

Electronic Supplementary information (ESI). Crystal structure prediction of flexible pharmaceutical-like molecules: density functional tight-binding as an intermediate optimisation method and for free energy estimation.

Luca Iuzzolino, Patrick McCabe, Sarah L. Price and Jan Gerit Brandenburg

Table of Contents

1. APPLICATION OF THE WORKFLOW	2
1.1 Clustering procedure	2
1.2 Energy distributions of the crystal structures at each stage of the workflow	2
1.3 Re-optimisation of some key generated crystal structures with CrystalOptimizer	6
2. RESULTS OF THE WORKFLOW	7
2.1 Crystal energy landscapes	7
2.2 Reproduction of the experimental conformers	9
2.3 Reproduction of some key putative polymorphs (PPMs) from previous CSP studies	10
2.3.1 Correlation between DFTB-B3 energies and $\Psi_{\text{mol}}^{\text{PBE0+FIT}}$ energies for the PPMs	16
2.3.2 Reproduction of the PPMs of molecule XX produced with the RCM method	17
2.3.3 Optimisation of some $Z'=2$ putative crystal structures of molecule XX	17
2.4 Phonon calculations	18
2.5 Computational cost	20

1. APPLICATION OF THE WORKFLOW

1.1 Clustering procedure

The clustering procedure was used to find and remove duplicate crystal structures, to simplify both the calculations and the analysis of the final results. In this work, duplicates were found overlaying structures with the Crystal Packing Similarity tool available through the CSD Python API. Crystal structures were considered as duplicates if it was possible to match 15/15 molecules, with 20% distance and 20° angle tolerances, and if the root mean square deviation (RMSD₁₅) was smaller than 0.5 Å after the DFTB3-D3 optimisations and 1 Å after the DMACRYS optimisations; different thresholds were introduced to avoid removal after the intermediate optimisation of any structures that could have gone to a different E_{latt} minimum after the final stage. If the structures had a different number of molecules in the asymmetric unit (Z'), the RMSD₁₅ threshold was reduced to 0.1 Å, to save isostructural polymorphs with different Z' and in different space groups. When two structures were found to be duplicates, the lower-energy one was kept, and the higher-energy one was removed.

1.2 Energy distributions of the crystal structures at each stage of the workflow

For CrystalPredictor, an E_{latt} cut-off of 40 kJ·mol⁻¹ was imposed, hence only structures up to this level were considered and are shown in the plots. Horizontal lines indicate the energies of the structure(s) matching the experimental one(s) relative to the global minimum in E_{latt} at each stage of the workflow. The plot for molecule XXVI is shown as Figure 2 in the main paper.

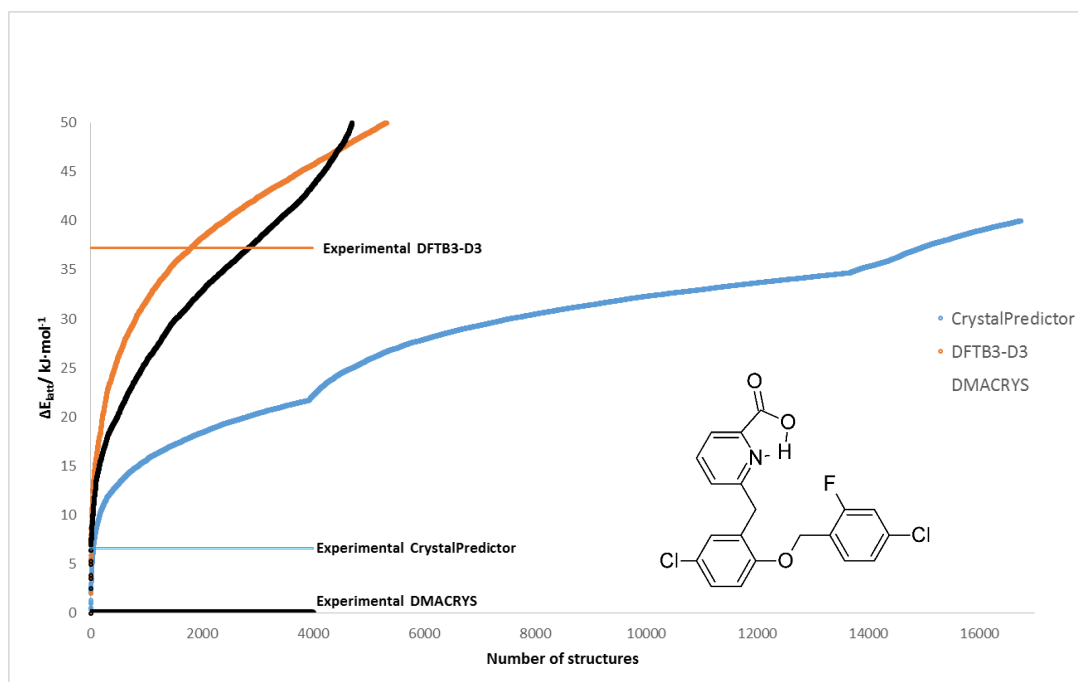


Figure 1: Plot showing the energy distribution of the computer-generated crystal structures of GSK269984B at the various stages of our CSP procedure.

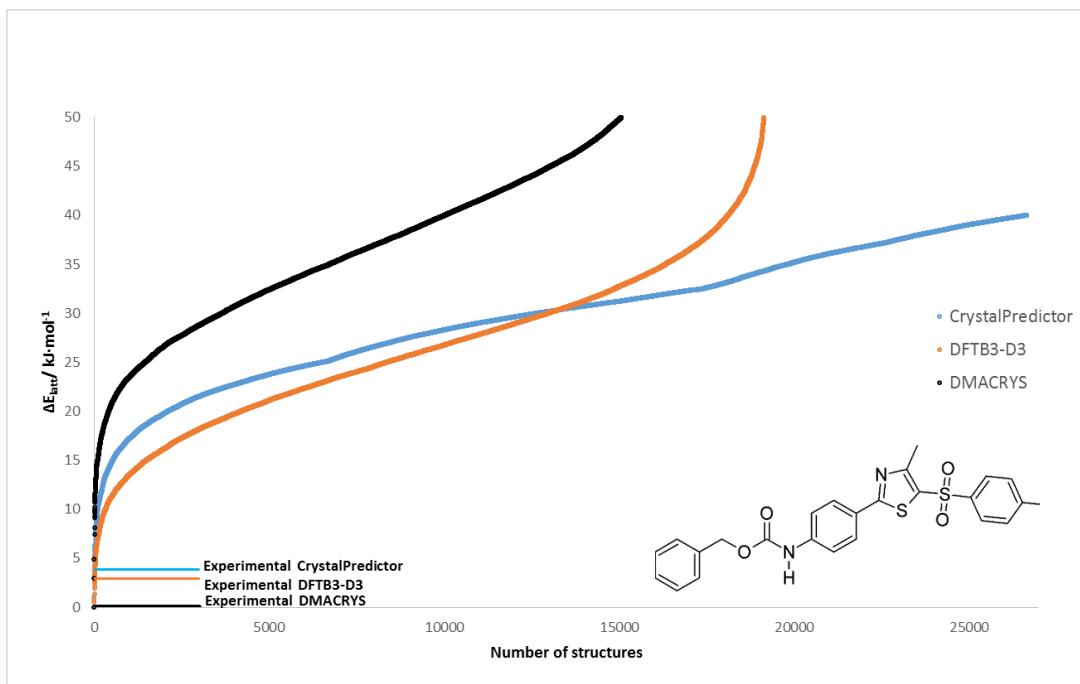


Figure 2: Plot showing the energy distribution of the computer-generated crystal structures of molecule XX at the various stages of our CSP procedure.

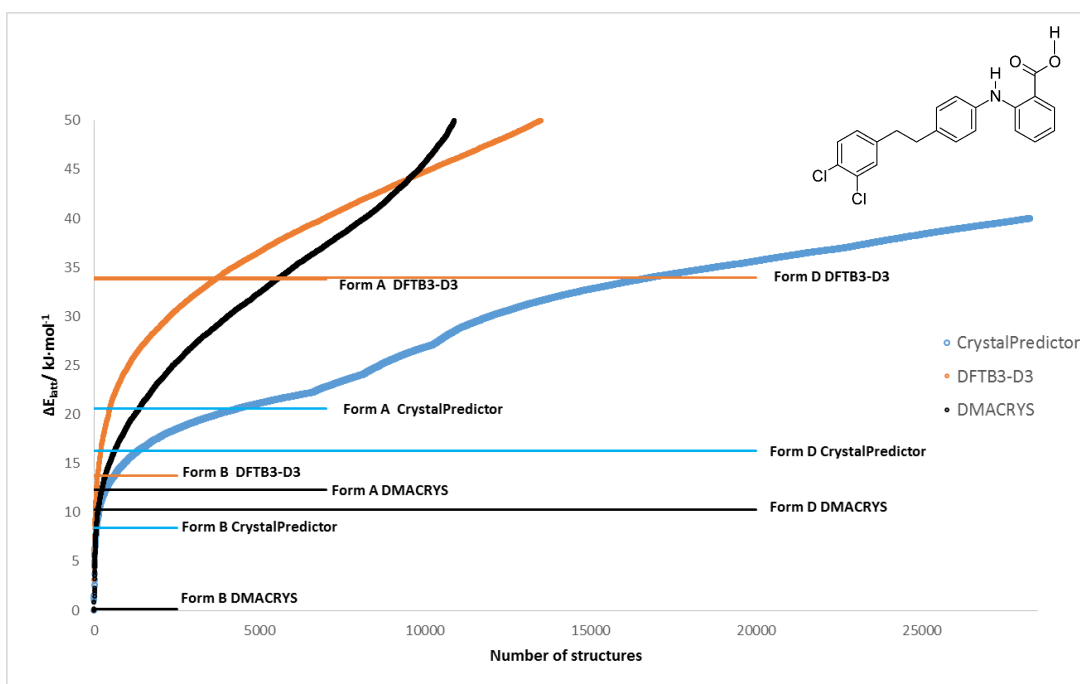


Figure 3: Plot showing the energy distribution of the computer-generated crystal structures of molecule XXIII at the various stages of our CSP procedure.

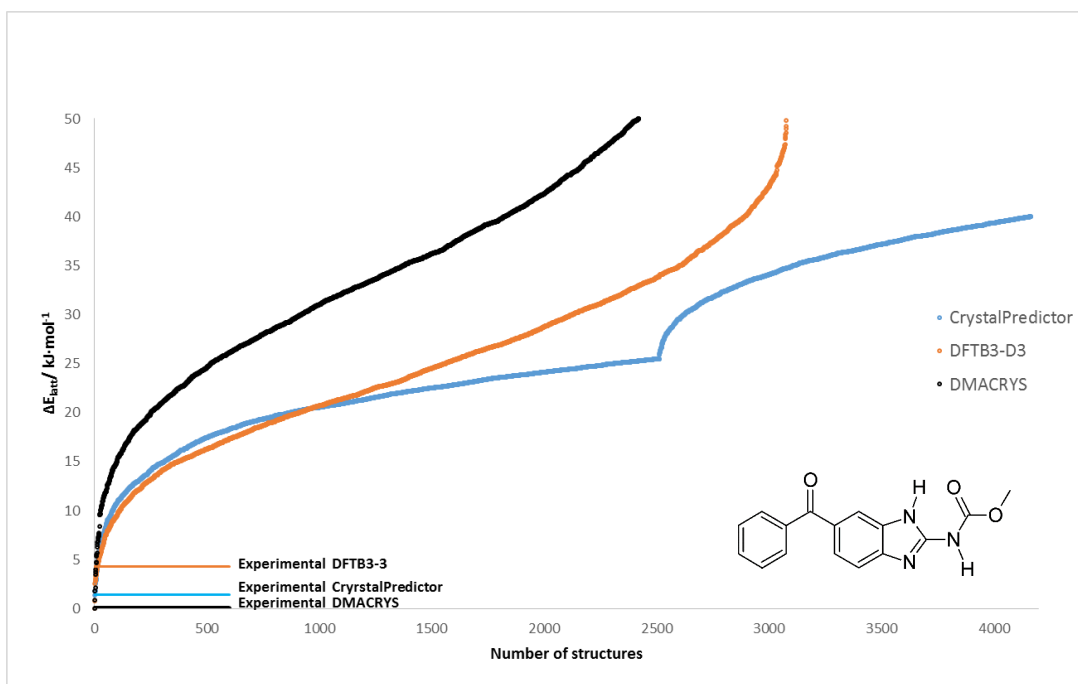


Figure 4: Plot showing the energy distribution of the computer-generated crystal structures of the A-tautomer of mebendazole at the various stages of our CSP procedure.

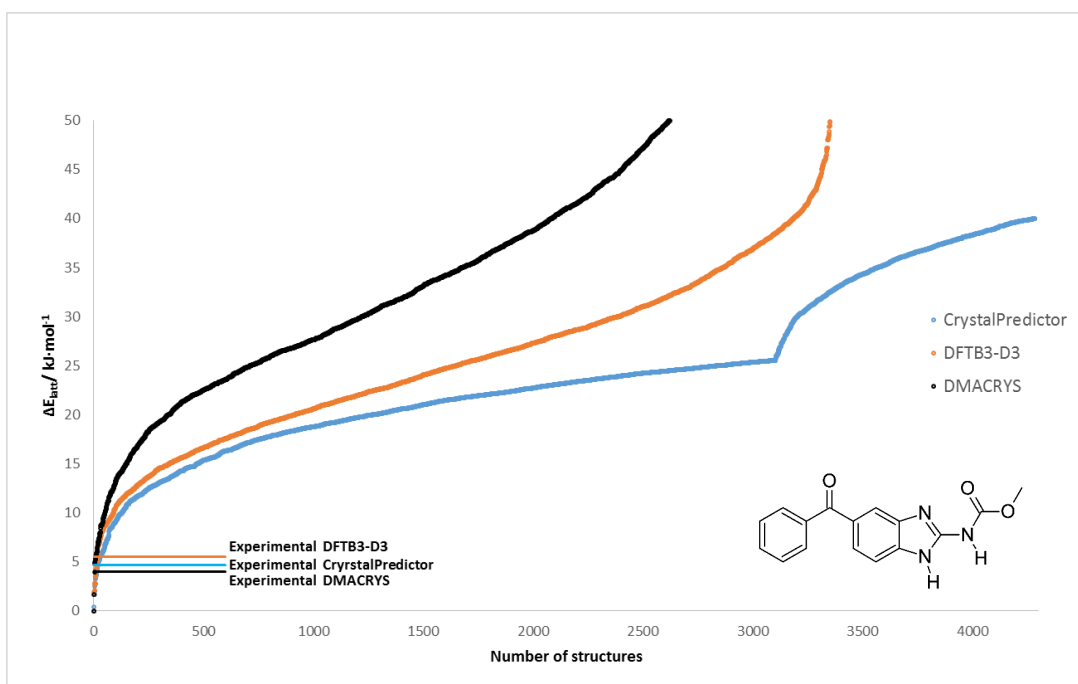


Figure 5: Plot showing the energy distribution of the computer-generated crystal structures of the C-tautomer of mebendazole at the various stages of our CSP procedure.

Table 1 shows how the DFTB3-D3 step reduced the number of structures to be considered, breaking down the various components of this reduction. The results of the clustering procedure described in section 1.1 are also broken down to differentiate between crystal structures that would have already been considered as duplicates before the DFTB3-D3

optimisation if the same clustering method had been employed and those that only became duplicates after the optimisations.

Table 1: Origin of the reduction in the number of crystal structures after the DFTB3-D3 optimisations.

	XXVI		GSK269984B		XXIII		XX		Mebendazole A		Mebendazole C	
	Number of structures	% of initial	Number of structures	% of initial	Number of structures	% of initial	Number of structures	% of initial	Number of structures	% of initial	Number of structures	% of initial
Original CrystalPredictor structures	9215	100.0	16744	100.0	28249	100.0	26650	100.0	4165	100.0	4284	100.0
Removing structures becoming duplicates after DFTB3-D3	8821	95.7	16569	99.0	27866	98.6	26278	98.6	4147	99.6	4267	99.6
Removing wrong molecules*	8583	93.1	16270	97.1	27866	98.6	26098	97.9	4124	99	4238	98.9
Clustering DFTB3-D3 with looser criteria	7534	81.8	13829	82.6	23682	83.8	19251	72.2	3083	74.0	3362	78.5
Applying the 50 kJ·mol⁻¹ cut-off to DFTB3-D3 energies	3346	36.3	5238	31.3	13490	47.8	19146	71.8	3078	73.9	3352	78.2

*In a few cases the DFTB3-D3 optimisation wrongly changed the covalent bonding of the molecule.

1.3 Re-optimisation of some key generated crystal structures with CrystalOptimizer

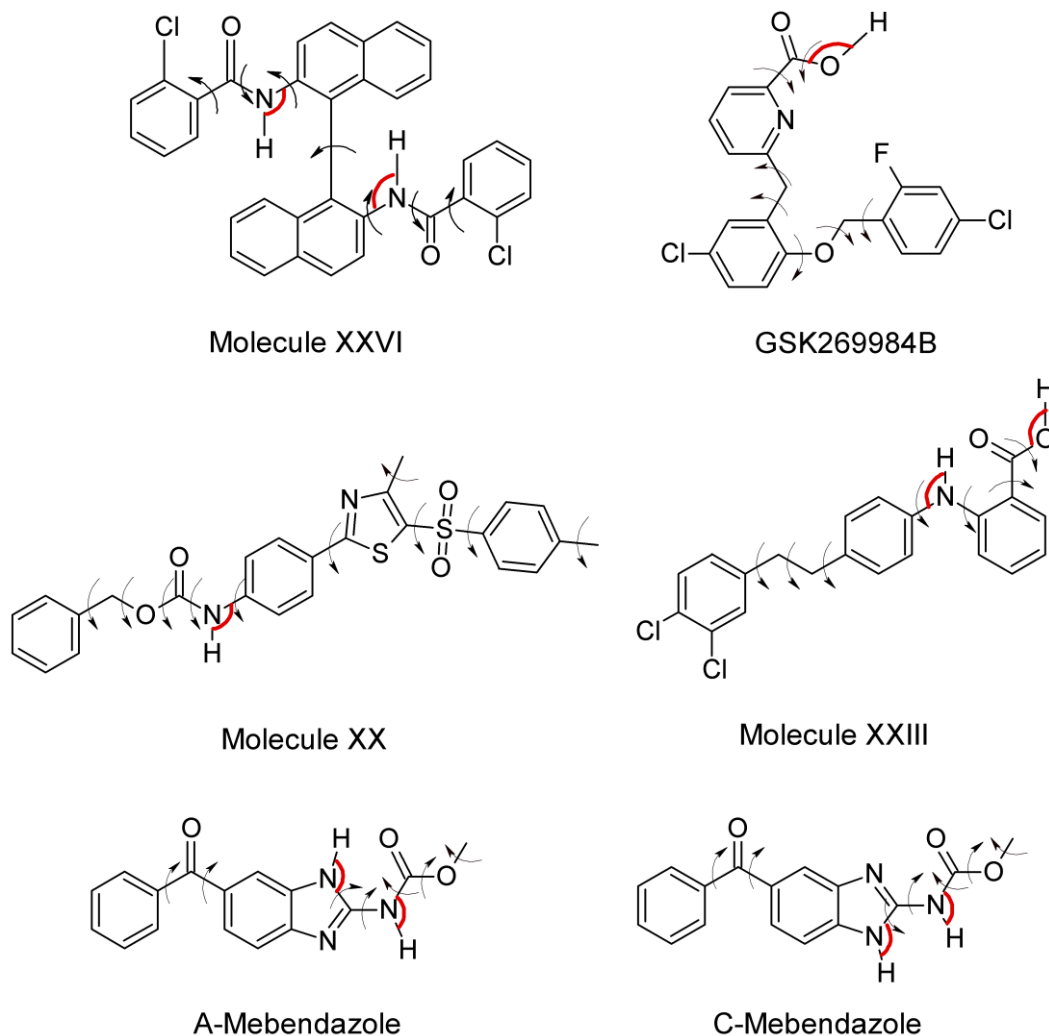
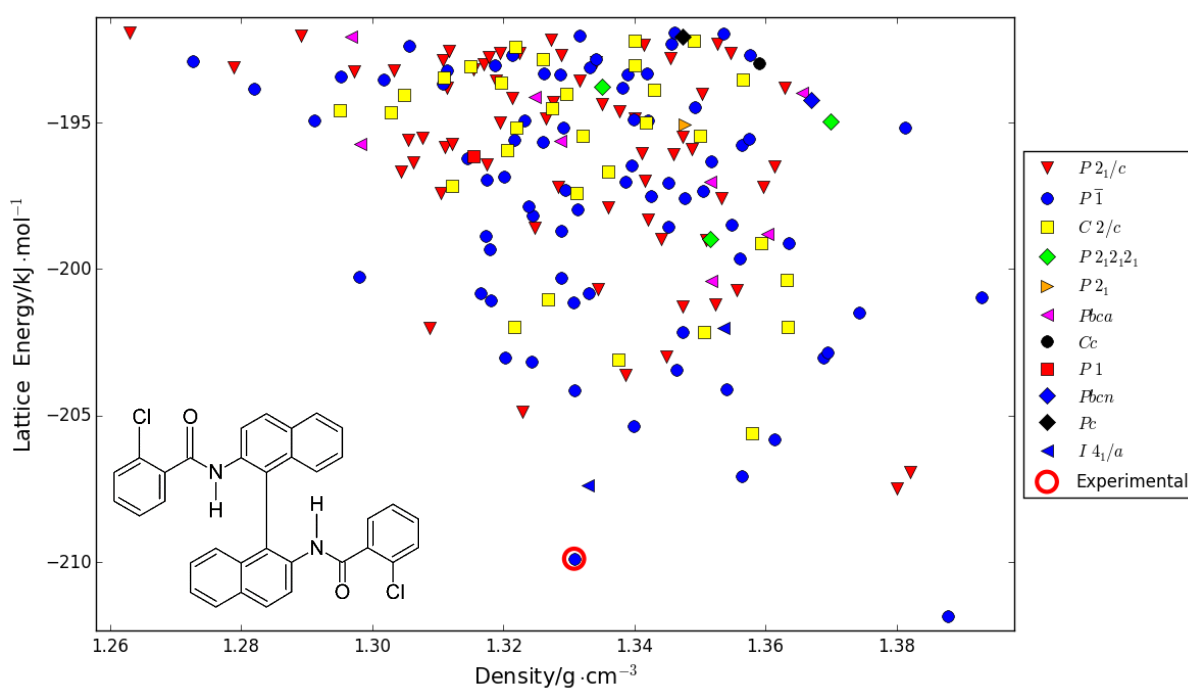


Figure 6: Chemical diagrams of each molecule, with indicated torsion-angles (black arrows) and bond-angles (red arcs) treated as independent degrees of freedom in the CrystalOptimizer optimisations of the key generated crystal structures.

2. RESULTS OF THE WORKFLOW

2.1 Crystal energy landscapes

Each point in the crystal energy landscapes corresponds to a distinct E_{latt} minimum found after the final DMACRYS optimisations. For each minimum, E_{latt} calculated by Gaussian and DMACRYS is plotted against the density. The structure(s) matching the experimental one(s) are also indicated. The crystal energy landscapes of molecules XX and XXIII are shown in Figure 3 in the main paper.



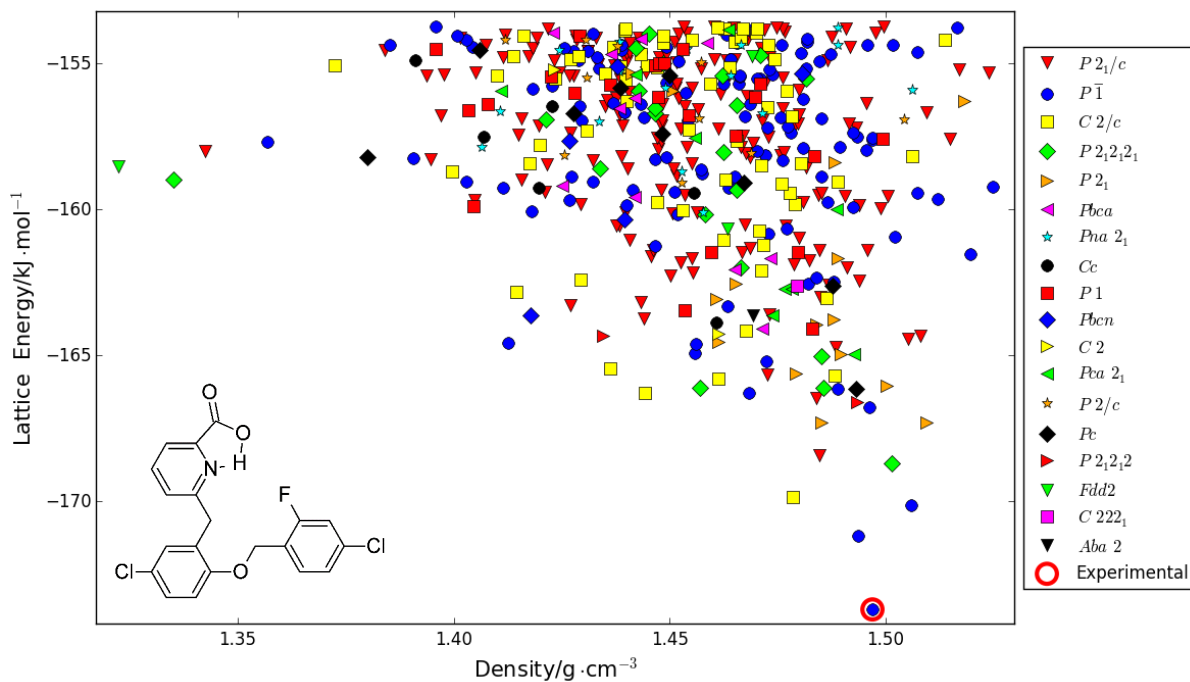


Figure 8: Crystal energy landscape for GSK269984B.

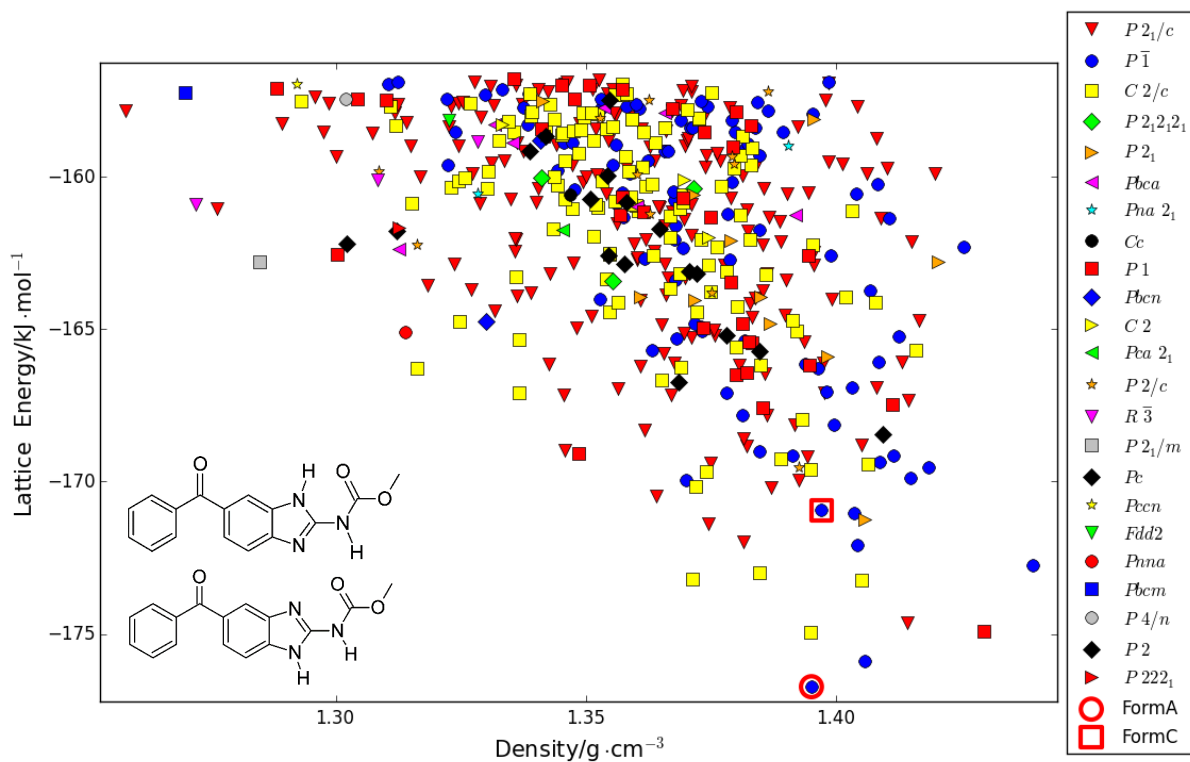
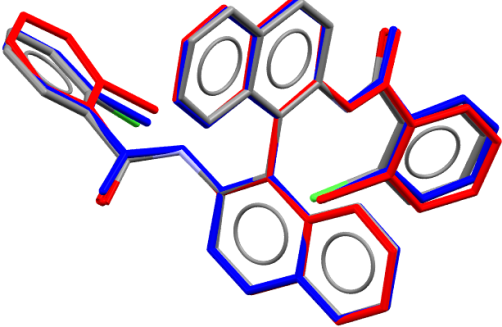
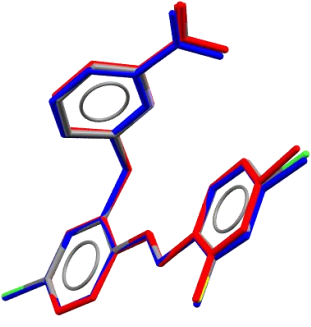
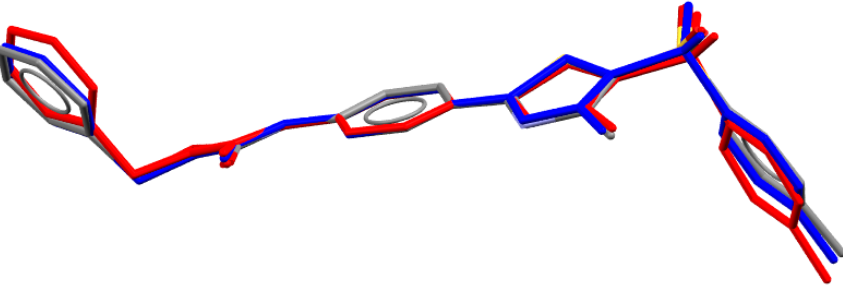
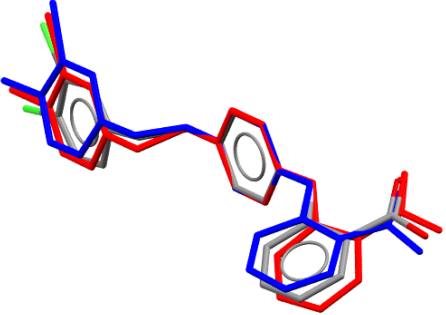
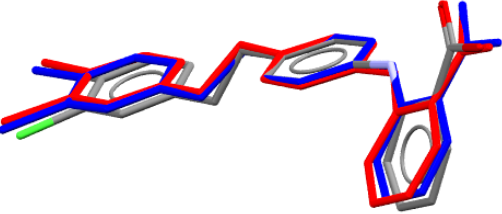
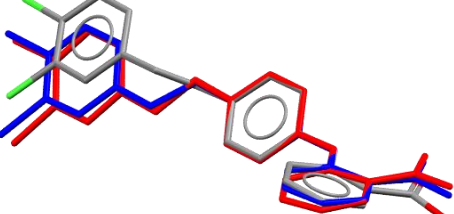
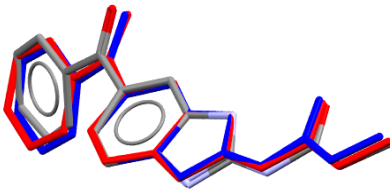
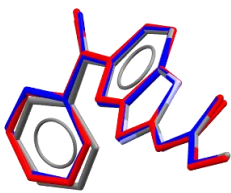


Figure 9: Crystal energy landscape for the two tautomers of mebendazole.

2.2 Reproduction of the experimental conformers

Table 2: For each molecule, overlay of the experimental conformer (coloured by element), the conformer within the CrystalPredictor-generated crystal structure corresponding to it and optimised with DFTB3-D3 (in red), and the conformer within the same structure optimised with CrystalOptimizer (in blue). The RMSD₁ for overlaying the experimental conformers and the optimised ones are also shown. Hydrogen atoms are not shown to facilitate the visualisation.

 <p>Molecule XXVI, DFTB3-D3 RMSD₁ = 0.133 Å, CrystalOptimizer RMSD₁ = 0.123 Å</p>	 <p>GSK269984B, DFTB3-D3 RMSD₁ = 0.071 Å, CrystalOptimizer RMSD₁ = 0.094 Å</p>
 <p>Molecule XX, DFTB3-D3 RMSD₁ = 0.137 Å, CrystalOptimizer RMSD₁ = 0.101 Å</p>	 <p>Molecule XXIII form a, DFTB3-D3 RMSD₁ = 0.216 Å, CrystalOptimizer RMSD₁ = 0.212 Å</p>
 <p>Molecule XXIII form b, DFTB3-D3 RMSD₁ = 0.180 Å, CrystalOptimizer RMSD₁ = 0.159 Å</p>	 <p>Molecule XXIII form d, DFTB3-D3 RMSD₁ = 0.278 Å, CrystalOptimizer RMSD₁ = 0.237 Å</p>
 <p>Mebendazole form A, DFTB3-D3 RMSD₁ = 0.176 Å, CrystalOptimizer RMSD₁ = 0.143 Å</p>	 <p>Mebendazole form C, DFTB3-D3 RMSD₁ = 0.067 Å, CrystalOptimizer RMSD₁ = 0.066 Å</p>

2.3 Reproduction of some key putative polymorphs (PPMs) from previous CSP studies

Tables 3-7 show the results of the analysis that was performed to determine whether the set of crystal structures optimised with our methodology contained each of the PPMs found in the previous CSP studies, and used in a previous work to test a new workflow for crystal structure generation. Note that for these comparisons, the structures from this study were overlaid with the optimised computer-generated ones from the previous studies, with the exception of form A of XXIII, which was not found in the original study. This explains some of the differences between the data shown here and in Tables 1 and 2 in the main paper. As a further comparison, the relative energies and geometries of the structures matching the PPMs (see section 2.3 in the main paper for details) were also compared with those found by re-optimising the same crystal structures, starting from the CrystalPredictor minima, with CrystalOptimizer at the PBE0 6-31G(d,p) level of theory, using the independent degrees of freedom shown in Figure 6. For the CrystalOptimizer-optimised structures, ΔE_{intra} was determined relative to the Gaussian09-minimised gas-phase minima of each molecule. The energies are shown relative to the lowest energy one within the CrystalOptimizer-optimised set.

In a few cases, the distance and angle tolerances had to be increased to 30% and 30° respectively to obtain a match between the crystal structures optimised with the $\Psi_{\text{mol}}^{\text{PBE0+FIT}}$ method and the PPMs or the CrystalOptimizer-optimised structures. In those cases, the RMSD₁₅ values for these structures are shown in italics.

Table 3: Reproduction and ranking of the most important PPMs from the original CSP study of molecule XXVI. The structure highlighted in green corresponds to the experimental structure. The structures which were in the original CSP and where a good match was not found in this work are classified: in blue where they were not found by the search, in turquoise where they had a poor match in the search, and in red where structures were missed despite having a good match in the search. The structures in orange were found after $\Psi_{mol}^{PBEO+FIT}$ at $\Delta E_{latt} > 20$ kJ·mol⁻¹.

Structure name	Found ?	Previous CSP ranking	Ranking after $\Psi_{mol}^{PBEO+FIT}$	ΔE_{latt} previous study/kJ·mol ⁻¹	ΔE_{latt} after $\Psi_{mol}^{PBEO+FIT}$ /kJ·mol ⁻¹	PPM and $\Psi_{mol}^{PBEO+FIT}$ optimised rmsd ₁₅	ΔE_{latt} after CrystalOptimizer optimisation/kJ·mol ⁻¹	CrystalOptimizer and $\Psi_{mol}^{PBEO+FIT}$ optimised rmsd ₁₅
3525	YES	1	7	0.00	6.03	0.48	0.23	0.447
1600	YES	2	2	0.49	1.96	0.528	0.00	0.541
675	YES	3	10	2.60	6.97	0.284	3.45	0.287
38	YES	4	4	4.15	4.45	0.195	5.17	0.209
421	YES	5	10	5.43	6.97	0.286	3.43	0.284
3104	YES	6	33	5.65	10.90	0.178	6.63	0.26
615	YES	7	91	6.28	16.12	0.37	6.79	0.271
239	YES	8	29	6.38	10.63	0.389	6.94	0.425
2930	YES	9	161	6.56	18.99	0.421	3.70	0.348
354	YES	10	147	6.88	18.58	0.398	8.31	0.372
851	YES (high energy)	11	272	7.04	22.54	0.359	7.43	0.366
6460	YES	12	31	7.11	10.79	0.931	8.26	0.577
6335	YES	13	100	7.45	16.41	0.452	8.15	0.379
221	YES	14	31	7.46	10.79	0.483	8.29	0.569
2231	YES	15	5	7.57	4.77	0.33	6.32	0.52
2496	NOT IN SEARCH	16	\	7.93	\	\	\	\
185	YES (high energy)	17	369	8.10	24.61	0.621	9.05	0.643
4201	POOR IN SEARCH	18	\	8.21	\	\	\	\
314	YES	19	23	8.22	9.83	0.322	10.55	0.334
508	YES	20	132	8.29	18.02	0.298	7.16	0.336
4946	YES	21	165	19.14	16.30	0.475	5.09	0.66
6879	YES	22	115	8.51	17.18	0.229	10.33	0.191
506	YES	23	59	8.62	13.98	0.481	10.05	0.421
4842	YES	24	46	8.83	12.85	0.645	7.03	0.591
43	YES	25	3	9.02	4.35	0.439	9.03	0.559
1236	YES	26	8	9.15	6.24	0.318	9.48	0.28
1537	YES	27	12	9.16	7.76	0.334	9.73	0.39
188	YES	28	14	9.41	8.39	0.942	9.47	0.925
5126	YES	29	180	10.05	19.77	0.513	10.54	0.428
444	YES	30	6	10.12	4.92	0.137	9.64	0.152
544	YES (high energy)	31	272	10.28	22.54	0.377	7.44	0.366
686	YES	32	25	10.34	9.86	0.211	10.78	0.166
89	POOR IN SEARCH	33	\	10.44	\	\	\	\
20	YES	34	70	10.69	14.78	0.932	\	no overlay
83	YES	35	88	10.82	15.93	0.398	10.74	0.34
2591	NO	132	\	17.03	\	\	\	\

Table 4: Reproduction and ranking of the most important PPMs from the original CSP study of GSK269984B. The structure highlighted in green corresponds to the experimental structure. The structures which were in the original CSP and where a good match was not found in this work are classified: in blue where they were not found by the search, in turquoise where they had a poor match in the search, and in red where structures were missed despite having a good match in the search. The structures in orange were found after $\Psi_{mol}^{PBEO+FIT}$ at $\Delta E_{latt} > 20$ kJ·mol⁻¹.

Structure name	Found ?	Previous CSP ranking	Ranking after $\Psi_{mol}^{PBEO+FIT}$	ΔE_{latt} previous CSP study/kJ·mol ⁻¹	ΔE_{latt} after $\Psi_{mol}^{PBEO+FIT}$ /kJ·mol ⁻¹	PPM and $\Psi_{mol}^{PBEO+FIT}$ optimised rmsd ₁₅	ΔE_{latt} after CrystalOptimizer optimisation/kJ·mol ⁻¹	CrystalOptimizer and $\Psi_{mol}^{PBEO+FIT}$ optimised rmsd ₁₅
180Intra10	YES	1	1	0.00	0.00	0.135	0.00	0.145
90InterB36	YES (high energy)	2	684	0.53	22.50	0.525	17.13	0.416
180InterA11	YES (high energy)	3	543	2.06	20.84	0.407	16.72	0.184
180InterA8	YES	4	17	2.76	7.59	0.581	10.83	0.223
180InterB6	YES (high energy)	5	820	3.26	23.91	0.34	19.79	0.204
180Intra8	YES	6	179	3.55	15.50	0.375	8.56	0.362
180Intra38	POOR IN SEARCH	7	\	3.62	\	\	\	\
180InterB9	YES	8	278	3.80	17.36	0.394	8.58	0.383
180Intra76	YES	9	98	4.24	13.15	0.47	8.51	0.494
180InterA22	YES	10	269	4.36	17.24	0.726	17.97	0.461
90InterB6	NOT IN SEARCH	11	\	4.61	\	\	\	\
180Intra19	YES	12	23	4.66	8.28	0.189	10.27	0.2
180Intra74	YES	13	34	5.00	9.38	0.366	9.42	0.373
180Intra4	YES	14	42	5.04	9.93	0.385	9.26	0.364
180InterA60	YES	15	404	5.08	19.19	0.161	16.73	0.135
180Intra2	YES	16	6	5.16	5.28	0.204	6.85	0.214
180InterA3	NOT IN SEARCH	17	\	5.32	\	\	\	\
180InterA30	YES	18	329	5.39	18.30	0.094	17.62	0.14
180Intra83	NOT IN SEARCH	19	\	5.40	\	\	\	\
180Intra56	POOR IN SEARCH	20	\	5.44	\	\	\	\
180InterA7	YES	21	417	5.81	19.34	0.428	17.44	0.497
90Intra31	YES	22	464	5.88	19.81	0.368	16.86	0.39
180Intra32	YES	23	169	6.15	15.25	0.18	12.68	0.199
180InterA18	POOR IN SEARCH	24	\	6.19	\	\	\	\
180Intra92	YES	25	49	6.35	10.40	0.932	8.34	0.827
180InterA12	YES	26	483	6.47	19.97	0.355	18.20	0.265
180InterA29	YES (high energy)	27	765	6.51	23.34	0.298	19.19	0.291
180InterB10	YES (high energy)	28	1418	6.53	29.12	0.471	21.50	0.465
90InterA14	NOT IN SEARCH	29	\	6.62	\	\	\	\
180Intra84	YES	30	61	6.67	11.15	0.258	8.54	no overlay
180Intra47	YES	31	77	6.72	12.05	0.162	11.28	0.192
180Intra65	YES	32	28	6.81	8.78	0.116	12.00	0.15
90InterA32	YES (high energy)	33	1412	6.89	29.09	0.26	24.66	0.277
180Intra5	YES	34	56	6.94	10.99	0.329	11.61	0.315
180Intra28	YES	35	21	7.08	8.06	0.208	11.36	0.203
180Intra26	YES	36	483	7.16	19.97	1.441	18.28	0.295
180InterB87	YES (high energy)	37	1051	7.18	26.08	0.265	20.29	0.306
180Intra57	YES	38	24	7.43	8.50	0.764	12.21	0.749

Table 5: Reproduction and ranking of the most important PPMs from the original CSP study of molecule XX. The structure highlighted in green corresponds to the experimental structure. The structures in red were missed despite having a good match in the search. The structures in orange were found after $\Psi_{mol}^{PBE0+FIT}$ at $\Delta E_{latt} > 20 \text{ kJ}\cdot\text{mol}^{-1}$.

Structure name	Found ?	Previous CSP ranking	Ranking after $\Psi_{mol}^{PBE0+FIT}$	ΔE_{latt} previous CSP study/ $\text{kJ}\cdot\text{mol}^{-1}$	ΔE_{latt} after $\Psi_{mol}^{PBE0+FIT}$ / $\text{kJ}\cdot\text{mol}^{-1}$	PPM and $\Psi_{mol}^{PBE0+FIT}$ optimised rmsd ₁₅	ΔE_{latt} after CrystalOptimizer optimisation/ $\text{kJ}\cdot\text{mol}^{-1}$	CrystalOptimizer and $\Psi_{mol}^{PBE0+FIT}$ optimised rmsd ₁₅
dfAa132	YES	1	1	0.00	0.00	0.308	0.00	0.222
dfAc102	YES	2	2	0.78	2.98	0.217	2.16	0.196
dfAa180	YES	3	5	2.38	8.08	0.357	3.01	0.275
dfAc14	YES	5	7	5.59	9.34	0.19	5.37	0.185
dfAc48	YES	10	3	6.15	4.91	0.429	8.83	0.61
dfAc19*	YES	6	6	6.42	9.13	0.658	5.87	0.321
dfAc7	YES	12	14	7.26	9.97	0.186	8.12	0.363
dfAc43	NO	14	\	7.69	\	\	\	\
dfAc17	YES	15	9	7.86	9.72	0.307	10.97	0.582
dfAc172	YES	16	25	7.97	11.53	0.528	7.76	0.634
dfAc29	YES	17	129	8.19	15.70	0.49	6.93	0.454
dfAb181	YES (high energy)	22	1135	9.12	24.00	0.671	10.14	0.654
dfAd152	YES	23	314	9.13	18.96	0.839	8.33	0.859
dfAc86	YES	24	13	9.36	9.92	0.329	8.91	0.399
dfAc67	YES	25	20	9.48	11.14	0.142	9.39	0.163
dfAa277	YES	27	43	9.70	12.85	0.16	9.12	0.163
dfAa4	YES	28	54	9.76	13.50	0.579	10.44	0.129
dfAa1	YES	29	115	9.78	15.51	0.419	11.51	0.248
dfAb161	YES	31	59	9.88	13.74	0.395	9.51	0.342
dfAb1	YES	32	34	9.90	12.41	0.245	10.20	0.192
dfAd79	YES	33	112	9.93	15.43	0.346	8.20	0.328
dfBa28	NO	47	\	11.44	\	\	\	\

*This crystal structure was not considered as a unique PPM in the previous work on the generation of key crystal structures.

Table 6: Reproduction and ranking of the most important PPMs from the original CSP study of molecule XXIII. The structures highlighted in green correspond to the experimental structures. The structures which were in the original CSP and where a good match was not found in this work are classified: in blue where they were not found by the search, in turquoise where they had a poor match in the search, and in red where structures were missed despite having a good match in the search. The structures in orange were found after $\Psi_{mol}^{PBE0+FIT}$ at $\Delta E_{latt} > 20$ $\text{kJ}\cdot\text{mol}^{-1}$.

Structure name	Found ?	Previous CSP ranking	Ranking after $\Psi_{mol}^{PBE0+FIT}$	ΔE_{latt} previous study/ $\text{kJ}\cdot\text{mol}^{-1}$	ΔE_{latt} after $\Psi_{mol}^{PBE0+FIT}$ / $\text{kJ}\cdot\text{mol}^{-1}$	PPM and $\Psi_{mol}^{PBE0+FIT}$ optimised rmsd ₁₅	ΔE_{latt} after CrystalOptimizer optimisation/ $\text{kJ}\cdot\text{mol}^{-1}$	CrystalOptimizer and $\Psi_{mol}^{PBE0+FIT}$ optimised rmsd ₁₅
A1361	YES	1	1	0.00	0.00	0.117	1.27	0.173
A70	YES	2	3	1.66	0.83	0.168	0.00	0.132
A6494	POOR IN SEARCH	3	\	2.13	\	\	\	\
A691	YES	4	5	3.38	1.46	0.293	0.21	0.278
A3457	YES	5	7	3.68	2.04	0.312	1.40	0.212
A72	YES	6	24	3.81	5.47	0.355	2.15	0.383
A424	YES	7	6	4.41	1.81	0.253	1.64	0.253
A771	YES	8	4	4.64	0.85	0.173	1.20	0.234
A191	NO	9	\	5.07	\	\	\	\
A4890	YES	10	66	5.46	8.35	0.703	9.04	0.39
A5191	NOT IN SEARCH	11	\	5.52	\	\	\	\
A272	YES	12	35	5.68	6.53	0.664	6.17	0.545
A63	POOR IN SEARCH	13	\	6.05	\	\	\	\
A118	YES	14	2	6.13	0.17	0.281	2.69	0.248
A75	YES	15	27	6.29	5.73	0.58	2.17	0.548
A1413	YES	16	12	6.33	3.59	0.133	2.28	0.189
A2457	YES	17	34	6.66	6.41	0.594	11.13	0.493
A587	YES	18	59	6.85	7.93	0.358	7.41	0.34
A2417	YES	19	40	6.97	6.91	0.434	8.25	0.392
A138	YES	20	111	7.17	10.02	0.594	5.36	0.497
A227	YES	21	52	7.34	7.64	0.3	5.20	0.281
A1949	YES	22	279	7.61	13.00	1.107	5.34	1.115
A3174	NOT IN SEARCH	23	\	7.76	\	\	\	\
A2054	NOT IN SEARCH	24	\	7.81	\	\	\	\
A3023	YES	25	153	7.85	10.89	0.27	6.96	0.241
A2311	YES	26	216	7.86	12.04	0.233	11.15	0.315
A3513	YES	27	82	7.97	9.09	0.52	8.47	0.171
A1109	YES	28	231	7.99	12.30	0.424	6.74	0.5
A894	POOR IN SEARCH	29	\	8.07	\	\	\	\
A1422	YES	30	68	8.15	8.44	0.603	9.77	0.147
A1127	YES	31	99	8.15	9.74	0.376	7.05	0.38
A6634	POOR IN SEARCH	32	\	8.34	\	\	\	\
A282	YES	33	155	8.81	10.93	0.213	13.68	0.341
A323	YES	34	85	8.85	9.16	0.865	8.96	0.811
A2715	YES	35	141	8.92	10.68	0.226	10.99	0.238
A24995	YES	36	55	8.98	7.79	0.248	7.23	0.141
A3746	NO	37	233	8.99	12.34	0.618	5.26	0.618
A368	YES	38	239	9.06	12.36	0.509	7.76	0.496
A6738	YES	39	1080	9.07	19.28	1.009	11.00	0.723
A4228	YES	40	93	9.08	9.58	0.533	4.59	0.524
A1752	YES	41	13	9.16	3.87	0.227	4.26	0.264
A113	YES	42	31	9.17	6.02	0.315	4.49	0.336
A3750	YES	43	87	9.19	9.21	0.198	12.60	0.197
A505	YES	44	217	9.27	12.06	0.314	11.16	0.316
A12658	YES	45	61	9.56	7.96	0.224	6.78	0.167
A1918	YES	46	37	9.64	6.69	0.797	7.99	0.367

A1411	YES	47	950	9.72	18.49	1.035	20.35	0.21
A5145	YES	48	711	9.92	16.92	0.674	7.48	0.654
A710	YES	49	302	9.98	13.37	0.677	12.31	0.225
B204	YES	66	401	10.93	14.45	0.672	5.39	0.672
B60	YES	83	160	11.65	11.04	0.441	8.80	0.416
B184	YES (high energy)	100	1379	12.36	20.69	0.717	10.22	0.698
Exptal A	YES	(167)	232	13.60	12.31	0.664	10.78	0.601

Table 7: Reproduction and ranking of the most important PPMs from the original CSP study of the two tautomers of mebendazole. The structures highlighted in green correspond to the experimental structures. The structures in orange were found after $\Psi_{mol}^{PBE0+FIT}$ at $\Delta E_{latt} > 20$ $\text{kJ}\cdot\text{mol}^{-1}$.

Structure name	Found ?	Previous CSP ranking	Ranking after $\Psi_{mol}^{PBE0+FIT}$	ΔE_{latt} previous study/ $\text{kJ}\cdot\text{mol}^{-1}$	ΔE_{latt} after $\Psi_{mol}^{PBE0+FIT}$ / $\text{kJ}\cdot\text{mol}^{-1}$	PPM and $\Psi_{mol}^{PBE0+FIT}$ optimised rmsd ₁₅	ΔE_{latt} after CrystalOptimizer optimisation/ $\text{kJ}\cdot\text{mol}^{-1}$	CrystalOptimizer and $\Psi_{mol}^{PBE0+FIT}$ optimised rmsd ₁₅
A788	YES	1	1	0.00	0.00	0.164	0.00	0.126
A19	YES	2	2	2.15	0.85	0.258	\	no overlay
C27	YES	3	3	2.54	1.79	0.128	2.66	0.159
C5	YES	4	15	2.54	5.78	0.142	2.51	0.193
C10	YES	5	6	2.63	3.51	0.103	2.31	0.114
A50	YES	6	10	3.15	4.65	0.795	4.69	0.312
A37	YES	7	5	4.07	2.11	0.234	4.41	0.393
C23	YES	8	36	4.33	7.89	0.255	6.12	0.135
C73	YES	9	51	4.33	9.63	0.219	7.95	0.092
C406	YES	10	93	4.68	11.62	0.172	12.75	0.152
A53	YES	11	11	4.81	4.75	0.347	4.67	0.277
C53	YES	12	35	5.48	7.74	0.121	5.25	0.113
C25	YES	13	18	5.50	6.57	0.097	5.22	0.119
A173	YES	14	126	5.66	12.80	0.333	10.69	0.109
A72	YES	15	101	5.75	11.90	0.214	5.79	0.215
A49	YES	16	17	5.79	6.52	0.88	6.18	0.948
A78	YES	17	12	5.93	5.34	0.097	5.73	0.102
A90	YES	18	7	5.97	3.53	0.193	5.67	0.101
A291	YES	19	23	6.14	7.13	0.273	6.85	0.281
C248	YES	20	29	6.17	7.47	0.166	6.03	0.158
A306	YES	21	91	6.27	11.53	0.413	6.53	0.377
C46	YES	22	25	6.30	7.20	0.106	7.08	0.128
C24	YES	23	22	6.36	7.04	0.228	6.69	0.349
C115	YES	24	37	6.47	7.91	0.32	6.50	0.339
C509	YES	25	72	6.61	10.60	0.374	6.81	0.574
C583	YES	26	448	6.62	19.25	0.304	14.60	0.274
A202	YES	27	20	6.72	6.78	0.764	7.78	0.717
C106	YES	28	53	6.72	9.65	0.506	6.62	0.413
A143	YES	29	28	7.19	7.36	0.19	6.94	0.195
A89	YES	30	40	7.33	8.42	0.17	7.33	0.214
C908	YES	31	43	7.42	8.77	0.296	7.51	0.247
CCis32	YES (high energy)	67	578	18.00	20.86	0.106	18.82	0.214

2.3.1 Correlation between DFTB-B3 energies and $\Psi_{mol}^{PBE0+FIT}$ energies for the PPMs.

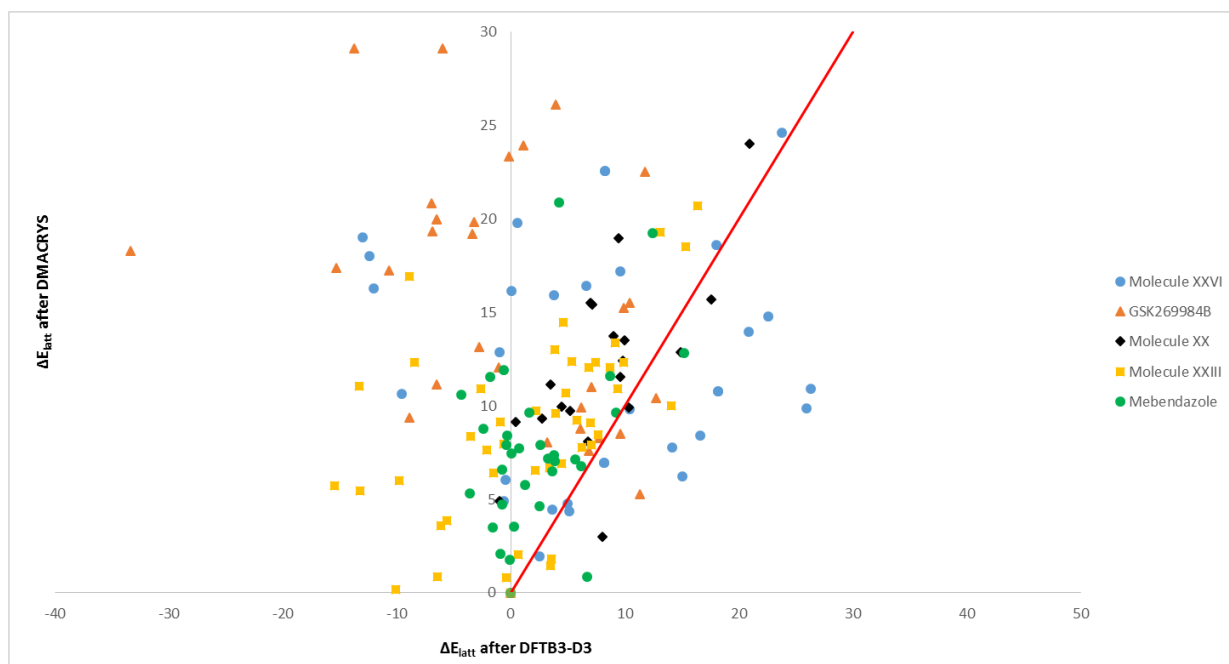


Figure 10: Relative lattice energies of each PPM calculated with DMACRYS using the $\Psi_{mol}^{PBE0+FIT}$ model versus the relative lattice energies calculated with DFTB3-D3. The energies are calculated relative to the structure that was the global minimum in E_{latt} with $\Psi_{mol}^{PBE0+FIT}$. If the relative energies were equal, they would lie on the red line. Note that the selection of structures for optimisation by $\Psi_{mol}^{PBE0+FIT}$ was that their DFTB3-D3 energies relative to the global minimum at DFTB3-D3 level was less than 50 kJ·mol⁻¹.

There is virtually no correlation between the relative energies for these sets of low energy but distinct crystal structures. The lines of best fit are:

$$\Delta E_{latt}(\Psi_{mol}^{PBE0+FIT}) = -0.01 \cdot \Delta E_{latt}(\text{DFTB3-D3}) + 12.43, \text{ with } R^2=0.0006 \text{ for molecule XXVI}$$

$$\Delta E_{latt}(\Psi_{mol}^{PBE0+FIT}) = -0.27 \cdot \Delta E_{latt}(\text{DFTB3-D3}) + 15.28, \text{ with } R^2=0.1459 \text{ for GSK269984B}$$

$$\Delta E_{latt}(\Psi_{mol}^{PBE0+FIT}) = 0.69 \cdot \Delta E_{latt}(\text{DFTB3-D3}) + 6.02, \text{ with } R^2=0.5233 \text{ for molecule XX}$$

$$\Delta E_{latt}(\Psi_{mol}^{PBE0+FIT}) = 0.32 \cdot \Delta E_{latt}(\text{DFTB3-D3}) + 8.26, \text{ with } R^2=0.2448 \text{ for molecule XXIII}$$

$$\Delta E_{latt}(\Psi_{mol}^{PBE0+FIT}) = 0.41 \cdot \Delta E_{latt}(\text{DFTB3-D3}) + 6.76, \text{ with } R^2=0.1639 \text{ for mebendazole}$$

2.3.2 Reproduction of the PPMs of molecule XX produced with the RCM method

Table 8: Reproduction and ranking of the 10 lowest-energy PPMs from the CSP study of molecule XX performed with the RCM method. The structure highlighted in green corresponds to the experimental structure. The structures in orange were found after $\Psi_{mol}^{PBE0+FIT}$ at $\Delta E_{latt} > 20$ kJ·mol⁻¹.

Structure name	Found ?	RCM ranking	Ranking after $\Psi_{mol}^{PBE0+FIT}$	ΔE_{latt} RCM/kJ·mol ⁻¹	ΔE_{latt} after $\Psi_{mol}^{PBE0+FIT}$ /kJ·mol ⁻¹	rmsd ₁₅ between PPM and produced with new method
frm_0001	YES	1	1	0	0.00	0.881
frm_0002	YES (high energy)	2	1752	0.52	25.97	0.644
frm_0003	YES	3	112	1.79	15.43	0.543
frm_0004	YES	4	214	1.79	17.61	0.519
frm_0005	YES	5	34	2.87	12.41	0.595
frm_0006	YES (high energy)	6	831	2.99	22.80	0.627
frm_0007	YES	7	72	3.3	14.51	0.394
frm_0008	YES	8	112	3.91	15.43	0.319
frm_0009	YES	9	275	3.98	18.46	0.665
frm_0010	YES (high energy)	10	1210	4.19	24.26	0.377

2.3.3 Optimisation of some Z'=2 putative crystal structures of molecule XX.

In the 5th Blind Test, Neumann *et al.* performed a Z'=2 search with GRACE and showed that some computer-generated Z'=2 crystal structures were competitive in E_{latt} with the Z'=1 observed form. Table 9 shows that an optimisation with DFTB3-D3 and $\Psi_{mol}^{PBE0+FIT}$ also finds some of those Z'=2 structures competitive with the observed form, with one being even slightly more stable.

Table 9: Relative lattice energies after DFTB3-D3 and $\Psi_{mol}^{PBE0+FIT}$ of the eight lowest energy Z'=2 crystal structures in the refined extended list submitted by Neumann *et al.* for the 5th Blind Test. The energies are calculated relative to the structure matching the experimental form, which is also the E_{latt} global minimum in our study.

Structure name	ΔE_{latt} after $\Psi_{mol}^{PBE0+FIT}$ /kJ·mol ⁻¹
Grace_Z2_1	18.65
Grace_Z2_2	-0.17
Grace_Z2_3	16.21
Grace_Z2_4	19.92
Grace_Z2_5	5.98
Grace_Z2_6	3.50
Grace_Z2_7	4.34
Grace_Z2_8	<i>failed optimisation</i>

2.4 Phonon calculations

Given the high computational cost of the phonon calculations, only a small sample of low-energy crystal structures underwent this step. Structures matching the experimentally-known forms and, if present, the unobserved E_{latt} global minima, were always selected; the rest of the selected crystal structures were representative of competitive hypothetical polymorphs, chosen for their energy competitiveness and/or their packing properties. Note that forms C and E of molecule XXIII were not present in this study, since they are $Z'=2$; hence they were optimised with the same method proposed in this paper starting from the experimental crystal structures and then the free energies were calculated with both methods.

Table 10: Structures selected for phonon calculations for each molecule. For each structure the identifier is indicated, together with its density and E_{latt} after the DMACRYS optimisations with $\Psi_{mol}^{PBE0+FIT}$, the variations in F_{vib} and in the Helmholtz free energies (A), calculated using the rigid body and the DFTB3-D3 models, relative to the global minimum in E_{latt} , and the supercell used to calculate the DFTB3-D3 phonons. Structures corresponding to the experimentally-known forms are outlined in green; as yet unobserved E_{latt} minima are in orange; note that forms C and E of XXIII were not present in the search, since they are $Z'=2$, and were optimised independently for comparison purposes. The first letter in the identifier of the mebendazole crystal structures indicates whether they contained the A or C tautomers.

Molecule XXVI							
Structure	Density/g·cm ⁻³	E_{latt} /kJ·mol ⁻¹	ΔF_{vib} rigid body/kJ·mol ⁻¹	ΔA rigid body/kJ·mol ⁻¹	ΔF_{vib} DFTB3-D3/kJ·mol ⁻¹	ΔA DFTB3-D3/kJ·mol ⁻¹	DFTB3-D3 supercell
C26_863	1.388	-211.84	0.00	0	0.00	0	222
C124_2	1.331	-209.88	-0.31	1.65	-3.36	-1.40	222
C1_134	1.333	-207.39	-0.24	4.21	-3.16	1.29	111
C41_619	1.361	-205.82	-0.87	5.15	-4.78	1.25	222
C805_7	1.393	-200.95	0.24	11.14	-6.25	4.64	222
GSK269984B							
Structure	Density/g·cm ⁻³	E_{latt} /kJ·mol ⁻¹	ΔF_{vib} rigid body/kJ·mol ⁻¹	ΔA rigid body/kJ·mol ⁻¹	ΔF_{vib} DFTB3-D3/kJ·mol ⁻¹	ΔA DFTB3-D3/kJ·mol ⁻¹	DFTB3-D3 supercell
C1_60	1.497	-173.71	0.00	0	0	0	222
C1_240	1.494	-171.17	-1.71	0.83	3.23	5.78	222
C1_1165	1.506	-170.14	-1.72	1.86	2.15	5.72	222
C1_685	1.479	-169.86	0.30	4.15	-3.18	0.67	211
C1_2_1	1.484	-165.01	2.09	10.79	0.01	8.71	141
Molecule XX							
Structure	Density/g·cm ⁻³	E_{latt} /kJ·mol ⁻¹	ΔF_{vib} rigid body/kJ·mol ⁻¹	ΔA rigid body/kJ·mol ⁻¹	ΔF_{vib} DFTB3-D3/kJ·mol ⁻¹	ΔA DFTB3-D3/kJ·mol ⁻¹	DFTB3-D3 supercell
C1_9	1.382	-218.52	0.00	0	0	0	111
C1_60	1.330	-215.54	-1.05	1.93	6.32	8.25	411
C1_250	1.393	-213.61	1.43	6.34	2.86	9.20	121
C78_1191	1.315	-211.12	1.63	9.03	2.37	11.41	121
C78_28	1.347	-210.45	-0.53	7.54	-2.60	4.94	111
C245_91	1.328	-199.71	-0.30	18.5	0.57	19.08	111
Molecule XXIII							
Structure	Density/g·cm ⁻³	E_{latt} /kJ·mol ⁻¹	ΔF_{vib} rigid body/kJ·mol ⁻¹	ΔA rigid body/kJ·mol ⁻¹	ΔF_{vib} DFTB3-D3/kJ·mol ⁻¹	ΔA DFTB3-D3/kJ·mol ⁻¹	DFTB3-D3 supercell
C1_13	1.387	-179.40	0.00	0	0	0	212
C1_60	1.394	-179.22	-0.25	-0.08	-0.95	-1.03	222
C1_43	1.402	-178.57	-0.80	0.03	-0.50	-0.47	231
C103_31	1.410	-178.55	0.70	1.55	0.86	2.41	311
C1_889	1.342	-169.14	-2.46	7.81	-2.79	5.02	221
C49_1002	1.411	-169.09	-0.60	9.71	-1.90	7.81	221
C103_847	1.345	-167.08	-1.15	11.17	0.86	12.02	221
Form C	1.402	-172.28	-1.21	5.91	-1.33	4.58	221
Form E	1.366	-170.88	-1.36	7.16	-2.44	4.72	221
Mebendazole							
Structure	Density/g·cm ⁻³	E_{latt} /kJ·mol ⁻¹	ΔF_{vib} rigid body/kJ·mol ⁻¹	ΔA rigid body/kJ·mol ⁻¹	ΔF_{vib} DFTB3-D3/kJ·mol ⁻¹	ΔA DFTB3-D3/kJ·mol ⁻¹	DFTB3-D3 supercell
A_C1_5	1.395	-176.73	0.00	0	0	0	221
A_C1_47	1.406	-175.87	0.14	1.00	-1.14	-0.14	211
A_C1_6	1.430	-174.91	1.43	3.25	0.77	4.02	111
C_C1_39	1.395	-174.94	0.63	2.42	-3.72	-1.29	111
C_C1_28	1.397	-170.94	0.28	6.06	-1.63	4.43	221
C_C1_2270	1.370	-169.96	-0.29	6.48	-3.38	3.10	211

2.5 Computational cost

Table 11: Breakdown of the computational cost for optimising and re-ranking the CrystalPredictor generated crystal structures for each molecule with the method outlined in this study. The cost of the phonon calculations is not included in the total, to have a more meaningful comparison with the original CSP studies for which only E_{latt} was calculated.

	XXVI	GSK269984B	XXIII	XX	Mebendazole
DFTB Optimisation cost/ hours	10,927	4,829	8,386	12,723	2,215
DFTB clustering cost/ hours	109	156	618	1,306	37
DMACRYS Optimisation cost/ hours	4,426	2,911	6,903	14,674	2,300
Total cost/ hours	15,462	7,896	15,907	28,703	4,552
DFTB3-D3 phonon calculations/hours	1,844	344	984	512	99
Rigid body phonon calculations/hours	5	3	9	6	7

Table 12: Breakdown of the computational cost for optimising and re-ranking the CrystalPredictor generated crystal structures for each molecule in the original CSP studies.

Step	XXVI	GSK269984B	XXIII	XX	Mebendazole
Cost of single iteration of CrystalOptimizer/ hours	138,080	<i>Not recorded</i>	<i>Not performed</i>	<i>Not performed</i>	1,758
Cost of single-point calculation with DMACRYS/ hours	<i>Not performed</i>	<i>Not recorded</i>	24,000	<i>Not performed</i>	<i>Not performed</i>
Cost of full CrystalOptimizer optimisation/ hours	96,299	<i>Not recorded</i>	35,000	100,000	4,847
Total cost/ hours	234,379	<i>Not recorded</i>	59,000	100,000	6,605

Table 13: Comparison between the computational costs needed to optimise and re-rank the search-generated structures with the method outlined here and in the previous CSP studies.

Molecule	CPU hours previous study	CPU hours new method	% difference
XXVI	234,379	15,462	-93
GSK269984B	<i>Not recorded</i>	7,896	/
XX	100,000	28,703	-71
XXIII	59,000	15,907	-73
Mebendazole	6,605	4,552	-31

These values are dependent on the computer hardware used. The current study was performed on a large cluster of Intel Xeon 2.3 GHz processors, which was used for the previous CSP studies only for mebendazole.