Solvates, Salts, and Cocrystals: A Proposal for a Feasible Classification System

E. Grothe,[†] H. Meekes,[†] E. Vlieg,[†] J. H. ter Horst,[‡] and R. de Gelder*,[†]

[†]Institute for Molecules and Materials, Radboud University, Heyendaalseweg 135, 6525 AJ Nijmegen, The Netherlands

[‡]EPSRC Centre for Innovative Manufacturing in Continuous Manufacturing and Crystallisation, Strathclyde Institute of Pharmacy and Biomedical Sciences, Technology and Innovation Centre, University of Strathclyde, 99 George Street, Glasgow G1 1RD, United Kingdom

S *Supporting Information*

ABSTRACT: The design of pharmaceutical cocrystals has initiated widespread debate on the classification of cocrystals. Current attempts to classify multicomponent crystals suffer from ambiguity, which has led to inconsistent definitions for cocrystals and for multicomponent crystals in general. Inspired by the work of Aitipamula et al. (*Cryst. Growth Des.* **2012**, *12*, 2147−2152), we present a feasible classification system for all multicomponent crystals. The present classification enables us to analyze and classify multicomponent crystal structures present in the Cambridge Structural Database (CSD). This reveals that all seven classes proposed are relevant in terms of frequency of occurrence. Lists of CSD refcodes for all classes are provided. We identified over 5000 cocrystals in the CSD, as well as over 12 000 crystals with more than two components. This illustrates that the possibilities for alternative drug formulations can be increased significantly by considering more than two components in drug design.
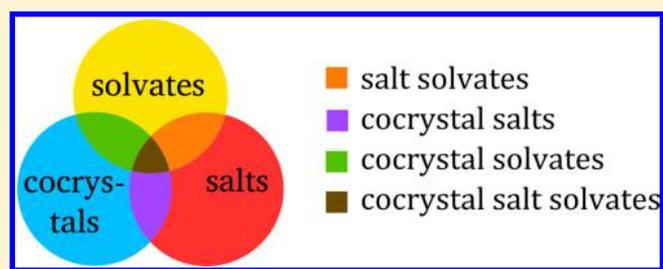
INTRODUCTION

The design of multicomponent crystals, such as solvates or salts, provides a means to alter the physicochemical properties of crystals without changing the chemical properties of the molecule of interest. This is particularly useful in the case of active pharmaceutical ingredients (APIs). For example, sulindac, a non-steroid anti-inflammatory drug, can be crystallized with different solvents to alter its dissolution rate.[1] Other APIs may exhibit increased water solubility as a salt, such as the sodium salt of naproxen.[2] Although solvates and salts are commonly used in this context, the formation of cocrystals has only recently been considered for APIs, which has significantly increased the solid forms available for formulation.[3] The number of solvents and counterions that can safely be included in pharmaceuticals is limited; safe coformers, on the other hand, are plentiful:[4] many are mentioned in the generally recognized as safe list (GRAS),[5] which lists hundreds of compounds, and even more are found in the everything added to food in the U.S. list (EAFUS),[6] which lists thousands of compounds suitable as food additives. Both of these lists are managed by the U.S. Food and Drug Administration (FDA).

In December 2011, the FDA announced a draft guidance on classification of cocrystals as "dissociable API-excipient molecular complexes".[7] This draft guidance provoked a response from the academic world in a paper urging for an alternative regulatory classification.[8] In the response, titled *Polymorphs, salts and cocrystals: what's in a name?*, Aitipamula et al. challenged the regulatory classification by proposing the merging of cocrystals with salts. They argue that the two should not be subject to separate sets of rules and regulations because cocrystals and salts are sometimes difficult to distinguish and because coformers, similar to ions, often act more like active ingredients than like excipients.

Apart from the many opinions and pharmaceutical issues that are addressed by the FDA, the industry, and academia, a practical question is how relevant the various classes are in terms of the number of crystal structures and how to use crystallographic databases in order to classify entries on the basis of concrete rules. This is the main goal and challenge of the present article.

PREVIOUS CLASSIFICATION

The *scientific* classification of multicomponent crystals into solvates, salts, and cocrystals is presented in Figure 1a. The aggregation states mentioned below refer to the pure un-ionized residues at room temperature. According to this classification, solvate crystals contain a solid residue and a liquid residue; salts contain two ions, one of which is a solid; and cocrystals contain two solid residues. These classes are not mutually exclusive and yield seven subclasses: the true solvate (yellow), salt (red), and cocrystal (blue) and the classes at the intersections of these circles (e.g., solvated cocrystal, green).
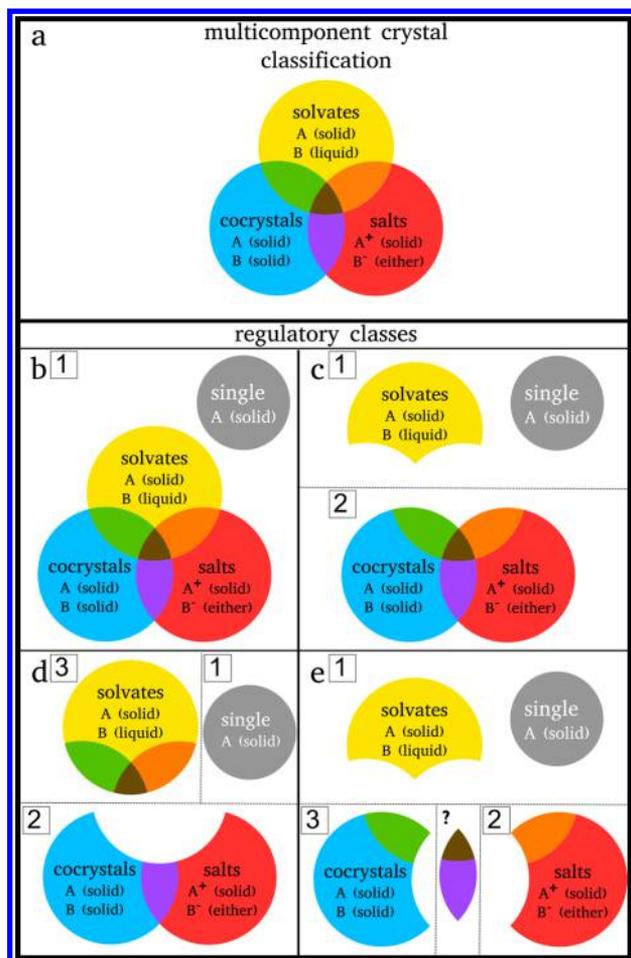
Figure 1. (a) Three overlapping multicomponent crystal classes, visualized as circles, adapted from Aitipamula et al.;[8] (b–e) dotted lines separate different regulatory classes proposed by[8] (b) one regulatory class for all solids; (c) two regulatory classes: one single-components and solvates and one remainder class; (d) three regulatory classes: one single components, one cocrystals and salts, and one solvates class (including cocrystal and salt solvates); and (e) FDA regulatory classes.

Aitipamula et al. suggest three alternative *regulatory* classifications in which cocrystals and salts belong to the same regulatory class. These three alternative regulatory classifcations are described as follows in the original paper:

(1) One class where all solid forms of APIs are classified together.

(2) Two classes: (1) single-component APIs, their polymorphs, and solvates and (2) salts, cocrystals, and their polymorphs or solvates.

(3) Three classes: (1) single-component APIs and their polymorphs; (2) salts, cocrystals, cocrystals of salts, binary salts, and their polymorphs; and (3) solvates and hydrates of single-component or multicomponent APIs and their polymorphs.

These classifications are presented in Figure 1b–d, respectively, where the different subclasses, represented by colored shapes, are combined in different regulatory classes, indicated by the black dotted lines. In spite of the recommendations by Aitipamula et al, the FDA chose not to alter the existing regulatory classification of solvates and salts[9] but to create a third class, "cocrystals"; this arrangement is

depicted in Figure 1e. The three FDA regulatory classes apply to (1) single-component crystals and their (pseudo)-polymorphs, (2) salts and their (pseudo)polymorphs, and (3) cocrystals and their (pseudo)polymorphs. (The FDA uses the term polymorph for pseudopolymorphs, so solvates are considered polymorphs in their guidances.) The FDA classification does not specify the overlap region of cocrystals and salts.

A comprehensive classification should be unambiguous, complete, and describe several mutually exclusive classes. The approach of Aitipamula et al. is consistent with current scientific thought and also a good starting point for that. In Figure 1a, however, the classes are not unambiguously defined and rules for classification are not proposed. For example, is it not clear how a crystal consisting of A^+ (solid), B^- (solid or liquid), and C (liquid) should be classified. If we want a clear and complete classification and accompanying nomenclature, then we have to expand the definitions of the three circles and define each of the seven subclasses created by the overlap of three circles according to the principles of set theory. Since the interest in multicomponent crystallization has shifted from fundamental research to a widely studied field in academia and industry, a proper classification of the different crystals of interest is important. It allows for proper regulatory classification and enables targeted exploration of classes that are promising for the optimization of APIs. Therefore, we propose concrete rules for the classification of multicomponent crystals and have applied this classification to the Cambridge Structural Database (CSD). Pharmaceutical salts and solvates represent two well-studied classes, in contrast to other multicomponent classes. The application to the CSD allows us to estimate the potential impact of these other multicomponent classes not only for cocrystals but also for crystals with more than two residues: If the number of components in a crystal is not restricted to one or two, then this will further increase the posibilities for alternative solids of APIs. In the next section of this article, we propose strict definitions for the seven subclasses found in Figure 1 and highlight issues that inevitably emerge. We then describe our approach to apply this classification to the CSD, followed in the Results and Discussion section by an example from the CSD for each subclass, up to a crystal classified as a cocrystal salt solvate. We also present the distribution of CSD crystals over the seven subclasses and provide lists of CSD refcodes for all subclasses.

■ CLASSIFICATION

Before exploring the multicomponent crystals in the CSD, we need to define all of the classes that we want to distinguish. For the classification to be unambiguous, all crystals must fall into exactly one subclass.

Multicomponent crystals can be defined as crystals with two or more different (different 2D structure, different elements, or different but not opposite chirality) residues in the crystal lattice. (A residue is considered to be a complete set of covalently bonded elements. In this article, covalent bonds will be defined by the connectivity records of the CSD datafile used.) We refer to the number of residues in the asymmetric unit as Z^R such that $Z^R > 1$ for multicomponent crystals and $Z^R = 1$ for single-component crystals.

We distinguish residues as either an **ion**, a **solvent**, or a **coformer**. These residue types are similar to the charged residue, the neutral liquid residue, and the neutral solid residue, respectively, that are used for classification by Aitipamula et al.

Some issues arise with the labeling of residues that we need to consider before defining our ion, coformer, and solvent residues. One issue is that of charge transfer: when is charge transfer considered significant? In principle, this question can be answered by performing an accurate charge density analysis, e.g., on the basis of X-ray diffraction or density functional calculations. In the CSD data used in this study, there is no concept of delocalized charge; charges are associated with specific atoms and are integral. This means that the issue of partial charge transfer can be solved in a pragmatic way for this study. We are aware of the fact that not all CSD entries were carefully analyzed with respect to charge transfer. Because only residues with a net overall charge are considered ions, zwitterions are considered coformers. On the other hand, the residues of ionic liquids would be considered ions, according to the above-mentioned standards, even though they are known to act as solvents.

Another issue is how to differentiate between solvents and coformers based on the crystallographic data. Before we define the two, we consider the definitions of true cocrystals found in scientific literature. The name *cocrystal* is mentioned in many different contexts, two of which stand out:

(1) Multicomponent crystals of neutral residues, individually solid at room temperature[10]
(2) Multicomponent crystals of neutral residues, not commonly used as solvents[11]

As an example, take a theophylline DMSO crystal (Figure 2) that was crystallized at a laboratory at MIT from a DMSO
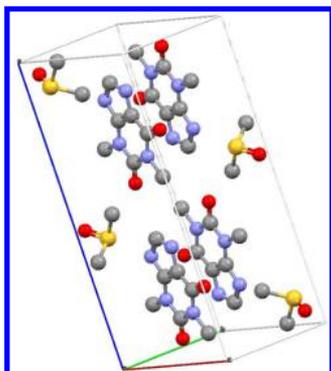
Figure 2. Unit cell of the theophylline DMSO solvate, from CSD (RIGYEM); hydrogens are omitted for clarity. DMSO is a borderline solvent, with a melting point of 19 °C.

solution and transported to a poorly heated classroom at the University of Calgary. Both above-mentioned cocrystal descriptions agree that this is a solvate since DMSO is a commonly used solvent with a melting point of 19 °C at atmospheric pressure and therefore a liquid in the lab. Now, regardless of the local room temperature in the Calgary classroom, the nature of the crystal is the same, even when pure DMSO may be solid in the colder class room and liquid in the lab. Hence, a strict definition is necessary to ensure that the solvate is always classified as such. We define a solvent as any neutral residue that is liquid under ambient conditions (T = 293.15 K, $P = 10^5$ Pa).

To summarize, our residue types are defined as follows.
ion: *a residue with a nonzero formal charge*
solvent: *a neutral residue that is liquid at ambient conditions*
coformer: *a neutral residue that is not a solvent*

Now, we can continue by defining the subclasses derived from these residue types. Our classification will follow the classification system proposed by Aitipamula et al. (Figure 1a): three main classes, named solvate, salt, and cocrystal, that form seven subclasses. Three of the subclasses belong to only one main class, three belong to two main classes, and one belongs to all three main classes. The seven subclasses will be called true salt, true solvate, true cocrystal, salt solvate, cocrystal solvate, cocrystal salt, and cocrystal salt solvate. First, however, we must expand the definitions of the main classes (solvate, salt, and cocrystal) because the classification scheme in Figure 1a is not complete.

This issue becomes evident when viewing the isonicotinamide (INA) pyromellitic acid hydrate in Figure 3: it contains an
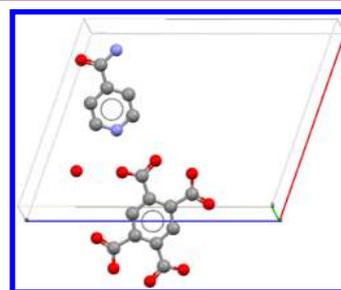
Figure 3. Asymmetric unit of protonated INA with a pyromellitic acid counterion and water, HEVPUU. This combination of a salt with a solvent molecule is not classifiable in the scheme presented in[8] Figure 1a.

ion pair (A^+ (solid), B^- (either)) and a solvent (C (liquid)). It seems intuitive to classify this as a salt solvate; however, according to Aitipamula et al, a solvate requires the presence of a neutral, solid residue (see Figure 1a). In fact, with their definitions, the combination of ions and a solvent is not classifiable. The same holds for other three component crystals, e.g., one coformer and two ions. In other words, the classification is not complete, and we are forced to reconsider the definitions of the main classes (salt, solvate, cocrystal) such that the INA crystal (Figure 3)—and actually *all* combinations of residues—can be classified, in line with set theory.

Therefore, we propose the following definitions of the main classes:

> **salt**: *a crystal containing at least two ions*
> **solvate**: *any crystal with a solvent molecule plus either a coformer or at least two ions*
> **cocrystal**: *a crystal with a coformer molecule plus either another coformer or at least two ions*

Table 1 shows the possible compositions of solvates, salts, and cocrystals in *main classes* on the basis of residue types ion, solvent, and coformer. The second part of the table, *subclasses*, summarizes the possible compositions of the seven subclasses that arise from the main classes. The seven subclasses are now defined as follows:

> **true solvate**: *one or more solvents and exactly one coformer, no ions*
> **true salt**: *only ions*
> **true cocrystal**: *only coformers*
> **salt solvate**: *one or more solvents and two or more ions, no coformers*
> **cocrystal solvate**: *one or more solvents and two or more coformers, no ions*

cocrystal salt: *one or more coformers, two or more ions, no solvents*

**cocrystal salt solvate**: *one or more solvents, two or more ions, one or more coformers*

The classification procedure is now straightforward: the INA crystal example holds two ions, one solvent, and no coformers, which means that it falls into the salt main class as well as the solvate main class and thereby classifies as the salt solvate subclass. Using the number of each residue type, one can directly determine the subclass.

Table 1. Number of Each Residue Type for the Main Classes and for the Subclasses Proposed Here[a]

class		ion	solvent	coformer
main classes				
solvate { or		≥ 0 / ≥ 2	≥ 1 / ≥ 1	≥ 1 / ≥ 0
salt		≥ 2	≥ 0	≥ 0
cocrystal { or		≥ 0 / ≥ 2	≥ 0 / ≥ 0	≥ 2 / ≥ 1
subclasses				
true solvate		0	≥ 1	1
true salt		≥ 2	0	0
true cocrystal		0	0	≥ 2
salt solvate		≥ 2	≥ 1	0
cocrystal solvate		0	≥ 1	≥ 2
cocrystal salt		≥ 2	0	≥ 1
cocrystal salt solvate		≥ 2	≥ 1	≥ 1
single		0	0	1

[a]A single-component crystal is listed for completeness.

By redefining the terms solvate, salt, and cocrystals, we have solved ambiguities that were present in classification attempts so far. In Figure 4, the three circles that visualize the classification are broken down into the seven newly defined subclasses. It can be seen that the true solvate, true salt, and true cocrystal have a minimum of two residues, whereas three intersection classes require at least three residues, and the cocrystal salt solvate requires at least four residues.
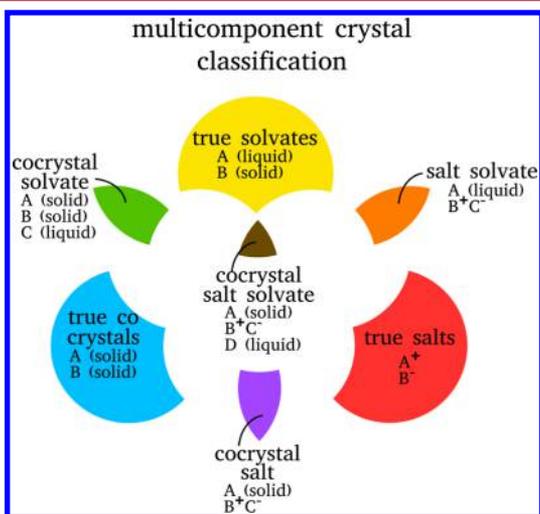
Figure 4. Visualization of the new classification scheme, inspired by the proposal of Aitipamula et al. Corresponding to Table 1, we see three main classes (represented by three colored circles) that create seven subclasses, which are visible as different colors.

Finally, it is noteworthy that in this approach the minimum requirements of a cocrystal salt, salt solvate, and cocrystal salt solvate are less than the sum of the corresponding true solvate, true salt, and true cocrystal requirements. We emphasize that the subclasses are created from the overlapping definitions of the solvate, salt, and cocrystal main classes.

■ APPLICATION TO THE CSD

Data Acquisition. ConQuest 1.17 was used to search the CSD 5.36 + update (February 2015). All organic entries containing 3D coordinates, no errors, and no disorder were included in our initial analysis. The reason for not including metal–organic crystals is that in these crystals the interaction between the metal and a nonmetal is treated as a covalent bond or as aromatic, π, or delocalized. Since we speak of a residue as a covalently bonded unit, this makes it difficult to assign elements of such entries to residues: sometimes a residue includes the metal, sometimes it does not; therefore, only organic entries were included. The refcode list of these entries was exported, as well as the Protein DataBase file (pdb) and the mol2 file (mol2).

Preprocessing. Following the presented classification rules, true polymorphs should be classified identically; therefore, we aim to include only one polymorph of each crystal. To this end, crystals with the same CSD refcode but different sequential number were excluded, and only the highest number was included in the data set. This procedure also excludes all but the newest redeterminations.

The atomic charges in the pdb file were replaced by the charges in the mol2 file (when exporting CSD entries to pdb using ConQuest, some information is lost and all charges are set to 0).

Processing. All residues, i.e., covalently bonded units, in a crystal were identified, and identical residues were defined as having identical connectivities, atom types, and atomic charges. All residues within one entry were indexed by our in-house software *Maruchi*. *Maruchi* distinguishes between connectivities of residues using an extended connectivity algorithm based on the Atomic Walk Counts.[12] This index is used to determine the number of unique (nonidentical) residues in a crystal, Z^R. Because the possible subclasses depend on Z^R, the results in the next section will be presented for $Z^R = 2$, $Z^R = 3$, and $Z^R \geq 4$ separately.

In each entry, the number of ions was determined by the number of residues with a net charge. All neutral liquid residues qualify as a solvent, but for practical purposes, we used a list of 80 known solvents, which are liquid under ambient conditions, 59 of which were published by Görbitz et al.[13] The CSD was searched for solvates of these 80 commonly used solvents in their un-ionized state using ConQuest. The number of coformers is determined by taking the total number of residues in an entry, as determined by *Maruchi*, and subtracting the number of ions and solvents (eq 1). Since the list of solvents does not include more exotic solvents, some residues are falsely identified as coformers.

$$n_{comp} = n_{cof} + n_{solv} + n_{ion} \tag{1}$$

After the number of each residue type has been determined for each entry, the multicomponent entries are labeled true solvate, true salt, true cocrystal, salt solvate, cocrystal solvate, cocrystal salt, or cocrystal salt solvate according to Table 1.

Out of the 759 627 CSD entries, 227 230 organic entries with 3D coordinates, no errors, and no disorder were studied

Table 2. Classification of Multicomponent Crystals in the CSD

	true solvate	true salt	true cocrystal	salt solvate	cocrystal solvate	cocrystal salt	cocrystal salt solvate	total
$Z^R = 2$	18 192 (33.4%)	18 422 (33.8%)	5685 (10.4%)					42 299 (77.6%)
$Z^R = 3$	937 (1.7%)	678 (1.2%)	88 (0.2%)	7164 (13.1%)	933 (1.7%)	1.114 (2.0%)		10 914 (20.0%)
$Z^R > 3$	32 (0.1%)	17 (0.0%)	1 (0.0%)	508 (0.9%)	104 (0.2%)	47 (0.1%)	583 (1.1%)	1292 (2.4%)
total	19 161 (35.2%)	19 117 (35.1%)	5774 (10.6%)	7.672 (14.1%)	1.037 (1.9%)	1.161 (2.1%)	583 (1.1%)	54 505 (100.0%)

here. The refcodes of the multicomponent CSD entries in this data set and the classes assigned can be found in the Supporting Information.

Postprocessing. Sometimes, 3D data of a solvent (i.e., coordinates and connectivity) are missing in a CSD entry, even when the presence of its structure is confirmed in 2D. In this case, n_{cof} cannot be correctly determined using eq 1. This inconsistency is especially common for water molecules, for which coordinates are often not determined. Because of the frequency of this problem for water and the simplicity of its molecular structure, we have checked and corrected for this problem in the case of water molecules.

Disorder and errors in the 3D data can cause erroneous classification, which is why we use ConQuest to include only crystals labeled as no disorder and no errors. Some entries in the CSD, however, do show disordered atoms or errors but are not labeled as such by ConQuest. Our data set will therefore contain some disorder and error entries. Disorder, partial absence of 3D data, and delocalized charges sometimes cause an erroneous classification by our software. We performed a targeted search on such entries and found that 339 entries were misclassified due to these data inconsistencies. We further ignored the rare and exotic cases of solvent solvates, that is, crystal structures composed of merely solvent residues.

A total of 146 $Z^R = 1$ crystals in true salts and true cocrystals were reassigned to the single-component class, 112 $Z^R = 2$ crystals in salt solvates and cocrystal solvates were reassigned to true solvates, 14 $Z^R = 2$ cocrystal salts were reassigned to true salt, and 67 $Z^R = 3$ cocrystal salt solvates were reassigned to the salt solvate class.

■ RESULTS AND DISCUSSION

The data set contains 54 505 organic multicomponent CSD entries (24.0% of all organic entries in the set), which were classified into seven subclasses, using our proposed scheme. Table 2 shows the distribution of crystals over the subclasses for different Z^R and for all multicomponent crystals. For each subclass, we will present one example from the CSD, all of which contain an INA residue. INA is a popular coformer[14−16] often used in cocrystallization screens. Its single-component crystal structure is shown in Figure 5a. All examples, except for the cocrystal salt solvate, are from the organic subset.

Crystals with Z^R = 2. Binary crystals ($Z^R = 2$) represent the majority of multicomponent crystals in the data set, with a total number of 42 299 (77.6%). These can only be true solvates, true salts, or true cocrystals, and the distribution along these is plotted in the histogram in Figure 7 (upper panel). The majority of binary crystals are a solvate (43.0%) or salt (43.6%); cocrystals represent one-eighth (13.4%) of binary crystals in the CSD. Figure 5b−d show binary INA crystals each of a different subclass. The solvate crystal (b) involves an OH···N interaction between acetic acid and the pyridine in INA (both residues are neutral).[17] The salt (d) is that of protonated INA and 3,5-dinitrosalicylate. The cocrystal of INA carbamazepine (c) has
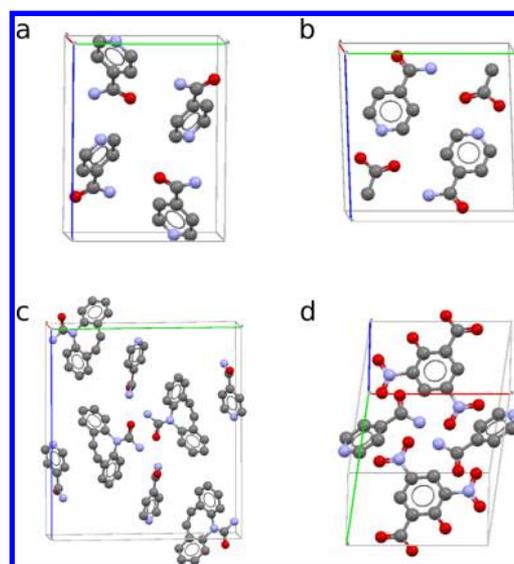

Figure 5. Single-component and three binary isonicotinamide (INA) crystals in the CSD. (a) Single-component crystal, EHOWIH05.[15] (b) True solvate: INA with acetic acid, JAWWAG.[17] (c) True cocrystal: INA with carbamazepine, LOFKIB.[18] (d) True salt: INA ion with 3,5-dinitrosalicylate, UQACIZ.[19]

two known polymorphs;[18] shown here is the stable variant of the two.

Crystals with Z^R = 3. The 10 914 ternary crystals ($Z^R = 3$) represent 20.0% of multicomponent entries. These crystals can be one of six classes: in addition to the three $Z^R = 2$ classes, there are salt solvates, cocrystal solvates, and cocrystal salts. The histogram in Figure 7 (middle panel) shows the distribution of ternary crystals along these six subclasses. The vast majority are salt solvates (65.6%), like the one shown in Figure 6a: an INA—pyromellitic acid ion pair with water. Figure 6b shows an INA—4,5-dichloro-phthalate ion pair with coformer 4,5-dichloro methyl phthalate, which is one of the 1114 cocrystal salts (10.2%) in the ternary subset. The fourth largest ternary class is the cocrystal solvate (8.5%), an example of which is shown in Figure 6c. This is an INA norfloxacin cocrystal pair with a chloroform solvent residue. Norfloxacin is a zwitterion, but it has no net charge and is therefore considered a coformer rather than an ion.

Crystals with Z^R ≥ 4. Crystals with four or more residues (Z^R ≥ 4) are more rare, making up less than 3% of the multicomponent crystal population. In addition to the six $Z^R = 3$ classes, these crystals can also be cocrystal salt solvates, which gives a total of seven possible subclasses for quaternary and higher order crystals. A cocrystal salt solvate contains all three residue types. The occurrence of each subclass for quaternary crystals is plotted in the histogram in Figure 7 (lower panel). The asymmetric unit of a cocrystal salt solvate is shown in Figure 6d, where INA is the coformer, the ion pair is formed by a positively charged zinc complex cation with a 1,5-naphtalene-
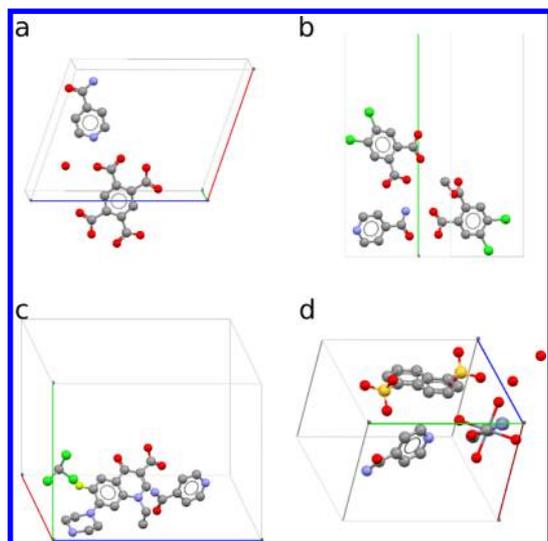
Figure 6. Four multicomponent isonicotinamide (INA) crystals in the CSD. (a) Salt solvate: INA^+, a pyromellitic acid counterion, water, HEVPUU.[20] (b) Cocrystal salt: INA^+, a phthalate counterion, a monomethyl-phthalic acid coformer, LORWOF.[21] (c) Cocrystal solvate: INA, norfloxacin, chloroform, VETVUM.[22] (d) Cocrystal salt solvate: a zinc complex, a 1,5-naphthalenedisulfonate counterion, an INA coformer, water, FUZGOX.[23]
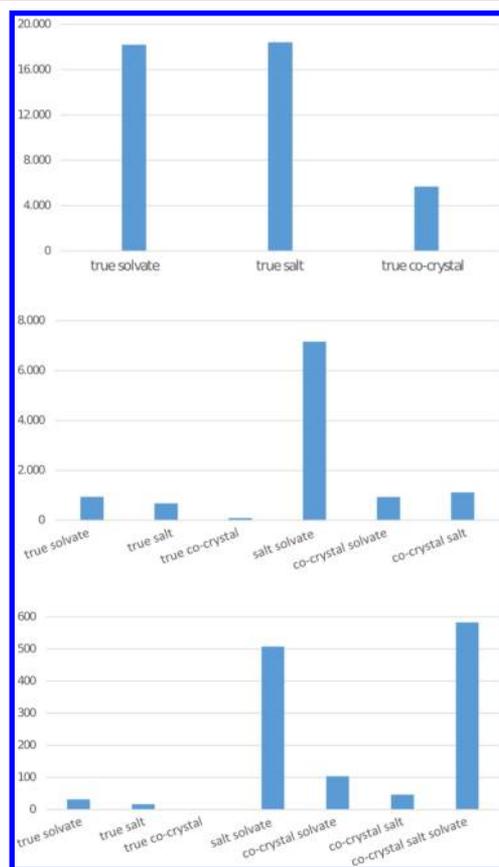
Figure 7. $Z^R = 2$, $Z^R = 3$, and $Z^R > 3$ distributions over the possible classes.

disulfonate anion (DNS), and the water molecules act as solvent. This is a metalorganic crystal, which is not part of the organic subset of the CSD, and is used for illustration purposes only. In the $Z^R \geq 4$ subset, most crystals are cocrystal salt

solvates (45.1%). Salt solvates are also abundant, 508 (39.3%), and cocrystal solvates, cocrystals salts, and true solvates occur frequently (104, 47, and 32 CSD entries, respectively), but true salts and cocrystals are rare, with only 17 and 1 quaternary crystals classified as such (both <2%).

Classification of All Multicomponent Crystals. The residue types (solvent, ion, coformer) in a multicomponent crystal determine its subclass according to Table 1. A clear distinction between solvents and coformers was applied on the basis of a list of 80 solvents that can be expanded. The distinction between ions and neutral residues is less straightforward, but it is overcome here by adopting the atomic charges provided by the CCDC. The ability to label residues uniquely allows us to unambiguously classify multicomponent crystals. In a recent paper,[24] Kelley et al. argue that a restrictive classification "will lead to misassumptions about physicochemical properties", especially in the case of crystals that contain both ionic and neutral residues or that show partial charge transfer. Our classification system deals with this issue by having subclasses specifically for these crystals of ionic and neutral residues: the cocrystal salt, salt solvate, and cocrystal salt solvate classes. The issue with partial charge transfer, however, is avoided rather than resolved.

With this classification system, we have successfully classified all multicomponent organic crystals in the CSD into seven subclasses that are mutually exclusive. All subclasses are relevant in terms of frequency of occurrence. We have found as many as 12 206 ternary and higher order crystals ($Z^R \geq 3$), very few of which are true solvates, salts, or cocrystals (1753). Classification attempts thus far have mainly dealt with binary crystals and thus only with true solvates, salts, and cocrystals. The high number of crystals with $Z^R \geq 3$ shows that multicomponent classification has to go beyond these traditional classes.

So far, most multicomponent API crystals are true solvates, true salts, or salt solvates. The number of acceptable counterions is in the range of 70 residues (excluding polymers), whereas the potential number of acceptable coformers is much higher: there are 304 substances on the GRAS lists and over 3000 substances on the EAFUS list. Therefore, the number of possible acceptable crystals (in terms of safety) for APIs increases considerably when coformers are considered for multicomponent crystallization. Table 2 shows that cocrystallization is already a commonly observed phenomenon: 5774 true cocrystals were found in the CSD (10.6%), indicating that a wide range of new possibilities exists for the pharmaceutical industry. Furthermore, cocrystal solvates, cocrystal salts, and cocrystal salt solvates become available by not limiting the amount of components in drug design to two salt solvates, totalling 19.2% of the data set studied here. In other words, the size of the subclasses at the intersection of the true classes (see Figure 1a) is significant, and combinations of ion pairs, solvents, and coformers are therefore viable options and can be considered for API crystallization.

The notion that cocrystals and salts are in some cases very similar to each other led Aitipamula et al.[8] to conclude that one regulatory class for cocrystals and salts should suffice. We agree that it is often difficult to make an unambiguous classification for a compound close to a border between two (sub)classes. However, while some cocrystals and salts are very similar, others are very dissimilar because the nature of interactions in cocrystals can differ greatly from the ionic interaction of salts. This is the main reason to choose two separate regulatory

classes, like the FDA did. Moreover, we see from Table 2 that one cocrystal exists for every three salts.

The results in this article were obtained using a biased collection of data, the CSD. The CSD does not represent a random subset of all possible crystals but, rather, a biased subset where researchers chose (combinations of) residues to study and publish. As a result, the collection of crystals in the CSD is affected not only by the limitations of crystallization but also by current scientific and commercial interests. Also, we note again the additional but limited bias in the cocrystal class, which will contain some false positives that are actually solvates, when a solvent is not on the used list of known solvents.

CONCLUSIONS

We have defined rules to come to a feasible classification system for multicomponent crystals, inspired by the earlier work of Aitipamula et al.[8] The present classification entails seven mutually exclusive subclasses and associated rules, based on the residue types in the crystal, thereby resolving the problem of overlapping classes present in the classification of Aitipamula et al. Testing the proposed classification on 54 505 multicomponent entries from the CSD showed a world of opportunities for multicomponent API crystallization: 5774 cocrystals were identified, as were 2781 cocrystal salts, cocrystal solvates, and cocrystal salt solvates. The number of acceptable counterions in pharmaceuticals is very limited, whereas the potential number of acceptable coformers is much higher (e.g., from the GRAS and EAFUS lists). Therefore, by including coformers and including more than two residues, the number of possible acceptable crystals (in terms of safety) increases considerably.

The significant number of entries found in the CSD for each subclass, without any overlap, shows that we have defined workable and useful rules for a classification of multicomponent crystals.

ASSOCIATED CONTENT

Ⓢ Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/acs.cgd.6b00200.

> List of 80 known solvents and lists of CSD refcodes of true solvates, true salts, true cocrystals, salt solvates, cocrystal solvates, and cocrystal salt solvates (PDF)

AUTHOR INFORMATION

Corresponding Author

*E-mail: r.degelder@science.ru.nl.

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

REFERENCES

(1) Tros de Ilarduya, M. C.; Martín, C.; Goñi, M. M.; Martínez-Uhárriz, M. C. *Drug Dev. Ind. Pharm.* **1997**, *23*, 1095−1098.

(2) Fini, A.; Fazio, G.; Feroci, G. *Int. J. Pharm. (Amsterdam, Neth.)* **1995**, *126*, 95−102.

(3) Steed, J. W. *Trends Pharmacol. Sci.* **2013**, *34*, 185−193.

(4) Wood, P. A.; Feeder, N.; Furlow, M.; Galek, P. T. A.; Groom, C. R.; Pidcock, E. *CrystEngComm* **2014**, *16*, 5839−5848.

(5) *Generally Recognized as Safe (GRAS)*; U.S. FDA: College Park, MD, 2011. http://www.fda.gov/Food/IngredientsPackagingLabeling/GRAS/default.htm (accessed April 4, 2016).

(6) *Everything Added to Food in the United States (EAFUS)*; U.S. FDA, 2011. http://www.fda.gov/food/ingredientspackaginglabeling/foodadditivesingredients/ucm115326.htm (accessed April 4, 2016).

(7) *Guidance for Industry: Regulatory Classification of Pharmaceutical Co-crystals*; U.S. FDA: Silver Spring, MD, 2011.

(8) Aitipamula, S.; et al. *Cryst. Growth Des.* **2012**, *12*, 2147−2152.

(9) *Guidance for Industry: Regulatory Classification of Pharmaceutical Co-crystals*; U.S. FDA: Silver Spring, MD, 2013. http://www.fda.gov/downloads/drugs/guidancecomplianceregulatoryinformation/guidances/ucm281764.pdf.

(10) Aakeröy, C. B.; Salmon, D. J. *CrystEngComm* **2005**, *7*, 439−448.

(11) Stahly, G. P. *Cryst. Growth Des.* **2009**, *9*, 4212−4229.

(12) Rücker, G.; Rücker, C. *J. Chem. Inf. Model.* **1993**, *33*, 683−695.

(13) Görbitz, C. H.; Hersleth, H.-P. *Acta Crystallogr., Sect. B: Struct. Sci.* **2000**, *56*, 526−534.

(14) Báthori, N. B.; Lemmerer, A.; Venter, G. A.; Bourne, S. A.; Caira, M. R. *Cryst. Growth Des.* **2011**, *11*, 75−87.

(15) Eccles, K. S.; Deasy, R. E.; Fábián, L.; Braun, D. E.; Maguire, A. R.; Lawrence, S. E. *CrystEngComm* **2011**, *13*, 6923−6925.

(16) Aakeröy, C. B.; Beatty, A. M.; Helfrich, B. A. *J. Am. Chem. Soc.* **2002**, *124*, 14425−14432.

(17) Bhogala, B. R.; Basavoju, S.; Nangia, A. *CrystEngComm* **2005**, *7*, 551−562.

(18) Habgood, M.; Deij, M. A.; Mazurek, J.; Price, S. L.; ter Horst, J. H. *Cryst. Growth Des.* **2010**, *10*, 903−912.

(19) Zulfiya, A.; Zhao, F.-H.; Che, Y.-X. *Chin. J. Struct. Chem.* **2010**, *29*, 1185.

(20) Aghabozorg, H.; Ghadermazi, M.; Sheshmani, S. *Anal. Sci.: X-Ray Struct. Anal. Online* **2006**, *22*, x251−x252.

(21) Smith, G.; Wermuth, U. D.; White, J. M. *Acta Crystallogr., Sect. C: Cryst. Struct. Commun.* **2009**, *65*, o103−o107.

(22) Basavoju, S.; Boström, D.; Velaga, S. P. *Cryst. Growth Des.* **2006**, *6*, 2699−2708.

(23) Zhao, N.; Lian, Z.; Liu, P. *Z. Kristallogr. - New Cryst. Struct.* **2010**, *225*, 461−462.

(24) Kelley, S. P.; Narita, A.; Holbrey, J. D.; Green, K. D.; Reichert, W. M.; Rogers, R. D. *Cryst. Growth Des.* **2013**, *13*, 965−975.