CRYSTAL GROWTH & DESIGN

Article

Analysis and ANN Prediction of Melting Properties and Ideal Mole fraction Solubility of Co-crystals

Rama Krishna Gamidi, and Åke C. Rasmuson

Cryst. Growth Des., **Just Accepted Manuscript** • DOI: 10.1021/acs.cgd.0c00182 • Publication Date (Web): 14 Jul 2020

is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Analysis and ANN Prediction of Melting Properties and Ideal Mole fraction Solubility of Co-crystals

Rama Krishna Gamidi [ab] and Åke. C. Rasmuson*[a]

Rama Krishna [gr_krishna2005@yahoo.co.in]

a Department of Chemical and Environmental Science, Synthesis and Solid State Pharmaceutical Centre, Bernal Institute, University of Limerick, Limerick, Ireland

b Organic Chemistry Division, CSIR-National Chemical Laboratory, Dr. Homi Bhabha Road, Pune, Maharashtra 411 008, India

Abstract: Different Artificial Neural Network models have been developed and examined for prediction of cocrystal properties based on pure component physical properties only. From the molecular weight, melting temperature, melting enthalpy and melting entropy of the pure compounds, the corresponding melting properties of the cocrystals and the cocrystal ideal solubility have been successfully predicted. Notably, no information, whatsoever about the cocrystals are needed, besides the identification of the two compounds from which the cocrystal is formed. In total, thirty co-crystal systems of eight different model components, namely, Theophylline, Piracetam, Gabapentin-lactam, Tegafur, Nicotinamide, Salicylic acid, Syringic acid and 4,4'-Bipyridine with distinct coformer's has been chosen as the model system's for the construction of ANN models. In all the cases, 70% of the data points has been used to train the model and the rest were used to test the capability of the model (as a validation set) as selected through a random selection process. The training process was stopped with overall r^2 values above 0.986. In particular, the models capture how the coformer structure influences on the targeted physical properties of cocrystals.

Key points: Melting enthalpy, Melting entropy, Melting point, Mole fraction solubility, Co-crystals, Artificial Neural Networks, Predictive models

Introduction

Co-crystals (CCs) are crystalline molecular complexes comprising two or more neutral components that in pure form are solid at room temperature. The components appear in the co-crystal in a specific stoichiometry, bonded by non-covalent interactions, especially hydrogen bonding. Co-crystals are of considerable interest to the pharmaceutical industry because of their ability to modify the physical properties of the active pharmaceutical ingredient (API), without affecting the molecular structure.[1-4] Co-crystal formation has indeed proved to be a useful tool to alter a wide range of properties including melting point, hygroscopicity, dissolution rate, thermal stability, solubility and hence bioavailability.[5-20] For a particular API a wide variety of molecules can be considered as coformers in the formation of a cocrystal, based on their hydrogen bonding functionalities. If the physical properties of the cocrystal could be predicted without actually having to manufacture the cocrystal, that would make the search for suitable coformers much more rapid and efficient.

Among the physical properties of primary interest in characterizing CCs are the melting properties. A higher melting point of the CC compared to the parent components demonstrates a higher thermal stability. The thermal stability of a particular API can often be improved by cocrystallization using coformers having a higher melting point and vice-versa.[21, 22] The melting temperature could also help us to explain the purity of the solid substances and also improve the possibility to set the chemical stability/degradation parameters at various humidity conditions based on its melting temperature.[23, 24]

The melting enthalpy ($\Delta H^{f}(T\mathrm{m})$) accounts for the difference in molecular binding and conformations between the solid and the molten state. The melting enthalpy is an important parameter to explain stability of the co-crystals[1] with respect to parent components, and in many cases the formation of CCs is enthalpy driven, *i.e*. the enthalpy change upon formation from the solid pure compounds is negative.[1] Accordingly, in virtual CC screening methods,[25, 26] the melting enthalpy is used as an indicator of stability of the CCs. Most recently, the ability to formation of CCs were explained on the basis of lattice energy calculations,[2, 3] and in most of the cases, CCs predicted to have lower lattice energy by a few kcal/mol than the corresponding individual components. Moreover, $\Delta H^{f}(T\mathrm{m})$ is useful to set the processing parameters in the pharmaceutical industry, especially, when dealing with thermodynamic processes and energetic transformations of the materials at constant pressure.[4] Such as the conditions for hot-melt extrusion to avoid chemical degradation,[5, 6] advancement of chemical reactions between the drug molecules and excipients,[7-10] and for conversion of pharmaceutical materials into the granules.[11, 12] In the field of crystallization, the $\Delta H^{f}(T\mathrm{m})$ is useful to determine crystalline percentage,[13] purity and thermal stability of the material.[14-16]

In addition to the $\Delta H^{f}(T\mathrm{m})$, the melting entropy is important for the understanding of the underlying reasons for CCs stability. Since, the true stability is of course given by the free energy. Recently, Si-Wei Zhang *et*.al[27] have found celecoxib-nicotinamide CC, and demonstrated that entropy controls the formation process of CCs and the CC is stable only above a transition temperature. Herein, the entropy increase is suggested to be due to volume expansion upon co-crystallization. The molecules form a tighter hydrogen bonding network, and as a result, hydrophobic functional groups will pack more loosely, which cause more rotational and vibrational freedom.[27] Also, the formation of the sulfamethazine–salicylic acid co-crystal appears to be entropy driven since the enthalpy actually increases quite substantially.[28]

The solid-liquid solubility in different solvents is the basis for the design and the operation of the crystallization process in which the CC is purified and given in its physical form. The solubility in gastrointestinal fluids is an important factor for the bioavailability of the drug. The solubility depends on the stability of the solid phase and on the solute-solvent interactions in the solution. In the standard chemical engineering treatment, the stability of the solid phase uses the pure compound melt as the reference state, and is thus represented by the Gibbs free energy of fusion at the temperature of interest. A thermodynamic analysis show that the activity of the solid can be determined according to eq. 1,[29] where $\Delta H^{f}(T\mathrm{m})$ is the melting enthalpy at the normal melting point, $T\mathrm{m}$, is the normal melting temperature and both can often be determined by Differential Scanning Calorimetry (DSC). However, with respect to CCs of course such determination requires that the cocrystal have been actually manufactured. ΔC_p is the heat capacity difference between the pure solute melt and the corresponding solid form.

$$\ln a^{s} = \frac{\Delta_{\mathrm{fus}} H\left(T_{\mathrm{m}}\right)}{R}\left(\frac{1}{T_{\mathrm{m}}} - \frac{1}{T}\right) + \frac{1}{R}\int_{T_{\mathrm{m}}}^{T} \frac{\Delta C_p}{T}\,\mathrm{d}T - \frac{1}{RT}\int_{T_{\mathrm{m}}}^{T} \Delta C_p\,\mathrm{d}T \qquad (1)$$

At equilibrium the activity of the solid, a^{s}, equals the activity of the solute in the solution. If a Raoults law activity coefficient γ is defined we can write:

$$x\gamma = a^{s} \qquad (2)$$

Where a^{s} is only dependent on the properties of the pure solute, while all the influence of the solvent on the mole fraction solubility, x, is captured by the activity coefficient, γ. In an ideal solution the

activity coefficient is equal to unity and the so called ideal solubility can be directly determined from the properties of the pure solute. In a simplified treatment the heat capacity terms are neglected (which is not always a valid approximation) and the ideal solubility can be estimated from the melting enthalpy and the melting temperature only as shown in eq. 3.[19, 20]

$$\ln x_{id} = \frac{\Delta H_m}{R}\left(\frac{1}{T_m} - \frac{1}{T}\right) \quad (3)$$

Normally, the activity coefficient is far from unity and thus needs to be accounted for in estimation of the actual solubility. Varieties of methods have been developed and are still under development[20, 25, 26, 30] for estimation of activity coefficients. However, as a necessary step in predicting solid-liquid equilibria the ideal solubility has to be characterized and this can be done approximately if we can predict the melting properties of the cocrystal.

Sofar, there is no simple method for prediction of $\Delta H^f(T\text{m})$, $\Delta S^f(T\text{m})$, and melting point (Tm), of the cocrystal and hence the ideal mole fraction solubility ($\ln x_{id}$) cannot be predicted prior to the synthesis of the CC. In our previous studies, a method has been developed for prediction of solid–state properties of the CCs,[31, 32]; melting point, crystal lattice energy and crystal density of the CCs using Artificial Neural Network (ANN) models. The model provides a high level of accuracy towards the prediction of the melting point of the CCs with an average relative error of 1.93%,[31, 32] but the model needs to incorporate the 1:1 MC–coformer gas phase binding energy as an input parameters. This binding energy were calculated using the material studio software programmed by appropriate selection of 1:1 MC–coformer synthon, based on crystal engineering principles.

In the present work, we develop a computational methodology which uses three properties of MC and three properties of coformer's, and predict the melting properties: $\Delta H^f(T\text{m})$, $\Delta S^f(T\text{m})$, Tm and the ideal mole fraction solubility of the CCs using a robust machine learning technique of the Matlab software program.

Methods and Modelling

Extraction of data from the literature

Eight different MCs, given in Fig. 1, have been selected for which $\Delta H^f(T\text{m})$ and Tm values were found in the literature, Table 1.[33-41] Additionally, all of these molecules have been studied in order to improve the physical properties of the solid form *via* formation of CCs using various GRAS coformers. Mostly, the studies succeeded to improve properties like hygroscopicity, tabletability, thermal stability, solubility and dissolution rate *etc*.[33-41] and crystallographic information files (cif) of the CCs are deposited and made available in the Cambridge Structural Database (CSD) software. In total, 30 CCs with known $\Delta H^f(T\text{m})$ and Tm values of these eight different MCs, were found and used. The properties of 22 coformers are given in Table 2.[33-41] Data for the cocrystals are given in the result presentation: $\Delta H^f(T\text{m})$ values in Table 6, and T_m values in Table 9. The $\Delta S^f(T\text{m})$ values of the CCs were calculated by dividing the $\Delta H^f(T\text{m})$ of the CCs with the respective Tm value of the CCs, and are given in Table 7. The thermodynamic parameters were determined by DSC analysis, and it is noteworthy that while melting temperatures for CCs are quite commonly found, melting enthalpies are much less frequently reported. This is the primary reason why the data bank had to limitiecd to 30 CC systems. Out of 30 co-crystal systems, 18 systems belong to 1:1 stochiometric ratio complexes, seven are forming 1:2 complexes, and the rest are 2:1 complexes.

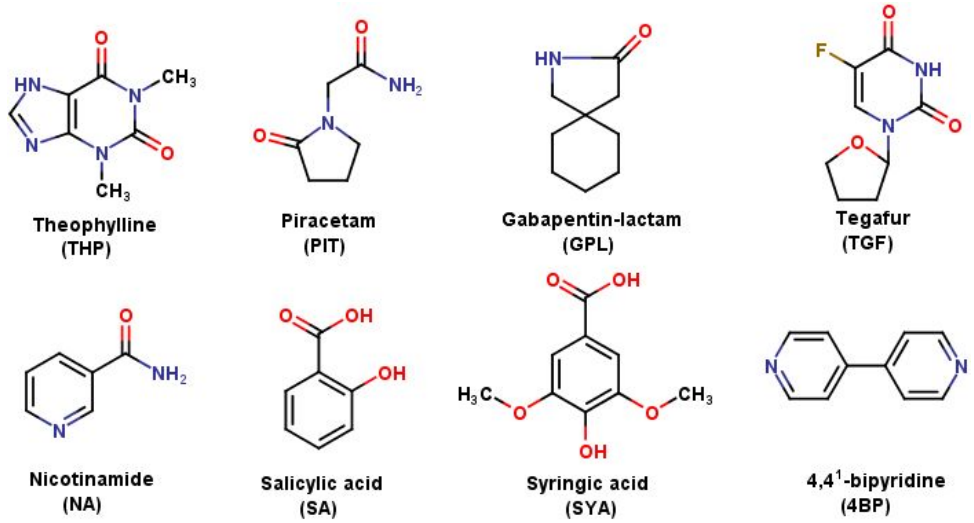


Figure 1. Chemical structures of all the eight model components which are used in this study.

Table 1. Model compounds (MCs) used in the study and the data used for the construction of models from ANN model–I to ANN model–V.

Name of MC	MW_{MC} (g/mol)	Tm_{MC} (K)	ΔH^{f}_{MC} (Tm) (kJ/mol)	ΔS^{f}_{MC} (Tm) (J/K/mol)
THP	180.16	544.15	29.60	54.40
PIT	142.15	425.15	41.11	96.70
GPL	153.2	359.65	19.00	52.83
TGF	200.16	445.15	20.26	45.50
NA	122.12	401.15	19.59	48.83
SA	138.12	432.15	24.60	56.92
SYA	198.17	480.15	37.75	78.62
4BP	156.19	387.15	24.70	158.14

Table 2. Coformers (CFs,) used in this study and the parameters used for the construction of models from ANN model–I to ANN model–V.

Name of the coformer (code)	MW_{CF} (g/mol)	Tm_{CF} (K)	ΔH^{f}_{CF} (Tm) (kJ/mol)	ΔS^{f}_{CF} (Tm) (J/K/mol)
Glutaric acid (GTA)	132.12	369.65	20.70	56.00
Diflunisal (DIF)	250.20	483.65	36.00	74.43
Diclofenac (DICFA)	296.15	430.65	40.90	94.97
Myricetin (MYCT)	318.23	630.15	137.80	218.68
L-(+)-Mandelic Acid (LMDA)	152.15	405.15	24.50	60.47
Citric acid (CTA)	192.12	429.15	40.32	93.95
L(+)-Tartaric acid (LTA)	150.09	442.15	36.31	82.12
Benzoic acid (BA)	122.12	395.56	18.60	47.02
4-aminobenzoic acid (4ABA)	137.14	460.15	27.20	59.11
4-Hydroxybenzoic acid (4HBA)	138.12	487.65	30.30	62.13
Fumaric acid (FUA)	116.07	560.15	31.80	56.77
Phloroglucinol (PG)	126.11	405.5	25.90	63.87
Isonicotinamide (INA)	122.12	429.15	17.19	40.07
Flufenamic acid (FFA)	281.23	408.15	26.70	65.42
Niflumic acid (NFA)	282.22	477.15	36.50	76.50
Tolfenamic acid (TFA)	261.71	480.15	41.00	85.39

Mefenamic acid (MFA)	241.28	503.65	38.70	76.84
Naproxen (NPX)	230.26	426.15	34.20	80.25
Febuxostat (FBX)	316.37	511.65	35.72	69.81
Isoniazid (INZ)	137.14	444.55	27.91	62.79
Urea (URE)	60.06	406.15	14.41	35.48
Fenamic acid (FA)	213.23	456.65	39.70	86.94

Analysis of the data bank

Cocrystal enthalpies and entropies are given per mole cocrystal stoichiometric complex; 1:1 or 1:2 or 2:1 moles of the coformer and the MC, and the number of molecules in the complex will depend on the stoichiometry of the cocrystal. Accordingly, e.g. for a 1:2 cocrystal the complex contains one molecule of MC and two molecules of the coformer, and the number of molecules in the complex equals three. The melting temperatures of the CCs vary from 357.05 K to 478.07 K, and among the eight different MCs, the melting point vary from 359.65 K to 544.15 K In Fig. 2, it is shown that there is no clear correlation between the melting point of the CCs and that of the MC or the CF, respectively. It should be noted though that the pure compound structures are not systematically related. As a comparison, Aakeröy *et. al*.[42] demonstrated that the length of the carbon chain of the diacid coformer ranging from succinic acid to dodecanedioic acid influence systematically on the melting point and the solubility of the CCs of Bis(pyridinecarboxamido)alkane. It was concluded that, the melting point and the aqueous solubility of the five CCs increase as the melting point of the coformer increases.

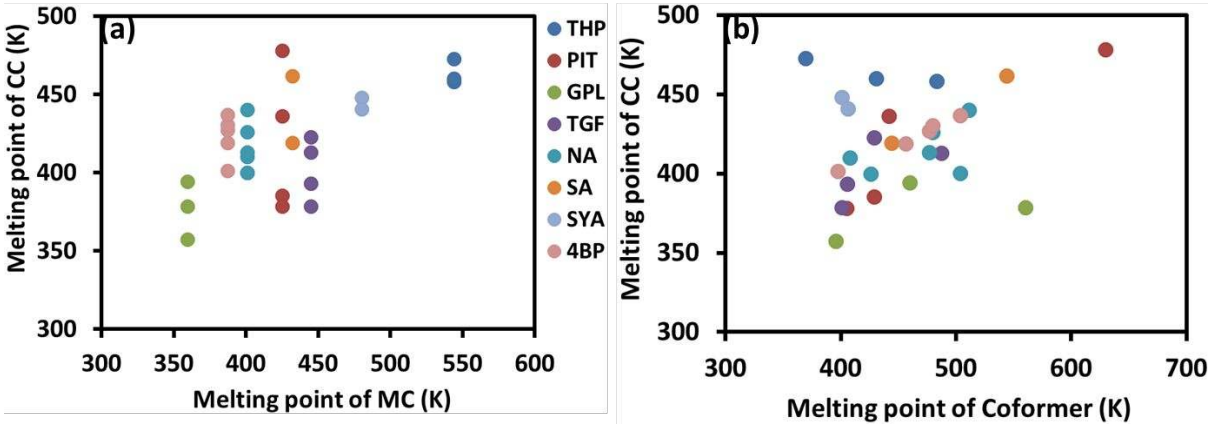


Figure 2. Correlation between the melting points of MC (K) *vs* the melting point of the CCs (K) given in (a), while the correlation between the melting point of the CF (K) *vs* the melting point of the CCs (K) given as in (b).

The cocrystal enthalpy of melting varies from 9.91 kJ/mol to 279.64 kJ/mol. Overall it is noted that the data tend to separate into two groups. All four CCs of PIT, both the CCs of syringic acid, all four of TGF and one out of five for NA have clearly high melting enthalpies in the range 179.99 kJ/mol to 279.64 kJ/mol. All the rest have melting enthalpies below 98.11 kJ/mole. In Fig. 3, it is shown that how the melting enthalpy is related to the melting point of the CC and whether there is a link to the melting enthalpies of the pure compounds. Different colours are used for the different MCs. In general, the $\Delta H_f^{\circ}(T\mathrm{m})$ and Tm relates to the packing pattern of the molecules and interactions which exist in–between the MC and CF of the CCs. Overall the diagram shows that there is very little correlation between the melting enthalpy and the melting temperature of the CCs. In diagrams b – d of Fig. 3 is shown how the melting enthalpy of the cocrystal relates to the melting enthalpy of the pure

compounds. In Fig 3 d, the melting enthalpy of the cocrystal is plotted against the stoichiometric sum of the pure component melting enthalpies. Obviously, there is no correlation between the cocrystal melting enthalpy and the melting enthalpies of the pure compounds. For the lower melting enthalpy cocrystals, the melting enthalpy is to some extent comparable to the stoichiometric sum of the melting enthalpies of the pure components, however not so for the higher melting enthalpy cocrystals.

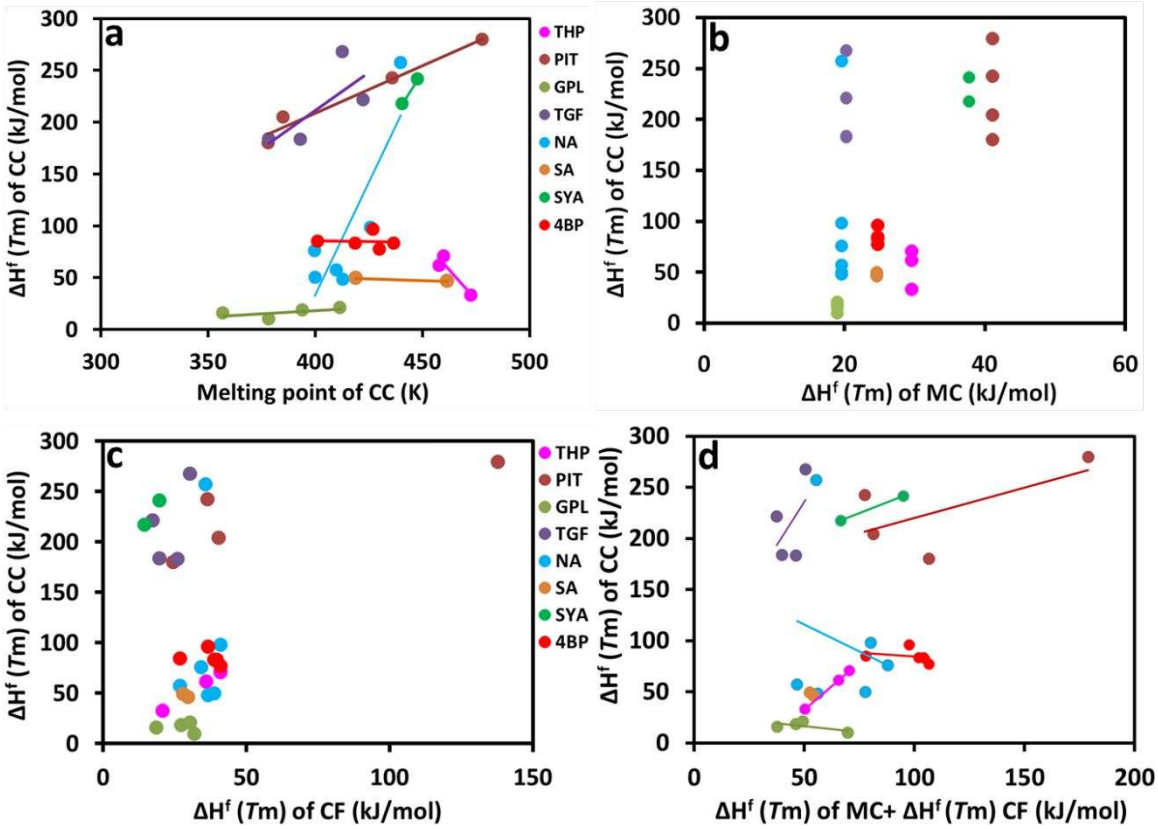


Figure 3. Analysis of cocrystal melting enthalpies given as kJ/mole of respectively the 1:1/1:2/2:1 cocrystal complex; a) cocrystal melting enthalpy *vs* cocrystal melting temperature; b) cocrystal melting enthalpy *vs* melting enthalpy of the MC; c) cocrystal melting enthalpy *vs* melting enthalpy of the coformer; d) cocrystal melting enthalpy *vs* sum of pure component melting enthalpies according to the stoichiometry.

From the Fig. 3, it is understood that, the melting enthalpy of MCs and CFs are within the range of 20 – 50 kJ/mol. But, the $\Delta H^f(Tm)$ of the CCs is higher by a factor from 2 to 10. Partly this stems from that the value is per mole of cocrystal complex. To some extent a higher value may reflect that the CCs can reach a more dense packing arrangement.

In Fig. 4 is presented the corresponding analysis of the CC melting entropies. Trendlines are used to distinguish the different MCs and have the same colour coding as in Fig. 3. The CC entropy change upon melting varies from 26.19 J/K/mol to 648.32 J/K/mol. Overall it is noted that the data tend to separate into the same two groups as for melting enthalpies. All four CCs of PIT, both the CCs of syringic acid, all four of TGF and one out of five for NA have clearly high melting entropies in the range from 465.83 J/K/mol to 648.32 J/K/mol. All the rest have melting entropies below 230.31 J/K/mol.

In diagrams a – c of Fig. 4, obviously there is very little systematic correlation between the cocrystal melting entropy and the melting entropy of either the MC or the coformer, besides the

observation above on the fact that there is a particular group of high melting entropy cocrystals not evenly distributed between the MCs. With the only exception of the cocrystals of gabapentin lactam the cocrystal melting entropy is higher than that of any of the individual components. Of course this is partly because the cocrystal is per mole stoichiometric complex. As shown in diagram c) for a majority of the cocrystals (the lower melting entropy group) the melting entropy is to some extent comparable to the sum of the melting entropies of the pure components.

In Fig. 4b, the trend is that $\Delta S^{f}(T\mathrm{m})$ of the CCs is increasing with increasing $\Delta S^{f}(T\mathrm{m})$ of CF. This observation is clearly noticed for the MCs: PIT, GPL, TGF, NA, SA and SYA, while not for THP and 4BP. In addition, $\Delta S^{f}(T\mathrm{m})$ of the CCs increases with increasing $T\mathrm{m}$ of the CCs, as found for the MCs: PIT, GPL, NA, SYA and 4BP, but not for remaining three: THP, TGF and SA. $\Delta S^{f}(T\mathrm{m})$ of the CCs is nearly equal to the sum of the $\Delta S^{f}(T\mathrm{m})$ of the MC and of the CF, especially, for GPL, 4BP, SA and three of six CCs of NA. For the other four MCs: THP, PIT, TGF, SYA and three of the CCs of NA have a clearly higher entropy than the corresponding sum of the individual components.

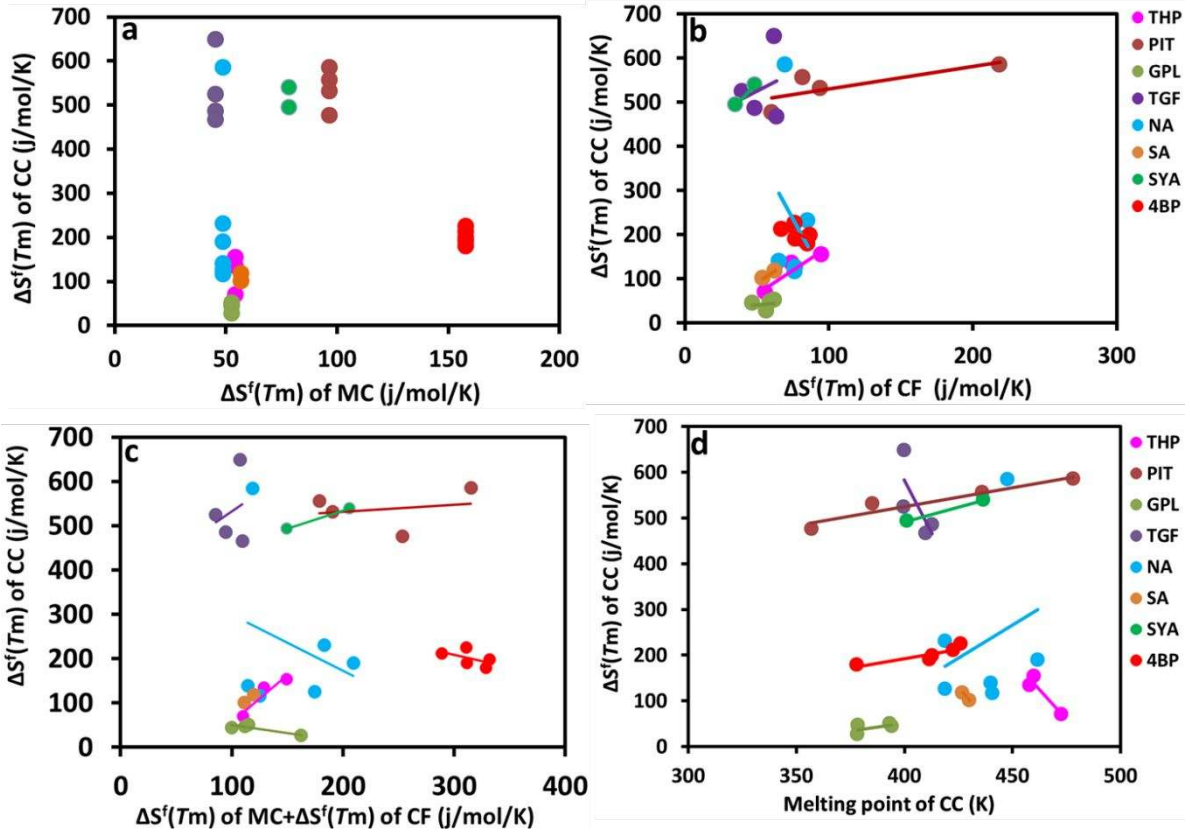


Figure 4. Analysis of cocrystal melting entropies given as J/K/mol for respectively the 1:1/1:2/2:1 cocrystal complex. a) cocrystal melting entropy *vs* melting entropy of the MC; b) cocrystal melting entropy *vs* melting entropy of the coformer; c) cocrystal melting entropy *vs* sum of pure component melting entropies according to the stoichiometry and d) cocrystal melting entropy *vs* cocrystal melting temperature.

In Fig. 5 and 6, melting entropies and enthalpies are compared: for MCs in Fig. 5a, for CFs in Fig. 5b and for cocrystals in Fig. 6. Obviously, there is overall a clear relation–between these two parameters, in accordance with the principle of so called enthalpy-entropy compensation. The stronger the bonding in the solid phase compared to the melt, the greater is the reduction of the solid phase entropy compared to the melt.

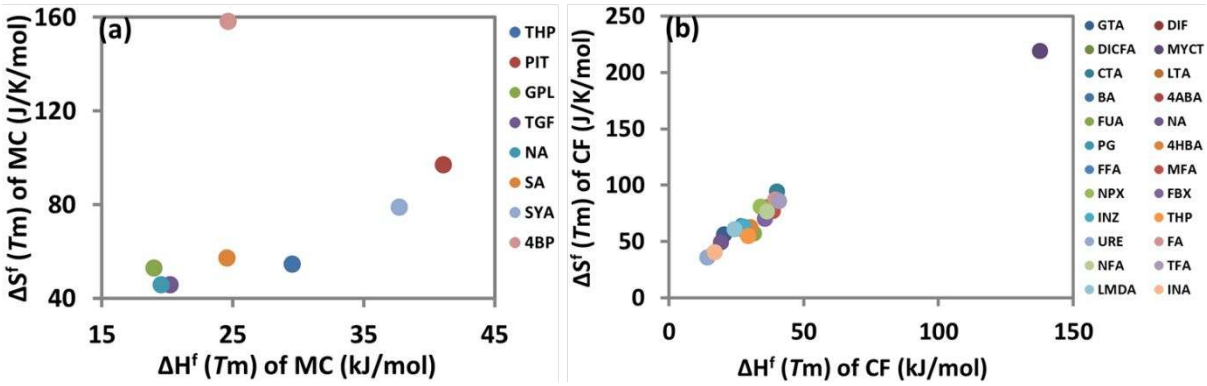


Figure. 5. Correlation of $\Delta H^f(Tm)$ of MC and CF againt the $\Delta S^f(Tm)$ of the MC (a), and CF (b), respectively.

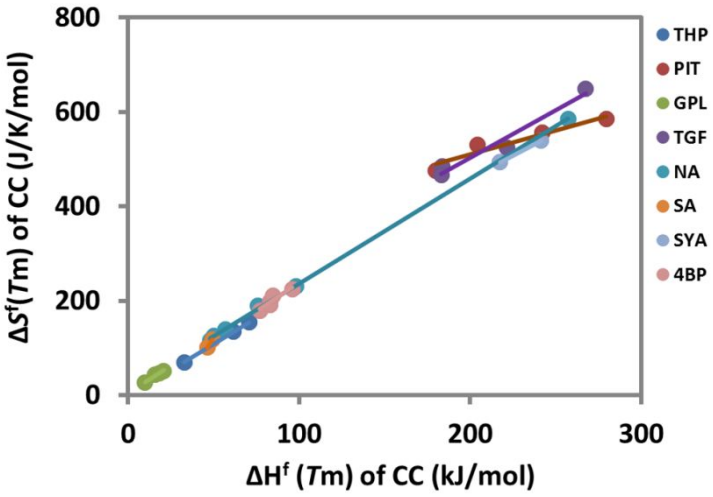


Figure 6. Correlation of $\Delta H^f(Tm)$ of CCs against the $\Delta S^f(Tm)$ of the CCs. The trend lines help to distinguish cocrystals of the same MC.

The ideal solubilities at 25 ℃, as calculated by eq 3, for the model compounds, coformers and the cocrystals are given in Table 3. In Fig. 7a, the ideal cocrystal solubility is plotted *vs* the cocrystal melting temperature.

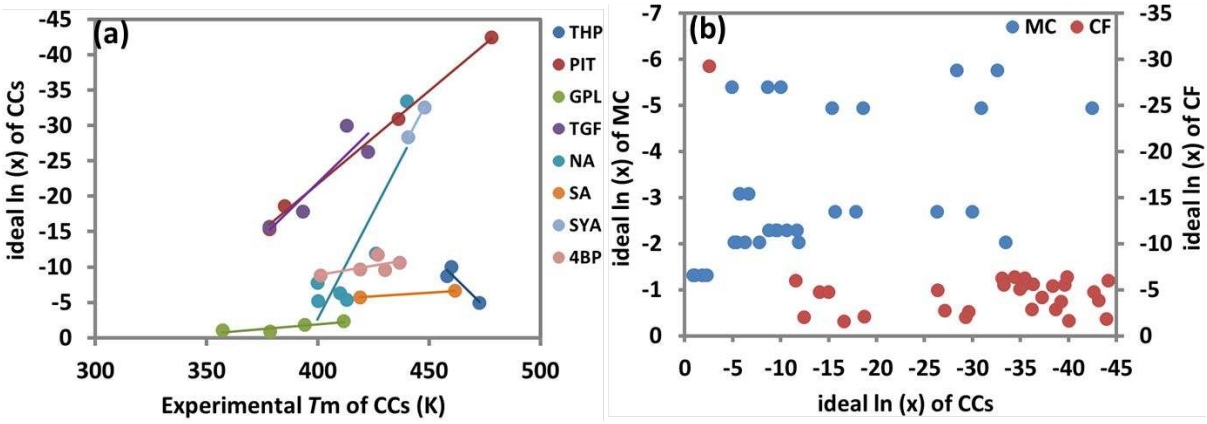


Figure 7. The solubility of CCs *vs* the experimental *T*m values of CCs given as (a) while the solubility of the CCs *vs* the solubility of the MCs/solubility of the CFs given as (b) in figure.

Table 3. The ideal mole fraction solubility at 25 °C. Values calculated from melting data for each respective form according to eq. 3.

Name of the CCs (ratio)	ln x of MC	x of MC	ln x of CF	x of MC	ln x of CCs	x of CCs
THP:GTA (1:1)	-5.40	0.81	-1.62	26.15	-4.91	0.00737
THP:DIF (1:1)	"	"	-5.57	0.95	-8.67	0.00017
THP:DICFA (1:1)	"	"	-5.08	1.84	-10.03	0.00004
PIT:MYCT (1:1)	-4.95	1.01	-29.29	$6.06\text{x}10^{-11}$	-42.46	$3.63\text{x}10^{-19}$
PIT:LMDA (2:1)	"	"	-2.61	11.19	-15.36	0.0000002
PIT:CTA (1:1)	"	"	-4.97	1.33	-18.63	$8.11\text{x}10^{-09}$
PIT:LTA (1:1)	"	"	-4.77	1.27	-30.94	$3.66\text{x}10^{-14}$
GPL:BA (1:1)	-1.31	41.34	-1.85	19.20	-1.05	0.34993
GPL:4ABA (1:1)	"	"	-3.86	2.89	-1.81	0.16365
GPL:4HBA (1:1)	"	"	-4.75	1.19	-2.33	0.097295
GPL:FUA (2:1)	"	"	-6.00	0.29	-0.85	0.42741
TGF:NA (1:1)	-2.70	13.45	-2.03	-2.03	-15.70	0.00000015
TGF:PG (1:1)	"	"	-2.77	7.90	-17.85	0.00000002
TGF:INA (1:1)	"	"	-2.12	14.66	-26.29	$3.82\text{x}10^{-12}$
TGF:4HBA (1:1)	"	"	-4.75	1.19	-30.00	$9.36\text{x}10^{-14}$
NA:FFA (1:1)	-2.03	16.04	-2.90	15.47	-6.28	0.00187
NA:NFA (1:1)	"	"	-5.52	1.13	-5.39	0.004561
NA:TFA (2:1)	"	"	-6.27	0.49	-11.88	0.000006
NA:MFA (2:1)	"	"	-6.37	0.41	-5.14	0.00585
NA:NPX (1:2)	"	"	-4.14	3.67	-7.77	0.00042
NA:FBX (1:1)	"	"	-6.01	0.78	-33.44	$3.00\text{x}10^{-15}$
SA:INZ (1:1)	-3.08	6.35	-3.71	3.36	-5.74	0.00321
SA:THP (1:1)	"	"	-5.40	0.81	-6.63	0.00132
SYA:NA (2:1)	-5.77	0.62	-2.03	16.04	-32.57	$7.16\text{x}10^{-15}$
SYA:URE (1:2)	"	"	-1.55	12.75	-28.37	$4.78\text{x}10^{-13}$
4BP:FA (1:2)	-2.29	15.82	-5.56	0.82	-9.66	0.00006
4BP:NFA (1:2)	"	"	-5.52	1.13	-11.70	0.000008
4BP:TFA (1:2)	"	"	-6.27	0.49	-9.54	0.000071
4BP:MFA (1:2)	"	"	-6.37	0.41	-10.63	0.000024
4BP:FFA (1:2)	"	"	-2.90	15.47	-8.77	0.000155

Obviously, the calculated ideal mole fraction solubility of CCs *i.e.*, ln (x_{id}) at 25 ℃ is reasonably correlated to the experimental melting point of the CCs. Here one mole of CC relates to one mole of CC complex as defined before. The higher the melting point the lower the ideal solubility. If the linear relation between melting entropy and melting enthalpy shown in Fig. 6 is accounted for the ideal solubility can be given as:

$$\ln x_{id} = \frac{1}{R}\left(\Delta S_m - \frac{\Delta H_m}{T}\right) = \frac{\Delta S_m}{R}\left(1 - \frac{1}{2.35T}\right)$$

In Fig. 7b, the ideal solubility of the CCs is plotted against the ideal solubility of the MC on the left hand axis (blue), and of the CF on the right hand axis (red). From the graph, it is concluded that, there is no correlation in general found between the solubility of CCs and the solubility of MCs or the solubility of CFs.

Artificial Neural Network Modelling

A series of five different ANN models have been constructed to predict physical properties of the co-crystals. The relation between the properties of the cocrystals and the properties of the pure compounds has been examined in a step wise manner to reach the ultimate goal, *i.e*., prediction of the cocrystal ideal solubility. The cocrystal physical property targeted by each ANN model is as follows:

ANN model-I: prediction of the $\Delta H^{f}(T\text{m})$
ANN model-II: prediction of ΔS^{f} (Tm)
ANN model–III: simultaneous prediction of $\Delta H^{f}(T\text{m})$ and ΔS^{f} (Tm)
ANN model-IV: prediction of Tm
ANN model-V: prediction of the ideal mole fraction solubility

All five ANN models has an input layer which provides input data for the neural network, and an output layer that contains the data to be predicted. In between, there is a hidden layer that connects the input and the output layer and is where all the computational work would be carried out. In our models the hidden layer contains six neurons (nine neurons for model II), weights, a sum function, *logsig* function as an activation function. The output layer contains $\Delta H^{f}(T\text{m})$, ΔS^{f} (Tm), Tm or the ideal mole fraction solubility, respectively for ANN model-I, ANN model-II, ANN model-IV and ANN model-V, as is represented in Fig. 8. In the case of ANN model-III, both the $\Delta H^{f}(T\text{m})$ and ΔS^{f} (Tm) were targeted as output neurons in the output layer, simultaneously. As input neurons, six affordable molecular descriptors of both MCs and CFs information have been included in the input layer of the neural network to obtain the reasonable level of accuracy. A feed–forward back–propagation network is used to transfer the information from the input layer to output layer until the convergence criterion is reached through the number of epochs (indicating the number of passes through the entire training dataset the machine learning algorithm has completed). The mean squared error between the experimental and the ANN predicted (output) values of the CCs were minimized by using the Gradient Descent method to adjust the number of weight parameters. A *logsig* function is used as an activation function to transfer the total sum of the input neurons into output neurons by passing through the number of hidden layers. Before training the model, all the six input neurons and output neurons were normalized by dividing with higher number value in each series of molecular descriptors, so that they fall within the interval range of 0–1, hence, the predicted output value and also expected standard deviation will also fall within the range of 0 - 1.[31, 32] Moreover, the obtained output values (refer Fig. 9, Fig. 11, Fig. 13, Fig. 15 and Fig. 17) will also set within the interval range of one for all the training, validation and test set of each MC. After the successful completion of the training process, the actual values of each molecular descriptor in each series were calculated back *via* denormalization process and the targeted values are plotted as experimental values against the predicted values by ANN model (refer Fig. 10, Fig. 12, Fig. 14, Fig. 16 and Fig. 18).

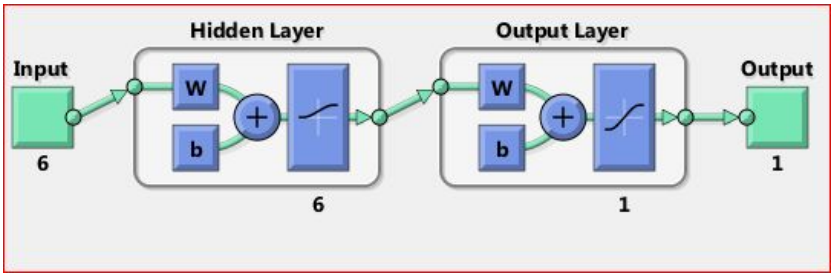


Figure 8. The architecture of the constructed ANN models are given in figure, the size of the model's is 6–6–1 (6 inputs–6 neurons in the hidden layer–one output) for model–I, model-IV and model-V. Whereas the size of the ANN model-II is 6–9–1 (6 inputs–9 neurons in the hidden layer–one output), and in model-III, both melting enthalpy and melting entropies were targeted simultaneously by remaining the number of neurons in the hidden layer as same so the size and architecture of the model is 6−6−2.

Table 4. Information about the number of models, the parameters used for the construction of models and targeted physical property of the CCs are given in table.

S. No	Model	Number of input parameters	Name of the input parameter	Targeted property
1	Model I	6	$MW_{MC,}$ $MW_{CF,}$ $Tm_{MC,}$ $Tm_{CF,}$ $\Delta H^{f}_{MC}(Tm)$ and $\Delta H^{f}_{CF}(Tm)$	$\Delta H^{f}(Tm)$
2	Model II	6	$MW_{MC,}$ $MW_{CF,}$ $Tm_{MC,}$ Tm_{CF} $\Delta S^{f}_{MC}(Tm)$ and $\Delta S^{f}_{CF}(Tm)$	ΔS^{f} (Tm)
3	Model III	6	same as 1	both $\Delta H^{f}(Tm)$ and ΔS^{f} (Tm), simultaneously
4	Model IV	6	same as 1	Tm
5	Model V	6	same as 1	mole fraction solubility (ln (x))

The overfitting of the models were avoided by selecting of one hidden layer, which consists of six neurons in the hidden layer of the constructed model–I, model–III, model–IV and model–V as shown in Fig. 8, while model–II consists of nine neurons in the hidden layer. The number of hidden layers has been chosen according to the Kolmogotov theorem, that demonstrates that less than two hidden layers are sufficient to construct a prediction model of the targeted property. A higher number than unity leads to overfitting and poor generalization capability of the model. At the same time the number of neurons in the hidden layer is systematically varied from lower to higher starting from two neurons. A too high number can lead to overfitting of the models, and as a consequence a reduced prediction capability (larger deviation between the experimental and the predicted property value). More details about the optimization of the number of hidden layers and hidden layer neurons is reported elsewhere.[31, 32] In the present study, an optimum result were obtained when the hidden layer contained six neurons being equal to the number of neurons in the input layer, except for model-II where the optimum was found for 9 hidden layer neurons. Thus, the size of the constructed neural network models for prediction of $\Delta H^{f}(Tm)$, $\Delta H^{f}(Tm)$ and $\Delta S^{f}(Tm)$ simultaneously, Tm and ln (x) of the CCs with a good generalization is 6–6–1, the generalized schematic representation of the models are as shown in Fig. 8. The performance of the model does not change beyond six neurons in the hidden layer, and consequently the training process has reached the convergence criteria with the usage of 68, 1000, 30 and 100 epochs, respectively, for ANN model–I, ANN model–III, ANN model–

IV and ANN model–V. Therefore, the constructed models are made up of six neurons as operational variables, such as, $MW_{MC,}$ $MW_{CF,}$ $Tm_{MC,}$ $Tm_{CF,}$ $\Delta H^{f}_{MC}(Tm)$ and $\Delta H^{f}_{CF}(Tm)$ in the input layer; six neurons in the hidden layer, and one output neuron in the output layer, *i.e*. $\Delta H^{f}(Tm)$, Tm and ln (x) of the CCs respectively for ANN model–I, ANN model–IV and ANN model–V. Whereas, two output neurons, such as both $\Delta H^{f}(Tm)$ and $\Delta S^{f}(Tm)$, simultaneously were targeted in the output layer for the construction of ANN model–III. Among the six operational variables, three input neurons were belonged to MCs such as $MW_{MC,}$ Tm_{MC} and ΔH^{f}_{MC} (Tm); and another three operational variables were belongs to CF such as $MW_{CF,}$ Tm_{CF} and $\Delta H^{f}_{CF}(Tm)$. In total, 30 CCs of eight different MC's were used for construction of all the five ANN models. Moreover, the 30 CCs data points were divided into three sets by selection of random selection process option in the neural network training process: such as i) 70% of the data used for the training set, and remaining of 30% of data used for both the validation set, and for the test set (as new data points for the prediction of targeted property of the CCs). Again, in which, 16% of the data points were used for the validation set and remaining 14% of the data points were used for the test set (randomly).

In case of ANN model–II, a good generalization of the model for prediction of $\Delta S^{f}(Tm)$ is obtained when nine neurons are incorporated in the hidden layer of the ANN model. Hence the size of the constructed model is 6–9–1 (6 neurons in the input layer–9 neurons in the hidden layer–$\Delta S^{f}(Tm)$ as an output neuron). Herein, the number of neurons in the hidden layer of ANN model–II has increased (in-comparison to rest of the four models) from 6 to 9, to improve the efficiency of the model with respect to prediction of melting entropy. Accordingly, the training process has been stopped by use of 92 epochs to reach the convergence criteria. Thus, the ANN model–II is made up of six operational variables in the input layer, nine neurons in the hidden layer and melting entropy as an output neuron. The six operational variables in the input layer are $MW_{MC,}$ $MW_{CF,}$ $Tm_{MC,}$ $Tm_{CF,}$ $\Delta S^{f}_{MC}(Tm)$ and $\Delta S^{f}_{CF}(Tm)$. Among six operational variables, three input neurons were belonged to MCs and the rest of the three input neurons were belonged to coformer. Thereby, the ANN model–II has constructed by dividing the data into three sets as explained above for construction of ANN model–I.

The training process

The Matlab program offers a machine learning Artificial Neural Network toolbox with an automated driving facility of the training process. The tools present in the toolbox were used to train and build the ANN models, the ANNmodel–I, ANN model–II, ANN model–III, ANN model–IV and ANN model–V are built to predict the $\Delta H^{f}(Tm)$, $\Delta S^{f}(Tm)$, $\Delta H^{f}(Tm)$ and $\Delta S^{f}(Tm)$ simultaneously, Tm and ln (x) of the CCs, respectively. The outline of the training process, which involved in the construction of all of the five models are explained as follows. During the training process, the six operational variables of training data fed into the input layer towards the forward direction using feed–forward back–propagation algorithm. In this process, each neuron in the hidden layer receives information from all the six input neurons in the input layer, which in turn multiplied with appropriate weights. Further, the total weighted sum of the input layer has transferred the information as output of the hidden layer in the form of nonlinear transformation. Similarly, the output layer gets the information from the hidden layer, which are multiplied with appropriate weights and then summed. The outcome of the output values of the output layer is the nonlinear transformation of the resulting sum obtained from the hidden layer. The obtained output values were compared with the targeted experimental output values. The deviation between the output values and targeted output values are calculated and propagated back through the back–propagation algorithm to update the connection weights, after completion of each training cycle. This process continues until it reaches to the convergence criterion with least possible mean square error percent. Notably, the weighted parameters

were not changed during the time of testing examples, such as, each test example will feed into the input layer in the feed–forward direction as similar like the feed–forward of the training data. Likewise, five models were constructed with a minimal amount of error towards the prediction of targeted property of the CCs. The obtained r^2 (overall) regression values of all the models are 0.999, 0.997, 0.986, 0.988 and 0.980 respectively, for model–I, model–II, model–III, model–IV and model–V. In general, a common problem that arises during the process of training is overfitting of the model through memorizing of the training set data points. This is the case where the model produces a very good fit to the training set, but a much larger deviation for the validation and the test set. Our final models give comparable regression values for the training set, validation set and test set. Thus, there is no overfitting of the models.

But in the case of ANN model–II towards prediction of $\Delta S^{f}(T\text{m})$, the best results were prevailed with consideration of nine neurons in the hidden layer of the neural network. Notably, the regression value has been improved from 0.952 to 0.999 (see Table 5) through increasing the number of neurons from six (used in the ANN model–I) to nine in the hidden layer of the neural network. The performance of the model does not change beyond nine neurons. Thus, the architecture of ANN model–II, for prediction of $\Delta S^{f}(T\text{m})$ is 6–9–1. Moreover, the model uses the $\Delta S^{f}_{MC}(T\text{m})$ and $\Delta S^{f}_{CF}(T\text{m})$ (not like other four models) as input neurons instead of $\Delta H^{f}_{MC}(T\text{m})$ and $\Delta H^{f}_{CF}(T\text{m})$. The training process was stopped with a regression value of 0.999 for the training set and 0.997 for the validation set as given in Fig. 11.

Table 5. Improvement in the performance of the ANN model–II through increasing the number of neurons in the hidden layer from six to nine, the obtained r^2 values of the training set, validation set, test set and overall an average regression value of the model has listed in the table.

size of the hidden layer	obtained r^2 value of the training set	obtained r^2 value of the validation set	obtained r^2 value of the test set	overall r^2 value of the model
6	0.982	0.873	0.930	0.952
7	0.966	0.988	0.986	0.975
8	0.987	0.980	0.996	0.989
9	0.999	0.997	0.993	0.997

The main difference between the models is that they, based on the same six operational input parameters of the parent components, target the prediction of different properties of the co-crystals, with model-II being an exception since here the enthalpies of the pure compounds are replaced by the corresponding entropies as input data. In addition, in all models the number of hidden layer neurons obtained in the model optimization is the same, i.e. six, except for in case of model-II where the optimum was found for nine hidden layer neurons. It is noticed that the prediction is somewhat better when the output layer only target one parameter. This not surprising, since the optimum for one single parameter is likely to be different from that of another single parameter, and if both are targeted at the same time the optimization has to find a compromise between the two.

Results and Discussion

ANN Model–I: prediction of $\Delta H^{f}_{CC}(T\text{m})$

The capability of the ANN model–I to predict the $\Delta H^{f}_{CC}(T\text{m})$ of the CCs from the molecular weight, the melting temperature and the melting enthalpy of the MC and the coformer is shown in Fig. 9. The experimental values (after normalization) are represented on the x–axis, whereas the predicted

$\Delta H^{f}(T\mathrm{m})$ values using ANN model–I is given on y-axis. The denormalised representation of graph of ANN model–I is shown in Fig. 10. The calculated mean squared error for the prediction of $\Delta H^{f}(T\mathrm{m})$ is 0.103 kJ/mol. The calculated deviation for each CC is given in Table 6. The smallest deviation of 0.01 kJ/moles obtained for the PIT CC with MYCT coformer (marked as a green color in Table 6), whereas the highest deviation is obtained for NA with NFA coformer with a value of –40.21 kJ/mol (marked as a red color in Table 6).

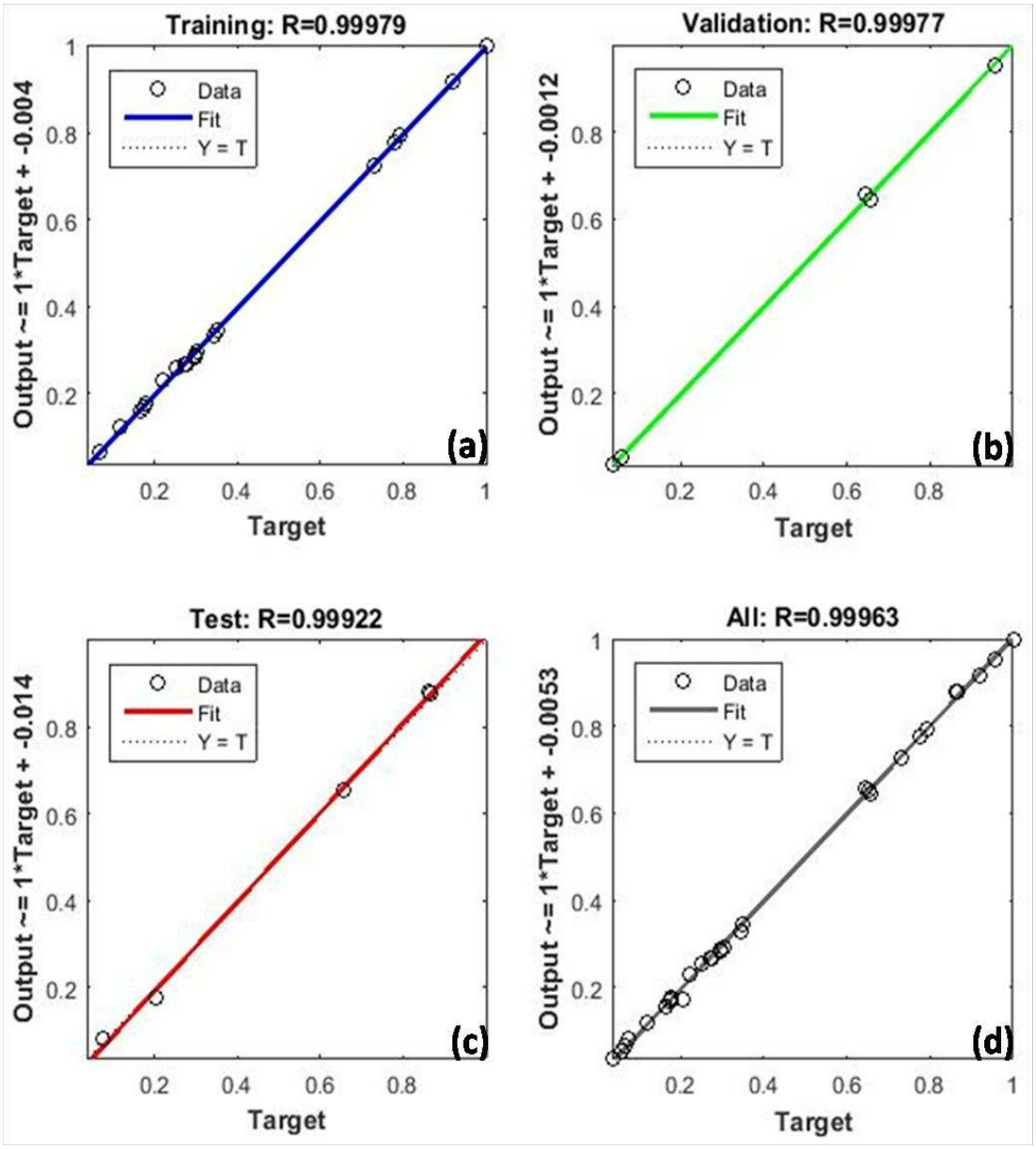


Figure 9. ANN model–I: prediction of $\Delta H^{f}(T\mathrm{m})$. normalised regressions, a) training set, b) validation set, c) test set and d) overall.

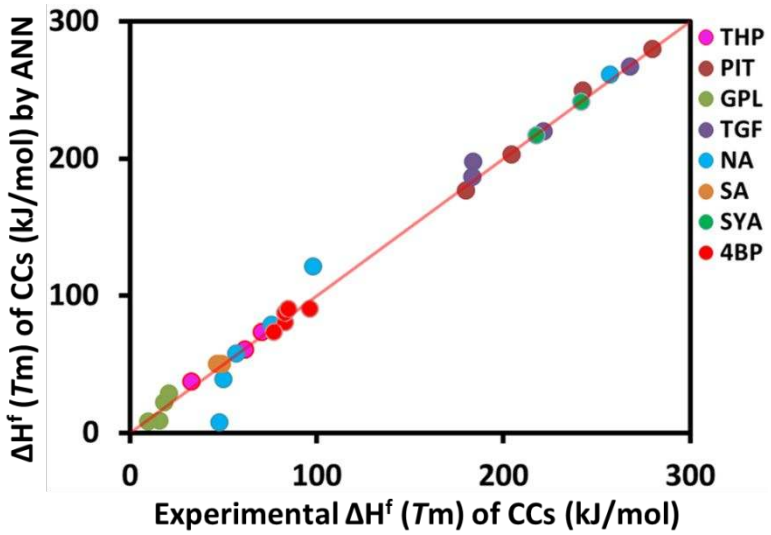


Figure 10. ANN model–I: quantitative overall performance including all data. Color coding same as in previous figures.

Table 6. Experimental *vs* predicted $\Delta H^{f}(T\text{m})$ values of the CCs obtained by ANN model–I.

Name of the CCs (ratio)	Experimental ΔH^{f}_{CC} (Tm) (kJ/mol)	Predicted ΔH^{f}_{CC}(Tm) by ANN (kJ/mol)	Error value (kJ/mol)
THP:GTA (1:1)	32.95	37.35	4.4
THP:DIF (1:1)	61.57	60.68	-0.89
THP:DICFA (1:1)	70.68	73.36	2.68
PIT:MYCT (1:1)	**279.64**	**279.65**	**0.01**
PIT:LMDA (2:1)	179.99	176.72	-3.27
PIT:CTA (1:1)	204.44	202.78	-1.66
PIT:LTA (1:1)	242.39	249.30	6.91
GPL:BA (1:1)	15.72	8.93	-6.79
GPL:4ABA (1:1)	18.42	22.05	3.63
GPL:4HBA (1:1)	20.92	28.60	7.68
GPL:FUA (2:1)	9.91	8.42	-1.49
TGF:NA (1:1)	183.60	197.44	13.84
TGF:PG (1:1)	183.14	186.65	3.51
TGF:INA (1:1)	221.39	220.13	-1.26
TGF:4HBA (1:1)	267.66	267.17	-0.49
NA:FFA (1:1)	57.07	57.81	0.74
NA:NFA (1:1)	**48.06**	**7.85**	**-40.21**
NA:TFA (2:1)	98.11	121.17	23.06
NA:MFA (2:1)	50.06	39.31	-10.75
NA:NPX (1:2)	75.78	78.76	2.98
NA:FBX (1:1)	257.16	261.20	4.04
SA:INZ (1:1)	49.35	50.25	0.9
SA:THP (1:1)	46.42	50.04	3.62
SYA:NA (2:1)	241.50	241.64	0.14
SYA:URE (1:2)	217.43	217.17	-0.26
4BP:FA (1:2)	83.10	80.70	-2.4
4BP:NFA (1:2)	96.11	90.59	-5.52
4BP:TFA (1:2)	77.09	73.58	-3.51
4BP:MFA (1:2)	83.10	87.51	4.41
4BP:FFA (1:2)	84.70	90.28	5.58

ANN Model–II: prediction of $\Delta S^f(T\mathrm{m})$

The capability of the ANN model–II to predict the $\Delta S^f(T\mathrm{m})$ of the CCs, from the molecular weight, the melting temperature and the melting entropy of the MC and the coformer is shown in Fig. 11. The experimental values (normalized) are given on the x-axis whereas the predicted $\Delta S^f(T\mathrm{m})$ values of ANN model–II is given on the y-axis. The denormalised result is shown in Fig. 12. The obtained average deviation is -1.68 J/K/mol. The calculated deviation for each CC is given in Table 7. The smallest deviation of 0.13 J/K/mol is obtained for GPL with FUA coformer (marked as a green color in Table 7), whereas the highest deviation is obtained for SA with INZ cocrystal with a deviation of –62.80 J/K/mol (marked as a red color in Table 7).

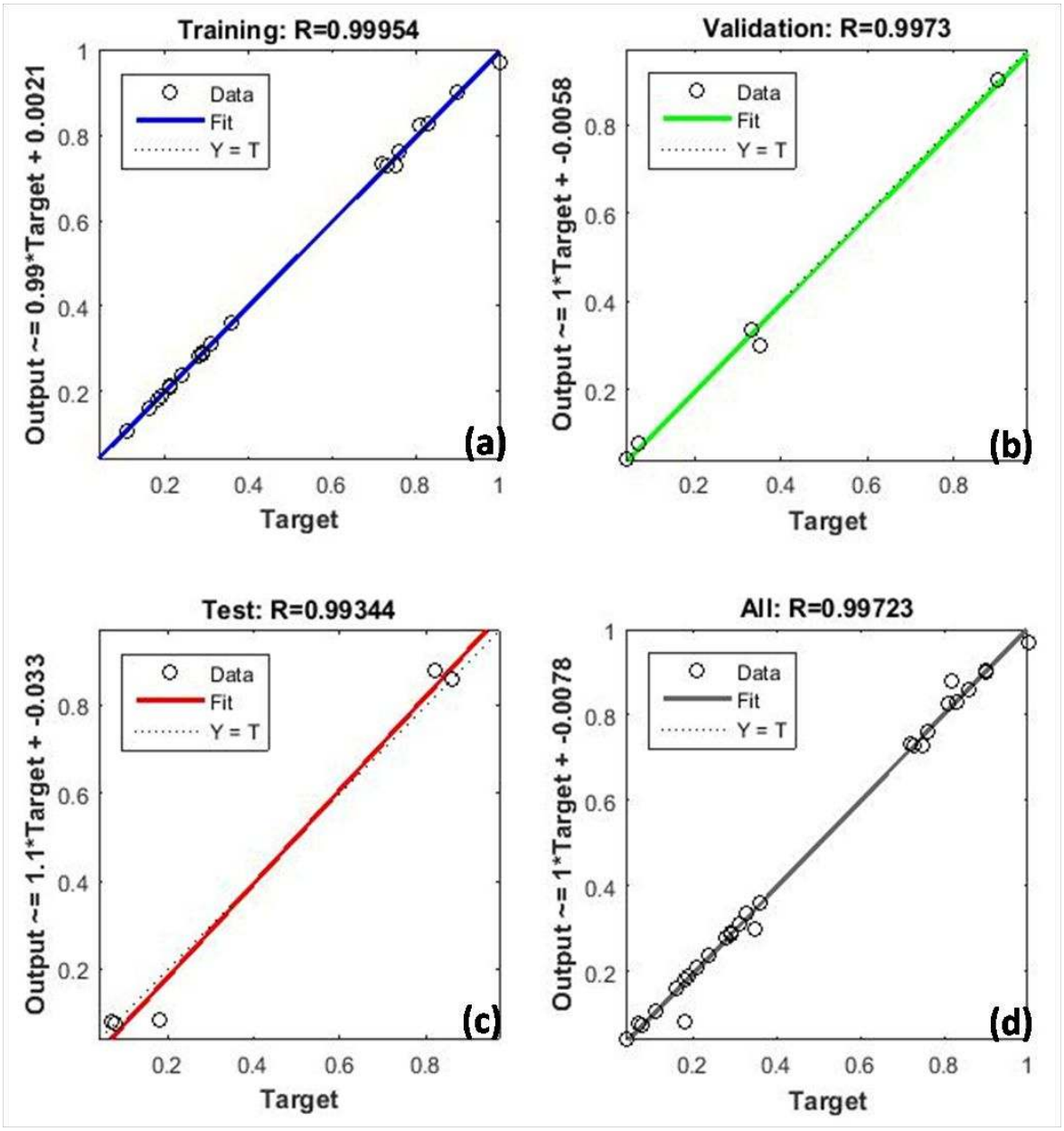


Figure 11. ANN model–II: prediction of $\Delta S^f(T\mathrm{m})$. normalised regressions, a) training set, b) validation set, c) test set and d) overall.

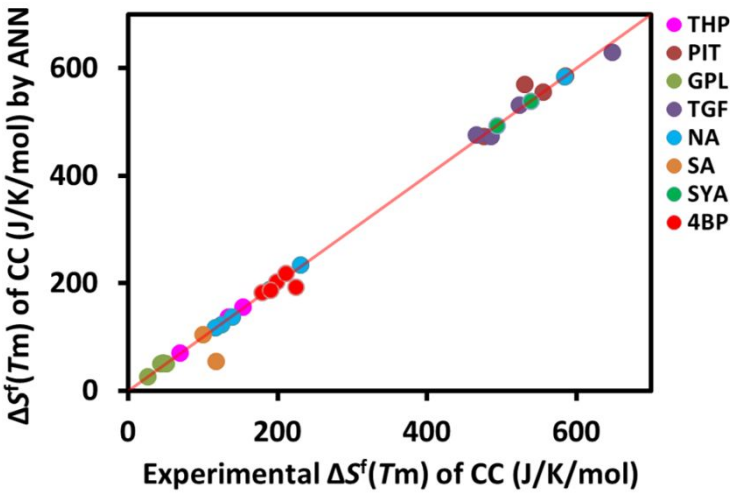


Figure 12. ANN model–II: quantitative overall performance including all data points. Color coding same as in previous figures.

Table 7. Experimental *vs* predicted $\Delta S^f(T\text{m})$ values of the CCs by ANN model–II, which are listed in table (J/K/mole of stoichiometric complex).

Name of the CCs (ratio)	Experimental ΔS^f_{CC} (Tm) (J/K/mol)	Predicted ΔS^f_{CC}(Tm) by ANN (J/K/mol)	Error value (J/K/mol)
THP:GTA (1:1)	69.71	69.55	-0.16
THP:DIF (1:1)	134.45	136.85	2.41
THP:DICFA (1:1)	153.64	155.21	1.57
PIT:MYCT (1:1)	584.94	584.76	-0.17
PIT:LMDA (2:1)	475.98	473.31	-2.67
PIT:CTA (1:1)	530.81	570.12	39.31
PIT:LTA (1:1)	555.75	555.98	0.23
GPL:BA (1:1)	44.03	50.64	6.61
GPL:4ABA (1:1)	46.73	52.15	5.42
GPL:4HBA (1:1)	50.82	49.74	-1.08
GPL:FUA (2:1)	**26.19**	**26.32**	**0.13**
TGF:NA (1:1)	485.26	473.11	-12.16
TGF:PG (1:1)	465.83	475.78	9.95
TGF:INA (1:1)	523.94	531.62	7.69
TGF:4HBA (1:1)	648.32	629.35	-18.97
NA:FFA (1:1)	139.20	136.37	-2.83
NA:NFA (1:1)	116.37	116.69	0.32
NA:TFA (2:1)	230.31	233.40	3.09
NA:MFA (2:1)	125.15	123.15	-2.00
NA:NPX (1:2)	189.57	187.90	-1.67
NA:FBX (1:1)	584.52	583.54	-0.98
SA:INZ (1:1)	**117.79**	**54.99**	**-62.80**
SA:THP (1:1)	100.55	103.77	3.22
SYA:NA (2:1)	539.19	537.96	-1.24
SYA:URE (1:2)	493.37	492.91	-0.47
4BP:FA (1:2)	198.40	202.29	3.89
4BP:NFA (1:2)	225.11	193.10	-32.01
4BP:TFA (1:2)	179.22	182.02	2.80
4BP:MFA (1:2)	190.31	186.06	-4.25
4BP:FFA (1:2)	211.14	217.65	6.51

ANN model–III: simultaneous prediction of $\Delta H^{\circ}(T\mathrm{m})$ and $\Delta S^{\circ}(T\mathrm{m})$

Model–III aims to predict both the $\Delta H^{\circ}(T\mathrm{m})$ and $\Delta S^{\circ}(T\mathrm{m})$, simultaneously using the same six input parameters which are used for construction of ANN model–I. The training process was stopped with the regression value of 0.994 for the training set, 0.963 for the validation set and 0.990 for the test set. As a whole, the regression value obtained for the prediction of both the $\Delta H^{\circ}(T\mathrm{m})$ and $\Delta S^{\circ}(T\mathrm{m})$ is 0.986. The result is shown in terms of normalized values in Fig. 13. In Fig. 14 is given the denormalized result, and specific values are listed in Table 8. In both the $\Delta H^{\circ}(T\mathrm{m})$ (Fig. 14a) and $\Delta S^{\circ}(T\mathrm{m})$ (Fig. 14b), most of the data points fall well onto the trendline. That means, the ANN model–III is capable of predicting $\Delta H^{\circ}(T\mathrm{m})$ and $\Delta S^{\circ}(T\mathrm{m})$, simultaneously in one and the same model. However, the regression value of ANN model–III *i.e.* 0.986, is lower than the values obtained for models targeting the individual parmeters only: model–I: $\Delta H^{\circ}(T\mathrm{m})$ (0.999) and model–II: $\Delta S^{\circ}(T\mathrm{m})$ (0.997).

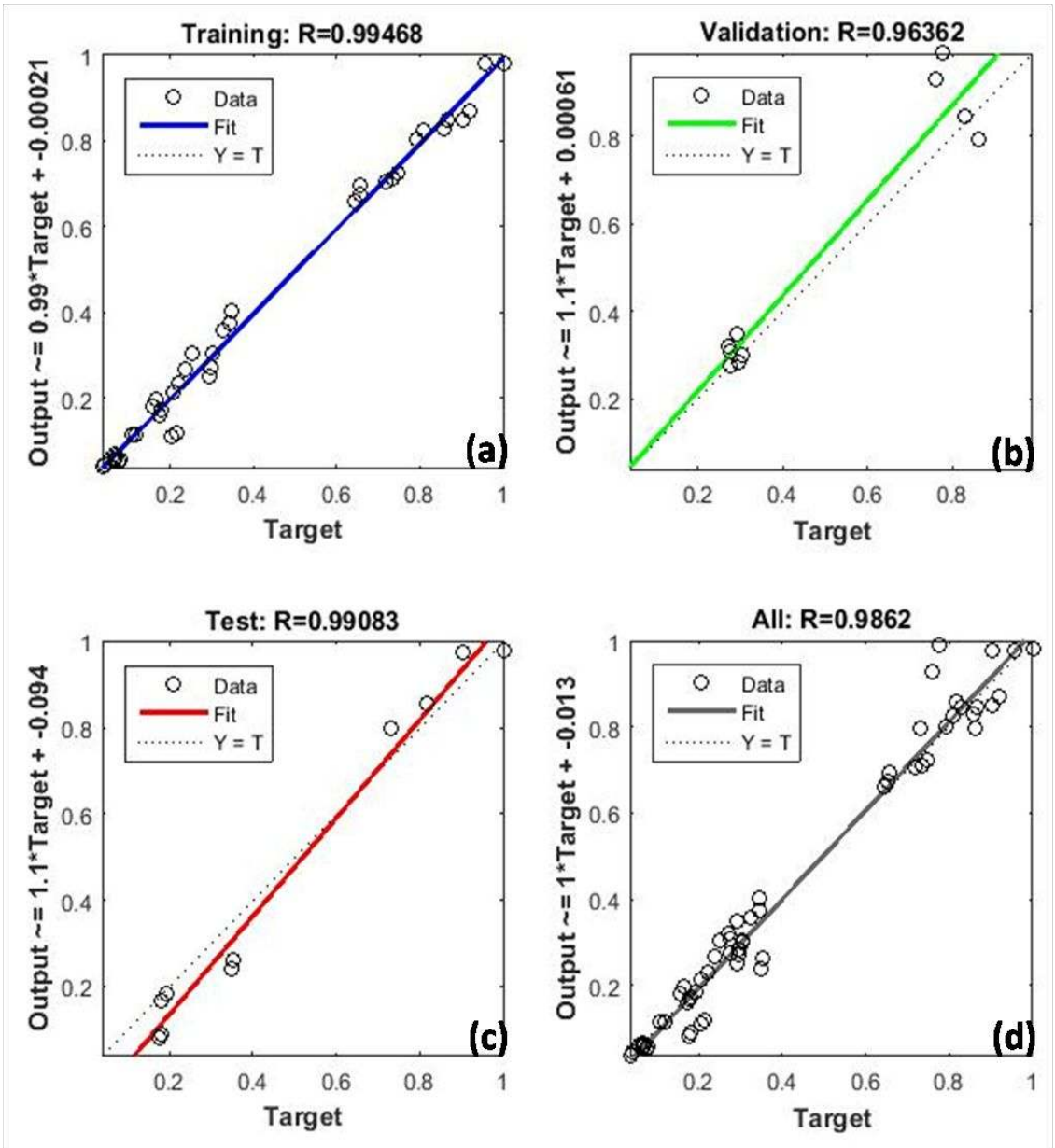


Figure 13. ANN model–III: prediction of $\Delta H^{\circ}(T\mathrm{m})$ and $\Delta S^{\circ}(T\mathrm{m})$ simultaneously, normalised regressions, a) training set, b) validation set, c) test set and d) overall.

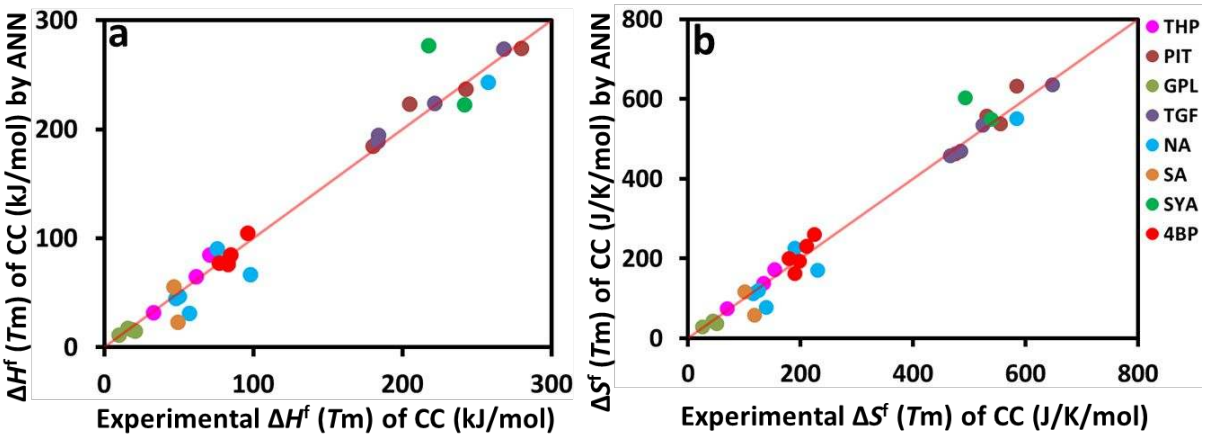


Figure 14. ANN model–III: quantitative overall performance to predict the $\Delta H^f(Tm)$ (a) and $\Delta S^f(Tm)$ (b) of the CCs.

Table 8. ΔH^f (Tm) and ΔS^f (Tm) values of CCs, obtained from ANN model–III.

Name of the CCs (ratio)	Exp ΔH^f_{CC} (Tm) (kJ/mol)	Predicted ΔH^f_{CC} (Tm) (kJ/mol)	Error (kJ/mol)	Exp ΔS^f_{CC} (Tm) (J/K/mol)	Predicted ΔS^f_{CC} (Tm) (J/K/mol)	Error (J/K/mol)
THP:GTA (1:1)	32.95	32.05	-0.9	69.71	74.10	4.39
THP:DIF (1:1)	61.57	64.68	3.11	134.45	137.38	2.93
THP:DICFA (1:1)	70.68	84.67	13.99	153.64	172.52	18.88
PIT:MYCT (1:1)	279.64	274.30	-5.34	584.94	632.37	47.43
PIT:LMDA (2:1)	179.99	184.42	4.43	475.98	461.80	-14.18
PIT:CTA (1:1)	204.44	222.85	18.41	530.81	556.39	25.58
PIT:LTA (1:1)	242.39	237.16	-5.23	555.75	537.20	-18.55
GPL:BA (1:1)	15.72	17.34	1.62	44.03	43.11	-0.92
GPL:4ABA (1:1)	18.42	16.22	-2.2	46.73	40.71	-6.02
GPL:4HBA (1:1)	20.92	14.68	-6.24	50.82	37.28	-13.54
GPL:FUA (2:1)	9.91	10.91	1	26.19	28.59	2.4
TGF:NA (1:1)	183.60	194.18	10.58	485.26	468.80	-16.46
TGF:PG (1:1)	183.14	188.90	5.76	465.83	457.00	-8.83
TGF:INA (1:1)	221.39	223.82	2.43	523.94	534.22	10.28
TGF:4HBA (1:1)	267.66	273.71	6.05	648.32	635.42	-12.9
NA:FFA (1:1)	57.07	31.01	-26.06	139.20	76.63	-62.57
NA:NFA (1:1)	48.06	44.69	-3.37	116.37	110.93	-5.44
NA:TFA (2:1)	98.11	66.64	-31.47	230.31	170.31	-60
NA:MFA (2:1)	50.06	46.81	-3.25	125.15	119.75	-5.4
NA:NPX (1:2)	75.78	90.13	14.35	189.57	225.62	36.05
NA:FBX (1:1)	257.16	242.92	-14.24	584.52	550.23	-34.29
SA:INZ (1:1)	49.35	22.82	-26.53	117.79	57.64	-60.15
SA:THP (1:1)	46.42	55.42	9	100.55	116.70	16.15
SYA:NA (2:1)	241.50	222.40	-19.1	539.19	548.29	9.1
SYA:URE (1:2)	217.43	276.87	59.44	493.37	602.68	109.31
4BP:FA (1:2)	83.10	79.89	-3.21	198.40	193.26	-5.14
4BP:NFA (1:2)	96.11	104.70	8.59	225.11	260.50	35.39
4BP:TFA (1:2)	77.09	77.04	-0.05	179.22	199.16	19.94
4BP:MFA (1:2)	83.10	75.92	-7.18	190.31	162.28	-28.03
4BP:FFA (1:2)	84.70	84.59	-0.11	211.14	231.00	19.86

ANN model–IV: prediction of *T*m of the CCs

In our previous studies,[31, 32] *T*m of CCs were predicted using eight parameters as input neurons in the input layer of the ANN model, including three properties for the MC and the CF each: MW, *T*m and binding energy. In addition, 1:1 binding energy of MC-coformer and the Δp*K*a value of the CCs where estimated and included. Herein, the accuracy of prediction of *T*m depends upon estimation of the MC-coformer binding energy. ANN model–IV in the present work is constructed based on six input neurons, the same used for the construction of ANN model–I, and ANN model–III. The training process was stopped with a regression value of 0.988 (overall), and the predicted *T*m values and obtained error values with respect to experimental values are listed in Table 9. To the best of our knowledge, ANN model–IV would be the best model to predict the *T*m of the CCs prior to the synthesis using only imformation for the pure MC and CF. Among the 30 CCs systems, the best prediction value is obtained for PIT with MYCT coformer (highlighted as green color in Table 9), whereas, the highest error is obtained for THP with DIF coformer (highlighted with red color in Table 9).

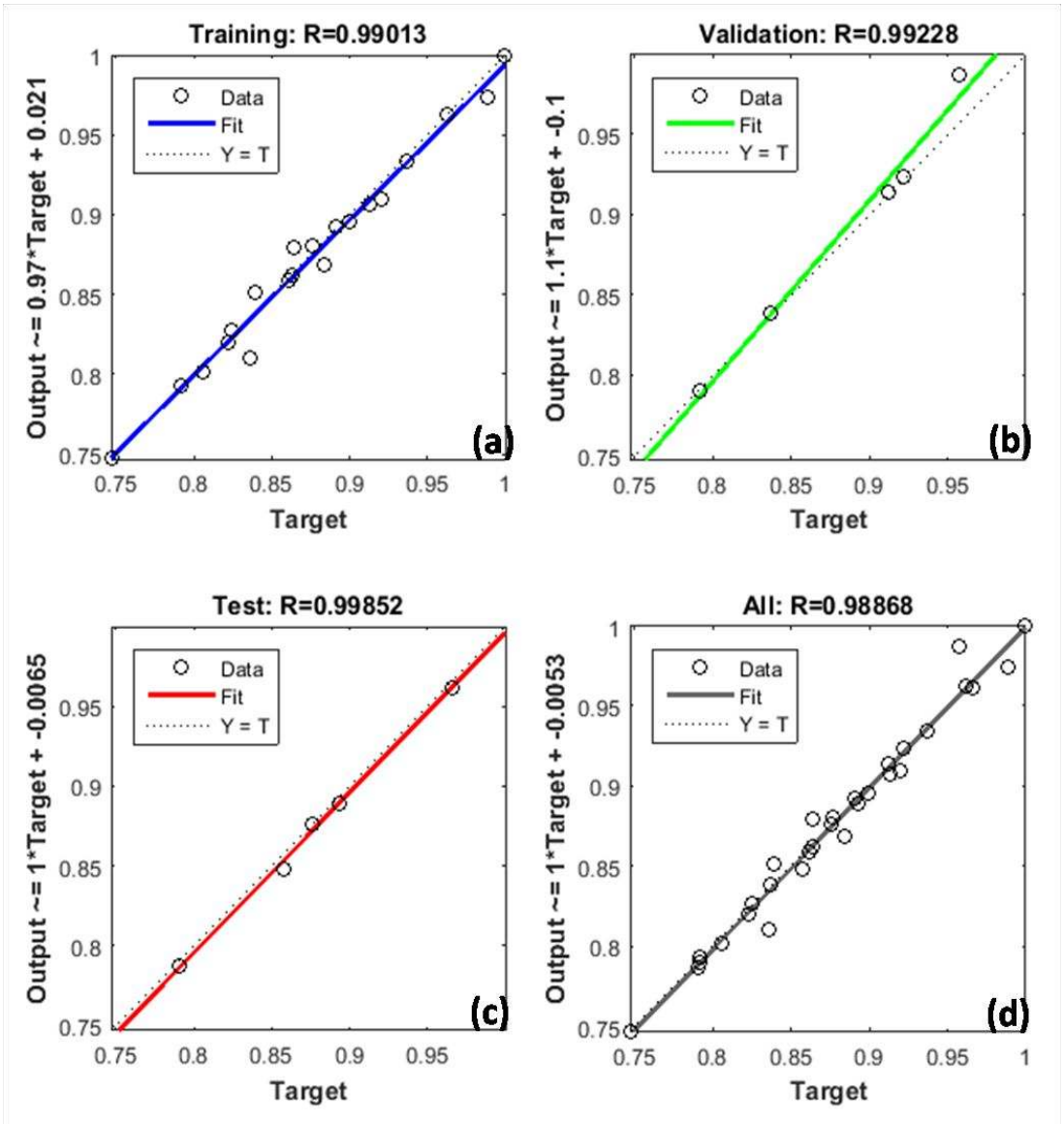


Figure 15. ANN model–IV: prediction of *T*m of CCs, normalised regressions, a) training set, b) validation set, c) test set and d) overall.

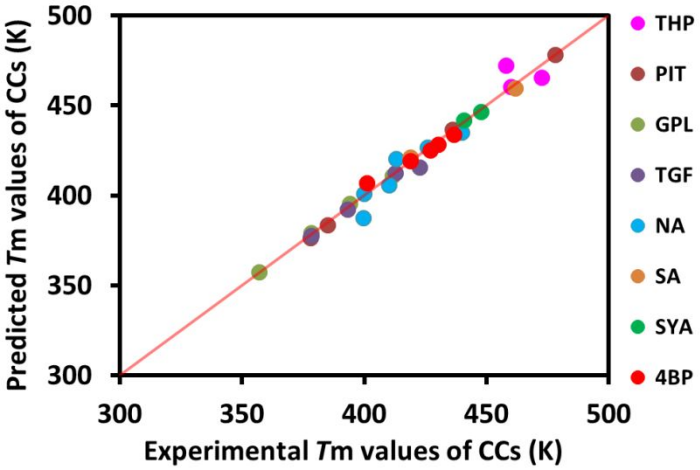


Figure 16. ANN model–IV: quantitative overall performance including all data. Color coding same as in previous figures.

Table 9. *T*m values of CCs, obtained from ANN model–IV.

Name of the CCs (ratio)	Melting point of CCs (*T*m) (K)	Predicted *T*m by ANN (K)	Error value (K)
THP:GTA (1:1)	472.65	465.468	-7.18
THP:DIF (1:1)	**457.95**	**471.9842**	**14.03**
THP:DICFA (1:1)	460.05	460.1663	0.12
PIT:MYCT (1:1)	**478.07**	**478.0557**	**-0.01**
PIT:LMDA (2:1)	378.15	376.3606	-1.79
PIT:CTA (1:1)	385.15	383.2209	-1.93
PIT:LTA (1:1)	436.15	436.65	0.50
GPL:BA (1:1)	357.05	357.1279	0.08
GPL:4ABA (1:1)	394.15	395.3591	1.21
GPL:4HBA (1:1)	411.65	410.6669	-0.98
GPL:FUA (2:1)	378.45	379.1764	0.73
TGF:NA (1:1)	378.35	377.6275	-0.72
TGF:PG (1:1)	393.15	391.9266	-1.22
TGF:INA (1:1)	422.55	415.3424	-7.21
TGF:4HBA (1:1)	412.85	412.1346	-0.72
NA:FFA (1:1)	410.00	405.5611	-4.44
NA:NFA (1:1)	413.00	420.1614	7.16
NA:TFA (2:1)	426.00	426.5006	0.50
NA:MFA (2:1)	400.00	400.8665	0.87
NA:NPX (1:2)	399.75	387.3084	-12.44
NA:FBX (1:1)	439.95	434.9624	-4.99
SA:INZ (1:1)	418.95	421.0602	2.11
SA:THP (1:1)	461.65	459.5735	-2.08
SYA:NA (2:1)	447.89	446.4457	-1.44
SYA:URE (1:2)	440.70	441.6315	0.93
4BP:FA (1:2)	418.85	419.0523	0.20
4BP:NFA (1:2)	426.95	425.0616	-1.89
4BP:TFA (1:2)	430.15	428.0209	-2.13
4BP:MFA (1:2)	436.65	433.5712	-3.08
4BP:FFA (1:2)	401.15	406.9332	5.78

In a related investigation,[43] a database analysis over melting points of CCs is presented and a correlation equation between the thermodynamic characteristics of CCs and those of the individual components is developed aiming to predict the melting point of the CCs. The method uses molecular descriptors as a basis, such as the molecular polarizability, sum of H-bond acceptor factors and sum of H-bond donor factors. Unfortunately, no correlation between the melting points of the CCs and the descriptors is observed.

ANN model–V: prediction of ideal mole fraction solubility of CCs

The calculated (eq 3) ideal mole fraction solubilities at 25 ℃ of the CCs varies from 0.4274 to $3.63x10^{-19}$ (see Table 3), and it wasn't possible to build a model for x as the output parameter. However, targeting ln(x) at 25 ℃ as the output parameter, the training process with 30 CCs systems using the same six input parameters as used in ANN models I, III and IV, has been done successfully. A very good model with a regression value of 0.990 and 0.9536 was obtained respectively for the training set, validation set and also for an overall performance of the model (R=0.980) as given in Fig. 17. The predicted ideal mole fraction solubility values by ANN model–V are compared with "calculated (eq 3) values" in Fig. 18. Obviously, in most cases the ideal solubility is well predicted by the model. The obtained deviations are listed in Table 10. Among the 30 CCs, the best prediction is obtained for PIT: MYCT cocrystal with only 0.01 deviation. The largest deviations are observed for some of the nicotinamide cocrystals.

The ideal solubility is entirely a property of the pure solid cocrystal, but as is shown by equation 2 is an important factor behind the real solubility in solution. As far as is known to us the ideal solubility has not been the target for prediction before, but there is previous work aiming for prediction of real cocrystal solubilities in solution.[44, 45, 46] Notably though these approaches involves input data for how the pure compounds dissolve in the same solvent, and accordingly are different to the work presented here both with respect to the complexity of the input data as well with respect to the property being the target of the prediction.

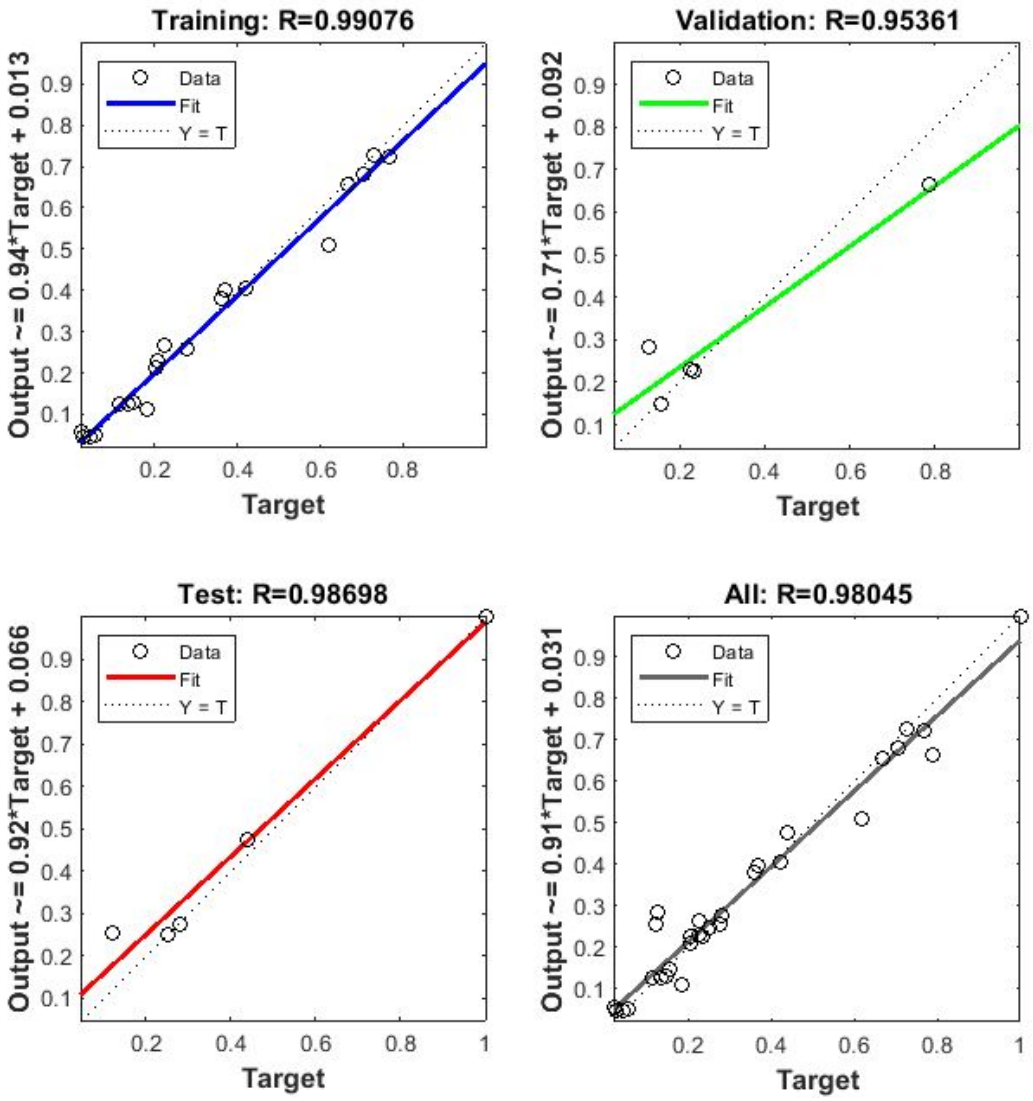


Figure 17. ANN model–V: prediction of ideal mole fraction solubility of CCs, a) training set, b) validation set, c) test set and d) overall.

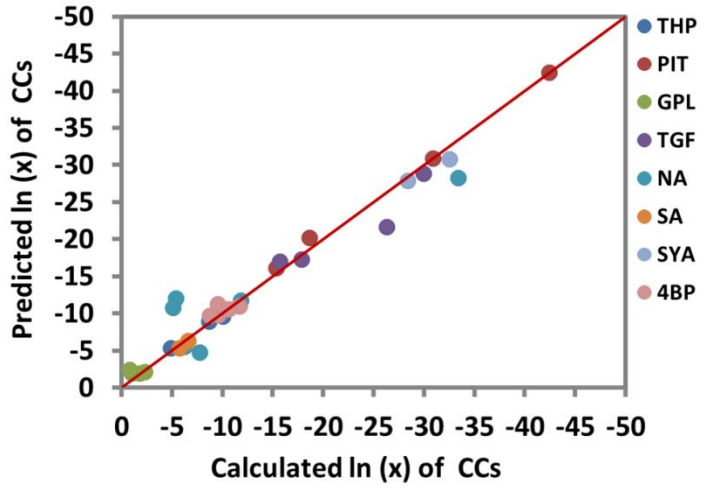


Figure 18. ANN model–V: quantitative overall performance including all data. Color coding same as in previous figures.

Table 10. Calculated ln (x_{id}) of CCs *vs* the predicted ln (x_{id}) of CCs obtained by ANN model–V.

Name of the CCs (ratio)	Calculated ln (x_{id}) of CCs	Predicted ln (x_{is}) of CCs	Error
THP:GTA (1:1)	-4.91	-5.28	-0.37
THP:DIF (1:1)	-8.67	-8.94	-0.27
THP:DICFA (1:1)	-10.03	-9.55	0.48
PIT:MYCT (1:1)	-42.46	-42.45	0.01
PIT:LMDA (2:1)	-15.36	-16.10	-0.74
PIT:CTA (1:1)	-18.63	-20.20	-1.57
PIT:LTA (1:1)	-30.94	-30.86	0.08
GPL:BA (1:1)	-1.05	-1.93	-0.88
GPL:4ABA (1:1)	-1.81	-1.93	-0.12
GPL:4HBA (1:1)	-2.33	-2.11	0.22
GPL:FUA (2:1)	-0.85	-2.43	-1.58
TGF:NA (1:1)	-15.7	-16.96	-1.26
TGF:PG (1:1)	-17.85	-17.23	0.62
TGF:INA (1:1)	-26.29	-21.62	4.67
TGF:4HBA (1:1)	-30	-28.87	1.13
NA:FFA (1:1)	-6.28	-5.49	0.79
NA:NFA (1:1)	-5.39	-12.03	-6.64
NA:TFA (2:1)	-11.88	-11.69	0.19
NA:MFA (2:1)	-5.14	-10.77	-5.63
NA:NPX (1:2)	-7.77	-4.73	3.04
NA:FBX (1:1)	-33.44	-28.23	5.21
SA:INZ (1:1)	-5.74	-5.33	0.41
SA:THP (1:1)	-6.63	-6.34	0.29
SYA:NA (2:1)	-32.57	-30.73	1.84
SYA:URE (1:2)	-28.37	-27.82	0.55
4BP:FA (1:2)	-9.66	-9.81	-0.15
4BP:NFA (1:2)	-11.7	-10.91	0.79
4BP:TFA (1:2)	-9.54	-11.26	-1.72
4BP:MFA (1:2)	-10.63	-10.53	0.10
4BP:FFA (1:2)	-8.77	-9.66	-0.89

Conclusions

Five ANN models have been developed to predict respectively: (I) the melting enthalpy, (II) the melting entropy, (III) the melting enthalpy and entropy, (IV) the melting temperature and (V) the ideal mole fraction solubility of co-crystals. Six operational variables were identified and used as being the most influential parameters for construction of all of five models, and includes molecular weight, melting point and melting enthalpy/melting entropy of the model compound and of the coformer. All five models gave the regression value between 0.986 and 0.999. Hence, the models are capable of predicting these physical properties of cocrystals prior to their synthesis. The models can serve as a tool for a first screening of coformers to be evaluated in the laboratory work, and thus make the process of finding suitable cocrystals more efficient. The models can also be used to access properties for CCs where the melting properties cannot be determined, in example when the cocrystal physically or chemically decomposes before reaching the melting point.

Acknowledgement

This publication has emanated from research supported in part by a research grant from Science Foundation Ireland (SFI), co-funded, under the European Regional Development Fund under Grant Number 12/RC/2275. G. R. K thanks DST-SERB for a Ramanujan fellowship (SB/S2/RJN-012/2017).

References

(1) Oliveira, M. A.; Peterson, M. L.; Davey, R. J. Relative Enthalpy of Formation for Co-Crystals of Small Organic Molecules. *Cryst. Growth Des*. **2011,** *11*, 449–457.

(2) Chan, H. C. S.; Kendrick, J.; Neumann, M. A.; Leusen, F. J. J. Towards ab initio screening of co-crystal formation through lattice energy calculations and crystal structure prediction of nicotinamide, isonicotinamide, picolinamide and paracetamol multi-component crystals. *CrystEngComm*. **2013,** *15*, 3799–3807.

(3) Taylor, C. R.; Day, G. M. Evaluating the Energetic Driving Force for Cocrystal Formation. *Cryst. Growth Des*. **2018,** *18*, 892−904.

(4) Kawakami, K. Reversibility of enantiotropically related polymorphic transformations from a practical viewpoint: thermal analysis of kinetically reversible/irreversible polymorphic transformations. *J. Pharm. Sci*. **2007,** *96*, 982–989.

(5) LaFountaine, J. S.; McGinity, J. W.; Williams III, R. O. Challenges and Strategies in Thermal Processing of Amorphous Solid Dispersions: A Review. *AAPS PharmSciTech*. **2016,** *17*, 43–55.

(6) Farahanchi, A.; Sobkowicz, M. J. Kinetic and process modeling of thermal and mechanical degradation in ultrahigh speed twin screw extrusion. *PolymDegradStab*. **2017,** *138*, 40–46.

(7) Gressl, C.; Brunsteiner, M.; Davis, A.; Landis, M.; Pencheva, K.; Scrivens, G.; Sluggett, G. W.; Wood, G. P. F.; Gruber-Woelfler, H.; Khinast, J. G.; Paudel, A. Drug–Excipient Interactions in the Solid State: The Role of Different Stress Factors. *Mol Pharm*. **2017,** *14*, 4560−4571.

(8) Crawford, D. E.; Miskimmin, C. K. G.; Albadarin, A. B.; Walker, G.; James, S. L. Organic synthesis by Twin Screw Extrusion (TSE): continuous, scalable and solvent-free. *Green Chem*. **2017,** *19*, 1507−1518.

(9) Crawford, D. E.; Casaban, J. Recent Developments in Mechanochemical Materials Synthesis by Extrusion. *Adv. Mater*. **2016***, 28*, 5747–5754.

(10) Crawford, D. E.; James, S. L.; McNally, T. Use of Batch Mixing To Investigate the Continuous Solvent–Free Mechanical Synthesis of OLED Materials by Twin-Screw Extrusion (TSE). *ACS Sustainable Chem. Eng*. **2018,** *6*, 193−201.

(11) Crowley, M. M.; Zhang, F.; Repka, M. A.; Thumma, S.; Upadhye, S. B.; Battu, S. K.; McGinity, J. W.; Martin, C. Pharmaceutical applications of hot-melt extrusion: part I. *Drug Dev. Ind. Pharm*. **2007,** *33*, 909–926.

(12) Liu, Y.; Thompson, M. R.; O'Donnell, K. P.; Ali, S. Heat assisted twin screw dry granulation. *AIChE Journal*. **2017,** *63*, 4748–4760.

(13) Liu, Y.; Bhandari, B.; Zhou, W. Glass Transition and Enthalpy Relaxation of Amorphous Food Saccharides: A Review. *J. Agric. Food Chem*. **2006,** *54*, 5701−5717.

(14) Zhou, D.; Zhang, G. G. Z.; Law, D.; Grant, D. J. W.; Schmitt, E. A. Physical Stability of Amorphous Pharmaceuticals: Importance of Configurational Thermodynamic Quantities and Molecular Mobility. *J. Pharm. Sci*. **2002,** *91*, 1863–1872.

(15) Svard, M.; Valavi, M.; Khamar, D–K.; Kuhs, M.; Rasmuson, Å. C.; Thermodynamic Stability Analysis of Tolbutamide Polymorphs and Solubility in Organic Solvents. *J. Pharm. Sci*. **2016,** *105*, 1901–1906.

(16) Solé, A.; Neumann, H.; Niedermaier, S.; Cabeza, L. F.; Palomo, E. Thermal Stability Test of Sugar Alcohols as Phase Change Materials for Medium Temperature Energy Storage Application. *Energy Procedia*. **2014,** *48*, 436–439.

(17) Maher, A.; Seaton, C. C.; Hudson, S.; Croker, D. M.; Rasmuson, Å. C.; Hodnett, B. K. Investigation of the Solid-State Polymorphic Transformations of Piracetam. *Cryst. Growth Des*. **2012,** *12*, 6223−6233.

(18) Espinell, J. R. H.; Lopez–Mejías, V.; Stelzer, T. Revealing Polymorphic Phase Transformations in Polymer–Based Hot Melt Extrusion Processes. *Cryst. Growth Des*. **2018,** *18*, 1995−2002.

(19) Svard, M.; Hjorth, T.; Bohlin, M.; Rasmuson, Å. C. Calorimetric Properties and Solubility in Five Pure Organic Solvents of N-Methyl-d-Glucamine (Meglumine). *J. Chem. Eng. Data.* **2016,** *61*, 1199−1204.

(20) Valavi, M.; Svard, M.; Rasmuson, Å. C. Improving Estimates of the Crystallization Driving Force: Investigation into the Dependence on Temperature and Composition of Activity Coefficients in Solution. *Cryst. Growth Des*. **2016,** *16*, 6951−6960.

(21) Schultheiss, N.; Newman, A. Pharmaceutical Cocrystals and Their Physicochemical Properties. *Cryst. Growth Des.* **2009,** *9*, 2950–2967.

(22) Batisai, E.; Ayamine, A.; Kilinkissa, O. E. Y.; Báthori, N. B. Melting point–solubility–structure correlations in multicomponent crystals containing fumaric or adipic acid. *CrystEngComm*. **2014,** *16*, 9992−9998.

(23) Wang, Z–Z.; Chen, J–M.; Lu, T–B. Enhancing the Hygroscopic Stability of S-Oxiracetam via Pharmaceutical Cocrystals. *Cryst. Growth Des*. **2012,** *12*, 4562−4566.

(24) Jutkus, R. A. L.; Li, N.; Taylor, L. S.; Mauer, L. J. Effect of Temperature and Initial Moisture Content on the Chemical Stability and Color Change of Various Forms of Vitamin C. *Int. J. Food Prop*. **2015,** *18*, 862–879.

(25) Tao, M.; Wang, Z.; Gong, J.; Hao, H.; Wang, J. Determination of the Solubility, Dissolution Enthalpy, and Entropy of Pioglitazone Hydrochloride (Form II) in Different Pure Solvents. *Ind. Eng. Chem. Res*. **2013,** *52*, 3036−3041.

(26) Grecu, T.; Hunter, C. A.; Gardiner, E. J.; McCabe, J. F. Validation of a Computational Cocrystal Prediction Tool: Comparison of Virtual and Experimental Cocrystal Screening Results. *Cryst. Growth Des*. **2014,** *14*, 165−171.

(27) Zhang, S–W.; Brunskill, A. P. J.; Schwartz, E.; Sun, S. Celecoxib–Nicotinamide Cocrystal Revisited: Can Entropy Control Cocrystal Formation?. *Cryst. Growth Des*. **2017,** *17*, 2836−2843.

(28) Ahuja, D.; Svärd, M.; Rasmuson, Å. C. Investigation of solid–liquid phase diagrams of the sulfamethazine–salicylic acid co-crystal. *CrystEngComm*. **2019,** *21*, 2863−2874.

(29) Nordström, F. L.; Rasmuson, Å. C. Determination of the activity of a molecular solute in saturated solution. *J. Chem. Thermodyn*. **2008,** *40*, 1684−1692.

(30) Valavi, M.; Svard, M.; Rasmuson, Å. C. Prediction of the Solubility of Medium–Sized Pharmaceutical Compounds Using a Temperature-Dependent NRTL-SAC Model. *Ind. Eng. Chem. Res*. **2016,** *55*, 11150−11159.

(31) Gamidi, R. K.; Rasmuson, Å. C. Estimation of Melting Temperature of Molecular Cocrystals Using Artificial Neural Network Model. *Cryst. Growth Des*. **2017,** *17*, 175–182.

(32) Gamidi, R. K.; Ukrainczyk, M.; Zeglinski, J.; Rasmuson, Å. C. Prediction of Solid State Properties of Cocrystals Using Artificial Neural Network Modeling. *Cryst. Growth Des*. **2018,** *18*, 133–144.

(33) Surov, A. O.; Voronin, A. P.; Manin, A. N.; Manin, N. G.; Kuzmina, L. G.; Churakov, A. V.; Perlovich, G. L. Pharmaceutical Cocrystals of Diflunisal and Diclofenac with Theophylline. *Mol. Pharmaceutics*. **2014,** *11*, 3707–3715.

(34) Zhang, S.; Chen, H.; Rasmuson, Å. C. Thermodynamics and crystallization of a theophylline–salicylic acid cocrystal. *CrystEngComm*. **2015,** *17*, 4125–4135.

(35) Viertelhaus, M.; Hilfiker, R.; Blatter, F. Piracetam Co–crystals with OH–Group Functionalized Carboxylic Acids. *Cryst. Growth Des*. **2009,** *9*, 2220–2228.

(36) Fábián, L.; Hamill, N.; Eccles, K. S.; Moynihan, H. A.; Maguire, A. R.; McCausl, L.; Lawrence, S. E. Cocrystals of Fenamic Acids with Nicotinamide. *Cryst. Growth Des*. **2011,** *11*, 3522–3528.

(37) Maheshwari, C.; André, V.; Reddy, S.; Roy, L.; Duarte, T.; Rodríguez–Hornedo, N. Tailoring aqueous solubility of a highly soluble compound via cocrystallization: effect of coformer ionization, pHmax and solute–solvent interactions. *CrystEngComm*. **2012,** *14*, 4801–4811.

(38) Aitipamula, S.; Chow, P. S.; Tan, R. B. H. Crystal Engineering of Tegafur Cocrystals: Structural Analysis and Physicochemical Properties. *Cryst. Growth Des*. **2014,** *14*, 6557−6569.

(39) Neurohr, C.; Marchivie, M.; Lecomte, S.; Cartigny, Y.; Couvrat, N.; Sanselme, M.; Subra–Paternault, P. Naproxen–Nicotinamide Cocrystals: Racemic and Conglomerate Structures Generated by CO2 Antisolvent Crystallization. *Cryst. Growth Des*. **2015,** *15*, 4616−4626.

(40) Thipparaboina, R.; Mittapalli, S.; Thatikonda, S.; Nangia, A.; Naidu, V. G. M.; Shastri, N. R. Syringic Acid: Structural Elucidation and Co–crystallization. *Cryst. Growth Des*. **2016,** *16*, 4679−4687.

(41) Wittering, K. E.; Agnew, L. R.; Klapwijk, A. R.; Robertson, K.; Cousen, A. J. P.; Cruickshank, D. L.; Wilson, C. C. Crystallisation and physicochemical property characterisation of conformationally-locked co-crystals of fenamic acid derivatives. *CrystEngComm*. **2015,** *17*, 3610−3618.

(42) Aakeröy, C. B.; Forbes, S.; Desper, J.; Altering physical properties of pharmaceutical co-crystals in a systematic manner. *CrystEngComm*. **2014**, *16*, 5870−5877.

(43) Perlovich, G. L.; Thermodynamic characteristics of cocrystal formation and melting points for rational design of pharmaceutical two–component systems. *CrystEngComm*. **2015**, *17*, 7019–7028.

(44) D. J. Good.; N. Rodríguez-Hornedo.; Cocrystal Eutectic Constants and Prediction of Solubility Behavior. *Cryst. Growth Des*. **2010**, *10*, 1028–1032.

(45) L. Lange.; S. Heisel.; G. Sadowski.; Predicting the Solubility of Pharmaceutical Cocrystals in Solvent/Anti–Solvent Mixtures. *Molecules*. **2016**, *21*, 593.

(46) A. Avdeef.; Cocrystal Solubility Product Prediction Using an in combo Model and Simulations to Improve Design of Experiments. *Pharm Res.* **2018**, *35*, 40.

For table of contents use only

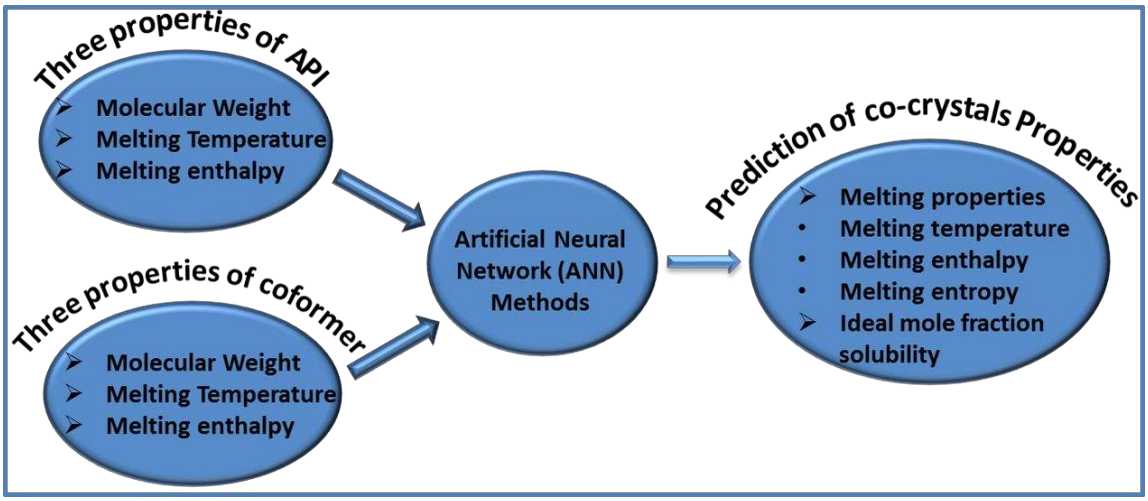
Analysis and ANN Prediction of Melting Properties and Ideal Mole fraction Solubility of Co-crystals

Rama Krishna Gamidi [ab] and Åke. C. Rasmuson*[a]

Rama Krishna [gr_krishna2005@yahoo.co.in]

a Department of Chemical and Environmental Science, Synthesis and Solid State Pharmaceutical Centre, Bernal Institute, University of Limerick, Limerick, Ireland

b Organic Chemistry Division, CSIR-National Chemical Laboratory, Dr. Homi Bhabha Road, Pune, Maharashtra 411 008, India



Different Artificial Neural Network models have been developed and examined for prediction of cocrystal properties based on pure component physical properties only. From the molecular weight, melting temperature, melting enthalpy and melting entropy of the pure compounds, the corresponding melting properties of the cocrystals and the cocrystal ideal solubility have been successfully predicted. The training process was stopped for all the five models with overall r2 values above 0.986. In particular, the models capture how the coformer structure influences on the targeted physical properties of cocrystals.