Supplementary information: Use of crystal structure informatics for defining the conformational space needed for predicting crystal structures of pharmaceutical molecules.

Luca Iuzzolino, Anthony M. Reilly, Patrick McCabe, Sarah L. Price

Table of Contents

1. DEFINING CONFORMATIONAL FLEXIBILITY

1.1 Initial testing on smaller molecules

1.1.1 Rotamer distributions

5-Formyluracil   Olanzapine   Tazofelone

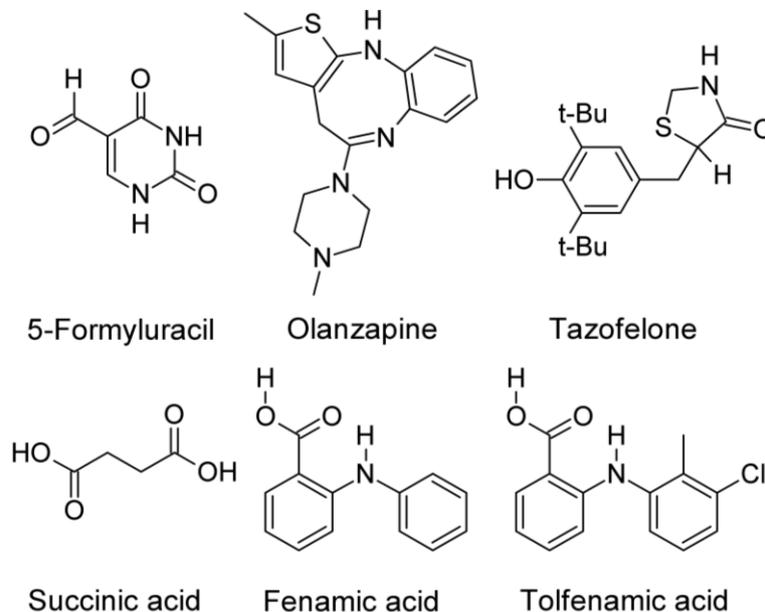Succinic acid   Fenamic acid   Tolfenamic acid

Figure 1: Chemical diagrams of the small molecules used to investigate the ability of CSD statistical information about conformational preferences to define the conformational space that could occur in crystal structures.

An initial test set of six small molecules (Figure 1), all of which had been subject to full CSP studies, was performed to illustrate some types of conformational behaviors and how CSD information on rotamer distributions could be applied in CSP. The rotamer distributions were retrieved from the CSD rotamer library, using a stand-alone program. On top of histograms, in order to assess the distributions more quantitatively, they were analyzed via kernel density estimation (KDE) with the Von Mises kernel, which produces smooth probability density functions (PDFs); the Matplotlib Python package was utilized to perform this analysis. The method proposed by McCabe *et al.* was followed, generating an estimate of the probability density function f(θ), with θ being the torsion angle ranging from 0 to 360°. $f(\theta)$ is given by the equation:

$$f(\theta) = \frac{1}{n(2\pi)I_0(v)} \sum_{i=1}^{n} exp[v\,cos(\theta - \theta_i)] \tag{1}$$

where n is the size of the sample, $I_0(c)$ is the 0th order modified Bessel function of the first kind and v is a smoothing parameter. The smoothing parameter v was once again calculated with the method proposed by McCabe *et al.*, which uses equations developed by Taylor and Fisher.

A range $35 \le \nu \le 500$ was imposed. The lower limit prevents the distribution from becoming too broad, the upper limit prevents the $I_0(\nu)$ from becoming too large making the calculation of $f(\theta)$ impossible. The lower limit is slightly larger than the one proposed by McCabe *et al.*, where a range of $10 \le \nu \le 500$ was suggested, because it guarantees a larger level of clustering around the mode and sharper peaks, which are more useful for the purpose of the PDFs in this work.

Figure 2 shows four examples of possible histograms and PDFs derived from rotamer distributions, which could be treated differently in a CSP study.
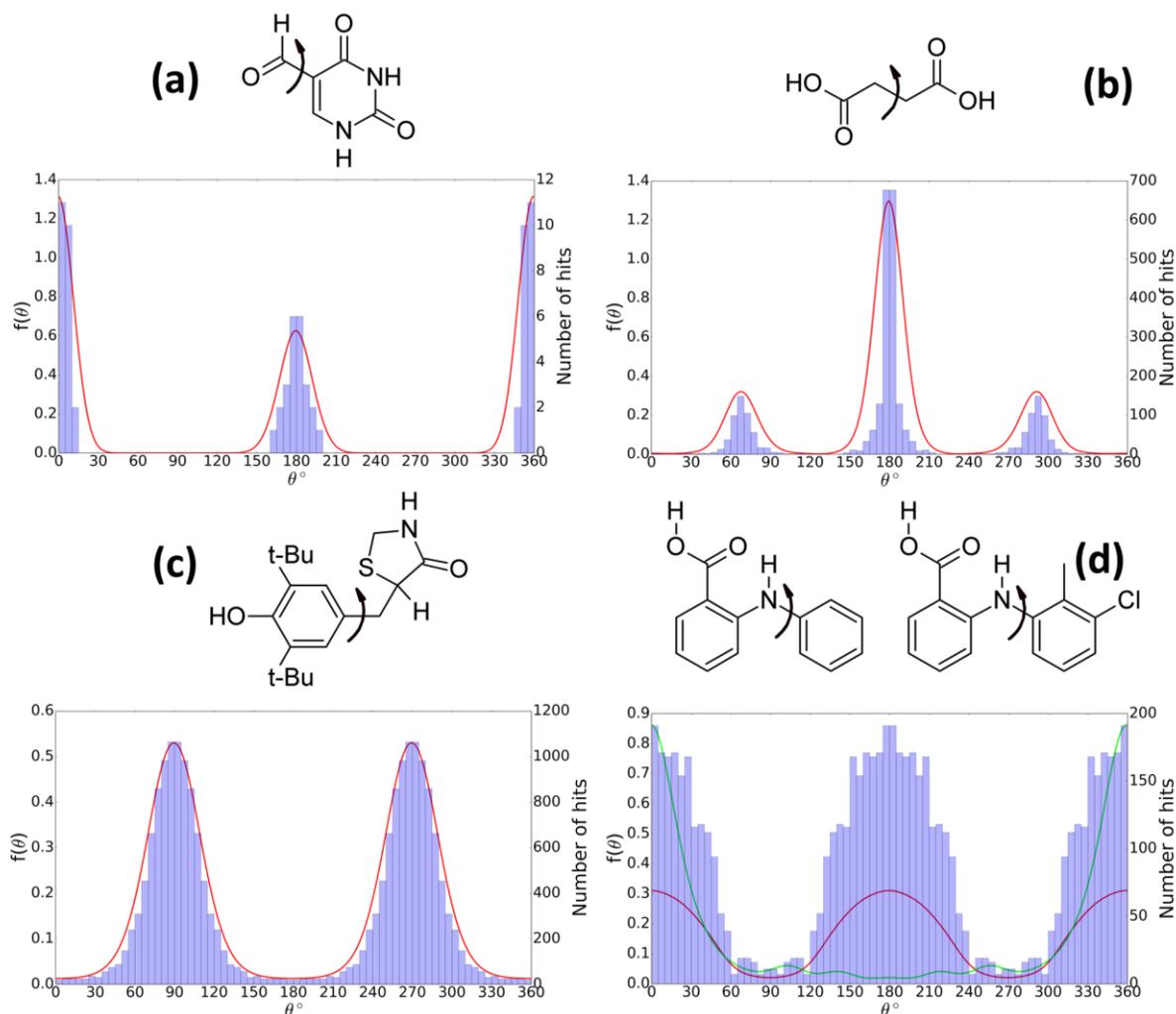
Figure 2: Histograms (light purple bars) and Von Mises kernel density approximation PDFs (red lines) for torsion angle distributions of the angles indicated on each molecular diagram of (a) 5-Formyluracil (0° in the diagram) (b) Succinic Acid (180° in the diagram) (c) Tazofelone (0° in the diagram) (d) Fenamic acid (0° in the diagram), with an overlay of the PDF for tolfenamic acid in green, showing the effect of the additional methyl and Cl substituents.

In all four cases, the insights from the rotamer distributions are similar to those derived from *ab initio* relaxed scans of the molecules, as done for the original CSP studies. For 5-formyluracil (Figure 2a), the distributions indicate that only two values are possible within very narrow ranges (i.e. 0 and 180°). The torsion angle of succinic acid (Figure 2b) can take three values within narrow ranges: 180°, which leads to a planar conformation, and two others at around 60° and 300°, generating folded conformations. In tazofelone, there is quite a wide spread of possible values around both 90 and 270° (Figure 2c), while the remainder have very low probabilities. In the final example of the fenamates, the only low-barrier torsion angle (Figure 2d) can adopt any angle in fenamic acid, but the methyl substitution in tolfenamic acid significantly reduces the probability of a crystalline conformation between 60 and 300° because of steric interactions.

1.1.2 CSD Conformer Generator

Successively the effectiveness of the CSD Conformer Generator (version 1.0), which is based on the rotamer libraries outlined above, to generate the experimental conformers of the molecules in Figure 1, was tested. The CG, with its default settings for molecular clustering and maximum number of unusual torsion angles and without any limit in number of conformations and probability scores, was used for each target molecule. The CG produced two conformations for 5-formyluracil, 37 for olanzapine, ten for tazofelone, 46 for succinic acid, 67 for fenamic acid and 83 for tolfenamic acid.

The full set of generated conformations was analyzed to verify whether it contained the conformer/s of each molecule observed in experimental crystal structures. The analysis was performed with the molecular comparison tool available in the CSD Python API 1.0.0.[45] All experimental crystalline conformers of the six small molecules were reproduced very well by a CG conformation, with $RMSD_1$ < 0.35 Å, as shown in Figure 3 and Table 1.
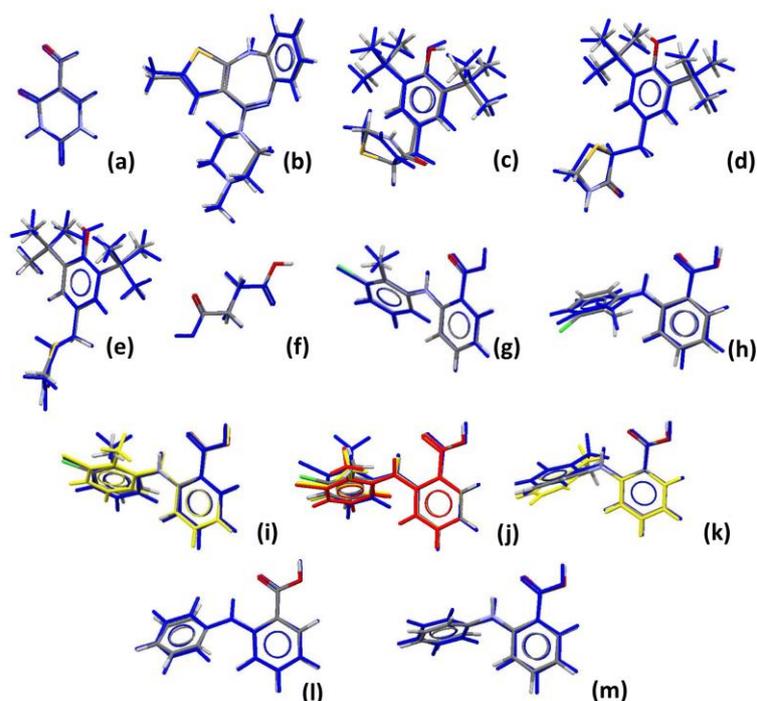
Figure 3: Overlays of the experimental conformers of the molecules in Figure 1 **with their most similar conformations in the CG set (in blue). If the same CG conformation was the closest match of each molecule in the asymmetric unit of Z' > 1 crystal structures, the extra experimental conformers are colored in red or in yellow. Polymorphs with identical conformers are only shown once.**

Table 1: Quantification of the ability of CG to reproduce the experimental conformations, as shown in Figure 3.

Label on Figure 3	Molecule	Generated conformers	Ranking of most similar conformer	RMSD(Å)
a	5-Formyluracil	2	2	0.019
b	Olanzapine forms I and II	37	1	0.151
c	Tazofelone forms I and II	10	5	0.337
d, e	Tazofelone form III (solid solution)	"	4, 1	0.260, 0.123
f	Succinic acid forms α and β	46	1	0.051
g	Tolfenamic acid form 1	83	2	0.087
h	Tolfenamic acid form 2	"	16	0.258
i	Tolfenamic acid form 3 (Z` = 2)	"	2, 2	0.219, 0.175
j	Tolfenamic acid form 4 (Z` = 3)	"	2, 2, 2	0.277, 0.314, 0.287
k	Tolfenamic acid form 5 (disordered)	"	15, 15	0.301, 0.228
l, m	Fenamic acid (Z` = 2)	67	2, 6	0.169, 0.185

This analysis reveals that the range of CG conformations covers, or even exceeds, the conformational space considered in the original search. Hence, a set of rigid searches would capture the entire flexibility ranges of the molecules. Additional analysis of the rotamer distributions (for example those in Figure 2) would make the choice of which torsion/s could be more efficiently considered as explicitly flexible in a search obvious. Nonetheless the data

5

shown in the main paper prove that this is not sufficient for an effective choice of conformational space for larger and more flexible molecules.

1.2 Preliminary analysis of the CG score
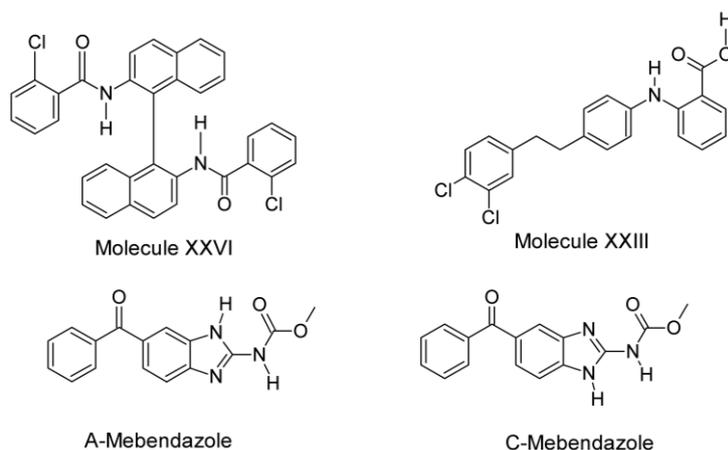
Figure 4: Diagrams of the four molecules used to perform a preliminary analysis of the CG score.

To test the quality of the CG score of conformations, the four molecules in Figure 4 were used. For each molecule, a maximum of 200 conformations having a probability score lower than 0.75 were generated with the CG. Only for molecules XXVI and XXIII were 200 conformations generated; 91 were produced for the A tautomer of mebendazole and 84 its C tautomer. All the generated conformations were optimized constraining each torsion angle and allowing the other degrees of freedom (i.e. bond angles and bond-lengths) to relax, at the PBE0 6-31G(d,p) level using Gaussian09. The results are shown in Figure 5.
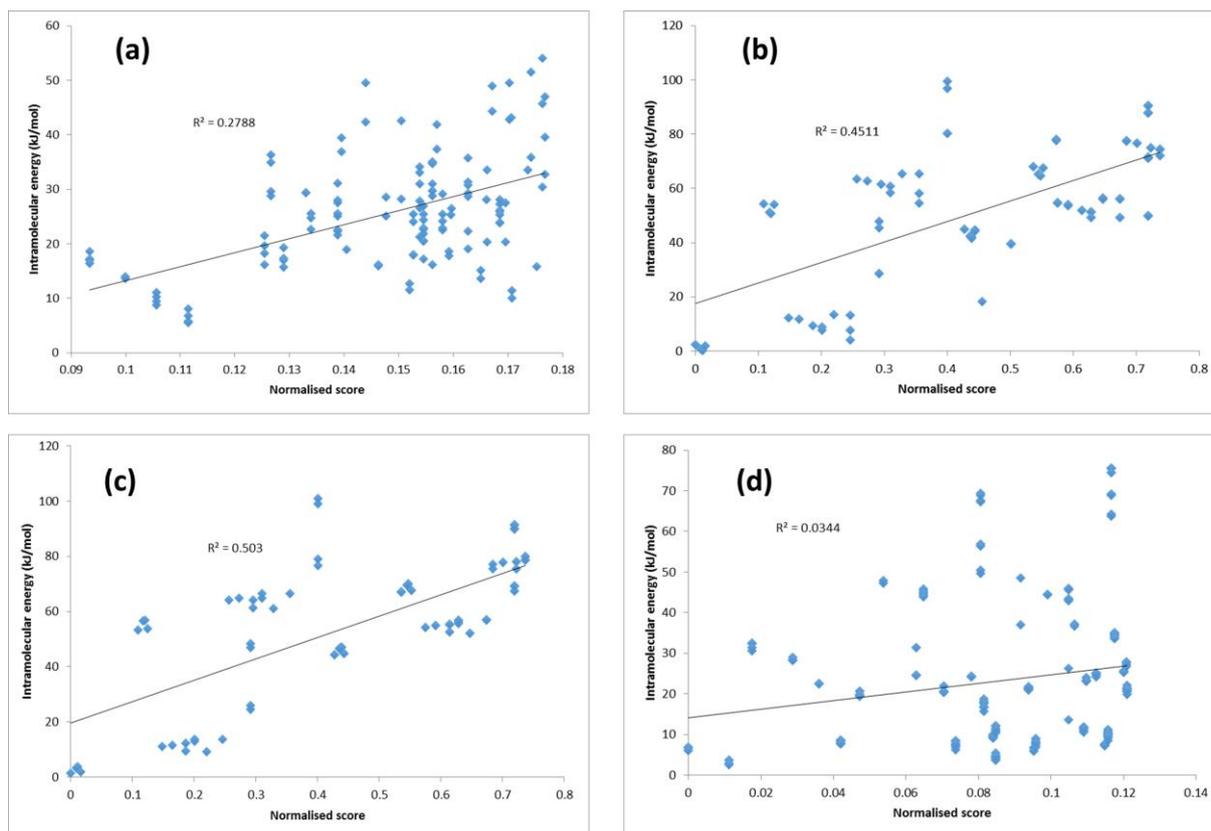
Figure 5: Plots of intramolecular energy versus CG probability scores of (a) Molecule XXVI (b) the A tautomer of Mebendazole (c) the C tautomer of Mebendazole (d) Molecule XXIII.

There seems to be a weak correlation between conformational energy and CG probability score. Nevertheless some high-probability conformations have unfeasible conformational energies for solid-state conformers. The reason is not clear, but it is probably because the CG analyzes rotamer distributions with a fragment based approach. This approach does not capture the interaction between different fragments. Hence conformations can be generated with high-probability values for the individual torsion angles, which nonetheless are high in conformational energy because the various fragments of the molecule interact in an unfavorable fashion, and vice versa.

In summary, this preliminary analysis reveals that probability alone is not suited to select important conformations, but some *ab initio* calculations are required.
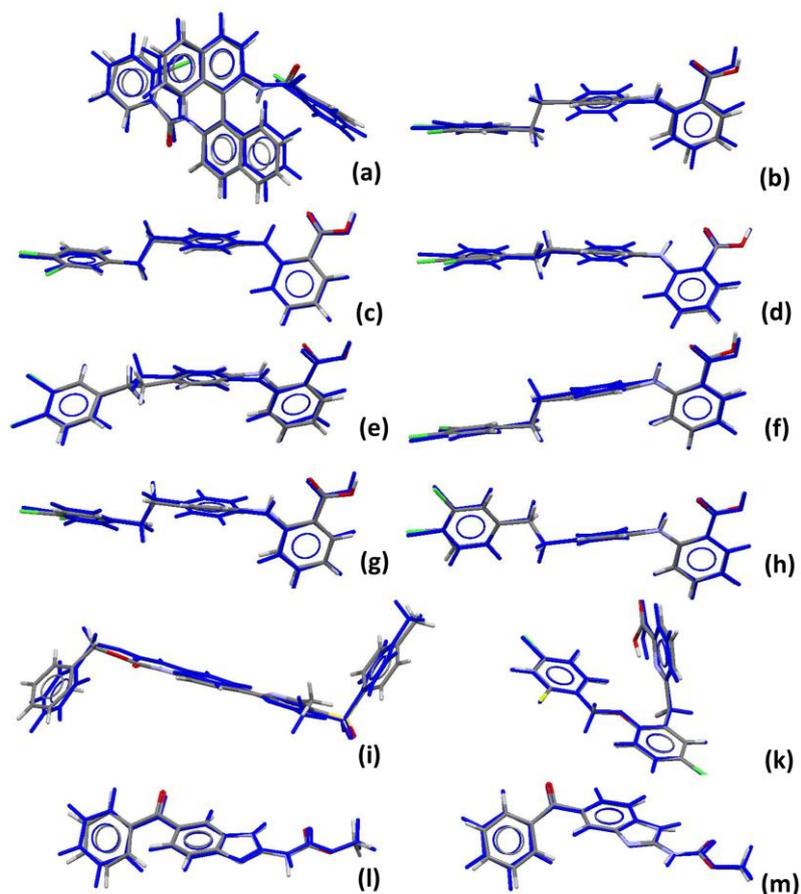
Figure 6: Overlays of the experimental conformers of the molecules in Figure 1 in the main paper (colored by elements) with their most similar conformations in the CG set (in blue).

Table 2: Quantification of the ability of CG to reproduce the experimental conformations, as shown in Figure 6.

Label on Figure 6	Molecule	Generated conformers	Ranking of most similar conformer	RMSD(Å)
a	XXVI	4947	685	0.386
b	XXIII form a	14269	440	0.683
c	XXIII form b	"	406	0.65
d, e	XXIII form c (Z` = 2)	"	657, 3160	0.66, 0.692
f	XXIII form d	"	406	0.192
g, h	XXIII form e (Z` = 2)	"	491, 411	0.235, 0.627
I	XX	17374	15	0.419
j	GSK269984B	9529	166	0.129
k	Mebendazole A	91	2	0.157
l	Mebendazole C	84	1	0.133

2. APPLICATION OF THE WORKFLOW TO THE SIX PHARMACEUTICAL-LIKE MOLECULES

Molecule XXVI

GSK269984B

Molecule XX

Molecule XXIII

A-Mebendazole

C-Mebendazole
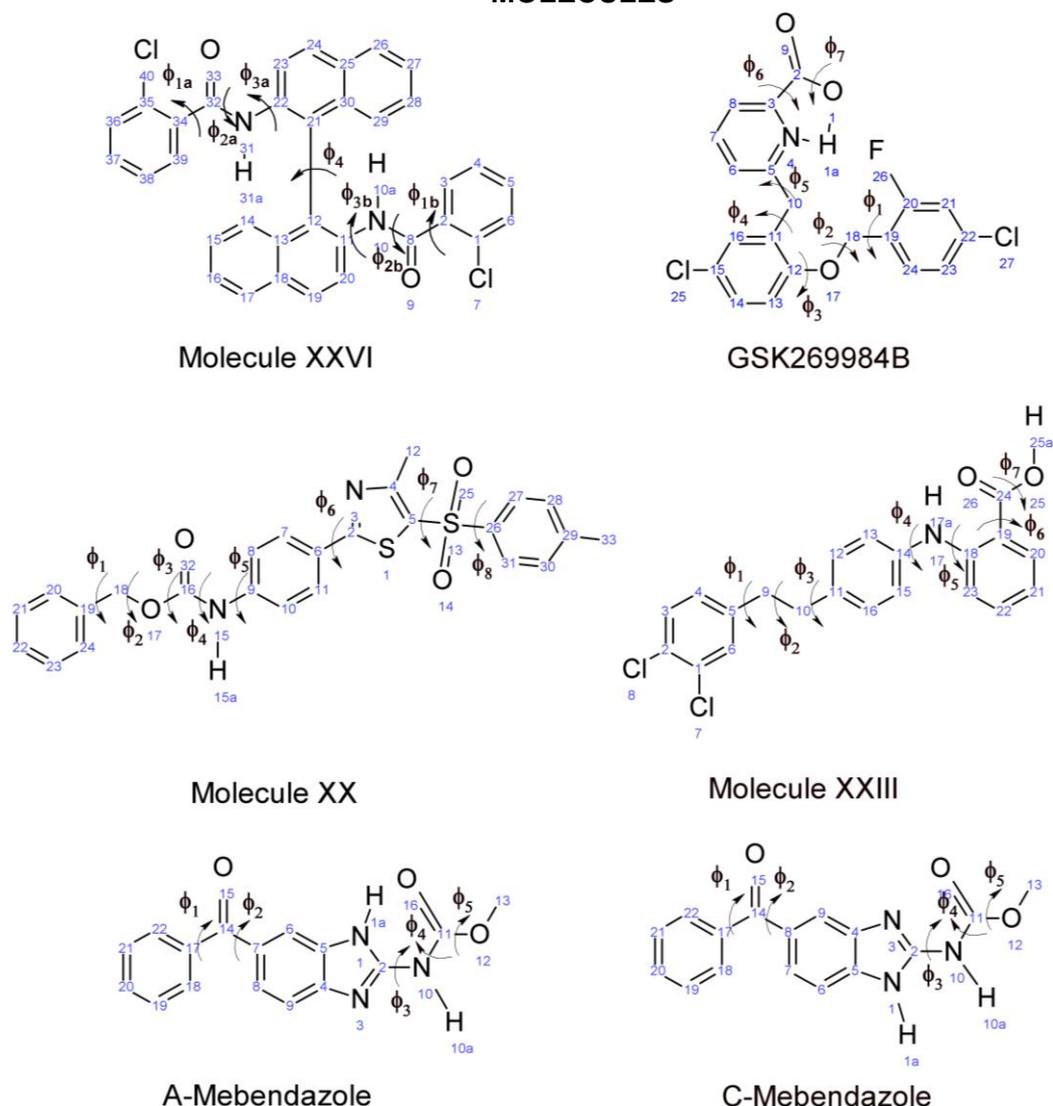
Figure 7: Chemical diagrams of the molecules on which the workflow outlined in Section 2.1 in the main paper was applied, adding the atomic numbering which precisely identifies the rotatable torsion angles in Table 3.

The PDFs of each torsion angle were calculated via Von Mises KDE and are shown in Figure 8- Figure 12. For each maximum in f(θ), the half width at half maximum (HWHM) was estimated with Python and is indicated by ± in Table 4. Maxima were only included if their $f(\theta)$ was larger than 0.2.

Torsion angles Φ_7 in GSD269984B and Φ_7 in molecule XXIII can only be defined with a hydrogen atom, and so they are not included in the rotamer libraries and the CG. Hence, they were scanned with *ab initio* methods (see section 2.5 for details).

Table 3: Atomic numbering definition of the key torsion angles shown in Figure 7.

Label for XXVI	Torsion Angle Definition
Φ_{1a}	35-34-32-33
Φ_{2a}	33-32-31-22
Φ_{3a}	32-21-22-21
Φ_{1b}	1-2-8-9
Φ_{2b}	9-8-10-11
Φ_{3b}	8-10-11-12
Φ_4	30-21-12-13
Label for GSK269984B	**Torsion Angle Definition**
Φ_1	17-18-19-20
Φ_2	12 -17-18-19
Φ_3	11-12-17-18
Φ_4	12-11-10-5
Φ_5	11-10-5-4
Φ_6	4-3-2-9
Φ_7	3-2-1-1a
Label for XX	**Torsion Angle Definition**
Φ_1	20-19-18-17
Φ_2	19-18-17-16
Φ_3	18-17-16-16
Φ_4	32-16-15-9
Φ_5	16-15-9-8
Φ_6	7-6-2-3
Φ_7	4-5-13-26
Φ_8	5-13-26-31
Label for XXIII	**Torsion Angle Definition**
Φ_1	6-5-9-10
Φ_2	5-9-10-11
Φ_3	9-10-11-12
Φ_4	13-14-17-18
Φ_5	14-17-18-19
Φ_6	18-19-26-26
Φ_7	19-24-25-25a
Label for Mebendazole A, C	**Torsion Angle Definition**
Φ_1	22-17-14-7 (A), 22-17-14-8 (C)
Φ_2	17-14-7-8 (A), 17-14-8-9 (C)
Φ_3	3-2-10-11
Φ_4	2-10-11-16
Φ_5	16-11-12-13

2.1 Torsion angle distributions of the molecules in Figure 3 in the main paper

Figure 8: Histograms (light purple bars) and Von Mises kernel density approximations (red lines) for torsion angle distributions of angles of molecule XXVI.

Figure 9: Histograms (light purple bars) and Von Mises kernel density approximations (red lines) for torsion angle distributions of angles of GSK269984B.

Figure 10: Histograms (light purple bars) and Von Mises kernel density approximations (red lines) for torsion angle distributions of angles of molecule XX.
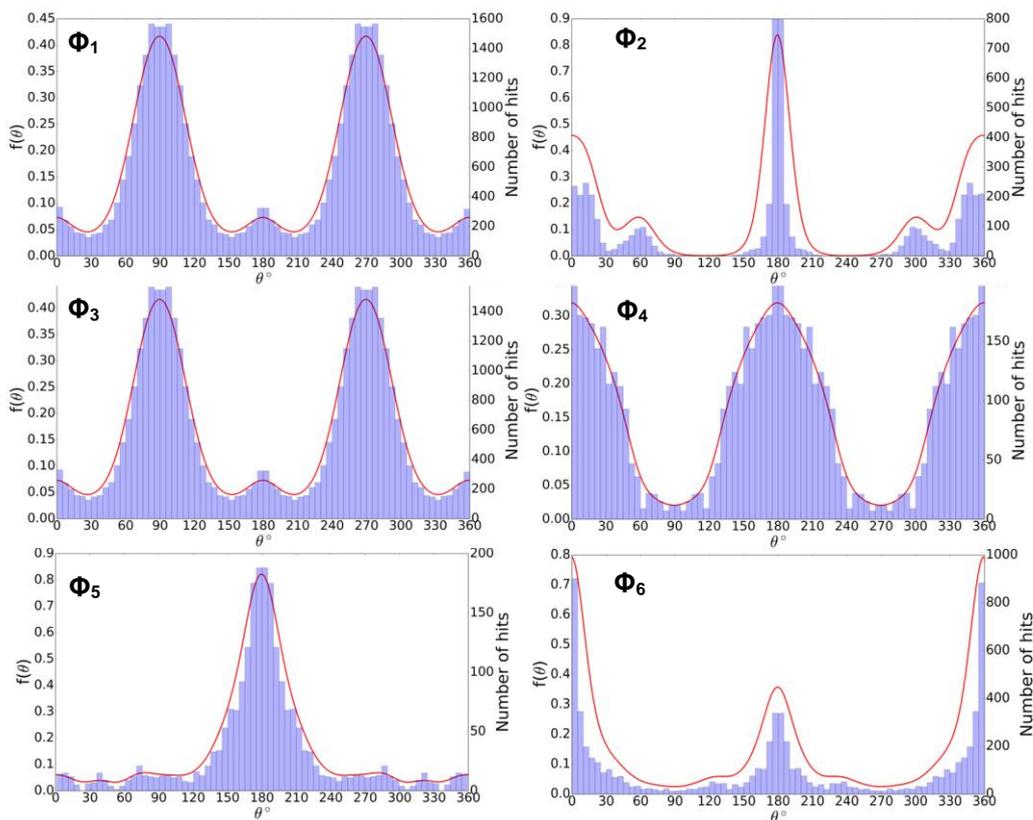
Figure 11: Histograms (light purple bars) and Von Mises kernel density approximations (red lines) for torsion angle distributions of molecule XXIII.

For Φ_1 and Φ_2 in both tautomers of mebendazole, the rotamer libraries failed to generate individual distributions since they are coupled around C14 (see Figure 7). Hence the Mogul distributions (Figure 17) were used and showed that they should be treated as explicitly flexible in the searches.
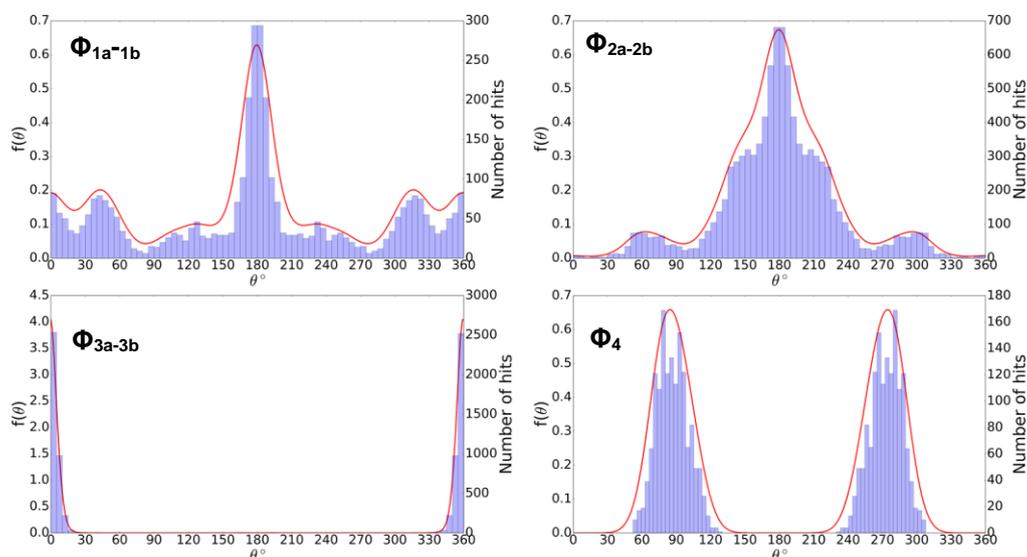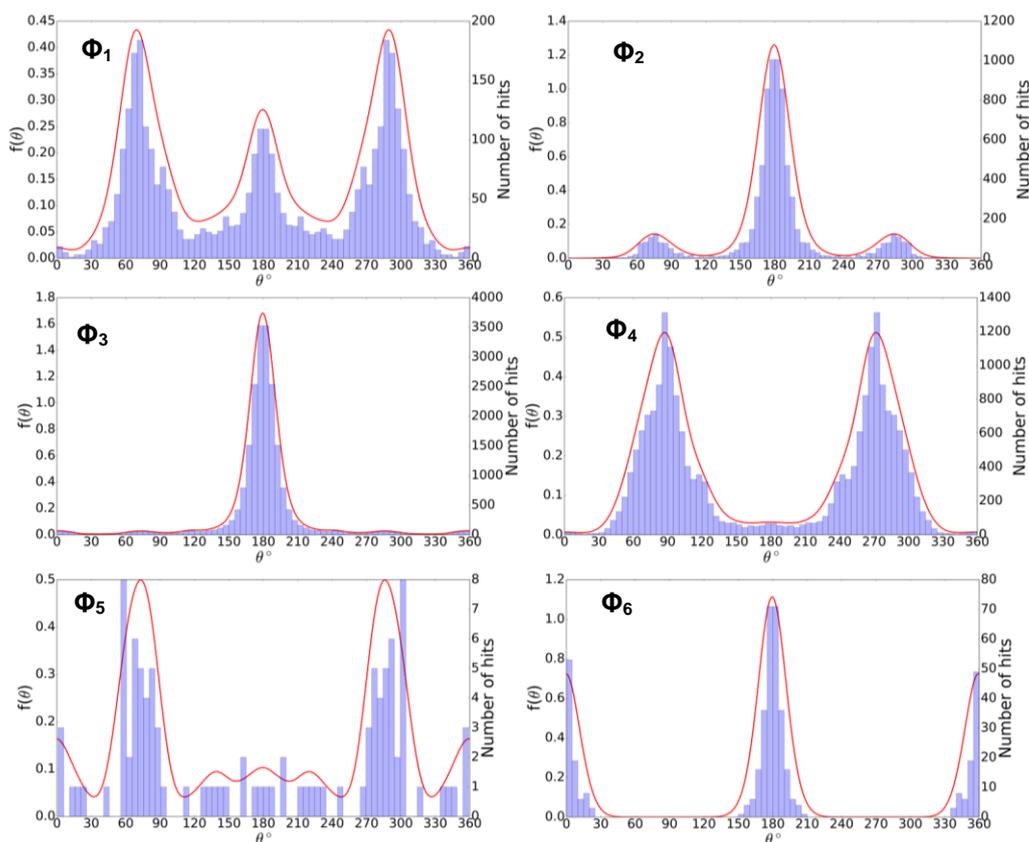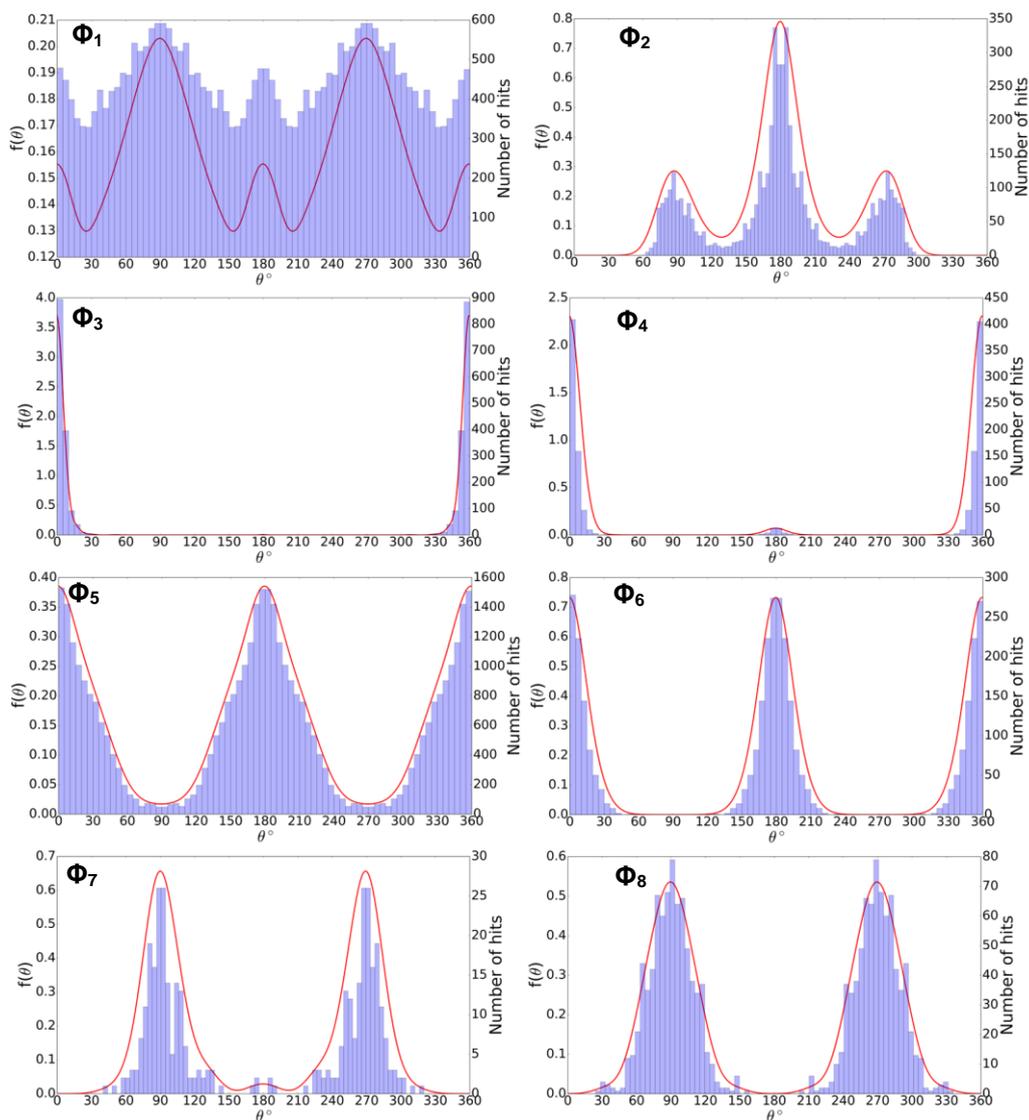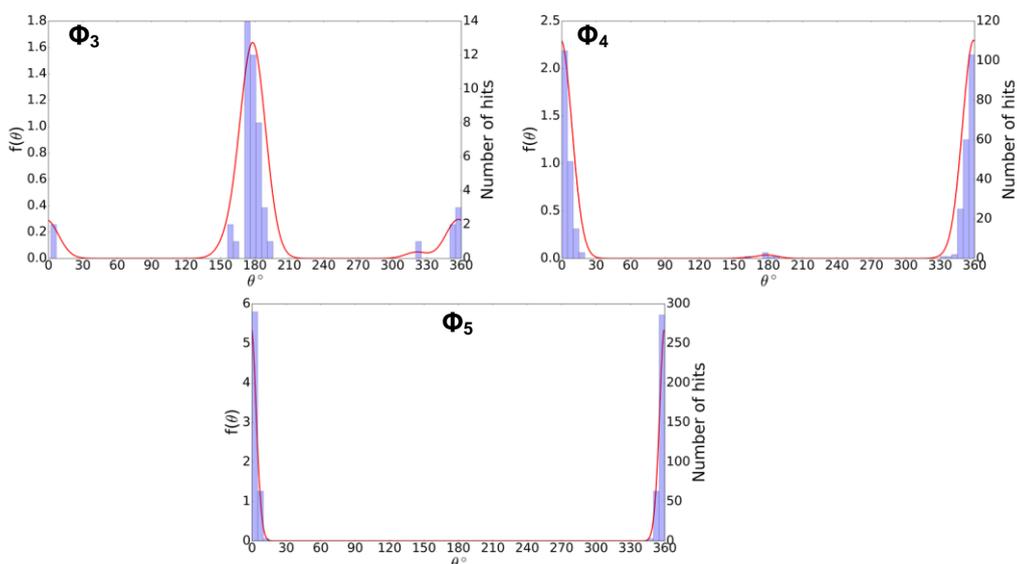

Figure 12: Histograms (light purple bars) and Von Mises kernel density approximations (red lines) for torsion angle distributions of angles of both tautomers of mebendazole.

Table 4: Summary of the Von Mises kernel density estimation for each key torsion angle. The number of maxima, their location and the HWHM, are shown. All these calculations were performed ensuring the angles were normalized to a 0-360° range. The maxima are given in bold if they are significantly higher than the others.

Label for XXVI	Distribution maxima ± HWHMs/°	Number of maxima
Φ_{1a} and Φ_{1b}	43 ± 22, **179 ± 17**, 315 ± 22	3
Φ_{2a} and Φ_{2b}	**359 ± 7**	1
Φ_{3a} and Φ_{4b}	**179 ± 35**	1
Φ_4	84 ± 21, 274 ± 21	2
Label for GSK269984B	**Distribution maxima ± HWHMs /°**	**Number of maxima**
Φ_1	69 ± 21, 179 ± 21, 289 ± 21	3
Φ_2	**179 ± 16**	1
Φ_3	**179 ± 13**	1
Φ_4	87 ± 24, 271 ± 24	2
Φ_5	73 ± 19, 285 ± 19	2
Φ_6	359 ± 15, **179 ± 14**	2
Φ_7	Unknown	/
Label for XX	**Distribution maxima ± HWHMs /°**	**Number of maxima**
Φ_1	89 ± 40, 269 ± 40	2
Φ_2	87 ± 19, **179 ± 19**, 271 ±19	3
Φ_3	**359 ± 7**	1
Φ_4	**359 ± 12**	1
Φ_5	359 ± 35, 179 ± 35	2
Φ_6	359 ± 18, 179 ± 19	2
Φ_7	90 ± 19, 268 ± 19	2
Φ_8	89 ± 25, 268 ± 25	2
Label for XXIII	**Distribution maxima ± HWHMs /°**	**Number of maxima**
Φ_1	89 ± 28, 269 ± 28	2
Φ_2	359 ± 25, **179 ± 13**	2
Φ_3	89 ± 28, 269 ± 28	2
Φ_4	359 ± 41, 179 ± 40	2
Φ_5	179 ± 24	1
Φ_6	**359 ± 17**, 179 ± 18	2
Φ_7	Unknown	/
Label for Mebendazole A and C	**Distribution maxima ± HWHMs /°**	**Number of maxima**
Φ_1	Unknown	/
Φ_2	Unknown	/
Φ_3	**178 ± 14**, 357 ± 13	2
Φ_4	**359 ± 12**	1
Φ_5	**359 ± 5**	1

2.2 Calculation of shape matches

2.2.1 Method

Shape effects were calculated by taking a reference conformation for a molecule and rotating each flexible torsion angle by 360° in 30° steps, comparing how much this change affected the overall shape of the molecule with respect to both the previous steps and the starting conformation. Ultra-fast shape recognition (USR) was used for this purpose. When each individual torsion angle was varied, the other ones were kept constant at the values they had in the starting conformation. Although USRCAT generates 60 moments for each molecular

conformation, only the first 12, which are identical to the standard USR moments, were considered. The shape match between two conformations was calculated as:

$$\% \ shape \ match = \frac{1}{1+\frac{1}{12}\sum_{l=1}^{12}\left|M_l^{ref}-M_l^{tar}\right|} \cdot 100 \qquad (2)$$

where M_l^{ref} and M_l^{tar} are the moment matrices of the target and the reference molecules and l is the moment number within each matrix. The torsion angles that determine the positions of polar hydrogens (Φ_7 in GSK269984B and Φ_7 in molecule XXIII) were not considered as barely affecting the shape of the molecule.

2.2.2 Shape matches of the 6 molecules used in the workflow decision trees

Table 5: Shape matches for varying each torsion angle of molecule XXVI with 30° steps with both the previous step and the starting conformation. The angles were normalized to a 0-360° range. The first value for each torsion angle corresponds to the initial value.

$\Phi_{1a,2a}$/°	% shape match with previous step	% shape match with original conformer	$\Phi_{2a,2b}$/°	% shape match with previous step	% shape match with original conformer
214.37	/	/	4.14	/	/
244.37	99.39	99.39	34.14	95.47	95.47
274.37	99.60	99.04	64.14	96.53	92.30
304.37	99.51	98.90	94.14	98.30	91.09
334.37	96.26	96.39	124.14	98.75	92.06
4.37	97.31	97.77	154.14	96.72	94.06
34.37	97.48	98.55	184.14	95.70	96.18
64.37	99.46	98.37	214.14	95.61	96.16
94.37	99.31	98.23	244.14	96.42	93.00
124.37	99.28	98.28	274.14	98.07	91.53
154.37	99.50	98.62	304.14	98.86	92.41
184.37	99.32	99.24	334.14	96.63	95.47
$\Phi_{3a,3b}$/°	% shape match with previous step	% shape match with original conformer	Φ_4/°	% shape match with previous step	% shape match with original conformer
222.46	/	/	258	/	/
252.46	98.71	98.71	287.97	91.41	91.41
282.46	98.41	97.31	317.97	93.60	86.04
312.46	98.14	95.83	347.97	93.32	82.99
342.46	98.19	94.40	17.97	96.53	80.93
12.46	98.65	93.58	47.97	88.74	81.81
42.46	98.90	93.54	77.97	88.27	82.76
72.46	96.24	95.21	107.97	92.50	85.50
102.46	98.39	94.86	137.97	76.01	77.20
132.46	98.70	95.15	167.97	77.46	71.13
162.46	96.79	97.82	197.97	93.17	73.82
192.46	98.75	98.98	227.97	83.49	85.93

Table 6: Shape matches for varying each torsion angle of GSK269984B with 30° steps with both the previous step and the starting conformation. The angles were normalized to a 0-360° range. The first value for each torsion angle corresponds to the initial value.

Φ_1/°	% shape match with previous step	% shape match with original conformer	Φ_2/°	% shape match with previous step	% shape match with original conformer
181	/	/	178.55	/	/
211	99.13	99.13	208.55	96.87	96.87
241	98.94	98.31	238.55	92.07	92.06
271	98.47	97.11	268.55	92.11	85.33
301	98.09	95.98	298.55	95.07	82.12
331	99.00	95.29	328.55	95.24	81.17
1	99.01	95.88	358.55	97.28	80.20
31	98.16	96.86	28.55	96.15	79.86
61	98.51	97.67	58.55	94.11	79.63
91	98.87	98.76	88.55	92.84	80.79
121	98.90	99.53	118.55	93.03	83.86
151	99.49	99.48	148.55	88.87	91.13

Φ_3/°	% shape match with previous step	% shape match with original conformer	Φ_4/°	% shape match with previous step	% shape match with original conformer
182.96	/	/	64.96	/	/
212.96	98.14	98.14	94.96	89.55	89.55
242.96	97.44	95.70	124.96	90.89	82.19
272.96	96.38	92.87	154.96	91.70	77.69
302.96	96.07	89.60	184.96	92.79	75.11
332.96	98.24	88.20	214.96	89.34	72.58
2.96	97.20	90.34	244.96	94.72	74.44
32.96	95.98	93.41	274.96	90.19	80.27
62.96	96.61	96.58	304.96	86.80	89.57
92.96	97.53	98.85	334.96	88.04	89.18
122.96	98.49	99.42	4.96	93.68	84.48
152.96	99.58	99.20	34.96	88.78	92.03

Φ_5/°	% shape match with previous step	% shape match with original conformer	Φ_6/°	% shape match with previous step	% shape match with original conformer
214.27	/	/	184.8	/	/
244.27	92.15	92.15	214.8	99.83	99.83
274.27	96.75	90.94	244.8	96.88	96.72
304.27	93.97	92.95	274.78	98.31	95.16
334.27	92.08	90.42	304.78	99.04	95.14
4.27	93.22	87.38	334.78	98.23	96.68
34.27	95.29	87.47	4.78	96.68	99.70
64.27	95.39	84.64	34.78	99.91	99.66
94.27	96.64	83.85	64.78	97.22	97.28
124.27	93.86	84.70	94.78	98.61	96.03
154.27	90.66	89.01	124.78	99.35	95.99
184.27	94.18	90.20	154.78	98.73	97.03

Table 7: Shape matches for varying each torsion angle of molecule XX with 30° steps with both the previous step and the starting conformation. The angles were normalized to a 0-360° range. The first value for each torsion angle corresponds to the initial value.

$\Phi_1/°$	% shape match with previous step	% shape match with original conformer	$\Phi_2/°$	% shape match with previous step	% shape match with original conformer
82.16	/	/	-105.82	/	/
112.16	96.73	96.73	-75.82	92.21	92.21
142.16	99.00	95.80	-45.82	89.91	87.53
172.16	98.38	96.51	-15.82	89.36	80.07
202.16	96.84	99.38	14.18	88.41	73.23
232.16	98.77	98.77	44.18	90.57	70.52
262.16	98.54	99.74	74.18	91.14	69.33
292.16	96.81	96.58	104.18	83.40	72.75
322.16	99.16	95.82	134.18	87.45	72.05
352.16	98.34	96.59	164.18	87.44	79.87
22.16	96.77	99.21	194.18	90.05	84.41
52.16	98.74	98.52	224.18	85.74	91.60

$\Phi_3/°$	% shape match with previous step	% shape match with original conformer	$\Phi_4/°$	% shape match with previous step	% shape match with original conformer
6.3	/	/	355.83	/	/
36.3	97.13	97.13	25.83	79.10	79.10
66.3	96.04	95.35	55.83	77.91	75.17
96.3	92.59	89.79	85.83	80.14	67.42
126.3	91.09	82.54	115.83	84.26	60.41
156.3	91.58	76.72	145.83	91.36	61.17
186.3	90.58	71.42	175.83	88.85	65.04
216.3	85.41	71.05	205.83	84.93	70.56
246.3	94.06	74.17	235.83	77.86	72.66
276.3	88.51	81.72	265.83	77.89	90.07
306.3	91.30	88.15	295.83	91.79	91.39
336.3	92.42	94.76	325.83	95.81	94.03

$\Phi_5/°$	% shape match with previous step	% shape match with original conformer	$\Phi_6/°$	% shape match with previous step	% shape match with original conformer
0.09	/	/	348.21	/	/
31.09	96.13	96.13	18.21	92.24	92.24
61.09	94.90	91.49	48.21	92.39	85.74
91.09	94.36	86.91	78.21	94.67	81.96
121.09	94.09	82.95	108.21	97.81	81.01
151.09	86.96	76.37	138.21	96.05	82.84
181.09	93.73	75.67	168.21	94.17	85.94
211.09	95.29	78.21	198.21	93.43	88.28
241.09	91.33	83.22	228.21	93.73	89.06
271.09	88.43	91.89	258.21	94.58	89.85
301.09	95.82	94.58	288.21	95.07	91.18
331.09	97.20	97.05	318.21	95.92	94.13

$\Phi_7/°$	% shape match with previous step	% shape match with original conformer	$\Phi_8/°$	% shape match with previous step	% shape match with original conformer
286.42	/	/	107.04	/	/
316.42	94.13	94.13	137.04	99.53	99.53
346.42	95.92	91.18	167.04	99.39	98.95
16.42	95.07	89.85	197.04	99.26	98.41
46.42	94.58	89.06	227.04	98.65	97.32
76.42	93.73	88.28	257.04	99.59	97.09
106.42	93.43	85.94	287.04	99.62	97.42
136.42	94.17	82.84	317.04	99.14	98.16
166.42	96.05	81.01	347.04	99.02	98.85
196.42	97.81	81.96	17.04	99.24	99.37
226.42	94.67	85.74	47.04	99.63	99.52
256.42	92.39	92.24	77.04	99.56	99.85

Table 8: Shape matches for varying each torsion angle of molecule XXIII with 30° steps with both the previous step and the starting conformation. The angles were normalized to a 0-360° range. The first value for each torsion angle corresponds to the initial value.

Φ_1/°	% shape match with previous step	% shape match with original conformer	Φ_2/°	% shape match with previous step	% shape match with original conformer
269.87	/	/	188.84	/	/
299.87	98.29	98.29	218.84	92.11	92.11
329.87	99.00	97.52	248.84	91.00	87.65
359.87	98.03	97.24	278.84	90.13	80.41
29.87	94.35	92.17	308.84	89.06	73.85
59.87	98.59	93.13	338.84	90.73	68.96
89.87	91.96	96.75	8.84	89.68	69.65
119.87	97.26	94.30	38.84	86.73	72.40
149.87	97.00	95.95	68.84	91.64	76.79
179.87	95.29	95.52	98.84	89.29	84.32
209.87	98.43	94.93	128.84	86.77	87.08
239.87	96.75	97.57	158.84	88.76	96.55

Φ_3/°	% shape match with previous step	% shape match with original conformer	Φ_4/°	% shape match with previous step	% shape match with original conformer
87.22	/	/	132.45	/	/
117.22	97.54	97.54	162.45	97.83	97.83
147.22	93.52	92.18	192.45	98.58	96.47
177.22	98.58	92.72	222.45	93.77	92.15
207.22	96.14	90.77	252.45	96.17	90.65
237.22	97.74	92.21	282.45	95.55	92.97
267.22	94.80	90.36	312.45	91.32	94.18
297.22	93.37	96.45	342.45	98.04	95.97
327.22	94.51	97.61	12.45	95.12	98.73
357.22	92.81	92.50	42.45	91.06	91.07
27.22	99.65	92.58	72.45	98.83	91.97
57.22	98.67	93.28	102.45	93.00	97.35

Φ_5/°	% shape match with previous step	% shape match with original conformer	Φ_6/°	% shape match with previous step	% shape match with original conformer
166.75	/	/	4.47	/	/
196.75	97.29	97.29	34.47	98.40	98.40
226.75	91.32	90.47	64.47	94.44	93.41
256.75	97.13	88.35	94.47	99.77	93.23
286.75	97.73	87.06	124.47	99.84	93.29
316.75	97.72	86.89	154.47	93.84	97.17
346.75	97.39	87.48	184.47	98.67	98.42
16.75	97.57	88.88	214.47	98.66	97.19
46.75	98.37	89.96	244.47	94.56	93.42
76.75	98.33	90.68	274.47	99.88	93.41
106.75	98.07	91.71	304.47	99.76	93.56
136.75	97.81	92.99	334.47	93.68	98.59

Table 9: Shape matches for varying each torsion angle of mebendazole with 30° steps with both the previous step and the starting conformation. The angles were normalized to a 0-360° range. The first value for each torsion angle corresponds to the initial value.

$\Phi_1/°$	% shape match with previous step	% shape match with original conformer	$\Phi_2/°$	% shape match with previous step	% shape match with original conformer
166.19	/	/	327.79	/	/
199.19	97.62	97.62	357.79	93.86	93.86
229.19	97.59	95.35	27.79	88.55	85.12
259.19	90.64	87.23	57.79	91.60	86.59
289.19	88.47	95.96	87.79	91.08	90.66
319.19	98.10	96.42	117.79	92.04	88.46
349.19	97.77	96.20	147.79	95.51	85.08
19.19	97.79	95.56	177.79	96.40	82.72
49.19	98.32	94.59	207.79	97.04	81.73
79.19	94.43	95.84	237.79	93.19	86.22
109.19	99.32	96.47	267.79	94.90	90.05
139.19	98.47	97.94	297.79	93.66	93.47

$\Phi_3/°$	% shape match with previous step	% shape match with original conformer	$\Phi_4/°$	% shape match with previous step	% shape match with original conformer
176.61	/	/	0.42	/	/
206.61	98.20	98.20	30.42	98.21	98.21
236.61	97.81	96.08	60.42	96.89	95.61
266.61	97.75	94.12	90.42	95.70	91.68
296.61	97.00	91.71	120.42	94.81	88.35
326.61	98.21	90.93	150.42	96.44	85.77
356.61	97.14	91.57	180.42	95.59	86.03
26.61	97.66	93.26	210.42	98.86	86.59
56.61	97.49	94.45	240.42	96.88	86.81
86.61	97.27	96.24	270.42	95.72	90.21
116.61	99.28	96.63	300.42	95.32	94.31
146.61	98.31	97.75	330.42	96.15	97.87

$\Phi_5/°$	% shape match with previous step	% shape match with original conformer
0.62	/	/
30.62	98.44	98.44
60.62	97.42	95.95
90.62	96.96	93.22
120.62	96.30	90.10
150.62	98.40	89.01
180.62	99.06	89.22
210.62	98.23	90.57
240.62	97.40	92.74
270.62	97.20	95.21
300.62	97.57	97.18
330.62	98.49	98.27

2.3 Mogul distributions for range of explicitly flexible torsion angles

The grid ranges of angles treated as explicitly flexible in the search were selected using Mogul distributions. Mogul only covers a 0-180° range for torsion angles, so the 180-360° range was generated by symmetry. The grids covered the range that contained any CSD entry, as the *ab initio* calculations would exclude inaccessible geometries from the searches. In these examples there were no outliers, but if there had been any angle represented by a single CSD entry without any crystal structures with neighboring values, these crystal structures should

be checked for their relevance to the CSP molecule. Finally, if certain ranges were completely absent in the CG generated conformations, because of some intramolecular clashes, they were excluded from the grids. The distributions shown in Figures 8-12 were plotted with the Matplotlib Python package.
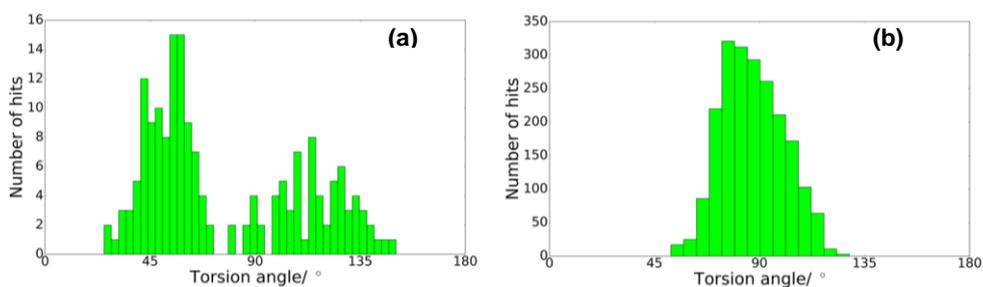
Figure 13: Mogul distributions used to select the ΔE$_{intra}$ grid dimensions for (a) Φ$_{1a-1b}$ and (b) Φ$_4$ of molecule XXVI.
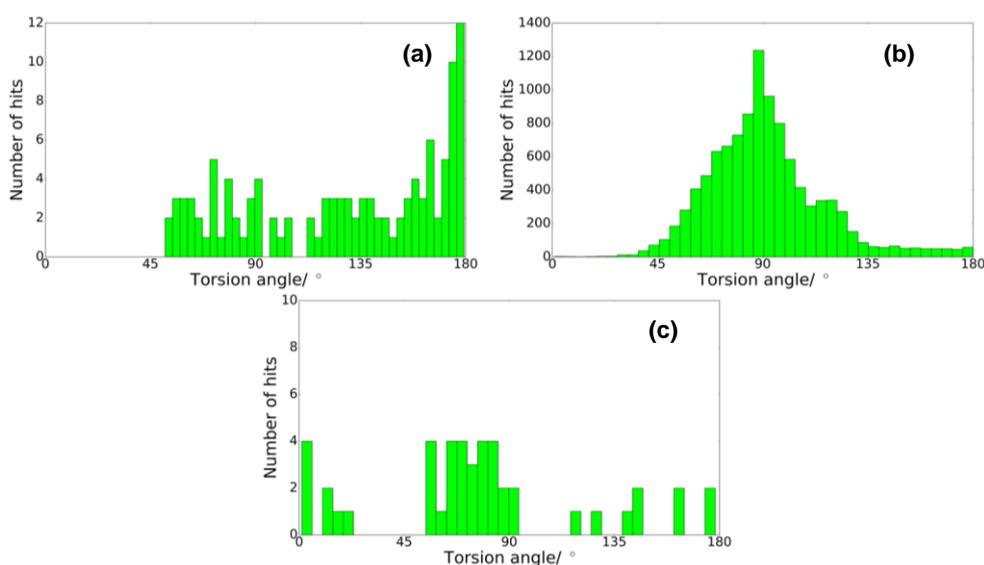
Figure 14: Mogul distributions used to select the ΔE$_{intra}$ grid dimensions for (a) Φ$_1$ (b) Φ$_4$ and (c) Φ$_5$ of GSK269984B
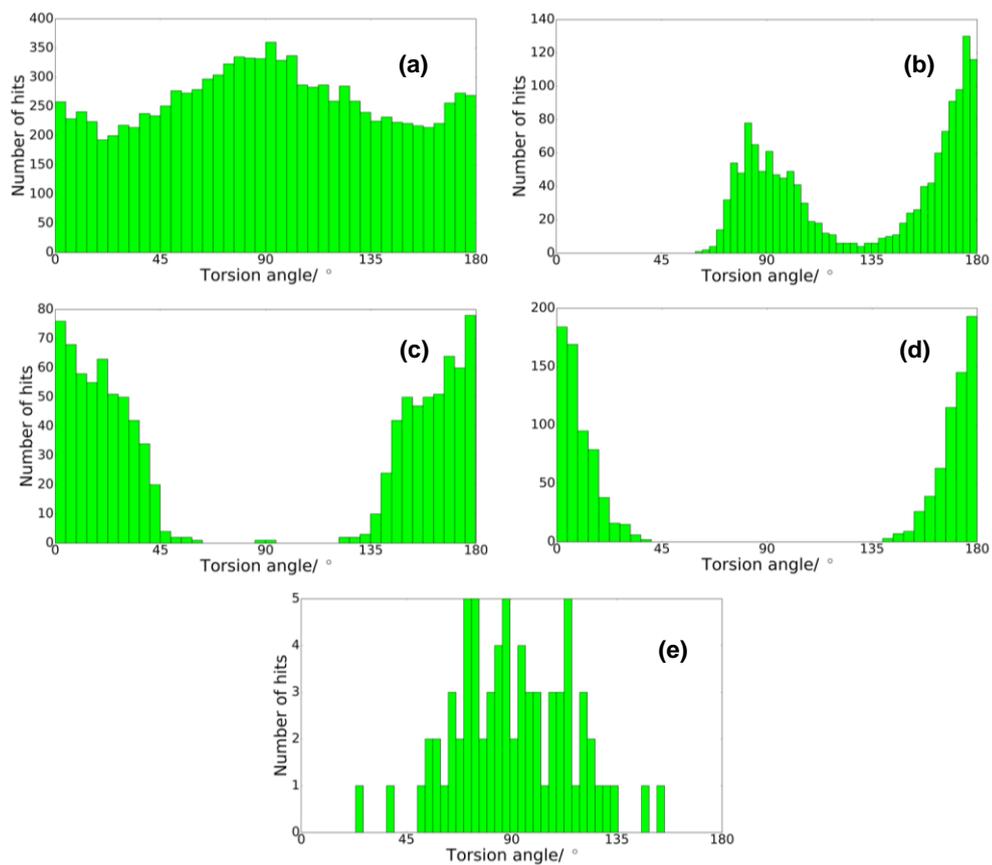
Figure 15: Mogul distributions used to select the ΔE_{intra} grid dimensions for (a) Φ_1 (b) Φ_2 (c) Φ_5 (d) Φ_6 and (e) Φ_8 of molecule XX.
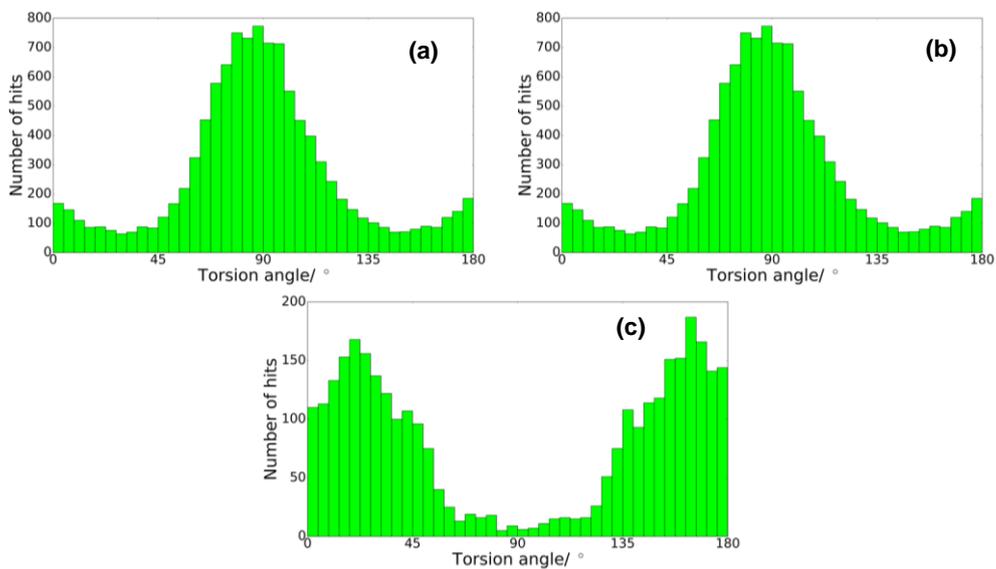
Figure 16: Mogul distributions used to select the ΔE_{intra} grid dimensions for (a) Φ_1 (b) Φ_3 (c) Φ_4 of molecule XXIII.
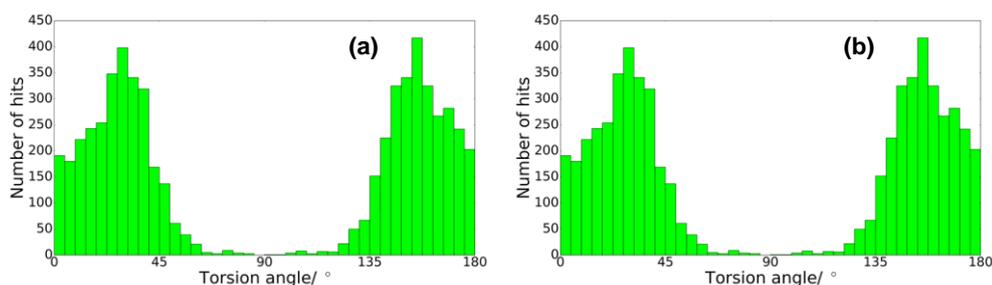
Figure 17: Mogul distributions used to select the ΔE_{intra} grid dimensions for (a) Φ_1 (b) Φ_2 of mebendazole.

2.4 Surrogate molecules for grid generation

In order to make the calculation of the ΔE_{intra} grids computationally cheaper, the molecules were represented by appropriate surrogate molecules, when it was reasonable to assume that the torsion angles were approximately independent. This assumption is too inaccurate for torsion angles that define positions of groups that interact strongly with one another. Hence, if explicitly flexible torsions were adjacent they were scanned together. The set of molecules used to calculate the grids for each explicitly flexible torsion angle are shown in Figure 18-Figure 22.
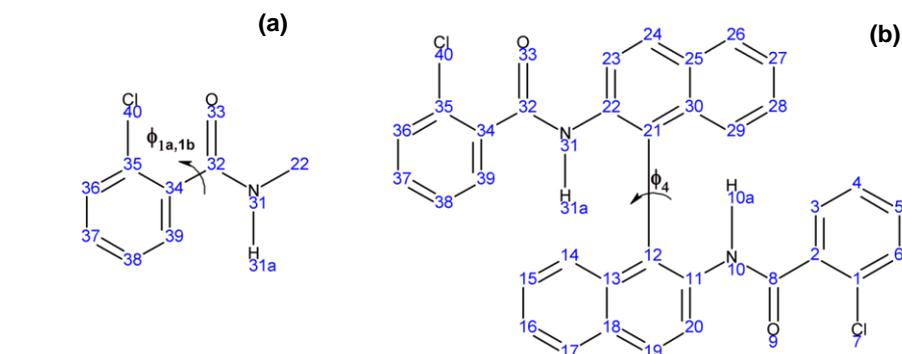
Figure 18: Surrogate molecules used to calculate the ΔE_{intra} grids of (a) Φ_{1a} and Φ_{1b} and (b) sterically congested Φ_4 of molecule XXVI.

22
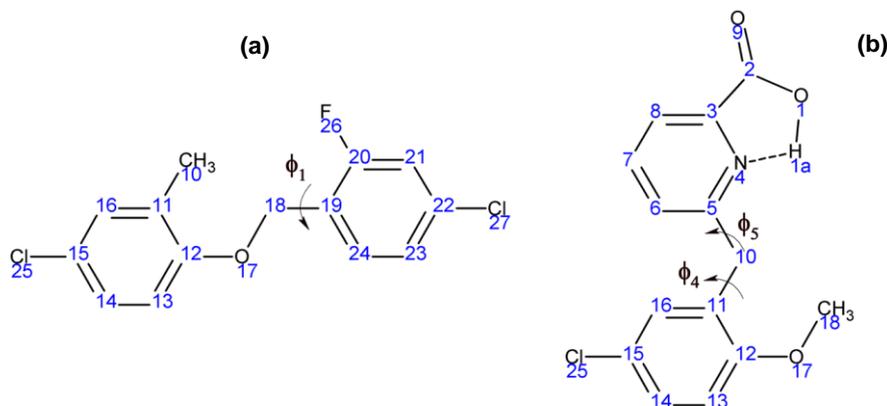
Figure 19: Surrogate molecules used to calculate the ΔE_{intra} grids of (a) Φ_1 (b) Φ_4 and Φ_5 of GSK269984B.
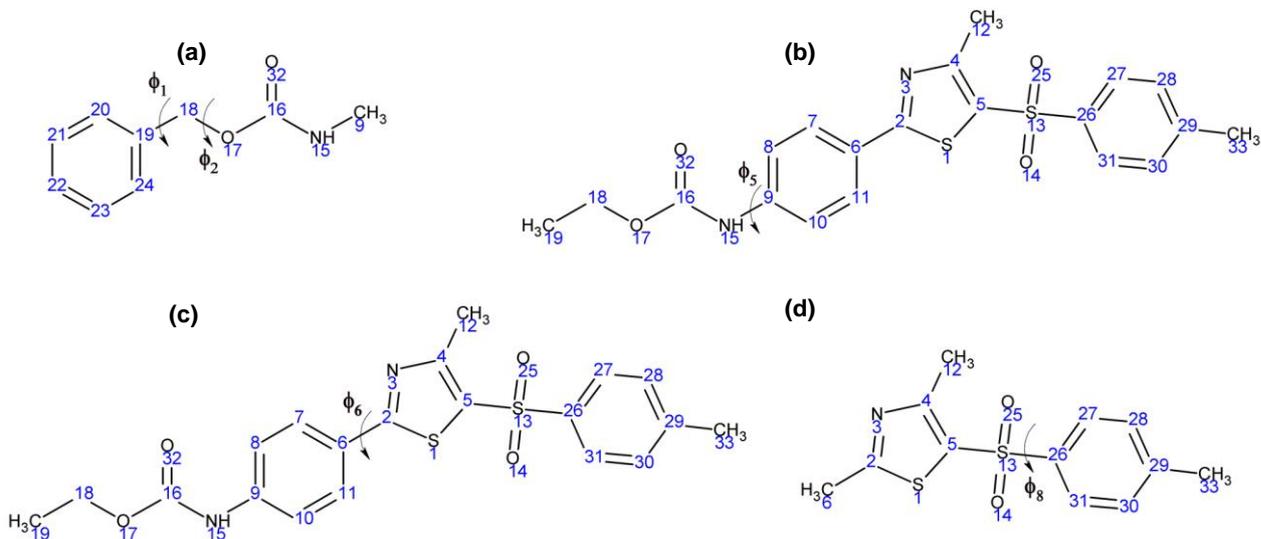
Figure 20: Surrogate molecules used to calculate the ΔE_{intra} grids of (a) Φ_1 and Φ_2 (b) Φ_5 (c) Φ_6 and (d) Φ_8 of molecule XX.
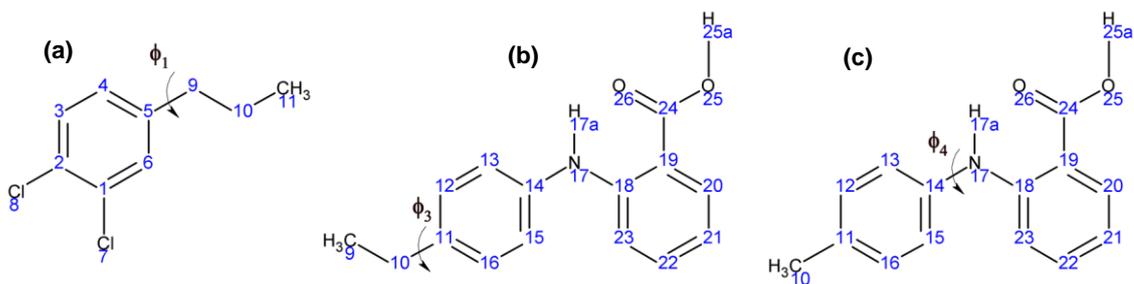
Figure 21: Surrogate molecules used to calculate the ΔE_{intra} grids of (a) Φ_1 (b) Φ_3 (c) Φ_4 of molecule XXIII.
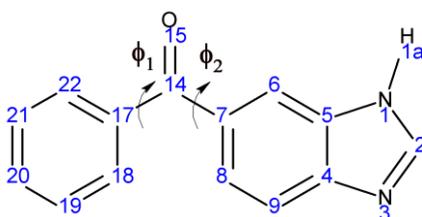
Figure 22: Surrogate molecule used to calculate the ΔE_{intra} grid of Φ_1 and Φ_2 of both tautomers of mebendazole.

2.5 Scans of torsions with terminal hydrogens

Since the CG and the related CSD distributions do not describe the position of torsion angles with terminal hydrogen atoms, given the inaccuracies in their experimental determinations, it was necessary to perform constrained angle scans, as it would be done in standard CSP studies for any torsion angles.

2.5.1 GSK269984B Φ_7

Two scans were performed, one with the OH group on the same side as the nitrogen atom, with which it can form an internal hydrogen bond, and one with the OH group on the opposite side of it.
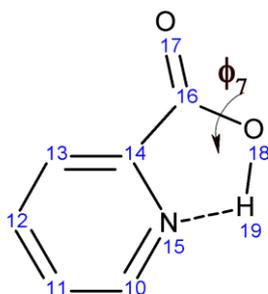
Figure 23: Fragment used to scan angle Φ_7 in GSK269984B.
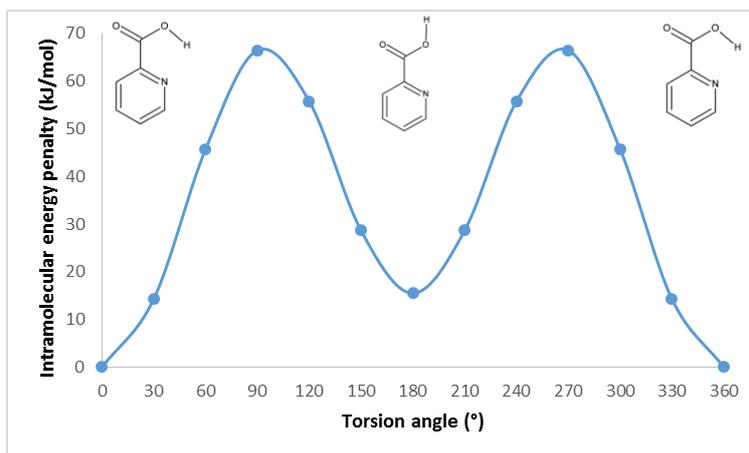
Figure 24: Constrained angle scan of Φ_7 in GSK269984B, with 30° steps, when the OH group is on the same side as the N atom in the fragment shown in Figure 23.
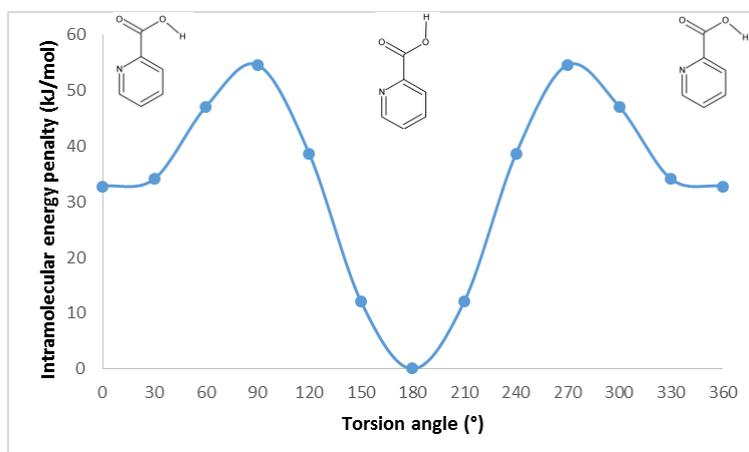
Figure 25: Constrained angle scan of Φ_7 in GSK269984B, with 30° steps, when the OH group is on the opposite side of the N atom in the fragment shown in Figure 23.

Looking at these two graphs, it was decided to constrain Φ_7 to two values (i.e. 0° and 180°) when the OH group is on the same side as the nitrogen atom, and one value (i.e. 180°) when it is on the opposite side. Although some minor variations in these values are possible, this conformational adjustment would occur upon full optimization of the generated crystal structures.

2.5.2 XXIII Φ_7

In this case, only one scan was performed, given that there was no possibility of forming an internal H-bond within the scanned group.
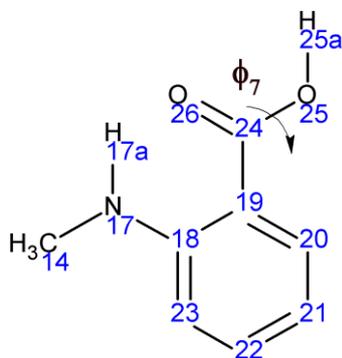
Figure 26: Fragment used to scan angle Φ_7 in molecule XXIII.
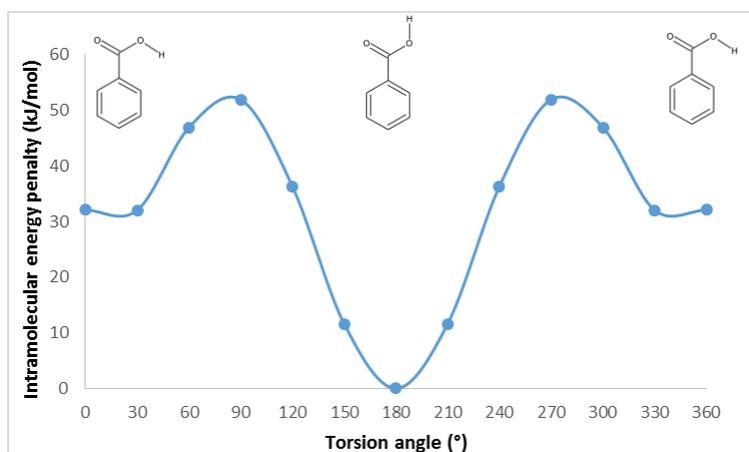
Figure 27: Constrained angle scan of Φ_7 in molecule XXIII, with 30° steps, using the fragment shown in Figure 26.

Upon looking at this graph, it was decided to constrain Φ_7 to of 180°. Once again, some minor oscillations around this values appear plausible but they should be recovered upon full optimization.

2.6 Selection of the conformational regions

Table 10 illustrates an example of how the CG-generated conformations are picked by the automated script depending on the separation thresholds (see section 2.1 of the main paper) for a hypothetical molecule. This removes CG conformations that are so similar to those already present to represent a distinct conformational region and whose inclusion would mean that the conformational space was covered twice.

Table 10. Example of the functioning of the selection algorithm. A hypothetical set of 10 CG generated conformations with three torsion angles suitable for keeping rigid during the CSP search was reduced to a set of 6 conformations defining separate conformational regions, given that the 3 torsion angles had separation thresholds of 25°, 45° and 90° respectively. The first conformation is always selected, and the other conformations are processed sequentially.

Conformation number	Torsion 1 (separation threshold of 25°) /°	Torsion 2 (separation threshold of 45°) /°	Torsion 3 (separation threshold of 90°) /°	Selected ?	Covered by conformation
1	180	180	180	YES	/
2	150	180	180	YES	/
3	180	210	180	NO	1
4	180	180	240	NO	1
5	180	180	0	YES	/
6	150	240	180	YES	/
7	180	240	180	YES	/
8	180	210	0	NO	5
9	180	240	0	YES	/
10	180	270	180	NO	7

Table 11-16 show for each molecule the CG selected conformations defining a separate conformational region, with the values of each constrained torsion angle. The calculated conformational energies of each region correspond to PBE0 6-31G(d,p) optimizations that were started with the explicitly flexible torsion angles at their minima on the respective grids.

Table 11: Selected conformational regions and calculation of the associated intramolecular energy penalty for molecule XXVI. Regions highlighted in blue were taken to the search stage (some conformations with $\Delta E_{Intra}^{CR} < 26$ kJ/mol were excluded because of an approximate molecular symmetry relationship; in those cases, the lower energy one was always chosen).

Conformer number	$\Phi_{2a}/°$	$\Phi_{3a}/°$	$\Phi_{2b}/°$	$\Phi_{3b}/°$	ΔE_{intra}
1	-4	178	9	179	1.92
25	0	212	-8	212	11.7
26	-4	151	-8	151	11.63
29	2	148	4	178	4.35
30 ~ 41 for symmetry	-2	212	-4	182	6.54
31 ~ 29 for symmetry	-2	178	2	148	4.47
37	2	209	8	148	7.93
38 ~ 37 for symmetry	-2	151	-8	212	8.67
41	2	178	2	209	5.54
60 ~ 124 for symmetry	0	178	8	238	15.77
61	0	182	-8	122	12.86
89	-3	66	-9	181	20.98
93	-4	209	8	118	16.42
94 ~ 198 for symmetry	4	151	-8	242	21.14
122 ~ 61 for symmetry	0	122	-4	182	12.93
124	-1	238	0	178	15.76
174 ~ 176 for symmetry	3	148	8	119	22.11
175	2	238	4	209	20.27
176	-2	122	-4	151	19.03
186 ~ 1502 for symmetry	3	177	5	89	25.78
187	-4	182	-4	272	26.34
190	2	268	9	179	25.54
198	2	238	2	148	19.63
199	-2	122	-2	212	16.07
218	-4	268	8	209	29.11
219	-4	148	8	88	31.66
224 ~ 175 for symmetry	-2	212	-4	242	22.61
232	0	65	0	212	23.61
241	-4	177	8	297	30.71
275	0	148	0	268	29.59
277	2	209	8	88	18.32
279	-2	212	-8	272	30.9
294	0	298	3	177	31.31
322	0	242	0	242	32.23
323	-4	122	-4	122	32.72
338	2	268	8	148	29.33
353 ~ 89 for symmetry	-4	179	8	59	22.16
364 ~ 396 for symmetry	-4	122	-8	242	25.54
375 ~ 805 for symmetry	-3	66	-8	242	21.01
396	-2	242	-8	122	25.53
407	0	148	8	297	33.26
417	-1	88	2	148	22.96
465	2	298	1	209	33.73
466	-2	212	-8	301	34.75
471	4	122	-8	272	36.93
472	-4	238	8	88	21.61
510	2	298	4	148	33.49
516	4	93	0	123	34.34
517	-4	272	-8	242	40.83
556	0	238	4	268	38.7
562	-2	272	-8	122	37.02
601	-4	58	2	148	38.17
609	4	123	-3	93	34.26
617	1	65	-8	272	31.18

646	0	152	-2	62	36.58
648	-4	239	8	297	43.3
696	-4	268	8	268	47.07
698 ~ 472 for symmetry	-2	92	-4	242	22.23
723 ~ 232 for symmetry	2	209	0	59	25.09
730 ~ 277 for symmetry	4	91	-2	212	23.57
741	1	298	4	239	43.27
742	0	122	0	301	41.14
754	-4	298	8	297	51.74
796	4	298	8	118	42.16
803	-4	117	8	59	37.67
805	4	241	-8	63	20.96
852	8	268	4	88	30.19
886	0	58	4	117	37.75
905	-10	61	-10	62	84.61
1160	2	298	4	268	51.42
1194	-2	272	-2	301	51.96
1214	4	91	-8	91	38.87
1502	-2	92	-9	181	23.87
1534	-4	178	178	143	23.2
1535	4	182	-178	217	30.07
1767	-178	178	9	179	31.86
1773	176	148	-5	182	27.63
1826	-179	216	-4	182	30.17
2000	-2	182	178	182	30.64
2099	179	144	2	148	35.22
2100	-179	216	-2	212	35.85
2101	-175	213	2	148	33.17
2102	174	147	-2	212	40.4
2380	-179	177	2	209	36.19
2381	179	183	-2	151	34.79
2533	-179	216	-8	242	47.54
2534	179	144	9	239	41.28
2535	179	144	8	118	38.82
2537	-174	213	8	118	40.85
2776	-2	242	174	146	49.44
2813	178	182	-8	242	46.44
2814	179	183	-8	122	45.51
2875	-4	212	175	236	43.66
2876	0	151	175	236	38.68
2954	-2	122	-180	213	40.05
2972	4	212	174	147	40.27
3058	2	238	-178	178	46.43
3060	2	118	-178	178	48.81
3188	2	209	-176	209	33.09
3189	-2	151	174	146	33.95
3192	1	123	179	147	45.88
3213	-179	216	0	92	41.97
3214	-179	216	0	272	54.2
3308	-180	146	5	267	49.48
3403	-180	146	5	91	43.72
3421	2	238	179	233	53.41
3445	-178	268	9	178	56.52
3489	-174	213	8	297	53.99
3518	-2	68	174	234	44.9
3596	2	177	-178	268	56.33
3682	-178	88	9	180	43.28
3716	-179	149	11	63	44.79
3717	-180	213	-8	63	43.97
3833	2	180	-177	89	43.14
3840	2	207	-178	270	64.2

3841	177	257	2	205	53.53
3842	178	272	-2	151	60.4
3912	-179	100	-8	211	39.19
4007	0	150	-177	121	34.63
4097	-175	92	2	148	41.38
4098	-178	270	8	121	67.03
4099	-178	271	8	239	78.72
4109	5	300	-178	236	59.1
4138	-4	122	-178	238	45.2
4228	-2	122	178	272	50.67
4229	-2	242	178	273	68.4
4310	178	91	-8	241	53.02
4395	2	238	-178	88	51.68
4396	5	95	-172	124	46.35
4774	2	148	-176	210	32.72
4819	179	144	178	143	40.77
4820	-179	216	-178	217	56.76
4821	-174	213	178	143	46
4822	174	147	-178	217	48.01
4872	6	237	-178	119	43.25
4899	2	209	-178	117	38.59
4928	-178	238	2	122	45.03
4941	-178	215	178	242	66.95

Table 12: Selected conformational regions and calculation of the associated intramolecular energy penalty for GSK269984B. The values of the polar hydrogen torsion angle highlighted in yellow were added manually, and those conformations that differ from the CG-generated ones only in the values of such torsions are indicated using a '_N' notation. Regions highlighted in blue were taken to the search (some conformations with $\Delta E_{Intra}^{CR} < 26$ kJ/mol were excluded because of an approximate molecular symmetry relationship; in those cases, the lower energy one was always chosen).

Conformer number	$\Phi_2/°$	$\Phi_3/°$	$\Phi_6/°$	$\Phi_7/°$	ΔE_{intra}
1	180	180	180	0	0.07
1_2	180	180	180	180	12.89
11	0	180	180	180	14.02
184	180	181	272	0	5.84
184_2	180	181	272	180	21.64
185 ~ 184 for symmetry	180	178	88	0	10.17
185_2 ~ 184_2 for symmetry	180	178	88	180	25.75
306	0	181	272	180	19.42
307 ~ 306 for symmetry	0	178	88	180	19.56
475	180	120	180	0	16.2
475_2	180	120	180	180	22.57
477	180	240	180	0	13.97
477_2	180	240	180	180	22.39
666	360	120	180	180	26.74
668	0	240	182	180	26.92
1694	180	75	181	0	23.31
1694_2	180	75	181	180	37.41
2018	180	285	179	0	20.48
2018_2	180	285	179	180	34.7
2302	360	75	181	180	43.91
2655	0	285	179	180	39.89
2730 ~ 2731 for symmetry	180	240	270	0	20.3
2730_2	180	240	270	180	35.67
2731	180	239	91	0	16.89
2731_2	180	239	91	180	28.31
2732	180	120	90	0	18.44
2732_2	180	120	90	180	33.64

3487	0	240	270	180	41.6
3488	0	121	269	180	42.66
3489	0	120	90	180	39.97
3491	360	238	91	180	31.57
5036	180	124	300	0	16.92
5036_2	180	124	300	180	32.57
5162	180	287	267	0	25.91
5162_2	180	287	267	180	37.21
5974	180	71	91	0	24.6
5974_2	180	71	91	180	39.48
6136	0	287	267	180	41.77
6810	180	120	240	0	24.51
6810_2	180	71	91	180	38.58
6875	0	71	91	180	45.49
9131	180	91	33	0	27.33
9131_2	180	91	33	180	34.39
9132	180	271	327	0	29.88
9132_2	180	271	327	180	42.18
9267	180	242	32	0	26.04
9267_2	180	242	32	180	38.83
9327	0	91	33	180	34.81
9329	360	271	327	180	46.87
9395	360	242	32	180	41.9
9458	0	123	327	180	60.49

Table 13: Selected conformational regions and calculation of the associated intramolecular energy penalty for molecule XX. Regions highlighted in blue were taken to the search stage (some conformations with ΔE_{Intra}^{CR} < 26 kJ/mol were excluded because of an approximate molecular symmetry relationship; in those cases, the lower energy one was always chosen).

Conformer number	Φ_3/°	Φ_4/°	Φ_7/°	ΔE_{intra}
1	180	180	270	1.45
25	180	180	240	5.25
78	180	180	300	3.34
92	180	180	180	12.33
245	180	357	270	11.68
273	180	180	210	8.48
413	180	180	330	15.50
515	180	3	240	15.50
745	180	357	300	13.79
826 ~ 273 for symmetry	179	180	150	8.46
1290	180	357	180	22.48
2911	180	357	330	25.84
3276 ~ 3739 for symmetry	180	357	150	18.56
3739	180	357	210	18.55
4480 ~ 245 for symmetry	180	357	90	11.69
4974 ~ 2911 for symmetry	180	357	30	25.87
7122	358	140	270	34.18
9765	358	140	240	38.30
10397	180	330	119	15.37
14548	233	177	120	26.23
16943	126	180	120	32.05

Table 14: Selected conformational regions and calculation of the associated intramolecular energy penalty for molecule XXIII. The values of the polar hydrogen torsion angle highlighted in yellow were added manually. Regions highlighted in blue were taken to the search stage (some conformations with $\Delta E_{Intra}^{CR} < 26$ kJ/mol were excluded because of an approximate molecular symmetry relationship; in those cases the lower energy one was always chosen).

Conformer number	$\Phi_2/°$	$\Phi_5/°$	$\Phi_6/°$	$\Phi_7/°$	ΔE_{intra}
1	180	164	360	180	0.57
9	360	164	0	180	25.13
17	180	164	180	180	16.66
33	0	164	180	180	41.2
49	60	164	0	180	1.15
51 ~ 49 for symmetry	300	164	0	180	2.17
65	0	229	360	180	32.68
103	180	214	0	180	3.34
145 ~ 147 for symmetry	60	164	180	180	17.56
147	300	164	180	180	17.02
172	0	229	180	180	45.8
231	180	161	240	180	35.84
232	180	161	120	180	43.69
235	180	214	180	180	18.11
256	180	117	360	180	16.07
290	60	229	360	180	9.3
309	180	270	0	180	32.67
328	0	199	120	180	59.34
329	0	199	240	180	67.29
402	0	199	60	180	63.72
403	0	199	300	180	54.51
434	360	90	360	180	53.8
464	180	64	2	180	47.28
500	300	214	0	180	4.94
512	180	117	180	180	27.25
563	60	229	180	180	22.58
613	180	269	180	180	38.59
645	180	199	60	180	37.47
646	180	199	300	180	44.14
658	0	296	358	180	71.63
712	0	117	180	180	51.65
831	0	143	240	180	62.23
893	60	199	120	180	36.3
894	60	199	240	180	43.95
897	300	199	120	180	48.44
919	300	214	180	180	19.44
949	60	117	360	180	16.42
950 ~ 949 for symmetry	300	117	360	180	17.91
1055	300	199	60	180	38.75
1056	300	199	300	180	32.54
1177	300	270	360	180	34.13
1580	300	64	2	180	47.64
1581	60	64	2	180	42.06
1640	0	138	90	180	76.89
1677	60	117	180	180	27.94
1678	300	117	180	180	27.17
1720	300	164	240	180	36.33
1957	300	270	180	180	39.94
2016	60	199	60	180	40.49
2017	60	199	300	180	31.37
2279	300	229	240	180	49.39
2334	180	61	213	180	45.36

2511	0	61	213	180	62.18
2512	0	299	147	180	69.56
2527	180	207	240	180	44.01
2530	180	229	120	180	39.45
2661	60	296	358	180	48.52
2915	180	131	300	180	46.17
2916	180	131	60	180	35.38
2926	300	146	90	180	52.46
2932	60	143	90	180	53.14
2933	60	143	270	180	55.14
3001	0	270	240	180	83.92
3020	0	117	300	180	76.46
3098	180	300	300	180	54.68
3099	180	61	60	180	54.68
3169	180	300	120	180	52.81
3170	180	63	118	180	66.95
3468	0	90	60	180	64.76
3469	0	270	300	180	72.65
3489	0	90	120	180	77.15
3882	0	270	60	180	85.38
3948 ~ 3950 for symmetry	120	164	0	180	15.94
3950	240	164	360	180	15.9
4283	359	36	60	180	70.45
4484	180	267	241	180	58.81
4810	360	63	269	180	82.6
4889	60	61	213	180	44.03
4891	60	299	147	180	46.9
4991	300	120	300	180	51.82
5452	300	61	213	180	39.86
5580	180	61	299	180	64.73
5590	180	299	61	180	65.36
5609	60	90	120	180	61.79
5741	120	164	180	180	32.03
5743	240	164	180	180	31.73
6049	300	90	60	180	44.92
6050	60	90	300	180	57.91
6155	300	90	120	180	53.02
6560	60	60	60	180	46.32
6563	60	299	300	180	55.85
6586	300	270	60	180	62.25
6587	300	270	300	180	49.82
6657	300	292	121	180	53.91
7010	330	325	300	180	67.45
7199	120	229	360	180	24.24
7242	300	35	60	180	55.32
7811	60	270	240	180	60.15
8717	240	214	0	180	18.84
8819	60	270	60	180	61.7
9063	120	229	180	180	37.37
9528	299	61	299	180	60.17
10475	240	161	120	180	58.81
10476	240	199	240	180	58.74
10557	240	214	180	180	33.5
10643	240	117	360	180	31.44
10644	120	117	0	180	31.05
10888	120	199	60	180	55.41
10889	240	161	60	180	46.51
10890	240	161	300	180	55.45
11600	240	270	0	180	48.03
11777	120	164	270	180	68.83
11924	120	64	2	180	62.08

11926	120	296	358	180	62.73
11928	240	64	2	180	62.18
12019	120	117	180	180	42.43
12024	240	117	180	180	42.54
12045	120	164	120	180	58.88
12267	240	270	180	180	53.82
12351	293	297	242	180	66.85
12628	240	229	120	180	54.52
13511	120	210	120	180	51.8
13512	120	210	240	180	59.61
13514	240	150	240	180	51.53
13625	120	150	60	180	47.08
13628	240	229	300	180	50.56
14103	240	214	60	180	57.34

Table 15: Selected conformational regions and calculation of the associated intramolecular energy penalty for the A-tautomer of mebendazole. Regions highlighted in blue were taken to the search stage.

Conformer number	$\Phi_3/°$	$\Phi_4/°$	$\Phi_5/°$	ΔE_{intra}
1	180	360	0	0.00
5	0	0.01	360	50.49
35	184	186	360	16.25
46	2	183	0	48.17

Table 16: Selected conformational regions and calculation of the associated intramolecular energy penalty for the C-tautomer of mebendazole. Regions highlighted in blue were taken to the search stage.

Conformer number	$\Phi_3/°$	$\Phi_4/°$	$\Phi_5/°$	ΔE_{intra}
1	180	0	0	3.27
5	0	0	0	56.54
33	184	186	0	21.08
41	2	183	360	52.89

Table 17. Summary of how each torsion angle was treated. For those angles defining separate conformational regions, the values at which they were constrained in at least one search are indicated. If the values in various selected conformations fluctuated around some common ones, the average ± the standard deviation is indicated, which were calculated normalizing the angles to a 0-360° range. Angles denoted by * were scanned in the same surrogate molecule.

Label for XXVI	Torsion Angle Definition	Treatment	Values/°
Φ_{1a} and Φ_{1b}	C_{ar}-C_{ar}-C=O	Explicitly flexible torsion angle	120 to 150 and 210 to 340 in 10° steps
Φ_{2a}	O=C-N-C_{ar}	CG constrained torsion angle	0 ± 3
Φ_{3a}	C-N-C_{ar}-C_{ar}	CG constrained torsion angle	66 ± 1, 88, 122, 150 ± 2, 179 ± 2, 210 ± 2, 239 ± 2 , 268
Φ_{2b}	O=C-N-C_{ar}	CG constrained torsion angle	1 ± 7, 178
Φ_{3b}	C-N-C_{ar}-C_{ar}	CG constrained torsion angle	63, 88 ± 1, 119 ± 2, 148 ± 3, 179 ± 1,
Φ_4	C_{ar}-C_{ar}-C_{ar}-C_{ar}	Explicitly flexible torsion angle	-120 to -60 in 20° steps
Label for GSK269984B	**Torsion Angle Definition**	**Treatment**	**Values/°**
Φ_1	O-C-C_{ar}-C_{ar}	Explicitly flexible torsion angle	50 to 310 in 20° steps
Φ_2	C_{ar}-O-C-C_{ar}	CG constrained torsion angle	91 ± 1, 180 ± 1, 240, 271 ± 2, 300
Φ_3	C_{ar}-C_{ar}-O-C	CG constrained torsion angle	73 ± 2, 121 ± 2, 181 ± 1, 240 ±1, 286 ± 1
Φ_4*	C_{ar}-C_{ar}-C-C_{ar}	Explicitly flexible torsion angle	50 to 320 in 30° steps
Φ_5*	C_{ar}-C-C_{ar}-C_{ar}	Explicitly flexible torsion angle	40 to 400 in 30° steps
Φ_6	N_{ar}-C_{ar}-C=O	CG constrained torsion angle	0, 180
Φ_7	C_{ar}-C-O-H	Scanned constrained torsion angle	0, 180
Label for XX	**Torsion Angle Definition**	**Treatment**	**Values/°**
Φ_1*	C_{ar}-C_{ar}-C-O	Explicitly flexible torsion angle	20 to 200 in 30° steps
Φ_2*	C_{ar}-C-O-C	Explicitly flexible torsion angle	60 to 300 in 30° steps
Φ_3	C-O-C-N	CG constrained torsion angle	180
Φ_4	O=C-N-C_{ar}	CG constrained torsion angle	0, 178 ± 2
Φ_5	C-N-C_{ar}-C_{ar}	Explicitly flexible torsion angle	-225 to 55 in 20° steps
Φ_6	C_{ar}-C_{ar}-C_{ar}-N_{ar}	Explicitly flexible torsion angle	-120 to 240 in 20° steps
Φ_7	C_{ar}-C_{ar}-S-C_{ar}	CG constrained torsion angle	119, 180, 210, 240, 270, 300, 330
Φ_8	C_{ar}-S-C_{ar}-C_{ar}	Explicitly flexible torsion angle	20 to 160 in 20° steps
Label for XXIII	**Torsion Angle Definition**	**Treatment**	**Values/°**
Φ_1	C_{ar}-C_{ar}-C-C	Explicitly flexible torsion angle	30 to 390 in 20° steps
Φ_2	C_{ar}-C-C-C_{ar}	CG constrained torsion angle	0, 60, 120, 180, 240, 300
Φ_3	C-C-C_{ar}-C_{ar}	Explicitly flexible torsion angle	30 to 390 in 20° steps
Φ_4	C_{ar}-C_{ar}-N-C_{ar}	Explicitly flexible torsion angle	90 to 450 in 20° steps
Φ_5	C_{ar}-N-C_{ar}-C_{ar}	CG constrained torsion angle	117, 164, 214, 229
Φ_6	N-C_{ar}-C=O	CG constrained torsion angle	0, 180
Φ_7	C_{ar}-C-O-H	Scanned constrained torsion angle	180
Label for Mebendazole A, C	**Torsion Angle Definition**	**Treatment**	**Values/°**
Φ_1*	C_{ar}-C_{ar}-C-C_{ar}	Explicitly flexible torsion angle	90 to 270 in 20° steps
Φ_2*	C_{ar}-C-C_{ar}-C_{ar}	Explicitly flexible torsion angle	0 to 360 in 20° steps
Φ_3	N_{ar}-C_{ar}-N-C	CG constrained torsion angle	182 ± 2
Φ_4	C_{ar}-N-C=O	CG constrained torsion angle	0, 183
Φ_5	O=C-O-C	CG constrained torsion angle	0

Table 18: List of the 59 space groups considered in the searches.

P1	P$\bar{1}$	P2$_1$	P2$_1$/c	P2$_1$2$_1$2	P2$_1$2$_1$2$_1$	Pna2$_1$	Pca2$_1$	Pbca	Pbcn
C2/c	Cc	C2	Pc	Cm	P2$_1$/m	C2/m	P2/c	C222$_1$	Pmn2$_1$
Cmc2$_1$	Aba2	Fdd2	Iba2	Pnna	Pccn	Pbcm	Pnnm	Pmmn	Pnma
Cmcm	Cmca	Fddd	Ibam	P4$_1$	P4$_3$	I$\bar{4}$	P4/n	P4$_2$/n	I4/m
I4$_1$/a	P4$_1$2$_1$2	P4$_3$2$_1$2	P$\bar{4}$2$_1$c	I$\bar{4}$2d	P3$_1$	P3$_2$	R3	P$\bar{3}$	R$\bar{3}$
P3$_1$21	P3$_2$21	R3c	R$\bar{3}$C	P6$_1$	P6$_3$	P6$_3$/m	P2$_1$3	PA$\bar{3}$	

3. RESULTS OF THE WORKFLOW

3.1 Crystal structure searches

3.1.1 Summary of the reproduction of the results of the previous CSP studies

Table 19 to Table 24 identify each significant crystal structure found to be low in energy in a previous CSP study, showing the conformation from Table 11 to Table 16 that generated that structure, the energy and rank after the full optimization (step 3) in the previous CSP study and the ranking after CrystalPredictor in the current search, as well as the change in the 15 molecule coordination sphere that would be needed during final full optimization.

Table 19: Comparison of the crystal structure search with the new method and the previous CSP results for molecule XXVI (Figure 6a in the main paper). The structure highlighted in yellow corresponds to the experimental structure, the one in red was not found with the new method. When RMSD$_{15}$ values are highlighted in blue, it indicates that the structure had been probably found in the search (i.e. RMSD$_{15}$ > 0.8 Å), but the amount of structural change required means that it could optimize to another nearby lattice energy minimum.

Structure name	Found?	Conformation number	Previous CSP ranking	Previous CSP lattice energy/kJ mol^{-1}	New method ranking after search	RMSD$_{15}$
3525	YES	29	1	-206.86	1	0.506
1600	YES	124	2	-206.37	251	0.695
675	YES	124	3	-204.25	61	0.257
38	YES	1	4	-202.71	122	0.597
421	YES	124	5	-201.43	61	0.295
3104	YES	805	6	-201.20	164	0.297
615	YES	41	7	-200.58	72	0.426
239	PROBABLY	1	8	-200.48	233	0.805
2930	YES	29	9	-200.30	51	0.37
354	YES	29	10	-199.98	71	0.377
851	YES	29	11	-199.82	1071	0.442
6460	YES	29	12	-199.74	7	0.666
6335	YES	29	13	-199.41	10	0.579
221	YES	29	14	-199.39	7	0.648
2231	PROBABLY	29	15	-199.29	56	0.809
2496	NO	/	16	-198.93	/	/
185	YES	805	17	-198.75	159	0.325
4201	PROBABLY	41	18	-198.65	670	1.223
314	YES	29	19	-198.63	13	0.406
508	YES	1	20	-198.57	31	0.505
4946	YES	29	21	-198.48	56	0.342
6879	YES	29	22	-198.35	632	0.43
506	YES	29	23	-198.23	20	0.42
4842	YES	29	24	-198.02	39	0.478
43	YES	41	25	-197.84	26	0.467
1236	YES	41	26	-197.71	33	0.387
1537	YES	1	27	-197.69	21	0.43
188	YES	61	28	-197.45	116	0.661
5126	YES	1	29	-196.81	855	0.504
444	YES	25	30	-196.74	42	0.683
544	YES	29	31	-196.57	1071	0.408
686	YES	29	32	-196.52	287	0.406
89	PROBABLY	41	33	-196.42	614	0.841

20	YES	29	34	-196.16	138	0.597
83	YES	29	35	-196.04	3	0.546
2591	YES	805	132	-189.83	1699	0.304

Table 20: Comparison of the crystal structure search with this new method and the previous CSP results for GSK269984B (Figure 6b in the main paper). The structure highlighted in yellow corresponds to the experimental structure, the ones in red were not found with the new method. When RMSD$_{15}$ values are highlighted in blue, it indicates that the structure had been probably found in the search (i.e. RMSD$_{15}$ > 0.8 Å), but the amount of structural change required means that it could optimize to another nearby lattice energy minimum.

Structure name	Found?	Conformation number	Previous CSP ranking	Previous CSP lattice energy/kJ mol^{-1}	New method ranking after search	RMSD$_{15}$
180Intra10	YES	1	1	-180.68	46	0.129
90InterB36	YES	306	2	-180.15	5072	0.626
180InterA11	YES	1_2	3	-178.62	776	0.374
180InterA8	YES	1_2	4	-177.92	225	0.543
180InterB6	YES	11	5	-177.42	354	0.32
180Intra8	YES	1	6	-177.13	31	0.142
180Intra38	PROBABLY	1	7	-177.06	68	1.175
180InterB9	YES	11	8	-176.88	942	0.282
180Intra76	YES	1	9	-176.44	1022	0.486
180InterA22	YES	1_2	10	-176.32	352	0.722
90InterB6	NO	/	11	-176.07	/	/
180Intra19	YES	1	12	-176.02	87	0.211
180Intra74	YES	1	13	-175.68	2217	0.341
180Intra4	YES	1	14	-175.64	38	0.355
180InterA60	YES	1_2	15	-175.60	1104	0.182
180Intra2	YES	1	16	-175.52	1	0.344
180InterA3	NO	/	17	-175.36	/	/
180InterA30	YES	1_2	18	-175.29	1320	0.15
180Intra83	NO	/	19	-175.28	/	/
180Intra56	PROBABLY	1	20	-175.24	53	0.977
180InterA7	YES	1_2	21	-174.87	180	0.298
90Intra31	YES	184	22	-174.80	7969	0.349
180Intra32	YES	1	23	-174.53	117	0.177
180InterA18	PROBABLY	1_2	24	-174.49	325	2.052
180Intra92	YES	1	25	-174.33	53	0.631
180InterA12	YES	1_2	26	-174.21	325	0.505
180InterA29	YES	1_2	27	-174.17	1499	0.172
180InterB10	YES	11	28	-174.15	489	0.507
90InterA14	NO	/	29	-174.06	/	/
180Intra84	PROBABLY	1	30	-174.01	926	1.437
180Intra47	YES	1	31	-173.96	125	0.294
180Intra65	YES	1	32	-173.87	50	0.118
90InterA32	YES	184_2	33	-173.79	14162	0.362
180Intra5	PROBABLY	1	34	-173.74	49	0.905
180Intra28	YES	1	35	-173.60	68	0.241
180InterA26	PROBABLY	1_2	36	-173.52	325	1.884
180InterB87	YES	11	37	-173.50	1749	0.192
180Intra57	YES	1	38	-173.25	114	0.156

Table 21: Comparison of the crystal structure search with this new method and the previous CSP results for molecule XX (Figure 6c in the main paper). The structure highlighted in yellow corresponds to the experimental structure. When RMSD$_{15}$ values are highlighted in blue, it indicates that the structure had been probably found in the search (i.e. RMSD$_{15}$ > 0.8 Å), but the amount of structural change required means that it could optimize to another nearby lattice energy minimum.

Structure name	Found?	Conformation number	Previous CSP ranking	Previous CSP lattice energy/kJ mol^{-1}	New method ranking after search	RMSD$_{15}$
dfAa132	YES	1	1	-218.73	10	0.386
dfAc102	YES	1	2	-217.95	161	0.264
dfAa180	YES	1	3	-216.35	61	0.618
dfAc14	YES	1	5	-213.14	70	0.308
dfAc48	YES	1	10	-212.58	278	0.46
dfAc7	YES	78	12	-211.47	14	0.365
dfAc43	YES	78	14	-211.04	1740	0.778
dfAc17	YES	1	15	-210.87	40	0.449
dfAc172	YES	1	16	-210.76	58	0.572
dfAc29	PROBABLY	1	17	-210.54	2154	1
dfAb181	YES	78	22	-209.62	2428	0.79
dfAd152	PROBABLY	78	23	-209.60	1596	1.012
dfAc86	YES	1	24	-209.37	13	0.532
dfAc67	YES	1	25	-209.25	94	0.167
dfAa277	YES	1	27	-209.03	134	0.546
dfAa4	YES	1	28	-208.97	2	0.376
dfAa1	YES	1	29	-208.95	7	0.376
dfAb161	YES	1	31	-208.86	49	0.26
dfAb1	YES	1	32	-208.83	1	0.123
dfAd79	YES	1	33	-208.80	602	0.444
dfBa28	YES	245	47	-207.29	302	0.65

Table 22: Comparison of the crystal structure search with this new method and the previous CSP results for molecule XXIII (Figure 6d in the main paper). The structures highlighted in yellow correspond to the experimental structures, the ones in red were not found with the new method. When RMSD$_{15}$ values are highlighted in blue, it indicates that the structure had been probably found in the search (i.e. RMSD$_{15}$ > 0.8 Å), but the amount of structural change required means that it could optimize to another nearby lattice energy minimum.

Structure name	Found?	Conformation number	Previous CSP ranking	Previous CSP lattice energy/kJ mol^{-1}	New method ranking after search	RMSD$_{15}$
A1361	YES	1	1	-212.68	11	0.222
A70	YES	1	2	-211.02	10	0.309
A6494	PROBABLY	1	3	-210.55	5375	1.394
A691	YES	1	4	-209.30	46	0.767
A3457	YES	1	5	-209.00	47	0.433
A72	YES	1	6	-208.87	163	0.384
A424	YES	1	7	-208.27	3	0.483
A771	YES	1	8	-208.04	1	0.237
A191	YES	103	9	-207.61	514	0.719
A4890	YES	1	10	-207.22	24	0.746
A5191	NO	/	11	-207.16	/	/
A272	YES	1	12	-207.00	778	0.749
A63	PROBABLY	1	13	-206.63	678	0.894
A118	YES	1	14	-206.55	70	0.449
A75	YES	1	15	-206.39	358	0.485
A1413	YES	1	16	-206.35	13	0.375
A2457	YES	1	17	-206.02	951	0.671
A587	YES	1	18	-205.83	2821	0.422
A2417	YES	1	19	-205.71	399	0.407
A138	PROBABLY	1	20	-205.51	782	1.11
A227	YES	1	21	-205.34	277	0.514
A1949	PROBABLY	1	22	-205.07	497	1.114
A3174	NO	/	23	-204.92	/	/
A2054	NO	/	24	-204.87	/	/
A3023	YES	103	25	-204.83	106	0.481
A2311	YES	1	26	-204.82	5	0.343
A3513	YES	1	27	-204.71	475	0.686
A1109	YES	1	28	-204.69	2	0.447
A894	PROBABLY	1	29	-204.61	1279	1.259
A1422	YES	1	30	-204.53	377	0.488
A1127	YES	1	31	-204.53	7	0.276
A6634	PROBABLY	1	32	-204.34	3394	1.474
A282	YES	1	33	-203.87	3838	0.322
A323	PROBABLY	1	34	-203.83	199	0.807
A2715	YES	1	35	-203.76	2489	0.537
A24995	YES	1	36	-203.70	2983	0.42
A3746	YES	1	37	-203.69	735	0.615
A368	YES	1	38	-203.62	82	0.606
A6738	NO	/	39	-203.61	/	/
A4228	PROBABLY	1	40	-203.60	2304	1.073
A1752	YES	1	41	-203.52	511	0.471
A113	YES	1	42	-203.51	125	0.275
A3750	YES	1	43	-203.49	584	0.256
A505	YES	1	44	-203.41	1262	0.37
A12658	YES	1	45	-203.12	626	0.31
A1918	YES	1	46	-203.04	802	0.757
A1411	PROBABLY	1	47	-202.96	350	0.855

A5145	PROBABLY	1	48	-202.76	262	0.872
A710	YES	1	49	-202.70	155	0.338
B204	PROBABLY	49	66	-201.75	465	1.543
B60	YES	49	83	-201.03	2472	0.427
B184	PROBABLY	49	100	-200.32	846	1.284
Exptal A	YES	103	(167)	-199.08	4218	0.377

Table 23: Comparison of the crystal structure search with this new method and the previous CSP results for mebendazole (Figure 6e in the main paper). The structures highlighted in yellow correspond to the experimental structures. When RMSD$_{15}$ values are highlighted in blue, it indicates that the structure had been probably found in the search (i.e. RMSD$_{15}$ > 0.8 Å), but the amount of structural change required means that it could optimize to another nearby lattice energy minimum.

Structure name	Found?	Conformation number	Previous CSP ranking	Previous CSP lattice energy/kJ mol^{-1}	New method ranking after search	RMSD$_{15}$
A788	YES	A1	1	-182.51	1	0.302
A19	YES	A1	2	-180.35	3	0.160
C27	YES	C1	3	-179.96	111	0.260
C5	YES	C1	4	-179.96	22	0.247
C10	YES	C1	5	-179.88	30	0.265
A50	PROBABLY	A1	6	-179.36	8	1.031
A37	YES	A1	7	-178.43	5	0.357
C23	YES	C1	8	-178.17	66	0.124
C73	YES	C1	9	-178.17	223	0.101
C406	YES	C1	10	-177.83	1313	0.159
A53	YES	A1	11	-177.70	12	0.590
C53	YES	C1	12	-177.03	123	0.242
C25	YES	C1	13	-177.01	60	0.241
A173	YES	A1	14	-176.84	119	0.418
A72	YES	A1	15	-176.76	6	0.312
A49	PROBABLY	A1	16	-176.72	4	1.066
A78	YES	A1	17	-176.58	46	0.310
A90	YES	A1	18	-176.54	23	0.326
A291	YES	A1	19	-176.37	284	0.289
C248	YES	C1	20	-176.33	129	0.210
A306	YES	A1	21	-176.24	87	0.311
C46	YES	C1	22	-176.21	201	0.243
C24	YES	C1	23	-176.15	111	0.266
C115	YES	C1	24	-176.04	105	0.440
C509	PROBABLY	C1	25	-175.90	128	1.857
C583	YES	C1	26	-175.89	360	0.353
A202	YES	A1	27	-175.79	51	0.603
C106	YES	C1	28	-175.78	56	0.235
A143	YES	A1	29	-175.31	178	0.429
A89	YES	A1	30	-175.18	29	0.247
C908	YES	C1	31	-175.08	206	0.513
CCis32	PROBABLY	C33	67	-164.51	1486	0.855

3.1.2 Efficiency data for the new method of defining conformational space.

Table 24: Comparison between the numbers of structures generated in this study for each molecule and in the previous CSP studies.

Molecule	~ structures in previous study/10^6	~ structures with new method/10^6	%difference
XXVI	1	2.2	+120
GSK269984B	0.4	1.9	+375
XX	2.8	2	-29
XXIII	1.9	2.1	+11
Mebendazole	4	0.8	-80

Table 25: **Comparison between the computational costs of generating crystal structures with the new method and the previous CSP studies.**

Molecule	CPU hours previous study	CPU hours new method	% difference
XXVI	27,300	13,355	-51.1
GSK269984B	*Not recorded*	6,462	/
XX	18,000	12,154	-32.5
XXIII	21,000	6,906	-67.1
Mebendazole	6,700	1,787	-73.3

Table 26: Breakdown of the computational cost of the previous CSP study up to crystal structure generation and of the computational cost with the new method for molecule XXVI.

		New method stages	CPU hours
Previous CSP stages	**CPU hours**	Conformers generation and selection	1
Flexibility analysis	11,000	*Ab initio* energy calculation	2,743
Grid Generation	11,000	Grid Generation	122
Crystal structure generation	5,300	Crystal structure generation	10,489
Total	27,300	Total	13,355

Table 27: Breakdown of the computational cost with the new method for GSK269984B. The cost of the previous study was not recorded.

New method stages	CPU hours
Conformers generation and selection	1
Ab initio energy calculation	1,068
Grid generation	174
Crystal structure generation	5,219
Total	6,462

Table 28: Breakdown of the computational cost of the previous CSP study up to crystal structure generation and of the computational cost with the new method for molecule XX.

		New method stages	CPU hours
		Conformers generation and selection	1
Previous CSP stages	**CPU hours**	*Ab initio* energy calculation	372
Grid generation	2,000	Grid generation	115
Crystal structure generation	16,000	Crystal structure generation	11,666
Total	18,000	Total	12,154

Table 29: Breakdown of the computational cost of the previous CSP study up to crystal structure generation and of the computational cost with the new method for molecule XXIII.

Previous CSP stages	CPU hours	New method stages	CPU hours
		Conformers generation and selection	1
Flexibility analysis	2,000	*Ab initio* energy calculation	420
Grid Generation	12,000	Grid generation	45
Crystal structure generation	7,000	Crystal structure generation	6,440
Total	21,000	Total	6,906

Table 30: Breakdown of the computational cost of the previous CSP study up to crystal structure generation and of the computational cost with the new method for both tautomers of mebendazole.

Previous CSP stages	CPU hours	New method stages	CPU hours
		Conformers generation and selection	1
Flexibility analysis	700	*Ab initio* energy calculation	33
Grid generation	200	Grid generation	153
Crystal structure generation	5,800	Crystal structure generation	1,600
Total	6,700	Total	1,787

4. A WORKED EXAMPLE

In this section, an example of how the new CSP methodology was applied to molecule XXIII is shown to aid adaption to different molecules. Each step listed in Section 2.1 in the main paper is shown for XXIII to go from the chemical diagram to a comprehensive spectrum of the possible solid-state crystal structures.
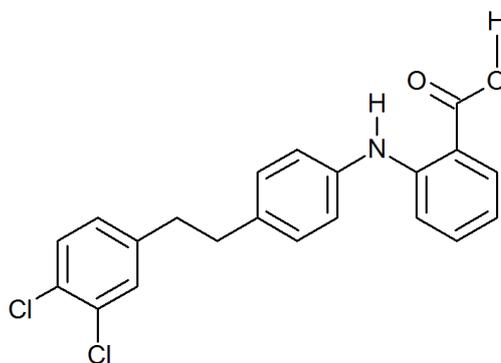
Figure 28: Chemical diagram of molecule XXIII

4.1 Methods

Firstly, Molden was used to draw the molecule and save it as a .mol2 file, which was passed through the CSD Conformer Generator (CG). The default settings for clustering similar conformations (i.e. molecule-based parameters) and for the maximum number of unusual torsion angles (two) were kept, and no limitations placed on the number of the generated conformations or the worst probability scores. This led to the generation of 14,269 individual distinct conformations. The the workflow was applied in the following steps:

1) The .mol2 file used as the CG input was analyzed with the CSD rotamer libraries. This generated a list of rotatable torsion angles (Φ_1-Φ_6, see Figure 7), which did not include Φ_7 because of its terminal hydrogen atom, and their raw CSD distributions. A Python program, using Equation 1, was used to generate plots of the histograms and the Von Mises kernel density approximations of each torsion distribution; they are shown in Figure 11. The RDKit and USRCAT Python packages were used to calculate the effect that each of the six main torsion angles had on the overall shape of the molecule. Each angle was rotated angle by 360° in 30° steps, and at each stage the level of shape match with both the previous step and the initial conformation was calculated via USR. The results are shown in Table 8. Φ_7 was scanned separately, from the fragment shown in Figure 26, using Gaussian 09 at the PBE0 6-31 (d,p) level of theory, from 0° to 360° in 30° steps.

2) Using the information generated in the step 1, the decision tree shown in Figure 4 in the main paper was used applied to the distributions in Figure 11 and the shape matches in Table 8 to decide how to treat angles Φ_1-Φ_6. The results were:

   - Φ_1: more than one maximum; no maximum with $f(\theta) > 0.55$ (see Figure 11a) ➔ **flexible treatment.**

   - Φ_2: the mode has a maximum with $f(\theta) > 0.8$, with the HWHM smaller than 25° ➔ **fixed in the search at CG values.**

   - Φ_3; no maximum with $f(\theta) > 0.55$ (see Figure 11c) ➔ **flexible treatment.**

   - Φ_4: no maximum with $f(\theta) > 0.55$ (see Figure 11d) ➔ **flexible treatment.**

   - Φ_5: the mode has a maximum with $f(\theta) > 0.8$, with the HWHM smaller than 25° ➔ **fixed in the search at CG values.**

   - Φ_6: there are maxima with $f(\theta) > 0.55$, none with $f(\theta) > 0.8$; less than three maxima, one maximum has a HWHM value larger than 25°, the shape match is always higher than 90% for a single step or 85% for the whole molecule ➔ **fixed in the search at CG values.**

3) a) For each angle that was chosen to be constrained at a set of CG values, the separation threshold to select those conformations that describe separate conformational regions was determined using the decision trees shown in Figure 5 in the main paper, using the distributions in Figure 11 and the shape matches in Table 8. The results were:

   - Φ_2: One value much more likely with $f(\theta)$ between 0.8 and 2, the maxima are separated by more than 90°, the HWHM at the mode is smaller than 25°, does have a shape match smaller than 90% with the starting conformer and with the previous steps for some angles changes (see Table 8) ➔ **45° separation threshold.**

- Φ_5: One value much more likely with $f(\theta)$ between 0.8 and 2, the maxima are separated by more than 90°, the HWHM at the mode is smaller than 25°, does have a shape match smaller than 90% with the starting conformer for some angles changes (see Table 8) ➔ **45° separation threshold.**
- Φ_6: No value is particularly favored, since $f(\theta)$ for each maximum is smaller than 0.8, the half width at half maximum of each peak is smaller than 25°, never has a shape match smaller than 85% both for single steps and with the starting conformer (see Table 8), the maxima are separated by more than 45° ➔ **45° separation threshold.**

A Python script was used to extract all the CG generated conformations differing in the values of these three angles by more than each specified threshold; Table 10 shows an example of how this selection worked in practice. 127 distinct conformations were selected. A summary of the results of this selection for XXIII can be seen in Table 14.

b) Three separate surrogate molecules were used to calculate the ΔE_{intra} grids for angles Φ_1, Φ_3 and Φ_4, which are shown in Figure 21. ΔE_{intra} was calculated with Gaussian 09 at the PBE0 6-31 (d,p) level of theory. The ranges were derived from the Mogul distributions shown in Figure 16, and are shown in Table 17.

c) The results of the *ab initio* scan of Φ_7 are shown in Figure 27. It indicates that this angle should be constrained at 180° in every conformational region. This value was added to each of the 127 selected conformations.

4) The relative intramolecular energies (ΔE_{Intra}^{CR}) of the 127 conformational regions were calculated via optimizations with Gaussian 09 at the PBE0 6-31 (d,p) level of theory, constraining Φ_2, Φ_5, Φ_6, and Φ_7 at the values determined in the previous steps and starting the optimizations with the values of Φ_1, Φ_3, and Φ_4 at their minima on the grids. A summary of the calculated energies can be found in Table 14.

5) 16 conformations with calculated ΔE_{Intra}^{CR} values smaller than 26 kJ/mol were used as starting points for partially flexible searches with CrystalPredictor 1.8. Only Φ_1, Φ_3, and Φ_4 were allowed to vary during the search. For 3 conformations, with calculated ΔE_{Intra}^{CR} values smaller than 4 kJ/mol, a maximum of 300,000 crystal structure minimizations were performed, for 6 conformations with ΔE_{Intra}^{CR} values between 4 and 17 kJ/mol a maximum of 150,000, and for 7 conformations with ΔE_{Intra}^{CR} values between 17 and 26 kJ/mol a maximum of 50,000. For some searches, the actual number of minimizations was lower than this maximum because the structures being generated were too high in energy. A total of approximately 2.1 million structures were generated, slightly more than in the blind test, where only 1.9 million minimizations had been performed within CrystalPredictor.

4.2 Results

An analysis performed with the Crystal Packing Similarity Tool, available through the CSD Python API, of the unique individual crystal structures with lattice energies, as calculated by CrystalPredictor, within 40 kJ/mol of the global minimum, revealed that only four out of the 53 crystal structures considered as 'significant' were not found with this method, which means there was a success rate of approximately 93%. In most cases, the matches were of good quality as exemplified by Figure 29: 15-molecule overlay between the experimental crystal structure of molecule XXIII form a (colored by elements) and the 4218th structure after CrystalPredictor with the new workflow (in green). The $RMSD_{15}$ calculated with the Crystal Packing Similarity tool is 0.791 Å..

All three Z' = 1 experimental forms were found, as well as the as yet unobserved global minimum in the crystal energy landscape of the fully optimized crystal structures calculated in the blind test. They all had calculated $RMSD_{15}$ values smaller than 0.8 Å, hence they were deemed to have been 'certainly found'. Despite the cheap energy model used by CrystalPredictor, these four structures were all found within 25 kJ/mol of the global minimum, so they would have been selected for further refinement in a full CSP study.
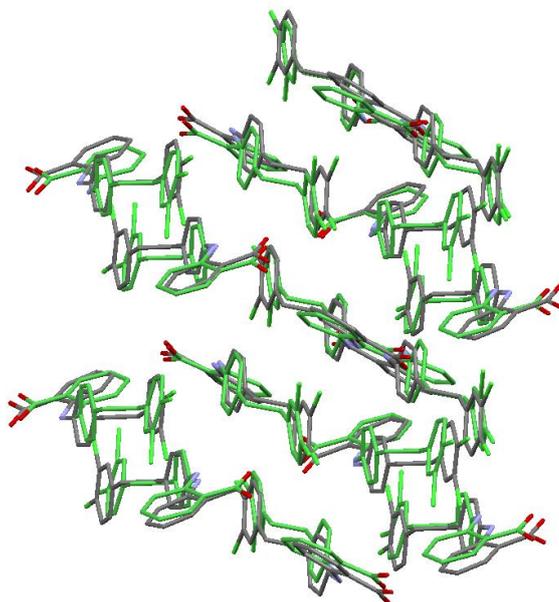

Figure 29: 15-molecule overlay between the experimental crystal structure of molecule XXIII form a (colored by elements) and the 4218th structure after CrystalPredictor with the new workflow (in green). The $RMSD_{15}$ calculated with the Crystal Packing Similarity tool is 0.791 Å.
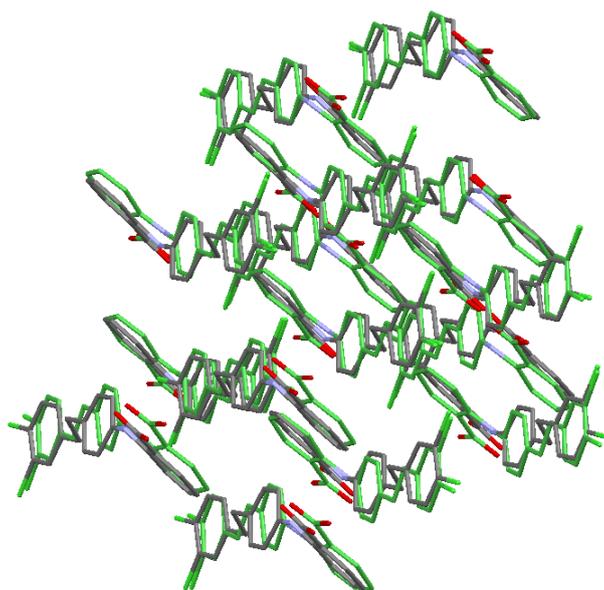
Figure 30: 15-molecule overlay between the experimental crystal structure of molecule XXIII form b (colored by elements) and the 10th structure after CrystalPredictor with the new workflow (in green). The RMSD$_{15}$ calculated with the Crystal Packing Similarity tool is 0.309 Å.
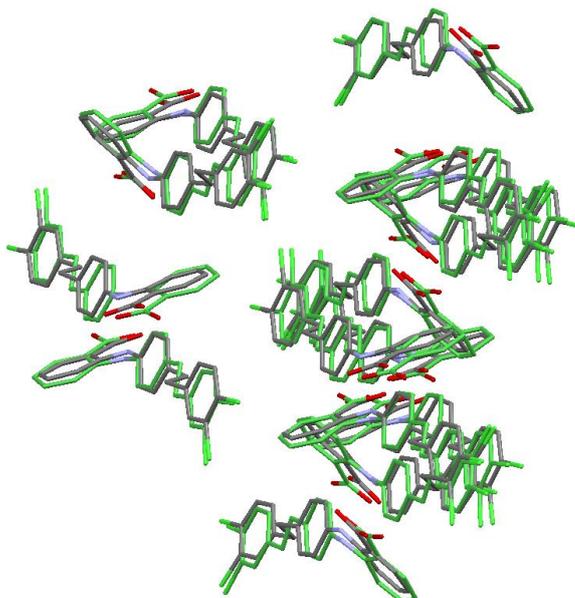

Figure 31: 15-molecule overlay between the experimental crystal structure of molecule XXIII form d (colored by elements) and the 1262nd structure after CrystalPredictor with the new workflow (in green). The RMSD$_{15}$ calculated with the Crystal Packing Similarity tool is 0.370 Å.
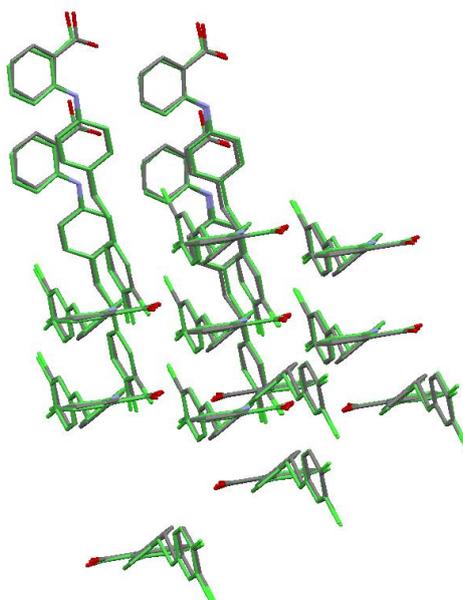
Figure 32: 15-molecule overlay between A1361, the global minimum found in the CSP study of molecule XXIII (colored by elements), and the 11th structure after CrystalPredictor with the new workflow (in green). The RMSD$_{15}$ calculated with the Crystal Packing Similarity tool is 0.222 Å.