

A prolific solvate former, galunisertib, under the pressure of crystal structure prediction, produces ten diverse polymorphs

Rajni M. Bhardwaj[§], Jennifer A. McMahon[§], Jonas Nyman^{§#}, Louise S. Price[†], Sumit Konar^{||}, Iain D.

H. Oswald[‡], Colin R. Pulham^{||}, Sarah L. Price[†] and Susan M. Reutzel–Edens^{§}*

[§]Small Molecule Design & Development, Eli Lilly and Company, Indianapolis, Indiana 46285, United States

[#]School of Pharmacy, University of Wisconsin – Madison, 777 Highland Ave, Madison, Wisconsin 53705, United States

[†]Department of Chemistry, University College London, 20 Gordon Street, London WC1H 0AJ, U.K.

[‡]Strathclyde Institute of Pharmacy and Biomedical Sciences, University of Strathclyde, 161 Cathedral St, Glasgow G4 0RE, U.K.

^{||}School of Chemistry, University of Edinburgh, David Brewster Road, Edinburgh, UK EH9 3FJ

Electronic Supporting Information

Table of Contents

Table of Contents.....	2
1 Solid Form Discovery.....	3
1.1 Solubility / Suspension Stability.....	3
1.2 Solution Crystallization Screen	4
1.3 Desolvation Screen	11
1.4 Crystallization under Pressure	15
1.5 Preparation of Neat GAL Polymorphs: Best Recipes	17
2 Solid-State Characterization	18
2.1 Single Crystal X-ray Structures.....	19
2.2 Powder X-ray Diffraction	25
2.3 Solid-State NMR Spectroscopy	34
2.4 Differential Scanning Calorimetry.....	37
2.5 Gravimetric Vapor Sorption Analysis.....	39
2.6 Thermodynamic Stability from Slurrying Experiments.....	41
3 Crystal Structure Prediction	42
3.1 CrystalPredictor/CrystalOptimizer (CSP1, Ψ_{mol})	42
3.1.1 Conformational Analysis and Construction of Conformational Energy Grids	42
3.1.2 Other CSP search parameters.....	45
3.1.3 Refinement of crystal structures	46
3.1.4 Inclusion of Experimental Crystal Structures, Desolvated Solvates and Evaluation of the Monohydrate Energy.....	47
3.1.5 An Analysis of CSP Generated Low Energy Structures.....	50
3.2 GRACE (CSP2, Ψ_{crys}).....	52
3.2.1 Structure Generation and Optimization	52
3.2.2 Identification of Candidate Structures using Experimental PXRD Patterns.....	53
3.3 Summary of CSP results	58
4 Crystal Structure Analysis / Solid Form Risk Assessment	60

4.1	Packing Similarity / Analysis.....	60
4.2	Hydrogen Bonding Propensity / Topology.....	61
4.3	Full Interaction Maps.....	64
4.4	Conformational Strain Energy / Molecular Surface Area	65
5	Abbreviations.....	67
6	References	68

1 Solid Form Discovery

The neat galunisertib (**GAL**) polymorphs are named in the order of their discovery (e.g., Form I, II, etc.). Solvates are named according to the solvent of crystallization and numbered in the order of their discovery.

1.1 Solubility / Suspension Stability

Solubility data were measured in commonly used processing solvents and selected aqueous-organic solvent mixtures according to a standardized slurry-solubility screening protocol. **GAL** monohydrate was suspended in diverse organic and aqueous-organic solvent mixtures. All suspensions were agitated at 250 RPM for 24 hours at 25 °C. After 2 and 24 hours of agitation, aliquots were withdrawn from the suspensions through 0.45 micron Millex PTFE filters, diluted with mobile phase and analyzed by HPLC. Residues were air dried and analysed by PXRD. The samples were analyzed using an Agilent 1100 HPLC with a Zorbax RX C-18 (15 cm × 4.6 mm, 5.0 micron particle size) column. The mobile phase consisted of methanol, 50% and 0.1% TFA/water, 50%. Instrument conditions were as follows: flow rate 1.0 mL/minute, wavelength of detection 235 nm, and column temperature 40 °C. A stock solution (1.0 mg/mL) of **GAL** was prepared in methanol and diluted with mobile phase to give concentrations suitable for quantification of the solubility samples. The results of the slurry-solubility screen are summarized in Table S1. The monohydrate was the predominant form, observed in the residues from all aqueous slurries, as well as those of some hygroscopic organic solvents. At least seven crystalline solvates were produced in this initial phase of the solid form screen.

Table S1. Solubility (N=1) of Galunisertib at 25 °C.^a

Solvent(s) (v/v)	a_w	Concentration (mg/mL)		Solid Form (PXRD)	Solvent(s) (v/v)	a_w	Concentration (mg/mL)		Solid Form (PXRD)
		2 h	24 h				2 h	24 h	
MeOH	-	8.51	12.11	MeOH solvate II	IPA-H ₂ O (1:3)	0.94	1.06	1.07	MH
EtOH	-	11.64	11.85	MH + III	IPA-H ₂ O (1:1)	0.89	6.43	15.69	MH
IPA	-	9.36	4.01	IPA solvate II	IPA-H ₂ O (3:1)	0.83	7.23	19.55	MH
1-butanol	-	14.06	3.66	1-BuOH solvate	acetone-H ₂ O (1:3)	0.94	1.06	1.03	MH
2-butanol	-	13.99	3.07	2-BuOH solvate	acetone-H ₂ O (1:1)	0.87	5.69	12.34	MH
acetone	-	6.99	9.76	MH	acetone-H ₂ O (3:1)	0.79	6.53	15.23	MH
ACN	-	5.19	4.20	IV	ACN-H ₂ O (1:3)	0.93	1.31	1.34	MH
1,4-dioxane	-	5.08	4.71	1,4-dioxane solvate I	ACN-H ₂ O (1:1)	0.91	5.61	10.63	MH
THF	-	6.96	6.83	THF solvate II	ACN-H ₂ O (3:1)	0.92	5.8	14.22	MH + peak
EtOAc (wet)	0.98	8.51	12.64	MH	acetic acid-H ₂ O (3:1)	0.91	10.97	>75	--
NMP	-	30.90	34.57	NMP solvate I	acetic acid-H ₂ O (1:1)	0.76	>75	>75	--
DMSO	-	>75	>75	--	acetic acid-H ₂ O (1:3)	0.52	>75	>75	--
DMF	-	>75	>75	--	DMSO-H ₂ O (1:3)	0.89	0.80	0.73	MH
H ₂ O	1.0	0.05	0.05	MH	DMSO-H ₂ O (1:1)	0.67	2.24	1.99	MH
acetic acid	-	>120	>120	--	DMSO-H ₂ O (3:1)	0.33	10.92	11.63	MH
MeOH-H ₂ O (1:3)	0.88	0.70	0.61	MH	DMAC-H ₂ O (1:3)	0.94	1.45	1.53	MH
MeOH-H ₂ O (1:1)	0.73	5.29	11.18	MH	DMAC-H ₂ O (1:1)	0.83	4.64	5.39	MH
MeOH-H ₂ O (3:1)	0.51	9.30	28.82	MeOH solvate II	DMAC-H ₂ O (3:1)	0.60	21.08	27.79	MH
EtOH-H ₂ O (1:3)	0.92	1.00	0.97	MH	THF-H ₂ O (1:3)	0.96	3.54	3.49	MH
EtOH-H ₂ O (1:1)	0.85	6.54	17.50	MH	THF-H ₂ O (1:1)	0.94	27.91	30.25	MH
EtOH-H ₂ O (3:1)	0.73	---	29.49	MH	THF-H ₂ O (3:1)	0.92	29.65	33.66	MH

^a Starting material: galunisertib monohydrate.

1.2 Solution Crystallization Screen

The solution crystallization screen of galunisertib encompassed solvent evaporation, cooling, antisolvent addition, vapor diffusion and slurry annealing experiments. Solvents were selected with the aid of solvent diversity mapping to increase the success rate of discovering new solid forms, based on the premise that crystallization from solution may be increased if solvents with diverse properties are surveyed. The specific conditions under which the compound could be recrystallized were determined not only by the solubility of **GAL**, but also by the suitability of specific solvents for the different types of recrystallization experiments. In addition to crystallization from solution, **GAL** was screened for hydrates specifically using moisture sorption analysis and slurry techniques, while thermal analysis (ramped temperature) and isothermal annealing were used to screen for high temperature forms and desolvates. Details are provided in the following sections.

Suspensions of the **GAL** forms were stirred in several solvents at different temperatures for varying amounts of time. Solid products were recovered by vacuum filtration, air-dried and analyzed by

PXRD as dry powders. The results are summarized in Table S2. These experiments produced many more **GAL** solvates, and also identified stable forms in the different solvents.

Table S2. Summary of Slurry Screening Experiments.

Starting Form(s)	Experimental Conditions		Final Form(s)
	Solvent(s)	Temp (°C)	
MH	MeOH	5 or RT	MeOH solvate I, MeOH solvate III
MH	3:1 MeOAc-MeOH	30, 50	MeOH solvate I
MH	EtOH	RT	EtOH solvate I, III
MH	4:1 EtOH-heptane	50	EtOH solvate I
MH	1-PrOH, 1:4 1-PrOH-heptane	RT	1-PrOH solvate
MH	IPA	25	IPA solvate I
MH	1-BuOH	RT	1-BuOH solvate I
MH, II, IV	S-(+) or RS-2-BuOH	RT	2-BuOH solvate I
MH	iAmOH	60,30,25	iAmOH solvate I
MH	acetone	RT	IV, VIII, acetone solvate I, acetone solvate II
MH	MEK	RT	II, MEK solvate I, MEK solvate II
MH	MIBK	RT	MH
MH	MIPK	RT	IV
MH, II, IV, VI, VII	ACN	RT	ACN solvate I
MH, II, IV	EtOAc	RT	EtOAc solvate I, EtOAc solvate II
MH	EtOAc	78	IV
MH, VI	nBuOAc	RT-100	IV
IV, MH	THF	RT	THF solvate I, THF solvate II, II
MH	DCM	5	DCM solvate III
II, IV, MH	DCM	RT	DCM solvate II
IV	6:1 DCM-DMSO	40	II
II	heptane	80	IV
II, IV	nitromethane	RT	nitromethane solvate I
II, IV	toluene	RT	toluene solvate I
MH	morpholine	60,30,25	morpholine solvate
MH	morpholine	65	morpholine solvate II
MH	1,4-dioxane	25	1,4-dioxane solvate I
MH, II, III, IV, 1,4-dioxane solvate I	1,4-dioxane	RT	1,4-dioxane solvate III
MH, IV	1,4-dioxane	65	1,4-dioxane solvate IV
MH	NMP	RT	NMP solvate II
MH	NMP	65	NMP solvate III
II, III, IV, V, VI	H ₂ O	RT	MH
MH	DMSO-toluene	RT	DMSO solvate

Suspensions/solutions of **GAL** were heated with stirring to dissolve the starting material, then filtered through 0.45 μm syringe filters into clean vials. The vials, covered with parafilm perforated with a pinhole or left loosely capped, were either placed in a fume hood to allow the solvents to slowly evaporate at RT, stored in the refrigerator at 5 $^{\circ}\text{C}$, or transferred to a temperature controlled hotplate for the solutions to evaporate at 50 $^{\circ}\text{C}$. Solutions from failed cooling and antisolvent addition crystallization experiments were also evaporated to induce crystallization. Residues were analyzed after either evaporation of the solvent to dryness or isolation from solution by vacuum filtration. The evaporative crystallization results are summarized in Table S3.

Table S3. Summary of Evaporative Crystallization Experiments.

Experimental Conditions		Final Form (PXRD)	Experimental Conditions		Final Form (PXRD)
Solvent	Temp ($^{\circ}\text{C}$)		Solvent	Temp ($^{\circ}\text{C}$)	
MeOH	RT	MH	70:1 MeOH-H ₂ O	RT	MeOH solvate I
IPA	RT	IPA solvate I	70:1 EtOH-H ₂ O	RT	MH
1-PrOH	RT	1-PrOH solvate	70:1 IPA-H ₂ O	RT	IPA solvate I
1-butanol	RT	1-BuOH solvate	70:1 1-PrOH-H ₂ O	RT	1-PrOH solvate II
2-BuOH	RT	MH + 2-BuOH solvate	70:1 1-PrOH-H ₂ O	RT	1-PrOH solvate I
RS-2butanol	RT	RS-2-BuOH solvate	1:3 1-PrOH-H ₂ O	RT	MH
iAmOH	RT	iAmOH solvate	70:1 2-BuOH-H ₂ O	RT	2-BuOH solvate
acetone	5	IV	70:1 2-BuOH-H ₂ O	RT	MH + 2-BuOH solvate
MEK	RT	MEK solvate I	70:1 acetone-H ₂ O	RT	MH
1:10 DMF-MIPK	RT	MH	70:1 ACN-H ₂ O	RT	MH
ACN	5	ACN solvate I	70:1 1,4-dioxane-H ₂ O	RT	1,4-dioxane solvate I
ACN	RT	MH	70:1 THF-H ₂ O	RT	MH
ACN	RT	IV	3:1 acetic acid-H ₂ O	RT	acetic acid solvate
THF	RT	THF solvate II	wet EtOAc	RT	amorphous
THF	RT	THF solvate II	1,4-dioxane	RT	1,4-dioxane solvate I
2-methylTHF	5, RT	2-MeTHF solvate I	1,4-dioxane	RT	1,4-dioxane disolvate + hemi-1,4-dioxane hydrate
cyclohexane	RT	film	DMSO	RT	MH
3:1 MeOH-H ₂ O	RT	MH	DCM	RT	DCM solvate II

Solvents were chosen for the cooling crystallization screen based on the melting and boiling points and solvent diversity. Programmed cooling crystallization experiments were performed using a Crystal16™ parallel crystallizer, equipped with programmable heating/cooling, magnetic stirring and turbidity sensors. The suspensions were stirred at 700 rpm and heated to 50–70 $^{\circ}\text{C}$ at 0.4–0.5 $^{\circ}\text{C min}^{-1}$, equilibrated for 15 minutes, then cooled to 10 $^{\circ}\text{C}$ at 0.1 or 0.2 $^{\circ}\text{C min}^{-1}$. Temperature cycling experiments were conducted for a subset of solvents using two or three heat-cool cycles. The results of the cooling crystallization screen are summarized in Table S4.

Table S4. Summary of Cooling / High Temperature Crystallization Experiments.

Solvent(s) (v/v)	Cooling Rate (°C/min)	Isothermal Hold Temp (°C)	Solid Form(s)
EtOH	0.5		III, EtOH solvate I
MeOH	0.5		MeOH solvate I
ACN	0.1		IV
acetone	0.1		Acetone solvate II
1:4 1-PrOH-MIPK	natural		1-PrOH solvate
2.5:1 t-BuOH-MeOH	natural		t-BuOH solvate
1:2 ACN-toluene			II + peaks
2:1 ACN-toluene	natural		II, ACN solvate
1:2 EtOH-H ₂ O	natural		MH
2:1 THF-H ₂ O	natural		MH
1.6:1 THF-H ₂ O	natural		MH
5:1 acetone-H ₂ O	natural		MH
2:1 acetone-H ₂ O	natural		MH
2.5:1 acetone-H ₂ O	natural		MH
2:1 ACN-H ₂ O	natural		MH
17:1 nBuOAc-DMSO	natural		nBuOAc solvate II (new)
20:1 iPrOAc-DMSO	0.1		IV, iPrOAc solvate I
17:1 nBuOAc-DMSO	0.1		IV, VI, IX, nBuOAc solvate I, nBuOAc solvate II
15:1 MIBK-DMSO	0.1		IV
18:1 MIBK-DMSO	0.1		IV
17:1 nBuOAc-DMSO	natural		IV, nBuOAc solvate II
20:1 MIBK-DMSO	1		IV, VI, IX, MIBK solvate I, MIBK solvate II
20:1 MIBK-DMSO		0, 20	IV
20:1 MIBK-DMSO		40	IV, IX
20:1 MIBK-DMSO		60	IV, VI
20:1 MIBK-DMSO		80	IV

Form IX, rarely seen in the solid form screen, was produced on a few occasions at small scale from 17:1 n-butyl acetate (nBuOAc) – dimethyl sulfoxide (DMSO). Many attempts were made to reproduce the original recrystallization procedures at larger scale without success. Attempts to seed solutions prepared using amorphous **GAL** with Form IX crystals were only occasionally successful, see Table S5.

Table S5. Summary of Seeded Crystallization Experiments.

Method ^a	Experimental Conditions			Final Form(s) (PXRD)
	Solvent(s) (v/v)	Seed Form	Seed Temp (°C)	
AA	nBuOAc // pentane	V	RT	II + pk @ 10.8
CC	20:1 MIBK-DMSO	VI	35	IV, VI
CC	ACN	VII	50	IV, VI
CC	17:1 nBuOAc-DMSO	IX	30	IX
CC	17:1 nBuOAc-DMSO	IX	40	IX
CC	17:1 nBuOAc-DMSO	IX	80	IV
CC	17:1 nBuOAc-DMSO	IX	40	IV
Slurry ^b	heptane	IX	RT	MH

^a AA: antisolvent addition, CC: cooling crystallization; ^bSM: amorphous suspension.

The crystallization behaviour of **GAL** from 17:1 nBuOAc – DMSO solutions as a function of nucleation temperature and level of supersaturation was examined; the results are shown in Figure S1. Forms IV, VI and IX were obtained, but inconsistently so. No procedure emerged to reliably produce Form IX, despite hundreds of attempts. When Form IX was observed, it was in phase pure form. However, in the presence of any of the other polymorphs, it rapidly disappeared.

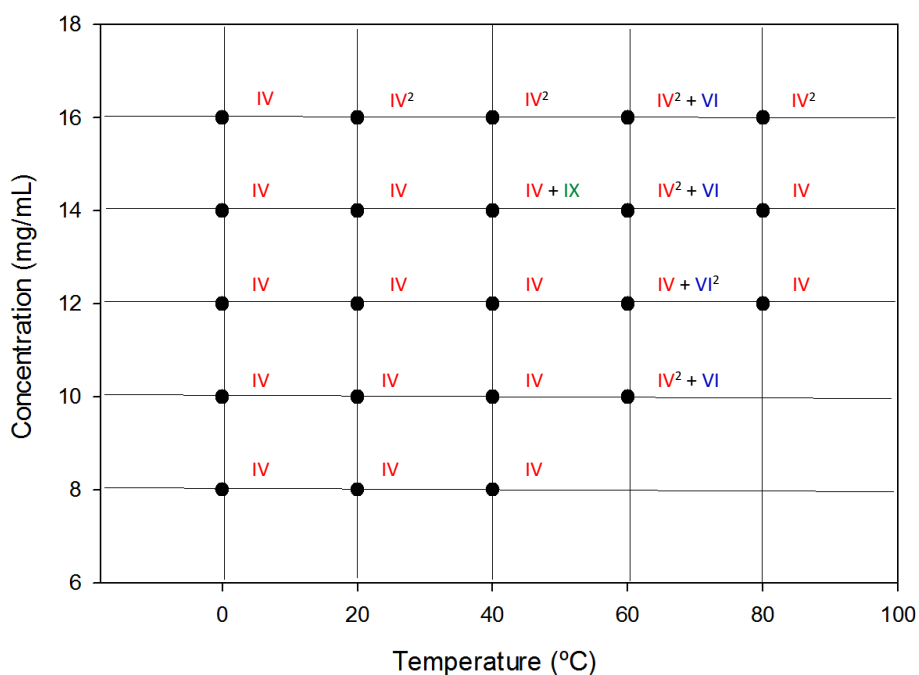


Figure S1. Results of isothermal 'one solvent' crystallization screen from 17:1 nBuOAc – DMSO. The superscript denotes the number of observations.

The antisolvent addition crystallization screen was designed from solvent systems that were selected based on the requirements of solvent/antisolvent miscibility, while ensuring solvent diversity. **GAL** was weighed into vials, then solvents were added by pipette. To ensure complete dissolution of the

drug substance prior to antisolvent addition, the suspensions were usually agitated on a J-KEM shaker block at 250 RPM, then filtered. Standard antisolvent addition experiments were conducted at RT. Antisolvent was added dropwise until either persistent clouding was observed or the maximum antisolvent volume (20 mL) was dispensed. Solutions, wherein precipitation was not observed, were subsequently evaporated. The solid products were isolated by vacuum filtration, then air-dried at RT. Reverse antisolvent addition experiments were also performed, where **GAL** solutions were filtered into flasks or vials containing antisolvent. The work-up procedures were the same as those used in the standard antisolvent addition experiments.

Vapor-liquid diffusion experiments were set up at ambient temperature by placing the **GAL** solutions in closed chambers containing antisolvent. The solid products were recovered by decantation of the mother liquor and air-dried. Vapor diffusion onto solid amorphous **GAL** was also performed. **GAL** was rendered amorphous by either quench cooling a melt or lyophilization. In the absence of solvent, no crystallization was observed (by PXRD) from purely amorphous **GAL** over a time period of 3 months. The results of the antisolvent addition and vapor diffusion crystallization screens are summarized in Table S6.

Table S6. Summary of Antisolvent Addition/Vapor Diffusion Experiments.

Solvent	Antisolvent	Method ^a	Solid Form(s)
MeOH	H ₂ O	VL	MH
EtOH	hexane	VL	EtOH solvate I
1-BuOH	pentane	VL	1-BuOH solvate I
iAmOH	pentane	VL	iAmOH solvate I
ACN	CHCl ₃	VL	IV, VI
ACN	hexane	VL	IV, VI
ACN	toluene	VL	MH
EtOAc	hexane	VL	EtOAc solvate II
EtOAc	toluene	VL	II
nBuOAc	pentane	VL	IV, VI, MH
DCM	hexane	VL	DCM solvate I, DCM solvate III
DCM-IPA	heptane	AA	II , IPA solvate I
DMSO	MIPK	AA	MH , DMSO solvate
1:2 DMSO-acetone	H ₂ O	VL	MH
2-BuOH	heptane	AA	2-BuOH solvate
ACN	hexane	RAA	IV, VI
acetone	toluene	RAA	IX , toluene solvate I
acetone	hexane	RAA	IV, VIII , acetone solvate III
acetone	hexane	VL	acetone solvate III
acetone	toluene	VL	IV, VI
MEK	hexane	VL	MEK solvate I
MIBK	heptane	VL	IV
MIBK	toluene	VL	IV, VI

MIBK	heptane	RAA	IV
MIBK	toluene	RAA	IX , toluene solvate I
17:1 nBuOAc-DMSO	toluene	RAA	IV , nBuOAc solvate I, DMSO solvate I
17:1 nBuOAc-DMSO	hexane	RAA	IV
17:1 nBuOAc-DMSO	heptane	VL	IV
17:1 nBuOAc-DMSO	toluene	VL	IV, VI
THF	hexane	VL	THF solvate I
toluene	hexane	VL	toluene solvate I
morpholine	hexane	VL	morpholine solvate II
1,4-dioxane	toluene	VL	mono 1,4-dioxane solvate
acetic acid	H ₂ O	VL	MH
1:5 acetic acid-hexane	hexane	VL	acetic acid solvate
NMP	CHCl ₃	VL	amorphous
NMP	heptane	VL	mono-NMP solvate
NMP	H ₂ O	VL	MH , NMP solvate II
DMSO	H ₂ O	VL	MH
DMAC	H ₂ O	VL	MH
--	nBuOAc	VS	IV , nBuOAc solvate I
--	MIBK	VS	VI
--	MeOH	VS	MeOH solvate III
--	1,4-dioxane	VS	1,4-dioxane solvate I, 1,4-dioxane solvate III
--	ACN	VS	IV
--	n-pentane	VS	IV

^aAA: antisolvent addition, RAA: reverse antisolvent addition, VL: vapor liquid diffusion, VS: vapor solid diffusion onto amorphous SM.

Eight neat polymorphs (II-IX) and more than 50 solvates of **GAL** were produced in the solution phase of the experimental solid form screen. Some of the neat forms (IV, VI and IX) crystallized directly from solution, while others (II, III, V, VII and VIII) were produced upon isolating and drying various product solvate crystals. The phase compositions of all materials were established by PXRD. All experiments that yielded new phases were repeated at larger scale to produce more material for further characterization and to fund desolvation studies. Multiple attempts were made to produce the neat forms as pure phases (Section 1.5). Single crystals suitable for structure determination were recovered, when possible. SCXRD analysis confirmed the identity of some of the neat polymorphs (Forms II, III, IV, VI) and many of the solvates. The remainder were characterized by their unique PXRD patterns (Figure S2) and ssNMR spectra. The solvent in the solvate crystals was identified, and in many cases quantified, using a combination of thermogravimetric analysis and solution ¹H NMR spectroscopy.

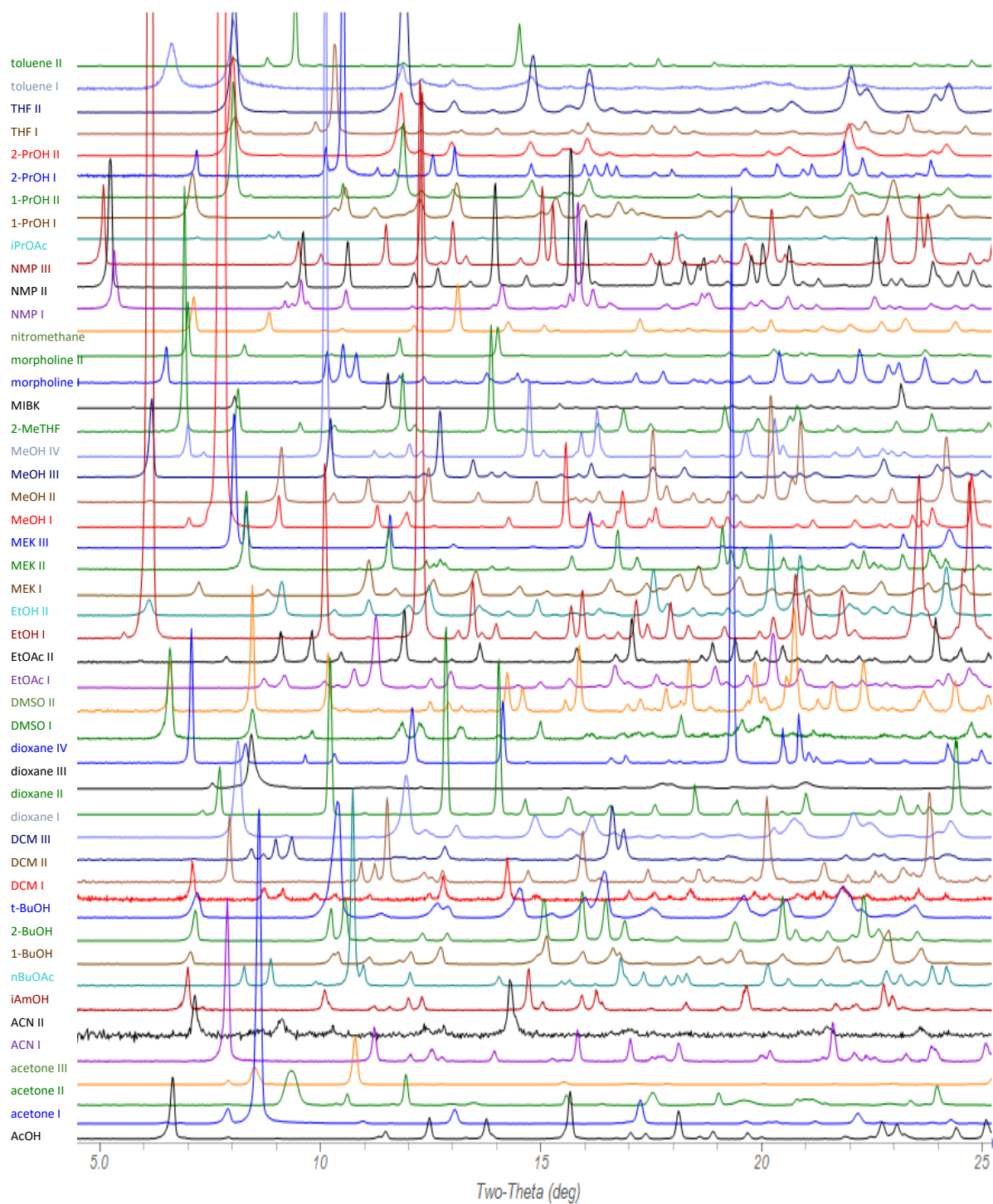


Figure S2. PXRD patterns of GAL solvates.

1.3 Desolvation Screen

Neat polymorphs were observed throughout the solid form screen, some directly crystallized from solution and others produced on isolation from the mother liquors or after air drying various **GAL** solvates. A systematic desolvation screen was conducted to search for new anhydrous polymorphs

and establish desolvation pathways that might lead to specific crystal forms. The results of the desolvation screen of 27 **GAL** solvates are summarized in Table S7. Mild drying by exposure of the solvates to vacuum at room temperature revealed which solvates were appreciably stable in the solid state. The most aggressive drying condition, slurrying in heptane at 90 °C, produced mainly Form IV, which suggests this form is most stable at this temperature. Intermediate drying conditions produced a range of neat **GAL** polymorphs, oftentimes as mixtures. The appearance of a few of the forms seems to be linked to specific parent solvates, e.g. monohydrate to Form I, MeOH solvate I to Form V, acetonitrile solvate I to Form VII, and acetone solvate I to Form VIII. It is noteworthy that under harsher drying conditions, these parent solvates will transform to primarily Forms IV and VI.

Table S7. Summary of Desolvation Experiments.

Starting Form	TG-DTA (1 °C/min)	TG-DTA (20 °C/min)	Heptane slurry (90 °C)	Parent solvent slurry (elevated temp)	Vac Dry (ambient)	Vac Dry (120 °C)
Monohydrate	◇+✕+★	◇+✕+◇+★	◇	◇	○	✕+◇+★
MeOH solvate I	◆	◇	◇	S _{MeOH} I	S _{MeOH} I	◆
MeOH solvate III	-	-	-	-	-	✕
EtOH solvate I	✕ + pks	✕ + pks	◇	S _{EtOH} I	S _{EtOH} II	✕
1-PrOH solvate	-	-	-	-	-	◇+★
IPA solvate I	◇	◇	◇	S _{IPA} I	S _{IPA} I	◇
nBuOH solvate I	◇	◇	◇	S _{nBuOH} I	S _{nBuOH} I	◇
2-BuOH solvate II	◇	◇	◇	◇	S _{2-BuOH} II	◇
iAmOH solvate I	◇	◇	◇	-	S _{iAmOH} I	◇
MEK solvate I	◇+★	◇+★	◇	S _{MEK} I + ◇ _{tr}	S _{MEK} I	◇+★
MEK solvate III	◇	◇	-	-	-	◇
THF solvate II	◇	◇	◇	S _{THF} II	◇	◇
2-MeTHF solvate	S _{2-MeTHF}	S _{2-MeTHF}	-	-	-	★
DCM solvate I	-	-	-	-	-	◇+★
DCM solvate II	★	★	◇	S _{DCM} II	S _{DCM} II	◇+★
morpholine solvate I	◇+★	◇+★	◇	S _{Morph} II	S _{Morph} I	◇+★
morpholine solvate II	★	★+◇ _{min}	◇	-	S _{Morph} II	★
EtOAc solvate II	◇	◇	◇	◇	S _{EtOAc} II	◇
NMP solvate II	◇+★	◇+★ _{tr}	◇+SNMP II _{tr}	-	S _{NMP} II	◇
NMP solvate III	◇	◇+SNMP III _{tr}	◇	-	S _{NMP} III	◇+pk
ACN solvate I	▼	▼	◇+pk	◇	S _{ACN} I	▼
acetone solvate I	★	★	◇+S _{acet} I + S _{acet} II + □ (60 °C)	S _{acet} I + □ (45 °C)	S _{acet} I + □	★+□ (80 °C)
acetone solvate II	◇	◇	◇	S _{acet} I	S _{acet} II	◇
1,4-dioxane disolvate	◇+◇ _{tr}	◇	◇	S _{diox} IV	◇+◇ _{tr}	◇+◇ _{tr}
1,4-dioxane solvate IV	★	★	◇	S _{diox} IV	S _{diox} IV	★
toluene solvate	◇+★ _{min}	★+◇ _{tr}	◇	◇+★+?	◇+★	◇
nitromethane solvate I	◇+★	◇	◇+pk	S _{Nitro} I + ?	◇+★	◇+★

○ = Form I; ◇ = Form II; ✕ = Form III; ◇ = Form IV; ◆ = Form V; ★ = Form VI; ▼ = Form VII; □ = Form VIII.

GAL monohydrate was dried at 40, 50 and 60 °C to evaluate the effect of temperature and vacuum on its dehydration kinetics and crystal form conversion pathways, Figure S3. The water of crystallization in **GAL** monohydrate could be partially removed at 40 °C over a period of hours (dark blue curve). Only with 60 °C vacuum drying did the monohydrate appear to be completely

dehydrated, with a total water weight loss of $\approx 4.6\%$ (theory for monohydrate = 4.65%) being observed. Ambient temperature moisture sorption analysis of the [partially] dried monohydrate revealed that in all instances, except 60 °C vacuum drying, the water rapidly returned to reform the monohydrate crystal structure at 5% RH. The fully dehydrated material, on the other hand, did not sorb water to reform the monohydrate until the RH was raised to 15% RH.

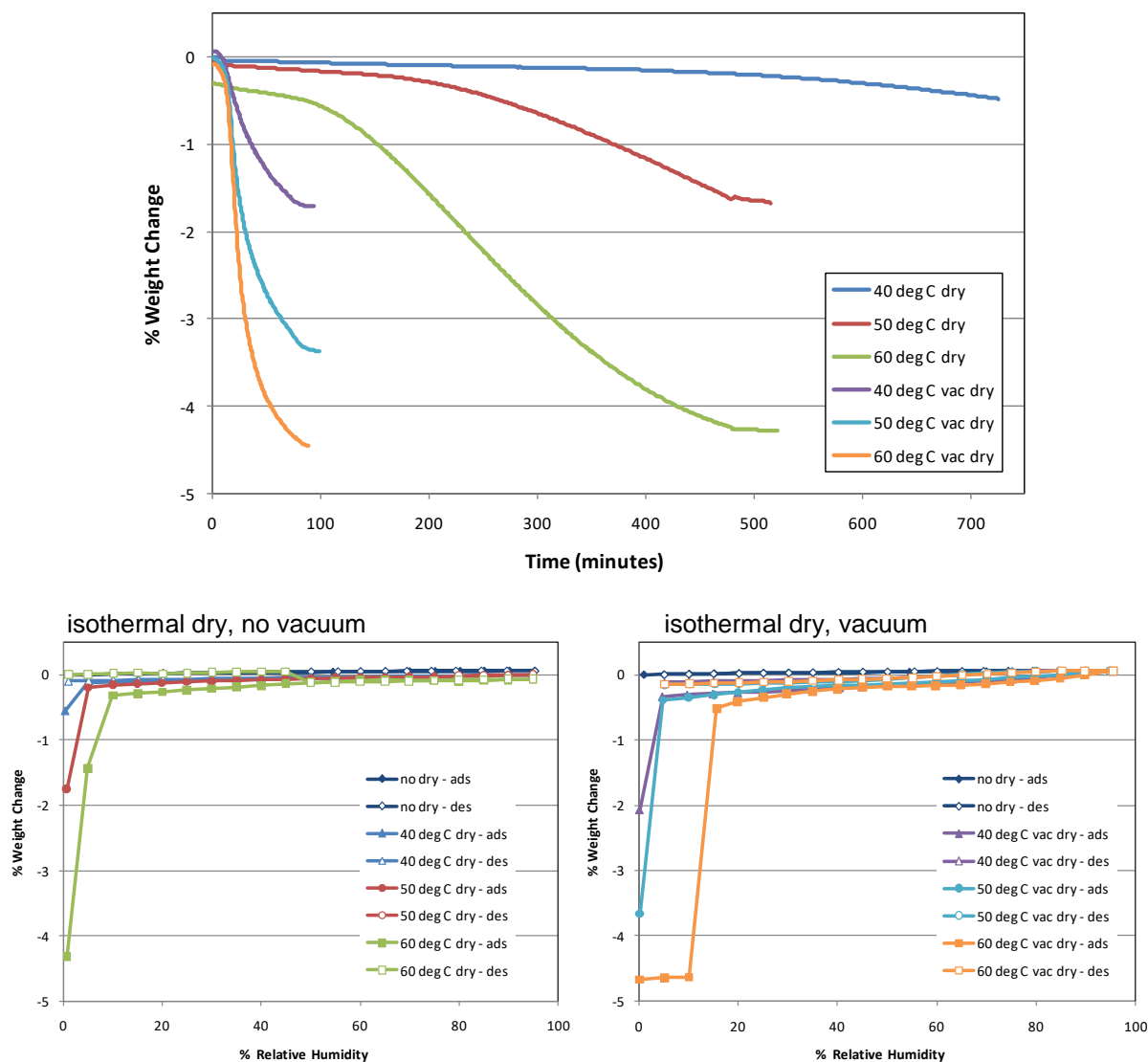


Figure S3. Top: Isothermal drying curves of GAL monohydrate with and without vacuum; bottom: 25 °C moisture sorption isotherms after drying the monohydrate at 40, 50 and 60 °C with (right) and without (left) vacuum.

Following the observation that the monohydrate could be completely dehydrated within a couple of hours at 60 °C *in vacuo* and would not immediately rehydrate, variable temperature PXRD was used to identify the dehydration product(s) at elevated temperature, as well as after re-equilibration at ambient temperature and RH. Material changes were observed by *in situ* PXRD analysis of the monohydrate at elevated temperatures, but the PXRD pattern changed only slightly and anisotropic lattice expansion on heating could not be ruled out. Dehydration of the monohydrate to Form I

(which should occur according to the TGA results) appeared to be reversible as the original pattern was observed when the sample was returned to ambient temperature and RH. At 60 °C, a change in dehydration mechanism (dehydration to Form IV instead of Form I) was confirmed. Whereas **GAL** monohydrate reversibly dehydrates within hours at 50 to 80 °C, above ≈ 100 °C, it irreversibly dehydrates to Forms III, IV and/or VI, Figure S4. Only with drying at 100 °C for 1-2 hours were all three neat forms, III, IV and VI, detected by PXRD post-analysis.

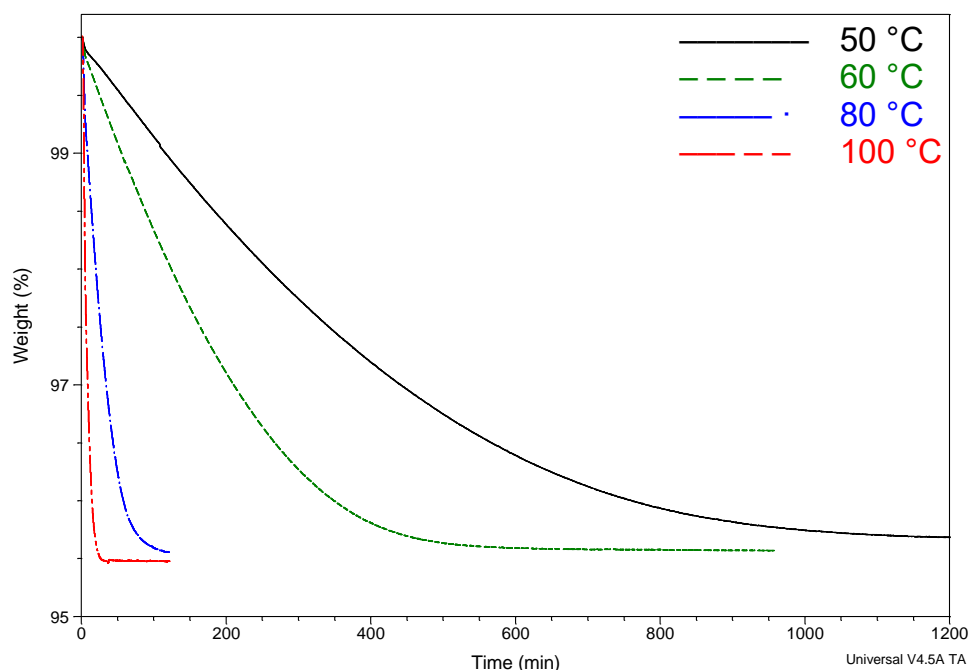


Figure S4. Isothermal drying curves of **GAL** monohydrate.

GAL monohydrate, as the thermodynamically stable crystal form at ambient temperature above at least $\approx 14\%$ RH, will not spontaneously dehydrate at RT to one or more of the anhydrous polymorphs. Above RT, however, the monohydrate is expected to become less stable relative to the neat polymorphs. Therefore to evaluate the effect that Forms III, IV and/or VI could have on the dehydration kinetics and pathways of **GAL** monohydrate, analogous drying experiments were conducted in the presence of Form III, IV and VI. A physical mixture of monohydrate and Forms III, IV and VI was first prepared by annealing **GAL** monohydrate at 100 °C, then subjected to the same isothermal drying conditions as the pure monohydrate. Forms III, IV and VI seed crystals appeared to have no effect on the course of the 50 °C drying experiment, as the phase composition was the same before and after drying. Seed crystals of Forms III, IV and VI were, on the other hand, found to accelerate the irreversible dehydration of the monohydrate at temperatures as low as 60 °C. These results, in showing that the neat forms, if present, may compromise the stability of the monohydrate during drying, confirm the greater thermodynamic stability of Forms III, IV and VI relative to the monohydrate at temperatures as low as 60 °C.

1.4 Crystallization under Pressure

The results from the CSP studies (Section S3.1) imply the possible existence of stable, higher density forms, and this was the motivation for an experimental exploration of the solid form landscape using both pressure and temperature, on the basis that previous studies have demonstrated that higher density forms can often be found at elevated pressures.¹ High-pressure crystallization experiments were performed using a Merrill-Bassett diamond anvil cell² (DAC) from Almax EasyLab [half-opening angle of 40°], equipped with 600 μm cut diamonds and a tungsten carbide backing disc. Small pieces of ruby were loaded in the DAC near the sample as a pressure calibrant with a measurement precision of 0.05 GPa. The pressure within the gasket hole was determined by the ruby fluorescence method³ using a Jobin-Yvon LabRam 300 spectrometer equipped with a 50 mW He-Ne laser of wavelength 632.8 nm. The attempts to crystallize **GAL** under pressure are summarized in Table S8.

Table S8. Summary of the Experiments Conducted at Higher Pressures.

Direct Compression				
Pressure (GPa)	Temp (°C)	Starting Form	PTM	Result
0 – 6	RT	II	1:1 isopentane: n-pentane	No phase transition
			4:1 methanol : ethanol	Solvate formation upon addition of PTM; No pressure-induced phase changes of solvate
		IV	4:1 methanol : ethanol	Solvate formation upon addition of PTM; No pressure-induced phase changes of solvate
		VI	1:1 isopentane: n-pentane	No phase transition
			4:1 methanol : ethanol	Solvate formation upon addition of PTM; No pressure-induced phase changes of solvate
Crystallization from saturated solution				
Pressure (GPa)	Temp (°C)	Starting Form	Solvent	Result
0 – 4.5	≈200	IV	ethanol	Form III crystallized directly from ethanol after release of pressure to ambient pressure
1	≈150	V	acetone	Single crystal of Form IV started to grow from clear solution at 1 GPa; upon cooling the pressure dropped to ambient pressure
Crystallization from melt				
Pressure (GPa)	Temp (°C)	Starting Form	Result	
Ambient	≈170	V	Single crystal of Form VI was obtained	
0.4	170	II	Polycrystalline Form VI was obtained	
	170	V		
0.4	240	V	Single crystal of Form X was obtained; recovered to ambient pressure	

Initial experiments focused on direct compression of selected forms (II, IV, and VI) at ambient temperature using different pressure-transmitting media (PTM), monitored by PXRD on the high-pressure beamline at the ESRF. When methanol/ethanol was used as the pressure-transmitting medium, the formation of putative solvates was observed at ambient pressure. With mixed pentanes, no phase transitions were observed under these conditions up to 6 GPa. This behavior is not unexpected and has been observed for acetaminophen where direct compression of Form I at ambient temperature results in only partial formation (no transformation by SCXRD) of Form II (the thermodynamically favourable form at elevated pressures) on account of the substantial kinetic barrier associated with molecular rearrangement in the solid state.⁴ It is likely that such barriers will be substantially greater for larger, more conformationally complex molecules such as **GAL**.

For this reason, a series of experiments were conducted that involved crystallization from solution at elevated pressure. In a typical experiment, a diamond-anvil cell (DAC) was loaded with a saturated solution of **GAL** dissolved in a suitable solvent. On pressurization, the reduction in solubility can result in precipitation of a high-pressure form as polycrystalline material. By judicious temperature cycling it is then often possible to grow a single crystal *in situ* that is suitable for single crystal X-ray diffraction. This methodology has previously been demonstrated as an effective technique for obtaining new high-pressure forms.⁵ While in principle this is a straightforward technique, complications can arise in practice from pressure-induced freezing of the solvent, slow nucleation kinetics caused by the high viscosity of pressurized solutions, and challenges associated with precise control of temperature, pressure and associated solubility. On crystallization from acetone solution at an initial pressure of ≈ 1 GPa, a single crystal of Form IV was obtained and identified by *in situ* single crystal X-ray diffraction. However, it should be noted that the pressure in the DAC dropped substantially during the crystallization process and so it is not possible to be certain of the exact pressure at which the crystal was grown. Attempts to crystallize from ethanol solution at elevated pressures (≈ 2 GPa) were initially unsuccessful – no nucleation occurred at elevated pressures. However, reduction of the pressure in the DAC to almost ambient pressure resulted in the growth of a single crystal that could be recovered from the DAC and which was subsequently identified as Form III using single crystal X-ray diffraction. It is likely that the reduction of pressure reduced the viscosity of the solution, thereby encouraging nucleation and crystal growth.

Despite numerous attempts to grow crystals from solution at elevated pressures, it proved very difficult to grow high-quality single crystals and no new high-pressure phases were obtained using this technique. Instead, we chose to investigate the effects of temperature on compressed samples of **GAL** in the absence of any solvent, on the basis that the input of thermal energy could overcome any kinetic barriers to molecular rearrangement in the solid state. Clearly one potential danger is thermal decomposition of the compound, but past experience with other molecular compounds indicates that decomposition of such compounds can often be inhibited at elevated pressures.⁶

Finely ground samples of Form II and Form V were loaded into DACs in the absence of any liquid and were pressurized to 0.4 GPa. PXRD patterns were recorded over the temperature range 25-190 °C using Beamline I15 at the Diamond Light Source. For both samples, at 170 °C the 2D-patterns started to become more textured, possibly indicating partial melting and recrystallization. At 190 °C, both patterns indicated the complete transition of the starting forms to Form VI. Of particular note was the lack of any obvious sign of thermal decomposition, in marked contrast to heating experiments conducted on the compound at ambient pressure and exposed to air. This prompted a subsequent experiment that involved melting of a sample of Form V to 170 °C under a nitrogen atmosphere at

ambient pressure. No signs of decomposition were observed and on slow cooling, single crystals were obtained that were identified as Form VI by single crystal X-ray diffraction.

In a final experiment, a finely ground sample of Form V placed in a 250 μm sized hole drilled in a pre-indented stainless steel gasket (thickness = 250 μm) was loaded into a diamond anvil cell. No pressure transmitting medium was employed to avoid any potential interaction with the sample. The sample was pressurized to 0.4 GPa and the DAC was left for 1 h at that pressure to equilibrate. The sample was then heated to 240 $^{\circ}\text{C}$, above the melting point of Form V, but below that of several of the other polymorphs, using a hot air gun; the temperature was monitored using a K-type thermocouple attached to the gasket. Changes inside the sample chamber during heating were observed using an optical microscope. On reaching 240 $^{\circ}\text{C}$, the sample melted and the cell was allowed to slowly cool to ambient temperature. A single crystal grew out of the melt to fill the gasket hole (Figure S5), during which time a small drop in pressure (to 0.3 GPa) was also observed. From *in situ* single-crystal X-ray diffraction data collected on beamline I19 at the Diamond Light Sources, it was possible to solve and refine the crystal structure as a new polymorph, Form X. On disassembling the DAC, the crystal survived decompression and it was therefore possible to collect at ambient pressure full sets of diffraction data for Form X at both 298 K and 100 K.

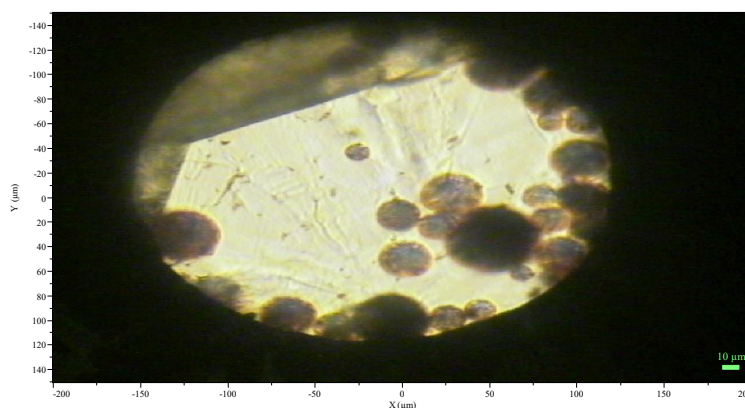


Figure S5. Single crystal of **GAL** Form X grown at 0.4 GPa in a diamond anvil cell, along with ruby spheres for pressure calibration.

1.5 Preparation of Neat **GAL** Polymorphs: Best Recipes

Form I: **GAL** monohydrate (25 mg) packed in a 0.7 mm borosilicate glass capillary was stored at 0% RH for more than 3 weeks until the sample weight stabilized. The capillary was rapidly sealed and PXRD data of the dehydration product collected. Highly metastable Form I stored at 0% RH was also packed in a ceramic rotor for solid-state NMR spectroscopy analysis and hermetically-sealed in an aluminium pan for DSC analysis.

Form II: Form IV (100 mg) was slurried in tetrahydrofuran (1 mL) for 2 days. The product tetrahydrofuran solvate was isolated by vacuum filtration and dried at ambient temperature under vacuum overnight, yielding Form II. Form II crystals were also grown by vapor diffusion of toluene into an ethyl acetate solution (prepared by dissolving 20 mg of **GAL** monohydrate in 500 μL ethyl acetate) at room temperature and by slow evaporation of a filtered 1,4-dioxane solution (containing 100 mg of monohydrate dissolved in 10 mL of 1,4-dioxane) at room temperature.

Form III: GAL monohydrate (220 mg) was suspended in ethanol (5 mL), then placed into a Crystalline PVM reactor (Technobis, Alkmaar, The Netherlands). The slurry was heated to 60 °C (3 °C/min) with stirring (300 rpm), held isothermally for 45 minutes, cooled to 5 °C (1 °C/min), then held at this temperature for 18 hours. The solids were isolated via vacuum filtration and air-dried. The product ethanol solvate I (+ trace monohydrate) was desolvated in a 120 °C vacuum oven for 3 hours.

Form IV: GAL monohydrate (200 mg) was slurried in acetonitrile (1.5 mL) at 60 °C for up to 24 hours. The solid product was isolated by vacuum filtration and air-dried.

Form V: GAL monohydrate (260 mg) was suspended in methanol (4 mL), heated with stirring to 60 °C, then cooled to RT and allowed to stir over a period of approximately 4 hours. The product methanol solvate was vacuum filtered, then dried at 130 °C for approximately 1.5 hours.

Form VI: Form IV (200 mg) was suspended in dichloromethane (1 mL) for 4 hours. The product dichloromethane solvate was isolated by vacuum filtration, air-dried, and annealed in a convection oven at 120 °C for 4 hours, yielding Form VI in pure form.

Form VII: Form IV (200 mg) was slurried in acetonitrile (2 mL) for 7 days. The solid product was isolated by vacuum filtration, air-dried, then annealed in a convection oven at 120 °C for 5 hours. Note: this process often yielded mixtures of Forms IV, VI and VII.

Form VIII: GAL monohydrate (487 mg) was suspended in acetone (9 mL) and allowed to stir at RT for 2 days. The slurry was transferred to a 50 °C heated block, stirred overnight, then cooled to RT. The solid product, acetone solvate I, was dried in a vacuum oven at 70 °C for approximately 5.5 hours. Conversion to mostly Form VIII (with a trace of acetone solvate I) was confirmed via PXRD. All attempts to produce phase pure Form VIII were unsuccessful.

Form IX: GAL Form IX was produced on multiple occasions at small scale by cooling 17:1 n-butyl acetate : dimethyl sulfoxide solutions containing 7-10 mg/mL of amorphous galunisertib at cooling rates of 0.1-20 °C/min in a Crystal16 parallel reactor system (Technobis, Alkmaar, The Netherlands). The starting material was prepared as follows: GAL was rendered amorphous by dissolving the monohydrate (100.2 mg) in 3:1 acetonitrile:H₂O (20 mL), filtering the solution into a clean, dry round bottom flask, then lyophilizing for approximately 3 hours. Note: This procedure was not reproducible, with many failed attempts instead yielding an n-butyl acetate solvate or mixtures of Forms IV and IX.

Form X: A finely ground sample of Form V was loaded into a DAC in the absence of any pressure-transmitting medium and was pressurized to 0.4 GPa. The sample was then heated to 240 °C, at which point the sample had completely melted. On slow cooling to ambient temperature, a single crystal grew to fill the gasket hole.

2 Solid-State Characterization

The atom numbering schemes used in the experimental and CSP1 / CSP2 crystal structures (.cifs) are summarized in Figure S6.

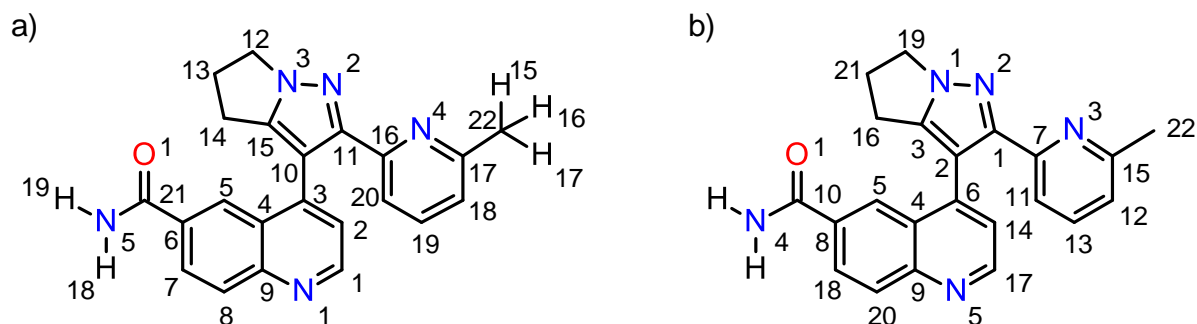


Figure S6. Atomic numbering in the a) experimental and CSP1 and b) CSP2 **GAL** crystal structures. CH hydrogen atoms are omitted for clarity, with the exception of those used in conformational analysis.

2.1 Single Crystal X-ray Structures

Single crystals of the **GAL** forms were grown by slow evaporation, cooling crystallization or vapor diffusion using various solvent systems.

Three-dimensional X-ray diffraction data (ϕ - and ω -scans) were collected on a Bruker three-circle diffractometer coupled to a Bruker Photon-1000 CMOS detector using either Mo K α radiation ($\lambda = 0.71073 \text{ \AA}$) from a fine focus sealed tube equipped with a graphite monochromator or Cu K α radiation ($\lambda = 1.54178 \text{ \AA}$) from an μ S microsource. All data were collected at 100 K using an Oxford Cryosystems Cryostream. Unit cell refinement and data reduction were performed using SAINT⁷ v6.2/v8.32B, while scaling and absorption corrections were performed with SADABS or TWINABS. All structures were solved by dual-space methods using the program SHELXT and refined against F^2 on all data with SHELXL-2014/SHELXL-2018 using established refinement techniques.⁸ All non-hydrogen atoms were refined anisotropically. All carbon-bound hydrogen atoms were placed in geometrically calculated positions and refined using a riding model while constraining their U_{iso} to 1.2 times the U_{eq} of the atoms to which they are bound (1.5 times for methyl groups). Coordinates for hydrogen atoms bound to nitrogen or oxygen were taken from the difference Fourier synthesis maps and those hydrogen atoms were subsequently refined semi-freely with the help of distance restraints (target values 0.84(2) \AA for O—H and 0.88(2) for N—H distances).

Form X: High-pressure diffraction data were collected using synchrotron radiation of wavelength $\lambda = 0.4859 \text{ \AA}$ at room temperature on a Newport IS4CCD (4 circle) diffractometer with a Pilatus 300 K detector at beamline I19 at the Diamond Light Source, Harwell Science and Innovation Campus.⁹ Integrations were performed using the program CrysAlisPro¹⁰ and absorption corrections using the program ABSPACK.¹¹ The structure was partially solved using XT¹² and full atom connectivity was found through cycles of full-matrix least squares refinement and addition of atoms from a Fourier difference map. RIGU and AFIX restraints were used to aid the molecular connectivity.¹³

The diamond anvil cell was slowly depressurized by loosening the screws and the crystal was extracted from the top of the one of the diamond surface. The crystal was protected from moisture and air by immersing it into paraffin oil. Ambient pressure data were collected on a Rigaku Oxford Diffraction SuperNova diffractometer with Cu K α radiation. An Oxford Cryosystems Cryostream 700+ low-temperature device was used to maintain a crystal temperature of 100 K. Refinements were carried out using the high pressure structure as the starting model. The structure was refined with version 2016/6 of SHELXL interfaced with OLEX2.¹⁴ All non-hydrogen atoms were refined using

anisotropic displacement parameters. Carbon-bound hydrogen atoms were placed in calculated positions geometrically and refined using the riding model as implemented in OLEX2.

A summary of crystal properties and data/refinement statistics is given for 5 neat polymorphs, the monohydrate and 20 solvates in Table S9.

Table S9. Single Crystal Data of Galunisertib Crystal Forms.

	Form II	Form III	Form IV	Form VI	Form X
Empirical formula	C ₂₂ H ₁₉ N ₅ O	C ₂₂ H ₁₉ N ₅ O	C ₂₂ H ₁₉ N ₅ O	C ₂₂ H ₁₉ N ₅ O	C ₂₂ H ₁₉ N ₅ O
Formula weight (g/mol)	369.42	369.42	369.42	369.42	369.42
Temperature (K)	100(2)	100(2)	100(2)	100(2)	100(1)
Wavelength (Å)	1.54178	1.54178	1.54178	1.54178	1.54178
Crystal size /mm	0.27 × 0.04 × 0.01	0.46 × 0.20 × 0.08	0.27 × 0.07 × x0.02	0.25 × 0.20 × 0.15	0.12 × 0.08 × 0.01
Crystal system	triclinic	triclinic	monoclinic	monoclinic	monoclinic
Space group	<i>P</i> -1	<i>P</i> -1	<i>P</i> 2 ₁ / <i>n</i>	<i>P</i> 2 ₁ / <i>n</i>	<i>C</i> 2/ <i>c</i>
<i>a</i> (Å)	8.2116(11)	7.7119(14)	18.9981(11)	10.7270(3)	27.938(4)
<i>b</i> (Å)	11.5104(14)	8.9111(17)	9.8532(6)	14.4711(5)	8.0817(7)
<i>c</i> (Å)	11.8746(18)	14.617(2)	21.8829(13)	12.7050(4)	22.004(3)
α (°)	102.001(10),	81.906(13)	90	90	90
β (°)	105.368(11)	78.799(14)	113.219(3)	107.1120(17)	133.55(3)
γ (°)	102.349(9)	75.258(15)	90	90	90
Volume (Å ³)	1014.6(3)	948.5(3)	3764.5(4)	1884.91(10)	3600.7(13)
Molecules per cell (Z)	2	2	8	2	8
Calculated density (g cm ⁻³)	1.209	1.293	1.304	1.302	1.363
Absorption coeff (mm ⁻¹)	0.848	0.666	0.671	0.670	0.045
<i>F</i> ₀₀₀	388	388	1552	776	1552
Data collection θ range (°)	4.023 to 72.693	3.096 to 72.365	2.617 to 72.305	4.754 to 67.677	8.11 to 112.57
Index ranges	-10 ≤ <i>h</i> ≤ 10 -13 ≤ <i>k</i> ≤ 14 -14 ≤ <i>l</i> ≤ 14	-9 ≤ <i>h</i> ≤ 9 -8 ≤ <i>k</i> ≤ 10 -18 ≤ <i>l</i> ≤ 17	-23 ≤ <i>h</i> ≤ 23 -12 ≤ <i>k</i> ≤ 12 -27 ≤ <i>l</i> ≤ 26	-12 ≤ <i>h</i> ≤ 12 -16 ≤ <i>k</i> ≤ 17 -15 ≤ <i>l</i> ≤ 15	-30 ≤ <i>h</i> ≤ 30 -8 ≤ <i>k</i> ≤ 8 -23 ≤ <i>l</i> ≤ 23
Meas/independent/ obsd [<i>I</i> > 2 σ (<i>I</i>)] reflections	21129/3946/1847	10787/3532/3049	91186/7426/6150	19870/3340/2743	8913/2345/-
Completeness (%)	99.3	95.7	99.8	98.1	99.2
Data/restraints/parameters	3946 / 0 / 255	3532 / 0 / 255	7426 / 0 / 508	3340 / 79 / 263	2345/0/242
Goodness-of-fit on <i>F</i> ²	1.026	1.139	1.008	1.088	1.068
Final R indices [<i>I</i> > 2 σ (<i>I</i>)]	R1 = 0.1506 wR2 = 0.3149	R1 = 0.0598 wR2 = 0.1348	R1 = 0.0416 wR2 = 0.0947	R1 = 0.0521 wR2 = 0.1391	R1 = 0.0680 wR2 = 0.1533
Final R indices (all data)	R1 = 0.2542 wR2 = 0.3669	R1 = 0.0701 wR2 = 0.1421	R1 = 0.0541 wR2 = 0.1011	R1 = 0.0632 wR2 = 0.1489	R1 = 0.1077 wR2 = 0.1760
Largest peak diff/hole (e Å ⁻³)	1.122/-0.412	0.269/-0.239	0.453/-0.271	0.222/-0.252	0.28/-0.29

Table S9 - cont. Single Crystal Data of Galunisertib Crystal Forms.

	Form X	Form X (0.3 GPa)	Monohydrate	MeOH Solvate I	EtOH Solvate I
Empirical formula	C ₂₂ H ₁₉ N ₅ O	C ₂₂ H ₁₉ N ₅ O	C ₂₂ H ₂₁ N ₅ O ₂	C ₂₃ H ₂₃ N ₅ O ₂	C ₂₄ H ₂₅ N ₅ O ₂
Formula weight (g/mol)	369.42	369.42	387.44	401.46	415.49
Temperature (K)	290.1(2)	293	100(2)	100(2)	296(2)
Wavelength (Å)	1.54178	0.4859	0.71073	1.54178	1.54178
Crystal size /mm	0.12 × 0.08 × 0.01	0.15 × 0.15 × 0.12	0.16 × 0.03 × 0.03	0.23 × 0.17 × 0.04	0.22 × 0.18 × 0.15
Crystal system	Monoclinic	monoclinic	monoclinic	monoclinic	triclinic
Space group	C2/c	C2/c	<i>P</i> 2 ₁ / <i>n</i>	<i>P</i> 2 ₁ / <i>n</i>	<i>P</i> -1
<i>a</i> (Å)	27.879(5)	27.654(13)	9.0505(7)	8.0458(4)	8.1100(4)
<i>b</i> (Å)	8.3092(10)	8.108(8)	11.3624(9)	22.6024(10)	11.6569(6)
<i>c</i> (Å)	22.025(4)	21.934(11)	19.1323(16)	11.2097(5)	12.7599(6)
α (°)	90	90	90	90	90.026(4)
β (°)	133.44(3)	133.41(8)	94.954(3)	106.274(2)	99.248(4)
γ (°)	90	90	90	90	101.238(4)
Volume (Å ³)	3704.6(17)	3573(5)	1960.1(3)	1956.85(16)	1167.15(10)
Molecules per cell (<i>Z</i>)	8	8	4	4	2
Calculated density (g cm ⁻³)	1.325	1.374	1.313	1.363	1.182
Absorption coeff (mm ⁻¹)	0.044	0.046	0.088	0.727	0.625
<i>F</i> ₀₀₀	1552	1552	816	848	440
Data collection θ range (°)	8.108 to 112.488	4.04 to 28.12	3.007 to 26.296	4.551 to 67.828	3.511 to 68.015
Index ranges	-30 ≤ <i>h</i> ≤ 30	-27 ≤ <i>h</i> ≤ 27	-11 ≤ <i>h</i> ≤ 11	-9 ≤ <i>h</i> ≤ 9	-9 ≤ <i>h</i> ≤ 9
	-8 ≤ <i>k</i> ≤ 8	-4 ≤ <i>k</i> ≤ 5	-14 ≤ <i>k</i> ≤ 14	-26 ≤ <i>k</i> ≤ 27	-13 ≤ <i>k</i> ≤ 13
	-23 ≤ <i>l</i> ≤ 23	-21 ≤ <i>l</i> ≤ 21	-23 ≤ <i>l</i> ≤ 23	-13 ≤ <i>l</i> ≤ 13	-15 ≤ <i>l</i> ≤ 14
Meas/independent/ obsd [<i>I</i> > 2 σ (<i>I</i>)] reflections	9183/2395/-	4412/1072/-	18939/3946/2930	21629/3485/3155	15304/4025/2652
Completeness (%)	98.9	57.2	99.7	98.0	95.1
Data/restraints/parameters	2395/0/242	1072/237/242	3949 / 0 / 272	3485 / 0 / 285	4025 / 0 / 276
Goodness-of-fit on <i>F</i> ²	0.997	1.105	1.032	1.075	1.053
Final R indices [<i>I</i> > 2 σ (<i>I</i>)]	R1 = 0.0899	R1 = 0.1157	R1 = 0.0356	R1 = 0.0468	R1 = 0.0992
	wR2 = 0.1876	wR2 = 0.2699	wR2 = 0.0907	wR2 = 0.1210	wR2 = 0.2630
Final R indices (all data)	R1 = 0.1895	R1 = 0.2209	R1 = 0.0481	R1 = 0.0501	R1 = 0.1317
	wR2 = 0.2442	wR2 = 0.3500	wR2 = 0.1051	wR2 = 0.1247	wR2 = 0.2947
Largest peak diff/hole (e ⁻ Å ⁻³)	0.23/-0.36	0.23/-0.30	0.425/-0.295	0.237/-0.331	0.681/-0.708

Table S9 - cont. Single Crystal Data of Galunisertib Crystal Forms.

	EtOH Solvate II	1-PrOH Solvate	IPA Solvate I	MEK Solvate III	Acetic Acid Solvate III
Empirical formula	C ₂₄ H ₂₅ N ₅ O ₂	C ₂₅ H ₂₇ N ₅ O ₂	C ₂₅ H ₂₇ N ₅ O ₂	C ₂₆ H ₂₇ N ₅ O ₂	C ₂₄ H ₂₃ N ₅ O ₃
Formula weight (g/mol)	415.49	429.51	429.51	441.52	429.47
Temperature (K)	100(2)	100(2)	100(2)	101(2)	103(2)
Wavelength (Å)	0.71073	1.54178	1.54178	1.54178	1.54178
Crystal size /mm	0.40 × 0.17 × 0.06	0.10 × 0.08 × 0.03	0.17 × 0.05 × 0.03	0.10 × 0.10 × 0.02	0.12 × 0.06 × 0.05
Crystal system	triclinic	triclinic	triclinic	monoclinic	triclinic
Space group	P-1	P-1	P-1	P2 ₁ /c	P-1
a (Å)	8.1074(10)	8.0480(8)	7.7720(11)	12.1495(14)	9.5331(11)
b (Å)	9.2583(12)	11.4825(13)	11.8319(18)	8.4511(9)	10.9792(13)
c (Å)	15.519(2)	12.8727(13)	12.8267(17)	20.111(2)	11.6291(17)
α (°)	73.278(4)	89.280(7)	89.030(9)	90	66.668(12)
β (°)	89.994(4)	80.416(7)	74.293(8)	93.551(7)	65.944(8)
γ (°)	84.496(5)	79.313(7)	82.225(10)	90	78.915(9)
Volume (Å ³)	1110.0(2)	1152.4(2)	1124.8(3)	2061.0(4)	1019.9(2)
Molecules per cell (Z)	2	2	2	4	2
Calculated density (g cm ⁻³)	1.243	1.238	1.268	1.423	1.399
Absorption coeff (mm ⁻¹)	0.082	0.649	0.665	0.743	0.775
F ₀₀₀	440	456	456	936	452
Data collection θ range (°)	2.838 to 26.518	3.483 to 72.555	3.58 to 72.889	3.645 to 72.438	4.388 to 72.770
Index ranges	-10 ≤ h ≤ 9 -11 ≤ k ≤ 11 -19 ≤ l ≤ 19	-9 ≤ h ≤ 9 -14 ≤ k ≤ 14 -15 ≤ l ≤ 15	-9 ≤ h ≤ 9 -14 ≤ k ≤ 14 -15 ≤ l ≤ 15	-14 ≤ h ≤ 14 -10 ≤ k ≤ 10 -24 ≤ l ≤ 24	-11 ≤ h ≤ 11 -13 ≤ k ≤ 13 -14 ≤ l ≤ 14
Meas/independent/ obsd [I > 2σ(I)] reflections	32945/4557/4012	24662/4508/3052	30708/4395/3105	65214/4070/3248	27197/3976/2501
Completeness (%)	99.7	99.3	99.3	100	99.2
Data/restraints/parameters	4557/0/294	4508 / 0 / 304	4395/ 0 / 296	4070 / 4 / 284	3976/ 2 / 295
Goodness-of-fit on F ²	4.975	1.451	1.041	1.067	0.957
Final R indices [I > 2σ(I)]	R1 = 0.0464 wR2 = 0.0951	R1 = 0.1318 wR2 = 0.3763	R1 = 0.0576 wR2 = 0.1152	R1 = 0.0861 wR2 = 0.12038	R1 = 0.0869 wR2 = 0.2061
Final R indices (all data)	R1 = 0.0533 wR2 = 0.0957	R1 = 0.1675 wR2 = 0.3959	R1 = 0.0927 wR2 = 0.1280	R1 = 0.1047 wR2 = 0.2165	R1 = 0.1389 wR2 = 0.2359
Largest peak diff/hole (e ⁻ Å ⁻³)	0.481/-0.423	0.495/-0.449	0.216/-0.272	2.964/-1.191	0.290/-0.666

Table S9 - cont. Single Crystal Data of Galunisertib Crystal Forms.

	NMP Solvate	1-BuOH Solvate	RS-2BuOH Solvate	S-2BuOH Solvate	iAmOH Solvate I
Empirical formula	C ₂₇ H ₂₈ N ₆ O ₂	C ₂₆ H ₂₉ N ₅ O ₂	C ₂₆ H ₂₉ N ₅ O ₂	C ₂₆ H ₂₉ N ₅ O ₂	C ₂₇ H ₃₁ N ₅ O ₂
Formula weight (g/mol)	468.55	443.54	443.54	443.54	457.57
Temperature (K)	100(2)	100(2)	100(2)	100(2)	100(2)
Wavelength (Å)	1.54178	1.54178	1.54178	1.54178	1.54178
Crystal size /mm	0.53 × 0.13 × 0.09	0.16 × 0.16 × 0.05	0.29 × 0.16 × 0.12	0.20 × 0.07 × 0.06	0.15 × 0.08 × 0.03
Crystal system	monoclinic	triclinic	triclinic	triclinic	triclinic
Space group	P2 ₁ /c	P-1	P-1	P1	P-1
a (Å)	9.9660(6)	8.1145(8)	8.1840(8)	8.1676(2)	8.0879(6)
b (Å)	33.730(2)	12.0611(13)	12.0219(12)	11.9112(3)	12.3578(9)
c (Å)	7.3404(5)	12.5401(13)	12.6533(12)	12.7014(3)	12.8909(9)
α (°)	90	88.454(5)	89.543(6)	89.4745(11)	85.722(3)
β (°)	196.892(2)	79.842(4)	81.238(6)	81.2360(14)	77.875(3)
γ (°)	90	76.813(5)	75.349(6)	75.9211(11)	74.707(3)
Volume (Å ³)	2361.0(3)	1176.1(2)	1189.8(2)	1184.05(5)	1214.83(15)
Molecules per cell (Z)	4	2	2	2	2
Calculated density (g cm ⁻³)	1.318	1.252	1.238	1.244	1.251
Absorption coeff (mm ⁻¹)	0.693	0.651	0.644	0.647	0.645
F ₀₀₀	992	472	472	472	488
Data collection θ range (°)	2.620 to 72.247	3.581 to 72.386	3.536 to 72.781	3.522 to 68.019	3.507 to 72.436
Index ranges	-12 ≤ h ≤ 12 -41 ≤ k ≤ 41 -9 ≤ l ≤ 9	-10 ≤ h ≤ 10 -14 ≤ k ≤ 14 -15 ≤ l ≤ 15	-10 ≤ h ≤ 10 -14 ≤ k ≤ 14 -15 ≤ l ≤ 15	-9 ≤ h ≤ 9 -14 ≤ k ≤ 14 -15 ≤ l ≤ 15	-10 ≤ h ≤ 9 -15 ≤ k ≤ 15 -15 ≤ l ≤ 15
Meas/independent/ obsd [I > 2σ(I)] reflections	77127/4642/4409	38535/4571/4163	24039/4601/3719	25401/7271/6566	17050/4695/4240
Completeness (%)	99.9	99.3	98.1	95.6	98.4
Data/restraints/parameters	4642 / 0 / 319	4571 / 0 / 305	4601 / 0 / 302	7271 / 3 / 609	4695 / 0 / 323
Goodness-of-fit on F ²	1.072	1.012	1.066	1.047	1.006
Final R indices [I > 2σ(I)]	R1 = 0.0757 wR2 = 0.1918	R1 = 0.0416 wR2 = 0.1053	R1 = 0.1359 wR2 = 0.3899	R1 = 0.0487 wR2 = 0.1271	R1 = 0.0369 wR2 = 0.0908
Final R indices (all data)	R1 = 0.0783 wR2 = 0.1937	R1 = 0.0454 wR2 = 0.1083	R1 = 0.1508 wR2 = 0.4002	R1 = 0.0545 wR2 = 0.1337	R1 = 0.0405 wR2 = 0.0934
Largest peak diff/hole (e ⁻ Å ⁻³)	1.882/-0.840	0.368/-0.432	0.783/-0.847	0.260/-0.264	0.251/-0.212

Table S9 - cont. Single Crystal Data of Galunisertib Crystal Forms.

	DCM Solvate II	THF Solvate II	2-MeTHF Solvate	1,4-dioxane Disolvate	Hemi-1,4-dioxane Hydrate
Empirical formula	C ₂₃ H ₂₁ Cl ₂ N ₅ O	C ₂₆ H ₂₇ N ₅ O ₂	C ₂₇ H ₂₈ N ₅ O ₂	C ₃₀ H ₃₅ N ₅ O ₅	C ₂₄ H ₂₅ N ₅ O ₃
Formula weight (g/mol)	454.35	441.52	454.54	545.63	431.49
Temperature (K)	101(2)	100(2)	100(2)	100(2)	100(2)
Wavelength (Å)	1.54178	1.54178	1.54178	1.54178	1.54178
Crystal size /mm	0.11 × 0.09 × 0.02	0.20 × 0.09 × 0.03	0.09 × 0.03 × 0.01	0.26 × 0.22 × 0.08	0.17 × 0.13 × 0.07
Crystal system	triclinic	triclinic	triclinic	triclinic	triclinic
Space group	P-1	P-1	P-1	P-1	P-1
a (Å)	8.4961(8)	9.8533(7)	8.5823(11)	10.6011(8)	7.7332(3)
b (Å)	10.3766(10)	11.2079(6)	11.0257(18)	11.6654(9)	11.4956(4)
c (Å)	12.7043(12)	11.5743(6)	12.7778(16)	12.5667(9)	12.1986(4)
α (°)	103.239(5)	101.869(2)	101.925(12)	69.035(3)	85.299(2)
β (°)	91.904(5)	105.813(2)	92.429(11)	86.603(2)	84.505(2)
γ (°)	94.505(5)	105.560(2)	97.579(12)	81.573(2)	85.518(2)
Volume (Å ³)	1085.37(18)	1129.90(12)	1169.7(3)	1435.49(19)	1073.11(7)
Molecules per cell (Z)	2	2	2	2	2
Calculated density (g cm ⁻³)	1.390	1.298	1.291	1.262	1.335
Absorption coeff (mm ⁻¹)	2.899	0.678	0.670	0.711	0.736
F ₀₀₀	472	468	482	580	456
Data collection θ range (°)	3.579 to 72.293	4.162 to 72.159	3.544 to 65.630	3.767 to 72.211	3.650 to 72.277
Index ranges	-10 ≤ h ≤ 10 -12 ≤ k ≤ 12 -15 ≤ l ≤ 15	-12 ≤ h ≤ 12 -13 ≤ k ≤ 13 -14 ≤ l ≤ 14	-10 ≤ h ≤ 10 -14 ≤ k ≤ 13 -14 ≤ l ≤ 15	-13 ≤ h ≤ 13 -14 ≤ k ≤ 14 -15 ≤ l ≤ 15	-9 ≤ h ≤ 9 -14 ≤ k ≤ 14 -14 ≤ l ≤ 15
Meas/independent/ obsd [I > 2σ(I)] reflections	26445/4233/3502	31096/4398/3934	19039/3995/1984	48029/5595/5313	35272/4199/3666
Completeness (%)	99.1	99.2	98.7	99.2	99.3
Data/restraints/parameters	4233/ 2 /309	4398 / 0 / 299	3995 / 4 / 280	5595 / 0 / 363	4199 / 0 / 299
Goodness-of-fit on F ²	1.003	0.999	1.030	1.006	1.028
Final R indices [I>2σ(I)]	R1 = 0.0522 wR2 = 0.1218	R1 = 0.0489 wR2 = 0.116	R1 = 0.1452 wR2 = 0.3410	R1 = 0.0349 wR2 = 0.0871	R1 = 0.0377 wR2 = 0.0923
Final R indices (all data)	R1 = 0.10650 wR2 = 0.1303	R1 = 0.05494 wR2 = 0.1209	R1 = 0.2389 wR2 = 0.4020	R1 = 0.0364 wR2 = 0.0882	R1 = 0.0451 wR2 = 0.0965
Largest peak diff/hole (e ⁻ Å ⁻³)	0.478/-0.468	0.815/-0.661	0.996/-0.467	0.301/-0.209	0.251/-0.229

Table S9 - cont. Single Crystal Data of Galunisertib Crystal Forms.

	Morpholine Solvate I	Hemi-EtOAc Solvate II	MEK Solvate I
Empirical formula	C ₂₆ H ₂₈ N ₆ O ₂	C ₂₄ H ₂₃ N ₅ O ₂	C ₂₆ H ₂₇ N ₅ O ₂
Formula weight (g/mol)	456.54	413.47	441.52
Temperature (K)	100(2)	100(2)	101(2)
Wavelength (Å)	1.54178	1.54178	1.54178
Crystal size /mm	0.15 × 0.10 × 0.08	0.28 × 0.20 × 0.18	0.23 × 0.08 × 0.06
Crystal system	triclinic	monoclinic	triclinic
Space group	P-1	P2 ₁ /c	P-1
a (Å)	8.6946(4)	14.7096(5)	8.3929(7)
b (Å)	10.9640(4)	16.9380(5)	11.7452(9)
c (Å)	12.9344(5)	16.8272(5)	12.0347(10)
α (°)	77.7370(16)	90	92.483(3)
β (°)	88.5612(17)	95.6389(12)	106.415(4)
γ (°)	79.1558(19)	90	97.071(4)
Volume (Å ³)	1183.21(8)	4172.2(2)	1125.52(16)
Molecules per cell (Z)	2	8	2
Calculated density (g cm ⁻³)	1.281	1.316	1.303
Absorption coeff (mm ⁻¹)	0.676	0.699	0.680
F ₀₀₀	484	1744	468
Data collection θ range (°)	3.497 to 67.723	3.019 to 67.716	3.805 to 72.479
Index ranges	-10 ≤ h ≤ 10 -13 ≤ k ≤ 13 -15 ≤ l ≤ 15	-17 ≤ h ≤ 17 -20 ≤ k ≤ 20 -20 ≤ l ≤ 20	-10 ≤ h ≤ 9 -14 ≤ k ≤ 14 -14 ≤ l ≤ 14
Meas/independent/ obsd [<i>I</i> > 2σ(<i>I</i>)] reflections	30172/4150/3735	13383/7533/7235	24288/4410/3813
Completeness (%)	96.8	99.7	99.3
Data/restraints/parameters	4150 / 0 / 320	7533 / 0 / 589	4410 / 0 / 309
Goodness-of-fit on F ²	1.080	1.046	1.039
Final R indices [<i>I</i> > 2σ(<i>I</i>)]	R1 = 0.0654 wR2 = 0.1692	R1 = 0.0548 wR2 = 0.1354	R1 = 0.0442 wR2 = 0.109
Final R indices (all data)	R1 = 0.0695 wR2 = 0.1772	R1 = 0.0560 wR2 = 0.1366	R1 = 0.0521 wR2 = 0.115
Largest peak diff/hole (e Å ⁻³)	0.301/-0.434	0.470/-0.289	0.405/-0.307

2.2 Powder X-ray Diffraction

Polycrystalline samples of various forms of **GAL** were lightly ground in a mortar, loaded into a 0.7 mm borosilicate glass capillary and mounted on a PANalytical Empyrean powder X-ray diffractometer equipped with a primary monochromator (Cu K $\alpha_{1,2}$, $\lambda = 1.54180$ Å) and PIXEL 3D detector. Data were collected in the rotating capillary at room temperature using a fixed count scheme. For each dataset, 5 or 8 scans were collected from 2-70° 2 θ with a step size of 0.07° and 500 or 800 s per step. The diffraction intensities were merged after data collection.

For Forms I, V and IX, the RT unit cell parameters were obtained by indexing the first 20 peaks in their PXRD patterns with the DICVOL program as implemented in DASH.¹⁵ For Form VII and VIII, good indexing solutions were not obtained because of phase impurities. In these cases, PXRD data were used to identify model structures from the CSP searches (Section S3.2.2), as a starting point for Rietveld refinement. The candidate crystal structures were also used to calculate ssNMR shielding parameters (Sections S.2.3). Unit cell parameters of Form I, V, VII, VIII and IX obtained by indexing or Pawley refinement of lattice parameters of the 0 K structural models obtained from CSP against the best available RT PXRD data are summarized in Table S10. Space groups were assigned from volume considerations, a statistical consideration of the systematic absences¹⁶ and after a check using the ADDSYM function in PLATON.¹⁷ A Pawley-type fit¹⁸ of unit cell and space group against the background-subtracted powder data was carried out in DASH.

Table S10. Room Temperature Unit Cells of Neat **GAL** Polymorphs

Form	Space Group	<i>a</i> (Å)	<i>b</i> (Å)	<i>c</i> (Å)	α (°)	β (°)	γ (°)
I	P2 ₁ /n	19.4346	11.3271	9.1555	90	95.3184	90
V	P2 ₁ /c	8.2670	19.6223	12.6953	90	116.5930	90
VII*	P-1	8.09128	12.43746	18.0375	86.4773	99.9738	84.2989
VIII*	P-1	8.5563	11.3754	11.5725	65.7811	105.4995	77.3124
IX	P2 ₁ /c	11.2994	20.8444	8.1057	90	98.8813	90

*Obtained from Pawley-type refinement of predicted unit cell from CSP against the mixed phase PXRD data

Simulated annealing (SA) was used to optimize the **GAL** Form I, V, and IX structures against the diffraction data set in direct space. An internal coordinate description (z-matrix) of **GAL** was derived from one of the experimental crystal structures. The O—H, N—H and C—H distances were normalized to 0.90, 0.90 and 0.95 Å, respectively. The structure was solved using 800 simulated annealing runs of 2.5×10^7 moves per run, as implemented in DASH, in which the position, orientation and conformation of the internal coordinate description of the **GAL** molecules were varied. Each **GAL** molecule was allowed 15 degrees of freedom (9 external and 6 internal). Each best solution, returning a χ^2 ratio less than 5 (χ^2 SA / χ^2 Pawley), no significant misfit to the data and a chemically reasonable packing arrangement, was subsequently refined with data in the range of 2–70° 2 θ using a rigid Rietveld refinement¹⁹ as implemented in TOPAS V4.2.²⁰ The rigid body description was derived from the z-matrix used in the simulated annealing runs.

Final fits from the rigid-body Rietveld refinement of Forms I, V, and IX are shown in Figures S7-S9. For Form I, simulated annealing has sampled both conformations of the five membered ring. To confirm the ring conformation, the DFT-optimized crystal structure from simulated annealing was used for Rietveld refinement. For Form I, the final refinement included a total of 61 parameters (40 profile, 4 cell, 1 scale, 1 isotropic temperature factor, 9 preferred orientation, 3 position and 3 rotation yielding a final R_{wp} = 9.42 (Figure S7). For Form V, the final refinement included a total of 58 parameters (38 profile, 4 cell, 1 scale, 1 isotropic temperature factor, 8 preferred orientation, 3 position and 3 rotation yielding a final R_{wp} = 4.54 (Figure S8). For Form IX, the final refinement

included a total of 57 parameters (37 profile, 8 preferred orientation, 4 cell, 1 scale, 1 isotropic temperature factor, 3 position and 3 rotation) yielding a final R_{wp} = 4.35 (Figure S9).

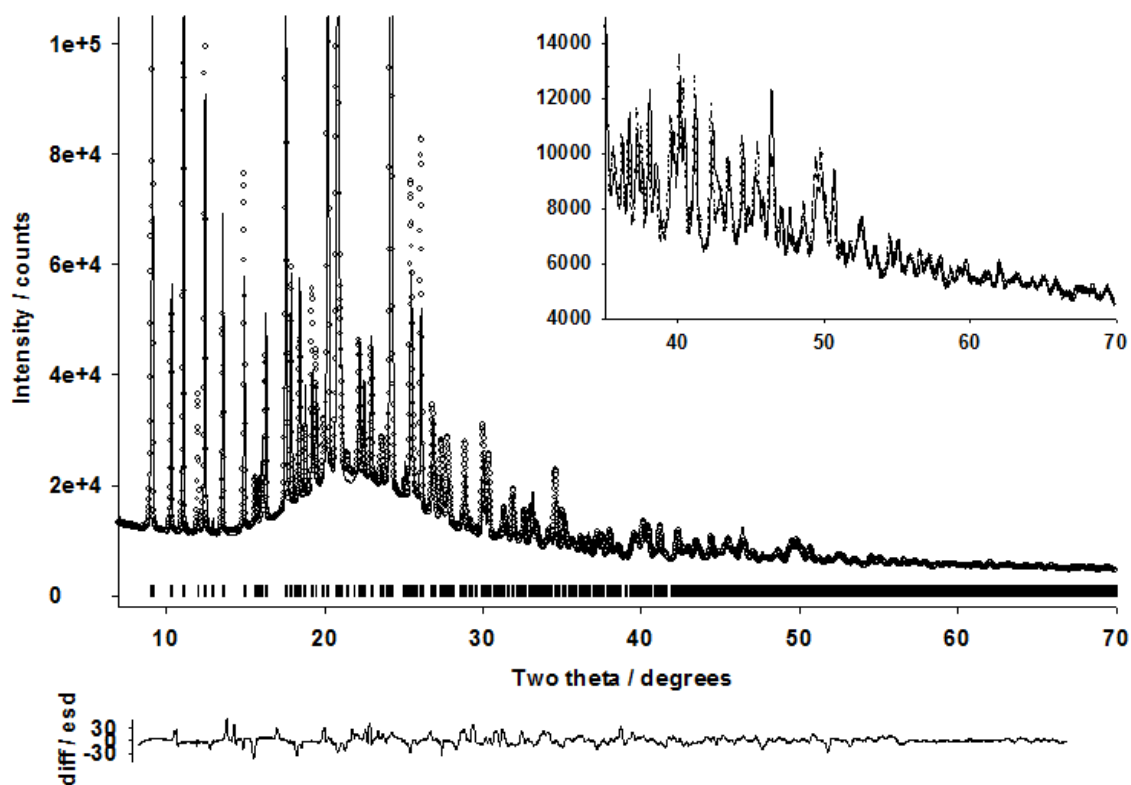


Figure S7. Final fit showing the observed yobs (lines), calculated ycalc (points) and difference [(yobs–ycalc)/sigma(yobs)] profiles for the rigid-body Rietveld refinement of the best SA solution of Form I.

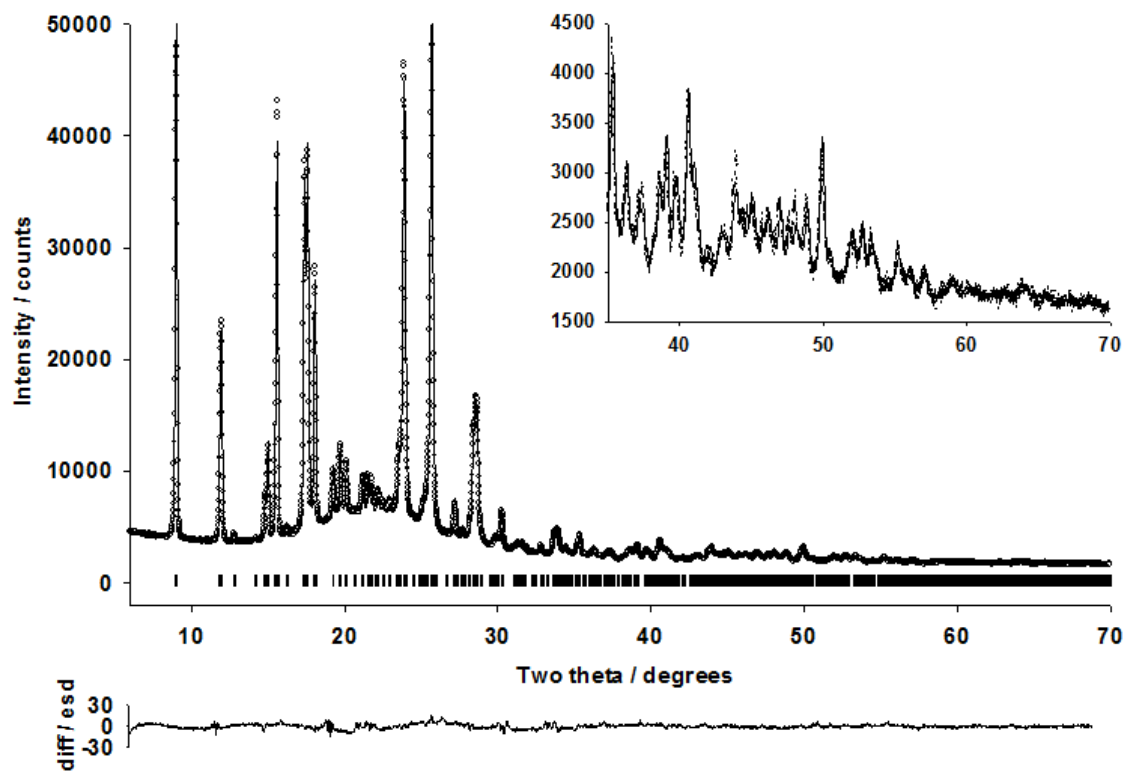


Figure S8. Final fit showing the observed y_{obs} (lines), calculated y_{calc} (points) and difference $[(y_{obs}-y_{calc})/\sigma(y_{obs})]$ profiles for the rigid-body Rietveld refinement of the best SA solution of Form V.

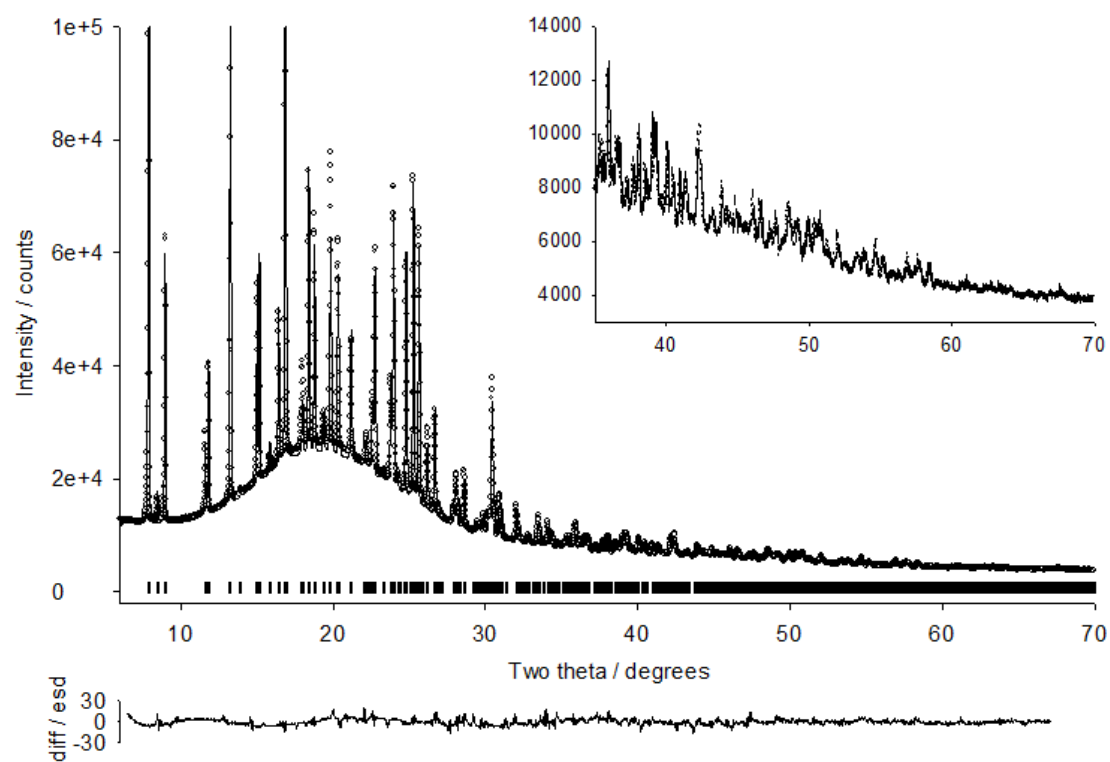


Figure S9. Final fit showing the observed y_{obs} (lines), calculated y_{calc} (points) and difference $[(y_{obs}-y_{calc})/\sigma(y_{obs})]$ profiles for the rigid-body Rietveld refinement of the best SA solution of Form IX.

For Form VII, a starting structure model (rank 2) was obtained by PXRD matching to simulated patterns of GRACE predicted structures (Section 3.2.2). Pawley refinement of this structure was not in good agreement with the PXRD pattern ($R_{wp} = 3.833$) (Figure S10). Careful examination of the unaccounted peaks suggested the sample was contaminated with Form IV. Mixed phase (Form VII + Form IV) Pawley refinement considerably improved the fit to the experimental powder pattern ($R_{wp} = 1.362$).

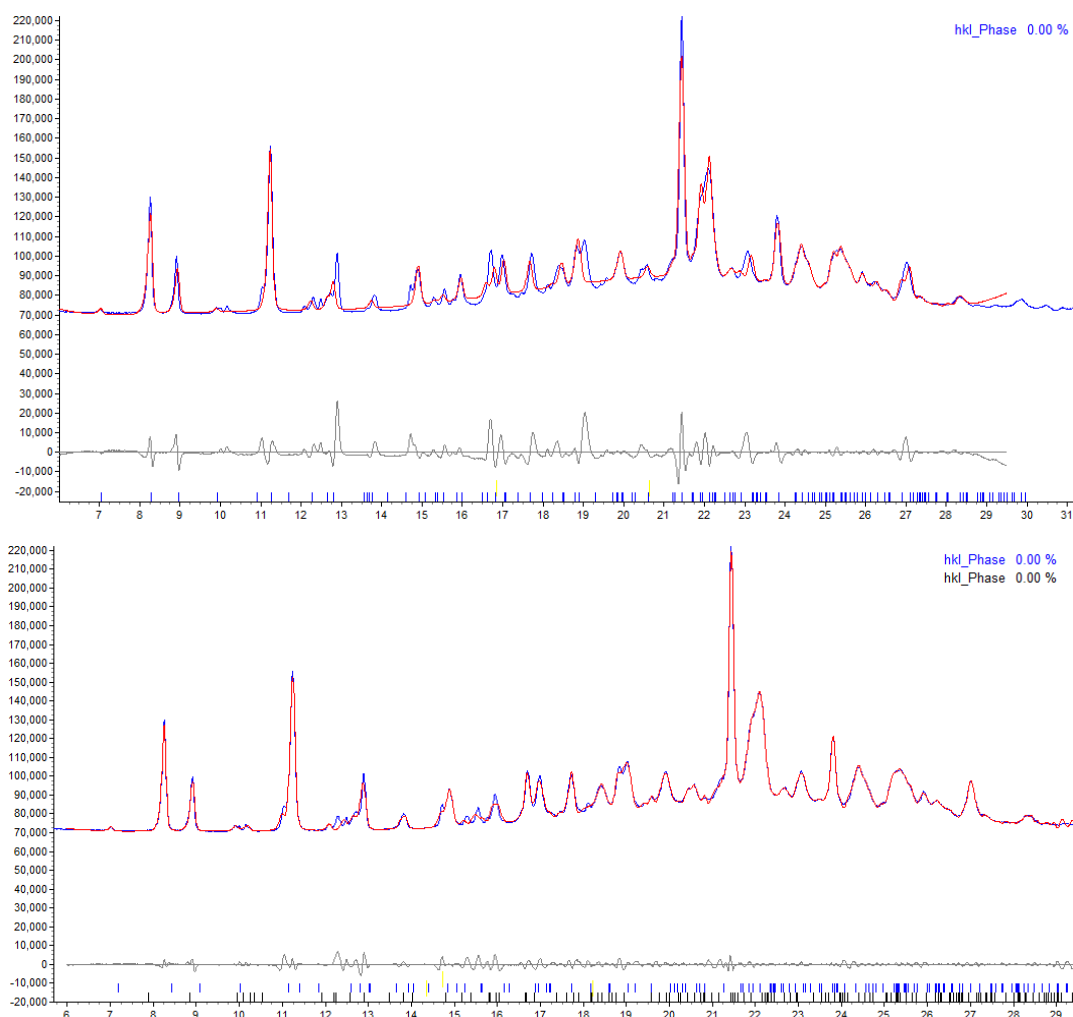


Figure S10. Pawley-type refinement of Form VII as a pure phase (top, $R_{wp} = 3.833$) and mixed phase Pawley-type refinement of Form VII with Form IV (bottom, $R_{wp} = 1.362$). Blue ticks denote peak positions of Form VII, black ticks denote peak positions of Form IV.

The CSP-derived structural model of Form VII (with lattice parameters modified by Pawley-type refinement against mixed phase data from Forms VII and IV) was taken as a starting structure for a multi-phase restrained Rietveld refinement in TOPAS V4.2. In the course of the refinement, the Form IV unit cell and peak shape parameters were allowed to vary, while all atomic coordinates were fixed. All atomic positions (including H atoms) for the Form VII structure were refined, subject to a

series of restraints on bond lengths, bond angles and planarity. The final restrained Rietveld refinement included a total of 273 parameters (25 profile, 6 cell, 1 scale, 1 isotropic temperature factor, 15 preferred orientation + 225 positions yielding a final Rwp= 4.47, Figure S11).

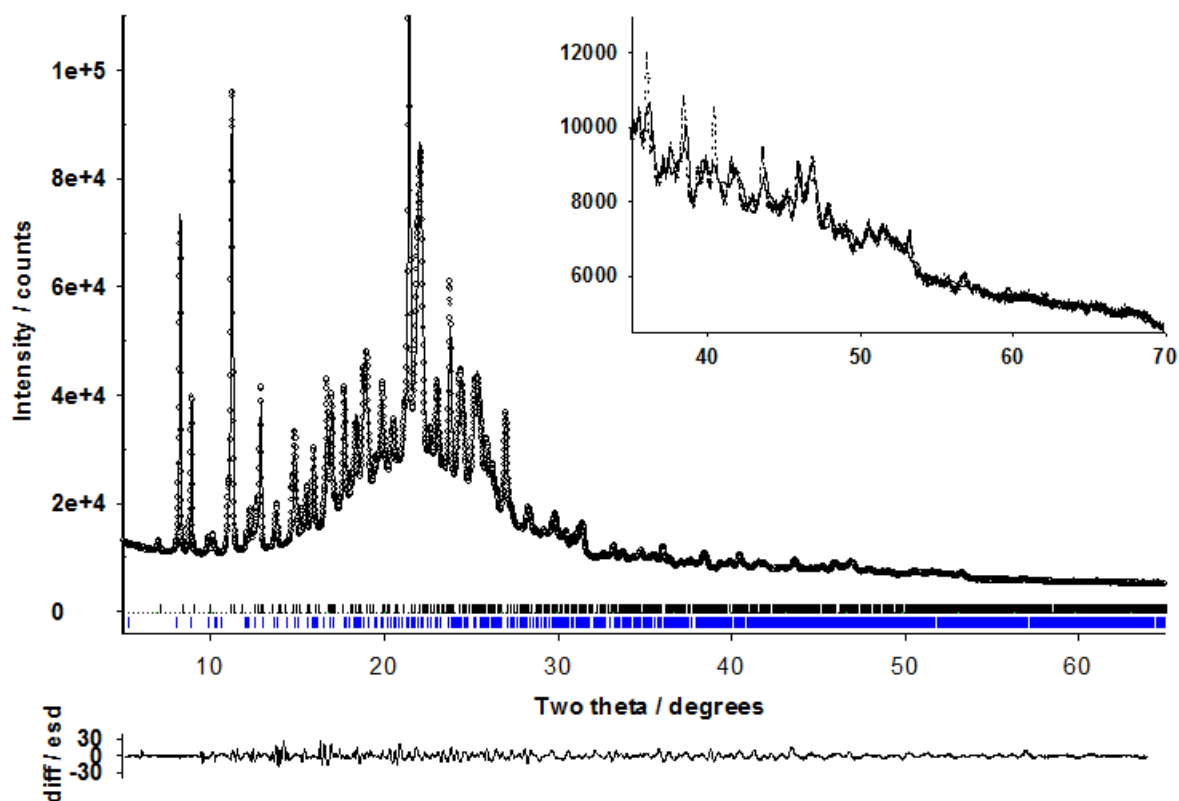


Figure S11. Final fit showing the observed yobs (lines), calculated ycalc (points) and difference $[(yobs-ycalc)/\sigma(yobs)]$ profiles for the mixed phase restrained Rietveld refinement of Form VII.

For Form VIII, multiple structural candidates were obtained by PXRD matching to simulated patterns of GRACE-predicted structures (Section 3.2.2). Pawley refinement of the unit cell parameters of the structural models against the best available PXRD data of Form VIII, along with space group symmetry and solid state NMR calculations, were used to find the best candidate, rank 39 (Figure S12).

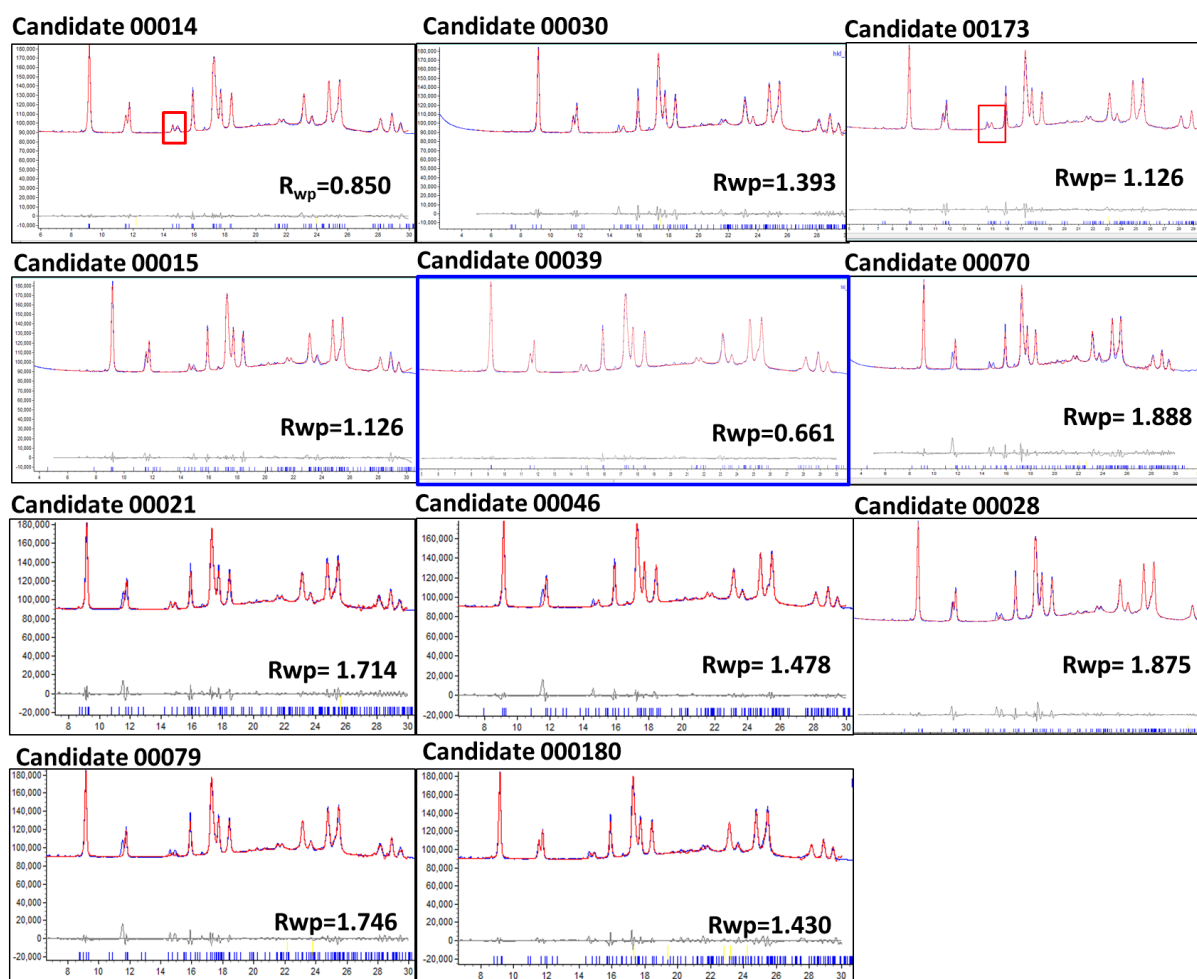


Figure S12. Pawley-type refinement of unit cell parameters of plausible predicted candidates for Form VIII against the best available experimental PXRD data of Form VIII. Candidates 14, 46 and 173 can be ruled out as they have peaks that do not match the experimental data. Candidate 39 best matches the experimental data.

The CSP-derived structural model of Form VIII (with lattice parameters modified by Pawley-type refinement against mixed phase data from Forms VIII and IV) was taken as a starting structure for a multi-phase restrained Rietveld refinement in TOPAS V4.2. In the course of the refinement, the Form IV unit cell and peak shape parameters were allowed to vary, while all atomic coordinates were fixed. All atomic positions (including H atoms) for the Form VIII structure were refined, subject to a series of restraints on bond lengths, bond angles and planarity. The final restrained Rietveld refinement included a total of 203 parameters (39 profile, 6 cell, 1 scale, 1 isotropic temperature factor, 15 preferred orientation + 141 positions) yielding a final Rwp= 3.074 (Figure S13).

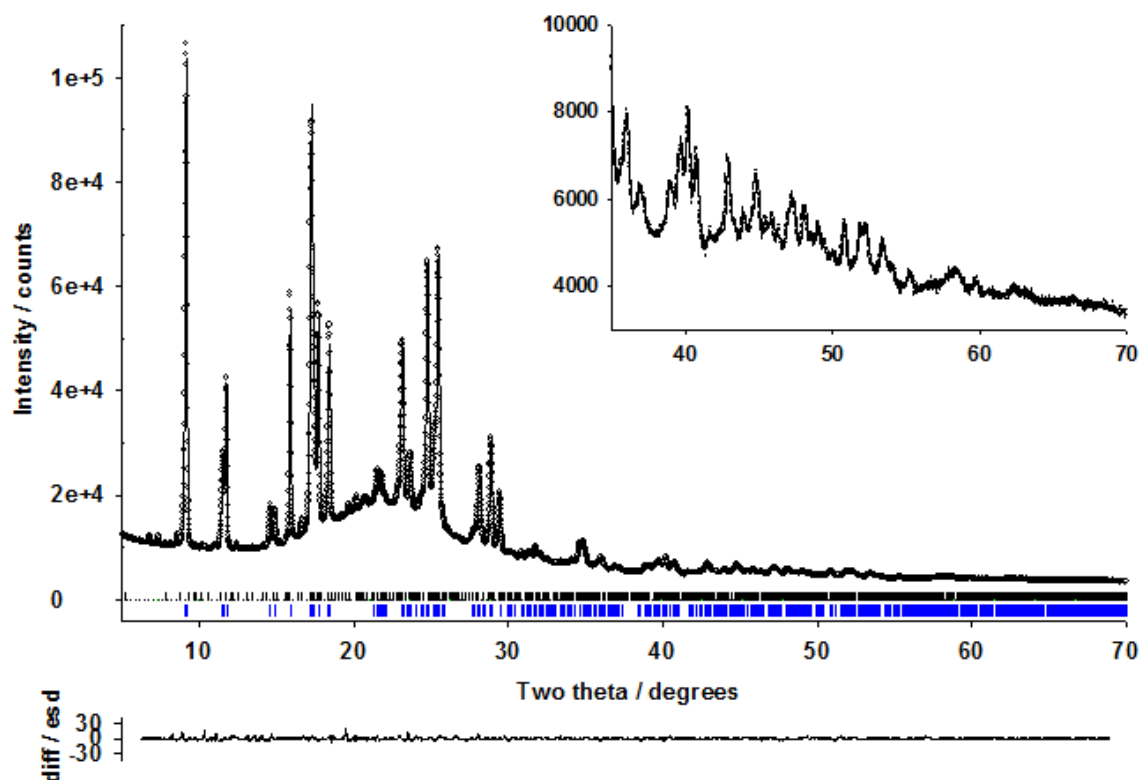


Figure S13. Final fit showing the observed yobs (lines), calculated ycalc (points) and difference [(yobs–ycalc)/sigma(yobs)] profiles for the restrained Rietveld refinement of Form VIII. Black and blue tick marks represent Form VIII and Form IV, respectively.

A summary of the powder structure solutions for **GAL** Forms I, V, VII, VIII and IX is compiled in Table S11.

Table S11. Powder Structure Solutions for **GAL** Crystal Forms.

	Form I	Form V	Form VII	Form VIII	Form IX
Empirical formula	C ₂₂ H ₁₉ N ₅ O	C ₂₂ H ₁₉ N ₅ O	C ₂₂ H ₁₉ N ₅ O	C ₂₂ H ₁₉ N ₅ O	C ₂₂ H ₁₉ N ₅ O
Formula weight (g/mol)	369.42	369.42	369.42	369.42	369.42
Temperature (K)	298	298	298	298	298
Radiation type	Cu K α ₁	Cu K α ₁	Cu K α ₁	Cu K α ₁	Cu K α ₁
Wavelength (Å)	1.54180	1.54180	1.54180	1.54180	1.54180
Crystal system	monoclinic	monoclinic	triclinic	triclinic	monoclinic
Space group	P2 ₁ /n	P2 ₁ /c	P-1	P-1	P2 ₁ /c
a (Å)	19.4305(17)	8.2599(3)	8.55610 (2)	8.5537 (2)	11.3011(3)
b (Å)	11.3245(11)	19.6072(14)	11.36763(2)	11.3718 (4)	20.8473 (5)
c (Å)	9.1511 (9)	12.6913(6)	11.56100(1)	11.5674 (5)	8.10637 (17)
α (°)	90	90	65.82061 (1)	65.784 (3)	90
β (°)	95.302(5)	116.545(2)	105.49234 (3)	105.495 (3)	98.879 (2)
γ (°)	90	90	77.30878 (4)	77.315 (2)	90
Volume (Å ³)	2005.0(3)	1838.72(17)	1786.75 (2)	916.89 (7)	1886.95(8)
Molecules per cell (Z)	4	4	4	2	4
Specimen shape/size (mm)	cylinder, 12 × 0.7	cylinder, 12 × 0.7	cylinder, 12 × 0.7	cylinder, 12 × 0.7	cylinder, 12 × 0.7
Absorption coeff (mm ⁻¹)	0.079	0.08	0.08	0.069	0.084
Data collection					
Diffractometer	PANalytical Empyrean				
Specimen mounting	0.7 mm borosilicate capillary				
Data collection mode	transmission				
Scan method	step				
2 θ range/step size (°)	20-70/0.014				
Refinement					
R factors	Rp = 0.098 Rwp = 0.094 Rexp = 0.011	Rp = 0.036 Rwp = 0.045 Rexp = 0.016	Rp = 0.04 Rwp = 0.045 Rexp = 0.01	Rp = 0.030 Rwp = 0.037 Rexp = 0.011	Rp = 0.039 Rwp = 0.044 Rexp = 0.011
Goodness-of-fit	8.603	2.783	4.30	3.16	3.86
Parameters	61	58	273	203	57
Preferred orientation	spherical harmonics-based correction applied with TOPAS during the Rietveld refinement				
Profile function	fundamental parameters with axial divergence correction				
Background function	Chebychev				

2.3 Solid-State NMR Spectroscopy

GAL monohydrate and Forms I-VIII were characterized by solid-state ^{13}C NMR (ssNMR) spectroscopy. Neither Form IX nor Form X were prepared in sufficient quantities to permit this type of analysis. Cross polarization (CP)/magic angle spinning (MAS) NMR spectra were obtained on a Bruker Avance III 400 wide-bore NMR spectrometer operating at ^1H and ^{13}C frequencies of 400.131 and 100.622 MHz, respectively, and using either a Bruker 4 mm triple resonance probe or a Bruker 4 mm double resonance probe. The MAS rate was set to 10 kHz using a Bruker MAS-II controller; spinning speeds were maintained within 2 Hz of the set point. SPINAL64 decoupling at a proton nutation frequency of 100 kHz was used for heteronuclear decoupling. Spinning sidebands were eliminated by a five-pulse total sideband suppression (TOSS) sequence. A RAMP100 shaped H-nucleus CP pulse was used on the ^1H channel to enhance CP efficiency, with a contact time for transferring magnetization from protons to carbons of 4.0 ms. The acquisition time was set to 34 ms and spectra were acquired over a spectral width of 30 kHz with a recycle delay of 60 s. The sample temperature was regulated to 297 ± 1 K in order to minimize frictional heating caused by sample spinning. The ^{13}C chemical shifts were externally referenced (± 0.05 ppm) to the proton-decoupled ^{13}C peak of neat (liquid) tetramethylsilane via the high-field resonance of adamantane ($\delta = 29.5$ ppm)

^{13}C CP/MAS NMR spectra of the monohydrate and anhydrous polymorphs of **GAL** are shown in Figure S14. The solid-state NMR spectra feature relatively sharp resonances commensurate with the short-range order present in these materials, and polymorphism (or hydrate formation) are inferred from the unique chemical shifts of the intense, isotropic ^{13}C resonances in the different ssNMR spectra. Despite the peak congestion in the aromatic region (110-175 ppm), a maximum of one ^{13}C peak was seen for every carbon atom in **GAL** monohydrate and Forms I, II, III, V, VI, and VIII, suggesting one molecule is present in the respective crystallographic asymmetric units. Two inequivalent molecules, most clearly seen near 110 ppm, were observed in Forms IV and VII.

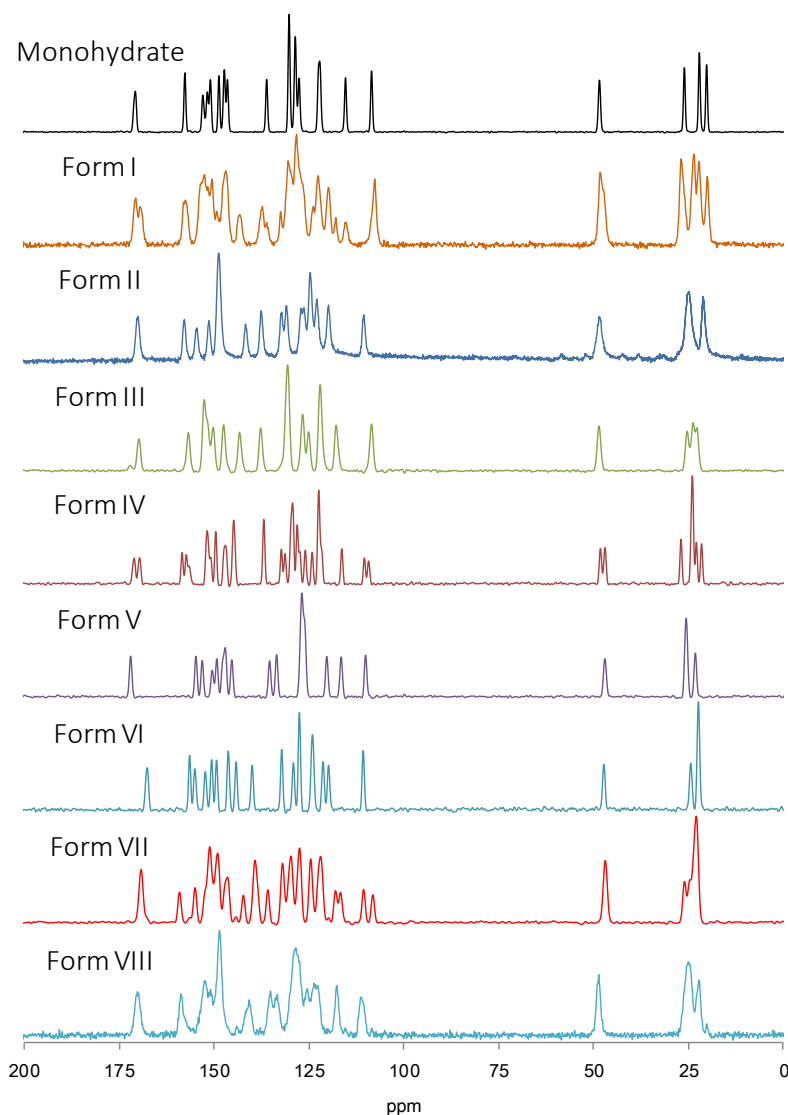


Figure S14. Experimental ^{13}C CP/MAS NMR spectra of **GAL** monohydrate and Forms I-VIII.

^{13}C Chemical shifts were calculated for PBE-D3BJ optimized crystal structures of the neat **GAL** polymorphs using the GIPAW (Gauge Including Projector Augmented Waves) plane-wave pseudopotential approach implemented in VASP (version 5.4.4). The structures were geometry optimized, relaxing all atoms, but fixing the unit cell parameters to their RT values (Table S10). For the shielding tensor calculations, the plane-wave cutoff energy was 700 eV, a \mathbf{k} -point grid with approximately $2\pi \times 0.032 \text{ \AA}^{-1}$ spacing was used and the convergence criterion on the total electronic energy was 10^{-7} eV. The computed shielding constants, σ_{calc} , were converted to chemical shifts, δ_{calc} , from the y-axis intercept and slope of the least-squares fit of the calculated shielding vs. experimental chemical shift plot ($\sigma_{\text{calc}} = -x \times \delta_{\text{exp}} + \sigma_{\text{ref}}$), Figure S15. Experimental ssNMR shifts (except C3, which was not well predicted across all forms) from Forms II-VI and the monohydrate (Figure S14) were used in a single determination of x and σ_{ref} ; the RMSD in the calculated chemical shifts was 1.2 ppm.

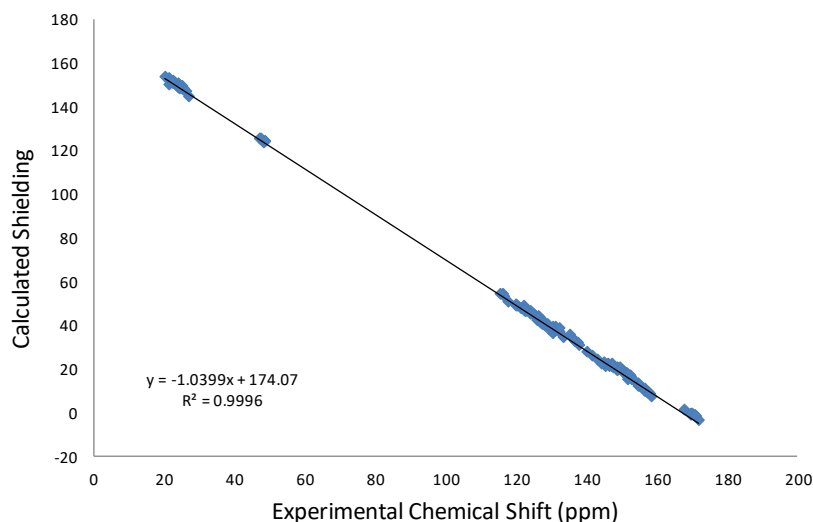


Figure S15. Correlation between calculated NMR shielding values and experimental ssNMR chemical shifts.

The excellent correlation between the calculated shieldings and the experimental chemical shifts in the monohydrate and Forms II-VI suggested that chemical shifts could be reliably estimated from the calculated shieldings. Shown in Figure S16 are calculated ^{13}C ssNMR spectra of **GAL** Forms I-X, including the best candidate structures for Forms VII and VIII, and the monohydrate plotted as a Lorentzian kernel density (HWHM = 0.4 ppm). The quality of the agreement between the experimental and calculated ssNMR spectra (Figure S14 v. Figure S16) suggested that ssNMR spectroscopy might be useful to identify the best CSP candidate structures among close matches to the known forms (Section 3.2.2) and also to verify PXRD structure solutions.

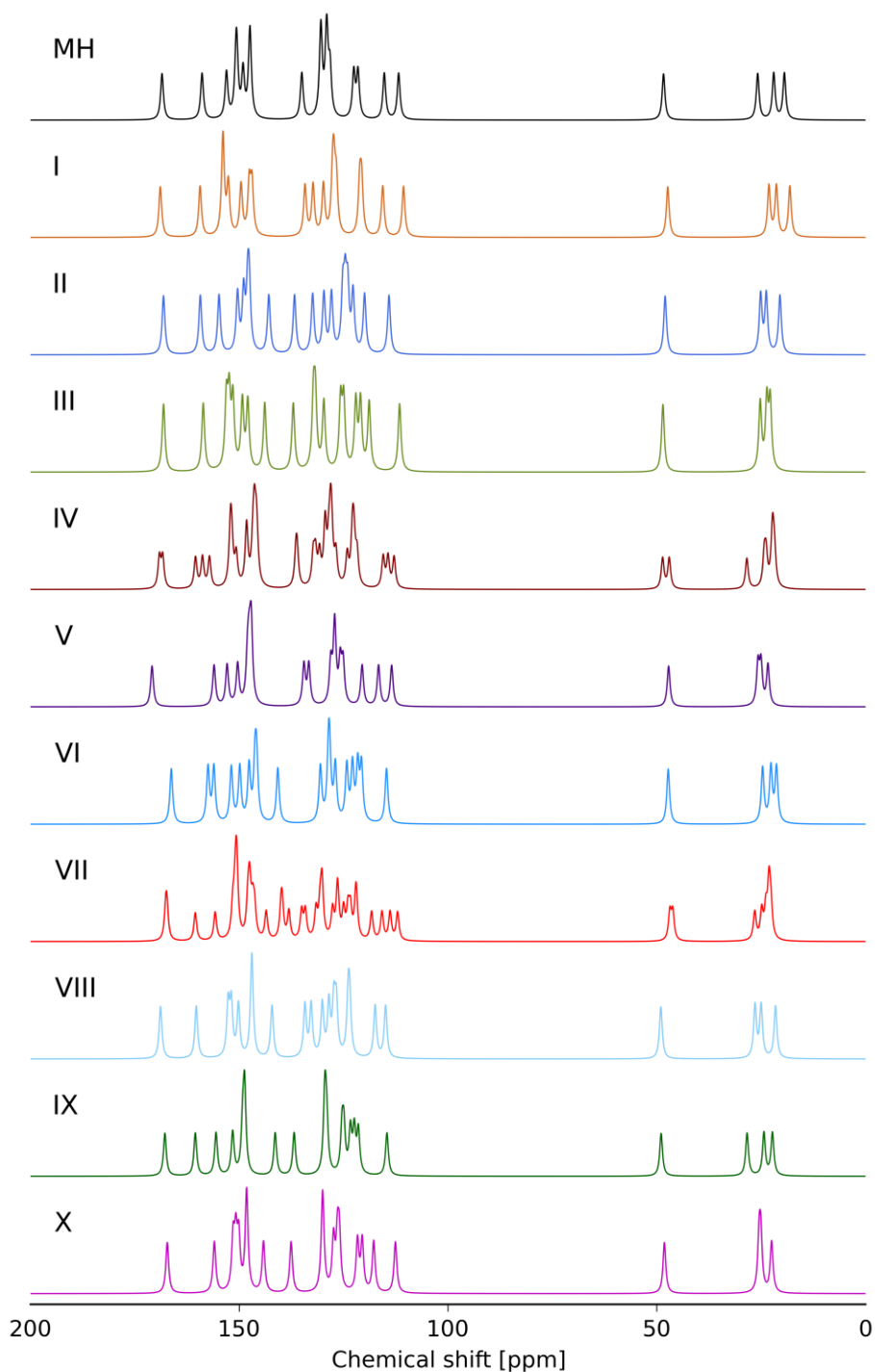


Figure S16. Calculated ^{13}C CPMAS NMR spectra of **GAL** Forms I-X and monohydrate.

2.4 Differential Scanning Calorimetry

Differential scanning calorimetry (DSC) data were collected for the neat crystal forms to determine their melting points and heats of fusion for an initial assessment of their thermodynamic stability relationships. DSC analyses was conducted using a Q1000 DSC from TA Instruments. Calibration was performed using sapphire discs, as recommended by the vendor. The temperature and the cell

constant of the cell were calibrated at 50 °C/min, using approximately 0.8 mg indium in a hermetically sealed pan with an inverted lid on the sample platform and an empty hermetically sealed pan with an inverted lid on the reference platform. The hermetically sealed pan and lid were previously cleaned with water and organic solvents and dried in an 80 °C oven for 30 minutes. Approximately 1 mg **GAL** was used for each DSC run. Each sample was scanned from -90 to 300 °C at 50 °C/min. A baseline scan, with the empty sealed pan/inverted lid on both platforms, was also collected for background subtraction.

Assessing the relative stability of the **GAL** polymorphs by DSC was complicated by the chemical instability at higher temperatures. To minimize decomposition at higher temperatures, a relatively fast (50 °C/min) heating rate was used. This allowed the melting onset temperatures of the neat polymorphs (Forms I-VII) to be clearly observed (by shifting solid-state transitions to higher temperatures), but lead to greater uncertainty in the measured enthalpies. Shown in Figure S17 are the DSC thermograms for the most representative samples of **GAL** Forms I-VII. It should be noted that Form VII, in its purest form, appears to be contaminated with Form IV (Section 2.2).

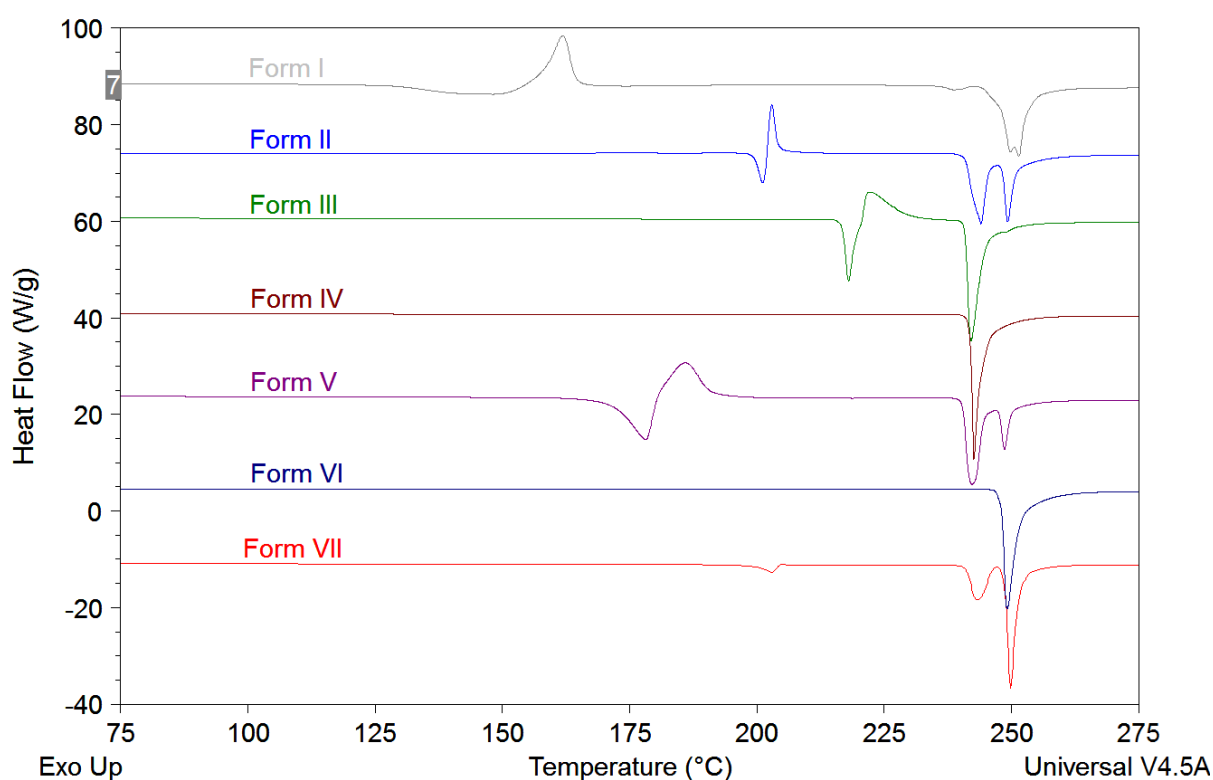


Figure S17. DSC thermograms of neat **GAL** Forms I-VII, measured at 50 °C/min.

With improved methods to generate the different polymorphs in highly crystalline, phase pure form and an optimized DSC method, the melting enthalpies could be accurately determined. Before heat integration, the heat signals from the background were subtracted from those of the sample. The subtracted heat signals were then normalized with respect to the sample weight before heat integration of the thermal event. When multiple thermal events were observed, the heat integration was performed separately on each (or the overlapped) event. If multiple integrations were needed,

the total heat of melting (Hess' law) was the sum of the melting heats from each individual thermal event. The results of the thermal analysis of **GAL** Forms II-VII are summarized in Table S12.

Table S12. Summary of **GAL** DSC results at 50 °C/min.

Crystal Form	Melting temperature; T_m (°C)			Integrated heat flow; W (J g ⁻¹)			ΔH_f (kJ mol ⁻¹)
	T_m (onset)	Mean	σ	W	Mean	σ	Mean
Form II	199.4	199.2	<0.1	107.9	108.2	0.8	40.0
	199.5			109.4			
	198.5			107.4			
Form III	217.0	217.3	0.2	107.3	107.4	0.2	39.7
	217.2			107.2			
	217.5			107.6			
Form IV	241.8	241.8	<0.1	115.5	117.0	1.2	43.2
	241.9			117.2			
	241.8			118.4			
Form V	171.5	171.5	<0.1	117.3	118.6	1.6	43.8
	171.6			117.6			
	171.5			120.9			
Form VI	247.6	247.7	0.1	117.2	117.3	0.1	43.3
	247.6			117.4			
	247.8			117.2			
Form VII	200.2	200.1	0.2	118.4	117.4	1.5	43.4
	200.3			115.3			
	199.9			118.4			

2.5 Gravimetric Vapor Sorption Analysis

The moisture sorption properties of **GAL** monohydrate and Forms II-VI were evaluated using gravimetric water vapor sorption analysis. For the neat polymorphs, a VTI flow moisture balance was used and analyses were performed at 25 °C over a 5-95% RH adsorption/desorption range in 5% RH step intervals. The minimum equilibrium criterion was 0.020% weight gain in 30 minutes for a maximum of 180 minutes. The monohydrate was analyzed using a vacuum moisture balance with initial drying at 60 °C *in vacuo* and the following equilibrium criteria: 0.0000% weight gain in 2 minutes for a maximum of 300 minutes and 0.0330% weight gain in 15 minutes for a maximum of 180 minutes.

The ambient temperature moisture sorption isotherms are compared in Figure S18. Forms IV-VI were nonhygroscopic, sorbing less than 0.2% water up to 95% RH, and physically stable over a wide RH range. No phase changes were detected by PXRD analysis of the sample residues. Forms II and III also showed minimal water uptake up to \approx 90% RH; however, unlike Forms IV-VI, a sharp weight increase of \approx 4.6% and 3.0% was observed, respectively, for these forms above 90% RH. The sorbed water was not lost during the desorption phase (95-5% RH). The irreversible weight gain at high RH

in the Form II/III adsorption profiles was confirmed by PXRD residue analysis to be [partial] conversion of Forms II and III to the monohydrate.

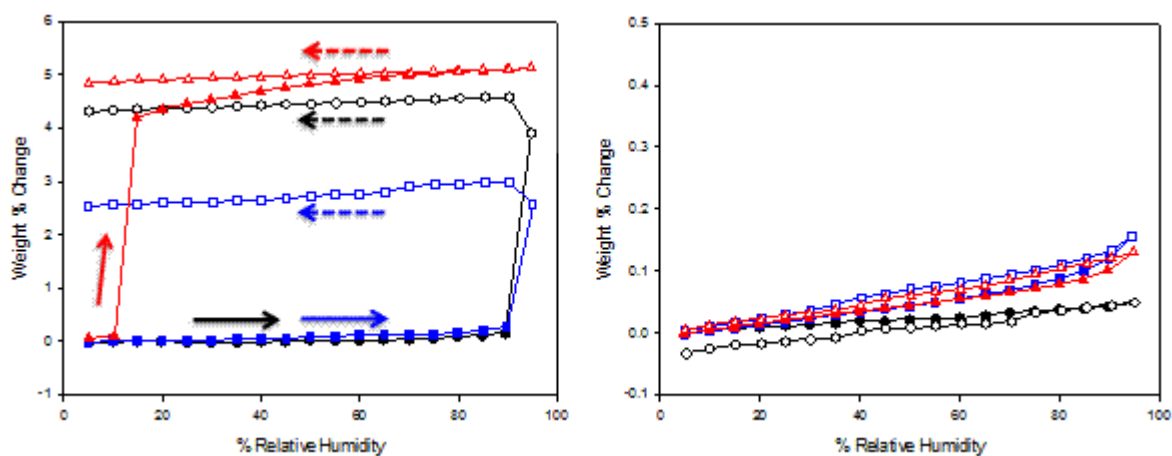


Figure S18. Gravimetric water vapor sorption isotherms of **GAL monohydrate**, Form II and **Form III** (left) and **GAL Form IV**, **Form V** and **Form VI** (right).

The GVS results confirmed the greater relative stability of the monohydrate than Forms II and III at high RH, however, without any observable phase changes for Forms IV and VI and irreversible form conversions for Forms II and III, these data were not sufficient to evaluate the thermodynamic stability relationships with respect to the monohydrate. Physical mixtures of **GAL** monohydrate and Form VI were therefore stored over a wide RH range for 2+ weeks in an attempt to establish their relative thermodynamic stability. Form VI is among (along with Form IV) the most stable neat polymorphs. Unfortunately, no evidence of form conversion was observed in the solid state over the 2+ weeks of storage at ambient temperature, Table S13. The slow transformation kinetics meant that solution-mediated techniques would be required to assess the thermodynamic stability of **GAL** monohydrate and the neat polymorphs.

Table S13. Summary of RH Annealing Experiments.

Starting Forms (PXRD)	Experimental Conditions			Final Form (PXRD)
	a_w	Temp	Time (d)	
MH, VI	0	RT	15	no change
MH, VI	0.22	RT	15	no change
MH, VI	0.58	RT	15	no change
MH, VI	0.76	RT	15	no change
MH, VI	0.98	RT	15	MH

2.6 Thermodynamic Stability from Slurrying Experiments

The relative thermodynamic phase stability of **GAL** Form VI and the monohydrate was assessed in aqueous suspensions over a range of water activities (a_w); the relative stability of Forms IV and VI was examined as a function of temperature. Two series of slurry experiments are summarized in Table S14. The first group of experiments examined the suspension stability of Form VI + monohydrate mixtures in n-butyl acetate and various acetone-water mixtures, while the second series explored the Form IV + VI stability in n-butyl acetate over a range of temperatures (with $a_w = 0$ to avoid conversion to the monohydrate). The results confirmed that at least above $a_w = 0.14$ ($\approx 14\%$ RH), the monohydrate is thermodynamically more stable than Form VI (and presumably all of the other neat polymorphs) at ambient temperature. Form IV was found to be more stable (less soluble) than Form VI, at least between 7 and 100 °C. With **GAL** having appreciable solubility (see Table S1 for reference) in each of the acetone-water compositions used in the slurry studies, the observation of the monohydrate, instead of Form IV, in the slurry residues suggests that Form IV is also less stable than the monohydrate over a wide RH range.

Table S14. Summary of Phase Stability (Slurry) Evaluations.

Starting Forms (RXRD)	Solvents (v/v)	Experimental Conditions			Final Form (PXRD)
		a_w	Temp (°C)	Time (d)	
MH + VI	nBuOAc	0	RT-100		IV
MH + VI	89:1 acetone-H ₂ O	0.14	RT	3	MH
MH + VI	27:1 acetone-H ₂ O	0.37	RT	3	MH
MH + VI	13:1 acetone-H ₂ O	0.55	RT	3	MH
MH + VI	6:1 acetone-H ₂ O	0.71	RT	3	MH
IV + VI	nBuOAc	-	7	6	IV
IV + VI	nBuOAc	-	RT	6	IV
IV + VI	nBuOAc	-	40	6	IV
IV + VI	nBuOAc	-	60	6	IV

The DSC, GVS and slurry experiments collectively revealed the thermodynamic stability relationships of interest for selecting the best solid form of **GAL** to develop and understanding the risks of form conversions during processing and storage. The crystalline monohydrate is the stable form at ambient temperature over a wide RH range, with desolvate Forms II and III being sufficiently metastable so as to convert in the solid state to the monohydrate at high RH and Form VI converting to the monohydrate in suspensions of at least $a_w = 0.14$. Form IV is the thermodynamically most stable neat polymorph up to at least 100 °C, but it is enantiotropically-related to the highest melting Form VI. The transition temperature below which Form IV is most stable and above which Form VI is most stable was not precisely determined, however, these studies confirm it to be above 100 °C. From the standpoint of thermodynamic stability, the monohydrate is clearly the preferred form for developing into a drug product. Drying studies (Section 1.3) did show, however, the risks of dehydration during typical drying operations.

3 Crystal Structure Prediction

The original CSP study of **GAL** was performed in 2014 using the Ψ_{mol} approach,^{21,22} which requires decisions about the range of conformations that could occur in the crystal structures (Section S3.1.1), and the conformational variables which can differ between the isolated molecule and the crystalline structures (Section S3.1.3.2). This search originally inspired the commissioning of the pressure work, and was used to help solve the $Z' = 1$ structures of Forms II, III, V and IX from PXRD (the structures of Forms II and III were later solved by SCXRD). The assumptions behind this CSP search, particularly the manual analysis of the possible conformational flexibility of the **GAL** molecule, were evaluated as the experimental forms were found and solved.

Late into the project, the commercial CSP program GRACE²³ became available. This uses an alternative approach, automatically generating a molecule-specific force field for the search and using periodic DFT-D (the Ψ_{crys} approach²¹) for the final lattice energy evaluations. This search was undertaken to cover $Z' = 2$, after it was established that this would be needed to solve the structure of Form VII. The algorithm implemented in GRACE for using experimental powder patterns to find potential matches amongst the CSP structures was first validated by the solved structures and then applied to Forms VII and VIII. The use of two of the leading approaches to CSP for pharmaceutical molecules allowed some analysis of the future use of CSP in polymorph screening.

3.1 CrystalPredictor/CrystalOptimizer (CSP1, Ψ_{mol})

The Ψ_{mol} approach splits the lattice energy into the intermolecular interactions within the crystal, U_{inter} , and the energy penalty for adopting a conformation that is less stable than the (*ab initio*) most stable for the molecule in isolation, ΔE_{intra} , with $E_{\text{latt}} = U_{\text{inter}} + \Delta E_{\text{intra}}$. The molecular wavefunction calculation provides both ΔE_{intra} and a distributed multipole representation of the molecular charge density, the latter which is used in evaluating the electrostatic contribution to the lattice energy U_{inter} . An initial conformational analysis defines the conformational variation that is explored in the generation of the crystal structures by CrystalPredictor version 1.6,²⁴ which uses grids of *ab initio* ΔE_{intra} estimates.

3.1.1 Conformational Analysis and Construction of Conformational Energy Grids

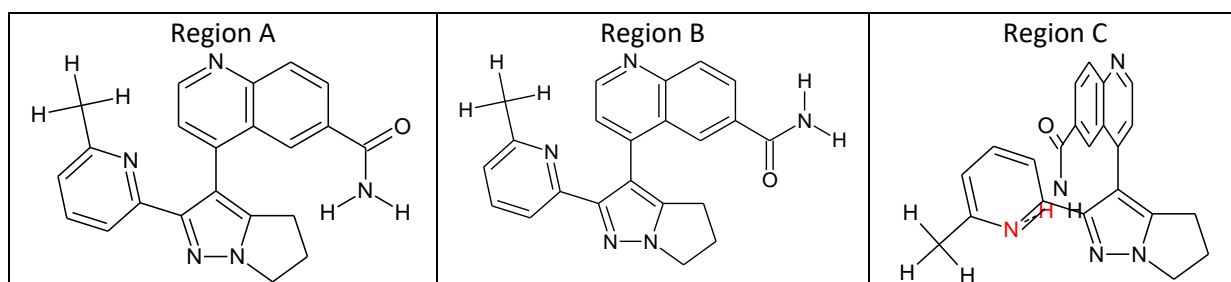


Figure S19. Schematic diagrams of the conformational regions. The atoms involved in the internal hydrogen bond in Region C are colored red.

Relaxed scans of the main torsion angles were carried out at the HF level of theory with the 6-31G(d,p) basis set prior to the search. The central C4_C3_C10_C11 angle was scanned with all four possible combinations of the pyridine N4_C16_C11_C10 and amide O1_C21_C6_C5 angles (Figure

S20) and each of pyridine N4_C16_C11_C10 and amide O1_C21_C6_C5 angles was scanned with the two possible values of the other (Figure S21). No degrees of freedom were constrained in any scans with the exception of the methyl angle (to speed up optimization) and the pyridine N4_C16_C11_C10 angle in one of the scans of central C4_C3_C10_C11 (see Figure S20 caption for details).

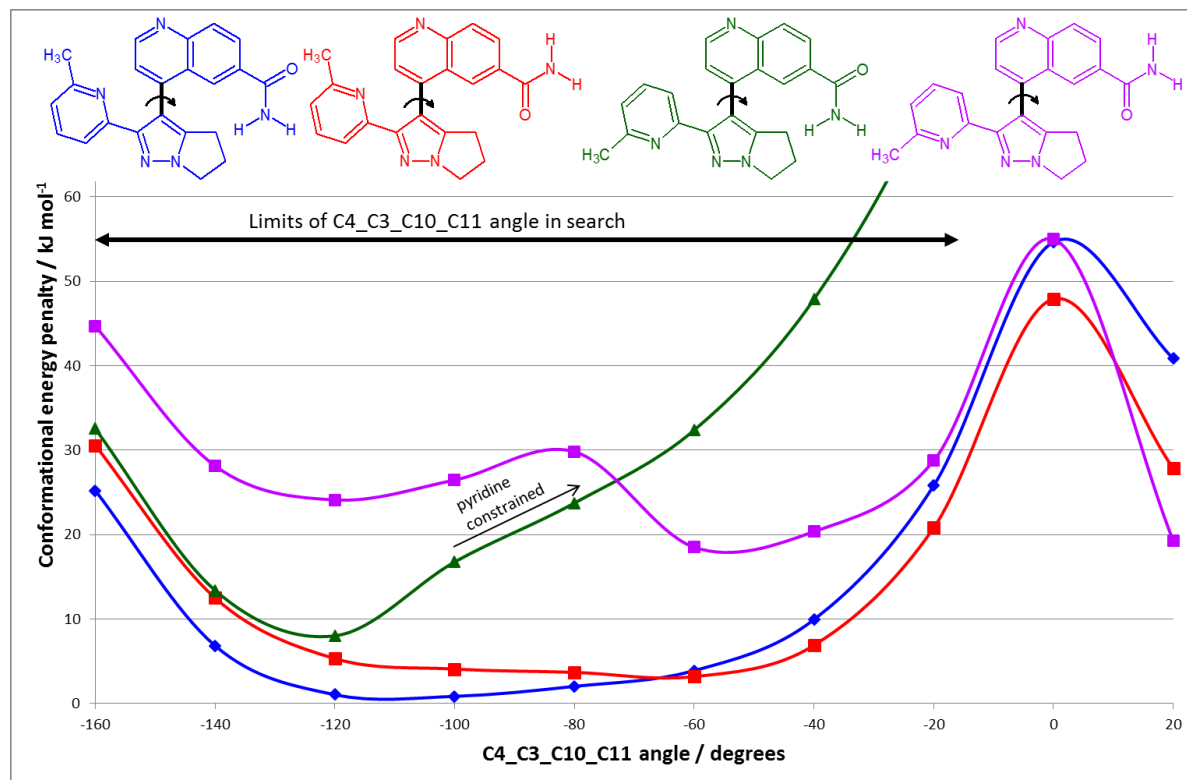


Figure S20. Torsion angle scans about C4_C3_C10_C11 at the HF/6-31G(d,p) level of theory. Each scan had different configurations of the amide and pyridine groups, as shown by the embedded colored molecular diagrams (shown with C4_C3_C10_C11 = 0°). With the exception of the scan in green, all scans were unconstrained. The configuration of the pyridine group changed at the maximum at C4_C3_C10_C11 = 0° in the red, blue and purple scans. For the scan in green, the angle between the pyridine and pyrrole groups was constrained after -100°, as indicated by the arrow on the chart. The black arrow shows the range of the angle used in the CrystalPredictor search.

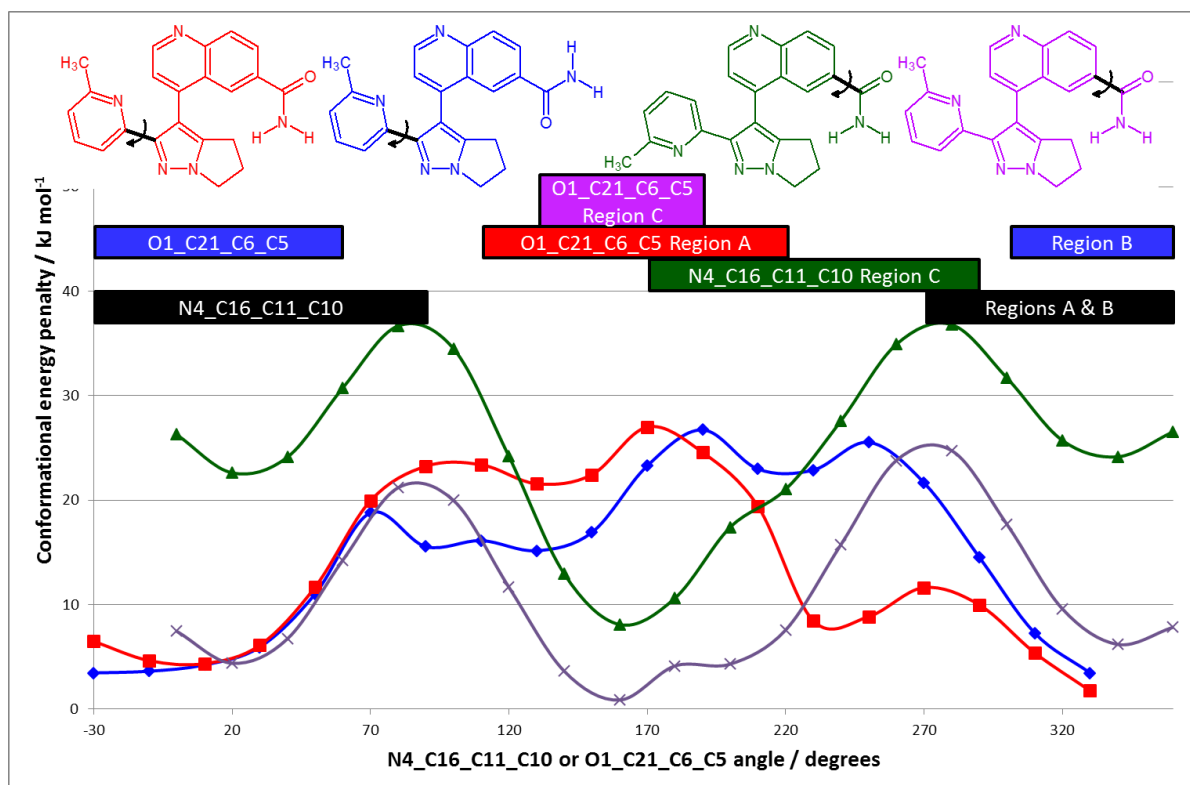


Figure S21. Torsion angle scans about N4_C16_C11_C10 and O1_C21_C6_C5 at the HF level of theory with the 6-31G(d,p) basis set. Each scan had 2 possible configurations of the other angle, as shown by the embedded colored molecular diagrams. The diagrams show the scanned angle in black. Only the methyl rotation was constrained in the scan, and all other degrees of freedom were allowed to relax freely. The colored blocks show the ranges of the angles used in the search, black for the N4_C16_C11_C10 angle in Regions A and B, green for the N4_C16_C11_C10 angle in Region C, red for the O1_C21_C6_C5 angle in Region A, blue for the O1_C21_C6_C5 angle in Region B and purple for the O1_C21_C6_C5 angle in Region C.

Following the search, it was clear that the 5-membered ring also had some conformational flexibility. This was subsequently scanned at the PBE0 level of theory and 6-31G(d,p) basis set. The starting point for this scan was the conformation within the search structure that corresponded to Form II. All bond lengths, bond angles and torsion angles were constrained at the starting conformation, apart from the variables either side of C13 and describing all six hydrogen atoms in this vicinity. The PBE0 method was used (as in the CrystalOptimizer stage from which the starting conformation was extracted), with the 6-31G(d,p) basis set. The result of this scan (Figure S22) shows that there is a wide variation with very little energy difference or barrier to flipping in the isolated molecule. However, for the molecule within various crystal structures, the overall energy of the crystal structure changes markedly with variation in this angle.

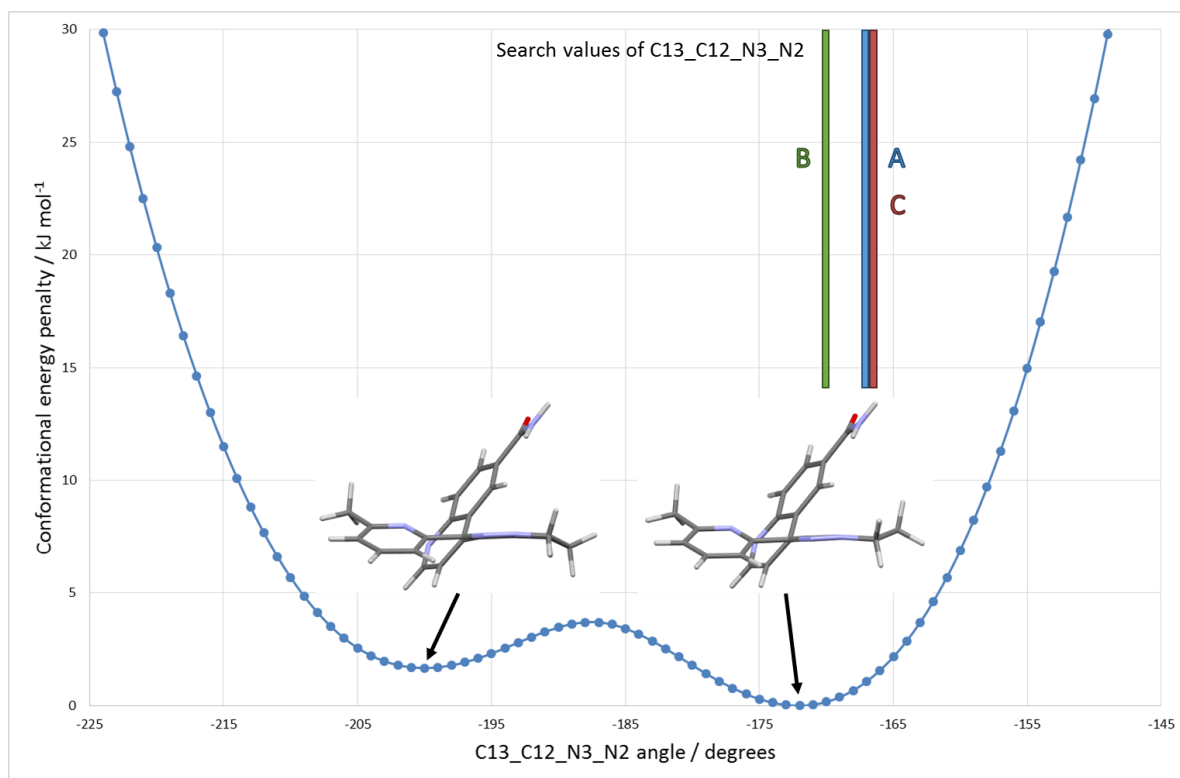


Figure S22. Torsion angle scan about C13_C12_N3_C2 at the PBE0 level of theory with the 6-31G(d,p) basis set. The conformations of the two minima are shown, and the difference in the orientation of the 5-membered ring can be seen at the right-hand end of the molecules. The three vertical lines show the value of this angle in the three conformational regions in CSP1.

3.1.2 Other CSP search parameters

The starting molecular conformation was that of the fully optimized local minimum at the PBE0/6-31G(d,p) level of theory, as calculated using GAUSSIAN.²⁵ Static point charges were used in the search, calculated for this local minimum energy conformation, with GAUSSIAN at the PBE0/6-31G(d,p) level of theory. The FIT potential²⁶ was used for the repulsion/dispersion contributions throughout. The search covered the 59 space groups: *P1*, *P-1*, *P2₁*, *P2₁/c*, *P2₁2₁2*, *P2₁2₁2₁*, *Pna2₁*, *Pca2₁*, *Pbca*, *Pbcn*, *C2/c*, *Cc*, *C2*, *Pc*, *Cm*, *P2₁/m*, *C2/m*, *P2/c*, *C222₁*, *Pmn2₁*, *Cmc2₁*, *Aba2*, *Fdd2*, *Iba2*, *Pnna*, *Pccn*, *Pbcm*, *Pnnm*, *Pmmn*, *Pnma*, *Cmcm*, *Cmca*, *Fddd*, *Ibam*, *P4₁*, *P4₃*, *I-4*, *P4/n*, *P4₂/n*, *I4/m*, *I4₁/a*, *P4₁2₁2*, *P4₃2₁2*, *P-42₁c*, *I-42d*, *P3₁*, *P3₂*, *R3*, *P-3*, *R-3*, *P3₁21*, *P3₂21*, *R3c*, *R-3c*, *P6₁*, *P6₃*, *P6₃/m*, *P2₁3*, *Pa-3*, with one molecule in the asymmetric unit cell ($Z' = 1$).

In regions A and B, close to 1,000,000 crystal structures were generated by CrystalPredictor1.6,²⁴ and in region C, 100,000 crystal structures were generated. This was sufficient to establish that structures with an internal hydrogen bond were not competitive in lattice energy. Following removal of duplicate structures, there were 7823 unique structures in Region A, 7826 unique structures in Region B, and 842 unique structures within Region C within 35 kJ mol⁻¹ of the lowest energy structure. All structures within 35 kJ mol⁻¹ of the global minimum were passed to the next stage. Structure names were given at this point, as Xyy, where X is the search region and yy the energy rank.

At this stage of the study, the only checks on the completeness of the search that were possible, were to establish that it had found the newly characterized Form VI, and the computationally desolvated solvates.

3.1.3 Refinement of crystal structures

3.1.3.1 Preliminary refinement – DMAflex-Quick

Each of the unique crystal structures within 35 kJ mol⁻¹ of the global minimum structure, was optimized with DMAflex-Quick. This uses the same intramolecular energy grid that was calculated for CrystalPredictor, but allows the use of multipoles. The multipoles for the fully optimized molecule at the local minimum of each region were calculated using GDMA2.2 (ref 27) from the PBE0/6-31G(d,p) wavefunction. For each crystal structure these multipoles were rotated and NEIGHCRY2.0.8RC1 and DMACRY2.0.8RC1 (ref 26) were used to minimize the energy of the structure.

Any duplicates were removed and all structures within 30 kJ mol⁻¹ of the lowest energy structure in their respective regions were carried forward to the next step, giving approximately two thousand structures in regions A and B and two hundred in region C. It was noted that the introduction of a distributed multipole model, capable of representing the electrostatic effects of π -electrons and lone pairs, instead of the charge-only model in CrystalPredictor, made a significant difference to the relative lattice energies.

3.1.3.2 Flexible refinement of crystal structures – CrystalOptimizer²⁸

CrystalOptimizer was used to optimize the molecular conformation and the crystal structure simultaneously, to model the effect of the crystalline environment on the conformation. The molecular conformation, distributed multipoles, and ΔE_{intra} were calculated at the PBE0/6-31G(d,p) level of theory. Databases of the *ab initio* parameters were built up for each conformational region.

Initially, chemical intuition suggested that only 10 torsion angles within **GAL** were likely to change from the isolated molecule values in response to the packing forces (Figure S23). This selection appeared satisfactory when tested on the very limited set of available crystal structures, which were four solvate structures (monohydrate, methanolate, 2-propanolate and S-2-butanolate) allowing for uncertainties in modelling the solvent interactions.

However, when the structure of Form VI became available, it became apparent that more flexible degrees of freedom needed to be included in the CrystalOptimizer refinement. CrystalOptimizer refinement of a small number of search structures with respect to all possible torsion and bond angles led to the selection of the 37 degrees of freedom (Figure S23) that changed the most (greater than 5° or 1° for torsion and bond angles, respectively) for refinement of the rest of the search structures. Comparison of the CrystalOptimizer(10), CrystalOptimizer(37) and CrystalOptimizer(all) refinements for Form VI (Figure S24) showed that this selection of degrees of freedom is appropriate. This emphasizes how small changes in the bond and torsion angles can have a significant effect on the overall shape of larger flexible pharmaceuticals, affecting their intermolecular interactions and hence lattice energies.

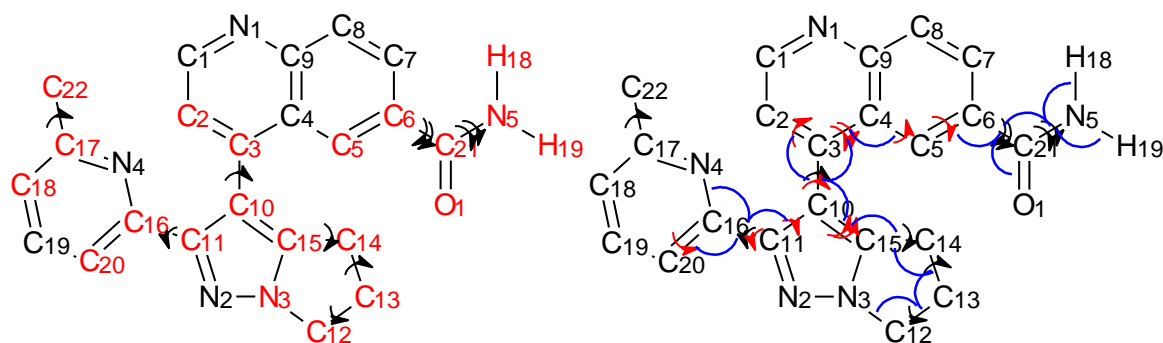


Figure S23. (Left) the degrees of freedom optimized within the initial CrystalOptimizer(10) model (red atom labels show exact definitions) and (right) the degrees of freedom optimized within the CrystalOptimizer(37) model with the additional torsion angles shown in red and the additional bond angles in blue.

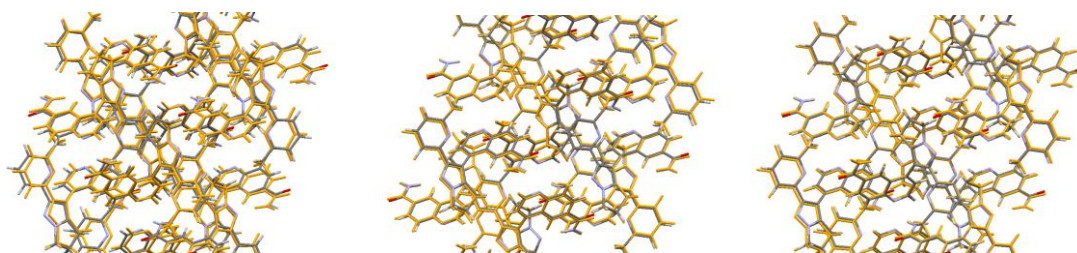


Figure S24. Left, overlay of CrystalOptimizer(10) minimized Form VI (orange) with experimental Form VI (colored by element) ($\text{RMSD}_{15} = 0.597 \text{ \AA}$), middle, overlay of CrystalOptimizer(37) minimized Form VI (orange) with experimental Form VI (colored by element) ($\text{RMSD}_{15} = 0.163 \text{ \AA}$), and right, overlay of CrystalOptimizer(all) Form VI (orange) with experimental Form VI (colored by element) ($\text{RMSD}_{15} = 0.158 \text{ \AA}$).

Search structure A85 corresponded to Form VI, but even after the CrystalOptimizer minimization with all possible degrees of freedom, there was a difference in the conformation of the 5 membered ring between A85 and Form VI minimized with the CrystalOptimizer model, corresponding to a lattice energy difference of 1 kJ mol^{-1} . The search input had only one conformation of this ring and CrystalOptimizer did not always change the configuration of this ring (Figure S22). Later work on Form IX showed the same change in the 5-membered ring resulted in a lattice energy difference of about 9 kJ mol^{-1} , showing the sensitivity of the energy to the local packing environment of this ring. This problem of the barriers to conformational change being environment dependent is a physical one manifest in the experimental occurrence of dynamic disorder for ring flips or methyl rotations. Hence periodic DFT-D optimizations starting from the CSP1 structure did not always correct the ring conformation.

3.1.4 Inclusion of Experimental Crystal Structures, Desolvated Solvates and Evaluation of the Monohydrate Energy

The CSP results were compared with the emerging results from the experimental screen (Figure S25). Structures were identified as starting models for solution of Forms II, III and V from the indexed powder patterns. Each of the experimental crystal structures that had been solved were minimized with CrystalOptimizer, using the same database of conformation dependent intramolecular energies and multipoles. Computational desolvation calculations were also carried out by doing corresponding optimization of the solvated solvate structures, with solvent molecules edited out. This procedure does not mimic experimental desolvation, where the solvent molecule disrupts the lattice

as it moves out, and the rearrangement of the **GAL** molecules will not necessarily proceed with the crystallographic symmetry, which is enforced in the calculations. However, such calculations do provide some insight into the arrangement of the molecules that could be conserved on desolvation, as well as providing other structures which should have been generated in the search.

The stability of the monohydrate was crudely estimated on the assumption that the lattice energy of the monohydrate must be less than the energy of the separate neat **GAL** and ice. Ice has many polymorphs (we have used ordered models of ice Forms II, VIII, IX, XI, XIII, XIV and XV in this investigation). The stability of the monohydrate relative to the neat forms is represented on the crystal energy landscapes as 7 entries, where the energies of each of the individual ice polymorphs has been subtracted from the overall energy of the monohydrate.

Table S15. Key and notes on the experimental structures shown on the summaries of the CSP results.

Polymorphs*		Density (g cm ⁻³)	Energy (kJ mol ⁻¹)	RMSD₂₀[#] (Å)
Form IV	Z' = 2 so could not be found in search	1.27	-180.44	0.208
Form VI	Found as A85 (-181.14)	1.28	-181.01	0.160 <i>0.577</i>
Form I	A995 (high E after CrystalOptimizer)	1.24	-154.57	0.245
Form VII	Z' = 2 so could not be found in search	1.32	-184.40	0.152
Form VIII	Found as A474 (-181.76)	1.32	-181.73	0.232 <i>0.224</i>
Form IX	Found as A44 (-169.63)	1.28	-176.32	0.320 <i>0.480</i>
Form X	Found as A385 (-180.61)	1.29	-180.64	0.315 <i>0.326</i>
Proposed PXRD matches				
Form II	A1207 (-167.27)	1.18	-172.49	0.237 <i>0.268</i>
Form III	A829 (-173.82)	1.27	-173.59	0.300 <i>0.247</i>
Form V	A58 (-172.05)	1.31	-171.88	0.204 <i>0.210</i>
Computationally desolvated				
THF solvate	A1207 (-167.27)	1.1265	-165.11	0.495 <i>1.406</i>
methyl ethyl ketone solvate form III	A8 (-173.92)	1.26	-177.12	5 molecules <i>5 molecules</i>
didioxane solvate	Not found	0.88	-131.13	2.311
DCM solvate form II	A928 (high E after CrystalOptimizer)	1.13	-156.07	0.587
morpholine solvate	A961 (-168.16)	1.22	-170.72	5 molecules <i>5 molecules</i>
2-methyl THF solvate	A961 (-168.16)	1.21	-170.86	5 molecules <i>5 molecules</i>
acetic acid solvate	A2613 (high E after CrystalOptimizer)	1.18	-163.03	0.369
hemitoluene disordered	A2613 (high E after	1.18	-163.11	1.04

solvate	CrystalOptimizer)			
dioxane solvate form IV	A961 (-168.16)	1.21	-170.98	5 molecules <i>5 molecules</i>
toluene solvate	Not found	1.14	-156.96	5 molecules
methanol solvate	A29 (-173.36)	1.27	-173.09	0.327 <i>0.402</i>
monohydrate	A995 (high E after CrystalOptimizer)	1.24	-154.57	0.245
hemi ethyl acetate solvate	Z' = 2 so could not be found in search	1.17	-172.43	0.368
ethanol solvate form III	B754 (high E after CrystalOptimizer)	1.16	-146.23	1.160
propan-1-ol solvate	B754 (high E after CrystalOptimizer)	1.16	-146.28	1.15
propan-2-ol solvate	B754 (high E after CrystalOptimizer)	1.13	-147.90	9 molecules
butan-1-ol	B754 (high E after CrystalOptimizer)	1.13	-147.92	11 molecules
butan-2-ol	B754 (high E after CrystalOptimizer)	1.16	-146.26	1.33
i-amyl alcohol solvate	B754 (high E after CrystalOptimizer)	1.16	-146.25	1.42
methyl ethyl ketone solvate form I	B847 (high E after CrystalOptimizer)	1.19	-164.37	0.331
NMP solvate	B2023 (high E after CrystalOptimizer)	1.00	-135.12	0.681
hemidioxane hydrate	B6389 (high E after CrystalOptimizer)	1.14	-141.27	0.718
ethanol solvate form I	A6486 (high E after CrystalOptimizer)	1.14	-146.01	0.935
Relative monohydrate stability				
Energy(monohydrate)- Energy(ice)	Energy span of 7 ice structures is 2.9 kJ mol ⁻¹	1.30	-185.69 to -182.75	

* Structures marked as matching are those where a CrystalOptimizer minimized experimental (or desolvate) structure matches search CrystalOptimizer minimum. Where there is no match, the experimental (or desolvate) structure's CrystalOptimizer minimum is compared with search CrystalPredictor minima.

RMSD₂₀ is calculated between the computationally minimized experimental structure and the experimental structure. Where available, the number below (*in italics*) is calculated between the search structure and the experimental structure. Where a number of molecules is given, this is the largest number that could be matched with 80% distance and 80° angle tolerances.

Figure S25 shows that the search was successful in locating the Z' = 1 structures that were known experimentally or derived from the solvates, but shows that there were many unobserved structures that were more stable and potentially more stable than the monohydrate. In some cases, a reasonably good structural match was found, but there was a significant difference in energy (Table S15) from small conformational differences or flat potential energy surfaces with multiple closely related minima.

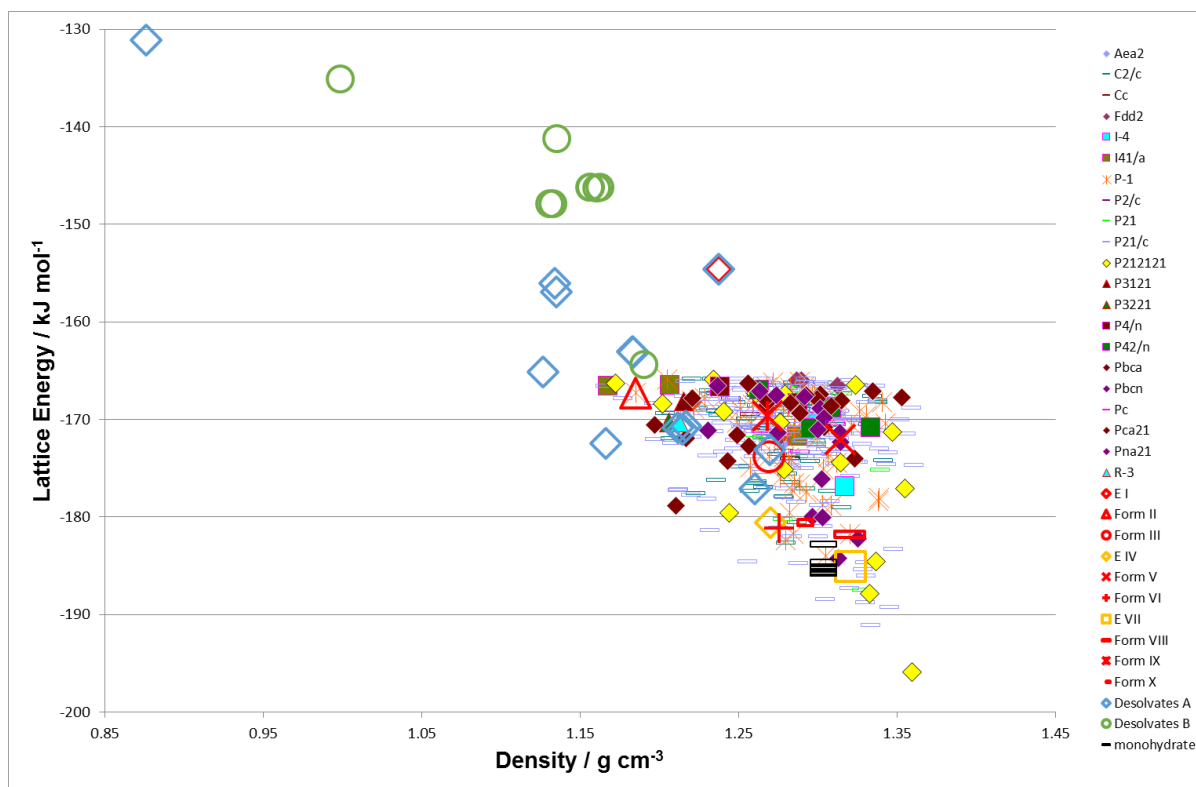


Figure S25. The crystal energy landscape of **GAL**, and the computationally desolvated structures, following CrystalOptimizer refinement. The open red (or orange for $Z' = 2$) symbols either denote structures optimized starting from the experimentally observed forms (EI, EIV and EVII) or CSP1 generated structures that corresponded to experimental forms. Open green (Region B) or blue (Region A) symbols denote desolvated solvate structures (see Table S15). Open black lines denote the monohydrate structure with the ice energies subtracted.

3.1.5 An Analysis of CSP Generated Low Energy Structures

The CSP study showed many alternative structures to those that had been experimentally identified. The structures that are most likely to appear as metastable polymorphs are those that have a significant barrier to transformation to the more stable forms, either from the need to change conformation or hydrogen bonding. Hence the diversity of the low energy structures was analysed as shown in Figure S26, showing that there were a large range of different hydrogen bonding combinations of donors of acceptors in each conformational region.

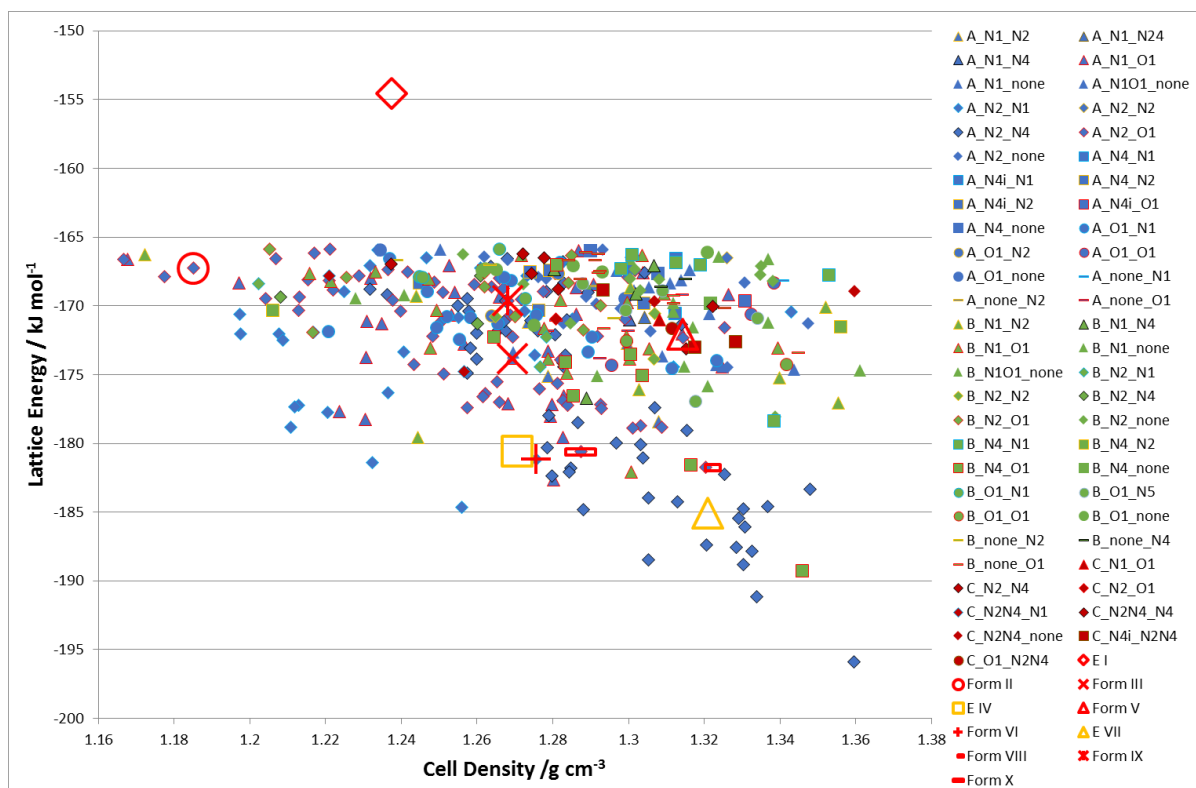


Figure S26. The hydrogen bonding and motif classification of the structures generated in the Ψ_{mol} search CSP1 (cf. Figure S25). The main color of each symbol denotes the conformational region; the shape of each symbol denotes the hydrogen bond acceptor of the H18 atom; the outline color of each shape denotes the hydrogen bond acceptor of the H19 atom. Hydrogen bonds were assessed in Mercury, with anything up to van der Waals radius plus 0.3 Å being considered a hydrogen bond.

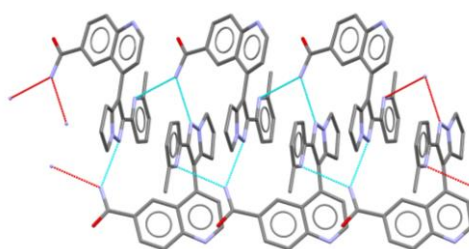


Figure S27. The hydrogen bonding motif in the lowest energy structure (GM) on Figure S26. Compare with Figure S33 for observed neat forms.

At the mid-stage of the project, the CSP1 study was consistent with the experimental observations. Forms IV and VI were close in energy, in agreement with the observed concomitant crystallization. The CSP had generated structures that had similar powder patterns to Forms II, III and V, which were metastable with respect to the two forms (IV and VI) that could be solved from single crystal X-ray diffraction, consistent with them being harder to crystallize. Hence, the CSP prediction that there

were many alternative possible polymorph structures, some of which could be more stable than the observed forms and possibly competitive with the monohydrate, had to be taken seriously.

The most important structure to target was the global minimum structure (GM, A997), which was also the densest and most stable packing of a hydrogen bonded chain seen in many low energy hypothetical structures. It is $P2_12_12_1$ *i.e.* it occurs in a chiral space group, which is unusual, but it is not a polar space group and so there is no net dipole moment in the unit cell. Hence, the most promising experiment for finding this form that could be suggested from this structure was to do crystallizations under pressure. It was computationally confirmed that A997 remained the most stable structure under pressure.

3.2 GRACE (CSP2, Ψ_{crys})

3.2.1 Structure Generation and Optimization

GRACE (version 2.4.235) was used to predict crystal structures of **GAL**. First a custom-made force field was parametrized, then crystal structures were generated by parallel tempering using that force field. The structures were then re-optimized with periodic DFT-D. All calculations were performed on Lilly's Brainiac cluster. The most accurate default options were set in GRACE in order to fit the most reliable force field within the assumed functional form. For fitting repulsion-dispersion terms, the **GAL** molecule was automatically split into a set of fragments, each fragment being a chemically sound molecule. GRACE then performed DFT-D calculations on individual fragments and crystal structures of those fragments. These DFT-D calculations were performed either in Turbomole (for individual fragments) or VASP 5.4.1 (for crystal structures), using the PBE functional and the Neumann-Perrin dispersion correction. For periodic calculations, a plane wave basis was used with a kinetic energy cutoff of 520 eV and a gamma-centered \mathbf{k} -point grid with approximately $2\pi \times 0.07 \text{ \AA}^{-1}$ spacing. For fragment calculations, the local aug-cc-pVQZ basis set was used. The force field parameters were then fitted to the results of these calculations. Both bonded and non-bonded interaction terms were fitted simultaneously in order to obtain a self-consistent force field. A 9-6 Lennard-Jones type potential was used to model repulsion-dispersion interactions. Atomic point charges were fitted to the electrostatic potential on a grid of points outside of molecules.

The custom-made force field was used to generate crystal structures by parallel tempering. Some of the generated structures were automatically submitted to periodic DFT-D calculations, and the results from those calculations were used by GRACE to further improve the bespoke force field in a refitting process. Those space groups that are the most common for crystal structures that do not have molecules situated on special positions were sampled. For $Z' = 1$, 38 space groups were included, and for $Z' = 2$, 19 space groups were included. In total, this corresponds to 99.9% of the CSD statistics, an unusually wide search.

For $Z' = 1$, the 38 most common space groups were sampled: $P1$, $P-1$, $P2_1$, $C2$, Pc , Cc , $P2/c$, $P2_1/c$, $C2/c$, $P2_12_12$, $P2_12_12_1$, $C222_1$, $Pca2_1$, $Pna2_1$, $Aba2$, $Fdd2$, $Iba2$, $Pcca$, $Pccn$, $Pbcn$, $Pbca$, $Fddd$, $P4_1$, $I4$, $I4_1$, $I-4$, $P4_2/n$, $I4_1/a$, $P4_12_12$, $I4_1cd$, $P-42_1c$, $P3_1$, $R3$, $R-3$, $P3_12_1$, $R3c$, $P6_1$, $P6_122$.

For $Z' = 2$, 19 space groups were searched: $P1$, $P-1$, $P2_1$, $C2$, Pc , Cc , $P2/c$, $P2_1/c$, $C2/c$, $P2_12_12_1$, $Pca2_1$, $Pna2_1$, $Iba2$, $Pccn$, $Pbcn$, $Pbca$, $P4_1$, $I4_1$, $R3$.

The structures were optimized with periodic DFT-D, using the PBE functional with the Neumann-Perrin dispersion correction. The optimizations were converged such that the forces on all atoms were smaller than 0.3 eV/Å and the total energy was converged to within 0.2 kJ mol⁻¹.

3.2.2 Identification of Candidate Structures using Experimental PXRD Patterns

GRACE produced a set of 1783 $Z' = 1$ and 2 structures at the reranking stage, i.e., at DFT-D level of theory, but with loose convergence criteria. To identify predicted structures corresponding to the known **GAL** polymorphs, cross-correlation methodology²⁹ coded in the GRACE program was used to search for matches to the experimental PXRD patterns. Pattern matching was performed over the 5-35° 2θ range using background-corrected experimental (transmission) powder patterns, except for Form X, where a simulated pattern from the experimental SCXRD structure was used. The pattern matching program allows the predicted structure (unit cell dimensions, molecular position and orientation) to adjust slightly to maximize the pattern similarity. The agreement between the experimental and calculated PXRD patterns was assessed by a figure of merit (*fom*), a number between 0 and 1. Normally, a score of about 0.96 is expected for 'perfect' matches, while scores less than 0.8 are clearly not matching.

Only Form I was not found among the GRACE CSP structures. This is due to poor convergence of the landscape at higher energies. Candidate structures with *fom* values of 0.8 and above were found for all other forms, Figure S28. For some of the forms, there was a unique match among the CSP structures, while in other cases, multiple candidates gave high *fom* values. The two structure matches to Form IV are identical apart from different 5-membered ring puckering, which given the low energy barrier separating them, may suggest disorder of this ring.

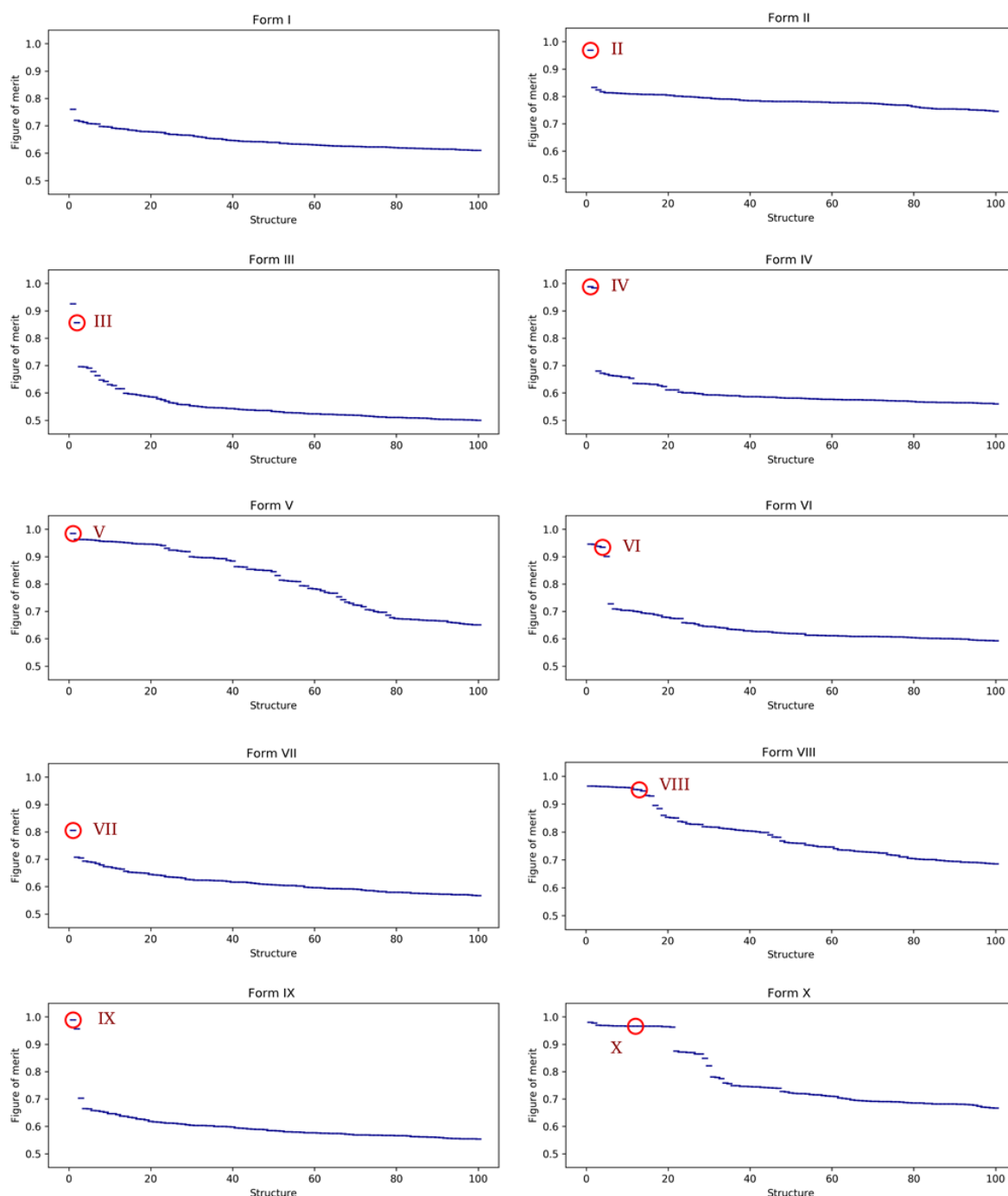


Figure S28. PXR D pattern similarities for the 100 best matches to experimental **GAL** Forms I-X.

The crystal structures of all of the neat **GAL** polymorphs, except Forms VII and VIII, were solved by either SCXR D or PXR D. Unfortunately, neither Forms VII and VIII could be grown as single crystals or isolated as pure phases for their PXR D patterns to be readily indexed. However, the closest match from CSP2 to the Form VII PXR D pattern helped to identify Form IV as the main phase impurity in Form VII material. Armed with this information, the PXR D pattern was successfully indexed and the structure ultimately solved from PXR D (Section S2.2.1).

For Form VIII, there were many matches from CSP2. To help identify the best one, NMR shieldings were calculated for D3BJ optimized Form VIII CSP candidate structures using the GIPAW method as described in Section S2.3. The correlation between the calculated shieldings and the experimental chemical shifts of the other forms were used to convert the calculated shieldings of the Form VIII candidates to chemical shifts. Carbon-13 calculated isotropic chemical shifts are plotted with a Lorentzian lineshape (0.4 ppm HWHM) to facilitate comparison to the experimental ssNMR spectrum, Figure S29. There is no obvious peak splitting in the experimental ssNMR spectrum of Form VIII to suggest it is a $Z' > 1$ structure. The only $Z' = 1$ structure among the Form VIII candidates is #39.

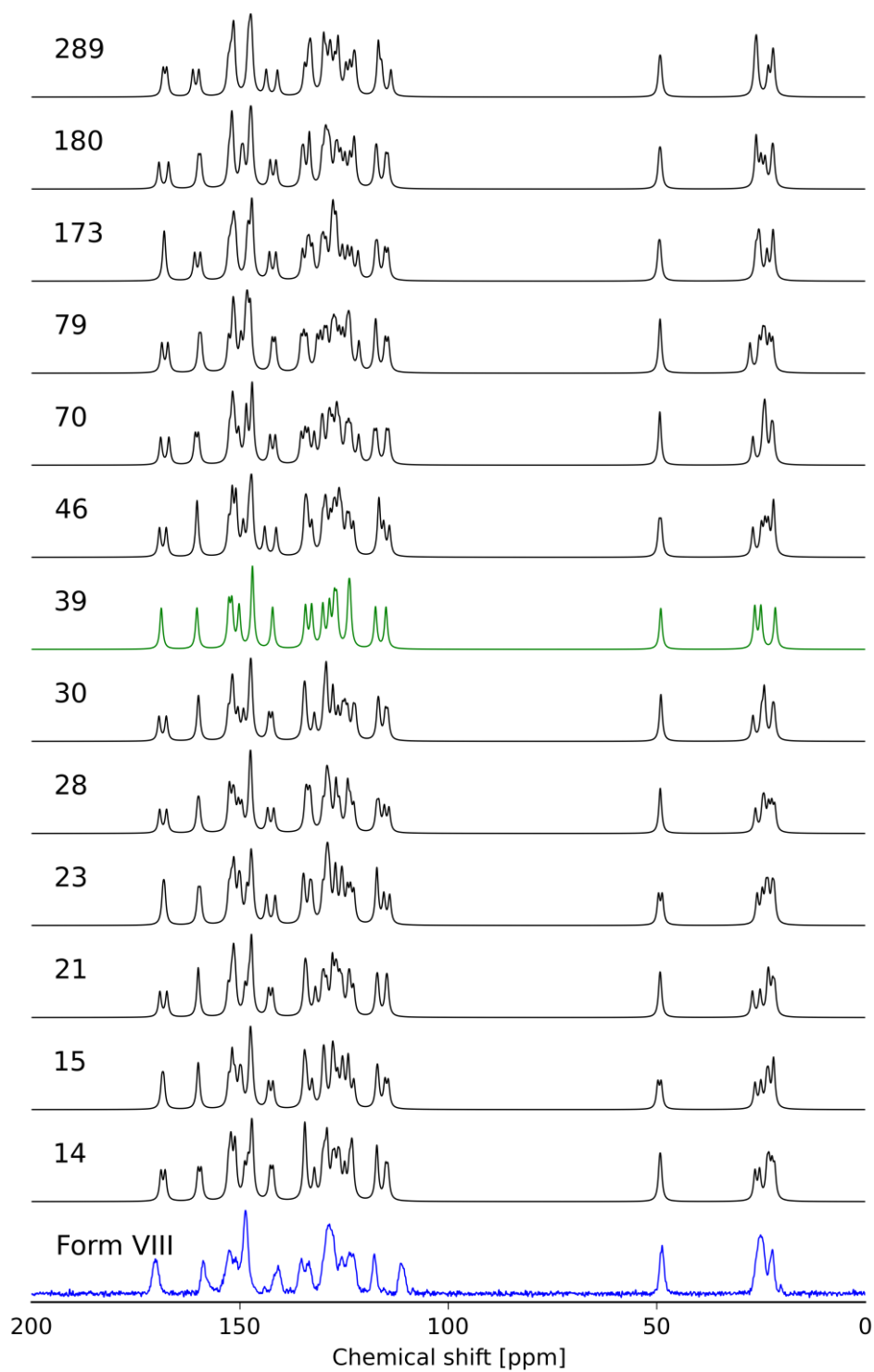


Figure S29. Computed ^{13}C ssNMR spectra of **GAL** Form VIII candidate structures, along with the experimental Form VIII ssNMR spectrum (blue). Structure 39 (green) is the only $Z' = 1$ structure among the Form VIII candidates.

The best structure matches to all of the experimental forms (except Form I) from the PXRD matching with GRACE are summarized in Table S16. Matches to SCXRD and PXRD structure solutions were also confirmed using the Packing Similarity tool in Mercury. Shown in Figure S30 are overlays of the experimental crystal structures solved from single crystals at 100 K or powders at room temperature and the corresponding PBE+D3BJ minimized structures. Excellent agreement ($\text{RMSD}_{20} < 0.200 \text{ \AA}$) between the atomic coordinates in the experimental and computer-generated DFT-D structures indicates that the models are appropriate for **GAL**.³⁰

Table S16. Summary of PXRD Matching of GRACE Structures to Experimental PXRD Patterns.

Structure	GRACE rank	Relative Energy (kJ mol ⁻¹)	fom	RMSD ₂₀ (Å)
Form II	366	8.46	0.969	0.136
Form III	543	9.43	0.857	0.152
Form IV	5	0.91	0.987	0.081
Form V	78	5.24	0.962	0.127
Form VI	47	4.72	0.933	0.091
Form VII*	2	0.41	0.805*	0.158
Form VIII	39	3.83	0.965	0.194
Form IX	6	0.95	0.989	0.172
Form X	13	3.29	0.967	0.156

* The experimental sample of Form VII was contaminated with Form IV, why this figure of merit is substantially smaller.

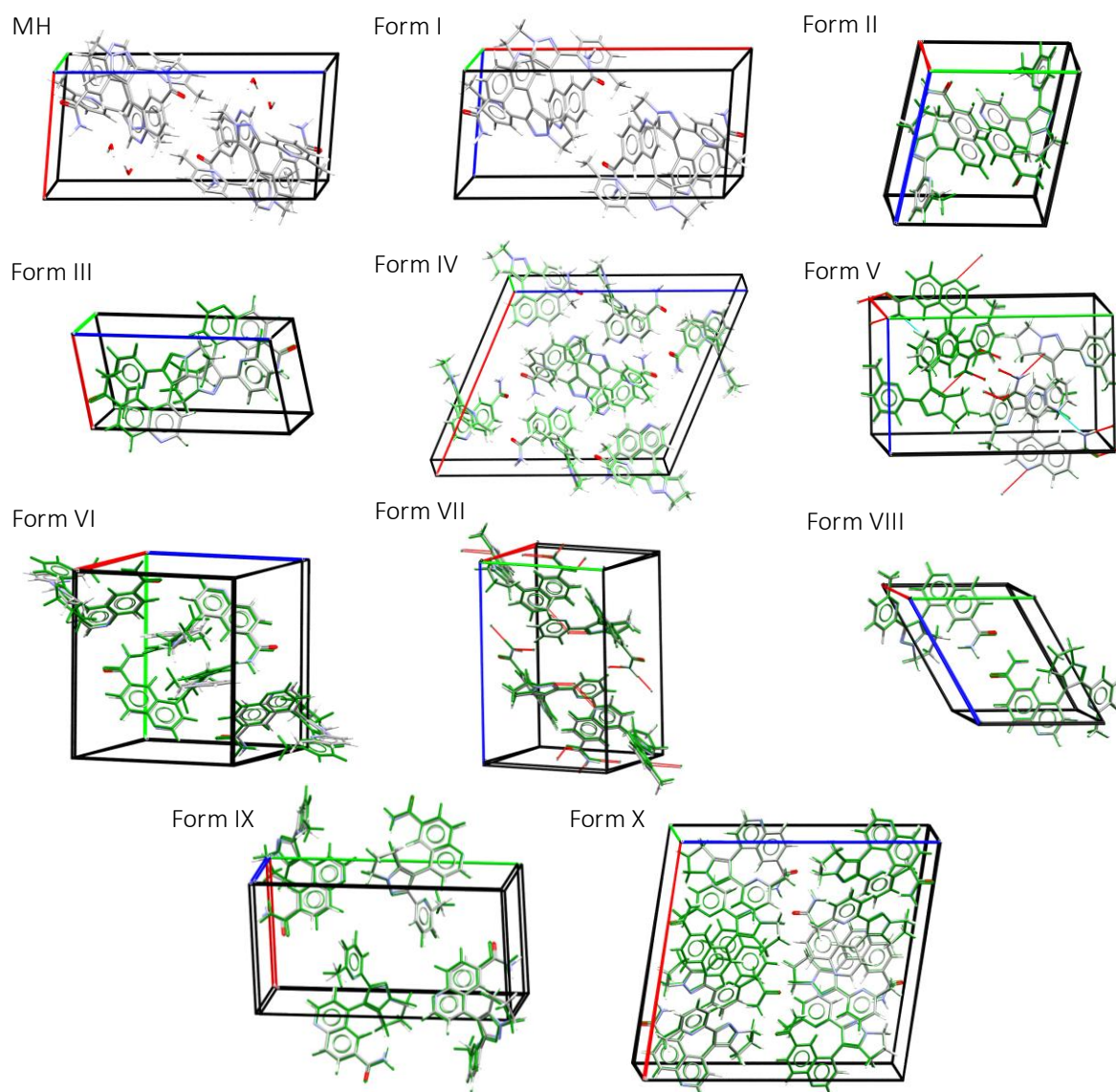


Figure S30. Overlay of the experimental (colored by element) and GRACE-generated (green) crystal structures of **GAL** monohydrate (experimental only), Form I (experimental only) and Forms II-X. Experimental structures were solved from either single crystals at 100K (monohydrate, Forms II, III, IV, VI, X) or powder diffraction data (Forms I, V, VII, VIII, IX) at 296 K.

3.3 Summary of CSP results

The most striking difference between CSP1 and CSP2 is the large number of unobserved structures that are significantly more stable than the observed forms in CSP1 (Figure S31). The GM and the vast majority of these structures have H18...N2 and H19...N4 hydrogen bonds with the A conformation, and one dense structure with the B conformation has a H18...N4 hydrogen bond (Figure S26). The statistically unlikely hydrogen bonding to N4 (see Figure S34) appears to be over stabilizing these structures, which could be an artefact of the same empirical repulsion-dispersion parameters being

used for all N atoms. However, it is notable that CSP2, which uses periodic DFT+D calculations has the same GM structure and the low energy $Z' = 1$ structures are common to the two searches.

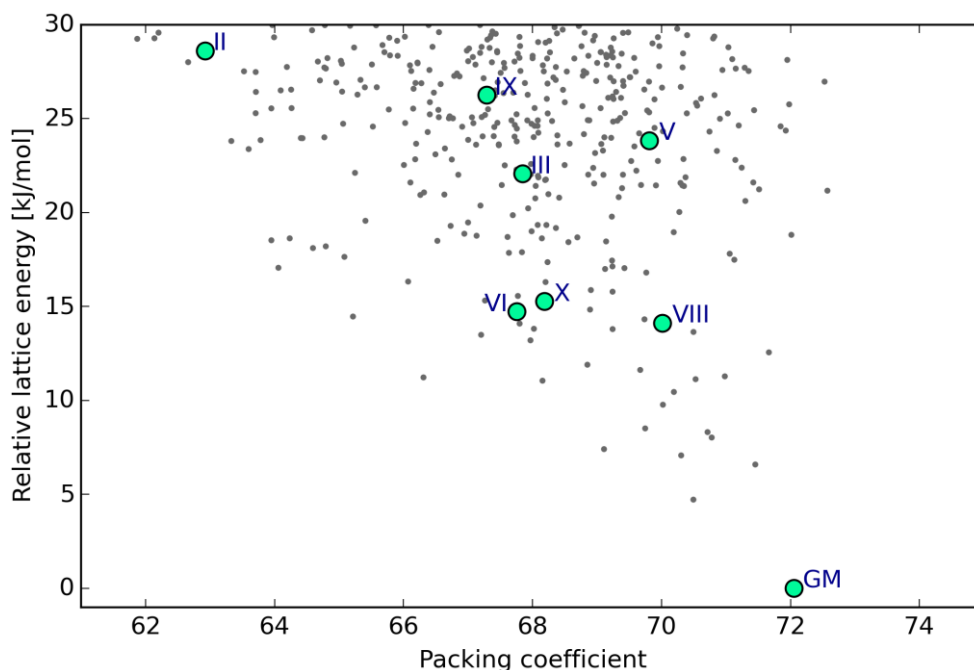


Figure S31. The output of CSP1, presented for ease of comparison with Figure 4 of manuscript.

Another key learning point for Ψ_{mol} methods is that more angles need to be optimized in response to the packing forces to give an acceptable overlay of the structures (Table S15) than would be required for smaller molecules, because the size of the molecule means that small changes in central angles can give rise to significant shifts in the position of peripheral interaction sites.³¹

All CSP methods have the problem that discarding “duplicate” structures may result in minima that do not differ markedly in the overall packing, (e.g. just in the configuration of the 5-membered ring), being lost.

The choice of CSP method is determined by many factors, including how the results might influence the project.

The computer-generated structures and CrystalOptimizer minimized neat experimental structures can be found in the Supporting Information CIF file Galunisertib_CSP1.cif, with energy range up to +30 kJ mol⁻¹ for CSP1 (390 structures, named Xyy_z_eee, where X is the conformational region, yy the rank at the CrystalPredictor stage, z the rank at the CrystalOptimizer stage and -eee the lattice energy in kJ mol⁻¹). The atom numbering used in the CSP1 CIF is shown in Figure S6(a).

Structures generated in CSP2 are included in the supplementary CIF file Galunistertib_CSP2.cif. The 638 structures within 10 kJ/mol of the global minimum are included, they are named 00nnn_e.ddd, where ‘nnn’ denotes the rank number, and ‘e.ddd’ denotes the lattice energy in kJ mol⁻¹ relative to

the global minimum structure (absolute lattice energies were not calculated in CSP2). The atom numbering used in the CSP2 CIF is shown in Figure S6(b).

4 Crystal Structure Analysis / Solid Form Risk Assessment

4.1 Packing Similarity / Analysis

In total, crystal structures of 10 neat polymorphs, the monohydrate and 22 solvates from diverse solvents were solved. Isostructurality was evident for some solvate pairs on inspection of the SCXRD structures and also comparison of PXRD patterns (for solvates lacking SCXRD structures). The crystal packing of **GAL** in the neat and solvated crystal forms was compared by two methods: Mercury Crystal Packing Similarity and XPac. Table S17 gives the results of all the pairwise matches observed in XPac. The structures fell into two distinct groups, according to the gross shape of the molecule.

Table S17. XPac Structural Matches between Experimental **GAL** Crystal Structures.

	Monohydrate	Form I	1-BuOH solvate	1-PrOH solvate I	EtOH solvate III	iAmOH solvate	2-PrOH solvate I	2-BuOH solvate	MEP solvate I	NMP solvate II	Form III	EtOH solvate I	hemi-1,4-dioxane hydrate	Form X	EtOAc hemisolvate	Form IV	2-MeTHF solvate	DCM solvate I	1,4-dioxane solvate IV	morpholine solvate II	toluene solvate I	acetic acid solvate II	heptane hemisolvate	1,4-dioxane disolvate	Form IX	THF solvate I	Form II	MEK solvate III	Form VI	Form VII	Form VIII	Form V	MeOH solvate I	Global minimum		
Monohydrate	x	3	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Form I	3	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
1-BuOH solvate	-	x	3	3	3	3	3	3	2	1	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
1-PrOH solvate I	-	-	3	x	3	3	3	3	2	0	1	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
EtOH solvate III	-	-	3	3	x	3	3	3	2	1	0	1	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
iAmOH solvate	-	-	3	3	3	x	3	3	2	0	1	1	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
2-PrOH solvate I	-	-	3	3	3	3	x	3	1	0	1	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
2-BuOH solvate	-	-	3	3	3	3	3	x	2	0	1	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
MEP solvate I	-	-	2	2	2	2	1	2	x	1	-	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
NMP solvate II	-	-	1	0	1	0	0	0	1	x	0	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Form III	-	-	1	1	0	1	1	1	-	0	x	2	1	-	-	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
EtOH solvate I	-	-	1	1	1	1	1	0	0	0	0	2	x	1	1	0	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
hemi-1,4-dioxane hydrate	-	-	1	1	1	0	0	0	0	0	1	1	x	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
Form X	-	-	-	-	-	0	-	-	-	-	-	1	-	x	0	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	
EtOAc hemisolvate	-	-	-	-	-	-	-	-	-	-	0	-	0	-	x	0	-	-	-	-	-	-	-	-	-	-	-	0	-	0	-	-	-	-	-	
Form IV	-	-	-	-	-	-	-	-	1	0	-	-	0	-	x	0	0	0	0	0	0	0	0	0	1	0	0	0	0	0	0	0	-	-	-	-
2-MeTHF solvate	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	x	3	3	3	2	2	1	1	1	1	1	1	1	1	1	1	0	-	-	-	-
DCM solvate I	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	3	x	3	3	2	2	1	2	1	1	1	1	1	1	1	0	-	-	-	-	-
1,4-dioxane solvate IV	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	3	3	x	3	2	2	1	1	1	1	1	1	1	1	1	0	-	0	-	-	-
morpholine solvate II	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	3	3	3	x	2	2	1	1	1	1	1	1	1	1	0	1	0	-	-	-	-
toluene solvate I	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	2	2	2	2	x	2	2	2	1	1	1	1	1	1	0	0	1	0	-	-	-
acetic acid solvate II	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	2	2	2	2	2	x	3	1	1	1	1	1	1	1	1	0	-	-	-	-	-
heptane hemisolvate	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	1	1	1	2	3	x	1	1	1	1	1	1	0	-	-	-	-	-	-	-
1,4-dioxane disolvate	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	2	1	1	2	1	1	x	1	2	2	1	1	0	1	0	-	-	-	-	
Form IX	-	-	-	-	-	-	-	-	-	-	-	-	-	-	1	1	1	1	1	1	1	1	1	x	1	1	1	1	0	0	0	0	-	-	-	-
THF solvate I	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	1	1	1	1	1	1	1	2	1	x	3	1	1	-	-	0	-	-	-	
Form II	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	1	1	1	1	1	1	1	2	1	3	x	3	1	0	1	0	0	-	-	
MEK solvate III	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	1	1	1	1	1	1	1	1	1	1	3	x	0	1	2	0	-	-	-	
Form VI	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	1	1	1	1	0	1	0	1	1	0	x	0	-	0	-	-	-	-	-	-	
Form VII	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	0	0	0	0	0	0	-	0	0	-	0	1	0	x	0	-	-	-	-	
Form VIII	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	-	-	-	1	1	-	-	1	0	-	1	2	-	0	x	3	1	-	-		
Form V	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	0	-	0	-	0	0	-	0	0	-	0	-	3	x	1	-	
MeOH solvate I	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	0	0	-	0	-	0	-	1	1	x	-		
Global minimum	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	x		

3 = 3D structural similarity; 2 = 2D structural similarity; 1 = 1D structural similarity; 0 = 0D structural similarity. Only 100 K structures were used, with the exception of ethanol solvate form I where only a room temperature structure is available. Solvent molecules were removed for the comparisons.

There are clear structure relationships between Forms II, III and V and their parent solvates, Figure S32. Form II uniquely porous, the lowest density of the observed **GAL** polymorphs. None of the other known polymorph structures shows solvent-accessible pores.

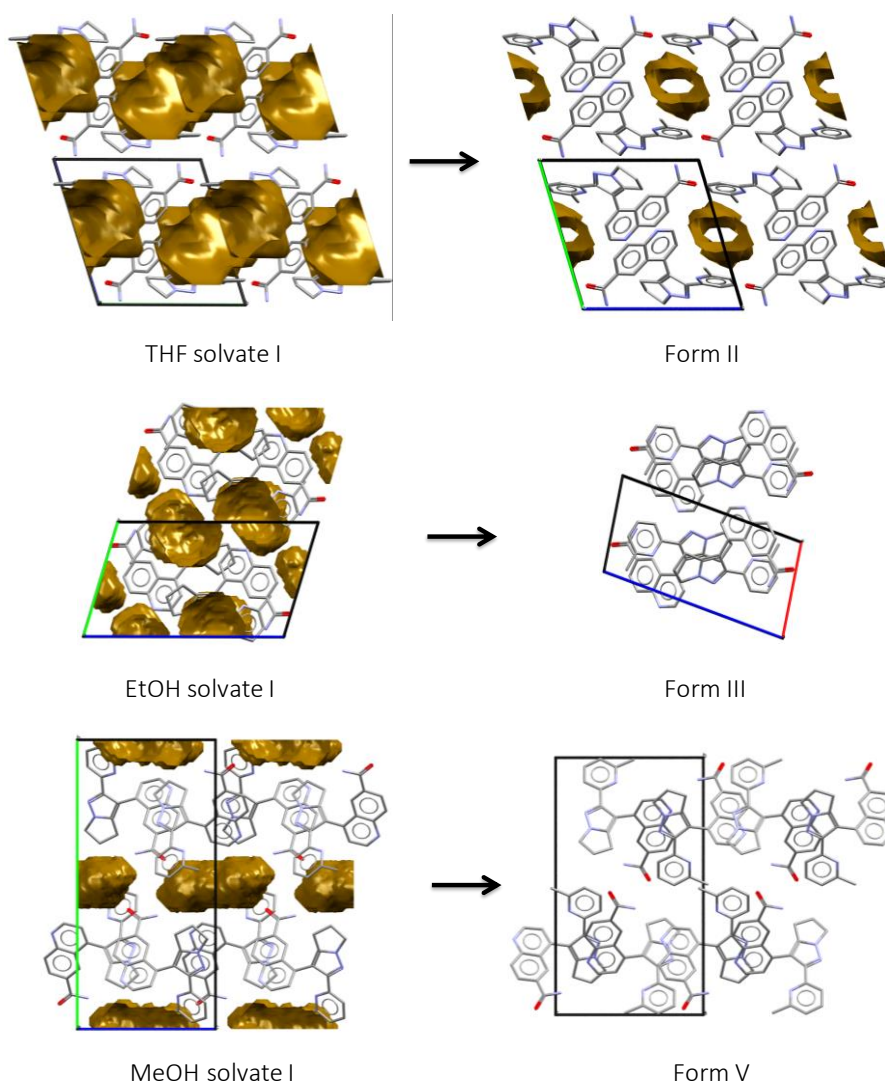


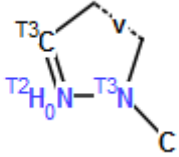
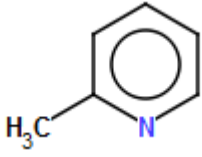
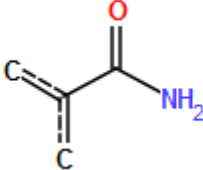
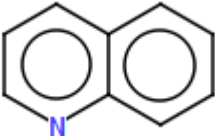
Figure S32. Structure relationships between parent solvates and product neat polymorphs. Solvent pores for solvate structures with the solvents removed and the neat polymorphs were calculated with a 1.2 Å probe radius using the Mercury program.

4.2 Hydrogen Bonding Propensity / Topology

The logit hydrogen bond propensity (HBP) tool³² was used to assess the likelihood of pairwise hydrogen bonding interactions based on a statistical analysis of hydrogen bonds in relevant structures in the Cambridge Structural Database (CSD version 5.39 + Nov 2017 update).³³ The hydrogen bond donor-acceptor pairings in the neat polymorph crystal structures and in the calculated global minimum structure, GM, were analyzed. Pairwise hydrogen bonding interactions were modelled for the amide NH donor and quinolone N1, pyrazoline N2, pyridine N4 and amide O1 acceptors in the **GAL** molecule, Table S18. The best HBP model (AUC = 0.803) was derived from a

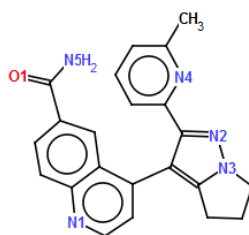
training set comprised of 2390 relevant structures (3D coordinates determined, R factor $\leq 0.075\%$, not disordered, no errors, not polymeric, only organics) retrieved from the CSD. Each of the four acceptors was equally represented by 650 structures in the training set. Hydrogen bond propensity scores for all intermolecular interactions (no intramolecular hydrogen bonds were predicted) derived from the model and using default hydrogen bond geometry definitions are summarized in Table S19.

Table S18. Functional Group Definition for CSD Substructure Searches.

Search Fragments			
pyrazoline_2		Me_pyridine	
ar_amide		quinoline	

Key: Tn = atom makes n bonds; Hn = n bonded hydrogen atoms; v = any bond order

Table S19. Hydrogen Bond Propensity Analysis of Neat **GAL** Polymorphs and the GM Structure.



Donor	Amide N5			
Acceptor	Quinoline N1	Pyridine N4	Amide O1	Pyrazoline N2
Propensity	0.88	0.77	0.76	0.54
Bounds	(0.84,0.91)	(0.69,0.83)	(0.70,0.80)	(0.45,0.64)
Form I			yes	
Form II			yes	yes
Form III	yes		yes	
Form IV*	yes (A)		yes (A/B)	yes (A)
Form V	yes		yes	
Form VI	yes			yes
Form VII*	yes (A)	yes (A)	yes (A)	yes (B)
Form VIII			yes	yes
Form IX			yes	yes
Form X			yes	yes
GM		yes		yes

* Z' = 2 structure, A and B refer to the independent molecules in each structure.

Figure S33 shows many different hydrogen bonding motifs. H18 is observed to hydrogen bond to N1 (Form III, Form IV – molecule 2 and Form V) or N2 (Form II, Form IV – molecule1 and Form VI), while H19 is observed to hydrogen bond to O1 (Form II, Form III, Form IV and Form V) or N1 (Form VI). Form III and Form V have the same graph sets, yet Form III forms a chain and Form V forms a sheet. This is because the $R_2^2(8)$ double hydrogen bond interactions take place all on the same side of the $C_1^1(8)$ chains in Form III, but on opposite sides of the $C_1^1(8)$ chains in Form V, allowing further chains to form through the molecules on the opposite sides of the dimers. This highlights the pitfalls of relying solely on the graph set analysis. Although Forms III and V have the same combinations of hydrogen bond donors and acceptors, Form III has these arranged in a 1-dimensional chain and Form V has these arranged in a sheet.

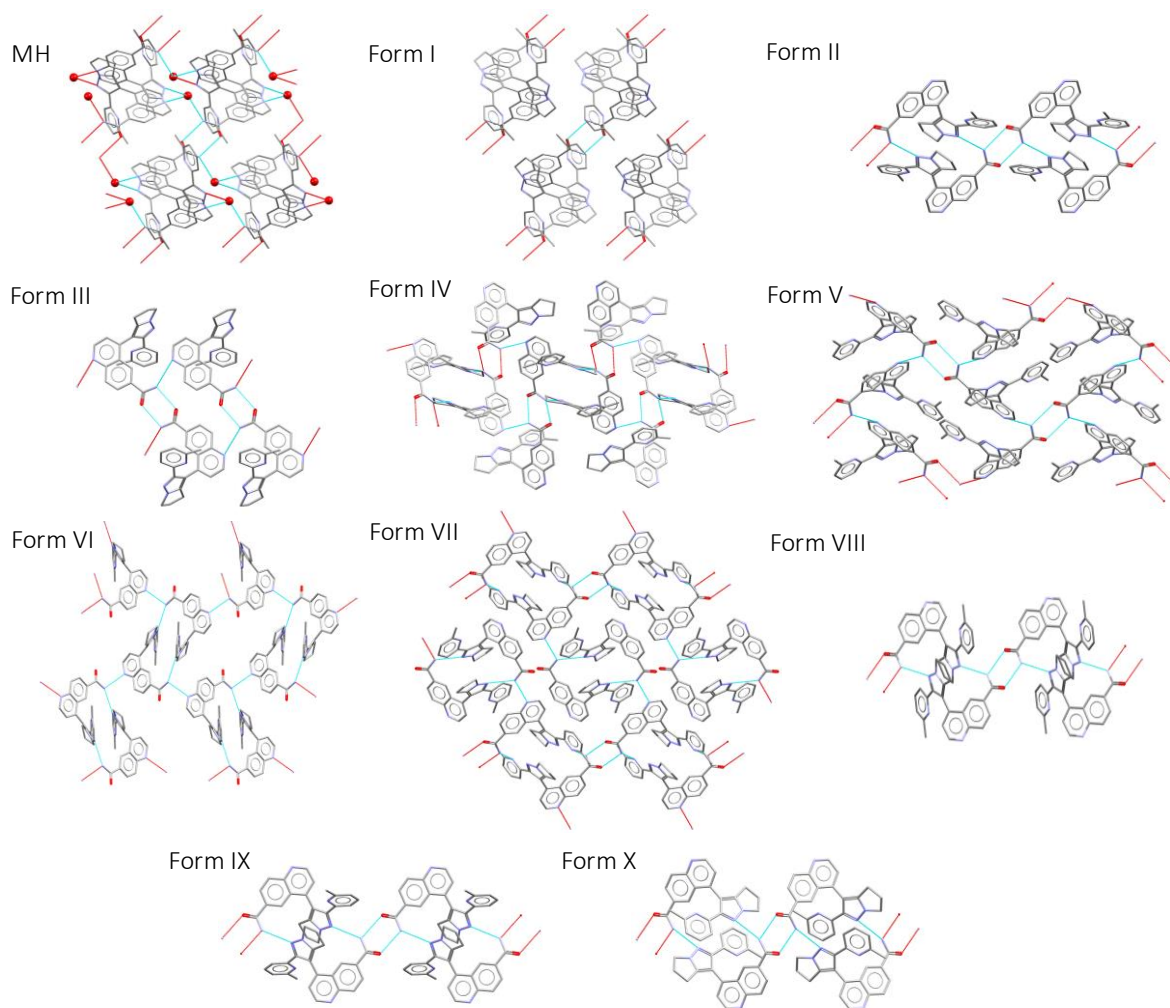


Figure S33. Hydrogen bonded 1D ribbons and 2D layers of **GAL** in monohydrate (MH) and Forms I-X.

4.3 Full Interaction Maps

Insights into the *quality* of the hydrogen bonding interactions were provided by full interaction map (FIM)³⁴ visualization of the interaction preferences in relation to the hydrogen bonding groups of near neighbor **GAL** molecules. FIMs are intermolecular interaction density maps generated by superimposing 3D Isostar scatterplots of individual functional groups onto whole molecules or clusters of molecules, taking into account steric crowding factors. Based on statistical analysis of the CSD, they represent the probability of finding an interaction above the random chance that it will occur. By taking into account the geometry of the hydrogen bonding interactions, FIM visualization complements the HBP analysis. The FIMs of isolated **GAL** molecules from each crystal structure are shown in Figure S34. Interaction preferences about the isolated **GAL** molecule show the expected donor and acceptor peaks for the amide, quinolone and pyrazoline groups. Except for the Form VII_B conformation, where the pyridine-N is outward facing, a hotspot near this acceptor atom is missing due to steric hindrance. A hydrogen bond is identified to the pyridine-N atom in the GM structure, however, the geometry of this interaction is severely compromised. This pyridine acceptor was used in a number of the other low energy structures on CSP1 (Figure S26), yet the FIMs of each of those structures also showed that this acceptor was inaccessible.

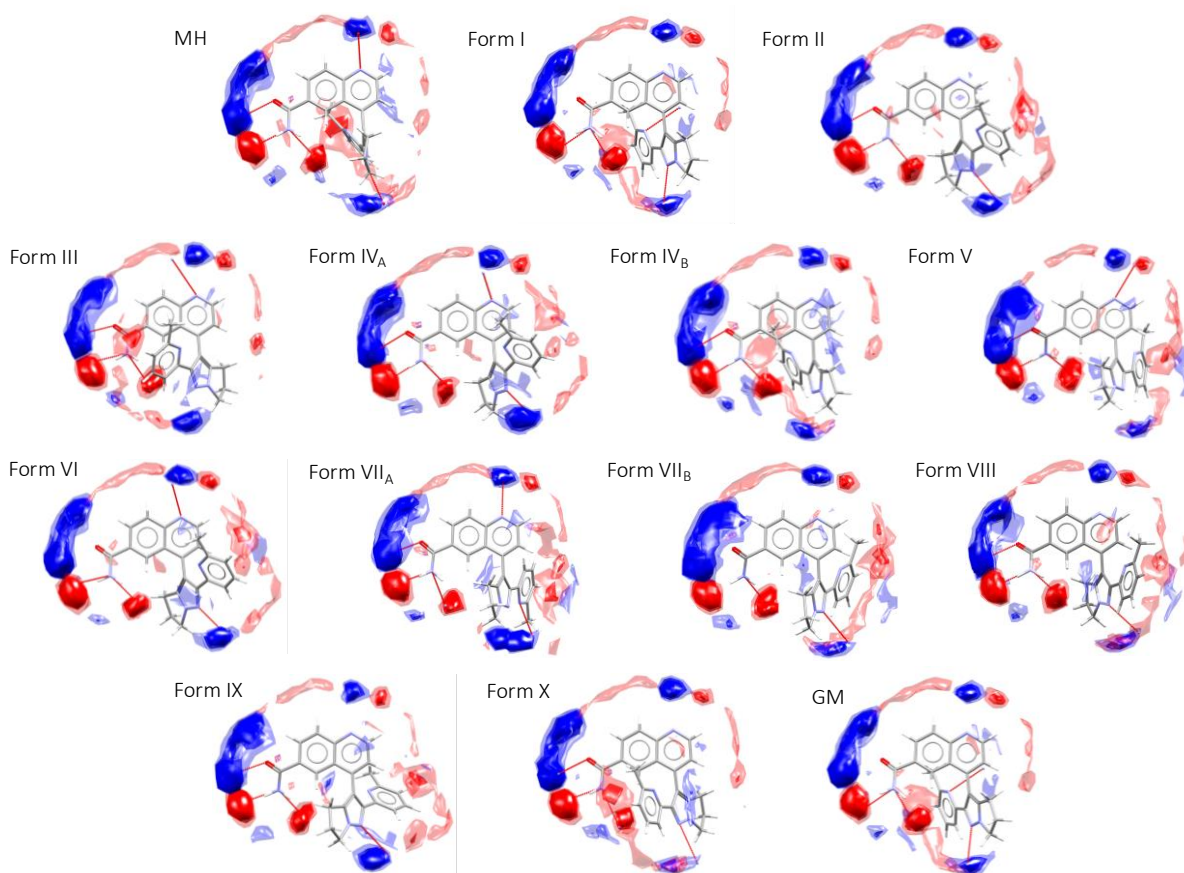


Figure S34. Full interaction maps of **GAL** in the monohydrate (MH), Forms I-X and global minimum (GM) structure. Hydrogen bond donor and acceptor hotspots were surveyed using carbonyl O acceptor (red) and uncharged NH donor (blue) probes.

4.4 Conformational Strain Energy / Molecular Surface Area

Molecular strain energies were calculated by computationally extracting molecules from the GRACE crystal structures and optimizing them in vacuum using Gaussian16. The strain energy was calculated as the difference in B3LYP+D3BJ/6-311G(d,p) energy between the in-crystal geometry and the relaxed conformer in vacuum. In-crystal geometries were obtained by constrained optimization of the bond lengths/angles, fixing the torsion angles, while relaxed conformers were obtained by full relaxation of the structures. The molecular surface areas of the optimized in-crystal conformers were calculated using the Atom Volumes and Surfaces tool in the BIOVIA Materials Studio 2018 program, setting the Connolly radius to 1.2 Å, solvent surface to 0 Å, and using a grid interval of 0.15 Å. An overlay of the crystal conformers, as well as the local energy minima (relaxed conformers), are shown in Figure S35. The results of the strain energy and molecular surface area calculations are summarized in Table S20.

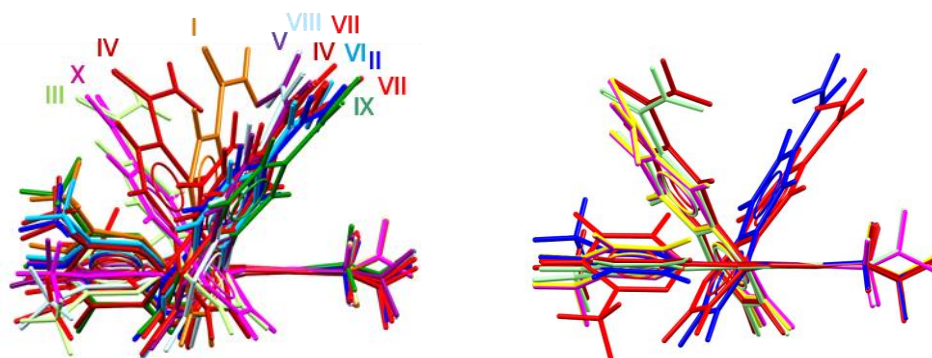


Figure S35. GAL crystal conformers in Forms I-X (left) and local conformational minima (right). Crystal conformers in Forms II, IV (conformer B), V, VI, VII (conformer B), VIII and IX minimized to the same local minimum (shown in blue).

Table S20. Conformational Strain Energies and Molecular Surface Areas in Forms I-X and the GM Structure.

Structure/Conformer	Strain Energy (kJ mol ⁻¹)	Local Minimum Relative Energy (kJ mol ⁻¹)	Molecular Surface Area (Å ²)
III	9.39	7.84	348.30
VII 1	22.54	26.42	352.77
VII 2	15.90	12.39	352.22
X	14.31	0	341.65
II	12.75	12.39	349.18
V	7.43	12.39	354.32
IV 1	16.94	7.91	342.57
IV 2	7.46	12.39	352.28
VI	12.25	12.39	350.44
GM	25.43	0.32	346.08
IX	15.80	12.39	353.98
VIII	11.25	12.39	355.40

5 Abbreviations

AA	antisolvent addition	MEK	methyl ethyl ketone
AUC	Area under the curve	MeOAc	methyl acetate
BuOH	butanol	MeOH	methanol
ACN	acetonitrile	MeTHF	methyl tetrahydrofuran
CP	cross-polarization	MH	monohydrate
CSD	Cambridge structure database	MIBK	methyl isobutyl ketone
CSP	crystal structure prediction Grimme's third dispersion correction with Becke-Johnsson's damping	MIPK	methyl isopropyl ketone
D3BJ	function	nBuOAc	n-butyl acetate
DAC	diamond anvil cell	NMP	N-methylpyrrolidinone
DCM	dichloromethane	NMR	nuclear magnetic resonance
DFT	Density functional theory	NP	Neumann-Perrin dispersion correction Perdew-Burke-Ernzerhof density functional
DFT-D	dispersion-corrected DFT	PBE	
DMAC	N,N-dimethyl acetamide	PBE0	
DMF	N,N-dimethyl formamide	PrOH	1-propanol
DMSO	dimethyl sulfoxide	PTFE	polytetrafluoroethylene (teflon)
DSC	differential scanning calorimetry European Synchrotron Radiation Facility	PTM	pressure-transmitting medium
ESRF		PXRD	Powder X-ray diffraction
EtOH	ethanol	RAA	reverse antisolvent addition
FIM	Full interaction map	RH	relative humidity
FIT	A force field	RMSD	root mean square deviation
GAL	galunisertib	RPM	revolutions per minute
GM	global minimum	RT	room temperature
GVS	gravimetric vapor sorption	SCXRD	single crystal X-ray diffraction
HBP	hydrogen bond propensity	SM	starting material
HF	Hartree-Fock	ssNMR	solid state nuclear magnetic resonance
HPLC	high pressure liquid chromatography	tBuOH	tert-butanol
HWHM	Half width at half maximum	TGA	thermogravimetric analysis
iAmOH	isoamyl alcohol	THF	tetrahydrofuran
IPA	isopropyl alcohol	VASP	Vienna ab initio simulation package
iPrOAc	isopropyl acetate	VL	vapor liquid diffusion
MAS	magic angle spinning	VS	vapor solid diffusion

6 References

- ¹ Oswald, I. D. H.; Chataigner, I.; Elphick, S.; Fabbiani, F. P. A.; Lennie, A. R.; Maddaluno, J.; Marshall, W. G.; Prior, T. J.; Pulham, C. R.; Smith, R. I., Putting pressure on elusive polymorphs and solvates. *CrystEngComm* 2009, 11 (2), 359-366
- ² Merrill, L.; Bassett, W. A., Miniature diamond anvil pressure cell for single crystal x-ray diffraction studies, *Review of Scientific Instruments* 1974, 45, 290-294.
- ³ Piermarini, G. J.; Block, S.; Barnett, J. D.; Forman, R. A., Calibration of the pressure dependence of the R1 ruby fluorescence line to 195 kbar, *Journal of Applied Physics* 1975, 46 2774-2780.
- ⁴ Boldyreva, E. V.; Shakhtshneider, T. P.; Ahsbahs, H.; Sowa, H.; Uchtmann, H., Effect of High Pressure on the Polymorphs of Paracetamol. *Journal of Thermal Analysis and Calorimetry* 2002, 68, (2), 437-452.
- ⁵ Fabbiani, F. P. A.; Allan, D. R.; David, W. I. F.; Moggach, S. A.; Parsons, S.; Pulham, C. R., High-pressure recrystallisation - a route to new polymorphs and solvates. *CrystEngComm* 2004, 6, 504-511.
- ⁶ Connor, L. E.; Morrison, C. A.; Oswald, I. D. H.; Pulham, C. R.; Warren, M. R., Carbon dioxide binary crystals via the thermal decomposition of RDX at high pressure. *Chemical Science* 2017, 8 (7), 4872-4878.
- ⁷ SAINT: Area detector control and data integration and reduction software, Bruker AXS: Madison, WI, USA, 2001.
- ⁸ Sheldrick, G. M., A short history of SHELX. *Acta Crystallogr. Sect. A* 2008, 64, 112-122.
- ⁹ Nowell, H.; Barnett, S. A.; Christensen, K. E.; Teat, S. J.; Allan, D. R., I19, the small-molecule single-crystal diffraction beamline at Diamond Light Source, *Journal of Synchrotron Radiation*, 2012, 19, 435-441.
- ¹⁰ CrysAlisPro 1.171.38.41k (Rigaku OD, 2015)
- ¹¹ ABSPACK, CrysAlis CCD and CrysAlis RED, 1.171; Oxford Diffraction Ltd: Abingdon, Oxfordshire, England, 2006.
- ¹² Sheldrick, G. M., Crystal structure refinement with SHELXL. *Acta Crystallographica Section C-Structural Chemistry* 2015, 71, 3-8.
- ¹³ Thorn, A.; Dittrich, B.; Sheldrick, G. M., Enhanced rigid-bond restraints. *Acta Crystallographica Section A* 2012, 68, 448-451.
- ¹⁴ Dolomanov, O. V.; Bourhis, L. J.; Gildea, R. J.; Howard, J. A. K.; Puschmann, H., OLEX2: a complete structure solution, refinement and analysis program. *Journal of Applied Crystallography* 2009, 42, 339-341.
- ¹⁵ David, W. I. F., Shankland, K., van de Streek, J., Pidcock, E., Motherwell, W. D. S., Cole, J. C. DASH: a program for crystal structure determination from powder diffraction data. *J. Appl. Cryst.* 2006, 39, 910-915.
- ¹⁶ Markvardsen, A. J., David, W. I. F., Johnson, J. C., Shankland, K. A probabilistic approach to space-group determination from powder diffraction data. *Acta Crystallogr. A* 2001, 57, 47-54.
- ¹⁷ PLATON, A set of tools for the interpretation of structural results. *Acta Crystallogr. Sect D: Biol. Crystallogr.* 2009.
- ¹⁸ Pawley, G. S. Unit-cell refinement from powder diffraction scans. *J. Appl. Cryst.* 1981, 14, 357-361.
- ¹⁹ Rietveld, H. M. A profile refinement method for nuclear and magnetic structures. *J. Appl. Cryst.* 1969, 2 65-71.
- ²⁰ Coelho, A. A., Topas: Coelho Software: Brisbane, 2012.
- ²¹ Pantelides C.C., Adjiman C. S., Kazantsev A.V., General Computational Algorithms for Ab Initio Crystal Structure Prediction for Organic Molecules. In *Topics in Current Chemistry*, Atahan-Evrenk S., A.-G. A., Ed. Springer, Cham: 2014; Vol. 345.
- ²² Price, S. L., Is zeroth order crystal structure prediction (CSP_0) coming to maturity? What should we aim for in an ideal crystal structure prediction code? *Faraday Discussions* 2018, 211, 9-30.
- ²³ GRACE (version 2.4), Avant-garde Materials Simulation Deutschland GmbH.
- ²⁴ Karamertzanis, P. G.; Pantelides, C. C., Ab initio crystal structure prediction. II. Flexible molecules *Molecular Physics* 2007, 105 (2-3), 273-291.
- ²⁵ Frisch, M. J.; Trucks, G. W.; Schlegel, H. B.; Scuseria, G. E.; Robb, M. A.; Cheeseman, J. R.; Scalmani, G.; Barone, V.; Mennucci, B.; Petersson, G. A.; Nakatsuji, H.; Caricato, M.; Li, X.; Hratchian, H. P.; Izmaylov, A. F.; Bloino, J.; Zheng, G.; Sonnenberg, J. L.; Hada, M.; Ehara, M.; Toyota, K.; Fukuda, R.; Hasegawa, J.; Ishida, M.; Nakajima, T.; Honda, Y.; Kitao, O.; Nakai, H.; Vreven, T.; Montgomery, J. A., Jr; Peralta, J. E.; Ogliaro, F.; Bearpark, M.; Heyd, J. J.; Brothers, E.; Kudin, K. N.; Staroverov, V. N.; Kobayashi, R.; Normand, J.; Raghavachari, K.; Rendell, A.; Burant, J. C.; Iyengar, S. S.; Tomasi, J.; Cossi, M.; Rega, N.; Millam, J. M.; Klene, M.; Knox, J. E.; Cross, J. B.; Bakken, V.; Adamo, C.; Jaramillo, J.; Gomperts, R.; Stratmann, R. E.; Yazyev, O.; Austin, A. J.; Cammi,

R.; Pomelli, C.; Ochterski, J. W.; Martin, R. L.; Morokuma, K.; Zakrzewski, V. G.; Voth, G. A.; Salvador, P.; Dannenberg, J. J.; Dapprich, S.; Daniels, A. D.; Farkas, Ö.; Foresman, J. B.; Ortiz, J. V.; Cioslowski, J.; Fox, D. J. Gaussian 09, Revision D.01, 2009.

²⁶ Price, S. L.; Leslie, M.; Welch, G. W. A.; Habgood, M.; Price, L. S.; Karamertzanis, P. G.; Day, G. M., Modelling organic crystal structures using distributed multipole and polarizability-based model intermolecular potentials. *Physical Chemistry Chemical Physics* 2010, 12 (30), 8478-8490.

²⁷ Stone, A. J., Distributed Multipole Analysis: Stability for Large Basis Sets. *Journal of Chemical Theory and Computation* 2005, 1 (6), 1128-1132

²⁸ Kazantsev, A. V.; Karamertzanis, P. G.; Adjiman, C. S.; Pantelides, C. C., Efficient Handling of Molecular Flexibility in Lattice Energy Minimization of Organic Crystals. *Journal of Chemical Theory and Computation* 2011, 7 (6), 1998-2016.

²⁹ Habermehl, S.; Mörschel, P.; Eisenbrandt, P.; Hammer, S. M.; Schmidt, M. U., Structure determination from powder data without prior indexing, using a similarity measure based on cross-correlation functions. *Acta Crystallographica Section B* 2014, 70 (2), 347-359.

³⁰ Macrae, C. F.; Bruno, I. J.; Chisholm, J. A.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Rodriguez-Monge, L.; Taylor, R.; van de Streek, J.; Wood, P. A., Mercury CSD 2.0 - new features for the visualization and investigation of crystal structures. *Journal of Applied Crystallography* 2008, 41, 466-470.

³¹ Iuzzolino, L., Expanding crystal structure prediction to larger and more flexible molecules of pharmaceutical interest, Eng. D. thesis, (2018), University College London.

³² Galek, P. T. A.; Allen, F. H.; Fabian, L.; Feeder, N., Knowledge-based H-bond prediction to aid experimental polymorph screening. *CrystEngComm* 2009, 11, (12), 2634-2639.

³³ Allen, F. H., The Cambridge Structural Database: A quarter of a million crystal structures and rising. *Acta Crystallographica Section B - Structural Science* 2002, 58, (3), 380-388.

³⁴ Wood, P. A.; Olsson, T. S. G.; Cole, J. C.; Cottrell, S. J.; Feeder, N.; Galek, P. T. A.; Groom, C. R.; Pidcock, E., Evaluation of molecular crystal structures using Full Interaction Maps. *CrystEngComm* 2013, 15 (1), 65-72.