Supporting Information

Computational screening for prediction of co-crystals: method comparison and experimental validation

Fateme Molajafari[1], Tianrui Li[2], Mehrnaz Abbasichaleshtori[3], Moein Hajian Z.D.[1], Anthony F. Cozzolino[3], Daniel R. Fandrick[2], Joshua D. Howe[1,*]

[1]Department of Chemical Engineering, Texas Tech University, Lubbock, TX, 79409, USA

[2]Agios Pharmaceuticals Inc., Cambridge, MA 02139, USA

[3]Department of Chemistry and Biochemistry, Texas Tech University, Lubbock, TX, 79409, USA

[*]Correspondence: joshua.d.howe@ttu.edu (J.D.H.)

Table of Contents

Theory and Model

COSMO-RS

COSMO-RS calculations begin with the computation of molecular surface screening charge densities through quantum mechanical calculations. To obtain the screening charge densities, the continuum solvation model COSMO is employed by COSMO-RS to model a virtual conductor environment for the molecule of interest. Next, the solute molecule in a virtual conductor environment induces a polarization charge density (σ) on the molecular surface; subsequently, these charges generate a more polarized electron density than in a vacuum around the solute. The screening charge density serves as an appropriate descriptor of the molecular surface polarity[2].

Solvent screening and solubility prediction

The COSMO-RS approach is formulated for liquid bulk systems, however, with the available thermodynamic data from external sources, the COSMO-RS method can be extended to treat solid systems[3]. COSMO-RS predicts the solubility of solid compounds in various solvents using equations below:

$$\Delta G_{fus} = \Delta H_{fus} - T\frac{\Delta H_{fus}}{T_m}$$
(1)

In the equation above, ΔG_{fus}, ΔH_{fus}, and T_m denote the free energy of fusion, the enthalpy of fusion, and the melting point, respectively.

$$\ln(x_S) = [\mu_X - \mu_S - \Delta G_{fus}]/RT$$
(2)

Where $\mu_X, \mu_S, R, and\ T$ denote the chemical potential of the pure compound and the compound in solution, gas constant, and temperature, respectively. $\mu_X\ and\ \mu_S$ are calculated with high accuracy via COSMO-RS. COSMO-RS prerequires the ΔG_{fus} to estimate the solubility of a compound in a solvent (using equation (2)). ΔG_{fus} may be obtained either experimentally (through measuring T_m and ΔH_{fus} by DSC and applying eq. (1)), through working with experimental reference solubilities and COSMO-RS chemical potential estimation (using eq. (2) and solving for ΔG_{fus}), or via a Quantitative Structure-Property Relationship (QSPR) model available in the COSMO-RS implementations.

Co-crystal screening

COSMO-RS predicts the viability of co-crystal formation using equations below:

$$\Delta G_{CC} = \Delta H_{mix} - T\Delta S_{mix} - \Delta\Delta G_{fus} \approx \Delta H_{mix}$$

(3)

Where, ΔS_{mix} and $\Delta\Delta G_{fus}$ denote the entropy of mixing, and the difference between the free energy of fusion of the co-crystal and the reactants, respectively. We note that neglecting $\Delta\Delta G_{fus}$ is an approximation that we expect to result in greater errors than neglecting ΔS_{mix} because the former assumption neglects any enthalpic changes associated with the solid state order[4]. Accordingly, COSMO-RS starts from a virtually supercooled liquid mixture of the API and coformer and computes the mixing enthalpy, ΔH_{mix}, as a closely related phenomenon to the strength of the interactions in the co-crystal phase compared to the pure reactants[4]:

$$\Delta H_{mix} = H_{API-coformer} - x_m H_{API} - x_n H_{coformer} = H_{ex}$$

(4)

In the equation above, $H_{API-coformer}$, H_{API}, $H_{coformer}$, and x denote the molar enthalpy of supercooled co-crystal, the molar enthalpies of API and coformer in the pure states, and the mole fraction, respectively. The enthalpic preference of the supercooled liquid phase is also valid in a mixed crystal; therefore, the use of the liquid phase H_{ex} as a guide for co-crystal screening, is applicable[5]. In COSMO-RS theory, two compounds with $H_{ex} < 0$ are treated as miscible; the negative value of H_{ex} indicates that the molecular interactions in the supercooled liquid phase are thermodynamically more spontaneous compared to the pure forms.

Machine learning

In this section, we provide brief descriptions of the three ML models and highlight the range of their configurations used for hyper-parameter optimization.

Random Forest (RF). RF is a conventional ML algorithm that operates as an ensemble learning method, comprising multiple decision trees. In the context of classification tasks, RF outputs the class most frequently selected by these trees.

Support Vector Machine (SVM). SVM is another conventional ML algorithm employed for both classification and regression (supervised learning model). SVM maps training samples to points in a multi-dimensional space, with the objective of maximizing the gap width between the two classes. This approach allows for new samples to be projected into the same space and classified based on which side of the gap they fall.

Deep Neural Network (DNN). DNN is composed of dense layers. In DNN, complex patterns within datasets are captured by a stack of interconnected neurons. Interconnected layers of neurons are utilized as functional units for capturing intricate patterns within the feature distribution. The input features are initially passed through the input layer, then traversed through a series of hidden layers, and ultimately generate output through the output layer[6].

Hyper-parameter optimization

To search for the optimal hyper-parameters for the RF, SVM, and DNN models, grid search with 5-fold cross-validation was utilized.

Table S1. The optimized range of hyper-parameters used for ML models.

Model	Hyper-parameters range
RF	'bootstrap': [True, False] 'n_estimators': [10, 20, 30, 40, 50, 100, 150, 200, 300, 500] 'max_depth': [None, 5, 10, 15, 20, 30] 'min_samples_split': [2,3,4, 5,6,7,8,9, 10,12,15] 'min_samples_leaf': [1, 2, 3, 4, 5,10,15] 'max_features': [2, 5, 10, 15, 20, 30]
SVM	'C': [0.1, 1, 10, 100] 'kernel': ['linear', 'rbf', 'poly', 'sigmoid'] 'degree': [2, 3, 4] 'gamma': [0.0001, 0.001, 0.01, 0.1, 1]
DNN	'Optimizer': Adam 'Activation function': ReLU, sigmoid 'Epoch': [10, 25, 50, 100, 150, 200, 300, 500] 'number of hidden layers': [1, 2, 3, 4, 5] ' Size of hidden layers ': [32, 64, 128] 'learning_rate': [0.1, 0.01, 0.001, 0.0001] 'l2_reg': [0.1, 0.01, 0.001]

Table S2. Coformers used for co-crystal screening by COSMOtherm and the ML/QSAR model (**independent test set**).

Compound	Co-crystal	Reference
1-Butanol	Unknown	-
2,3,5,6-Tetramethylpyrazine	Unknown	-
2,5-Dihydroxybenzoic acid	Yes	Bučar et al.[7]
2-Picolonic acid	Unknown	-
4-Cyanopyridine	Unknown	-
4-toluamide	No	Wicker et al.[8]
6-Hydroxynicotinic acid	Unknown	-
Acetamide	Unknown	-
Acetic acid	Yes	Trask et al.[9]
Aconitic acid	Unknown	-
Adipic acid	Yes	Bučar et al.[10]
Ammonium alginate	Unknown	-
Anisole	Unknown	-
Anthranilamide	Yes	Wicker et al.[8]
Aspartame	Unknown	-
Aspartic acid	Unknown	-
Aspirin	Unknown	-
Benzamide	Unknown	-
Benzoic acid	Yes	Bučar et al.[11]
Betaine	Unknown	-
Butylated hydroxytoluene	Unknown	-
Caffeic acid	Yes	Satyanarayana et al.[12]
Caprolactam	Unknown	-
Caprylic acid	Unknown	-
Catechin	Yes	Syed et al.[13]
Catechol	No	Syed et al.[13]
Celecoxib	Unknown	-
Chloral hydrate	Unknown	-
Chlorobenzene	Unknown	-
Chloroform	No	This work
Cholic Acid	Unknown	-

Cinnamic acid	Yes	Kumar et al.[14]
Citric acid	Yes	Karki et al.[15]
Cyclohexane	Unknown	-
d-10-Camphorsulfonic acid	Unknown	-
D-Alanine	Unknown	-
Dichloromethane	No	This work
Dimethyl sulfoxide	Unknown	-
DL-phenylalanine	Unknown	-
Embonic acid	Unknown	-
Ferulic acid	Unknown	-
Fumaric acid	Unknown	-
Gallic acid	Yes	Clarke et al.[16]
Glutaric acid	Yes	Trask et al.[17]
Glycerol	Unknown	-
Glycine	Unknown	-
Glycol stearate	Unknown	-
Glycolic acid	Yes	Alvarez-Lorenzo et al.[18]
Hexamethylenetetramine	Unknown	-
Hippuric acid	Unknown	-
Isoniazid	Unknown	-
Isonicotinamide	Unknown	-
Lactic acid	Unknown	-
L-Pyroglutamic acid	Unknown	-
L-Tartaric acid	Yes	Friščić et al.[19]
Maleic acid	Yes	Trask et al.[20]
Malic acid	Yes	Leyssens et al.[21]
Malonic acid	Yes	Trask et al.[20]
Mannitol	Unknown	-
Methyl 4-hydroxybenzoate	Unknown	-
MTBE	Unknown	-
N,N-dimethylacetamide	Unknown	-
Naringenin	Yes	Cui et al.[22]
n-Butyl acetate	Unknown	-
Nicotinamide	Unknown	-
Nicotinic acid	Unknown	-
Orcinol	Unknown	-
Oxalic acid	Yes	Trask et al.[20]

Para-aminobenzoic acid	No	Wicker et al.[8]
Phosphoric acid	Yes	This work
Picolinamide	Unknown	-
Piperazine	Unknown	-
Proline	Unknown	-
Propanone	Unknown	-
Propyl gallate	Unknown	-
Propylene glycol	Unknown	-
Propylparaben	Unknown	-
Pterostilbene	Yes	Schultheiss et al.[23]
Pyrazinamide	Unknown	-
Pyrazine	Unknown	-
Pyrogallol	Yes	This work
Quercetin	Yes	Smith et al.[24]
Resorcinol	No	This work
Saccharin	Yes	Padrela et al.[25]
Salicylic acid	Yes	Wicker et al.[8]
Sebacic acid	Unknown	-
Serine	Unknown	-
Sorbitol	Unknown	-
Suberic acid	Unknown	-
Succinic acid	Yes	Friščić et al[26]
Sulfur dioxide	Unknown	-
Tert-butanol	Unknown	-
Theobromine	Unknown	-
Theophylline	Yes	Eddleston et al.[27]
Urea	Yes	MacFhionnghaile et al.[28]
Vanillic acid	Yes	Jacobs et al.[29]
Vitamin C	Unknown	-
1-Butanol	Unknown	-

Table S3. Positive and negative co-crystal samples for caffeine in the training and test dataset.

Compound	Co-crystal	Reference
1-Hydroxy-2-naphthoic acid	Yes	Bučar et al.[30]
1-Naphthoxyacetic acid	Yes	Kumar et al.[31]
2,3-Difluorobenzoic acid	Yes	Bučar et al.[10]
2,3-Dihydroxybenzoic acid	Yes	Bučar et al.[7]
2,4-Dihydroxybenzoic acid	Yes	Bučar et al.[7]
2,4-Dinitrobenzoic acid	Yes	Ghosh et al.[32]
2,5-Difluorobenzoic acid	Yes	Bučar et al.[10]
2-Aminobenzoic acid	Yes	Wicker et al.[8]
2-Chloro-5-nitroaniline	Yes	Ghosh et al.[32]
2-Fluoro-5-nitro-aniline	Yes	Ghosh et al.[32]
2-Fluoro-5-nitrobenzoic acid	Yes	Ghosh et al.[32]
2-Fluorobenzoic acid	Yes	Wicker et al.[8]
2-Hydroxy-1-naphthoic acid	Yes	Bučar et al.[30]
2-Iodo-4-nitroaniline	Yes	Ghosh et al.[32]
2-Methoxybenzamide	No	Wicker et al.[8]
2-Methoxybenzoic acid	Yes	Wicker et al.[8]
2-Methylbenzamide	No	Wicker et al.[8]
2-Methylbenzoic acid	Yes	Wicker et al.[8]
2-Naphthoxyacetic acid	Yes	Kumar et al.[31]
2-Nitrobenzamide	No	Wicker et al.[8]
2-Nitrobenzoic acid	Yes	Wicker et al.[8]
3,4-Dihydroxybenzoic acid	Yes	Bučar et al.[7]
3,5-Dihydroxybenzoic acid	Yes	Bučar et al.[7]
3-Aminobenzamide	No	Wicker et al.[8]
3-Aminobenzoic acid	No	Wicker et al.[8]
3-Fluorobenzamide	Yes	Wicker et al.[8]
3-Fluorobenzoic acid	Yes	Wicker et al.[8]
3-Hydroxy-2-naphthoic acid	Yes	Bučar et al.[30]
3-Hydroxybenzoic acid	Yes	Wicker et al.[8]
3-Methoxybenzamide	No	Wicker et al.[8]
3-Methoxybenzoic acid	Yes	Wicker et al.[8]
3-Methylbenzamide	No	Wicker et al.[8]

3-Methylbenzoic acid	Yes	Wicker et al.[8]
3-Nitrobenzamide	Yes	Wicker et al.[8]
3-Nitrobenzoic acid	Yes	Wicker et al.[8]
4-Aminobenzamide	No	Wicker et al.[8]
4-Chloro-3-nitroaniline	Yes	Ghosh et al.[32]
4-Chloro-3-nitrobenzoic acid	Yes	Ghosh et al.[32]
4-Fluoro-3- nitroaniline	Yes	Ghosh et al.[32]
4-Fluorobenzamide	No	Wicker et al.[8]
4-Fluorobenzoic acid	No	Wicker et al.[8]
4-Hydroxybenzamide	Yes	Wicker et al.[8]
4-Hydroxybenzoic acid	Yes	Wicker et al.[8]
4-Iodo-3-nitroaniline	Yes	Ghosh et al.[32]
4-Methoxybenzamide	Yes	Wicker et al.[8]
4-Methoxybenzoic acid	No	Wicker et al.[8]
4-Methylbenzoic acid	No	Wicker et al.[8]
4-Nitroaniline	Yes	Ghosh et al.[32]
4-Nitrobenzamide	Yes	Wicker et al.[8]
4-Nitrobenzoic acid	No	Wicker et al.[8]
5-Chlorosalicylic acid	Yes	Shefter et al.[33]
6-Hydroxy-2-naphthoic acid	Yes	Bučar et al.[30]
Acetyl-L-carnitine HCl	Yes	Zaworotko et al.[34]
Baicalein	Yes	Zaworotko et al.[34]
Chlorogenic acid	Yes	Mukherjee et al.[35]
Creatine	Yes	Zaworotko et al.[34]
Cyanuric acid monohydrate	Yes	Mukherjee et al.[35]
Diadzein	Yes	Zaworotko et al.[34]
Diaminodiphenyl sulfone	Yes	Todaro et al.[36]
Diethylbarbituric acid	Yes	Craven et al.[37]
Ellagic acid monohydrate	Yes	Zaworotko et al.[34]
Eosin	Yes	Kumar et al.[14]
Epicatechin gallate	Yes	Zaworotko et al.[34]
Ethyl gallate (dihydrate)	Yes	Mukherjee et al.[35]
Ethylvanillin	Yes	Zaworotko et al.[34]
Febuxostat	Yes	Chava et al.[12]
Ferulic acid	Yes	Zaworotko et al.[34].
Formic acid	Yes	Ghosh et al.[32]
Furosemide	Yes	Goud et al.[38]

Gallocatechin	Yes	Zaworotko et al.[34]
Genistein	Yes	Zaworotko et al.[34]
Glucosamine HCl	Yes	Zaworotko et al.[34]
Hesperetin	Yes	Chadha et al.[39]
Hesperidin	Yes	Zaworotko et al.[34]
Lamotrigine	Yes	Leksič, et al.[40]
Luteolin	Yes	Zaworotko et al.[34]
Melatonin	Yes	Zaworotko et al.[34]
Mesaconic acid	Yes	Leyssens et al.[21]
Methyl gallate	Yes	Sun et al.[41]
Methylvanillin	Yes	Zaworotko et al.[34]
Niclosamide	Yes	Sanphui et al.[42]
Norfloxacin	Yes	Chava et al.[12]
O-formylphenoxy acetic acid	Yes	Kumar et al.[43]
p-Coumaric acid	Yes	Schultheiss et al.[44]
P-formylphenoxy acetic acid	Yes	Kumar et al.[43]
P-formylphenoxypropionic acid	Yes	Kumar et al.[43]
Piceatannol	Yes	Chava et al.[12]
Propionic acid	Yes	Karan et al.[45]
Pyridoxine	Yes	Zaworotko et al.[34]
Resveratrol	Yes	Chava et al.[12]
Rosmarinic acid	Yes	Zaworotko et al.[34].
Rutin hydrate	Yes	Zaworotko et al.[34]
Salicylamide	Yes	Wicker et al.[8]
Silibinin	Yes	Zaworotko et al.[34]
Sinapic acid	Yes	Chava et al.[12]
Sulfacetamide	Yes	Leger et al.[46]
Sulfaproxyline	Yes	Gosh et al.[47]
Syringic acid (tetrahydrate)	Yes	Mukherjee et al.[35]
Trifluoroacetic acid	Yes	Trask et al.[9]
Vitamin B1 hydrochloride	Yes	Zaworotko et al.[34]
Vitamin B2	Yes	Zaworotko et al.[34]
Vitamin B9	Yes	Zaworotko et al.[34].

Results

Table S4. Results of the ML models coformers' screening and ranking on the independent test set. (Experimental reports on the coformers listed below can be found in Table S2.)

Compound	RF		SVM		DNN	
	Co-crystal	Ranking	Co-crystal	Ranking	Co-crystal	Ranking
1-Butanol	1	8	1	46	1	53
2,3,5,6-Tetramethylpyrazine	1	81	1	82	1	82
2,5-Dihydroxybenzoic acid	1	27	1	59	1	47
2-Picolonic acid	1	79	1	70	1	77
4-Cyanopyridine	1	63	1	51	1	56
4-toluamide	0	94	0	97	0	97
6-Hydroxynicotinic acid	1	12	1	76	1	78
Acetamide	1	49	1	53	1	50
Acetic acid	1	43	1	38	1	28
Aconitic acid	1	29	1	40	1	22
Adipic acid	1	68	1	85	1	60
Ammonium alginate	1	39	1	13	1	1
Anisole	0	93	1	83	0	91
Anthranilamide	1	91	0	89	0	93
Aspartame	1	51	1	39	1	16
Aspartic acid	1	24	1	63	1	34
Aspirin	1	15	1	74	1	46
Benzamide	0	97	0	96	0	96
Benzoic acid	1	84	1	80	1	83
Betaine	1	58	1	34	1	41
Butylated hydroxytoluene	1	20	1	35	1	12
Caffeic acid	1	13	1	69	1	66
Caprolactam	0	92	1	55	1	68
Caprylic acid	1	83	1	71	1	72
Catechin	1	6	1	16	1	13
Catechol	1	72	1	44	1	84
Celecoxib	1	70	1	25	1	7

Chloral hydrate	1	64	1	22	1	11
Chlorobenzene	1	53	1	57	1	64
Chloroform	1	55	1	18	1	30
Cholic Acid	1	23	1	20	1	3
Cinnamic acid	1	76	1	86	1	76
Citric acid	1	33	1	24	1	14
Cyclohexane	1	86	1	37	1	54
d-10-Camphorsulfonic acid	1	48	1	31	1	9
D-Alanine	1	34	1	72	1	51
Dichloromethane	1	45	1	23	1	21
Dimethyl sulfoxide	1	59	1	32	1	33
DL-phenylalanine	1	90	0	95	1	87
Embonic acid	1	19	1	15	1	4
Ferulic acid	1	65	1	73	1	62
Fumaric acid	1	42	1	27	1	18
Gallic acid	1	1	1	12	1	49
Glutaric acid	1	46	1	67	1	48
Glycerol	1	38	1	11	1	39
Glycine	1	18	1	58	1	36
Glycol stearate	1	30	1	17	1	8
Glycolic acid	1	14	1	26	1	23
Hexamethylenetetramine	1	89	1	29	1	29
Hippuric acid	1	32	1	87	1	73
Isoniazid	1	87	0	93	0	94
Isonicotinamide	1	78	0	90	0	90
Lactic acid	1	56	1	79	1	57
L-Pyroglutamic acid	1	50	1	62	1	65
L-Tartaric acid	1	7	1	9	1	31
Maleic acid	1	41	1	28	1	19
Malic acid	1	16	1	30	1	26
Malonic acid	1	11	1	3	1	6
Mannitol	1	36	1	6	1	37
Methyl 4-hydroxybenzoate	1	88	0	92	0	89
MTBE	1	5	1	61	1	74
N,N-dimethylacetamide	1	21	1	81	1	81
Naringenin	1	2	1	8	1	15

n-Butyl acetate	1	82	1	78	1	85
Nicotinamide	1	85	0	94	0	92
Nicotinic acid	1	75	1	75	1	79
Orcinol	1	71	1	5	1	70
Oxalic acid	1	26	1	14	1	10
Para-aminobenzoic acid	0	96	0	88	0	88
Phosphoric acid	1	74	1	19	1	5
Picolinamide	0	95	0	91	0	95
Piperazine	1	61	1	49	1	58
Proline	1	47	1	41	1	67
Propanone	1	28	1	54	1	71
Propyl gallate	1	10	1	42	1	35
Propylene glycol	1	35	1	52	1	43
Propylparaben	1	52	1	77	1	80
Pterostilbene	1	67	1	48	1	25
Pyrazinamide	1	80	1	84	1	86
Pyrazine	1	31	1	47	1	55
Pyrogallol	1	3	1	2	1	75
Quercetin	1	9	1	10	1	2
Resorcinol	1	40	1	1	1	69
Saccharin	1	44	1	50	1	20
Salicylic acid	1	57	1	33	1	63
Sebacic acid	1	60	1	60	1	42
Serine	1	25	1	36	1	40
Sorbitol	1	37	1	7	1	38
Suberic acid	1	77	1	68	1	52
Succinic acid	1	17	1	45	1	27
Sulfur dioxide	1	73	1	21	1	17
Tert-butanol	1	4	1	56	1	59
Theobromine	1	54	1	66	1	32
Theophylline	1	69	1	65	1	44
Urea	1	66	1	43	1	24
Vanillic acid	1	62	1	64	1	61
Vitamin C	1	22	1	4	1	45

Table S5. Co-crystallization experiments attempted in this study.

Caffeine-coformer	Sample name	Condition	Result
Caffeine-Resorcinol	Caff-Res-H$_2$O	Caff. (50 mg, 0.26 mmol) and Res. (28.35 mg, 0.26 mmol) dissolved in 1 ml H$_2$O, stirred for 48 h	SC
	Caff-Res-DMF	Caff. (30 mg, 0.15 mmol) and Res. (17.01 mg, 0.15 mmol) dissolved in 1ml DMF, stirred for 48 h	N/A
	Caff-Res-DCM	Caff. (50 mg, 0.26 mmol) and Res. (28.35 mg, 0.26 mmol) dissolved in 2 ml DCM, stirred for 48 h	N/A
	Caff-2Res-DCM	Caff. (50 mg, 0.26 mmol) and Res. (56.70 mg, 0.51 mmol) dissolved in 2 ml DCM, stirred for 48 h	N/A
Caffeine-Pyrogallol	Caff-Py-H$_2$O	Caff. (25 mg, 0.13 mmol) and Py. (16.24 mg, 0.13 mmol) dissolved in 1 ml H$_2$O, stirred for 48 h	A
	Caff-Py-DMF	Caff. (30 mg, 0.15 mmol) and Py. (19.48 mg, 0.15 mmol) dissolved in 1 ml DMF, stirred for 48 h	A
	Caff-Py-DMF	Caff. (50 mg, 0.26 mmol) and Py. (32.47 mg, 0.26 mmol) dissolved in 2 ml DMF, stirred for 48 h	A
	Caff-Py-DCM	Caff. (50 mg, 0.26 mmol) and Py. (32.47 mg, 0.26 mmol) dissolved in 2 ml DCM, stirred for 48 h	N
	Caff-2Py-DCM	Caff. (50 mg, 0.26 mmol) and Py. (64.94 mg, 0.51 mmol) dissolved in 2 ml DCM, stirred for 48 h	N
Caffeine-Phosphoric acid	Caff-PA-H$_2$O	Caff. (100 mg, 0.51 mmol) and PA. (50.46 mg, 0.51 mmol) dissolved in 2 ml H$_2$O, stirred for 48 h	A
	Caff-PA-DMF	Caff. (88.32 mg, 0.45 mmol) and PA. (44.57 mg, 0.45 mmol) dissolved in 2 ml DMF, stirred for 48 h	A
	Caff-PA-DCM	Caff. (220 mg, 1.13 mmol) and PA. (111.01 mg, 1.13 mmol) dissolved in 2 ml DCM, stirred for 48 h	A
	Caff-2PA-H$_2$O	Caff. (100 mg, 0.51 mmol) and PA. (100.93 mg, 1.03 mmol) dissolved in 2 ml H$_2$O, stirred for 48 h	A
	Caff-2PA-DCM	Caff. (220 mg, 1.13 mmol) and PA. (222.02 mg, 2.27 mmol) dissolved in 2 ml DCM, stirred for 48 h	N
	Caff-2PA-DCE	Caff. (220 mg, 1.13 mmol) and PA. (222.02 mg, 2.27 mmol) dissolved in 2 ml DCE, stirred for 48 h	N
	Caff-2PA-TCM	Caff. (220 mg, 1.13 mmol) and PA. (222.02 mg, 2.27 mmol) dissolved in 2 ml TCM, stirred for 48 h	N

* A = amorphous solid, N = new crystalline phase, N/A = no solids formed, and SC = solids co-former isolated.

Table S6. Experimental logarithmic solubilities of caffeine in 13 solvents measured by HPLC.

Solvent	Log S (g/l)
Dichloromethane (DCM)	1.89
Dimethylacetamide (DMA)	1.46
Dimethylformamide (DMF)	1.44
Dimethyl sulfoxide (DMSO)	1.44
Acetonitrile (ACN)	1.35
Water	1.33
Acetone	0.97
Methanol	0.95
Toluene	0.72
Ethanol	0.61
2-Propanol	0.52
2-Butanone	0.32
Tetrahydrofuran (THF)	0.12

Table S7. Deviation of solubility predictions by three COSMO-RS implementations from experimental solubility measurements by HPLC presented as RMSD.

Implementation	Input type	Reference solvent	RMSD* (log S(g/l))
ADF COSMO-RS	Experimental fusion data	-	0.64
COSMOtherm	Experimental fusion data	-	0.51
COSMOtherm	Experimental reference solubility	water	0.50
COSMOquick	Experimental fusion data	-	0.57
COSMOquick	Experimental reference solubility	water, ethanol, acetone, and tetrahydrofuran	0.49

*The RMSD is calculated based on solubilities measured in g/l in a logarithmic scale.
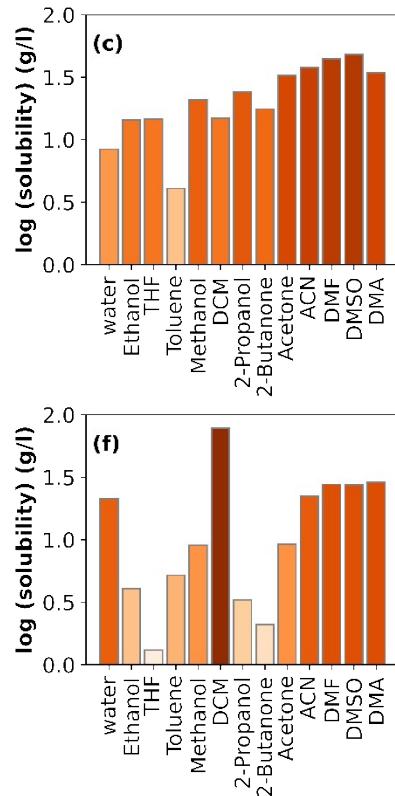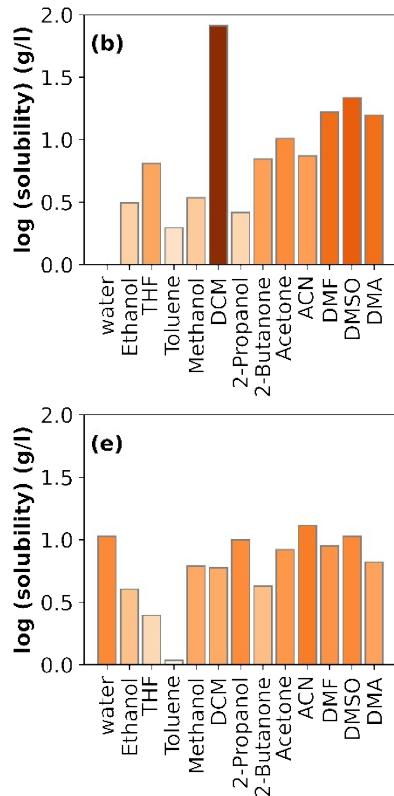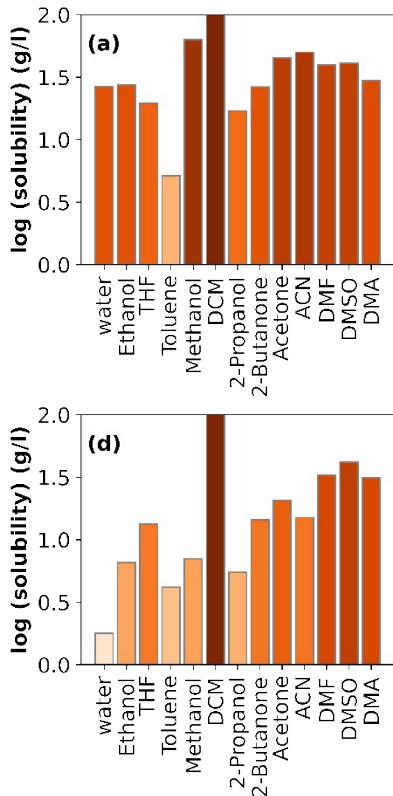
Figure S1. Caffeine solvents rankings by (a) ADF COSMO-RS (exp. fusion data), (b) COSMOtherm (exp. fusion data), (c) COSMOquick (exp. fusion data), (d) COSMOtherm (exp. sol. ref.), (e) COSMOquick (sol. ref.), and (f) experimental solubility by HPLC.
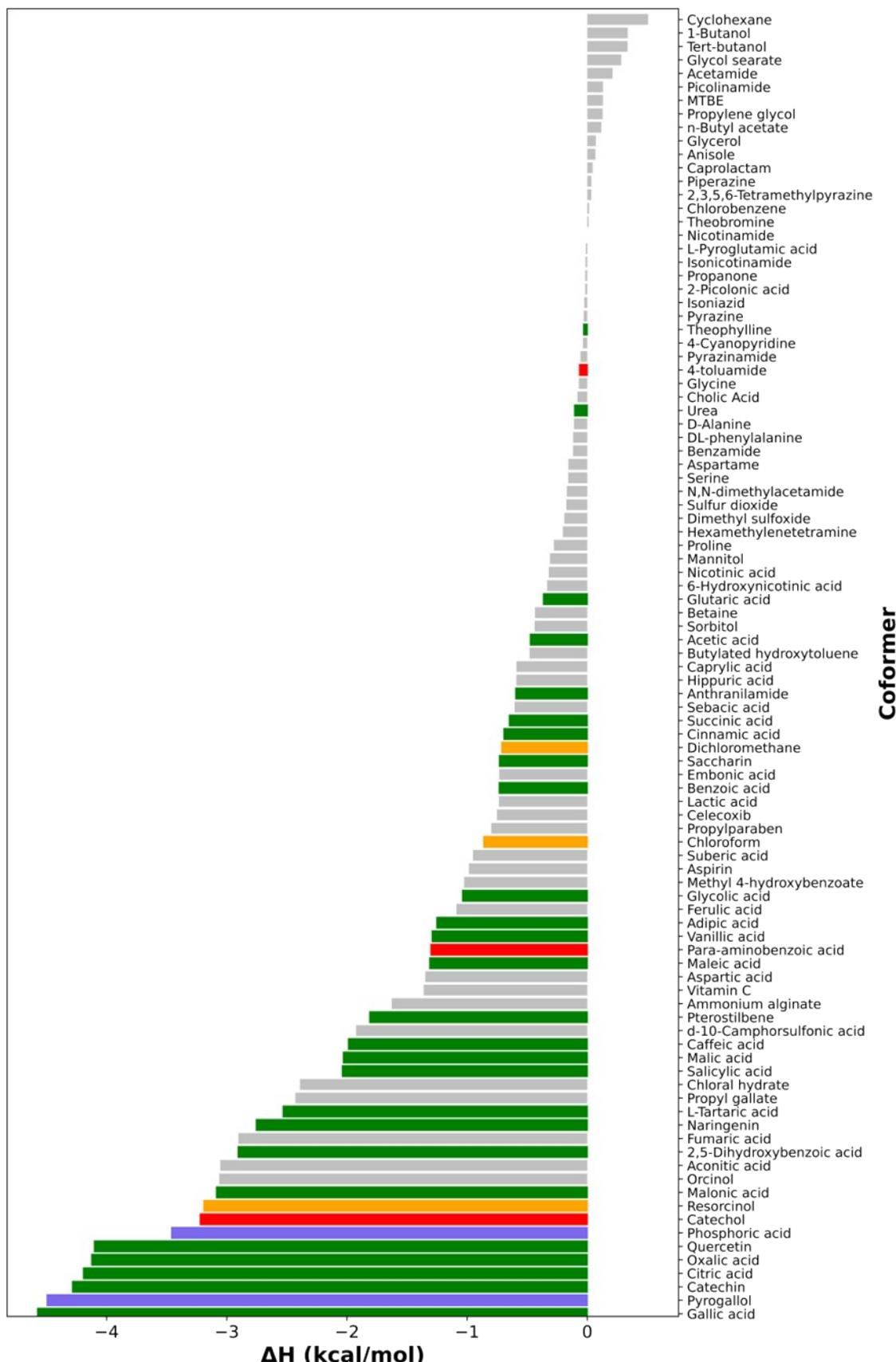
Figure S2. Caffeine co-crystal screening by COSMOtherm. Known caffeine coformers are shown in green and non-coformers are in red. Newly synthesized co-crystals in this study are highlighted in purple, while coformers that did not result in co-crystal formation are marked in orange.
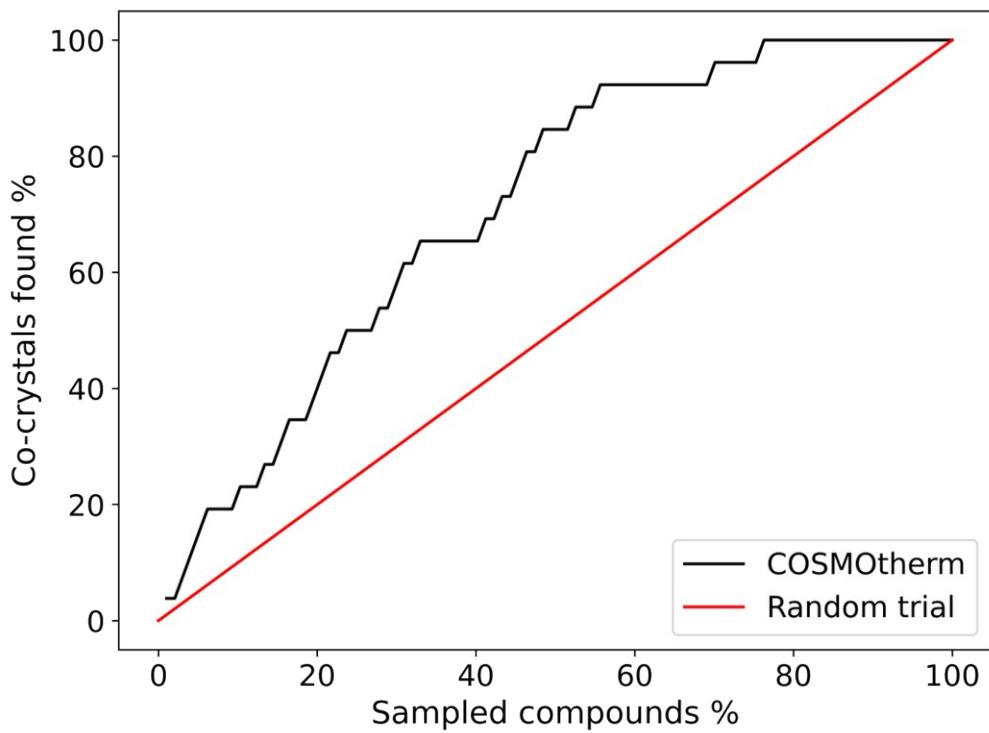
Figure S3. Enrichment plot of caffeine co-crystal screening by COSMOtherm. The dataset encompasses 26 co-crystal formers, three non-coformers, and 68 unknown coformers. (Non-coformers and unknown coformers were assigned a value of 0.)
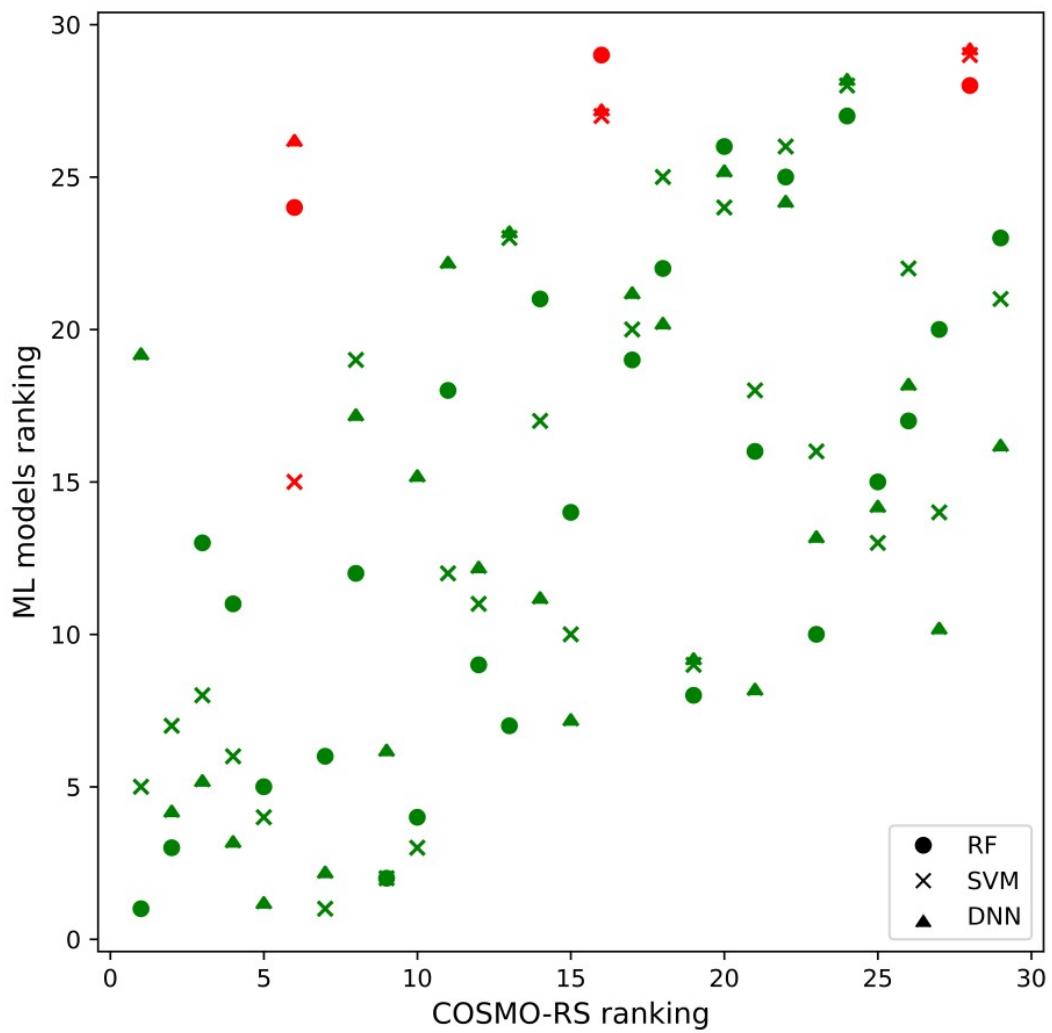
Figure S4. Ranking of 29 experimental data points within the independent test set, arranged according to probability estimates from the RF, SVM, and DNN models. Green symbols denote successful co-crystals, and red symbols represent unsuccessful coformers.
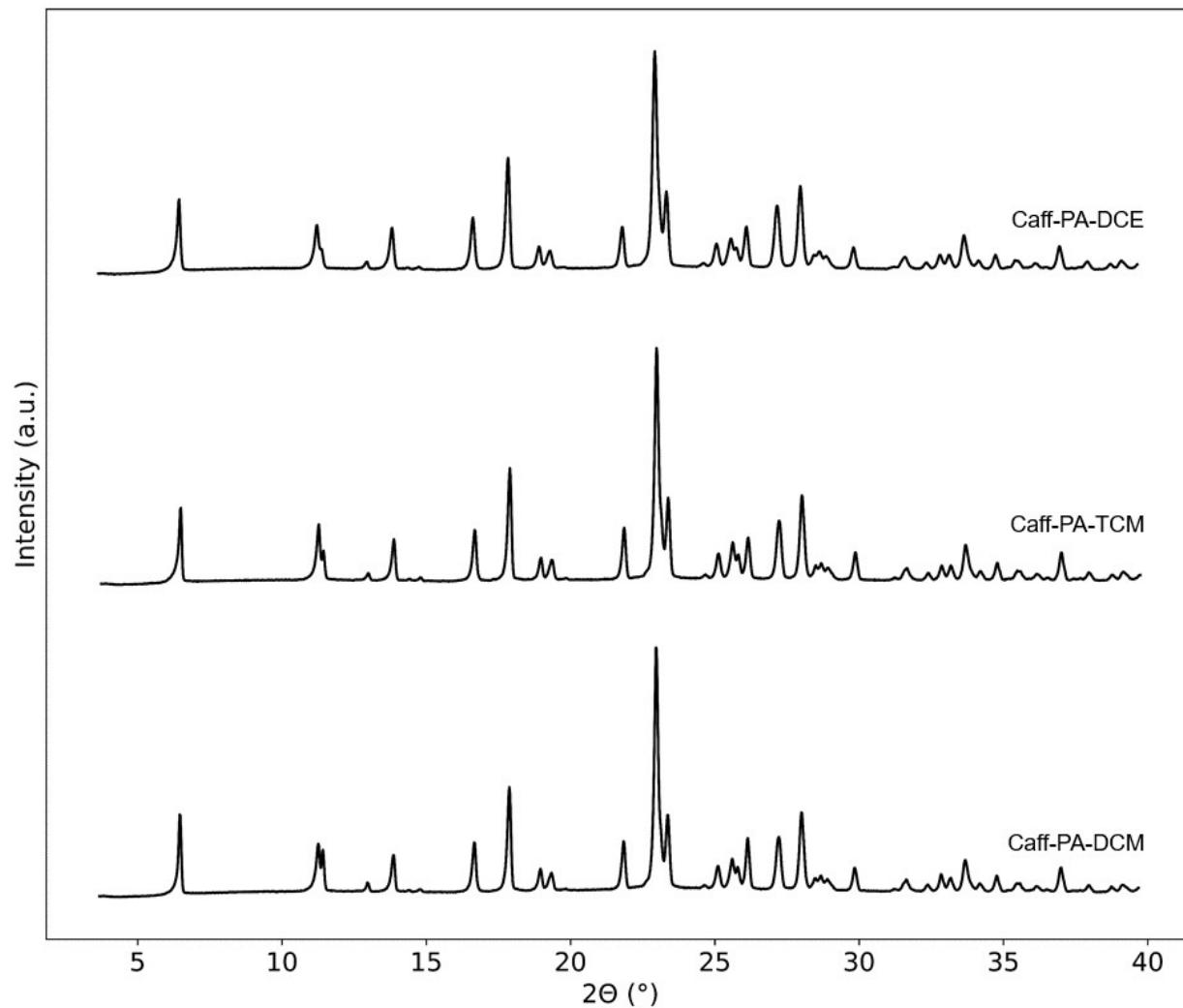
Figure S5. PXRD patterns of Caff-PA co-crystal from dichloroethane (DCE), chloroform (TCM), and dichloromethane (DCM).
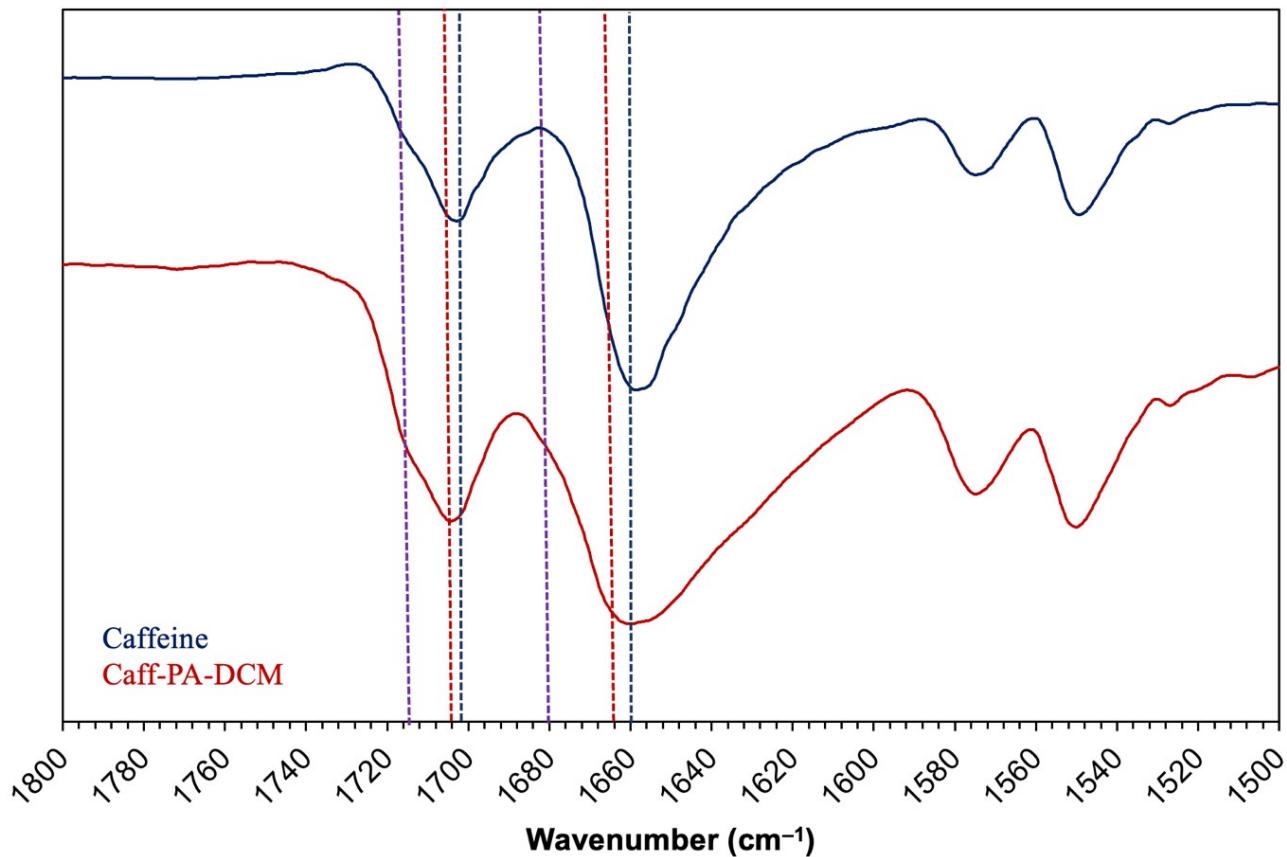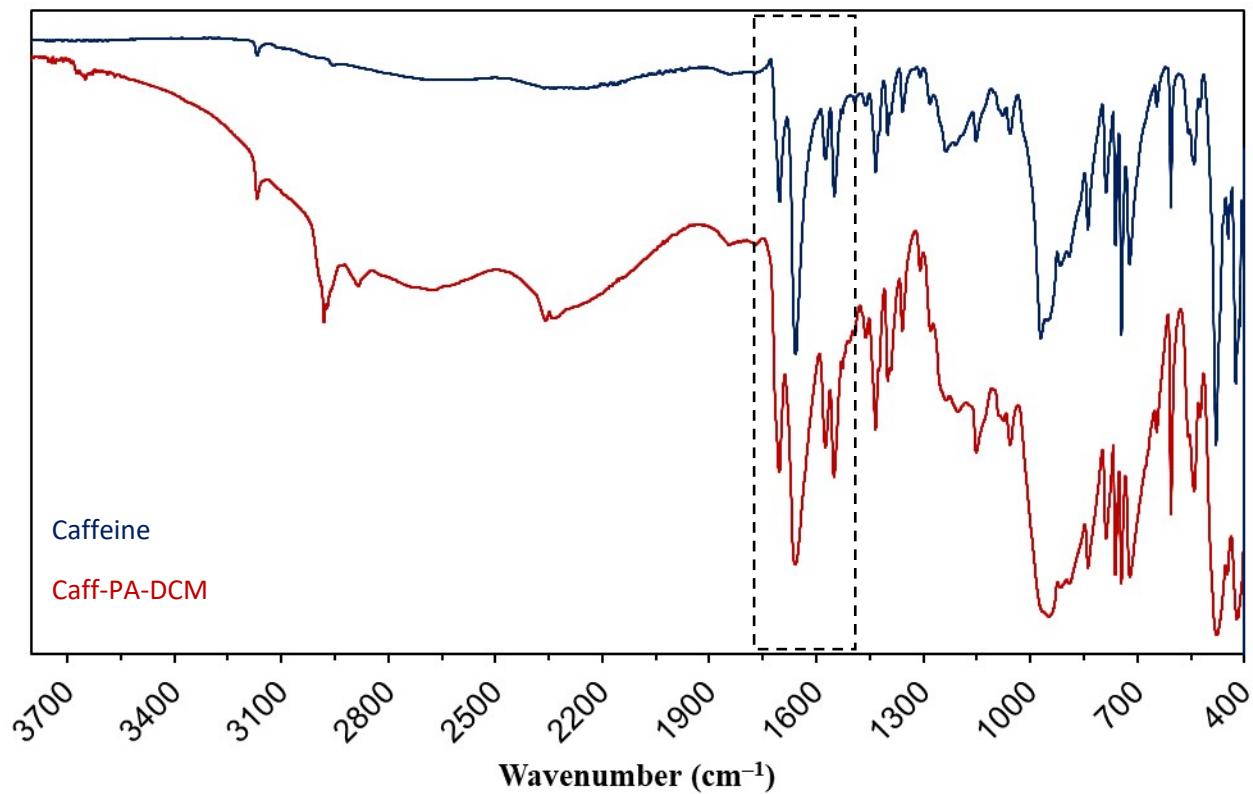
Figure S6. (Top) Di-ATR FTIR spectra of caffeine (blue) and Caff-PA-DCM (red), and (bottom) enlarged Di-ATR FTIR spectra of caffeine (blue) and Caff-PA-DCM (red) from 1500-1800 cm[-1]. The hashed lines represent the ν_{CO} shifts in caffeine (blue), Caff-PA-DCM (red) and the caffeine HCl salt (purple)[1].
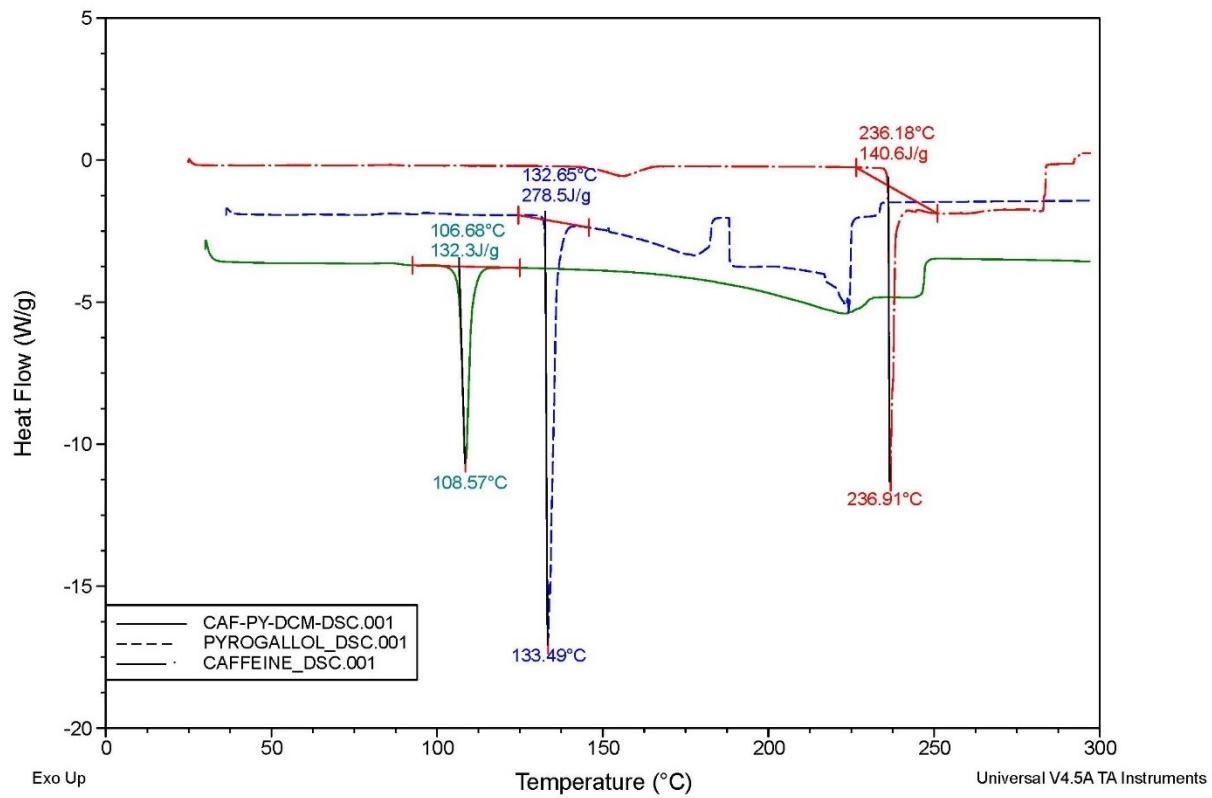
Figure S7. DSC thermogram (overlay) of caffeine, pyrogallol, and Caff-Py-DCM co-crystal.
Color scheme: caffeine (red), pyrogallol (blue), and Caff-Py co-crystal (green).
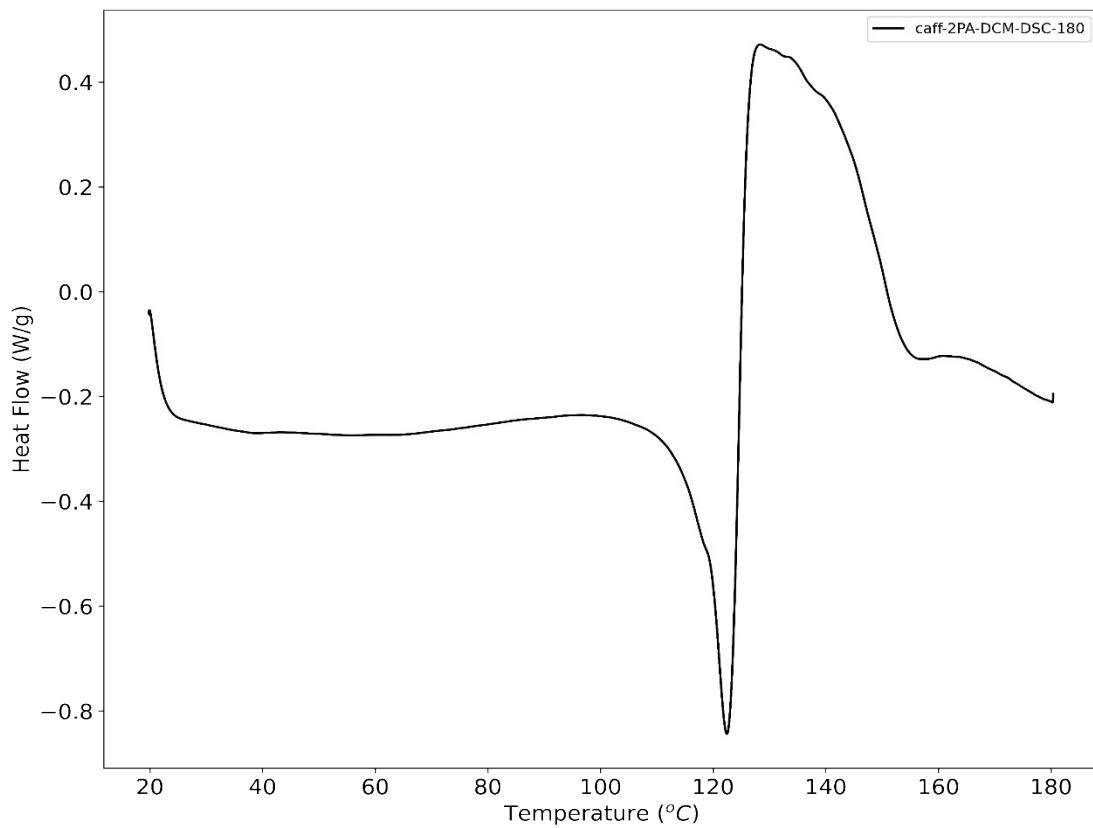
Figure S8. DSC thermogram of Caff-PA-DCM co-crystal.
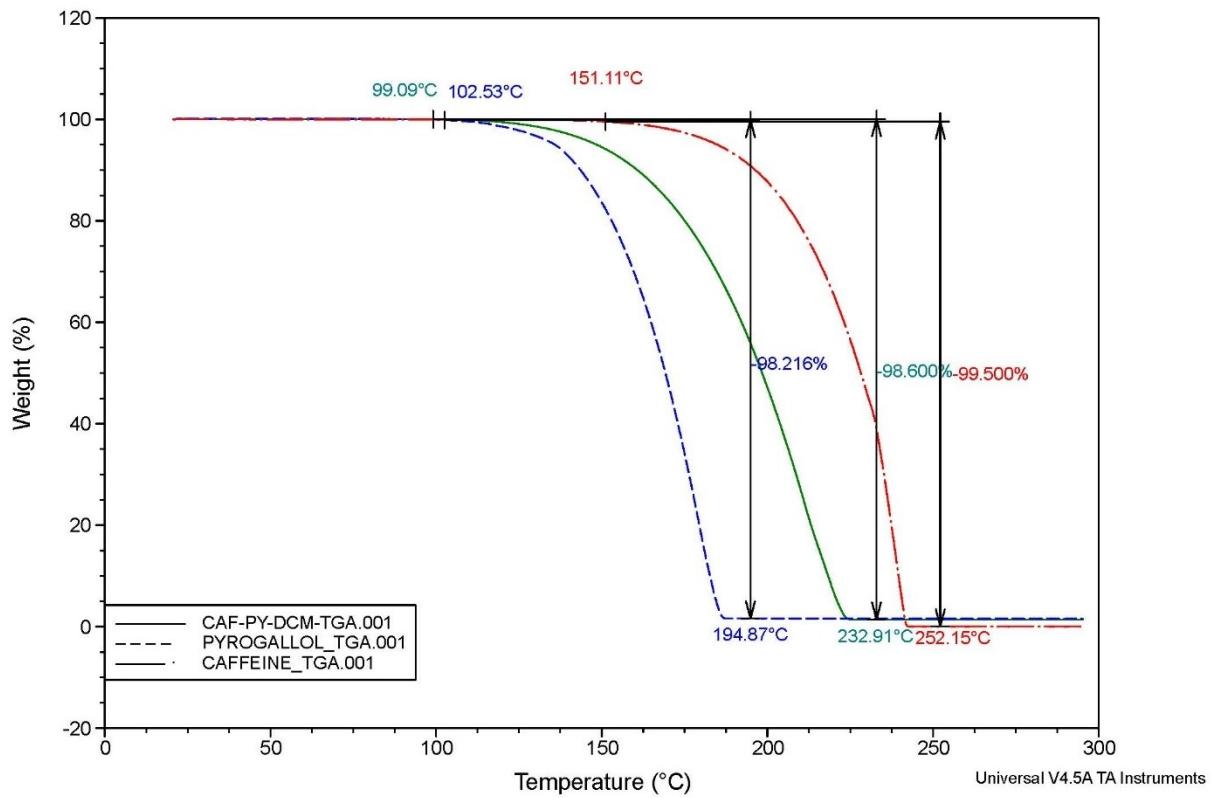
Figure S9. TGA (overlay) of caffeine, pyrogallol, and Caff-Py-DCM co-crystal.
Color scheme: caffeine (red), pyrogallol (blue), and Caff-Py co-crystal (green).
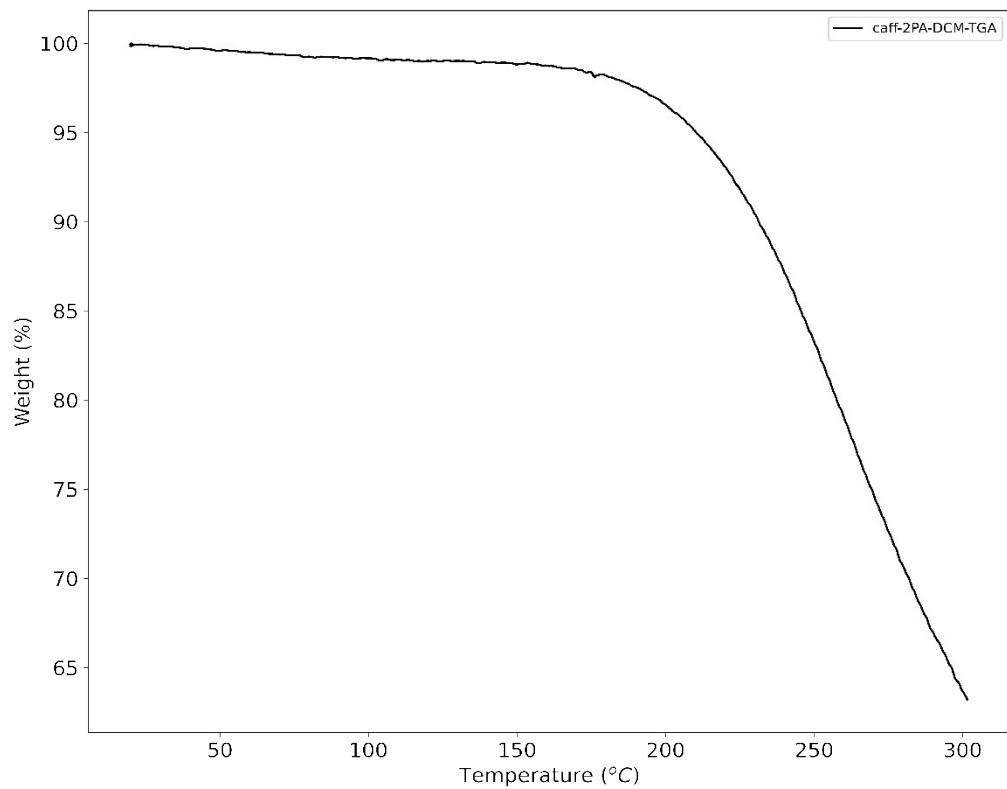
Figure S10. TGA of Caff-PA-DCM co-crystal.
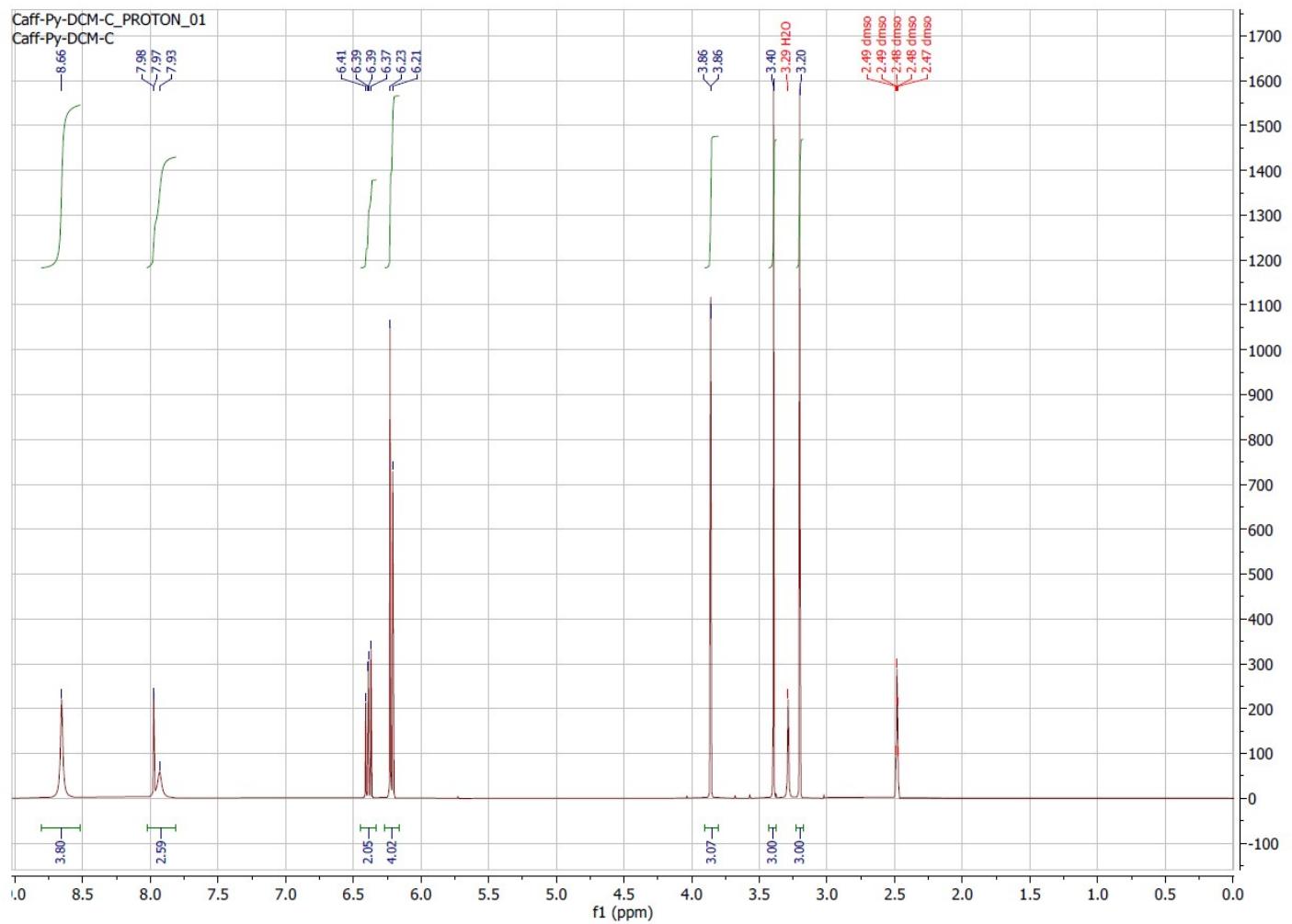
Figure S11. ¹H NMR spectrum of Caff-Py-DCM co-crystal collected in deuterated DMSO.
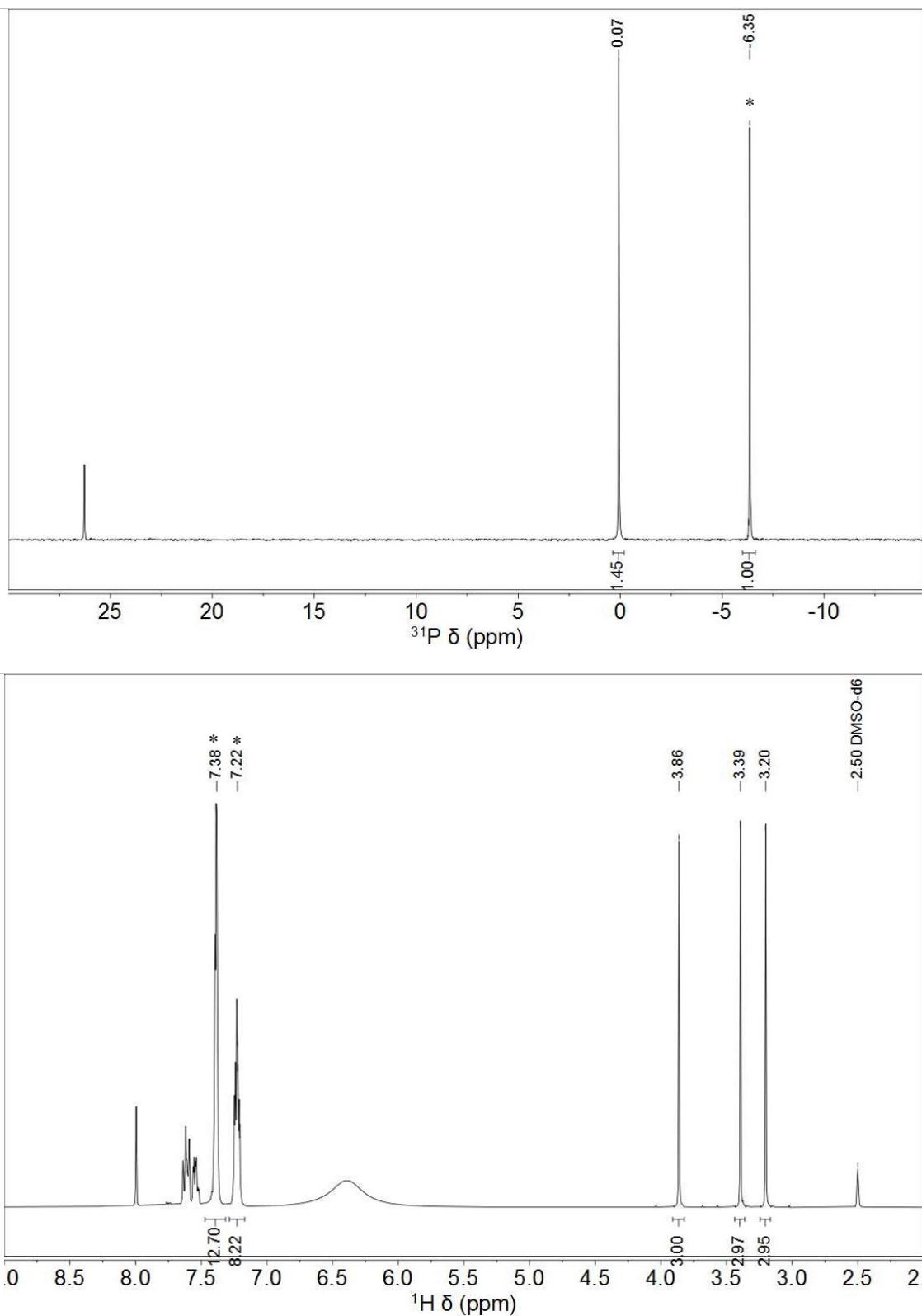
Figure S12. ^{31}P NMR (top) and ^{1}H NMR (bottom) spectra of Caff-PA-DCM co-crystal collected in deuterated DMSO with triphenylphosphine added as ^{1}H/^{31}P relative integration standard. The * marked peaks correspond to the internal integration standard, triphenylphosphine.
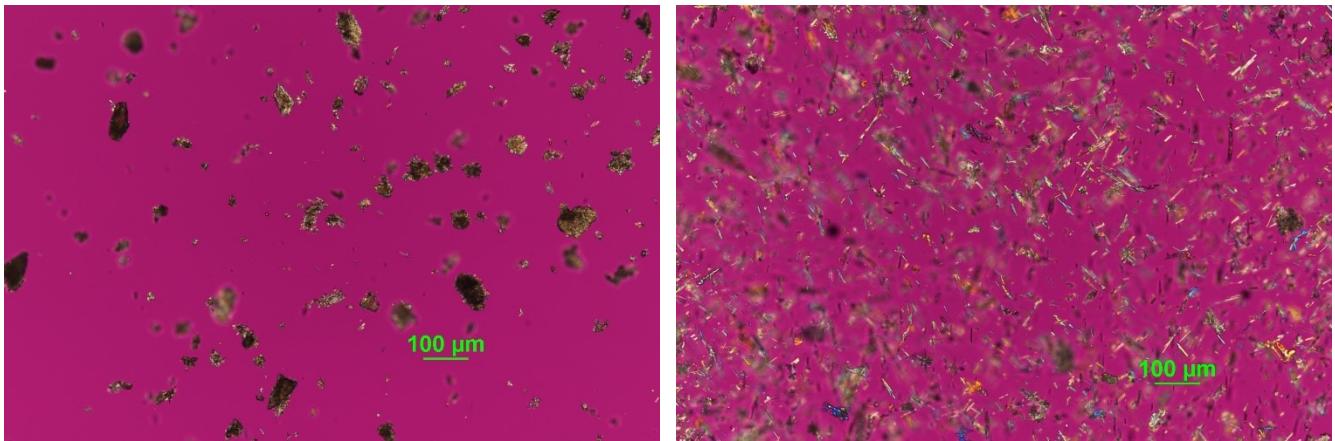
Figure S13. Polarized light microscopy of (a) caffeine and (b) Caff-Py-DCM co-crystal.
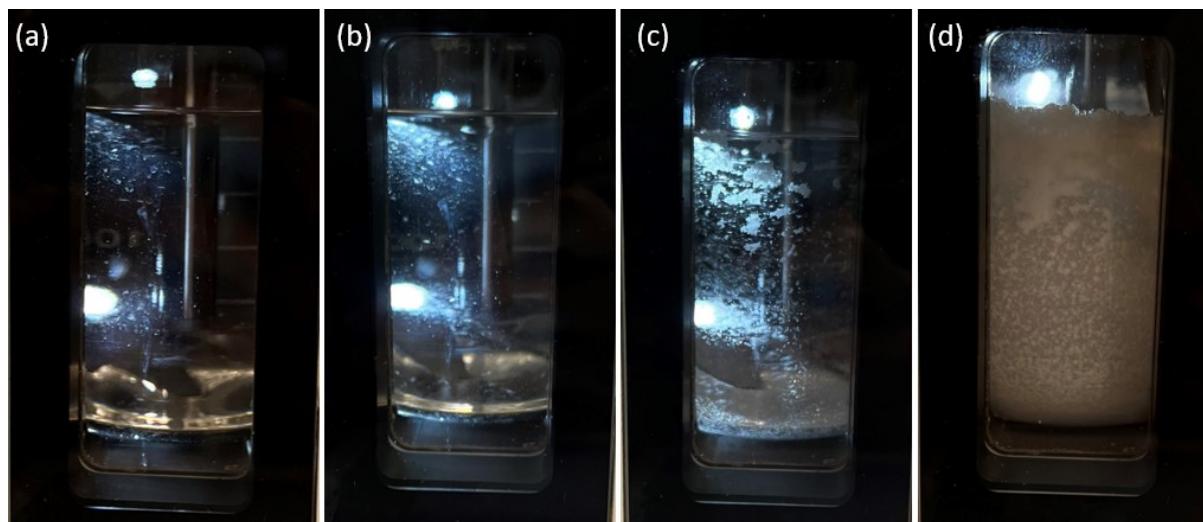

Figure S14. Visualization of Caff-Py-DCM co-crystal scale-up process by EasyMax402 at (a) t=0, (b) t= 30 min, (c) t= 120 min, and (d) t= 24 h.
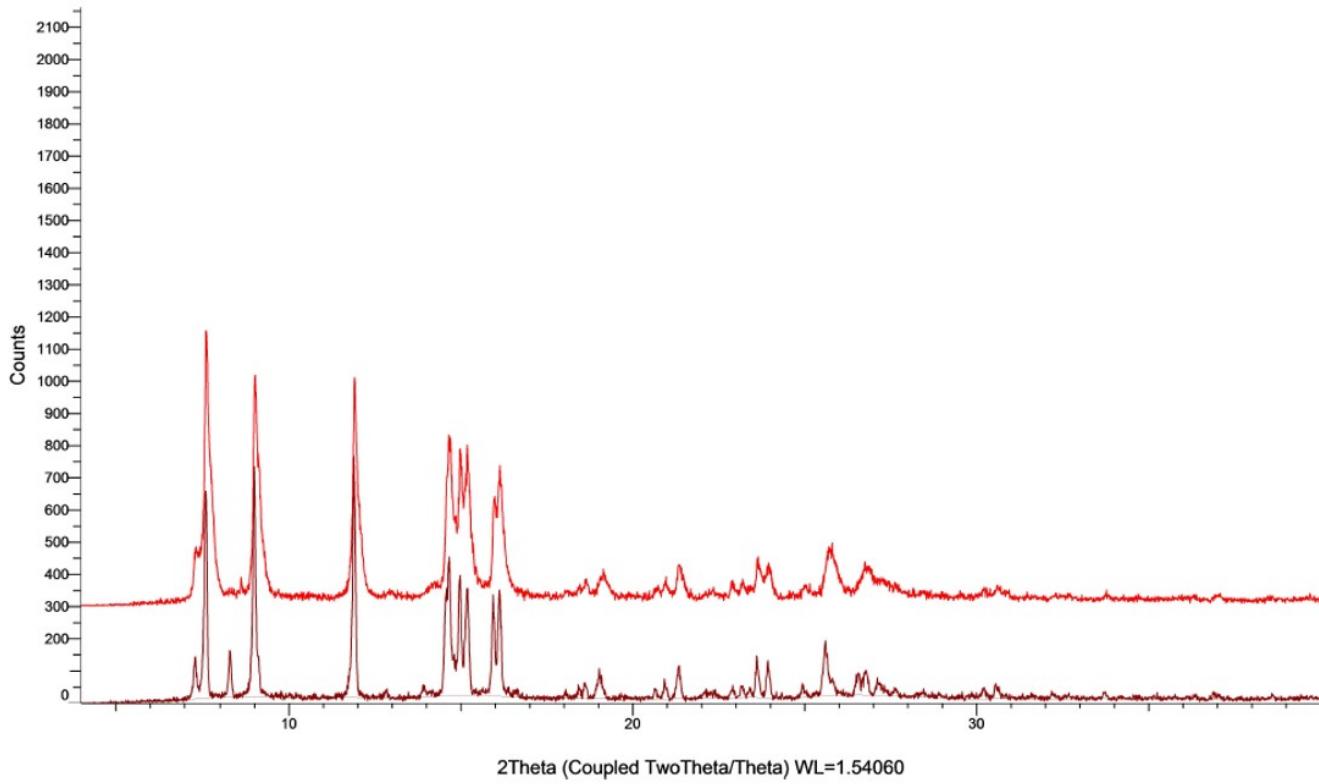
Figure S15. PXRD patterns of Caff-Py-DCM co-crystal (red) and Caff-Py-DCM scale-up (brown).
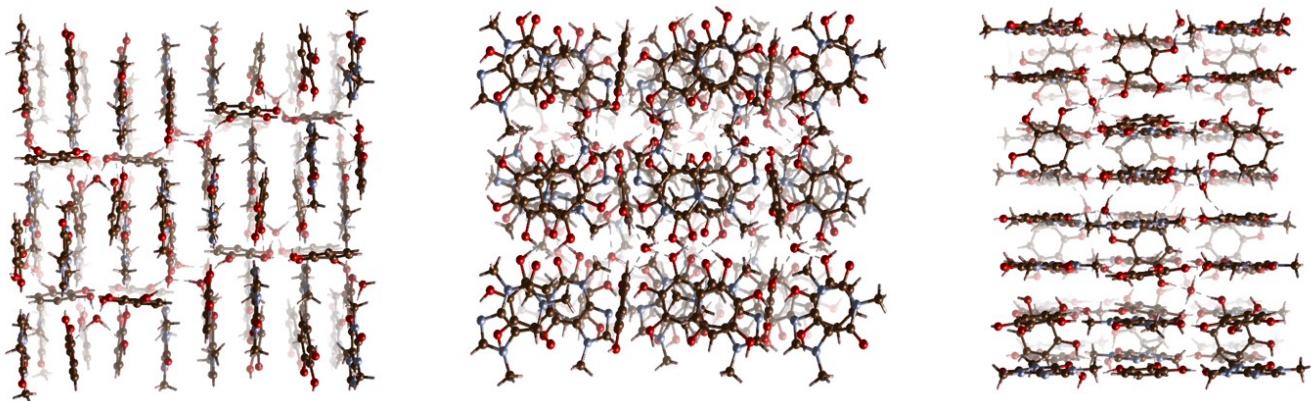
Figure S16. The crystal packing of Caff-Py co-crystal view along the *a*, *b*, and *c* axis, respectively. Color scheme: oxygen (red), carbon (brown), nitrogen (light blue), and hydrogen (white).
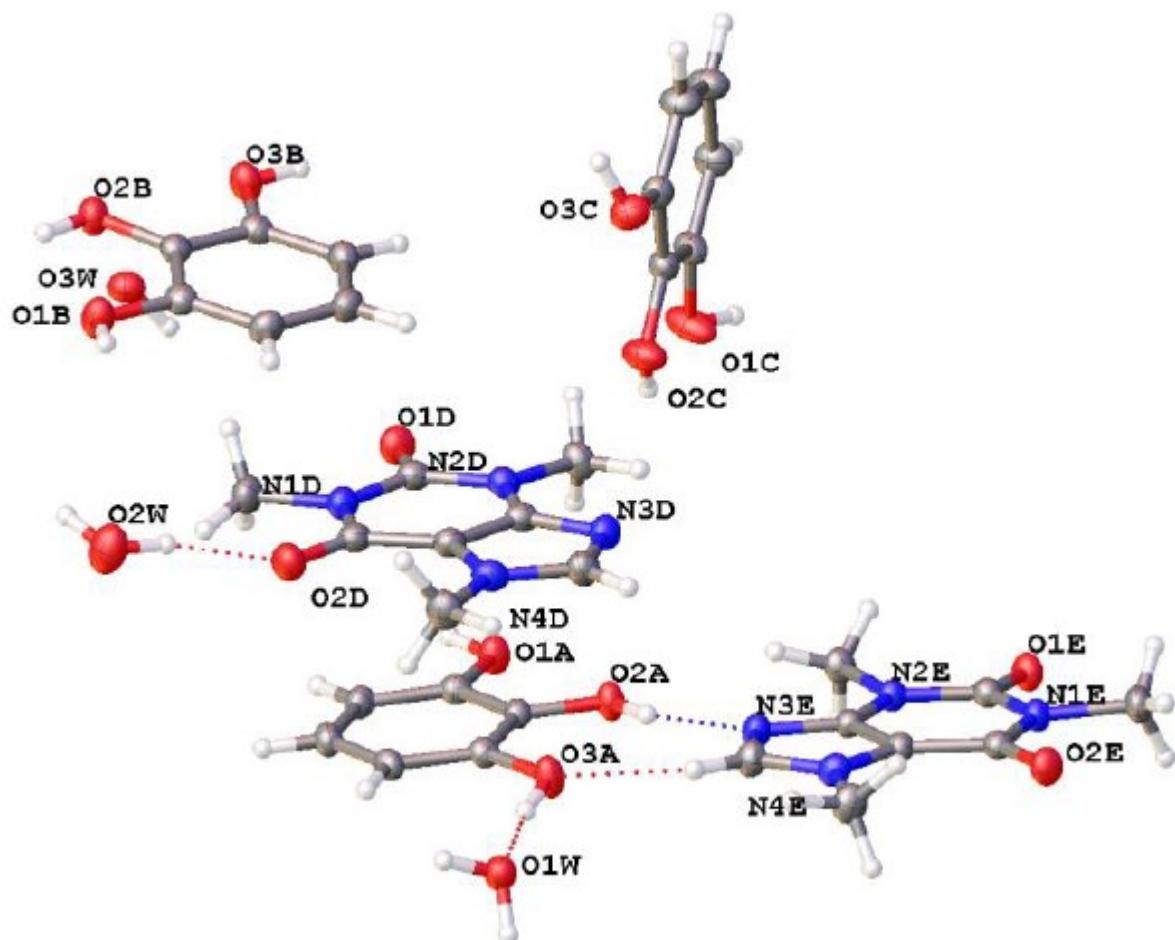

Figure S17. Asymmetric unit cell of Caff-Py co-crystal comprising two caffeine, three pyrogallol, and 2.5 water molecules. Color scheme: oxygen (red), carbon (gray), nitrogen (blue), and hydrogen (white).

Table S8. Crystallographic data for Caff-Py co-crystal.

Formula	$C_{34}H_{43}N_8O_{15.5}$
$D_{calc}.(g\ cm^{-3})$	1.464
μ/mm^{-1}	0.997
Formula weight	811.76
Colour	Colourless
Shape	Plate-shaped
Size(mm^3)	$0.14 \times 0.09 \times 0.02$
T(K)	100.00 (10)
Crystal System	Orthorhombic
Space Group	Pccn
a, b, c (Å)	13.97060 (16), 26.3924 (3), 19.9773(3)
α, β, γ (°)	90
V/Å^3	7365.95(15)
Z, Z'	8, 1
Wavelength (Å)	1.54184
Radiation type	Cu K$_\alpha$
$\Theta_{min}, \Theta_{max}$ (°)	3.349, 80.452
Measured Refl's.	41596
Indep't Refl's	7897
Refl's I> 2 σ (I)	6403
R_{int}	0.0475
Parameters	581
Restraints	0
Largest Peak	0.504
Deepest Hole	-0.262
GooF	1.033
wR$_2$ (all data)	0.1203
WR$_2$	0.1127
R$_1$ (all data)	0.0533
R$_1$	0.0433

TableS9. Hydrogen Bond information for Caff-Py co-crystal.

D	H	A	d(D-H)(Å)	d(H-A)(Å)	d(D-A)(Å)	D-H-A(°)
C3D	H3D	O1B[1]	0.95	2.45	3.005(2)	117.2
C3E	H3E	O3A	0.95	2.38	2.998(2)	122.4
O1A	H1A	O1E[2]	0.90(3)	1.80(3)	2.7029(16)	177(2)
O2A	H2A	N3E	0.90(3)	1.93(3)	2.7858(18)	160(3)
O3A	H3A	O1W	0.89(2)	1.82(2)	2.6909(17)	167(2)
O1B	H1B	O2W[3]	0.96(3)	1.66(3)	2.5989(18)	167(3)
O2B	H2B	N3D[4]	0.89(3)	2.02(3)	2.8881(19)	163(2)
O2B	H2B	O1B	0.89(3)	2.29(2)	2.7094(17)	108.3(19)
O3B	H3B	O1D[5]	0.87(3)	1.78(3)	2.6411(17)	176(3)
O1C	H1C	O3W1	0.90(3)	1.79(3)	2.6889(17)	172(3)
O2C	H2C	O1C	0.87(3)	2.11(3)	2.6409(18)	119(3)
O3C	H3C	O1W[3]	0.89(3)	1.90(3)	2.7554(18)	160(3)
O1W	H1WA	O2E[6]	0.90(3)	1.82(3)	2.7159(17)	172(2)
O1W	H1WB	O2A[7]	0.88(3)	2.03(3)	2.8516(17)	155(3)
O2W	H2WA	O2C[4]	0.93(3)	1.99(3)	2.903(2)	165(3)
O2W	H2WB	O2D	0.91(3)	1.82(3)	2.706(2)	164(3)
O3W	H3W	O3B[5]	0.92(3)	1.86(3)	2.7688(17)	175(3)

[1]+x,1/2-y,-1/2+z; [2]3/2-x,+y,1/2+z; [3]1/2-x,1/2-y,+z; [4]+x,1/2-y,1/2+z; [5]3/2-x,1/2-y,+z; [6]1/2-x,+y,1/2+z; [7]-1/2+x,1-y,1/2-z

In COSMO-RS, the estimation of molecular interactions involves analyzing the local contacts of COSMO surfaces and the polarization charge densities; σ and σ′ quantify the interactions between two surface segments of the COSMO surfaces[4]. The σ-profile is divided into three parts: H-bond donor (σ<-0.0084), hydrophobic (-0.0084<σ<+0.0084), H-bond acceptor (σ>+0.0084) regions. The hydrogen atom bound to the imidazole ring has an acidic character, therefore, caffeine acts as a donor in pair formation with proton-accepting molecules. While in the presence of proton-donating molecules, caffeine acts as a hydrogen bond acceptor compound through the basic center of the nitrogen atom in the imidazole ring[48, 49].
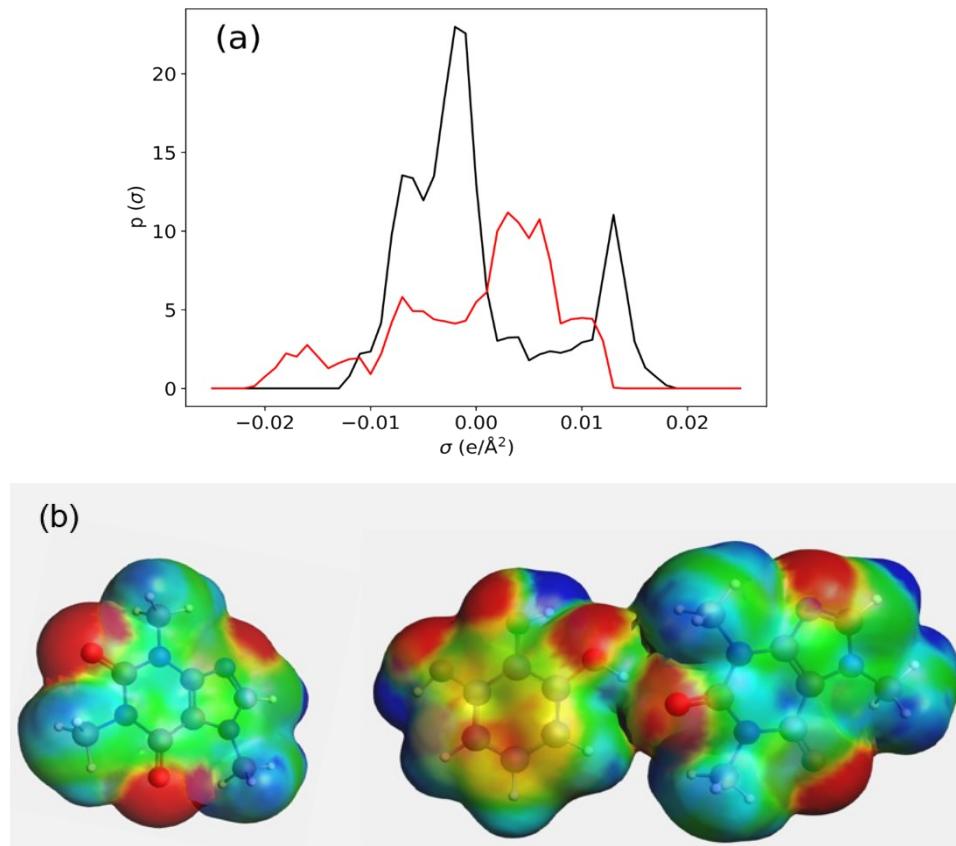

Figure S18. (a) σ-profile characteristics of caffeine (black) and pyrogallol (red) and (b) caffeine and caffeine-pyrogallol COSMO surfaces; the red part represents positive COSMO charge density (the underlying molecular charge is negative), and the blue part, negative COSMO charge density (the underlying molecular charge is

References:

(1) Cook, D.; Regnier, Z. R. The Infrared spectra of Caffeine salts. *Can. J. Chem.* **1967**, *45* (23), 2895-2897.
(2) Klamt, A. Conductor-like screening model for real solvents: a new approach to the quantitative calculation of solvation phenomena. *J. Phys. Chem.* **1995**, *99* (7), 2224-2235.
(3) Klamt, A.; Smith, B. J. Challenge of drug solubility prediction. *Methods Princ. Med. Chem.* **2008**.
(4) Loschen, C.; Klamt, A. Solubility prediction, solvate and cocrystal screening as tools for rational crystal engineering. *J. Pharm. Pharmacol.* **2015**, *67* (6), 803-811.
(5) Abramov, Y. A.; Loschen, C.; Klamt, A. Rational coformer or solvent selection for pharmaceutical cocrystallization or desolvation. *J. Pharm. Sci.* **2012**, *101* (10), 3687-3697.
(6) LeCun, Y.; Bengio, Y.; Hinton, G. Deep learning. *nature* **2015**, *521* (7553), 436-444.
(7) Bucar, D.-K.; Henry, R. F.; Lou, X.; Duerst, R. W.; MacGillivray, L. R.; Zhang, G. G. Cocrystals of caffeine and hydroxybenzoic acids composed of multiple supramolecular heterosynthons: Screening via solution-mediated phase transformation and structural characterization. *Cryst. Growth Des.* **2009**, *9* (4), 1932-1943.
(8) Wicker, J. G.; Crowley, L. M.; Robshaw, O.; Little, E. J.; Stokes, S. P.; Cooper, R. I.; Lawrence, S. E. Will they co-crystallize? *CrystEngComm* **2017**, *19* (36), 5336-5340.
(9) Trask, A. V.; van de Streek, J.; Motherwell, W. S.; Jones, W. Achieving polymorphic and stoichiometric diversity in cocrystal formation: Importance of solid-state grinding, powder X-ray structure determination, and seeding. *Cryst. Growth Des.* **2005**, *5* (6), 2233-2241.
(10) Bučar, D.-K.; Henry, R. F.; Lou, X.; Borchardt, T. B.; Zhang, G. G. A "hidden" co-crystal of caffeine and adipic acid. *Chem. Commun.* **2007**, (5), 525-527.
(11) Bučar, D.-K.; Day, G. M.; Halasz, I.; Zhang, G. G.; Sander, J. R.; Reid, D. G.; MacGillivray, L. R.; Duer, M. J.; Jones, W. The curious case of (caffeine)·(benzoic acid): how heteronuclear seeding allowed the formation of an elusive cocrystal. *Chem. Sci.* **2013**, *4* (12), 4417-4425.
(12) Satyanarayana, C.; Damarla, S.; Gorantla, S.; Muppidi, V. Novel caffeine co-crystals and their polymorphic forms. *WO* **2017**, *9813*, A1.
(13) Syed, T. A.; Ansari, K. B.; Banerjee, A.; Wood, D. A.; Khan, M. S.; Al Mesfer, M. K. Machine-learning predictions of caffeine co-crystal formation accompanying experimental and molecular validations. *J. Food Process Eng.* **2023**, *46* (2), e14230.
(14) Kumar, G. S.; Seethalakshmi, P.; Bhuvanesh, N.; Kumaresan, S. Studies on the syntheses, structural characterization, antimicrobial-, and DPPH radical scavenging activity of the cocrystals caffeine: cinnamic acid and caffeine: eosin dihydrate. *J. Mol. Struct.* **2013**, *1050*, 88-96.
(15) Karki, S.; Friščić, T.; Jones, W.; Motherwell, W. S. Screening for pharmaceutical cocrystal hydrates via neat and liquid-assisted grinding. *Mol. Pharmaceutics* **2007**, *4* (3), 347-354.
(16) Clarke, H. D.; Arora, K. K.; Bass, H.; Kavuru, P.; Ong, T. T.; Pujari, T.; Wojtas, L.; Zaworotko, M. J. Structure–stability relationships in cocrystal hydrates: Does the promiscuity of water make crystalline hydrates the nemesis of crystal engineering? *Cryst. Growth Des.* **2010**, *10* (5), 2152-2167.
(17) Trask, A. V.; Motherwell, W. S.; Jones, W. Solvent-drop grinding: green polymorph control of cocrystallisation. *Chem. Commun.* **2004**, (7), 890-891.
(18) Alvarez-Lorenzo, C.; Castiñeiras, A.; Frontera, A.; García-Santos, I.; González-Pérez, J.; Niclós-Gutiérrez, J.; Rodríguez-González, I.; Vílchez-Rodríguez, E.; Zaręba, J. Recurrent motifs in pharmaceutical cocrystals involving glycolic acid: X-ray characterization, Hirshfeld surface analysis and DFT calculations. *CrystEngComm* **2020**, *22* (40), 6674-6689.

(19) Friščić, T.; Fábián, L.; Burley, J. C.; Jones, W.; Motherwell, W. S. Exploring cocrystal–cocrystal reactivity via liquid-assisted grinding: the assembling of racemic and dismantling of enantiomeric cocrystals. *Chem. Commun.* **2006**, (48), 5009-5011.

(20) Trask, A. V.; Motherwell, W. S.; Jones, W. Pharmaceutical cocrystallization: engineering a remedy for caffeine hydration. *Cryst. Growth Des.* **2005**, *5* (3), 1013-1021.

(21) Leyssens, T.; Tumanova, N.; Robeyns, K.; Candoni, N.; Veesler, S. Solution cocrystallization, an effective tool to explore the variety of cocrystal systems: caffeine/dicarboxylic acid cocrystals. *CrystEngComm* **2014**, *16* (41), 9603-9611.

(22) Cui, W.; He, Z.; Zhang, Y.; Fan, Q.; Feng, N. Naringenin cocrystals prepared by solution crystallization method for improving bioavailability and anti-hyperlipidemia effects. *AAPS PharmSciTech* **2019**, *20* (3), 1-12.

(23) Schultheiss, N.; Bethune, S.; Henck, J.-O. Nutraceutical cocrystals: utilizing pterostilbene as a cocrystal former. *CrystEngComm* **2010**, *12* (8), 2436-2442.

(24) Smith, A. J.; Kavuru, P.; Wojtas, L.; Zaworotko, M. J.; Shytle, R. D. Cocrystals of quercetin with improved solubility and oral bioavailability. *Mol. Pharmaceutics* **2011**, *8* (5), 1867-1876.

(25) Padrela, L.; Rodrigues, M. A.; Velaga, S. P.; Fernandes, A. C.; Matos, H. A.; de Azevedo, E. G. Screening for pharmaceutical cocrystals using the supercritical fluid enhanced atomization process. *J. Supercrit. Fluids* **2010**, *53* (1-3), 156-164.

(26) Friscic, T.; Trask, A. V.; Motherwell, W.; Jones, W. Guest-directed assembly of caffeine and succinic acid into topologically different heteromolecular host networks upon grinding. *Cryst. Growth Des.* **2008**, *8* (5), 1605-1609.

(27) Eddleston, M. D.; Patel, B.; Day, G. M.; Jones, W. Cocrystallization by freeze-drying: preparation of novel multicomponent crystal forms. *Cryst. Growth Des.* **2013**, *13* (10), 4599-4606.

(28) MacFhionnghaile, P.; Crowley, C. M.; McArdle, P.; Erxleben, A. Spontaneous solid-state cocrystallization of caffeine and urea. *Cryst. Growth Des.* **2020**, *20* (2), 736-745.

(29) Jacobs, A.; Noa, F. M. A. Co-crystals and co-crystal hydrates of vanillic acid. *CrystEngComm* **2015**, *17* (1), 98-106.

(30) Bučar, D.-K.; Henry, R. F.; Duerst, R. W.; Lou, X.; MacGillivray, L. R.; Zhang, G. G. A 1: 1 cocrystal of caffeine and 2-hydroxy-1-naphthoic acid obtained via a slurry screening method. *J. Chem. Crystallogr.* **2010**, *40* (11), 933-939.

(31) Kumar, G. S.; Seethalakshmi, P.; Sumathi, D.; Bhuvanesh, N.; Kumaresan, S. Syntheses, structural characterization, and DPPH radical scavenging activity of cocrystals of caffeine with 1-and 2-naphthoxyacetic acids. *J. Mol. Struct.* **2013**, *1035*, 476-482.

(32) Ghosh, S.; Reddy, C. M. Co-crystals of caffeine with substituted nitroanilines and nitrobenzoic acids: Structure–mechanical property and thermal studies. *CrystEngComm* **2012**, *14* (7), 2444-2453.

(33) Shefter, E. Structural studies on complexes II. Crystal and molecular structure of a 1: 1 caffeine and 5-chlorosalicyclic acid complex. *J. Pharm. Sci.* **1968**, *57* (7), 1163-1168.

(34) Zaworotko, M. J.; Clarke, H.; Kapildev, A.; Kavuru, P.; Shytle, R. D.; Pujari, T.; Marshall, L.; Ong, T. T. Nutraceutical co-crystal compositions. **2019**.

(35) Mukherjee, S. *Crystal engineering of pharmaceutical cocrystals*; University of South Florida, 2011.

(36) Todaro, V.; Worku, Z. A.; Cabral, L. M.; Healy, A. M. In situ cocrystallization of dapsone and caffeine during fluidized bed granulation processing. *AAPS PharmSciTech* **2019**, *20*, 1-13.

(37) Craven, B.; Gartland, G. The 2: 1 crystal complex of 5, 5-diethylbarbituric acid (barbital) and caffeine. *Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry* **1974**, *30* (5), 1191-1195.

(38) Goud, N. R.; Gangavaram, S.; Suresh, K.; Pal, S.; Manjunatha, S. G.; Nambiar, S.; Nangia, A. Novel furosemide cocrystals and selection of high solubility drug forms. *J. Pharm. Sci.* **2012**, *101* (2), 664-680.

(39) Chadha, K.; Karan, M.; Bhalla, Y.; Chadha, R.; Khullar, S.; Mandal, S.; Vasisht, K. Cocrystals of hesperetin: structural, pharmacokinetic, and pharmacodynamic evaluation. *Cryst. Growth Des.* **2017**, *17* (5), 2386-2405.

(40) Lekšić, E.; Pavlović, G.; Mestrovic, E. Cocrystals of lamotrigine based on coformers involving carbonyl group discovered by hot-stage microscopy and DSC screening. *Cryst. Growth Des.* **2012**, *12* (4), 1847-1858.

(41) Sun, C. C.; Hou, H. Improving mechanical properties of caffeine and methyl gallate crystals by cocrystallization. *Cryst. Growth Des.* **2008**, *8* (5), 1575-1579.

(42) Sanphui, P.; Kumar, S. S.; Nangia, A. Pharmaceutical cocrystals of niclosamide. *Cryst. Growth Des.* **2012**, *12* (9), 4588-4599.

(43) Kumar, G. S.; Seethalakshmi, P.; Bhuvanesh, N.; Kumaresan, S. Cocrystals of caffeine with formylphenoxyaliphatic acids: Syntheses, structural characterization, and biological activity. *J. Mol. Struct.* **2013**, *1034*, 302-309.

(44) Schultheiss, N.; Roe, M.; Boerrigter, S. X. Cocrystals of nutraceutical p-coumaric acid with caffeine and theophylline: polymorphism and solid-state stability explored in detail using their crystal graphs. *CrystEngComm* **2011**, *13* (2), 611-619.

(45) Maninder, K.; Kunal, C.; Renu, C.; Anupam, S. Cocrystal of caffeine with propionic acid: preliminary characterization and stability evaluation. *Journal of Pharmacy Research* **2012**, *5* (4), 2022-2026.

(46) Leger, J.-M.; Alberola, S.; Carpy, A. Structure cristalline du complexe 1: 1 sulfacétamide–caféine. *Acta Crystallographica Section B: Structural Crystallography and Crystal Chemistry* **1977**, *33* (5), 1455-1459.

(47) Ghosh, M.; Basak, A.; Mazumdar, S.; Sheldrick, B. Structure and conformation of the 1: 1 molecular complex sulfaproxyline–caffeine. *Acta Crystallographica Section C: Crystal Structure Communications* **1991**, *47* (3), 577-580.

(48) Jeliński, T.; Kubsik, M.; Cysewski, P. Application of the Solute–Solvent Intermolecular Interactions as Indicator of Caffeine Solubility in Aqueous Binary Aprotic and Proton Acceptor Solvents: Measurements and Quantum Chemistry Computations. *Materials* **2022**, *15* (7), 2472.

(49) Jeliński, T.; Cysewski, P. Quantification of Caffeine Interactions in Choline Chloride Natural Deep Eutectic Solvents: Solubility Measurements and COSMO-RS-DARE Interpretation. *Int. J. Mol. Sci.* **2022**, *23* (14), 7832.