STRUCTURAL SCIENCE
CRYSTAL ENGINEERING
MATERIALS

Volume 72 (2016)

Supporting information for article:

Transferable force field for crystal structure predictions, investigation of performance and exploration of different rescoring strategies using DFT-D methods

Anders Broo and Sten O. Nilsson Lill

Table S1 van der waals parameters. Atom type as assigned by the MacroModel atom typing process.

Atom type	r_i (in Å)	ε_i (in kcal/mol)
As	3.80	0.500
B	3.60	0.050
BT	3.60	0.050
Br	3.47	0.470
C356	3.55	0.070
C3M	3.55	0.076
C3T	3.50	0.066
C4	3.75	0.105
C4M	3.55	0.076
C4T	3.50	0.066
C5	3.55	0.070
C5A	3.55	0.070
C5X	3.55	0.070
C55	3.55	0.070
C56A	3.55	0.070
C56B	3.55	0.070
C56X	3.55	0.070
C57	3.55	0.070
C5B	3.55	0.070
C5BB	3.55	0.070
C5BC	3.55	0.076
C5BD	3.55	0.070
C5M	3.55	0.070
C5N	3.55	0.070
C7	3.55	0.070

C	3.75	0.105
CA	3.55	0.070
CA5	3.55	0.070
CAA	3.55	0.070
CAB	3.55	0.070
CB	3.55	0.070
CD	3.75	0.105
CDS	3.75	0.105
CDX	3.75	0.105
CDZ	3.75	0.105
CG	3.75	0.105
CHZ	3.75	0.105
CIN	3.65	0.150
CM	3.55	0.076
CMA	3.55	0.076
CN56	3.55	0.070
CN	3.55	0.070
CNA	3.50	0.080
CO	3.50	0.066
CO1	3.50	0.066
CO3	3.75	0.105
CO4	3.75	0.105
CP1	3.50	0.066
CQ	3.55	0.070
CR	3.55	0.070
CR3	3.55	0.070
CRA	3.55	0.070
CRAB	3.50	0.080
CRB	3.55	0.070

CRP	3.55	0.070
CT	3.50	0.066
CT1	3.50	0.066
CT2	3.50	0.066
CT3	3.50	0.066
CTA	3.50	0.066
CTC	3.50	0.066
CTD	3.50	0.066
CTE	3.50	0.066
CTF	3.50	0.066
CTG	3.50	0.066
CTH	3.50	0.066
CTI	3.50	0.066
CTJ	3.50	0.066
CTL	3.50	0.066
CTHO	3.50	0.066
CTO	3.50	0.066
CTP	3.50	0.066
CTQ	3.50	0.066
CTR	3.75	0.105
CTS	3.50	0.066
CTU	3.50	0.066
CZ	3.30	0.066
CZA	3.65	0.150
Cl	3.40	0.300
Cm	4.18	0.118
F	2.94	0.061
H	0.50	0.030
HA	2.42	0.030

HC	2.50	0.030
HO	0.50	0.030
HP	2.50	0.030
HS	0.50	0.030
HSi	2.50	0.030
HWS	0.50	0.030
I	3.75	0.600
N*	3.25	0.170
N3T	3.30	0.170
N4	3.25	0.170
N4T	3.30	0.170
N55	3.25	0.170
N56P	3.25	0.170
N5A	3.28	0.170
N5B	3.28	0.170
N5M	3.42	0.170
N5P	3.25	0.170
N	3.25	0.170
NA	3.25	0.170
NA5	3.25	0.170
NA5B	3.25	0.170
NAA	3.28	0.170
NAB	3.28	0.170
NAP	3.24	0.170
NAX	3.24	0.120
NB	3.25	0.170
NC	3.25	0.170
NC56	3.25	0.170
NE	3.30	0.170

NG	3.25	0.170
NH	3.28	0.170
NHZ	3.30	0.170
NI	3.28	0.170
NIN	3.18	0.250
NIP	3.25	0.170
NIX	3.25	0.120
NM	3.42	0.170
NMX	3.42	0.170
NN	3.25	0.170
NNA	3.25	0.170
NNB	3.25	0.170
NNC	3.25	0.170
NND	3.25	0.170
NNE	3.25	0.170
NNF	3.25	0.170
NO	3.25	0.120
NO2	3.25	0.170
NO3	3.25	0.120
NP	3.25	0.170
NR3	3.25	0.170
NS	3.25	0.170
NST	3.28	0.170
NT	3.30	0.170
NTX	3.25	0.120
NTH	3.30	0.170
NTH0	3.30	0.170
NTH2	3.30	0.170
NX	3.25	0.170

NX2	3.25	0.170
NZ	3.28	0.170
NZA	3.18	0.250
NZC	3.18	0.250
NZP	3.28	0.170
NZT	3.18	0.250
O2Z	2.98	0.200
O3T	2.90	0.140
O4T	2.90	0.140
O	2.96	0.210
OA	2.90	0.140
OC3	2.96	0.210
OH	3.12	0.170
OH5	3.12	0.170
OM	3.15	0.250
ON	2.96	0.170
ONX	2.96	0.170
OS	2.90	0.140
OS1	2.90	0.140
OWS	3.17	0.156
OY	2.96	0.170
OZ	2.70	0.280
P1	3.74	0.200
P2	3.74	0.200
PM	3.74	0.200
PP	3.74	0.200
PR	3.74	0.200
S	3.60	0.355
SA	3.56	0.395

SD	4.10	0.395
SH	3.60	0.425
SM	4.25	0.500
SM2	4.25	0.500
ST	4.25	0.500
SX	4.10	0.395
SX6	4.10	0.395
SY	3.55	0.250
SZ	3.55	0.395
Se	3.65	0.560
Si	3.83	0.402
F9	2.94	0.061
N2	3.25	0.170
Cl-	4.02	0.710
Br-	4.28	0.710

Table S2 Force field parameters for the hydrogen bonding type of interactions

Type (A…D)	Req (Å)	D (kcal/mol)
O…HO	2.70	5.0
N…HO	2.78	6.9
O…HN	2.85	1.9
N…HN	2.85	3.1
F…HO	2.80	1.0
F…HN	2.80	1.0
Cl^-…HO	3.20	13.5
Cl^-…HN	3.10	13.5

Table S3 The energy minimized crystal structures compared to the observed structure in CSD. $RMSD_{15}$ is reported in the table and used to judge the deviation from the observed crystal structure. In cases where a cluster of 15 molecules could not be overlaid with the RMSDx value is reported, where X denotes the size of the cluster that was possible to overlay. 4 different charge models were used together with the COMPASS II force field. FF=force field internal charges, Qeq charges were determined with the Qeq method, Gasteiger=charges were determined with the Gasteiger method, ESP=charges were the same ESP charges that was used with the AZ-FF force field

CSD id	Form	Z´	Compass II FF	Compass Qeq	Compass Gasteiger	Compass ESP	AZ-FF	PBE-TS DND	PBE-TS OTF	PBE-TS NC	PBE-GNC	PBE-NP
SUTHAZ26	IV	1	$RMSD_{11}$ 0.589	$RMSD_{8}$ 0.766	$RMSD_{11}$ 0.633	0.377	0.211	0.189	0.049	0.043	0.093	0.143
SUTHAZ25	III	2	$RMSD_{10}$ 0.635	0.464	$RMSD_{10}$ 0.744	0.399	0.202	0.175	0.074	0.054	0.097	0.109
SUTHAZ24	II	1	$RMSD_{11}$ 0.724	0.428	$RMSD_{9}$ 0.496	$RMSD_{12}$ 0.511	0.186	0.239	0.086	0.086	0.154	0.075
SUTHAZ28	I	2	$RMSD_{7}$ 0.95	$RMSD_{11}$ 0.584	$RMSD_{4}$ 0.834	$RMSD_{9}$ 0.768	$RMSD_{12}$ 0.583	0.261	0.125	0.121	0.175	0.111
SUTHAZ27	V	2	$RMSD_{6}$ 1.028	$RMSD_{4}$ 1.037	$RMSD_{8}$ 0.768	$RMSD_{3}$ 1.126	0.278	0.395	0.107	0.106	0.094	0.161

.

Special cases where original OPLS-aa torsion parameters was modified.

Rule 1: When the Taylor expansion contains 3 non-zero values for V_{ijkl} the largest of the three is selected and P_{ijkl} is set to 3.

Rule 1.1: For torsion between two aromatic ring systems that is described with 3 terms set P_{ijkl} to 2

Rule 2: When the Taylor expansion contains 2 non-zero values for V_{ijkl} the largest of the two is selected and P_{ijkl} is set to the periodicity of the largest term in the expansion.

Rule 3: Aliphatic hydroxyl group; scale V_{ijkl} with a factor of 10.

Rule 4: Carboxylic acid; scale V_{ijkl} with a factor of 4

Rule 5: Morpholino, Piperidine and Piperazine (CT-NE-CA-X); scale V_{ijkl} with a factor of 2.

Rule 6: Aromatic Nitro group (X-CA-NO2-ON); scale V_{ijkl} with a factor of 3.

Rule 7: Aromatic methoxy (CA-CA-OS-CT); scale V_{ijkl} with a factor of 3.

Rule 8: Two aromatic rings with V2 term only (CA-CA-NA-N5A or CA-CA-NA-C56A); scale V_{ijkl} with a factor of 3.

Rule 9: Pyrimidine (X-NB-CQ-NB); scale V_{ijkl} with a factor of 5.

Rule 10: CA-CA-CO4-CA and CA-CA-CO4-O; scale V_{ijkl} with a factor of 10.

Rule 11: X-CHZ-NI-N; scale V_{ijkl} with a factor of 3.

Rule 12: CHZ-NI-N-CT; scale V_{ijkl} with a factor of 5.

Rule 13: Aromatic amide; scale V_{ijkl} with a factor of 4.

Top 20 space groups

*P*21/*c*, *P*-1, *P*212121, *P*21, *C*2/*c*, *Pbca*, *Pnma*, *Pna*21, *Pbcn*, *P*1, *Cc*, *C*2, *Pca*21, *P*21/*m*, *C*2/*m*, *P*21212, *P*2/*c*, *R*-3, *Pc*, *Pccn*

Top 7 chiral space groups

*P*21, *P*212121, *P*1, *C*2, *P*21212, *P*43, *C*2221

Linear correlation equations

Free energy of hydration ΔG_{hyd}

$\Delta G_{hyd}(exp) = -14.45045 + 0.8039476*\Delta G_{hyd}(kJ/mol)$ PBF

Logarithm of intrinsic solubility S_0

LogS exp = -1.558283 + 0.6738871*LogS Norm TS

Free energy of sublimation ΔG_{sub}

ΔG_{sub} $(exp)^b$ kJ/mol = 22.958037 + 0.5241986*ΔG_{sub} (PBE-G)

ΔG_{sub} $(exp)^b$ kJ/mol = 12.630554 + 0.6878778*ΔG_{sub} (PBE-TS).

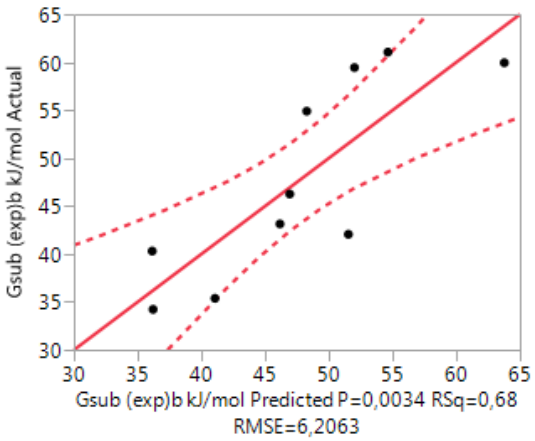


Figure A

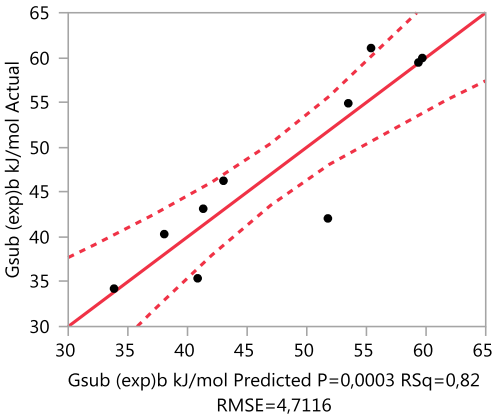


Figure B

Figure A: Free energy of sublimation $\Delta G_{sub}(exp)$ vs ΔG_{sub}(PBE-TS)

Figure B: Free energy of sublimation $\Delta G_{sub}(exp)$ vs ΔG_{sub}(PBE-G)

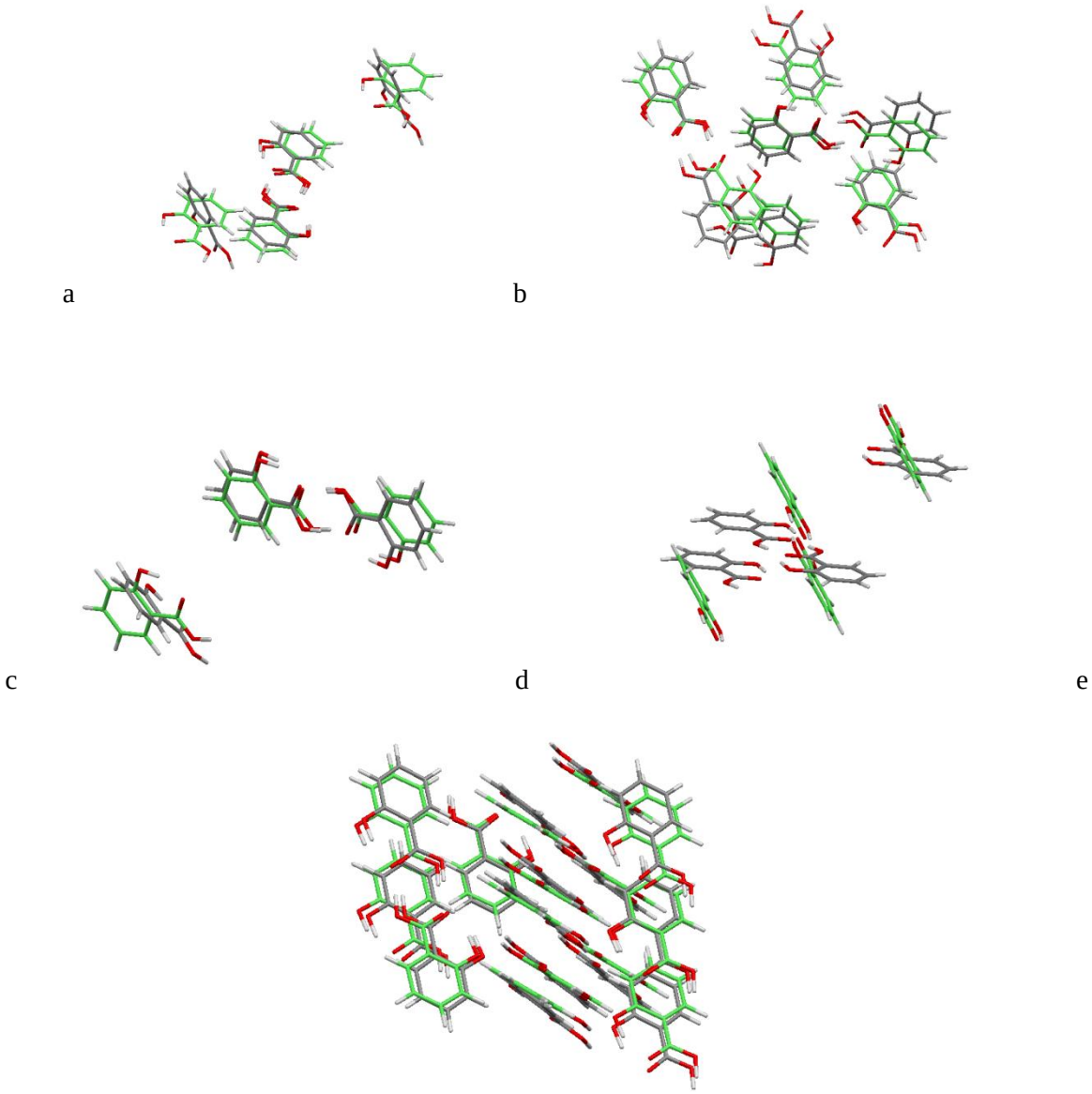


Figure S1 Overlays of DFT-D optimized AZ-FF predicted crystal structures of salicylic acid (SALIAC). a) AZ-FF rank 6, DFT-D rank 4 4/15 R_{msd}=1.89 Å^2. b) AZ-FF rank 8, DFT-D rank 1 7/15 R_{msd} =1.08 Å^2. c) AZ-FF rank 9, DFT-D rank 3, 3/15 R_{msd} =1.55 Å^2. d) AZ-FF rank 12, DFT-D rank 2, 4/15 R_{msd} =2.21 Å^2. e) AZ-FF rank 347, DFT-D rank 5 15/15 R_{msd} =0.55 Å^2.

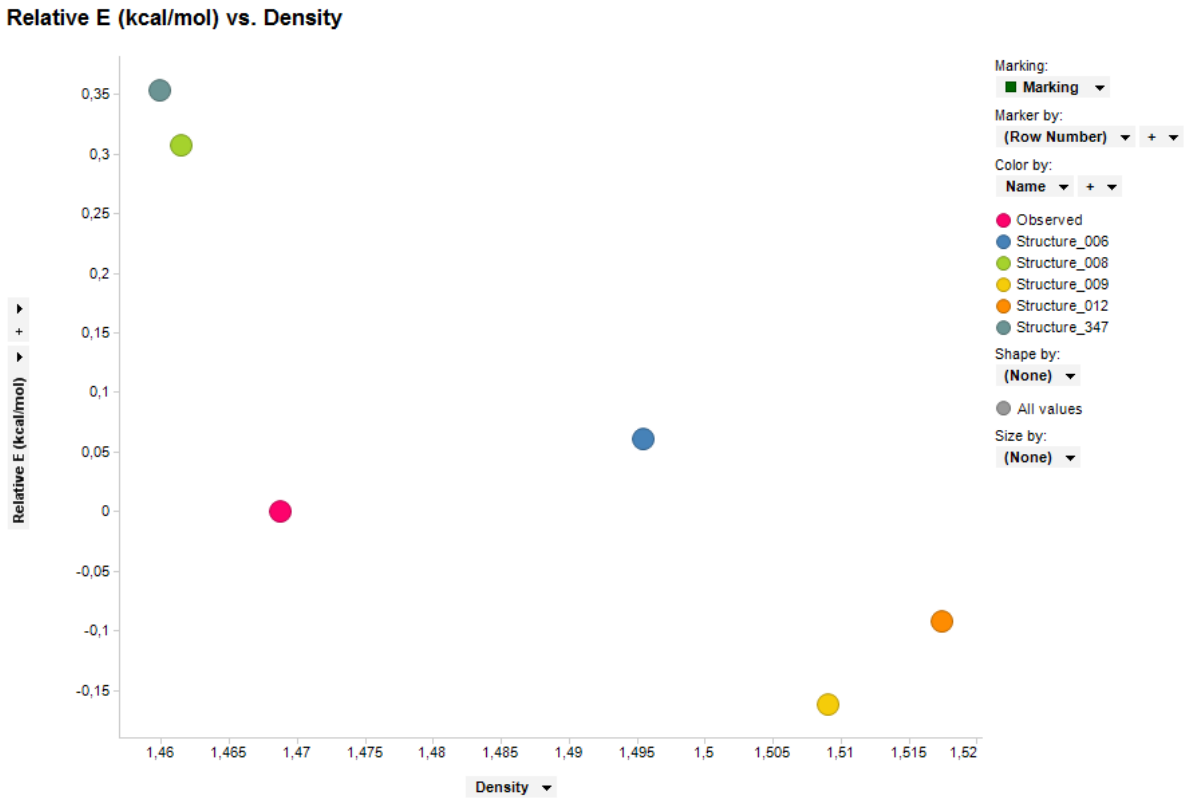
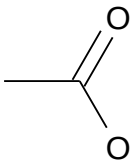
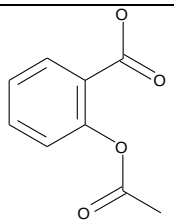
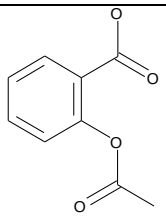
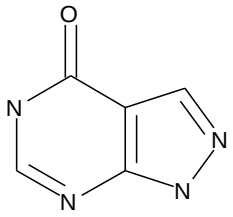
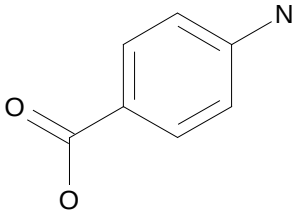


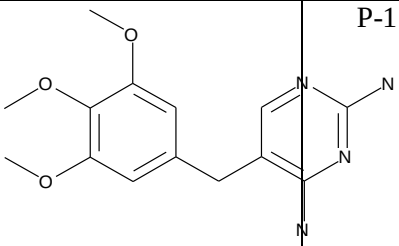
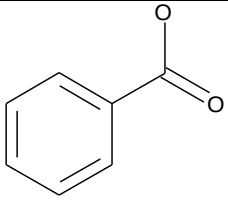
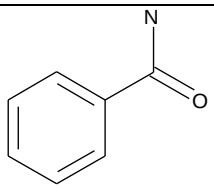
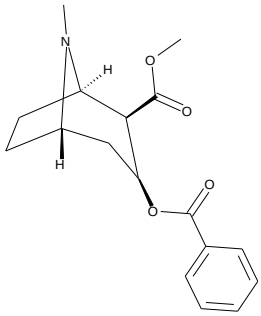
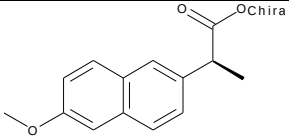
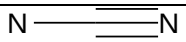
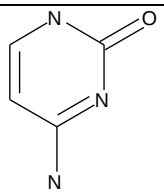
Figure S2 Polymorph landscape for salicylic acid. Structure_347 agrees best with the observed structure and 15/15 molecules are overlaid with an R_{msd}=0.33 Å^2 when comparing PBE-TS optimized structures.

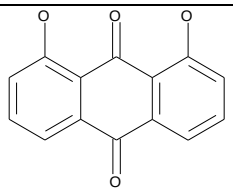
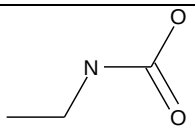
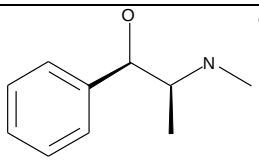
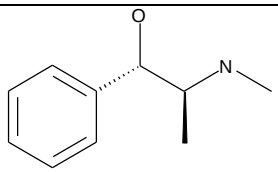
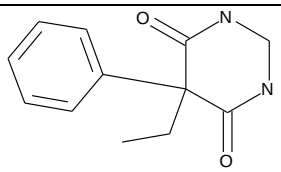
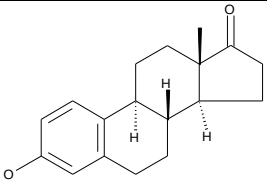
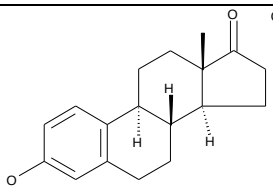
Table S4 Comparison of cell data for PBE-NP optimized tentative crystal structures for acetyl salicylic acid

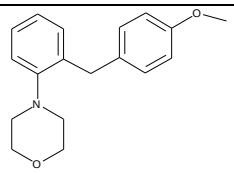
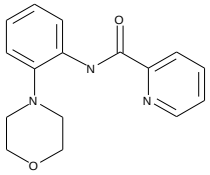
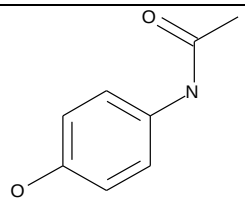
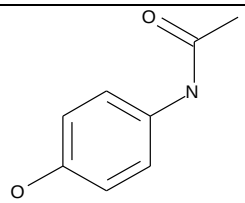
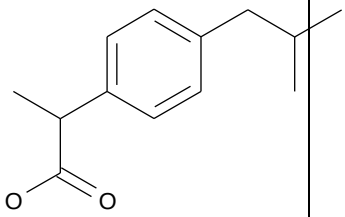
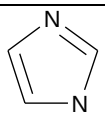
Name	a	b	c	Beta	X/15	Rmsd (Å^2)
SALIAC	4.92	11.21	11.52	90.83		
Rank 8 / 1	6.28	8.83	11.63	102.949	4	1.89
Rank 12 / 2	3.85	12.15	13.87	106.591 (beta)	7	1.08
Rank 9 / 3	3.81	6.81	24.08	91.91	3	1.55
Rank 6 / 4	3.77	7.02	23.98	97.86	4	2.21
Rank 347 / 5	4.35	11.87	12.12	91.17	15	0.55

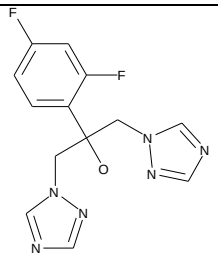
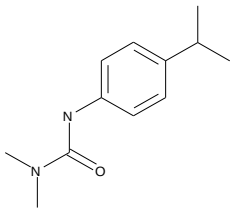
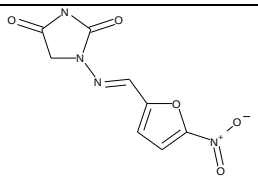
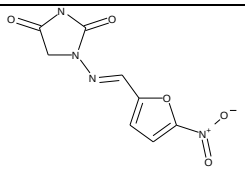
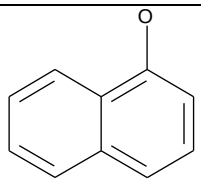
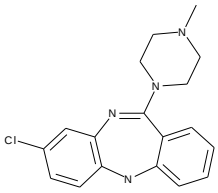
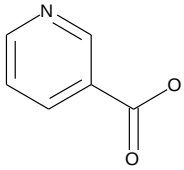
Table S5 Summary of the CSP using the AZ-FF for the 47 polymorphs Z´=1 in the validation set. $RMSD_{15}$ is reported unless a smaller cluster than 15 could be overlaid. In those cases a RMSDx value is reported. A Root Mean Square Cartesian Deviation (RMSCD) was calculated by optimizing the observed crystal structure and comparing the final coordinates with the initial coordinates.

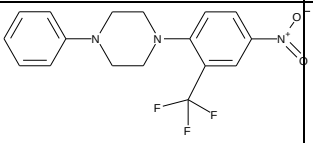
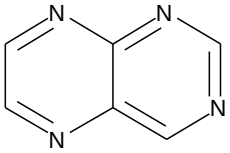
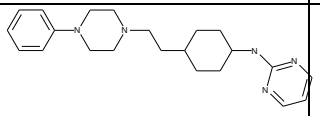
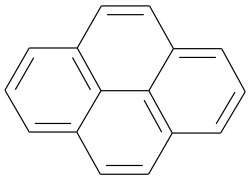
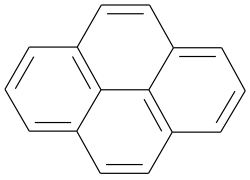
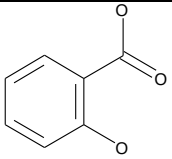
Nr	Name	CSD id	Structure	Space group	Space groups searched	Rank	$RMSD_{15}$	RMSCD AZ-FF opt
1	Acetic acid	ACETAC01		Pna21	Top 20	2	0.384	0.140
2	Aspirin II	ACSALA19		P21/c	Top 20	4	0.418	
3	Aspirin I	ACSALA21		P21/c	Top 20	13	0.436	0.210
4	Allopurinol	ALOPUR		P21/c	Top 20	3	0.237	0.205
5	Aminobenzoic Acid form IV	AMBNAC04		P21/n	Top 20	12	0.321	0.143

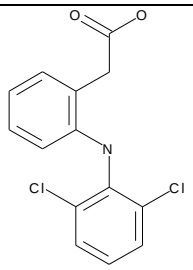
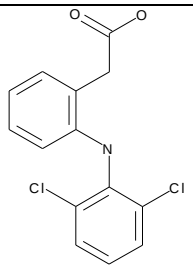
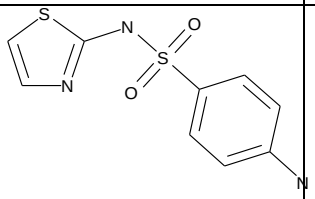
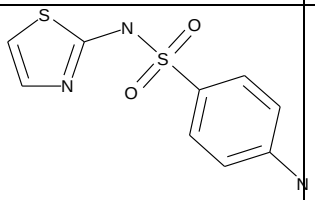
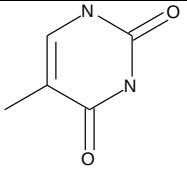
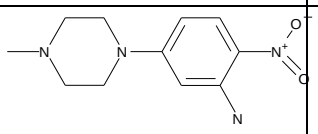
Nr	Name	CSD id	Structure	Space group	Space groups searched	Rank	$RMSD_{15}$	RMSCD AZ-FF opt
6	Trimethoprim	AMXBPM10		P-1	Top 20	56	$RMSD_9$ 0.712	0.569
7	Benzoic Acid	BENZAC02		P21/c	Top 20	23	0.473	0.256
8	Benzamide	BZAMID02		P21/c	Top 20	12	0.555	0.514
9	Cocaine	COCAIN10		P21	Top 7 chiral	40	0.288	0.265
10	Naproxen	COYRUD11		P21	Top 7 chiral	21	0.273	0.212
11	Cyanamide	CYANAM01		Pbca	Top 20	154	0.391	0.281
12	Cytosine	CYTSIN01		P212121	Top 20	3	0.179	0.132

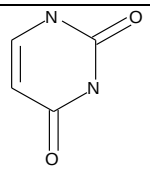
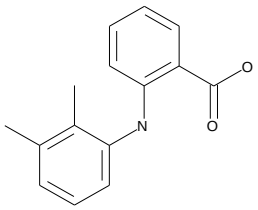
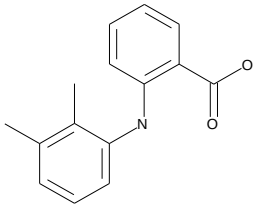
Nr	Name	CSD id	Structure	Space group	Space groups searched	Rank	$RMSD_{15}$	RMSCD AZ-FF opt
13	Danthron form IV	DHANQU06		P21/n	Top 20	191	RMSD12 0.952	0.407
14	Ethylcarbamate	ECARBM10		P-1	Top 20	19	0.630	0.293
15	Ephedrine	EPHEDR01		P212121	Top 7 chiral	105	0.201	0.285
16	PseudoEphedrine	PSEPED		P212121	Top7 chiral	41	0.201	0.345
17	Primidone	EPHPMO		P21/c	Top 20	17	0.487	0.339
18	Estrone form I	ESTRON11		P212121	Top 7 chiral	1	0.232	0.101
19	Estrone form II	ESTRON14		P212121	Top 7 chiral	2	0.145	

Nr	Name	CSD id	Structure	Space group	Space groups searched	Rank	$RMSD_{15}$	RMSCD AZ-FF opt
20	4-(2-(4-methoxybenzyl)phenyl)morpholine	GAMNUE		P21/c	Top 20	1	0.425	0.350
21	N-(2-(Morpholin-4-yl)phenyl)pyridine-2-carboxamide	HOBQEW		P212121	Top 20	29	0.484	0.299
22	Acetaminophen form I	HXACAN07		P21/n	Top 20	23	0.304	0.175
23	Acetaminophen form II	HXACAN08		Pbca	Top20	3	0.118	
24	Ibuprofen form I racemate	IBPRAC01		P21/c	Top20	49	0.277	0.305
25	Imidazole	IMAZOL13		P21/c	Top20	4	0.314	0.195

Nr	Name	CSD id	Structure	Space group	Space groups searched	Rank	$RMSD_{15}$	RMSCD AZ-FF opt
26	Fluconazole	IVUQOF		P-1	Top20	2	0.426	0.305
27	Isoproturon	JODTUR01		Pbca	Top20	1	0.309	0.277
28	Nitrofurantoin Alpha form	LABJON01		P-1	Top20	18	$RMSD_{11}$ 0.684	0.420
29	Nitrofurantoin Beta form	LABJON02		P21/n	Top20	1	0.563	
30	1-Naphthol	NAPHOL01		P21/c	Top20	2	0.597	0.224
31	Clozapine	NDNHCL01		P212121	Top20	48	0.318	0.295
32	Nicotinic Acid	NICOAC02		P21/c	Top20	1	0.229	0.224

Nr	Name	CSD id	Structure	Space group	Space groups searched	Rank	$RMSD_{15}$	RMSCD AZ-FF opt
33	1-(4-Nitro-2-(trifluoromethyl)phenyl)-4-phenylpiperazine	OMOJIJ		P21/n	Top20	2	$RMSD_{10}$0.289 The tool crashed when calculating RMSD15	0.369
34	Pteridine	PTERID11		Pna21	Top20	2	0.305	0.288
35	trans-(4-(2-(4-Phenylpiperazin-1-yl)ethyl)-cyclohexyl)pyrimidin-2-ylamine	PUDGOK		P21/n	Top20	64	$RMSD_{13}$ 0.523	
36	Pyrene form I	PYRENE02		P21/a	Top20	42	0.599	
37	Pyrene form II	PYRENE07		P21/a	Top20	25	0.181	0.165
38	Salicylic Acid	SALIAC		P21/a	Top20	13	$RMSD_{6}$ 0.24	0.147

Nr	Name	CSD id	Structure	Space group	Space groups searched	Rank	$RMSD_{15}$	RMSCD AZ-FF opt
39	Diclofenac form I	SIKLIH01		C2/c	Top20	1	0.405	0.224
40	Diclofenac form II	SIKLIH02		P21/c	Top20	7	0.441	
41	Sulfathiazole form II	SUTHAZ30		P21/c	Top20	34	0.181	
42	Sulfathiazole form IV	SUTHAZ36		P21/c	Top20	4	0.197	0.180
43	Thymine	THYMINE01		P21/c	Top20	90	0.749	0.414
44	5-(4-Methylpiperazin-1-yl)-2-nitroaniline	TUNPIC		P21/c	Top20	118	0.376	0.327

Nr	**Name**	**CSD id**	**Structure**	**Space group**	**Space groups searched**	**Rank**	**$RMSD_{15}$**	**RMSCD AZ-FF opt**
45	Uracil	URACIL		P21/a	Top20	89	0.422	0.341
46	Mefenamic Acid form I	XYANAC		P-1	Top20	4	0.32	0.191
47	Mefenamic Acid form III	XYANAC03		P-1	Top20	3	0.709	