Electronic Supplementary Material (ESI) for CrystEngComm

Supplementary Information for

One in half a million: A Solid Form Informatics study of a pharmaceutical crystal structure

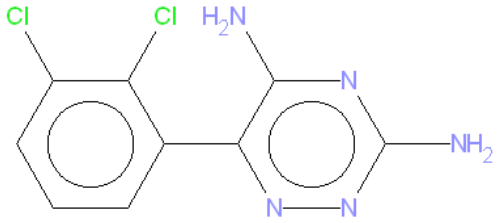
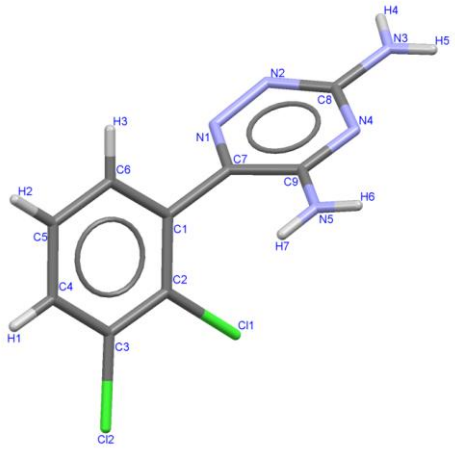
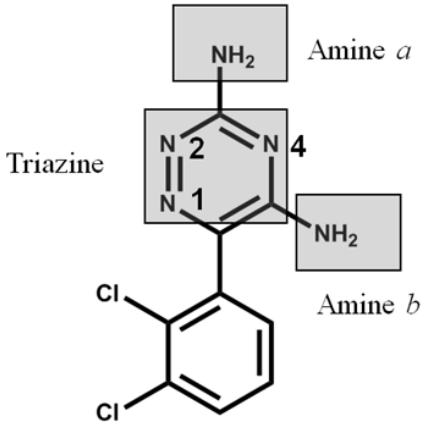
Peter T. A. Galek, Elna Pidcock, Peter. A. Wood, Ian J. Bruno & Colin R. Groom

Contents

1. Chemical Analysis

1.1. Basic chemical information

Table S1.

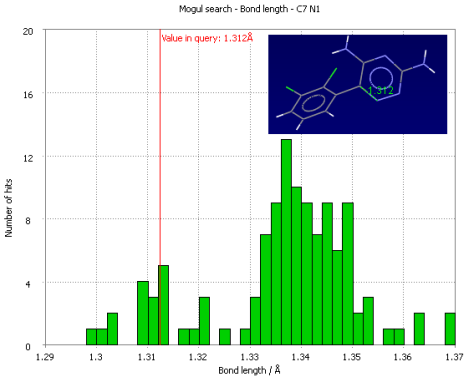
Common Name; Synonyms	Lamotrigine; Lamictal
Systematic Name	6-(2,3-Dichlorophenyl)-1,2,4-triazine-3,5-diamine
Chemical Diagram	
Empirical Formula	C9 H7 Cl2 N5
SMILES String	Clc2c(Cl)c(c1nnc(nc1N)N)ccc2
Std. InChI	InChI=1S/C9H7Cl2N5/c10-5-3-1-2-4(6(5)11)7-8(12)14-9(13)16-15-7/h1 -3H,(H4,12,13,14,16)
Bioactivity information:	Antiepileptic drug; Used in the treatment of bipolar I disorder; Blocks voltage-gated sodium and calcium channels and inhibits release of excitatory neurotransmitters such as glutamate
3D crystal conformation with labelled atoms	
Assignment of hydrogen bonding groups	

1.2. Molecular geometry analysis

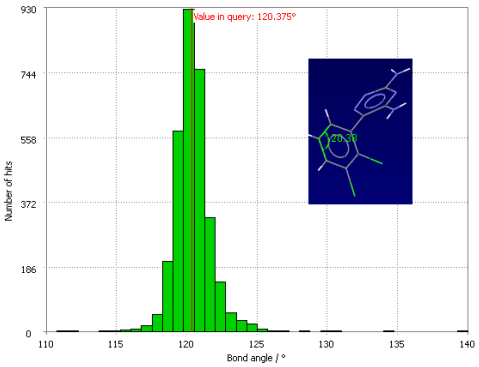
Table S2. Molecular Geometry Summary, excluding H atoms

Bond data	17 bonds. 0 **unusual** (sufficient data)
Angle Data	24 angles. **0 unusual** (sufficient data)
Torsion Data	4 Flexible Torsion angles. **0 Unusual**
Ring Geometry Data	2 Rings. **0 Unusual**.

Figure S1. Selected geometries compared with corresponding CSD distributions. a) C7-N1 bond, showing largest Z-score b) C9-N4 bond c)C6-C5-C4 angle showing smallest z-score. Metal-organic structures were excluded. Similar fragments were included when the number of exact matches was <15 for bonds, angles and rings, and < 40 for torsions.



a)



b)

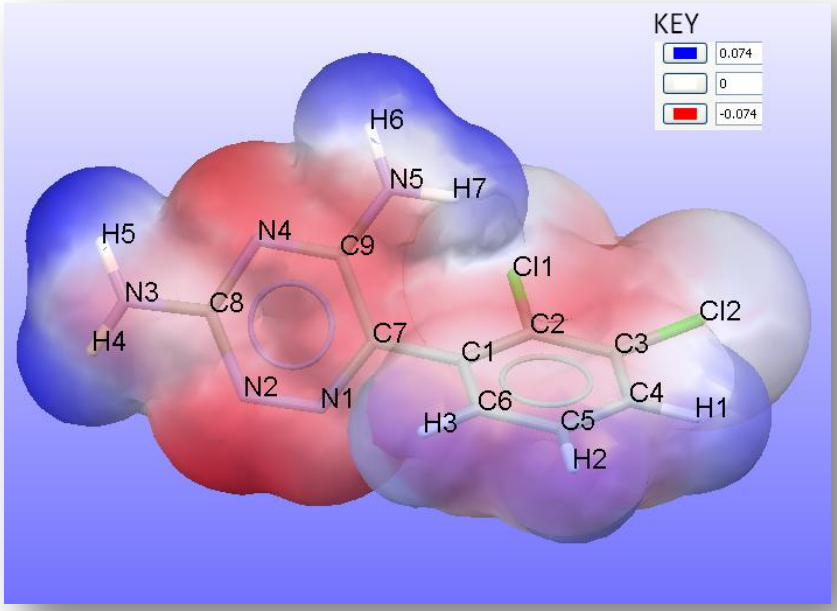
1.3. Molecular descriptors

Table S3.

Property	**Value**	**Property**	**Value**
Molecular Mass	256.091	Cyclicity	0.7059
Atom count	20	Enclosing box ratio L:S	1.730
Surface Area (marvin calc. 3D)	236.75 Å^2	Enclosing box ratio M:S	1.200
Polar Surface Area (PSA)	90.71 Å^2	Enclosing box ratio L:M	1.441
Fraction PSA	38.3 %	ALogP	2.426
# Rotatable Bonds	1	Log D (pH 7.4)	2.422
# Donor H Atoms	4	Heat of Formation (MOPAC)	84.2068 kcal/mol
#H-bond Acceptors	5	Dipole (magnitude)	2.145 debye
Fraction N,O atoms	0.3125	Dipole (x, y, z)	-1.431, -1.493, 0.568
Fraction Cl, Br, I, F	0.125		

1.4. Calculated electrostatics

Figure S2. Electrostatic potential (eV) mapped onto molecular van-der-Waals surface. The electropositive amine hydrogen donors are easily visible, as are the electronegative N atoms of the triazine ring. Some variation in the potential around the dichlorobenzene ring is also apparent, suggesting Cl can be a weak H acceptor in this case, or that CH could be a weak H donor.



2. Intermolecular Analysis

2.1. Crystal Lattice Data

Table S4.

Unit Cell		Z	Z: 8 Z': 1
a	19.136(3)	R-factor(%)	2.81
b	8.6409(12)	Density (calculated)	1.607 (1.607)
c	13.5549(18)	Packing coefficient	0.72
α	90	%void space (1.2Å probe)	0.0
β	109.172(2)		
γ	90		
Space Group	C2/c		
Cell Volume	2117.02		

2.2.Hydrogen bond geometries

Figure S3. Hydrogen bonds as dashed lines in EFEMUX01. 4 molecules are shown making up a motif arranged about the 2-fold screw axis. The donor and acceptor atoms are labelled.

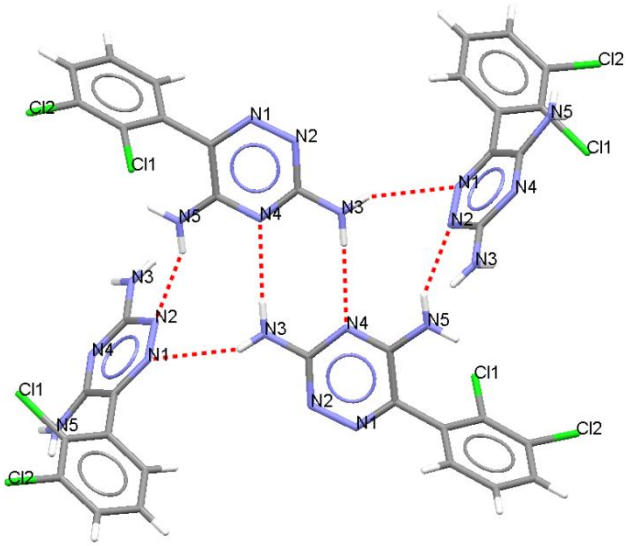


Table S5. a) H-bond geometries, b) Short contact geometries in EFEMUX01

a)

Donor D	Acceptor A	Symm. op. 1	Symm. op. 2	H•••A Distance	H•••A Distance (r_{VdW} corrected)	D-H•••A Angle
N3-H4	N1	x,1-y,1/2+z	1.5-x,1.5-y,-z	2.539	-0.211	155(2)
N3-H5	N4	x,y,z	1.5-x,1.5-y,-z	2.33	-0.42	177(2)
N5-H6	N2	x,y,z	x,1-y,1/2+z	2.399	-0.351	151(2)

b)

Atom X	Atom Y	Symm. op. 1	Symm. op. 2	X•••Y Distance	X•••Y Distance (r_{VdW} corrected)	X-H•••Y Angle
C6-H3	N4	x,y,z	1.5-x,1/2-y,-z	2.597	-0.153	134.6(1)
N5-H7	Cl2	x,y,z	2-x,y,1/2-z	2.648	-0.302	151(2)
C5-H2	N4	x,y,z	x,-1+y,z	2.721	-0.029	169.1(1)
C5	C3	x,y,z	2-x,-y,-z	3.319	-0.081	N/A
C6	Cl2	x,y,z	2-x,-y,-z	3.336	-0.114	N/A
C4	C2	x,y,z	2-x,-y,-z	3.345	-0.055	N/A
C3	C3	x,y,z	2-x,y,1/2-z	3.357	-0.043	N/A

2.3.CSD Contact distributions

Figure 4a. Scatterplots showing density of short-contact atoms in the vicinity of the lamotrigine molecule, coloured by Isostar propensity.

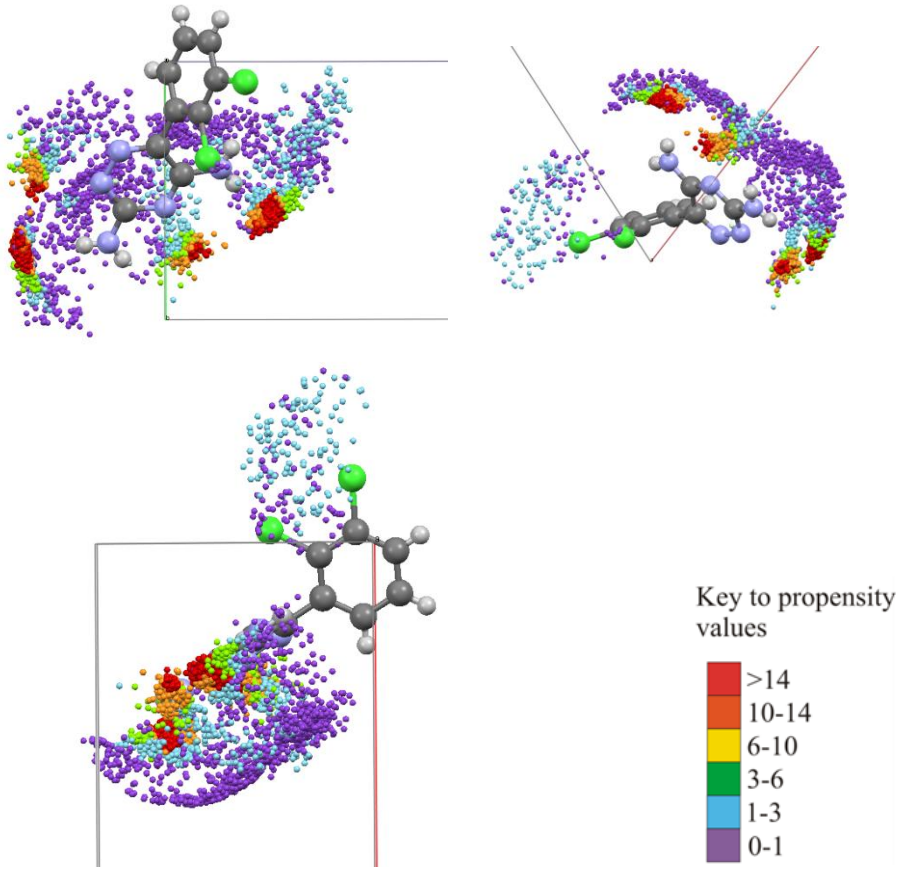


Figure 4b. Geometric overlay of observed hydrogen bonds in EFEMUX01 with scatterplots of observed close contacts between the same atom types from the CSD, coloured by Isostar propensity. A) aromatic N interaction with NH_2 group. B) Pyridizine fragment interaction with NH_2 groups, C) aromatic-Cl group interaction with NH_2 group.

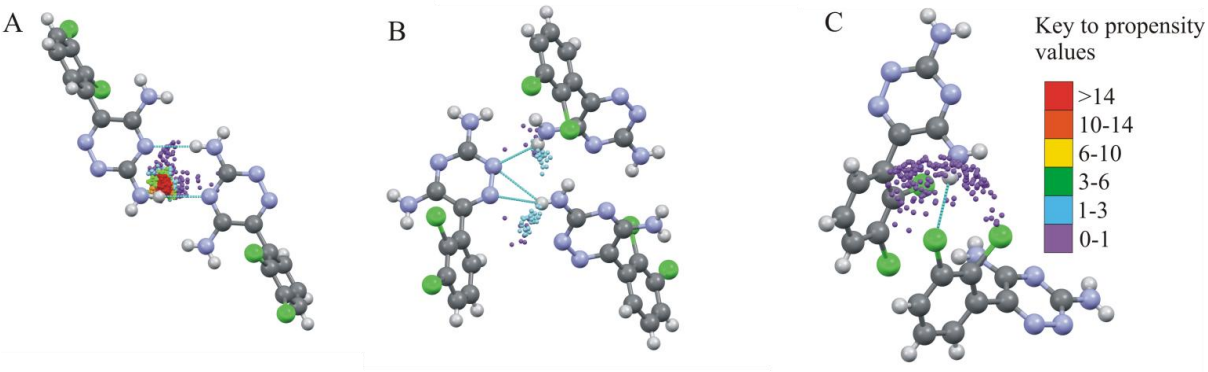
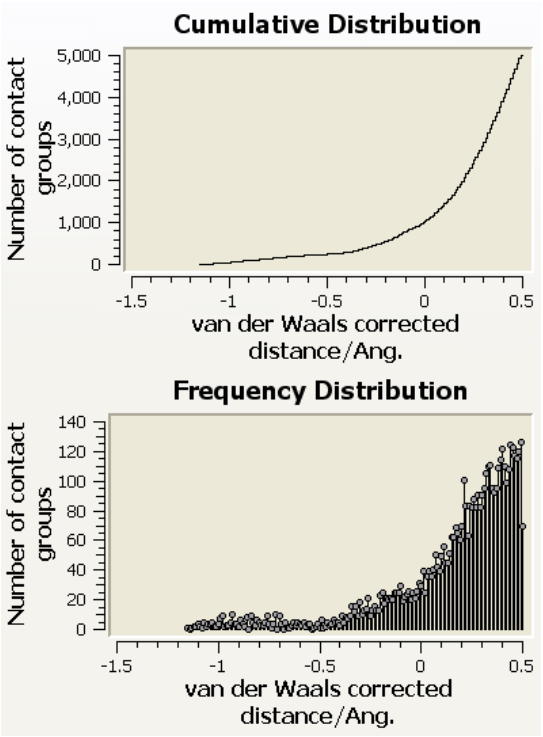
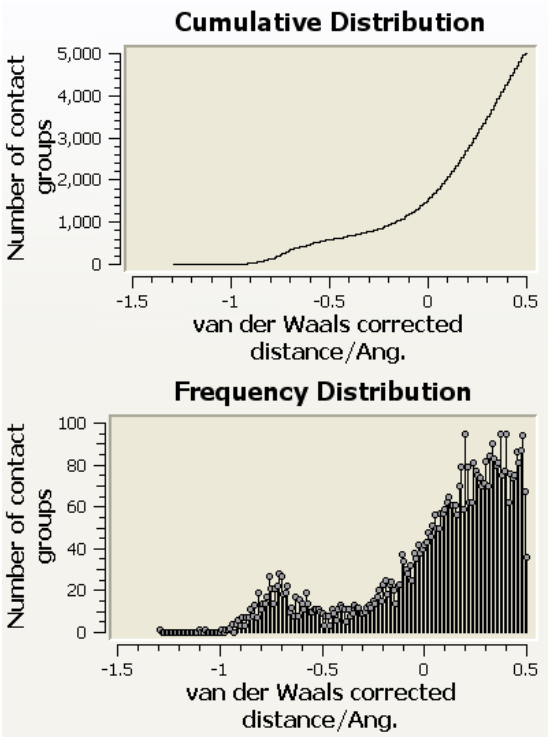


Figure 4c. Histograms showing the distribution of atom-atom contact distances in the CSD by atom type as found in lamotrigine i) Aromatic Chloro ii) Aromatic amino planar, iii)Aromatic N, in 6-rings iv) nitrogen in N(ar)-CH-N(ar). The peak below sum van der Waals radii in ii) indicates a preference for close contacts at this donor.

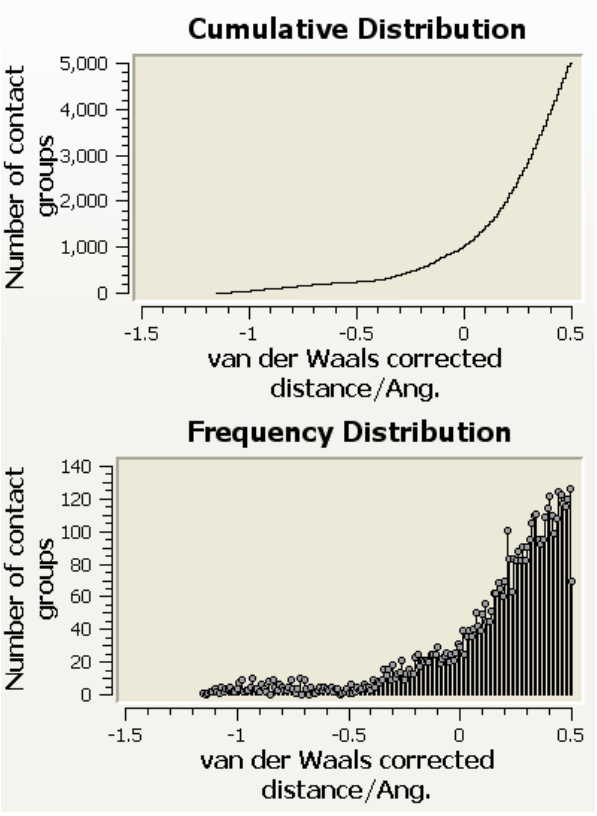
i)Aromatic Chloro



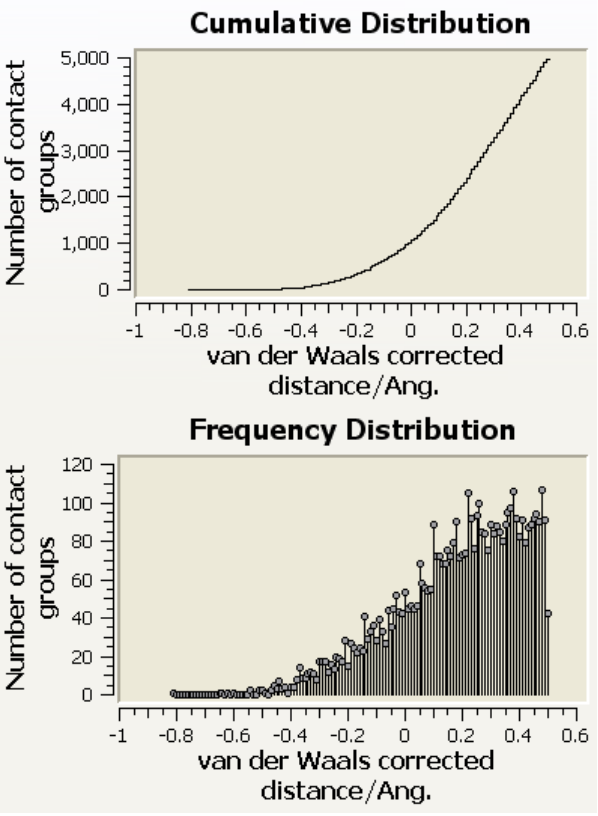
ii) Aromatic amino planar



ii)Aromatic N, in 6-rings

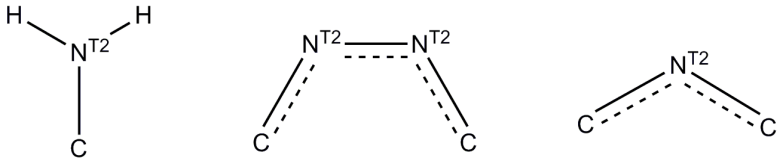


iv) N(ar)-CH-N(ar)



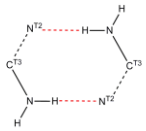
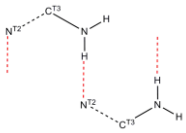
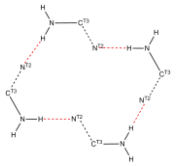
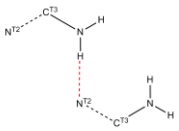
2.4.Motif Analysis

Scheme 1 - Functional group representation used in interaction motif analysis. Atom super-script labels refer to the coordination number of the specific atoms.



Amine **1,2-diazine** **Pyridine**

Table S6. H-bond motif statistics.

(a) Aminopyrdine homo-motif		No. of structures	Frequency of observation	Proportion of observed motifs
R2 ring dimer		631	43.3%	75.3%
Infinite chain		256	17.6%	30.5%
Larger ring (R3, R4, R6)		29	1.9%	3.5%
Discrete contact		26	1.7%	3.1%

(b) Group A hetero-motif	Examples	No. of contacts	Frequency of occurrence	Proportion of observed interactions
Oxygen (1-coordinate)	O=C, O=N, O=S	575		43.3%
Oxygen (2-coordinate)	H-O-H, H-O-C, C-O-C	443		33.4%
Nitrogen (2-coordinate)	C-N=C, aromatic N	188		14.2%
Nitrogen (1-coordinate)	N=C	47		3.5%
Nitrogen (3-coordinate)	H_2N-C, H-NC_2, NC_3	20		1.5%
Sulfur (1-coordinate)	S=C	20		1.5%
Other	F, Cl, Br, I	32		2.4%

3. Supramolecular Analysis

3.1. Crystal Lattice data and descriptors

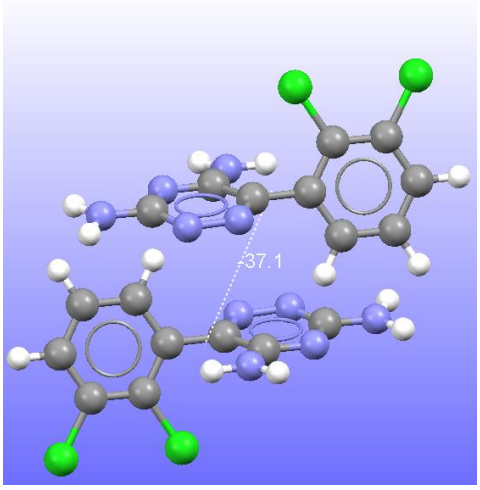
Table S7.

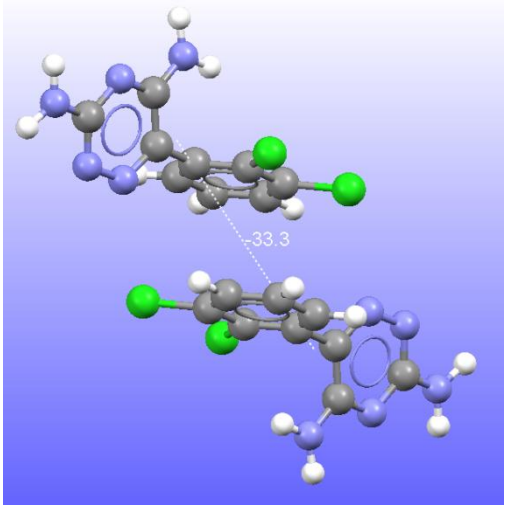
	Value	Structure Property	Value
a	19.136(3) Å	*R*-factor (%)	2.81
b	8.6409(12) Å	Density (reported)	1.607 g cm^{-3}
c	13.5549(18) Å	Density (calculated)	1.607 g cm^{-3}
α	90°	Temperature	294 K
β	109.172(2)°	Crystal colour	colourless
γ	90°	Habit	block
Cell volume	2117.02 Å^3	Packing coefficient	0.72
Space group	*C* 2/*c*	Percentage void space	0.0 %
Z values	*Z*: 8 *Z*': 1		

3.2.Intermolecular energy calculations and morphology

.

Figure S5. Six highest intermolecular interaction potentials computed from a central molecule and 14 nearest neighbours. Using the UNI intermolecular force-field available *Mercury*. The value of the computed potential energy is shown in each case.

a 

b 

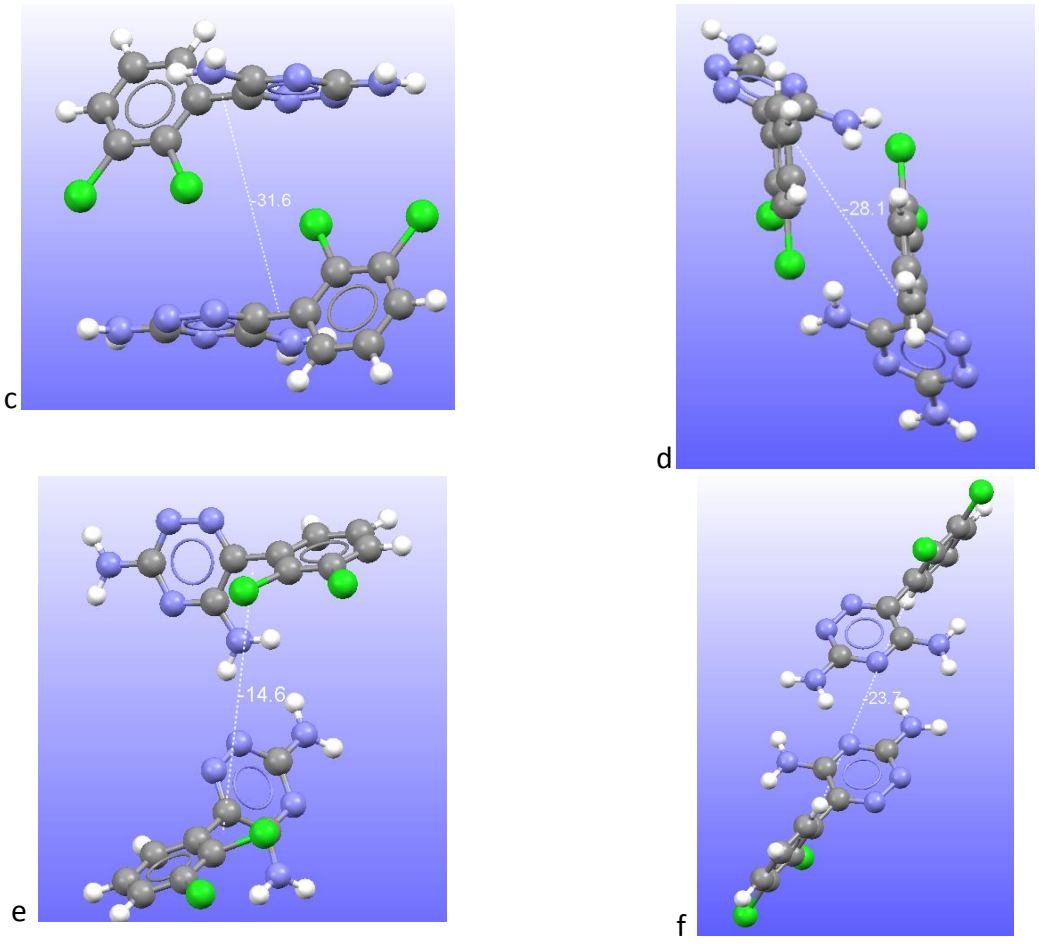


Figure S6a. Comparison of Bravais-Friedel-Donnay-Harker (BFDH) morphology prediction of the EFEMUX01 structure and the HABIT morphology prediction. In addition to unit cell data that BFDH relies on, HABIT includes energetic considerations. A blocky morphology can be expected.

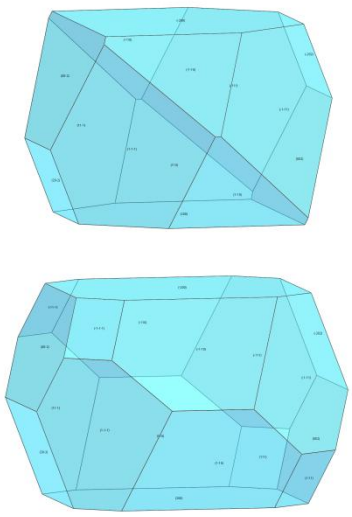


Figure S6b. HABIT morphology prediction of the EFEMUX01 structure superimposed with a molecule cluster showing the H-bonding network. Molecules are colored by interaction energy with the central (grey) molecule. Color = red > 30 kJmol-1, gold= 21-30 kJmol-1 and green = 10-20 kJmol^{-1}. The strong interactions which contribute to the slow growth rate and hence morphological importance of the (200) face can be seen.

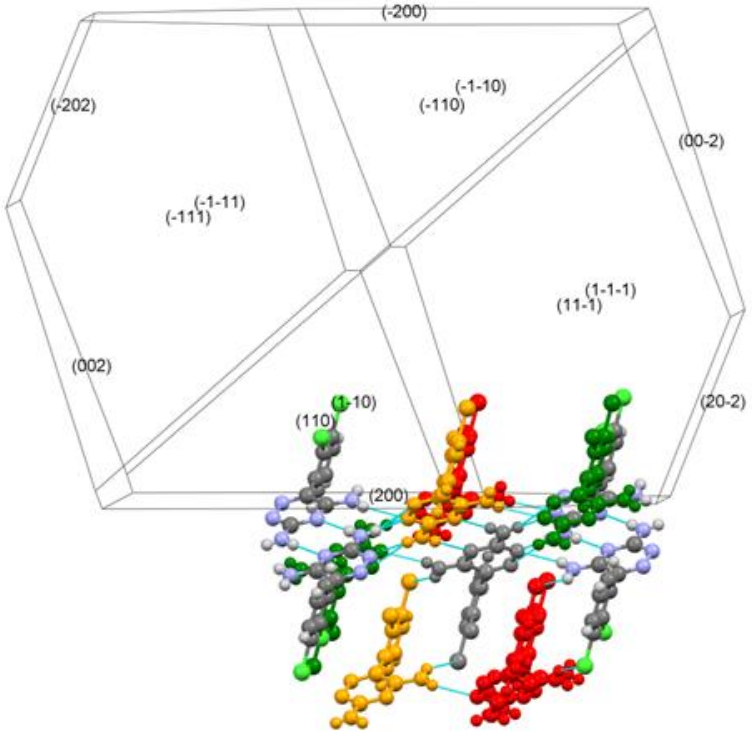


Figure S6c. Separate display of intermolecular energies within a cluster categorized by size. Color = red > 30 kJmol^{-1}, gold = 21-30 kJmol^{-1} and green = 10-20 kJmol^{-1}.

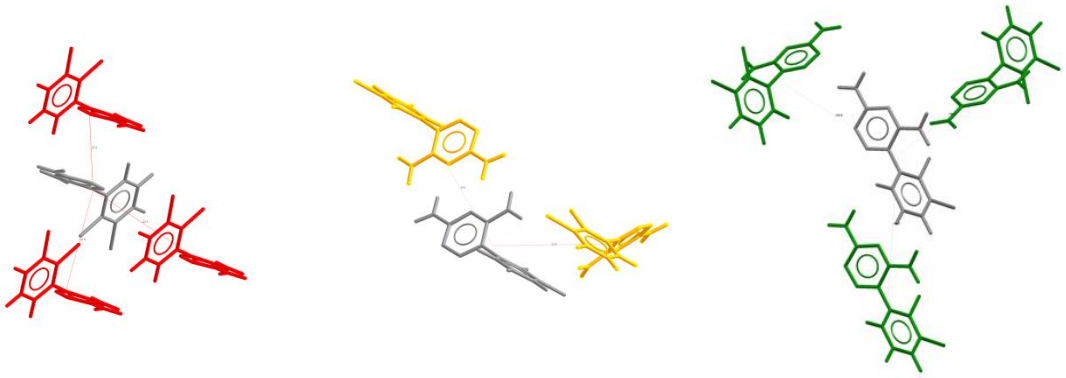


Figure S7a. Interaction energies between molecules of the 200 crystal face. Dotted lines display the interaction vector between molecule centroids. These are coloured red to blue from most to least energetic (~35 $kJMol^{-1}$ to ~10 $kJMol^{-1}$).

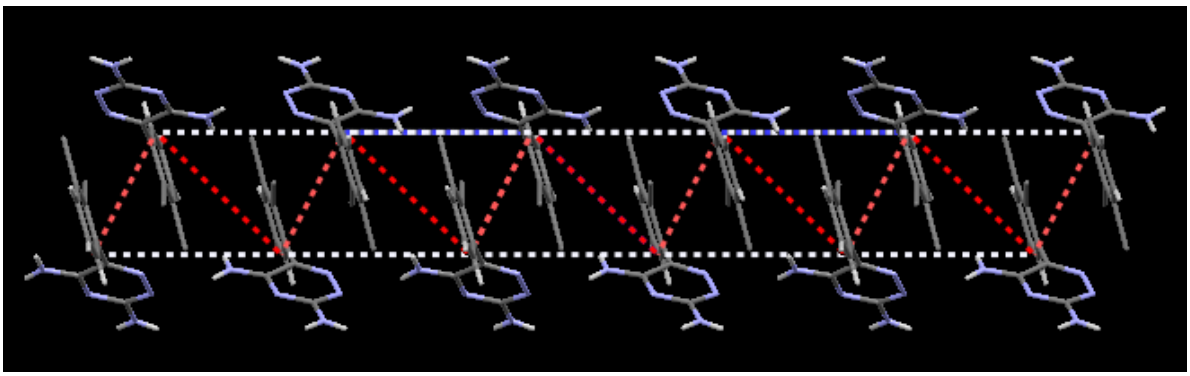
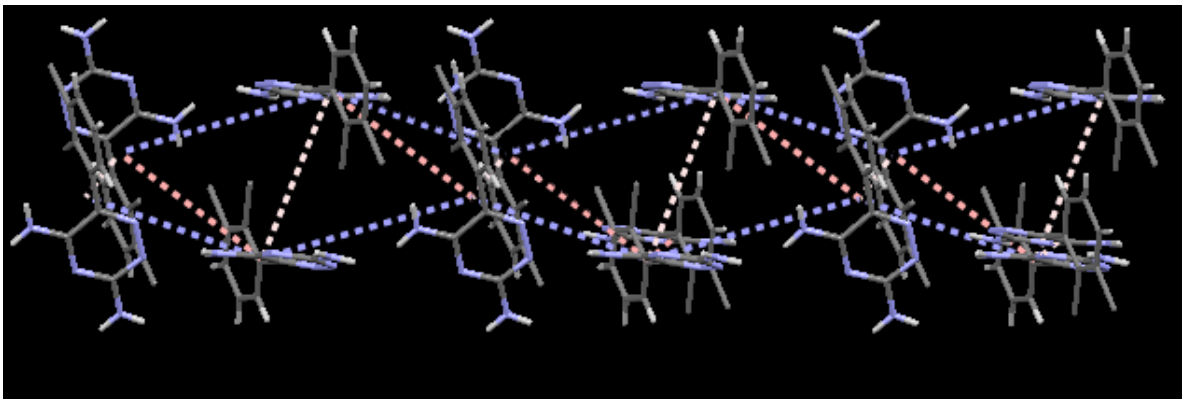
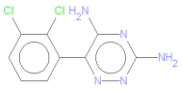
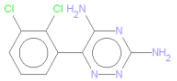
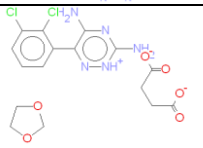
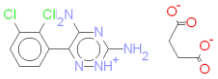
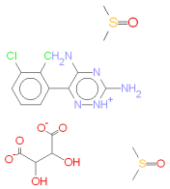
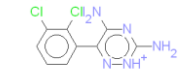
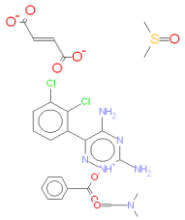
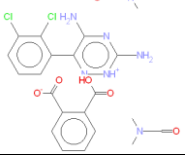
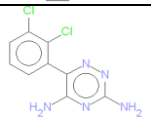
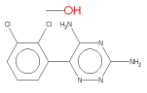
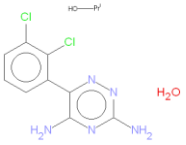
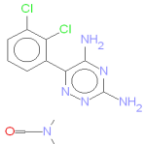


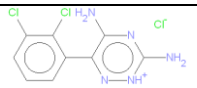
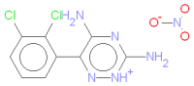
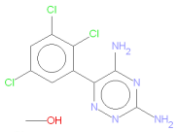
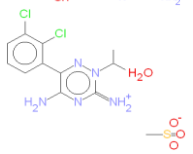
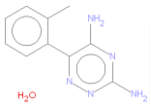
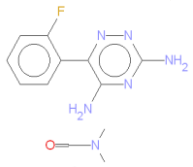
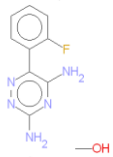
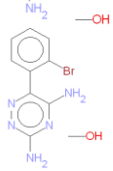
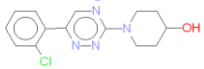
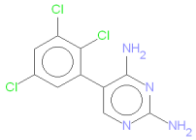
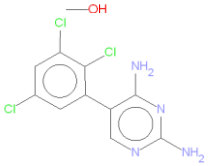
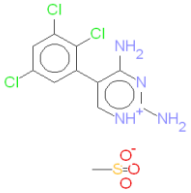
Figure S7b. Interaction energies between molecules of the 301 crystal face.



3.3. Structural Relationships in the CSD

TableS8. Related Structures containing a) Lamotrigine b) derivatives of Lamotrigine*

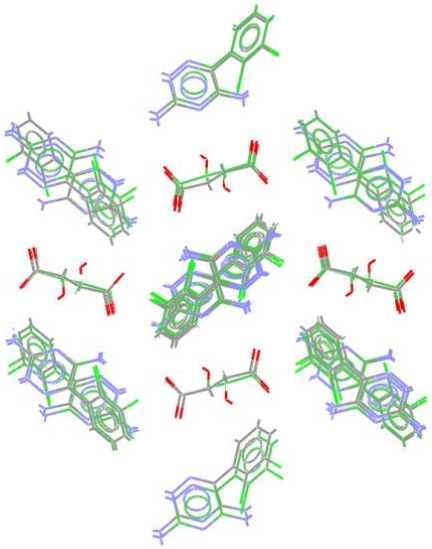
a CSD Refcode	Structure Type	Secondary components	Diagram
EFEMUX	Polymorph	n.a.	
EFEMUX01		n.a.	
FOXMAH	Solvated Salt	hemikis(succinate) 1,3-dioxolane	
FOXLUA		hemikis (succinate) dimethylsulfoxide	
FOXMEL		hemikis(dl-tartrate) dimethylsulfoxide	
FUHVOU		hemikis(fumarate) dimethylsulfoxide	
GAVLEV		benzoate N,N'-dimethylformamide	
YEXFUD		phthalate N,N'-dimethylformamide	
KADPAG	Solvate	methanol	
IJAHOR		isopropanol	
XUVLOP		hydrate	
YERTAR		N,N'-dimethylformamide	

YUCRAQ	Salt	chloride	
YUCREU		nitrate	
b CSD Refcode	Tanimoto Similarity Score	Compound	Diagram
BEGZUI	0.991	3,5-Diamino-6-(2,3,5-trichlorophenyl)-1,2,4-triazine methanol solvate	
XUVLUV	0.727	2-Isopropyl-3-iminium-5-amino-6-(2,3-dichlorophenyl)-1,2,4 triazine mesylate monohydrate	
LINFOD	0.663	3,5-Diamino-6-(2-methylphenyl)-1,2,4-triazine hemihydrate	
CIQLET	0.604	3,5-Diamino-6-(2-fluorophenyl)-1,2,4-triazine dimethylformamide	
WINMIP	0.604	3,5-Diamino-6-(2-fluorophenyl)-1,2,4-triazine methanol solvate	
TEKWAH	0.598	3,5-Diamino-6-(2-bromophenyl)-1,2,4-triazine methanol solvate	
SAPYUD	0.582	1-(6-(2-Chlorophenyl)-1,2,4-triazin-3-yl)piperidin-4-ol	
HOBVEA	0.514	5-(2,3,5-Trichlorophenyl)-2,4-diaminopyrimidine methanol solvate	
HOBVIE	0.514	5-(2,3,5-Trichlorophenyl)-2,4-diaminopyrimidine	
TIRNOY	0.514	5-(2,3,5-Trichlorophenyl)-2,4-diaminopyrimidinium methanesulfonate	

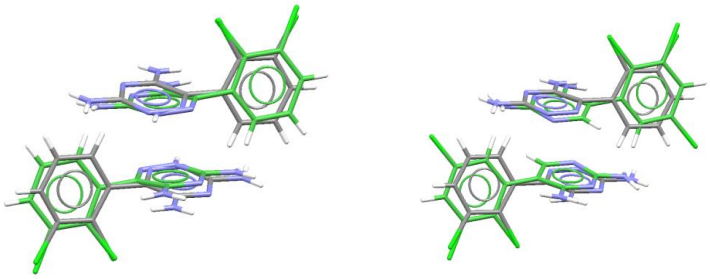
*Listing CSD structures whose compounds are chemically similar by fingerprint comparison with Tanimoto coefficient > 0.5.(ref.)

3.3.1. Common Packing Features

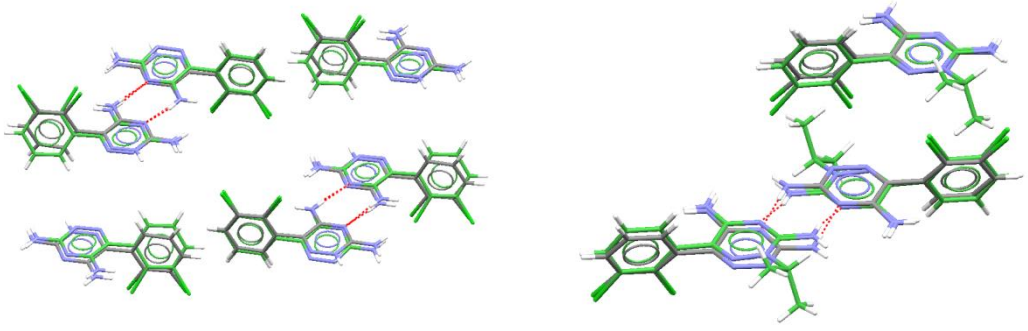
Figure S9. a) Overlaid 15 molecule shell in the FOXLUA and FOXMEL structures (viewed down the crystallographic *a* axes, which also coincide). The carbon atoms of the latter are colored green. Good spatial alignment of the chemically different counter ions can be seen. The third DMSO component in each structure is omitted for clarity. b) Common dimer involving aromatic stacking found in EFEMUX01, FOXLUA and HOVBEA. c) common amide-triazine H-bonded dimer motif found in GAVLEV and YERTAR, and EFEMUX01, LINFOD and XUVLUV.



a



b



c

4. Assessing a potential co-crystal formulation

4.1.Shape descriptors for co-former selection.

Table S9. CSD-based cocrystal former screening a) descriptor values of API/co-formers b) screening protocol applied to co-formers with respect to lamotrigine. See text for definitions of descriptor abbreviations.

a)

Compound	FNO	Dipole	S	S/L	M/L
Lamotrigine	0.313	4.189	6.778	0.552	0.663
3,4 butylated hydroxyanisole	0.154	1.705	6.43	0.567	0.735
methylparaben	0.273	5.094	4.268	0.416	0.676
L-malic acid	0.556	3.543	5.423	0.716	0.974
nicotinamide	0.333	1.678	3.401	0.372	0.73
nicotinic acid	0.333	2.885	3.401	0.37	0.725

b)

Coformer screening	|FNOa-FNOb|	|dipole a-dipole b|	|Sa-Sb|	|Sa/La-Sb/Lb|	|Ma/La-Ma/Lb|	**Verdict**
(Pass Mark)	**<0.294**	**<5.94**	**<3.23**	**<0.275**	**<0.310**	
3,4 butylated hydroxyanisole	0.159	2.484	0.348	0.015	0.072	**Pass**
methylparaben	0.04	0.905	2.51	0.136	0.013	**Pass**
L-malic acid	**0.243**	0.646	1.355	0.164	**0.311**	**Borderline**
nicotinamide	0.02	2.511	**3.377**	0.18	0.067	**Fail**
nicotinic acid	0.02	1.304	**3.377**	0.182	0.062	**Fail**

4.2.H-bond prediction

Figure S10. Chemical diagram of lamotrigine and methylparaben with labelled functional groups.

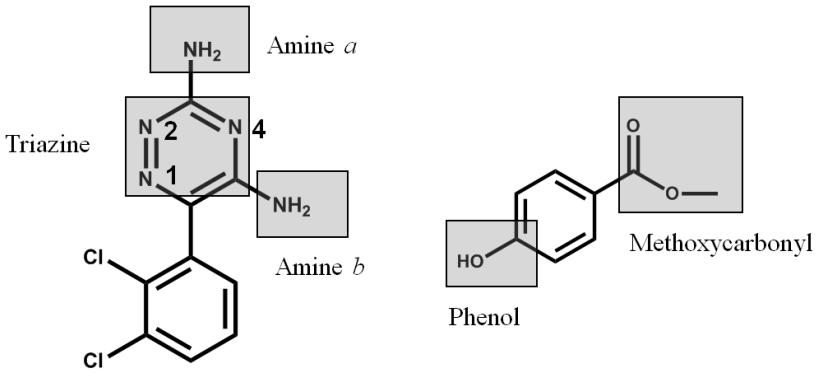


Table S10. H-bond propensity predictions for a) pure lamotrigine, b) lamotrigine buytlated hydroxyanisole cocrystal, c) lamotrigine methylparaben cocrystal. Functional groups are labelled according to Fig. S10. Lower and upper bounds calculated at the 95% confidence level based on a χ^2 distribution.

Key to Bond Formed column: ; *Green* ✓: observed, *Red* ×: not observed. p = observed in pure form, c = unique to cocrystal. Note as no hydroxyanisole cocrystals have been reported, in table b the final column is not annotated with ✓or × .

a

Donor	Acceptor	propensity	Lower bound	Upper bound	H-Bond Observed?
amine NH2 a	triazine N1	0.928	0.913	0.940	✓
amine NH2 b	triazine N1	0.909	0.890	0.926	×
amine NH2 a	triazine N2	0.905	0.884	0.923	×
amine NH2 b	triazine N2	0.881	0.855	0.904	✓
amine NH2 a	triazine N4	0.702	0.630	0.764	✓
amine NH2 b	triazine N4	0.647	0.569	0.719	×
amine NH2 a	Cl8	0.419	0.377	0.461	×
amine NH2 b	Cl8	0.360	0.317	0.405	×
amine NH2 a	Cl7	0.353	0.311	0.398	×
amine NH2 b	Cl7	0.299	0.258	0.343	×

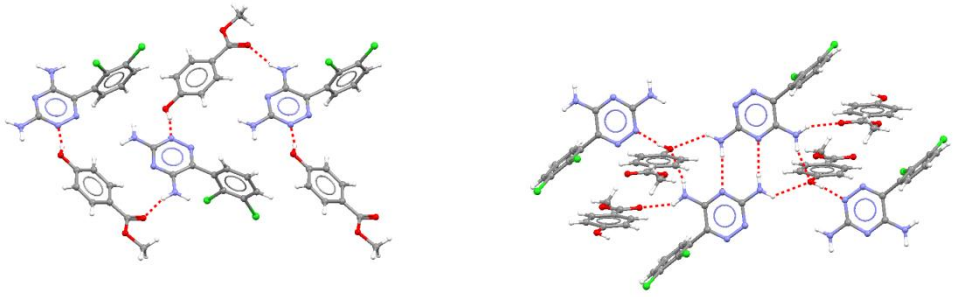
b

Donor	Acceptor	propensity	Lower bound	Upper bound	H-Bond Observed?
amine NH_2 a	triazine N1	**0.932**	0.916	0.945	p
amine NH_2 b	triazine N1	**0.922**	0.903	0.937	
amine NH_2 a	triazine N2	**0.915**	0.893	0.932	
amine NH_2 b	triazine N2	**0.902**	0.876	0.923	p
phenol OH	triazine N1	**0.793**	0.745	0.834	c
phenol OH	triazine N2	**0.749**	0.692	0.800	c
amine NH_2 a	triazine N4	**0.608**	0.515	0.693	p
amine NH_2 a	phenol OH	**0.588**	0.501	0.669	c
amine NH_2 a	OMe	**0.582**	0.493	0.665	c
amine NH_2 b	triazine N4	**0.571**	0.475	0.662	
amine NH_2 b	phenol OH	**0.550**	0.461	0.636	c
amine NH_2 b	OMe	**0.544**	0.454	0.632	c
amine NH_2 a	Cl5	**0.385**	0.341	0.431	
amine NH_2 b	Cl5	**0.350**	0.304	0.399	
amine NH_2 a	Cl6	**0.333**	0.288	0.381	
phenol OH	triazine N4	**0.302**	0.221	0.397	c
amine NH_2 b	Cl6	**0.300**	0.255	0.349	
phenol OH	phenol OH	**0.284**	0.215	0.366	c
phenol OH	OMe	**0.279**	0.209	0.363	c
phenol OH	Cl5	**0.149**	0.120	0.183	c
phenol OH	Cl6	**0.122**	0.097	0.153	c

c

Donor	Acceptor	propensity	Lower bound	Upper bound	H-Bond Observed?
amine NH_2 a	triazine N1	**0.904**	0.881	0.923	× p
amine NH_2 b	triazine N1	**0.886**	0.858	0.909	×
amine NH_2 a	triazine N2	**0.883**	0.853	0.908	×
amine NH_2 b	triazine N2	**0.862**	0.826	0.892	✓p
amine NH_2 a	C=O	**0.793**	0.725	0.848	✓c
OH	triazine N1	**0.787**	0.734	0.832	×
amine NH_2 b	C=O	**0.761**	0.684	0.823	✓c
OH	triazine N2	**0.748**	0.685	0.802	✓c
amine NH_2 a	OMe	**0.693**	0.553	0.804	× c
amine NH_2 a	OH	**0.684**	0.601	0.757	✓c
amine NH_2 b	OMe	**0.651**	0.505	0.773	×c
amine NH_2 b	OH	**0.641**	0.551	0.723	✓c
OH	C=O	**0.602**	0.504	0.692	× c
amine NH_2 a	triazine N4	**0.563**	0.463	0.657	✓p
amine NH_2 b	triazine N4	**0.515**	0.414	0.615	×
OH	OMe	**0.470**	0.324	0.620	× c
OH	OH	**0.459**	0.365	0.556	× c
OH	triazine N4	**0.335**	0.245	0.440	× c
amine NH_2 a	Cl15	**0.316**	0.273	0.362	×
amine NH_2 b	Cl15	**0.277**	0.234	0.323	×
amine NH_2 a	Cl16	**0.274**	0.232	0.321	×
amine NH_2 b	Cl16	**0.238**	0.197	0.284	×
OH	Cl15	**0.154**	0.123	0.190	× c
OH	Cl16	**0.129**	0.101	0.163	× c

Figure S10. Hydrogen bonds as dashed lines in the lamotrigine butylated methylparaben cocrystal a) form I b) form II.



a b