Journal of Molecular Graphics and Modelling 68 (2016) 23–28

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/JMGM

Efficacy of bi-component cocrystals and simple binary eutectics screening using heat of mixing estimated under super cooled conditions

Piotr Cysewski

Chair and Department of Physical Chemistry, Pharmacy Faculty, Collegium Medicum of Bydgoszcz, Nicolaus Copernicus University in Toruń, Kurpińskiego 5, 85-950 Bydgoszcz, Poland

ARTICLE INFO

Article history:
Received 29 March 2016
Received in revised form 6 June 2016
Accepted 7 June 2016
Available online 8 June 2016

Keywords:
Cocrystals
Simple binary eutectics
Theoretical screening
Heat of mixing
Confusion matrix
COSMO-RS
COSMOtherm

ABSTRACT

The values of excess heat characterizing sets of 493 simple binary eutectic mixtures and 965 cocrystals were estimated under super cooled liquid condition. The application of a confusion matrix as a predictive analytical tool was applied for distinguishing between the two subsets. Among seven considered levels of computations the BP-TZVPD-FINE approach was found to be the most precise in terms of the lowest percentage of misclassified positive cases. Also much less computationally demanding AM1 and PM7 semiempirical quantum chemistry methods are likewise worth considering for estimation of the heat of mixing values. Despite intrinsic limitations of the approach of modeling miscibility in the solid state, based on components affinities in liquids under super cooled conditions, it is possible to define adequate criterions for classification of coformers pairs as simple binary eutectics or cocrystals. The predicted precision has been found as 12.8% what is quite accepted, bearing in mind simplicity of the approach. However, tuning theoretical screening to such precision implies the exclusion of many positive cases and this wastage exceeds 31% of cocrystals classified as false negatives.

1. Introduction

New solid state materials obtained as a result of cocrystallization [1,2] become increasingly more attractive [3] both due to deepening insight into the fundamental nature [4–6] of intermolecular contacts and rapidly increasing range of practical applications [7,8]. The former aspect inspired many theoretical approaches [9–22] for cocrystal screening and the latter resulted in synthesis of a variety of materials [23–28]. Indeed, there were proposed many theoretical approaches for prediction of miscibility in the solid state. In general they can be classified into two broad categories depending on the considered crystal lattice details. One group of methods, sometimes termed liquid-based approaches, completely ignores the information of molecules arrangements in the crystal lattice while the other class crucially relying on such details includes explicitly the crystal packing motifs [12,13]. The former have the advantage of relatively inexpensive scans of potential cocrystallization landscape, while the latter, assuming in advance the possible crystal structures, screen out low probable solutions. Unfortunately, this approach, as belonging to general group of theoretical models [11] utilized in crystal engineering [29], is quite computationally demanding.

On the other hand, it is often surprising how efficient indirect approaches can be in offering valuable predictions at modest computational costs. For example, the information of the shape of electrostatic potential surface of the molecule [21] can be used for identification of the most likely contacts between components [17]. The mixing enthalpy of super cooled liquid coformers with a given stoichiometry can be used for screening of cocrystallization potential [15,18,19]. Also Hansen solubility parameters [30] can be applied for cocrystals screening [14], taking advantage only of the knowledge of the chemical structure of interacting components. Besides, the supramolecular phenomena [10] expressed in terms of homo- or heterosynthons [31] proved to be a valuable guidance for practical applications [9]. Alternatively, the semiquantitative models for predicting cocrystallization probability were formulated in terms of statistical analysis distributions [16] of a variety of molecular descriptors [20,22].

The plethora of methods developed for cocrystal screening is justified by the significance of the obtained new materials. Especially in the pharmaceutical domain the modulation of many physical properties of such new formulations is appreciated [32,33]. The multicomponent pharmaceutical cocrystals involving

E-mail address: piotr.cysewski@cm.umk.pl

http://dx.doi.org/10.1016/j.jmgm.2016.06.003
1093-3263/

active pharmaceutical ingredients (API) are often characterized by bioavailability modulation for increased solubility [34,35], alteration of dose response [34], enhancement of mechanical properties [36], improved hygroscopicity [26,37], affecting also the stability [27,28] and prolonged life shelf [25,34]. On the other hand, systems that are not miscible in the solid state are also interesting. For example, one of the way of administering drugs is transdermal dosage [38,39] of deep eutectic mixtures obtained by mixing of solid components. Recently, great potential of eutectic mixtures in extraction processes [40,41] was appreciated, demonstrating the usefulness of new sustainable solvents in a variety of extraction techniques. Also scientific and technological activities within energy storage domain [42,43] take advantage of a variety of eutectic mixtures.

Keeping in mind the importance of both types of systems, the aim of this paper is related to precision of quantification of theoretical cocrystals screening performed with an aid of the distributions of excess heat computed for binary liquid system mixed under super cooled conditions. To author's best knowledge such a validation was not offered so far, despite the fact that this particular method was used for pharmaceutical cocrystal predictions [15,18,19]. To achieve such quantification the criteria of classification of coformers pairs as miscible or immiscible in the solid state were formulated and tested. The actual validation was done against an extended set of training binary systems belonging to either experimentally observed cocrystals or simple binary eutectics systems documented by corresponding phase diagrams.

2. Methods

2.1. Confusion matrix approach

In predictive analytics [44] a contingency table [45] is defined based on a 2×2 confusion matrix. It offers a way of analyzing uncertainties in prediction performance of a classifier by enumerating items coming from the two classes. Since in the most classification problems the considered groups of positives and negatives overlap, the cut off criterion is to be adjusted for optimizing misclassified cases. In general, one can report the number of cases categorized as either true positives (TP), true negatives (TN), false positives (FP) or false negatives (FN). There are many different parameters, which can be used for quantifying the classification efficacy and among them the Matthews correlation coefficient [46] has the advantage of not being related to the difference in a size of the data sets. The following formula quantifies this parameter:

$$MCC = \frac{TP \cdot TN - FP \cdot FN}{\sqrt{(TP+FP)\cdot(TP+FN)\cdot(TN+FP)\cdot(TN+FN)}}$$

The value of MCC, as the whole contingency table, is crucially related to the classification criterion. Here, the dataset that is to be classified is simply a collection of negative and positive cases of cocrystallization trials. The former were identified as simple binary eutectic systems based on patterns of the phase diagrams found in the literature. The formation of such simple eutectic mixtures is typically confirmed by the presence of a single eutectic point suggesting the lack of molecular complex formation. Hence, the negative counterpart of the bi-component database comprises pairs immiscible in the solid state. The positive part collects cases of cocrystals and was taken either from CSD [47] or from literature reports if cocrystals were confirmed but their structures were unsolved. However, not all known binary cocrystals were included from the set of 7688 records of deposited in CSD. Only such systems were considered, which were formed by one of the compounds involved in pairs assigned to the negative counterpart of the database. Although this limits the number of potential cocrystals, it makes the two counterparts more coherent. For this project there were considered 493 simple eutectic mixtures and 965 cocrystals. The detailed definitions of all binary systems used in this paper are provided in Supporting materials in Tables S1–S3. The last one compiles all chemicals involved in pairs enumerated in Tables S1 and S2, which collect eutectics and cocrystals, respectively. For each pair the values of mixing enthalpy, H_{mix}, were computed using COSMOtherm [48] program and were used as a classification criterion within the confusion matrix formalism. It is expected that the related trends of Matthews correlation coefficient show a maximum and provide a clue for the most accurate separation of considered cases. The MCC returns a value within the range <−1,+1> and stands for a correlation coefficient between the observed and predicted binary classifications. It is worth mentioning that MCC is a contingency matrix method of calculating the Pearson product-moment correlation coefficient and has essentially the same interpretation [45]. In the case of −1 value the total disagreement between prediction and observation is obtained. On the contrary, the +1 value represents a perfect match. Finally, a zero denotes no better than random prediction. Hence, the maximization of MCC can be an useful criterion for quantification of theoretical approaches used for cocrystal screening based for example on excess heat computations.

Alternatively, one can estimate the percentage of true (TP%) or false (FP%) positive cases included in the prediction at a given value of selection criterion. In an ideal situation all true positive cases and none of negative ones should be included in cocrystal or simple eutectics subsets. However, due to similarities between many pairs they might be misclassified. As a measure of such improper grouping a falsehood parameter can be defined according to the following formula:

$$FA\% = \sqrt{(100\% - TP\%)^2 + FP\%^2}$$

It can be interpreted as a distance between TP% or FP% plots computed between points for a given value of excess enthalpy. It sums up both the misclassified cases and the ones omitted in the final prediction. The lower the FA% values the more precise classification. Interestingly, FA% trends exhibit a minimum for a particular value of H_{mix} and are typically very close to the maximum of MCC. Hence, the characteristics of such extreme points can be used for quantification of accuracy and wastage of different theoretical approaches. Finally, from the perspective of practical cocrystals screening, one can expect that minimization of FP% can also be considered as an important factor characterizing the quality of classification.

2.2. Computational details

The classification of a given pair of coformers either as cocrystals or simple eutectics is done based on the values of heat of mixing. To obtain these data typically four types of computations are indispensable. In the initial step the conformational analysis is used for searching of all thermodynamically important structures. This was done using Conformers module in Material Studio 8.0. The Boltzmann jump search method was used for collecting 200 conformers, which were further optimized in COMPASSII forcefield with Gasteiger atomic charges and ultra-fine quality options. Conformers were classified according to their total energy and RMS difference of Cartesian coordinates. After selection of no more than 10 dominant structures their geometries were fully optimized in a gas phase and in a polar medium using conductor-like screening model (COSMO) [49], in which the solvent is simplified as a polarized continuum with an infinite permittivity. In the final stage, the obtained geometries were used for σ-profiles generation. Because the systematic errors yielded by different quantum chemical methods can affect the final accuracy of the pairs classification, the adequacy of utilized theoretical approaches were tested