JOURNAL OF APPLIED CRYSTALLOGRAPHY

ISSN: 1600-5767

journals.iucr.org/j

Improved crystal structure solution from powder diffraction data by the use of conformational information

Elena A. Kabova, Jason C. Cole, Oliver Korb, Adrian C. Williams and Kenneth Shankland

J. Appl. Cryst. (2017). **50**, 1421–1427

JOURNAL OF APPLIED CRYSTALLOGRAPHY
ISSN 1600-5767

Improved crystal structure solution from powder diffraction data by the use of conformational information

Elena A. Kabova,[a]* Jason C. Cole,[b] Oliver Korb,[b] Adrian C. Williams[c] and Kenneth Shankland[c]

[a]School of Pharmacy, University of Reading, Whiteknights Campus, Reading, Berkshire RG6 6AD, UK, [b]Cambridge Crystallographic Data Centre, 12 Union Road, Cambridge, CB2 1EZ, UK, and [c]School of Chemistry, Food and Pharmacy, University of Reading, Whiteknights, Reading, Berkshire RG6 6AP, UK. *Correspondence e-mail: e.kabova@reading.ac.uk

Received 10 July 2017
Accepted 1 September 2017

Edited by Th. Proffen, Oak Ridge National Laboratory, USA

Keywords: crystal structure determination; powder diffraction; simulated annealing; conformational information; crystal structure database.

The effect of introducing conformational information to the *DASH* implementation of crystal structure determination from powder diffraction data is investigated using 51 crystal structures, with the aim of allowing increasingly complex crystal structures to be solved more easily. The findings confirm that conformational information derived from the Cambridge Structural Database is indeed of value, considerably increasing the chances of obtaining a successful structure determination. Its routine use is therefore encouraged.

1. Introduction

Global optimization (GO)-based methods of crystal structure determination from powder diffraction data (SDPD) make explicit use of a significant amount of chemical knowledge; well characterized bond lengths and bond angles are typically held as fixed values throughout the optimization. Furthermore, certain torsion angles (such as those spanning a double bond) and certain components of molecules (for example, cyclic groups) are often treated as fixed entities. Any remaining torsion angles around which atoms are free to rotate are treated as variables to be determined by the GO procedure and are allowed to vary freely in the range of 0–360°. Thus the conformational space of a molecule under study is treated as a continuum, rather than as a sequence of isolated conformations. This work sets out to improve the performance of SDPD by implementing conformational restraints or conformational bias, derived from observed crystal structures stored in the Cambridge Structural Database (CSD; Allen, 2002; Groom *et al.*, 2016), to these freely varying torsion angles.

The utility of conformational information as constraints during crystal structure solution has long been recognized and has found particular application in macromolecular crystallography. For example, a protein molecule from a known crystal structure is often used as a start point for the crystal structure refinement of a distinct but closely related structure; see for example Scapin (2013) and DiMaio *et al.* (2011) and references therein. However, in the area of small-molecule crystallography, and in particular SDPD, conformational information has not been routinely employed, despite the fact that some work has demonstrated that it can be beneficial (CCDC, 2015; Florence *et al.*, 2005; Middleton *et al.*, 2002; Cole *et al.*, 2014). Generally though, this evidence base is not strong, consisting of a few 'one-off' demonstrations and lacking

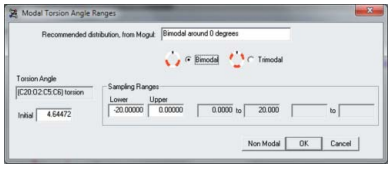

Table 1
The 51 crystal structures used in this work.

Code	Compound name	CSD refcode	Reference
A7	Zopiclone	CUHNEY10	Borea *et al.* (1987)
A16	Tolbutamide	ZZZPUS02	Donaldson *et al.* (1981)
A18	Pigment orange 36 (PO 36)	HOYVOH	van de Streek *et al.* (2009)
A19	{4′-[2-(*p*-Tosylamino)benzylideneamino]-2,3-benzo-15-crown-5}-isothiocyanato-lithium	RIFVEI	Dorokhov *et al.* (2007)
A20	Famotidine	FOGVIG03	Florence *et al.* (2003)
A21	Sotalol hydrochloride	SOTALC	Gadret *et al.* (1976)
A22	Glipizide	SAXFED	Burley (2005)
A23	Diltiazem hydrochloride	CEYHUJ01	Kojicprodic *et al.* (1984)
A24	Zopiclone dihydrate	UCUVET	Shankland *et al.* (2001)
A25	Capsaicin	FABVAF01	David *et al.* (1998)
A26	Pigment yellow (PY 181 polymorph β)	GITWUC	van de Streek *et al.* (2009)
A28	Sodium 4-[(*E*)-(4-hydroxyphenyl)diazenyl] benzene sulfonate dihydrate	YAYWUQ	Kennedy *et al.* (2001)
A30	Carbamazepine:indomethacin 1:1	LEZKEI	Majumder *et al.* (2013)
A31	2-[3-(2-Phenylethoxy)propyl sulfonyl] ethyl benzoate	BIFREO	Florence *et al.* (2005)
A32	*S*-Ibuprofen	JEKNOC10	Freer *et al.* (1993)
A33	Ampicillin trihydrate	AMPCIH01	Burley *et al.* (2006)
A34	Verapamil hydrochloride	CURHOM	Carpy *et al.* (1985)
A35	Amodiaquinium dichloride dihydrate	SENJIF	Llinàs *et al.*, 2006)
A36	Nifedipine (polymorph C)	BICCIZ01	Bortolotti *et al.* (2011)
A37	*N*-[2-(4-Hydroxy-2-oxo-2,3-dihydro-1,3-benzothiazol-7-yl)ethyl]-3-[2-(2-naphthalen-1-ylethoxy)ethylsulfonyl]propylaminium benzoate	PAHFIO	Johnston *et al.* (2004)
A39	Cyheptamide	TEVSOD01	Florence *et al.* (2008)
A40	Ornidazole	NETRUZ	Shin *et al.* (1995)
B21	Bis{4′-[2-(*p*-tosylamino)benzylideneamino]-2,3-benzo-15-crown-5-*N*,*N*′,*O*}copper(II)	RIFVAE	Dorokhov *et al.* (2007)
B27	4-(Phenyldiazenyl)naphthalen-1-amine hydrochloride	QIJCAN	Yatsenko *et al.* (2001)
B31	Telmisartan (polymorph A)	XUYHOO01	Dinnebier *et al.* (2000)
B34	Clarithromycin (polymorph I)	NAVSUY02	Noguchi *et al.* (2012)
B35	Pigment orange 62(PO 62)	HOYVUN	van de Streek *et al.* (2009)
B36	Pigment yellow (PY 151)	HOYWAU	van de Streek *et al.* (2009)
B37	Pigment yellow (PY 154 polymorph α)	HOYWEY	van de Streek *et al.* (2009)
B38	Pigment yellow 194 (PY 194)	HOYWIC	van de Streek *et al.* (2009)
B39	2,4-Dinitro-*N*-phenyl-6-(phenylazo)benzamide	IHESUJ	Chernyshev *et al.* (2002)
B40	*N*-Methyl-2,4-dinitro-*N*-phenyl-6-(phenylazo)benzamide	IHETEU	Chernyshev *et al.* (2002)
B42	Trihexyphenidyl hydrochloride	KUZDIT	Maccaroni *et al.* (2010)
B43	*N*-(2-Methoxyphenyl)-2-(2-methoxyphenylazo)-4,6-dinitrobenzamide	IHETAQ	Chernyshev *et al.* (2002)
B44	Nimustine hydrochloride	WAWZAX	Bekö *et al.* (2012)
B45	(*R*)-1-Phenylethylammonium (*R*)-2-phenylbutyrate (polymorph II)	PBUPEA01	Fernandes *et al.* (2007*a*)
B46	(*R*)-1-Phenylethylammonium (*R*)-2-phenylbutyrate (polymorph III)	PBUPEA02	Fernandes *et al.* (2007*b*)
B47	Tetracaine hydrochloride	XISVOK	Nowell *et al.* (2002)
B48	α/β-Lactose	LAKKEO	Lefebvre *et al.* (2005)
B49	*N*-(6-Phenylhexanoyl)glycyltryptophanamide	FEFNOV	Bushmarinov *et al.* (2012)
B50	Pigment yellow 183 (PY183 polymorph α)	HOMMEC01	Ivashevskaya *et al.* (2009)
B51	Pigment yellow 191 (PY191 polymorph α)	HOMMIG01	Ivashevskaya *et al.* (2009)
B52	Pigment yellow 191 (PY191 polymorph β)	HOMMOM01	Ivashevskaya *et al.* (2009)
B53	Lisinopril dihydrate	GERWUX01	Sorrenti *et al.* (2013)
B54	Prednisolone succinate	KIXDEB01	Nishibori *et al.* (2008)
B55	Cytenamide (polymorph II)	SODNOP	Florence *et al.* (2008)
B56	Carvedilol dihydrogen phosphate propan-2-ol solvate	PUJTOE	Chernyshev *et al.* (2010)
B57	Ritonavir	YIGPIO01	Bauer *et al.* (2001)
B59	D-Sorbitol	GLUCIT03	Rukiah *et al.* (2004)
B60	Chlorothiazide *N*,*N*-dimethylformamide solvate	NILSEH	Fernandes *et al.* (2007)
B61	1,2,3,-Tris(nonadecanoyl)glycerol (polymorph β)	MEZNAG	Helmholdt *et al.* (2002)

detailed quantitative assessment of any performance gains achieved.

Of particular interest is the easily accessible[1] conformational information obtainable from the nearly 900 000 crystal structures deposited in the CSD. All the tools necessary to search for, retrieve and analyse the structures from which relevant molecular geometry information can be derived are provided with the Cambridge Structural Database System (CSDS). With increasingly complex crystal structures being attempted by SDPD, it is therefore timely to re-visit the potential of exploiting conformational information in a more systematic and wide-ranging study, to provide a definitive report of its benefits.

[1] In the sense that there is no need to perform additional practical experiments, *e.g.* solid-state NMR.

2. Materials and methods

2.1. Selection and composition of powder X-ray diffraction data sets

A recent publication (Kabova *et al.*, 2017) described the optimization of the key simulated annealing (SA) parameters in *DASH* (David *et al.*, 2006) using 101 crystal structures. All those structures for which a success rate of less than 60% was

Table 2
Summary of software used in this work.

Software	Version	Application	Reference
DASH	3.3.2	Indexing† Space-group determination‡ Pawley refinement Structure solution	David *et al.* (2006)
MDASH	3.1	Structure solution	Griffin *et al.* (2009)
TOPAS	4.2	Indexing Pawley refinement Rietveld refinement	Coelho (2003)
CSD	5.36	Model building	Allen (2002)
MarvinSketch	6.0.5	Model building	ChemAxon (2011)
ConQuest	1.17	Structure mining of CSD	Bruno *et al.* (2002)
Mercury	3.3	Structure visualization	Macrae *et al.* (2008)
Mogul	1.6	Structure verification	Bruno *et al.* (2004)
Minitab	17.1.0.0	Statistical analysis	Minitab (2010)

† *Via* interface to *DICVOL91* (Boultif & Louër, 1991). ‡ With *ExtSym* as implemented in *DASH*.

Table 3
Summary of hardware used in this work.

PC	CPU	RAM	Operating system
1	Intel Core 2 Quad Q9400 (2.66 GHz)	4 GB	Windows 7 Enterprise (64 bit)
2	2 × Intel Xeon E5520 (2.270 GHz)	32 GB	Windows Server 2008 R2 Datacenter (64 bit)
3	2 × Intel Xeon E5-2630 v2 (2.60 GHz)	16 GB	Windows 7 Professional (64 bit)
4	2 × Intel Xeon E5-2630 (2.30 GHz)	16 GB	Windows 7 Enterprise (64 bit), Ubuntu 13.04 (32 bit)

obtained in that work [using the default *DASH* SA parameters of cooling rate (CR) = 0.02, $N_1 = 20$ and $N_2 = 25$] were selected for subsequent evaluation in this current work. Details of these 51 structures are given in Table 1.

2.2. Software and hardware

The software and hardware employed in this work are summarized in Tables 2 and 3.

2.2.1. *Mogul*. *Mogul* (Bruno *et al.*, 2004) is a knowledge-based library of molecular geometries derived from the CSD. It acts as a source of information on preferred molecular geometries and as such can be used to validate the geometry of a solved structure. In the case of SDPD, our main interest is in obtaining information on preferred conformations of a molecule under study, based on the torsion angle distribution information contained in *Mogul*.[2] Taking the C6—C5—O2—C20 torsion angle of verapamil hydrochloride (structure A34) as an example, the distribution of structurally closely related torsion angles (based on *circa* 11 100 CSD-deposited crystal structures) is shown in Fig. 1. The distribution clearly shows that this torsion angle is likely to adopt a value in the range

[2] A full *Mogul* geometry check of any input molecular model constructed for GO-based SDPD is also highly recommended.

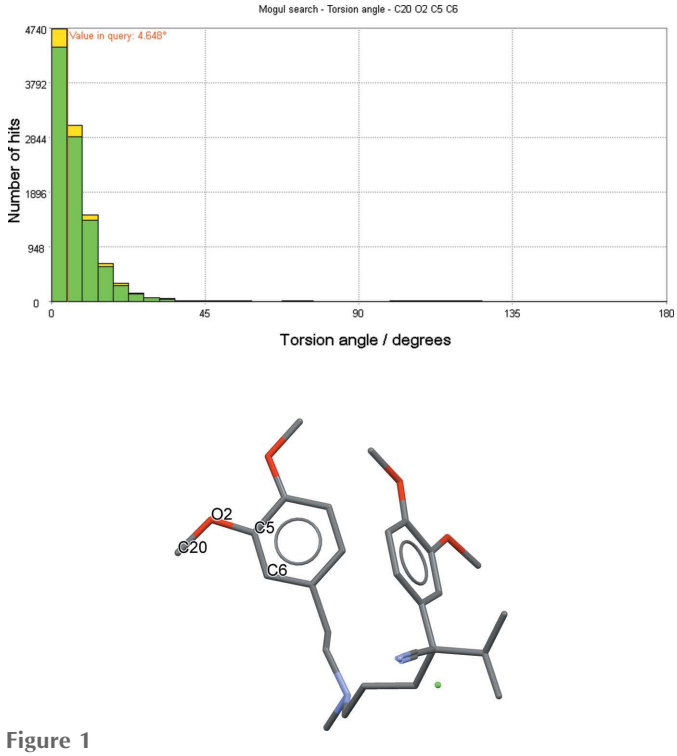

Figure 1
The *Mogul*-derived distribution (top) for the C6—C5—O2—C20 torsion angle of verapamil hydrochloride (A34, bottom).

−20° to +20°, as approximately 95% of deposited crystal structures with this structural feature fall into this range.

2.2.2. Use of *Mogul* distributions as hard constraints within *DASH*. During the GO process, any torsion angle that is free to rotate in the molecule under study can be subjected to a *Mogul* query by pressing the 'Modal' button on the 'parameter bounds' window of *DASH* (Fig. 2). From the results of the *Mogul* query, a set of discrete constraints is derived. For example, in the case of the C6—C5—O2—C20 torsion angle of verapamil hydrochloride, only values in the ranges of 0° to +20° and −20° to 0° are permitted (Fig. 3). The reduction in search space from 360° to only 40° for one torsion angle is not expected to have a notable impact on the overall success rate

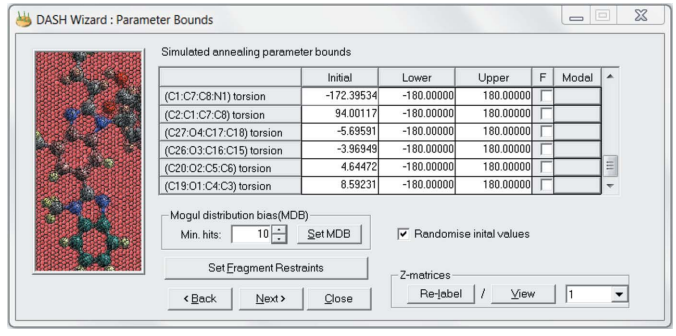

Figure 2
The 'parameter bounds' window of *DASH*. *Mogul* constraints are applied individually to each torsion angle in the form of modal distributions (see Fig. 3) using the 'Modal' option adjacent to each torsion angle description. In contrast, the MDB distributions are calculated automatically for all torsion angles when the 'Set MDB' button is pressed.

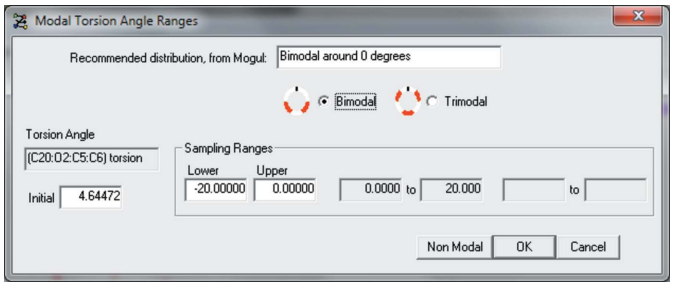

Figure 3
The modal ranges suggested for the C6—C5—O2—C20 torsion angle of verapamil hydrochloride (A34) based on the distribution shown in Fig. 1. Note that the user has the option to change the values of the ranges if desired. Note also that, for this particular torsion angle, the two ranges are adjacent and so effectively form a unimodal range spanning −20° to +20°, restricting the methyl group to be either 20° above or 20° below the plane of the benzene ring.

in solving verapamil hydrochloride, but if similar *Mogul*-derived restrictions are applied to all of the 14 variable torsion angles in the molecule, then the total search space reduction becomes more significant.

2.2.3. Use of *Mogul* distributions to bias parameter space sampling within *DASH*. *Mogul* distribution bias (MDB) is an alternative method of exploiting the *Mogul*-derived conformational information. In contrast to the *Mogul*-derived constraint approach, the MDB approach still samples the full 0–360° range for a torsion angle according to a Maxwellian distribution. However, here the Maxwellian is binned and multiplied by the value of the corresponding bin in the *Mogul* distribution, to generate a new distribution which favours moves to torsion angles in regions of space that are heavily populated in the *Mogul* distribution. MDB is invoked within *DASH* by pressing the 'Set MDB' button in the 'parameter bounds' window (Fig. 2). This automatically performs the necessary *Mogul* searches on all torsion angles that are free to rotate in the molecule under study. Considering again the C6—C5—O2—C20 torsion angle of verapamil hydrochloride, the *Mogul* search returns a string [hidden to the interactive user, but visible in the *DASH* batch file (DBF) that is generated for a batch run]: `4.6355 MDB -180 180 18 8072 2245 446 113 34 18 9 10 10 9 15 18 28 15 14 16 14 14`, where the initial torsion angle value[3] is listed first, followed by the instruction to use the MDB approach, the minimum and maximum angular values, the number of bins in the probability histogram, and finally the number of observations in each bin.

The MDB and *Mogul* hard constraints approaches introduce the same underlying information in different ways, resulting in a different exploration of χ^2 search space during the SA.

2.3. Performance analysis

An empirical log-of-the-odds (ELO) analysis was performed in order to evaluate any increase in the success rate (SR) as a result of the conformational information introduced by the use of *Mogul*.

The ELO, described by Cox & Snell (1989), can be written as

$$\mathrm{ELO} = \ln\left(\frac{\mathrm{SR}_i + 0.5}{100 - \mathrm{SR}_i + 0.5}\right). \quad (1)$$

Logistic regression was performed using *Minitab*. Full details are given by Kabova *et al.* (2017).

3. Experimental

Following the protocol established previously, 50 SA runs were initially executed on all structures, using optimized SA control parameters of CR = 0.27, N_1 = 73 and N_2 = 56 (Kabova *et al.*, 2017). Each run was set to perform 1×10^7 SA moves followed by a short simplex calculation. A χ^2 multiplier of 1 ensured the full number of SA steps was always carried out and the SA was not prematurely terminated. Five hundred SA runs of 5×10^7 moves were performed for compounds for which no successful solution was observed in the initial 100 SA runs. The *MDASH* utility (Griffin *et al.*, 2009) was used to distribute these longer runs over ten CPU cores. The option to manually alter any of the torsion angle ranges suggested by *Mogul* was not used. For consistency and to facilitate comparison of success rates, the same values of the random seeds in *DASH* were used throughout.

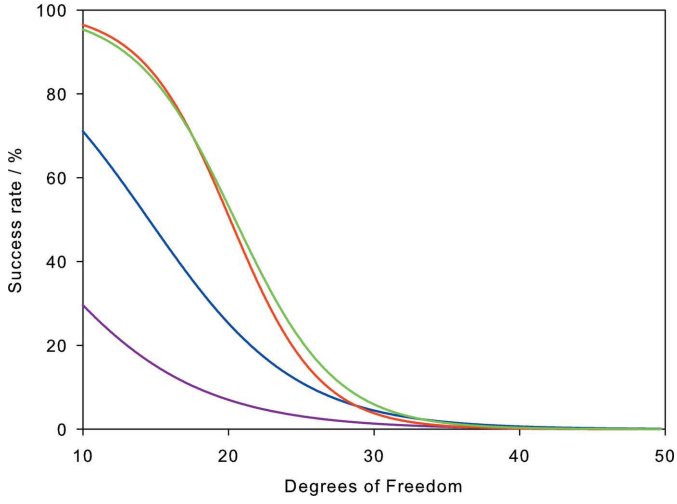

Figure 4
A comparison of the performance models based on the ELO regressions. The performance of the default *DASH* parameterization (CR = 0.02, N_1 = 20, N_2 = 25) with no conformational information is shown in purple, whilst the green line shows the performance of the best performing *DASH* parameterization (CR = 0.27, N_1 = 73, N_2 = 56) with the use of MDB. The red line shows the best performing parameterization with *Mogul* constraints, whilst the blue is the best performing with no conformational information. Note that success rates shown here cannot be directly compared with those reported by Kabova *et al.* (2017), as the data sets used in the current work are those that exhibited low success rates in the parameter tuning work reported by Kabova *et al.* (2017).

[3] This is the value of the torsion angle in the input model used by *DASH*.

Table 4
Results obtained from all *DASH* runs, expressed as percentage success rate.

DoF_{tot} = total number of degrees of freedom; DoF_{tor} = number of torsional degrees of freedom; N_{corr} = number of torsions in crystal structure lying within the constrained range; N_{incorr} = number of torsions in crystal structure lying outside the constrained range; N_{NR} = number of torsions for which no *Mogul* recommendation was made; N_{filter} = number of torsion angles filtered by *Mogul*.

							Success rate (%)			
No.	DoF_{tot}	DoF_{tor}	N_{corr}	N_{incorr}	N_{NR}	N_{filter}	0.02/ 20/25	0.27/ 73/56	*Mogul*	MDB
A7	10	4	3	0	1	0	48	78	98	100
A16	13	7	2	2	0	3	42	74	92	94
A18	14	8	7	0	0	1	4	6	24	32
A19	14	8	4	0	2	2	14	12	18	44
A20	15	9	6	0	3	0	34	88	92	88
A21	15	6	5	1	0	0	56	78	96	86
A22	16	10	8	2	0	0	28	74	94	98
A23	16	7	6	0	1	0	54	92	98	94
A24	16	4	2	1	1	0	50	84	44	48
A25	17	11	7	1	2	1	2†	24	4	6
A26	17	11	8	2	0	1	1†	10	30	2
A28	18	3	2	0	1	0	8	40	54	48
A30	18	6	4	0	1	1	34	56	56	20
A31	18	12	6	0	3	3	16	20	16	12
A32	20	8	8	0	0	0	18	54	74	70
A33	20	5	3	1	1	0	14	40	62	84
A34	22	14	12	1	1	0	4	36	60	28
A35	24	6	0	1	1	4	14	48	16	70
A36	24	12	5	5	2	0	46	72	80	98
A37	25	13	7	0	3	3	0	1†	0‡	0.4‡
A39	28	4	1	3	0	0	1†	4	2†	4
A40	30	12	12	0	0	0	0.2‡	4	0.2‡	0.4‡
B21	10	4	0	0	2	2	44	60	96	96
B27	12	3	3	0	0	0	44	78	100	100
B31	13	7	3	1	3	0	58	50	98	96
B34	14	8	6	0	2	0	50	100	100	100
B35	14	8	8	0	0	0	14	48	100	100
B36	14	8	7	0	0	1	4	12	84	44
B37	14	8	7	0	0	1	12	30	100	100
B38	14	8	7	0	0	1	36	76	100	100
B39	14	8	7	0	1	0	4	14	98	96
B40	14	8	7	0	1	0	8	26	100	96
B42	14	5	1	2	2	0	20	44	42	6
B43	16	7	4	2	1	0	12	32	92	86
B44	16	7	1	0	2	4	8	48	56	66
B45	16	4	3	1	0	0	14	54	56	46
B46	16	4	3	1	0	0	4	70	80	60
B47	18	9	5	1	3	0	14	54	94	98
B48	20	8	5	3	0	0	4†	12	4	10
B49	20	14	11	1	2	0	0‡	1†	1†	14
B50	21	6	3	1	1	1	0.2‡	0†	10	8
B51	21	6	3	1	1	1	1†	6	86	78
B52	24	6	4	0	1	1	9.4‡	18	68	56
B53	25	13	7	2	0	4	2	22	0†	46
B54	26	14	7	1	0	6	2‡	1†	4	8
B55	28	4	0	4	0	0	4	90	78	90
B56	28	10	2	3	0	5	0‡	0‡	0.8‡	0.4‡
B57	28	22	19	2	1	0	0‡	0.4‡	0‡	1†
B59	33	15	15	0	0	0	0	0.4‡	1.2‡	2‡
B60	42	6	6	0	0	0	0.4‡	1†	4.6‡	5‡
B61	49	43	37	3	3	0	0‡	0‡	0‡	0‡

4. Results

The results are summarized in Table 4, which demonstrates that considerable improvements in SR were observed when the *Mogul* and MDB approaches outlined above were utilized. The best fit lines from the ELO analysis performed on all the results obtained in this work are shown in Fig. 4.

Table 4 also includes information on how well the *Mogul*-derived distributions described the torsion angles present in the molecules of the crystal structures being solved in this work. Of the 453 variable torsion angles present:

(*a*) 309 were constrained to ranges that spanned the values observed in the reference crystal structures.

(*b*) 49 were constrained to ranges which did not span the values observed in the reference crystal structures.

(*c*) 95 could not be constrained on the basis of their *Mogul* distributions, but were still available for use by MDB (with the exception of the *Mogul*-filtered torsions).

5. Discussion

The overall benefit of using MDB or *Mogul* constraints during the SA process is most clearly shown by the ELO analysis (Fig. 4) performed on all the results obtained in this work. The shift to the right of the fitted curves, relative to the conventional *DASH* approach using optimized SA parameters, shows the increased probability of solving a crystal structure owing to the inclusion of *Mogul*-derived information. These gains, which were obtained in combination with *DASH*'s optimized SA parameters (Kabova *et al.*, 2017), were also realized when *DASH*'s default SA parameters were used [results not shown here; see Kabova (2016) for full details].

Interestingly, the use of torsion angle constraints which did not span the torsion angle values seen in the final crystal structure did not necessarily preclude obtaining a good solution; the simplex minimization employed at the end of the SA ignores the *Mogul* constraints and gives the possibility that the correct torsion angle values can be recovered. For example, with structure A36, five of the 12 torsion angle constraints did not span the crystal structure values, but the mean absolute angle difference was only 12.5°, and an increase in the SR with both constraints and MDB was still seen. Unsurprisingly, the MDB approach deals better with such cases than the constraint-based approach; an MDB distribution does not explicitly preclude a parameter taking an unlikely value during the SA itself, it merely makes it less likely that it is sampled.

Because of the chemical diversity of structures in the CSD, and the numerous factors that influence their packing into a crystal structure, it is inevitable that novel crystal structures will possess torsion angles that are not well represented in the CSD. Even though there are nearly 900 000 crystal structures in the CSD,[4] some torsion angles are only found in a very small number of structures and so either their MDB influence on the SA is minimal or no valid *Mogul*-derived constraints can be derived. Table 4 shows, for each structure studied, the number of torsion angles flagged as 'no recommendation' by *Mogul*, when either the torsion angle of interest is poorly represented in the CSD or the torsion angle distribution is nearly uniform.

[4] February 2017 updated CSD v5.38.

Such cases (around 10% of the total torsion angles of this work) are treated as fully flexible by *DASH*. In the case of B57 [ritonavir form II, 28 degrees of freedom (DoF)], the complex molecule was reported to have an unexpected conformation in the crystal structure, as a result of a strong hydrogen-bonding network (Bauer *et al.*, 2001). Whilst the use of conformational information did not allow *DASH* to solve the structure with the use of its default SA settings, a solution was obtained with the optimized SA parameters using both MBD and *Mogul*-derived constraints, although the latter required the use of 500 SA runs.

When using *Mogul*-derived torsion angle information, it is important to consider the way in which torsion angles are defined in the input model *Z*-matrices.[5] If, for example, the input *Z*-matrix contains a torsion angle that is defined using at least one hydrogen atom, then no *Mogul* distribution is generated and potential information is lost.[6] Taking B56 (a structure for which no solution was obtained in the absence of torsion angle information) as an example, five of the ten torsion angles in the *Z*-matrix (as automatically generated by *DASH*) are described with the use of a hydrogen atom and as such are ineligible for inclusion in the *Mogul*/MDB distributions. This represents a considerable loss of information, but despite this, two correct solutions were found using MDB with 500 SA runs. A simple workaround for this hydrogen-related issue, when using a CIF input model, is to manually re-order atoms in the CIF, such that all the hydrogen atoms appear at the end of the atom list. Future releases of *DASH* may address this issue by changing its *Z*-matrix generating code, or by giving the option to include filtered results in a distribution.

The reduction in the SR observed for a small number of compounds (*e.g.* A37, A40 and B60) must be addressed. In the cases of A40 and B60, the DoF in the problem are largely positional and orientational (18 out of 30 for A40, 36 out of 42 for B60), and as such they are not so heavily influenced by the introduction of conformational information. Interestingly, even when the correct conformations of the three independent molecules of A40 are used as input, and held fixed throughout the SA, *DASH* fails (with the default settings) to solve the structure within 50 SA runs, indicating the extent of the positional/orientational challenge for this structure.

6. Conclusions

This work represents a comprehensive study of the effects of including conformational information, derived from the CSD, on SDPD using the *DASH* program. The results provide strong evidence that such information should be routinely employed when faced with complex structures; the necessary tools are already in place (the fully automated MDB option is particularly convenient) and there is no significant computational overhead involved in its use. It is likely that other GO-based approaches to SDPD can benefit from this type of information and the tools provided in the CSDS are extremely valuable in this regard.

[5] *DASH* uses a *Z*-matrix description (Shankland, 2005) of the molecule internally, with flexible torsion angles flagged as variables for optimization by the SA. If a model is input to *DASH* as a CIF, MOL, MOL2, RES or PDB, it is automatically converted to a *Z*-matrix.

[6] The angle is 'filtered', on the basis that H-atom positions are often fixed geometrically, are 'assumed' or are otherwise unreliable

7. Availability and documentation

Details of *DASH*'s availability can be found at https://www.ccdc.cam.ac.uk/solutions/csd-materials/components/dash/.

Acknowledgements

EAK thanks the University of Reading and the Cambridge Crystallographic Data Centre (CCDC) for funding. We thank Mark Spillman and David Edgeley for their help with various computational matters pertaining to the rapid execution of *DASH*, and Wei Dong for his help in implementing MDB within *DASH*. We are also grateful to the University of Reading Chemical Analysis Facility for local powder diffraction facilities.

References

Allen, F. H. (2002). *Acta Cryst.* B**58**, 380–388.
Bauer, J., Spanton, S., Henry, R., Quick, J., Dziki, W., Porter, W. & Morris, J. (2001). *Pharm. Res.* **18**, 859–866.
Bekö, S. L., Urmann, D., Lakatos, A., Glaubitz, C. & Schmidt, M. U. (2012). *Acta Cryst.* C**68**, o144–o148.
Borea, P. A., Gilli, G., Bertolasi, V. & Ferretti, V. (1987). *Mol. Pharmacol.* **31**, 334–344.
Bortolotti, M., Lonardelli, I. & Pepponi, G. (2011). *Acta Cryst.* B**67**, 357–364.
Boultif, A. & Louër, D. (1991). *J. Appl. Cryst.* **24**, 987–993.
Bruno, I. J., Cole, J. C., Edgington, P. R., Kessler, M., Macrae, C. F., McCabe, P., Pearson, J. & Taylor, R. (2002). *Acta Cryst.* B**58**, 389–397.
Bruno, I. J., Cole, J. C., Kessler, M., Luo, J., Motherwell, W. D. S., Purkis, L. H., Smith, B. R., Taylor, R., Cooper, R. I., Harris, S. E. & Orpen, A. G. (2004). *J. Chem. Inf. Comput. Sci.* **44**, 2133–2144.
Burley, J. C. (2005). *Acta Cryst.* B**61**, 710–716.
Burley, J. C., Streek, J. van de & Stephens, P. W. (2006). *Acta Cryst.* E**62**, o797–o799.
Bushmarinov, I. S., Dmitrienko, A. O., Korlyukov, A. A. & Antipin, M. Yu. (2012). *J. Appl. Cryst.* **45**, 1187–1197.
Carpy, A., Léger, J.-M. & Melchiorre, C. (1985). *Acta Cryst.* C**41**, 624–627.
CCDC (2015). *Solving the Powder Pattern of Verapamil Hydrochloride*. https://www.ccdc.cam.ac.uk/support-and-resources/ccdcresources/Verapamil_DASH.pdf.
ChemAxon (2011). *Marvin 5.4.1.1*, https://www.chemaxon.com/.
Chernyshev, V. V., Kukushkin, S. Y. & Velikodny, Y. A. (2010). *Acta Cryst.* E**66**, o613.
Chernyshev, V. V., Yatsenko, A. V., Kuvshinov, A. M. & Shevelev, S. A. (2002). *J. Appl. Cryst.* **35**, 669–673.
Coelho, A. (2003). *TOPAS User Manual*. Bruker AXS GmbH, Karlsruhe, Germany.
Cole, J. C., Kabova, E. A. & Shankland, K. (2014). *Powder Diffr.* **29**, S19–S30.
Cox, D. R. & Snell, E. J. (1989). *The Analysis of Binary Data*, 2nd ed. London: Chapman and Hall.
David, W. I. F., Shankland, K., Shankland, N. & Shankland, N. (1998). *Chem. Commun.* pp. 931–932.

David, W. I. F., Shankland, K., van de Streek, J., Pidcock, E., Motherwell, W. D. S. & Cole, J. C. (2006). *J. Appl. Cryst.* **39**, 910–915.
DiMaio, F., Terwilliger, T. C., Read, R. J., Wlodawer, A., Oberdorfer, G., Wagner, U., Valkov, E., Alon, A., Fass, D., Axelrod, H. L., Das, D., Vorobiev, S. M., Iwaï, H., Pokkuluri, P. R. & Baker, D. (2011). *Nature*, **473**, 540–543.
Dinnebier, R. E., Sieger, P., Nar, H., Shankland, K. & David, W. I. F. (2000). *J. Pharm. Sci.* **89**, 1465–1479.
Donaldson, J. D., Leary, J. R., Ross, S. D., Thomas, M. J. K. & Smith, C. H. (1981). *Acta Cryst.* B**37**, 2245–2248.
Dorokhov, A. V., Chernyshov, D. Y., Burlov, A. S., Garnovskii, A. D., Ivanova, I. S., Pyatova, E. N., Tsivadze, A. Y., Aslanov, L. A. & Chernyshev, V. V. (2007). *Acta Cryst.* B**63**, 402–410.
Fernandes, P., Florence, A., Shankland, K., Karamertzanis, P. G., Hulme, A. T. & Anandamanoharan, P. (2007*a*). *Acta Cryst.* E**63**, o247–o249.
Fernandes, P., Florence, A. J., Shankland, K., Karamertzanis, P. G., Hulme, A. T. & Anandamanoharan, R. P. (2007*b*). *Acta Cryst.* E**63**, o202–o204.
Fernandes, P., Shankland, K., Florence, A. J., Shankland, N. & Johnston, A. (2007). *J. Pharm. Sci.* **96**, 1192–1202.
Florence, A. J., Baumgartner, B., Weston, C., Shankland, N., Kennedy, A. R., Shankland, K. & David, W. I. F. (2003). *J. Pharm. Sci.* **92**, 1930–1938.
Florence, A. J., Shankland, K., Gelbrich, T., Hursthouse, M. B., Shankland, N., Johnston, A., Fernandes, P. & Leech, C. K. (2008). *CrystEngComm*, **10**, 26–28.
Florence, A. J., Shankland, N., Shankland, K., David, W. I. F., Pidcock, E., Xu, X., Johnston, A., Kennedy, A. R., Cox, P. J., Evans, J. S. O., Steele, G., Cosgrove, S. D. & Frampton, C. S. (2005). *J. Appl. Cryst.* **38**, 249–259.
Freer, A. A., Bunyan, J. M., Shankland, N. & Sheen, D. B. (1993). *Acta Cryst.* C**49**, 1378–1380.
Gadret, M., Goursolle, M., Leger, J. M., Colleter, J. C. & Carpy, A. (1976). *Acta Cryst.* B**32**, 2757–2761.
Griffin, T. A. N., Shankland, K., van de Streek, J. & Cole, J. (2009). *J. Appl. Cryst.* **42**, 360–361.
Groom, C. R., Bruno, I. J., Lightfoot, M. P. & Ward, S. C. (2016). *Acta Cryst.* B**72**, 171–179.
Helmholdt, R. B., Peschar, R. & Schenk, H. (2002). *Acta Cryst.* B**58**, 134–139.
Ivashevskaya, S. N., van de Streek, J., Djanhan, J. E., Brüning, J., Alig, E., Bolte, M., Schmidt, M. U., Blaschka, P., Höffken, H. W. & Erk, P. (2009). *Acta Cryst.* B**65**, 212–222.
Johnston, A., Florence, A. J., Shankland, K., Markvardsen, A., Shankland, N., Steele, G. & Cosgrove, S. D. (2004). *Acta Cryst.* E**60**, o1751–o1753.
Kabova, E. A. (2016). PhD thesis, University of Reading, UK.
Kabova, E. A., Cole, J. C., Korb, O., López-Ibáñez, M., Williams, A. C. & Shankland, K. (2017). *J. Appl. Cryst.* **50**, https://doi.org/10.1107/S1600576717012602.
Kennedy, A. R., Hughes, M. P., Monaghan, M. L., Staunton, E., Teat, S. J. & Smith, W. E. (2001). *J. Chem. Soc. Dalton Trans.* pp. 2199–2205.
Kojicprodic, B., Ruzictoros, Z., Sunjic, V., Decorte, E. & Moimas, F. (1984). *Helv. Chim. Acta*, **67**, 916–926.
Lefebvre, J., Willart, J.-F., Caron, V., Lefort, R., Affouard, F. & Danède, F. (2005). *Acta Cryst.* B**61**, 455–463.
Llinàs, A., Fábián, L., Burley, J. C., Streek, J. van de & Goodman, J. M. (2006). *Acta Cryst.* E**62**, o4196–o4199.
Maccaroni, E., Malpezzi, L. & Masciocchi, N. (2010). *Acta Cryst.* E**66**, o2511.
Macrae, C. F., Bruno, I. J., Chisholm, J. A., Edgington, P. R., McCabe, P., Pidcock, E., Rodriguez-Monge, L., Taylor, R., van de Streek, J. & Wood, P. A. (2008). *J. Appl. Cryst.* **41**, 466–470.
Majumder, M., Buckton, G., Rawlinson-Malone, C. F., Williams, A. C., Spillman, M. J., Pidcock, E. & Shankland, K. (2013). *CrystEngComm*, **15**, 4041–4044.
Middleton, D. A., Peng, X., Saunders, D., Shankland, K., David, W. I. F. & Markvardsen, A. J. (2002). *Chem. Commun.* pp. 1976–1977.
Minitab (2010). *Minitab 17 Statistical Software*, http://www.minitab.com/en-us/.
Nishibori, E., Ogura, T., Aoyagi, S. & Sakata, M. (2008). *J. Appl. Cryst.* **41**, 292–301.
Noguchi, S., Miura, K., Fujiki, S., Iwao, Y. & Itai, S. (2012). *Acta Cryst.* C**68**, o41–o44.
Nowell, H., Attfield, J. P., Cole, J. C., Cox, P. J., Shankland, K., Maginn, S. J. & Motherwell, W. D. S. (2002). *New J. Chem.* **26**, 469–472.
Rukiah, M., Lefebvre, J., Hernandez, O., van Beek, W. & Serpelloni, M. (2004). *J. Appl. Cryst.* **37**, 766–772.
Scapin, G. (2013). *Acta Cryst.* D**69**, 2266–2275.
Shankland, K. (2005). *IUCr Commission on Crystallographic Computing Newsletter*, No. 5, pp. 92–102.
Shankland, N., David, W. I. F., Shankland, K., Kennedy, A. R., Frampton, C. S. & Florence, A. J. (2001). *Chem. Commun.* pp. 2204–2205.
Shin, H. S., Song, H., Kim, E. & Chung, K. B. (1995). *Bull. Korean Chem. Soc.* **16**, 912–915.
Sorrenti, M., Catenacci, L., Cruickshank, D. L. & Caira, M. R. (2013). *J. Pharm. Sci.* **102**, 3596–3603.
Streek, J. van de, Brüning, J., Ivashevskaya, S. N., Ermrich, M., Paulus, E. F., Bolte, M. & Schmidt, M. U. (2009). *Acta Cryst.* B**65**, 200–211.
Yatsenko, A. V., Chernyshev, V. V., Paseshnichenko, K. A. & Schenk, H. (2001). *Acta Cryst.* C**57**, 295–297.