An efficient treatment of ring conformations during molecular crystal structure determination from powder diffraction data

ARTICLE

An efficient treatment of ring conformations during molecular crystal structure determination from powder diffraction data

Received 00th January 20xx,
Accepted 00th January 20xx

DOI: 10.1039/x0xx00000x

Mark J. Spillman,[*a] Norman Shankland [b] and Kenneth Shankland [c]

An effective and efficient method for dealing with ring systems whose conformations are not known in advance, during global optimisation-based crystal structure determination from powder diffraction data, is described. The method can also be used to deal with compounds containing multiple centres of unknown chirality.

Introduction

Global optimisation (GO) methods for crystal structure determination from powder diffraction data (SDPD) are widely accepted as providing efficient routes to obtaining the crystal structures of small molecules. The use of such techniques is becoming increasingly necessary due to the popularity of synthetic methods based on mechanochemistry, where obtaining crystals suitable for single crystal diffraction analysis is extremely difficult[1]. Despite their many successes, GO methods become less effective as the crystal structures under study become more complex, reflecting the increasing complexity of the hypersurface that describes the agreement of the observed and calculated diffraction data. Significant effort has gone into counteracting this drop-off in effectiveness, with the result that crystal structures with large numbers of degrees of freedom (DoF) are now tractable on reasonable compute timescales[2-4]. In this work, this increased effectiveness is utilised to solve crystal structures that involve ring systems of unknown conformation and structures with multiple stereocentres.

Background to the approach

In general, when solving a crystal structure by a GO-based approach, one seeks to minimise the number of DoF in the problem in order to maximise the chances of success. For many features of the 3D molecular model(s) that comprise the start point for the optimiser, this is straightforward; bond lengths and angles are normally held fixed, as are known torsion angles, with only torsion angles that are free to rotate allowed to vary during optimisation. Other commonly encountered features, however, are more problematic.

Ring systems

The popular Z-matrix molecular representation does not allow for the optimisation of ring conformations. As such, for problems where ring conformations are not known *a priori*, the use of conformer generators is common, though not always efficient. Conformer generators[5-8] may provide a plurality of likely ring conformations, each of which needs to be tested in a separate GO run, and historically they have had difficulty in producing accurate conformations for macrocyclic rings and complex fused ring systems[6, 9]. This has led some to adopt strategies which do not require conformations to be defined in advance. The program *FOX*[10] has shown success by replacing the Z-matrix representation of molecules with a system of distance, angle and torsion restraints that allow molecular geometry to be conserved whilst optimising the positions of each individual atom in the structure. This increases the DoF, but allows the ring conformation to adjust. Another strategy available in *FOX* is the use of molecular dynamics routines to modify ring conformations during SDPD. Whilst shown to be effective, this comes at the cost of approximately doubling the time taken to process the GO runs[11]. A less commonly used approach that still makes use of the Z-matrix representation is to allow optimisation of a ring conformation during the SDPD process by breaking a bond (henceforth referred to as 'cutting') within it, converting it into a flexible chain of atoms[12]. Depending upon the number of atoms in the ring, this can significantly increase the number of DoF, thus increasing the difficulty of obtaining a solution. However, as the length of the broken bond is well-defined, it can be used as the basis of a restraint which forces the distance between the two atoms involved to refine to the known bond length during optimisation[13]. This limits the set of possible conformations that can be adopted by the flexible chain of atoms to only those that reform the ring, without placing any restriction on the conformation adopted by the ring.

[a.] *Nuclear Department, HMS Sultan, Gosport PO12 3BY, UK*
[b.] *CrystallografX Ltd, 2 Stewart Street, Glasgow, Strathclyde G62 6BW, Scotland*
[c.] *School of Pharmacy, University of Reading, Reading RG6 6AD, UK*
Electronic Supplementary Information (ESI) available: details of the implantation of restraints in GALLOP, and how they differ from those of other SDPD programs, is available. See DOI: 10.1039/x0xx00000x

Stereocentres

The relative stereochemistry of structures containing *n* (> 1) chiral centres can obtained by building separate models for each PXRD-distinguishable‡ combination of *R* and *S* and attempting to solve the structure with each model. The maximum number of models to be tested is thus 2^{n-1}, significantly increasing the work needed to obtain a solution. However, if two of the bonds at every chiral centre are cut, groups of atoms can then move independently, allowing them to be optimised to their correct relative stereochemical configurations. Whilst this process introduces several additional DoF, these additional DoF can again be restricted by using restraints based on known bond lengths.

Implementation

Distance, angle and torsion angle restraints (full details in the ESI§) have been incorporated into the local optimisation component of *GALLOP,* an open-source program for SDPD[4], in the form of penalty terms that are minimised along with the diffracted intensity χ^2. To ensure that the values of the penalty terms and χ^2 are appropriately weighted, all restraint penalties are first globally scaled by the value of χ^2 at each iteration, with individual restraints then being subject to further scaling by user-supplied weighting terms. This allows the user to set the importance of each restraint relative to χ^2. A weight of 1 assigns equal importance to χ^2 and the restraint; weights less than or greater than 1 decrease or increase the relative importance of the restraints respectively. In this work, the weights of all restraints were set to 1. No restraints are applied during the particle swarm optimisation step, which allows structures to escape the additional local minima introduced by the restraints.

Experimental

Four crystal structures, previously solved from PXRD data (Table 1), were chosen as examples for the applicability of the restraint-based approach. The published structures were first validated by periodic dispersion-corrected DFT (DFT-D) calculations, following the approach of van de Streek[14]. PXRD data for all structures, obtained from their respective publications, were Pawley fitted using *DASH*[2, 15-17] and the resultant fit files used as input for *GALLOP*. Z-matrices for all structures were derived from their CSD[18] entries§§. To produce the cut models, *Mercury*[19] was used to delete the bonds shown in Figures 1 to 4. Distance restraints applied to the cut bonds were derived from the original bond-lengths measured in *Mercury* and rounded to two decimal places (see ESI). Double bonds within the macrocycle of IJUXUI and fused rings of YIXSII were treated as flexible, though torsion angle restraints were applied to favour the expected geometry. The DoF for the original and cut models, as well as the number of restraints applied during the *GALLOP* runs are also reported in Table 1. *GALLOP* was run using cloud-based virtual machines, each equipped with a single Nvidia Tesla A100 GPU (40 GB VRAM). The virtual machines were accessed using two cloud computing providers: Google Compute Engine and vast.ai.

Table 1 Molecular and crystallographic details of the structures used in this study

CSD REFCODE	IJUXUI	LAQSON01	YIXSII	IQISAE01
Reference	[20]	[21]	[22]	[23]
Feature	M/cycle	M/cycle	Fused	5 × S/C
Sp. Grp.	$P2_12_12_1$	$P2_12_12_1$	$P2_1/c$	$P2_12_12_1$
a / Å	4.718	15.710	13.116	19.772
b / Å	19.152	18.893	21.322	15.074
c / Å	22.942	15.036	11.756	7.6743
β / °	90.000	90.000	113.99	90.000
V / Å^3	2073	4463	3004	2287
Z′	1	1	2	1
λ / Å	1.54056	1.30000	1.2525	1.54056
$2\theta_{max}$	41.99	37.96	41.50	45.00
N_{ref}	157	368	570	195
Res. / Å	2.15	2.99	1.77	2.01
DoF_{pos} : orig/cut	3 / 3	12 / 12	6 / 6	3 / 15
DoF_{ori} : orig/cut	3 / 3	3 / 3	6 / 6	3 / 12
DoF_{tor} : orig/cut	2 / 14	8 / 29	0 / 24	11 / 12
DoF_{tot} : orig/cut	8 / 20	23 / 44	12 / 36	17 / 39
Dist. restraints	1	3	6	6
Tors. restraints	2	0	2	0
Figure	1	2	3	4

M/cycle = macrocycle; Fused = fused ring system; S/C = stereocentres; N_{ref} = number of reflections extracted in Pawley fit; Res = resolution of Pawley fit; DoF_{xxx}= number of degrees of freedom (where xxx indicates: pos = positional, ori = orientational, tor = torsional, tot = total) in each problem, for the original models based on the CSD REFCODES, and the models where bonds have been cut to open up rings or deconstruct stereocentres.

For each structure, runs were carried out using the original models, cut models without restraints, and cut models with restraints applied. Each set of runs was repeated with 10 and 20 *GALLOP* iterations. *GALLOP* was set to use 500 local optimisation steps followed by a single particle swarm step per iteration. The size of each independent swarm was set to 1000 particles; hence either 5×10^6 or 1×10^7 χ^2 function evaluations were carried out by each independent swarm. For speed and GPU-memory efficiency, only the coordinates of the non-hydrogen atoms were used in the χ^2 calculations. Due to differences in space group, number of reflections, degrees of freedom and number of non-H atoms, the number of swarms (*N*) that could be accommodated in the GPU memory varied with each structure. To ensure that results were obtained from at least 100 independent swarms, repeat runs were required for most of the structures: IJUXUI – 1 run of *N*=100 swarms/GPU; LAQSON01 – 5 runs of *N*=20 swarms/GPU; YIXSII – 3 runs of *N*=34 swarms/GPU; IQISAE01 – 4 runs of *N*=25 swarms/GPU.

The best structure obtained by each independent swarm was compared to the published crystal structure using the 'Crystal Packing Similarity' tool in *Mercury*, run with the default distance and torsion angle tolerances. Only structures with 15/15 molecules in common with the reference crystal structure were considered to be correct solutions. In the case of IJUXUI, the solution's intensity χ^2 was also used as an additional criterion for determining a solution; results that gave 15/15 overlap but

with a substantially higher χ^2 than that of the rigid-macrocycle model solutions were deemed to have not solved.

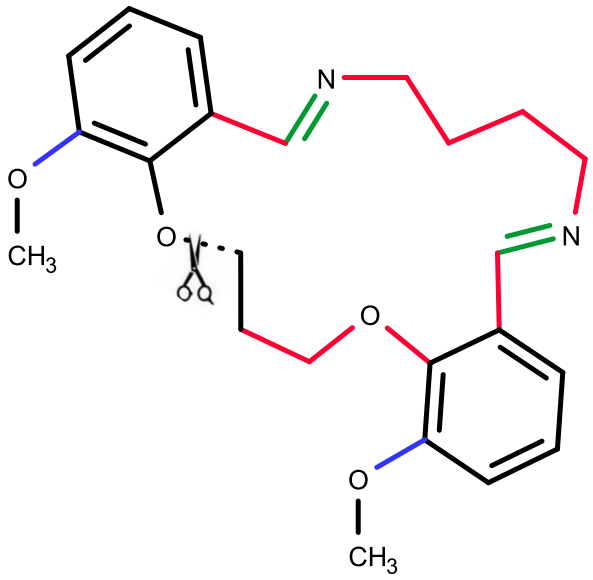

Fig 1 The molecular structure of IJUXUI (Z′ = 1), with scissors indicating the bond that is cut in order to introduce conformational flexibility to the macrocycle. Bonds coloured blue indicate those around which free rotation can occur when the macrocycle is intact; bonds coloured red indicate the additional torsional DoF introduced by bond cutting; bonds coloured green indicate those around which restrained rotation can occur after bond cutting i.e. torsional restraints are applied to these bonds to limit deviation from planarity. Overall, bond cutting increases the number of torsional DoF from 2 to 14.

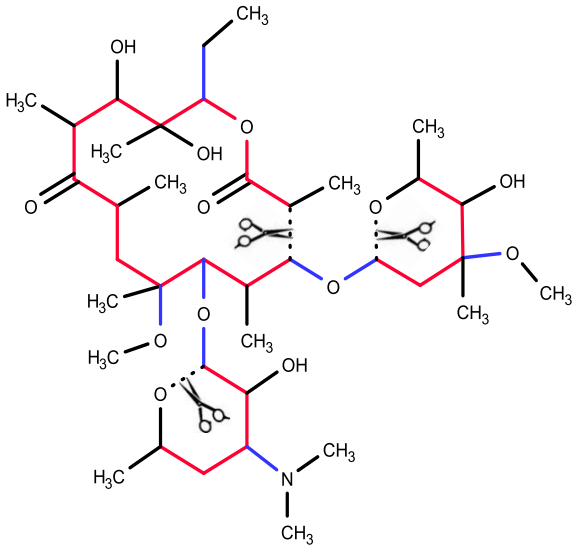

Fig 2 The structure of the largest molecular component of LAQSON01 (Z′ = 1, trihydrate), with scissors indicating the bonds that are cut in order to introduce conformational flexibility to the macrocycle and the 6-membered rings. Bonds coloured blue indicate those around which free rotation can occur when the macrocycle and rings are intact; bonds coloured red indicate the additional torsional DoF introduced by bond cutting. Overall, bond cutting increases the number of torsional DoF from 8 to 29.

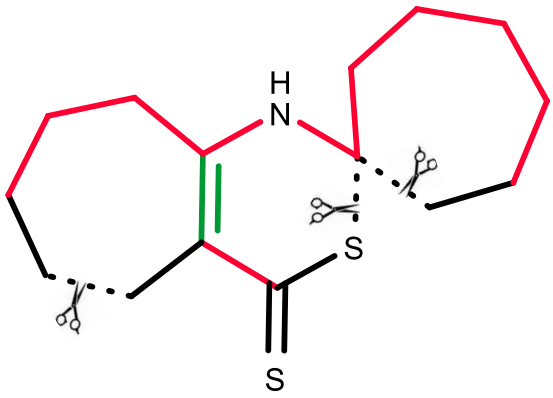

Fig 3 The molecular structure of YIXSII (Z′ = 2), with scissors indicating the bonds that are cut in order to introduce conformational flexibility to each ring. Bonds coloured red indicate the additional torsional DoF introduced by bond cutting; restrained rotation can occur around the bond coloured green after bond cutting i.e. a torsional restraint is applied to this bond to limit deviation from planarity. Overall, bond cutting increases the torsional DoF, per molecule, from 0 to 12.

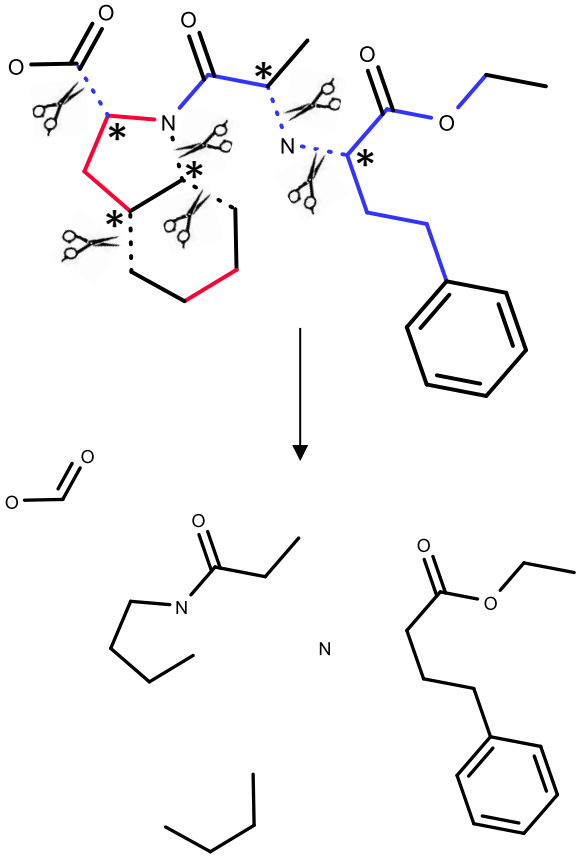

Fig 4 The molecular structure of IQISAE (Z′ = 1, H-atoms omitted for clarity), with scissors indicating the bonds that are cut in order to create the 5 fragments shown underneath. Bonds coloured blue indicate those around which free rotation can occur when the molecule is intact; bonds coloured red indicate the additional torsional DoF introduced by bond cutting. Overall, bond cutting increases the number of torsional DoF from 11 to 12 and the number of external DoF from 6 to 27. Each asterisk denotes a chiral centre in the intact molecule.

Table 2 Results obtained using 20 *GALLOP* iterations

Structure	Cut	Restr.	Succ. / %	t / min	RMSD/ Å	
					Avg.	Best
IJUXUI	✗	-	100	25	0.07	0.07
1 × *N*=100	✓	✗	48	25	0.14	0.11
	✓	✓	100	25	0.14	0.11
LAQSON01	✗	-	100	21	0.08	0.08
5 × *N*=20	✓	✗	0	21	-	-
	✓	✓	10	21	0.12	0.11
YIXSII	✗	-	86	17	0.05	0.05
3 × *N*=34	✓	✗	0	21	-	-
	✓	✓	100	21	0.14	0.11
IQISAE01	✗	-	44	16	0.12	0.12
4 × *N*=25	✓	✗	0	17	-	-
	✓	✓	17	17	0.19	0.17

Cut = did model incorporate cut bonds?; Restr. = did model incorporate restraints?; Succ. = frequency of success in achieving correct structure solution; t = time taken for a single *GALLOP* run with *N* swarms; RMSD = root mean square deviation of answer from known REFCODE structure, obtained by *Mercury* (15/15 molecules in common, 20% tolerances; H atoms ignored) rounded to 2 decimal places, *N*=number of independent swarms that were simultaneously accommodated in GPU memory.

Results

Table 2 reports the results for runs where 20 *GALLOP* iterations were performed by each independent swarm; the results for the 10 iteration runs are included in the ESI. Inclusion of restraints significantly improves the probability of obtaining a solution when using the cut models. Three of the structures (LAQSON01, YIXSII, IQISAE01) failed to solve using the cut models without restraints, whilst with restraints included, *GALLOP* rapidly returned multiple high-quality solutions with RMSDs low enough for subsequent Rietveld refinement and DFT-D optimisation to proceed without difficulty. The results obtained for runs with 10 *GALLOP* iterations per swarm show similar trends, though unsurprisingly with fewer solutions and slightly higher average RMSDs.

The simplest structure solved in this work, IJUXUI, is solved easily with the cut model even in the absence of restraints. However, the inclusion of restraints results in performance comparable to that of the original (rigid macrocycle) model in terms of success rate, as well as a slightly improved average RMSD for the final structures. The torsion angle restraints applied to the double bonds allows for some deviation from planarity (e.g. one double bond in the published structure has a torsion angle of -173°) whilst ensuring that incompatible conformations are avoided.

The cut model of LAQSON01 is the most complex structure tested in this work in terms of number of DoF. However, despite the relatively low number of χ^2 evaluations performed by each swarm, high-quality solutions were obtained with ease when restraints were included.

In the case of YIXSII, despite the cut models having three times the number of DoF of the original rigid models, more solutions were obtained from runs with cut restrained models than the runs with the original rigid models. Examination of the unsuccessful rigid model results showed that most had located a deep local minimum on the χ^2 hypersurface; this minimum being characterised by the positions of the two molecules (which have different conformations) in the asymmetric unit being swapped relative to those in the correct structure. Using the cut models, the correct conformation of each molecule can be obtained regardless of where it is located within the unit cell.

The five chiral centres in IQISAE01 result in the requirement to test up to 16 possible full models if no chirality information is known. Using instead a single cut model consisting of restrained fragments, the correct structure is obtained in about 1-in-5 runs, only a factor of ca. 2.5 times worse than using the full model with the correct stereochemical configuration.

The use of restraints adds a small computational overhead and a small increase in GPU-memory use; the latter reduces the total number of particles (and hence independent swarms) that can be simultaneously optimised. However, such trade-offs are insignificant compared to the high success rates that result from the use of restraints.

Conclusions

We have demonstrated an efficient and effective common approach to dealing with ring systems, macrocycles and multiple stereocentres during global-optimisation based SDPD that should greatly simplify solving crystal structures that possesses such features. The impact of this approach is greatest in the case of systems featuring macrocyclic rings, where obviating the need to employ other, often complex molecular modelling software packages to generate multiple likely conformers for testing, is a major advantage. Whilst the benefits for systems with multiple stereocentres are less significant, when such systems have a large number of chiral centres, the potential time savings afforded by this single cut-model approach are still considerable.

The distance, angle and torsion restraints described in this work have been included in an update to *GALLOP*, which is freely available at https://github.com/mspillman/GALLOP. *GALLOP* can be operated through a browser-based graphical user interface which includes the ability to specify restraints. Alternatively, a Python API can also be used, which allows for scripting and batch-running different combinations of restraints and weights. The ready availability of free and low-cost providers of cloud-based GPUs means that interested users do not require local GPU hardware to take advantage of this approach.

Author Contributions

MJS: Conceptualization, Data curation, Investigation, Methodology, Project administration, Software, Validation, Writing – original

draft, Writing – review & editing; NS: Data curation, Methodology, Validation, Writing – review & editing, KS: Methodology, Validation, Writing – original draft, Writing – review & editing

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

We thank the UK Materials & Molecular Modelling Hub for computational resources, which is partially funded by EPSRC (EP/P020194/1 and EP/T022213/1), for DFT-D calculations.

Notes and references

‡ Standard PXRD experiments cannot distinguish between enantiomers.
§ Other SDPD packages such as *FOX*[10, 11, 24] and *TALP* [25, 26] also allow users to define restraints. Whilst closely related, our implementation differs in a number of ways, which are discussed further in the article ESI.
§§ CSD-based models are used in this work to maintain the focus on the structure solution implementation, and to avoid introducing confounding model-building-related variables e.g. level of theory used to optimise a gas-phase model.

1. O. Al Rahal, M. Majumder, M. J. Spillman, J. van de Streek and K. Shankland, *Crystals*, 2020, **10**.
2. E. A. Kabova, J. C. Cole, O. Korb, M. Lopez-Ibanez, A. C. Williams and K. Shankland, *J. Appl. Crystallogr.*, 2017, **50**, 1411-1420.
3. E. A. Kabova, J. C. Cole, O. Korb, A. C. Williams and K. Shankland, *J. Appl. Crystallogr.*, 2017, **50**, 1421-1427.
4. M. J. Spillman and K. Shankland, *Crystengcomm*, 2021, **23**, 6481-6485.
5. R. Taylor, J. Cole, O. Korb and P. McCabe, *J Chem Inf Model*, 2014, **54**, 2500-2514.
6. S. Z. Wang, J. Witek, G. A. Landrum and S. Riniker, *J Chem Inf Model*, 2020, **60**, 2044-2058.
7. M. J. Vainio and M. S. Johnson, *J Chem Inf Model*, 2007, **47**, 2462-2474.
8. C. Scharfer, T. Schulz-Gasch, J. Hert, L. Heinzerling, B. Schulz, T. Inhester, M. Stahl and M. Rarey, *Chemmedchem*, 2013, **8**, 1690-1700.
9. J. C. Cole, O. Korb, P. McCabe, M. G. Read and R. Taylor, *J Chem Inf Model*, 2018, **58**, 615-629.
10. V. Favre-Nicolin- and R. Cerny, *Z. Kristallogr.*, 2004, **219**, 847-856.
11. R. Cerny, V. Favre-Nicolin, J. Rohlicek and M. Husak, *Crystals*, 2017, **7**.
12. A. J. Florence, in *International Tables for Crystallography*, 2019, DOI: https://doi.org/10.1107/9780955360206000959, pp. 433-441.
13. C. Cuocci, N. Corriero, F. Baldassarre, M. Dell'Aera, A. Falcicchio, R. Rizzi and A. Altomare, *J. Appl. Crystallogr.*, 2022, **55**, 411-419.
14. J. van de Streek and M. A. Neumann, *Acta Crystallogr. Sect. B: Struct. Sci.*, 2014, **70**, 1020-1032.
15. W. I. F. David, K. Shankland, J. van de Streek, E. Pidcock, W. D. S. Motherwell and J. C. Cole, *J. Appl. Crystallogr.*, 2006, **39**, 910-915.
16. A. J. Florence, N. Shankland, K. Shankland, W. I. F. David, E. Pidcock, X. L. Xu, A. Johnston, A. R. Kennedy, P. J. Cox, J. S. O. Evans, G. Steele, S. D. Cosgrove and C. S. Frampton, *J. Appl. Crystallogr.*, 2005, **38**, 249-259.
17. T. A. N. Griffin, K. Shankland, J. V. van de Streek and J. Cole, *J. Appl. Crystallogr.*, 2009, **42**, 360-361.
18. C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, *Acta Crystallogr. Sect. B: Struct. Sci.*, 2016, **72**, 171-179.
19. C. F. Macrae, I. Sovago, S. J. Cottrell, P. T. A. Galek, P. McCabe, E. Pidcock, M. Platings, G. P. Shields, J. S. Stevens, M. Towler and P. A. Wood, *J. Appl. Crystallogr.*, 2020, **53**, 226-235.
20. A. D. Khalaji, S. H. Ghoran, J. Rohlicek and M. Dusek, *J. Struct. Chem.*, 2015, **56**, 259-265.
21. J. van de Streek, *Acta Crystallogr C*, 2012, **68**, O369-+.
22. E. E. Avila, A. J. Mora, G. E. Delgado, R. R. Contreras, A. N. Fitch and M. Brunelli, *Acta Crystallogr. Sect. B: Struct. Sci.*, 2008, **64**, 217-222.
23. J. W. Reid, J. A. Kaduk and M. Vickers, *Powder Diffr.*, 2016, **31**, 205-210.
24. V. Favre-Nicolin and R. Cerny, *J. Appl. Crystallogr.*, 2002, **35**, 734-743.
25. J. Rius, O. Vallcorba, C. Frontera and C. Miravitlles, *Acta Crystallogr. Sect. A: Foundations and Advances*, 2012, **68**, S125-S125.
26. O. Vallcorba, J. Rius, C. Frontera and C. Miravitlles, *J. Appl. Crystallogr.*, 2012, **45**, 1270-1277.