. Powder diffraction patterns are calculated for a large number of trial structures and compared to the experimental powder pattern. The trial structures are generated using a Monte Carlo / simulated annealing algorithm. This approach is designed to locate the global minimum of the weighted Rietveld parameter R_{wp} , which is a measure for the similarity of the experimental powder data and the simulated powder patterns. The lattice symmetry is explicitly taken into account. Each molecular fragment in the asymmetric unit cell is build up of one or more rigid bodies. Rigid bodies that belong to the same fragment can rotate with respect to each other around the bond between them. The conformation of the rigid bodies has to be determined prior to structure solution by potential energy calculations or by using structural information from related crystal structures. In the Monte Carlo search, only the translational, rotational and torsional degrees of freedom of the molecular fragments are considered. Structure solution has to be attempted separately for all relevant space groups. The final structure refinement can be performed using POWDER SOLVE as a rigid body Rietveld refinement tool.

This tutorial is divided into several consecutive parts. Files containing the mayor results of each part are provided with this tutorial, so that it is possible to leave out parts that the reader may be less interested in. It is assumed that the subdirectory 'TCHC' containing all files is located in the 'tutorial' directory where Cerius² is running. The powder diffraction data supplied with this tutorial has been obtained from the original data by subtraction of the non-linear background.

The tutorial consists of the following steps:

Part 1: Indexing, profile fitting and space group selection

- A. Indexing
- B. Profile fitting
- C. Selecting space groups for structure solution

Part 2: Conformer analysis

- A. Sketching the tetracycline ion
- B. Optimizing the tetracycline ion with the COMPASS force field
- C. Searching low-energy conformers by Dynamics/Simulated Annealing
- D. Optimizing low energy conformations with MOPAC

Part 3: Structure solution

- A. Setting up a POWDER SOLVE run
- B. Analyzing the results of a POWDER SOLVE run

Part 4: Structure refinement

Part 1: Indexing, profile fitting and space group selection

A. Indexing

1. Loading the powder pattern

If Cerius² is not already running, open a UNIX window, go to the directory 'tutorial' and type in **cerius2** followed by <ENTER>.

To load the powder diffraction pattern, find the **POWDER INDEXING** card on the **POWDER SOLVE** stack and click on **Experimental Data** to bring up the **1-D Experimental Data** control panel. Set the data file format to **XY-GRAPH**.

The powder diffraction data can be found under **tutorial/TCHC/tetracycline_hydrochloride_xrd.grf**. Load the data and close the **1-D Experimental Data** control panel.

2. Generating a peak list and setting the wavelength

We will use the Automatic Peak Search tool to generate the peak list.

Click on **Peak Positions** on the **POWDER INDEXING** card to bring up the **Peaks** panel. In the **Peaks** panel, click on **Preferences** to open the **Automatic Peak Search** panel. Choose **SAVITZKY-GOLAY** as the **peak detection method**. Change the **low amplitude cutoff** to 5 %. Click on **RUN** to start the peak search.

Now we have to make sure that all peaks are correctly placed and that no peaks have been missed in the low angle part of the diffraction pattern. Therefore, we have to magnify the diffraction pattern until we see the statistical noise. Then scan over the diffraction pattern horizontally, looking at only a small number of diffraction peaks at a time. In this particular case, no modification of the automatically generated peak list should be necessary.

To magnify the diffraction pattern, move the **mouse horizontally and vertically** while pressing **SHIFT** and the **middle mouse button**. To scan over the diffraction pattern, **move the mouse** while pressing the **middle mouse button** only.

Finally, we have to set the wavelength. The experimental data has been measured at a synchrotron radiation source at a wavelength of 0.692 Å.

In the **Peaks** panel, click in the wavelength input field and change the wavelength to 0.692 Å.

3. Indexing the powder pattern

We will first try to index the powder pattern with TREOR90.

Click on **Run** on the **Powder Indexing** card to bring up the **Run Indexing** panel. In the **Run Indexing** panel, click on the **Run TREOR90** button to start TREOR90. By default, all seven crystal systems are searched to find the unit cell. TREOR90 will not be able to find the unit cell, and a window with the message 'Sorry - cell parameters were not found' will open. Click in that window with the left mouse button to close it.

Now we will try to index the powder pattern with DICVOL91. By default, the search is only carried out for crystal systems of high symmetry, and we first have to change the preferences to look for orthorhombic, monoclinic and triclinic space groups as well.

In the **Run Indexing** panel, click on **Preferences** next to DICVOL91. In the **DICVOL91 Preferences** panel, enable the search for orthorhombic, monoclinic and triclinic cells. Close the **DICVOL91 Preferences** panel and click on the **RUN DICVOL91** button to start DICVOL91. The calculation may take several minutes.

We will now examine the DICVOL91 results and create a model with an empty unit cell that corresponds to the best solution found.

Click on the **Analysis** button that is related to DICVOL91 to bring up the **DICVOL91 Analysis** panel. This panel shows a list of solutions ranked according to their agreement with the peak positions in the peak list. The best solution is highlighted, and the corresponding cell parameters are shown. Click on **Create empty unit cell from solution** and close the **DICVOL91 Analysis** panel.

B. Profile fitting

To make sure that the result of the indexing procedure is in agreement with the experimental powder pattern and to prepare the structure solution step, we will now run through the profile fitting procedure.

Select the **POWDER FIT** card from the **POWDER SOLVE** stack and click on **Run** to open the **Powder Fit** panel.

Before we start refining the profile, we have to set the wavelength again.

In the **Powder Fit** panel, click on **Radiation** to open the **Radiation** panel. Change the radiation source from **COPPER** to **SYNCHROTRON** and set wavelength (1) to 0.692 Å. Close the **Radiation** panel.

In the profile fitting procedure, we will determine background parameters, profile parameters, peak intensities, cell parameters and the zero point shift of the diffractogram. A modified Pawley algorithm is used to provide a high degree of stability, and it is usually possible to refine all parameters simultaneously.

1) In the **Powder Fit** panel, click on **Range** to bring up the **Profile Range** panel. Set the upper limit of the profile range to 20°. Cutting of the high angle part of the powder pattern helps to reduce calculation times. Close the **Profile Range** panel.

2) In the **Powder Fit** panel, click on **Profiles** to bring up the **Peak Profile Refinement** panel. Change the constant peak width w from 0.02° to 0.002°. The default value of 0.02° is too large for powder data measured with synchrotron radiation. Switch on the refinement of all profile parameters. Close the **Peak Profile** panel.

3) In the **Powder Fit** panel, click on **Cell** to bring up the **Cell Parameter Refinement** panel. Switch on the refinement of the **cell parameters a, b and c**. Close the **Cell Parameter Refinement** panel.

4) In the **Powder Fit** panel, click on **Background** to open the **Background Refinement** panel. Switch on the refinement of the **zero parameter**. The refinement of the background parameters is switched on by default

5) The refinement of all peak intensities is switched on by default. Click on **RUN** in the **Powder Fit** panel to refine all parameters simultaneously. Continue refining all parameters simultaneously by clicking on the **RUN** button until the R_{wp} parameter does not improve any further. The R_{wp} parameter is shown in the graphs window and in the text window.

Now we have to compare the simulated and experimental powder patterns under high magnification in order to check if all experimental peaks are well described by the simulated profile.

Magnify the powder pattern and scan over it as described in section A of this tutorial.

There should be no significant mismatch between the simulated and calculated powder pattern. We have successfully determined the unit cell of TCHC. When this tutorial was written, an R_{wp} parameter of 6.35% was obtained at this point. The unit cell is orthorhombic, with $a=15.7325 \text{ \AA}$, $b=12.8523 \text{ \AA}$ and $c=10.9799 \text{ \AA}$.

C. Selecting space groups for structure solution

Certain space groups appear more frequently than others. According to the Cambridge Crystallographic Database, about 76 % of all organic and organometallic compounds crystallize in only 5 space groups and about 90% of all organic and organometallic crystal structures are covered by the 17 most frequent space groups [5]. In general, many space groups can be discarded based on chirality, density considerations or systematic absences. It is common practice to start with the most frequent possible space group in the structure solution step. If the crystal structure can not be solved in this space group, one has to work down the list of possible space groups in the order of decreasing probability.

The most frequent orthorhombic space group is $P 2_1 2_1 2_1$. A convenient way to investigate the effect of systematic absences is to repeat the profile fitting procedure in this space group.

Select the **CRYSTAL BUILDER** card from the **BUILDERS 1** stack, click on **Symmetry** and choose **Space Groups** from the pull down menu. Enter 'P 21 21 21' in the Space Group input field of the **Space Groups** panel.

Click on **RUN** in the **Powder Fit** panel.

A R_{wp} factor of 6.38% should be obtained at this point, compared to 6.36% in space group $P 1$. Symmetry allowed reflections are indicated by green tick marks in the graphs window. In $P 1$, three reflections of very small or no intensity occur below 4° , while they are symmetry-forbidden in $P 2_1 2_1 2_1$. The difference of the experimental and the simulated powder pattern shown in the graphs window does not reveal any missing peaks of significant intensity in $P 2_1 2_1 2_1$. It can thus be concluded that the space group $P 2_1 2_1 2_1$ is in agreement with the experimental powder diffraction pattern. In part 3 of this tutorial, we will see that the crystal structure of TCHC can indeed be solved in $P 2_1 2_1 2_1$.

Finally, we save the result of the first part of this tutorial for later use.

In the **Visualizer** control panel, change the model name from 'index' to 'PF_P212121'

Select **File/Save Model** from the **Visualizer** menu bar to open the **Save Model** panel. Enter 'PF_P212121' in the input field situated below the file browser and press **SAVE** to save the unit cell and the outcome of the profile fitting procedure.

Part 2: Conformer Analysis

A flexible carbon ring system constitutes the major part of the tetracycline ion (see Fig. 1). Successful structure solution is only possible if the flexibility of the ring system is appropriately taken into account. One possibility would be to separate the flexible ring system into several rigid bodies, but such an approach bears the disadvantage of increasing the total number of degrees of freedom significantly, resulting in much longer calculation times. Instead, we will carry out a thorough conformational analysis to determine all possible conformations of the carbon ring system. In the structure solution step, the whole ring system will be defined as a single rigid body, and structure solution has to be attempted for each conformation of the ring system separately, until the crystal structure is solved.

Since a conformer search is computationally demanding, an initial set of low-energy conformations will be determined using the COMPASS force field. In a second step, these conformations will be further optimized by semi-empirical calculations with MOPAC using the PM3 Hamiltonian.

A. Sketching the tetracycline ion

1. Start Cerius²

If Cerius² is not already running, open a UNIX window, go to the directory 'tutorial' and type in **cerius2** followed by <ENTER>.

2. Sketch the ion

In the **Visualizer** main panel, select **Build/3D-Sketcher**.

The 3D-Sketcher control panel appears.

Use the Sketcher to construct a model of the tetracycline ion. The structure of the tetracycline ion is shown in Fig. 1. When you have finished, click and hold the **Clean** button until the molecule stops moving. After cleaning the molecule, choose the Selection Mode icon (the arrow at the top left) in the **Sketcher** control panel. Then close the **Sketcher** control panel.

If you do not want to spend time sketching the molecule, you can load the molecular structure.

Select **File/Load Model** from the **Visualizer** menu bar. Load the file **tutorial/TCHC/tetracycline.msi**.

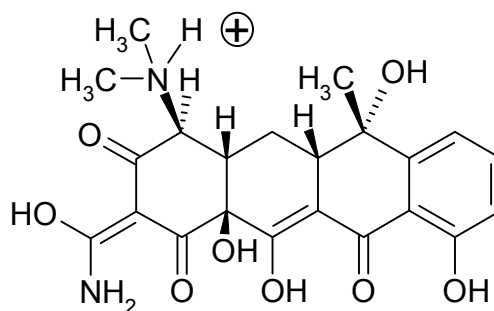


Fig. 1: Structure of the tetracycline ion.

B. Optimizing the tetracycline ion with the COMPASS force field

1. Force field setup

Select the **OPEN FORCE FIELD** card from the **OFF SETUP** stack and click on **Load**. Load the **compass** force field.

Note that the Compass force field comes with its own charges. Charges are attributed to atoms when the atom typing is performed. Compass should only be used with its own charges.

2. Optimize the molecular geometry

Select **Run** from the **MINIMIZER** card on the **OFF METHODS** stack.

This opens the Energy Minimization control panel.

In the **Energy Minimization** control panel, click on the **Minimize Energy** button.

During the minimization process, energy changes are reported in the graphics window and the text window.

C. Searching low-energy conformers by Dynamics/Simulated Annealing

1. Dynamics Simulation

Select the **DYNAMICS SIMULATION** card from the **OFF METHODS** stack and click on **Run**. This opens the **Dynamics Simulation** control panel.

By default the NVE ensemble is selected for dynamics simulations. The temperature is allowed to vary, but the atomic velocities may be scaled periodically to keep the temperature within a specified range.

In a complete simulated annealing cycle, the temperature is gradually increased from an initial value to a mid-cycle temperature and then slowly decreased back to the initial value again. At the end of each cycle, the energy of the molecule is minimized. The next annealing cycle starts using the final structure and the final velocities from the previous cycle.

Check the **Anneal Dynamics** check box under **Dynamics Methods**. Click on the **Preferences** pushbutton next to Anneal Dynamics. Set **Number of Annealing Cycles** to 15, **Initial Temperature** to 300, **Mid-cycle Temperature** to 2000, **Temperature Increment** to 50 and **Steps of Dynamics per Increment** to 250. Turn on **Minimize Model After Each Cycle**. Close the **Anneal Dynamics Preferences** control panel.

Before we start the dynamics simulation, we have to make sure that the optimized structures obtained at the end of each simulated annealing cycle are saved into a trajectory file.

Click on **Trajectory** in the **Dynamics Simulation** control panel. Turn on **Create Annealing Dynamics Trajectory**, and enter 'tetracycline_dynamics' in the **Filename Prefix** box. Close the **Dynamics Output Preference** control panel.

Now we can start the anneal dynamics simulation.

In the **Dynamics Simulation** control panel, click on **Run Dynamics**.

Running the simulation will take some time. Therefore, a file with the results of the calculation is provided for the next section of this tutorial, and we can stop the calculation.

Click on the **INTERRUPT** button in the **Cerius2 Interrupt** window. Choose **Stop current process ASAP**.

2. Analyze the output

Select the **ANALYSIS** card from the **OFF METHODS** stack and click on **Input**. Load the file **tutorial/TCHC/tetracycline_dynamics.atrj**. Cerius² will inform you that the current model is not empty. Choose **OVERWRITE**.

Conformation 1 of the trajectory file is displayed in the model window.

Select **Analyse/Statistics** from the **ANALYSIS** card. Under **Select Properties**, highlight **Total Energy**. Press the **PROFILE** button to show a graph with the energies of all conformers in the trajectory file.

There should be 15 conformers saved in the trajectory file. The different conformers can be selected by clicking on the corresponding data points in the graphics window.

Use the **Analysis Statistics** panel to search for the **LOWEST 15** frames.

In the text window, the 15 conformers are listed in the order of increasing energy. Several conformers have exactly the same energy, indicating that they are identical. In total, there are 5 different conformations. For use in the next section, we copy the most stable conformation into a new model space.

Select **Show Frames** from the **ANALYSIS** card and set the current frame to 6. Choose **Edit/Copy** from the **Visualizer** menu bar. Click on the **plus** sign in the **Visualizer** control panel to create a new model space. Select **Edit/Paste** from the **Visualizer** menu bar. Change the name of the new model space.

D. Optimizing low-energy conformations with MOPAC

To improve the geometries of the five conformations and to obtain a more accurate energy ranking, we are now going to optimize all conformers with MOPAC using the PM3 Hamiltonian. Looking at the five conformations, it seems that the tetracycline ion may be further stabilized by making and intramolecular hydrogen bond between the hydroxyl group of the OH-C-NH₂ fragment and an oxygen atom connected to the carbon ring system. However, such a hydrogen bond is present in none of the conformations. Since the absence of the intramolecular hydrogen bond may be related to a force field error, we will make and break this hydrogen bond manually and carry out two separate structure optimizations for each of the original five conformations.

In the case of tetracycline, it turns out that the default minimization algorithm in MOPAC has convergence problems, and we have to change the minimization algorithm to Eigenvector Following.

Select the **MOPAC** card from the **QUANTUM 1** stack and choose **Run**. In the **MOPAC Run** panel, change the **Task** to **Geometry Optimization** and the **Method** to **PM3**. Enter a new **File Prefix** and set the **Charge** to **1**. Click on the **Task** related **More** button. In the **MOPAC Task Options** panel, change the minimization algorithm to **Eigenvector Following**. Close the **MOPAC Task Options** panel. Click on **RUN** in the **MOPAC Run** panel.

An xterm window appears providing information on the progress of the job. The calculation is finished when "MOPAC done" is printed at the end of the output file.

To check the **heat of formation**, select **Analyse/Files** from the **MOPAC** card.

After geometry optimization, the intramolecular hydrogen bond is closed. We now reopen the hydrogen bond and repeat the geometry optimization.

1) Select **BUILD/Edit H-bonds** from the **Visualizer** menu bar and click on **CALCULATE** in the **Edit Hydrogen Bonding** panel. Check on **Enable automated recalculation**.

2) Select **Move/Bond Geometry** from the **Visualizer** menu bar. In the **Bond Geometry** panel, change **LAST** to **FIRST**. In the model window, select a series of 4 neighboring atoms, starting with the hydrogen atom of the hydroxyl group belonging to the OH-C-NH₂ fragment. Click on the **Measure current torsion** button. Use the torsion angle entry field to change the torsion angle by 180°.

3) Select the **MOPAC** card from the **QUANTUM 1** stack and choose **Run**. Enter a new **File Prefix** and click on **RUN**.

This time, the intramolecular hydrogen bond is still open after the geometry optimization. According to the results obtained with the PM3 Hamiltonian, closing the intramolecular hydrogen bond decreases the total potential energy from -126.15 kcal/mol to -138.46 kcal/mol. We now have to repeat the geometry optimizations for the other four conformations. To reduce the time required for working through this tutorial, these calculations have already been carried out, and the results are summarized below.

In all cases, closing the intramolecular hydrogen bond decreases the total potential energy by about 10 kcal/mol. The five original conformations with closed hydrogen bonds minimize towards only three separate potential energy minima. The

corresponding structures *tc_conf1_PM3*, *tc_conf2_PM3* and *tc_conf3_PM3* can be found in the directory *TCHC* and have energies of -138.46 kcal/mol, -137.75 kcal/mol and -133.94 kcal/mol, respectively.

Part 3: Structure Solution

Whenever there is ambiguity with respect to the space group or the conformation of rigid bodies, it is likely that structure solution has to be attempted several times, and one usually starts with the most likely space group and the most stable conformation. When the crystal structure of tetracycline hydrochloride was originally solved with POWDER SOLVE, it turned out that the right space group was indeed the most frequent space groups $P 2_1 2_1 2_1$. However, two unsuccessful attempts were made with the two most stable conformations before the crystal structure was finally solved with least stable of the three conformations. In this tutorial, we will directly start with the conformation that allows for structure solution.

A. Setting up a POWDER SOLVE run

1. Start Cerius²

If Cerius² is not already running, open a UNIX window, go to the directory 'tutorial' and type in **cerius2** followed by <ENTER>. Otherwise, select **File/New Session** from the menu bar and **confirm**.

2. Load the powder diffraction pattern

To load the powder diffraction pattern, find the **POWDER SOLVE** card on the **POWDER SOLVE** stack and click on **Experimental Data** to bring up the **1-D Experimental Data** control panel. Set the data file format to **XY-GRAPH**.

Load the file **tutorial/TCHC/tetracycline_hydrochloride_xrd.grf** and close the **1-D Experimental Data** control panel.

3. Load the result of profile fitting in $P 2_1 2_1 2_1$

Select **File/Load Model** from the **Visualizer** menu bar. Load the file **tutorial/TCHC/PF_P212121.msi**.

4. Load a conformer of the tetracycline ion and define rigid bodies.

Select **File/Load Model** from the **Visualizer** menu bar. Load the file **tutorial/TCHC/ tetracycline_conf3_PM3.msi**

Now we have to define rigid bodies so that all important torsional degrees of freedom can be varied. Since X-ray scattering from hydrogen atoms is relatively weak, we will neglect scattering from hydrogen atoms in the structure solution step. Therefore, we will not define rigid bodies for torsional degrees of freedom that involve the rotation of hydrogen atoms only. In total, we will divide the molecular fragment into two rigid bodies: the CH₃-NH-CH₃ group and the remaining atoms of the molecule. All atoms should be part of one and only one rigid body.

Click on **Simulation Setup** in the **POWDER SOLVE** card and choose **Rigid bodies** from the pull down menu to open the **Rigid Bodies** control panel.

Select all atoms that belong to the first rigid body. To do this, enclose all atoms of the first rigid body in a selection box by clicking in the model window with the left mouse button and dragging the mouse to the lower right while holding the mouse button pressed. If you cannot select all atoms this way, add the remaining atoms to the group of selected atoms by clicking on them while holding the <**SHIFT**> key.

In the **Rigid Bodies** control panel, click on the **Define Rigid Body** button and then on the **Color Rigid Bodies** button.

Deselect all atoms by clicking in some empty space in the model window.

Repeat the procedure for the second rigid body.

At this point, the molecular ion should consist of two parts of different color.

5. Create a model containing a chlorine ion and define rigid a rigid body.

Create a new model space by clicking on + in the **Visualizer** control panel. Change the name of the new model to **Cl ion**. To do so, click on the name of the model in the **Visualizer** control panel. Type in the new name and press <ENTER>.

Select **Build/3D-Sketcher** from the **Visualizer** menu bar. In the **Sketcher** control panel, type in 'Cl' in the text input field to the right of **sketch with** and press <ENTER>. Click in the model window to create a chlorine atom. Choose the **Selection Mode** icon (the arrow at the top left) in the **Sketcher** control panel. Then close the **Sketcher** control panel.

Click on **Simulation Setup** in the **POWDER SOLVE** card and choose **Rigid bodies** from the pull down menu to open the **Rigid Bodies** control panel. In the **Rigid Bodies** control panel, click on the **Locate Rigid Bodies** button and then on the **Color Rigid Bodies** button. Close the **Rigid Bodies** control panel.

6. Copy the ions into the unit cell

In the **Visualizer** control panel, click on the lozenge to the left of **PF_P212121** to make this model the active model. Click on the squares to the right of **tc_conf3_PM3** and **Cl ion** to show all three models.

The model window should now be divided into 6 parts with PF_P212121 being shown in the large model space.

Click on some empty space in the **tc_conf3_PM3** model with the left mouse button, hold it down and drag the **tc_conf3_PM3** model into the **PF_P212121** model. Release the mouse button.

Four symmetry copies of the molecular ion should appear in the PF_P212121 model.

Click on the empty space in the **Cl ion** model space with the left mouse button, hold it down and **drag** the **Cl ion** model into the **PF_P212121** model space. Release the mouse button.

Four symmetry copies of the Cl ion should appear in the PF_P212121 model.

In the **Visualizer** control panel, click on the green squares to the right of **tc_conf3_PM3** and **Cl ion** to show only the model **PF_P212121**.

7. Calculate the density

Select **Unit Cell** -> **Cell Contents** from the **CRYSTAL BUILDER** card on the **BUILDERS 1** stack. Examine the **Cell Contents** panel.

With one ion pair in the asymmetric unit, tetracycline hydrochloride has a density of 1.44 g/cm³. This is a very plausible value, and we conclude that there can not be more than one ion pair in the asymmetric unit, if the space group is P 2₁ 2₁ 2₁.

8. Choose the angular range of the powder diffraction pattern to be used for structure solution

The CPU time required for a POWDER SOLVE run is approximately proportional to the number of reflections in the powder diffraction pattern. By cutting off the high angle part the diffraction pattern, calculation times can be reduced significantly. However, care has to be taken that the remaining part of the diffraction pattern contains enough information to solve the crystal structure.

First, we will determine the number of degrees of freedom.

On the **POWDER SOLVE** card, click on **Simulation Setup** and choose **Degrees of Freedom** from the pull down menu. **Find** the number of degrees of freedom in the **Degrees of Freedom** panel.

There are 10 degrees of freedom. Now, we find out how many reflections can be found below 15°.

On the **POWDER FIT** card, click on **Refinable Variables** and choose **Intensities** from the pull down menu. **Scroll** down the list of reflections and count the number of reflections below 15°.

There are about 80 reflections below 15° , corresponding to 8 reflections per degree of freedom. In general, this is sufficient for structure solution. Therefore, we change the upper limit of the powder diffraction pattern to 15°.

On the **POWDER FIT** card, choose **Run** to open the **Powder Fit** control panel. In the **Powder Fit** control panel, click on **Range** and change **Max** in the **Profile Range** panel to **15.00**. Close the **Profile Range** panel.

We have to optimize the profile fitting for the angular range up to 15.00°.

Click on **RUN** in the **Powder Fit** control panel. Select **Files/Save Model** from the menu bar of the **Visualiser** panel and save the current model.

By saving the model, we have stored our definition of rigid bodies and the results of the last profile fitting.

Examine the result of the profile fitting in the graphics window. During profile fitting, all intensities are refinable parameters, while they are related to the unit cell content in the structure solution step. Therefore, the R_{wp} factor of 6.9 % obtained by profile fitting is a lower limit for the best possible outcome of the structure solution step.

9. Start the POWDER SOLVE run

Select **Run** from the **POWDER SOLVE** card to open the **Powder Solve** control panel.

By default, Powder Solve will perform 5 simulated annealing cycles and set the start temperature and the end temperature automatically. Scattering from hydrogen atoms will be neglected. We only have to set the number of steps per cycle and to choose a filename for the output.

Click on the **Propose no. of steps** button in the **Powder Solve** control panel.

The required number of Monte Carlo steps is estimated from the number of degrees of freedom. The function used to calculate the number of Monte Carlo steps has been calibrated in such a way that a solution is usually obtained within 5 Monte Carlo cycles, if structure solution is possible for the chosen combination of space group and rigid bodies.

Click on **Output** in the **Powder Solve** control panel. In the **Output Options** panel, set the **File Seed** to **PS_conf3**. Close the **Output Options** panel. In the **Powder Solve** control panel, click on **RUN**.

Look at the output in the graphics window. The structure solution of TCHC requires a certain amount of CPU time. If you do not want to wait, stop the calculation and continue with the tutorial.

Click on the **INTERRUPT** button in the **Cerius2 Interrupt window**. Choose **Stop current process ASAP**.

B. Analyzing the results of a POWDER SOLVE run

We now have to decide, if the crystal structure has indeed been solved in the POWDER SOLVE run. Visually comparing the simulated and experimental powder patterns, it is often obvious whether or not the best crystal structure found by

Powder Solve actually matches the true crystal structure. Sometimes, however, it is difficult to make this decision, since moderate intensity differences may be related to an incorrect crystal structure as well as preferred orientation or inaccurate conformations of rigid bodies. In the case of TCHC, the whole carbon ring system has to be defined as a single rigid body, and slight deviations from the true conformation cause significant intensity differences. In order to distinguish between correct and incorrect solutions, it can be helpful to determine close contacts, hydrogen bonding schemes and voids in the crystal structure.

1. Extract the best solution from the POWDER SOLVE output file

1) Select **Analysis/Input** from the **POWDER SOLVE** card. Load the file **tutorial/TCHC/PS_conf3.trj** and choose **OVERWRITE**. Close the **Analysis Input** window.

2) Select **Analysis/Analyze** from the **POWDER SOLVE** card. In the **Analysis Statistics** window, change **HIGHEST 10** to **LOWEST 1** and click on the **Search for** button. The number of the crystal structure with the lowest R_{wp} factor is shown in the text window.

3) Select **Analysis/Show frames** from the **POWDER SOLVE** card. Change the frame number from 1 to 13 (best solution found). When you press **<ENTER>**, the corresponding crystal structure appears in the model window. To project all symmetry copies of the molecule into the unit cell, choose **Crystal Building** from the **CRYSTAL BUILDER** card on the **BUILDER 1** stack. Click on **UNBUILD CRYSTAL** and then on **BUILD CRYSTAL**. Close the **Crystal Building** control panel.

2. Compare the simulated and the experimental powder diffraction pattern

Choose **Run** on the **POWDER SOLVE** card. In the **Powder Solve** control panel, set the task to **SHOW**. Click on **RUN** to calculate a powder pattern for the current model.

The comparison between the simulated and experimental powder patterns shows some important intensity mismatches, but the overall intensity distribution is correct. Since the number of reflections in the diffraction pattern is significantly higher than the number of degrees of freedom, it is unlikely that the good agreement in the overall intensity distribution is just accidental. We conclude that the current model is probably close to the true crystal structure, but needs further refinement.

In cases without flexible ring systems, successful structure solution typically leads to R_{wp} factors that differ from the R_{wp} factor obtained with Powder Fit (over the same angular range) by no more than a factor of two. In the present case, the two values differ by almost a factor of four (24.5 % compared to 6.9 %), but since we were expecting difficulties due to the flexibility of the carbon ring system, we are willing to accept this large relative difference in the R_{wp} factors.

3. Calculate close contacts

- 1) Select **Simulation Setup/Rigid Bodies** from the **Powder Solve** card. Click on some empty space in the model window to make sure that no atoms are selected. Click on **Remove Rigid Bodies** in the **Rigid Bodies** control panel. Choose **OK** when asked if you want to remove all rigid bodies. Click on **Color Rigid Bodies** in the **Rigid Bodies** control panel.
- 2) Select **Geometry/Close contacts** from the **Visualizer** menu bar. Click on **Monitor close-contacts**. To examine the close contacts, rotate the model by pressing the right mouse button and dragging the mouse.

There are some close contacts that cannot be attributed to hydrogen bonding, but we do not observe any serious overlap between molecular fragments. The $CH_3-NH-CH_3$ fragment participates in several close contacts that would not be tolerable for a completely refined crystal structure, but that can be accepted at this stage. The $O-H..H-C$ distance of 1.91 Å is clearly too short, but the hydrogen positions were not determined in the structure solution step, and by adjusting the torsion angle of the methyl group it is possible to eliminate this close contact. We conclude that the crystal structure does not have to be rejected because of atomic overlap.

Click on the **Turn off close-contact monitoring** button and close the **Close Contacts** panel.

4. Examine hydrogen bonding

Select **Build/Edit H-Bonds** from the **Visualizer** menu bar. Click on **CALCULATE**. Rotate the model to examine the hydrogen bonding pattern.

The hydrogen bonding pattern looks very reasonable. All Hydrogen bonding donors and acceptors participate in hydrogen bonding.

5. Search for voids in the crystal structure

Select **Geometry/Free Volumes** from the **Visualizer** menu bar. Change the **Volume to Calculate** from **TOTAL** to **OCCUPIABLE** and press **CALCULATE**. Close the **Free Volume** panel.

No surfaces should appear in the model window, indicating that there are no significant voids in the crystal structure. Finally, we conclude that the current crystal structure is a very good candidate for the correct crystal structure and save it for further refinement.

Select **File/Save Model** from the **Visualizer** menu bar and save the current model.

Part 4: Structure Refinement

The approximate crystal structure found in the structure solution step can be further refined using POWDER SOLVE as a rigid body Rietveld refinement tool. A larger molecule like tetracycline hydrochloride may be cut into several rigid bodies, refining the position and the orientation of each part independently. If the information content of the powder diffraction pattern is sufficient, all atoms can be defined as individual rigid bodies. It is also possible to repeat the profile fitting procedure with intensities calculated from the crystal structure, and after one or two iterations between structure refinement and profile fitting, all parameters are usually converged.

Refining the crystal structures of larger molecules, one frequently ends up in local minima on the R_{wp} hypersurface, and it may be necessary to use the Monte Carlo/simulated annealing approach again to get closer to the global minimum. In this part of the tutorial, we will take a different approach and show how molecular mechanics can be used to improve the crystal structure.

In the crystalline state, the conformation of a molecule may differ from the conformation of an isolated molecule as a consequence of the interactions between the molecule and its crystal environment. Therefore, performing a force field based geometry optimization of the preliminary crystal structure obtained by structure solution may significantly improve the molecular conformation. In the case of tetracycline hydrochloride, several different hydrogen bonding patterns are possible depending on the orientation of the hydroxyl groups. When the crystal structure of tetracycline hydrochloride was originally refined, the relative stability of the different hydrogen bonding schemes was investigated using the COMPASS force field, and it turned out that the orientation of the hydroxyl groups after structure solution corresponded to the most stable hydrogen bonding pattern.

1. Start Cerius²

If Cerius² is not already running, open a UNIX window, go to the directory 'tutorial' and type in **cerius2** followed by <ENTER>. Otherwise, select **File/New Session** from the menu bar and **confirm**.

2. Load the powder diffraction pattern, the starting structure and the COMPASS force field

To load the powder diffraction pattern, find the **POWDER SOLVE** card on the **POWDER SOLVE** stack and click on **Experimental Data** to bring up the **1-D Experimental Data** control panel. Set the data file format to **XY-GRAPH**.

Load the file **tutorial/TCHC/tetracycline_hydrochloride_xrd.grf** and close the **1-D Experimental Data** control panel.

Select **File/Load Model** from the **Visualizer** menu bar. Load the file **tutorial/TCHC/ PS_solution.msi**.

Select **Load** from the **OPEN FORCE FIELD** card on the **OFF SETUP** stack. Load the **compass** force field.

3. Increase the angular range

In the structure solution step we have limited the angular range to a portion of the powder diffraction pattern to reduce calculation times. For the structure refinement, we are going to use the full powder diffraction pattern.

Select **Simulation Setup -> Range** from the **POWDER FIT** card on the **POWDER SOLVE** stack. Set the upper limit (**Max**) of the powder diffraction pattern to 30.0 and close the **Profile Range** panel.

Choose **Run** on the **POWDER FIT** card and click on **RUN** in the **POWDER FIT** panel to perform profile fitting over the full angular range.

Profile fitting should result in a R_{wp} factor of about 5.88%. This value represents a lower limit for the R_{wp} factor after structure refinement, and the relative deviation from this value is a good indicator for the quality of the structure refinement.

4. Take into account scattering from hydrogen atoms

Select **Run** from the **POWDER SOLVE** card on the **POWDER SOLVE** stack. Switch on **Include H atoms**. Set the task to **SHOW** and press **RUN**.

The R_{wp} value in the graphs window should be about 24.25%. This is the starting value for our structure refinement.

6. Optimization using molecular mechanics

1. Choose the **MINIMIZER** card on the **OFF METHODS** stack. Select **Constraints** -> **Cell** and switch off the optimization of cell parameters.
2. Select one chlorine atom in the models window. Select **Constraints** -> **Atoms** and click on the **Fix Atomic Position** button followed by the **Color Atoms by Constraint** button.
3. Choose **Run** on the **MINIMIZER** card and click on **RUN**.
4. Make sure that the task in the **Powder Solve** panel is set to **SHOW** and click **RUN**.

The new R_{wp} factor should be about 20.07%.

7. Repeat the profile fitting with intensities calculated from the crystal structure

In the Powder Solve panel, click on the Export Intensities to Powder Fit button. Choose **Run** on the **POWDER FIT** card and click on **RUN** in the **POWDER FIT** panel to repeat profile fitting over the full angular range with fixed peak intensities.

The R_{wp} factor should go down to about 19.48%.

8. Repeat Rietveld refinement with two rigid bodies

Select **Simulation Setup** -> **Rigid Bodies** on the **POWDER SOLVE** card. Click on the **Locate Rigid Bodies** button, followed by the **Color Rigid Bodies** button.

Select **Run** on the **POWDER SOLVE** card, change the task to **REFINE** and click on **RUN**.

The R_{wp} factor should go down to about 15.32%.

Cutting the molecular ion into several independent rigid bodies and continuing the structure refinement, the R_{wp} factor can be brought down to 9.35%. To examine the final result of the structure refinement, load the file tutorial/TCHC/tchc_final_result.msi. Fig. 2 shows a superposition of the crystal structure obtained with Powder Solve and a crystal structure determined by single crystal X-ray diffraction [6]. The main difference between the two structures is a deviation of the unit cell parameters related to different experimental temperatures.

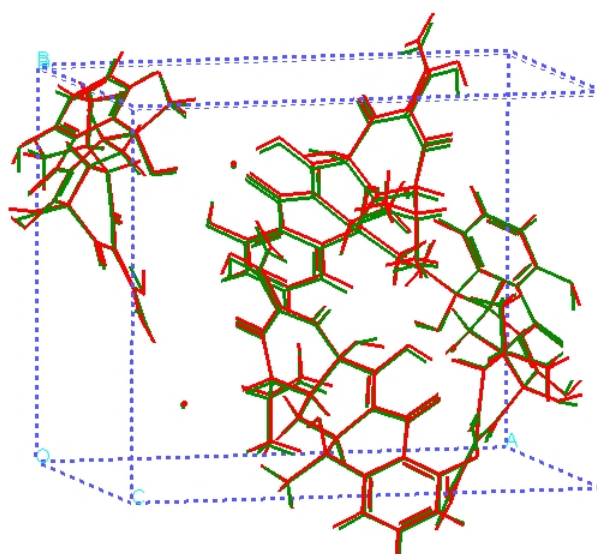


Fig. 2: Superposition of the crystal structure of tetracycline hydrochloride obtained with Powder Solve and a crystal structure determined by single crystal X-ray diffraction [6].

Conclusion

When crystal structures of organic compounds have to be determined, the molecular structure is usually known prior to structure solution. Using potential energy calculations for conformational analysis and geometry optimization, it is possible to reduce the total number of degrees of freedom. Individual atomic coordinates can be replaced by a small number of parameters, describing the rotation of rigid bodies with respect to each other as well as the position and orientation of connected groups of rigid bodies in the unit cell. Taking into account the knowledge about the molecular geometry, fairly complex crystal structures can be solved directly from powder diffraction data by Monte Carlo / simulated annealing techniques. With the appearance of readily available software packages for molecular modeling and structure solution, crystal structure determination from

powder data has already become a routine task for compounds like tetracycline hydrochloride

Acknowledgements

MSI would like to thank Prof. Armel le Bail for supplying the powder diffraction pattern of tetracycline hydrochloride and Prof. W. Clegg for providing the single crystal structure.

Literature References

- [1] G. E. Engel, S. Wilke, O. König, K. D. M. Harris and F. J. J. Leusen, *Powder Solve - a complete package for crystal structure solution from powder diffraction patterns*, J. Appl. Cryst. **32**, 1169-1179 (1999)
- [2] <http://www.cristal.org/SDPDRR/>
- [3] L. Eriksson and M. Westdahl, *TREOR, a semi-exhaustive trial-and-error powder indexing program for all symmetries*, J. Appl. Cryst. **18**, 367-370 (1985)
- [4] A. Boulton and D. Louër, *Indexing of Powder Diffraction Patterns for Low-Symmetry Lattices by the Successive Dichotomy Method*, J. Appl. Cryst. **24**, 987-993 (1991)
- [5] <http://wwwchem.tamu.edu/services/crystal/tables/allspg.html>
- [6] W. Clegg and S. J. Teat, *Acta Crystallogr. Sect. C*, 2000, in press.

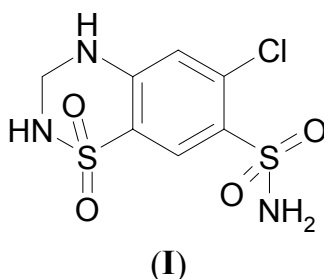
DASH Tutorial 1¹

Kenneth Shankland & William I. F. David

*ISIS Facility
Rutherford Appleton Laboratory
Chilton, Didcot
Oxon OX11 0QX
United Kingdom*

Introduction

The object of this tutorial is to guide you step by step through the process of structure solution, using the molecule hydrochlorothiazide (**I**) as an example. As this is the first tutorial example, it goes through the process in considerable detail – subsequent tutorial examples will be more concise, but will introduce other, new aspects of the structure solution process. This tutorial will take a novice user about 2 hours to complete. Note that experienced powder crystallographers will take considerably less time and after familiarisation with DASH, an entire ‘data-to-structure’ session will take less than 15 minutes.



Data

The data set “hct20.xye” is a synchrotron X-ray diffraction data set collected at 20K on Beamline X7A of the Brookhaven National Synchrotron Light Source. The incident wavelength was 1.1294Å and the sample was held in a 0.7mm glass capillary. We gratefully acknowledge the help of Dr. Dave Cox who help set up the experiment.

¹ Copyright CCLRC / CCDC

Stage 1: Reading the data

Open up the DASH package by double clicking on the DASH icon. Use the directory selector to choose the directory where the data resides e.g.

C:\Program Files\Dash\Tutorials\Hydrochlorothiazide (*note that the exact location will depend upon how the program has been set up for the tutorials!!*)

The DASH Wizard will guide you through the structure solution process, which is performed in a series of steps.

- Select the first option, 'View data / determine peak positions' and click 'Next'.
- Click the 'Browse...' button.
- Select the file 'hct20.xye' - the diffraction data will be loaded into DASH.
- Click on Finish to dismiss the Wizard.

Stage 2: Examining the data

If you are familiar with handling powder diffraction data, you can probably skip to Stage 3, as this section is purely descriptive.


The data spans 5 to 44° 2 θ . If you open the file 'hct20.xye' in an ASCII file editor (such as Wordpad, the file is too large for the current version of Notepad to open) you'll see that it consists of three columns.

```
5.000      81.96      10.952
5.004      71.25      10.284
.
<bulk of data omitted for clarity>
.
43.996     69.55      3.572
44.000     68.28      3.540
```

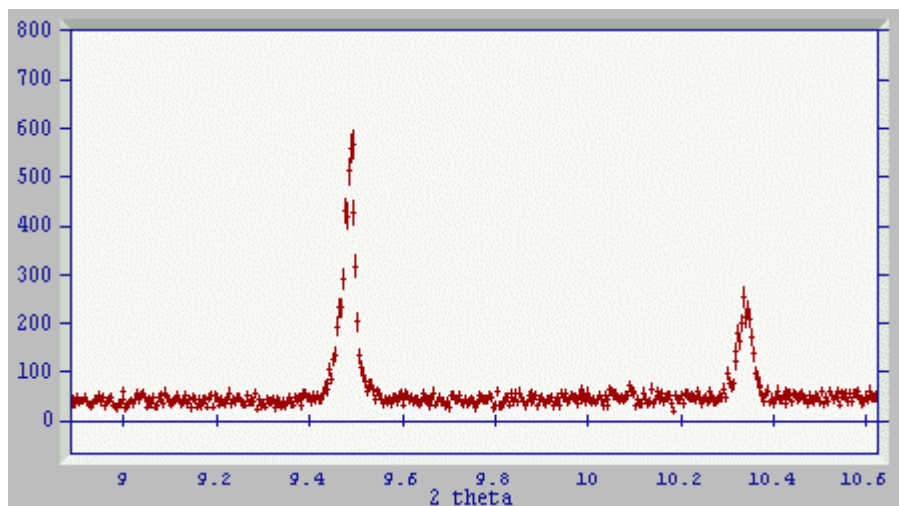
Column 1 = 2 θ position

Column 2 = diffracted intensity (counts)

Column 3 = estimated standard deviation of the intensity

If you zoom in on the diffracted data as it is displayed in DASH, you'll see that DASH displays both the intensity and the error bars. The simplest way to zoom is to use the left mouse button ; ensure that you are in Zoom mode (this is the default mode) by selecting 'Default' from the 'Mode' menu, or depressing the  icon on the menu bar. Click and hold the left mouse button and drag out a rectangle around the area that you want to zoom in on. To zoom out, simply

press the 'Home' key on the keyboard (*Note that there are other ways to zoom the data – see the DASH Interface Guide for details*). Try zooming in on the two peaks that lie just either side of $10^\circ 2\theta$.



You can use the left and right cursor keys to move up and down the data in 2θ . Some other useful keyboard shortcuts whilst examining data are


Shift + UpArrow : Zoom in
Shift + DownArrow : Zoom out
Control+ UpArrow : Rescale the y-axis to the max. in the current range

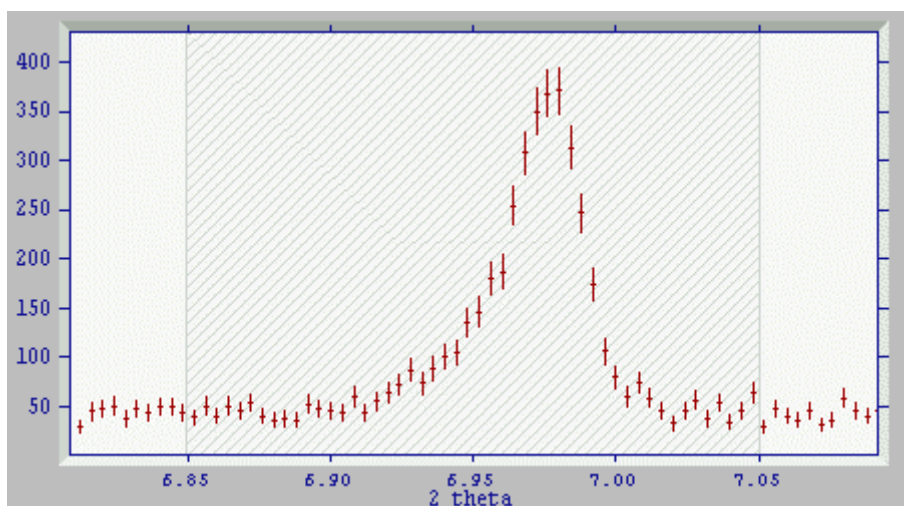
Whilst browsing the data, note the following features

- The peak asymmetry (elongated tails to the left hand side of the peaks) in the low angle peaks, due to axial divergence.
- The flat background indicative of a lack of amorphous content.
- The sharp peaks, indicating a good crystalline sample.
- The excellent instrumental resolution. See for example, the doublet of peaks around 12.17° two theta.
- The use of a small step size commensurate with the instrumental resolution and the narrow peaks i.e. plenty of points across each peak
- The fall off in diffracted intensity with increasing angle due to the Lorentz effect and thermal effects.
- The increasing number of peaks per unit angle with increasing angle.
- The excellent signal to noise ratio, even at the maximum diffraction angle i.e. peaks can still be clearly discriminated from background.

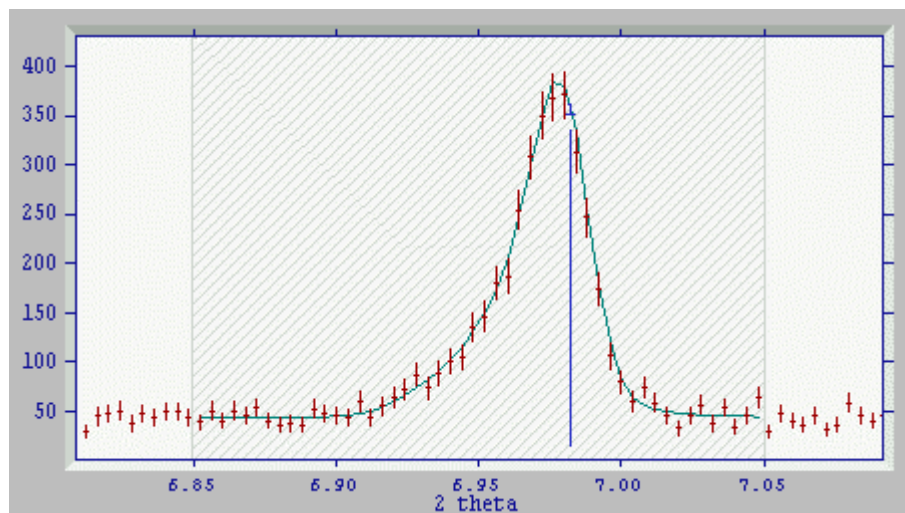
Stage 3. Fitting the peaks to determine the exact peak positions

We need accurate estimates of the 2θ positions of the first 20 or so peaks in the diffraction pattern in order to index the diffraction data i.e. determine the unit cell and hence the Laue class of the crystal. DASH makes this process quick and easy by fitting entire peaks accurately. ***It is important to emphasise that we are interested only in peak positions, not peak intensities, at this stage, so weak peaks are every bit as important in indexing as strong ones.*** The first peak in the diffraction pattern is at just under 7 deg. To fit it ;

- Zoom in to the area around the peak
- Switch to peak fitting mode by selecting 'Peak Fitting' from the 'Mode' menu or by depressing the  button on the menu bar
- Sweep out an area using the right mouse button i.e. move to about 6.85 deg two theta, click right and hold down as you sweep right to about 7.05 degrees before releasing the right button. The hatched area now covers the peak and enough background either side to allow an accurate estimate of the peak parameters. If you are not happy with the area that you've swept out (e.g. your finger slipped as you were sweeping!), simply put the cursor inside the hatched area and press the 'Delete' key on the keyboard to remove the current selected area, then try again.



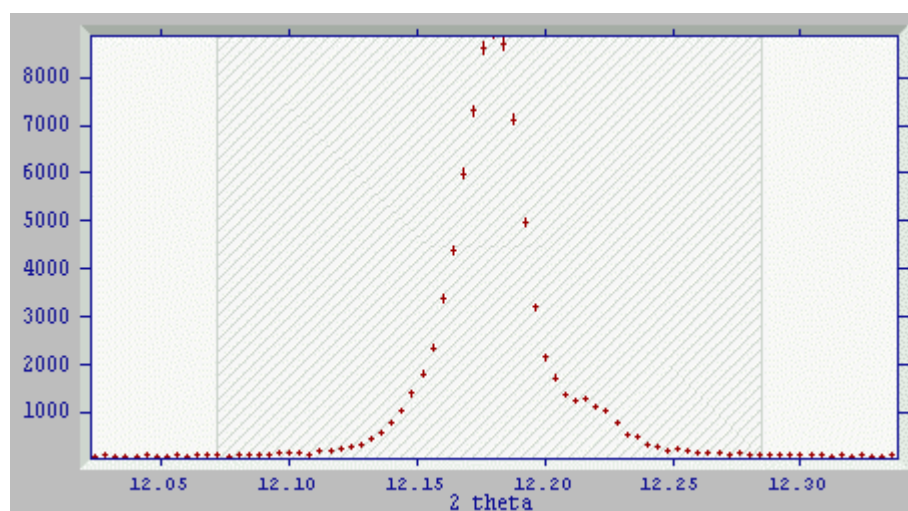
- With the cursor in the hatched area, press the Return or Enter key to fit the peak



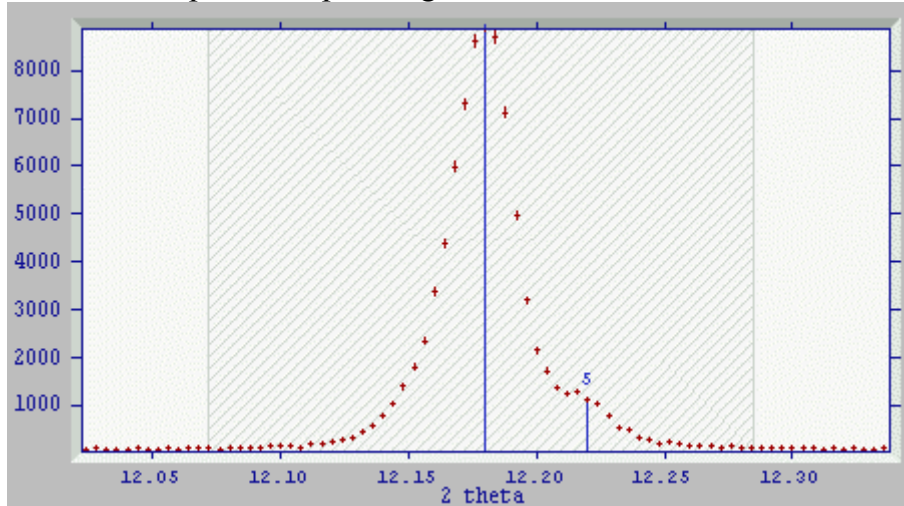
- The solid green line indicates the fit to the data, whilst the vertical blue line indicates the peak position. Selecting 'View' from the 'Peak Positions' shows the exact peak position

Diffraction Data		Peak Positions	
	Position	Esd	Tic
1	6.9822	0.0030	

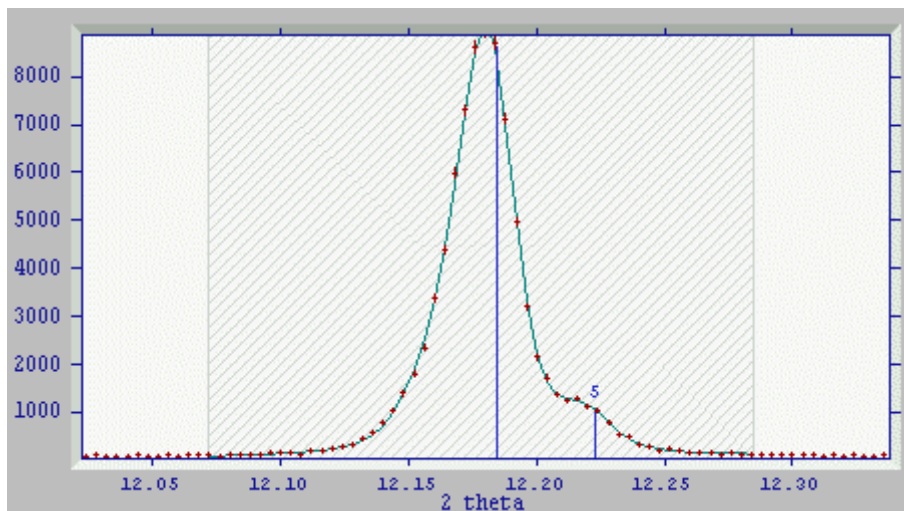
- Don't worry if you don't have the exact same position! It should be very close though.
- If your peak fit (and the resultant peak position) does not look anything like the fit shown above, please ask for advice.
- Fit the next two peaks at around 9.5 and 10.3 ° 2θ in the same way.
- Zoom in on the doublet at 12.17 ° 2θ. It is clear from the shape of the peak that there are two contributing reflections here.
- In peak fitting mode, sweep out an area covering the two peaks using the right mouse button



- Now, you need to give two initial estimates for the peak position. This is easily done by moving the cursor close to the top of the first peak, and pressing '1' on the keyboard to insert the first estimate, then moving to the top of the second peak and pressing '2' to insert the second estimate



- Then, with the cursor inside the hatched area, press 'Return' or 'Enter' as before to fit the two peaks



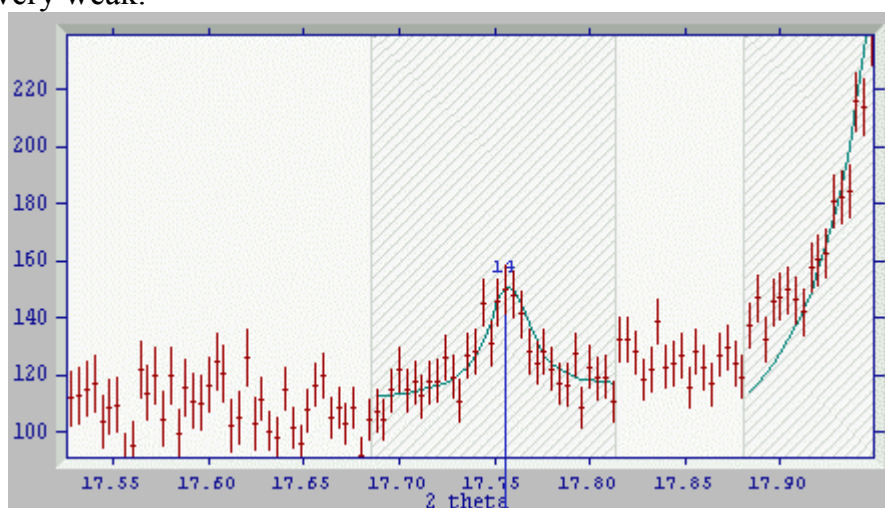
- Now, it's simply a case of repeating this until we have 20 accurate peak positions. Listed below are 2θ ranges and the number of peaks contained in them as a guide.

Region / ° 2θ	# peaks in region	Cumulative # peaks
13.5 to 14.6	3	8
14.6 to 16.4	4	12
16.4 to 17.5	1	13
17.5 to 18.5	2	15
18.5 to 20	4	19
20 to 20.9	3	22
20.9 to 21.3	2	24

- Here are the peak positions as returned by DASH

6.9822	9.4942	10.3453	12.1847	12.2228
13.6925	13.7905	14.0003	15.2696	15.6883
15.7753	15.9581	16.8146	17.7552	18.0107
19.0501	19.1452	19.3479	19.7249	20.5468
20.6314	20.7735	21.0639	21.1688	

- The only peak you might have struggled to see was the one at ~17.75 ° 2θ, as it is very weak.



Stage 4. Indexing

You can easily get the peak positions out of DASH and into a file by

- selecting 'View' from the 'Peak Positions' menu and then clicking on the word 'Position' at the top of the peak position column. This selects the entire column.
- Use 'Ctrl+C' to copy the entire column to the clipboard
- inside an appropriate editor such as Notepad or Wordpad use 'Ctrl+V' to paste the column into a file.
- Save the line positions into a file with the correct format for your favourite autoindexing program.
- You can exit DASH for the moment.

There's an example of a completed Dcivol input file for this data in the Appendix. Your indexing program should return a monoclinic unit cell of volume $\sim 576 \text{ \AA}^3$. Dcivol returns

$$a=9.9388 \text{ \AA} \quad b=8.49954 \text{ \AA} \quad c=7.31875 \text{ \AA} \quad \beta=111.19^\circ \quad V=576.453 \text{ \AA}^3$$

Figures of merit : $M(24)=131$ $F(24)=446$

With figures of merit as good as these, there is little doubt that the cell has been correctly indexed. It is possible to change this cell into one with a 'conventional' setting, but for the moment, we will proceed with the cell as it is returned by Dcivol

Stage 5. Stop and think

Does the cell make sense? In this case, we estimate the molecular volume to be about 290 \AA^3 and so a cell of this size would accommodate two molecules. Given that the cell is monoclinic, a likely space group is $P2_1$.

Stage 6. Checking the cell and determining the space group

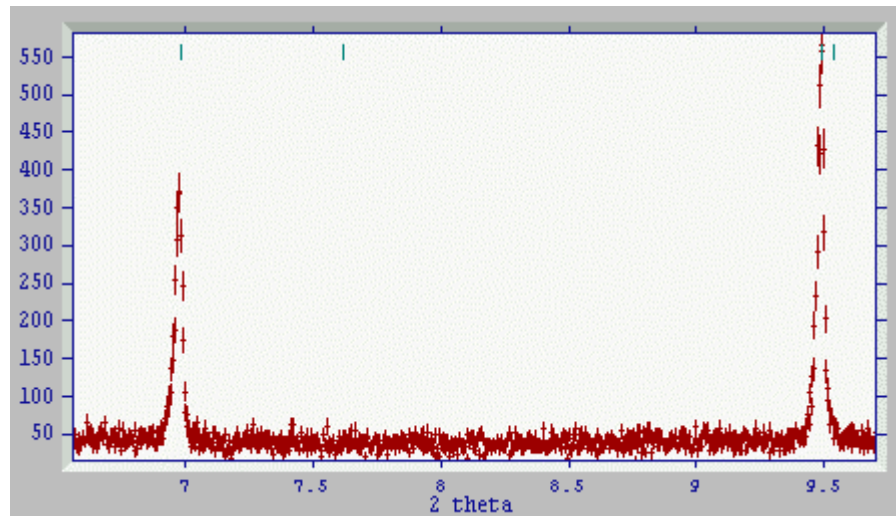
Start up DASH as before and select 'Preparation for Pawley refinement' from the Wizard menu.

- Enter the lattice constants you obtained from indexing into the dialog box, hitting the Tab or using the mouse to move between fields. (*Note that you have to hit Tab or click anywhere except space group before the crystal system is picked up. You cannot click directly on Space Group*)
- Once you've entered the information, you'll see that DASH selects the monoclinic space group $P2$ as the starting space group, as the lattice constants indicate a monoclinic cell and space group $P2$ has no systematic absences.
- Click Next to advance to the next dialog and select 'Synchrotron radiation' ; enter the wavelength of 1.1294 \AA and browse for the file hct20.xye as before.
- Once the data have been loaded, click on 'Close'.

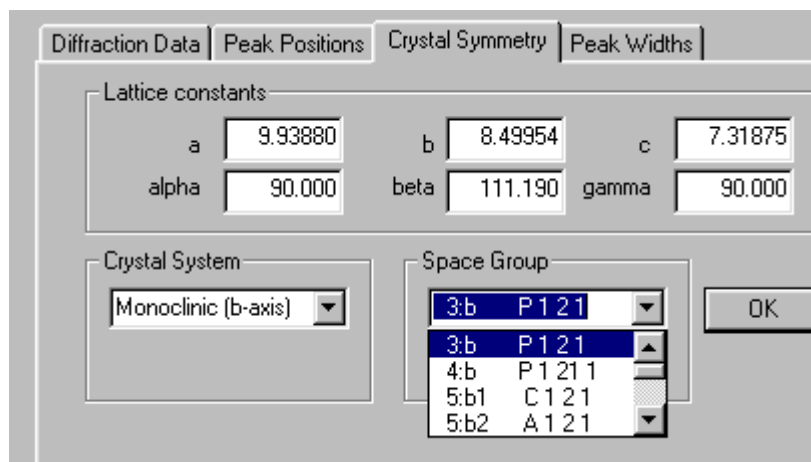
You'll see the data displayed as before, but this time, there are a series of tick marks at the top of the plot to indicate where the Bragg reflections corresponding to the input cell occur.

- The first thing to do is to ensure that in general, the tick marks correspond to peaks within the pattern.
- Any unaccounted for diffraction peaks are a warning that the determined unit cell might not be correct, or that there is an impurity phase present.

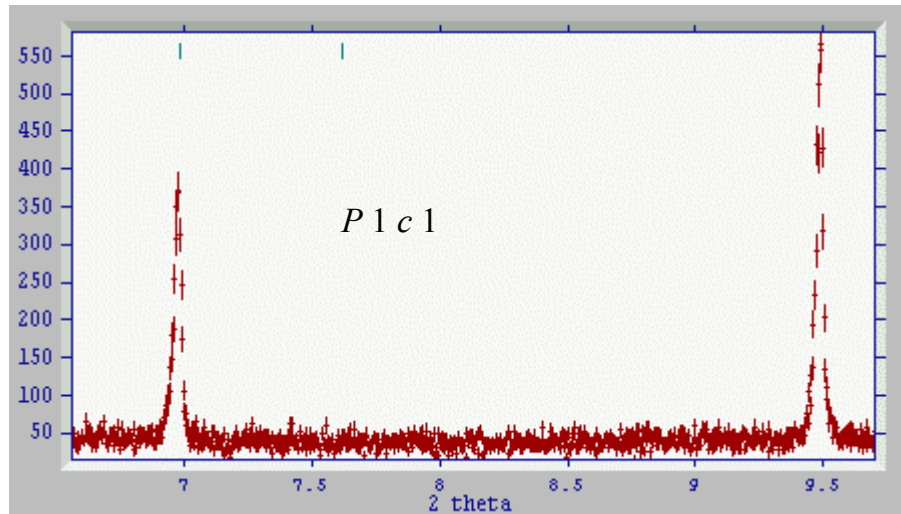
- A quick glance at the 'hct20' pattern shows no unaccounted for peaks, but a few excess tick marks. For example...



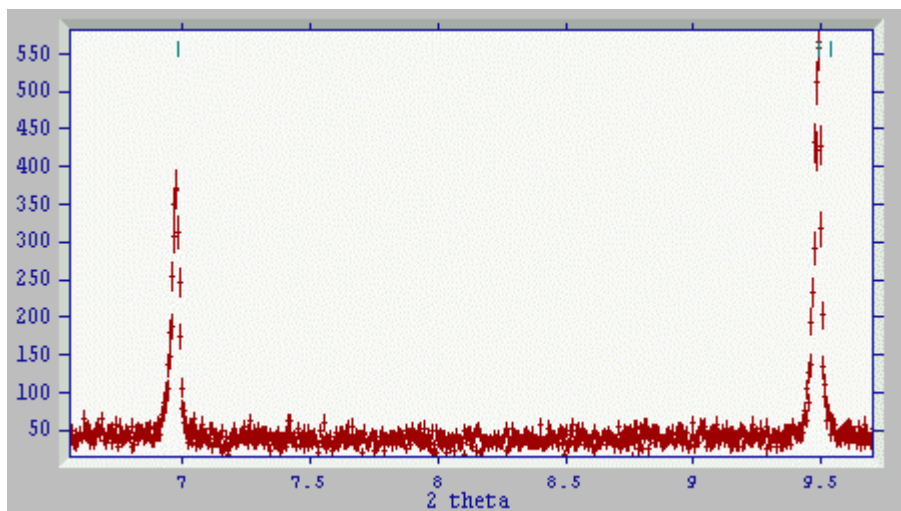
- The tick at just over $7.5^\circ 2\theta$ does not appear to correspond to anything other than background intensity, which means that it probably corresponds to a systematic absence for the true space group of the crystal.
- The tick at just over $9.5^\circ 2\theta$ may be another absence, although there is just a hint of a shoulder present on the stronger peak.
- We already guessed that a likely space group is $P2_1$, so let's see if increasing the symmetry from $P2$ to $P2_1$ eliminates likely absences whilst leaving no unaccounted for peaks.
- Bring up the crystal symmetry dialog by selecting 'Crystal Symmetry' from the 'View' menu.
- Click on the Space Group pull down menu



- Using the down arrow cursor key, move down the list, and watch the tick marks update to show the reflection positions corresponding to the currently selected space group. Alternatively, you can use the mouse to scroll down and select individual space groups.
- It's pretty obvious that choices such as $P 1 c 1$ eliminate major peaks and clearly can't be correct.



- Alternatively, $P 1 2_1 1$ eliminates the tick at $7.5^\circ 2\theta$ whilst leaving one at just over 9.5° .



- Examining the rest of the pattern, the correspondance between tick marks and peaks is excellent and we can conclude that the peak at just over $9.5^\circ 2\theta$ is a very weak diffraction feature of a crystal whose space group is $P2_1$, b axis unique.

Stage 7. Extracting intensities


This is initially much like the indexing phase. We are aiming to model the entire diffraction pattern and so we need to be able to fit peaks. We are confident that we have a reasonably accurate cell and the correct space group. The criteria for peak fitting are however, slightly different from the ones used in indexing.

- We need to fit a number of preferably isolated reflections
- We need to sample peaks across the pattern in order to parameterise the peak shape across the pattern
- We need to ensure that we model any peak asymmetry at the start of the pattern

Here are some suitable peaks for this pattern. Fit them (ie by entering Peak Fitting mode and sweeping out areas over the peaks with the right mouse button as before) in the order they are given

Peak	Approx Location	Note
1	6.97	
2	9.49	
3	14.0	¹ Option to Pawley refine
4	16.8	
5	20.78	
6	22.75	² L/parameter refine starts
7	27.8	
8	31.85	³ Pawley window appears

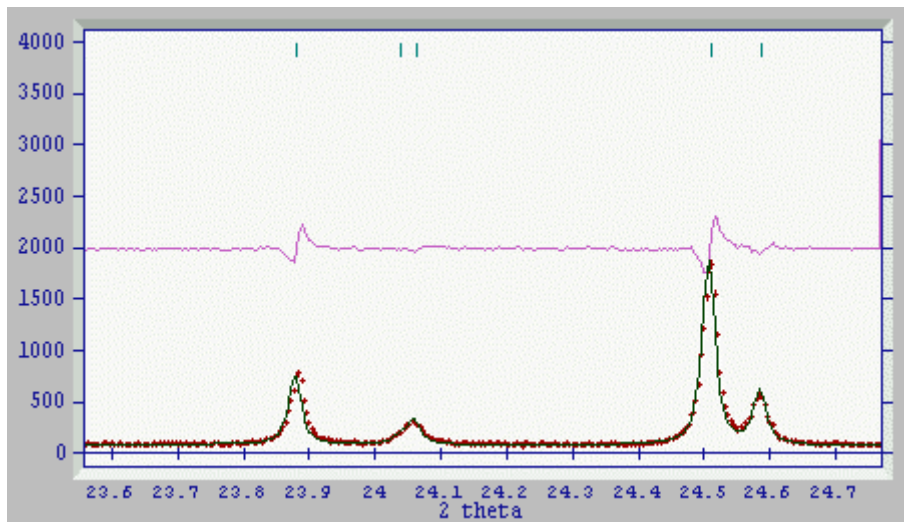
Notes

1. NB: After three peaks have been fitted, you are given the chance to go directly to profile refinement by pressing the  icon or choosing 'Pawley refinement' from the 'Mode menu'. Ignore this option for the moment.
 2. After six peaks have been fitted, DASH has sufficient information to allow a lattice parameter (4 parameters + zero point) refinement. The results of the refinement can be seen by choosing 'Peak positions' from 'View' menu. This improves the lattice constants in the majority of cases and greatly improves the starting position for the Pawley refinement.
 3. After eight peaks have been fitted, DASH has determined that the peak shape has been sufficiently well defined to allow a full Pawley refinement to be performed.
- In the initial Pawley refinement, only the terms describing the background and the terms corresponding to individual reflection intensities are refined, using the previously refined unit cell and zero point.
 - When you press Refine, 3 cycles of least squares are performed.

- This should return figures similar to the ones below

204 reflections 9751 points $R_{wp}=22.25$ $R_e=9.38$ $\chi^2=5.6$

- Press 'Accept' to accept the results of this refinement - the fit is displayed
- Now click in the main window and press 'Home' to see how well the data are fitted. The (obs minus calc) plot is shown in pink and emphasises any misfit in the data. If you look closely at the data, you are likely to see something like this

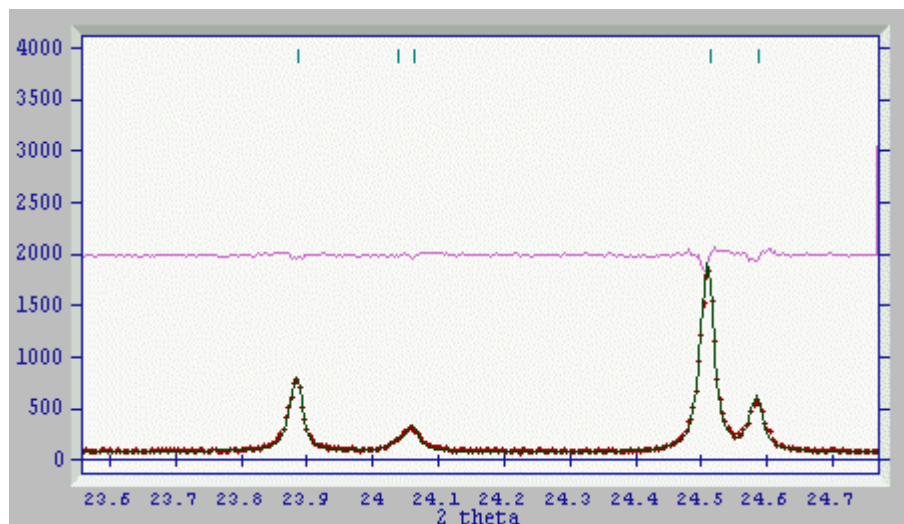


The fit is very good, but the tell-tale sinusoidal misfit indicates that the unit cell and zero point are in need of some further refinement.

- Going back to the Pawley window, note that the program has anticipated this and has flagged the unit cell and zeropoint for refinement.
- Press Refine in order to perform a Pawley refinement in which the background, intensities, unit cell and zero point are refined. The fit should improve to something like

204 reflections 9751 points $R_{wp}=16.2$ $R_e=9.38$ $\chi^2=3.0$

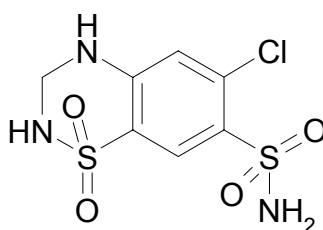
- Note that the figures of merit have improved - press 'Accept' to see the improvement in the fit



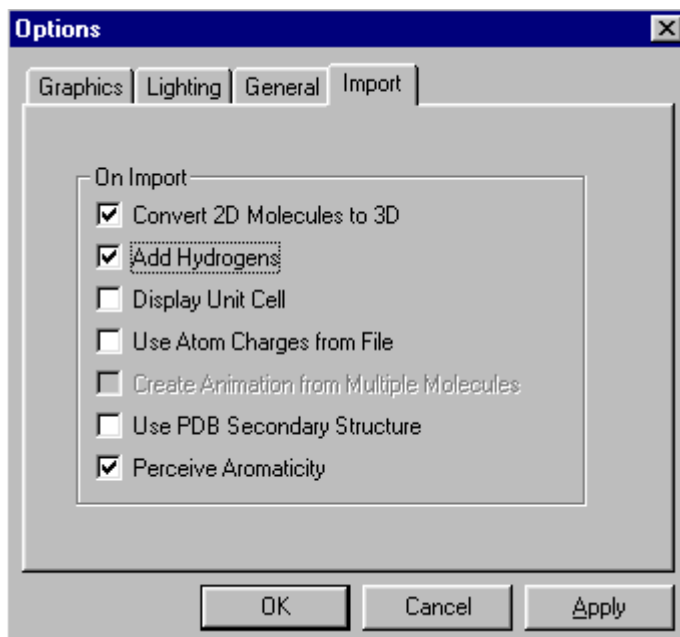
- Examine the whole profile. If you have achieved a χ^2 of around 3, the fit to the data will be excellent. Click Save to save the refinement results to disk as a DASH project (.sdi) file called 'hct20.sdi'.
- You can exit DASH for the moment

Stage 8 : Molecule construction

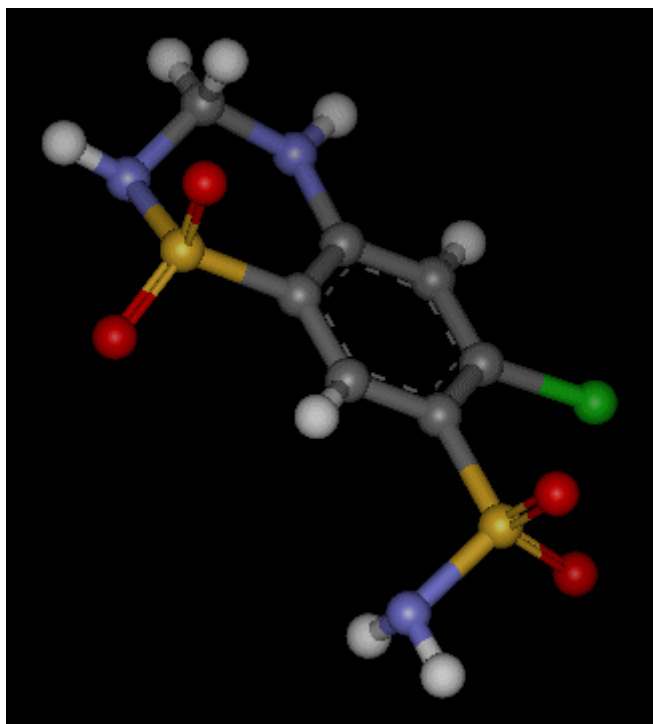
Here again is the molecule that we are trying to solve the crystal structure of



- You need to construct a 3D molecular description of the above molecule using your favourite modelling software and save it in PDB, MOL or MOL2 format.
- For the purposes of the tutorial, we'll assume that the molecule was sketched (as indeed it was) with the freely available ISIS/Draw sketching package
- Furthermore, we'll assume the 2D to 3D conversion will be performed using the widely available WebLab ViewerLite.
- Once the molecule is sketched within ISIS/Draw, select the whole molecule and copy it into memory using 'Ctrl+C'
- Within WebLab ViewerLite, ensure that the following 'Import' options (accessed from the View menu) are enabled




- Paste the 2D model into WebLab ViewerLite using 'Ctrl+V'
- Upon pasting the molecule into WebLab ViewerLite, the 2D chemical sketch is converted into a 3D molecular model, that should look like (in ball & stick display mode)



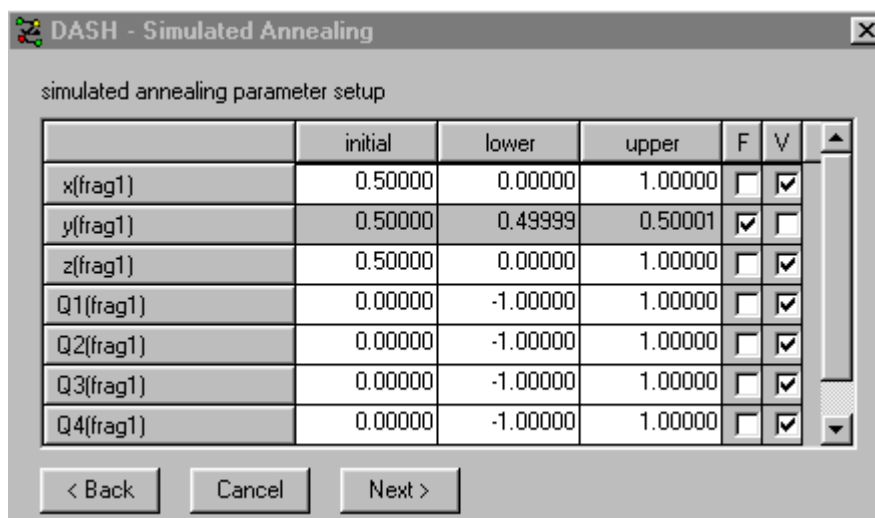
- Save the molecule coordinates to disk in MDL Mol format as 'hct.mol'.

Stage 9 : Setting up the Structure Solution Run

- Start DASH as before and select 'Simulated Annealing Structure Solution' from the Wizard
- Browse for the DASH project file that you saved at the end of Stage 7 and load the file that you saved, hct20.sdi
- Press 'Import... ' and follow the on-screen instructions to read in the hct.mol file
- DASH will tell you that it has successfully generated the internal format (Z-matrix) that it uses to describe the molecular conformation
- DASH analyses the molecule and automatically selects rotatable torsions. In this case, the bond connecting the benzene ring to the SO₂NH₂ group is the only rotatable torsion in the molecule.
- Read in the new created Z-matrix by clicking on the  icon in the Simulated Annealing window and selecting the file 'hct.zmatrix'.
- Note that DASH has determined that there are 8 variables to be determined if the crystal structure is to be solved i.e. 3 positional coordinates for the centre of mass of the molecule, 4 parameters describing the orientation of the molecule within the unit cell, and 1 internal torsion angle describing the molecular conformation.
- DASH now has the information it needs concerning the molecule, so Click Next.


The following menu gives you the chance to fix or bound parameters. In this particular example, we are allowed to fix the *y* coordinate of the centre of mass of the molecule at any position, as *P2*₁ is a polar space group.

- Do this by clicking on 'F' (short for 'fix', 'V' short for 'vary') in the line corresponding to the *y* coordinate of the molecular fragment.



- Click Next>

The Simulated anneal protocol menu that follows need not concern us here. In most cases, the default values will suffice. See the DASH Interface Guide for more details.

- Click Solve>
- You're ready to start the run. Just press the play () button.

Stage 10 : Monitoring structure solution progress

Full details of all the output from the structure solution run are given in the DASH Interface Guide. For the moment, you need only watch two things.

1. The Profile χ^2
2. The (obs - calc) plot i.e. *the difference plot, shown by default in pink*

The profile χ^2 is on the same scale as the Pawley fit profile χ^2 that you obtained in when fitting the data in Stage 7. So if the current profile χ^2 is close to the value of the Pawley profile χ^2 , you've probably solved the structure.

NB: DASH currently runs the SA process until the user intervenes by pressing one of the following buttons on the SA output panel :



Pauses until you hit play again (useful to free up the processor!)



Fast rewind, stops SA and returns you to the molecule parameter control

screen



Rewind, stops SA and returns you to the SA protocol control screen



Eject stops the SA and returns you to the sdi / zmatrix setup screen



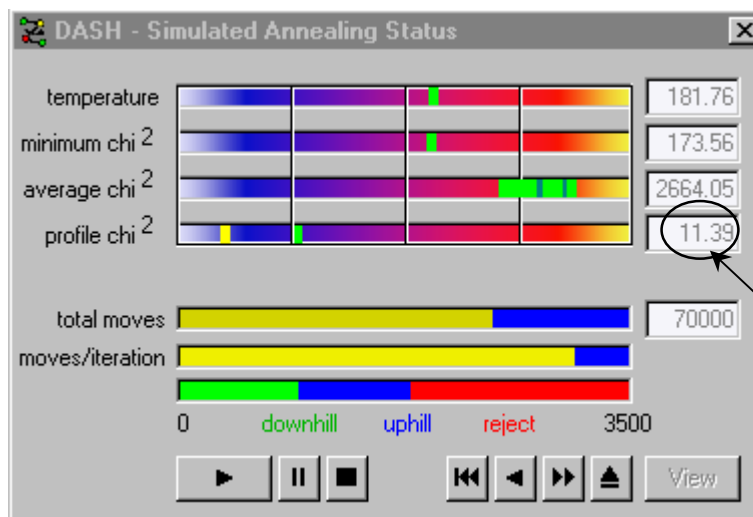
Fast forward, invokes simplex optimisation



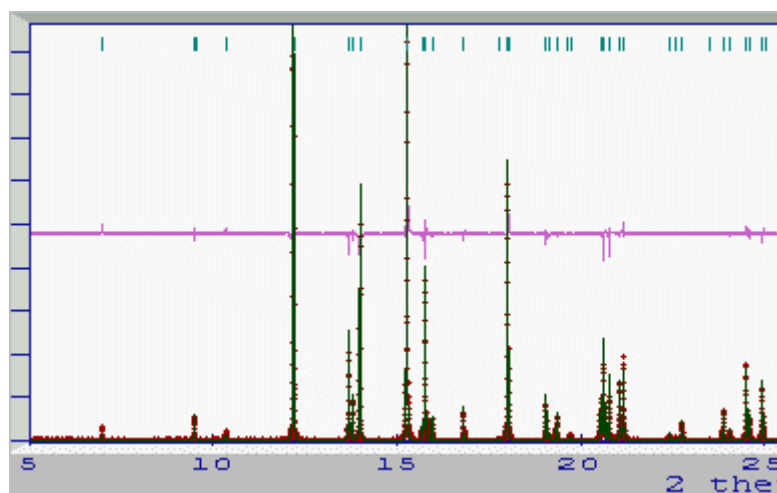
Stop, currently inactive, reserved for future use

On a modestly specified PC (e.g. Pentium III 300MHz) the structure solution process should take less than 30 seconds to reach a profile χ^2 of around 12, by which point the structure is solved to a high degree of accuracy.

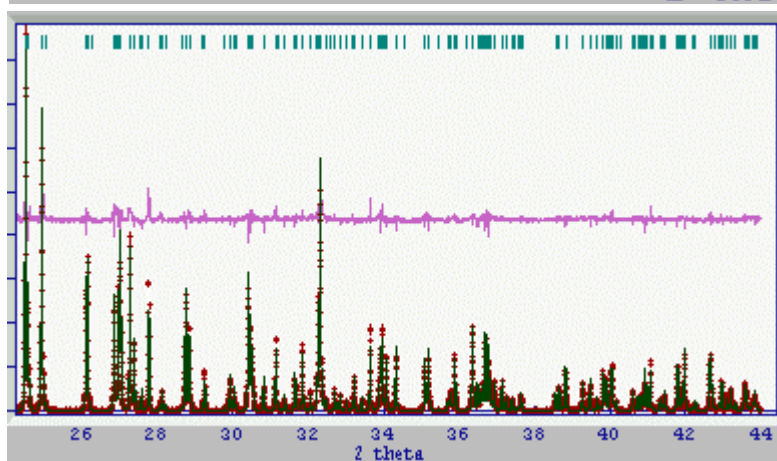
Profile χ^2 only about four times that of the Pawley profile χ^2 i.e. *certainly solved*



Examine the difference plot.



Low angle range



High angle range

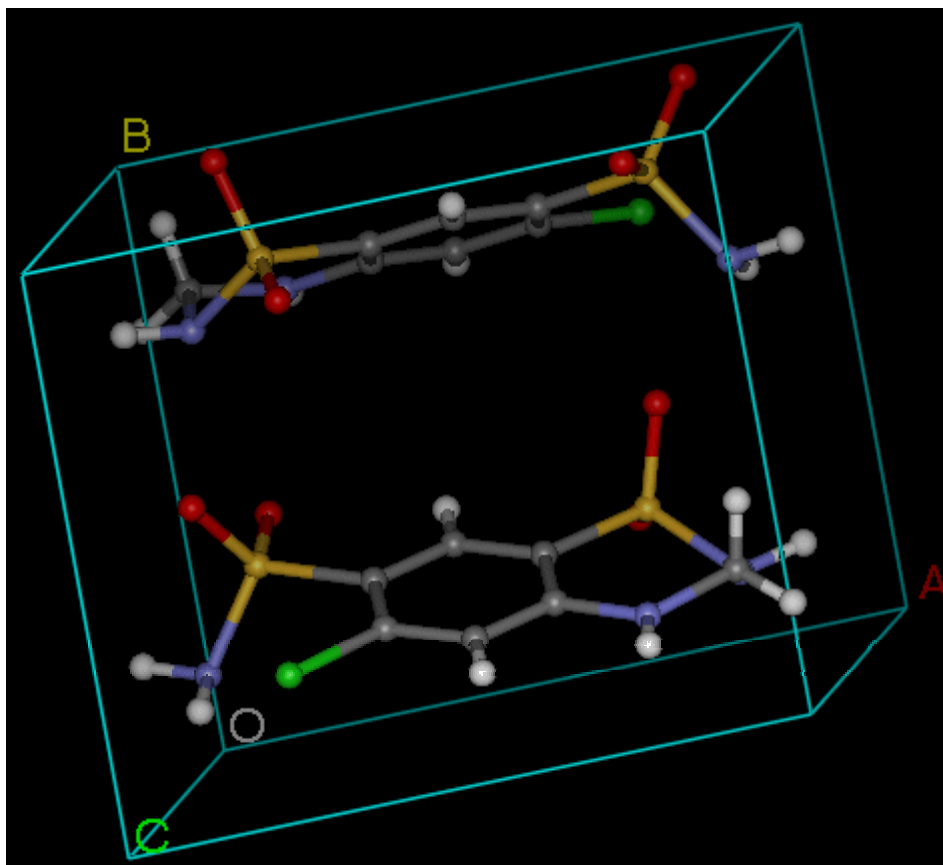
- The fit is excellent, even at high angle
- Remember also that we've effectively only refined a scale factor to get to this point! The structure is clearly solved.

Stage 11 : Examining the output structure

DASH outputs 3 coordinate files describing the final answer output from simulated annealing. Here, we assume that the project filename was hct20.sdi

1. Hct20.pdb : protein data bank format file containing a Cartesian coordinate description of the SA solution
2. Hct20.cssr : CSSR format file containing a fractional co-ordinate description of the SA solution
3. Hct20.ccl : Cambridge Crystallographic Subroutine Library format file containing a fraction coordinate description of the SA solution.

Your answer should look like this



NB: Remember that the exact location of your molecule along 'b' depends upon where you anchored the molecule. In the above picture, the molecule was fixed at 'y=0.3'. The solution obtained is in excellent agreement with that reported for hydrochlorothiazide at room temperature by Dupont & Dideberg, Acta Cryst B (1972) **28** 2348.

Troubleshooting

Please ask any of the demonstrators for help if you run into problems with DASH.

WebLab ViewerLite Version 3.20 (12/8/98) is Copyright 1998 Molecular Simulations Inc.

DICVOL91:

LOUER, D. & LOUER, M. (1972). J. APPL. CRYST. 5, 271-275.

BOULTIF, A. & LOUER, D. (1991). J. APPL. CRYST. 24, 987-993

Appendix A : Example of a completed dicvol91 input file for hydrochlorothiazide

```
*** HCT ***
24 2 0 0 0 1 1 0
0. 0. 0. 0. 0. 0. 0.
1.1294 0. 0. 0.
0. 0. 0.
6.9822
9.4942
10.3453
12.1847
12.2228
13.6925
13.7905
14.0003
15.2696
15.6883
15.7753
15.9581
16.8146
17.7552
18.0107
19.0501
19.1452
19.3479
19.7249
20.5468
20.6314
20.7735
21.0639
21.1688
```

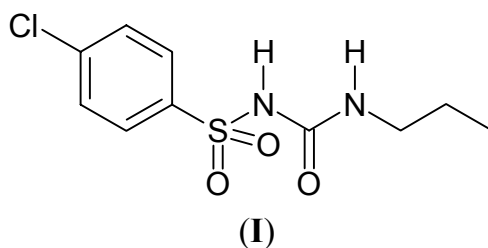

DASH Tutorial 2¹

Harriott Nowell & Kenneth Shankland

*ISIS Facility
Rutherford Appleton Laboratory
Chilton, Didcot
Oxon OX11 0QX
United Kingdom*

Introduction

The object of this tutorial is to guide you through the structure solution of chlorpropamide (**I**). It assumes that you are already familiar with DASH and in particular with Tutorial 1. The stages in this tutorial correspond exactly to the stages in Tutorial 1, so it is always possible to refer back to Tutorial 1 at any time for a more detailed description. In it, you will learn how to (a) handle structures that are more flexible than hydrochlorothiazide, (b) solve a structure from a low resolution data set, and (c) see one of the potential pitfalls of global optimisation i.e. local minima.



Data

The data set “cp.xye” is synchrotron X-ray diffraction data collected on BM16 at the European Synchrotron Radiation Facility, $\lambda = 0.800077\text{\AA}$.

Stage 1: Reading the data

Open DASH and select the directory in which the chlorpropamide data resides. Select the ‘View data / determine peak positions’ option, then select the “cp.xye” data file using the ‘Browse’ button.

¹ Copyright CCLRC / CCDC

Stage 2: Examining the data

The data spans -4 to $22^\circ 2\theta$. Why are there two theta values less than zero? Of course, there is no reason that data should not be collected on both sides of the beam stop! In this particular case, it is a function of the multiple crystal-analyser detector used on station BM16. This data adds nothing to the structure solution process and so it is necessary to edit the data file and remove points with 2θ values less than zero. Since there are no Bragg peaks apparent in the positive 2θ data until approximately $3.4^\circ 2\theta$, it is possible to reduce the data further still. Here we will use data in the range 2° - $22^\circ 2\theta$. To make life simple, we've provided a file, "cp_2.xye" that contains this range and is ready to use. If you want to see the procedure for creating it, see the short section at the end of the tutorial.

Reopen DASH, remembering to select the directory in which the chlorpropamide data resides, select the 'View data / determine peak positions' option, then select the "cp_2.xye" data file using the 'Browse' button. Note that this data set was collected quickly at the end of a day's beamtime, and so only extends to $22^\circ 2\theta$. Hence the data set extends to a resolution of only $\sim 2\text{\AA}$.

Stage 3. Fitting the peaks to determine the exact peak positions

Select the first twenty peaks using the method described in Tutorial 1, Stage 3.

Here is a guide to the positions ($^\circ 2\theta$) of the first twenty peaks:

3.4383	6.1080	6.8792	8.5344	8.9466
9.4316	9.9800	10.1033	10.2499	10.3269
10.7041	11.1635	11.3767	11.5027	12.2579
12.3053	13.3092	13.4047	13.5143	13.5696

Stage 4. Indexing

Copy the twenty peak positions from DASH, using the 'Peak positions' option in the 'View' menu, and paste into a file with the correct format for your favourite autoindexing program, such as Dicvol.

Your indexing program may reveal a number of possible unit cells. The unit cell with the highest figures of merit should be orthorhombic with volume $\sim 1266\text{\AA}^3$. Dicvol, for example, returns an orthorhombic cell with $a = 26.66826\text{\AA}$, $b = 9.08435\text{\AA}$, $c = 5.22571\text{\AA}$ and volume = 1265.999\AA^3 with figures of merit $M(20)=107.1$ and $F(20)=506.6$

Closer inspection of the other unit cells that are suggested by the indexing program will reveal that many of them are slight monoclinic distortions of the above unit cell, with almost identical volumes and lattice parameters and $\beta \approx 90^\circ$. Other suggestions generally have much lower figures of merit and can be ruled out immediately.

Considering that the orthorhombic unit cell has the best figures of merit, and that it is usually best to try the simplest option first, we will go ahead to the next stage assuming an orthorhombic unit cell, with the lattice parameters given above.

Stage 5. Stop and think

Does the cell make sense? In this case we estimate the molecular volume to be $\sim 290 \text{ \AA}^3$, from the fact that there are 17 non-Hydrogen atoms in the molecule. Therefore, given the unit cell volume of $\sim 1266 \text{ \AA}^3$ we know from this very rough approximation that the cell is most likely to accommodate 4 molecules. At this point, your knowledge of space group frequencies should suggest to you that $P2_12_12_1$ is a strong possibility.

Stage 6. Checking the cell and determining the space group

Reopen DASH and select the 'Preparation for Pawley refinement' option. Enter the lattice constants, the space group P222 will automatically be selected. Go to the next step of the wizard and select the "cp_2.xye" file, synchrotron radiation and enter 0.800077 \AA for the wavelength.

Close the wizard and ensure that the tick marks generally correspond to peaks in the diffraction pattern. The presence of some excess tick marks indicates probable systematic absences, this means that a space group of higher symmetry might be more appropriate. Go to 'Crystal symmetry' in the 'View' menu and scroll through some of the possible space groups. You will see that some of the space groups can be ruled out immediately; for example, face centred and body centred lattices leave some peaks unaccounted for. Many of the primitive lattice space groups appear likely from the tick mark positions. In this situation, where more than one possible space group exists, it is logical to begin with the most frequently occurring space group (a table of frequency of occurrence of space groups is given in the DASH manual). In this case, the most frequently occurring orthorhombic space group is $P2_12_12_1$, so select this (number 19), confirm visually that it matches the data and click on OK.

Stage 7. Extracting intensities

Choose 7 isolated peaks from across the pattern. Fit these peaks using the method described in Tutorial 1 Stage 7, then carry out the Pawley refinement. The initial 3 cycles of least squares refinement only involve the terms corresponding to the background and to the individual reflection intensities, accept these three cycles. The next 5 cycles of least squares refinement involve the terms describing background, intensities, unit cell and zero point. These refinement details will be suggested automatically by DASH.

When these cycles are complete check the difference line; this should be almost flat by this point. The final Pawley χ^2 should be between about 3 and 4.

Accept this Pawley fit and save it as “cp.sdi”.

Stage 8 : Molecule construction

Construct a 3D molecular description of the molecule using your favourite modelling software and save it in PDB, MOL or MOL2 format. This can be done, for example, by importing an ISIS/Draw sketch into WebLab ViewerLite. For further details, see Tutorial 1 Stage 8. Save this as cp.pdb, cp.mol or cp.mol2.

Stage 9 : Setting up the Structure Solution Run

- Reopen DASH and select the 'Simulated Annealing Structure Solution' option.
- Select the “cp.sdi” file
- Press the Import button and select cp.pdb, cp.mol or cp.mol2 (the file that you created in Stage 8); a Z-matrix file called cp.zmatrix will be generated automatically.
- Read in the cp.zmatrix file.

Note that as $Z = 4$ for $P2_12_12_1$, it follows that $Z' = 1$ because we know from Stage 5 that the cell is most likely to accommodate 4 molecules. Therefore, only one Z-matrix needs to be read in.

At this point, DASH will confirm that there are 13 variable parameters. These parameters are listed when you click on ‘Next’. There are 3 parameters describing the positional coordinates, 4 describing the molecular orientation within the unit cell and 6 variable torsion angles. All ‘v’ (short for vary) boxes are ticked by default, indicating that all 13 parameters are allowed to vary during the structure solution. Click ‘Next’, then ‘Solve’, then the ‘Play’ button to begin

the simulated annealing. *NB: Keen chemists should resist the urge to restrict the torsional rotations pertaining to the two bonds around the carbonyl group! The uses of, and advantages of, restricted rotations in relation to this urea type linkage are discussed in Tutorial 3.*

Stage 10 : Monitoring structure solution progress

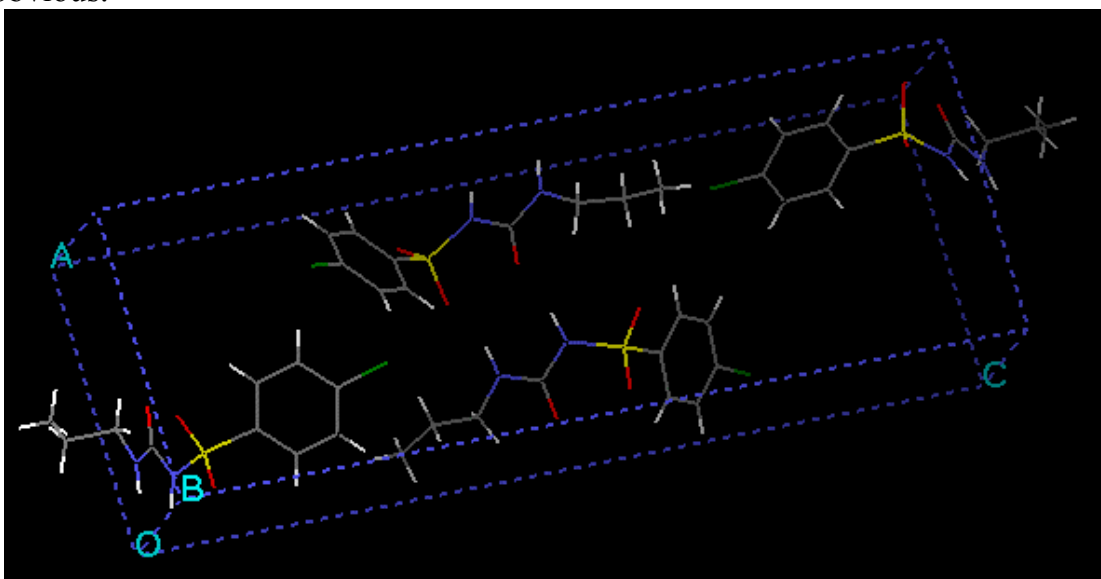
The progress of the structure solution can be followed by monitoring the profile χ^2 and the difference plot. Remember that the Simplex button ('fast forward' button) can be pressed at any time to accelerate the search in the vicinity of the current minimum.

Once a profile χ^2 of approximately 10 or less is reached, you can be sure that a very good structure has been found, as this value is only ~ 3 times the Pawley χ^2 value. Finalise the solution by pressing the Simplex button and accepting the answer.

If your final profile χ^2 is a bit higher than 10, you are clearly close and perhaps only a single atom at the end of the chain is slightly misplaced. Take a close look at the output structure and read the section below.

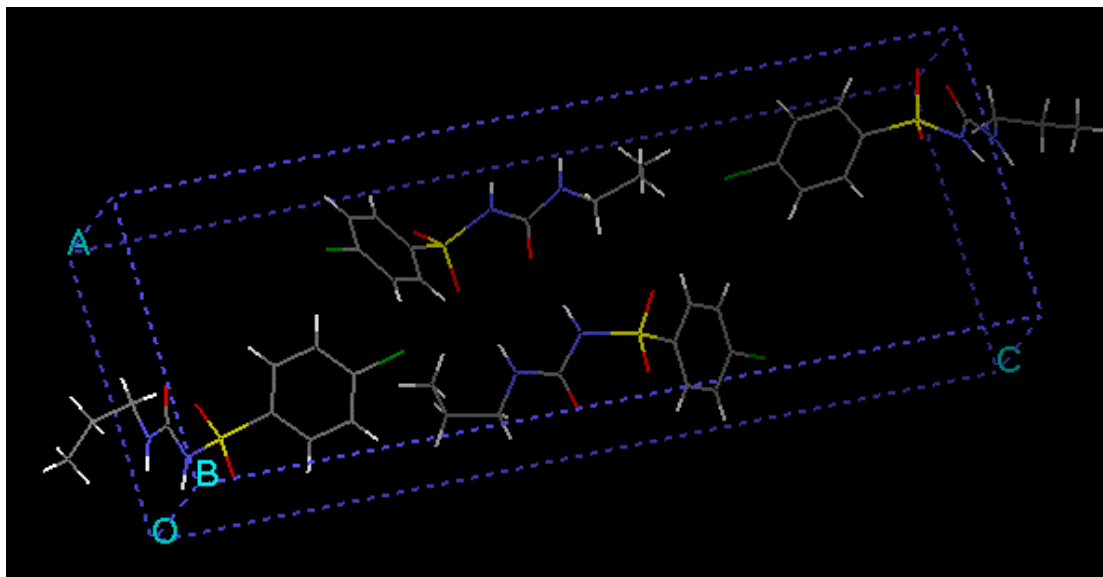
Stage 11 : Examining the output structure

View the molecule and the unit cell using the .pdb file that has been created by DASH. The structure should be chemically reasonable in terms of molecular conformation and intermolecular distances. The potential for H-bonding is obvious.



Unrefined DASH solution with profile $\chi^2 < 10$. Note the potential for hydrogen bonding between the symmetry related molecules.

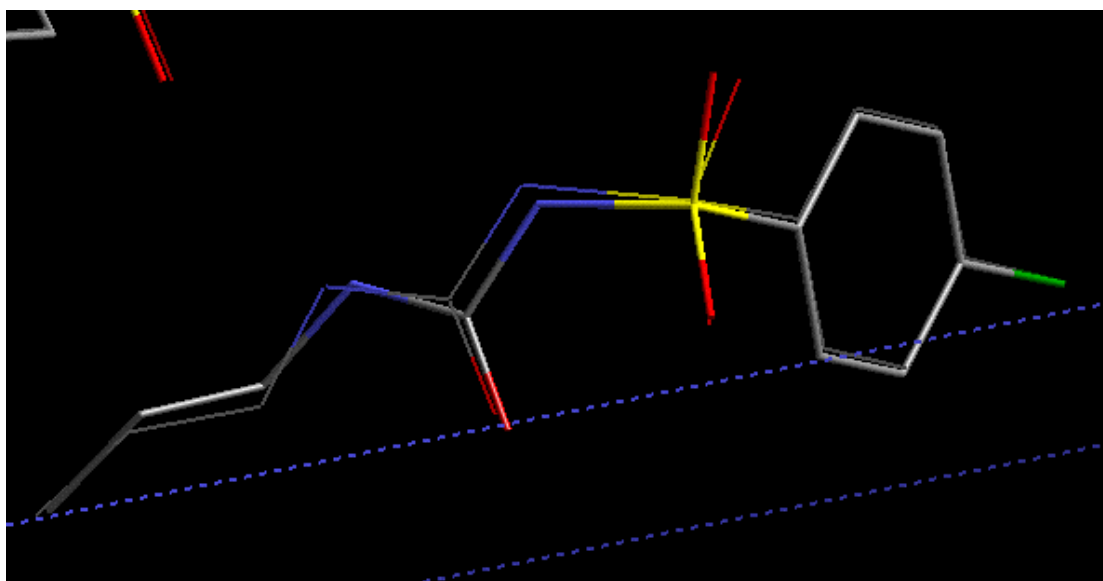
However, it is entirely possible that you have obtained a structure solution with a χ^2 value close to, but not less than 10. Is it correct? Consider the output structure below, taken from a solution with a final profile χ^2 of approximately 10.7.



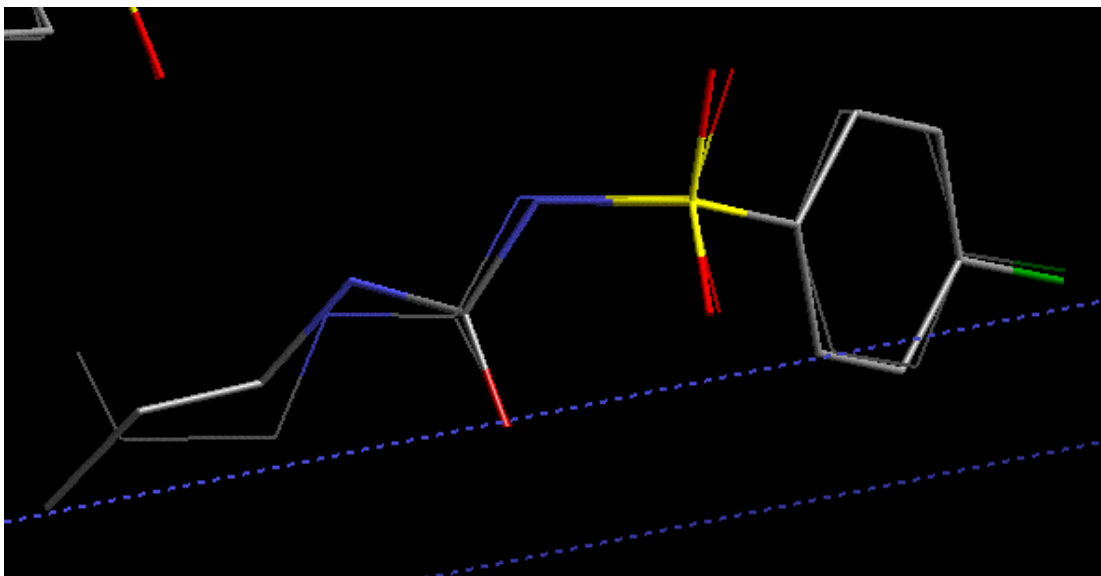
Spot the difference – an unrefined DASH solution with profile $\chi^2 = 10.7$.

In this case, it is a structure that differs *only slightly* from the correct structure, giving rise to a local minimum with a profile χ^2 slightly higher than that of the correct crystal structure.

The subtle differences become clearer when both initial structures are overlaid with a molecule of chlorpropamide determined from a single crystal experiment .



Unrefined DASH solution with $\chi^2 < 10$ (stick), overlaid upon a single crystal solution (cylinder). The agreement is excellent.



Unrefined DASH solution with $\chi^2 > 10$ (stick), overlaid upon a single crystal solution (cylinder). The agreement is still good and the differences could be overcome in a Rietveld refinement, but nevertheless, you can do better.

Stage 12 : Take home message

Global optimisation processes may locate local minima, particularly if (a) the molecule under study is highly flexible (b) $Z' > 1$ or (c) the data are of limited resolution. Looking at the above example of a false minimum, it is clear that superficially, it can look chemically plausible. This is hardly surprising, as it lies at a point on the χ^2 hypersurface very close to the global minimum of the crystal structure. Accordingly, it is always prudent to run a structure solution multiple times (with different random numbers of course...) to ensure that a consistent minimum has been reached.

Troubleshooting

Please ask any of the demonstrators for help if you run into problems with DASH.

WebLab ViewerLite Version 3.20 (12/8/98) is Copyright 1998 Molecular Simulations Inc.

DICVOL91: LOUER, D. & LOUER, M. (1972). J. APPL. CRYST. 5, 271-275.
BOULTIF, A. & LOUER, D. (1991). J. APPL. CRYST. 24, 987-993

Procedure for editing a data file

First, create a copy of the “cp.xye” file. Open this copy in an ASCII file editor, such as Wordpad, delete the data between -4 and $1.998^\circ 2\theta$ and save the file as “cp_2.xye”. Remember that column 1 in the data file corresponds to the 2θ value, column 2 to the diffracted intensity and column 3 to the estimated standard deviation of the intensity.

Combined Method for "*Ab Initio*" Structure Solution from Powder Diffraction Data: The ENDEAVOUR^{1,2} Software

H. Putz

Crystal Impact

1. Introduction

The determination of the atomic structure of crystalline solids from powder diffraction data^{3,4,5} generally consists of 6 steps:

1. Determination of peak positions from the raw diffraction data
2. Indexing (calculation of lattice parameters from peak positions)
3. Extraction of intensities (Le Bail or Pawley method)
4. Space group determination (if possible)
5. Structure solution (creation of a structural model with approximate atomic positions)
6. Rietveld⁶ refinement (Refinement of the atomic positions)

The steps 1, 3, 4 and 6 are fairly routine nowadays with a number of computer programs available. The main problems associated with structure determination from powder frequently arise at points 2 and 5. Despite of a large variety of

¹ H. Putz, J.C. Schön, M. Jansen, *J. Appl. Cryst.* **32**, 864 (1999).

² Crystal Impact, Endeavour 1.0, Internet: <http://www.crystalimpact.com/endeavour>, Email: info@crystalimpact.com, Bonn 2000.

³ K.D.M. Harris, M. Tremayne, *Chem. Mater.* **8**, 2554 (1996).

⁴ D. Louër, *Acta Cryst. A* **54**, 922 (1998).

⁵ A. Meden, *Croatica Chemica Acta* **71**(3), 615 (1998).

⁶ H.M. Rietveld, *J. Appl. Cryst.* **2**, 65 (1969).

available indexing programs, this task may nevertheless be quite challenging (especially with low quality data and/or in the presence of impurities). Though this is preliminary to structure solution, this exercise will focus on point 5, the actual structure solution step, assuming that the unit cell parameters are already known.

Why is the structure solution step actually necessary ? One of the main problems associated with Rietveld's method is its implicit use of a **local** optimization of the atomic positions. Thus, it depends on the availability of a reasonable structural model as a starting point, i.e. an approximate idea of what the atomic arrangement should be.

Providing such a model from a powder diffraction pattern remains a difficult task in general. In some cases, the analogy to existing compounds and their structures allows a relatively simple construction of a model, but in many other cases this is not possible.

With the more or less routine application of so-called "direct methods" in the structure solution for single crystals in mind one may wonder why this problem is so severe with powder diffraction data. One of the basic differences between single crystal- and powder diffraction is the projection of the 3-dimensional diffraction information using single crystals into just one dimension when dealing with powders. Because of this, reflections which belong to more or less the same d-spacing in the crystal(s) by accident or because of symmetry may overlap. Hence, in many cases it is no longer possible to determine their intensities with sufficient accuracy, so that direct methods (which are based on the knowledge of the structure factor or intensity for each reflection) may become difficult to apply.

In the past 10 years or so, the so-called "direct-space methods" have been developed. The basic advantage is that they propose a structural model independent of the powder diffraction diagram^{3,4,7,8,9}. These models are subsequently optimized with respect to the difference of the calculated and the experimental powder pattern concerning $I(2\theta)$. These values are readily obtained from experimental powder diffraction data of sufficient quality.

In general, direct-space methods work as follows: Beginning with some arbitrarily (e.g. random) chosen starting configuration, the difference between the calculated and the measured diffraction pattern is minimized through repeated change of the atomic arrangement while the unit cell is kept fixed.

If no other constraints are introduced, this straightforward prescription (also

⁷ Y.G. Andreev, G.S. MacGlashan, P.G. Bruce, *Phys. Rev. B* **55**(18), 12011 (1997).

⁸ K.D.M. Harris, R.L. Johnston, B.M. Kariuki, *Acta Cryst.* **A54**, 632 (1998).

⁹ K. Shankland, W.I.F. David, T. Csoka, *Z. Krist.* **212**, 550 (1997).

known as the "Reverse Monte Carlo"-method^{10,11}) is in practice limited by the fact that for most crystalline solids the system is quickly getting trapped in some minimum that does not correspond to a physically reasonable atomic arrangement. The reason for this lies in the landscape of the cost function which includes numerous deep local minima¹¹, again a direct consequence of the loss of information in the diffraction experiment: If this loss is so severe that there are more parameters than observations, the problem becomes ambiguous, i.e. there may be a large variety of global minima. Besides this, even minor inaccuracies in intensity values can give rise to the phenomenon that the correct structure solution does no longer belong to the global but just to a deep local minimum of the hypersurface of the cost function.

In order to avoid these traps, certain additional constraints are required which restrict the available configuration space to physically reasonable atomic arrangements. Fortunately, the physically reasonable region of configuration space has been the subject of investigation in the field of crystal structure prediction^{12,13}. Nearly all of these methods use optimization of the atomic arrangement with regard to elaborate cost functions (mostly potential energy, but also deviation of crystal chemistry rules and others) in order to detect the global minimum and hence the crystal structure of the system.

Unfortunately, most methods for structure prediction still suffer from the lack of really adequate cost functions. Nevertheless, they might be of use even at their current state concerning the problem of crystal structure determination from powder diffraction data, if these data and the knowledge of the unit cell and its content are regarded as additional information for the global optimization.

2. Method

2.1 Concept

In order to avoid the traps mentioned above, our concept for crystal structure solution from powder patterns consists of a combined global optimization ("Pareto¹⁴ optimization") of the difference between the calculated and the

¹⁰ R. Kaplow, T.A. Rowe, B.L. Averbach, *Phys. Rev.* **168**, 1068 (1968).

¹¹ R.L. McGreevy, "Reverse Monte Carlo Methods for Structural Modelling", in: *Computer Modelling in Inorganic Crystallography*, ed. C.R.A. Catlow, Academic Press, San Diego 1997.

¹² C.R.A. Catlow, R.G. Bell, J.D. Gale, *J. Mater. Chem.* **4**(6), 781 (1994).

¹³ J.C. Schön, M. Jansen, *Angew. Chem. Int. Ed. Engl.* **35**, 1286 (1996).

¹⁴ The idea of a combined optimization of two (or even more) cost functions ("multi-objective" or "multicriteria" optimization) was first developed by V. Pareto in the field of economics in the late 19th century.

measured diffraction pattern and of the potential energy of the system. Size, shape and content of the unit cell are assumed to be known from experiment.

"Merging" both hypersurfaces¹⁵ weakens or even eliminates the minima that belong to only one of the two surfaces, and strengthens those which belong to both of them. Therefore, a sufficiently long global optimization run should sooner or later reach the global minimum of the system corresponding to the correct crystal structure.

The combined cost function C is calculated as

$$C = \alpha \cdot R_B + (1 - \alpha) \cdot E_{\text{pot}}$$

where R_B denotes the so-called R-value³ frequently used for the comparison of calculated and experimental diffraction patterns and E_{pot} is the potential energy of the atomic arrangement.

α is the Pareto parameter which weights the contributions of the two parts of the cost function. The best choice for α may vary from problem to problem. This parameter can be used to weaken the influence of either a low quality diffraction pattern or a very crude potential on the overall cost function. Thus, a powder pattern where the intensity values are suspected to be not very accurate may be balanced by a good potential by choosing a smaller α -value, and vice versa. Nevertheless, a value of $\alpha = 0.5$ has shown to be a good starting point for most systems.

In fact, this combination of structure prediction and powder diffraction analysis has both synergetic and complementary aspects, depending on the particular problem: If both an accurate potential function and a high quality powder diffraction pattern are available, the optimization is very fast and yields the correct crystal structure with high probability.

However, if only a very approximate potential is available, this can be compensated for to a large degree by a high precision powder diffraction diagram. One might then use the potential only for ensuring reasonable distances between the various atom types, thus enforcing the optimization according to the difference in powder diffraction patterns.

If, on the other hand, only few and not very reliable diffraction data are available for the compound under investigation, a high quality potential can nevertheless lead to reasonable structures. Here, the effect of the R_B -Term lies in the enhancement of one of the many structures that constitute local minima of the potential energy.

¹⁵ Both cost functions depend on all atomic coordinates in the fixed unit cell.

One should note that this method can also be of use in cases where a high crystal symmetry leads to very few diffraction peaks and thus to few parameters, limiting the complexity of the atomic arrangement that can be solved by conventional methods. The additional information contained in the potential can help to solve even complex highly symmetric crystal structures from their powder diffraction pattern.

2.2 Simulated Annealing

As has been mentioned above, the atomic arrangement representing the solved crystal structure is supposed to correspond to the global minimum of the multidimensional hypersurface of the cost function. Hence, a global optimization method operating on the atomic coordinates should be able to solve the structure. One of the most common algorithms for global optimization is the so-called "simulated annealing"¹⁶. The great advantage of this method lies in the relative ease of implementation and the very general applicability, more or less regardless of the specific optimization problem. Simulated annealing is a Monte Carlo method based on the Metropolis algorithm¹⁷ which implements a weighted random walk through configuration space.

Starting from a current configuration i , a neighbouring configuration $i+1$ is chosen at random according to a set of rules ("moveclass"). If the cost function C_{i+1} is below or equal to C_i , the move is always accepted, i.e. $i+1$ becomes the new current configuration. Otherwise, the move is only accepted with probability $e^{-\frac{C_{i+1}-C_i}{K}}$, where K is a control parameter of the random walk. Thus, during a sequence of such MC steps, the system can climb over barriers of the hypersurface of the cost function, depending on the control parameter K . In analogy to the parameter "temperature" in the annealing of a real material, starting from a relatively high value the control parameter is slowly reduced according to a certain schedule (the so-called "temperature program"), until the system ends up in a deep-lying minimum. It can be shown^{18,19} that an ergodic system will reach the global minimum during such a simulated annealing run for $t \rightarrow \infty$.

¹⁶ S. Kirkpatrick, C. Gelatt, M.P. Vecchi, Science **220**, 671 (1983).

¹⁷ N. Metropolis, A. Rosenbluth, M. Rosenbluth, A. Teller, E. Teller, J. Chem. Phys. **21**, 1087 (1953).

¹⁸ P.J.M. van Laarhoven, E.H.L. Aarts, "Simulated Annealing", Reidel, Dordrecht 1987.

¹⁹ P.J.M. van Laarhoven, PhD-thesis, De Erasmus Universiteit Rotterdam 1988.

2.3 Potential energy calculation

In principle any method for the calculation of the potential energy E_{pot} can be employed, but the global optimization usually makes ab initio energy calculations not feasible. With regard to those test structures which can more or less be considered as ionic compounds, a simple empirically parametrized two-body potential is chosen:

$$E_{\text{pot}} = \frac{1}{2} \sum_{i,j} \frac{e^2}{4\pi\epsilon_0} \cdot \frac{q_i \cdot q_j}{d_{ij}} + \left(\frac{d_{ij}^{\text{min}}}{d_{ij}} \right)^{12}$$

The Coulomb interactions resulting from atomic charges q_i are calculated using the Ewald-sum. d_{ij} is the distance between the two atoms no. i and j , and d_{ij}^{min} is a parameter indicating the minimum observed distance between the two atom types. These values may be readily obtained from the ICSD^{20,21} or even common chemical knowledge, thus leading to a pronounced simple and transparent parametrization of the potential.

Alternatively, better potentials like Lennard-Jones may be used in addition to the Coulomb interactions, provided that adequate potential parameters are available for the compound under investigation. Missing parameters can be obtained by fitting to known crystal structures of similar compounds using the **ENDEAVOUR** builtin fitting tool or e.g. the well-known crystal lattice code GULP²². A better and more accurate potential function makes the solution of the crystal structure easier for compounds for which only relatively raw diffraction data are available.

For most intermetallic compounds, a simple repulsion function parametrized with minimum interatomic distances d_{ij}^{min} can be employed for the purpose of structure solution, too.

$$E_{ij}^{\text{rep}} = \begin{cases} 0 & : d_{ij} \geq d_{ij}^{\text{min}} \\ \left(\frac{d_{ij}^{\text{min}}}{d_{ij}} \right)^6 & : d_{ij} \leq d_{ij}^{\text{min}} \end{cases}$$

This simple "potential" may also be used as a "penalty" type function prohibiting unphysically short atom-atom-distances for any chemical system if no adequate

²⁰ G. Bergerhoff, R. Hundt, R. Sievers, I.D. Brown, J. Chem. Inf. Comput. Sci. 23, 66 (1983).

²¹ K. Brandenburg, masters thesis, University of Bonn, Bonn 1989.

²² J.D. Gale, GULP (General Utility Lattice Program), Royal Institution of Great Britain, 1993.

potential is known, requiring, however, that a relatively high quality diffraction pattern is available.

3. The Program ENDEAVOUR

3.1 Implementation

The method described in the preceding section has been implemented in our program **ENDEAVOUR**, which is dedicated especially to the solution of inorganic crystal structures. The basic goal of the implementation was to provide an easy-to-use tool for structure solution from powder data, even for non-experts in this field. Hence, the program provides an intuitive Windows user interface with a large variety of structure visualization capabilities. Most parameters of the structure solution process must not be set by the user. Instead, their values have been optimized in extensive test calculations or are determined automatically during the structure solution calculation. The input of the necessary data is guided by a so-called "structure solution wizard" which asks for the data step by step.

Even the input of the parameters needed for the calculation of the potential energy is made as easy as possible: The input of the minimum interatomic distances needed by the default simple repulsion potential is guided by distance histograms obtained from the ICSD database (Fig. 1). Parameters for the Lennard-Jones potential which may be used alternatively can be generated using the built-in fitting tool (Fig. 2).

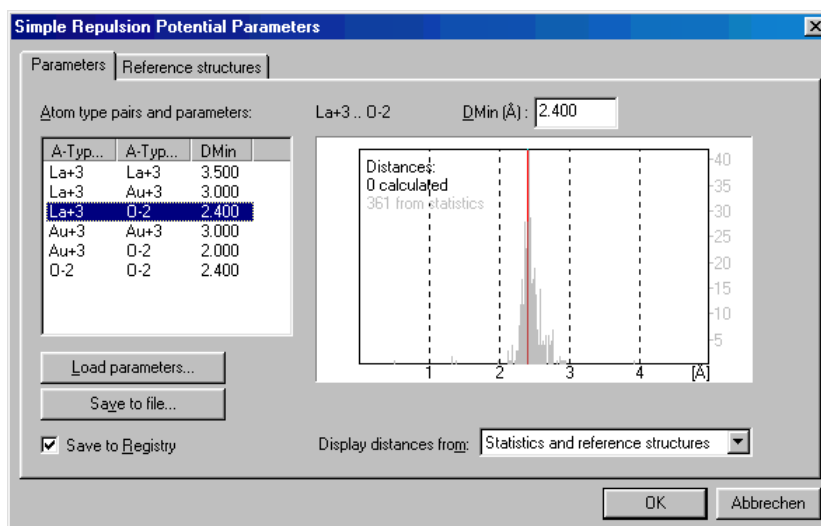


Fig. 1: Input dialog for minimum interatomic distances which are used as parameters in the simple repulsion potential

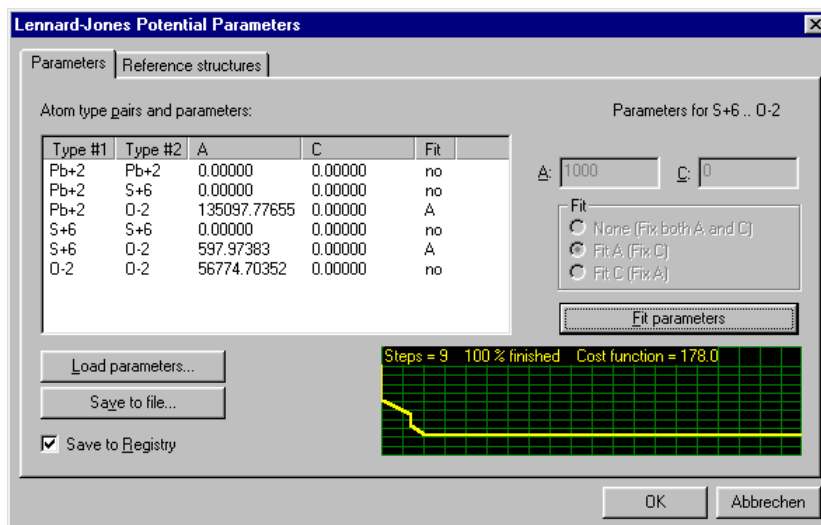


Fig. 2: Integrated fitting tool for Lennard-Jones potential parameters

The current version of **ENDEAVOUR** does not yet provide tools for the handling of raw diffraction data. Hence, the following steps have to be performed prior to the usage of the program: The first step when beginning with the raw data set obtained from the diffractometer is to locate the peaks and extract their intensities as a function of the diffraction angle 2θ , followed by a correction of the zero-point. Afterwards, this peaklist is submitted to an indexing program in order to obtain the unit cell parameters. All this can generally be done using the standard software provided with the powder diffractometers. Besides this, the composition of the compound under investigation has to be known, e.g. from X-ray fluorescence spectroscopy.

The following list shows the minimum required data to be present before Endeavour can be used:

- Peak list file (integrated intensities at the Bragg angles 2θ or $|F(hkl)|$, no profile step-scan data!)
- Unit cell parameters
- Composition

Besides this, it can be useful if the space group is known from systematic absences, however, this is not obligatory. Although the determination of the space group from the systematic absences of certain reflections in the indexed powder diffraction pattern is often not overly difficult when dealing with crystals of organic molecules, this is frequently not the case when the structures of inorganic compounds have to be solved.

Hence, though **ENDEAVOUR** can of course use a given space group if it can be determined, it is nevertheless perfectly possible to solve small-medium sized structures (up to 40-50 atoms/unit cell) in triclinic crystal symmetry (space group P1). Once a reasonable structural model has been found, the correct space group can be determined using the **ENDEAVOUR**-builtin symmetry finder SFND²³/RGS²⁴.

Finally, the structural data can be exported in a large variety of data formats. Afterwards, the resulting model should be ready for a successful converging Rietveld refinement yielding the crystal structure.

The structure solution itself may last between a few seconds for small structures like rutile to a couple of days for large unit cells containing many dozens of independent atoms. During the structure solution calculation (and of course also once it has finished), the structural model can be visualized (automatically or manually) using the integrated **DIAMOND**^{25,26,27} visualisation technology.

Our experience has shown, however, that certain basic requirements have to be fulfilled for a successive application of **ENDEAVOUR**:

- Though conventional laboratory X-ray diffraction data are generally sufficient, the quality of the powder pattern has nevertheless to be high enough so that enough accurate peak positions are available to allow a reliable indexation resulting in the correct unit cell.
- The peaks should not be too broad, in order to allow the extraction of fairly accurate intensities at the peak positions.
- The content of the unit cell should be known accurately.
- The approximate description of the structure of the compound under investigation must be possible using a simple empirically parametrized potential, i.e. the desired structure should exhibit a low potential energy, preferably constituting a deep minimum of the energy landscape.

If these conditions are not fulfilled, we have found that in some cases a diffraction pattern or potential of very poor quality can block the optimization process. However, we expect that improvements especially in both the accuracy and the general applicability of the potential energy calculation methods will

²³ R. Hundt, J.C. Schön, A. Hannemann, M. Jansen, *J. Appl. Cryst.* **32**, 413 (1999).

²⁴ A. Hannemann, R. Hundt, J.C. Schön, M. Jansen, *J. Appl. Cryst.* **31**, 922 (1998).

²⁵ G. Bergerhoff, M. Berndt, K. Brandenburg, *J. Res. Natl. Inst. Stand. Technol.* **101**, 221 (1996).

²⁶ W.T. Pennington, *J. Appl. Cryst.* **32**, 1028 (1999).


²⁷ Crystal Impact, Internet: <http://www.crystalimpact.com/diamond>, Email: info@crystalimpact.com, Bonn 2000.

make more problems amenable to solution.

3.2 Sample Calculation²⁸: Structure Solution for RuS₂

Let's start with a simple example so that you can draw your attention to the usage of **ENDEAVOUR** rather than to the actual structure solution problem. Suppose you have made a new compound; X-ray fluorescence measurements show that its sum formula is RuS₂. The X-ray powder diffraction pattern shows a few sharp lines; indexation indicates a cubic unit cell with $a = 5.6095 \text{ \AA}$. With regard to the many overlapping reflections you decide not to commit yourself to a certain space group at this early stage. Supposing that you would not be able to generate a structural model from scratch, you decide to use **ENDEAVOUR**.

The first thing you have to do is prepare the file that contains the diffraction data: The current version of **ENDEAVOUR** uses the integrated intensity values at the peak positions (not the profile). These data are generally provided by the software of your diffractometer. Reformat the corresponding file such that each line contains one peak, first the 2θ , then the intensity value. For your convenience, we have prepared this file "rus2.dif" in the "Examples\RuS2" directory which is present in your **ENDEAVOUR** program directory. If you want to have a look at it, you can read it with any conventional ASCII text editor. In any case, at least be sure that you have it available.

Now start **ENDEAVOUR** by double-clicking on the corresponding icon on your desktop or by selecting "Endeavour" from the "Start-Programs-Endeavour" windows menu. After a few seconds, **ENDEAVOUR**'s (still empty) main window will appear. From the "File"-menu choose "New" (or press the corresponding toolbar button ). In the dialog window, be sure that "Crystal Structure" is chosen, then press OK. A new empty **ENDEAVOUR** document appears on the screen (Fig. 3).

²⁸ The sample calculation shown in this text has been performed using a pre-version of Endeavour 1.0.

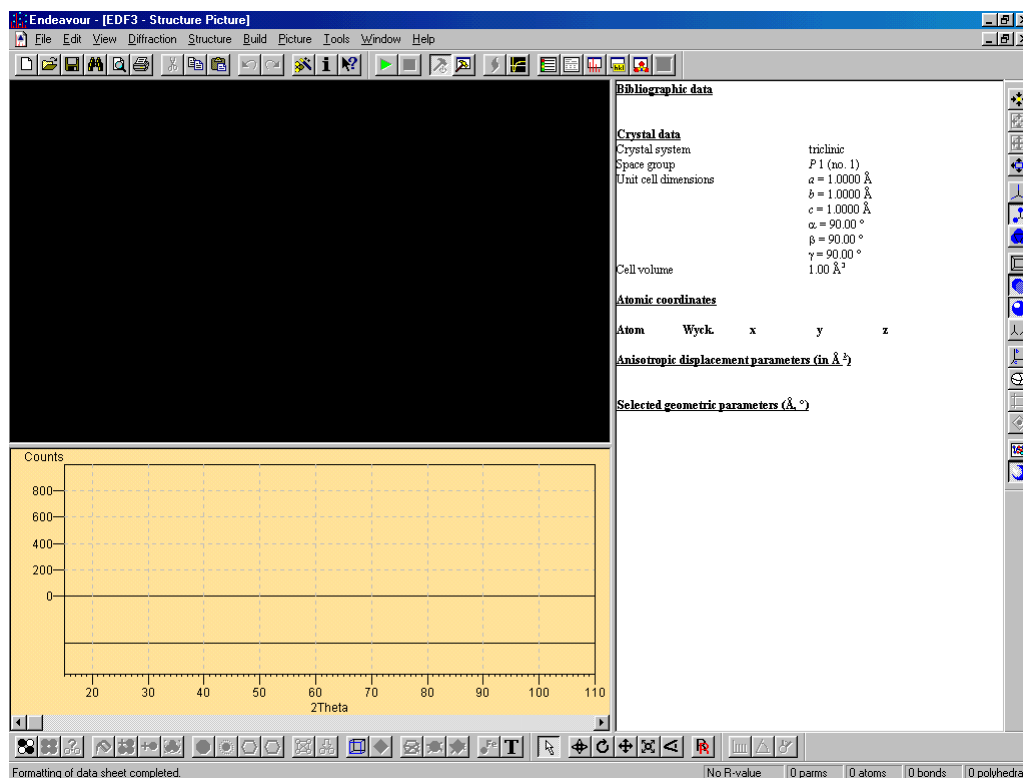




Fig. 3: Empty ENDEAVOUR document

As you see, the screen is by default divided into three parts: In the upper left area, the structures will be displayed. The powder pattern is shown in the lower left area, and on the right hand side textual output in various forms can be displayed.

Please assure that the entry "Data sheet" in menu "View" is checked (or press the corresponding toolbar button ) , so that you can see the crystal structure data on the textual part of the screen to the right.

Now you have to enter your data. ENDEAVOUR provides a so-called "Structure Solution Wizard" which makes this process very easy. To start with the structure solution, simply click on the  button in the toolbar or select "Start Solution..." from the "Structure" menu. The first page of the Wizard appears (Fig. 4).

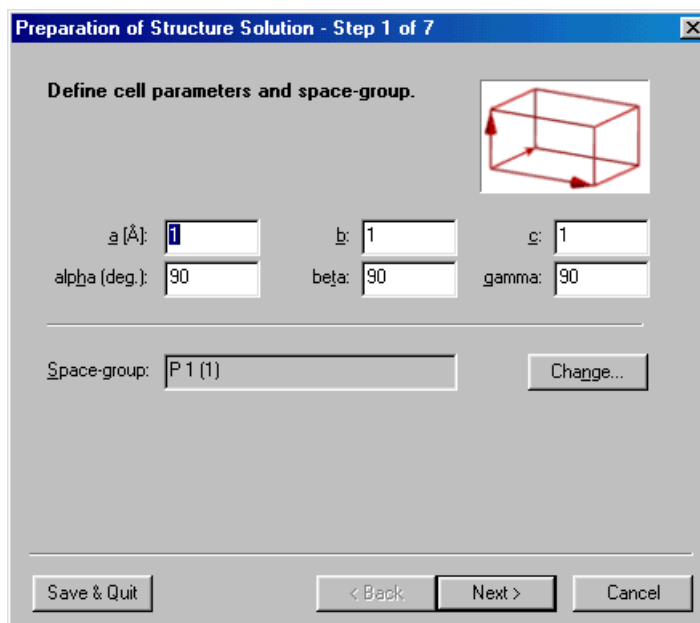


Fig. 4: Structure Solution Wizard page 1 of 7

On this first wizard page you enter the results of the indexation of the diffraction pattern, i.e. the unit cell constants and eventually the space group. Please enter 5.6095 in the fields "a", "b" and "c" as well as 90 in the fields "alpha", "beta" and "gamma". In this example, we assume that it was not possible to determine the space group from the diffraction data by analysis of the Laue group and the systematic absences of reflections. Hence, we will have to perform the structure solution in P1. This is the default space group, so just leave it as it is (Fig. 5).

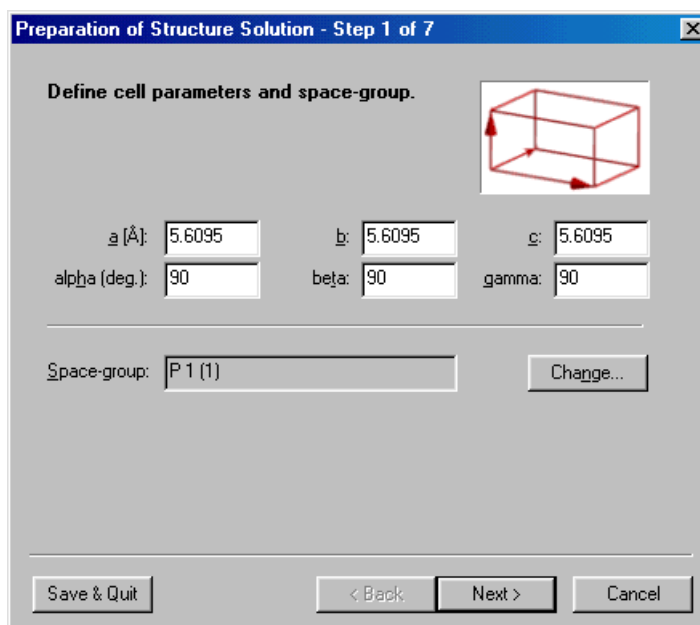


Fig. 5: Structure Solution Wizard p. 1: Unit Cell Parameters entered

Proceed to the next wizard page by clicking "Next". Page 2 of the wizard is dedicated to the input of rigid body molecules (Fig. 6).

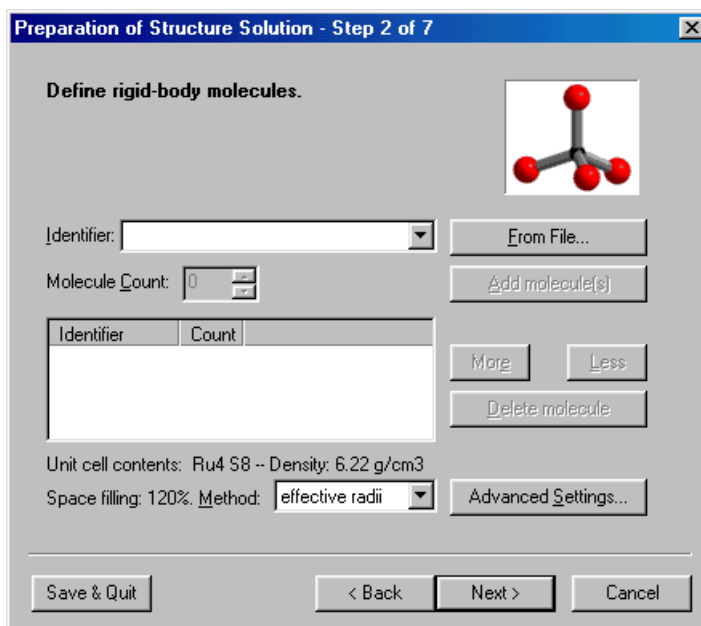


Fig. 6: Structure Solution Wizard p. 2: Declaration of rigid body molecules

Since you assume that your compound consists of individual atoms or ions, respectively, you can skip this page by pressing "Next" again. Page 3 of the structure solution wizard appears (Fig. 7).

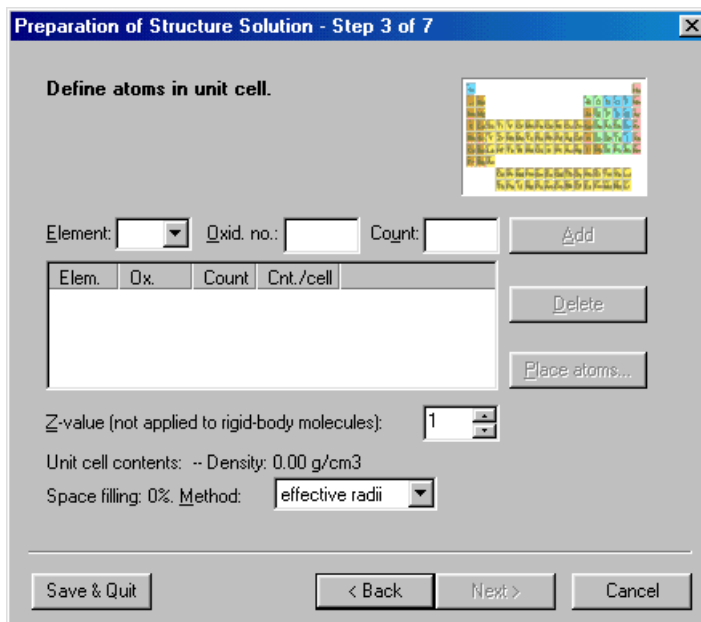


Fig. 7: Structure Solution Wizard p. 3: Definition of unit cell contents

On page 3 you have to enter the unit cell contents, i.e. the sum formula and the

number of formula units per unit cell Z . The sum formula is entered atom by atom. Since the formula of your compound is "RuS₂", write "Ru" in the input field next to "Element:", then press "tab" to advance to the next field "Oxid. no.:". Assuming that RuS₂ is a compound basically consisting of ions, you give the oxidation state of Ru (4 in this case). Press "tab" again to advance to "Count:". The composition is 1*Ru, 2*S, so please enter a 1. Afterwards, click on "Add" to add the newly created atom type to the list (Fig. 8).

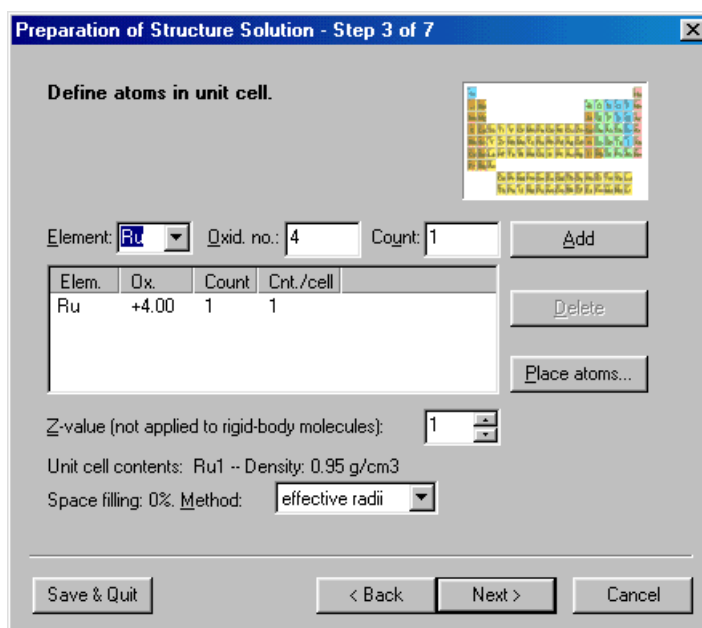


Fig. 8: One Ru⁴⁺-ion is present in the formula unit

Now proceed by entering the data for S (S,-2,2). Finally, you have to give the number of formula units per unit cell Z . **ENDEAVOUR** offers a variety of options to assist in the calculation of this value. Please adjust the Z -value in order to best match an experimental density of 6.2 g/cm³. (If you are in doubt, simply enter the correct value of 4 in the appropriate field below the atom list.) The wizard page should now look as Fig. 9.

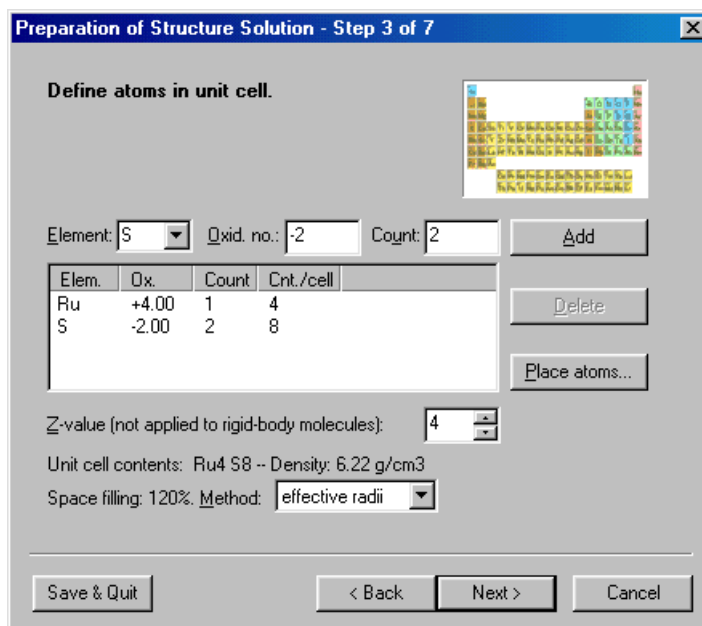


Fig. 9: One Ru^{4+} and two S^{2-} -ions in the formula unit; 4 formula units in the unit cell

Please press "Next" to advance to page 4 of the wizard which is dedicated to the input of the diffraction data (Fig. 10).

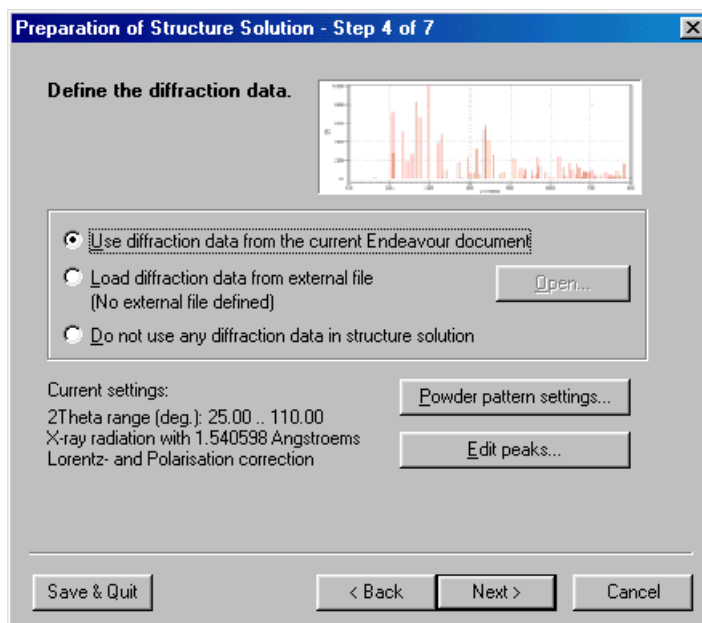


Fig. 10: Structure Solution Wizard page 4 of 7

First, you have to open the prepared file mentioned above containing the diffraction data as a list of 2θ - and corresponding intensity values. Select "Open...". A conventional Windows file dialog box appears with the file type set to "*.dif" which is the default file extension for this kind of diffraction data

representation. Select "rus2.dif" from the "Examples\RuS2" directory in your **ENDEAVOUR** program directory. Now you are returned to the first wizard page; the filename you just selected appears in the line after "Load diffraction data from external file" (Fig. 11).

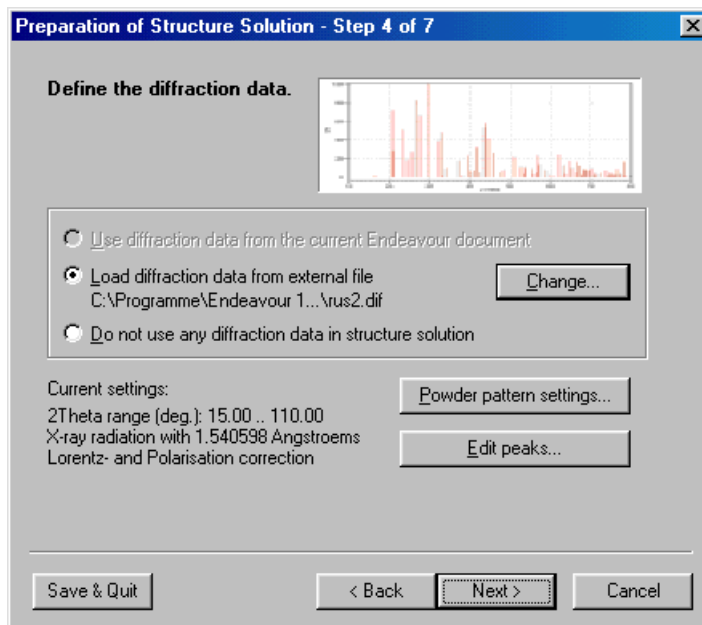


Fig. 11: Structure Solution Wizard page 4: Diffraction data loaded

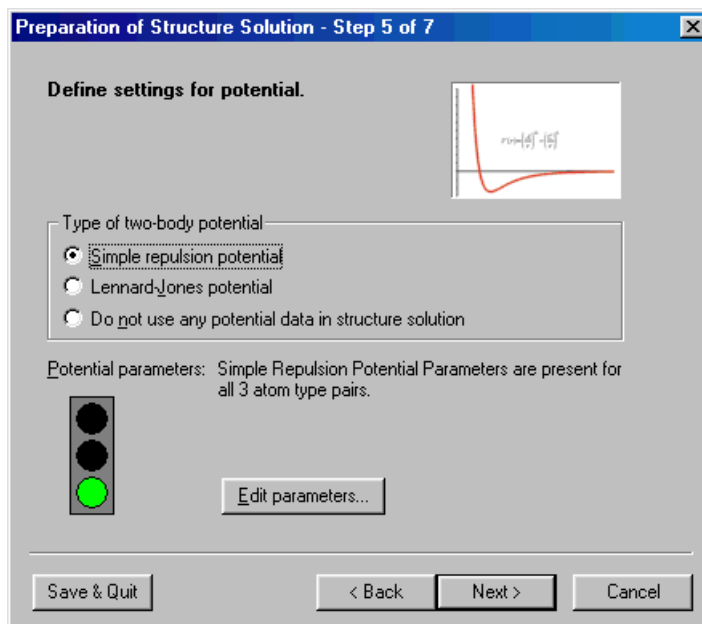


Fig. 12: Structure Solution Wizard p. 5: Which potential shall be used in the calculation of the potential energy ?

The remaining two buttons allow the modification of the experimental parameters

(wavelength etc., "Powder pattern settings...") and the grouping and/or assignment of peak positions to the possible positions calculated from the unit cell parameters ("Edit peaks..."). However, in our simple example you don't have to worry about them; just press "Next" to proceed to step 5 (Fig. 12).

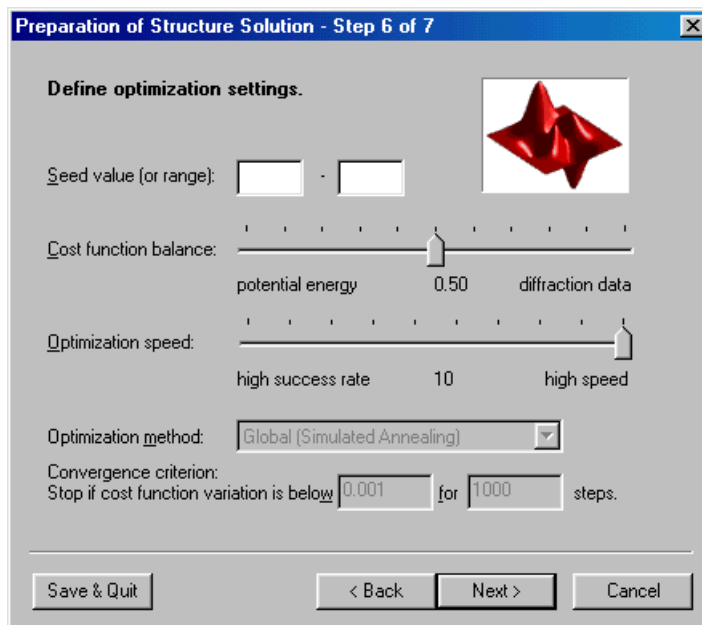


Fig. 13: Wizard p. 6: There are a variety of options concerning the optimization

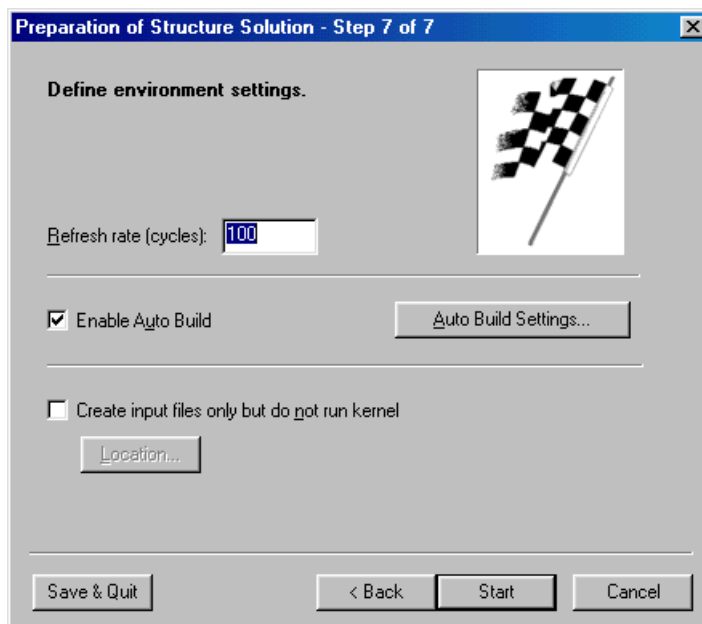


Fig. 14: Wizard p. 7: Final settings: Do you want to visualize intermediate structures ?

This page is dedicated to the input of data needed for the calculation of the potential energy. For our current example, appropriate values are already present

(as may be seen by the "traffic light" showing green), so don't care about this at the moment and just press "Next" to advance to page 6 of the wizard (Fig. 13).

Page 6 offers a lot of options concerning the optimization; we simply use the default values for our first example. Please press "Next" again to advance to the 7th and final wizard page (Fig. 14).

Since **ENDEAVOUR** is not only a program for structure solution but also for structure visualisation, you can view the intermediate structures during the structure solution process. On wizard page 7 you can tell the program how this should be done; however, don't care about this at the moment. Just press "Start" to finish the data input and start the calculation. The structure solution wizard vanishes, and a small window called "Structure Solution Progress Viewer" is displayed instead. The screen should look similar to Fig. 15.

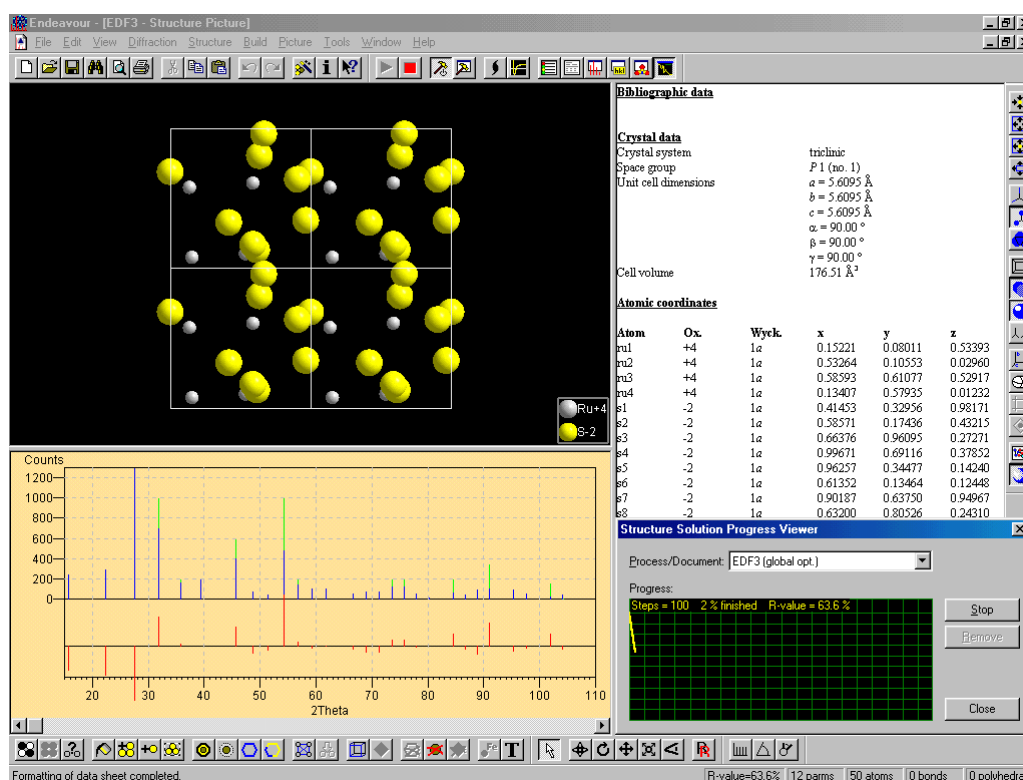



Fig. 15: The structure solution calculation for RuS₂ has just been started

The calculation will last approx. 40 seconds on a Pentium II 400 MHz processor. In the meantime, relax and watch the atoms move around in the unit cell while the program tries to find the global optimum of the cost function. If you look carefully you can see that the Ru-atoms reach their final arrangement way before the S-atoms due to their larger number of electrons and hence their larger X-ray scattering power. In the progress viewer you can watch the R-Value decrease and finally reach a limit of around 7-9% at the end of the calculation. Wait until "Optimization finished." is displayed in the progress viewer.

At this point, the crystal structure already has been solved, however, the crystallographic description is incomplete due to the missing space group information (remember that the calculation was performed in P1). **ENDEAVOUR** provides a builtin combination of tools called SFND²³/RGS²⁴ which detect symmetry elements in the triclinic atomic arrangement from which the correct space group is determined automatically. Please press  in the toolbar or select "Find Symmetry..." from the "Structure" menu; this brings up a dialog window in which certain tolerance parameters for the symmetry element detection may be adjusted. Please adjust the second parameter to a value of 0.4 as in Fig. 16.

Afterwards, press "Finish". Now have a look at the data sheet in the textual window to the right of your screen: The space group has changed from "P1" to "P a -3 ", and the list of atoms has become much shorter, as the atoms have been placed on special positions of the space group.

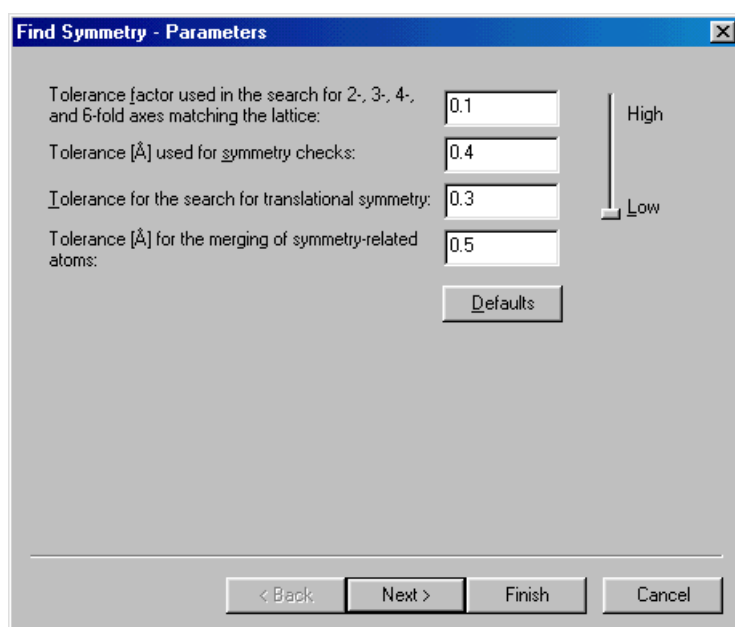


Fig. 16: Setting of tolerance parameters for the determination of the space group

At this stage, the structure solution is complete; all crystal structure data necessary for the Rietveld refinement have been determined. You may now export your crystal structure in a variety of common formats (formats (e.g. CIF, CRYSTIN, CSD, PDB, SHELX) and/or save or print the data sheet and advance to the Rietveld refinement. Besides this, you may use **ENDEAVOUR**'s large amount of visualisation capabilities to display the structure the way you like (e.g. using polyhedra like in Fig. 17).

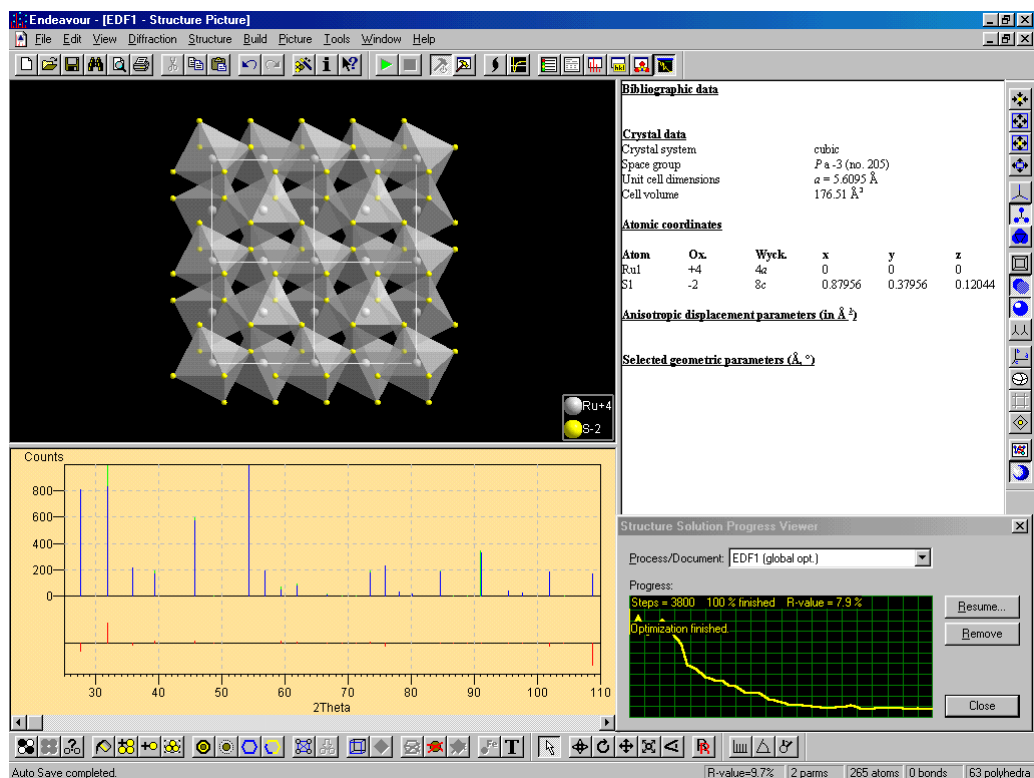


Fig. 17: The structure solution for RuS_2 has been finished; polyhedra have been drawn around the Ru atoms. The crystallographic data are displayed in the textual part of the screen to the right

Structure Determination From Powder Data with TOPAS

Arnt Kern

*Bruker AXS GmbH
Östliche Rheinbrückenstraße 50
76187 Karlsruhe
Germany*

1 INTRODUCTION

Structure determination from powder data (SDPD) has become more and more attractive in recent years. The increasing amount of new materials in material sciences and pharmaceuticals combined with the need of knowing the structure for more precise and enhanced scientific research demands to go new ways in analytical sciences.

Single crystal x-ray diffraction is one of the most powerful tools in scientific research. Unfortunately not all phases can be grown as single crystals and as a consequence the usage of x-ray powder methods becomes more and more attractive.

X-ray powder methods have two major disadvantages compared with single crystal methods:

- 3 dimensional data of the reciprocal space collapse onto 1 dimension
Consequences are:
 - Systematic and accidental peak overlap resulting in information losses
 - An accurate determination individual integrated peak intensities is impeded
- Less amount of measured reflections
The consequence is:
 - Smaller information content

The application of most common single crystal structure determination methods (Direct methods, Patterson synthesis) often fails because of these substantial problems. Therefore new methods have to be considered in order to overcome these facts.

Structures of inorganic compounds, small molecules or even large proteins have some facts in common: Reasonable packing, bond distances and coordination.

All these points are the expression of the bonding situation between its components and they are following physical and chemical principles.

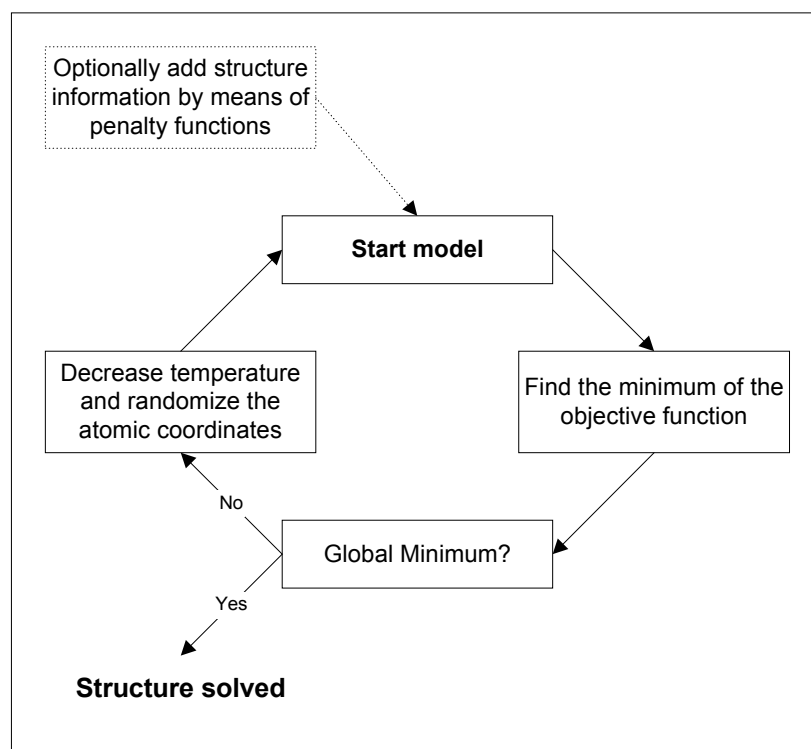
Hint! Although TOPAS has been designed for powder data work, structure determination from single crystal data is fully supported as well based on structure factors instead of step intensity data (see section 0).

2 SDPD USING STEP INTENSITY DATA

TOPAS offers the possibility of structure determination from powder data as part of a Rietveld refinement with respect to the global principles mentioned in section 1 (Coelho, 2000). Simulated annealing as well as penalty function techniques have become part of the Rietveld refinement process (Fig. 1).

In this approach penalty functions can be optionally applied to characterize the interactions between the atoms/molecules in an unknown structure in order to achieve faster convergence of the simulated annealing process. All other features of TOPAS, as a normal Rietveld refinement program, are still available. The structure determination process is done in direct space using step intensity data. It is therefore not necessary to extract integrated intensity for Direct/Patterson solution methods or usage of Fourier maps. Different types of penalty functions, for example Lennard-Jones or Born Mayer energy potential functions (used for energy minimization) and Anti Bump algorithms, are supplied with the program package and can be used. But any other function describing the bonding situation in a better way can be created and supplied by the user. This e.g. includes the possibility to use rigid bodies with all parameters refineable, which can be linked and rotated about any user defined axis. This way all information about a structure that a user can access can be added to the structure determination process to get the highest probability of success.

Fig. 1:
Schematic representation of the simulated annealing approach as implemented in TOPAS.



The TOPAS ab-initio structure determination approach from step intensity has been introduced the first time in the workshop "A New Fundamental Parameters Approach" accompanying the VI. EPDIC meeting in Budapest, Hungary, 24.08.1998. For details please refer to Coelho, 2000.

3 TUTORIAL

The following two examples demonstrate how to use TOPAS for structure determination from step intensity data. Note, that this approach is only available in Launch Mode

The structure determination approach requires the following information:

- Approximate cell parameters of the structure
- Rough knowledge about the cell content, e.g. chemical elements, amount of formula units in cell (density), molecule fragments
- A space group guess. In principle it is possible to start with P1.

Any additional information about the structure, for example from any spectroscopic methods, may be helpful and should be therefore taken into account.

3.1 Structure determination of PbSO₄

This example is the inorganic compound PbSO₄. In this well known structure the asymmetric cell consists of five sites, one Pb, one S, and three oxygen sites. For this tutorial the number of sites, however, is assumed to be unknown. The structure will be solved from scratch.

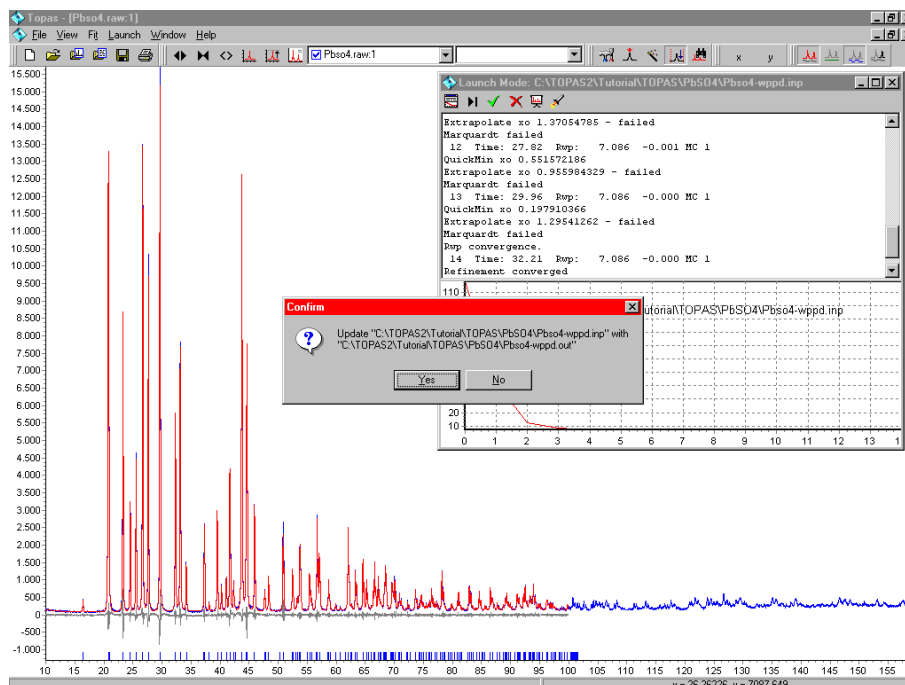
The lesson is divided into two parts:

- I. Whole Powder Pattern Decomposition (Pawley method) to obtain reasonable background, lattice and profile parameters
- II. Structure determination

Whole Powder Pattern Decomposition

For the first calculation follow these steps:

1. Start TOPAS.
2. In the Launch menu define the following input file: PBSO4-WPPD.INP. By default this file is located in C:\TOPAS2\TUTORIAL\TOPAS\PBSO4. Inspect the input file.
3. Start the refinement. Refined background, lattice and profile parameters will be used in the next example files.



Structure determination

4. In the Launch menu define the following input file: PBSO4-1.INP. Inspect the input file, which is based on the output file of the previous refinement.

The input file consists of the following new commands compared to the WPPD example:

- The block *STR(...)* with the space group, unit cell parameters and site parameters has been added.
- Profile and background parameters are fixed to the values from the WPPD result.
- *chi2_convergence_criteria 0.01*: Calculation stops when chi2 varies under 0.01.
- *dont_walk*: Instructs the refinement to randomize the atomic coordinates based on the best R_{WP} value found.
- *continue_after_convergence*: Activates the structure determination mode.
- *Temperature_Regime { 2 .7 .7 .7 }*: Simulated annealing temperatures.
- *rand_xyz*: Shift parameters which is multiplied by the simulated annealing temperatures.
- *num_posns*: Outputs the multiplicity of the site.
- *penalties_weighting_K1*: Weighting factor for penalties.
- *swap_sites*: A process invoked after a particular temperature sequence is processed. Pb sites are swapped with and and O sites and S sites are swapped with O sites.
- *append_bond_lengths*: The bond distances and angles will be calculated and added at the end of the output file.
- *append_fractional*: Very helpful function, calculates all fractional atomic coordinates for each site within the unit cell.

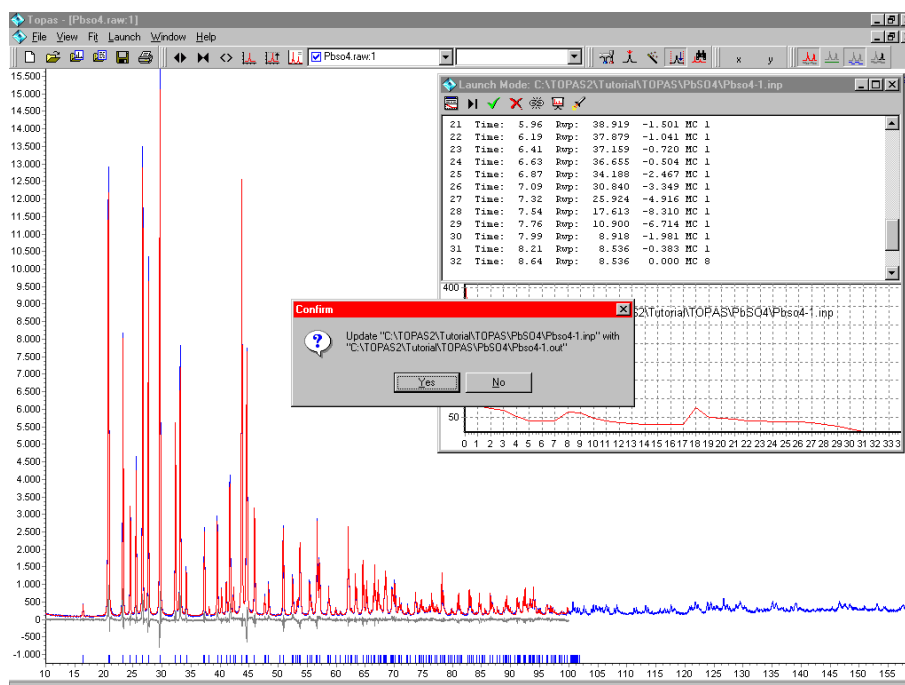
The space group Pbnm generates 8 equivalent positions per site. It is expected that there are 4 Pb atoms, 4 S atoms and 16 O atoms in the unit cell. Thus Pb and S must lie on special positions. We will assume that there is only one Pb and one S site. The task is to find the special position for these sites. This is accomplished by setting the starting coordinates to a general position. Thus there will be 8 Pb and 8 S atoms within the cell. This represents too much scattering power as the unit cell should only contain 4 atoms of each. To maintain the expected scattering power the occupancies has been set to 0.5. Four oxygen sites have been included each with an occupancy of 0.5; this maintains the expected scattering power of 16 O atoms within the unit cell.

Starting positions of the atoms were chosen to be almost zero. *rand_xyz* values has been set to correspond to expected mean bond lengths; *rand_xyz* for Pb and oxygen has been set to 2 Å. For further working (especially for unknown structures) it is necessary to get information about the possible

positional properties of the respective space group. It is therefore advisable to use the *International Tables for Crystallography*.

5. Start the refinement.

- Step after step calculations are performed and the ongoing results can be graphically viewed with the Animated Fitting option. In the Fit Window the current refinement R_{WP} can be displayed. Note, that the graphical display slows down the refinement.
- Stop the calculations after some hundred cycles when the R_{WP} does not get lower than about 8.6%.
- The program chooses the best refinement result and then calculates bond lengths/angles.



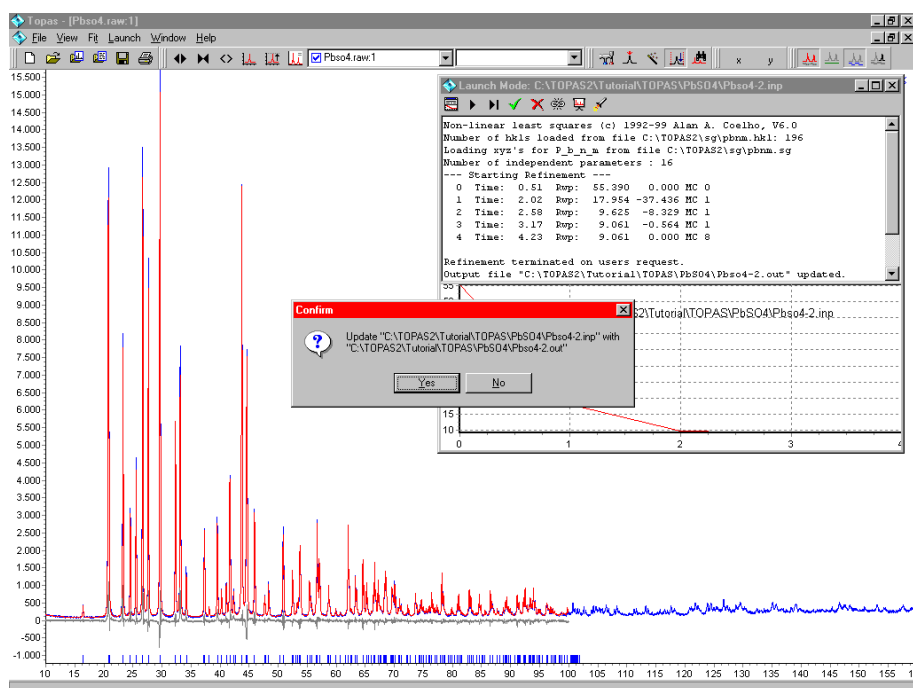
6. Open the file PbSO4-1.out. The values for the positional parameters have changed. *num_positions* now is 8 (general position multiplicity of this space group) and at the end of this file the new bond lengths/angles are summarized.

- The shortest bond lengths for the six sites are:

Pb	Pb	0.19485
S	S	0.15067
O1	O4	0.52480
O2	O2	0.22152
O3	O3	0.02091
O4	O1	0.52480

where the values are in Angstroms.

- It is seen that S-S and Pb-Pb bond distances are near zero. Inspecting the refined coordinates shows that the *z*-coordinates are near values required for a special position and thus the special position for these sites have been found. Set the occupancies to 1 and the *z*-coordinates to the special position values of 0.25 and 0.75 for Pb and S respectively and stop refining on the *z*-coordinates. *num_positions* for these sites will be updated to a value of 4 the next time the program is run. Note, special positions with recurring values such as 0.3333 and 0.6667 should be entered as equations such as $z = 1/3$ and $z = 2/3$;
 - It is also noticed that the O2-O2 and O3-O3 bond distances are near zero and that the *z*-coordinates are close to special positions. Note, that your refinement may show different oxygen sites as the process has a random element to it; in any case two oxygen sites (regardless of their labels) should have one of their equivalent positions as their nearest neighbour. Set the *z*-coordinate values of these oxygen sites to 0.75 and stop refining on the *z*-coordinates.
 - Also seen is that O1 and O4 are relatively close to each other, or, too close to make physical sense, unless of course they represent a split site. Assuming they do not represent a split site, then they must be only one site instead of the two. Thus remove one of the sites and change the occupancy of the remaining site to 1 to maintain the scattering power.
7. PbSO4-2.inp contains the above mentioned changes. Load this file and launch it. The final result exhibits a very good agreement between the calculated and observed pattern. All R-values are good and the bond distances are reasonable.



Comments and ideas for further working:

- Use more or different Temperature steps.
- The input file can be used to set-up structure determination calculations for other compounds. Use the examples supplied with the program.

3.2 Structure determination of KCP

The example Cyclopentadienylpotassium KCP (Dinnebier *et al.*, 1999) represents a small molecular structure, which has been selected to demonstrate the following techniques:

- Use of simple penalty functions only (no energy minimization)
- Use of the "pseudo atom method"
- Use of user-defined rigid bodies

In the pseudo atom method organic groups (such as the CP ligand in this example) are replaced by pseudo atoms. The pseudo atom should consist of almost the same amount of electrons as the organic ligand. Applying a very high temperature factor will approximate the electron density of the molecule to be replaced.

The pseudo atom method can assist in locating molecular fragments in the cell in particular if several ligands are present (e.g. solvent, disordered ligand). In addition it requires less calculation time.

An important disadvantage of this method to be considered is the fact, that

pseudo atoms are spherical symmetric and thus do not properly describe most ligands such as rings or chains.

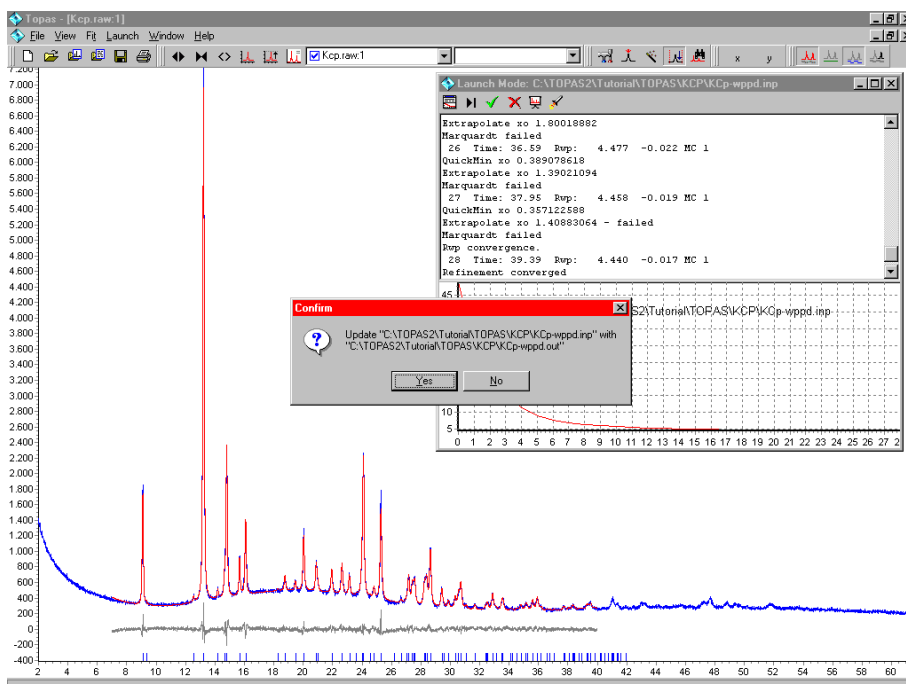
Note: To solve simple structures such as the KCP structure the pseudo atom method is not required at all. Nevertheless this approach will be described here as it might be useful elsewhere.

The following lesson is divided into four parts:

- I. Whole Powder Pattern Decomposition (Pawley method) to obtain reasonable background, lattice and profile parameters
- II. Partial structure determination using the pseudo atom method
- III. Structure determination using an user-defined rigid body for the CP ring
- IV. Rietveld structure refinement using a rigid body

Part I: Whole Powder Pattern Decomposition:

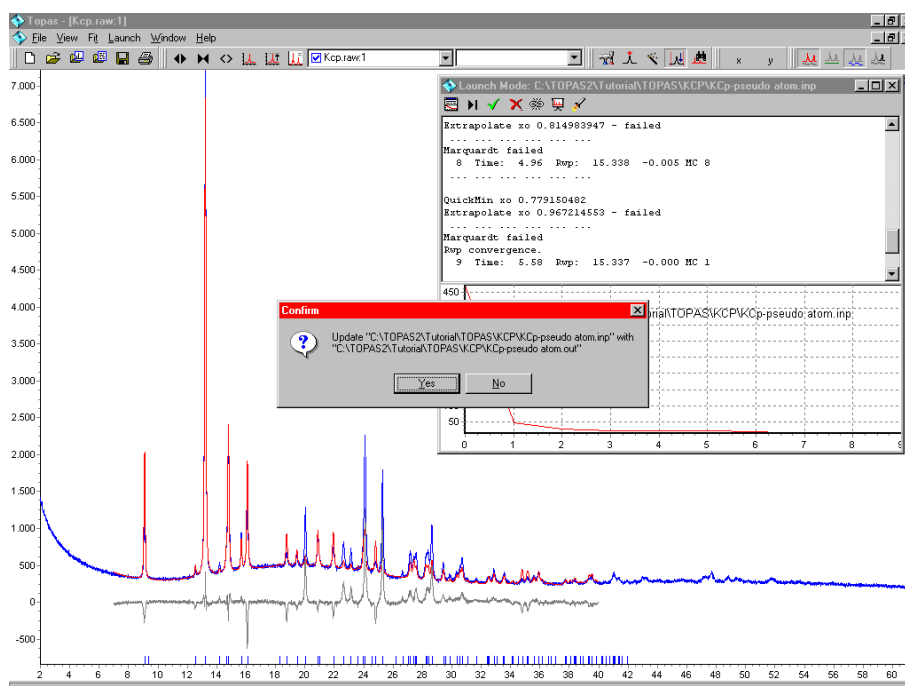
1. Start TOPAS.
2. In the Launch menu define the following input file: KCP-WPPD.INP. By default this file is located in C:\TOPAS2\TUTORIAL\TOPAS\KCP. Inspect the input file. The space group and start values for lattice parameters have been taken from Dinnebier (1999). It is sufficient to use only the data up to $40^\circ 2\theta$.
3. Start the refinement. Refined background, lattice and profile parameters will be used in the next example files.



In the next part the use of the pseudo atom method for locating molecular fragments will be described. As this method is not required to solve the KCP structure you can also proceed directly to part III (step 6) in order to solve the structure completely from scratch.

Part II: Structure determination using the pseudo atom method:

4. In the Launch menu define the following input file: KCP-WPPD.INP. Inspect the input file. The structure consists of just a K and a Br atom (the pseudo atom). A high temperature factor has been assigned to Br in order to model the CP ring. The Anti_Bump macros apply penalty functions to the K - K, K - Br, and Br - Br sites to avoid distances less than 4, 2, and 4 Å, respectively.
5. Start the refinement. After some 7 to 10 cycles the refinement calculation can be stopped (when R_{WP} drops below 16%).
 - Result: The positional parameters of the heavy scatterers (Potassium and pseudo atom) have been determined and can be used for the next steps.

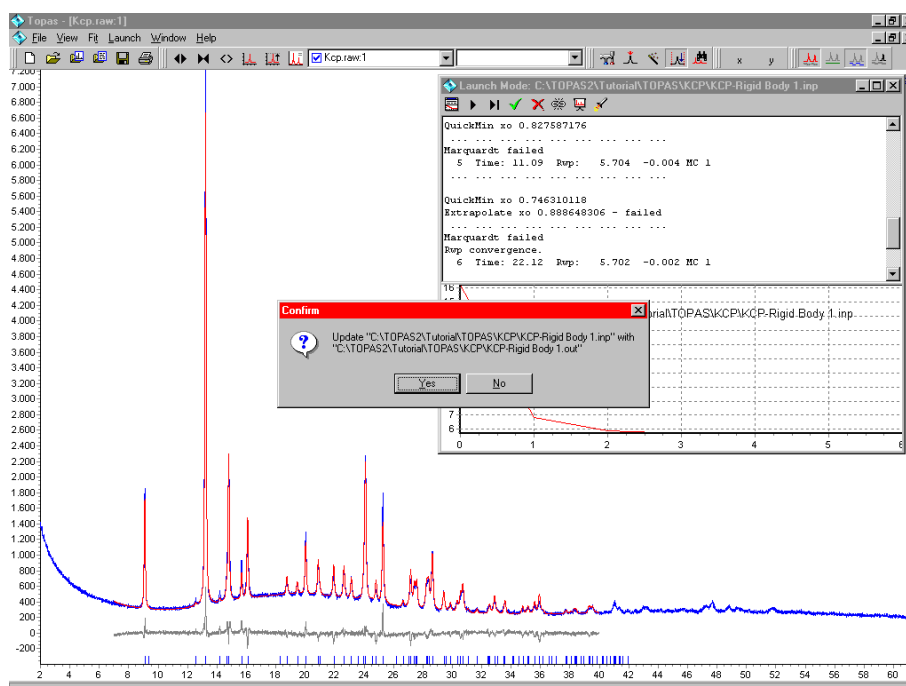


Part III: Structure determination using an user-defined rigid body for the CP ring

6. Introducing an rigid body for the CP ring will greatly assist in structure determination. The rigid body definition is provided in the include file "MYRIGIDBODY.INP", the procedure how to setup this rigid body is outlined in section 4.8 in detail.

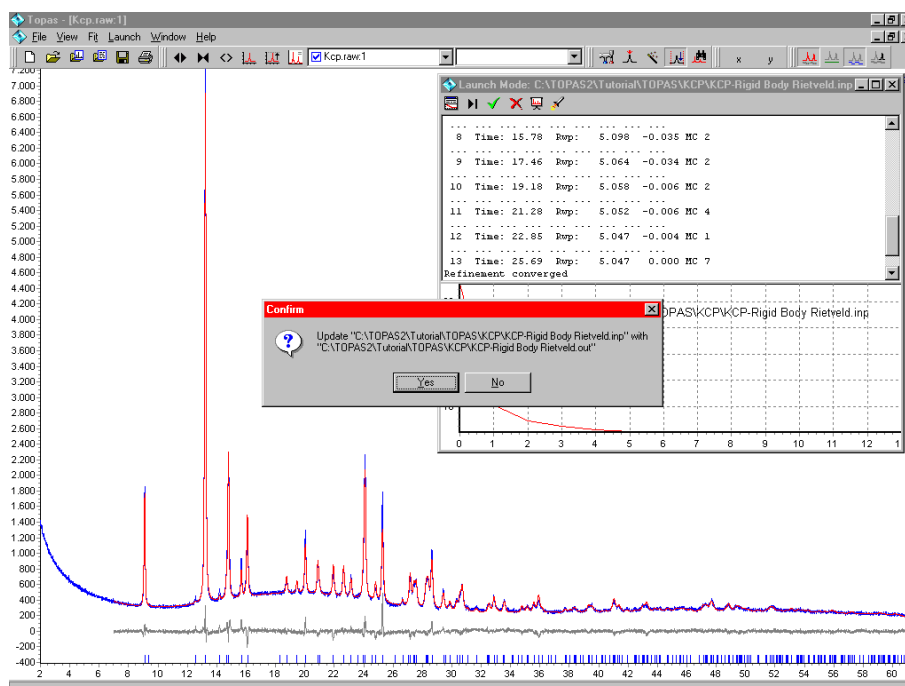
- Structure determination using pseudo atoms:
To solve the structure the rigid body has to replace the pseudo atom. Inspect the input file KCP-RIGID BODY 1.INP. Note the arbitrary A0 site, which defines the center the rigid body. The translate macro puts the center of the rigid body to the refined pseudo atom position.
 - Structure determination from scratch:
Inspect the input file KCP-RIGID BODY 2.INP. Note, that all sites have been set close to zero.
7. In the Launch menu define one of these input files and start the refinement.
- Structure determination using pseudo atoms:
The structure will be solved instantly in 1 -2 cycles.
 - Structure determination from scratch:
The structure will be solved almost instantly in about 7 cycles.

The output file can be used for final structure refinement.



Part IV: Rietveld structure refinement using a rigid body

8. In the Launch menu define the following input file: KCP-RIGID BODY RIETVELD.INP. Inspect the input file, which is based on an output file obtained in part III. For final Rietveld structure refinement the background, lattice and profile parameters have to be refined.
9. Start the refinement. R_{WP} and R_{Bragg} will drop down to about 5% and 3%, respectively.



Comments and ideas for further working:

- Use larger and smaller data ranges.
- Use a different atom type as a pseudo atom with much more or much less electrons. Apply a different temperature factor.
- Modify the parameters of the rigid body (mean distances)
- Modify the weighting and distance parameters of the Anti_Bump restraint.
- Use a different bondlength restraint but Anti_Bump.

4 REFERENCES

Coelho, A.A. (2000): *Whole-profile structure solution from powder diffraction data using simulated annealing.* - J. Appl. Cryst., **33**, 899 - 908.

Dinnebier R. E., Behrens U, Olbrich F (1997): *Solid state structures of cyclopentadienyllithium, -sodium, and -potassium. Determination by high-resolution powder diffraction.* - Organometallics, **16 (17)**, 3855-3858.

Energy Minimization Techniques

Martin U. Schmidt

Clariant GmbH,

Laboratory for Crystal Engineering and Polymorphism Studies

Pigment Technology Research, G 834,

D-65926 Frankfurt am Main, Germany

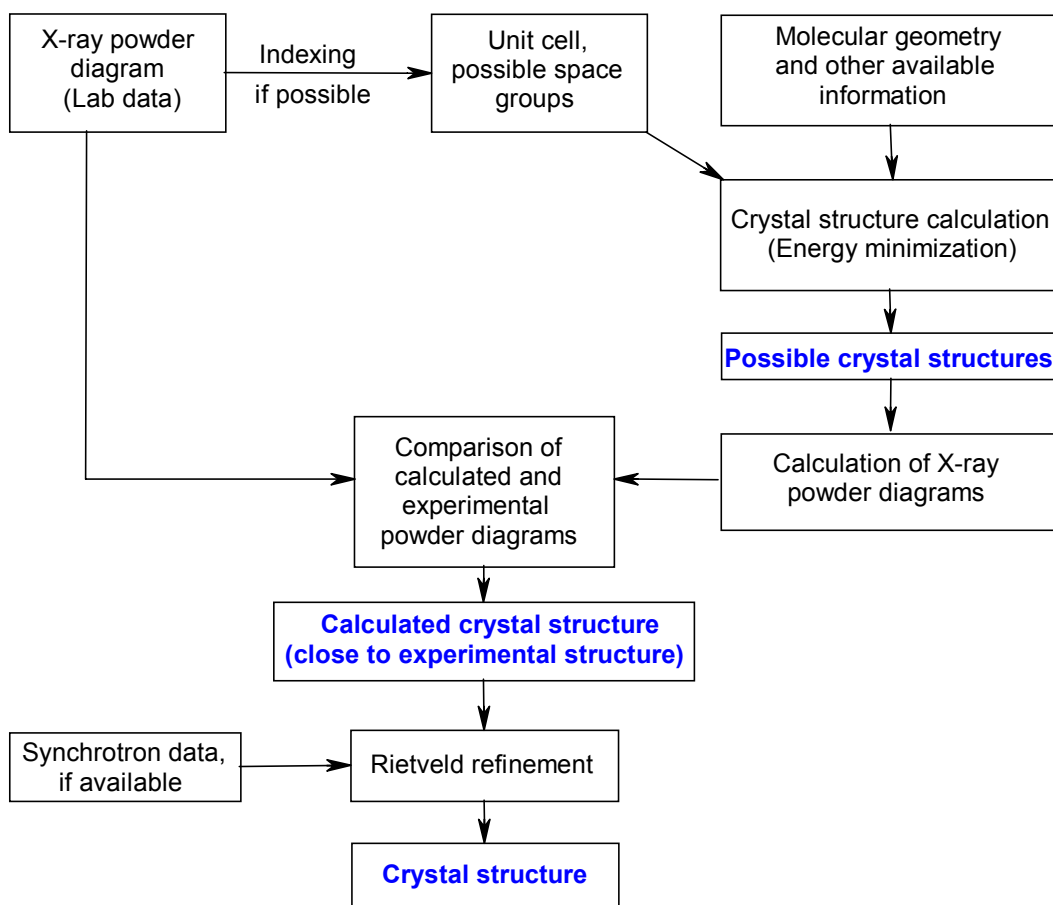
Tel. ++49 (0)69 305-13426,

Fax ++49 (0)69 331749

E-mail: martinulrich.schmidt@clariant.com

Introduction

Crystal structures of molecular compounds can be solved from X-ray powder data by energy minimization.



The procedure consists of six steps:

1. Indexing of the powder diagram and deduction of the possible space groups. If indexing is not possible, energy minimization can be performed as well, but larger calculation times are needed.
2. Set-up of the molecular geometry.
3. Calculation of the possible crystal structures by lattice energy minimization.
4. Calculation of the powder patterns for the possible crystal structures.
5. Selection of the correct solution by comparing the calculated with the experimental powder diagrams.
6. Fit onto the full powder diagram by Rietveld refinement

Lattice Energy Minimization: Method and Approximations [1,3]

Energy terms

The most sophisticated method for calculating the energy of a molecular crystal would be a high-level quantum mechanical calculation, taking into account the periodicity of the crystal lattice. Ab initio calculations on crystals of medium-sized molecules are not possible yet with a sufficiently high accuracy. Therefore force field methods are used.

Generally speaking, three assumptions are made for crystal structure calculations by energy minimization:

1. Entropic effects are neglected. The free energy of a crystal lattice, given by

$$F = U - T \cdot S, \quad (1)$$

is approximated by a temperature-independent energy E . The entropic term $T \cdot S$ is not small, but similar for different packings of a given molecule: Under the assumption, that the molecular geometry does not change drastically, the intramolecular contributions to the sum of states remain almost constant; furthermore the intermolecular contributions change only slightly, since in all packings the molecules are surrounded by other molecules. The entropic term differs mostly in the order of $T \cdot \Delta S = 0$ to 10 kJ/mol between polymorphic forms. The energy E includes an averaged entropic term; the force field parameters are adjusted in order to reproduce crystal structures at ambient temperature. By neglecting the entropy, the temperature of phase transitions etc. cannot be calculated.

2. It is assumed, that the experimental crystal structure corresponds to the absolute minimum of energy. In fact, experimental crystal structures can correspond to either the global or a local minimum of the free energy. Energy

differences between different polymorphic forms are mostly in the order of $\Delta H = 0$ to 10 kJ/mol . Thus for a prediction of all possible polymorphic forms one should take into account not only the structure with the lowest energy, but also all packings having slightly higher energies. The energy range, which has to be considered, depends on the reliability of the force field, and on the other assumptions made.

3. The intermolecular energy is calculated by the atom-atom potential method [4]. In this approximation the interactions of the molecules are divided into a sum over individual atom-atom interactions:

$$E = \frac{1}{2} \sum_i \sum_j E_{ij}(r_{ij}) \quad (2)$$

The sums run over all atoms i of a reference molecule and all atoms j of all other molecules. The energy contributions of 3-body interactions are fitted empirically onto the energy of 2-body interactions. In the atom-atom potential approximation the atom-atom energy E_{ij} depends only on the distance r_{ij} . This works well for van der Waals and Coulomb interactions. In a first approximation, hydrogen bonds can be calculated with this approach, too [5].

It should be mentioned, that instead of energy terms also statistically derived potentials can be used [6,7].

Packing Parameters

The crystal structure of a molecular compound can be described by the molecular geometry, the crystal symmetry, and a set of packing parameters, i.e. the unit-cell dimensions $a, b, c, \alpha, \beta, \gamma$, and the positions and orientations of each symmetry-independent molecule. For the position of a molecule the fractional coordinates x, y, z of the centre of gravity can be used. The spatial orientation of a molecule is described by three angles φ_x, φ_y and φ_z (Unfortunately each program uses its own definition of angles).

Symmetry

In crystal structure calculations the space group symmetry is generally included from the beginning. For applications like structure prediction or search for possible polymorphic forms, where no *a priori* space group information is available, possible space groups have to be tested separately. Fortunately only a very limited number of crystal symmetries are common ($P2_1/c, Z=4$; $P2_12_12_1, Z=4$; $P\bar{1}, Z=2$; $P2_1, Z=2$; $P2_1/c, Z=2$; $Pbca, Z=8$; $P2_1/c, Z=8$; $C2/c, Z=8$). It is recommended to respect also space groups known to occur in similar compounds. If the molecule has internal symmetry, or may adopt internal symmetry during the minimization, supergroups of the tested space groups can be reached (e.g. $Pnma$,

$Z=4$ from $P2_1/c$, $Z=4$ for molecules with mirror planes, or $P2_1/c$, $Z=2$ from $P2_1/c$, $Z=4$ for molecules with inversion centres).

Energy Minimization

The expression for the lattice energy as a sum of several thousand individual interactions is too complicated to be minimized analytically. Therefore the minimization must be performed by numerical methods. In the last years a variety of different methods has been applied, like steepest descent [2], conjugate gradient, Newton-Raphson, truncated Newton, simulated annealing [8], molecular dynamics [9], diffusion-equation [10] and cluster methods [11]. Frequently combinations of these methods are used [12,14]. Since the energy hypersurface has a large number of local minima, the 'classical' minimization methods like steepest descent require several hundred runs starting from different points. These starting points can be randomly chosen [2], systematically varied [13,14], or calculated previously [12]. A review on different methods and their use to predict possible crystal polymorphs is given by Verwer and Leusen [15].

The accuracy of a calculated crystal structure depends on the force field. Even if the force field parameters are carefully adjusted, the calculated crystal structures are less accurate than crystal structures determined by single crystal X-ray diffraction. Typical deviations between calculated and experimental crystal structures are in the order of $0.3 \text{ \AA} / 1^\circ$ for lattice parameters and 0.1 to 0.2 \AA for intermolecular distances. This is about 100 times larger than the standard deviations coming out of a single crystal structure analysis. The calculation difficulties are mostly generated by the weak forces between the molecules, not by the strong, directed forces within them.

The following method is implemented in the program CRYSCA [1,2,16]: The minimizations start from random packings of the molecules; i.e. all packing parameters are assigned random values inside a user defined range. If the lattice parameters are known, they may be used as well. In contrast to several other methods the minimization procedure allows calculations in all space groups with molecules occupying every kind of special position, including 'complicated' cases like $\text{Ni}(\text{CO})_4$ (space group $Pa\bar{3}$ with molecules on the $\langle 111 \rangle$ axes) [2]. Disorder and non-crystallographic symmetries can be handled as well. Hitherto, one of the most complex cases was the calculation of the disordered structure of $\text{Si}[\text{Si}(\text{CH}_3)_4]_4$ using a molecule consisting of one fully occupied and 624 partially occupied atomic positions per asymmetric unit [17]. The energy is minimized by a special steepest-descent procedure. After the minimization has located an energy minimum, new random values are generated for all free packing parameters. This procedure is repeated, until the best minima are found several times from different starting points. The reproducibility is $<0.001 \text{ \AA}$, which is by far better than the precision of the force fields. The minima are sorted according

to energy and checked for higher symmetries, meaningful molecular conformations and reliable intermolecular interactions. The packing having the lowest energy corresponds to the 'predicted' crystal structure, other minima having slightly higher energies are possible polymorphic forms.

Prediction of Crystal Structures

The prediction of a crystal structure before the synthesis of the compound is very valuable, especially for the search of compounds having desired solid state properties. Examples include materials with non-linear optical properties (structures without inversion centres), explosives (crystals with high densities), pigments (coloured, stable, insoluble crystals) and pharmaceuticals (stable, bioavailable compounds). Furthermore new polymorphic forms can be patented. Therefore much work has been done with the aim to predict crystal structures from a given molecular structure; and it is not by fortune, that a considerable amount of this research has been done in the industry [10,18-20] or in collaboration with it. The prediction of crystal structures without reference to experimental data is sometimes called "*ab initio* prediction". This term might be misleading in the future, if quantum mechanical *ab initio* methods will be used for crystal structure predictions.

Are crystal structures predictable?

Angelo Gavezzotti said "no" [21]. Other authors say "yes" or "sometimes" [13,22,23]. Apart from the approximations concerning the molecular geometry and the energy terms, there are two additional problems hindering a successful prediction of molecular crystal structures:

1. The calculations are incomplete: Most methods require either the space group symmetry and the number of independent molecules, or the number of molecules per unit cell as input. The number of space groups is limited, but there are unlimited packing possibilities, if molecules are disordered or if the crystal contains more than one independent molecule. Packings with 'exotic' crystal symmetries like in α -CD₄ (space group $P\bar{4}m2$, $Z = 32$, with 9 independent molecules [24]) would probably not be generated, unless the user of the program explicitly performs calculations with the corresponding symmetry operators.

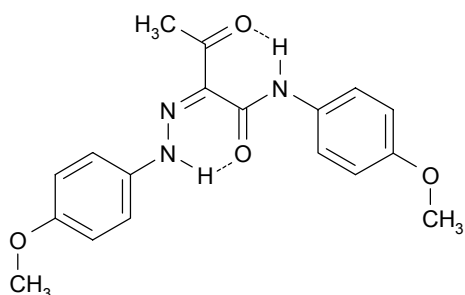
2. Most organic compounds can crystallize in different polymorphic forms. In these cases it is not possible to predict "*the*" crystal structure, one can only try to calculate possible polymorphic forms.

3. Often several minima with comparable energies are found. (In the case of acetic acid about 100 minima were found within 5 kJ/mol above the minimum energy; the number of structures could be reduced by removing space group symmetry constraints, or by molecular dynamics, but many possible structures

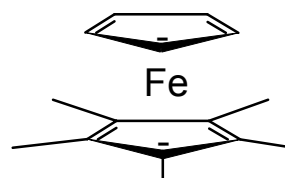
remained [25]). It is difficult to predict, which possible crystal structures may be realised experimentally, and how this could be achieved.

Predicted crystal structures

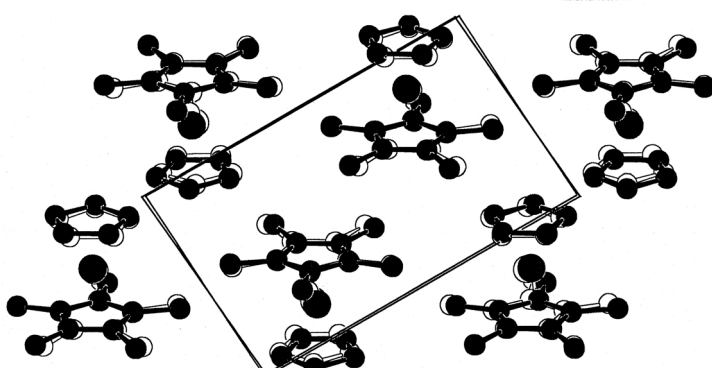
The following examples show two of the rare cases, where the crystal structure was predicted successfully. (Calculations by CRYSCA).



Azo-Pigment [26]



Pentamethylferrocene [1]



Calculated (full circles) [27] and experimental (open circles) [28] crystal structures of pentamethylferrocene (SCHAKAL plot [29])

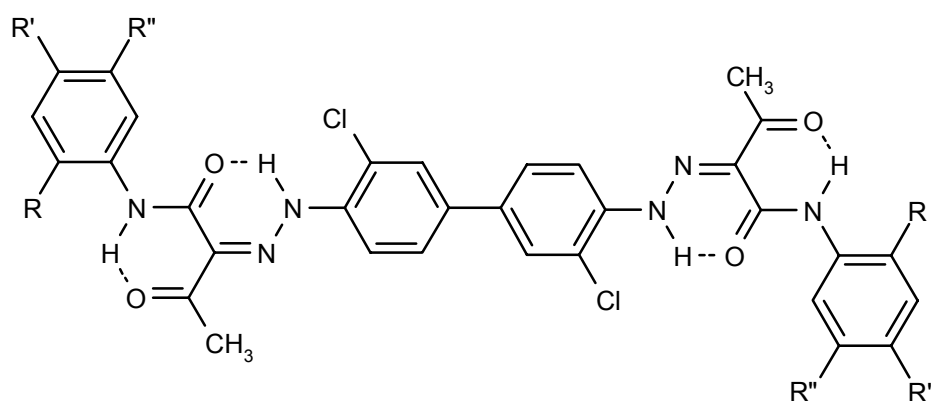
Round robin on crystal structure prediction [30]

In 1999 a blind test on crystal structure prediction was conducted on a selection of 4 small organic compounds and 11 different programs representing a wide range of methodologies (various force field methods, as well as statistical methods). The predicted crystal structures were afterwards compared with the unpublished experimental crystal structures determined by single crystal X-ray analyses. Some crystal structures were predicted correctly by some methods. No computer program was able to predict all structures successfully, but in most cases the experimental crystal structures were found in the list of calculated possible polymorphic forms. All energy calculations led to many possible structures within a few *kJ/mol* of the global minimum, and hence did not lead to unequivocal prediction of a particular structure within specified confidence

limits. The fine detail of the force-field methodology and parametrization profoundly affected the order of ranking by energy within each method.

Examples of crystal structures solved from powder data by energy minimization

The following crystal structures were solved from X-ray powder data by energy minimization using the program CRYSCA[16], and refined by Rietveld methods [31].

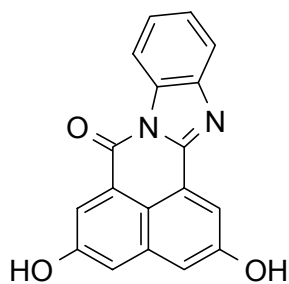


Pigment Yellow 12 ($R=R'=R''=H$) [32]

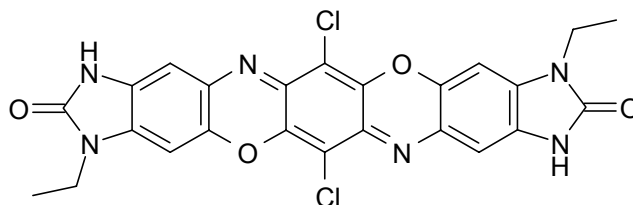
Pigment Yellow 13 ($R=R'=CH_3$, $R''=H$) [32]

Pigment Yellow 14 ($R=CH_3$, $R'=R''=H$) [32]

Pigment Yellow 83 ($R=R''=OCH_3$, $R'=Cl$)



Perinone pigment [33]



Benzimidazolono-dioxazine pigment [34]

References

-
1. Schmidt, M.U. and Englert, U. (1996) Prediction of crystal structures, *J. Chem. Soc. Dalton Trans.* **1996**, 2077-82.
 2. Schmidt, M.U. (1995) *Kristallstrukturberechnungen metallorganischer Molekülverbindungen*, Verlag Shaker, Aachen, Germany.
 3. Timofeeva, T.V., Chernikova, N.Yu. and Zorkii, P.M. (1980) Theoretical calculation of the spatial distribution of molecules in crystals, *Vcnexu Xumuu* **49**, 966-997; *Russian Chemical Reviews* **49**, 509-525.
 4. Pertsin, A.J. and Kitaigorodsky, A.I. (1987) *The Atom-Atom Potential Method*, Springer-Verlag, Berlin Heidelberg New York.
 5. Gavezzotti, A. and Filippini, G. (1994) Geometry of the intermolecular X-H·Y (X, Y = N, O) hydrogen bond and the calibration of empirical hydrogen-bond potentials, *J. Phys. Chem.* **98**, 4831-37.
 6. Hofmann, D.W.M. and Lengauer, T. (1997) A discrete algorithm for crystal structure prediction of organic molecules, *Acta Cryst.* **A53**, 225-235.
 7. Motherwell, W.D.S. (1998) Crystal structure prediction and the Cambridge crystallographic database, International School of Crystallography: "Implications of Molecular and Materials Structure for New Technologies", Erice, 28 May - 7 June 1998, Poster Abstracts P. 40.
 8. Gdanitz, R.J. (1992) Prediction of molecular crystal structures by Monte Carlo simulated annealing without reference to diffraction data, *Chem. Phys. Lett.* **190**, 391-96.
 9. Tajima, N., Tanaka, T., Arikawa, T., Sakurai, T., Teramae, S. and Hirano, T. (1995) A heuristic molecular-dynamics approach for the prediction of a molecular crystal structure, *Bull. Chem. Soc. Jpn.* **68**, 519-527.
 10. Wawak, R.J., Gibson, K.D., Liwo, A. and Scheraga, H.A. (1996) Theoretical prediction of a crystal structure, *Proc. Natl. Acad. Sci. USA* **93**, 1734-36.
 11. Gavezzotti, A. (1991) Generation of possible crystal structures from the molecular structure for low-polarity organic compounds, *J. Am. Chem. Soc.* **113**, 4622-29.
 12. Karfunkel, H.R. and Gdanitz, R.J. (1992) Ab initio prediction of possible crystal structures for general organic molecules, *J. Computat. Chem.* **13**, 1171-83.
 13. van Eijck, B.P., Mooij, W.T.M. and Kroon, J. (1995) Attempted prediction of the crystal structures of six monosaccharides, *Acta Cryst.* **B51**, 99-103.

-
14. Williams, D.E. (1996) Ab initio molecular packing analysis. *Acta Cryst.* **A52**, 326-328.
 15. Verwer, P. and Leusen, F.J.J. (1998) Computer simulation to predict possible crystal polymorphs, *Reviews in Computational Chemistry* **12**, 327-365.
 16. Schmidt, M.U. and Kalkhof, H. (1997) CRYSCA, Program for crystal structure calculations of flexible molecules, Clariant GmbH, Frankfurt am Main, Germany.
 17. Dinnebier, R.E., Dollase, W.A., Helluy, X., Kümmerlen, J., Sebald, A., Schmidt, M.U., Pagola, S., Stephens, P.W. and van Smaalen, S. (1999) Order - disorder phenomena determined by high-resolution powder diffraction: the structures of tetrakis(trimethylsilyl)methane $C(Si(CH_3)_3)_4$ and tetrakis(trimethylsilyl)silane $Si(Si(CH_3)_3)_4$, *Acta Cryst.* **B**, submitted.
 18. Perlstein, J. (1994) Molecular self-assemblies. 2. A computational method for the prediction of the structure of one-dimensional screw, glide, and inversion molecular aggregates and implications for the packing of molecules in monolayers and crystals, *J. Am. Chem. Soc.* **116**, 455-70.
 19. Frank, D. (1997) Hoechst AG, Corporate Research and Technology, Frankfurt am Main, Germany, unpublished results.
 20. Erk, P. (1998) BASF AG, Colourants Laboratory, Ludwigshafen, Germany, unpublished results.
 21. Gavezzotti, A. (1994) Are crystal structures predictable? *Acc. Chem. Res.* **27**, 309-314.
 22. Chaka, A.M., Zaniewski, R., Youngs, W., Tessier, C., Klopman, G. (1996). Predicting the crystal structure of organic molecular materials, *Acta Cryst.* **B52**, 165-183.
 23. Leusen, F.J.J. (1994) Ab initio prediction of possible crystal structures, *Z. Krist. Suppl.* **8**, 161
 24. Prokhvatilov, A.I., Isakina, A.P. (1980) An X-ray powder diffraction study of crystalline α -methan- d_4 , *Acta Cryst.* **B36**, 1576-80.
 25. Mooij, W.T.M., van Eijck, B.P., Price, S.L., Verwer, P. and Kroon, J. (1998) Crystal structure predictions for acetic acid, *J. Computat. Chem.* **19**, 459-474.
 26. Schmidt, M.U. and Englert, U. (1995) Prediction of Crystal Structures for Organic and Organometallic Compounds, European Crystallographic Meeting 16, Lund/Sweden, 6.-11.8.1995.

-
27. Englert, U., Herberich, G.E. and Schmidt, M.U. (1993, submitted 1992) Vorhersage der Kristallstruktur von Pentamethylferrocen, *Z. Krist. Suppl.* **7**, 44.
28. Занин, И.Е., Антипин, М.Ю., Стручков, Ю.Т., Кудинов, А.Р. и Рыбинская, М.И. (1992) Молекулярная и кристаллическая структура пентаметилферроцена ($\eta^5\text{-C}_5\text{H}_5$)Fe($\eta^5\text{-C}_5\text{Me}_5$) в интервале 153-293 К. Анализ теплового движения в кристалле по рентгеновским дифракционным данным, *Металлорганическая Химия* **5**, 579-89. (Zanin, I.E., Antipin, M.Yu., Struchkov, Yu.T., Kudinov, A.R. and Rybinskaya, M.I. (1992) Molecular and crystal structure of pentamethylferrocene ($\eta^5\text{-C}_5\text{H}_5$)Fe($\eta^5\text{-C}_5\text{Me}_5$) in the interval 153-293 K. Analysis of the thermal motion in the crystal using X-ray diffraction data, *Metallorganicheskaya Khimiya* **5**, 579-89).
29. Keller, E. (1997). SCHAKAL97, Kristallographisches Institut der Universität, Freiburg, Germany.
30. Lommerse, J.P.M., Motherwell, W.D.S., Ammon, H.L., Dunitz, J.D., Gavezzotti, A., Hofmann, D.W.M., Leusen, F.J.J., Mooij, W.T.M., Price, S.L., Schweizer, B., Schmidt, M.U., van Eijck, B.P., Verwer, P., Williams, D.E. (2000): A test of crystal structure prediction of small organic molecules, *Acta Cryst.* **B**, in print.
31. Rietveld, H.M. (1969) A profile refinement method for nuclear and magnetic structures, *J. Appl. Cryst.* **2**, 65-71.
32. Schmidt, M.U. and Dinnebier, R.E. (1999) Crystal structures of organic pigments determined from X-ray powder data, IUCr-Workshop "Structure Solution from Powder Diffraction Data", Glasgow, Scotland, 4 August 1999.
33. Schmidt, M.U. and Dinnebier, R.E. (1999) Combination of energy minimizations and rigid body Rietveld refinement: The structure of 2,5-dihydroxy-benzo[*de*]benzo[4,5]imidazo[2,1-*a*]isoquinolin-7-one, *J. Appl. Cryst.* **32**, 178-186.
34. Schmidt, M.U. (1998) Crystal Structures of Organic Pigments Determined from X-ray Powder Data. Lecture on the Colour Science '98 conference, Harrogate/England, 30 March - 1 April 1998; Published in: Colour Science '98, Vol. 1: Dye and Pigment Chemistry (1999), 72-81.

Maximum Entropy Method Applied to Crystallographic Problems

Martin Schneider and Sander van Smaalen

Laboratory of Crystallography, University of Bayreuth, 95440 Bayreuth, Germany

Abstract

A short Introduction to the Maximum Entropy Method is given and applications to several crystallographic problems are discussed. The advantages over classical Fourier-transformation are shown.

Introduction

The principle of Maximum Entropy (MaxEnt) has found applications in a number of very different crystallographic problems [1]. It has been used in conjunction with direct methods to determine the phases of the structure factors from the experimental intensity data in x-ray diffraction [2]. MaxEnt has been used to derive accurate values for the integrated intensities of Bragg reflections [3]. Finally, it has been used to determine accurate electron density maps from phased reflections [4]. Classical structural analysis based on a model and a set of parameters, refined against the experimental data. Usually the model consists of independent spherical atoms, which are deformed to ellipsoidal shape through the use of anisotropic temperature factors. The electron density contained in chemical bonds is not considered by the model, therefore the model gives no information about it. Also disorder must be introduced by incorporation of additional parameters. There are several methods to extend the model of spherical atoms and anisotropic temperature factors with other properties [5]. With each special feature of a model new parameters are needed, thus extending the space of possible solutions and introducing the correlation between the parameters, while the deformations in the electron density are small. The correlation will have an adverse impact in the refined values of the parameters and their accuracies. A method to determine the electron density, that does not suffer from these adverse effects is provided by the Maximum Entropy Method (MEM)

The crystallographic Maximum Entropy Method

Definition of the entropy and the constraints

The Maximum Entropy Method is based on the Bayesian statistical theory. The entropy is defined as

$$S = -\sum_{j=1}^N \rho_j \log\left(\frac{\rho_j}{r_j}\right) \quad (1)$$

where \langle_j is the electron density at position j within the unit cell ($j=1\dots N$), r_j is the reference density. The maximum of S will give the most probable electron density of the crystal. The reference density r_j can incorporate any preknowledge about the problem. If there is no further information available, it is chosen flat which means the density values of all voxels are the same. With no information from the measured data the entropy will be maximised if \langle_j equals r_j , a quite reasonable result. The experimental data are introduced in so-called constraints. All constraints have the form $C_i=0$ and are additive terms to S , using the undetermined Lagrange multipliers technique. Note the MEM is model free at this point, but there were several model incorporated with the constraints.

The first constraint is the number of electrons c in the unitcell:

$$C_0 = \sum_{j=1}^N \rho_j - c = 0 \quad (2)$$

Reflections with known magnitude and phase are combined in the so-called F-constraint:

$$C_1 = \left(\sum_{j=1}^{N_1} \frac{|F_{MEM}(\vec{H}_j) - F_{obs}(\vec{H}_j)|^2}{\sigma_j^2} \right) - N_1 = 0 \quad (3)$$

with N_1 being the number of reflections contributing to the sum.

Reflections with the phases not known, or groups of overlapping reflections are collected into the G-constraint:

$$C_2 = \left(\sum_{j=1}^{N_2} \frac{|G_{MEM}^j - G_{obs}^j|^2}{\sigma_j^2} \right) - N_2 = 0 \quad \text{with} \quad G^j = \sqrt{\frac{\sum_{i=1}^{n_j} m_i |F(\vec{H}_i)|^2}{\sum_{i=1}^{n_j} m_i}} \quad (4)$$

m_i denoting the multiplicity of a reflection. In powder diffraction both types of constraints are needed. There is almost no chance to get a good result without using the F-constraint, since phase information is needed to pin the electron density to atoms. But it is advisable to use all available information, not only the F-constraint [6].

The complete function to maximise is then:

$$L = S + \lambda \sum_{i=1}^2 C_i \quad (5)$$

using a unified Lagrange multiplier for the constraints containing data. The iteration procedure starts from the reference density r_j , calculates the resulting structure factors F_{MEM} and computes the needed variation of the density in the asymmetric unitcell according to C . This procedure is repeated until the constraint C is satisfied.

Preparation of experimental data

The MEM expects the given F_{obs} as the Fouriertransformation of the electron density of **one** unitcell. Obviously the raw data (integrated intensities), have to be corrected and normalised to match that condition. These corrections implicitly introduce features of models into the MEM. Usually they are:

- Geometrical properties of the experiment (Lorentz and polarisation effects)
- absorption correction
- secondary extinction
- anomalous scattering
- absolute scale of the data

While the absorption correction depends on the atoms contained in the unit cell, secondary extinction depends on the refined mosaicity, anomalous scattering depends on the structure and therefore all those corrections depend on the model. All these corrections are necessary to achieve a meaningful comparison of the Fouriertransform of the electron density of one unitcell with the experimental data.

In order to determine the phases of the structure factors an approximate model is used. The obtained F_{calc} of the best model are used to apply all corrections in a single step [10]:

$$F_{obs}^{electron}(\vec{H}) = \frac{F_{calc}^{electron}(\vec{H})}{|F_{calc}(\vec{H})|} |F_{obs}(\vec{H})| \quad (6)$$

where the superscript 'electron' denotes values equal to the Fouriertransform of the corresponding electron density.

The deviations between the F_{calc} and the real/observed structure factors are contained in their magnitude, since the phases determine the positions of the atoms in the unit cell mainly. Therefore a good fitting model should give the proper phases, even if it does not cover all properties of the structure. It follows the MEM can be used favourable to determine structural features beyond that of a reasonable fitting model.

For the calculations we used the programs MEED [7] and BAYMEM[11]. Both programs are based on the same MaxEnt algorithm, a modified Gull-Daniell algorithm [7], and therefore have identical results.

Applications

Silicon

Based on a high precision measurement of 30 structure factors a MEM calculation of the density was carried out. Originally a non-nuclear maximum of the electron density was found at the mid-positions of the chemical bonds [4]. After a controversial discussion [8,9] the maxima in the middle of the bonds are considered as nonexistent. Figure 1 shows the Fouriertransform of F_{obs} at about bonding level. No electron density in the bonds is evident, due to the series termination effects. There are even positive walls and negative ditches between the atoms, where the density of the bonding should show up. The usual way to overcome the problem is to calculate a difference Fouriermap, which is shown in figure 2. The atoms and artefacts should cancel out, leaving the bonds in the map. But even here are negative ranges in the same order as the positive density! Both maps were computed with JANA 98 [12].

In comparison to the two Fouriermaps, the MEM-map in figure 3 is smooth, without any negative range and much more detailed. Artefacts are not critical to the interpretation of the bonding, even though they are still present as wavy overlay to the lines. The MEM is therefore much more suited to determine the electron density in the bonds and other fine structure of a electron density than Fourieranalysis.

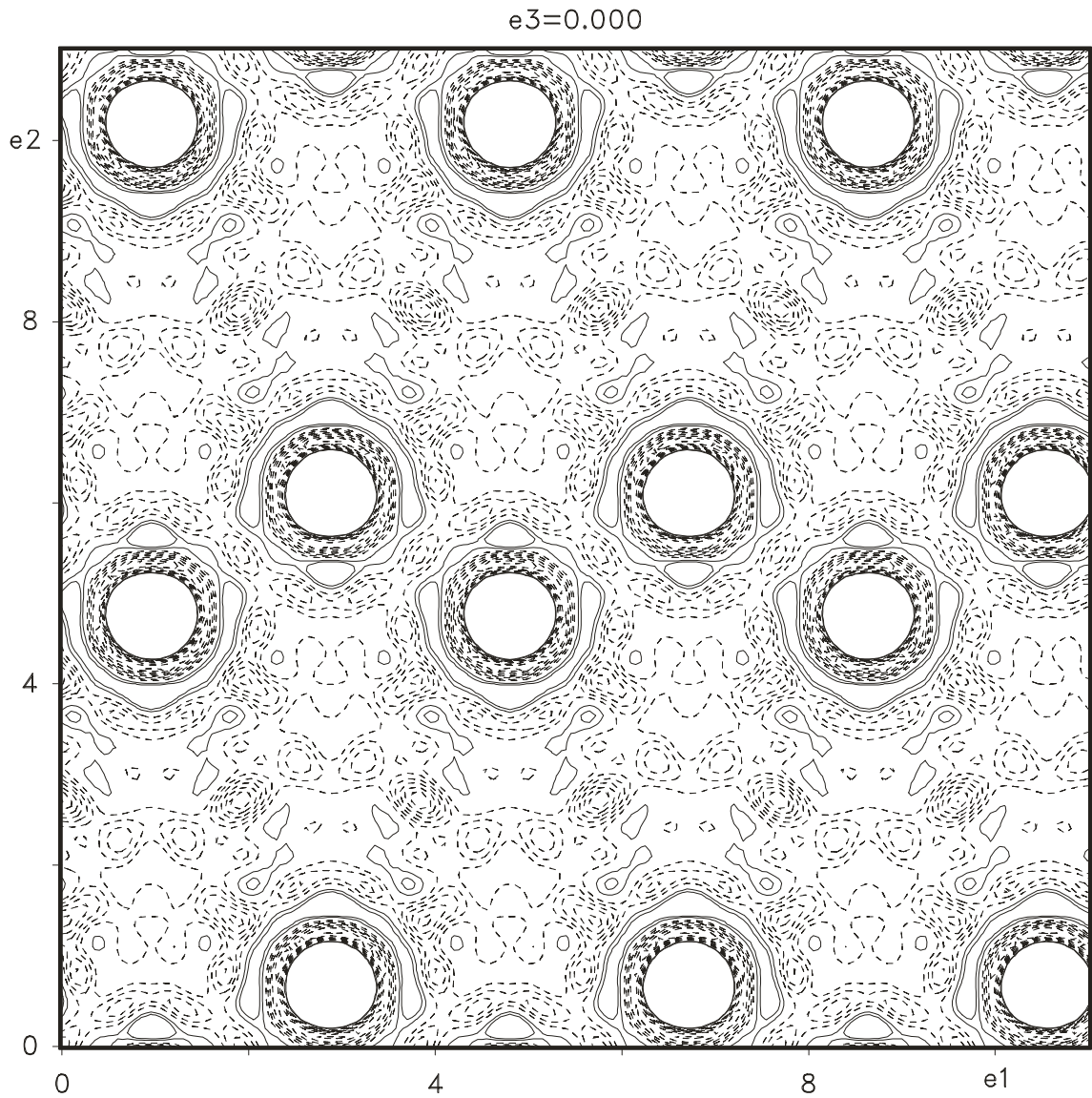


Figure 1: Fouriermap of Silicon along (110) showing the zig zag lines of Silicon atoms. No bonds are visible. Contourlevel is $0.5 \text{ e}/\text{\AA}^3$, cutoff at -4 and $1 \text{ e}/\text{\AA}^3$.

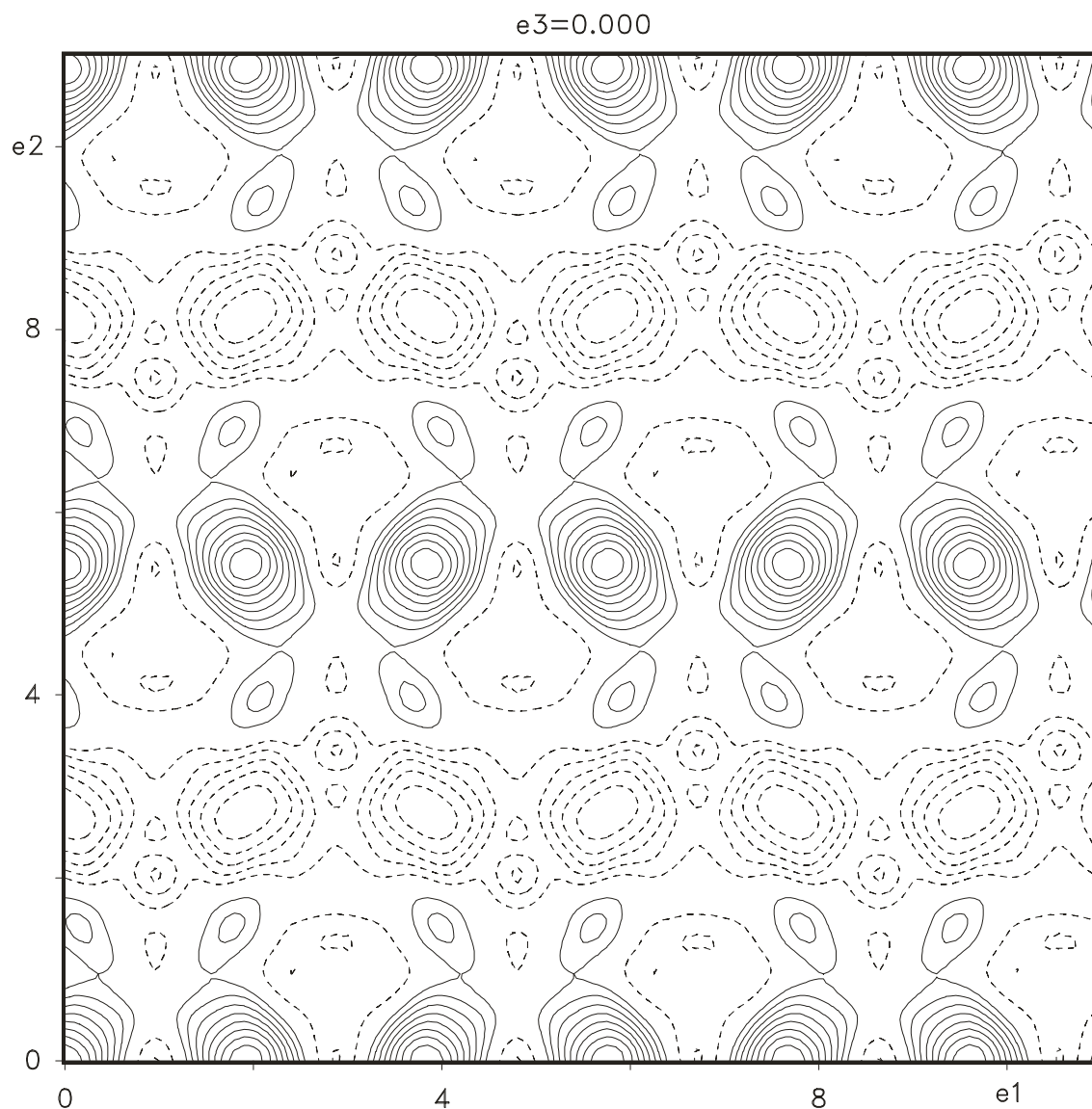


Figure 2: Difference Fourier map of Silicon along (110) showing the bonds of the Silicon atoms. Contourlevel is $0.015 \text{ e}/\text{\AA}^3$, cutoff at -0.1 and $0.15 \text{ e}/\text{\AA}^3$.

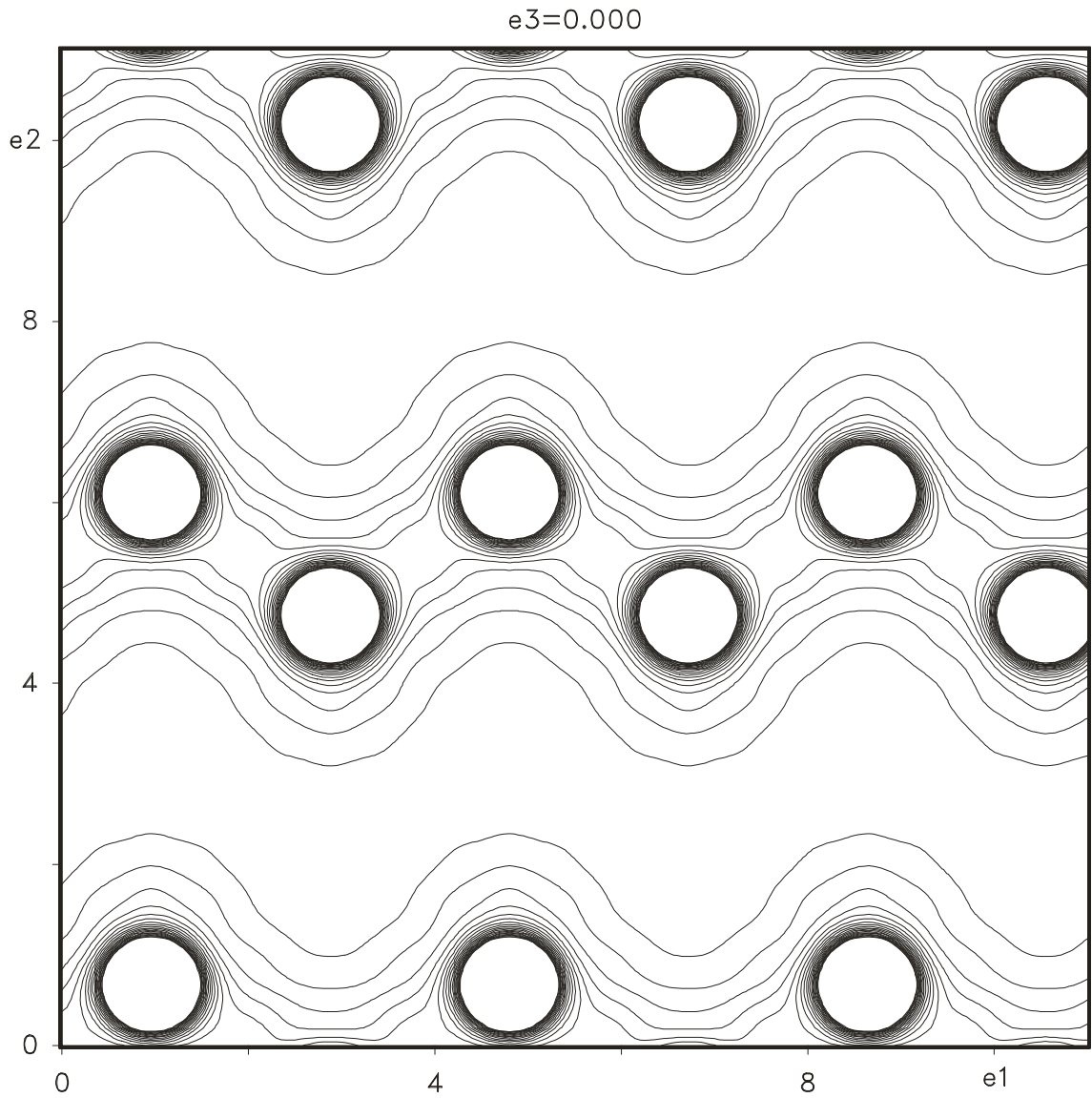


Figure 3: MEM-map of Silicon along (110) showing the zig zag lines of Silicon atoms with their bonds between them. Contourlevel is $0.1 \text{ e}/\text{\AA}^3$, cutoff at $1 \text{ e}/\text{\AA}^3$.

Disorder in LiCp*

LiCp* is an organometallic compound of lithium and the molecule pentamethylcyclopentane C₅(CH₃)₅ (Cp*). It crystallises in space group *R3m* with lattice parameters $a=14.7711 \text{ \AA}$ and $c=3.82206 \text{ \AA}$. The structure consists of chains of alternately Li and Cp*. X-ray powder diffraction gave the intensities of 33 reflections and intensities of 60 groups of overlapping reflections [6]. Rietveld refinement did not lead to an entirely satisfactory fit. The model in space group *R3*, in which the Cp* can rotate free around the chain axis, gave a better fit than using space group *R3m*, which pins the Cp* into one of three possible directions. In the *R3* refinement the Cp* was only 2° rotated out of the position of *R3m* with a standard deviation of 3°! Obviously the refinement cannot discriminate between the two symmetries.

The MEM reproduced the model in *R3* when 157 F_{calc} values were used in a F-constraint. The experimental data correspond to a F-constraint with 33 reflection and a G-constraint with 60 groups of reflections. With a flat initial density, MEED led to an electron density containing artefacts. Apparently, the number of 33 available phases was not sufficient to lead the iteration towards the true minimum. With the electron density of the model, either *R3* or *R3m*, as initial value, MEED converged smoothly towards a density showing the three-fold disorder of the Cp* molecule. With the phases and initial electron density from the *R3* refinement, MEED shifted the electron density towards the *R3m* symmetry. It was concluded, that the MEM gave evidence for *R3m* contrary to *R3*. Figure 4 presents the MEM density of three Cp* molecules with the threefold disorder as overlay.

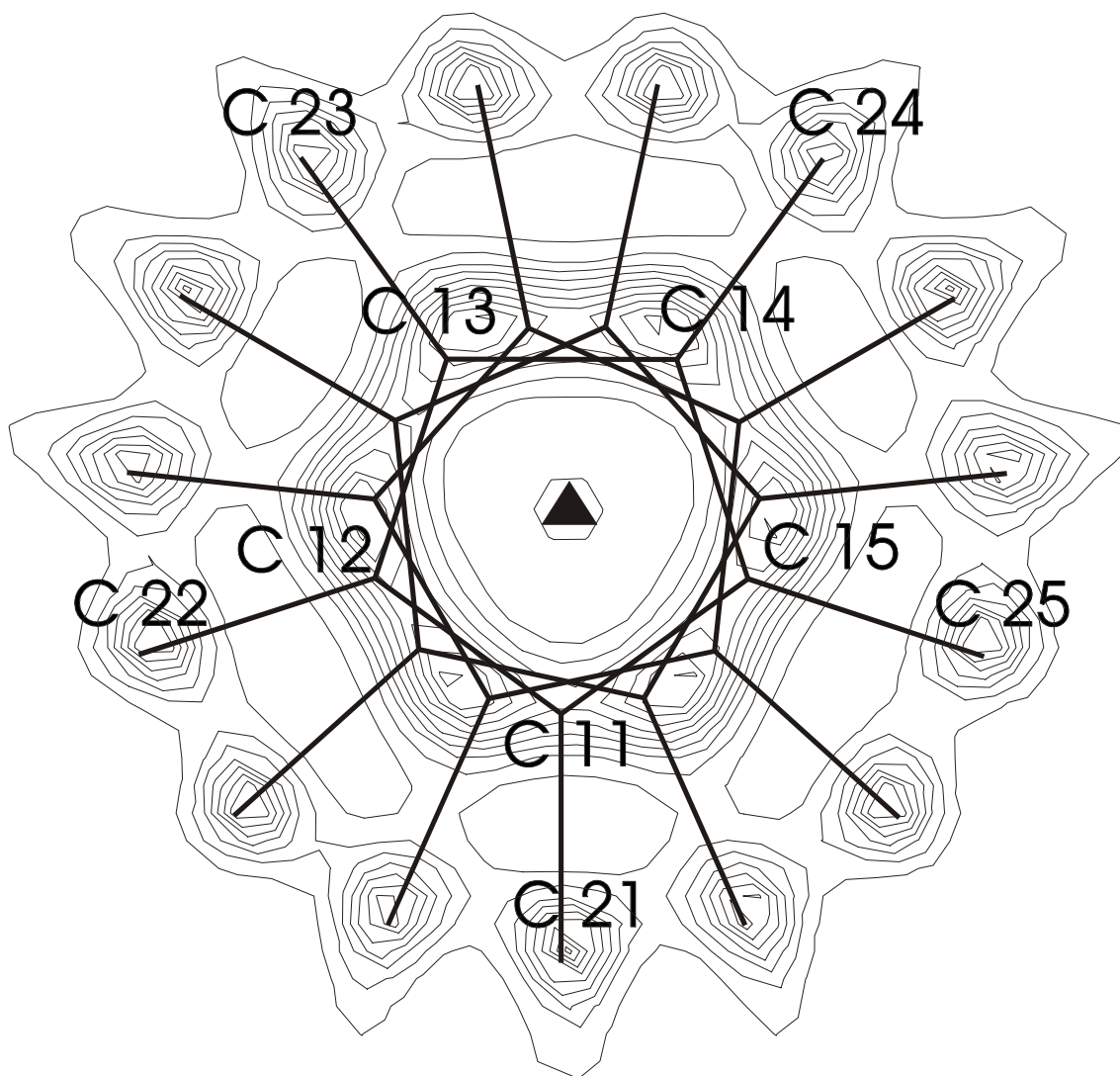


Figure 4: The threefold disorder of the Cp molecule in LiCp*. The 15 C-atoms of the outer ring are clearly visible, while the inner ring is smeared so that no individual atoms are shown. Superimposed are three models of the Cp*, one has its atoms named.*

References:

1. C. Gilmore, "Maximum entropy and bayesian statistics in crystallography: A review of practical applications", *Acta Cryst.*, **A52**, pp. 561-589, 1996.
2. C. Bricogne, "Direct phase determination by entropy maximisation and likelihood ranking: Status report and perspectives", *Acta Cryst.*, **D49**, pp. 37-60, 1993.
3. D. Sivia and W. David, "A bayesian approach to extracting structure-factor amplitudes from powder diffraction data", *Acta Cryst.*, **A50**, pp. 703-714, 1994
4. M. Sakata and M. Sata, "Accurate structure analysis by the Maximum-Entropy Method", *Acta Cryst.*, **A46**, pp. 263-270, 1990
5. C. Giacovazzo, ed., *Fundamentals of Crystallography*, 1990
6. R. Dinnebier, M. Schneider, S. van Smaalen, F. Olbrich and U. Behrens, "Disorder determined by high resolution powder diffraction: The structure of pentamethylcyclopentadienyllithium", *Acta Cryst.*, **B55**, pp. 35-44, 1999
7. S. Kumuzawa, Y. Kubota, M. Takata and M. Sakata, "MEED: A program package for electron-density-distribution calculation by the Maximum-Entropy Method", *J. Appl. Cryst.*, **26**, pp. 453-457, 1993
8. R.Y. de Vries, W. Briels, and D. Feil, "Critical analysis of non-nuclear electron-density maxima and the Maximum-Entropy Method", *Phys. Rev. Letters*, **77**, pp. 1719-1722, 1996.
9. M. Takata and M. Sakata, "The influence of the completeness of the data set on the charge density obtained with the Maximum-Entropy Method. A re-examination of the density distribution in Si", *Acta Cryst.*, **A52**, pp. 287-290, 1996
10. B. Bagautdinov, J. Lüdecke, M. Schneider and S. van Smaalen, "Disorder in the Crystal Structure of Cs₂HgCl₄ Studied by the Maximum Entropy Method", *Acta Cryst.*, **B54**, pp. 626-634, 1998
11. M. Schneider and S. van Smaalen, "BayMEM: A program for electron density determination of modulated crystals by the Maximum Entropy Method", in preparation
12. JANA 98, <http://www-xray.fzu.cz/jana/jana.html>

BRASS: The Bremen Rietveld Analysis and Structure Suite

*Reinhard X. Fischer**, *Thomas Messner**, *Dethart Kassner*⁺

**Fachbereich Geowissenschaften, Universität Bremen,
Klagenfurter Straße, D-28359 Bremen,
rfischer@uni-bremen.de*

⁺ *Glashütten-Oberems*

1 Introduction

After Rietveld's ALGOL program [1,2,3] has been introduced in the late 1960s it became readily available for the diffraction community when Hewat [4] and Wiles & Young [5] distributed FORTRAN versions of the program. At present, there are numerous programs, both public domain and commercial software, which represent a wide variety of options for the analysis and refinement of crystal structures using the Rietveld method. Lists of programs together with short descriptions of the available options are published on several websites, e.g., <http://www.iucr.org/iucr-top/comm/cpd/rietveld.html> which also contains some links to Rietveld related subjects and other fields of crystallography.

The Rietveld method is essentially a refinement procedure of profile and crystal structure parameters based on single step intensities. However, the results of the least squares procedure allow a detailed analysis of the crystal chemical properties of the material, and many Rietveld programs have a direct link to structure determination programs including Fourier codes for the calculation of electron density maps. An analysis of the peak profiles with respect to general or anisotropic peak broadening yields some clues on crystallite sizes, internal stress and strain, and disorder effects. Therefore, the Rietveld refinement method is often termed just "Rietveld analysis" which covers all the features interacting with the least squares procedure. This term is used here in its broader sense to describe the suite of programs.

The program package *BRASS* has been developed to fill in some gaps existing in current state of the art programs, especially concerning the crystal chemical interpretation of the results. However, it is impossible to implement all features and powerful tools of the available Rietveld software in the first release of our program. Nevertheless, we do hope that the Bremen Suite will complement the existing programs and will achieve a high standard in its future developments. All programs in *BRASS* run under Windows NT or Windows 95 (or higher).

2 The Rietveld kernel

The Rietveld kernel program is based on the PC-Rietveld plus program [6], extensively modified and extended with several new options. The roots of the program date back to the LHPM program [7] which is a further development of the DBW code [5]. The most important extension in the new program is the introduction of distance least squares (DLS) calculations which are applied to the interatomic bonds predefined by the user or automatically set by the program. Alternatively, an input file can be generated for the standalone DLS76 program [8]. Weights can be applied to the distance restraints which can be released successively during the refinement. The application of geometric restraints to the crystal structure refinement is especially useful when parts of the crystal structure are known and the analysis can be focused on the determination of unknown parameters, atoms, molecules, or structural units.

The background is modeled by spline interpolations between adjacent setpoints. The height of the setpoints, not their 2θ -positions, can be refined together with the profile and structural parameters, thus permitting a high flexibility to adjust to highly undulated background intensities due to incoherent scattering from amorphous components or sample holder contributions. The 2θ -positions are fixed to avoid any interactions with the Bragg intensities which could compensate the misfit between the observed intensities and the intensities calculated from the model.

Several parts of the old program [6] are completely rewritten. The symmetry generation routine has been replaced by the code described in ref. [9] which is much more stable and independent of standard settings of the space group symbol. However, the default setting of the unit cell can be changed by entering a transformation matrix. The whole input/output handling is reorganized accounting for various input and output formats.

A sophisticated online display shows the progress in each step of the refinements. The full history of the results in each of the refinement cycles is stored which permits an analysis of the individual shifts and convergence behavior of the variable parameters.

Observed and calculated structure factors are passed to the Fourier program to calculate electron density maps. Alternatively, the program performs a grid search analysis where the positional parameters of a dummy atom are refined on a grid in the asymmetric unit of the unit cell together with its occupancy. The occupancy factors and the residuals are then plotted in xy-layers of the unit cell. In this fashion, atom positions can be located that are difficult to obtain from Fourier or difference Fourier maps. This procedure was introduced and described in ref. [10], successfully applied to the determination of non framework atoms in zeolites [11].

The program kernel in its first release is written in FORTRAN which communicates with the Windows user interface via DLLs.

3 Electron density and grid search maps

Electron density, or (pseudo) nuclear density maps from neutron diffraction data, are calculated from the structure factors passed after the last refinement cycle from the Rietveld program to the Fourier program. Maxima and minima are interpolated and sorted by height. Peak positions are directly returned to the Rietveld program and entered into the program calculating distances and angles. The results can be plotted in layers of the unit cell as contour maps (Fig. #.1) or in a three dimensional representation as shown in Fig. #.2. The same graphics routine is used for the representation of the grid search results displaying occupancies and residuals in layers of the unit cell.

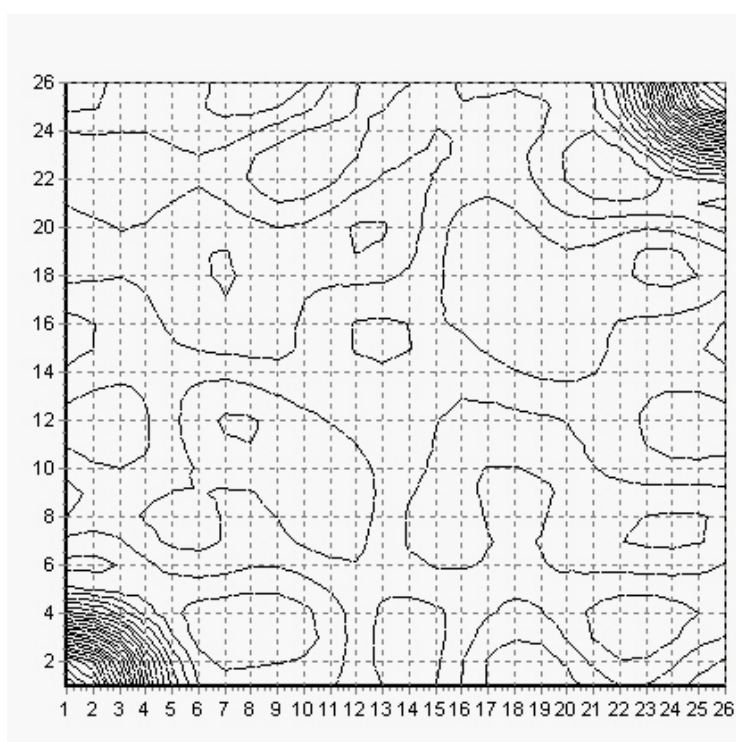


Fig. 1: Contour map of the residual electron density distribution from a difference Fourier analysis of Cr-Mullite [12]; $z = 0$; $0 < x,y < 0.5$. The maxima in $0, 0, 0$ and $\frac{1}{2}, \frac{1}{2}, \frac{1}{2}$ indicate the position of the Cr atom.

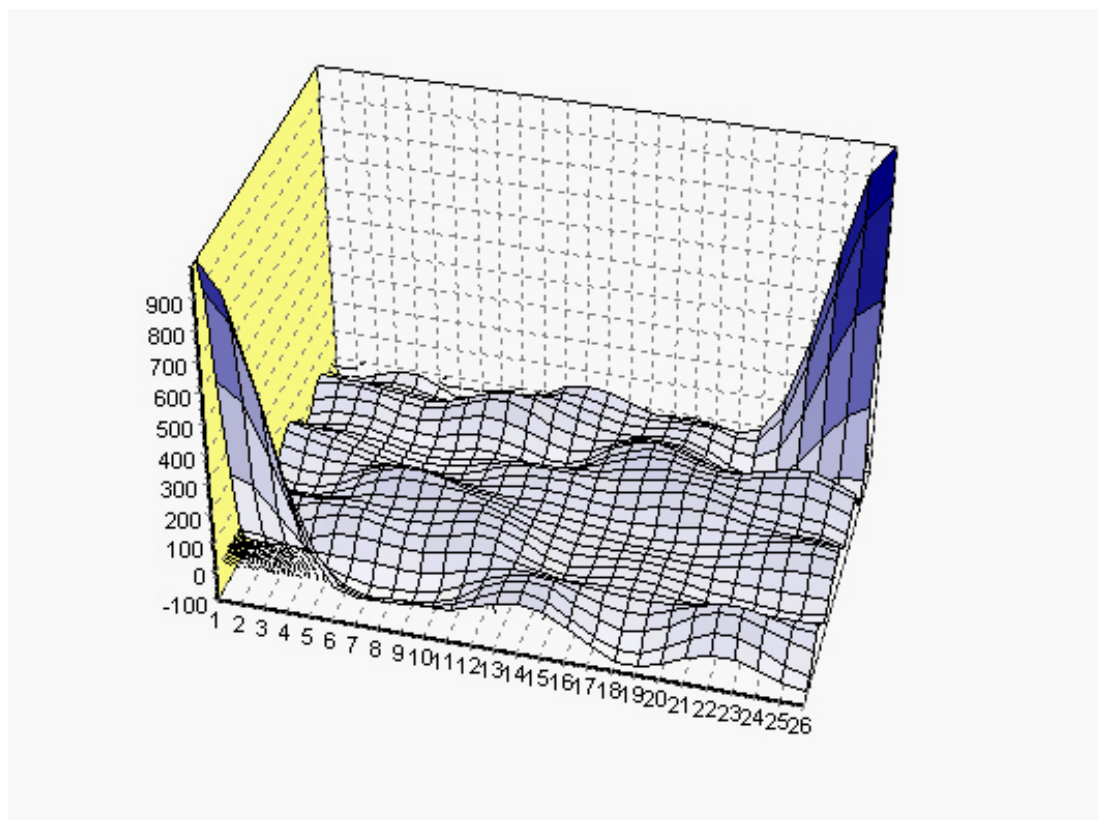


Fig. 2: The same layer as shown in Fig. #.1 in three-dimensional representation.

4 Powder pattern display

Observed and calculated intensities are displayed as dotted lines with small crosses indicating the single step intensities or they are drawn as solid lines. They can be superimposed or vertically shifted. Multiphase diagrams can be decomposed into the single phase contributions. The usual zoom and scroll options are available. Multiple diagrams can be displayed in superposition mode or in a three-dimensional representation. Angle positions, d-values, intensities, and indices are shown indicating the actual values at the cursor position. The difference plot (observed minus calculated step intensities) is shown underneath. The difference curve can be linear or it can be scaled by the weighting factor used in the least squares refinement. The latter mode has the advantage of representing the actual least squares fit which is based on the minimization of the square root of the sum of the weighted differences of the squared intensities. Initial setpoints for the background definition are set by simple mouse clicks at the respective positions.

5 Crystal chemical calculations

Comprehensive options are implemented for crystal chemical calculations. Interatomic distances and angles are listed for each of the atoms in the asymmetric unit. They are compared with the sum of the effective ionic radii [13] of the two neighboring atoms forming the bond. If more atoms occupy the same position, their contributions to the bond are weighted by their occupancies. In the case of aluminosilicates, the statistical distribution of Si and Al atoms on one site is calculated from the observed interatomic distance and compared with the given or refined occupancy factors.

Coordination sequences are calculated for all atoms in the asymmetric unit for the first ten levels. In framework compounds, especially zeolite like compounds, the coordination sequences are calculated also among the T-atoms ignoring the oxygen contacts. These values are used to identify zeolite like compounds irrespective of their real symmetry and space group settings.

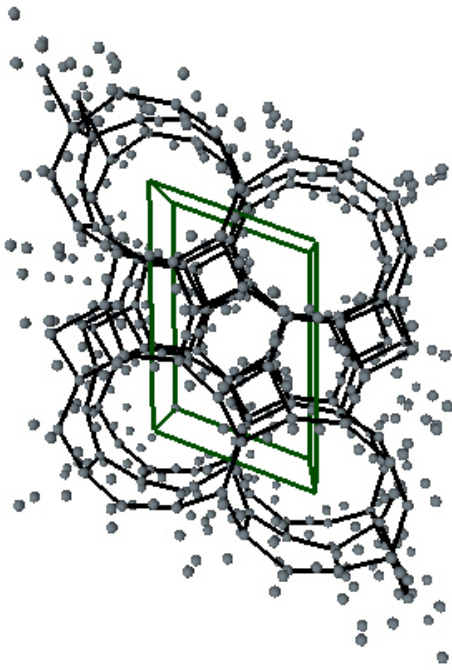
The program searches for structural units in framework compounds and identifies polyhedral units from the catalogue of units described in ref. [14]. Zeolite type structures can be transformed to the standard setting of zeolite structures as given in [15] for the 130 different zeolite types currently known [16].

Optical properties like electronic, ionic, and dielectric polarizabilities, and mean refractive indices are calculated using the ion additivity rule and the data from ref. [17] and [18].

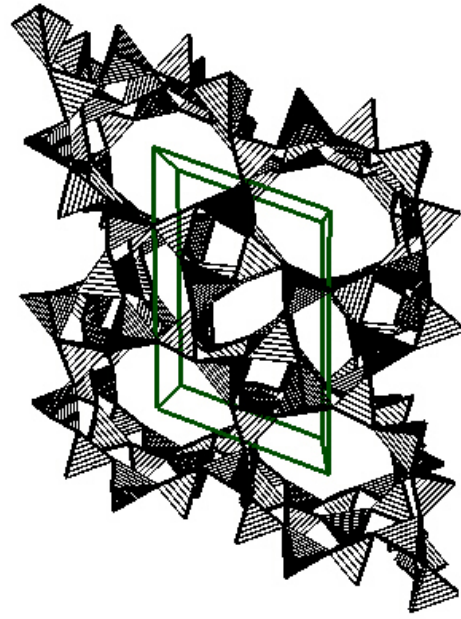
For those elements which have lone pair electrons, the stereochemical influence of the lone-pair electrons on the geometry of the coordination polyhedron is calculated using the approach of ref. [19].

6. Crystal structure drawings

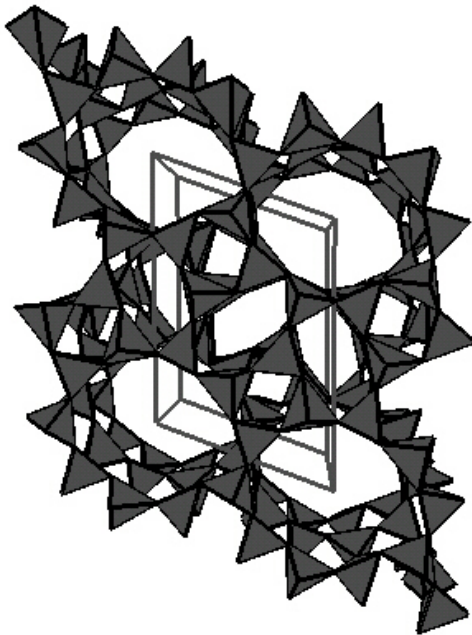
BRASS contains a new program for the display of crystal structures in ball and stick mode or in polyhedral representation. The program is based on the old code of the STRUPLO program [20], but completely rewritten in Delphi and reorganized to conform to the Windows type programs. Sample plots are shown in Fig. #.3. A special feature is the online selection mode of structural units like atoms and polyhedra. The units are copied to a separate window by mouse click on the respective unit. There, the selected units, especially complex polyhedra, can be viewed and rotated without the overlapping neighbors. That way, the layout of single units (color, hatching etc.) can be modified easily. Interatomic distances are calculated upon mouse click on two atoms. All single objects (atoms, bonds, polyhedra) can be deleted by mouse click on the object. A comprehensive redo option allows the recovery of all or just individual actions.



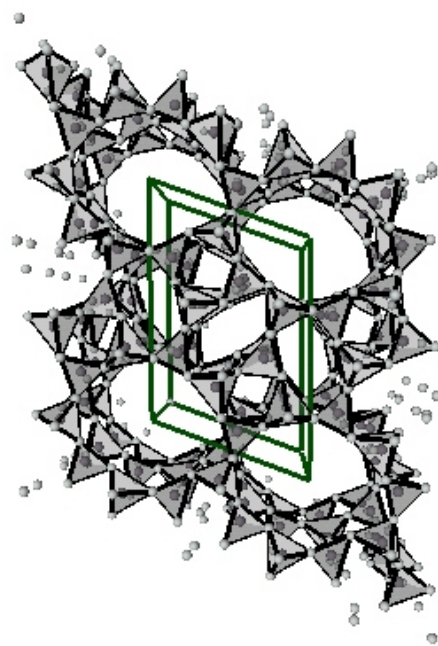
a) Framework atoms are connected by solid lines. Non framework atoms are shown as spheres. Perspective view.



b) TO₄-tetrahedra are represented as solid polyhedra with hatched faces. Perspective view.



c) TO₄-tetrahedra are shown as solid polyhedra



d) TO₄-tetrahedra are shown as semi-transparent polyhedra.

Fig. 3: Crystal structure drawings of AlPO₄-41 [21] in various representations.

7. Availability and Acknowledgement

The program will be distributed free of charge after finishing the beta-testing in 2001. Support from Philips Analytical, Almelo (Netherlands) is gratefully acknowledged.

8. References

- [1] Rietveld, H.M.: Line profiles of neutron powder-diffraction peaks for structure refinement. *Acta Cryst.* **22** (1967) 151-152.
- [2] Rietveld, H.M.: An ALGOL program for the refinement of nuclear and magnetic structures by the profile method. Reactor Centrum Nederland, Research Report 1967.
- [3] Rietveld, H.M.: A profile refinement method for nuclear and magnetic structures. *J. Appl. Cryst.* **2** (1969) 65.
- [4] Hewat, A.W.: The Rietveld computer program for the profile refinement of neutron diffraction powder patterns modified for anisotropic thermal vibrations. Atomic Energy Research Establishment, Harwell 1973.
- [5] Wiles, D.B., Young, R.A.: A new computer program for Rietveld analysis of X-ray powder diffraction patterns. *J. Appl. Cryst.* **14** (1981) 149-151.
- [6] Fischer, R.X., Lengauer, C., Tillmanns, E., Ensink, R.J., Reiss, C.A., Fantner, E.J.: PC-Rietveld plus, a comprehensive Rietveld analysis package for PC. *Materials Science Forum* **133-136** (1993) 287-292.
- [7] Hill, R.J., Howard, C.J.: A computer program for Rietveld analysis of fixed wavelength X-ray and neutron powder diffraction patterns. Australian Atomic Energy Commission Research Establishment, Research Report M112 (1986).
- [8] Bärlocher, C., Hepp, A., Meier, W.M.: DLS-76, a program for the simulation of crystal structures by geometric refinement. ETH Zürich (1977).
- [9] Burzlaff, H., Hountas, A.: Computer program for the derivation of symmetry operations from the space-group symbols. *J. Appl. Cryst.* **15** (1982) 464-467.
- [10] Baur, W.H., Fischer, R.X.: Recognition and treatment of background problems in neutron powder diffraction refinements. In: *Advances in X-ray analysis*, Vol. 29 (Eds.: Barrett, C.S., Cohen, J.B., Faber, J., Jenkins, R., Leyden, D.E., Russ, J.C., Predecki, P.K.). Plenum Publishing Corporation (1986) 131-141.
- [11] Fischer, R.X., Baur, W.H.: New techniques of background corrections and atom search methods applied to zeolite RHO. *Fortschr. Mineral.* **64**, Beiheft 1, (1986) 48.
- [12] Fischer, R.X., Schneider, H.: Crystal structure of Cr-Mullite. *Amer. Mineral.* **85** (2000) 1175-1180.

- [13] Shannon, R.D.: Revised effective ionic radii and systematic studies of interatomic distances in halides and chalcogenides. *Acta Cryst.* **A32** (1976) 751-767.
- [14] Smith, J.V.: Tetrahedral frameworks of zeolites, clathrates, and related materials. Landolt-Börnstein Group IV, Physical Chemistry. Vol. 14A (Eds.: Baur, W.H., Fischer, R.X.) Springer (2000).
- [15] Baur, W.H., Fischer, R.X.: Zeolite-type crystal structures and their chemistry. Landolt-Börnstein Group IV, Physical Chemistry. Vol. 14B (Eds.: Baur, W.H., Fischer, R.X.). Springer (2000). (Vols. 14C to 14E in preparation).
- [16] Meier, W.M., Olson, D.H., Baerlocher, C.: Atlas of zeolite structure types. Web site of the Structure Commission of the International Zeolite Association at the ETH Zürich, <http://www.iza.ethz.ch/IZA-SC/Atlas/AtlasHome.html>).
- [17] Shannon, R.D.: Dielectric polarizabilities of ions in oxides and fluorides. *J. Appl. Phys.* **73** (1993) 348-366.
- [18] Shannon, R.D., Shannon, R.C., Fischer, R.X.: Empirical electronic polarizabilities in oxides and fluorides. In preparation.
- [19] Wang, X., Liebau, F.: Influence of lone-pair electrons of cations on bond-valence parameters. *Z. Krist.* **211** (1996) 437-439.
- [20] Fischer, R.X.: STRUPLO84, a Fortran plot program for crystal structure illustrations in polyhedral representation. *J. Appl. Cryst.* **18** (1985) 258-262.
- [21] Kirchner, R.M., Bennett, J.M.: The structure of calcined AlPO_4 -41: a new framework topology containing one-dimensional 10-ring pores. *Zeolites* **14** (1994) 523-528.

High-Pressure Powder Diffraction

Ross J. Angel

*Bayerisches Geoinstitut,
Universität Bayreuth,
D95440 Bayreuth, Germany*

Introduction

Diffraction studies at high pressures provide an opportunity to probe the behaviour of the chemical bonding of solids as a function of decreasing inter-atomic separation, without the complications introduced by changing chemistry. In this contribution to the workshop, the methods available for performing high-pressure powder diffraction are briefly introduced along with a number of important cautions for the experimentalist new to high-pressure diffraction.

In general, powder diffraction methods at high pressures yield data that is of lower quality than that which is obtainable from the same sample measured at ambient conditions on the same instrument. The reasons for this are several. High-pressure apparatus generally only allow small sample volumes while the containment of the sample absorbs both the incident and diffracted beams, all of which reduce the intensities of the diffracted beams. Secondly, the containment of the sample can give rise to scattered radiation that appears as background in the detector, either by diffraction from pressure-cell components (especially the gasket) or other scattering processes including Compton scattering from the diamonds of diamond-anvil cells (DACs). The lower signal levels can only be overcome by longer counting times and more intense radiation sources. The higher levels of background are best addressed by appropriate shielding and/or collimation at the time of the experiment.

Thus, the most important differences between performing a powder diffraction experiment at high pressure and one at ambient conditions are in the data collection process. Some important points are reviewed here in some detail. Then, a few brief remarks are made about the refinement of high-pressure powder diffraction data, which is essentially no different from refinement of data collected at ambient conditions. Finally, a brief review of how to analyse pressure-volume data in the form of Equations of State (EoS) is presented, along with a program to do the necessary least-squares fitting.

Experimental methods

X-ray powder diffraction.

Synchrotrons have been the X-ray source of choice for some time for high-pressure powder diffraction, which can be performed, with care, to pressures in excess of 100 GPa. A wide variety of diamond-anvil cell designs have been developed for high-pressure powder diffraction, but many are based on the very simple principle of applying a moderate force to the relatively large surface-area of the thread of a drive screw, and then transferring this force onto the small area of the tips of two opposed diamond-anvils. An alternative drive mechanism is used in the “gas-membrane cells”, in which the force is generated by inflating a metal membrane whose expansion drives the anvils together. DACs that have been specifically designed for single-crystal diffraction can also be successfully used for powder diffraction. An extensive review of DAC designs, including their relative advantages and disadvantages, can be found in Hazen (2000).

Most early X-ray powder diffraction studies at high pressures employed energy-dispersive diffraction, a method which is limited in resolution because of the restricted energy resolution of the solid-state detectors involved. The advent of image-plate detectors and improved DAC designs with larger opening angles to allow X-ray access to the sample has resulted in angle-dispersive diffraction becoming the standard method. The data quality from image plates is also greatly increased because the entire diffraction cone can be collected and therefore effects due to, for example, sample texture can be readily identified before integration of the data into a conventional 1-dimensional Intensity vs. 2θ data-set used for refinement. Currently, data quality is such that reliable unit-cell parameters can be obtained from high-pressure powder diffraction as well as structural data in more simple systems. The recent introduction of in-situ read-out from image plates that allows data to be collected and processed on a ~ 1 minute cycle (Thoms et al., 1998) makes these detectors competitive with other area detectors such as CCD-based systems for real-time studies of phase transitions, and rapid measurements of compressibility.

Neutron powder diffraction.

Neutron diffraction is the method of choice for studies of materials containing both light and heavy atoms. For precise studies up to 0.5 GPa there are a wide variety of gas-pressure cells suitable for both angle-dispersive and time-of-flight diffraction. For slightly higher pressures there are a variety of clamp cells, the latest

developments of which can reach pressures of 3.5 GPa and temperatures in excess of 800 K (e.g. Knorr et al. 1997, 1999). Scaled-up opposed-anvil cells equipped with sapphire anvils have been used to pressures of at least 3 GPa (e.g. Kuhs et al. 1996). For higher pressures there is the Paris-Edinburgh cell which is capable of developing pressures of up to 25 GPa (Besson et al. 1992, Klotz et al. 1998).

Pressure Media

It is very important to perform high-pressure diffraction experiments under well-defined conditions of applied stress. In effect, this means that a hydrostatic pressure medium must be used to enclose the crystal, because non-hydrostatic stresses in a high-pressure device are very difficult to quantify. The effects of non-hydrostatic stresses include, but are probably not limited to -

- Broadening of diffraction peaks and consequently reduction in signal-to-noise ratios.
- Incorrect measurement of pressures (see the example of the ruby pressure sensor, below).
- Displacement of the transition pressures of ferroelastic or coelastic structural phase transitions (e.g. Decker et al. 1979).

A 4:1 mixture by volume of methanol:ethanol remains hydrostatic to just over 10 GPa (Eggert et al. 1992) and is convenient and suitable for many studies. If the sample dissolves in alcohols, then a mixture of pentane and iso-pentane which remains hydrostatic to ~ 6 GPa (Nomura et al. 1982), or a solidified gas such as N₂, He, or Ar can be employed. Water appears to remain hydrostatic to about 2.5 GPa at room temperature, just above the phase transition from ice-VI to ice-VII (Angel, unpublished data). The solid pressure media such as NaCl or KCl favoured by spectroscopists are very non-hydrostatic even at pressures below 1 GPa. Similarly, the “fluorinert” material used in many neutron diffraction experiments because of its low neutron scattering power becomes significantly non-hydrostatic at ~1.3 GPa. At pressures in excess of the hydrostatic limit of the solidified gas and fluid pressure media, the non-hydrostatic stresses can be relaxed after each change in pressure by annealing the sample chamber, either by laser-heating or an external resistance furnace. For example, heating a cell in which the ethanol:methanol mixture is the pressure fluid to 150-200 °C for about 1 hour is sufficient to relax the non-hydrostatic stresses developed above 10 GPa (Sinogeikin and Bass, 1999).

Ideally each grain of the sample should be completely surrounded by the pressure fluid, and not make contact with any other grain because grain-grain contacts can lead to the development of non-hydrostatic stresses. This criterion can not be achieved perfectly, except by resort to a single-crystal experiment! But in loading a high-pressure cell for powder diffraction, it is important not to over-fill the sample chamber with sample - for DACs and X-ray diffraction the author's experience is that the lightly-compressed powder should occupy approximately one-half of the volume of the hole in the gasket.

Pressure Measurement

The ruby fluorescence method is the most commonly used to determine pressure in diamond-anvil cell measurements. It is based upon the observation that a pair of electronic transitions in the Cr^{3+} dopant atoms in Al_2O_3 change in energy as the Al_2O_3 lattice is compressed. The fluorescence in the red area of the optical spectrum is strong and easily excited by blue/green laser light, and the shift is quite large, approximately $3.6\text{\AA}/\text{GPa}$. Unfortunately, the fluorescence wavelength is also very sensitive both to temperature, such that a 5° temperature change gives rise to a shift equivalent to 0.1 GPa (Barnett et al. 1973, Vos and Schouten 1991). It is also sensitive to the c/a ratio of the Al_2O_3 host lattice (Sharma and Gupta 1991). As a result, non-hydrostatic stresses increase the observed shift of the stronger R_1 component of the doublet, and can yield an apparent pressure that is higher than the true pressure (Gupta and Shen 1991, Chai and Brown 1996). Other fluorescence sensors have also been employed; for reviews see Holzappel (1997) and Miletich et al. (2000). Measurement of optical fluorescence is relatively fast, and is extremely useful for setting the approximate pressure in a DAC prior to a diffraction measurement. With the proper precautions it can yield pressures as precise as 0.1 GPa, provided temperature fluctuations are completely excluded. In reality, these and other factors often mean that 0.3 GPa is a more realistic estimate of the precision. For more precise pressure determination internal diffraction standards can be used in DACs, while this is essential for completely enclosed cells, such as the Paris-Edinburgh cell. The pressure is then determined from the unit-cell volume of the standard and its EoS. The precision in pressure then depends upon the precision of the volume measurement and the bulk modulus of the material; the softer the standard the more precise the pressure determination. Materials in common use as standards at pressures up to 10 GPa include NaCl (Brown 1999), quartz (Angel et al. 1997) and fluorite (Hazen and Finger 1981, Angel 1993), while metals such as gold (e.g. Heinz and Jeanloz 1984) have been used at higher pressures. It is important to note that there is no absolute pressure standard measurement above 2.5 GPa, so all EoS and all pressure scales are provisional and subject to revision in the light of

improved calibrations. As an example, the pressure scale based upon the EoS of NaCl which was introduced by Decker (1971) and developed by Birch (1986) was recently shown to be in significant error by Brown (1999).

Experimental Strategy.

Because all pressure cells affect the intensities in the diffraction pattern of the sample, and imperfect alignment and positioning of the cell with respect to the incident beam and the detector can influence the positions of the diffracted beams, it is strongly recommended that all high-pressure studies should commence with a measurement of the diffraction pattern of the sample in the pressure cell, but at ambient pressure. Comparison of the result with a refinement of the diffraction pattern measured outside of the cell will provide an indication of the systematic errors present in the high-pressure results. For EoS determination, the room pressure determination of the unit-cell volume also provides a strong constraint on the EoS parameters. In all subsequent analysis, the room pressure datum collected from within the cell should be used as the room-pressure reference. Use of data collected outside of the cell as this reference often leads to false conclusions being drawn about the high-pressure evolution of structure or unit-cell parameters (e.g. Hazen and Finger, 1989).

If the only intention of the experiment is to derive the parameters of the EoS of the sample, then the number of data points required to obtain a given precision in the EoS parameters can be estimated from the known experimental parameters. The details are given in Angel (2000a). To summarise, the number of data required can be estimated from the bulk modulus of the sample (usually known approximately, or can be estimated), the precision of the volume and pressure measurements and the maximum pressure achievable.

Refining high-pressure powder diffraction data

As noted above, the refinement of powder-diffraction data by the Rietveld method is not intrinsically different from the refinement of data collected under any other conditions. Just a few notes are given here, based on the author's own, but somewhat limited, experience.

The reduction of the data collected by the detector, whether time-of-flight neutron powder diffraction data or a 2-dimensional image from an area detector used in X-ray powder diffraction, is usually performed by software provided by the facility. This data reduction usually includes, if appropriate, corrections for spatial

distortion and calibration of the detector (e.g. Fit2D, Hammersley et al. 1995, 1996) and normalisation to the incident beam spectrum. An important consideration for X-ray powder diffraction is the calibration of the wavelength and the sample-to-detector distance, normally obtained by performing diffraction from a standard sample. Repeat measurements of the standard over the period of the experiment will provide an estimate of the reproducibility and variation of these parameters.

Because the background levels in a high-pressure diffraction pattern are much higher than in an ambient pressure measurement, the question of whether the background should be subtracted before refinement rears its ugly head. The statistical arguments for and against such a process have been made before in the literature. My personal view is that the background should be fitted as far as possible, although specific regions in which the background changes sharply may be excluded entirely from the fitting process. Within GSAS, the best background function for most high-pressure data appears to be the shifted-Chebyshev polynomial. For some cases, the refinement becomes more stable if a set of fixed background points are determined to provide a basis for the refinement process. Diffraction peaks from the cell components can either be excluded (if they are few and they do not interfere with the sample), or they can be refined as further phases. If these phases are refined, it is quite normal to have to resort to a Le-Bail fit to obtain a reasonable approximation to the observed intensities, as one expects cell components to have strong preferred orientation and possibly an exceedingly poor powder average. The same cautions apply to the diffraction pattern from the sample itself, especially as the application of pressure can often lead to the development of strong preferred orientation. Lastly, the peak widths of the samples should be monitored carefully from pattern to pattern in order to detect the possible onset of non-hydrostatic conditions. However, this is not a guaranteed test for non-hydrostatic conditions as it is possible that these can affect a pressure sensor such as ruby at pressures *below* that at which broadening is first detected in the sample.

In summary, refinements of high-pressure data should, at minimum, include refinement of the background, together with the unit-cell parameters, scale factors and peak-widths of each phase in the sample. If image-plate data is being fitted, the 2θ zero of the pattern should also be refined as the integration methods do not locate this to better than about 1 pixel in the 2-dimensional image. Refinement of further parameters, including structural parameters, depends on the quality of the pattern.

Equations of state

Formulations of Equations of State.

The unit-cell parameters and unit-cell volume of a solid normally vary in a non-linear way with pressure because as the volume of the solid becomes smaller the inter-atomic forces opposing further compression become stronger. The “stiffness” of a solid is characterised by the bulk modulus, defined as $K = -V \partial P / \partial V$ which will generally increase with increasing pressure. Different assumptions can then be made about how K varies with P , or how V varies with P . Each set of assumptions then leads to a relationship between P and V known as an “Equation of State” or EoS. Note that, unlike “ideal gases” there is no absolute thermodynamic basis for specifying the correct form of the EoS of solids, although for simple solids such as the NaCl structure a direct relationship between the inter-atomic potential and the EoS can be derived (see Anderson 1995 for a thorough review).

Measured equations of state are usually parameterized in terms of the values of the bulk modulus and its pressure derivatives, $K' = \partial K / \partial P$ and $K'' = \partial^2 K / \partial P^2$, evaluated at zero pressure. These zero-pressure moduli are normally denoted by a subscript “0”, thus: $K_0 = -V_0 (\partial P / \partial V)_{P=0}$, $K'_0 = (\partial K / \partial P)_{P=0}$, and $K''_0 = (\partial^2 K / \partial P^2)_{P=0}$. High-pressure diffraction measurements are isothermal measurements, so in the following all references to bulk modulus, K_0 , and its derivatives K'_0 and K''_0 , refer to isothermal values and all compression values, $\eta = V / V_0$, and variables such as finite strain f derived from them, are similarly isothermal quantities. The relationship between the isothermal bulk modulus, more generally denoted K_T , and the adiabatic bulk modulus K_S that describes compression in a thermally closed system (at constant entropy) is $K_S = K_T (1 + \alpha \gamma T)$ where α is the volume thermal expansion coefficient and γ is the Gruneisen parameter. At room temperature the factor $\alpha \gamma T$ is normally in the range of 0.01-0.02 for most ceramic solids.

The EoS most commonly used for fitting isothermal (i.e. P - V datasets) are listed briefly below. Further details of the derivations and limitations can be found in, for example, Anderson (1995) and Angel (2000a).

Murnaghan. This can be derived from the assumption that the bulk modulus varies linearly with pressure, $K = K_0 + K'_0 P$; K'_0 being independent of pressure. Integration yields the P - V relationship:

$$V = V_0 \left(1 + \frac{K'_0 P}{K_0} \right)^{-1/K'_0} \quad (1)$$

This EoS (Murnaghan, 1937) both reproduces P - V data and yields correct values of the room pressure bulk modulus for compressions up to about 10% (i.e. $V/V_0 > 0.9$), and has the advantage of algebraic simplicity over other formulations such as the Vinet or Birch-Murnaghan EoSs (e.g. Anderson 1995, Angel 2000a) which should be used if the range of compression is greater than 10%. The Murnaghan EoS can also be re-arranged to provide a direct expression for pressure in terms of compression:

$$P = \frac{K_0}{K'_0} \left[\left(\frac{V_0}{V} \right)^{K'_0} - 1 \right] \quad (2)$$

Birch-Murnaghan. This is a “Finite strain EoS”, and is based upon the assumption (e.g. Birch 1947) that the strain energy of a solid undergoing compression can be expressed as a Taylor series in the finite strain, f . There are a number of alternative definitions of f , each of which leads to a different relationship between P and V . The Birch-Murnaghan EoS (Birch 1947) is based upon the Eulerian strain, $f_E = [(V_0/V)^{2/3} - 1]/2$. Expansion to fourth-order in the strain yields an EoS:

$$P = 3K_0 f_E (1 + 2f_E)^{5/2} \left(1 + \frac{3}{2}(K' - 4)f_E + \frac{3}{2} \left(K_0 K'' + (K' - 4)(K' - 3) + \frac{35}{9} \right) f_E^2 \right) \quad (3)$$

If this EoS is truncated at second-order in the energy, then the coefficient of f_E must be identical to zero, which requires that K' has the fixed value of 4 (higher-order terms are ignored). The third-order truncation, in which the coefficient of f_E^2 is set to zero yields a three-parameter EoS (with V_0 , K_0 and K') with an implied value of K'' given by (Anderson 1995):

$$K'' = \frac{-1}{K_0} \left((3 - K')(4 - K') + \frac{35}{9} \right) \quad (4)$$

Natural strain. Poirier and Tarantola (1998) developed an EoS based upon the “natural” or “Hencky” measure of linear strain, $f_N = \ln(l/l_0)$ which, for hydrostatic compression, may be written as $f_N = 1/3 \ln(V/V_0)$. This yields a pressure-volume relationship expanded to fourth-order in strain of:

$$P = 3K_0 \left(\frac{V_0}{V} \right) f_N \left[1 + \frac{3}{2}(K' - 2)f_N + \frac{3}{2} \left(1 + K_0 K'' + (K' - 2) + (K' - 2)^2 \right) f_N^2 \right] \quad (5)$$

Examination of Equation (5) shows that truncation of this “Natural strain” EoS at second-order in the strain implies a value of $K' = 2$, different from that of the second-order Birch-Murnaghan EoS. For truncation at third-order in the strain, the implied value of K'' is given by:

$$K'' = \frac{-1}{K_0} \left[1 + (K' - 2) + (K' - 2)^2 \right] \quad (6)$$

This value for K'' is normally substantially larger than that implied by the truncation of the 3rd-order Birch-Murnaghan EoS (Eqn. 4), and this may result in significantly different values of K_0 being obtained from fits of the two equations to the same P - V data, as in the worked example with quartz P - V data, given below.

Vinet. The finite-strain EoS do not accurately represent the volume variation of most solids under very high compression ($\eta < 0.6$), so Vinet et al. (1986, 1987) derived an EoS from a general inter-atomic potential. For simple solids under very high compressions the resulting Vinet EoS provides a more accurate representation of the volume variation with pressure:

$$P = 3K_0 \frac{(1 - f_V)}{f_V^2} \exp \left(\frac{3}{2} (K' - 1)(1 - f_V) \right) \quad (7)$$

where $f_V = (V/V_0)^{1/3}$. There is no theoretical basis for truncation of the EoS to lower order, although examination of Equation (7) shows that such truncation yields an implied value for K' of 1. The value of K'' implied by Equation (7) is given by Jeanloz (1988) as:

$$K'' = \frac{-1}{K_0} \left[\left(\frac{K'}{2} \right)^2 + \left(\frac{K'}{2} \right) - \left(\frac{19}{36} \right) \right] \quad (8)$$

Expansions of the Vinet EoS to include a refineable K'' have been proposed but are not required to fit most experimental P - V data of simple solids. Despite being often called a “Universal EoS” (e.g. Vinet et al. 1986, 1987) it should be noted that the Vinet EoS is not intended for materials with significant degrees of internal structural freedom such as bond-bending (Jeanloz, 1988).

Fitting Equations of State: EOSFIT.

Because of the algebraic form of EoSs, least-squares fitting of P - V data leads to high correlations between the refined parameters V_0 , K_0 and K' . Great care must therefore be taken in fitting EoS to avoid unintended bias of the resulting parameters by incorrect weighting schemes, incorrect fixing of parameters or outliers in the dataset. And in assessing the final refined parameter values the covariance must be considered. Further details about the methods of data analysis and assessment of the results are provided in Angel (2000a). Here the program EOSFIT, distributed with the course software, is used to illustrate the fitting of a P - V dataset (Table 1).

Because all of the EoS listed above, except the Murnaghan, can be written with pressure as a function of volume and not *vice-versa*, the EOSFIT program performs least-squares fit of P - V data with pressure as the dependent variable. The dataset (see the quartzpv.dat file as an example) must be an ASCII file with one data point per line, delimited by commas. At minimum, each line must include a pressure and a volume. In addition, each line may also include a value of $\text{esd}(P)$ and/or a value for $\text{esd}(V)$. If both are provided, $\text{esd}(P)$ must be written before $\text{esd}(V)$.

Assignment of weights. In any P - V dataset, both the pressures and the volumes have experimental uncertainties associated with them. EOSFIT provides the user with the opportunity to perform the least-squares fit with either the data unweighted, or with weights derived from either the estimated uncertainties in the pressures, or in the volumes, or both. If one or both uncertainties are not present in the datafile, then the choice of weighting scheme is restricted by the program. Uncertainties in volumes are converted into uncertainties in pressure by the effective variance method (e.g. Orear 1982):

$$\sigma^2 = \sigma_p^2 + \sigma_v^2 \cdot \left(\frac{K}{V}\right)^2 \quad (9)$$

Because the bulk modulus at the pressure of each datum appears on the right-hand-side of this equation, the EOSFIT program recalculates the weights before each least-squares cycle.

Refinement strategy. Examination of the equations of all isothermal EoS (Eqns. 1-8) shows that they are non-dimensional; they can all be written in terms of P/K_0 and V/V_0 . Therefore K_0 and V_0 have the same units as the experimental pressures and volumes respectively and are the scaling parameters of an EoS. In particular, V_0 is a quantity that is dependent upon the calibration of the technique used to measure the volumes. For example, in monochromatic angle-dispersive powder diffraction, the volumes obtained from fitting the powder pattern will depend upon the alignment of

the monochromator and the value of the resulting X-ray wavelength. Errors in calibration of the sample-to-detector distance will also strongly affect the value of V_0 . Similarly, in energy-dispersive diffraction the volume is dependent upon the energy calibration of the detector. In all of these cases the volumes measured at high pressures may be on a different scale from some high-accuracy value of V_0 determined by another technique. As demonstrated by Hazen and Finger (1989), the fixing of V_0 to such an inappropriate value can lead to incorrect estimates of the other EoS parameters being obtained from the least-squares refinement to high-pressure volume data.

The parameters V_0 and K_0 thus have the largest influence on the calculated pressure and should always be refined. For isothermal data sets the first stage of refinement should therefore be the refinement of V_0 and K_0 alone in a second-order EoS, with the next higher order term, K' set to its implied value. For fitting the quartz data in Table 1 with a Birch-Murnaghan EoS, we proceed as follows with the EOSFIT program (a summary of the results that you should obtain is given in Table 2):

Start the program in a DOS box from the directory in which the datafile (quartzpv.dat) is stored. The program first requests the name for a log file, and then the name of the data file:

```
INPUT NAME OF PRINT FILE: quartz.prt
INPUT NAME OF DATA FILE: quartzpv.dat
```

The next display is the main menu, to which the program always returns after fitting an EoS:

```
SELECT ACTION:
-1: EXIT
0: SELECT NEW DATA FILE
1: FIT EOS
2: FIT LINEAR DATA
ENTER SELECTION: 1
```

There now follow two menus that initialise the fitting of the EoS. At the first menu, select the Birch-Murnaghan P,V fit by entering “3”.

The next menu allows you to select the weighting scheme to be used in performing the least-squares process. It is normal to use weights derived from the uncertainties in both pressure and in volume, so select this option by entering “4”.

The least squares program requires some initial estimates of the values of the parameters to be refined, together with the values of those to be fixed in the refinement. For a second-order Birch-Murnaghan EoS the value of K' must be fixed at 4, the value implied by the truncation (see above). A reasonable guess for V_0 is 113\AA^3 , and 45 GPa for K_0 (these values are not critical). Therefore,

INPUT INITIAL VALUES OF PARAMETERS, V_0 , K, K_p : 113,45,4.0

The program now calculates and prints the implied value for K_0'' from your input parameters.

You now choose which parameters to refine at the next prompt: 1 for refine, 0 to fix. Thus, to refine a 2nd-order Birch-Murnaghan EoS:

INPUT 4 REFINEMENT FLAGS (1=YES, 0=NO): 1,1,0,0

The program performs the least-squares fit and terminates when the total sum of the parameter shifts divided by their esd's becomes insignificant. The parameter values after each least-squares cycle are printed to the log file. The parameter values after the last cycle are printed to the screen, together with their esd's (a summary is given in Table 2). Note also that the implied values for the unrefined parameters are also provided. The next pages (listed by entering a <CR>) list the observed and calculated values of the pressures, the difference $P_{obs}-P_{calc}$, and the weight of each data point. This is followed by a screen showing a number of fit parameters (all of this information is printed to the log file). Note the large value of 128 for χ_w^2 , together with the maximum misfit, $|P_{obs} - P_{calc}|_{max}$, more than ten times larger than the esd in an individual data point indicates that this EoS does not represent the data.

The EoS must therefore be expanded by a further parameter by refining K_p as well. This can be achieved by answering “Y” to:

FURTHER CALCULATIONS (Y/N)? y

which returns you to the main menu. Select EoS fit again by entering “1”, and proceed as above until the prompt to input the refinement flags. This time, refine the value of K' as well:

INPUT 4 REFINEMENT FLAGS (1=YES, 0=NO): 1,1,1,0

This expansion of the EoS to third-order reduces χ_w^2 to 0.95, indicating a significant improvement to the fit. The same conclusion would be drawn from the other indicators; the refined value of the additional parameter K' (5.99) differs by 50 esd's from the previously implied value of $K' = 4$, the esd's of V_0 and K_0 have decreased, the maximum misfit is similar to the estimates of the uncertainties in pressure estimated directly from the experiment, and the value of V_0 is identical to that determined experimentally.

As a final step, proceed through the menus again and refine a fourth-order Birch-Murnaghan EoS by setting all of the refinement flags to 1, so as to refine the value of K'' . This yields only a marginal improvement in χ_w^2 , because the refined value of K'' only differs marginally (1.2 esd's) from the value implied by the 3rd-order truncation of the EoS. Note also that the esd's of K_0 and K' have increased significantly in this last refinement due to their strong correlation (93.6% and -99.2% respectively) with K'' . For practical purposes, therefore, the 3rd-order Birch-Murnaghan EoS would be considered to yield an adequate representation of the data-set.

The steps in the refinement of the Natural Strain EoS to the same data-set (Table 2) are similar, except for the choice of termination of the refinement process. In this case further expansion of the Natural Strain EoS to 4th order results in a significant decrease in χ_w^2 from 1.15 to 0.93 as a result of the value of K'' deviating by more than 4 esd's from the value implied by the 3rd-order truncation (Eqn. 6).

Table 1. Cell parameters of quartz with pressure from Angel et al. (1997)

P: GPa	a: Å	c: Å	V: Å ³
10 ⁻⁴	4.91300(11)	5.40482(17)	112.981(2)
0.429(9)	4.89295(29)	5.38861(22)	111.725(14)
0.794(10)	4.87657(12)	5.37563(12)	110.711(6)
1.651(9)	4.84201(15)	5.34856(14)	108.597(7)
1.845(9)	4.83461(39)	5.34284(37)	108.150(19)
1.933(9)	4.83136(17)	5.34135(17)	107.974(8)
2.628(12)	4.80593(16)	5.32266(15)	106.467(8)
3.299(9)	4.78306(18)	5.30679(16)	105.141(9)
3.468(12)	4.77750(27)	5.30341(22)	104.831(12)
3.778(12)	4.76798(22)	5.29692(22)	104.285(10)
4.026(12)	4.75970(27)	5.29116(28)	103.810(13)
4.553(11)	4.74411(16)	5.28128(14)	102.939(7)
4.827(14)	4.73671(25)	5.27699(23)	102.534(12)
5.212(11)	4.72561(21)	5.27072(19)	101.933(10)
5.416(12)	4.71973(17)	5.26617(17)	101.592(08)
5.736(11)	4.71137(25)	5.26150(21)	101.143(11)
6.203(14)	4.69710(32)	5.25385(32)	100.385(15)
6.478(13)	4.69089(33)	5.25027(30)	100.051(15)
6.751(12)	4.68392(18)	5.24622(20)	99.677(09)
7.191(15)	4.67228(27)	5.23993(21)	99.064(12)
7.898(8)	4.65612(30)	5.23058(28)	98.204(14)

8.449(15)	4.64333(15)	5.22416(32)	97.545(16)
8.905(13)	4.63253(38)	5.21863(35)	96.989(17)

Table 2. EoS parameters fitted to the quartz P-V data of Angel et al. (1997)

	$V_0 : \text{\AA}^3$	$K_0 : \text{GPa}$	K'	$K'' : \text{GPa}^{-1}$	χ_w^2	$ P_{obs} - P_{calc} _{max}$ GPa
BM2	112.97(2)	41.5(3)	[4.0]	[-.094]	128	0.32
BM3	112.981(2)	37.12(9)	5.99(5)	[-.265]	0.95	0.025
BM4	112.981(2)	36.89(22)	6.26(24)	-.41(12)	0.93	0.026
NS2	112.95(5)	46.5(6)	[2.0]	[-.022]	580	0.65
NS3	112.982(2)	36.39(11)	6.91(7)	[-.825]	1.15	0.026
NS4	112.981(2)	36.90(24)	6.25(29)	-.39(11)	0.93	0.026
Vinet	112.981(2)	37.02(9)	6.10(4)	[-.319]	0.90	0.025
Murn.	112.981(2)	37.63(10)	5.43(4)	[0]	1.57	0.033

Note: Numbers in parentheses represent esd's in the last digit. Numbers in square brackets are the implied values of the parameters.

Fitting high-pressure lattice parameters.

As for volume variations with pressure, there is no fundamental thermodynamic basis for specifying the form of cell parameter variations with pressure. It is therefore not unusual to find in the literature cell parameter variations with pressure fitted with a polynomial expression such as $a = a_0 + a_1P + a_2P^2$, even when the P - V data have been fitted with a proper EoS function. Use of polynomials in P is not only inconsistent, it is also unphysical in that a linear expression implies that the material does not become stiffer under pressure, while a quadratic form will have a positive coefficient for P^2 , implying that at sufficiently high pressures the material will expand with increasing pressure. A consistent alternative is provided by using the same EoS as that used to fit the P-V data, but substituting the cube of the lattice parameter for the volume in the EoS, and this is the method implemented in the EOSFIT program. The cubing of the lattice parameter and the transformation of it's

esd is performed by the program when the user selects the “Fit linear data” option. (The datafile format is the same as for P - V data, except that the lattice parameter and its esd should be entered in the datafile). The refined parameter value and its esd is transformed back from volume to unit-cell parameter, but note that the printed variance-covariance matrix contains entries for the lattice parameter cubed. Note also, that the value of “ K_0 ” obtained from fitting the cell parameters in this way is related to the zero-pressure compressibility β_0 of the axis by $-1/3K_0 = \beta_0 = a_0^{-1}(\partial a/\partial P)_{P=0}$ in which a_0 is the length of the unit-cell axis at zero pressure.

For crystals with higher than monoclinic symmetry the definition of the axial compressibilities in this way fully describes the evolution of the unit-cell with pressure because the tensor describing the strain arising from compression is constrained by symmetry from rotating. In the monoclinic system, however, one unit-cell angle may change, and in triclinic crystals all three unit-cell angles may change. The full description of the change in unit-cell shape in these cases must therefore include the full definition of the strain tensor resulting from compression. A computer program, originally written by Ohashi (1972) is available to calculate the components and principal axes of strain tensors. The calculation method of Ohashi (1972), further developed by Schlenker et al. (1975) and Jessen and Küppers (1991), is explicitly based upon a finite difference approach. The strain is evaluated from the change in lattice parameters between one data point and the next. Thus the resulting strain tensor represents an average strain over this interval in pressure or temperature. This is a sound approach for crystals of orthorhombic symmetry, or higher, because the orientation of the strain ellipsoid is fixed by symmetry. But for triclinic and monoclinic crystals the strain ellipsoid may rotate with changing P or T . The finite difference calculation of strain then represents an average not only the magnitudes of the principal axes of the strain ellipsoid, but also an average of their orientation over the finite interval in P or T . An alternative approach which avoids this problem and employs the calculation of the continuous derivatives of the unit-cell parameters with respect to T (or P) has been developed by Paufler and Weber (1999).

Fortunately, in monoclinic systems the strain tensor often does not rotate significantly with pressure. Then it may be appropriate to fit quantities such as $a\sin\beta$ against pressure with an EoS function, or the β angle separately as a polynomial function of pressure (e.g. Angel et al., 1999). The important criterion is that the resulting expressions provide not only a good fit to the data, but are reliable in extrapolation to further pressures of interest (e.g. when studying phase transitions; see Angel, 2000b). The reliability of these extrapolations can always be tested by

parallel calculations with different functions (e.g. Boffa-Ballaran et al. 2000). A further internal check on the robustness of the extrapolations can be obtained by comparing the unit-cell volumes obtained from the lattice parameters extrapolated to a given pressure with those predicted by the EoS function fitted to the unit-cell volume.

Program release notes.

The EOSFIT program is distributed on a non-commercial basis and the author would appreciate its use being acknowledged by reference to Angel (2000a) in any publications. I plan to re-write and expand the program with a GUI, incorporating a more flexible input-file format and the calculation of strain tensors. If you would like to receive program updates (including bug fixes), please register with me as a user by e-mail (ross.angel@uni-bayreuth). If you discover apparent bugs in the program, please send me both the input file, the output file and a full description of the problem by e-mail.

References

- Anderson OL (1995) Equations of state of solids for geophysics and ceramic science. Oxford University Press, Oxford.
- Angel RJ (1993) The high-pressure, high-temperature equation of state of calcium fluoride, CaF₂. *J Phys: Cond Matter* 5:L141-L144
- Angel RJ (2000a) Equations of state. In Hazen RM (Ed), *Comparative Crystal Chemistry*. MSA Reviews in Mineralogy 39.
- Angel RJ (2000b) High-pressure structural phase transitions. In Redfern SAT (Ed), *Transformation Processes in Minerals*. MSA Reviews in Mineralogy 40.
- Angel RJ, Allan DR, Miletich R, Finger LW (1997) The use of quartz as an internal pressure standard in high-pressure crystallography. *J Appl Cryst* 30:461-466
- Angel RJ, Kunz M, Miletich R, Woodland AB, Koch M, Xirouchakis D (1999) High-pressure phase transition in CaTiOSiO₄ titanite. *Phase Transitions* 68:533-543
- Barnett JD, Block S, Piermarini GJ (1973) An optical fluorescence system for quantitative pressure measurement in the diamond-anvil cell. *Rev Sci Instrum* 44:1-9

- Besson JM, Nelmes RJ, Hamel G, Loveday JS, Weill G, Hull S (1992) Neutron diffraction above 10 GPa. *Physica B* 180 & 181:907-910
- Birch F (1947) Finite elastic strain of cubic crystals. *Phys Rev* 71:809-824
- Birch F (1986) Equation of state and thermodynamic parameters of NaCl to 300 kbar in the high-temperature domain. *J Geophys Res* 91:4949-4954
- Boffa-Ballaran T, Angel RJ, Carpenter MA (2000) High-pressure transformation behaviour of the cummingtonite-grunerite solid solution. *Euro J Miner*, accepted
- Brown, JM (1999) The NaCl pressure standard. *J Appl Phys* 86:5801-5808
- Chai M, Brown JM (1996) Effects of non-hydrostatic stress on the R lines of ruby single crystals. *Geophys Res Letts* 23:3539-3542
- Decker DL (1971) High-pressure equations of state for NaCl, KCl, and CsCl. *J Appl Phys* 42:3239-3244
- Decker DL, Petersen S, Debray D, Lambert M (1979) Pressure-induced ferroelastic phase transition in $\text{Pb}_3(\text{PO}_4)_2$: A neutron diffraction study. *Phys Rev B* 19:3552-3555
- Eggert JH, Xu L-W, Che R-Z, Chen L-C, Wang J-F (1992) High-pressure refractive index measurements of 4:1 methanol: ethanol. *J Appl Phys* 72:2453-2461
- Gupta YM, Shen XA (1991) Potential use of the ruby R_2 line shift for static high-pressure calibration. *Appl Phys Letts* 58:583-585
- Hammersley AP, Svensson SO, Thompson A, Graafsma A, Kvick A, Moy JP (1995) Calibration and correction of distortions in 2D detector systems. *Rev Sci Instr* 66:2729-2733
- Hammersley AP, Svensson SO, Hanfland M, Fitch AN, Häusermann D (1996) Two-dimensional detector software: from real detector to idealised image or two-theta scan. *High Press Res* 14:235-248
- Hazen RM (2000) Comparative Crystal Chemistry. *MSA Reviews in Mineralogy* 39.
- Hazen RM, Finger LW (1981) Calcium fluoride as an internal pressure standard in high-pressure/high-temperature crystallography. *J Appl Crystallogr* 14:234-236
- Hazen RM, Finger LW (1989) High-pressure crystal chemistry of andradite and pyrope: revised procedures for high-pressure diffraction experiments. *Am Min* 74:352-359
- Heinz DL, Jeanloz R (1984) The equation of state of the gold calibration standard. *J Appl Phys* 55:885-893

- Holzappel WB (1997) Pressure determination. In: High-Pressure Techniques in Chemistry and Physics (WB Holzappel, NS Isaacs, eds) Oxford University Press, Oxford: pp 47-55
- Jeanloz R (1988) Universal equation of state. *Phys Rev B* 38:805-807
- Jessen SM, Küppers H (1991) The precision of thermal-expansion tensors of triclinic and monoclinic crystals. *J Appl Cryst* 24:239-242
- Klotz S, Besson JM, Hamel G, Nelmes RJ, Marshall WG, Loveday JS, Braden M (1998) *Rev High Pressure Sci Technol* 7:217-220
- Knorr K, Fütterer K, Annighöfer B, Depmeier W (1997) A heatable large volume high pressure cell for neutron powder diffraction: The Kiel-Berlin Cell I. *Rev Sci Instrum* 68:3817-3822
- Knorr K, Annighöfer B, Depmeier W (1999) A heatable large volume high pressure cell for neutron powder diffraction: The Kiel-Berlin Cell II. *Rev Sci Instrum* 70:1501-1504
- Kuhs W, Bauer FC, Hausmann R, Ahsbahr H, Dorwarth R, Hölzer K (1996) Single crystal diffraction with X-rays and neutrons: High quality at high pressure? *High Press Res* 14:341-352
- Miletich R, Allan DR, Kuhs WF (2000) High-pressure single-crystal techniques. In Hazen RM (Ed), *Comparative Crystal Chemistry*. MSA Reviews in Mineralogy 39.
- Murnaghan FD (1937) Finite deformations of an elastic solid. *Am J Math* 49:235-260
- Nomura M, Nishizaka T, Hirata Y, Nakagiri N, Fujiwara H (1982) Measurement of the resistance of Manganin under liquid pressure to 100 kbar and its application to the measurement of the transition pressures of Bi and Sn. *Jap J Appl Phys* 21:936-939
- Ohashi (1972) Program Strain. Program listing provided in Hazen and Finger (1982) *Comparative Crystal Chemistry*, John Wiley and Sons, New York.
- Orear J (1982) Least squares when both variables have uncertainties. *Am J Phys* 50:912-916
- Paufler PP, Weber T (1999) On the determination of linear thermal expansion coefficients of triclinic crystals using X-ray diffraction. *Euro J Min* 11:721-730
- Poirier J-P, Tarantola A (1998) A logarithmic equation of state. *Phys Earth Planet Int* 109:1-8

- Schlenker JL, Gibbs GV, Boisen MB (1975) Thermal expansion coefficients for monoclinic crystals: a phenomenological approach. *Am Min* 60:828-833
- Sharma SM, Gupta YM (1991) Theoretical analysis of R-line shifts of ruby subjected to different deformation conditions. *Phys Rev B* 43:879-893
- Sinogeikin SV, Bass JD (1999) Single-crystal elasticity of MgO at high pressure. *Phys Rev B* 59:R14141-R14144
- Thoms M, Bauchau S, Häusermann D, Kunz M, LeBihan T, Mezouar M, Strawbridge D (1998) An improved X-ray detector for use at synchrotrons. *Nucl Inst Meths Phys Res A* 413:175-184
- Vinet P, Ferrante J, Smith JR, Rose JH (1986) A universal equation of state for solids. *J Phys C: Solid State* 19:L467-L473
- Vinet P, Ferrante J, Rose JH, Smith JR (1987) Compressibility of solids. *J. Geophys Res* 92:9319-9325
- Vos WL, Schouten JA (1991) On the temperature correction to the ruby pressure scale. *J Appl Phys* 69:6744-6746

Application of X-Ray Powder Diffraction in Pharmaceutical Sciences

Peter Sieger

Boehringer Ingelheim Pharma KG

Abstract

The tendency for pharmaceutical solids to crystallize in multiple crystal forms and the significance of this phenomenon (polymorphism) have been demonstrated (ref. 1, 2). Since polymorphism can affect the chemical, biological and pharmaceutical properties of a drug, it is very important to detect polymorphic, solvated or amorphous forms of the drug substance. Solid state properties, like crystallinity, thermal and hygroscopical behavior as well as the tendency to form different crystalline modifications (polymorphism) under varying crystallization conditions is of special interest for the developability of a new active pharmaceutical compound.

X-ray powder diffraction is a very powerful and widely used analytical tool among the variety of existing techniques to study these properties. X-ray powder diffraction is used in simple routine experiments to measure crystallinity and/or phase purity, in temperature and/or humidity dependent experiments to study solid state phase transformations under the influence of varying temperature and/or humidity and in collecting highly resolved diffraction data for elucidation of structures which can not be solved from single crystal data. The last issue is certainly the most challenging task because pharmaceutical solids normally crystallize in large molecular structures (sometimes with several thousand Å³ volume) with low symmetry (more than 90 % of the compounds crystallize in monoclinic or triclinic space groups) and in addition the scattering properties are normally also very low.

Pharmaceutical solids are normally organic compounds forming molecular crystals with weak intermolecular binding forces among which are mainly van-der-Waals attractions and hydrogen bonds. Very often pharmaceutical solids tend to crystallize in multiple crystal forms. For this behavior also the term polymorphism is frequently used. Formation of different crystalline modifications is also observed by inclusion of water (hydrates) or organic solvents (solvates) in the crystal structures. This phenomena is also known as pseudopolymorphism. More than 80 % of all marketed drugs show polymorphism or pseudopolymorphism. Beside crystalline forms also amorphous materials are frequently encountered in pharmaceutical sciences. Especially drug

substances with rather flexible molecules are hard to crystallize and tend to form amorphous forms upon precipitation.

The variability of the solid state properties may have an impact on the physico-chemical, biological and pharmaceutical properties of a drug substance. Among the physicochemical properties are the melting point, the stability and the solubility which strongly depend on the solid state of a drug. Solubility, for example, is one of the limiting factors for the bioavailability of a pharmaceutical compound. A drug can only permeate through the membranes in the gastrointestinal tract if it is dissolved, so a fast dissolution rate and high saturation solubility are important for a good bio-availability, which itself influences strongly the efficacy of a drug. Stability can affect the safety of a drug because the rate of decomposition, which could be different for polymorphs and especially for amorphous forms, will lower the amount of active ingredient in a pharmaceutical dosage form upon storage time. Among the pharmaceutical properties an easy processability is one of the major aspects. For this issue the crystal morphology is for example of great importance. A nightmare for large scale pharmaceutical processes are drug substances which precipitate in long needle-like crystals. Such compounds have very unfavorable processing properties, like almost no flowing properties, very low tapped density, they are extremely hard to filter, they exhibit very long drying time due to inclusion of solvent and they strongly tend to electrostatic charging.

Due to this knowledge authorities like the FDA (= Food and Drug Administration) or the EMEA (= European Medicinal product Evaluation Agency) established guidelines for pharmaceutical industry how to handle this issue. The guidelines state that the crystal form of a drug substance has to be controlled and therefore appropriate analytical procedures should be established to detect polymorphic, hydrated/solvated or amorphous forms of a drug substance. These guidelines give also lists which analytical procedures are appropriate. In these lists X-ray powder diffraction is always mentioned first, followed by thermoanalytical methods (DSC, TG), microscopy, IR-spectroscopy, solid state NMR-spectroscopy and solution calorimetry. Therefore, X-ray powder diffraction is found nowadays in almost every big pharmaceutical company, at least in the ones which have their own R&D facilities.

At Boehringer Ingelheim Pharma KG X-ray powder diffraction has several applications. Most of the time this technique is used for simple routine experiments to check crystallinity and phase purity (fingerprint) of a compound. In order to study the thermal and hygroscopical behaviour temperature and humidity dependent experiments are performed. For this purpose special instrumentation is necessary to collect X-ray powder diffraction data under controlled temperature and humidity conditions. In some cases high resolution data is desirable for structure elucidation studies. First choice for structure analysis is certainly a single crystal X-ray diffraction study. However, for several pharmaceutical compounds adequate single crystals are not available and

therefore structure analysis is only possible from powder data. On the other hand, there are several limitations for pharmaceutical solids which make the structure elucidation from powder data very challenging. Pharmaceutical solids normally crystallize in large molecular structures (sometimes with several thousand Å³ volume) with low symmetry (more than 90 % of the compounds crystallize in monoclinic or triclinic space groups) and in addition the scattering properties are normally also very low.

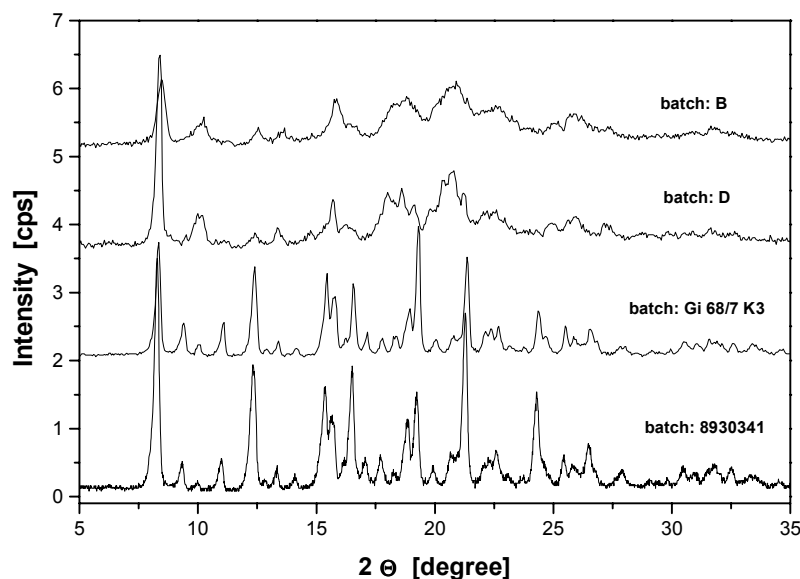


Figure 1: X-ray powder diffractograms of batches of a drug from an optimization campaign for crystallization conditions.

For a better understanding of the different applications of X-ray powder diffraction in pharmaceutical sciences some typical examples are given below.

In the field of routine experiments very often analytical support is given to the optimization of crystallization processes in the scale-up of drug substance synthesis. In Figure 1 typical X-ray diffraction patterns of a synthesis campaign of a drug substance are shown where the crystallization conditions had to be optimized. The series of diffractograms indicate that from batch B, start of the campaign, to batch 8930341, end of the campaign, a significant increase in crystallinity was obtained. The batch size increased also in this campaign from approx. 50 g for batch B to about 50 kg for batch 8930341.

A second typical example among the routine experiments are polymorphism screening studies where a drug substance is recrystallized from different organic solvents. The recrystallized samples are analyzed by X-ray powder diffraction in order to detect other crystalline modifications of the compound. These studies are performed very early in development to get an idea if a new drug substance tends to form different polymorphs and if care has to be taken if changes in the last crystallisation step of the synthesis have to be applied.

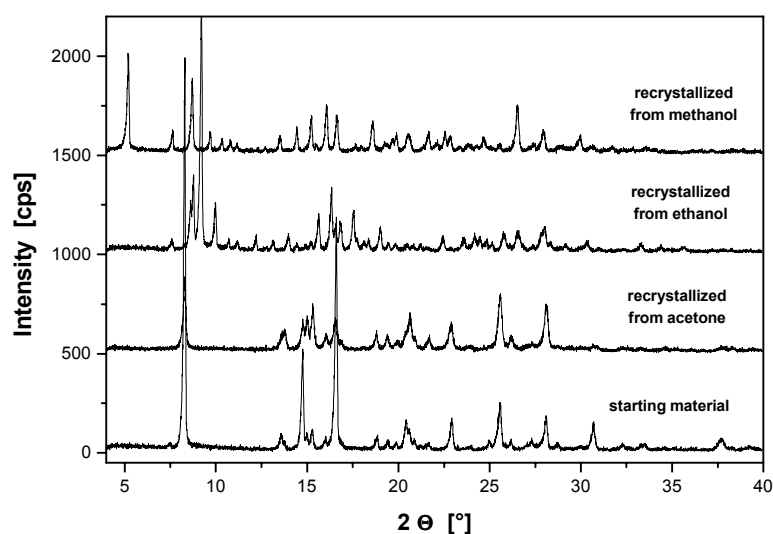


Figure 2: X-ray powder diffractograms of a polymorphism screening study.

In Figure 2 a typical example of such a polymorphism screening study is shown. For this compound different polymorphs are obtained if the material is recrystallized from ethanol or methanol. For these two solvents solvated forms are obtained whereas from acetone the same solvent free form recrystallized which was used as starting material in this study.

Another very often asked question is if pharmaceutical solids change their solid state properties upon intensive milling. At Boehringer Ingelheim pharmaceutical solids are very often micronised in order to reduce particle size and to increase surface area. This procedure is of special interest for drug substances for which powder inhalation is the desired way of application, which is frequently used for the treatment of airway diseases (asthma, COPD). For this way of application the particle size distribution of the drug substance has to be in a very narrow range (1 – 5 µm) in order to reach the target organ lung. To obtain such fine material intensive milling in so called air-jet mills is necessary. This type of milling is a high energy process which could alter the solid state properties of a drug substance. Very often partial amorphisation is observed upon this intensive milling step, especially if the drug substance crystallizes in a layered structure. Upon storage, the amorphous components tend to crystallize inducing particle size growth.

In Figure 3 an example is shown where the effect of intensive milling is clearly visualised in the X-ray powder diffraction patterns. In the diffractograms of the samples which were milled with 3.5 and 6.0 bar milling pressure a very broad hump could be observed in the background, indicating a certain amount of amorphous components in the material. For the sample micronised with 6.0 bar milling pressure the "amorphous halo" seems to be even more pronounced. In order to avoid uncontrolled crystallization of the amorphous component inducing uncontrolled particle size growth, a conditioning step is included in the

processing of the micronised material. Conditioning is performed under elevated temperature and humidity. The last X-ray powder pattern in Figure 3 shows an example of such a material which was conditioned in a climatic chamber right after the micronisation process. The "amorphous halo" clearly vanished in this diagram indicating that the amorphous components are reduced by the conditioning step at least below the limit of detection of X-ray powder diffraction. However, this example shows also the limitations of this method because X-ray powder diffraction is not capable to detect very low amounts of amorphous components ($< 10\%$) in an otherwise crystalline matrix. For this question more sensitive analytical methods have to be applied as for example solution calorimetry. The heat of solution, which can be nowadays measured very sensitively with commercially available calorimeters, depends on the amount of amorphous components present in a sample. Using this more sensitive method allows a reproducible determination of amorphous components in a crystalline matrix down to about $1 - 2\%$.

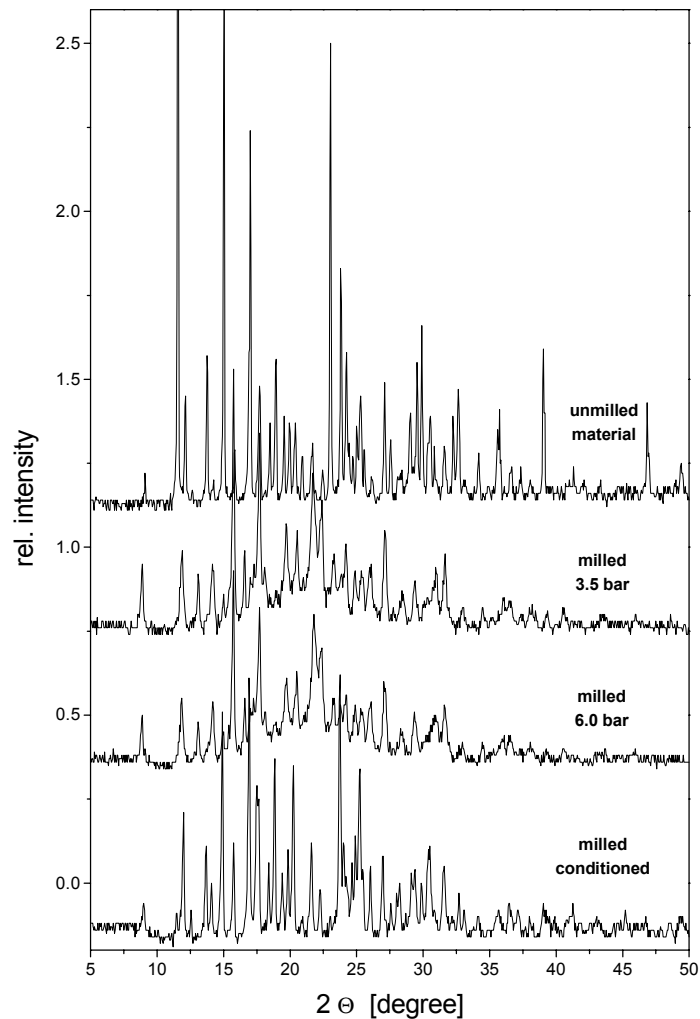


Figure 3: Partial amorphisation upon micronisation X-ray powder diffraction study

It is quite common for pharmaceutical solids that they change their crystal structure under the influence of varying temperature or humidity. Solid state phase transitions under the influence of temperature can be easily detected by running DSC (= Differential Scanning Calorimetry) measurements because changes in the crystal structure of a compound induced by temperature are accompanied by a heat flow which is sensitively measured in the DSC experiment. For pharmaceutical solids both types of solid state phase transitions, enantiotropic (fully reversible) and monotropic (irreversible) transitions are encountered. By using temperature dependent X-ray powder diffraction such solid state phase transitions can be also easily followed. Hygroscopic compounds change their crystal structure under the influence of varying humidity by forming hydrated forms. Running water sorption isotherms is an elegant way to detect hygroscopicity. In Figure 4 an example of such a water sorption isotherm is shown. In this case the compound changes its crystal structure from an anhydrous form to a monohydrate while increasing the relative humidity above 70 % r.h. A clear hysteresis could be observed in the desorption experiment.

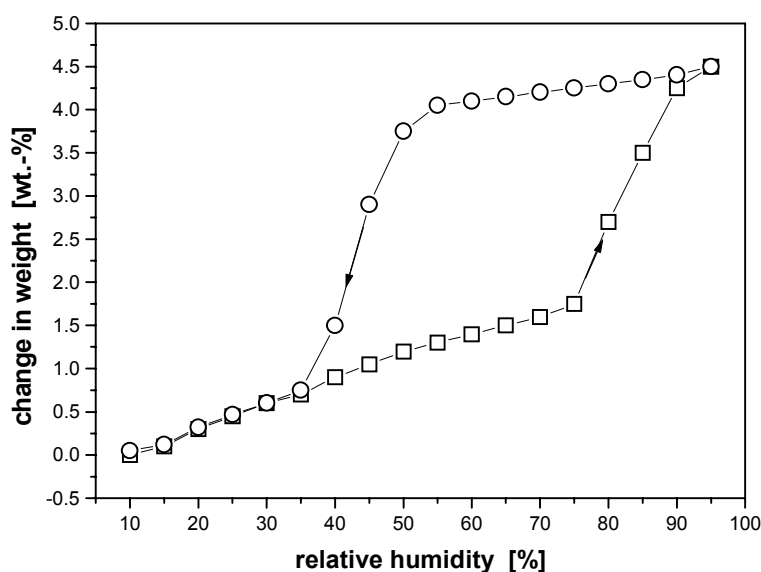


Figure 4: Water sorption isotherm at room temperature.

Transformation back to the anhydrous form is observed at lower relative humidity (40 – 50 % r.h.). In order to follow the changes in the crystal structure of this solid state transition X-ray powder diffraction patterns were recorded under varying humidity. In Figure 5 on the next page a series of diffractograms are shown which were taken from the same compound under controlled humidity conditions in a special humidity chamber. Changes in the patterns could be observed in the same humidity range where strong weight changes are indicated

in the water sorption isotherms. Upon rising the humidity, transformation to the monohydrate

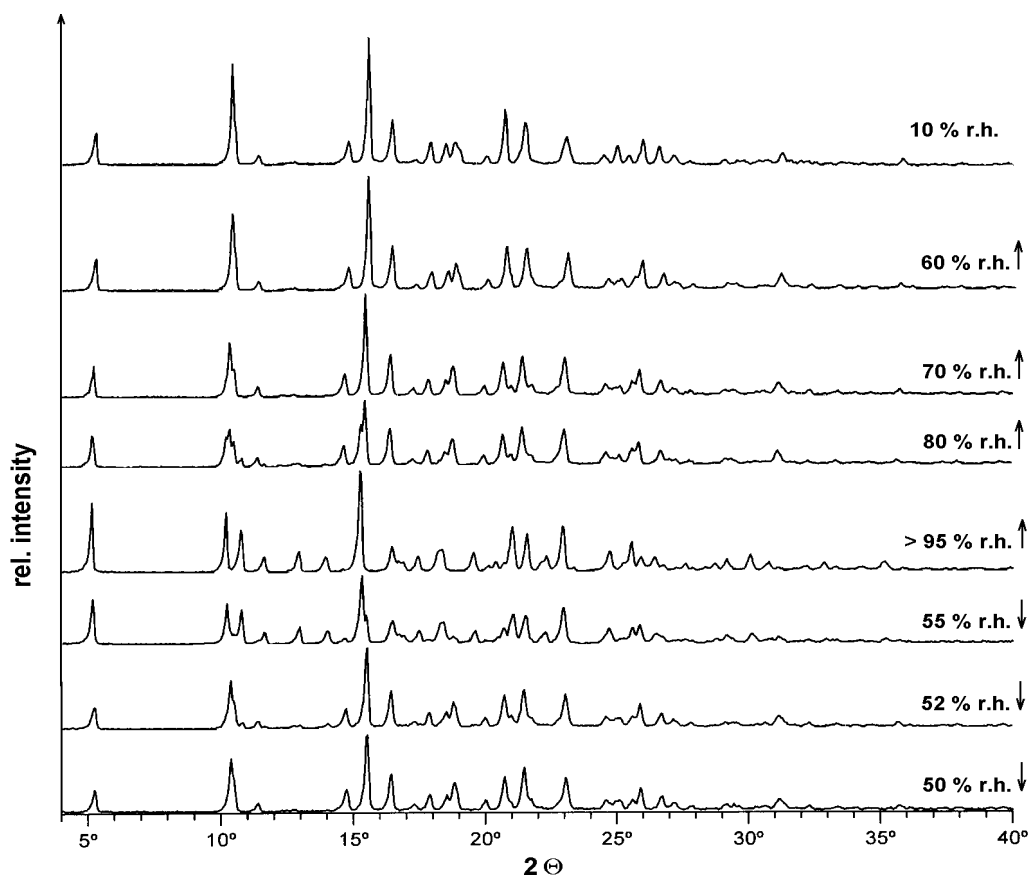


Figure 5: RH-dependent X-ray powder diffraction patterns of a solid state transformation induced by varying humidity

is clearly indicated above 70 % r.h. In the desorption process upon lowering the humidity the transformation back to the anhydrous form is observed in a very narrow range between 55 and 50 % r.h. This experiment nicely show that the hygroscopical behavior of pharmaceutical solids can be also studied by X-ray powder diffraction.

The most interesting but also most challenging application of X-ray powder diffraction in pharmaceutical sciences is certainly the possibility of structure analysis from high resolution powder data. This for pharmaceutical industry rather new technology attracts more and more attention because the classical way of structure analysis by single crystal X-ray diffraction is often not possible because single crystals of appropriate size and quality are not available. As mentioned above, pharmaceutical solids have often undesirable properties (large unit cells, low symmetry, low scattering properties) which complicate the alternative of structure elucidation from powder data. Nevertheless, this area has made tremendous progress in the last couple of years by using smart software

tools (simulating annealing techniques) which allows nowadays also the analysis of more complicated structures.

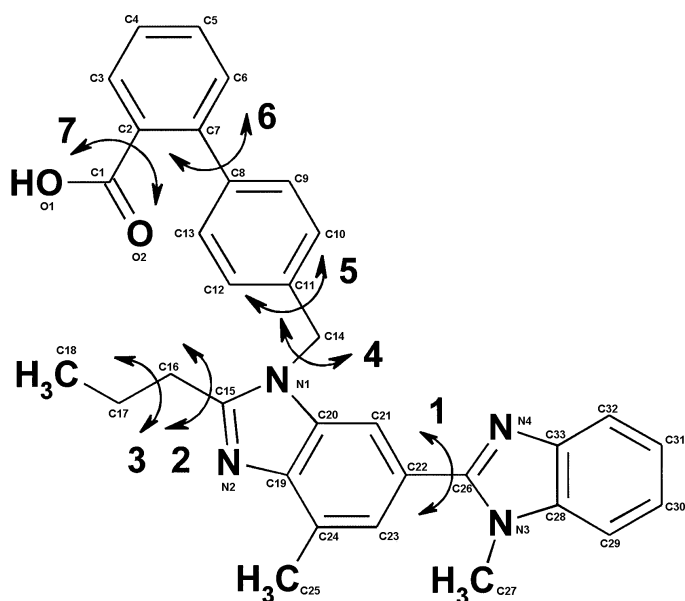


Figure 6: Structural formula of Telmisartan with the 7 torsion angles which were varied during the simulated annealing process

To test the capabilities of this new technology the structure of the Boehringer Ingelheim compound Telmisartan (see figure 6) was tried to solve using different software packages. Telmisartan is an orally active non-peptide angiotensin II receptor Antagonist for control of blood pressure. Telmisartan is known to exist in at least three different polymorphic forms, two anhydrous forms and a solvated form containing formic acid and water molecules in the structure. The structure of the solvated form could be solved from single crystal data. The two anhydrous forms were not available in large enough crystals of appropriate quality. Therefore structure analysis could be only done from powder data. High resolution X-ray powder diffraction patterns were collected at the beamline X3B1 at the National Synchrotron Light Source, Brookhaven National Laboratory (see figure 7, next page). Both powder patterns are characterized by a rapid fall off of intensity beyond $\sin\Theta/\lambda \approx 0.17 \text{ \AA}^{-1}$. Indexing of the powder patterns of polymorphs A and B led to primitive monoclinic cells with the following lattice parameters: polymorph A crystallizes in space group $P2_1/c$ (# 14), $Z = 4$, with unit cell parameters $a = 18.7798(3) \text{ \AA}$, $b = 18.1043(2) \text{ \AA}$, $c = 8.00578(7) \text{ \AA}$, $\beta = 97.066(1)^\circ$ and $V = 2701.31 \text{ \AA}^3$, polymorph B crystallizes in space group $P2_1/a$ (# 14), $Z = 4$, with unit cell parameters $a = 16.0646(5) \text{ \AA}$, $b = 13.0909(3) \text{ \AA}$, $c = 13.3231(3) \text{ \AA}$, $\beta = 99.402(1)^\circ$ and $V = 2764.2(1) \text{ \AA}^3$. For the solvated form C the following also monoclinic lattice parameters were obtained:

space group C2/c (# 15), $Z = 8$, with unit cell parameters $a = 30.990(5) \text{ \AA}$, $b = 13.130(3) \text{ \AA}$, $c = 16.381(3) \text{ \AA}$, $\beta = 95.02(2)^\circ$ and $V = 6639(2) \text{ \AA}^3$.

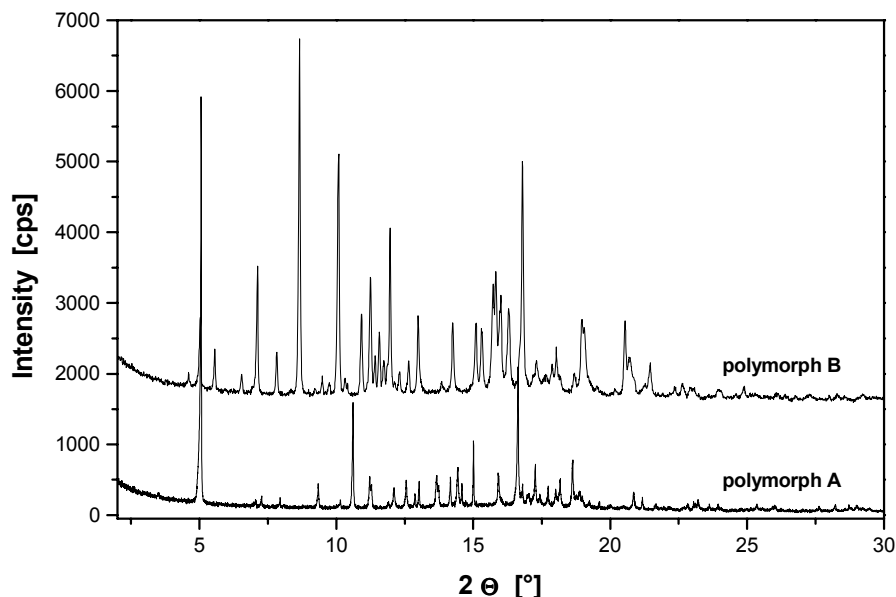


Figure 7: Synchrotron X-ray diffraction diagram $\lambda = 1.14991 \text{ \AA}$

For the structure solutions of polymorphs A and B, 13 degrees of freedom (3 translational, 3 orientational, 7 torsion angles (see also figure 6 above)) were determined in approximately two hours of computer time. Due to the very limited numbers of observations for structure determination the result of a structure analysis from X-ray powder data has also limited information. It is indisputably possible to determine the molecular conformations and the packing of the molecules with high precision, but it is usually not possible to determine individual bond lengths and bond angles within the molecules. However, for pharmaceutical compounds this is not a real problem because for organic compounds typical bond lengths and bond angles are very well established and the molecular structures are routinely solved from spectroscopic data (NMR and MS). For the understanding of bulk solid state properties of a pharmaceutical compound the overall crystal structure, which means molecular conformation and packing of the molecules which indicates intra- and intermolecular binding forces, is of main interest.

For Telmisartan, for example, it was very interesting to learn that the observed polymorphism and pseudopolymorphism is a result of different possible conformations of the molecule which causes a completely different packing in the different crystalline modifications of this compound. For more details on this subject see ref. 3.

References:

- (1) S. Byrn, R. Pfeiffer, M. Ganey, C. Hoiberg, G. Poochikian: Pharm. Res., Vol. 12(7), 1995
- (2) S. Byrn: Solid-State Chemistry of Drugs. Academic Press: New York, 1982
- (3) R.E. Dinnebier, P. Sieger, H. Nar, K. Shankland, W.I.F. David: Structural characterization of three crystalline modifications of Telmisartan by single crystal and high resolution X-ray powder diffraction. J. Pharm. Sci., in press.

Structure Determination from X-Ray Powder Data of Notoriously Difficult Materials

Hermann Gies

*Institut für Geologie, Mineralogie und Geophysik,
Ruhr-Universität Bochum, Germany*

0 Abstract

Crystal structure solution from powder data sometimes fails. These structures are difficult structures, however, the reason for the failure might be different in every material. The article highlights the most frequently occurring reasons for failure and examples with suggestions how one possibly could extend the limits of the technique.

1 Introduction

What are difficult materials where we struggle to solve their crystal structure? So far we have learned to tackle structure solution as a purely technical crystallographic problem: we have to have the latest structure solving programs used with the latest computer hardware and required diffraction data of highest quality. In general, this will always be the best starting point for solving a crystal structure from powder data, however, there are numerous examples of unsolved problems where all advances in technical development failed so far to resolve the enigma. Those who develop programs for solving and refining crystal structure also depend on deeper insight in the structure solving process. Much of the progress we have experienced in the last decade, in particular in powder crystallography, is the result of the interaction of those who develop the technique and those who solve more and more complicated problems and refine the structure with more and more sophisticated methods.

The intention of the article is to address some of the obstacles met more frequently in the structure solving process and show in examples taken from the literature how one can address one or the other problem. It is also an attempt to show that the knowledge of the properties of the material and the careful characterisation of the material with as many techniques as are available in the laboratory, complementary to diffraction, are most useful in solving those difficult structures. Last but not least, solving problems takes time. Since solving crystal structures has developed to a analytical tool, the time needed to solve a

structure is, in most cases, a fraction of a minute, even in powder crystallography. However, whenever the routine for solving a structure fails, you will need time to think about the problem, technically and methodologically. This challenge is indeed the fun of the game of solving crystal structures where crystallographers, chemists and physicists show their skill as scientists. Do not give up and leave it all to a computer program.

2 Which structures are difficult?

Solving crystal structures from powder diffraction data underlies the same conditions as solving crystal structures from single crystal diffraction data. Using direct methods atomic resolution of the diffraction data set is wanted, the space group symmetry should be known and all classes of symmetrically inequivalent reflections should be represented in the data set. The advancement in direct methods and in computational power has significantly loosened the requirements for structure solution which makes it feasible to use the limited powder diffraction data successfully. Computational methods to predict structures and to calculate energy minimised models also contributes significantly to solving crystal structures whenever the molecules or fragments of the structure are known. Here, the quality of the diffraction data set is not as important but complementary knowledge on the material is most useful.

There are different categories of reasons which might render a problem difficult. When the **crystallinity of the material** leads to peak broadening indexing of the powder pattern will become exceedingly difficult. This material property can't be overcome by high-resolution diffraction experiments and, inherently, causes ambiguities for the symmetry analysis and space groups determination. In cases one succeeds to index the powder pattern and to assign a space group, direct method or molecular modelling techniques can be applied. However, because of the line broadening, signal overlap for reflections in the 2Θ -range from $30 - 40^\circ$ aggravates the assignment of intensities. At higher angles but still below atomic resolution, signal to noise in the powder pattern is generally very reduced due to the poor scattering properties of moderately crystalline materials. Therefore, the chances for an ab initio structure solution are limited and complementary information on the material is required.

Another group of **materials** crystallises in space group symmetries **with systematic overlap of symmetrically non-equivalent reflections** (hemimorphic space groups). Here, although high resolution diffraction data might be available, the one-dimensional projection of the three-dimensional reciprocal space in the powder diffraction data set again is an inherent property of the method. This might lead to failure in direct method calculations and to refinement inaccuracies in the Rietveld data fit.

Since all intensity scattered is collected in the powder trace all materials with deviations in the periodicity on the atomic scale, here summarised as **disordered materials**, create specific problems. The type of disorder might be static or dynamic, constitute slight deviations in symmetry, or give rise to incompatibilities in unit cells. Indications on disorder are, in the case of stacking disorder, varying peak half width and peak anisotropy, shift of the peak maximum for certain peaks off the Bragg position, and increased scattering contribution from diffuse intensities to the background. If there is dynamic disorder the diffuse scattering intensity is part of the background signal. Commensurate or incommensurate superstructures lead to weak superstructure reflections or to satellite reflections which only careful analysis of the powder data set might reveal.

Finally, there is a limit to **complexity of the material** which is directly related to the size of the unit cell and the number of atoms in the asymmetric unit. Large structures are difficult structures for methods using reciprocal space as well as for methods using direct space techniques. Today, new polycrystalline materials with 20 atoms in the asymmetric unit for inorganic framework structures and 20 non-hydrogen atoms in organic or inorganic molecular structures can usually be solved with direct methods or molecular modelling assisted solution techniques routinely. The challenge to solve ever more complicated structures has pushed the limits to a degree one would not have expected several years ago. The availability of synchrotron light sources and the use of high resolution instruments in combination with a short wavelength has helped the development of the technique steadily.

There is much recent literature on structure solution from powder data. The proceeding of the EPDIC 7 (1) and ECM 19 (2), two crystallographic meetings held in 2000 contain a wealth of information on structure solution from powder data. Since state of the art and complicated structures goes hand in hand the interested reader is referred to the contribution of K.M.D. Harris and Chr. Baerlocher (EPDIC 7) and L.B. McCusker (ECM 19) and the references cited therein.

3 Materials with poor crystallinity leading to powder pattern of moderate resolution

Many materials in applications such as catalysis, adsorption, pigments, ceramics etc. only exist as microcrystalline powder, and, therefore, powder crystallography is the most powerful technique for their structural characterisation. In many cases all attempts to grow good single crystals for structure analysis failed and recourse must be made to the less powerful powder diffraction experiment. On the other side, the fact that those materials exist only as micro crystals is one of their properties, which, in many cases, is very important. These materials diffract poorly and give diffraction pattern of moderate resolution. Broad reflections in X-ray powder patterns often holds off crystallographers and prevents the crystal structure of these materials to be solved. In Fig. 1 the X-ray powder pattern of the microporous material RUT (3) is shown where the scattering strongly falls off above $30^\circ 2\theta$ using synchrotron radiation ($\lambda = 1.1\text{\AA}$; equivalent to $\sim 45^\circ 2\theta$ Cu $K\alpha$ -radiation). In addition, the peak half-width leads to severe overlap which makes indexing very difficult. Since intensity extraction from those peaks to atomic resolution is impossible standard direct methods must fail for the solution of the crystal structure.

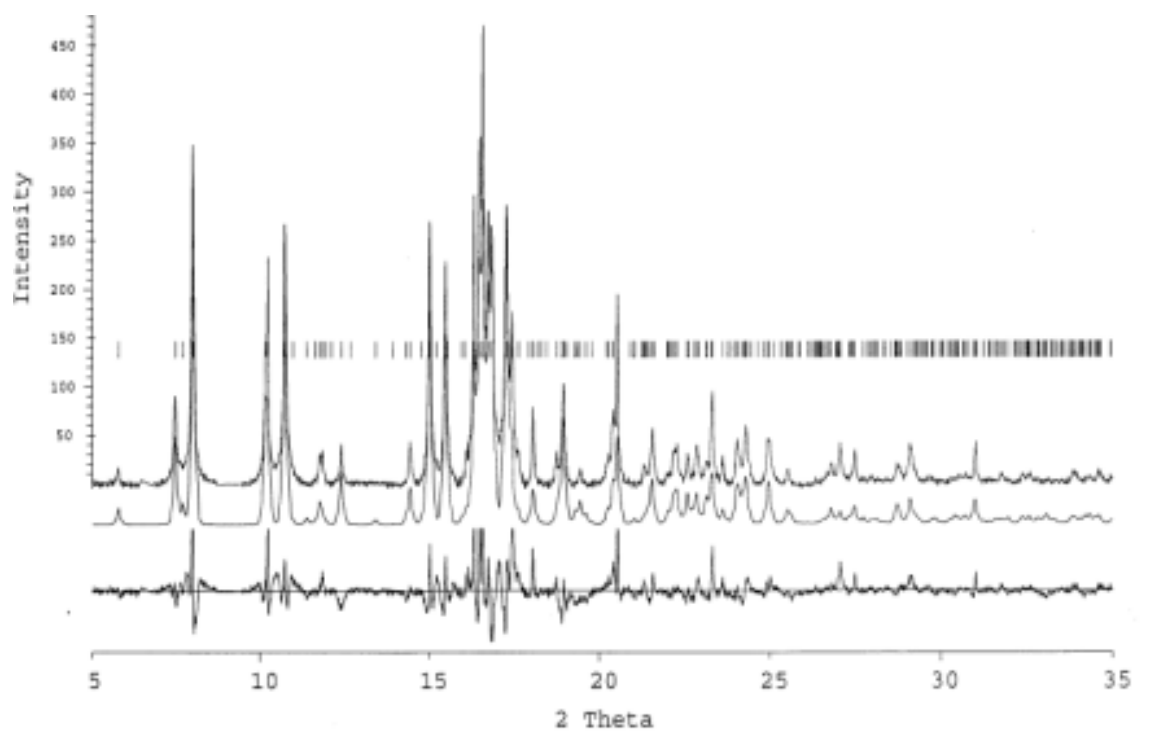


Fig. 1: Synchrotron powder pattern of zeolite RUB-10 ($\lambda = 1.1\text{\AA}$): The broad FWHM of the diffraction peaks is due to imperfections of the material and can't be overcome with high resolution diffractometers.

3.1 Indexing of powder pattern of moderate resolution

In the structural characterisation of materials of moderate resolution every single step in the analysis is at the limit of the techniques and, therefore, more difficult. In order to arrive at a quantitative characterisation as many as possible complementary techniques should be used in addition to the careful chemical and physical analysis of the composition, stoichiometry and density of the material, and the study of the optical properties using polarised light leading to the morphology of the crystal, the refractive indices and the double refraction. For an unknown powder indexing is the first step where pure phase materials is advantageous. For preliminary inspection an optical microscope is an invaluable aid. Colour differences, the morphology, refractive indices and double refraction might distinguish the compounds. DTA and TGA also help to identify impurities. If the material is stable in the electron beam the most useful information is obtained from **electron diffraction**. Here, the diffraction pattern clearly identify different phases and lead to reliable lattice parameters. This combination of techniques was used in the case of the zeolite RUB-10 of which the diffraction pattern is shown in Fig. 1. With the approximate values of the lattice parameters it was possible to completely index and refine the X-ray powder pattern. In cases where the material is not stable in the electron beam optical microscopy can provide restrictions on the possible symmetry. If it is possible to study the morphology and the optical properties at least the point group symmetry might be derived which sets limits on the number of likely space groups. Most indexing programs can use this information actively for their unit cell search. Chemical intuition and knowledge of the crystal chemistry of the particular material might also be of great help. In cases where structural subunits lead to regular increments of unit cell parameters typical d-spacings might be identified and again used as restriction in lattice parameters actively for the indexing.

3.2 Intensity extraction from X-ray pattern of moderate resolution and direct method structure solution

Technically, the extraction of integrated intensities from powder pattern with moderate resolution is without specific difficulties. Based on the indexing and the subsequent unit cell refinement, the profile fit leads straight forwardly to the quantities needed for direct method structure solution. However, the limiting restriction of the interpretation of the powder pattern is the percentage of overlapping peaks. The overlap is increasing with increasing diffraction angle and most severe for low symmetry structures. The broadened peak in poorly crystalline materials make the situation even worse. In many cases there is little overlap in the low angle part of the diffraction pattern which allows for the unambiguous indexing of the intensity maxima. However, the quality of the unit

cell refinement as result of the indexing depends largely on the highest angle reflection which can be assigned without doubt. This ambiguity has consequences for the extraction of intensities from a powder pattern leading to resolution limits in reciprocal space and thus restricting the power of the method. For materials with moderately resolved powder pattern, the ambiguities in the indexing and assignment of intensities arise from accidental and systematic overlap of reflections and, in addition, from the uncertainty in the unit cell refinement.

Generally, atomic resolution in the diffraction data set is required to have optimum conditions for direct methods. In order to overcome this precondition techniques have been developed for powder data (4) with moderate resolution which include additional information from e.g. the Patterson method. The scattering contribution from molecules or rigid fragments such as $[\text{PO}_4]^-$ or $[\text{SiO}_4]$ -polyhedra is used in the phase optimisation with the tangent formula releasing the resolution limit to less than 2\AA corresponding to ca. $45^\circ 2\Theta$ for Cu $K\alpha$ -radiation. The interpretation of the electron densities obtained from direct method calculations is far from trivial and requires an intimate knowledge of the crystal chemistry and the specific bonding scheme of the fragments in the compound investigated. E-map interpretation is in some way similar to macromolecular crystallography where a combination of techniques only allows for the structure analysis and refinement.

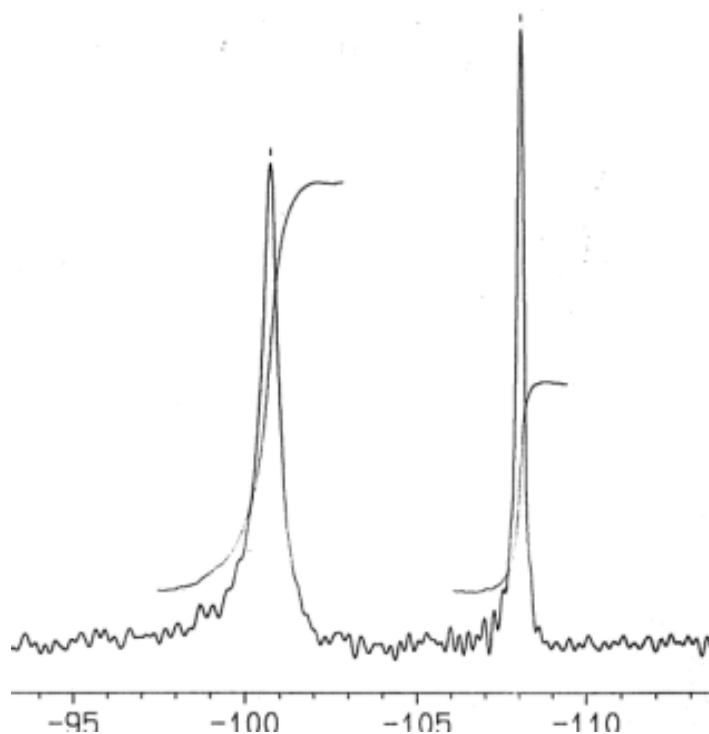


Fig. 2: ^{29}Si MAS NMR spectrum of RUB-15. The Q^3 and Q^4 intensities have a ratio of 2:1.

Again, solid state magic angle spinning NMR (SS MAS NMR) is one of the most useful complementary techniques. Since spectrometers are available nowadays in almost every materials laboratory it is no longer a privilege to fortunate users. In organic crystals the ^{13}C and ^1H spectra might show the integrity of the molecule in its specific stereo-chemistry. In cases where the local order is high its might also be possible to distinguish between different molecules in the asymmetric unit. In inorganic crystals with invariant building blocs such as $[\text{SiO}_4]$, $[\text{PO}_4]$, $[\text{AlO}_4]$, $[\text{GeO}_4]$ etc. SS MAS NMR might yield information on the connectivity of the units, the intensity ratio of units with different connectivity, and their site occupancy in the particular space group.

As an example, the ^{29}Si MAS NMR spectrum of the layer silicate RUB-15 (5) is shown (Fig. 2). The 2 signals correspond to two differently connected $[\text{SiO}_4]$ -units in the ratio 2:1. The low field signal (~ -101 ppm) is characteristic for a $[\text{SiO}_4]$ -tetrahedron with three Si-O-Si bonds and one Si-O-H bond (Q^3 -silicon), the high field signal (~ -108 ppm) is for one with four Si-O-Si bonds (Q^4 -silicon). Since the morphology of the material is typical for layered structures it was concluded that the structure consists of silicate layers surface- Q^3 units are connected with each other via framework Q^4 -units. In addition, ^{13}C MAS NMR showed that tetramethyl ammonium cations are intercalated to balance the charge of the silicate layer. Combining the information obtained from NMR and microscopy with the results of the direct method calculations the E-map (Fig. 3) has been interpreted unambiguously.

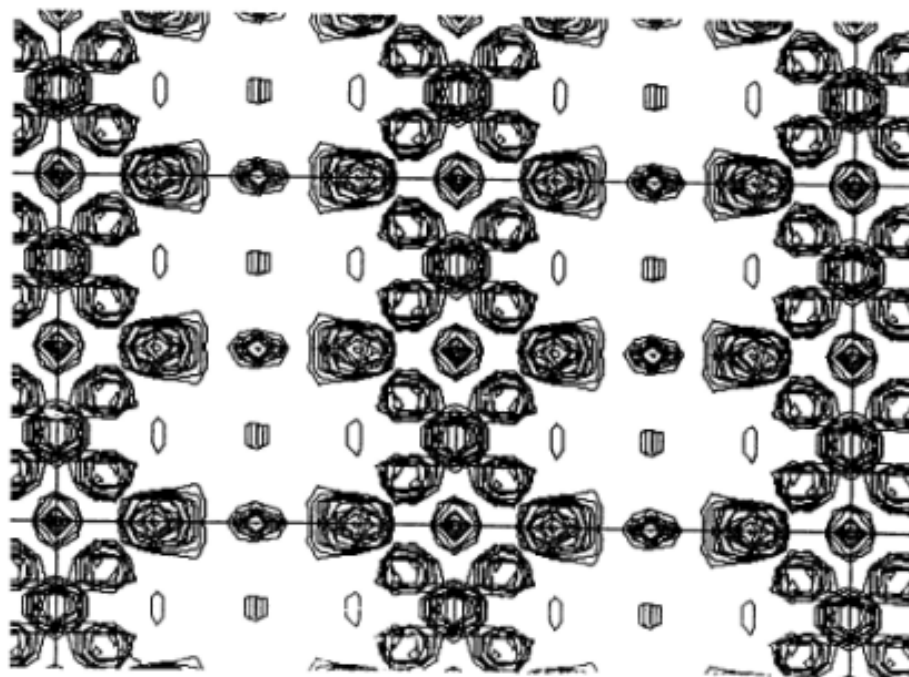


Fig. 3: E-map calculated from powder data of moderate resolution. Here, the fragments $[\text{TX}_4]$ of the silicate show up clearly and allow for the construction of a structure model.

4 Materials with high crystallinity

Compounds which crystallise with high internal order can usually be grown as crystals big enough for single crystal structure analysis. In particular since micro crystal diffractometers are available at synchrotron light sources the crystal volume required for data collection has decreased by a factor of 1000 and more. However, there are still materials with good crystallinity which exist only as powders. In these cases recourse must be made to powder diffraction. For the structure solution from powder data of materials with high crystallinity one should aim at the highest possible resolution in the diffraction data set. In general, synchrotron light sources have high resolution powder diffractometers installed which should be used for such purposes. In addition, shorter wave length could be used leading to better S/N for high order reflections thus improving the resolution of the experiment. The first step in data analysis is the indexing of the powder pattern. The extraction of precise peak positions of all low angle peaks usually leads to successful indexing with standard computer programs. The high resolution powder pattern also yields high precision cell parameters and reliable integrated intensities from the powder pattern. In fortunate cases where the complexity of the crystal structure is moderate (25 non-hydrogen atoms in the asymmetric unit) and accidental or systematic overlap is low structure solution with direct methods is routine.

For reciprocal space methods, more complicated materials require again additional information which is generally intended to reconstruct information lost through the overlap problem. Most of the techniques have already been covered in the different lectures presented in the course of the workshop. They include the use of multiplet relations (6), Patterson recycling (7), direct method approach to retrieve intensity information on systematically and accidentally overlapping reflections (8) and, finally, the use of high resolution textured powder pattern for the reconstruction of a three dimensional intensity data set (9). When known, fragments of the structure should be used in the structure solution process as discussed above for materials with moderate resolution powder pattern. This information is accessible with spectroscopic techniques such as SS MAS NMR, microscopic techniques, and knowledge in crystal chemistry.

Computational techniques which explore direct space have also been used to solve crystal structures from this group of materials, with or without the active use of information from the experimental powder pattern. Monte Carlo simulations (10), simulated annealing (11), generic algorithm (12) and grid search (13) are the most successful approaches, but pattern recognition in maps calculated from Patterson synthesis (14) and direct methods FOCUS (15) have also been applied successfully. In all these techniques and strategies it is most useful to use the experimental powder pattern as "penalty function" for the evaluation and the development of the proposed solution.

In all cases where high resolution data are available, difficult structures for direct methods are those where the accidental and systematic overlap is high (more than 30% up to $\sim 1.5 \text{ \AA}$) and the number of non-hydrogen scatterers is high (more than 30 atoms in the asymmetric unit). For computational techniques all those structures are complicated where the number of "degree's of freedom" is high (flexible molecules in molecular crystals, high number of non-equivalent fragments in inorganic materials) and/or the number of molecules per asymmetric unit is more than one.

In the case of well crystallised materials which give high resolution powder pattern there is no general rule on how to proceed to solve a problem structure. The different techniques all exploit the experimental diffraction data to the most. In the future the combination of those techniques which are complementary are most promising to solve even more complicated structures of this kind. A first attempt is the implementation of Monte Carlo modelling into a direct method program EXPO2000 which will face severe tests in the near future.

5 Materials with superstructure, commensurate and incommensurate modulated structures, dynamic disorder, stacking disorder

Materials with structural disorder and pseudo symmetry are very difficult to deal with unless their property is obvious. Since diffraction experiments average in time and space, an averaged structure is represented in the Bragg peaks of the diffraction pattern. If one suspects disorder or pseudo symmetry in a materials crystal structure, electron diffraction is the most suitable technique to use. Superstructure reflections, satellites, diffuse intensities and diffuse streaks are easy to detect and observe in the electron diffraction diagram if the material is stable in the electron beam. The TEM images, in addition, give insight in the local structure and the distribution of the disorder. It is recommended then to transfer the information obtained from the electron diffraction experiment concerning the disorder to powder diffraction.

Since the powder diagram is a projection of the three-dimensional reciprocal space onto one dimension, all non-Bragg intensities are simultaneously present in the powder trace and must be taken into account for the detailed analysis of the crystal structure. Hints for these phenomena mentioned above are increased diffuse background intensities and/or weak extra-reflections which are not accounted for in the indexed unit cell and which are not related to impurities. In the case of a **superstructure** the doubling of a unit cell parameter would explain the extra reflections, however, the structure solution problem would become more difficult since the number of atoms in the asymmetric unit will increase whereas the number of observations has only grown slightly. Since weak superstructure reflections are indicators for small deviations of atoms or

fragments from the high symmetry arrangement in the unit cell **structure solution should be attempted in the pseudo-symmetric space group** with the reduced number of parameters. An instructive example is given by Estermann et al. for SAPO-40 (16). The microporous alumophosphate framework structure has Al and P alternating on tetrahedral positions. For the structure solution the ordering of the T-atoms was neglected increasing the symmetry and at the same time reducing the number of independent atoms from 24 to 12. Only in the subsequent Rietveld refinement the true symmetry describing the ordering of the T-atoms Al and P was considered.

Commensurate and incommensurate structures will show satellite reflections separated from the position of the Bragg reflection depending on the modulation periodicity. Similar to what was already explained for superstructure reflections the **satellites should be neglected in the structure determining process**. Once the averaged structure is known a model for the modulation of a fragments can be worked out and included in the refinement process. As an example the incommensurately modulated structure of tridymite is given (17) Explicit details are also discussed in the lecture of S. van Smaalen in the course of the summer school.

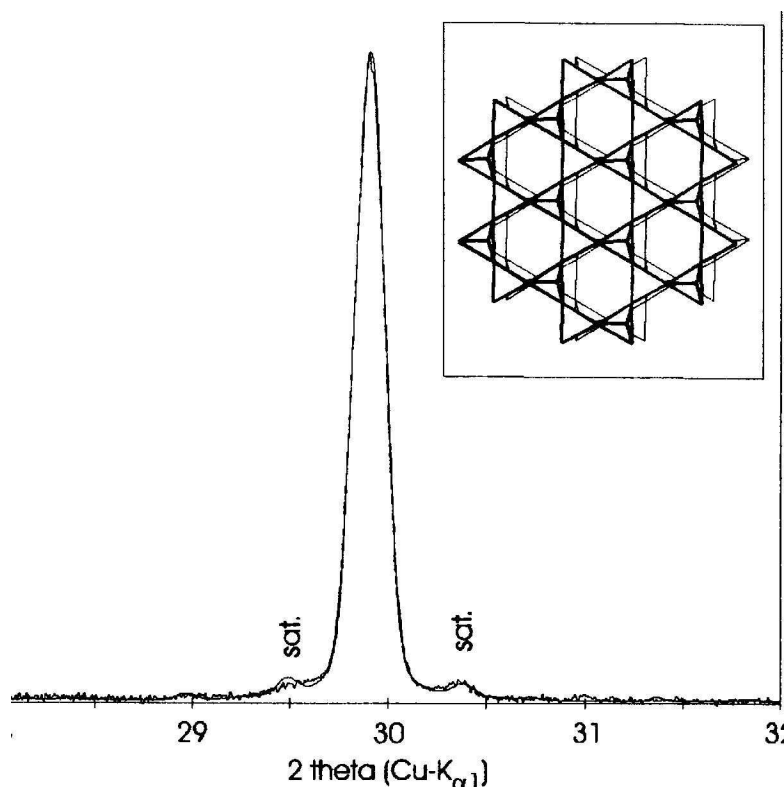


Fig. 4: Bragg peak and satellites in the powder pattern of modulated Tridymite at 160°C.

In crystal structures with **static and/or dynamic disorder** of fragments of the structure the scattering contribution of the fragment is restricted to low angle

reflections. Contribution of diffuse scattering to the Bragg reflections is apparent in a broadening of the base of the peak. This might lead to difficulties in fitting an analytical profile function to the reflection. However, intensity extraction is only little affected and **direct methods usually have no particular problems** with data sets from materials with static and/or dynamic disorder. For the study of disorder NMR is also a valuable tool. Since the technique probes the local geometry and is sensitive to motion in general, the static spectrum of the mobile fragment should show considerable motional narrowing in the resonance peaks. This can be exploited for structure determination; for the Rietveld refinement information on dynamic disorder is of great value. As example the structure solution of the alumosilicate zeolites MCM-61 is given (18). Here, the inorganic three-dimensional framework contains disordered organic templates in cage-like voids. These templates are dynamically disordered, however, structure solution with FOCUS (direct methods – pattern search) (15) wasn't hampered.

There are many materials where various packing or bonding motives can occur. The most familiar disorder family of structures are the dense packed metals where, under certain circumstances, the hexagonal AB stacking of layers randomly mixes with the cubic ABC. This leads to an undefined periodicity along the stacking direction leading to continuous scattering intensities along the correspondent reciprocal axis. Materials with **stacking disorder** of two- or one-dimensional invariant structure slabs such as the hexagonal layer in dense packed metals give rise to very unusual powder pattern. Depending on the degree of disorder sharp and broad reflections are present in the powder diagram (Fig. 5). In addition the 2Θ position of the peak maximum of the broad peaks is not on a Bragg angle. Therefore indexing of the powder pattern must fail and also structure solution is not possible. In those cases where one observes varying peak half width, the indexing of the powder pattern fails, and when without doubt the material is a pure phase, electron microscopy and diffraction is the most powerful, if not the only tool to solve and analyse the structure of the material. However, if the degree of disorder is small and in the order of a few percent the powder pattern is largely dominated by the major structural component. Therefore, it might be possible to index the powder pattern completely. The refinement of the cell parameters will show unusually high ESD's but will be close enough to the dominant ordered end member structure that direct methods still work and yield a model for the dominant structure type. As an example the structure solution of zeolite beta (Fig. 5) is given (19). The citation also gives a historical overview on the theory of disorder in solids and its implications on the diffraction pattern. The structure of zeolite beta was solved by a combination of electron diffraction, high resolution electron microscopy and powder diffraction. The simulated powder pattern showed that the stacking of the invariant silicate layer is almost random with a 50:50 probability of two different stacking variants.

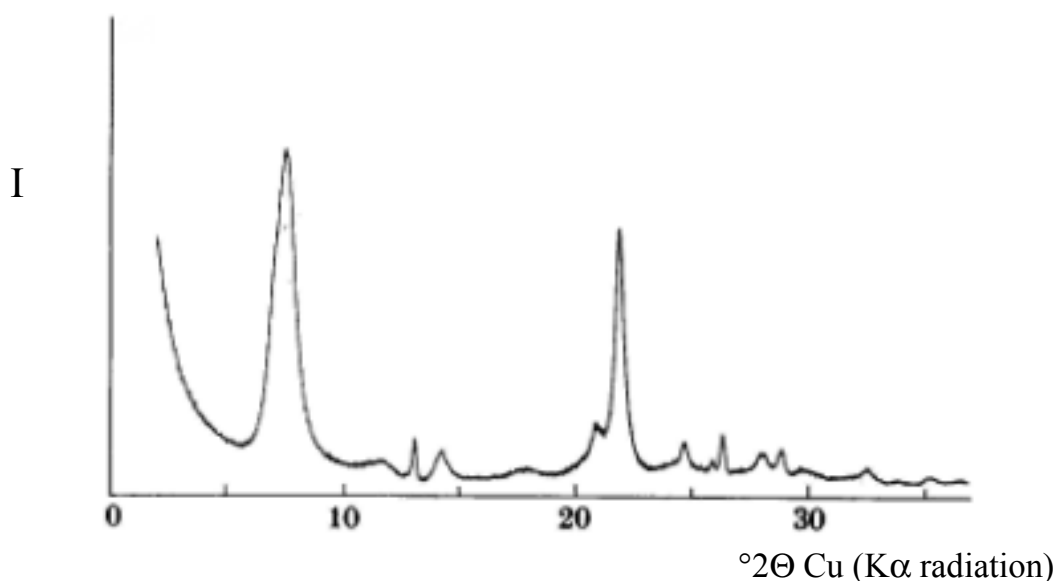


Fig. 5: Powder diagram of zeolite beta. Broad and sharp reflections indicate stacking disorder of the silicate framework. The simulation of the powder diagram revealed 50/50 stacking disorder.

6 Conclusions

The contribution summarises the most frequently met difficulties solving crystal structures from powder data. In selected examples a brief introduction in the basic understanding of the complication is given. Those who are interested in a more complete discussion of one or the other problem the cited literature might help in particular the proceedings and abstract of the latest crystallographic meetings, where the progress in the technique is presented and difficult examples are abundant.

7 Literature:

- 1 Proceedings of EPDIC 7
- 2 Collection of abstracts, ECM 19
- 3 H. Gies, J. Rius, Z. Kristallogr. **210**, 475-480(1995).
- 4 a) J. Rius, Powder Diffraction **14**, 267-273(1999).
b) A. Altomare et al., J. Appl. Cryst. **32**, 339-340(1999).
c) R. Spengler et al., Acta Cryst B **50**, 578-582(1994)
- 5 U. Oberhagemann et al., Angew. Chemie **108**, 3041-3043(1996).
- 6 J. Jansen et al., Z. Kristallogr. **206**, 33-43(1993).
- 7 M Estermann, V. Gramlich, J. Appl. Cryst. **26**, 396-404(1993).
- 8 a) J. Rius et al., J. Appl. Cryst. **32**, 89-97(1999).
b) J. Rius et al., Eur. J. Mineral. **12**, 581-588(2000).
- 9 T. Wessels et al., Science **284**, 477-479(1999).
- 10 M. Tremayne et al., J. Appl. Cryst. **29**, 211-214(1996).
- 11 W.I.F. David et al. Chem. Commun. **1998**, 931-932(1998).
- 12 K. Shankland et al., Z. Kristallogr. **212**, 550-552(1997).
- 13 V.V. Chernyshev H. Schenk, Z. Kristallogr. **213**, 1-3(1998).
- 14 H. Gies, J. Rius, Z. Kristallogr. **210**, 475-480(1995).
- 15 R.W. Grosse-Kunstleve et al., J. Appl. Cryst. **30**, 985-995(1997).
- 16 M. Estermann et al., J. Appl. Cryst. **25**, 539-543(1992).
- 17 H. Graetsch, Z. Kristallogr., **Suppl. 16**, 83(1999).
- 18 D.F. Shantz et al., Micropor. Mesopor. Mat. **31**, 61-73(1999).
- 19 a) J.M. Newsam et al., Proc. R. Soc. Lond. A **420**, 375-405(1988).
b) M.M.J. Treacy et al.: Proc. R. Soc. Lond. A **433**, 499-520(1991).

The Importance of the Inorganic Crystal Structure Database ICSD for the Application of Rietveld-Methods

Rudolf Allmann

*Institut für Mineralogie, Hans Meerwein Str., D35043 Marburg
E-mail: napieral@mail.uni-marburg.de*

As soon as for an intended Rietveld-calculation the cell, the space group, and the chemical cell contents are known, it is worthwhile to go into the corresponding crystal structure database and search for isotypic compounds. If such a structure can be found, its atomic parameters deliver a sufficient starting set for a successful Rietveld refinement.

Since my retirement in 1996 I help the inorganic database ICSD to find errors and missing data. At first I checked the about 10,000 mineral structures contained in ICSD. For about 4,000 of these, the mineral names were missing as well as localities and measured densities. As a 2nd mineral name the group name as in PDF was added. In the soon expected WINDOWS program the 'mineral group' will be a field by its own as well as the 'structure type'. The references to the Powder Diffraction File were added as PDF numbers.

Then I started to check the volumes of Structure Reports for structures missing in ICSD. The years 1951 – 1990 are done with the result of 6,000 structures newly added to ICSD. More than 3000 existing entries were corrected. This work is going on. I thank all authors who helped me to complete their structures in ICSD.

With the 2nd update 1999 53,373 entries are included, of which 2,893 are new and 2,264 corrected or completed. For these entries a reference to the Powder Diffraction File PDF is given.

Definition of an inorganic structure to be incorporated into ICSD

Inorganic structures are structures which don't contain any C-H or C-C bonds in any residue and contain at least one of the elements:

H						He
	B	C	N	O	F	Ne
		Si	P	S	Cl	Ar
		Ge	Sb	Se	Br	Kr
				Te	I	Xe
					At	Rn

In addition metal carbides and inorganic frameworks including organic residues (e.g. *Zeolites* including ethanol) are allocated to inorganic structures. Exceptions of above constraints hold for minerals and elements, which are included all.

The database contains all structures of which three-dimensional atomic coordinates have been determined directly or indirectly and if the coordinates are published or deposited. Other deposited data are included too if possible (e.g. anisotropic temperature factors).

Coordinates for hydrogen atoms or vagabonding atoms like Na in zeolites may be missing. Structures described as isotypic to known structures but without determination of free parameters are omitted. Parameter free structures, e.g. of the NaCl-type, should be included, but were often overlooked in the beginning of this database. If super- and substructures have been described ICSD contains both of them.

What you as author can do for ICSD

Authors can easily help to complete the database ICSD! Just check your **publication list** against the entries in ICSD going with your name and send me reprints of the missing data or tell me printing errors in your publications. Do **anisotropic temperature factors** exist, which are not published? If your publication list contains more than 50 inorganic structures, I am willing to do this for you: just send me your publication list. About 10% of your published structures are probably missing in ICSD.

Another problem are **unpublished structures**, which were only presented as **posters** and the corresponding meeting abstracts are too short to contain a list of atomic parameters. If the meeting abstracts can be cited, you may send me the parameter list together with a copy of the abstract and I will include the structure in ICSD. Nearly half of all structures presented as posters are never published in full and your valuable data are **lost for ever**.

Errors in publications

Obvious printing errors may be detected by too short atomic distances or unlikely calculated densities D_m . Most of these errors could be corrected in ICSD. Frequent printing errors are:

- Missing signs
- Missing leading zeros
- Interchanged digits
- Wrongly doubled digits (e.g. .113 instead of .133), easily overlooked in proof-reading

More difficult are substantial errors as:

- Unusual origin of space group (most often for $P2_12_12_1$)
- Wrong space group
- Missing or complementary angle β of a monoclinic cell
- Wrong constraints for anisotropic temperature factors (for trigonal or hexagonal cells)

