FIDDLE. Simultaneous Indexing and Structure Solution from Powder Diffraction Data using a Genetic Algorithm and Correlation Functions

Carmen Guguta, Jan M.M. Smits, René de Gelder*

Radboud University, Institute for Molecules and Materials, Heyendaalseweg 135, 6525 AJ Nijmegen, The Netherlands. E-mail: r.degelder@science.ru.nl

ABSTRACT

The usual process for crystal structure determination from powder diffraction data consists of (1) indexing of the powder pattern, (2) space group determination, (3) structure solution and (4) structure refinement. Despite the success of methods for powder indexing, there are many cases in which the very first step in a structure determination fails or is far from straightforward. Due to a number of fundamental and experimental problems, like peak broadening, the presence of impurity phases, dominant zones and geometrical ambiguities, powder indexing will remain difficult in many cases, thereby hampering the next steps in the structure determination.

We present a method for the determination of crystal structures from powder diffraction data that circumvents the difficulties associated with separate indexing. Structure determination from powder diffraction data can be seen as a process of global optimization of all model parameters, including the unit cell parameters. For the simultaneous optimization of the parameters that describe a crystal structure a genetic algorithm is used together with a pattern matching technique based on auto and cross correlation functions. We show that this "one-pot" strategy for indexing and structure determination can successfully be used for cases for which indexing is problematic.

INTRODUCTION

Nowadays, single-crystal X-ray diffraction is a standard method for the determination of molecular structures and for the elucidation of intra and intermolecular interactions in various materials. Often, however, crystals suitable for single-crystal diffraction cannot be obtained. Compounds that show polymorphism or solvate (hydrate) formation frequently cause difficulties in the crystallization of one of the modifications. In addition, most of the times recrystallization of unstable compounds that are initially obtained as microcrystalline powders is not feasible. A possible route for crystal structure analysis of "problematic" compounds is to make optimal use of powder diffraction data. These data contain less information than single crystal diffraction data but for a lot of compounds powder diffraction has allowed for a full crystal structure determination.

The complexity of crystal structures determined from powder diffraction data has steadily increased through further development of "traditional" methods for structure determination in reciprocal space and application of global optimization algorithms in direct space (Harris et al., 2001; David et al. 2002; Favre-Nicolin et al., 2002; Altomare et al., 2004; Altomare et al., 2004; Tremayne, 2004; David and Shankland, 2008). The published number of crystal structures

determined by powder diffraction is still rather limited compared to the enormous number of structures determined by single-crystal diffraction. Clearly, structure determination from powder diffraction data is not a generally applicable method yet.

The lack of success in the indexing step has become one of the major bottlenecks in structure determination from powder diffraction data. There are a number of fundamental and experimental problems that can make powder indexing problematic or simply impossible: peak broadening (leading to a loss in resolution), increasing peak density at higher angles, peak shifts, systematic or accidental absences of reflections, the presence of impurity phases, inaccuracies in the experimental measurements, dominant zones and geometrical ambiguities. Problems that originate from the use of laboratory instruments, like peak broadening caused by the instrumental profile and experimental inaccuracies, can be reduced by measuring on synchrotron facilities. However, problems originating from specific powder imperfections (impurities, crystallite size, strain, lattice mistakes) cannot be solved by going to more sophisticated instruments. This means that for powder patterns measured on laboratory as well as synchrotron facilities the problem of limited quality can exist.

It is clear that particularly for compounds for which no suitable single crystals can be grown the powder method is an attractive alternative. Often, however, the problematic crystallization behavior of these compounds is also reflected by the moderate quality of their microcrystalline powders and their corresponding powder diffraction data.

Despite the success of standard methods for powder indexing, like ITO (Visser, 1969), TREOR (Werner, 1985) and DICVOL91 (Boultif, 1991), there are many cases in which the very first step in structure determination fails or is far from straightforward. New strategies and enhancements will certainly extend the possibilities of powder indexing methods (Coelho, 2003; Neumann, 2003; Altomare et al., 2000). However, the problems associated with the nature of powder diffraction data (absences, increasing peak density etc.) together with the usual strategy of applying a separate indexing step, which may lead to geometrical ambiguities, will be a fundamental limitation for improvement.

A method called OCEANA was reported for structure determination from powder diffraction data without prior indexing (Padgett et al., 2007). OCEANA is a grid search based method which uses a genetic algorithm to adjust the cell parameters, potential packing energy and R_{wp} to locate the global minimum in a given space group. The method is limited to $Z' \leq 1$.

In this paper a different methodology for the determination of crystal structures from powder diffraction data is presented that circumvents the difficulties associated with separate indexing. The method, implemented in the program *FIDDLE*, has the possibility of searching through the most common space groups, while varying Z' and without taking into account any energy or packing considerations. By using correlations functions for pattern matching the method is able to optimize cell and structural parameters simultaneously using a genetic algorithm, without any assumptions on cell parameters or density. Results from several tests of the method on various organic compounds are presented to demonstrate the effectiveness of the concept. Moreover, using *FIDDLE*, three previously unknown crystal structures, cases for which indexing was problematic, could be determined successfully.

COMPUTATIONAL METHODOLOGY

Program description

There is no fundamental reason to separate the process of unit cell determination and the process of structure solution. If one is able to describe the main part of a crystal structure by a model consisting of a discrete set of parameters (cell parameters {a, b, c, α, β, γ}, space group number S and - using information on molecular structures - a limited set of structural parameters describing the positional parameters {$(x, y, z)_j$}), then structure determination from powder diffraction data can be seen as a process of (global) optimization of all crystal structure parameters, including the unit cell parameters. This is the strategy implemented in the *FIDDLE* program. Structure solution then means that in the end the calculated powder diffraction pattern corresponding to the complete set of optimized parameters, including the cell parameters, must match the experimental pattern, assuming that there is one unique structure which has this property.

The *FIDDLE* method is a direct space method. Prior knowledge of the structure of molecular fragments (and the internal degrees of freedom) is needed to define a model of the crystal structure in terms of structural parameters that can be optimized. Like in other direct space methods *FIDDLE* reduces the set of positional parameters to a set of rotation angles {θ, ϕ, ψ}, which define the orientation of a molecular fragment in the unit cell, a translation vector {x, y, z}, which defines the position of the molecular fragment in the cell and a set of n torsion angles {τ_1, τ_2, ..., τ_n}, which describe the intramolecular geometry of the fragment.

Part of the parameters, the cell parameters, define the positions of the peaks in the powder pattern. The other parameters, space group and positional parameters, define the intensities of the peaks. Together, they define the total pattern (not regarding zero-point shifts and peak shapes). A suitable measure for similarity between experimental and calculated pattern must be used to be able to optimize both sets of parameters simultaneously. *FIDDLE* uses weighted auto and cross correlation functions for pattern matching and the mathematics for this fitness function are described below.

For space group determination *FIDDLE* uses a strategy that is also applied in crystal structure prediction: the frequency of the occurrence of space groups in the CSD (Cambridge Structural Database) is used and systematically the most common space groups are explored. About 79% of all organic and organometallic compounds crystallize in only 5 space groups: $P2_1/c$, $P\text{-}1$, $P2_12_12_1$, $C2/c$ and $P2_1$ (Allen, 2002). Chiral molecules (non-superimposable on their mirror image) crystallize in Sohncke space groups and this further reduces the number of possible space groups. *FIDDLE* explores the space groups $P2_1/c$, $P\text{-}1$, $P2_12_12_1$, $P2_1$, $C2/c$, *P*bca, *P*nma, *P*bcn, and *P*1 and varies Z'. Prior knowledge about Z' obtained from methods like solid-state NMR can reduce the problem. Fig. 1 shows the main steps in the *FIDDLE* procedure.

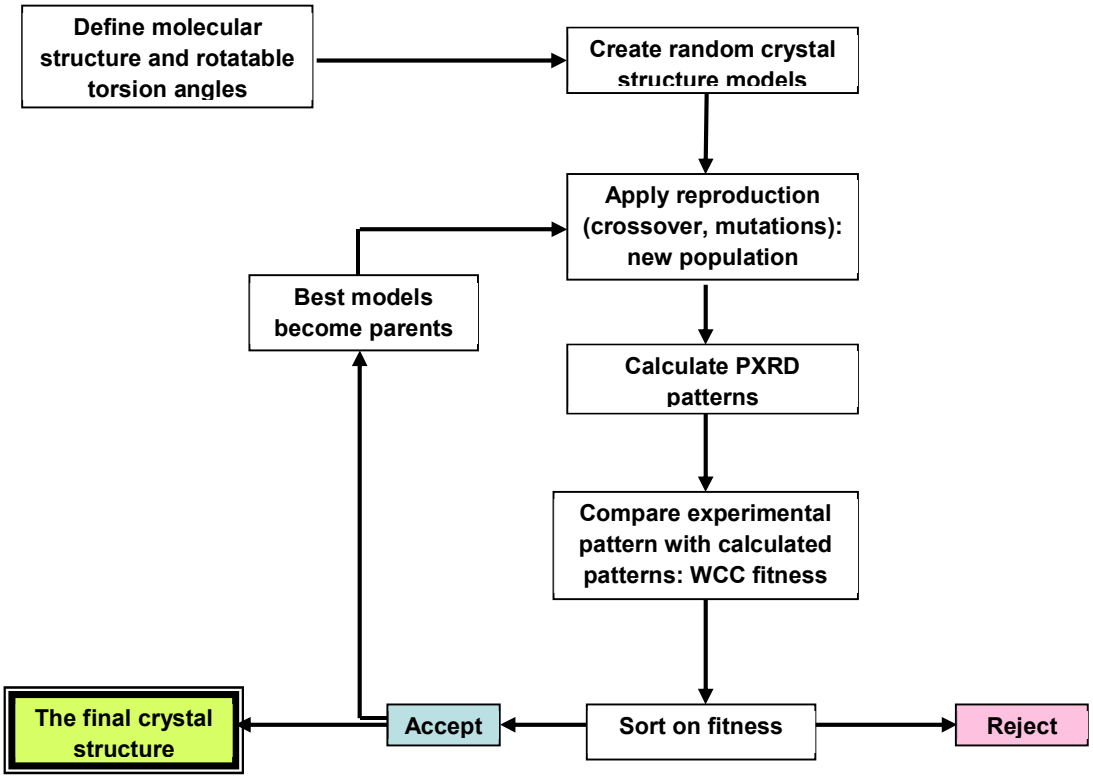


Figure 1. Overview of the main steps in the *FIDDLE* procedure

The Genetic Algorithm

A Genetic Algorithm (GA) is a global optimisation method based on evolution principles (Harris *et al.*, 1998). Evolutionary operations such as mating, mutation and "natural selection" are applied, through which members with the highest fitness value within a population survive and procreate. Genetic algorithms can be applied to any problem in which the quantity to be optimized can be written as a function of a set of variables (Goldberg, 1989; Cartwright, 1993; Keane, 1996).

The settings of the Genetic Algorithm implemented in the *FIDDLE* program were determined by trial and error and were initially based on experience with the determination of molecular constants from rovibronic spectra with genetic algorithms (Hageman *et al.*, 2000). Table 1 shows the genetic algorithm settings used in *FIDDLE*.

Table 1. Genetic algorithm settings in *FIDDLE*

Settings	Value
Maximum number of generations	300
Population size	400
Elitism	200
Crossover type	Two-point crossover
Unit cell edges boundaries	3-40 Å
Unit cell angles boundaries	60-120°
Orientation angles boundaries	0-360°
Torsion angles boundaries	0-360°
Mutation type	New random value within boundaries
Mutation probability	0.02
Selection type	Probabilistic
Fitness type	Weighted cross correlation function

The unit cell parameters (cell edges and angles) and the positional parameters such as orientation angles and the torsion angles are randomly set within a range of values during initialization (see Table 1). This also holds for the translation vector(s) but the boundaries for these parameters are dependent on the space group.

Comparison of powder diffraction patterns in FIDDLE: the fitness function

The *RMS* (Root Mean Square) criterion is a well known criterion that is based on the sum of the squared differences between *n* observed and *n* calculated data values, respectively $y_i(obs)$ and $y_i(calc)$. A well-known criterion for powder pattern similarity is R_{wp} (R-weighted pattern). This criterion is used for the refinement of crystal structures on XRPD data and also for the structure solution from XRPD data. R_{wp} is similar to an *RMS* criterion but the squared differences are weighted according to the standard deviations in the observed intensities.

$$R_{wp} = \left(\sum_{i=1,\,n} w_i \left(y_i(obs)-y_i(calc)\right)^2 \right)^{1/2} / \left(\sum_{i=1,\,n} w_i\, y_i(obs)^2 \right)^{1/2}$$

$$w_i = 1/y_i(obs)$$

R_{wp} can only be used when the peak positions in the observed and calculated powder patterns are the same or at least very close. It is therefore useless for the comparison of powder patterns corresponding to structures with significantly different unit cells.

The *WCC* (Weighted Cross Correlation) criterion was developed to deal with cases where two patterns are different with respect to peak positions. It is based on correlation functions (de Gelder *et al.*, 2001):

$$WCC = \int w_{fg}(r)c_{fg}(r)\mathrm{d}r \,/\, (\int w_{ff}(r)c_{ff}(r)\mathrm{d}r \int w_{gg}(r)c_{gg}(r)\mathrm{d}r)^{1/2}$$

$$c_{fg}(r) = \int f(x)g(x+r)\mathrm{d}x$$

The weighting function, which extracts information from the correlation functions, can be adapted to influence the sensitivity for shifts in peak positions, in XRPD as a result of lattice parameter variations. This is done by changing the value of width *l*:

$$w_{fg}(r) = 1 - |\, r \,| \,/\, l \qquad \text{if } |\, r \,| < l$$

$$w_{fg}(r) = 0 \qquad \text{if } |\, r \,| \geq l$$

$$w_{ff}(r) = w_{gg}(r) = w_{fg}(r)$$

The *WCC* criterion is always normalized and scaling is unnecessary since this is done implicitly. Comparison of deformed patterns, caused by unit cell variations, is possible with the *WCC* criterion since it recognizes shifted peaks. When we divide the intensities of the powder patterns by the standard deviations in the observed intensities and calculate *WCC* at $r = 0$, we obtain the relation between *WCC* and R_{wp}, a relation between a sine and cosine function:

$$WCC_{r=0} = (1 - R_{wp}^2)^{1/2}$$

So, (only) for $r = 0$ there is a direct and clear relation between R_{wp} and *WCC*.

EXPERIMENTAL

16 compounds covering a range of structural complexity, number of torsion angles, different Z' and different space groups were selected from the CSD in order to test the methodology implemented in the *FIDDLE* program (Table 2).

The known crystal structure of morphine anhydrate was re-determined from laboratory X-ray powder diffraction data with the *FIDDLE* program. This re-determination was carried out to test the performance of the *FIDDLE* method for experimental data.

The crystal structures of three compounds (ethinyl estradiol anhydrate, naltrexone monohydrate and creatine anhydrate) were determined from laboratory X-ray powder diffraction data using the *FIDDLE* program. Indexing the three powder patterns with the most commonly used indexing programs, such as: DICVOL91, ITO and TREOR, was not successful and therefore the determination of these three crystal structures was of particular interest.

Table 2. Compound name and corresponding molecular formula

Code	*Compound name*	*Molecular formula*	*CSD Refcode*
1	Dihydroqinghaosu	$C_{15}H_{24}O_5$	COTYAL
2	3-Methoxy-11-methyl-4a,5,9,10,11,12-hexahydro-6H-benzofuro(3a,3,2-ef)(2)benzazepin-6-ol	$C_{17}H_{21}NO_3$	SIBHAM
3	17β-Hydroxy-7α-methyl-androst-5-en-3-one	$C_{20}H_{30}O_2$	HMANDR
4	(-)-S-2-(2-Chlorophenyl)-2-methylaminocyclohexanone	$C_{13}H_{16}ClNO$	XURCAO
5	(+)-R-2-(2-Chlorophenyl)-2-methylaminocyclohexanone	$C_{13}H_{16}ClNO$	XURCES
6	2',4'-Difluoro-4-hydroxy-3-biphenylcarboxylic acid	$C_{13}H_8F_2O_3$	FAFWIS02
7	(R)-(-)-3-Acetoxy-quinuclidine methiodide	$C_{10}H_{18}NO_2^+I^-$	ACQUIN
8	2-Amino-6-ethyl-4,5,7,8-tetrahydro-6H-oxazolo(5,4-d)azepine	$C_9H_{15}N_3O$	BERFUZ
9	3-(2-Hydroxyphenyl)-4-methyl-1,2,4-Δ^2-triazoline-5-thione	$C_9H_9N_3OS$	HPMTZT
10	(R)-(+)-2-(2-Oxo-3-piperidinyl)-1,2-benzisothiazol-3(2H)-one-1,1-dioxide	$C_{12}H_{12}N_2O_4S$	SUCROS01
11	Sucrose	$C_{12}H_{22}O_{11}$	CAVGOV
12	2-(2,6-Dichloro-3-methylphenyl)-aminobenzoic acid	$C_{14}H_{11}Cl_2NO_2$	MECLOF10
13	(3aα,5α,6α)-4,4,6-Trichloro-7-methoxy-3,4,5,6-tetrahydro-2H-3a,5-methanoazulene	$C_{12}H_{13}Cl_3O$	RUWLOK
14	2-Phenyl-7-methyl-8-azahypoxanthine	$C_{11}H_9N_5O$	AZHPXA
15	10,11-Dihydro-10-hydroxy-5(H)-dibenz(b,f)azepine-5-carboxamide	$C_{15}H_{14}N_2O_2$	KAMCEG
16	(3-(((2-((Aminoiminomethyl)amino)-4-thiazolyl)methyl)thio)-N-(aminosulfonyl)propanimidamide)-nickel(ii)	$C_8H_{13}N_7NiO_2S_3$	BEVQAV

The X-ray powder diffraction measurements for the structure solutions were performed using a Bruker D8 AXS Advance X-ray Diffractometer. The D8 was equipped with a Johansson type monochromator with a focusing curved Ge 111 crystal. A VÅNTEC-1 detector was used with an

effective angular region of 2°. The data were collected in transmission capillary geometry using monochromatic Cu Kα_1 radiation. The most important instrumental and data collection parameters are presented in Table 3.

Table 3. Instrumental and data collection parameters

Typical instrument settings	
System	Bruker D8 AXS Advance $\theta/2\theta$
Generator	40 kV, 40 mA
Measuring circle (mm)	435
Radiation (Å)	Cu Kα_1, λ=1.54056Å
Monochromator	Primary, focusing curved Ge 111
Geometry	Transmission capillary configuration
Sample holder	0.5 mm glass capillary tube
Detector	VÅNTEC-1
Typical measuring conditions	
Range (°2θ)	2-50
Step size (°2θ)	0.0084696
Step time (s)	50
Total data collection time (h)	ca 74
Spinning (rpm)	15

TESTS APPLIED TO CALCULATED POWDER DIFFRACTION DATA (A)

The crystal structures of 16 compounds selected from the CSD database were determined with *FIDDLE* to verify the effectiveness of the program and the methodology, for cases with various symmetries and structural complexities (Table 2 and 4). The input molecular models were also obtained from the CSD. The calculations were performed on 6 AMD Dual-Core Opteron 280 machines (two dual core 2.4GHz processors each), all running under the LINUX operating system. The total number of seeds used for the various crystal structure determinations was between 2400 and 4800. The maximum 2θ range used was 40°. For each compound the tests were done in the correct space group and for the true value for Z'. On the best *FIDDLE* solutions having the highest fitness values a full final Rietveld refinement was carried out using Topas 3. Fig. 2 shows a representative overlay between the solution with the highest fitness value obtained from *FIDDLE* and the known crystal structure in the CSD for one of the studied compounds (FAFWIS02). It is clear that the final result obtained with *FIDDLE* is close to the true crystal structure of this compound. The compound has one torsion angle and crystallizes in the orthorhombic space group $P2_12_12_1$ with unit cell parameters: a=39.700(1) Å, b=14.129(2) Å, c=3.835(6) Å, V=2151.51 Å^3, Z'=2. From the data presented in Table 4 it can clearly be seen that when the correct space group and Z' is used, the correct unit cell parameters are always found

for the top twenty solutions. Always the first solution with the highest fitness value appeared to be the correct crystal structure.

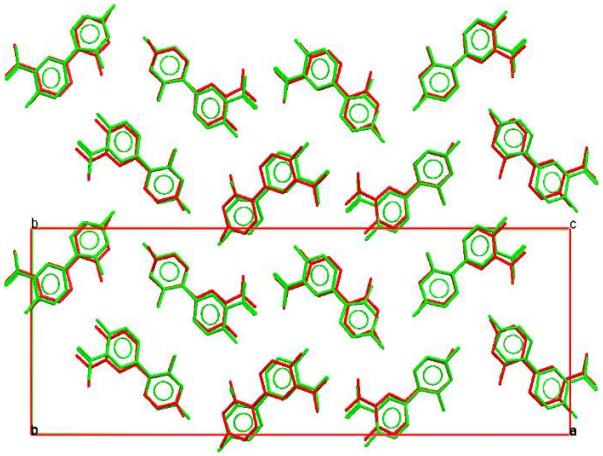


Figure 2. The overlay between the solution with the highest fitness value obtained from *FIDDLE* (red) and the known crystal structure in the CSD (green) for FAFWIS02

Table 4. CSD compounds selected for testing of *FIDDLE*

CSD refcode	*Space group*	*Z'*	*No. torsion angles*	*Width (Å)**	*No. seeds*	*Best Fitness value*	*No. correct unit cells in the first 20*	*No. correct solutions in the first 20*
COTYAL	$P2_12_12_1$	1	0	1.5	2400	0.8369	20	20
SIBHAM	$P2_12_12_1$	1	1	1.5	2400	0.8236	20	20
HMANDR	$P2_12_12_1$	1	0	1.5	2400	0.8058	20	20
XURCAO	$P2_12_12_1$	1	2	1.5	2400	0.7946	20	20
XURCES	$P2_12_12_1$	1	2	1.5	2400	0.7973	20	20
FAFWIS02	$P2_12_12_1$	2	4	0.5	4800	0.7477	20	2
ACQUIN	$P1$	1	2	1.5	2400	0.9796	19	19
BERFUZ	$P2_1/c$	1	1	1.5	4800	0.9491	20	19
HPMTZT	$P2_1/c$	1	1	1.5	4800	0.9631	20	20
SUCROS01	$P2_1$	1	5	1.5	2400	0.9619	20	20
CAVGOV	$P2_1$	1	1	1.5	4800	0.9706	20	20
MECLOF10	P-1	1	3	1.5	4800	0.9510	17	17
RUWLOK	$Pna2_1$	1	1	1.5	4800	0.9763	20	20
AZHPXA	$C2/c$	1	1	1.5	4800	0.9664	20	20
KAMCEG	$Pbca$	1	1	1.5	4800	0.9730	20	20
BEVQAV	Pbcn	1	2	1.5	4800	0.9797	20	20

*See Tests applied to calculated powder diffraction data (B)

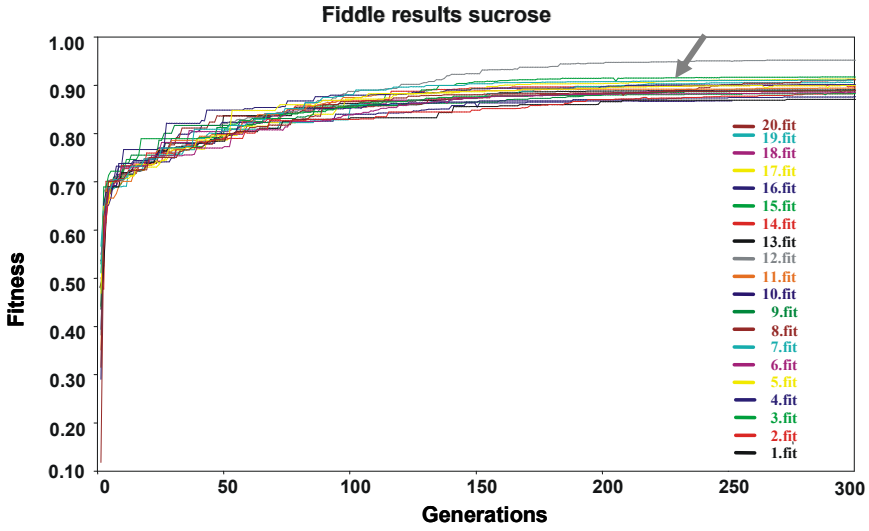


Figure 3. The different fitness values *vs.* the number of generations for sucrose. The overlay between the solution with the highest fitness value and the known crystal structure in the CSD (green) is presented in Fig. 4 as well as a comparison between its corresponding pattern and the theoretical pattern for sucrose.

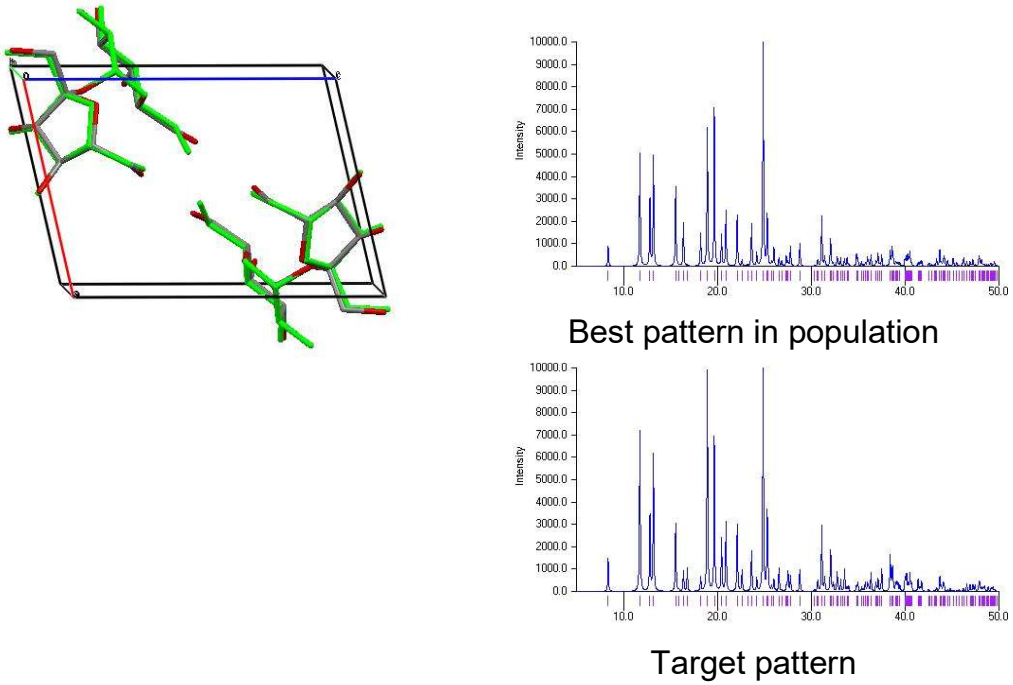


Figure 4. The overlay between the solution with the highest fitness value obtained from *FIDDLE* and the known crystal structure in the CSD (green) for sucrose as well as a comparison between the best pattern in the population and the theoretical pattern for sucrose

To illustrate that in general the correct crystal structure is the one having the highest fitness value and that the correct unit cell parameters are more often found than a complete solution, a simple test on sucrose data was performed. Sucrose is a compound crystallizing in the monoclinic space group $P2_1$, with unit cell parameters: a=7.7235(5) Å, b=8.6786(7) Å, c=10.824(1) Å, □=102.982(3)°, V= 706.98 Å^3, Z'=1 and having five torsion angles. For this compound 20 optimizations were performed. The graph showing the different fitness values *vs.* the number of generations for sucrose is presented in Fig.3. The simple test for sucrose showed that the correct unit cell parameters are found for all 20 optimizations and that the solution with the highest fitness value corresponds to the correct crystal structure.

TESTS APPLIED TO CALCULATED POWDER DIFFRACTION DATA (B)

Another experiment was carried out on the sucrose data in order to check the influence of the value of the width of the weighting function. For this test the space group and Z' were again fixed. 2400 Seeds and a maximum 2θ value of 40° were used. Fig. 5 shows the width *vs*. the number of correct solutions found in the top 80. There is a tendency to a smaller number of correct solutions for smaller values of the width, which indicated again that going to narrow weighting functions, with R_{wp} as an extreme case, reduces the performance of the genetic algorithm. On the other hand there is a wide range for the value of the width for which good results are obtained. Experience showed that a width of 0.7 Å is a good compromise between calculation speed and efficiency.

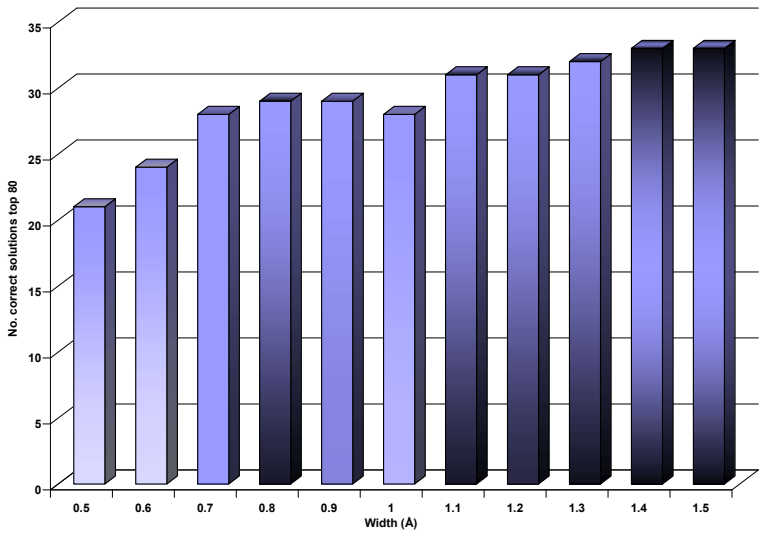


Figure 5. The number of correct solutions *vs.* the width

Searching through all ten space groups implemented in *FIDDLE* while varying Z' from 1 to 3 and using 1.5 Å for the width, the correct crystal structure of sucrose was again found, the best solution having a fitness value of 0.9690 (see Fig. 6). This experiment mimics the actual situation in a structure solution from powder diffraction data where in principle space group and Z' are unknown.

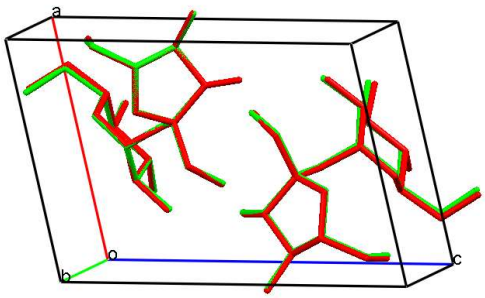


Figure 6. The overlay between the solution with the highest fitness value obtained from *FIDDLE* (red), obtained after varying the space group and Z', and the known crystal structure in the CSD (green) for sucrose

TEST APPLIED TO EXPERIMENTAL DATA

When re-determining the crystal structure of morphine anhydrate, the effectiveness of the program was proved once again. For this test the space group and Z' were set to their correct value. The number of seeds used was 9600. A width of 0.7 Å and a maximum 2θ of 50° was used. The first 8 solutions having the highest fitness value returned the correct crystal structure, although the correct unit cell parameters were found within the first 37 solutions. The quality of the best solution returned by *FIDDLE* could be used directly for Rietveld refinement. The full Rietveld refinement proceeded smoothly to reach a minimum characterized by a good fit to the diffraction profiles presented in Fig. 7 (R_{wp}=4.31). Morphine anhydrate crystallizes in the orthorhombic space group $P2_12_12_1$ with unit cell parameters: a=13.8616(5) Å, b=12.7775(4) Å, c=7.6903(3) Å, V= 1362.07 Å^3. A comparison between the best solution obtained from *FIDDLE* and the final structure obtained after Rietveld refinement is presented in Fig. 7.

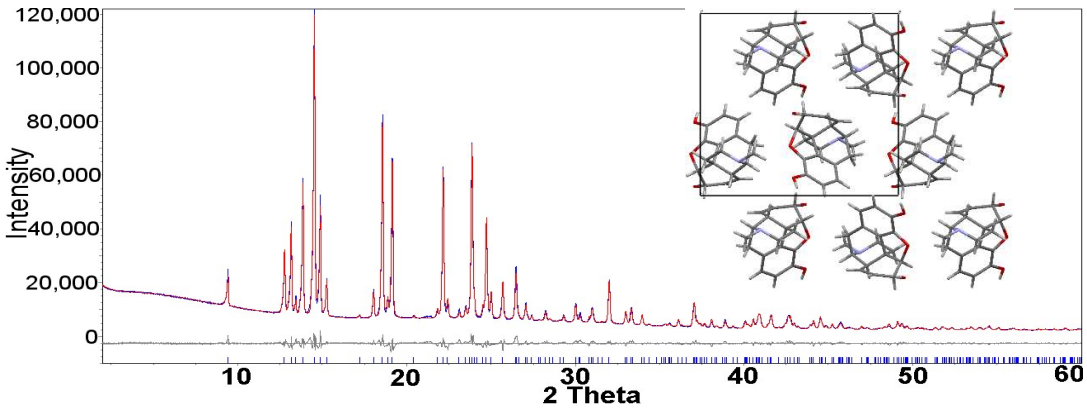


Figure 7. The fit after the final Rietveld refinement for morphine anhydrate

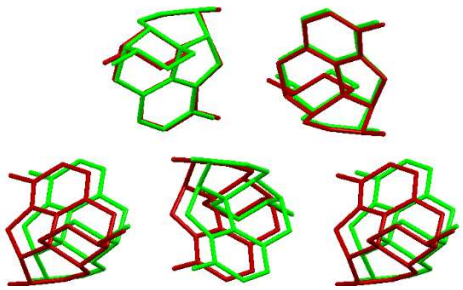
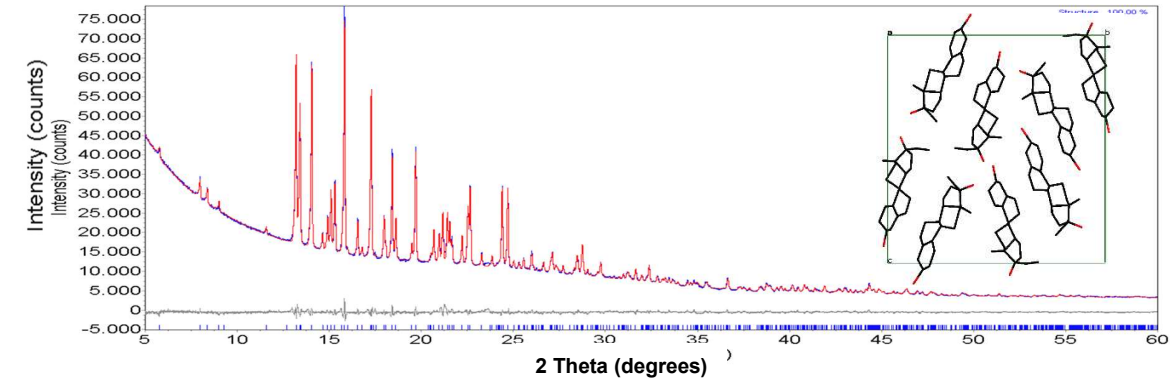


Figure 8. Comparison between the best solution obtained from *FIDDLE* (red) and the final structure obtained after Rietveld refinement (green) for morphine anhydrate

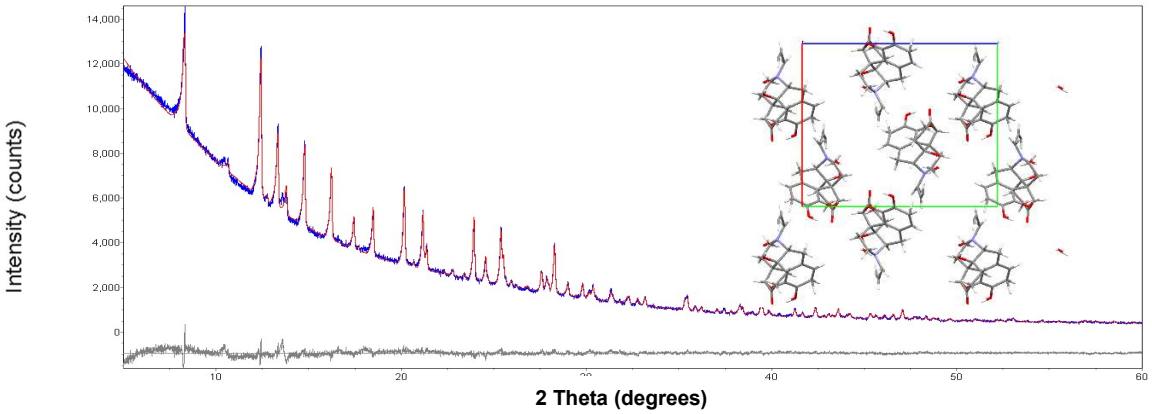
Of particular interest were the crystal structures of ethinyl estradiol anhydrate, naloxone monohydrate and creatine anhydrate. After collecting the X-ray powder diffraction data for the three compounds, indexing using DICVOL91, ITO and TREOR failed. The crystal structures of ethinyl estradiol, naloxone monohydrate and creatine anhydrate could successfully be determined using *FIDDLE*. Prior information regarding the number of molecules per asymmetric unit was obtained from solid-state NMR and used as input in *FIDDLE*, in order to reduce the computing time. The input molecules were obtained from related structures in the CSD. Searching through all ten space groups implemented in *FIDDLE*, using the information about Z' obtained from solid-state NMR, and using a maximum 2θ of 40° and 0.7 for the width, the final crystal structures of the three compounds were determined successfully. 4800 Seeds were used in total for all space groups. For ethinyl estradiol the first 40 solutions obtained from *FIDDLE* indicated the correct unit cell. In the case of naloxone monohydrate and creatine anhydrate the first 5, respectively 4 solutions indicated the correct unit cell. On the best *FIDDLE* solutions having the highest fitness value a full final Rietveld refinement was carried out using Topas 3. The full Rietveld refinements proceeded smoothly to reach a minimum characterized by a good fit to the diffraction profiles presented in Fig. 9 for all three cases. Comparisons between the best solution obtained from *FIDDLE* for ethinyl estradiol anhydrate, naloxone monohydrate and creatine anhydrate and the final structures obtained after Rietveld refinement are presented in Fig. 10.

Initially, the crystal structure of ethinyl estradiol was determined with the use of four programs: *IsoQuest*, *FIDDLE*, *DASH* and *TOPAS 3*. The *IsoQues*t program was used in order to get information about the unit cell parameters from isostructural compounds in the CSD. *IsoQuest* is a program, which can use as input the X-ray powder diffraction pattern and can search for isostructural compounds in the CSD (de Gelder *et al.*, 2006). As a result an isostructural compound (CSD refcode - EYHENO) was found. *IsoQuest* has an extremely low computing time and therefore is a good choice for obtaining information about the space group and Z'. Therefore the computing time of *FIDDLE* can be reduced significantly. The information about the space group obtained from *IsoQuest* together with the information obtained from solid-state NMR about Z' was used in the *FIDDLE* program in order to determine the crystal structure of ethinyl estradiol anhydrate. After running 2400 seeds and a width of 1.5 Å the unit cell of ethinyl estradiol anhydrate was obtained. Using the unit cell from *FIDDLE*, the *DASH* program was applied, followed by a full Rietveld refinement (using *TOPAS 3*). Afterwards, several tests were performed on ethinyl

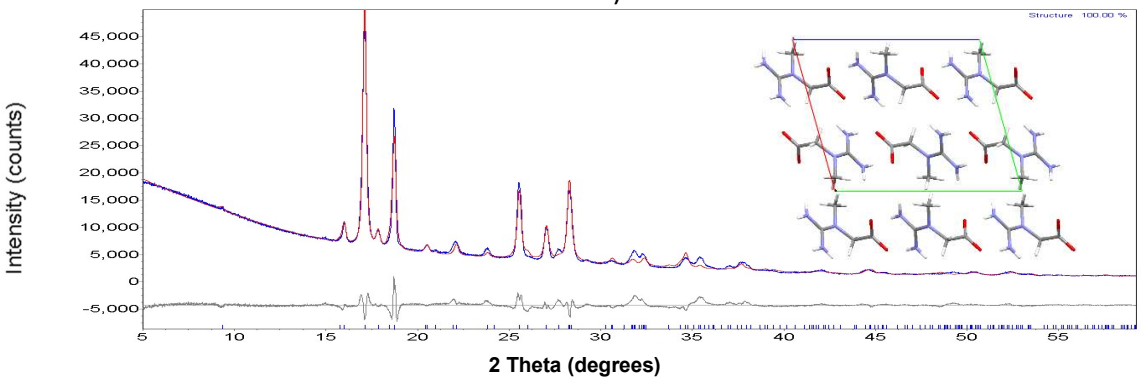
estradiol anhydrate data using the *FIDDLE* program only. It was shown that ethinyl estradiol anhydrate could also be determined with *FIDDLE* alone (R_{wp}=2.15). Therefore, when *IsoQuest* does not give any useful results, performing a full determination with *FIDDLE* always leads to a correct result although computing time is higher.



a)



b)



c)

Figure 9. The fit after the final Rietveld for: a) ethinyl estradiol anhydrate; b) naloxone monhydrate; c) creatine anhydrate

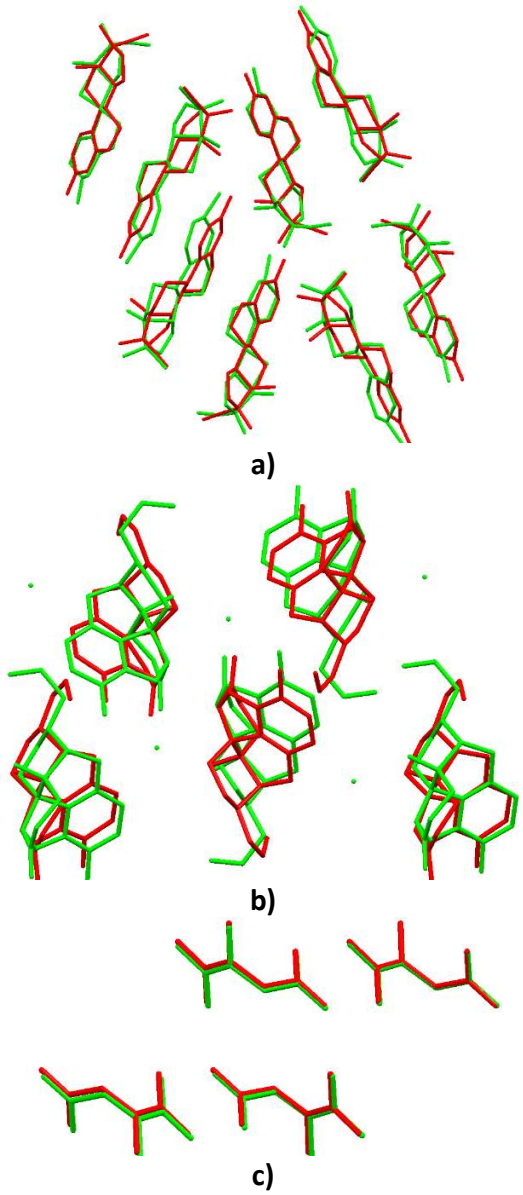


Figure 10. Comparison between the best solution obtained from *FIDDLE* (red) and the final structure obtained after Rietveld refinement (green) for: a) ethinyl estradiol anhydrate; b) naloxone monhydrate; c) creatine anhydrate

Ethinyl estradiol anhydrate crystallizes in the orthorhombic space group $P2_12_12_1$ with unit cell parameters: a=6.918(2) Å, b=21.109(5) Å, c=22.147(6) Å, V=3234.5 Å^3. Naloxone monohydrate (R_{wp}=2.901) crystallizes in the orthorhombic space group $P2_12_12_1$ with unit cell parameters: a= 13.902(7) Å, b=7.257(3) Å, c=16.640(8) Å, V= 1678.98 Å^3. Creatine anhydrate was successfully determined with *FIDDLE* and refined using Topas3 (R_{wp}=7.20). Creatine anhydrate crystallizes in the monoclinic space group $P2_1/c$ with unit cell parameters: a=9.838(8) Å, b=5.831(5) Å, c=11.679(9) Å, β=74.31(2)°, V= 645.062 Å^3. Creatine monohydrate, which is obtained after hydration of creatine anhydrate, crystallizes in the monoclinic space group $P2_1/a$ with unit cell

parameters: a=12.159(1) Å, b=5.038(2) Å, c= 12.491(2) Å, β= 108.87(1)°, V=724.038 Å^3. From a comparison between the anhydrate form of creatine and the monohydrate it is clear that dehydration takes place with an anisotropic shrinkage of the unit cell of the monohydrate (Fig. 7.11).

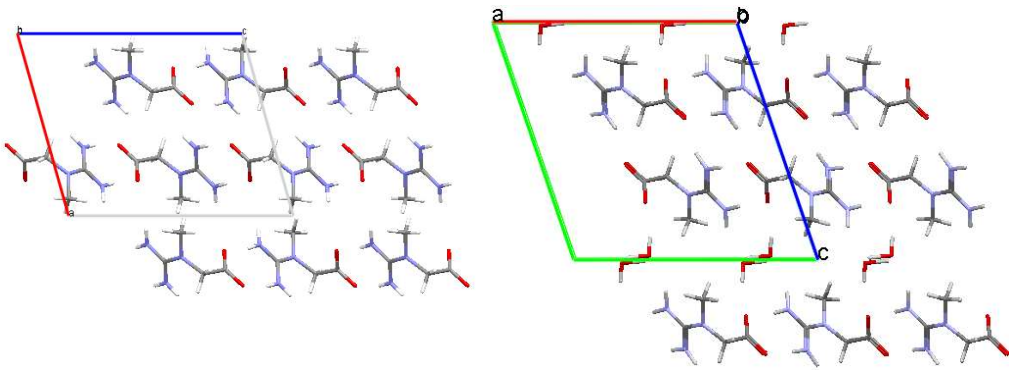


Figure 11. Comparison between the anhydrate (left) and the monohydrate (right) forms of creatine

Although the quality of the powder diffraction patterns recorded for ethinyl estradiol anhydrate and naloxone monohydrate was high enough for crystal structure determination, standard indexing proved not to be successful. After determining the crystal structure of the two forms, the *FIDDLE* results pointed at impurity peaks appearing at low 2θ angles, peaks that probably hampered the indexing process. The impurity in the case of ethinyl estradiol corresponds to the hemi-hydrate form and for naloxone monohydrate to the anhydrate form. Eventually we were able to obtain a pure form of ethinyl estradiol anhydrate and therefore the data presented here do not show any impurity corresponding to the hemi-hydrate form. This is not the case for naloxone monohydrate for which the purest form of the monohydrate still had an impurity peak corresponding to the anhydrate form. In the case of creatine anhydrate, the quality of the X-ray powder diffraction pattern was low and therefore the indexing step was impeded by severe peak overlap. Even using the knowledge obtained from *FIDDLE* about the unit cell parameters, the quality of the recorded data restricted a Pawley refinement and subsequent structure solution with other programs.

CONCLUSIONS

Simultaneous indexing and structure solution is possible using a global optimization approach. *FIDDLE* was successfully used to determine known and unknown crystal structures and can be applied to compounds for which the indexing step is impeded, for several reasons. Indexing with *FIDDLE* is computationally cheaper than complete structure determination, meaning that when a structure is not fully solved the program can still deliver the correct unit cell. Any information related to the number of molecules per asymmetric unit obtained from solid-state NMR or knowledge about the chirality of the compounds may be input into the program in order to reduce the computational time. Nevertheless, *FIDDLE* is capable of solving structures for which unit cell, space group and Z' are unknown.

ACKNOWLEDGMENT
This research was supported by the Technology Foundation STW, applied science division of NWO and the technology program of the Ministry of Economic Affairs.

NOTES
The *FIDDLE* method was first described in the doctoral thesis of Carmen Guguta that was published in **2009** (https://repository.ubn.ru.nl/handle/2066/75814). The method was presented in a lecture during the XXI IUCr Congress, Osaka, Japan, August **2008**: *FIDDLE*: A method for simultaneous indexing and structure solution from powder diffraction data. R. de Gelder, C. Guguta, J.M.M. Smits. Acta Crystallographica (**2008**), A64, C149. It was also presented during the 25th European Crystallography Meeting, Istanbul, Turkey, August 2009 at the poster session: *FIDDLE*: powder pattern indexing and structure solution hand in hand. J.M.M. Smits, C. Guguta and R. de Gelder. Acta Crystallographica (**2009**), A65, s331.

REFERENCES
Altomare A., Caliandro R., Camalli M., Cuocci C., da Silva I., Giacovazzo C., Moliterni A. G. G., Spagna R. *J. Appl. Cryst.*,**2004**, *37*, 957-966.
Altomare A., Caliandro R.,Camalli M.,Cuocci C., Giacovazzo C., Moliternia A.G.G., Rizzia R. *J. Appl. Cryst.*, **2004**, *37*, 1025-1028.
Altomare A., Giacovazzo C., Guagliardi A., Moliterni A.G.G., Rizzi R., Werner P.-E. *J. Appl. Cryst.*, **2000**, *33*, 1180-1186.
Boultif A., Louër D. *J.Appl.Cryst.*, **1991**, *24*, 987-993.
Cartwright H.M. *Applications of Artificial Intelligence in Chemistry*, **1993**, Oxford University Press.
Coelho A.A. *J. Appl. Cryst.*, **2003**, *36*, 86-95.
David W.I.F., Shankland K., McCusker L.B., Baerlocher Ch. *Structure determination from powder diffraction data*, **2002**, Oxford University Press, Oxford.
David W.I.F., Shankland K. *Acta Cryst.*, **2008**, *A64*, 52-64.
Favre-Nicolin V., Cerny R. *J. Appl. Cryst.*, **2002**, *35*, 734–743.
de Gelder R., Wehrens R., Hageman J.A. *J.Comp.Chem.*, **2001**, *22*, 273-289.
de Gelder R. *IUCr CompComm Newsletter*, **2006**, 59-69.
Goldberg D.E. *Genetic Algorithms in Search,Optimization and Machine Learning*, **1989**, MA: Addison Wesley.
Hageman J.A., Wehrens R., de Gelder R., Leo Meerts W., Buydens L.M.C. *J.Chem. Phys.*, **2000**, *113*, 7955-7962.
Harris K.D.M., Johnston R.L., Kariuki B.M. *Acta Cryst.*, **1998**, *A54*, 632-645.
Harris K.D.M., Tremayne M., Kariuki B.M. *Angew. Chem. Int. Ed.*, **2001**, *40*, 1626–1651.
Keane A. J. *Modern Heuristic Search Methods*, **1996**, 255-272, New York: Wiley.
Neumann M.A. *J. Appl. Cryst.*, **2003**, *36*, 356-365.
Padgett C.W., Arman H.D., Pennington W.T. *Cryst.Growth Des.*, **2004**, *2*, 367-372.
Tremayne M. *Phil. Trans. R. Soc. Lond. A.*, **2004**, 362, 2691–2707.
Visser J. W. *J. Appl. Cryst.*, **1969**, *2*, 89-95.
Werner P.-E., Eriksson L., Westdahl M. *J. Appl. Cryst.*, **1985**, *18*, 367-370.