Université catholique de Louvain (UCL)
Institute of Condensed Matter and Nanosciences (IMCN)

Insights and Advances in Cocrystal Screening

A focus on Levetiracetam/Etiracetam with achiral coformers

Thèse présentée en vue de l'obtention du grade de docteur en Sciences par

Fanny George

Louvain-la-Neuve
2012-2016

Jury members

Professor Y. Filinchuk
Université catholique de Louvain

Doctor K. Robeyns
Université catholique de Louvain

Professor J. ter Horst
University of Strathclyde

Professor J. Wouters
Université de Namur

Professor D. Peeters (President)
Université catholique de Louvain

Professor T. Leyssens (Supervisor)
Université catholique de Louvain

Remerciements

La fin de mon doctorat pointe à l'horizon et je me sens emplie de gratitude envers toutes les personnes qui ont rendu ces quatre années si riches et mémorables. Selon moi, poursuivre un doctorat s'apparente à un défi lancé à soi-même de tenir la distance et de se dépasser, pour en sortir grandi et avec une meilleure connaisance de soi. Et quand les résultats ne sont pas au rendez-vous, qu'on se sent découragé ou pas à sa place, ce sont souvent les proches qui sont là pour vous aider à ne pas perdre l'objectif de vue et ne pas baisser les bras, ou tout simplement pour vous faire retrouver le sourire en détournant votre attention pour quelques minutes ou heures bien précieuses.

Dans mon cas, le défi avait en fait commencé un peu plus tôt, lorsque j'ai décidé de réorienter mon cursus universitaire de la gestion vers la chimie ; un choix que je n'ai jamais regretté mais qui n'a pas toujours été facile à assumer. Mes premières pensées vont donc naturellement à M. Tinant sans qui je n'en serais pas là aujourd'hui. Merci d'avoir cru en moi, d'avoir conçu une passerelle sur mesure, de m'avoir donné l'impulsion dont j'avais besoin, de m'avoir accompagnée moralement pendant toutes ces années et de m'avoir donné l'opportunité de garder un pied dans l'enseignement via les cours d'été. Je me sens terriblement chanceuse de vous avoir croisé sur mon chemin !

Par rapport à cette période de transition, je suis également très reconnaissante envers Corentin et Florence d'une part et Madeleine, Nathalie, Youlia et Thuy d'autre part, qui m'ont accueillie à bras ouverts dans leurs promotions respectives en me fournissant leurs notes de cours mais surtout leur soutien et de solides points d'ancrage. Merci Coco pour ton amitié fidèle et sans frontière depuis lors, pour tous tes passages en chimphy, pour toutes ces heures devant Fringe/HIMYM etc. Je vais regretter mon partenaire attitré de badminton ;) Merci les filles pour votre gentillesse et tous nos rdv lunch au fil des années, en bioch, chimphy, au Galilée, à la crêperie, chez l'une de nous ou même à Uttrecht ! J'espère qu'à l'avenir, nous trouverons toujours un moyen pour perpétuer ces réunions joyeuses.

En ce qui concerne mon doctorat, c'est bien entendu à mon promoteur, Tom Leyssens, que je dois le plus. Merci Tom de m'avoir donné la liberté dont j'avais besoin pour oser et me trouver, tout en me soutenant psychologiquement au fur et à mesure des évènements plus ou moins importants et sensibles qui ont ponctué ma vie privée ces dernières années.

Merci aussi d'être cette personne si ouverte, flexible et bienveillante ; je mesure bien ma chance d'avoir pu travailler au sein de ton équipe. Je te remercie enfin de m'avoir donné l'opportunité de passer 4 mois au MIT ; cette période restera probablement l'une des meilleures de ma vie à de nombreux égards.

Ce doctorat, je n'aurais pas non plus pu le mener à bien sans mes collègues adorés. Il y a bien sûr eu d'abord la dream team déjà formée pendant mon mémoire, avec Thomas (Thomyyyyy) et Géraldine (Geeert), et qui a beaucoup contribué à mon envie d'entreprendre une thèse. Quelle joie d'avoir partagé votre bureau, de nombreux restos et verres, des conversations sans fin, ces après-midi en musique etc. Cette période était bénie et j'en suis souvent nostalgique. Merci Géraldine d'avoir également eu la patience de me former et de répondre à toutes mes questions. Tes élèves ont beaucoup de chance ! Merci ensuite à Natalia, Ricky, Bram, Chloé, J-Bap, Gabriel, Vanessa mais aussi à Fabrice, Iurii, Nikolay, Voraksmy, Antoine, Julien, Lisa, Rafaël et encore à M. Peeters et Tinant, Yaroslav, Koen, Laetitia, Christine et Aurore d'avoir assuré la relève et partagé tous ces dîners, tartes, crêpes, fromages, galettes des rois ; ces matchs de ping-pong, babyfoot, bowling ou descente de Lesse. Votre compagnie m'a procuré beaucoup de plaisir et j'espère que mes futurs collègues seront à la hauteur ;) Merci en particulier à Natalia et Bram d'avoir égayé mes voyages en conférences et à Ricky pour son dévouement inconditionnel à la vie du labo et de l'institut, et nos whatsup matinaux. J'espère que tu prendras confiance en toi et tes qualités d'expérimentateur, Ricky, car on sait tous que tu en es capable !

Merci aussi aux membres de l'ACIM, toujours prêts à faire une réunion au soleil et à boire un verre en terrasse, mais surtout bien présents pour représenter l'institut, le distraire et le rassembler. Je pense notamment à Sara et Fabrice qui ont été des présidents exemplaires, mais aussi à Ricky, Geert, Coco, Nath, JF, Benoît…

Merci aussi au Prof. Johan Wouters et à Bernadette qui, depuis Namur, m'ont aidée tout au long de ma thèse grâce à leurs analyses rapides et efficaces, leur disponibilité et leur accueil. Merci aussi à toi, Koen, pour tes services réguliers, qu'il s'agisse de résoudre des structures cristallines, de m'installer le dernier software ou de corriger ma thèse et mes articles avec autant de minutie. Je remercie également Laurent Collard et Pascal Van Veltem pour m'avoir formée et aidée lors des mes expériences en HPLC et DSC respectivement, et d'avoir excusé mes manœuvres malheureuses ☺

J'ai également beaucoup de gratitude pour le professeur Allan S. Myerson pour m'avoir accueillie dans son laboratoire au MIT pendant ces 4 mois

mémorables. Il est heureux de constater que des personnes de votre calibre restent si disponibles et bienveillantes. Merci également à Samir Kulkarni de m'avoir encadrée pendant ce stage de recherche et à Yuqing, Li, Chris & Jiqong pour l'accueil et l'ambiance dans notre petit bureau sans fenêtre.

Je voudrais ensuite remercier les membres de mon jury de thèse : les professeurs Joop ter Horst, Koen Robeyns, Johan Wouters, Yaroslav Filinchuk et Daniel Peeters pour leurs remarques et questions pertinentes visant à améliorer la qualité du présent document.

Je ne voudrais pas non plus passer à côté de l'occasion de remercier les membres de l'équipe des cours d'été et en particulier Catherine, Nadine et Anne pour leur dévouement et leur gentillesse.

Je remercie finalement le FRIA pour avoir financé ma recherche.

Si mes collègues directs et plus éloignés ont grandement contribué à l'aboutissement de ma thèse, je ne peux cependant pas sous-estimer l'apport de ma famille et de mes amis extérieurs. Je remercie du fond du cœur mes parents pour leur amour et leur soutien inconditionnels dans toutes les étapes de ma vie. Votre présence à mes côtés me rend plus forte et je ne saurais inventorier toutes les raisons qui justifient mon immense gratitude envers vous. Merci aussi à mes frères et soeurs, mes grands parents, tantes, oncles et cousins pour tous nos échanges et les bons moments passés ensemble. J'ai cependant une pensée particulière pour toi, Quentin, qui m'a énormément soutenue la dernière année, pour ton écoute de qualité et ton empathie, et pour avoir été mon taxi attitré vers l'aéroport si souvent fréquenté dernièrement ;)

Je voudrais ensuite remercier mes amis fidèles, en commençant bien entendu par la meilleure. Merci Jen d'être toujours là pour moi et pour tout ce qu'on partage depuis maintenant 15 ans. Je remercie également Jagna, Poe, Gaëlle, Kent, Nico, Adri & Caro, Martin & Adé, John & Cha, les akapellistes au complet mais aussi Alice & Seb, Sarah & Wes, Chris & Alice et la team de Boston (Alice, Henri, Camille & Sacha, Althea & Nathan, Antoine, Lucie…) pour tous les bons moments passés et à venir.

Last but no least, merci à toi Jan d'avoir partagé ces deux dernières années avec moi, de m'avoir donné confiance en moi et en la vie, de m'aider à aller de l'avant et d'être une source illimitée de joie et de rires au quotidien. Grâce à toi, je me sens plus heureuse que jamais et je me réjouis déjà de nos prochaines aventures ! <3

List of publications

Fanny George, Nikolay Tumanov, Bernadette Norberg, Koen Robeyns,Yaroslav Filinchuk, Johan Wouters and Tom Leyssens. "Does Chirality Influence the Tendency toward Cocrystal Formation?", *Crystal Growth & Design*, 2014, 14 (6), pp 2880–2892. DOI: 10.1021/cg500181t

Fanny George, Bernadette Norberg, Johan Wouters and Tom Leyssens, "Structural Investigation of Substituent Effect on Hydrogen Bonding in (S)-Phenylglycine Amide Benzaldimines", *Crystal Growth & Design*, 2015, 15 (8), pp 4005–4019. DOI: 10.1021/acs.cgd.5b00621

Fanny George, Bernadette Norberg, Koen Robeyns, Johan Wouters, and Tom Leyssens, "The peculiar case of Levetiracetam and Etiracetam α-ketoglutaric acid cocrystals: obtaining a stable conglomerate of Etiracetam", *Crystal Growth & Design*, 2016. DOI: 10.1021/acs.cgd.6b00819

Abstract

Cocrystals designate organic multicomponent crystals, containing a stoichiometric ratio of at least two components interacting through directional contacts, and that are not simple solvates or salts (at least one component is non ionized). Cocrystals have been especially developed in pharmaceutical sciences for their ability to modulate the physico-chemical properties of a drug product showing non-optimal parameters. Selecting likely coformers to form a cocrystal with a given active pharmaceutical ingredient (API) among the millions of potential candidates is however not trivial, as there are many factors coming into play in varying manners. This step is thus critical as the number of experiments that can be performed in practice is often quite limited due to materials, time and/or cost restrictions.

In this thesis, we thus investigated ways of improving the selection of coformers to limit the trial-and-error proportion of the procedure. This was done by studying overall trends in cocrystal formation on one hand and by considering new screening methodologies on the other hand.

For that matter, we first performed an extended and systematic experimental cocrystal screen on the enantiopure and racemic versions of a selected API, Levetiracetam. By comparing the results of their respective cocrystal screening, we first showed a tendency for enantiopure and racemic versions of a selected API to form cocrystals with identical non-chiral partners. Accordingly, we suggested a new procedure to identify cocrystals more efficiently especially when a limited amount of this compound is available.

From this set of results, we also discovered a stable conglomerate of cocrystals, which is, to the best of our knowledge, only the second report of such a case. Conglomerates are indeed very rare in comparison with racemic crystals. Yet they are intensely researched as various chiral resolution techniques are conditioned by their existence. Notably, this conglomerate is formed with the lactol tautomer of alpha-ketoglutaric acid, which had never been isolated in the solid-state up to now. The existence of a stable conglomerate in this system was put in relation with the enantiospecificity of the niary Levetiracetam cocrystals. More generally, by comparing the peculiarities of the system in hand to the general behavior of cocrystallizing chiral systems with and without zwitterionic coformers, we suggested that for a pseudoquaternary cocrystal to exist, the pseudoternary combinations

should exist and the enantiomers of the two compounds should form a diastereomeric pair at the binary level, rather than behave enantiospecifically.

In a second phase, we examined the possibility of using Isothermal Titration Calorimetry (ITC) to measure interactions between an API and complexing agents in solution, to determine whether these are indicative of successful cocrystal formation. This technique is often used in biochemistry to measure macromolecule/ligand binding and kinetic interactions, but has never been employed to screen for cocrystals before. We showed that interactions in solution between non-charged compounds, despite being quite small, can be detected by ITC but that they are not sufficient to identify cocrystal formers of a given API, as one needs to also consider the feasibility of an efficient tridimensional packing involving the two molecular partners.

Similarly, we considered the use of some of the CSD solid-form modules to validate the results of the experimental cocrystal screenings of Levetiracetam and Paracetamol. Doing so, we demonstrated how CSD knowledge-based informatics can help take decisions concerning coformer selection and experimental prioritization, and how one can optimally apply them to a given system. Finally, we suggested an overall methodology based on these modules, while highlighting that pertinence and complementarity of the different analyses varies from one system to another, depending on the relative strength of the interactions/influences in play.

Résumé

Un cocristal est un cristal organique contenant deux ou plusieurs composés en quantités stoechiométriques interagissant au travers de contacts directionnels, et qui ne sont ni de simples solvates ou sels (au moins un des composés n'est pas ionisé). L'intérêt des cocristaux pour l'industrie pharmaceutique réside principalement dans leur capacité à moduler favorablement les propriétés physico-chimiques de médicaments. Identifier les composés qui pourraient potentiellement former un cocristal (appelés coformers) avec un principe actif pharmaceutique (API) donné n'est cependant pas trivial, étant donné le nombre important de paramètres influençant la nucléation et la croissance du cocristal. Cette étape de sélection est pourtant critique puisque le nombre d'expériences pouvant être réalisées en pratique est souvent assez limité en raison de contraintes de matériel, temps et/ou budget.

Dans le cadre de ce doctorat, nous avons donc investigué plusieurs façons d'améliorer le screening de cocristaux, d'une part en mettant en exergue des tendances générales dans la formation des cocristaux et d'autre part en testant de nouvelles méthodologies de screening.

Dans un premier temps, nous avons testé de façon systématique un ensemble important de coformers non chiraux afin d'identifier les cocristaux d'un API choisi, le Levetiracetam, et de son équivalent racémique. En comparant les résultats de leurs screenings respectifs, il apparaît que les versions énantiopures et racémiques d'un composé donné ont tendance à former des cocristaux avec les mêmes partenaires. En conséquence, nous proposons une nouvelle procédure permettant d'identifier plus efficacement les cocristaux d'un composé énantiopur; celle-ci étant particulièrement pertinente lorsque le composé d'intérêt n'est disponible qu'en faible quantité.

L'analyse de ces résultats a également conduit à la découverte d'un conglomérat de cocristal, ce qui est assez rare puisqu'il n'y avait qu'une référence à ce type de composé jusqu'à présent dans la littérature. Les conglomérats sont pourtant activement recherchés en raison de leur nécessité dans certains procédés de séparation chirale. Le conglomérat étudié est d'autant plus remarquable qu'il est formé entre les énantiomères du Levetiracetam et le tautomère lactol de l'acide alpha-ketoglutarique (AKGA); ce dernier n'ayant jamais été isolé à l'état solide auparavant.

L'existence d'un conglomérat stable dans ce système est mise en relation avec l'énantiospécificité des cocristaux binaires formés entre le Leviteracetam et l'AKGA. De façon plus générale, en comparant les particularités de ce système avec le comportement général des cocristaux chiraux formés avec ou sans coformers zwittérioniques, nous suggérons que la formation d'un cocristal pseudoquaternaire est conditionnée par l'existence des cocristaux ternaires correspondants et par la présence d'une paire diastéréomérique au niveau binaire (absence d'énantiospécificité).

Dans un second temps, nous avons examiné la possibilité d'utiliser la titration calorimétrique isotherme (ITC) pour mesurer les interactions existant en solution entre un API et des agents de complexation, en vue de déterminer si celles-ci sont indicatrices de la capacité de cocristallisation de ces agents. Cette technique est couramment utilisée en biochimie afin de mesurer les interactions entre une macromolécule et un ligand ainsi que leur cinétique, mais n'avait jamais été employée auparavant dans le cadre de screening de cocristaux. Les résultats indiquent que les interactions en solution entre des composés non chargés peuvent être mesurées grâce à l'ITC, bien qu'elles soient très faibles, mais qu'elles ne sont pas suffisantes pour identifier les coformers effectifs d'un API d'intérêt, étant donné la nécessité de considérer en parallèle la faisabilité d'un packing tridimensionnel efficace impliquant les deux composés.

De façon alternative, nous avons considéré l'utilisation de certains outils informatiques développés par la CSD, afin de valider les résultats des screenings expérimentaux de cocristaux du Levetiracetam et du Paracetamol. Ce faisant, nous démontrons l'utilité générale de ces outils dans le cadre de la sélection de coformers et de la priorisation des expériences, et comment les utiliser de façon optimale sur un système donné. Enfin, nous suggérons une procédure en plusieurs étapes impliquant ces différents modules, tout en insistant sur le fait que la pertinence et la complémentarité des différentes analyses varient d'un système à l'autre en fonction de l'importance relative des différentes interactions/influences en jeu.

Abbreviations

3NBA	3-nitrobenzoic acid
4NBA	4-nitrobenzoic acid
A atoms	HB-Acceptor atoms
ABA	4-aminobenzoic acid
Ac	acetone
ACN	acetonitrile
AKGA	α-ketoglutaric acid
API	Active Pharmaceutical Ingredient
AUC	Area Under the Curve
CC	cocrystal
CCA	citraconic acid
CCDC	Cambridge Crystallographic Database Center
CL	Coordination Likelihood
CSD	Cambridge Structural Database
CSP	Crystal Structure Prediction
D atoms	HB-Donor atoms
DHBA	2,4-dihydroxybenzoic acid
DIOX	1,4-dioxane
DMMA	dimethylmalonic acid
DMSA	2,2- dimethylsuccinic acid
EtAc	ethyl acetate
Eti	Etiracetam
FRA	ferulic acid
Freq	frequency
HB	hydrogen bond
HBA	3-hydroxybenzoic acid
HBP	Hydrogen Bond Propensity
ITC	Isothermal Titration Calorimetry
LAG	Liquid Assisted Grinding
Levi	Levetiracetam
LSAM	Long-range Synthon Afbau Module
MC Score	Multi-Component Score
MeOH	methanol
MLA	maleic acid
MORPH	morpholine
MSA	mesaconic acid
NIA	5-nitroisophthalic acid
NMR	Nuclear Magnetic Resonance

NOE	Nuclear Overhauser Effect
OXA	oxalic acid
PIP	piperazine
PTA	phthalic acid
RC	Reaction Crystallization
(S)-PGA	(S)-phenylglycine amide
SDG	Solvent Drop Grinding
SLA	salicylic acid
XRD	X-Ray Diffraction
XRPD	X-Ray Powder Diffraction

Table of Contents

Part I. Introductions and objectives

1. Cocrystal definition 3
2. Cocrystal applications 4
3. Crystal engineering 5
    - 3.1 Data mining 6
    - 3.2 Graph-sets 6
    - 3.3 Supramolecular synthons 7
    - 3.4 Other approaches 11
    - 3.5 Crystal Structure Prediction 11
4. Cocrystal screening techniques 13
    - 4.1 Solution cocrystallization 13
    - 4.2 Grinding cocrystallization 16
    - 4.3 Comparison of screening techniques 18
    - 4.4 Cocrystal characterization 20
5. Thesis outline 21
6. References 23

Part II. Methodology

1. Cocrystal screening 31
2. Single crystal preparation 32
3. Slurry preparation 33
4. References 34

Part III. Results and Discussion

Chapter 1. Does Chirality Influence the Tendency towards Cocrystal Formation?

1.1 Introduction 39
1.2 Experimental Section 40
1.3 Results 43

1.3.A Cocrystal Screening 43
1.3.B Crystal Structure Analysis 46

1.4 Discussion 61
1.5 Conclusion 63
1.6 References 64

Chapter 2. The peculiar case of Leviteracetam and Etiracetam α-ketoglutaric acid cocrystals: obtaining a stable conglomerate of Etiracetam

2.1 Introduction 69
2.2 Experimental Section 70
2.3 Results 72

2.3.A Cocrystal identification 72
2.3.B Selectivity and stability analyses of the Eti-AKGA system 74
2.3.C Structural analysis 76

2.4 Discussion 79

2.4.A Lactol formation 80
2.4.B Absence of Levi-AKGA cocrystal with keto-AKGA 81
2.4.C Conglomerate increased stability 84

2.5 Conclusion 88
2.6 References 90

Chapter 3. Are intermolecular interactions in solution predictive of cocrystal formation?

3.1 Introduction 97
3.2 Experimental Section 98
3.3 Results 99

3.3.A Selection of the complexing agents 99
3.3.B Measurement of solution interaction data 100

3.4 Discussion 104
3.5 Conclusion 107
3.6 References 109

Chapter 4. Using CSD solid-form informatics to screen *in silico* for cocrystals of Levetiracetam and Paracetamol

4.1 Outline 115
4.2 Methodology 116
4.2.A Motifs search 116
4.2.B Logit-Hydrogen-bonding propensity model 117
4.2.C Coordination number 122
4.2.D Molecular complementarity 125
4.2.E Packing feature searches 127
4.3 Practical methodology details 128
4.3.A Motifs search 128
4.3.B HBP models 128
4.3.C Packing Feature searches 129
4.4 Results 130
4.4.A Levetiracetam cocrystal screening 130
4.4.B Paracetamol cocrystal screening 147
4.5 Discussion 166
4.5.A Motifs searches 166
4.5.B Molecular complementarity tests 167
4.5.C HBP & CL models 169
4.5.D Packing feature searches 171
4.5.E Recommended methodology 174
4.6 Conclusion and perspectives 176
4.7 References 177

Part IV. Conclusion and Perspectives

1. Conclusion 183
2. Perspectives 185
3. References 189

Part V. Appendices

Structural investigation of substituent effect on hydrogen bonding in (S)-phenylglycine amide benzaldimines

1. Introduction 195
2. Experimental Section 196
3. Results 198
3.1 Type I 202
3.2 Type II 206
3.3 Type III 210
3.4 Type IV 213
3.5 Type V 216
4. Discussion 217
5. Conclusion 224
6. References 226

Supporting Information of Chapter 1 229
Supporting Information of Chapter 4 237

Part I

Introduction and Objectives

1. Cocrystal definition

Despite an ongoing debate concerning the definition of cocrystals (or co-crystals), there is a general consensus on some of their characteristics. In particular, are recognized as cocrystals "all solids that are crystalline single phase materials composed of two or more different molecular and/or ionic compounds generally in a stoichiometric ratio".[1]

The controversy concerns the overlap of this definition with the historically entrenched categories of salts and solvates and the necessity or not to distinguish them. Solvates are often consensually excluded, as Aakeröy *et al.* suggested that cocrystals embrace only crystals with components that are solid at room temperature.[2] But opinions on salts are divided. On one hand, the FDA stipulates that "cocrystals components are in a neutral state and interact via nonionic interactions" and demand a factual (calculation of ΔpK_a between the donor and acceptor[a]) or experimental (spectroscopic measurements) proof of it.[3] On the other hand, experts in the field consider that such debate is meaningless as, among other reasons, clinical performance is what really matters.[1]

Moreover, it has been demonstrated that for acid-base complexes with ΔpK_a between 0 and 3, the extent of proton transfer in the solid state varies with the crystalline environment and nature of the acid/base system.[4] In this so called 'salt-cocrystal continuum', ΔpKa values are thus not reliable as decision criterion.

In the context of this work, cocrystals designate organic multicomponent crystals, containing a stoichiometric ratio of at least two components interacting through directional contacts, and that are not simple solvates or salts (at least one component is non ionized), without further consideration on proton transfer (Table 1).

Table 1. Multi-component crystals classification

	Salts	**Cocrystals**	**Solvates**
A:B	A^+B^-	A:B	A:B
A	solid	solid	solid
B	solid	solid	liquid

[a] $\Delta pK_a = pK_a(\text{base})-pK_a(\text{acid})>=1$ for salts and <1 for cocrystals.

2. Cocrystal applications

Cocrystals have been especially developed in pharmaceutical sciences for their ability to modulate the physico-chemical properties of a drug product showing non-optimal parameters. Solubility, in particular, is a highly important matter as, in 2010, it was estimated that 40% of the marketed drugs and 80-90% of the drug candidates in the R&D pipeline had critically low solubility.[5]

Improvements in drug formulation thus concern solubility[6–8] and dissolution rate,[9] but also bioavailability,[10] handling properties[11]... Drugs that suffer from polymorphism that is complex to handle may also benefit from cocrystal formation.

As an example, let's cite the case of the well-known stimulant caffeine for which at least two cocrystals exhibit significant and desirable property changes. The first one, formed with oxalic acid, is completely stable with respect to humidity over weeks while pure caffeine rapidly forms a hydrate.[12] Alternatively, its cocrystal with gentisic acid shows slower dissolution and could thus be formulated as chewable tablets.[13]

Such cocrystals, generally grouped under the term "pharmaceutical cocrystals",[14–16] thus involve one active pharmaceutical ingredients (API) and one neutral component, which can be the excipient or some added compound, intermolecularly interacting with the API while being tolerated in the body. This latter is designated as the coformer (co-former or cocrystal former) and may be selected from the GRAS (Generally Recognized as Safe) list edited by the FDA, which contains hundreds of compounds, or it may be another safe drug in sub-therapeutic amount (e.g. aspirin).

Note that the interest in cocrystals by pharmaceutical companies also lies in their suitability for patent protection. They indeed easily satisfy the three criteria for patentability (novelty, non-obviousness and utility); non-obviousness characterizing the identification of matching coformers, as explained later.

Apparently, the first description of a cocrystal synthesis in the literature occurred in 1844,[17,18] while the first patents on pharmaceutical cocrystals were issued by F. Hoffmann-La Roche in 1924 and 1934.[19]However, very few pharmaceutical cocrystals have been marketed up to now; itriconazole-succinic acid[11] and dapagliflozin propylene glycol monohydrate (Forxiga in

Europe and Farxiga in the US)[20,21] appearing as exceptions. But this may not be the case anymore in the future as cocrystallization continues to find new applications in very different areas on top of drug formulation. These include non-linear optics,[22] photographic film formulation, cocrystallization as a purification step[23] and chiral resolution tool.[24–26]

3. Crystal engineering

As all APIs possess hydrogen bond forming moieties, there is an inherent potential to synthesize cocrystals of any API, which is not the case for salts as these latter require the presence of ionizable groups on the API. Cocrystals are thus attractive alternatives in case of salt formulation failure or when the isolated forms (free drug or salts) do not exhibit the required properties. Besides, salts are less easy to control due to their solvation tendency. Finally, contrary to salt formation, cocrystals are not limited to binary combinations of compounds.[27,28]

Cocrystal design is thus a matter of finding partners with compatible functional groups able to form sufficiently strong interactions with the API and to fine-tune its properties. Besides, the choice of acceptable coformers for a given cocrystal is conditioned by its intended use. While coformers should be safe for human consumption in pharmaceutical cocrystals, this is not a restriction in academia, where prevailing criteria are rather the ease and cost of acquisition. There are thus tens of millions ("known" organic compounds) or even 10^{60} (synthetically feasible organic compounds) potential candidates for cocrystal screening. [29]

This contrasts with the number of experiments that can be performed in practice due to materials, time and/or cost limitations. Hence there is a strong need for *in silico* methods that could predict or at least reduce the number of suitable coformers and limit the trial-and-error proportion of the procedure. This is the objective of the crystal engineering approach, which has raised much interest in recent years. [17,30–35]

Crystal engineering was defined by Desiraju in 1989 as "the understanding of intermolecular interactions in the context of crystal packing and in the utilization of such understanding in the design of new solids with desired physical and chemical properties".[36]

In practice, the design of new materials through crystal engineering is thus a stepwise process. The first one consists in the analysis of hydrogen-bond patterns in very diverse crystal structures, with the objective to identify recurring schemes in terms of H-bond topology and functional group selectivity. The second step is the application of such knowledge in the systematic architecture of organic crystals with attractive properties.

3.1 Data mining

To complete the first step and identify statistically significant interaction patterns, one needs to have at disposal a sufficiently large set of structures. For that matter, the Cambridge Structural Database (CSD)[37] soon appeared as an essential resource in the practice of crystal engineering. This international database was created in 1965 to record organic molecular crystal structures and rapidly showed its potential. At its early development, it notably permitted to suggest and then prove the existence of hydrogen bonds.[38]

The database has been continuously incorporating new structures ever since and now nears 800 000 validated 3D crystal structures, from both X-ray and neutron diffraction. The number of potential comparisons and analyses now thus seems infinite and important discoveries may remain to be made.

3.2 Graph-sets

Comparative analyses of hydrogen-bond patterns are facilitated by the use of graph sets notation. In this context, graph-sets are alphanumeric characters that describe, in a very simple and efficient manner, the assembly of molecules in a crystal structure. This mathematical formalism was applied to crystal engineering for the first time in 1980 by Kuleshova & Zorky,[39] but the notation progressively evolved. The actual version is due to M. Etter that emphasized the chemical aspect,[40] but one may refer to the contribution of Bernstein *et al.* for a detailed description of its recommended usage.[41]

Graph-sets assignation is performed according to a stepwise procedure that first decomposed the overall network into H-bond motifs containing only one specific type of hydrogen bond, and then combined them in an ordered fashion. For a given motif, graph-sets specify the number of donor and acceptor atoms and the total number of atoms involved, in addition to their topology (*i.e.* chains, rings, dimers and intermolecular patterns, Figure 1). Hence, through a focus on the topology of intermolecular interactions rather

than on their exact geometries (*i.e.* cutoff distances and bond angles), graph-sets allowed to compare intermolecular networks of structures containing very different molecules, while questioning the definition of hydrogen bonds.

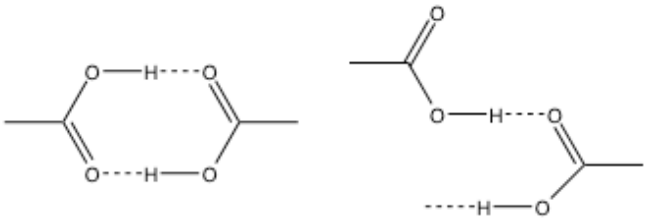


Figure 1. Representative patterns of hydrogen-bonds between molecules with carboxylic acid moieties showing one ring motif with $R^2_2(8)$ graph set (left) and one catemer/chain motif represented by C(4) graph-set (right).

3.3 Supramolecular synthons

A further simplification of the analysis was introduced by Desiraju as he drew a systematic parallel between organic synthesis and crystal engineering (Table 2).[32] He compared the covalent bonds that lead to the building of molecules from atoms, to intermolecular interactions connecting molecules into crystals, which can in turn be considered as solid-state supermolecules. This analogy is based on the fact that both processes are kinetically driven and on the apparent interchangeability of hydrogen-bond patterns within families of crystals, as it is the case for functional groups in organic synthesis. Indeed, the definition of functional groups originates from the empirical observation of discriminatory reactivities.

Table 2. Comparative elements between organic synthesis and crystal engineering

Comparative units	**Organic synthesis**	**Crystal engineering**
Target	Molecules	Crystals
Structural units	Atoms	Molecules
Cement	Covalent bonds	Intermolecular interactions
Transferable units between targets	Functional groups	H-bond patterns

From that point, he suggested the use of the retrosynthesis approach in a supramolecular way*, i.e.* to work backward from a target crystal and identify molecules that could be assembled through conceivable operations involving strategic intermolecular contacts, named supramolecular synthons. Supramolecular synthons are structural units based on the recognition of

both chemical and geometrical features of the molecules and are thus not simple intermolecular interactions. Two main categories of synthons are encountered: synthons formed between identical functional groups, called homosynthons, and synthons created around two different moieties, named heterosynthons (Figure 2).

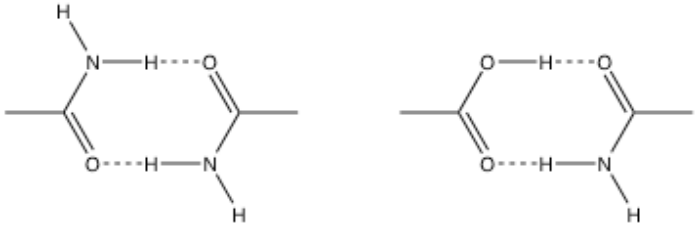


Figure 2. Representative amide-amide homosynthon and amide-acid heterosynthon, both characterized as $R^2_2(8)$ ring motif.

As in organic retrosynthesis, synthon determination is subjective and many synthons can be recognized in a given structure. But one aims in particular to identify statistically recurrent synthons in crystal structures of molecules displaying the same functional groups. Indeed, these "robust" synthons can be used for prediction and design of crystals with these complementary functional groups. Robustness is expected in particular for synthons based on multipoint recognition of functional groups (Figure 3). But the utility of synthons in the simplification of an intermolecular network and in its retrosynthesis also depends on their representativeness with respect to the overall crystal.

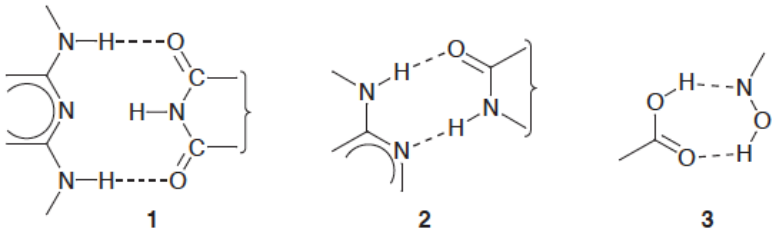


Figure 3. Three of the most robust bimolecular ring motifs involving hydrogen bonds formed with oxygen or nitrogen acceptors.[42]

The synthonic approach rapidly spread thanks to its ease of use and in particular to its absence of reference to crystallographic jargon. But its success is mainly due to its utility in pharmaceutical cocrystals design as these materials rapidly demonstrated their values. The advantage of knowing the existence and robustness of specific heterosynthons was indeed rapidly exploited to design binary,[43] ternary[27] and even quaternary cocrystals.[28]

But the synthonic strategy also showed its limits. These are due to the limit of the comparison between molecular synthesis and crystal growth. The synthesis of an organic compound often involves several steps and it is the responsibility of the chemist to choose the step order and the experimental conditions (temperature, solvent, stirring...) for each step, to optimize the entire process. In crystal engineering, however, the assembly of synthons in a 3D network is not controlled by the experimentalist. Indeed, the number of fragments (types of molecules) is limited (one for single component crystal and generally up to 4 for cocrystals) and their addition is simultaneous (*i.e.* all the fragments are present in solution at the time of nucleation), which reduces the control that could be exercised on the final product(s).

This is particularly problematic when (1) several functional groups coexist on the molecule(s), which is highly frequent for APIs, and/or when (2) both steric and electrostatic factors influences are prevalent, which seems to be almost always the case.[44]

The first complication is named synthons interference, by opposition to synthon insulation that ensures transferability from one structure to another (also called modularity). In practice, individual synthons are in fact compromised. Structural robustness and synthons hierarchy are thus a primary concern.[42,45] But the complexity of the process is really comprehended when one realizes that conventional hydrogen bonds are not the only influent intermolecular interactions and that synthons based on other types of weaker and less directional contacts also have a predictive power. These involve halogen atoms,[46–49] hydrocarbon fragments (pi stacking, C-H..X bonds…) but also more exotic atoms such as boron[50–53] or sulfur (Figure 4). Note however that this increased complexity enlarges the horizon of crystal engineering with the potential to design new materials based on these unconventional interactions.

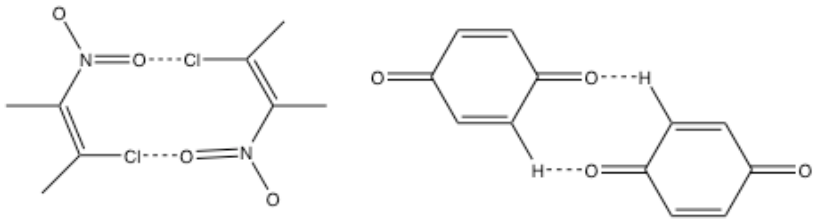


Figure 4. Bimolecular synthons involving halogen atoms (left) or C-H..X bonds (right).

The second difficulty is often referred to as the dichotomy between geometrical and chemical factors. Indeed, crystallization is often described as a kinetic event following the Curtin-Hammet principle, while everyone recognizes the application of the close-packing model proposed by Kitaigorodskii and based on geometrical arguments.[54] But this contradiction is only apparent. It seems indeed that both factors play a role but at different stages of the process: kinetics initially dominates, *i.e.* the strongest (directional) interactions are formed in priority in solution, while thermodynamics seems prevalent at the later stages of nucleation; the complexity arising from the fact that their interplay vary from one system to another.[55]

In practice, this results in the formation of nearly close-packed structures. However, Desiraju insisted on the fact that "*it is the small deviations from close-packing that are of the greatest importance, because these small deviations from close-packing owe to chemical factors and in turn lead to the formation of crystal structures that can be engineered in a systematic manner*".[36]

Besides, the interactions responsible for close packing at the late stages of crystal assembly are very weak. But recent discoveries suggest they could be critical in determining the final outcome of any crystallization experiment. This was already suspected by the late M. Etter in her pioneering work on crystal engineering: "*It may be that very weak intermolecular interactions have a disproportionally large effect on the crystal growth process compared to their interaction energies. (…) Such interactions may be important determinants in controlling molecular aggregation and the possibility of such weak interactions should not be overlooked in interpreting known crystal structures or in designing new ones".*[56]

This explains the initial focus on the molecules chemistry to the detriment of the geometrical factor, for which neither systematic retrosynthesis nor identification through energy minimization seem feasible or useful. Concerning cocrystals design, however, the importance of the geometric factor is clearly illustrated by the fact that (1) some compounds possess identical functional groups prone to the same synthons, but differ with regard to their ability to cocrystallize with an API of interest and that (2) some compounds crystallize in the same lattice without any explicit interaction between them.

3.4 Other approaches

In parallel to the synthonic approach, other descriptive properties were thus investigated to account for all the crystal engineering subtleties.

For example, researchers became progressively aware of the importance of coordination in complex formation. In 1994, Ermer stated that "*complementary (in the number and directionality of HB donors and acceptors) represents the crucial prerequisite supporting favourable interactions between the molecular partners*" and to achieve HB maximization.[57] Similarly, Issa suggested that the driving force for cocrystallization is substantial (> 10 kcal/mol) only when one partner cannot hydrogen-bond to itself due to a lack of hydrogen bond donors/acceptors.[58,59] And more recently, Galek showed that coordination can be used as a criterion to rank polymorphs by order of stability.[60]

Following the work of Pidcock, Motherwell and Galek on packing and shape considerations,[61–64] Fabian also emphasized that cocrystallization is more likely between coformers sharing similar molecular dimensions (especially planarity) and polarity, and suggested molecular descriptors that can be used to assess compatibility between two coformers.[55,65]

The practice of supramolecular retrosynthesis is thus highly conditioned by the knowledge of the properties of all the intermolecular interactions that may take place in a crystal and of the nucleation route(s) that will lead to the final product(s). These points hence logically represent the main interests in current crystal engineering efforts.

3.5 Crystal Structure Prediction

The use of computational methods to calculate crystal energy landscape of single components and predict polymorphism occurrence have proven efficient in solving the blind tests organized by the Cambridge Crystallographic Database Center (CCDC)[66] and very useful for the pharmaceutical industry.[67]

Conventional crystal structure prediction (CSP) uses force fields and a space group as input to generate structures by applying common proper symmetry operations; the quality of the results being often limited by the appropriateness of the force fields.

These methods were thus logically applied to multi-component systems with the hope to predict their existence and suggest packing preferences. These methods indeed present an advantage over the other crystal engineering approaches relying on qualitative synthons as they simultaneously and quantitatively take into account all the potential interactions in hand.[58]

However, probing the energy landscape of multi-component systems is intrinsically more challenging than for single components as the relative position and orientation of the diverse independent molecules need to be represented by additional variables, thereby increasing the search space and the computational cost. Hence advanced methodologies are multistage and involved exploring first the conformational space of the component pair through high-level *ab initio* calculations and then holding these optimized conformations rigid during the generation of putative structures of low lattice-energy.[59] Besides, these are sometimes informed by a priori statistical analysis of the CSD to identify a small set of preferred conformations that should be considered.[68]

Concerning cocrystals, the position of the acidic proton (in the salt -cocrystal continuum) is also problematic as it influences the ordering of the putative structures by favouring different hydrogen bond motifs.[69]

Besides, as for single-component studies, models usually predict a considerable number of plausible putative crystal structures (typically 100-200) near the global minimum in terms of energy.[31]

In practice, the experimental structures are often found to be amongst the most thermodynamically stable forms. However, the cocrystal stabilization is quite small in comparison with (a) the weighted sum of the lattices energies of the component structures (< 10 kcal/mol), (b) the approximations in the calculations and (c) the usual polymorphic differences. Hence, cocrystal formation cannot be foreseen without doubt with such methods in a realistic screening procedure.[58] Up to now, it remains quicker and easier to proceed experimentally.[44]

Moreover, these methodologies do not take into account the kinetic factors, such that there is no guarantee that all the promising arrangements are really in competition during the nucleation step.

Nonetheless, CSP may be useful as a complementary method in assessing the risk of polymorphism or crystal disorder.[44] Besides, it could help to solve crystal structures from powder X-ray data.[59,70]

4. Cocrystal screening techniques

If serendipity prevailed in early cocrystallization developments, rationalization permitted by the advances of crystal engineering allow for a more focused screening. However, there remain grey areas in the understanding of the nucleation and growth processes and prediction of suitable coformers for a target molecule is still limited. High-throughput methods are thus required in practice and allow sampling the myriad of variables that play a role during a cocrystallization experiment.

There are two main types of techniques that allow performing efficient cocrystal screenings: grinding and solution-mediated crystallization. These will be briefly described and compared. Other cocrystallization methods have been developed, such as cocrystallization from the melt, using hot-stage microscopy[71,72] or with compressed CO_2 as antisolvent,[73] to name but a few. But they won't be discussed here, as their use is more occasional and usually requires more sophisticated devices.

4.1 Solution cocrystallization

Producing cocrystals in solution implies taking into account the effect of the solvent on the crystallization outcome. In particular, one aims to identify the conditions for which the cocrystal is the only stable solid phase, and the paths that can lead to its crystallization. In other words, one has to find conditions where supersaturation is sufficient to overcome the kinetic barrier to cocrystal nucleation, but where supersaturation of separate components is prevented.

For that matter, ternary phase diagrams are helpful,[24,25,74–77] especially when optimization or scale-up is envisaged. But their complete definition may be quite time-consuming. However, recent advances in the understanding of the thermodynamics of solution cocrystallization suggested the use of alternative binary phase-diagrams, which allow finding the cocrystal stability zone with the knowledge of only a few specific solubility points. These are based on the description of the cocrystal solubility as the product of the components concentrations and solution complexation constants.[78]

Two types of methods are used to screen for cocrystals in solution, depending on the starting conditions (stoichiometric or non-stoichiometric solutions of the reactants). The choice of one method in particular is

conditioned by the relative solubility of both components in the crystallization solvent.

4.1.A Stoichiometric solutions as starting point

When coformers (A and B) have comparable and sufficient solubilities ($x_A^* \approx x_B^*$) in the crystallization solvent (S), the corresponding ternary phase diagram is (almost) symmetric and the cocrystal shows congruent dissolution (*i.e.* coformers are more soluble/less stable than the cocrystal at 1:1 composition). In this case, evaporation or cooling of an equimolar solution of the reactants leads to selective crystallization of the AB cocrystal.

The evaporation process (E) is illustrated on the solution binary diagram (Figure 5, right), by an arrow starting from the origin of the chart and following the 1:1 line until reaching the AB zone. On the ternary diagram, the origin corresponds to the solvent (S) apex but the path is similar.

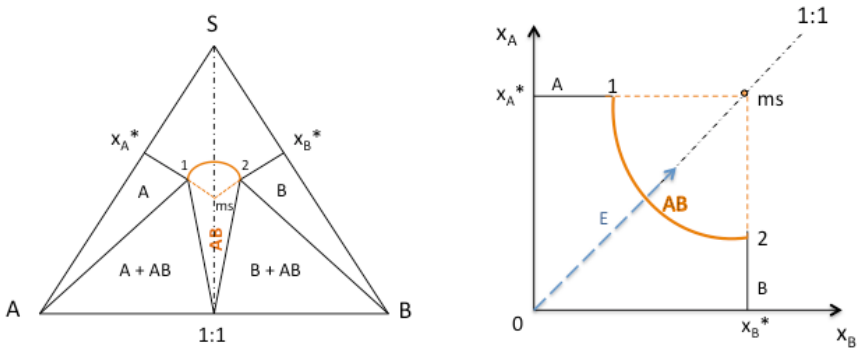


Figure 5. Ternary and binary phase diagrams characterizing the thermodynamics of two components (A & B) in a given solvent (S) at fixed temperature, with formation of a cocrystal AB showing congruent dissolution.[79]

4.1.B Non-stoichiometric solutions as starting point

When coformer solubilities are too disparate and cocrystal dissolution is incongruent (*i.e.* coformers are less soluble/more stable than cocrystal in pure solvent), evaporation or cooling of equimolar solutions is inappropriate. Indeed, this would lead to supersaturation with respect to the more stable coformer or both the coformer and the cocrystal (Figure 6) and selective crystallization of the cocrystal is not guaranteed anymore. In such case, one has to consider alternative pathways to generate cocrystal supersaturation.

These are based on the fact that if a cocrystal is more stable in one solvent/mixture of solvents than its separate components, it will be less soluble than their combination. Consequently, a solution that is simultaneously at saturation with respect to both coformers (maximum saturation, ms) will be supersaturated in the cocrystal.

This property may be exploited in different ways. The first one has been proposed by Rodriguez-Hornedo *et al.*[79,80] and consists in initiating nucleation of the cocrystal from a solution that is (nearly) saturated in one coformer (the most soluble), through the progressive addition of the other coformer (the least soluble).[b] This technique is referred as Reaction Crystallization (RC) and corresponds to the S path on the phase diagrams of Figure 6. It is particularly useful when coformer solubilities are drastically different and when a 1:1 ratio solution would never lead to successful cocrystallization. This was evidenced by Childs *et al.* for the 2:1 carbamazepine-succinic acid cocrystal.[79] Solubility of succinic acid is 1000 time higher than that of carbamazepine such that the acid concentration should be at least 40 times above the stoichiometric ratio to allow cocrystal supersaturation.

Note that the mixing regime applied to the solution prepared accordingly and left for equilibration, may affect the polymorphic outcome of the cocrystallization procedure.[81]

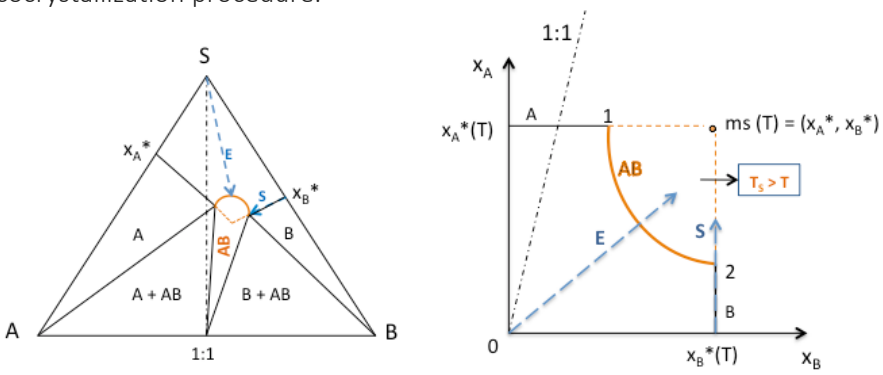


Figure 6. Ternary and binary phase diagrams characterizing the thermodynamics of two components (A & B) in a given solvent (S) at fixed temperature, with formation of a cocrystal AB showing incongruent dissolution.[79]

[b] In case of pharmaceutical cocrystals, the coformer is usually more soluble and more readily available than the API and should be used to prepare the initial saturated solution.

ter Horst *et al.* suggest a different procedure for an effective cocrystal screening.[82] The first step consists in the determination of coformer solubilities at a reference temperature (room temperature being convenient), $x_A^*(T)$ and $x_B^*(T)$, which can be automated with the use of standard laboratory equipment. In a second step, one prepares a suspension of both coformers characterized by coformer concentrations equal to their reference temperature solubilities, $ms(T)=(x_A^*(T), x_B^*(T))$, by first dissolving the solid phase at higher temperature and then recrystallizing it to ensure the potential cocrystal formation. Then, the saturation temperature T_S of the suspension (*i.e.* the point where it is all dissolved upon heating) is measured and compared to the reference temperature T. If $T_S > T$, cocrystal formation is suspected, as more stable phases show lower solubilities.

Note that if solubility measurements are not readily feasible, one can prepare the equivalent suspensions in a few steps. First, suspensions of each component are prepared separately at a given temperature T. Then, a similar volume of each filtrated solutions are sampled and mixed. The resulting solution has a concentration equal to half of the solubilities of each component ($x_A^*(T)/2$, $x_B^*(T)/2$) and evaporating half of its solvent volume, following the E direction on Figure 6, brings it in the desired region of the phase diagram.

4.2 Grinding cocrystallization[83–85]

Another way to produce cocrystals consists in grinding the coformers together.[86] There are in particular two types of grinding crystallizations. The first one is referred to as neat grinding by opposition to the second type that involves addition of a drop of solvent in the mixture. This latter has a lot of different denominations: solvent drop grinding (SDG), liquid assisted grinding (LAG), kneading, mechanochemistry and wet cogrinding.

In such experiments, grinding is performed either manually through the use of a mortar and pestle, or in an automated way, with an electrical miller (ball milling), which often allows performing several experiments in parallel.

In case of cocrystal screening, one generally starts by grinding the 1:1 mixture of coformers. But grinding of other compositions can be useful if cocrystals with other stoichiometries are suspected.

Rationalization of the process and the outcome of a mechanochemical reaction requires calling upon multiple mechanisms that may operate in

parallel depending on the type of grinding performed and on the physical state of the coformers. Variation between the different mechanisms concerns the nature of the intermediate phase produced during the experiment (*i.e.* a gas, a liquid, or an amorphous solid) that increases the mobility and/or energy of the coformers in the bulk. Four mechanisms are notably suggested. These are briefly described here, as they are the subject of several detailed contributions.[83–85]

The first one involves molecular diffusion of species through a vapor phase during grinding, which proceeds in continuously removing the product from the surface to expose new reactants. This process is at stake when neat grinding a mixture of reactants where at least one shows significant vapor pressure in the solid.

The second type of grinding crystallization is mediated via a liquid phase. In case of neat grinding, the liquid phase is progressively depleted as the reaction proceeds and may arise from the use of a reactant that is liquid at room temperature. Such a mechanism is also used to describe experiments in which molecules of solvent get incorporated in the crystalline phase (solvate formation).

LAG also proceeds according to this mechanism even though, by contrast with neat grinding, the liquid phase is conserved during the experiment and supposedly acts as a catalyst improving the kinetics of cocrystallization. However, there is currently no certainty concerning the influence of the liquid nature on the grinding outcome. Indeed, its nature doesn't seem to matter in some cases while there are systems for which different solvents lead to distinct products. In the first scenario, the solvent thus seems to have only a physical effect (lubrication) while in the second, its influence is more chemical (molecular recognition and templating). Nonetheless, it appears that the rate of cocrystal formation by grinding is increased when one or both reactants are soluble in the added solvent.

Milling time also seems to affect the cocrystallization outcome of systems with heteroditopic partners (*i.e.* displaying two distinct binding sites). In several cases,[87,88] grinding for less than 5 minutes produces a kinetic intermediate cocrystal based on discrete assemblies and/or the strongest supramolecular interactions between the coformers; while further grinding (around 20 minutes) lead to the formation of a more stable polymorph characterized by a larger number of weaker H-bonds per formula unit. Competition of intermolecular interactions thus also occurs in mechanosynthesis.

The next cocrystal forming mechanism operates through the formation of metastable eutectic: in some systems, reactants melt at their interface, which allows cocrystal nucleation and subsequent formation of solid product.

In the last mechanism, which solely concerns neat grinding, cocrystallization is facilitated by the formation of an amorphous phase. This mechanism accounts for cocrystallization of reactants that are neither volatile nor liquid at ambient conditions, which represents a large proportion of the commonly used pharmaceutical compounds.

4.3 Comparison of screening techniques

Several comparative studies suggest a superiority of grinding approaches over solution-mediated techniques in the screening of pharmaceutical forms. The main advantage is the avoidance of limitations due to solubility disparities between the coformers. Independence with respect to solubilities comes from the fact that the amount of solvent included in the process is so small that the overall composition of the system lies at the bottom of the phase diagram, where the cocrystal is the only solid stable form. Another common interpretation invokes the continuous saturation of the reactants in the liquid phase to justify the absence of solubility effects.

LAG also permits to obtain pure phase materials when no control over polymorphism can be exercised using solution crystallization.[89] Besides, grinding methods allow designing three-component cocrystals, whose formation is rather unlikely and challenging in solution due to the multiple competing paths (Figure 7).

Other benefits include reduced costs and environmentally friendly character of the process. Finally, recent advances demonstrated the possibility of using mechanochemical reactions at industrial scale by replacing laboratory ball millers with twin screw extrusion devices.[90]

However, there is one major limitation of mechanochemical synthesis in comparison with solution-based approaches: its incapacity to generate single crystals suitable for structural analysis. But this is less of a concern now due to the recent advances in X-ray powder diffraction (XRPD) methodologies (synchrotron faciities). Besides, the powder samples produced by LAG can be used as seeds to ease crystal growth in solution.

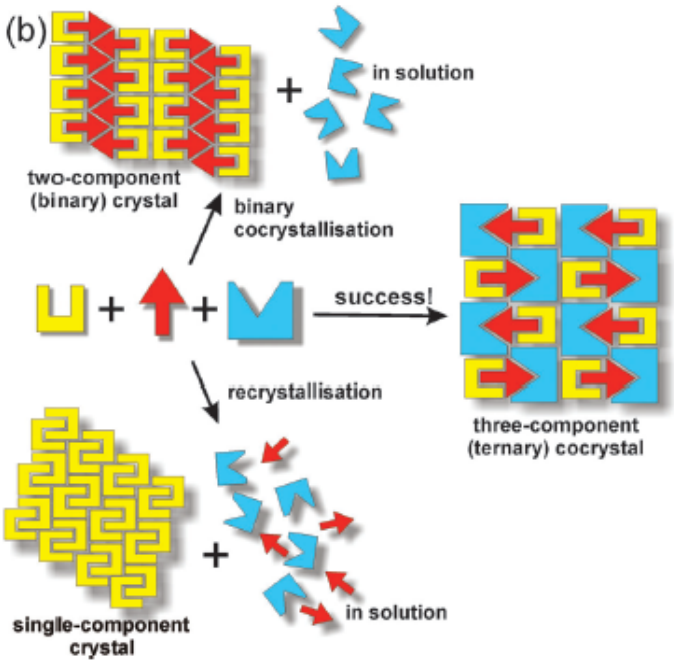


Figure 7. Synthesis of three-component cocrystals is challenging in solution as components may also crystallize separately aside or in binary combinations.[84]

LAG thus appears as the solution of choice, especially when an electric miller is available. Indeed, manual grinding may be tiresome and heterogeneous and no control can be exercised over milling intensity and temperature. Moreover, it is rather difficult to estimate the required milling time as it varies from one system to another due to physico-chemical and mechanical differences between the coformers. For that reason, manual grinding often lead to partial conversion of the starting materials.[81] It is thus recommended to use an electric miller during a sufficient period of time (at least 20 minutes) for reproducibility. Note that when such a device is not directly available, standard vortex mixers, may be used alternatively to produce the same results.[91]

Finally, the fact that different polymorphs may be obtained using different screening techniques suggests their concomitant application when the objective lies in finding the maximum number of polymorphs.

In practice, however, the most popular technique is evaporation (Figure 8).[81] But this may change in a near future, as the other techniques that are more recent, get more and more popularized.

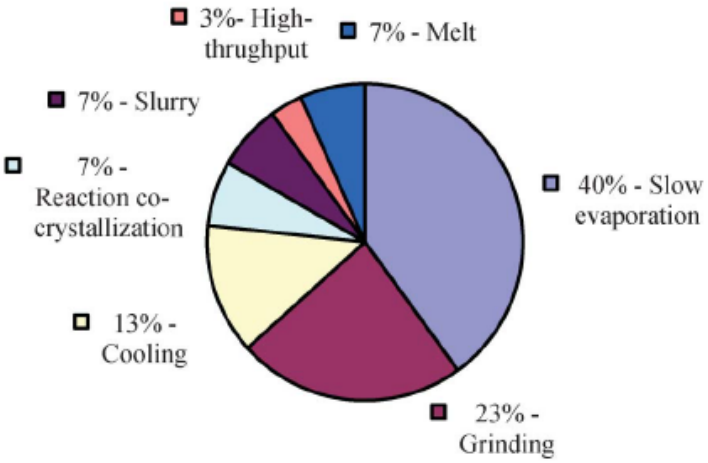


Figure 8. Breakdown of the experimental cocrystal screening techniques according to their frequency of use as recorded in the litterature.[81,92]

4.4 Cocrystal characterization

Presence of a cocrystal, produced either by solution or by grinding crystallizations, is usually attested by comparing the XRPD pattern of the filtrated residues or ground materials resp., with the spectra of the pure components and their polymorphs. But other analytical techniques may be used to characterize the cocrystallization experiment outcome or to monitor the conversion process in situ (IR, Raman, UV, DSC...).[93]

The existence of a new solid phase may then be confirmed by growing the corresponding single crystal in solution. To obtain crystalline material of sufficient size, low supersaturation and low nucleation rate are required, which is not the case at the screening stage.

Finally, note that, as for single component crystallization, polymorphism may be encountered (e.g. XRPD pattern does not match the pattern generated from the single crystal).[94]

5. Thesis outline

As highlighted in this brief introduction, selecting likely coformers to form a cocrystal with a given API is not trivial, as there are many factors coming into play in varying manners. Within this context, we decided to look for ways of improving the selection procedure, by studying overall trends in cocrystal formation on one hand and by considering new screening methodologies on the other hand. We were in particular interested in answering the following questions:

- How frequent are enantiopure cocrystals in comparison with their racemic equivalent when one of the two coformers is not chiral?
- Is it possible to isolate a stable conglomerate of cocrystals that could further be used for preferential cocrystallization[75,99,100] or Viedma ripening?[101–110]
- Are the interactions responsible for cocrystal formation and cohesion detectable from dilute solutions containing both coformers?
- Can we use some of the knowledge-based softwares developed by the CCDC in a complementary fashion to easily select promising coformers?

To address the first two questions, we performed an extended and systematic experimental cocrystal screen on the enantiopure and racemic versions of a selected API, Levetiracetam, and compare the structural outcomes from each screening. Doing so, we suggested an approach that could lead to a more optimal cocrystal screening of an enantiopure compound, especially when a limited amount of this compound is available. This analysis is presented in the first chapter of this thesis, which corresponds to a publication in Crystal Growth & Design. From this set of results, we also discovered a stable conglomerate of cocrystals, which is, to the author's best knowledge, only the second report of such compound. In a second chapter, we analyzed in details the peculiarities of this complex system to evaluate if they could be reproduced to easily generate new cocrystal conglomerates. The answer was unfortunately negative but the system in hand remained a very interesting case study that has just been accepted for publication in Crystal Growth & Design.

In a second phase, we examined the possibility of using Isothermal Titration Calorimetry (ITC) to measure interactions between an API and complexing agents in solution, to determine whether these are indicative of successful cocrystal formation. This technique is often used in biochemistry to measure macromolecule/ligand binding and kinetic interactions,[111] but has never

been employed to screen for cocrystals before. In fact, only a limited amount of work suggests using solution interactions for that matter. To assess the potential of ITC in this context, we selected coformers used in the above-mentioned cocrystal screening (some that successfully formed a cocrystal and some that did not) and observed if this method could validate the previous results. This investigation was performed during a research stay at the laboratory of Prof. Myerson at the MIT and is discussed in a third chapter.

Similarly, we decided to consider using some of the CSD solid-form modules, which were developed over the years by the CCDC to tackle problems such as polymorphism, to validate the results of the experimental cocrystal screenings of Levetiracetam and Paracetamol. We debate notably their capacity and limits in this context, how to draw maximum benefit from them and some potential sources of improvement. Each tool is first discussed separately and then, an overall methodology is suggested.

In a last contribution, we present a detailed structural analysis of twenty-three new crystal structures of (S)-phenylglycine amide benzaldimines with various substituents. We discussed in particular the interplay of steric and electronic effects of the substituents on the resulting bonding patterns, conformational features and packing, and confirmed the usefulness of Hirshfeld surfaces to probe the secondary interactions. This analysis was published in Crystal Growth & Design but is placed in the appendices of this thesis as it does not directly concern cocrystal engineering, which is the main topic of this work.

6. References

(1) Aitipamula, S.; Banerjee, R.; Bansal, A. K.; Biradha, K.; Cheney, M. L.; Choudhury, A. R.; Desiraju, G. R.; Dikundwar, A. G.; Dubey, R.; Duggirala, N.; Ghogale, P. P.; Ghosh, S.; Goswami, P. K.; Goud, N. R.; Jetti, R. R. K. R.; Karpinski, P.; Kaushik, P.; Kumar, D.; Kumar, V.; Moulton, B.; Mukherjee, A.; Mukherjee, G.; Myerson, A. S.; Puri, V.; Ramanan, A.; Rajamannar, T.; Reddy, C. M.; Rodriguez-Hornedo, N.; Rogers, R. D.; Row, T. N. G.; Sanphui, P.; Shan, N.; Shete, G.; Singh, A.; Sun, C. C.; Swift, J. A.; Thaimattam, R.; Thakur, T. S.; Kumar Thaper, R.; Thomas, S. P.; Tothadi, S.; Vangala, V. R.; Variankaval, N.; Vishweshwar, P.; Weyna, D. R.; Zaworotko, M. J. *Cryst. Growth Des.* **2012**, *12*, 2147–2152.

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

(3) FDA. *Fda* **2013**, 1–5.

(4) Childs, S. L.; Stahly, G. P.; Park, A. *Mol. Pharm.* **2007**, *4*, 323–338.

(5) Babu, N. J.; Nangia, A. *Cryst. Growth Des.* **2011**, *11*, 2662–2679.

(6) Schultheiss, N.; Bethune, S.; Henck, J.-O. *CrystEngComm* **2010**, *12*, 2436.

(7) Thakuria, R.; Delori, A.; Jones, W.; Lipert, M. P.; Roy, L.; Rodríguez-Hornedo, N. *Int. J. Pharm.* **2013**, *453*, 101–125.

(8) Rodríguez-hornedo, N. .

(9) Gagniere, E.; Mangin, D.; Puel, F.; Rivoire, A.; Monnier, O.; Garcia, E.; Klein, J. P. *J. Cryst. Growth* **2009**, *311*, 2689–2695.

(10) Jung, M.-S.; Kim, J.-S.; Kim, M.-S.; Alhalaweh, A.; Cho, W.; Hwang, S.-J.; Velaga, S. P. *J. Pharm. Pharmacol.* **2010**, *62*, 1560–1568.

(11) Remenar, J. F.; Morissette, S. L.; Peterson, M. L.; Moulton, B.; MacPhee, J. M.; Guzmán, H. R.; Almarsson, Ö. *J. Am. Chem. Soc.* **2003**, *125*, 8456–8457.

(12) Trask, A. V.; Motherwell, W. D. S.; Jones, W.; Samuel Motherwell, W. D.; Jones, W. *Cryst. Growth Des.* **2005**, *5*, 1013–1021.

(13) Higuchi, T.; Pitman, I. H. *J. Pharm. Sci.* **1973**, *62*, 55–58.

(14) Vishweshwar, P.; McMahon, J. a; Bis, J. a; Zaworotko, M. J. *J. Pharm. Sci.* **2006**, *95*, 499–516.

(15) Sekhon, B. S. *ARS Pharm.* **2009**, *50*, 99–117.

(16) *Pharmaceutical Salts and Co-crystals*; Wouters, J.; Quéré, L., Eds.; RSC Drug Discovery; Royal Society of Chemistry: Cambridge, 2011.

(17) Zaworotko, M. J. *Cryst. Growth Des.* **2007**, *7*, 4–9.

(18) Wöhler, F. *Annalen* **1844**, 153.

(19) Viertelhaus, M.; Hilfiker, R.; Blatter, F.; Neuburger, M. *Cryst. Growth Des.* **2009**, *9*, 2220–2228.

(20) Bronson, J.; Black, A.; Dhar, T. G. M.; Ellsworth, B. A.; Merritt, J. R. In *Annu. Rep. Med. Chem*; 2013; pp. 471–546.
(21) Gougoutas, J. Z.; Lobinger, H.; Ramakrishnan, S.; Deshpande, P. P.; Bien, J. T.; Lai, C.; Wang, C.; Riebel, P.; Grosso, J. A.; Nirschl, A. A.; Singh, J.; Dimarco, J. D. WO. Pat. 2008002824, 2008.
(22) Huang, K.-S.; Britton, D.; Margaret, L.; C. Etter, T.; Byrn, S.; R. *J. Mater. Chem.* **1997**, *7*, 713.
(23) Billot, P.; Hosek, P.; Perrin, M. **2013**.
(24) Springuel, G.; Leyssens, T. *Cryst. Growth Des.* **2012**, *12*, 3374–3378.
(25) Springuel, G.; Collard, L.; Leyssens, T. *CrystEngComm* **2013**, *15*, 7951.
(26) Springuel, G. Chirality and cocrystal systems : from fundamental understanding to development of a novel industrial chiral resolution technique, Université catholique de Louvain, 2014.
(27) Aakeröy, C. B.; Desper, J.; Smith, M. M. *Chem. Commun.* **2007**, 3936.
(28) Dubey, R.; Mir, N. A.; Desiraju, G. R. *IUCrJ* **2016**, *3*, 102–107.
(29) Wood, P. a.; Feeder, N.; Furlow, M.; Galek, P. T. a.; Groom, C. R.; Pidcock, E. *CrystEngComm* **2014**, *16*, 5839.
(30) Desiraju, G. R. *Angew. Chemie - Int. Ed.* **2007**, *46*, 8342–8356.
(31) Desiraju, G. R. *J. Am. Chem. Soc.* **2013**, *135*, 9952–9967.
(32) Desiraju, G. R. *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 2311–2327.
(33) Mukherjee, A. *Cryst. Growth Des.* **2015**, *15*, 3076–3085.
(34) Aaltonen, J.; Allesø, M.; Mirza, S.; Koradia, V.; Gordon, K. C.; Rantanen, J. *Eur. J. Pharm. Biopharm.* **2009**, *71*, 23–37.
(35) Vishweshwar, P.; McMahon, J. A.; Peterson, M. L.; Hickey, M. B.; Shattock, T. R.; Zaworotko, M. J. *Chem. Commun. (Camb).* **2005**, 4601–4603.
(36) Desiraju, G. R. *Crystal Engineering: The design of organic solids*; Elsevier.; Amsterdam, 1989.
(37) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.* **2016**, *72*, 171–179.
(38) Feeder, N.; Pidcock, E.; Reilly, A. M.; Sadiq, G.; Doherty, C. L.; Back, K. R.; Meenan, P.; Docherty, R. *J. Pharm. Pharmacol.* **2015**, *67*, 857–868.
(39) Kuleshova, L. N.; Zorky, P. M. *Acta Crystallogr. Sect. B Struct. Crystallogr. Cryst. Chem.* **1980**, *36*, 2113–2115.
(40) Etter, M. C.; MacDonald, J. C.; Bernstein, J. *Acta Crystallogr. Sect. B Struct. Sci.* **1990**, *46*, 256–262.
(41) Bernstein, J.; Davis, R. E.; Shimoni, L.; Chang, N. *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 1555–1573.
(42) Allen, F. H.; Samuel Motherwell, W. D.; Raithby, P. R.; Shields, G. P.; Taylor, R. *New J. Chem.* **1999**, *23*, 25–34.
(43) Aakeröy, C. B.; Desper, J.; Helfrich, B. A. *CrystEngComm* **2004**, *6*, 19–

24.
(44) Issa, N.; Barnett, S.; Mohamed, S.; Braun, D.; Copley, R.; Tocher, D.; Price, S. *CrystEngComm* **2012**.
(45) Moragues-Bartolome, A. M.; Jones, W.; Cruz-Cabeza, A. J. *Crystengcomm* **2012**, *14*, 2552–2559.
(46) Mukherjee, A.; Desiraju, G. R. *Cryst. Growth Des.* **2011**, *11*, 3735–3739.
(47) Allen, F. H.; Goud, B. S.; Hoy, V. J.; Howard, J. A. K.; Desiraju, G. R. *J. Chem. Soc. Chem. Commun.* **1994**, 2729.
(48) Pedireddi, V. R.; Sarma, J. A. R. P.; Desiraju, G. R. *J. Chem. Soc., Perkin Trans. 2* **1992**, 311–320.
(49) Cinčić, D.; Friščić, T.; Jones, W. *Chem. - A Eur. J.* **2008**, *14*, 747–753.
(50) Tarakeshwar, P.; Lee, S. J.; Lee, J. Y.; Kim, K. S. *J. Phys. Chem. B* **1999**, *103*, 184–191.
(51) Tian, S. X.; Li, H.-B.; Bai, Y.; Yang, J. *J. Phys. Chem. A* **2008**, *112*, 8121–8128.
(52) Sarma, R.; Baruah, J. B. *J. Mol. Struct.* **2009**, *920*, 350–354.
(53) Madura, I. D.; Czerwińska, K.; Sołdańska, D. *Cryst. Growth Des.* **2014**, *14*, 5912–5921.
(54) Beevers, C. A. *Acta Crystallogr.* **1962**, *15*, 622–623.
(55) Fábián, L.; Frišcic, T. In *Pharmaceutical Salts and Co-crystals*; 2011; pp. 89–109.
(56) Etter, M. C. *J. Phys. Chem.* **1991**, *95*, 4601–4610.
(57) Ermer, O.; Eling, A. *J. Chem. Soc. Perkin trans. 2* **1994**, 925–944.
(58) Issa, N.; Karamertzanis, P. G.; Welch, G. W. A.; Price, S. L. *Cryst. Growth Des.* **2009**, *9*, 442–453.
(59) Karamertzanis, P. G.; Kazantsev, A. V.; Issa, N.; Welch, G. W. A.; Adjiman, C. S.; Pantelides, C. C.; Price, S. L. *J. Chem. Theory Comput.* **2009**, *5*, 1432–1448.
(60) Galek, P. T. a; Chisholm, J. a; Pidcock, E.; Wood, P. a. *Acta Crystallogr. B. Struct. Sci. Cryst. Eng. Mater.* **2014**, *70*, 91–105.
(61) Pidcock, E.; Motherwell, W. D. S. *Cryst. Growth Des.* **2004**, *4*, 611–620.
(62) Pidcock, E.; Motherwell, W. D. S. *Cryst. Growth Des.* **2005**, *5*, 2322–2330.
(63) Motherwell, W. D. S. *CrystEngComm* **2010**, *12*, 3554.
(64) Galek, P. T. a. *CrystEngComm* **2011**, *13*, 841.
(65) Fábián, L. *Cryst. Growth Des.* **2009**, *9*, 1436–1443.
(66) Day, G. M.; Cooper, T. G.; Cruz-Cabeza, A. J.; Hejczyk, K. E.; Ammon, H. L.; Boerrigter, S. X. M.; Tan, J. S.; Della Valle, R. G.; Venuti, E.; Jose, J.; Gadre, S. R.; Desiraju, G. R.; Thakur, T. S.; Van Eijck, B. P.; Facelli, J. C.; Bazterra, V. E.; Ferraro, M. B.; Hofmann, D. W. M.; Neumann, M.

A.; Leusen, F. J. J.; Kendrick, J.; Price, S. L.; Misquitta, A. J.; Karamertzanis, P. G.; Welch, G. W. A.; Scheraga, H. A.; Arnautova, Y. A.; Schmidt, M. U.; Van De Streek, J.; Wolf, A. K.; Schweizer, B. *Acta Crystallogr. Sect. B Struct. Sci.* **2009**, *65*, 107–125.
(67) Price, S. L. *Adv. Drug Deliv. Rev.* **2004**, *56*, 301–319.
(68) Karamertzanis, P. G.; Price, S. L. *J. Phys. Chem. B* **2005**, *109*, 17134–17150.
(69) Mohamed, S.; Tocher, D. A.; Price, S. L. *Int. J. Pharm.* **2011**, *418*, 187–198.
(70) Tremayne, M.; Grice, L.; Pyatt, J. C.; Seaton, C. C.; Kariuki, B. M.; Tsui, H. H. Y.; Price, S. L.; Cherryman, J. C. *J. Am. Chem. Soc.* **2004**, *126*, 7071–7081.
(71) Berry, D. J.; Seaton, C. C.; Clegg, W.; Harrington, R. W.; Coles, S. J.; Horton, P. N.; Hursthouse, M. B.; Storey, R.; Jones, W.; Friščić, T.; Blagden, N. *Cryst. Growth Des.* **2008**, *8*, 1697–1712.
(72) Lemmerer, A.; Esterhuysen, C.; Bernstein, J. *J. Pharm. Sci.* **2010**, *99*, 4054–4071.
(73) Neurohr, C.; Marchivie, M.; Lecomte, S.; Cartigny, Y.; Couvrat, N.; Sanselme, M.; Subra-Paternault, P. *Cryst. Growth Des.* **2015**, *15*, 4616–4626.
(74) Chiarella, R. A.; Davey, R. J.; Peterson, M. L. *Cryst. Growth Des.* **2007**, *7*, 1223–1226.
(75) Coquerel, G. In *Top Curr Chem*; 2006; pp. 1–51.
(76) Leyssens, T.; Springuel, G.; Montis, R.; Candoni, N.; Veesler, S. *Cryst. Growth Des.* **2012**, *12*, 1520–1530.
(77) Jayasankar, A.; Reddy, L. S.; Bethune, S. J.; Rodríguez-Hornedo, N. *Cryst. Growth Des.* **2009**, *9*, 889–897.
(78) Nehm, S. J.; Rodriguez-Spong, B.; Rodriguez-Hornedo, N. *Cryst. Growth Des.* **2006**, *6*, 592–600.
(79) Childs, S. L.; Rodríguez-Hornedo, N.; Reddy, L. S.; Jayasankar, A.; Maheshwari, C.; McCausland, L.; Shipplett, R.; Stahly, B. C. *CrystEngComm* **2008**, *10*, 856.
(80) Rodríguez-Hornedo, N.; Nehm, S. J.; Seefeldt, K. F.; Pagán-Torres, Y.; Falkiewicz, C. J. *Mol. Pharm.* **2006**, *3*, 362–367.
(81) Rahim, S. A.; Hammond, R. B.; Sheikh, A. Y.; Roberts, K. J. *CrystEngComm* **2013**, *15*, 3862.
(82) ter Horst, J. H.; Deij, M. A.; Cains, P. W. *Cryst. Growth Des.* **2009**, *9*, 1531–1537.
(83) Friščić, T.; Jones, W. *Cryst. Growth Des.* **2009**, *9*, 1621–1637.
(84) Friščić, T. *Chem. Soc. Rev.* **2012**, *41*, 3493.
(85) Delori, A.; Friščić, T.; Jones, W. *CrystEngComm* **2012**, *14*, 2350.
(86) Karki, S.; Friscic, T.; Jones, W.; Motherwell, W. D. S. *Mol. Pharm.*

2007, *4*, 347–354.
(87) Friscić, T.; Fábián, L.; Burley, J. C.; Jones, W.; Motherwell, W. D. S.; Friščić, T.; Fábián, L.; Burley, J. C.; Jones, W.; Motherwell, W. D. S. *Chem. Commun.* 2006, 5009–5011.
(88) Cinčić, D.; Friščić, T.; Jones, W. *J. Am. Chem. Soc.* 2008, *130*, 7524–7525.
(89) Trask, A. V.; Motherwell, W. D. S.; Jones, W. *Chem. Commun.* 2004, 890.
(90) Medina, C.; Daurio, D.; Nagapudi, K.; Alvarez-Nunez, F. *J. Pharm. Sci.* 2010, *99*, 1693–1696.
(91) Stojaković, J.; Farris, B. S.; MacGillivray, L. R. *Chem Commun* 2012, *48*, 7958–7960.
(92) Sheikh, A. Y.; Rahim, S. A.; Hammond, R. B.; Roberts, K. J. *CrystEngComm* 2009, *11*, 501–509.
(93) Brittain, H. G. *Cryst. Growth Des.* 2011, *11*, 2500–2509.
(94) Aitipamula, S.; Chow, P. S.; Tan, R. B. H. *CrystEngComm* 2014.
(95) Gu, C.-H.; Young, V.; Grant, D. J. W. *J. Pharm. Sci.* 2001, *90*, 1878–1890.
(96) Miller, J.; Collman, B.; Greene, L.; Grant, D.; Blackburn, A. *Pharm. Dev. Technol.* 2005, *10*, 291–297.
(97) Zhang, G. G. Z.; Henry, R. F.; Borchardt, T. B.; Lou, X. 2007, *96*, 990–995.
(98) Takata, N.; Shiraki, K.; Takano, R.; Hayashi, Y.; Terada, K. *Cryst. Growth Des.* 2008, *8*, 3032–3037.
(99) Levilain, G.; Coquerel, G. *CrystEngComm* 2010, *12*, 1983.
(100) Eicke, M. J.; Levilain, G.; Seidel-Morgenstern, A. *Cryst. Growth Des.* 2013, *13*, 1638–1648.
(101) Viedma, C.; Ortiz, J. E.; de Torres, T.; Izumi, T.; Blackmond, D. G. *J. Am. Chem. Soc.* 2008, *130*, 15274–15275.
(102) Viedma, C.; Verkuijl, B. J. V; Ortiz, J. E.; de Torres, T.; Kellogg, R. M.; Blackmond, D. G. *Chemistry* 2010, *16*, 4932–4937.
(103) Viedma, C.; Noorduin, W. L.; Ortiz, J. E.; de Torres, T.; Cintas, P. *Chem. Commun. (Camb).* 2011, *47*, 671–673.
(104) Viedma, C.; Cintas, P. *Chem. Commun. (Camb).* 2011, *47*, 12786–12788.
(105) Noorduin, W. L.; Izumi, T.; Millemaggi, A.; Leeman, M.; Meekes, H.; Van Enckevort, W. J. P.; Kellogg, R. M.; Kaptein, B.; Vlieg, E.; Blackmond, D. G. *J. Am. Chem. Soc.* 2008, *130*, 1158–1159.
(106) Noorduin, W. L.; Kaptein, B.; Meekes, H.; van Enckevort, W. J. P.; Kellogg, R. M.; Vlieg, E. *Angew. Chem. Int. Ed. Engl.* 2009, *48*, 4581–4583.
(107) Van Der Meijden, M. W.; Leeman, M.; Gelens, E.; Noorduin, W. L.;

Meekes, H.; Van Enckevort, W. J. P.; Kaptein, B.; Vlieg, E.; Kellogg, R. M. *Org. Process Res. Dev.* 2009, *13*, 1195–1198.
(108) Noorduin, W. L.; Van Der Asdonk, P.; Bode, A. A. C.; Meekes, H.; Van Enckevort, W. J. P.; Vlieg, E.; Kaptein, B.; Van Der Meijden, M. W.; Kellogg, R. M.; Deroover, G. *Org. Process Res. Dev.* 2010, *14*, 908–911.
(109) Spix, L.; Alfring, A.; Meekes, H.; Van Enckevort, W. J. P.; Vlieg, E. *Cryst. Growth Des.* 2014, *14*, 1744–1748.
(110) Spix, L.; Meekes, H.; Blaauw, R. H.; van Enckevort, W. J. P.; Vlieg, E. *Cryst. Growth Des.* 2012, *12*, 5796–5799.
(111) Freyer, M. W.; Lewis, E. A. *Methods Cell Biol.* 2008, *84*, 79–113.

Part II

Methodology

1. Cocrystal screening

Cocrystals were synthesized by solvent-drop grinding of equimolar mixtures of Levetiracetam or Etiracetam and a suitable coformer. In practice, about 90 mg of API are placed in a 2 mL eppendorf with an equimolar amount of coformer, with addition of 5-10 µL of methanol and 3 stainless steel balls (RETSCH – 5 to 10 mm). Samples are then ground in a RETSCH Mixer Mill MM 400 for 90 min with a beating frequency of 30 Hertz to ensure complete conversion. Note that when mixtures resulted in formation of an amorphous phase, dry grinding was performed instead, using the same conditions.

The resulting powders are characterized using XRPD. Comparison of the resulting diffractogram with the diffraction patterns of the pure phases is used to indicate cocrystal formation. Existence of a cocrystal is suspected in particular when there are peaks on the cocrystal diffractogram that do not correspond to peaks of the pure API/coformer, usually at small 2θ angles (Figure 1). For that matter, all possible known forms of the pure phases should be considered, to avoid confusing cocrystal formation with other phase transformations (e.g. solvate formation, polymorphism).

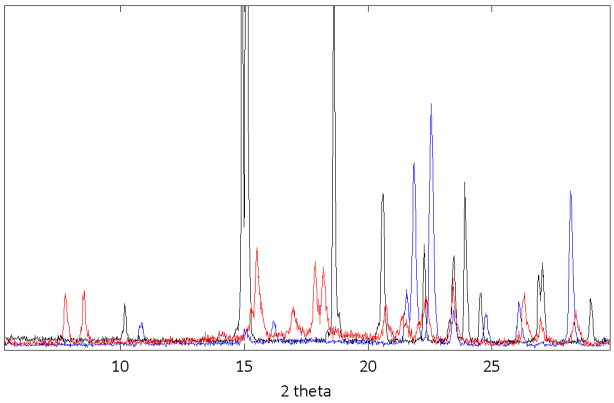


Figure 1. Comparison of the diffractogram corresponding to Levetiracetam (black), alpha-ketoglutaric acid (blue) and their ground mixture. Existence of a cocrystal is attested by the presence of isolated red peaks at small angles.

Conversion of the starting materials into a cocrystal is complete when there are peaks of the API/coformer with no correspondance on the cocrystal diffractogram. Note however that when there is still traces of only one pure phase in the cocrystal mixture, this may attest the existence of stoichiometrically diverse cocrystals. In such cases, different stoichiometries may be tested until finding the one leading to full conversion.

Note finally that some compounds may deteriorate during the grinding and lead to new phases that are no cocrystal. In consequence, it is recommended to check by routine NMR that the new compound formed is still made up of the initial materials only. If one can make an educated guess on the decomposition products, one can also compare the XRPD diffractograms of the new phase to the ones of these products (if available).

To confirm the existence of any cocrystal, attempts are then made to obtain a single crystal sufficiently large for structural XRD analysis.

2. Single crystal preparation

Different techniques may be used to grow single crystals from solution. In the context of this thesis, we mainly performed slow evaporation or cooling experiments, in 2mL vials using different starting conditions. It is indeed recommended to run several trials in parallel to increase the likelihood of getting a single crystal of sufficient quality, especially when there is no a priori information about coformer solubilities. Variations may notably concern the stoichiometric ratio (equimolar or not, as described in section 4.1 of the introduction) and the solvent nature; acetonitrile, ethyl acetate, methanol/ethanol and acetone being suggested in particular for organic compounds.

When solubilities are known, one procedure is especially efficient. It consists in preparing solutions characterized by coformer concentrations equal to their solubilities at 35°C, heat them to dissolve all the material and then let them cool at room temperature (or below if necessary) to allow crystal growth.

Coformer solubilities may be accessed by performing solubility curves. For that matter, one has to prepare suspensions of different concentrations at room temperature and identify the temperature at which all material is dissolved, by progressively increasing the temperature of the heater on which they are positioned (either automatically, using a Crystal16 device or equivalent, or manually using a standard thermoshaker).

In all types of crystallization, seeding the supersaturated solution with very small amounts of the corresponding cocrystal obtained by grinding is often necessary to provoke its nucleation.

3. Slurry preparation[1–4]

Slurry crystallization, also called solution-mediated phase transformation or aging experiments, may be used to evaluate if a cocrystal shows congruent dissolution in a given solvent or to identify the most stable phase in case of a polymorphic system.[a] In practice, it consists in suspending a small amount of the cocrystal in a solvent for a period of time typically ranging from 12h to 2 weeks.

If there are no traces of the corresponding coformers after this period, it is said that the cocrystal dissolves congruently in the solvent. This happens when coformers have comparable solubilities in a solvent. In that case, one may use stoichiometric solutions to grow the associated single crystal.

In case of polymorphism, conversion of the less stable form is due to the lower solubility of the most stable polymorph in the solvent. Indeed, the suspension of the metastable polymorph is supersaturated with respect to the more stable one and crystallization of this latter will lead to dissolution of the less stable form.

This process, however, is limited by the free energy barrier for nucleation of the stable phase, which is influenced by both internal (molecular flexibility and crystal structure) and external parameters (nature of the solvent, temperature, agitation, presence of impurities and/ additives…).[3] Among these factors, solubility has been proved to be of primary importance.[2] In routine experiments, solvents with various polarities are thus often tested in parallel and mixture of solvents may be useful when solubility is too low to ensure high enough induction time.

This technique may also be used to confirm the absence of any cocrystal for a given system. For that matter, small amounts of both coformers have to be suspended in a solvent in 1:1 ratio or in a composition corresponding to the suspected form. Then, if a cocrystal exists, its nucleation and subsequent growth will deplete the amount of pure coformers from the solution and lead to their dissolution. Ultimately, this process will lead to total conversion of the coformers into the cocrystal if the cocrystal shows congruent dissolution. If no cocrystal exists, on the contrary, only the separated coformers remain.

[a] Note that the relative stability of cocrystal polymorphs may also be assessed by comparing their melting points using differential scanning calorimetry.

4. References

(1) Gu, C.-H.; Young, V.; Grant, D. J. W. *J. Pharm. Sci.* 2001, *90*, 1878–1890.

(2) Miller, J.; Collman, B.; Greene, L.; Grant, D.; Blackburn, A. *Pharm. Dev. Technol.* 2005, *10*, 291–297.

(3) Zhang, G. G. Z.; Henry, R. F.; Borchardt, T. B.; Lou, X. 2007, *96*, 990–995.

(4) Takata, N.; Shiraki, K.; Takano, R.; Hayashi, Y.; Terada, K. *Cryst. Growth Des.* 2008, *8*, 3032–3037.

Part III

Results and Discussion

Chapter 1. Does Chirality Influence the Tendency towards Cocrystal Formation?

George, F.; Tumanov, N.; Norberg, B.; Robeyns, K.; Filinchuk, Y.; Wouters, J.; Leyssens, T. *Cryst. Growth Des.* **2014**, *14*, 2880–2892.

Abstract: We performed a systematic cocrystal search for the enantiopure and racemic version of a selected API, expecting that a coformer giving a cocrystal with a single enantiomer will also interact with the racemic mixture since they present identical functional groups prone to cocrystallization. We identified several novel cocrystals of Levetiracetam and its racemic equivalent, Etiracetam, using a wide variety of non-chiral coformers. 14 novel cocrystals of the enantiopure compound were obtained, whereas 18 of the racemic compound were identified. Out of these, 13 share a common coformer. A structural analysis indicates that in most cases the strongest hydrogen bonding interactions occur both in the enantiopure, as well as the racemic cocrystal, whereas Vander Waals interactions, or less strong secondary hydrogen bonding interactions, lead to a differentiation of the final structure. Based on our work we suggest an approach that could lead to a more optimal cocrystal screening of an enantiopure compound, especially when a limited amount of this compound is available. Starting with a screen of the racemic compound, the set of possible coformers for the enantiopure screen can be limited to those yielding a positive hit in the former screen. Doing so, our example shows an increase in efficiency from 10% to 72%.

1.1 Introduction

One of the goals of a crystallization process is to obtain a pure form, especially when working with active pharmaceutical ingredients (APIs).[1] Different forms of these latter may be accessible, among them polymorphs of the pure compound,[2,3] salts,[4,5] hydrates, [6,7] cocrystals.[8–10] Forms may differ significantly from one another in their physico-chemical properties, such as solubility or bioavailability,[11,12] justifying the efforts towards finding API forms with optimal properties without altering chemical identity.[13,14]

A multitude of drug substances are marketed as salts but for some APIs, especially ones without ionizable groups, no viable salts can be isolated. In this case, formation of cocrystals is an interesting alternative.[15–18] In the context of this work, cocrystals are considered as solids that are crystalline single phase materials composed of two or more different molecular and/or ionic compounds, called coformers, generally in a stoichiometric ratio, with components being solid under ambient conditions. They are therefore neither solvates nor salts.[19]

Identifying appropriate coformers able to cocrystallize with a given API may be time consuming (trial and error process), and compound consuming, which is particularly an issue at early stages of drug research. Consequently, attempts are made to predict matching coformers for a given compound, to improve cocrystal screening efficiency. Within this context, Springuel *et al.* showed that molecules presenting relatively similar chemical structures have an increased likelihood to form cocrystals with the same coformers.[20] In this contribution, we use a similar approach, investigating whether or not the likelihood of forming a cocrystal between an enantiopure API and a coformer is increased when this coformer was already shown to be effective towards the racemic compound. For salts this does not seem surprising, as salt formation is mainly governed by pK_a differences between base and acid. Hence an acid (or base) interacting with a given enantiopure API is also expected to interact with the racemic API. For cocrystals, this is less straightforward, as these compounds mostly depend on hydrogen-bonding interactions, which are directional in nature. In a recent contribution,[21] cocrystals were shown to behave differently with respect to salts when both components are chiral: whereas for salts, formation of a diastereomeric pair seems to be the general rule, cocrystal systems are mostly enantiospecific.

To investigate how chiral APIs respond in cocrystallization towards non-chiral coformers, we performed an experimental cocrystal screen on an

enantiopure and racemic version of a pharmaceutically active compound that does not form a salt. As a model compound we chose Levetiracetam, a nootropic drug used as an anticonvulsant in the treatment of epilepsy.[22,23] Levetiracetam is the biologically active enantiomeric form of (RS)-2-(2-oxopyrrolidin-1-yl) butanamide, also called Etiracetam (Figure 1). Of the former, only one solid form is reported,[24] whilst for the racemic compound two enantiotropically related polymorphs and one hydrate have been studied extensively. [25–29]

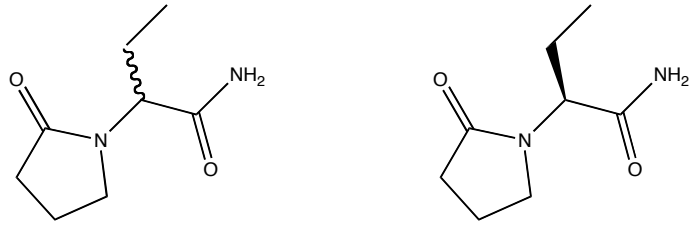


Figure 1. Chemical structures of Etiracetam (left) and Levetiracetam (right).

Previous work showed Levetiracetam to be a potent candidate for cocrystal formation.[20,30] In the current contribution, cocrystals of Levetiracetam and Etiracetam were identified using solvent drop grinding. X-ray powder diffraction (XRPD) was used to detect cocrystals while a structural comparison was carried out to compare the hydrogen-bonded network topology of enantiopure and racemic cocrystals, using crystal structures obtained from single crystal diffraction or synchrotron powder diffraction data.

1.2 Experimental Section

Starting Materials. S-2-(2-oxopyrrolidin-1-yl)butanamide (Levetiracetam, LEVI) was purchased from Xiamen Top Health Biochem Tech. Co., Ltd. 4-nitrobenzoic acid (4NBA); adipic acid; oxalic acid (OXA) and acetoacetamide were purchased from Acros Organics. 1,3,5-benzenetricarboxylic acid; 2,2-dimethylsuccinic acid (DMSA); 5-hydroxylisophthalic acid; oxaloacetic acid; methyl 3,4,5-trihydroxybenzoate and 4,4-bipyridine were purchased from Alfa Aesar. 3-methylbutanamide was purchased from Maybridge. 1H-Pyrazole-3,5-dicarboxylic acid monohydrate; 2,4-dihydroxybenzoic acid (DHBA); 3-nitrobenzoic acid (3NBA); 5-nitroisophtalic acid; anthranilamide; 4-chlorobenzaldehyde was purchased from Sigma-Aldrich. Citraconic acid and gallic acid ethyl ester were purchased from TCI. These materials were used as received, without further purification. (RS)-2-(2-oxopyrrolidin-1-yl)butanamide (Etiracetam, ETI) was prepared by racemization of S-2-(2-

oxopyrrolidin-1-yl)butanamide. 10g of S-2-(2-oxopyrrolidin-1-yl)butanamide together with catalytic amount (0.05 eq) of MeONa were added to 10 ml of MeOH. The solution was kept at reflux under continuous stirring for 24h, and then cooled to room temperature. The compound crystallizes spontaneously. After filtration, the compound was washed twice with MeOH. The recovered compound was used as such.

Cocrystal Screen. Cocrystals were synthesized by solvent-drop grinding of equimolar mixtures of Levetiracetam or Etiracetam and a suitable coformer, with addition of 10 µL of methanol. Samples were ground in a RETSCH Mixer Mill MM 400 for 90 min with a beating frequency of 30 Hertz. The resulting powders were characterized using XRPD. Comparison of the resulting diffraction pattern with the diffraction patterns of the pure phases was used to indicate cocrystal formation. All possible known forms of the pure phases were considered, to avoid confusing cocrystal formation with other phase transformations (e.g. solvate formation polymorphism).

When mixtures resulted in formation of an amorphous phase, neat (dry) grinding was performed instead. For all cocrystals identified, attempts were made to obtain a single crystal sufficiently large for structural XRD analysis.

X-ray Powder Diffraction (XRPD). X-ray diffraction measurements were performed on a Siemens D5000 diffractometer equipped with a Cu X-ray source operating at 40 kV and 40 mA and a secondary monochromator allowing selection of the Kα radiation of Cu (λ = 1.5418 Å). A scanning range of 2θ values from 2° to 72° at a scan rate of 0.6° min^{-1} was applied.

Synchrotron Radiation Powder X-ray Diffraction. Data for Levetiracetam - 2,4-dihydroxybenzoic acid and Levetiracetam- 2,2-dimethylsuccinic acid cocrystals were collected at the MS-X04SA beamline at the Swiss Light Source (PSI, Switzerland) using a 1D microstrip detector MYTHEN II. The wavelength was 0.775045 Å. Wavelength and zero-shift were calibrated using a NIST 640d Si standard sample. Patterns were indexed using FOX.[31] Le Bail fit was done using FOX[31] and Fullprof[32] programs. Space group for each pattern was suggested taking into account analysis of systematic absences and considering that space groups have to be chiral due to presence of a chiral API molecule in the unit cell. The structure was solved by global optimization in direct space using the FOX program.[31] Molecular model of Levetiracetam,[24] 2,2-dimethylsuccinic acid[33] and 2,4-dihydroxybenzoic acid (using the corresponding fragment in Etiracetam-2,4-dihydroxybenzoic acid cocrystal structure) were imported from structural models determined from single-crystal diffraction. Constraints for interatomic distances and angles

were introduced: rigid group restraints were created to keep proper geometry of planar rings, amide and carboxylic acid groups. Antibump restraints were not applied. Position, orientation and conformation of molecules were optimized. In average, one out of ten trials leads to a reasonable solution, so 200-300 runs were performed to find the optimal solution. At the final stages of solution, torsion angle restraints were applied to hydrogen atoms involved in hydrogen bonding to achieve a reasonable geometry for the hydrogen bonds. The final refinement was done by the Rietveld method using the Fullprof suite,[32] keeping the orientation of the molecules fixed. Unit cell parameters, coordinates of two separate molecules, scale factor and overall temperature factor were refined. The background was described by linear interpolation between selected points determined from Le Bail fit. The final discrepancy factors are: R_I = 12.37 %, R_p = 1.33 %, R_{wp} = 1.53 %, R_{cp} = 17.03 % and R_{cwp} = 18.22 % for the Levetiracetam -2,4-dihydroxybenzoic acid cocrystal and R_I = 8.75 %, R_p = 1.25 %, R_{wp}= 1.55 %, R_{cp} = 16.91 % and R_{cwp} = 17.37 % for the Levetiracetam-2,2-dimethylsuccinic acid cocrystal. Their refinement profiles are shown in the Supporting Information.

Using XRPD data it is nearly impossible to differentiate cocrystals from salts. However, in the context of this work, we used a compound that does not form salts. This resolves all issues concerning the position of the hydrogen atoms.

Single Crystals. Single crystals were grown by slow evaporation from an equimolar solution of starting materials or by cooling of the corresponding supersaturated solution in suitable solvents. Different solvents were used to increase the probability of identifying single crystals large enough for analysis.

Single Crystal X-ray Diffraction. Single crystal X-ray diffraction was performed on a Gemini Ultra R system (4-circle kappa platform, Ruby CCD detector) using Cu Kα radiation (λ = 1.54056 Å). Cell parameters were estimated from a pre-experiment run and full data sets collected at room temperature. The structures were solved by direct methods with the SHELXS-97 program and then refined on $|F|^2$ using SHELXL-97 software.[34] Non-hydrogen atoms were anisotropically refined, and the hydrogen atoms (not implicated in H-bonds) in the riding mode with isotropic temperature factors were fixed at 1.2 times U(eq) of the parent atoms (1.5 times for methyl groups). Hydrogen atoms implicated in H-bonds were localized in the Fourier difference maps.

1.3 Results

Hydrogen bonding is the main driving force behind cocrystal formation. As most of the typical cocrystal synthons are not affected by chirality, one could expect that a coformer yielding a cocrystal with a given chiral compound would also interact with its racemic version, as similar hydrogen bonding synthons are expected. However, in case of cocrystals between two chiral partners[21], it was shown that the Van der Waals or π-π stacking interactions could be as important as the hydrogen bonding interactions, and influence cocrystal stability.

To investigate the propensity towards cocrystal formation between a non-chiral partner and enantiopure or racemic compound respectively, we performed a cocrystal screen for both Etiracetam (racemic) and Levetiracetam (enantiopure) with a set of 152 achiral coformers (see the appendices for the corresponding list). These latter were selected to have at least one alcohol, aldehyde, amide, amine or carboxylic acid function.

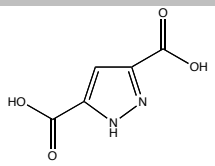
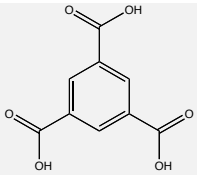
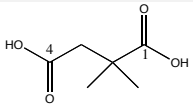
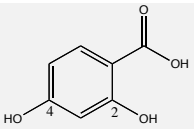
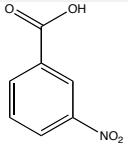
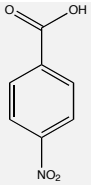
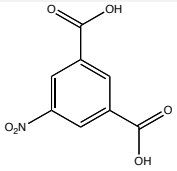
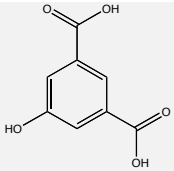
1.3.A Cocrystal Screening

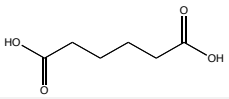
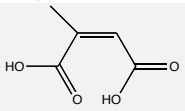
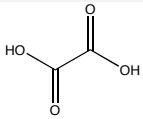
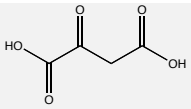
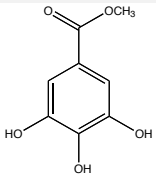
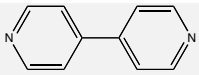
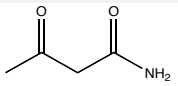
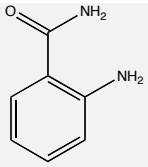
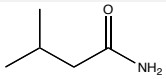
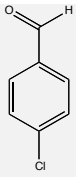
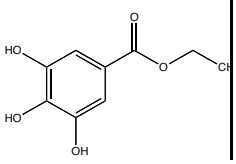
In total, 14 novel cocrystals of Levetiracetam were identified, which gives a success rate for cocrystal formation of about 10%. A similar number of cocrystals was also encountered for the racemic Etiracetam compound, yielding a total of 18 novel cocrystals (Table 1). Strikingly, out of these, 13 involve a common coformer.[a]

Although a negative test through grinding does not necessarily imply the cocrystal phase does not exist, solvent-drop grinding has been shown to be an efficient method for a fast and effective detection of the majority of cocrystals.[35] It is however possible that some of the coformers that were excluded in the initial screen do lead to cocrystal formation under different conditions (eg. solvent screening, supercritical fluid crystallization, *etc*). Similarly, in some cases stoichiometrically diverse cocrystals can exist. In the context of this work, we were not concerned with identifying all possible cocrystal forms between two given compounds. However, when it was evident that grinding of an equimolar mixture, led to excess amount of one of the two components, different stoichiometries were tested (eg. 1:2 mixing was done for 3-nitrobenzoic acid and 4-nitrobenzoic acid).

[a] These results do not include the cocrystals formed between ETI and LEVI with alpha-ketoglutaric acid, which are described in the next chapter of this thesis.

Table 1. Overview of Etiracetam and Levetiracetam Cocrystals.

Coformers		ETi Cocrystals	LEVI Cocrystals
1H-Pyrazole-3,5-dicarboxylic acid monohydrate		1 : 1	---[a]
1,3,5-benzenetricarboxylic acid		1 : 1	---[a]
2,2-dimethylsuccinic acid (DMSA)		1 : 1	1 : 1
2,4-dihydroxybenzoic acid (DHBA)		1 : 1	1 : 1
3-nitrobenzoic acid (3NBA)		1 : 1	1 : 2
4-nitrobenzoic acid (4NBA)		1 : 2	1 : 2
5-nitroisophtalic acid		* [b]	1 : 1
5-hydroxylisophthalic acid		*[c]	*[c]

Adipic acid		1 : 1	--- [c]
Citraconic acid		1 : 1	1 : 1
Oxalic acid (OXA)		1 : 1	1 : 1
Oxaloacetic acid		*	*
Methyl 3,4,5-trihydroxybenzoate		---	*
4,4-bipyridine		* [c]	--- [a]
Acetoacetamide		*	*
Anthranilamide		1 : 1	--- [c]
3-methylbutanamide		*	*
4-chlorobenzaldéhyde		*	*
Gallic acid ethyl ester		* [c]	* [c]

--- no cocrystal, * cocrystal identified, structure not yet determined. [a] absence of cocrystal confirmed by slurrying experiment. [b] solvate found. [c] cocrystal confirmed by binary melting phase diagram.

Out of the 32 novel cocrystals identified in this work, 15 have been structurally characterized, either through single crystal analysis or through XRPD structure determination. A pair-wise comparison of enantiopure and racemic cocrystals can yield insight into the question, whether or not similar synthons are observed in both cases.

1.3.B Crystal Structure Analysis

The structural analysis below focuses on the five coformers for which a cocrystal was found for both Etiracetam and Levetiracetam, allowing a comparison of the number and types of H bonding patterns formed in a pair of cocrystals. Crystallographic parameters for those ten cocrystals are displayed in Table 2.

ETI-DMSA, ETI-OXA, ETI-3NBA and ETI-4NBA cocrystals were synthesized by slow evaporation at room temperature of an equimolar, under-saturated solution of both coformers. ETI-DHBA, LEVI-OXA, LEVI-3NBA and LEVI-4NBA were obtained by cooling crystallization to 3°C. Finally, structures of LEVI-DMSA and LEVI-DHBA cocrystals were determined from synchrotron powder diffraction data.

Table 2. Crystallographic Parameters for Cocrystal Pairs.

Cocrystals	ETI-DMSA	LEVI-DMSA	ETI-DHBA	LEVI-DHBA	ETI-OXA	LEVI-OXA
Structural formula	$(C_8H_{14}N_2O_2)$ $(C_6H_{10}O_4)$	$(C_8H_{14}N_2O_2)$ $(C_6H_{10}O_4)$	$(C_8H_{14}N_2O_2)$ $(C_7H_6O_4)$	$(C_8H_{14}N_2O_2)$ $(C_7H_6O_4)$	$(C_8H_{14}N_2O_2)$ $(C_2H_2O_4)$	$(C_8H_{14}N_2O_2)$ $(C_2H_2O_4)$
Formula weight (g/mol)	316.35	316.35	324.33	324.33	260.25	260.25
Space system	monoclinic	monoclinic	monoclinic	orthorhombic	monoclinic	monoclinic
Space group	$P2_1/c$	$P2_1$	$P2_1/n$	$P2_12_12_1$	$P2_1/c$	$P2_1$
a (Å)	12.9082	6.23877(7)	6.2828(3)	6.60584(3)	5.7253(2)	5.7646(4)
b (Å)	11.5417	11.55362(12)	13.3843(6)	13.11919(7)	20.2528(7)	19.8963(14)
c (Å)	12.5179	11.34360(12)	18.7984(8)	18.80723(8)	10.9457(5)	11.0690(8)
α (°)	90	90	90	90	90	90
β (°)	116.342	90.9259(9)	96.638(4)	90	100.085(4)	99.331
γ (°)	90	90	90	90	90	90
V ($Å^3$)	1671.3	817.544	1570.18	1629.9	1249.58	1252.75
Z	4	2	4	4	4	4
R-factor (%)	4.39	/	3.99	/	4.13	4.57
d ($g.cm^{-3}$)	1.257	1.285	1.372	1.322	1.383	1.380
Technique	SC XRD[a]	P XRD[b]	SC XRD[a]	P XRD[b]	SC XRD[a]	SC XRD[a]

[a] Single Crystal XRD. [b] Powder XRD

Table 2. continued

Cocrystals	ETI-3NBA	LEVI-3NBA	ETI-4NBA	LEVI-4NBA
Structural formula	$(C_8H_{14}N_2O_2)$ $(C_7H_5NO_4)$	$(C_8H_{14}N_2O_2)$ $2(C_7H_5NO_4)$	$(C_8H_{14}N_2O_2)$ $2(C_7H_5NO_4)$	$(C_8H_{14}N_2O_2)$ $2(C_7H_5NO_4)$
Formula weight (g/mol)	337.33	504.45	504.45	504.45
Space system	triclinic	orthorhombic	monoclinic	triclinic
Space group	*P*-1	$P2_1\,2_1\,2_1$	$P2_1/c$	*P*1
a (Å)	5.8455(2)	8.0511(4)	24.2697(8)	7.0988(9)
b (Å)	11.6845(12)	13.0536(6)	7.4683(2)	7.3613(10)
c (Å)	12.5762(10)	23.0243(9)	13.3230(5)	12.1428(13)
α (°)	78.856(8)	90	90	77.496(10)
β (°)	82.443(6)	90	104.663(4)	88.250(10)
γ (°)	84.035(7)	90	90	74.232(12)
V ($Å^3$)	832.734	2419.76	2336.19	595.906
Z	2	4	4	1
R-factor (%)	5.26	4.12	5.26	5.4
d ($g.cm^{-3}$)	1.345	1.385	1.434	1.406
Technique	SC XRD[a]	SC XRD[a]	SC XRD[a]	SC XRD[a]

[a] Single Crystal XRD. [b] Powder XRD

1.3.B.1 2,2-dimethylsuccinic acid as coformer

Etiracetam- 2,2-dimethylsuccinic acid (1:1) (ETI-DMSA) The ETI-DMSA cocrystal crystallizes from acetonitrile in the monoclinic space group $P2_1/c$. The asymmetric unit contains one molecule of Etiracetam (ETI) and one molecule of 2,2-dimethylsuccinic acid (DMSA).

An amide-carboxylic acid heterosynthon, described in graph set notation as $R^2_2(8)$,[36] is formed between the noncyclic amide of ETI and the carboxylic acid function in position 4 of DMSA (Figure 2a). In addition, the carboxylic acid function in position 1 links two molecules of ETI through a $C^2_2(11)$ hydrogen-bonded infinite chain motif: the acid function of DMSA acts as an acceptor of the second proton of the noncyclic amide function (the one not included in the ring motif mentioned above) and as a donor to the cyclic carbonyl function of ETI. This chain propagates along the c-axis (Figure 2b).

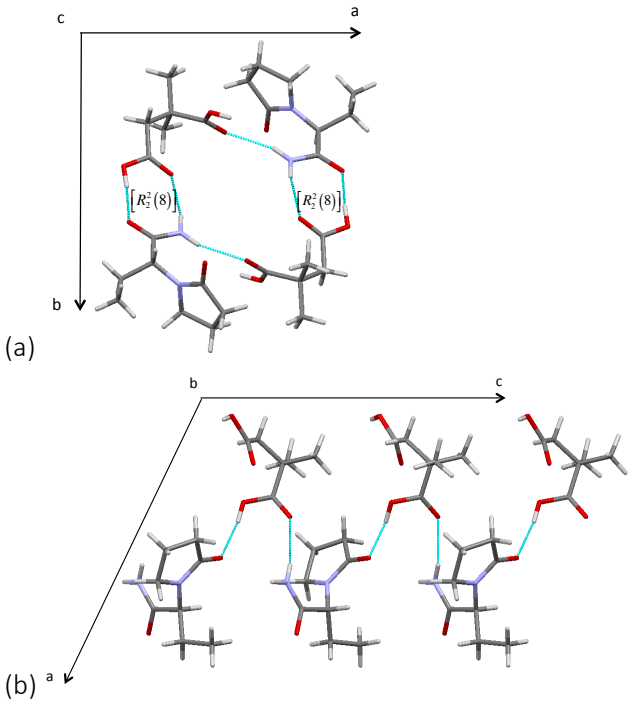


Figure 2. (a) Etiracetam and 2,2-dimethylsuccinic acid form an amide-carboxylic acid heterosynthon $R^2_2(8)$ ring motif and (b) a $C^2_2(11)$ infinite chain motif in ETI-DMSA cocrystal.

On the whole, the ETI-DMSA cocrystal exhibits a layered stacking; all the hydrogen bonds being located inside each layer. Secondary interactions hold the layers together due to the presence of the ethyl groups of ETI molecules and the carbonated chains of DMSA oriented outwards from the layer planes.

Levetiracetam- 2,2-dimethylsuccinic acid (1:1) (LEVI-DMSA) Structure of the LEVI-DMSA cocrystal was determined from synchrotron powder diffraction data in the monoclinic space group $P2_1$. The asymmetric unit contains one molecule of Levetiracetam (LEVI) and one molecule of 2,2-dimethylsuccinic acid (DMSA).

One molecule of DMSA is linked to four molecules of LEVI through 4 H-bonds, two with each carboxylic acid function. The carboxylic acid in position 1 has the carbonyl group bound to a hydrogen of the non-cyclic amide of a first molecule of LEVI, while the hydroxyl is connected to the cyclic carbonyl function of a second molecule of LEVI (Figure 3 a), leading to a similar $C^2_2(11)$ hydrogen-bonding infinite chain motif as observed for the racemic

ETI-DMSA cocrystal (Figure 3 b). In this first cocrystal comparison, similar motifs can hence be encountered between enantiopure and racemic cocrystals. In turn the carboxylic acid group in position 4 accepts a H-bond from the non-cylic amide of a third LEVI molecule while donating one to the non-cyclic amide carbonyl group of a fourth LEVI molecule, creating a second infinite chain motif $C^2_2(8)$, along the a-axis (Figure 3 c). In contrast with the racemic ETI-DMSA cocrystal, no amide-carboxylic acid heterosynthon $R^2_2(8)$ is observed, implying that not all hydrogen bonding patterns are comparable in both cases.

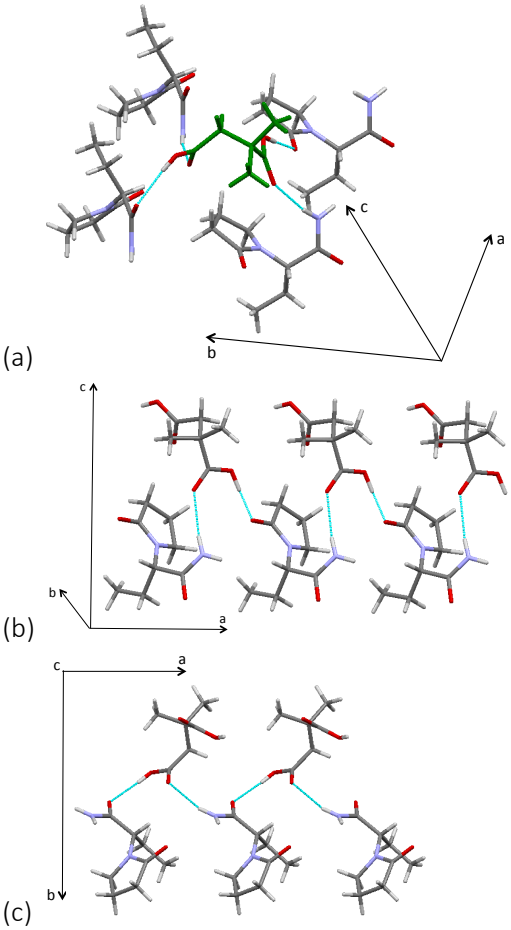


Figure 3. (a) One molecule of DMSA (green) is linked to four molecules of LEVI through four H-bonds in LEVI-DMSA cocrystal. LEVI-DMSA cocrystal shows (b) a $C^2_2(11)$ infinite chain motif and (c) an infinite chain with $C^2_2(8)$ motif.

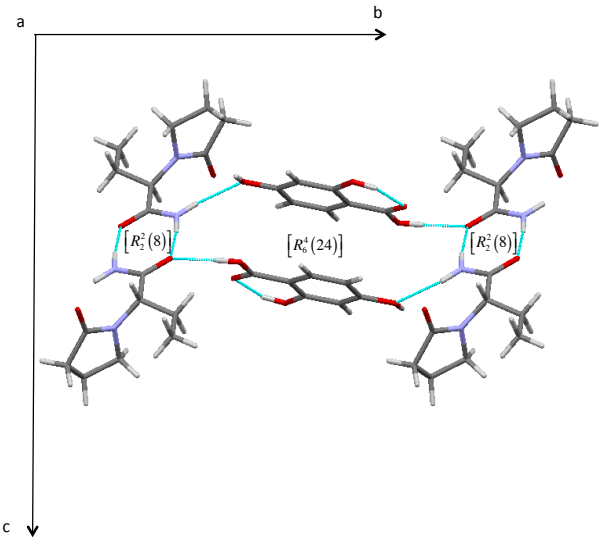
Overall, molecules of LEVI and DMSA are also organized in layers, which stack along the c axis.

1.3.B.2 2,4-dihydroxybenzoic acid as coformer

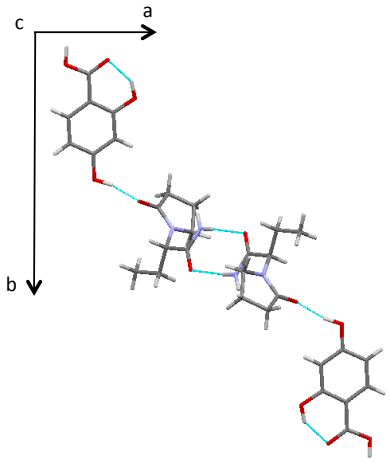
Etiracetam-2,4-dihydroxybenzoic acid (1:1) (ETI-DHBA) The ETI-DHBA cocrystal crystallized from acetonitrile and belongs to the monoclinic space group $P2_1/n$. One molecule of Etiracetam (ETI) and one molecule of 2,4-dihydroxybenzoic acid (DHBA) occupy the asymmetric unit.

Each DHBA molecule is locked in a near planar conformation through the formation of an intra-molecular H-bond between the carboxylic acid's carbonyl group and the hydrogen of the nearest hydroxyl group (in position 2).

An amide-amide homosynthon $R^2_2(8)$ is formed between the noncyclic amides of two ETI molecules (Figure 4 a). A Eti-DHBA dimer motif is also observed, based on a hydrogen bond between the DHBA hydroxyl group (in position 4) and the cyclic oxygen of ETI (Figure 4 b).



(a)



(b)

Figure 4. (a) The ETI-DHBA cocrystal shows an amide-amide homosynthon $R^2_2(8)$ between the non-cyclic amides of two ETI and a $R^4_6(24)$ ring motif and (b) a dimer motif links ETI to DHBA.

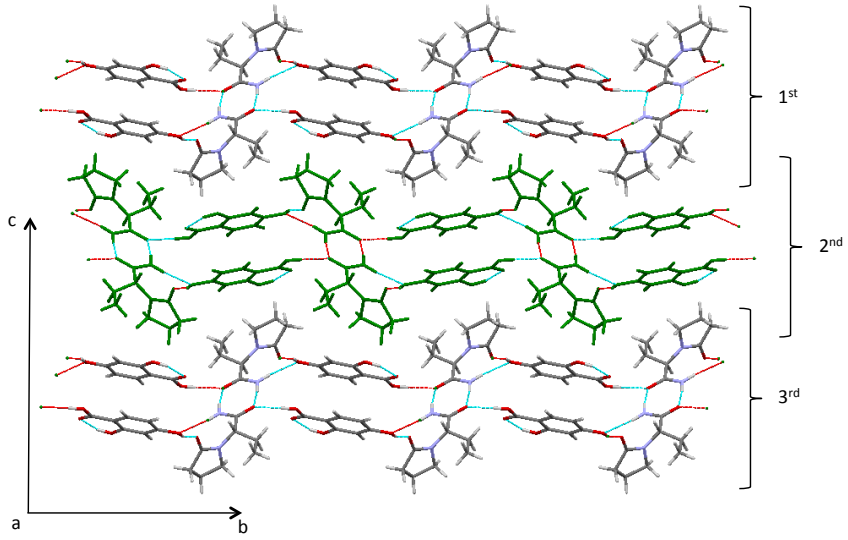


Figure 5. Stacking of three layers in ETI-DHBA cocrystal, showing infinite chain motifs $C^2_2(12)$ along the c axis and π stacking of molecules of DHBA acid in the *bc* plane.

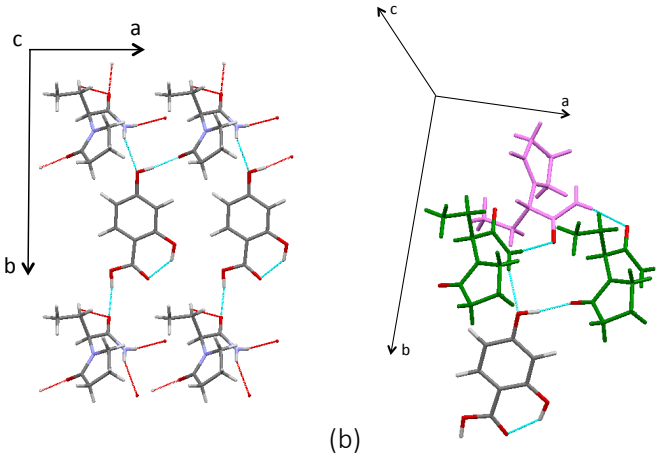
Here again, a $C^2_2(12)$ infinite chain motif is observed. This motif is formed through a first hydrogen bond donated by the DHBA carboxylic acid function to the oxygen atom of the non-cyclic amide. The second hydrogen bond

joins the noncyclic ETI amide with the DHBA hydroxyl group in position 4 (Figure 5).

A final structural element of interest is the π stacking of DHBA molecules along the c axis, with each successive molecule rotated by 180°C. Consequently, two consecutive infinite chains link through a $R^4_6(24)$ ring motif (Figure 4 a) In this motif, there are an uneven number of hydrogen acceptors and donors as the ETI oxygen atom of the noncylic amide is involved in two different hydrogen bonds. On the whole, a close stacking of layers along the c axis is observed, held together by hydrophobic interactions, due to the presence of non-polar groups oriented outwards.

Levetiracetam- 2,4-dihydroxybenzoic acid (1:1) (LEVI-DHBA) The LEVI-DHBA cocrystal was determined *ab initio* from XRPD data and belongs to the orthorhombic space group $P2_12_12_1$. The same cocrystal was crystallized from solution by Springuel *et al.* [20] with almost identical structural parameters, confirming the validity of our results originating from synchrotron radiation data analysis. Four molecules of Levetiracetam (LEVI) and four molecules of 2,4-dihydroxybenzoic acid (DHBA) occupy the unit cell.

As in the racemic ETI-DHBA cocrystal, each DHBA molecule is locked in a near planar conformation through the formation of an intra-molecular H-bond. Likewise, a similar dimer motif connects the DHBA hydroxyl group in position 4 to the LEVI cyclic amide (Figure 6 a). Furthermore, the structure also shows an infinite chain motif $C^2_2(12)$ although constructed from slightly different hydrogen bonds. A first hydrogen bond is formed between the donating DHBA carboxylic acid function and the LEVI noncyclic amide carbonyl function, while a second hydrogen bond joins the LEVI noncyclic amide to the DHBA oxygen of the hydroxyl in position 4 (Figure 6 a). This chain allows the growth of the network along the b axis and shows alternating LEVI and DHBA molecules.



(a) (b)

Figure 6. (a) An infinite chain motif $C^2_2(12)$ and a dimer motif links LEVI to DHBA acid in LEVI-DHBA cocrystal and (b) another dimer motif joins two LEVI (one pink, one green).

Although a lot of structural elements are once more similar between racemic and enantiopure cocrystals, contrary to the ETI-DHBA cocrystal, no homosynthon is found between the non-cyclic amides of two LEVI molecules. Instead a supplementary hydrogen bond exists between the LEVI non-cyclic carbonyl group and the proton of the non-cyclic amide of a neighboring LEVI molecule (Figure 6 b). As was the case for the racemic cocrystal, the carbonyl of the noncyclic amide is involved in two hydrogen bonds.

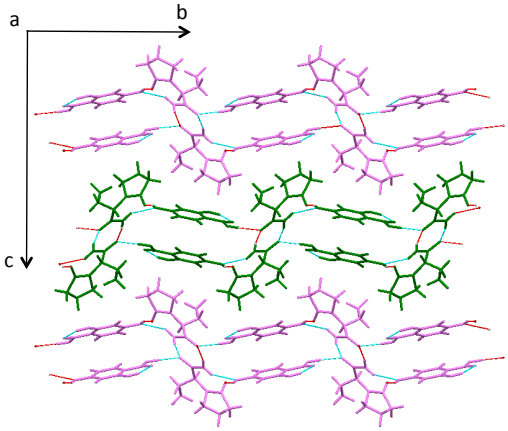


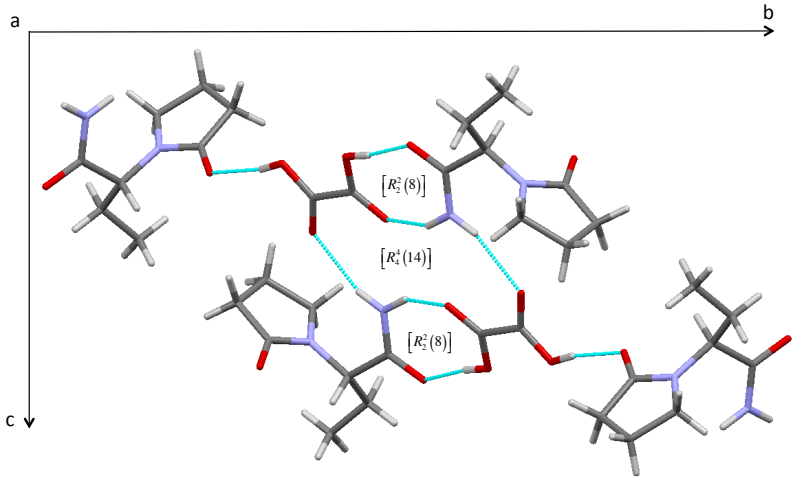
Figure 7. Stacking of three layers in LEVI-DHBA cocrystal, showing infinite chain motifs $C^2_2(12)$ along the b axis.

On the whole, a close stacking of layers along the c axis, can be observed. In a given layer, there is no π stacking of molecules of DHBA since those are slightly staggered along the a axis. Each layer has a thickness of two molecules of DHBA or LEVI, depending on the position along the b axis. Layers are held together by secondary interactions, due to the presence of the non-polar parts of LEVI molecules oriented outwards, facing the DHBA aromatic cycles. Figure 7 shows the stacking of three layers in the *bc* plane.

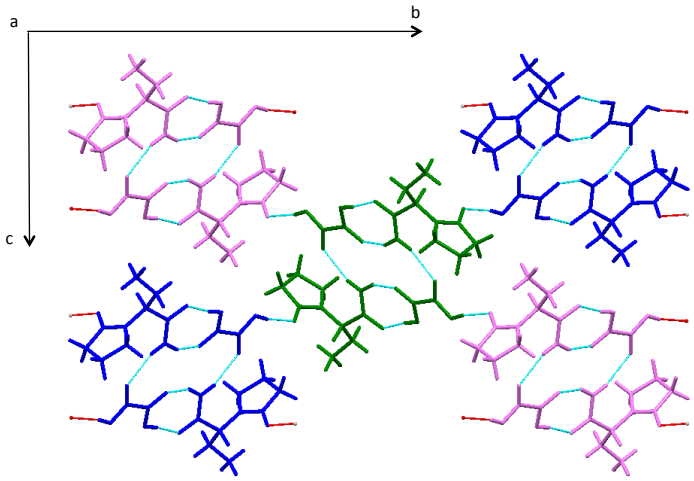
1.3.B.3 Oxalic acid as coformer

Etiracetam - oxalic acid (1 :1) (ETI-OXA) The ETI-OXA cocrystal crystallizes in the monoclinic space group $P2_1/c$. The asymmetric unit contains one molecule of Etiracetam and one molecule of oxalic acid (OXA).

An amide-carboxylic acid heterosynthon $R^2_2(8)$ is formed between the non-cyclic amide of ETI and one of the OXA carboxylic acid functions. In addition, the second hydrogen of the ETI non-cyclic amide acts as a hydrogen bond donor for the second carboxylic acid function. These two interactions link two ETI molecules with two OXA molecules forming a $R^4_4(14)$ ring motif (Figure 8 a).



(a)

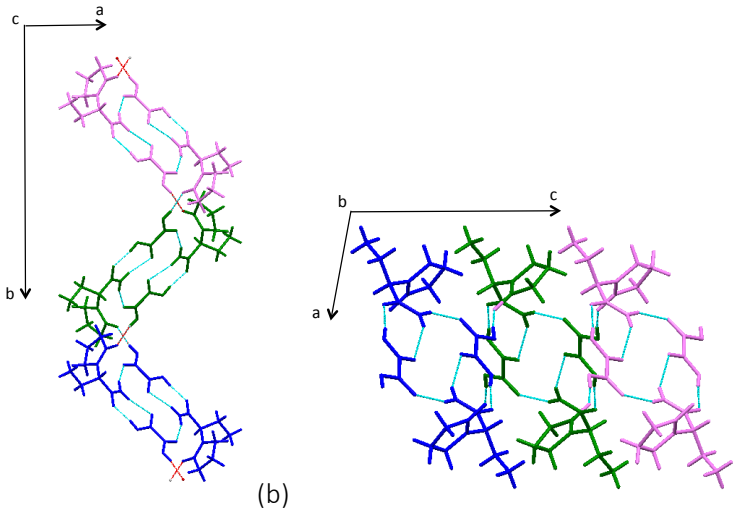


(b)

Figure 8. (a) An amide-carboxylic acid heterosynthon $R^2_2(8)$ is formed between ETI and OXA and a $R^4_4(14)$ ring motif links two ETI with two OXA, in ETI-OXA cocrystal. (b) One $R^4_4(14)$ ring motif (green) is linked to four identical ones (blue and pink) through hydrogen bonds.

This latter ring motif binds to four similar ring motifs through a hydrogen bond between the ETI cyclic amide carbonyl group and the OXA carboxylic acid hydroxyl function not participating in the aforementioned heterosynthon (Figure 8 b).

On the whole, the ETI-OXA structure is arranged in separate layers stacked along the a-axis. Layers are once more held together thanks to the hydrophobic groups of ETI pointing outwards. Figures 9 a and b show one layer in the *ab* plane and in the *ac* plane respectively.



(a) (b)

Figure 9. (a) One layer in the *ab* plane and in (b) *ac* plane in ETI-OXA cocrystal.

Levetiracetam- oxalic acid (1 :1) (LEVI-OXA) The enantiopure LEVI-OXA cocrystal crystallizes from acetonitrile in the monoclinic space group $P2_1$. The asymmetric unit contains one molecule of Levetiracetam (LEVI) and one molecule of oxalic acid (OXA).

LEVI-OXA cocrystal parameters are almost identical to those of the ETI-OXA cocrystal. Similar hydrogen bonding motifs and tridimensional patterns are observed (Figure 10).

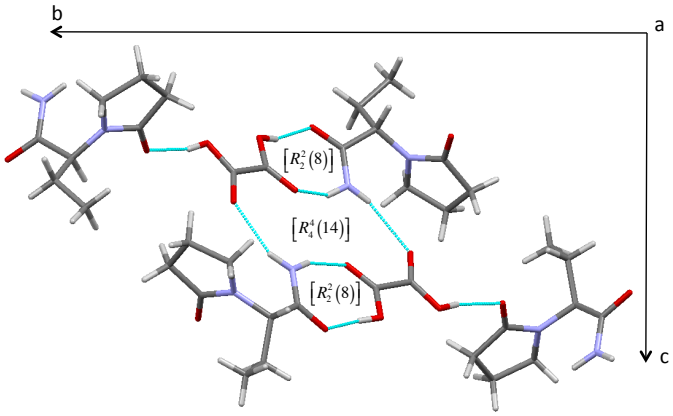
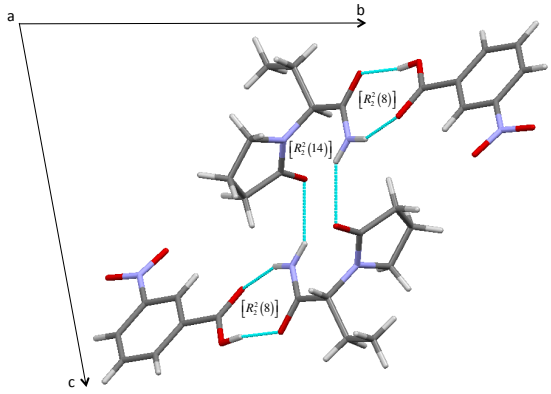


Figure 10. An amide-carboxylic acid heterosynthon $R^2_2(8)$ is formed between LEVI and OXA and a $R^4_4(14)$ ring motif links two LEVI with two OXA in LEVI-OXA cocrystal.

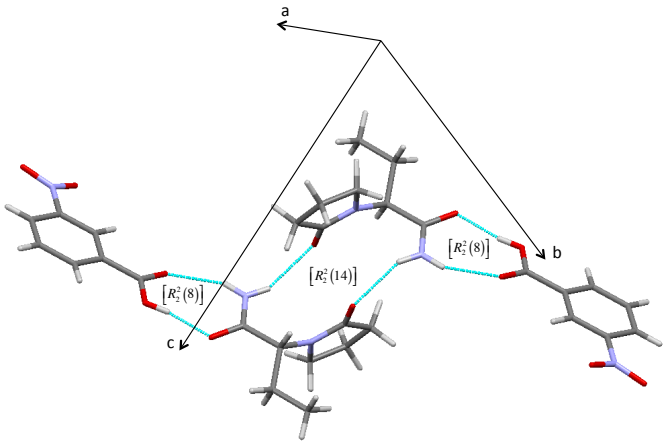
1.3.B.4 3-nitrobenzoic acid as coformer

Etiracetam- 3-nitrobenzoic acid (1:1) (ETI-3NBA) The ETI-3NBA cocrystal, isolated from acetonitrile, was solved in the triclinic space group *P*-1. Two molecules of Etiracetam (ETI) and two molecules of 3-nitrobenzoic acid (3NBA) occupy the unit cell, forming a tetramer.

In the unit cell, one observes the presence of an amide-carboxylic acid $R^2_2(8)$ heterosynthon between the ETI noncyclic amide and the 3NBA carboxylic acid (Figure 11 a). Furthermore, two ETI molecules form an amide $R^2_2(14)$ homosynthon ring motif: the cyclic carbonyl of one molecule acts as an acceptor to the noncyclic NH of a second, resulting in the folding of the tetramer into an S shape (Figure 11 b).



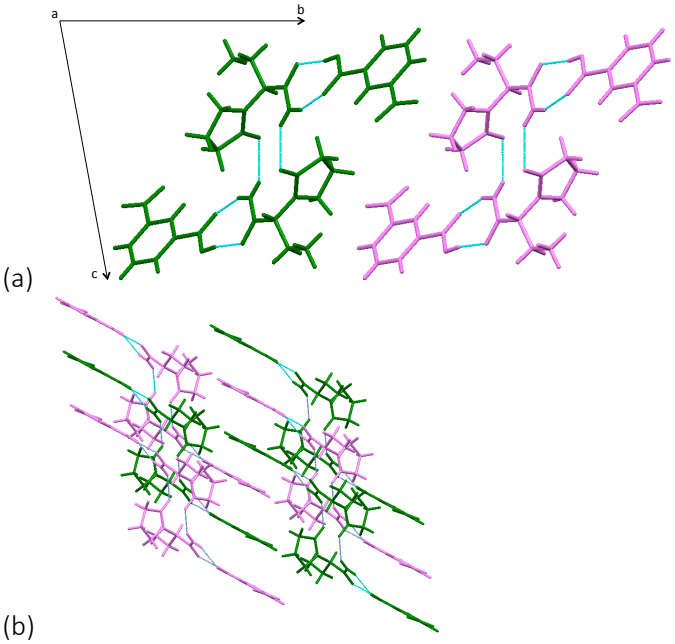
(a)



(b)

Figure 11. (a) The ETI-3NBA cocrystal exhibits an amide-carboxylic acid heterosynthon $R^2_2(8)$ between ETI and 3NBA and an amide homosynthon of $R^2_2(14)$ ring motif between two ETI (b) An S-shape tetramer formed by two ETI and two 3NBA.

Tetramers are stacked in staggered rows and are linked to one another by weak hydrophobic interactions between the phenyl and pyrrolidone groups directed outwards (Figure 12).



(a)

(b)

Figure 12. Stacking of tetramers (a) in the *bc* plane and (b) in staggered rows, in ETI-3NBA cocrystal.

Leviteracetam- 3-nitrobenzoic acid (1:2) (LEVI-3NBA) The LEVI-3NBA cocrystal was obtained by cooling a saturated 1:2 solution (ratio suspected by XRPD) of coformers in acetonitrile, and solved in the orthorhombic space group $P2_12_12_1$. One molecule of Levetiracetam (LEVI) and two molecules of 3-nitrobenzoic acid (3NBA) coexist in the asymmetric unit.

The carboxylic acid functions of these two molecules of 3NBA are involved in different hydrogen bonding motifs with one LEVI: as for the racemic version, a $R^2_2(8)$ heterosynthon is observed for the carboxylic acid function of a first molecule while that of the second takes part in a $R^2_2(11)$ ring motif (Figure 13) not observed in the enantiopure version. This motif is formed by a proton of the LEVI noncyclic amide acting as a H donor to the 3NBA carbonyl. The motif is completed by a cyclic amide carbonyl group binding to the 3NBA acid proton. These H bonding interactions result in the formation of a V shaped trimer.

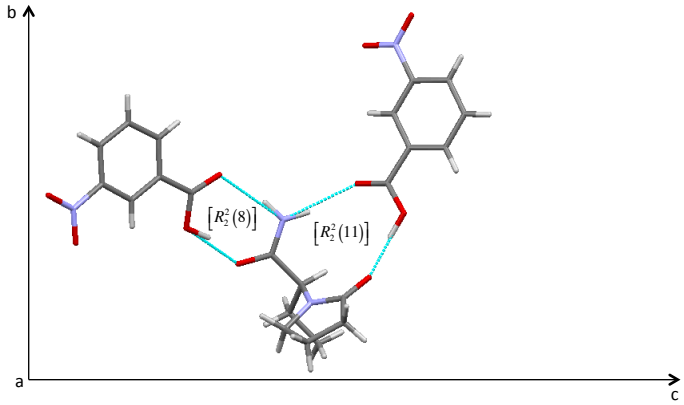


Figure 13. A $R^2_2(8)$ heterosynthon and a $R^2_2(11)$ ring motif are formed by 3NBA and the two LEVI molecules in the asymmetric unit cell of LEVI-3NBA cocrystal, resulting in a V shaped trimer.

Overall, trimers are held together by π stacking of 3NBA molecules and hydrophobic interactions, leading to the final 3D network.

1.3.B.5 4-nitrobenzoic acid as coformer

Etiracetam- 4-nitrobenzoic acid (1 : 2) (E4NBA) The ETI-4NBA co-crystal was obtained by evaporation of a 1 : 1 solution of coformers in acetonitrile, and solved in the monoclinic space group $P2_1/c$. One molecule of Etiracetam (ETI) and two molecules of 4-nitrobenzoic acid (4NBA) coexist in the asymmetric unit.

Those two 4NBA molecules are involved in different kinds of hydrogen bonds with the ETI: the carboxylic function of the first forms a $R^2_2(8)$ heterosynthon with the noncylcic amide of ETI while the second take part in a zigzag $C^2_2(16)$ infinite chain motif (Figure 14). In this motif, a first hydrogen bond is formed between the noncyclic amide of ETI and the nitro group of 4NBA, while a second hydrogen bond joins the ETI cyclic amide carbonyl function to the acid of 4NBA.

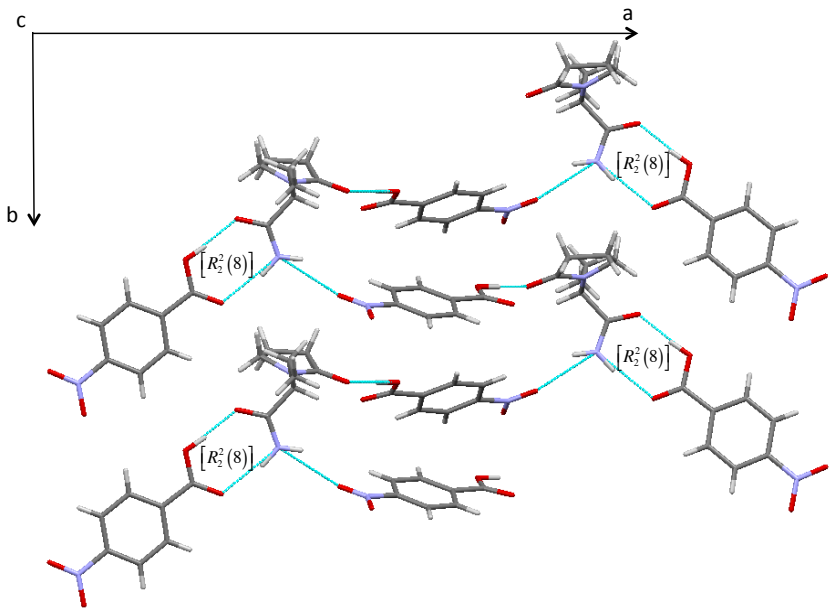


Figure 14. The $R^2_2(8)$ acid-amide heterosynthon and the $C^2_2(16)$ infinite chain motif formed in the ETI-4NBA cocrystal.

Overall, these bonding patterns create thin layers, densely stacked along the three directions.

Levetiracetam- 4-nitrobenzoic acid (1 : 2) (LEVI-4NBA) The LEVI-4NBA cocrystal was obtained by cooling a saturated 1:1 solution of coformers in ethyl acetate, and solved in the triclinic space group *P*1. One molecule of Levetiracetam (LEVI) and two molecules of 4-nitrobenzoic acid (4NBA) coexist in the unit cell.

The carboxylic acid functions of these two molecules of NBA are involved in different hydrogen bonding motifs with one LEVI molecule. As for the racemic version, a $R^2_2(8)$ acid-amide heterosynthon is observed for the carboxylic acid function of a first molecule, with the LEVI noncyclic amide. The second takes part in a zigzag $C^2_2(11)$ infinite chain motif different from the one observed in the enantiopure version (Figure 15). In this motif, the 4NBA accepts a H-bond from the LEVI noncyclic amide and acts as a hydrogen bond donor to the LEVI cyclic carbonyl; allowing the growth of the chain along the b-axis. The nitro group is not involved in any hydrogen bond.

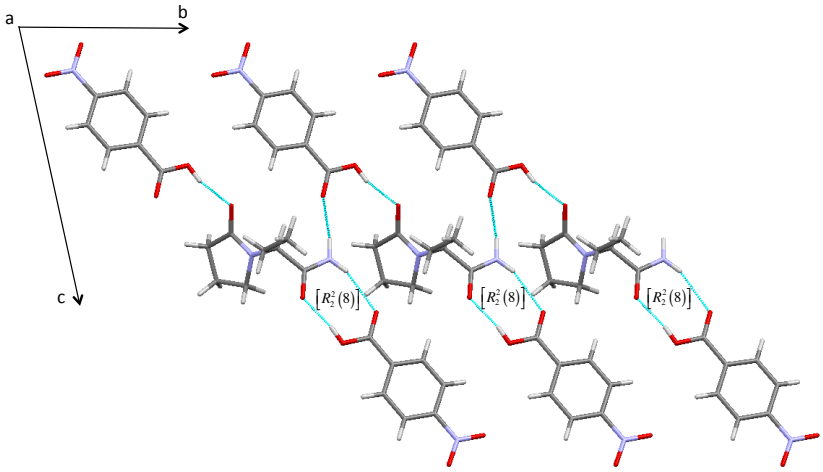


Figure 15. The $R^2_2(8)$ acid-amide heterosynthon and the $C^2_2(11)$ infinite chain motif present in the LEVI-4NBA cocrystal.

Overall, those bonding patterns form thin layers, densely stacked along the three directions, as in the ETI-4NBA cocrystal.

1.4 Discussion

In this contribution, we show that enantiopure and racemic versions of a selected API, Levetiracetam, have a tendency to form cocrystals with identical non-chiral partners. This implies that cocrystal screening of an enantiopure molecule can be performed more effectively if the cocrystal screening of the racemic compound has already been performed. In early stages of drug research, the racemic compound is often more readily available, compared to the enantiopure compound. At this stage, it thus seems more interesting to perform an extended screen using the racemic compound, followed by a reduced screen with the enantiopure compound using the coformers that led to positive hits in the first screen.

Our results show that if 152 coformers are tested both on the racemic as well as on the enantiopure compound, a comparable success rate of about 10% (18/152 and 14/152 respectively) is observed. On the other hand, if only the reduced set of 18 coformers which gave a positive hit during the racemic screen were to be used for the enantiopure screen, a success rate of up to 72 % (13/18) is achieved. Although one out of the 14 enantiopure cocrystals would not have been identified using this approach, 90% less

compound is required for the enantiopure cocrystal screen, which can be a significant advantage, as well as a speed-up for cocrystal studies in early stages of drug development.

A structural investigation into five racemic-enantiopure cocrystal pairs shows that, although not all structural elements are identical, similarities with respect to hydrogen bonding synthons can be found between the enantiopure and racemic cocrystals.

In particular the amide-acid heterosynthon is commonly encountered. In a CSD search, Steiner[37] states that in 10% of cocrystal structures containing an amide and a carboxylic acid functional group, the $R^2_2(8)$ acid-amide heterosynthon is present. Similarly, Vishweshwar et al.[38] established that this percentage increases to approximately 50% when the implied amide is a primary amide. In our structures, this element is also frequently encountered. Etiracetam cocrystals display a $R^2_2(8)$ acid-amide heterosynthon with all coformers but 2,4-dihydroxybenzoic acid. For this latter, a $R^2_2(8)$ amide homosynthon is found instead. In Leviteracetam cocrystals structures, these motifs are only present in LEVI-OXA, LEVI-3NBA and LEVI-4NBA cocrystals. In the LEVI-DMSA cocrystal, the $R^2_2(8)$ heterosynthon is replaced by a $C^2_2(8)$ heterosynthon formed between one molecule of Levetiracetam and two molecules of 2,2-dimethylsuccinic acid (or conversely). This type of heterosynthon was already found in the literature[20] in the Levetiracetam- D-tartaric acid or S-methylsuccinic acid cocrystals.

Preponderance of Etiracetam cocrystals over Levetiracetam stands *a priori* in favor of the Wallach's rule[39] which states that crystals of racemic molecules show higher densities than those of enantiomeric forms; being thus more stable.[40,41] This rule, however, no longer holds when the number of hydrogen bonds is important, as is the case here. Table 2 indeed shows that for some coformers, the racemic form is denser while for other coformers, the opposite is observed.

Another striking common element found in almost all cocrystals studied in this contribution (with exception of 3NBA based cocrystals) is the formation of layered 3D networks, with layers held together by hydrophobic interactions. This shows that, although hydrogen bonding interactions are the main driving force towards cocrystal formation, possible steric hindrance and hydrophobic interactions (π-π and van der Waals) also need to be taken into account when considering the overall structure.[42–44]

1.5 Conclusion

In this article, we show that enantiopure and racemic versions of a selected API tend to form cocrystals with identical non-chiral coformers. Indeed, using 152 non-chiral coformers, we identified 14 novel cocrystals of Levetiracetam, an enantiopure compound and 18 of Etiracetam, its racemic version. Out of these, 13 share a common coformer. A structural investigation into five racemic-enantiopure cocrystal pairs demonstrated that, although not all structural elements are identical, they often show similar hydrogen bonding synthons.

Hence, in early stages of drug research, when a racemic compound is often more readily available than its enantiopure counterpart, we suggest to perform an extended screen using the racemic compound, followed by a focused screen with the enantiopure one, using the coformers that led to positive hits in the first screen. In our case, this two-step approach would allow a success rate of 72 % instead of 10% and would require 90% less compound and significantly speed up cocrystal studies.

1.6 References

(1) Billot, P.; Hosek, P.; Perrin, M.-A. *Org. Process Res. Dev.* 2013, *17*, 505–511.

(2) Bauer, J.; Spanton, S.; Henry, R. F.; Quick, J.; Dziki, W.; Porter, W.; Morris, J. *Pharm. Res.* 2001, *18*, 859–866.

(3) Fabbiani, F. P. A.; Allan, D. R.; Parsons, S.; Pulham, C. R. *CrystEngComm* 2005, *7*, 179.

(4) Kumar, V.; Malhotra, S. V. *ACS Symp. Ser.* 2010, *1038*, 1–12.

(5) Neau, S. H. In *Water-Insoluble Drug Formulation*; Rong, L., Ed.; CRC Press: Boca Raton, 2000; pp. 405–425.

(6) Karki, S.; Friscic, T.; Jones, W.; Motherwell, W. D. S. *Mol. Pharm.* 2007, *4*, 347–354.

(7) Khankari, R. K.; Grant, D. J. W. *Thermochim. Acta* 1995, *248*, 61–79.

(8) Sanphui, P.; Kumar, S. S.; Nangia, A. *Cryst. Growth Des.* 2012, *12*, 4588–4599.

(9) Viertelhaus, M.; Hilfiker, R.; Blatter, F.; Neuburger, M. *Cryst. Growth Des.* 2009, *9*, 2220–2228.

(10) Espinosa-Lara, J. C.; Guzman-Villanueva, D.; Arenas-García, J. I.; Herrera-Ruiz, D.; Rivera-Islas, J.; Román-Bravo, P.; Morales-Rojas, H.; Höpfl, H. *Cryst. Growth Des.* 2013, *13*, 169–185.

(11) Cheney, M. L.; Shan, N.; Healey, E. R.; Hanna, M.; Wojtas, L.; Zaworotko, M. J.; Sava, V.; Song, S.; Sanchez-Ramos, J. R. *Cryst. Growth Des.* 2010, *10*, 394–405.

(12) Liao, X.; Gautam, M.; Grill, A.; Zhu, H. J. *J. Pharm. Sci.* 2010, *99*, 246–254.

(13) Rasenack, N.; Müller, B. W. *Int. J. Pharm.* 2002, *245*, 9–24.

(14) Aaltonen, J.; Allesø, M.; Mirza, S.; Koradia, V.; Gordon, K. C.; Rantanen, J. *Eur. J. Pharm. Biopharm.* 2009, *71*, 23–37.

(15) Aakeröy, C. B.; Grommet, A. B.; Desper, J. *Pharmaceutics* 2011, *3*, 601–614.

(16) Thakuria, R.; Delori, A.; Jones, W.; Lipert, M. P.; Roy, L.; Rodríguez-Hornedo, N. *Int. J. Pharm.* 2013, *453*, 101–125.

(17) Tilborg, A.; Michaux, C.; Norberg, B.; Wouters, J. *Eur. J. Med. Chem.* 2010, *45*, 3511–3517.

(18) *Pharmaceutical Salts and Co-crystals*; Wouters, J.; Quéré, L., Eds.; RSC Drug D.; Royal Society of Chemistry: Cambridge, UK, 2011.

(19) Aitipamula, S.; Banerjee, R.; Bansal, A. K.; Biradha, K.; Cheney, M. L.; Choudhury, A. R.; Desiraju, G. R.; Dikundwar, A. G.; Dubey, R.; Duggirala, N.; Ghogale, P. P.; Ghosh, S.; Goswami, P. K.; Goud, N. R.; Jetti, R. R. K. R.; Karpinski, P.; Kaushik, P.; Kumar, D.; Kumar, V.;

Moulton, B.; Mukherjee, A.; Mukherjee, G.; Myerson, A. S.; Puri, V.; Ramanan, A.; Rajamannar, T.; Reddy, C. M.; Rodriguez-Hornedo, N.; Rogers, R. D.; Row, T. N. G.; Sanphui, P.; Shan, N.; Shete, G.; Singh, A.; Sun, C. C.; Swift, J. A.; Thaimattam, R.; Thakur, T. S.; Kumar Thaper, R.; Thomas, S. P.; Tothadi, S.; Vangala, V. R.; Variankaval, N.; Vishweshwar, P.; Weyna, D. R.; Zaworotko, M. J. *Cryst. Growth Des.* 2012, *12*, 2147–2152.

(20) Springuel, G.; Norberg, B.; Robeyns, K.; Wouters, J.; Leyssens, T. *Cryst. Growth Des.* 2012, *12*, 475–484.

(21) Springuel, G.; Robeyns, K.; Norberg, B.; Wouters, J.; Leyssens, T. *Cryst. Growth Des.* 2014.

(22) Sekhon, B. S. *ARS Pharm.* 2009, *50*, 99–117.

(23) Hurtado, B.; Koepp, M. J.; Sander, J. W.; Thompson, P. J. *Epilepsy Behav.* 2006, *8*, 588–592.

(24) Herman, C.; Vermylen, V.; Norberg, B.; Wouters, J.; Leyssens, T. *Acta Crystallogr. B. Struct. Sci. Cryst. Eng. Mater.* 2013, *69*, 371–378.

(25) Herman, C.; Haut, B.; Aerts, L.; Leyssens, T. *Int. J. Pharm.* 2012, *437*, 156–161.

(26) Herman, C.; Haut, B.; Douieb, S.; Larcy, A.; Vermylen, V.; Leyssens, T. *Org. Process Res. Dev.* 2012, *16*, 49–56.

(27) Herman, C.; Leyssens, T.; Debaste, F.; Haut, B. *J. Cryst. Growth* 2012, *342*, 57–64.

(28) Herman, C.; Haut, B.; Halloin, V.; Vermylen, V.; Leyssens, T. *Org. Process Res. Dev.* 2011, *15*, 774–782.

(29) Herman, C.; Leyssens, T.; Vermylen, V.; Halloin, V.; Haut, B. *J. Chem. Thermodyn.* 2011, *43*.

(30) Springuel, G.; Leyssens, T. *Cryst. Growth Des.* 2012, *12*, 3374–3378.

(31) Favre-Nicolin, V.; Černý, R. *J. Appl. Crystallogr.* 2002, *35*, 734–743.

(32) Rodríguez-Carvajal, J. *Phys. B Condens. Matter* 1993, *192*, 55–69.

(33) Özcan, Y.; Osmanoglu, S.; Ide, S. *Anal. Sci.* 2003, *19*, 1221–1222.

(34) Sheldrick, G. M. *Acta Crystallogr. A.* 2008, *64*, 112–122.

(35) Delori, A.; Friščić, T.; Jones, W. *CrystEngComm* 2012, *14*, 2350.

(36) Etter, M. C.; MacDonald, J. C.; Bernstein, J. *Acta Crystallogr. Sect. B Struct. Sci.* 1990, *46*, 256–262.

(37) Steiner, T. *Angew. Chemie Int. Ed.* 2002, *41*, 48–76.

(38) Vishweshwar, P.; McMahon, J. a; Bis, J. a; Zaworotko, M. J. *J. Pharm. Sci.* 2006, *95*, 499–516.

(39) Wallach, O. *Justus Liebig's Ann. der Chemie* 1895, *286*, 90–118.

(40) Friscić, T.; Fábián, L.; Burley, J. C.; Reid, D. G.; Duer, M. J.; Jones, W. *Chem. Commun. (Camb).* 2008, 1644–1646.

(41) Brock, C. P.; Schweizer, W. B.; Dunitz, J. D. *J. Am. Chem. Soc.* 1991, *113*, 9811–9820.

(42) Baures, P. W.; Rush, J. R.; Wiznycia, A. V.; Desper, J.; Helfrich, B. A.; Beatty, A. M. *Cryst. Growth Des.* 2002, *2*, 653–664.
(43) Takata, N.; Shiraki, K.; Takano, R.; Hayashi, Y.; Terada, K. *Cryst. Growth Des.* 2008, *8*, 3032–3037.
(44) Vishweshwar, P.; Nangia, A.; Lynch, V. M. *Cryst. Growth Des.* 2003, *3*, 783–790.

Chapter 2. The peculiar case of Levetiracetam and Etiracetam α-ketoglutaric acid cocrystals: obtaining a stable conglomerate of Etiracetam

Fanny George †, Bernadette Norberg ‡, Koen Robeyns †, Johan Wouters ‡, Tom Leyssens †*

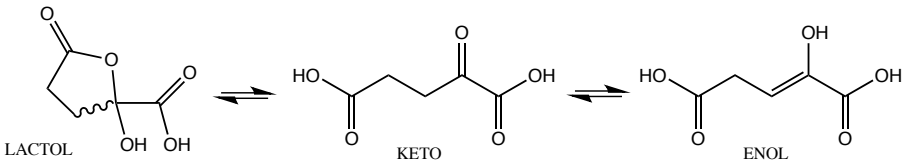
† Institute of Condensed Matter and Nanosciences, Université catholique de Louvain, 1348 Louvain-la-Neuve, Belgium
‡ Unité de chimie physique, théorique et structurale, University of Namur, Namur, Belgium

Abstract: This chapter is dedicated to the cocrystals formed between Levetiracetam and Etiracetam with alpha-ketoglutaric acid that are peculiar in many respects. We first demonstrate that it is possible to obtain the lactol tautomer of alpha-ketoglutaric acid (AKGA) in the solid-state by cocrystallizing it with Leviteracetam. Then, we show that a cocrystal can be isolated with the racemic counterpart of Levetiracetam, Etiracetam, in which AKGA stays in the keto-form. We also report the existence of a cocrystal conglomerate in the Etiracetam-AKGA system, which is more stable than the racemic cocrystal at room temperature. The existence of a stable conglomerate in this system is put in relation with the enantiospecificity of the Levetiracetam cocrystals, which is likely related to the ability of the Etiracetam enantiomers to stabilize one lactol tautomer at a time in solution, or to promote its formation by H-bonding. More generally, by comparing the peculiarities of the system in hand to the general behavior of cocrystallizing chiral systems with and without zwitterionic coformers, we suggest that for a pseudoquaternary cocrystal (i.e. cocrystal made up of two racemate compounds) to exist, the pseudoternary combinations (i.e. cocrystal made up of one racemate and an enantiomer of the second compound) should exist and the enantiomers of the two compounds should form a diastereomeric pair at the binary level, rather than behave enantiospecifically. We also evidence that a tautomeric equilibrium may be induced by grinding, without the requirement of any amount of solvent.

2.1 Introduction

Cocrystals are organic multicomponent crystals containing a stoichiometric ratio of at least two components interacting through directional contacts, and that are not simple solvates or salts (at least one component is non ionized).[1] Cocrystals have been developed extensively in recent years due to their interest in the pharmaceutical industry. Pharmaceutical cocrystals, containing an active pharmaceutical ingredient (API) and one neutral component called coformer, may indeed show very desirable properties in comparison with the isolated drug, such as improved solubility[2] and dissolution rate,[3] but also bioavailability,[4] and handling properties.[5,6] But their potential is not restricted to drug formulation and other applications have been proven, including non-linear optics,[7] and chiral resolution,[8–10] to name but a few.

α-Ketoglutaric acid (AKGA) is a very interesting compound for two reasons. First, it is the substrate of the enzymatic reaction producing L-glutamate.[11] Second, it is involved in three different equilibria in solution[11] due to its double identity.[12] Indeed, it is a α-ketocarboxylic acid and a γ-keto carboxylic acid at the same time and thus shows characteristics of both compounds: being a α-ketocarboxylic,[13,14] it is prone to enolization and hydration, while its γ-keto carboxylic identity suggests an ability to cyclize into a lactol form (Eq.1). These 3 forms thus coexist in solution, in relative proportions that depend on many factors, including pH and temperature of the solution. This has been the object of several researches.[12] Various NMR studies[11,12] showed notably that, in aqueous solution, the non-hydrated keto form is predominant and the lactol form inexistant at neutral pH and room temperature, while their percentage have been calculated to be comparable (35% and 30% resp.) at 29°C and pH 0. Besides, their interconversion has been shown to be extremely rapid on the ^{13}C NMR time scale.



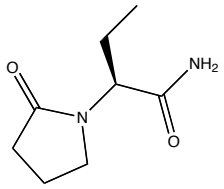
Equation 1. Keto-lactol and keto-enol tautomerisms of AKGA in solution.

In the solid-state, however, the cyclic form has never been isolated for this compound, contrary to certain o-acylbenzoic acids.[15] Indeed, there are only two entries for AKGA in the CSD[16] and in both AKGA exists in its open-chain

form. The first one (refcode COTPAC)[17] is the pure compound structure while the second corresponds to a complex with 1,3-bis((Pyrid-2-ylamino)carbonyl)adamantine (refcode RIZWUS).[18]

In this contribution, we demonstrate that it is possible to obtain the lactol form of AKGA by cocrystallizing it with Leviteracetam (Levi, scheme 1), which is a chiral (S) nootropic drug used to treat epilepsy. Besides, we show that a cocrystal can also be isolated with the racemic equivalent of Levetiracetam, Etiracetam (Eti, scheme 1), in which AKGA stays in the keto-form. Last but not least, we found that, depending on the experimental conditions, the racemic mixture (Eti + AKGA) may crystallize as a conglomerate found to be more stable than the racemic cocrystal at room temperature. To the author's best knowledge, this is only the second report of a cocrystal conglomerate; the first one being characterized by Neurohr *et al.* in 2015.[19]

Conglomerates are indeed very rare in comparison with racemic crystals, occurring in only 5-10% of racemic single-component crystallizations.[20] But the probability of finding a conglomerate is expected to be even smaller in case of cocrystallization, as matching coformers for a given compound are not yet predictable and often implies a high-throughput screening procedure.[21–24] Yet, conglomerates are intensely researched as various chiral resolution techniques, including Viedma ripening[25–34] and preferential crystallization,[35–38] are conditioned by their existence.



Scheme 1. Chemical diagram of S-Levetiracetam.

2.2 Experimental Section

Starting Materials. S-2-(2-oxopyrrolidin-1-yl)butanamide (Levetiracetam) was purchased from Xiamen Top Health Biochem Tech. Co., Ltd. 2-ketoglutaric acid was purchased from Acros Organics. These materials were used as received, without further purification. (RS)-2-(2-oxopyrrolidin-1-yl)butanamide (Etiracetam) was prepared by racemization of S-2-(2-oxopyrrolidin-1-yl)butanamide. Ten grams of S-2-(2-oxopyrrolidin-1-

yl)butanamide together with a catalytic amount (0.05 equiv) of NaOMe was added to 10 mL of MeOH. The solution was kept at reflux under continuous stirring for 24 h and then cooled to room temperature. The compound crystallizes spontaneously. After filtration, the compound was washed twice with MeOH. The recovered compound was used as such. R-2-(2-oxopyrrolidin-1-yl)butanamide cannot be purchased and was therefore obtained from Etiracetam using the chiral resolution procedure described by Springuel et al.11 A solution containing Etiracetam, R-mandelic acid and acetonitrile in molar percentages of respectively 4.36, 6.63 and 89 mol%, was kept at -10°C and seeded with the cocrystal formed between R-2-(2-oxopyrrolidin-1-yl)butanamide and R-mandelic acid. Under these conditions, this cocrystal is recovered, as it is the most stable phase in suspension. After filtration, R-2-(2-oxopyrrolidin-1-yl)butanamide was separated from R-mandelic acid with a reverse HPLC system Waters Alliance 2690 equipped with a PDA detector (Waters 2998). A Waters Atlantis T3 column (4.6mm x 50mm x 3.5 _m) has been used with CH3CN/H2O 50/50 v/v as dilution solvent. Contrary to its (S)-counterpart, this compound does not show any biological effect and hence does not have any common name. For clarity sake, it will however be referred as (R)-Levi in the following text. Besides, as it was at our disposal in very small quantity, this compound was only used to demonstrate enantiospecicity in the Levi-AKGA system.

Cocrystal Screen. Cocrystals were synthesized by grinding of equimolar mixtures of Levi or Eti and AKGA with or without a drop of solvent (see Table 1). Samples were ground in a RETSCH Mixer Mill MM 400 with a beating frequency of 30 Hertz for at least 10 min (the default grinding time being 90 minutes to ensure complete conversion). The resulting powders were characterized using XRPD. Comparison of the resulting diffraction pattern with the diffraction patterns of the pure phases was used to indicate cocrystal formation. All possible known forms of the pure phases were considered, to avoid confusing cocrystal formation with other phase transformations (e.g. solvate formation, polymorphism).

X-ray Powder Diffraction (XRPD). X-ray diffraction measurements were performed on a Siemens D5000 diffractometer equipped with a Cu X-ray source operating at 40 kV and 40 mA and a secondary monochromator allowing selection of the Kα radiation of Cu (λ =1.5418 Å). A scanning range of 2θ values from 2° to 72° at a scan rate of 0.6° min−1 was applied.

Single Crystal Preparation. Single crystals were grown in acetone and/or acetonitrile (one solvent at a time) using a non-stoichiometric ratio of both coformers.[22,39,40] First, suspensions of each component were prepared

separately at room temperature. Then, a similar volume of each supernatant solution was filtered and both solutions mixed, such that the resulting solution has a concentration equal to half of the solubilities of each component. Slow evaporating of about half the solvent volume initiated the selective crystallization of the cocrystals with sufficient size.

Single Crystal X-ray Diffraction. Single crystal X-ray diffraction was performed either on a Gemini Ultra R system (4-circle kappa platform, Ruby CCD detector) using Cu Kα radiation (λ = 1.54184 Å) or Mo Kα radiation (λ = 0.71073 Å), or on a Mar345 image plate (Xenocs Fox3D mirrors) using Mo Kα radiation. Cell parameters were estimated from a pre-experiment run and full data sets collected at room temperature. The structures were solved by direct methods with the SHELXS-97 program and then refined on $|F|^2$ using the SHELXL-2014 software.[41] The final reported R_1 value is calculated on |F| for observed reflections (I > 2 sigma(I)). Non-hydrogen atoms were anisotropically refined, and hydrogen atoms were placed at calculated positions and refined in riding mode with isotropic temperature factors fixed at 1.2 times U_{eq} of the parent atoms (1.5 times for methyl groups).

Slurry preparation. Supersaturated solutions of racemic Eti-AKGA 1:1 composition were prepared in two different solvents (acetone and acetonitrile) and using alternatively two different starting materials, for a total of four experiments. In the first two vials (one with each solvent), the suspension was initially composed of Eti-AKGA 1:1 powder obtained by liquid assisted grinding (conglomerate form), while in the last two vials (one with each solvent), the powder resulted from dry grinding of the corresponding mixture (racemic form). These suspensions were left to equilibrate overnight and then seeded with the alternative form (i.e. the form not used as starting material) to test the potential conversion from the starting material to the seeded product, in case of higher stability of the latter. Suspensions were stirred for one week at room temperature before analyzing the solid phase by XRPD to determine the outcome.

2.3 Results

2.3.A Cocrystal identification

To determine the existence of cocrystals between Levetiracetam/Etiracetam and AKGA, liquid assisted grinding (LAG) of the corresponding materials was performed, as this type of experiments has been shown to be among the most effective for cocrystal screening.[42–44] In particular, it enables polymorphism control[6] while avoiding solubility restrictions. The resulting

products were analyzed by XRPD and presence of a new phase was detected in both conditions. Then, attempts were made to grow the corresponding single crystals in solution for structural characterization as described in the experimental part. Two novel cocrystal structures were identified. The one with Etiracetam contains AKGA in the keto form while the Levetiracetam cocrystal was formed with the lactol AKGA tautomer. Even though cocrystallization has been shown to impact tautomerism,[45] in our case, the result is nevertheless surprising as the lactol tautomer has never been detected up to now in the solid-state.

Furthermore, whereas the XRPD diffractogram of the Levi-AKGA ground material overlapped with the simulated diffractogram obtained from the single crystal, the simulated XRPD diffractogram of the Eti-AKGA cocrystal did not match the XRPD pattern of the ground mixture ((Eti-AKGA_LAG). However, superposition of the diffractograms of Levi-AKGA and the ground mixture Eti-AKGA_LAG revealed identical patterns (Figure 1), indicating the formation of a conglomerate when performing liquid assisted grinding.

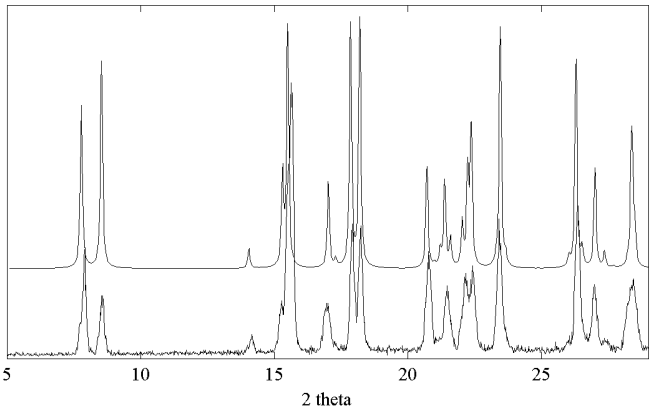


Figure 1. XRPD patterns (up) simulated from the Levi-AKGA cocrystal and (down) of ground Eti-AKGA material; the similarity of patterns indicating conglomerate formation.

Dry grinding was also performed on these systems as it is well documented that neat and liquid-assisted grinding may generate different outcomes.[43] In our case, dry grinding of Etiracetam with AKGA indeed led to the racemic compound for which the single crystal was obtained. Neat grinding of the Levetiracetam - AKGA mixture however led to the same product as the LAG experiment. This latter observation implies that the tautomeric transformation does not require added solvent to occur in grinding experiments.

2.3.B Selectivity and stability analyses of the Eti-AKGA system

As a conglomerate was obtained in the case of liquid assisted grinding of Eti-AKGA, while neat grinding led to the formation of the racemic compound, some complementary grinding experiments were performed to assess the kinetics of the system and to evaluate how Etiracetam is dismantled into a conglomerate during liquid assisted grinding (*i.e.* direct transformation, or intermediate transformation into the racemic compound which is then transformed into the conglomerate). For this, the duration of the grinding experiments on Eti - AKGA 1:1 mixtures was varied. Grinding was performed during 4 and 10 min with and without a drop of solvent. It appeared that the conglomerate was totally formed after only 4 minutes of kneading, while the conversion of the starting components into the racemic form through dry grinding was not complete at this stage, but well completed after 10 min. In all cases, there were no traces of the reciprocal forms at these smaller grinding times. Both forms are thus rapidly and selectively generated.

To establish the relative stability between the conglomerate and the racemic compound at ambient temperature, slurry experiments[46–49] were carried out. These consist in seeding a suspension of a given composition with all possible phases. The initial suspension progressively evolves toward the most stable phase, following Ostwald's rule of phases. Analysis of the resulting solid phase allows identification of the thermodynamic most stable phase. Solvents with different dielectric constants are often used to vary the solubility profiles and increase the chance of a quick and complete conversion.

In our case, we started by preparing separate suspensions of the conglomerate and the racemic compound in acetone and acetonitrile at room temperature. Each suspension was then seeded with the other solid state (*i.e.* the conglomerate suspension was seeded with the racemic compound, and *vice versa*). Suspensions were then left to equilibrate for a one-week period. In both cases, the solid phase contained only the conglomerate form, indicating the increased stability of the conglomerate over the racemate at room temperature. This was also confirmed by the fact that grinding the racemic compound with a drop of acetonitrile converts it into the conglomerate.

However, their melting points were found to be almost identical (ca. 94°C), suggesting a very slight stability preference in favor of the conglomerate. This situation corresponds to the binary phase diagram shown on Figure 2.

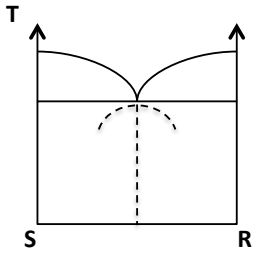


Figure 2. Binary phase diagrams of a system characterized by a stable conglomerate (plain curves) and a metastable racemic compound (dash curves).

Note that even if a stable conglomerate has been found for this system, it appears non-ideal for use in chiral resolutions techniques such as Viedma ripening and preferential crystallization. Indeed, the first method requires the chiral compound to be easily racemizable in solution, which is not the case of Etiracetam enantiomers. Concerning preferential crystallization, Coquerel carefully explained that the existence of an easily isolable metastable racemic compound, which is the case here, drastically reduces the performances of this method. Indeed, the similar stability of the conglomerate and the racemic compound means that the heterochiral and homochiral interactions are competitive in solution, which increases the probability of wrong docking and hinders the process.[35]

A summary of all the experiments that were performed in order to characterize the Eti/Levi-AKGA cocrystals can be found in Table 1.

Table 1. Summary of the experimental outcomes of cocrystallization between Eti/Levi and AKGA in various conditions.

Starting materials	**Experiments**	**Conditions** [a]	**Products**	**Solid-state outcomes**
Eti - AKGA 1:1	Grinding	ACN[b]/MeOH	powder	Conglomerate
	Grinding	dry[c]	powder	Racemic
Eti - AKGA 1:1	Evaporation	Ac	Single crystal	Racemic
Conglomerate	Slurry	ACN/Ac	powder	Conglomerate
Racemic	Slurry	ACN/Ac	powder	Conglomerate

Starting materials	**Experiments**	**Conditions** [a]	**Products**	**Solid-state outcomes**
(S)-Levi - AKGA 1:1	Grinding	ACN/MeOH	powder	(S)-Levi_(R)-AKGA
	Grinding	dry	powder	(S)-Levi_(R)-AKGA
(S)-Levi - AKGA 1:1	Evaporation	ACN/Ac	Single crystal	(S)-Levi_(R)-AKGA

Table 1. Continued

Starting materials	Experiments	Conditions [a]	Products	Solid-state outcomes
(R)-Levi - AKGA 1:1	Grinding	drop ACN	powder	(R)-Levi_(S)-AKGA
	Grinding	dry	powder	(R)-Levi_(S)-AKGA
(R)-Levi - AKGA 1:1	Evaporation	ACN	Single crystal	(R)-Levi_(S)-AKGA

[a] For all experiments performed in solution or using a drop of solvent, we used one solvent at a time. ACN = acetonitrile, Ac =acetone, MeOH = methanol. [b] Complete conversion of the starting components into the conglomerate was achieved after 4 min. [c] Conversion into the racemic form was not complete after 4 min, but well completed after 10 min.

2.3.C Structural analysis

The cocrystals of AKGA found with (S)-/(R)- and (RS)-racetam were structurally characterized through single crystal analysis. Their crystallographic parameters are displayed in Table 2 and followed by a comparative analysis of the main bonding patterns existing in these three cocrystals. Graph-sets were assigned to synthons that were judged to best represent the structures, using Etter's nomenclature.[50]

Table 2. Crystallographic Parameters of the three AKGA cocrystals

Co-crystals	(RS)-AKGA	(S)-AKGA	(R)-AKGA
Structural formula	$(C_8H_{14}N_2O_2)$ $(C_5H_6O_5)$	$(C_8H_{14}N_2O_2)$ $(C_5H_6O_5)$	$(C_8H_{14}N_2O_2)$ $(C_5H_6O_5)$
Formula weight (g/mol)	316,31	316,31	316,31
Space system	triclinic	monoclinic	monoclinic
Space group	*P*-1	$P\,2_1$	$P\,2_1$
a (Å)	5,5195(2)	11,6406(8)	11,7707(15)
b (Å)	11,1809(8)	5,5396(2)	5,5329(4)
c (Å)	13,5104(12)	12,7812(9)	12,706(2)
α (°)	112,193(8)	90	90
β (°)	99,168(5)	115,940(9)	115,831(18)
γ (°)	95,746(5)	90	90
V ($Å^3$)	750,326	741,152	744,80(16)
Z	2	2	2
R-factor (%)	4,43	6,67	7,35
Density	1,4	1,417	1,41
Radiation	Mo K(α)	Cu K(α)	Mo K(α)
Temperature (K)	150	293	293

Etiracetam- α-ketoglutaric acid (1 :1) ("Eti-AKGA" or "RS-AKGA"). The RS-AKGA co-crystal crystallizes from acetone in a triclinic system with space group *P*-1. The unit cell contains two molecules of Eti and two molecules of AKGA as keto tautomer.

An amide-carboxylic acid heterosynthon, described as $R^2_2(8)$, is constituted between the carbamoyl of Eti and the carboxylic acid function of AKGA adjacent to the ketone (Figure 3). In addition, two molecules of Eti and two molecules of AKGA form a tetramer including two hydrogen bonds of the $R^2_2(8)$ heterosynthon mentioned above and two others binding the AKGA ketone to the carbamoyl trans hydrogen of Eti. This tetramer may be characterized by either an "inner" $R^4_4(14)$ ring motif or an "outer" $R^4_4(18)$ ring motif (Figure 3 middle and right), depending on the $R^2_2(8)$ portion included. Finally, the second carboxylic group of AKGA is hydrogen bonded to the carbonyl of the cyclic amide (oxopyrrolidin) of Eti, forming a dimer oriented along the c-axis. However, as there is one extra acceptor in comparison with the total number of donors (two donors on each partner), the second carbonyl of AKGA is not involved in any strong H-bond. Each molecule of the co-crystal is thus involved in four H-bonds.

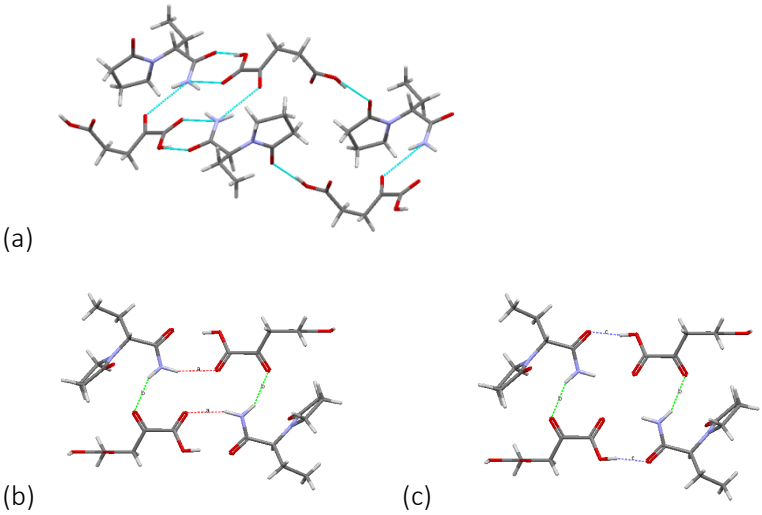


Figure 3. Hydrogen bonds network in RS- AKGA cocrystal showing all types of interactions (a), the inner $R^4_4(14)$ (b) and the outer $R^4_4(18)$ (c) ring motifs.

These motifs form stepwise columns directed along the *c*-axis and stacked along *a* and *b*-axes (Figure 4). Columns are two-molecules wide in the *b*-direction, with ethyl groups of Eti molecules pointing outward and facing equivalent groups in neighbouring columns.

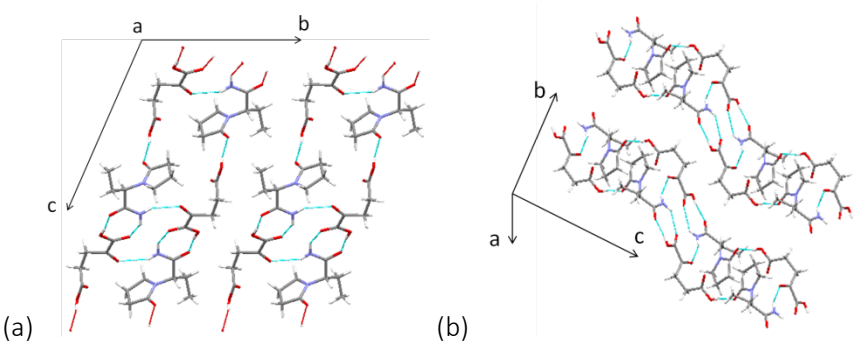


Figure 4. Columnar Stacking in RS- AKGA cocrystal (a) in the *bc*-plane and (b) showing the steps in the columns.

(S)-Levetiracetam- α-ketoglutaric acid (1 :1) ("S-Levi_R-AKGA" or "S-AKGA"). The S-AKGA co-crystal was successfully grown from both acetone and acetonitrile, in a monoclinic system with space group $P2_1$. The unit cell contains two molecules of Levi and two molecules of AKGA in lactol configuration with (R)-chirality exclusively.

Hydrogen bonding patterns in this cocrystal may be described by a $R^2_2(9)$ ring and one $C^2_2(11)$ chain (Figure 5). The ring motif is formed by the carbamoyl (C=O and adjacent H) of a first molecule of Levetiracetam with the carboxylic acceptor and adjacent hydroxyl of AKGA. The pyrrolidone carbonyl of a second molecule of Levi then accepts the carboxylic donor of AKGA, which in turn interacts with the trans carbamoyl hydrogen of a third molecule of Levi. This constitutes a chain that involves two parallel molecules of Levi and propagates along the *b*-axis.

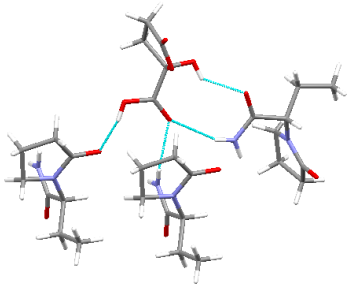


Figure 5. Views of main H-bonding in (S)-AKGA cocrystal showing the $R^2_2(9)$ ring motif and the $C^2_2(11)$ chain involving two parallel molecules of Levi.

Hence, as for the (RS)-AKGA cocrystal, each molecule takes part in four H-bonds. But while all donor and acceptor atoms of Levi are involved in H-bonding, the ester function of AKGA remains unoccupied. This is due to the bifurcation of the AKGA carboxyl that makes the ester acceptors redundant.

On the whole, S-AKGA exhibits a columnar stacking. All the hydrogen bonds are located inside the columns, forming a complicated and interlocked network. Columns are directed along the *b*-axis and stacked in the two other directions (Figure 6). In a column, ester groups of AKGA and hydrocarbonated groups of Levi point outward, such that Levi ethyl groups on one column interact with ester or ethyl groups from other columns, holding them together.

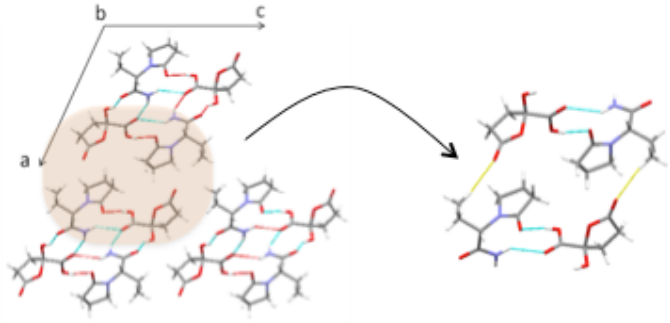


Figure 6. Columnar stacking along *b*-axis in (S)-AKGA cocrystal, showing C-H..O contacts (yellow) between the Levi ethyl and the AKGA ester of different columns.

(R)-Levetiracetam- α-ketoglutaric acid (1 :1) ("R-Levi_S-AKGA" or "R-AKGA"). This cocrystal crystallizes from acetonitrile in a monoclinic system with space group $P2_1$. As it is the mirror image of the (S)-Levi_(R)-AKGA cocrystal, it differs from this latter by the presence of AKGA in the (S)-enantiomeric lactol form. Structural parameters are almost identical to those of the (S)-Levi cocrystal. Identical hydrogen bonding motif and tridimensional patterns are observed.

2.4 Discussion

In view of these observations, three questions arose. First, why was it possible to isolate AKGA in lactol form at the solid state in presence of Levetiracetam? Second, why did we not find any traces of a cocrystal form involving (S)/(R)-Levi and AKGA in the keto form? Finally, is there any

apparent reason behind the formation of a stable conglomerate in this case? Several elements may be considered to address these questions and will be discussed sequentially.

2.4.A Lactol formation

A way to rationalize AKGA cyclization in the enantiopure cocrystals consists in paying attention to its mechanism. In particular, one has to consider the two types of factors that influence the intramolecular addition of the carboxyl to the carbonyl, leading to the cyclization.[51] The first type is electronic and refers to the presence of groups adjacent to the ketone and/or the γ-carboxy group that may affect their electro-/nucleo-philic character respectively and/or induce/prevent resonance stabilization. The second type is geometric and concerns the proximity in space of these two functional groups.[52] These factors were illustrated by Jones and Desio to justify the predominance of one isomer (acid or ester) in function of the nature and position of substituents in o-acylbenzoic acids.[15] Similarly, Winston *et al.*[53] found that for lactones prepared by trichloromethylation of anhydrides, the stability of the lactone form in solution depends on the distance between acids and trichloromethyl groups in the keto-forms. They also evoked internal rotation preventing coplanarity and ring-chain tautomerism in non-cyclic compounds.

Hence, if AKGA is able to cyclize, it is due to the fact that it possesses a carboxylic acid function in both *alpha* and *gamma* positions of the ketone. Indeed, the one in *alpha* increases the electrophilic character of the ketone while the one in *gamma* makes the 5-membered lactol stable in certain conditions. The reason for the absence of a crystal structure of pure AKGA as lactol may be due to its conformational flexibility. Indeed, there exist several cyclic γ-keto-carboxylic acids crystallized in the ring-form due to their geometrical restrictions.[15,51]

However, when Levetiracetam (or its R-equivalent) is added to the solution, it is suspected to selectively stabilize one enantiomer of the lactol form or promote its formation through H-bonding (Figure 7 a and b resp.); which thus influences the corresponding equilibrium. This was already suggested by Valter concerning the solvation effect on γ-aldehydo- and γ-keto-carboxylic acids: "intermolecular hydrogen bonds may play an important part in stabilizing a particular form, the cyclic isomer showing a greater tendency to form such bonds involving both carbonyl and hydroxyl groups".[51] The exact mechanism is nonetheless rather difficult to ascertain

as cocrystallization from solution is complex and involves different equilibria,[39,40] and was thus not investigated here.

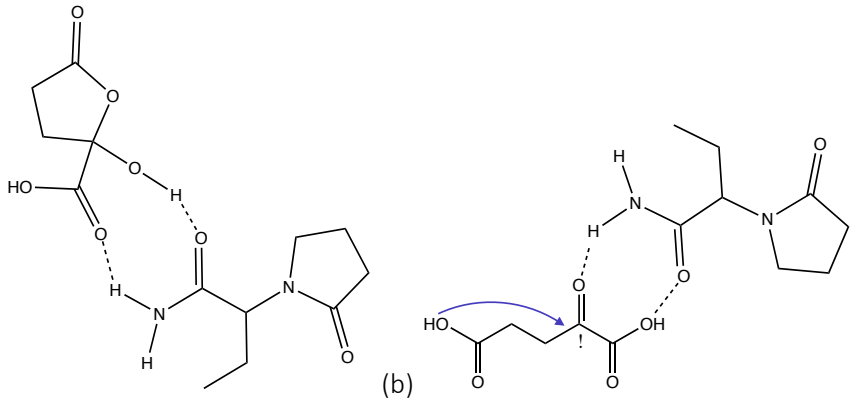


Figure 7. Chemical diagrams showing potential hydrogen bonds occurring in solution that could hypothetically (a) block AKGA in the lactol form or (b) promote its formation by activating the ketone for the intra-molecular attack by the carboxylic acid in gamma.

2.4.B Absence of Levi-AKGA cocrystal with keto-AKGA

Another interesting feature of this work, concerns the fact that thus far we observed no cocrystal of Levetiracetam with the keto form of AKGA, whereas this form does appear when cocrystallizing with Etiracetam. It could be that such form exists and that we simply did not encounter it yet. Even though one could argue that this cocrystal should exist under certain conditions, some structural considerations suggest a decreased stability of the potential Levi-AKGA(keto) cocrystal when compared to the existing racemic equivalent.

These considerations arise from the analysis of the Eti-/Levi-cocrystals with oxalic acid (OXA) ("RS-OXA and S-OXA, refcodes XOGPAM and XOGPEG resp.).[54] OXA and AKGA form similar H-bonding pattern with Eti in their respective cocrystals (Figure 8 b and c), due to their structural resemblance. Besides, OXA also successfully cocrystallizes with Levetiracetam, giving us an idea of what the Levi-AKGA cocrystal with keto-AKGA would look like. Hence, one could expect conclusions drawn from a (RS)-OXA and (S)-OXA comparison to be reasonably valid for the (RS)- and (S)-AKGA system.

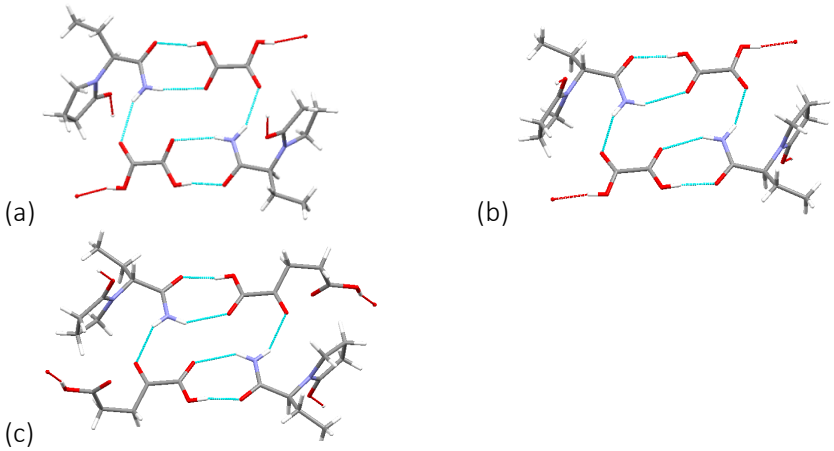


Figure 8. H-bonding rings in (S)-OXA (a), (RS)-OXA (b) and (RS)-AKGA (c).

(S)-OXA and (RS)-OXA cocrystals are characterized by identical H-bonding patterns (Figure 8 a and b), which results in relatively similar PXRD patterns. But, as the first cocrystal is racemic and the other one chiral, their arrangements must differ in some ways. In fact, the main difference between the two cocrystals lies in the conformation of the racetam. The (S)-OXA cocrystal has twice the number of molecules per asymmetric unit than the (RS)-OXA cocrystal, with the two symmetry-independent molecules of Levetiracetam adopting different conformations. In the first conformation, the hydrogen atom on the asymmetric carbon points in the same direction as the oxopyrridin carbonyl ("cis-conformation") while this is contrary in the second conformation ("trans-conformation") (Figure 9).

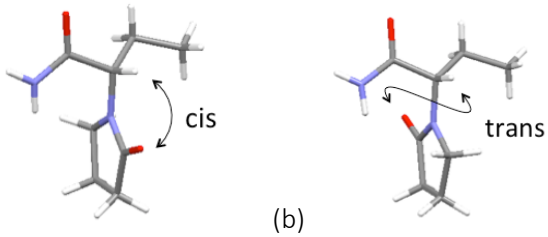


Figure 9. (a) Cis- and (b) trans-conformations of Levi in (S)-OXA cocrystal.

Doing so, the crystal structure of (S)-OXA mimics the centrosymmetric group $P2_1/c$ adopted by the (RS)-OXA cocrystal, with a pseudocenter of inversion relating the two independent molecules of Levetiracetam and a pseudo glide plane *c*. This imitation attempt suggests a particularly desirable and thus efficient packing of the racemic crystal.

Besides, among the 17 structures with either (S)-Levi or Eti recorded in the CSD (2015 release) (including the OXA cocrystals), the racetam always

adopts the cis-conformation; except in (S)-OXA where both are encountered. The cis-conformation is thus expected to be more stable than the trans-conformation as confirmed by their computed single-point energies showing the cis-conformation to be 14 kJ mol $^{-1}$ lower in energy.[a]

Moreover, the situation is highly similar to the one observed in the structures of the racemic and chiral theophylline – malic acid (tp-ma) cocrystals as described by Friscic *et al.*[55] In this system, the racemic and chiral cocrystals also show similar PXRD patterns and H-bond topologies. Besides, the same $R^4_4(18)$ motif is encountered in the racemic (tp).(DL-ma) cocrystal as in the (RS)-AKGA and (RS)-OXA cocrystals (Figure 10, left). However, as the asymmetric carbon of malic acid molecules is included in the H-bonded ring motif, a conformation change in malic acid molecules is not sufficient to accommodate a chiral space group, as observed for the Levi-OXA cocrystal. In this case, the packing that most mimics the racemic (tp).(DL-ma) involves different conformations and different connectivities (*i.e.* the two symmetry-independent molecules of malic acid do not interact with theophylline through the same moieties), which results instead in a $R^4_4(19)$ motif (Figure 10, right). The two conformations were also calculated to have a 15 kJ mol $^{-1}$ energy difference and the presence of the less stable conformer in (tp).(D-ma) was suggested to account for the difference in hydration stabilities and thermal behaviour between the racemic and chiral cocrystals.

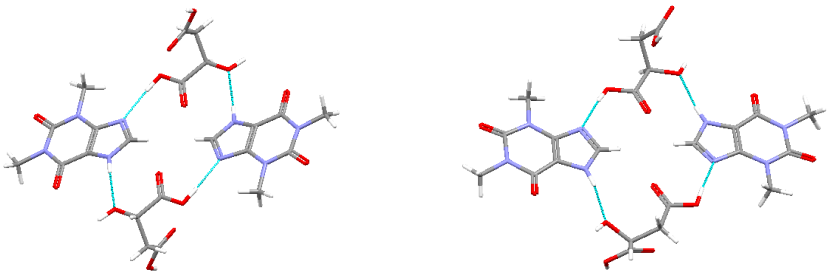


Figure 10. $R^4_4(18)$ and $R^4_4(19)$ ring motifs resp. in racemic (left, refcode CIZTAH) and chiral (right, refcode COCDOO) theophylline – malic acid cocrystals.

Hence, one might expect that the conformational features of Levetiracetam molecules in the virtual Levi- AKGA(keto) cocrystal would destabilize the overall structure, in comparison with the Eti-AKGA(keto) cocrystal. Indeed,

[a] Single point energies were computed using the Gaussian series program with B3LYP/6-31G(d,p) method, and provided a difference of 14.27 kJ mol^{-1} between the two conformations.

the energetic difference between the two conformations is of the same order of magnitude than the usual stabilization energy of cocrystals estimated by *ab initio* studies.[56–58] And the competition with the Levi-AKGA(lactol) form may make the keto version undetectable.

Besides, this analysis confirms the relevance of considering, for the design of multicomponent crystals, larger synthons than the ones commonly used.[59] In this case, the famous $R^2_2(8)$ motif is indeed little representative of the overall structures while the recurrence of the $R^4_4(18)$ motif in both chiral and racemic cocrystals seems to indicate a very efficient packing and possibly a greater predictive power of this motif.

2.4.C Conglomerate increased stability

The last point of this analysis concerns the existence of a conglomerate in this system and the origin of its higher stability.

A racemic solution may crystallize, by decreasing order of probability, as a racemic compound, a conglomerate or a solid solution. In a racemic compound, both enantiomers are present in the same quantity in the crystal lattice. In a conglomerate, the two enantiomers crystallize independently (*i.e.* they form homochiral crystals). In a solid solution for various compositions, both enantiomers are distributed at random in the same crystal network. Solid solution formation is rare for organic compounds and will not be considered in the following discussion.

Conglomerates are statistically disadvantaged in comparison with racemic compounds for two reasons.[60] The first one is thermodynamic. There are simply less packing arrangements available to accommodate chiral crystals. In consequence, as nicely expressed by Brock *et al.*, "it seems likely that the best of many possible racemic packing arrangements is to be preferred to the best of fewer possible chiral arrangements". The second reason concerns the crystallization kinetics. In a racemic solution, the rate of formation of chiral crystals is reduced in comparison with the nucleation of racemic crystals. Indeed, only half of the molecules (the "right" enantiomer) that come in contact with an existing chiral cluster will be suited for its development, whereas molecules of both enantiomers will find a matching site on the racemic clusters. This contrasts with polymorphism, for which less stable forms are usually kinetically favored, as poorer arrangements may be generated more easily. Conglomerates are thus often thermodynamically penalized and not really preferred kinetically either.

Conglomerates are thus expected to spontaneously occur only when their stability is favorable or when the energy difference between the racemic compound and the conglomerate is relatively small. This may be the case when the two structures display a high level of packing similarities, such as homochiral columns or layers, and thus comparable densities.[20]

This is however, not what happens in the system described here. Indeed, the racemic compound and the conglomerate found for this system differ by the nature of AKGA tautomer and are thus not directly related neither structurally, nor energetically. In fact, it is preferable to virtually separate this system (Etiracetam and AKGA) into two systems (Figure 11): one with AKGA in the keto form on one hand and one with AKGA in the lactol form on the other. In each system, the observed form (*i.e.* racemic compound or conglomerate) is more stable than the other possible outcome to such extent that the latter is not formed. The greater stability of the racemic compound in the first system involving a keto-AKGA has indeed been suggested in the previous paragraph and is likely due to the reduced stability of the hypothetical enantiopure crystals involving keto-AKGA. Concerning the second system (lactol-AKGA), we shall refer to our earlier work on enantiospecificity.

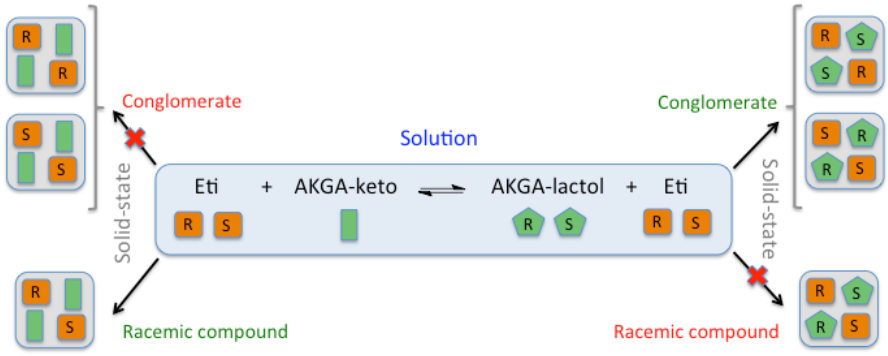


Figure 11. Schematic representation of the two virtual systems featuring Etiracetam and AKGA in solution, and their respective outcomes in the solid-state.

A pair of optically active compounds (represented as RS and DL) cocrystallizes enantiospecifically when only one of the two enantiomers of the first pair cocrystallizes with an enantiomer of the second pair. This behaviour is represented the binary level of situation A in Table 3 and may be opposed to the formation of diastereomeric pairs, which involves both enantiomers cocrystallizing with the chiral coformer (binary level of situation

B in Table 3). A recent contribution showed that enantiospecific cocrystallization is frequent between two optically active compounds in contrast with chiral salts, which usually form diastereomeric pairs.[61]

Table 3. Some possible interactions between two racemates (RS and DL) or the different combinations of their enantiomers.

	Binary	**Ternary**	**Quaternary**
A	*Enantiospecific* R + D ➔ RD R + L ➔ /	R + DL ➔ /	RS + DL ➔ /
B	*Diastereomeric* R + D ➔ RD R + L ➔ RL	R + DL ➔ RDL D + RS ➔ DRS	RS + DL ➔ RSDL

To explain this difference between cocrystals and salts, Springuel and coworkers[61] invoked the very weak stabilization energy of cocrystals with respect to salts. Indeed, it has been estimated by *ab initio* calculations that the energy difference between a cocrystal and its separate components is often inferior to 10 kJ/mol (which is similar to the energy difference between polymorphs),[56–58] while the stabilization free energy may amount to 400 kJ/mol in case of salt formation. For that matter, cocrystals are very sensitive to minor changes in secondary interactions (π-stacking, Van der Waals contacts...), as the one induced by a change of chirality, even when similar H-bonding occurs. Systems involving zwitterionic coformers lie in-between, as charge-assisted interactions are stronger than conventional H-bonds, but weaker than ionic interactions. In practice, they usually form diastereomeric pairs as well.[62] But one could also evoke the position of the chiral center with respect to the functional groups potentially interacting, to account for cocrystal enantiospecificity. For example, H-bonding motifs including the chiral center or close to it may be more sensitive to the steric profile of the coformer and be allowed with only one of its two enantiomers.

These authors also studied cocrystallization between racemic compounds and in particular the likelihood of forming pseudoternary (*i.e.* one racemic compound with one chiral coformer) and pseudoquaternary (*i.e.* two racemic compounds) cocrystals.[10] They analyzed nine different systems forming at least one binary cocrystal, either in an enantiospecic manner (7/9) or as a diastereomeric pair (2/9). Among these, they found only two pseudoternary cocrystals and one pseudoquaternary cocrystal. The system forming a pseudoquaternary cocrystal (Etiracetam with methylsuccinic acid) is also one of the two systems for which a ternary cocrystal is formed, as well

as for which a diastereomeric pair is observed. In other words, a pseudoquaternary cocrystal was encountered only when all the binary and ternary combinations were also successful (Table 3, situation B).

To rationalize this behaviour, one may proceed in stages. Ternary cocrystals are expected to be less likely with a chiral coformer than with an achiral one,[54] as the pseudoternary cocrystal RDL cannot adopt a centrosymmetric packing. This is the contrary in case of an achiral coformer A as for this latter the pseudoternary cocrystal ARS has high chances to display such an arrangement. Besides, pseudoternary cocrystallization would be more likely when one enantiomer of the first pair doesn't have a strong preference for an enantiomer of the second pair (*i.e.* when there is no enantiospecific cocrystallization at the binary level) and rather forms satisfactory interactions with both enantiomers. Hence the occurrence of pseudoternary cocrystals may suggest a preponderant role of the interactions in these systems and less strict requirements on packing. In these flexible systems, pseudoquaternary cocrystals are thus expected to be more likely as well.

This could explain why there is no racemic cocrystal of Etiracetam with AKGA-lactol. Indeed, AKGA is chiral in the lactol form and the corresponding cocrystal with Eti could thus be considered as a pseudoquaternary cocrystal. Yet there are no pseudoternary cocrystals in this system, as this would imply for either Etiracetam to crystallize with one of the lactol forms of AKGA or Leviteracetam with both lactol forms of AKGA, which is not observed. Furthermore, one can also consider that the binary combinations are enantiospecific, as Levi prefers to crystallize with the R-form of AKGA (and reciprocally for the (R)-Levi); even if in practice the enantiomers of AKGA were not introduced separately but created in situ. In fact, enantiospecificity in this case is probably due to the cyclization mechanism of AKGA. All these outcomes are summarized in Table 4.

Table 4. Crystalline outcomes of the different combinations of enantiomers of Eti and AKGA-lactol.

Binary	**Ternary**	**Quaternary**
Enantiospecific	Eti + (R)-AKGA → /	Eti + (RS)-AKGA →/
(S)-Levi + (R)-AKGA → (S)-Levi_(R)-AKGA	Eti + (S)-AKGA → /	
(S)-Levi + (S)-AKGA → /	(S)-Levi + (RS)-AKGA → /	

Even though this analysis is based on a very limited number of cocrystallizing systems, it can be placed in parallel with systems involving zwitterionic

coformers that easily form diastereomeric pairs and which in consequence easily generate all higher order combinations.[63]

Hence, the fact that a conglomerate has been easily isolated in this system should be put in relation to the ability of AKGA to cyclize in solution and likely to the ability of the enantiomers of Etiracetam to stabilize the lactol tautomer. Even though the peculiarities of this system are not very likely to occur for many other systems, they do show that formation of a conglomerate through cocrystallization is a possibility.

Besides, the conglomerate being more stable than the racemic compound in should also be taken with care, as the ring-chain tautomerism equilibrium has been proven to depend on the temperature, nature of the solvent and pH. The stability ranking and the ease of formation of the conglomerate with respect to the racemic cocrystal may therefore be affected by varying these parameters. A full study allowing for a comprehensive rationalization of this phenomenon goes however beyond the scope of this work.

2.5 Conclusion

In this contribution, we report three novel cocrystal structures involving AKGA and (S)-/(R)-Leviteracetam or (RS)-Etiracetam. Surprisingly, the Levetiracetam cocrystals are formed with the lactol tautomer of AKGA, which has never been isolated in the solid-state up to now. These cocrystals may be generated by neat or liquid-assisted grinding; indicating that solvent is not required for the tautomeric equilibrium to take place. On the contrary, the presence of Levetiracetam in the medium is suspected to influence the equilibrium by selectively stabilizing one enantiomer of the lactol form at a time (depending on the nature of the Levetiracetam enantiomer) or promote its formation by H-bonding.

Concerning the Etiracetam-AKGA system, it was established that two forms may be selectively isolated, depending on the experimental conditions. The first one is a racemic cocrystal, in which AKGA is present as the keto form and the second one is a cocrystal conglomerate, which corresponds to a physical mixture of the two (S)-/(R)-Levi_(R)-/(S)-AKGA(lactol) cocrystals, and which was proven to be more stable than the racemic compound at ambient temperature.

By comparison with two other related systems, it was suggested that the absence of any Levetiracetam-AKGA cocrystal with the keto tautomer of

AKGA is due to the presence of an unfavorable conformation of the racetam molecules in the potential cocrystal, which should significantly reduce its stability in comparison with the existing racemic compound.

Finally, the existence of a stable conglomerate in this system was put in relation with the chiral nature of the AKGA lactol form and the enantiospecificity of the Levetiracetam cocrystals, which is likely related to the ability of the Etiracetam enantiomers to stabilize the lactol tautomer in solution. In particular, it was suggested that for a pseudoquaternary cocrystal (*i.e.* cocrystal made up of two racemate compounds) to exist, the pseudoternary combinations (*i.e.* cocrystal made up of one racemate and an enantiomer of the second compound) should exist and the enantiomers of the two compounds should form a diastereomeric pair at the binary level, rather than behave enantiospecifically.

2.6 References

(1) Aitipamula, S.; Banerjee, R.; Bansal, A. K.; Biradha, K.; Cheney, M. L.; Choudhury, A. R.; Desiraju, G. R.; Dikundwar, A. G.; Dubey, R.; Duggirala, N.; Ghogale, P. P.; Ghosh, S.; Goswami, P. K.; Goud, N. R.; Jetti, R. R. K. R.; Karpinski, P.; Kaushik, P.; Kumar, D.; Kumar, V.; Moulton, B.; Mukherjee, A.; Mukherjee, G.; Myerson, A. S.; Puri, V.; Ramanan, A.; Rajamannar, T.; Reddy, C. M.; Rodriguez-Hornedo, N.; Rogers, R. D.; Row, T. N. G.; Sanphui, P.; Shan, N.; Shete, G.; Singh, A.; Sun, C. C.; Swift, J. A.; Thaimattam, R.; Thakur, T. S.; Kumar Thaper, R.; Thomas, S. P.; Tothadi, S.; Vangala, V. R.; Variankaval, N.; Vishweshwar, P.; Weyna, D. R.; Zaworotko, M. J. *Cryst. Growth Des.* 2012, *12*, 2147–2152.

(2) Schultheiss, N.; Bethune, S.; Henck, J.-O. *CrystEngComm* 2010, *12*, 2436.

(3) Gagniere, E.; Mangin, D.; Puel, F.; Rivoire, A.; Monnier, O.; Garcia, E.; Klein, J. P. *J. Cryst. Growth* 2009, *311*, 2689–2695.

(4) Jung, M.-S.; Kim, J.-S.; Kim, M.-S.; Alhalaweh, A.; Cho, W.; Hwang, S.-J.; Velaga, S. P. *J. Pharm. Pharmacol.* 2010, *62*, 1560–1568.

(5) Remenar, J. F.; Morissette, S. L.; Peterson, M. L.; Moulton, B.; MacPhee, J. M.; Guzmán, H. R.; Almarsson, Ö. *J. Am. Chem. Soc.* 2003, *125*, 8456–8457.

(6) Trask, A. V.; Motherwell, W. D. S.; Jones, W.; Samuel Motherwell, W. D.; Jones, W. *Cryst. Growth Des.* 2005, *5*, 1013–1021.

(7) Huang, K.-S.; Britton, D.; Margaret, L.; C. Etter, T.; Byrn, S.; R. *J. Mater. Chem.* 1997, *7*, 713.

(8) Springuel, G.; Leyssens, T. *Cryst. Growth Des.* 2012, *12*, 3374–3378.

(9) Springuel, G.; Collard, L.; Leyssens, T. *CrystEngComm* 2013, *15*, 7951.

(10) Springuel, G. Chirality and cocrystal systems : from fundamental understanding to development of a novel industrial chiral resolution technique, Université catholique de Louvain, 2014.

(11) Viswanathan, T. S.; Johnson, R. E.; Fisher, H. F. *Biochemistry* 1982, *21*, 339–345.

(12) Cooper, A. J. L.; Redfield, A. G. *J. Biol. Chem* 1975, *250*, 527–532.

(13) Cooper, A. J. L.; Ginos, J. Z.; Meister, A. *Chem. Rev.* 1983, *83*, 321–358.

(14) Kerber, R. C.; Fernando, M. S. *J. Chem. Educ.* 2010, *87*, 1079–1084.

(15) Jones, P. R.; Desio, P. J. *J.Org. Chem.* 1965, *1542*, 4–9.

(16) Groom, C. R.; Bruno, I. J.; Lightfoot, M. P.; Ward, S. C. *Acta Crystallogr. Sect. B Struct. Sci. Cryst. Eng. Mater.* 2016, *72*, 171–179.

(17) Lis, T.; Matuszewski, J. *Acta Crystallogr. Sect. C Cryst. Struct.*

Commun. 1984, *40*, 2016–2019.
(18) Karle, I. L.; Ranganathan, D.; Haridas, V. *J. Am. Chem. Soc.* 1997, *119*, 2777–2783.
(19) Neurohr, C.; Marchivie, M.; Lecomte, S.; Cartigny, Y.; Couvrat, N.; Sanselme, M.; Subra-Paternault, P. *Cryst. Growth Des.* 2015, *15*, 4616–4626.
(20) Jacques, J.; Collet, A.; Wilen, S. H. *Enantiomers, Racemates, and Resolutions*; J. Wiley &.; New York, Chichester, Brisbane, Toronto, 1981.
(21) Morissette, S. L.; Almarsson, Ö.; Peterson, M. L.; Remenar, J. F.; Read, M. J.; Lemmo, A. V.; Ellis, S.; Cima, M. J.; Gardner, C. R. *Adv. Drug Deliv. Rev.* 2004, *56*, 275–300.
(22) ter Horst, J. H.; Deij, M. A.; Cains, P. W. *Cryst. Growth Des.* 2009, *9*, 1531–1537.
(23) Luu, V.; Jona, J.; Stanton, M. K.; Peterson, M. L.; Morrison, H. G.; Nagapudi, K.; Tan, H. *Int. J. Pharm.* 2013, *441*, 356–364.
(24) Yamashita, H.; Hirakura, Y.; Yuda, M.; Terada, K. *Pharm. Res.* 2014, *31*, 1946–1957.
(25) Viedma, C.; Ortiz, J. E.; de Torres, T.; Izumi, T.; Blackmond, D. G. *J. Am. Chem. Soc.* 2008, *130*, 15274–15275.
(26) Viedma, C.; Verkuijl, B. J. V; Ortiz, J. E.; de Torres, T.; Kellogg, R. M.; Blackmond, D. G. *Chemistry* 2010, *16*, 4932–4937.
(27) Viedma, C.; Noorduin, W. L.; Ortiz, J. E.; de Torres, T.; Cintas, P. *Chem. Commun. (Camb).* 2011, *47*, 671–673.
(28) Viedma, C.; Cintas, P. *Chem. Commun. (Camb).* 2011, *47*, 12786–12788.
(29) Noorduin, W. L.; Izumi, T.; Millemaggi, A.; Leeman, M.; Meekes, H.; Van Enckevort, W. J. P.; Kellogg, R. M.; Kaptein, B.; Vlieg, E.; Blackmond, D. G. *J. Am. Chem. Soc.* 2008, *130*, 1158–1159.
(30) Noorduin, W. L.; Kaptein, B.; Meekes, H.; van Enckevort, W. J. P.; Kellogg, R. M.; Vlieg, E. *Angew. Chem. Int. Ed. Engl.* 2009, *48*, 4581–4583.
(31) Van Der Meijden, M. W.; Leeman, M.; Gelens, E.; Noorduin, W. L.; Meekes, H.; Van Enckevort, W. J. P.; Kaptein, B.; Vlieg, E.; Kellogg, R. M. *Org. Process Res. Dev.* 2009, *13*, 1195–1198.
(32) Noorduin, W. L.; Van Der Asdonk, P.; Bode, A. A. C.; Meekes, H.; Van Enckevort, W. J. P.; Vlieg, E.; Kaptein, B.; Van Der Meijden, M. W.; Kellogg, R. M.; Deroover, G. *Org. Process Res. Dev.* 2010, *14*, 908–911.
(33) Spix, L.; Meekes, H.; Blaauw, R. H.; van Enckevort, W. J. P.; Vlieg, E. *Cryst. Growth Des.* 2012, *12*, 5796–5799.
(34) Spix, L.; Alfring, A.; Meekes, H.; Van Enckevort, W. J. P.; Vlieg, E. *Cryst.*

Growth Des. **2014**, *14*, 1744–1748.
(35) Coquerel, G. In *Top Curr Chem*; 2006; pp. 1–51.
(36) Levilain, G.; Coquerel, G. *CrystEngComm* **2010**, *12*, 1983.
(37) Eicke, M. J.; Levilain, G.; Seidel-Morgenstern, A. *Cryst. Growth Des.* **2013**, *13*, 1638–1648.
(38) Lorenz, H.; Seidel-Morgenstern, A. *Angew. Chemie - Int. Ed.* **2014**, *53*, 1218–1250.
(39) Nehm, S. J.; Rodriguez-Spong, B.; Rodriguez-Hornedo, N. *Cryst. Growth Des.* **2006**, *6*, 592–600.
(40) Rodríguez-Hornedo, N.; Nehm, S. J.; Seefeldt, K. F.; Pagán-Torres, Y.; Falkiewicz, C. J. *Mol. Pharm.* **2006**, *3*, 362–367.
(41) Sheldrick, G. M. *Acta Crystallogr. A.* **2008**, *64*, 112–122.
(42) Karki, S.; Friscic, T.; Jones, W.; Motherwell, W. D. S. *Mol. Pharm.* **2007**, *4*, 347–354.
(43) Friščić, T.; Jones, W. *Cryst. Growth Des.* **2009**, *9*, 1621–1637.
(44) Friščić, T. *Chem. Soc. Rev.* **2012**, *41*, 3493.
(45) Tilborg, A.; Springuel, G.; Norberg, B.; Wouters, J.; Leyssens, T. *Cryst. Growth Des.* **2014**, *14*, 3408–3422.
(46) Gu, C.-H.; Young, V.; Grant, D. J. W. *J. Pharm. Sci.* **2001**, *90*, 1878–1890.
(47) Miller, J.; Collman, B.; Greene, L.; Grant, D.; Blackburn, A. *Pharm. Dev. Technol.* **2005**, *10*, 291–297.
(48) Zhang, G. G. Z.; Henry, R. F.; Borchardt, T. B.; Lou, X. **2007**, *96*, 990–995.
(49) Takata, N.; Shiraki, K.; Takano, R.; Hayashi, Y.; Terada, K. *Cryst. Growth Des.* **2008**, *8*, 3032–3037.
(50) Etter, M. C.; MacDonald, J. C.; Bernstein, J. *Acta Crystallogr. Sect. B Struct. Sci.* **1990**, *46*, 256–262.
(51) Valter, R. *Russ. Chem. Rev.* **1973**, *42*, 464–476.
(52) Winston, A.; Sharp, J. C.; Atkins, K. E.; Battin, D. E. *J. Org. Chem* **1966**, *32*, 2166–2171.
(53) Winston, A.; Bederka, J. P. M.; Isner, W. G.; Juliano, P. C.; Sharp, J. C. *J. Org. Chem* **1965**, *30*, 2784–2787.
(54) George, F.; Tumanov, N.; Norberg, B.; Robeyns, K.; Filinchuk, Y.; Wouters, J.; Leyssens, T. *Cryst. Growth Des.* **2014**, *14*, 2880–2892.
(55) Friscić, T.; Fábián, L.; Burley, J. C.; Reid, D. G.; Duer, M. J.; Jones, W. *Chem. Commun. (Camb).* **2008**, 1644–1646.
(56) Issa, N.; Karamertzanis, P. G.; Welch, G. W. A.; Price, S. L. *Cryst. Growth Des.* **2009**, *9*, 442–453.
(57) Karamertzanis, P. G.; Kazantsev, A. V.; Issa, N.; Welch, G. W. A.; Adjiman, C. S.; Pantelides, C. C.; Price, S. L. *J. Chem. Theory Comput.* **2009**, *5*, 1432–1448.

(58) Habgood, M.; Price, S. L. *Cryst. Growth Des.* 2010, *10*, 3263–3272.
(59) Dubey, R.; Mir, N. A.; Desiraju, G. R. *IUCrJ* 2016, *3*, 102–107.
(60) Brock, C. P.; Schweizer, W. B.; Dunitz, J. D. *J. Am. Chem. Soc.* 1991, *113*, 9811–9820.
(61) Springuel, G.; Robeyns, K.; Norberg, B.; Wouters, J.; Leyssens, T. *Cryst. Growth Des.* 2014.
(62) Tumanova, N.; Tumanov, N.; Robeyns, K.; Filinchuk, Y.; Wouters, J.; Leyssens, T. *CrystEngComm* 2014, *16*, 8185.
(63) Tilborg, A.; Springuel, G.; Norberg, B.; Wouters, J.; Leyssens, T. *CrystEngComm* 2013, *15*, 3341.

Chapter 3. Are intermolecular interactions in solution predictive of cocrystal formation?

F. George†, S. A. Kulkarni ‡, T. Leyssens†, Allan S. Myerson ‡*

† Institute of Condensed Matter and Nanosciences, Université catholique de Louvain, 1348 Louvain-la-Neuve, Belgium
‡ Novartis-MIT Center for Continuous Manufacturing and Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, Massachusetts 02139, United States

3.1 Introduction

Cocrystals screening of a given API may be time and compound consuming (trial and error process) and their detection with classical analytical techniques (PXRD, IR…) is not always conclusive. Moreover, high throughput cocrystal screening methods, such as liquid assisted grinding,[18–20] can lead to false negatives.[21]

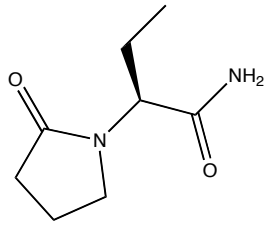
Consequently, attempts are made to improve cocrystal screening procedures in various ways.[22–27] Among these, one finds thermal methods,[23] such as hot stage microscopy to map the melt profile of a two-component system and identify whether there is a simple eutectic profile or not.[24] Other methods sort potential cocrystal formers of a given compound based on the hydrogen bond propensities of specific interactions existing in the CSD structures of compounds having similar functional groups,[25] or by computing the molecular complementarity index of the two partners.[26]

Moreover, even though an extended literature exists on models and techniques to rationalize or improve solution-mediated cocrystallization,[28–31] only a limited amount of work suggests using solution interactions as a way to screen cocrystal former candidates.[32–34]

In this contribution, we examine the possibility of using Isothermal Titration Calorimetry (ITC) to measure interactions between an API and complexing agents in solution, to determine whether these are indicative of successful cocrystal formation. Besides being a possible alternative tool for cocrystal identification, investigation of solution interactions could furthermore yield valuable information on molecular clustering events that take place prior to nucleation.

ITC measures the heat produced or absorbed by the interaction of two compounds in solution, as one is injected incrementally into a solution of the other one, using a power compensation design. ITC measurements yield information on the binding constant, the enthalpy and the stoichiometry of the solution interaction. This technique is often used in biochemistry to measure macromolecule/ligand binding and kinetic interactions.[35] Recently, Weber *et al.* suggested to use ITC to select a complexing agent able to extract an impurity from a solution containing a target compound and a structurally similar impurity, based on the measured interactions between the impurity and various complexing agents in solution.[36]

Levetiracetam, a nootropic drug used as an anticonvulsant in the treatment of epilepsy, was chosen as model API as it does not form a salt, but forms a multitude of cocrystals (Scheme 1).[37,38] We selected a series of complexing agents based on our earlier findings.[39] Five compounds known to form a cocrystal with Levetiracetam and six that do not, were used to study intermolecular interactions with Levetiracetam in solution.

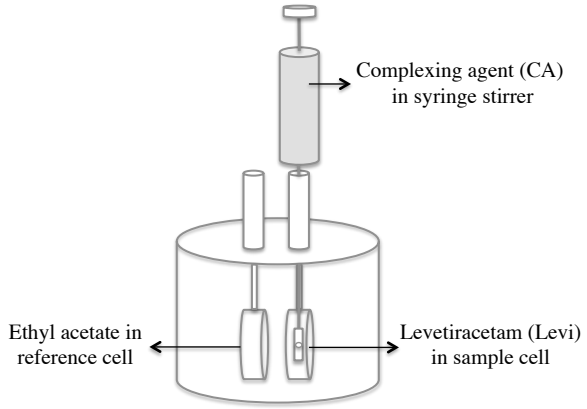


Scheme 1. Chemical structure of Levetiracetam

3.2 Experimental Section

Starting Materials. S-2-(2-oxopyrrolidin-1-yl)butanamide (Levetiracetam, LEVI) was purchased from Xiamen Top Health Biochem Tech. Co., Ltd. Acetylsalicylic acid, indole-3-acetic acid, maleic acid were purchased from Acros Organics. 2,2-dimethylsuccinic acid (DMSA) was purchased from Alfa Aesar. 2,4-dihydroxybenzoic acid (DHBA); 3-nitrobenzoic acid (3NBA) and 5-nitroisophtalic acid, p-coumaric acid, and sorbic acid were purchased from Sigma-Aldrich. Citraconic acid and 6-hydroxy-2naphtoic acid were purchased from TCI. These materials were used as received, without further purification.

Isothermal Titration Calorimetry (ITC). ITC experiments were performed on a TA Instruments Nano ITC calorimeter. The syringe was filled with a solution of the complexing agent (0.1M) while the sample cell was filled with a solution of Levetiracetam (0.01M), using ethyl acetate as solvent. The reference cell contained only pure ethyl acetate (Scheme 2). The complexing agent solution was injected at a dose of 10 µL every 300s in the Levetiracetam solution, for a total of 20 injections, with a stirring rate of 250 rpm, and produced a heat of solution $\Delta H_S^{CA(Levi)}$ (Eq. 1). The heat output of the corresponding blank experiment (complexing agent injected in pure ethyl acetate), $\Delta H_S^{CA(ethyl\ acetate)}$, was then substracted to obtain the heat of complexation ΔH (Eq.2). NanoAnalyze software was used to perform non-linear regression to fit the outcome to an independent binding model, assuming a 1 :1 interaction. To ensure reproducible results, at least three experiments were performed for each complexing agent.



Scheme 2. Schematic representation of the ITC device.

$\Delta H_S^{CA(Levi)}$; CA $_{solution}$ + Levi $_{solution}$ <=> (Levi: CA) $_{solution}$ (1)

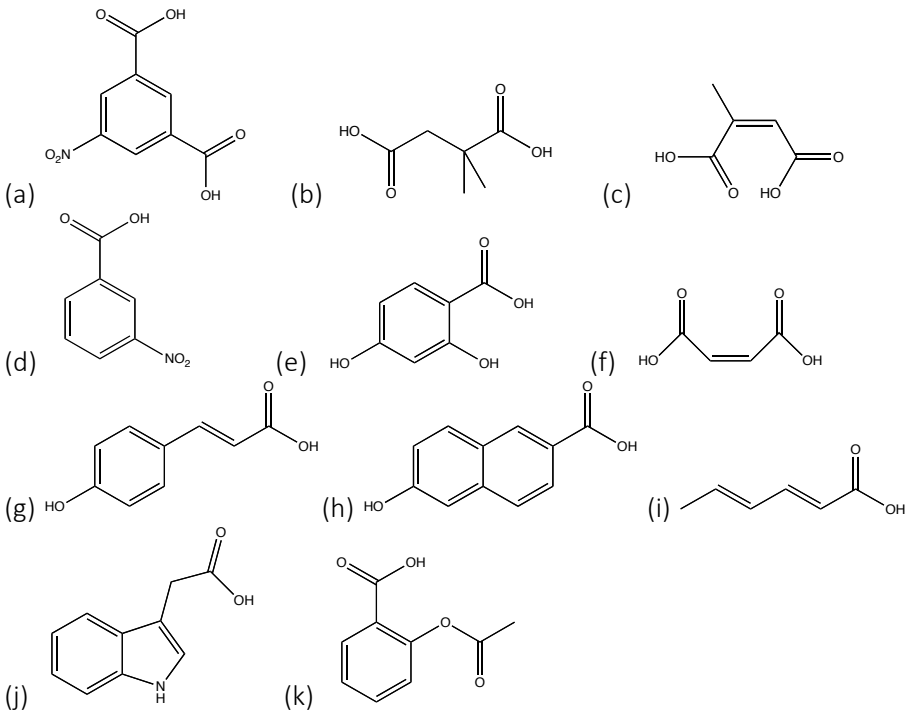
$$\Delta H = \Delta H_S^{CA(Levi)} - \Delta H_S^{CA(ethyl\ acetate)} \quad (2)$$

3.3 Results

3.3.A Selection of the complexing agents

Complexing agents tested in this contribution were selected on the basis of a previously performed cocrystal screen of Levetiracetam with a wide variety of non-chiral coformers.[39] To allow comparison between complexing agents, we chose compounds having at least one carboxylic acid group, which is known to form robust supramolecular synthons (*i.e.* hydrogen-bonding pattern found in a lot of cocrystals) with amides.[39–41] Among the complexing agents tested in the aforementioned screening, we selected five compounds that form a cocrystal with Levetiracetam and six compounds that do not (Scheme 3). Furthermore, care was taken that all coformers selected are soluble in ethyl acetate at a 0.1M concentration.

As the interaction extent was expected to be linked to the structural features of the compounds (see below), complexing agent having other functional groups (double bond, nitro/hydroxyl/second carboxylic acid group) in addition to the first carboxylic acid group were introduced, to evaluate the importance of these functions.



Scheme 3. Structures of the complexing agents selected in the context of this work. Complexing agents forming a cocrystal with Levetiracetam (a-e) and those that do not (f-k). (a) 5-nitroisophthalic acid, (b) 2,2-dimethylsuccinic acid, (c) citraconic acid, (d) 3-nitrobenzoic acid, (e) 2,4-dihydroxybenzoic acid, (f) maleic acid, (g) p-coumaric acid, (h) 6-hydroxy-2-naphthoic acid, (i) sorbic acid, (j) indole-3-acetic acid, (k) acetylsalycilic acid.

3.3.B Measurement of solution interaction thermodynamic data

In theory, the value of the association constant (K), the enthalpy (ΔH°) and the binding stoichiometry of the solution interaction can be drawn from a single experiment. However, in practice, due to the high sensibility of the technique to solution concentrations, at least three experiments were performed to ensure reproducibility, using newly prepared solutions for both the API and the complexing agent. Although one could physically derive the complexation stoichiometry from the overall data, this reduces the precision on the thermodynamic properties or even prevents fitting convergence in case of low affinity systems.[42] For this reason, we preferred assuming a 1:1 interaction so that only the K values and the enthalpies were derived from the experiments. The complexation entropy values were computed using the fitted values of those two parameters using Eq. 3 and 4.

$\Delta G° = -R.T.\ln K°$ (3)

$\Delta G° = \Delta H° - T.\Delta S°$ (4)

In tables 1 and 2, the complexing agents are separated in different groups according to the values of the three parameters.

Table 1 shows the K values to be overall quite small in comparison with the results obtained for salt interactions, which may be two orders of magnitude larger.[36] This is explained by the fact that the interactions investigated do not imply any charged species and are hence less important. Besides, the low precision of the data is due to the low solution concentrations used. However, these could not be increased, as the solubility of the complexing agents and the API in ethyl acetate is limited. We should note here that if this reduced solubility limits the amount of interaction detectable by ITC, it also prevents too strong solute-solvent interactions from occurring. Indeed, as Higuchi and Connors mentioned, lower solubilities favor solute-solute interactions.[43]

Table 1. Association constants and standard free energies between complexing agents and Levetiracetam using ITC at 298K in ethyl acetate.[a] Cocrystal formers are shaded grey.

Complexing agents	**K**	**ΔG° (kJ/mol)**
5-nitroisophtalic acid	8.9 (0.4)	-5.4
2,2-dimethylsuccinic acid	8.0 (1.4)	-5.1
Citraconic acid	7.1 (0.5)	-4.8
Maleic acid	7.1 (2.0)	-4.9
p-coumaric acid	6.7 (0.6)	-4.7
3-nitrobenzoic acid	6.5 (1.9)	-4.6
2,4-dihydroxybenzoic acid	6.3 (1.1)	-4.6
6-hydroxy-2naphtoic acid	5.9 (0.4)	-4.4
Sorbic acid	4.2 (1.0)	-3.6
indole-3-acetic acid	4.1 (0.4)	-3.5
Acetylsalicylic acid	2.9 (0.4)	-2.7

[a] Association constants are average of at least 3 experiments, with corresponding standard deviations displayed in brackets.

From the data presented above, one clearly notes that the magnitude of the K value depends on the nature and number of functional groups present in the complexing agent rather than their ability to form cocrystals. Indeed, all the compounds on the top of the table own two acidic functions while the

ones on the bottom have only one carboxylic acid moiety. The complexing agents possessing two carboxylic acid functions also contain other functionalities in some cases, but this does not seem to present a discriminating behavior. From this table, it therefore seems that the equilibrium interaction constant cannot be used to separate those agents leading to cocrystal formation from those that do not. S. Miller *et al.* reached the same conclusion when trying to predict the formation of inclusion complexes based on solution NMR shifts.[44]

Concerning the enthalpy values, Table 2 shows all of them to be exothermic for all the complexing agents tested. This implies that new hydrogen bonds are being formed when adding the complexing agent to the Levetiracetam solution and that their number and/or strength are more important than those of the hydrogen bonds existing in the pure compound solutions. As the carboxylic acid moiety is the strongest H-bond partner of all complexing agents, we may deduce that the heteromeric amide-acid interaction seems favored over the homomeric (acid-acid and amide-amide) interactions in solution. The amide-acid interaction is also a frequently encountered interaction in cocrystals. But the carboxylic acid function is not the only one responsible for the solution interaction. And it is clear from the data, that complexing agents having the least number of functional groups available for H-bonding present a smaller interaction enthalpy.

Table 2. Enthalpies and entropies of association equilibria between complexing agents and Levetiracetam using ITC at 298K in ethyl acetate.[a] Cocrystal formers are shaded grey.

Complexing agents	**ΔH° (kJ/mol)**	**ΔS° (J/(K mol))**
Acetylsalicylic acid	-25 (2)	-74
5-nitroisophtalic acid	-24 (1)	-62
Citraconic acid	-22 (3)	-58
Maleic acid	-19 (5)	-46
Sorbic acid	-18 (5)	-49
indole-3-acetic acid	-17 (2)	-46
2,2-dimethylsuccinic acid	-17 (4)	-41
6-hydroxy-2naphtoic acid	-16 (3)	-40
p-coumaric acid	-16 (3)	-38
2,4-dihydroxybenzoic acid	-13 (3)	-29
3-nitrobenzoic acid	-9 (3)	-14

[a] Standard enthalpies are average of at least 3 experiments, with corresponding standard deviations displayed in brackets.

This is the case for 2,4-dihydroxybenzoic acid. One possible explanation for this result can be found when studying the interactions present in the cocrystal structure. In the solid-state, an intramolecular hydrogen bond exists between the carbonyl and the adjacent hydroxyl, to form a 6 membered ring (Figure 1), preventing the hydroxyl group from participating in another interaction with the Levetiracetam molecules. Indeed, as Bilton emphasizes, when a donor hydrogen is involved in an intramolecular H bond, it is less likely to participate in an additional intermolecular contact. This is also the case in the chlorzoxazone- 2,4-dihydroxybenzoic acid cocrystal described by Childs and Hardcastle: the hydroxyl group adjacent to the carboxylic acid function takes part in the aforementioned intramolecular bond only.[45]

Hence, if the same bonding pattern exists in solution (*i.e.* the hydroxyl group in position 2 is involved only intramolecularly), it could explain the smaller enthalpy found for this complexing agent. This is highly likely since planar conjugated 6-membered rings are the most probable intramolecular H-bonding motifs and in structures where they are observable, these motifs have a very high likelihood to form.[46,47]

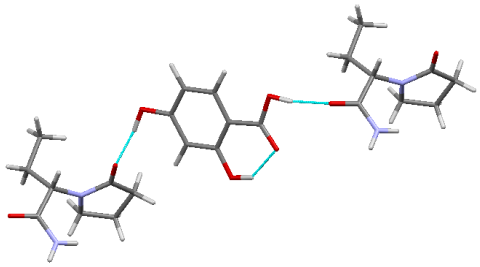


Figure 1. The Levetiracetam - 2,4-dihydroxybenzoic acid cocrystal shows the presence of an intramolecular H bond between the carboxylic acid's carbonyl and the nearest hydroxyl group and the absence of other interaction involving the carbonyl of 2,4-dihydroxybenzoic acid.

As for the enthalpy, the entropy values are all negative (Table 2), partially explained by the reduction in degrees of freedom as the molecules interact. As expected, the group of compounds having the most favorable enthalpy (strongest interaction) shows a stronger reduction in entropy, and reciprocally. This is known as the 'enthalpy-entropy compensation': as the number and strength of binding sites on the interacting molecules increase (tight complex), motions become restricted.[48,49] The final position in the equilibrium constant table is due to the extent of this compensation.

For example, 5-nitroisophtalic acid and acetyl salicylic acid display more negative enthalpies and entropies due to the multiplicity of interaction sites. In case of acetyl salicylic acid, one could attribute the even more negative entropy to the reduction of the ester mobility near the carboxylic acid binding site. Indeed, in the other molecules studied, there is no such proximity between the binding function and other functional groups.

On the contrary, 3-nitrobenzoic acid also displays very low enthalpy and entropy variations. This could be due to the presence of a nitro group, which is not a good hydrogen bond former and does not show any real flexibility. Similarly, the entropy change occurring in presence of 2,4-dihydroxybenzoic acid is quite low supposedly due to the existence of the above-mentioned intramolecular hydrogen bond that reduces the flexibility in the pure compound itself.

3.4 Discussion

Some comments may be formulated concerning the accuracy of the results and their interpretation.

The first one is related to the concentrations of the species and in particular their ratio. In order for the complex to be populated and generate recordable and accurate response, a parameter *c* has been suggested (Eq. 5) for optimal ITC measurement. The usual recommendation is that it should be situated between 10 and 100 (very good) or 5 and 500 (good).[35,50]

$$c = n.K_c.[M_{tot}] \quad (5)$$

where $[M_{tot}]$ is the total concentration in macromolecule (Leviteracetam in our case).

In practice, however, it is often not possible to reach such *c* values due to the limited solubilities of the interacting compounds; which is more problematic for low affinity systems. Both limitations are encountered here, with resulting c values rather situated between 0.03 and 0.1. But one could expect even smaller values for interactions between other functional groups than those studied here as both acid and amide are good hydrogen bond formers.

Turnbull and Daranas[50] have however studied low affinity systems (0.01< c <10) and proved that it is also possible to obtain valid results in those cases by considering other fraction of the isotherms for the derivation of the binding parameters. They suggest notably that the interval $0 \leq [X]/[M] \leq 2$ (X= ligand concentration) used for high affinity systems is inadequate for low affinity systems and rather recommend the injection of multiple equivalents of ligand at a time, which significantly improves the signal-to-noise ratio.

For that matter, Tellinghusien[51] provided more detailed recommendations and suggested to determine the ideal excess titrant over titrate after the final injection (for relative error on K and ΔH below 1%), R ($= [X_{tot}]/[M_{tot}]$), using the following empirical equation (Eq.6).

$$R = \frac{6.4}{c^{0.2}} + \frac{13}{c} \quad (6)$$

In the present case, this would have suggested final ratios included between 155 and 4613, rather than the used ca. 2:1 ratio, which was not feasible in practice due to solubility restrictions. However, one may adapt the concentration of each complexing agent to be as close as possible from its particular solubility rather than use one concentration for all, as it doesn't impact the validity of the thermodynamic parameters comparison. One should however keep in mind that the probability of higher order aggregates (e.g. 1:2 complexes) coexisting in solution with 1:1 complex increases along with concentrations,[52,53] which thus compromises the reliability of the fit, as 1:1 stoichiometry has to be assumed for low affinity systems.

Note also that error in the complexing agent concentration - which cannot be verified by the calculation of the stoichiometry of the interaction, n, in case of low affinity systems- may cause significant inaccuracies in the determination of K and ΔH, with bigger relative effects on K, especially when c is large.[51]

A second remark concerns the nature of the solvent used. Ethyl acetate was used for all systems studied here. Yet it can be expected that different solvents could lead to different conclusions as the competition between the solute-solute and solute-solvent interactions varies accordingly. The K value is indeed solvent dependent. It has furthermore been proven recently by spectroscopic techniques that the solvent nature, through its hydrogen-bonding capacity, is one of the determining factors that influences the nature of the association in solution and the polymorphic outcome of the corresponding crystallization.[54] And it is well known that some solvents

permit crystal growth more easily than others, depending on the system nature.

It seems that He *et al.* had this limitation in mind during their work on predicting multicomponent crystal formation through the measurement of molecular self-diffusivities by pulsed gradient spin-echo NMR,[34] as they chose to use, for each cocrystallizing system analyzed, the solvent in which the cocrystal was successfully grown. However, this removes the idea of predicting cocrystal formation in a given solvent. As one does not have *a priori* knowledge of solvents in which a cocrystal would successfully grow, solvents with different polarities and for which the species show significant solubility differences should ideally be tested, keeping in mind that lower solubilities increase the proportion of solute-solute interactions in the medium. They did not perform such experiments on the systems that do not show cocrystal formation. Hence, for these systems, their experiments may prove that the chances are low that a cocrystal would grow from the solvent studied, but give no information on the probability of occurrence of a cocrystal when using another solvent.

Note that the temperature effect has not been taken into account either.

But even taking these factors into account, one can still question the ability of such in solution analyses to inform on cocrystallization propensity. Indeed, our results clearly indicate that the ability to interact as relatively strong binary complexes may be a necessary condition for cocrystal formation, but is clearly not sufficient. In particular, the formation of one-directional synthons does not require any specific molecular topology while the geometric fit is of primary importance for the creation of a tridimensional closed-packed structure. For an efficient packing, one needs in addition the presence of interactions stabilizing the network in the other directions at later stages of nucleation, which may be quite weak when taken independently and inexistent at low concentrations.

This is in line with the inability of *ab initio* calculations, which are based on energy minimization, [55–59] and other analyses focusing only on changes in H-bonding motifs,[60] to predict cocrystal formation. For that matter, it seems more promising to use qualitative techniques, that allow distinguishing the various interactions at stake under given experimental conditions, and to study the effect of concentrations on these interactions.

As an example, Desiraju and coworkers showed that the presence of octameric synthons involving both hydrogen bonds and π-stacking (Figure

2), may be deduced by combining the results of various NMR experiments (1D ^{1}H NMR chemical shifts upon dilution, 1D difference NOE and ^{13}C longitudinal relaxation time constants).[61] The predictive power of such analyses remains however to be proven.

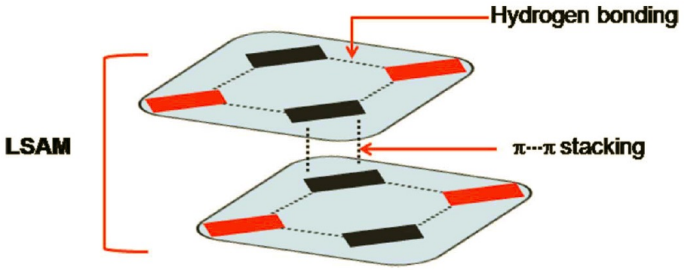


Figure 2. Long-range Synthon Afbau Module (LSAM) arising from the combination of H-bonded tetramers through pi-stacking.[61]

For these reasons, the success of He *et al.* in predicting multicomponent crystal formation[34] may be questioned, in particular due to small number of non-cocrystallizing systems studied. Indeed it seems reasonable to expect other systems showing strong hetero-interactions in solution (or at least hetero-interactions stronger than the corresponding homo-interactions) and not leading to cocrystal formation, as observed in the context of this contribution.

3.5 Conclusion

The interactions in solution between non-charged compounds are quite small but they can still be detected by isothermal titration calorimetry. The accuracy of the results however depends on the API/complexing agent concentrations ratio achievable with respect to the species solubilities and the desire to avoid uncontrolled occurrence of higher level aggregates. The nature of the solvent may also affect the conclusions and should thus be taken into account.

Besides, this technique cannot be used to identify cocrystal formers of a given API as it seems that the amount of interactions is related to the structural features of the complexing agents and not to their ability to form cocrystals. We showed however that the enthalpy values of the complexing equilibrium are negative for all the complexing agents tested, implying that the amide-acid heteromeric interaction is favored over the homomeric

(acid-acid and amide-amide) interaction in solution, as it is often the case in the solid-state.

These findings corroborate previous *ab initio* calculations showing the modest stabilization energy arising from cocrystal formation[56–58] and confirm the need to also consider the feasibility of an efficient packing, through the use of qualitative analytical techniques for example. Finally, the importance of studying non-cocrystallizing systems in parallel to the successful ones has once more been highlighted.

3.6 References

(1) Bauer, J.; Spanton, S.; Henry, R. F.; Quick, J.; Dziki, W.; Porter, W.; Morris, J. *Pharm. Res.* 2001, *18* (6), 859–866.

(2) Fabbiani, F. P. A.; Allan, D. R.; Parsons, S.; Pulham, C. R. *CrystEngComm* 2005, *7* (29), 179.

(3) Kumar, V.; Malhotra, S. V. *ACS Symp. Ser.* 2010, *1038*, 1–12.

(4) Neau, S. H. In *Water-Insoluble Drug Formulation*; Rong, L., Ed.; CRC Press: Boca Raton, 2000; pp 405–425.

(5) Khankari, R. K.; Grant, D. J. W. *Thermochim. Acta* 1995, *248*, 61–79.

(6) Sanphui, P.; Kumar, S. S.; Nangia, A. *Cryst. Growth Des.* 2012, *12* (9), 4588–4599.

(7) Viertelhaus, M.; Hilfiker, R.; Blatter, F.; Neuburger, M. *Cryst. Growth Des.* 2009, *9* (5), 2220–2228.

(8) Espinosa-Lara, J. C.; Guzman-Villanueva, D.; Arenas-García, J. I.; Herrera-Ruiz, D.; Rivera-Islas, J.; Román-Bravo, P.; Morales-Rojas, H.; Höpfl, H. *Cryst. Growth Des.* 2013, *13* (1), 169–185.

(9) Cheney, M. L.; Shan, N.; Healey, E. R.; Hanna, M.; Wojtas, L.; Zaworotko, M. J.; Sava, V.; Song, S.; Sanchez-Ramos, J. R. *Cryst. Growth Des.* 2010, *10* (1), 394–405.

(10) Liao, X.; Gautam, M.; Grill, A.; Zhu, H. J. *J. Pharm. Sci.* 2010, *99* (1), 246–254.

(11) Rasenack, N.; Müller, B. W. *Int. J. Pharm.* 2002, *245* (1-2), 9–24.

(12) Aaltonen, J.; Allesø, M.; Mirza, S.; Koradia, V.; Gordon, K. C.; Rantanen, J. *Eur. J. Pharm. Biopharm.* 2009, *71* (1), 23–37.

(13) Aakeröy, C. B.; Grommet, A. B.; Desper, J. *Pharmaceutics* 2011, *3* (3), 601–614.

(14) Thakuria, R.; Delori, A.; Jones, W.; Lipert, M. P.; Roy, L.; Rodríguez-Hornedo, N. *Int. J. Pharm.* 2013, *453* (1), 101–125.

(15) Tilborg, A.; Michaux, C.; Norberg, B.; Wouters, J. *Eur. J. Med. Chem.* 2010, *45* (8), 3511–3517.

(16) *Pharmaceutical Salts and Co-crystals*, RSC Drug D.; Wouters, J., Quéré, L., Eds.; Royal Society of Chemistry: Cambridge, UK, 2011.

(17) Aitipamula, S.; Banerjee, R.; Bansal, A. K.; Biradha, K.; Cheney, M. L.; Choudhury, A. R.; Desiraju, G. R.; Dikundwar, A. G.; Dubey, R.; Duggirala, N.; Ghogale, P. P.; Ghosh, S.; Goswami, P. K.; Goud, N. R.; Jetti, R. R. K. R.; Karpinski, P.; Kaushik, P.; Kumar, D.; Kumar, V.; Moulton, B.; Mukherjee, A.; Mukherjee, G.; Myerson, A. S.; Puri, V.; Ramanan, A.; Rajamannar, T.; Reddy, C. M.; Rodriguez-Hornedo, N.; Rogers, R. D.; Row, T. N. G.; Sanphui, P.; Shan, N.; Shete, G.; Singh, A.; Sun, C. C.; Swift, J. A.; Thaimattam, R.; Thakur, T. S.; Kumar Thaper,

R.; Thomas, S. P.; Tothadi, S.; Vangala, V. R.; Variankaval, N.; Vishweshwar, P.; Weyna, D. R.; Zaworotko, M. J. *Cryst. Growth Des.* 2012, *12* (5), 2147–2152.
(18) Karki, S.; Friscić, T.; Jones, W.; Motherwell, W. D. S. *Mol. Pharm.* 2007, *4* (3), 347–354.
(19) Friščić, T. *Chem. Soc. Rev.* 2012, *41* (9), 3493.
(20) Friščić, T.; Jones, W. *Cryst. Growth Des.* 2009, *9* (3), 1621–1637.
(21) Rahim, S. A.; Hammond, R. B.; Sheikh, A. Y.; Roberts, K. J. *CrystEngComm* 2013, *15* (19), 3862.
(22) Springuel, G.; Norberg, B.; Robeyns, K.; Wouters, J.; Leyssens, T. *Cryst. Growth Des.* 2012, *12* (1), 475–484.
(23) Lu, E.; Rodríguez-Hornedo, N.; Suryanarayanan, R. *CrystEngComm* 2008, *10* (6), 665–668.
(24) Berry, D. J.; Seaton, C. C.; Clegg, W.; Harrington, R. W.; Coles, S. J.; Horton, P. N.; Hursthouse, M. B.; Storey, R.; Jones, W.; Friščić, T.; Blagden, N. *Cryst. Growth Des.* 2008, *8* (5), 1697–1712.
(25) Wood, P. a.; Feeder, N.; Furlow, M.; Galek, P. T. a.; Groom, C. R.; Pidcock, E. *CrystEngComm* 2014, *16* (26), 5839.
(26) Fábián, L. *Cryst. Growth Des.* 2009, *9* (3), 1436–1443.
(27) Childs, S. L.; Rodríguez-Hornedo, N.; Reddy, L. S.; Jayasankar, A.; Maheshwari, C.; McCausland, L.; Shipplett, R.; Stahly, B. C. *CrystEngComm* 2008, *10* (7), 856.
(28) Leyssens, T.; Springuel, G.; Montis, R.; Candoni, N.; Veesler, S. *Cryst. Growth Des.* 2012, *12* (3), 1520–1530.
(29) Chiarella, R. A.; Davey, R. J.; Peterson, M. L. *Cryst. Growth Des.* 2007, *7* (7), 1223–1226.
(30) Jayasankar, A.; Reddy, L. S.; Bethune, S. J.; Rodríguez-Hornedo, N. *Cryst. Growth Des.* 2009, *9* (2), 889–897.
(31) Nehm, S. J.; Rodrı, B. 2006, No. 2.
(32) Higuchi, T.; Kristiansen, H. 1970, *59* (11), 1601–1608.
(33) He, G.; Jacob, C.; Guo, L.; Chow, P. S.; Tan, R. B. H. *J. Phys. Chem. B* 2008, *112* (32), 9890–9895.
(34) He, G.; Chow, P. S.; Tan, R. B. H. *Cryst. Growth Des.* 2009, *9* (10), 4529–4532.
(35) Freyer, M. W.; Lewis, E. A. *Methods Cell Biol.* 2008, *84*, 79–113.
(36) Weber, C. C.; Wood, G. P. F.; Kunov-Kruse, A. J.; Nmagu, D. E.; Trout, B. L.; Myerson, A. S. *Cryst. Growth Des.* 2014, *14* (7), 3649–3657.
(37) Sekhon, B. S. *ARS Pharm.* 2009, *50* (3), 99–117.
(38) Hurtado, B.; Koepp, M. J.; Sander, J. W.; Thompson, P. J. *Epilepsy Behav.* 2006, *8* (3), 588–592.

(39) George, F.; Tumanov, N.; Norberg, B.; Robeyns, K.; Filinchuk, Y.; Wouters, J.; Leyssens, T. *Cryst. Growth Des.* **2014**, *14* (6), 2880–2892.
(40) Aakeröy, C. B.; Hussain, I.; Forbes, S.; Desper, J. *CrystEngComm* **2007**, *9* (1), 46–54.
(41) Vishweshwar, P.; McMahon, J. A.; Peterson, M. L.; Hickey, M. B.; Shattock, T. R.; Zaworotko, M. J. *Chem. Commun. (Camb).* **2005**, No. 36, 4601–4603.
(42) Turnbull, W. B.; Daranas, A. H. *J. Am. Chem. Soc.* **2003**, *125* (48), 14859–14866.
(43) Higuchi, T.; Connors, K. A. In *Advances in analytical chemistry and instrumentation*; Nurnberg, H. W., Ed.; New York, 1965; Vol. 11, pp 117–212.
(44) Müller, S.; Afraz, M. C.; De Gelder, R.; Ariaans, G. J. a; Kaptein, B.; Broxterman, Q. B.; Bruggink, A. *European J. Org. Chem.* **2005**, No. 6, 1082–1096.
(45) Childs, S. L.; Hardcastle, K. I. *CrystEngComm* **2007**, *9* (5), 364.
(46) Bilton, C.; Allen, F. H.; Shields, G. P.; Howard, J. A. K. *Acta Crystallogr. Sect. B Struct. Sci.* **2000**, *56*, 849–856.
(47) Galek, P. T. a.; Fábián, L.; Allen, F. H. *Acta Crystallogr. Sect. B Struct. Sci.* **2010**, *66* (2), 237–252.
(48) Searle, M. S.; Williams, D. H.; Ah-tu, A. G. *J. Am. Chem. Soc.* **1992**, *114* (8), 10690–10697.
(49) Hunter, C. a. *Angew. Chem. Int. Ed. Engl.* **2004**, *43* (40), 5310–5324.
(50) Jelesarov, I.; Bosshard, H. R. *J. Mol. Recognit.* **1999**, *12*, 3–18.
(51) Tellinghuisen, J. *J. Phys. Chem. B* **2005**, *109* (42), 20027–20035.
(52) He, G.; Chow, P. S.; Tan, R. B. H. *Cryst. Growth Des.* **2010**, *10* (8), 3763–3769.
(53) Rekharsky, M.; Inoue, Y. *J. Am. Chem. Soc.* **2000**, *122* (44), 10949–10955.
(54) Kulkarni, S. a.; McGarrity, E. S.; Meekes, H.; ter Horst, J. H. *Chem. Commun.* **2012**, *48* (41), 4983.
(55) Issa, N.; Karamertzanis, P. G.; Welch, G. W. A.; Price, S. L. *Cryst. Growth Des.* **2009**, *9* (1), 442–453.
(56) Karamertzanis, P. G.; Kazantsev, A. V.; Issa, N.; Welch, G. W. A.; Adjiman, C. S.; Pantelides, C. C.; Price, S. L. *J. Chem. Theory Comput.* **2009**, *5* (5), 1432–1448.
(57) Habgood, M.; Price, S. L. *Cryst. Growth Des.* **2010**, *10* (7), 3263–3272.
(58) Habgood, M.; Deij, M. a.; Mazurek, J.; Price, S. L.; Ter Horst, J. H. *Cryst. Growth Des.* **2010**, *10* (2), 903–912.

(59) Issa, N.; Barnett, S.; Mohamed, S.; Braun, D.; Copley, R.; Tocher, D.; Price, S. *CrystEngComm* **2012**.
(60) Oliveira, M. A.; Peterson, M. L.; Davey, R. J. *Cryst. Growth Des.* **2011**, *11* (2), 449–457.
(61) Mukherjee, A.; Dixit, K.; Sarma, S. P.; Desiraju, G. R. *IUCrJ* **2014**, *1* (2003), 228–239.

Chapter 4. Using CSD solid-form informatics to screen *in silico* for cocrystals of Levetiracetam and Paracetamol

Fanny George & Tom Leyssens

4.1 Outline

Molecular recognition in the context of cocrystallization is a complex process that balances many factors and interaction types. For that matter, the CCDC created software tools based on the knowledge derived from all the structures recorded in the CSD (now nearing 800 000 entries), that could encompass all these influences in a quantitative and systematic manner.[1–6] Their initial effort concerned polymorphism and assessment of existing structures,[2] but they rapidly showed an interest in cocrystal prediction[7,8] and progressively developed new modules entirely dedicated to this problematic.[9–12] These are now fully integrated in the freshly released CSD-Discovery and Materials suites that are destined to both academic and industrial users interested in pharmaceutical development issues.

The current work presents two case studies showing the effectiveness of the CSD solid-form informatics to validate the results of two experimental cocrystal screenings. The APIs chosen to illustrate these procedures are Levetiracetam and Paracetamol. They were selected as both positive (*i.e.* identified cocrystals and their corresponding structures) and negative results (API-coformer combinations that did not give any cocrystal) concerning them were available, either from published studies[13–18] or in-house data[19] leaving us with the opportunity to rationalize the process.

Other research groups have occasionally proposed IT methodologies to improve coformer selection strategy, among which the work on virtual screening by Mesumeci[20] and computational prediction of cocrystals by Price and coworkers.[21–24] But the various solutions offered by the CCDC over the years are more complete and/or contain fewer approximations than others while proving efficient in a diversity of cases. They seemed thus more appropriate for a thorough research on this matter.

In particular, as some of these tools were not initially designed to help coformer selection (H-bond propensity and packing feature modules, see sections 4.2.B and 4.2.E resp.), we shall see their capacity and limits in this context, how to draw maximum benefit from them and some potential sources of improvement.

4.2 Methodology

In the methodology section we will briefly describe the different tools that are available in the CSD informatics package, and which will be used in the context of this work.

4.2.A Motifs search[8–10]

Using the motifs search option in Mercury, it is possible to record the frequency of occurrence of specific inter-/intramolecular contacts between defined functional groups in the CSD or any given structural subset. The identification of robust intermolecular interactions can then help the design of cocrystals.

The procedure involves 4 main steps. The first one consists in creating new motifs or to select pre-defined motifs in the software library (Figure 1). To create new motifs, one has to define the functional groups that will participate in the interactions. Similarly, these can be defined manually or selected in a library.

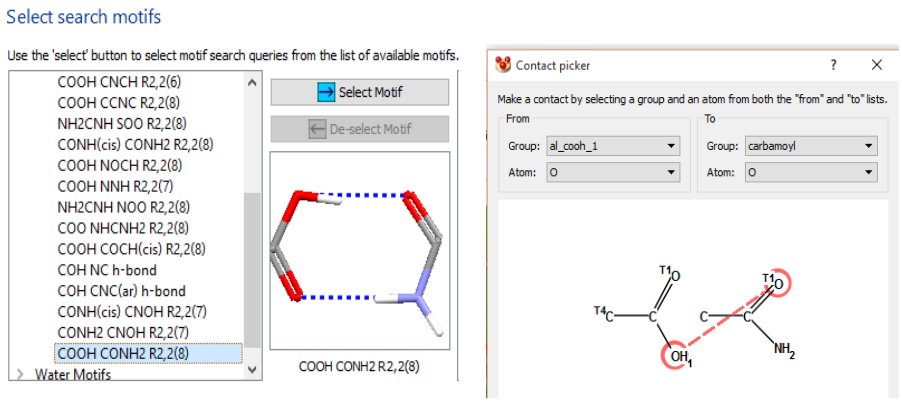


Figure 1. Contact defined between a carboxylic acid and a primary amide using a predefined motif ($R^2_2(8)$ ring, left) or by manually selecting the donor and acceptor atoms involved (right).

Then, for each pair of functional groups forming an interaction, the contact atoms have to be specified. Note that contacts are only defined by a maximum separation distance between atoms and that no angle is considered by the search algorithm.

Third, the type of motif (ring/chain/simple contact) to look for can also be specified.

Finally, the user has to choose the set of structures that will be used as source of information. The entire CSD can be selected but it is rather recommended to use one of 'the best representative refcode lists', as they exclude all redetermined structures from the CSD and thus remove any bias caused by the presence of duplicate entries. There are four of these lists: the best-hydrogen list, the best R-factor list, the best Room Temperature and the best low temperature list; depending on the criterium used to select the best representative structure(s) of each refcode family.

In the context of cocrystal screening, this tool enables eg. to compare the probabilities of homosynthons occurring when compared to the occurrence of heterosynthons. If a heterosynthon is more likely than a homosynthon, coformers containing the corresponding functional group represent good guesses for cocrystallization with the API of interest. (eg. if the API contains a carboxylic acid group, and the amide-carboxylic acid heterosynthon is more likely than the carboxylic acid-carboxylic acid homosynthon, amide containing coformers are considered promising for cocrystallization).

4.2.B Logit-Hydrogen-bonding propensity model[1,7,8,25,26]

The aim of this method is to compute probability scores, named H-bond propensities, for the formation of every potential H-bonding interaction of a target system. This is achieved by creation of a statistical model based on 2D-properties of a set of structures containing functional groups and molecules similar to the target. Such knowledge-based models rest on the assumption that the H-bonds observed in a structure that drived its crystal growth are the most likely ones among all potential donor-acceptor pairing in this structure.

To create a H-bond propensity (HBP) model and make predictions for a target system, one has to go through four steps. Note that prediction molecule(s) can be specified by loading the corresponding cif file or by drawing a 2D molecular diagram (Figure 2).

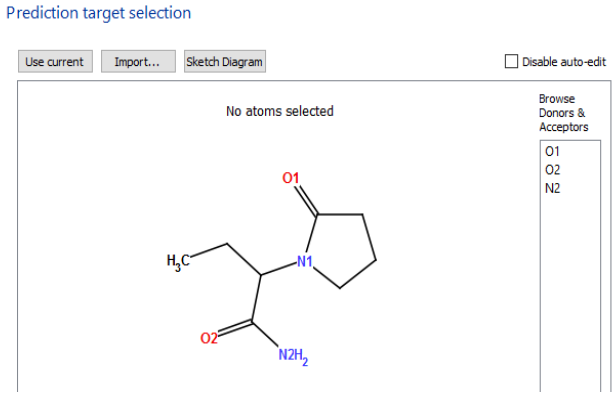


Figure 2. 2D molecular diagram of Levetiracetam

4.2.B.1 Functional group definition and dataset derivation

The first stage consists in extracting, from the CSD or any set of 3D structural data, a subset of crystal structures containing at least one of the functional groups of the target system (eg. the pyrrolidone group of Leviteracetam). Functional groups that will be used to retrieve relevant structures should thus be carefully defined to ensure the prediction reliability of the corresponding propensity model. But there is no unique way to define functional groups. The aim is to be as specific as possible about their chemical environment in the target molecule to maximize the probability of a discriminatory response, but not too specific to obtain a sufficiently large dataset. However the appropriateness of the definitions can only be evaluated after the model fitting and usually requires an iterative procedure. In the software, a library of functional groups is provided that can be used as starting point.

Structures are then combined, in the most even way (*i.e.* there should be similar number of occurrences for each functional group of the target molecules), to form a training dataset representing the possible interactions that might form between all sites of the target compound. Then a similarity search algorithm allows to select structures that are the most similar to the target and to remove duplicate entries due to redetermination. Common groups not present on the target molecule can also be specified at this stage and included in the dataset to later act as additional qualitative parameters and improve the model fitting. The recommended total size of the dataset should be around 1200 structures, depending on the number of defined functional groups.

4.2.B.2 Data extraction and model fitting

In a second stage, descriptive properties are extracted per each donor-acceptor pair for every structure of the dataset.

First, user-variable contact criteria (distances and angles) are used to categorize all pairs as being bonded or non-bonded (true/false observations). Default minimum and maximum donor-acceptor distance cut-offs correspond to the sum of the atomic van der Waals radii -5 and +0,1 Angstrom, respectively; while the minimum donor-hydrogen-acceptor angle is set to 120°. Indeed, Wood *et al.* drew from their analysis that less than 6% of hydrogen bonds in the CSD are characterized by an angle smaller than 120° and that these bonds may in fact be artifacts of stronger interactions or the result of poor structural determination.[27]

Then the model properties are computed. These consist in four molecular/chemical descriptors that can be used as explanatory variables for the fact that certain H-bond donor-acceptor pairs are statistically favored over others in the training set.

One parameter, the functional group categories (for both donor and acceptor), is qualitative but the three others are quantitative functions that account for the specific molecular environment of each D-A pair and guarantee the prediction is specific to the target system. They concern in particular the competition between donors and acceptors, the steric density function around the donor and the acceptor group, and the aromaticity of the donor and acceptor molecules. A detailed description of all parameters can be found in the referential work of Galek,[1] with a brief description given here.

Donor and acceptor functional groups. To each donor/acceptor atom corresponds a binary variable that acts as a switch: when the atom is involved in a H-bond pair, its value is 1 and the corresponding model coefficient influences the model equation to account for the relative strength of this donor/acceptor in comparison with others. When the function is not relevant to the donor-acceptor pair, the variable is set to 0 and is not part of the model anymore.

A supplementary variable, named “other”, is also systematically introduced to collectively account for all the potential donors or acceptors not taken as distinct variables due to their reduced occurrence in the training set.

Competition function: this function compares the number of H atoms available for donation and the number of acceptor lone pairs, D and A respectively, of a chosen donor-acceptor pair (i,a), to the total c in the asymmetric unit (Eq. 1). This function intends to take into account the potential advantage of a group owning several donor/acceptor sites when competing for a partner to form a H-bond. In particular, it allows considering the influence of those donor and acceptor atoms that are not explicitly introduced as separate qualitative variables.

$$K_c(i,a) = \frac{\sum_c D_c + \sum_c A_c}{D_i + A_a} \quad (1)$$

As an example, the competition parameter has been computed for the Levetiracetam H-bonding pair formed by the carbamoyl nitrogen (N2) and oxygen (O2) sites (Figure 3). On the molecule, N2 donor possesses 2 hydrogen atoms prone to H-bonding (D_c = 2 and D_c=D_i as N2 is the only donor) while the two carbonyls (O1 and O2) both have two lone pairs that can be used as H-bond acceptors (A_c= 4 and A_a=2 for O2 in particular). K_c (N2,O2) = (2+4)/(2+2) =6/4= 1,5.

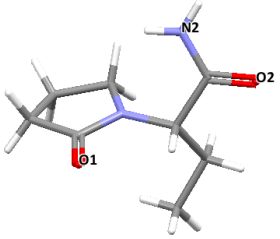


Figure 3. Levetiracetam molecule with labeled H-bonding partners.

Steric density function: for each donor/acceptor functional group, this function assesses the length of its surrounding hydrophobic region using a rather complex algorithm.

Aromaticity function: this function attempts to account for the reduced H-bond potential of a donor/acceptor atom due to the presence of π bonds on its molecule. For each donor/acceptor molecule, it corresponds to the fraction of covalent bonds that have π-character on the total number of non-terminal covalent bonds.

The resulting propensity model is a logit function with linear description of the variable parameters (x_k^i, Eq. 2). α corresponds to the randomly chosen baseline variable while the β_k coefficients are related to the other variables. Model fitting of the *alpha* and *beta* coefficients is carried out via a logistic

regression procedure, to best reproduce all true and false observations of the training dataset. For more information on this kind of models, refer to Hosmer *et al.*[28]

$$\log\left(\frac{P}{1-P}\right) = \alpha + \sum_k x_k^i \beta_k \qquad (2)$$

4.2.B.3 Model validation

Various statistics help to assess the quality of the model. One of these is the AUC (Area Under the Curve) that gives the percentage of well-classified true and false predictions (model sensitivity *vs* specificity) for the training set outcomes, using a variable cut-off distinguishing HB as likely or not. The default value of the cut-off criterion was chosen to be 0.35 in order to equalize the number of correct positive and negative predictions.

If the model AUC is greater than 0.5, it makes correct predictions more frequently than a random choice and if the AUC equals 1 the model is correct every time. An AUC above 0.8/0.9 indicates an excellent/outstanding model respectively, while an AUC lower than 0.7 is sub-optimal and suggests that some defined chemical groups behave too heterogeneously in the selected set of structures. In this case, it is recommended to go backward in the procedure and define these groups differently. To improve the AUC, the user can also remove from the model variables that proved to be non-significant, by deselecting them at the model fitting step. This is suggested for instance when low significance is due to correlation between variables.

Finally, as propensities result from statistical modeling, the size of the 95% confidence intervals built around the predicted values gives information on the quality of the model.

4.2.B.4 Target assessment

When the model is approved, predictions of H-bond propensities of the target molecule(s) can be made.

H-bond propensity values range from 0 to 1. A high H-bond propensity (say > 0.6) means that the H-bond formation is in agreement with expectations derived from the CSD and is not likely to be broken in favor of an alternative

pair in the structure. Conversely, the presence of an unlikely H-bonding pattern in a structure may reveal a stability issue.

A HB pairing score is also computed per putative structure by combining propensities over the corresponding set of HBs. This can be used to compare entire crystal structures.

Hence this tool is useful either to predict H-bond formation in a putative structure or to assess the stability of observed structures and thus finds various applications.

First, this method could help to rank known polymorphs or evaluate polymorphism risk by suggesting the existence of undiscovered stable forms with more favorable H-bond patterns.[25]

It can also be used for cocrystal screening[8] by comparing, for each binary cocrystal system A-B, propensity of the homo- (A with A and B with B) and hetero- (A with B) interactions: if the corresponding H-bond propensity confidence intervals overlap, one may conclude cocrystallization is likely. In this context, Wood *et al.* also suggest the computation of a Multi-Component Score (MC Score) that subtracts the propensity value of the most likely homointeraction from the propensity value of the most likely heterointeraction. According to him, "a positive value indicates that there is a strong hydrogen bonding-based drive towards a cocrystal, while a value close to zero suggests that either outcome is feasible".

A third potential application consists in avoiding solvate formation of a compound by selecting an optimal crystallization solvent. A solvent is considered as appropriate if the solvent-compound heterointeractions propensities are much lower than the propensities of the compound homointeractions.[7]

Finally, such calculations could also help to decrease the size of the putative structure landscape generated by computational crystal structure prediction methods, by identifying candidates with unlikely hydrogen-bonding patterns.[24]

4.2.C Coordination number[6,12]

The coordination number of a given donor/acceptor atom in a structure corresponds to the observed number of H-bonds involving this atom.

Theoretically, every donor/acceptor atom has a finite number of allowed coordinations, depending on their donating/accepting capacity. But in practice this number also varies according to the chemical environment in the structures. Moreover, some potential H-bonds are mutually exclusive.

Besides, according to Galek *et al.*, there is a relation between interaction commonality and stability of a crystal structure. Observed stable structures are thus expected to display (near-) optimal H-Bond coordination while metastable structures should be characterized by unusual coordination numbers. Hence, studying coordination number behaviour could be used to assess hydrogen-bonding efficiency and structural stability.

For that matter, they decided to statistically analyze the molecular interactions of 29 H-bond donor and 57 acceptor group types in a subset of the CSD. First, they record the frequencies of observation for each coordination number (below specified distance and angle thresholds) of these donor/acceptor atoms. As an example, results obtained for carbamoyl as H-bond acceptor are shown in Table 1.

Table 1. Statistics of hydrogen-bond coordination count for the carbamoyl group as acceptor in the CSD subset used for the coordination model creation.

	Coordination number n						
HB acceptor	**0**	**1**	**2**	**3**	**4**	**5**	**Total**
Carbamoyl	12	139	133	5	1	–	290

Then, they were able to generate predictive coordination models for every possible coordination number of each of these groups. Once more, a logit model appeared as the best option to describe the binary probability of a functional group to participate as donor or acceptor in n or more contacts or not. But, contrary to H-bond propensities that are pairwise, coordination likelihoods are related to atoms.

Next, the likelihood of a group to donate/accept exactly once is derived; the sum of all coordination likelihood being one. Coordination likelihoods of the carbamoyl of Levetiracetam in its pure form[a] are given as example in Table 2 (read p_c (1)= 0.13)

[a] The term "pure form" is an abuse of language used in this chapter to designate the crystalline form(s) of a compound containing solely this compound, by opposition to its cocrystals, which contain in addition a coformer.

[b] Very high (distances within +/- 10%, angles within +/-5 degrees) - High (distances within +/- 20%, angles within +/-12 degrees) - Medium (distances within +/- 30%, angles within +/-20 degrees) - Low (distances within +/- 40%, angles within +/-30 degrees) - Custom (editable tolerance values).

[c] From the original list of 152 coformers,[19] 14 coformers were not considered here because

Table 2. Coordination likelihoods for the three potential coordination number (0, 1 and 2) of the carbamoyl of Levetiracetam in its pure form. Underlined values correspond to the coordination number observed in the structure.

Atom	**0**	**1**	**2**
N2 of carbamoyl	0,00	0,13	<u>0,87</u>

Maximum coordination numbers, n_{max}, were chosen using separate cut-off criteria for donors and acceptors, such that both $p(m \geq n_{max})$ is non zero and unusual coordination numbers are excluded.

Eight descriptors were chosen as potential variables. These are the three-dimensional accessible surfaces for donor and acceptor,[3] Gasteiger charges,[29] existence of any five-, six- or seven-numbered intramolecular hydrogen bond involving donors and acceptors, non-H atom count, two-dimensional steric density and ratio of donor to acceptor counts. The presence of intramolecular interactions is taken into account in the computation as it decreases the probability of a donor to form an intermolecular contact.[4,5,30]

For any given donor/acceptor atom in a structure, it is thus possible to compute the probability of observing each coordination number, named coordination likelihood, and to derive the expected/preferred coordination number (maximum likelihood) of this functional group. This functionality is included in the H-bond propensity wizard in Mercury: coordination likelihoods for each donor/acceptor atom of a system are shown in a table, at the final step of the model generation procedure, along with computed H-bond propensity of the system. When evaluating a known structure, observed coordination numbers are highlighted and compared to expectations.

To assess the stability of an entire structure or any feasible hydrogen pattern, the coordination likelihoods of each donor and acceptor atom belonging to the compound(s) of interest are averaged to calculate a mean coordination score.

Finally, as the hydrogen-bond coordination analysis and hydrogen-bond propensity analysis give complementary information, the software also generate a chart that shows these values for all hypothetical hydrogen-bond sets of the target molecule; with optimal combinations of H-bonds in the lower-right hand corner of the chart. On this chart (solid-state landscape),

each point represents a putative structure; the one corresponding to the observed structure being coloured magenta (Figure 4).

Diverse applications can be considered using this methodology. First, it could enable to rank existing polymorphs by order of stability and to assess the probability of polymorphism. Indeed, polymorphism is not expected for compounds having only one feasible arrangement of H-bonds at the bottom right corner of the above-mentioned chart. Second, higher proclivity for cocrystallization/solvation could be revealed in structures having unsatisfied or over-coordinated donor/acceptor atoms. Indeed, the pursuit of the preferred coordination number could constitute the required driving force for crystallization in such cases.

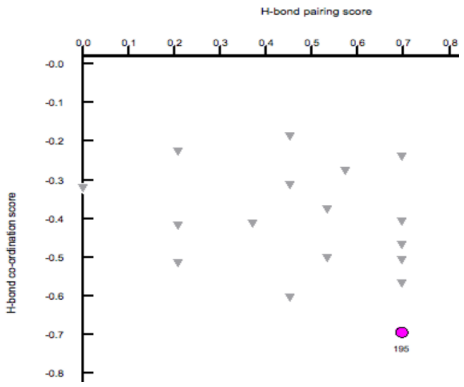


Figure 4. Solid-state landscape of oxalic acid in OXALAC07. The vertical axis corresponds to the mean propensity of the putative H-bond arrangements relative to a zero-base-line, while the horizontal one accounts for their mean coordination. The observed structure appears as the optimal H-bonding combination since both the H-bond propensity and coordination scores reach their maximum values.

4.2.D Molecular complementarity[8,10,31,32]

This tool intends to test coformers according to molecular descriptors beyond synthon matching, in order to differentiate between coformers with identical functional groups, or to discover cocrystallization potential in coformers which do not have common hydrogen-bonding functionalities and/or prone to interactions competitive with H-bonding (such as pi-stacking).

Complementarity between partners was statistically evaluated by studying a dataset of cocrystals in the CSD and high correlation was found for two main

factors: polarity and shape. Five numerical descriptors were selected to describe these factors.

Two are related to the polarity of the molecules. These are the dipole moment and the FNO parameter. This latter corresponds to the Fraction of Nitrogen and Oxygen atoms on the total number of heavy atoms in the molecule and can thus be directly computed from the molecular formula. Similar polarity between coformers is found to increase likelihood of cocrystallization. This can be related to competition between homo- and hetero-molecular interactions and reflects that "*intermolecular interactions between a strongly dipolar molecule and a less polar molecule are unlikely to compete successfully with the strong homomolecular association*".[33]

The shape factor is described by three descriptors chosen to account for the fact that elongated/flat molecules tend to cocrystallize with molecules showing similar shape; matching of absolute dimensions being less important. They refer to the size of a molecular enclosing box (Figure 5) as this simple model was proven efficient to describe molecular aggregation and to identify recurring packing patterns.[34] The length of the short axis of this box (S), the short/long axis ratio (S/L) and the medium/long axis ratio (M/L) were selected to be the more important descriptors. The first two are related to the flatness of the molecules, while the third one intends to spot elongated/rod-shaped molecules.

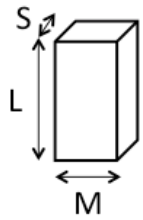


Figure 5. Schematic representation of a molecular enclosing box with definition of its dimensions.

In fact, shape correlation was noticed as an important factor for molecular recognition and close-packing a long time ago by Desiraju,[35] but the novelty of the Molecular Complementarity tool resides in the selection of relevant descriptors and the design of tests based on the pairwise differences in descriptor values to discard unlikely coformers. Threshold values for each descriptor difference were determined such that 90% of the cocrystals studied would be predicted as likely. A coformer is thus selected for further

virtual/experimental analysis if it passes all the tests (*i.e.* differ from the API by less than the corresponding upper limits).

For all the coformers passing the tests, Fabian and Friscic suggest to compute a dissimilarity score, to rank these coformers and prioritize the experiments accordingly, limiting as much as possible the time required for cocrystal discovery. For each descriptor D, the difference in the values computed for the API and the coformer are divided by the corresponding cut-off (δ_D); and the dissimilarity score is obtained by summing the contribution of each descriptor (Eq. 3).[33]

$$Score = \sum_D \frac{|x_{D,API} - x_{D,cof}|}{\delta_D} \quad (3)$$

This tool has been recently implemented in the Mercury software (3.7 version). A wizard devoted to it allows choosing molecular descriptors to be computed and to load and assess the compatibility of an API with a set of coformers and their potential conformations.

Note that this filtering methodology has proven efficient for neutral systems only.

4.2.E Packing feature searches[9,36]

This tool allows the CSD, or any structural data set, to be searched for packing features existing in a structure. Packing feature searches can be applied to assist coformer selection. Indeed, by searching for packing features existing in a known cocrystal structure, we may find new coformers able to realize the same hydrogen-bonding network with the API due to similar spatial disposition of donor/acceptor atoms. In other words, we are looking for compounds able to replace the coformer in the cocrystal framework. Note that the method is also applicable to salts and solvates.

To specify packing features from a "mother" crystal structure, one has to select two or more atoms on the guest molecule to be substituted. The packing feature will then be characterized by the constraints on these atoms specified by the user. The characterization concerns the element type, the required number of bonds and hydrogens for each atom, their bond type, atomic charge and cyclicity.

According to the number of atoms specified in the query, the search will consist in retrieving "child" structures having defined atoms similarly

positioned and oriented (if more than 3 atoms), within distance and angle thresholds. The specification of geometric tolerances falls to the user; 5 options being possible[b]. This method thus allows (sub-)structure comparison without the requirement of any information on unit cells or space-groups.[36]

The outcome is a list of structures with matching fragments found in a specified dataset. Along with each hit comes a root mean square deviation (RMSD) that informs on the similarity level between the mother and the child clusters, computed by superimposing them. One structure may have multiple matches corresponding to different RMSD values. All of them can be viewed but only the best match is displayed in the initial list.

4.3 Practical methodology details

All knowledge based tools used in this work were accessed through the *Materials* module of Mercury (3.6 version for all the analyses except the molecular complementarity tests, for which the 3.7 version was required).[37]

4.3.A Motifs search

Motifs were created using functional groups from the library, except for the oxopyrrolidin moiety, for which a less-specific definition was used (Scheme 1b in section 4.4.A.1) to maximize the number of structures with motifs that contain this moiety.

Simulations were made by forming motifs between the functional groups of the API and the main moieties represented in the screened coformer set, using the CSD 2015 best R-factor list[38] as training set.

4.3.B HBP models

During the HBP model generation procedure, the user needs to make choices concerning the origin of the target system and training set, functional group definitions and parameters used in the final models. Here is the approach adopted in this work.

[b] Very high (distances within +/- 10%, angles within +/-5 degrees) - High (distances within +/- 20%, angles within +/-12 degrees) - Medium (distances within +/- 30%, angles within +/-20 degrees) - Low (distances within +/- 40%, angles within +/-30 degrees) - Custom (editable tolerance values).

First, existing structures were chosen as target systems for all systems except for the Levetiracetam-non-cocrystallizing coformers mixes that had to be drawn, as there are no corresponding structures to be imported.

Second, functional groups were defined by improving the definition of the library groups suggested by the wizard. Specifications focused on the direct environment of the functional groups and were such that each functional group present on the target compounds was represented by at least 350 structures. A qualitative variable for « ether » groups was added to most of the models as this moiety was very frequent in the training sets formed around the target functional groups and helped to significantly improve the models AUC.

Third, training sets were auto-generated by the HBP wizard instead of manually creating a dataset using Conquest, to ensure speed and repeatability.

Fourth, all qualitative and quantitative variables were kept in the final models, as their removal do not help to significantly improve AUC or confidence intervals size.

Note that for each coformer showing polymorphism, results presented in the text concern only the forms having the highest coordination score, as this criterion is recommended to identify the most stable polymorph.[12]

4.3.C Packing Feature searches

For both Levetiracetam and Paracetamol screening, packing features were selected on cocrystallizing coformer molecules from their corresponding cocrystals structures. The number of selected atoms varies with packing type. No restrictions were made concerning the number and nature of bonds formed by the selected atoms, their charge or their cyclicity but for each search, it was specified that "all matched fragments must belong to the same molecule".

4.4 Results

4.4.A Levetiracetam cocrystal screening

Levetiracetam (Levi) is a non-planar molecule with two donors (carbamoyl hydrogens) and two acceptor atoms (carbonyl groups of both amides). These are sufficiently spaced such that 4 different molecules can interact with a given molecule of Levetiracetam, as is the case in the crystal structure of pure Leviteracetam (only one polymorph of Leviteracetam is known, Figure 7).

Levetiracetam has several rotatable bonds and can hence easily adapt its conformation (Figure 6), which leads to different types of packing as observed when Leviteracetam is involved in multi-component crystals (Figure 7).

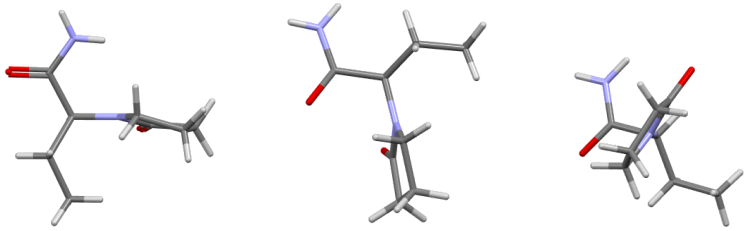


Figure 6. The three main conformations of Leviteracetam as observed: in the crystal structure of pure Leviteracetam (left), in the Levi-dimethylmalonic acid cocrystal (middle) and in the Levi- 3-nitrobenzoic acid cocrystal (right).

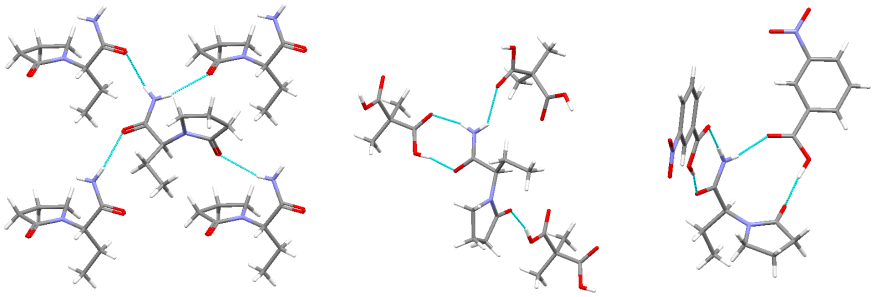
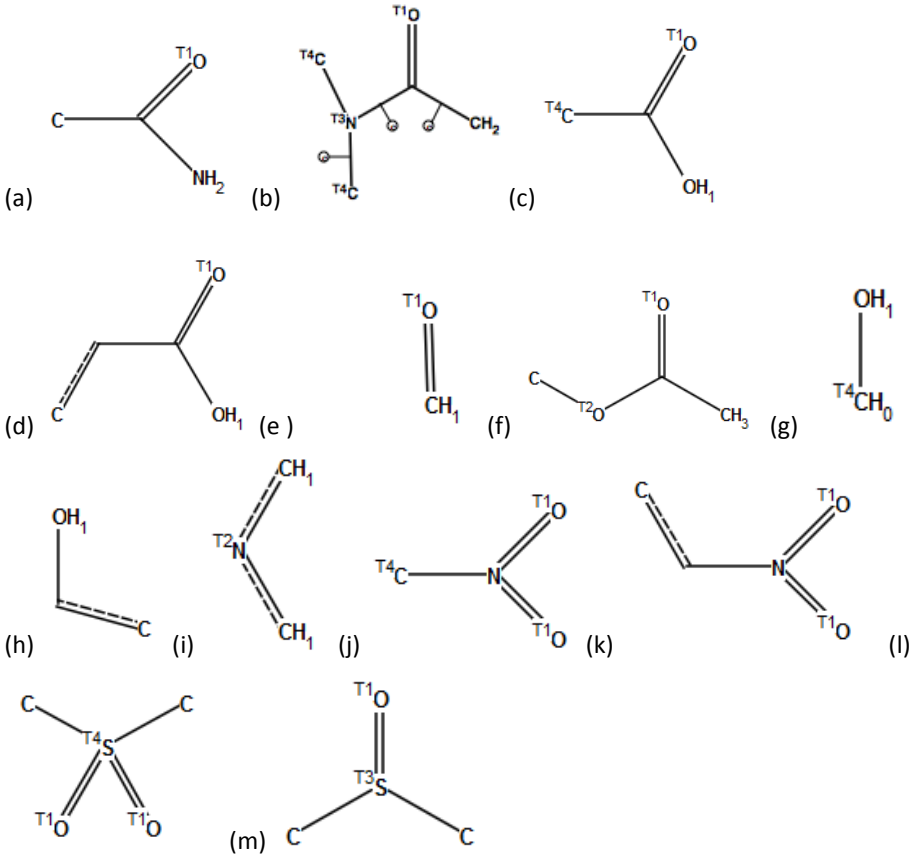


Figure 7. Packing around Leviteracetam in the pure crystal form (left), in Levi-dimethylmalonic acid cocrystal (middle) and in Levi- 3-nitrobenzoic acid cocrystal (right).

4.4.A.1 Motifs search

A search on the most frequent motifs formed by primary and cyclic ternary amides and common functional groups (Scheme 1) in the CSD indicates that they most likely interact with carboxylic acids and aliphatic alcohols (Table 3), both when acting as donor (D) and acceptor (A). Sulfoxides are also designated as very good acceptors for the carbamoyl hydrogens but a visual inspection of the structures showing this interaction reveals that in most cases, the sulfoxides belong to DMSO molecules. The corresponding frequency is thus biased by the presence of solvates in the dataset and from the data one can therefore not explicitly state that sulfoxide containing molecules should be considered as likely coformers. Hence, this result seems to indicate that one should focus on coformers that possess at least one carboxylic acid moiety when looking for Leviteracetam cocrystals.



Scheme 1. Functional groups defined (b) or existing in the library (all but b) to represent the functional groups of Levetiracetam (a-b) and potential coformers (c-n). (a) Carbamoyl, (b) cyclic_AmIII, (c) al_cooh_1, (d) ar_cooh_1,

(e) CH_aldehyde, (f) acetoxy, (g) al_hydroxy_5, (h) ar_hydroxy, (i) aromatic nitrogen, (j) al_nitro, (k) ar_nitro, (l) sulfone, (m) sulfoxide. Tx super-script labels on an atom indicate that it is covalently coordinated to x atoms; while "c" circled labels on bonds indicate these bonds are part of a cycle.

Table 3. CSD frequency of occurrence statistics (Freq) for intermolecular interactions involving the functional groups of Levetiracetam and common functionalities. Total # struct represents the total number of structures which contain both functional groups, and # struct is the sub-group containing a specific interaction between the D-A pair. Bold numbers correspond to competitive heterointeractions.

D	A	Freq	# struct	Total # struct
NH2 of carbamoyl	C=O of carbamoyl	63%	1431	2271
	C=O of AmIII_cycl	66%	27	41
	C=O of al_cooh_1	**79%**	134	170
	C=O of ar_cooh_1	**72.8%**	91	125
	C=O of acetoxy	40%	10	25
	C=O of aldehyde	48%	13	27
	OH of al_hydroxy_5	26%	12	47
	OH of ar_hydroxy	35%	69	196
	S=O of sulfone	56%	5	9
	S=O of sulfoxide	**92%**	24	26
	N=O of al_nitro	57%	4	7
	N=O of ar_nitro	36%	41	114

D	A	Freq	# struct	Total # struct
NH2 of carbamoyl	C=O of carbamoyl	63%	1431	2271
OH of al_cooh_1		57%	97	170
OH of ar_cooh_1		45.6%	57	125
OH of al_hydroxy_5		34%	16	47
OH of ar_hydroxy		28%	55	196

D	A	Freq	# struct	Total # struct
NH2 of carbamoyl	C=O of cyclic_AmIII	66%	27	41
OH of al_cooh_1		**78%**	38	49
OH of ar_cooh_1		47%	7	15
OH of al_hydroxy_5		**69%**	47	68
OH of ar_hydroxy		48%	13	27

If this criterion had been used to select coformers from the original list of 138 coformers investigated experimentally[c], 98 coformers would have been removed prior to performing the experiments. These 98 coformers have an amide, ester, aldehyde, ketone, amine, sulfoxide, aromatic hydroxyl moiety or a mix of these, sometimes in combination with a nitro group or a halogen atom and do not contain any carboxylic group. Indeed, none of these coformers led to cocrystal formation with Levetiracetam.

This first selection would have reduced our initial set to 40 coformers that contain at least one carboxylic acid function. Out of these, only two coformers possess an additional aliphatic hydroxyl group. This shows the interest of performing an a priori motif search, as it would have indicated us to include, more coformers containing such a group in the initial set, as they appear as good hydrogen donors for the ternary amide.

Among the 40 coformer candidates, 8 produced a cocrystal with Leviteracetam during the liquid-assisted grinding screening. The efficiency of this selection step can be visually summarized through a confusion matrix, which compares observed and predicted outcomes (Table 4). It emphasizes the fact that, for the Levi cocrystal screening, a motif search would have produced no false negative while significantly narrowing down the number of experiments to perform.

Table 4. Confusion matrix showing the motifs search results for the Leviteracetam cocrystal screening.

Motifs search		Predicted Outcomes	
		CC	No CC
Actual Outcomes	CC	8	0
	No CC	32	98

Finally, we would like to note that a motif search analysis can also be performed to evaluate the likelihood of solvation (Table 5). From the following table, it appears that incorporation of solvent in the crystal lattice of Leviteracetam cocrystals is not likely, as hetero-interactions between Leviteracetam and carboxylic acid coformers are more competitive. Up to now, there is indeed no known solvated cocrystal of Levetiracetam. However, Etiracetam, its racemic equivalent, possesses one hydrated form (CSD

[c] From the original list of 152 coformers,[19] 14 coformers were not considered here because either they formed an amorphous phase during the screening experiments so we are uncertain about their cocrystallizing ability or they were later found to be duplicates.

refcode OFIREC). This is coherent with the fact that water and carbamoyl are almost equivalent donors for the ternary amide (interaction frequency of 66% and 66.9% resp.). In the Etiracetam hydrate, interaction between water and the ternary amide carbonyl is indeed encountered.

Table 5. CSD frequency of occurrence statistics for intermolecular interactions involving the functional groups of Levetiracetam and common crystallization solvents.

D	A	Freq	# struct	Total # struct
NH2 of carbamoyl	OH2 of water	54.1%	220	407
	N of acetonitrile	39.1%	9	23
	OH1 of methanol	46.8%	22	47
	O of acetone	62.5%	10	16

D	A	Freq	# struct	Total # struct
OH2 of water	O of carbamoyl	56.0%	228	407
OH1 of methanol		40.4%	19	47

D	A	Freq	# struct	Total # struct
OH2 of water	O of AmIII_cycl_2CT4	66.9%	85	127
OH1 of methanol		53.8%	7	13

4.4.A.2 Molecular complementarity tests

Molecular complementarity analyses were then conducted to assess if such criteria can help to further reduce the number of potential cocrystallization partners for Leviteracetam. As this compound has shown to easily adapt its conformation, three different conformers were generated using the corresponding module ("Conformer Generation") in Mercury software and used for the subsequent analyses. Seven of the 40 coformers evaluated as promising according to the motif search are also known or suspected to show various conformations (eg. 2,2-dimethylsuccinic acid and dimethylmalonic acid). Conformers were therefore also produced to consider all possible scenarios. However, no significant difference between the conformers of these 7 coformers was established in this case.

A first test was performed using all the five numerical descriptors suggested by Fabian and coworkers.[31,33] But the results were not satisfying as 5 out of 8

cocrystallizing coformers would have been dismissed (Tables 6 and 7) by the software; the result being identical for all the tested conformers of Leviteracetam. Their failure is due to the fact that Leviteracetam has a cubic shape and these five coformers are judged too flat or too thin.

Table 6. Confusion matrix showing the results of the test of molecular complementarity performed with 5 descriptors for Levetiracetam and 40 potential coformers.

Molecular complementarity tests (5)		Predicted Outcomes	
		CC	No CC
Actual Outcomes	CC	3	5
	No CC	5	27

Table 7. Results of the molecular complementarity analysis for Levetiracetam and the five cocrystallizing coformers failing to at least one of the tests. See section 4.2.D for definitions of descriptor abbreviations and Scheme 2 for definition of compound abbreviations.

Coformers	**M/L**	**Results [a]**	**S (Å)**	**Results [b]**	**S/L**	**Results [c]**
LEVI	0,915		6,861		0,727	
DHBA	0,736	v	3,427	**x**	0,346	**x**
3NBA	0,744	v	3,78	v	0,357	**x**
4NBA	0,634	v	3,528	**x**	0,333	**x**
NIA	0,934	v	3,471	**x**	0,334	**x**
OXA	0,764	v	3,406	**x**	0,476	v

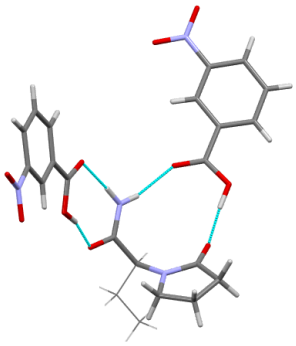
Coformers	**Dipole (Debye)**	**Results [d]**	**FNO**	**Results [e]**
LEVI	2,924		0,333	
DHBA	1,373	v	0,364	v
3NBA	1,46	v	0,417	v
4NBA	1,419	v	0,417	v
NIA	1,503	v	0,467	v
OXA	0,002	v	0,667	**x**

a: pass if delta < 0.31, b: pass if delta < 3.23; c: pass < 0.275; d: pass if delta < 5.94; e: pass if delta < 0.294. Passing mark v, failing mark x.

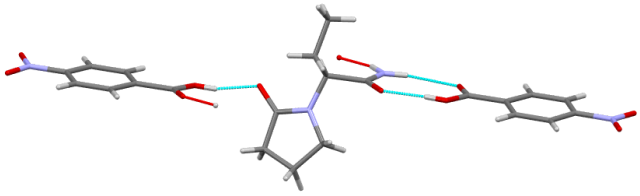
These dismissals can however be questioned. Indeed, it appears from the original study of Fabian that thickness similarity was proven only for planar molecules. In other words, planar molecules prefer to crystallize with other

planar molecules while there is no correlation concerning non-planar molecules. So this may signify that, contrary to planar molecules, non-planar species can accommodate partners with various shapes. Levetiracetam being non planar (S>5 and S/L >0.6), the two related shape descriptors thus seem non-relevant. Moreover, coformer planarity seems more important when the API possess moeities prone to pi-stacking, which is not the case here. By contrast, the M/L test appears pertinent as Levi shows a large value (0.915) for this ratio.

Besides, while not being planar, Levetiracetam moeities may form coplanar synthons. Indeed three of the four H-bonding atoms (those belonging to the primary amide) of any Levetiracetam molecule are automatically coplanar and the last one may be oriented approximately along the same plane due to the conformational flexibility of the molecule (Figure 8). This may in fact explain the tendency of Levetiracetam cocrystals to form 2D layers rather than 3D networks (see chapter 1).



(a)



(b)

Figure 8. Hydrogen bonding in Levi_3-nitrobenzoic acid (a) and in Levi_4-nitrobenzoic acid (b) cocrystal, showing two almost coplanar heterosynthons formed by the two Levetiracetam amides.

Hence, a new analysis was performed with only 3 descriptors (the M/L ratio and the two polarity descriptors). All the cocrystallizing coformers except oxalic acid[d] are then retained, but the same applies to almost all the non-

[d] Rejection of oxalic acid is due to a disparity of FNO parameters. This result reminds us that the descriptors of molecular complementarity express general trends and that exceptions can

cocrystallizing coformers selected. In this situation the molecular complementarity tool becomes useless, and does not substantially improve the screening efforts (Table 8).

Table 8. Confusion matrix showing the results of the test of molecular complementarity performed with 3 descriptors for Levetiracetam and 40 potential coformers.

Molecular complementarity tests (3)		Predicted Outcomes	
		CC	No CC
Actual Outcomes	CC	7	1
	No CC	30	2

be expected. In particular, Fabian notes that molecules with dissimilar polarity can gather provided that their corresponding packing show effective segregation of the polar and apolar groups,[31] which is indeed the case in the Levi- oxalic acid cocrystal.

4.4.A.3 Hydrogen-bond propensities and coordination likelihood models

An analysis of the hydrogen-bond propensities (HBP) of the interactions present in the known structure of Levetiracetam (OMIVUB) and of the coordination likelihoods (CL) of its atoms (Tables 9 and 10 resp.) show that the two observed interactions have very high propensities and that optimal coordination is reached by every atom. Besides, the H-bond pairing combination observed is optimal from HBP and CL points of view (Figure 9). The observed structure is thus inherently stable and the driving force for Levetiracetam to cocrystallize is not expected to be large.

Table 9. Hydrogen-bond propensities of the two interactions observed in Levetiracetam (OMIVUB).

Donor	Acceptor	Propensity
N2 of carbamoyl	O1 of cyclic_AmIII	0.90
N2 of carbamoyl	O2 of carbamoyl	0.82

Table 10. Coordination likelihoods for the donor/acceptor atoms in Levetiracetam (OMIVUB). Underlined values correspond to the observed coordination numbers in the structure.

D/A	Atom	0	1	2
D	N2 of carbamoyl	0,00	0,13	0,87
A	O1 of cyclic_AmIII	0,34	0,64	0,03
	O2 of carbamoyl	0,01	0,88	0,11

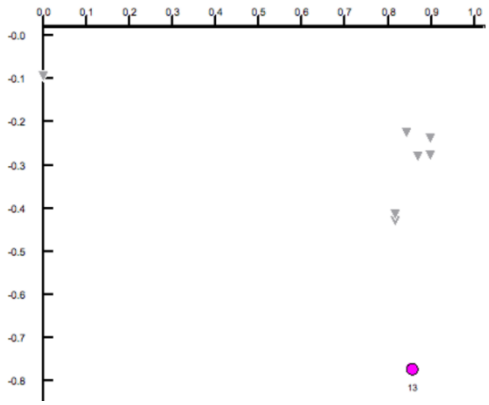
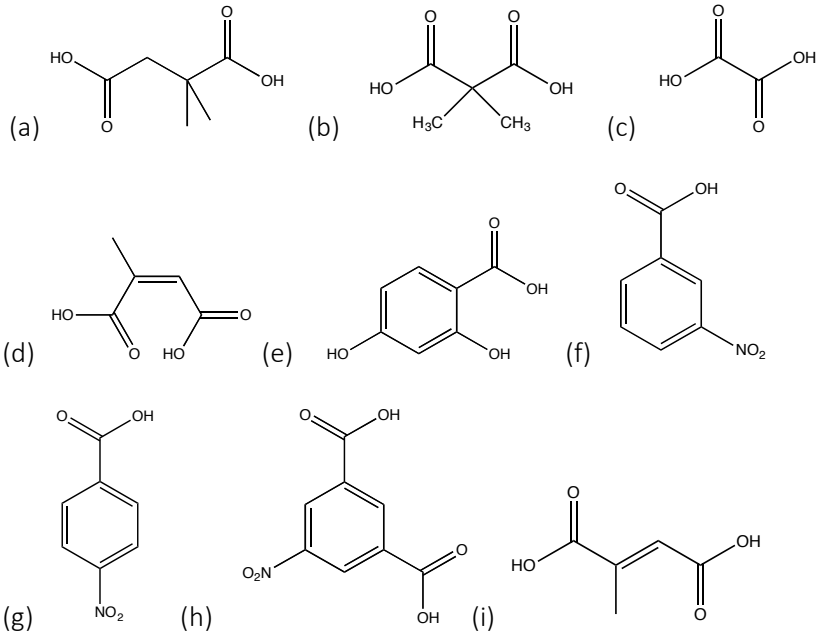
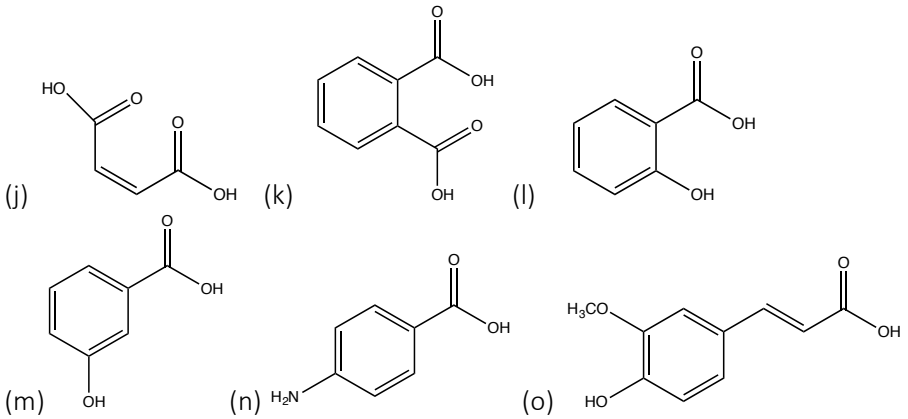


Figure 9. Chart showing the H-bond coordination score in function of the H-bond pairing score for all the putative combination of H-bond pairs in Levetiracetam (OMIVUB). The pink dot corresponds to the observed structure.

Similarly, HBP models were built for the 8 coformers previously identified as cocrystallizing (Scheme 2 a-h)[19], their Leviteracetam cocrystals and 7 non-cocrystallizing coformers (Scheme 2 i-o). These latter were selected among the 40 non-cocrystallizing coformers remaining after the motifs search, using the following criteria: first, only rigid molecules were selected to limit conformation issues (see section 4.5.C). Among the remaining candidates, coformers having structures with disorder, unresolved hydrogen atom positions and/or Z'>1 were dismissed. Then, priority was given to non-cocrystallizing coformers having structural similarities with cocrystallizing coformers, in order to facilitate the comparison of their properties and evaluate if one of these can be used to discriminate cocrystallizing coformers.

The analysis starts with the HBP parameters. From Table 11, it appears that the difference between the HBP of the best hetero-interaction and the best homo-interaction, Δ max HBP (also called MC Score by Wood, see section 4.2.B.4), is slightly negative (except for the Levi-ABA mix) or nul taking into account the statistical uncertainty for all the Levi-coformers pairs. Hence this parameter cannot be used to estimate the cocrystallization ability of a system.





Scheme 2. Molecular diagrams of the coformers selected for HBP-CL analysis. Coformers forming a cocrystal with Levi (a-h) and those that do not (i-o). (a) 2,2-dimethylsuccinic acid (DMSA), (b) dimethylmalonic acid (DMMA), (c) oxalic acid (OXA), (d) citraconic acid (CCA), (e) 2,4-dihydroxybenzoic acid (DHBA), (f) 3-nitrobenzoic acid (3NBA), (g) 4-nitrobenzoic acid (4NBA), (h) 5-nitroisophthalic acid (NIA), (i) mesaconic acid (MSA), (j) maleic acid (MLA), (k) phthalic acid (PTA), (l) salicylic acid (SLA), (m) 3-hydroxybenzoic acid (HBA), (n) 4-aminobenzoic acid (ABA), (o) ferulic acid (FRA).

In fact, by comparing the best homo-interaction in pure Levetiracetam and pure coformers (Max HBP, Table 12) with the values for the corresponding cocrystals (hetero max HBP, Table 11), we see that the maximum HBP of Leviteracetam (0.9, Table 12) is larger than the corresponding values in all its cocrystals and that the situation is reverse for the pure coformers. In other words, the carboxylic acid-amide heterointeraction is computed to be less likely than the amide-amide interaction in pure Levi, which is in contradiction with the results of the motif search. But the carboxylic acid-amide hetero-interaction is more likely than the carboxylic acid -carboxylic acid homo-interaction. And as the MC Score is negative, this means that, from the point of view of HBP, the benefit conferred to the coformers is offset by the disadvantage felt by the API.

Two other parameters were computed to describe the structures in their entirety instead of focusing on a single interaction: the hetero HBP Score, which corresponds to the average of the HBP of the interactions observed in

the cocrystals; and the homo HBP score, which is computed by averaging the HBP of the interactions observed in both partners.[e]

Table 11. Keys values for the analysis of H-bond propensities in Leviteracetam cocrystals. Non-cocrystallizing Levi-coformer mixes are shaded grey. HBP scores uncertainty is based on Chi^2 statistics and its size is related to the amount of contributory CSD data used to determine the HBP of the pairwise interactions. It usually amounts to 10 to 25% if we sum up the errors on both intervals.

Cocrystals/mixes	hetero max HBP	homo max HBP	Δ max HBP	Hetero HBP score	max?	Homo HBP Score	Δ HBP Score
DMSA	0,73	0,85	-0,12	0,63	No	0,61	0,02
DMMA	0,78	0,87	-0,09	0,71	No	0,68	0,03
OXA	0,76	0,8	-0,04	0,72	No	0,72	0
CCA	0,72	0,85	-0,13	0,67	No	0,65	0,02
MSA	0,71	0,84	-0,13	-	-	-	-
MLA	0,73	0,86	-0,13	-	-	-	-
PTA	0,75	0,84	-0,09	-	-	-	-
DHBA	0,78	0,8	-0,02	0,55	No	0,48	0,07
SLA	0,69	0,81	-0,12	-	-	-	-
HBA	0,71	0,86	-0,15	-	-	-	-
3NBA	0,66	0,79	-0,13	0,59	No	0,58	0,01
4NBA	0,66	0,79	-0,13	0,59	No	0,59	0
NIA	0,66	0,82	-0,16	0,61	No	0,54	0,07
ABA	0,9	0,82	0,08	-	-	-	-
FRA	0,76	0,84	-0,08	-	-	-	-
Lower bound	0,66	0,79	=MC Score	0,55		0,48	
Upper bound	0,9	0,87		0,72		0,72	

Contrary to the MC Score, the difference in HBP Scores (ΔHBP Score, Table 11) is slightly positive or null for all the Levi-coformers systems. We also noticed that all cocrystals have sub-optimal hetero HBP Scores. This emphasizes the importance of other factors for crystallization.

Concerning the coordination Scores (Table 13), there is almost no variation between the homo- and hetero-combinations (ΔCL Score) for all the cocrystallizing coformers. Hence, if there is a driving force toward cocrystallization, it does not come from an improvement of the CL Score either in this case. This was expected, as coordination seems of less importance for self-complementary functional groups such as carboxylic acids and amides, which possess both good donors and acceptors.

[e] Homo HBP score were calculated using the values obtained from the corresponding cocrystal models, for valid comparison with the hetero HBP Score (*i.e.* HBP have to originate from the same model).

Finally, note that half of the cocrystals show their most likely coordination and in case of sub-optimal values, non-optimality is mainly due to the presence of either a hydroxyl or a nitro group on the molecules.[f]

Some comment may also be formulated about the absolute value of the HBP and CL Scores in the coformer structures and in the various cocrystals.

In the pure coformers structures, both the HBP and the CL Scores are maximum or close to it (Table 12). But in some cocrystal structures (Table 11), we observe a HBP Score far from the maximum value (the difference between the two is > 0.2) while their CL Score is either maximum or very close to it (see the corresponding solid-state landscapes in the appendices). This suggests that achieving a high CL Score is a necessary condition for the stability of any structure and that any multi-component score should take it into account, which is not the case for the MC-score proposed by Wood.[8]

Table 12. Keys values to analyze the H-bonding pattern of coformers cocrystallizing with Levetiracetam. Non-cocrystallizing Levi-coformer mixes are shaded grey.

Coformers	Max HBP	HBP Score	max?	CL Score	max?
Levi	0,9	0,86	Yes	0,8	Yes
DMSA	0,46	0,44	Yes	0,76	Yes
DMMA	0,66	0,66	Yes	0,76	Yes
OXA	0,7	0,7	Yes	0,72	Yes
CCA	0,51	0,5	Yes	0,79	Yes
MSA	0,51	0,51	Yes	0,79	Yes
MLA	0,54	0,54	Yes	0,77	Yes
PTA	0,57	0,57	Yes	0,78	Yes
DHBA	0,35	0,29	No	0,76	+/-
SLA	0,32	0,32	Yes	0,81	Yes
HBA	0,26	0,24	No	0,71	+/-
3NBA	0,54	0,54	Yes	0,61	No
4NBA	0,57	0,57	Yes	0,61	No
NIA	0,51	0,51	Yes	0,69	Yes
ABA	0,7	0,42	No	0,53	No
FRA	0,56	0,35	No	0,72	No
Lower bound	0,26	0,24		0,53	
Upper bound	0,7	0,7		0,81	

[f] We noticed that the CL Score never reaches one and that values around 0,7 can be considered as high.

Table 13. Keys values for the analysis of coordination likelihoods (CL) in Levetiracetam cocrystals. Non-cocrystallizing Levi-coformer mixes are shaded grey.[g]

Cocrystals/mixes	Hetero CL Score	max?	Homo CL Score	Δ CL Score
DMSA	0,78	Yes	0,78	0
DMMA	0,78	Yes	0,79	-0,01
OXA	0,73	Yes	0,73	0
CCA	0,79	Yes	0,79	0
MSA	-	-	-	-
MLA	-	-	-	-
PTA	-	-	-	-
DHBA	0,7	No	0,76	-0,06
SLA	-	-	-	-
HBA	-	-	-	-
3NBA	0,66	No	0,66	0
4NBA	0,66	No	0,66	0
NIA	0,72	No	0,72	0
ABA	-	-	-	-
FRA	-	-	-	-
Lower bound	0,66			
Upper bound	0,79			

For all the cocrystals structures studied, the CL Score is superior to 0.65 (Table 13) and the HBP Score is greater than 0.55 (Table 11). Besides, there is no contact with a propensity lower than 0.35 in the cocrystals studied. All observed interactions are thus more likely than not to be formed. This is not the case though in some pure coformer structures (DHBA, ABA, FRA, SLA and HBA), that display very low propensity interactions (interactions realized by the OH and NH_2 group), but this does not prevent them from existing nor push them towards cocrystallization (as ABA, FRA, SLA and HBA do not form a cocrystal with Levi).

Finally, we observe that all the donor atoms are satisfied in all Leviteracetam cocrystals while this is not always the case for acceptor atoms (as in Levi-DMMA cocrystal). This is in accordance with the observation of Galek *et al.* that "*it is more favourable to satisfy good donors with the acceptors available, rather than reducing hydrogen-bond coordination at poor acceptors and leaving good donors unsatisfied*".[6]This is due to the fact that most organic molecules have a preponderance of acceptors over donors

[g] Statistical uncertainty on the CL Scores cannot be evaluated as no information is given concerning the size of the error on the coordination likelihoods computed by the software.

such that acceptors are more likely to remain unsatisfied.[33] This implies a competition between acceptors for donors and the importance for coformers to display good acceptors.

However, as Fabian noticed, such imbalance in partner molecules gives no indication on their cocrystallization probability. Indeed this disparity is likely to continue in the cocrystal, since an abundance of acceptors on one molecule is usually not compensated by its partner (which, in all likelihood, also has a majority of acceptors).[33]

Besides, as Musumeci *et al.* assumed in their work on virtual cocrystal screening:[20] "*If the number of HB donor and acceptor sites differ, the molecules can be arranged in such a way that the excess sites are not forced to make unfavorable contacts with each other, rather than find gaps or regions of low electrostatic potential such that they make no contribution to the total electrostatic interaction energy*". Hence, the presence of unemployed acceptors seems of low incidence as far as close-packing is concerned as, in most compounds, there will be outward hydrogen atoms to face them; while donors seem to be limited to acceptors as buffers. This also suggests that a compound with an excess of donors has more incentive to find a partner possessing acceptors than reciprocally for a compound having a majority of acceptors.

In conclusion, both HBP and coordination models do not permit to discriminate between coformers with a carboxylic acid function. Hence the ability of some coformers to cocrystallize must come from a difference in packing feasibility of the potential combinations of H-bonds. Indeed, both levels of analysis (intermolecular and supramolecular) have to be considered to evaluate cocrystallization likelihood, especially as close-packing may be in conflict with directional factors.[35]

4.4.A.4 Packing Feature Search

As of today, there is no knowledge-based tool in the CCDC Solid Form suite that can assess or suggest packing arrangements. However, if we know one or several structures of a compound (as if it is frequently the case in the CSD), the Packing Feature Search module can be useful as it starts from well-proven packings.

Among the 40 coformers (selected from the motifs search), the subsequent analysis was limited to coformers having a minimum of two carboxylic acids, as it turned out that the packing feature searches only properly work when

at least two acceptor atoms are selected. This left us with a set of 24 components including 5 coformers known to cocrystallize with Levetiracetam.

As packing features, C=O atoms of each carboxylic acid moiety were selected from the DMSA, DMMA, CCA, OXA and NIA molecules in their respective cocrystals (mother structures); the inclusion of carbon atoms in the selection ensured that the acceptor lone pairs pointed in the right directions. Medium tolerance on angles and distances (+/- 20 degrees and 30% respectively) was applied to all searches and it was specified that the number of hydrogens on the hit molecules have to match these of the selected atoms, to aim for carbonyls only. The screened dataset was made up of the 24 coformers and all the conformers generated for the 8 flexible coformers (mainly aliphatic carboxylic acids). The results of these analyses are shown in Table 14.

Table 14. Results of the Levetiracetam packing feature searches. Corresponding CSD refcodes are specified in brackets for cocrystal structures already recorded in the CSD. Underline refcodes corresponds to the pure form of the coformer used for the search in the mother structure.

Mother structures	Hits	Coformers	Cocrystals	RMSD
Levi-DMSA (XOGMOX)	OLENIC	2,2-dimethylsuccinic acid	Levi-DMSA	0,182
	NEDNOZ01	Citraconic acid	Levi- CCA	0,285
	RUCFAX	Citric acid	-	0,296
Levi-DMMA	RUCFAX	Citric acid	-	0,229
Levi-OXA (XOGPEQ)	OXALAC07	Oxalic acid	Levi-OXA	
Levi-CCA	ACSALA	Acetylsalicylic acid	-	0,252
	MESCON	Mesaconic acid	-	0,532
Levi-NIA	BTCOAC	1,3,5-benzenetricarboxylic acid	-	0,081
	VARJAA	Isophtalic acid	-	0,071

A few remarks may be formulated about the results. First, we notice that the pure forms of CCA, DMMA and NIA (refcodes: NEDNOZ01, MMALAC01 and COFDUW10 respectively) were not found to match the query started from their corresponding cocrystal. This is due to a significant conformational change between their two respective forms, meaning that the conformer generator module is not as effective as expected or that the compound conformation in the cocrystal is rather unlikely and could not be predicted.

Second, it appears that there is some similarity between the conformations of DMSA in its cocrystal and CCA in its pure form. However, when comparing the packing patterns of the two cocrystals, we observe that they are in fact

very different (Figure 10): in Levi-DMSA, there is no carboxylic acid-carbamoyl $R^2_2(8)$ ring between the two partners, contrary to what happens in the Levi-CCA cocrystal, and the second carboxylic acid of the coformer molecules (for which interactions are hidden) are oriented differently in the two cocrystals. This is in fact the case for all the cocrystallizing coformers used for this search, as it can be seen on Figure 11. These observations suggested the use of stricter search criteria.

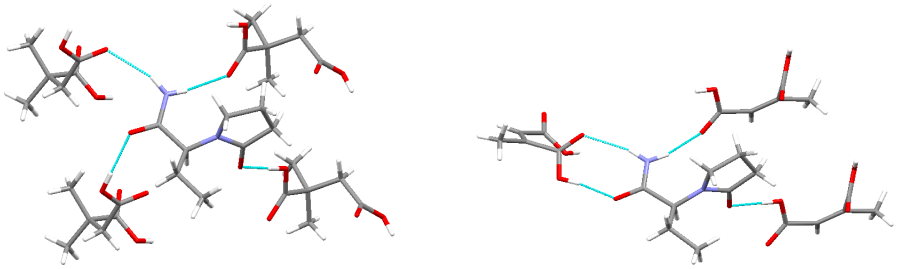


Figure 10. Packing patterns around Levetiracetam in Levi-DMSA (left) and Levi-CCA (right) cocrystals.

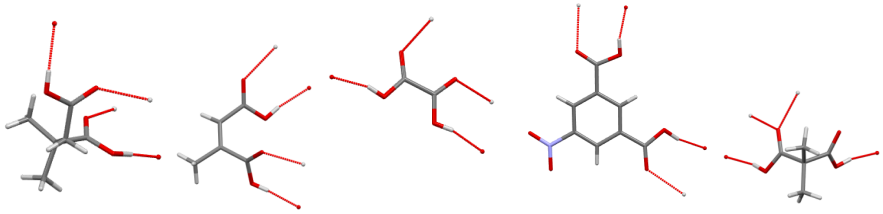


Figure 11. Carboxylic acids orientation for coformers in cocrystals. From left to right: DMSA in Levi-DMSA, CCA in Levi-CCA, OXA in Levi-OXA, NIA in Levi-NIA and DMMA in Levi-DMMA.

4.4.B Paracetamol cocrystal screening

Paracetamol (Para) is a planar molecule that owns two hydrogen-donors (one amide hydrogen and one hydroxyl group) and two hydrogen-acceptors (one carbonyl and one hydroxyl group). In pure Paracetamol, all of them are involved in hydrogen bonds such that each molecule is bound to 4 others (Figure 12) to form approximate planar layers.

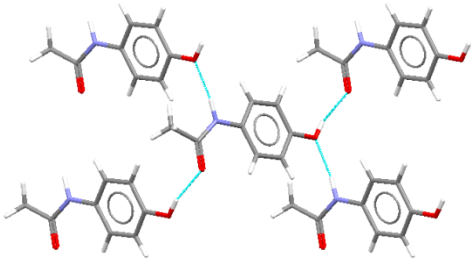


Figure 12. View of paracetamol (hxacan11) along the *b*-axis.

However, an analysis of the coordination numbers of these four groups in the three polymorphs of paracetamol (CSD refcodes HXACAN 11, 29 and 33)[h] suggests that the aromatic hydroxyl group prefers not to accept hydrogen bonds (Table 15). Besides, comparing bond lengths, it seems that the bond formed between the NH donor and the OH acceptor is weaker than the interaction between the OH donor and the amide acceptor.[17]

Table 15. Coordination likelihoods for the donor/acceptor atoms in polymorph I of Para (HXACAN 33). Underline values correspond to the observed coordination numbers in the structure.

D/A	Atom	0	1	2
D	N1 of trans_amide	0,02	0,94	0,04
	O2 of aromatic_hydroxyl	0,02	0,93	0,05
A	O1 of trans_amide	0,25	0,70	0,05
	O2 of aromatic_hydroxyl	0,79	0,20	0,01

The coordination number is not optimal for the hydroxyl group but the global coordination score reaches its maximum (Figure 13). Indeed, in Paracetamol, both acceptor groups have to accept in order to satisfy the two donors. This suggests the use of coformers having at least one acceptor atom that could satisfy the amide donor of Paracetamol and leave the

[h] The same hydrogen bonds are present in the three polymoprhs of Paracetamol; their difference rather lying in minor conformations changes and overall packing modifications.

phenol group donor unoccupied. We are thus looking for a coformer that owes an acceptor moiety able to compete with the aromatic hydroxyl group. In order to find such a functional group, a motifs search is performed.

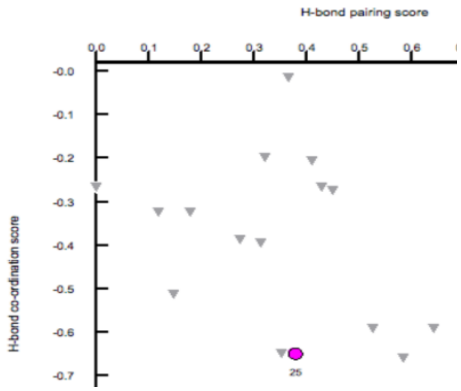
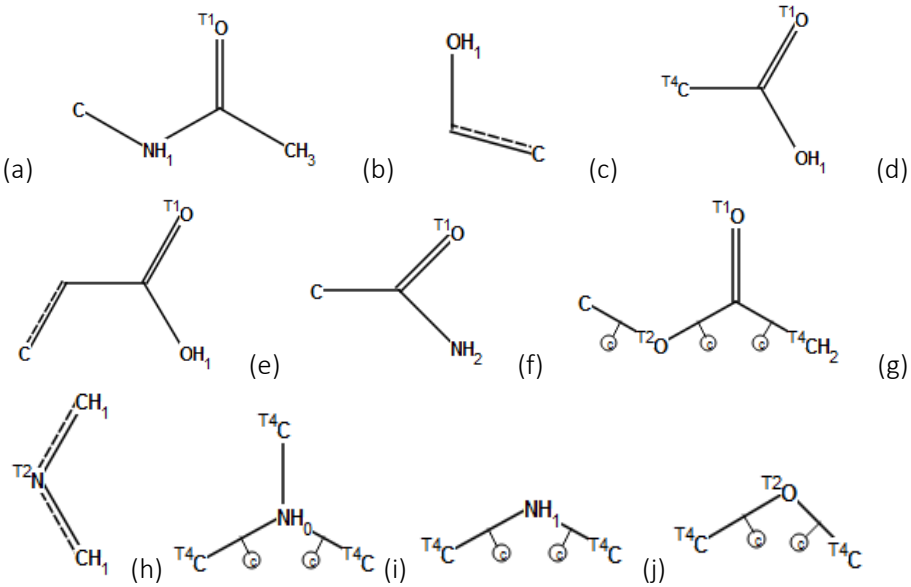


Figure 13. Chart displaying all the putative combinations of hydrogen bonds in polymorph I of Para (HXACAN 33). The horizontal and vertical axes refer to the HB pairing scores and the HB coordination scores respectively.

4.4.B.1 Motifs search

Searches have been performed to identify the frequency of occurrence of motifs involving either a secondary amide or a phenol group and selected functional groups (Scheme 3) in the CSD.



Scheme 3. Functional groups from the library used for the motifs searches: (a) Acetylamino_1, (b) ar_hydroxy (c) al_cooh_1, (d) ar_cooh_1, (e) carbamoyl, (f) al_cyclic_ester, (g) aromatic nitrogen, (h) al_cyclic_NC, (i) al_cyclic_NCH, (j) al_cyclic_ether. Tx super-script labels on an atom indicate that it is covalently coordinated to x atoms; while "c" circled labels on bonds indicate that these bonds are part of a cycle.

The analysis (Table 16) reveals that:
- Carboxylic acids and carbamoyls could compete with the two acceptor groups of Paracetamol for the N-H donor group.
- Aromatic nitrogen groups are the only competitors for the O-H donor group among the acceptors tested.
- Both aliphatic carboxylic acids and carbamoyls may compete with the secondary amide as donor for the carbonyl group.

Table 16. CSD frequency of occurrence statistics for intermolecular interactions involving the functional groups of Paracetamol and selected functionalities. Bold numbers correspond to competitive heterointeractions.

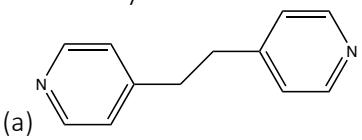
D	A	Freq	# struct	Total # struct
NH of acetylamino_1	O of acetylamino_1	37%	621	1665
	OH of ar_hydroxy	25%	27	110
	O al_cooh_1	**60%**	59	99
	O ar_cooh_1	**42%**	13	31
	O of carbamoyl	**40%**	21	52
	O of al_cyclic_ester	0%	0	2
	N aromatic nitrogen	25%	6	24
	NH0 of al_cyclic_NC	0%	0	36
	NH1 of al_cyclic_NH	0%	0	13
	O of al_cyclic_ether	3%	6	202

D	A	Freq	# struct	Total # struct
NH of acetylamino_1	O of acetylamino_1	37%	621	1665
OH of ar_hydroxy		53%	58	110
OH of al_cooh_1		**56%**	55	99
OH of ar_cooh_1		39%	12	31
NH2 of carbamoyl		**54%**	28	52
NH1 of al_cyclic_NH		23%	3	13

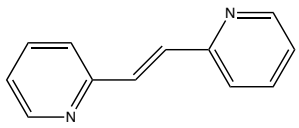
Table 16. Continued

D	A	Freq	# struct	Total # struct
OH of ar_hydroxy	O of ar_hydroxy	13%	2054	16121
	O of acetylamino_1	53%	58	110
	O al_cooh_1	28%	62	225
	O ar_cooh_1	16%	111	705
	O of carbamoyl	28%	55	196
	O of al_cyclic_ester	48%	32	67
	N aromatic nitrogen	**65%**	410	633
	NH0 of al_cyclic_NC	17%	130	757
	NH1 of al_cyclic_NH	15%	60	401
	O of al_cyclic_ether	19%	161	840

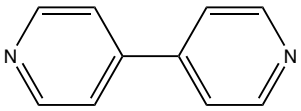
Comparing *a posteriori* the results of this motifs search with the motifs actually formed by Paracetamol with various coformers (Scheme 4) in its cocrystals (registered in the CSD with no disorder), we notice a strong consistency.



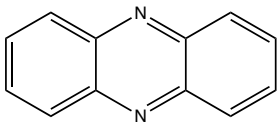
(a)



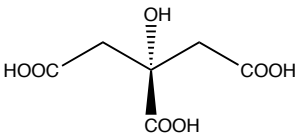
(b)



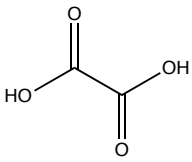
(c)



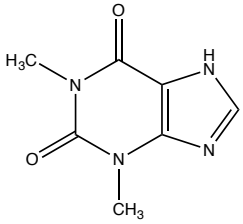
(d)



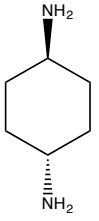
(e)



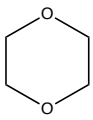
(f)



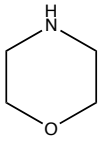
(g)



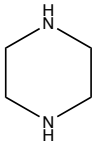
(h)



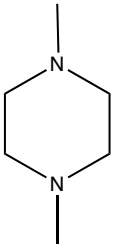
(i)



(j)



(l)



(m)

Scheme 4. Molecular diagrams of coformers cocrystallizing with paracetamol:[13,17,18,39,40] (a) 1,2-bis-4-pyridyl-ethane, (b) 1,4-di-4-pyridyl-ethylene, (c) 4,4-bipyridine, (d) phenazin, (e) citric acid, (f) oxalic acid, (g) theophyllin (h) trans-1,4-diaminocyclohexane, (i) 1,4-dioxane, (j) morpholine, (k) piperazine, (l) N,N-dimethyl-piperazine.

First, there are 4 cocrystallizing coformers that possess an aromatic nitrogen acceptor (Scheme 4, a-d) forming 3 cocrystals (MUPQAP, WIGBUL & LUJSOZ) and 4 cocrystal solvates (KETYUF, KETZAM, KETZEQ & WIGCAS). This moiety is found to accept the OH group of Paracetamol as predicted, but also the secondary amide hydrogen in some cases. In fact, it has been shown by structural and spectroscopic data that the O-H...N interaction formed in some of these cocrystals (in WIGBUL, WIGCAS, KETYUF and KETZAM) is of comparable strength to the strong OH...O=C bond existing in pure paracetamol.[17]

Second, we notice that two coformers own carboxylic acid functions (Scheme 4, e and f), which is in concordance with carboxylic acid ability to be a good donor and acceptor for all functional groups of Paracetamol. Hence, in the cocrystal formed with oxalic acid (LUJTAM, Figure 14), the carboxylic acid accepts both donors and donates to both acceptors. In the citric acid cocrystal (AMUBAM) however, carboxylic acid molecules preferably form homosynthons and the only interaction with Paracetamol consists in donating to the OH group.

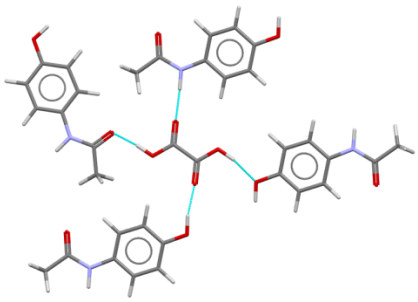


Figure 14. Interactions formed by oxalic acid in Para-OXA cocrystal (LUJTAM).

Third, amines and ethers (Scheme 4, h-n) were found to accept the hydrogen bond donated by the hydroxyl group in a lot of cocrystals/solvates of cocrystals.

Fourth, we observe that in most cocrystals of Paracetamol, there remain homo-interactions. Indeed, the secondary amide of Paracetamol is one of the best acceptors for both donors of other Paracetamol molecules and thus

occupies this role in most cocrystals. This confirms that one competitive hetero-interaction is sufficient for a cocrystal to exist. Consequently, motif searches should in priority target the weakest homo-interaction or functional groups not involved in hydrogen bonding in the original API.[41]

Finally, according to table 17, it appears that in the 12 cocrystals of Paracetamol studied here, coformers are more competitive as acceptors. Indeed heteromeric interactions outnumber homointeractions for both donors of Paracetamol in its cocrystals while the contrary prevails for donation to the amide. Besides, Paracetamol hydroxyl group is unoccupied in half of the cocrystals structures, which confirms the results of the coordination number analysis that suggested a poor acceptor ability of the phenol group.

Table 17. Number of homo- and hetero-interactions formed by the functional groups of Paracetamol in its cocrystals in comparison with the total number of such structures in which these functional groups interact.

		No. of cocrystals with		
	Functional groups	Homo interaction	Hetero Interaction	Total with interactions [a]
D	NH amide II	6	**7**	12
	OH ar_hydroxy	4	**9**	12
A	O of acetylamino_1	**9**	3	10
	O ar_hydroxy	3	3	6

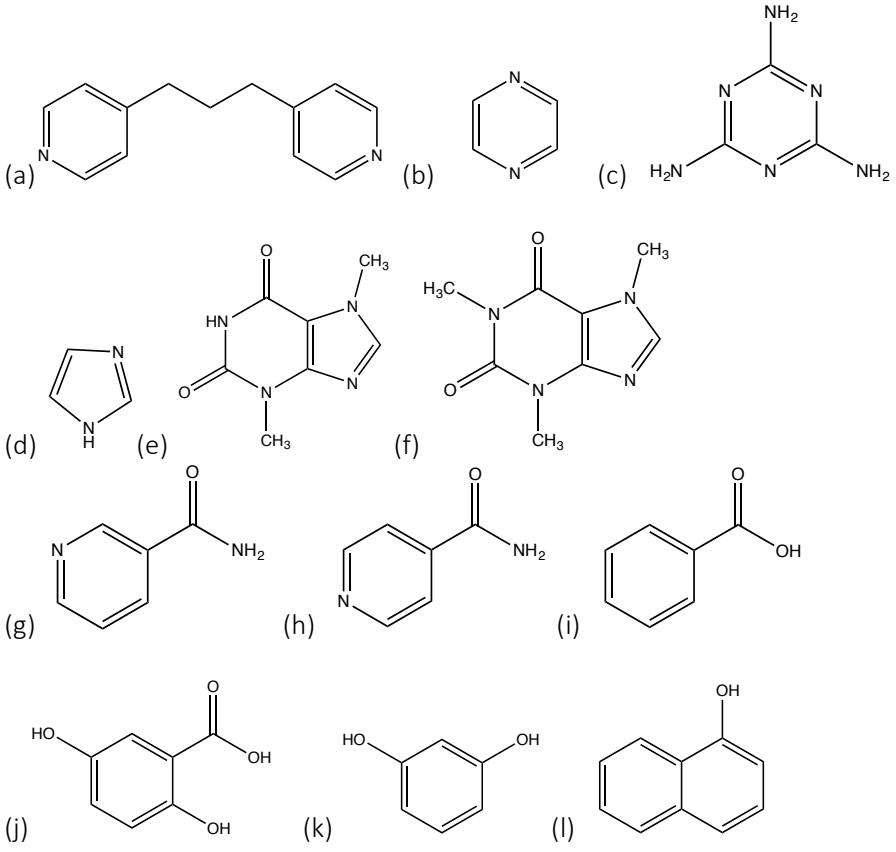
[a] The total number of cocrystals displaying the corresponding interactions is not necessarily the sum of homo- and hetero- interactions due to the occurrence of bifurcation and multiple types of H-bonding in some cocrystals and the absence of interactions for some moieties in others.

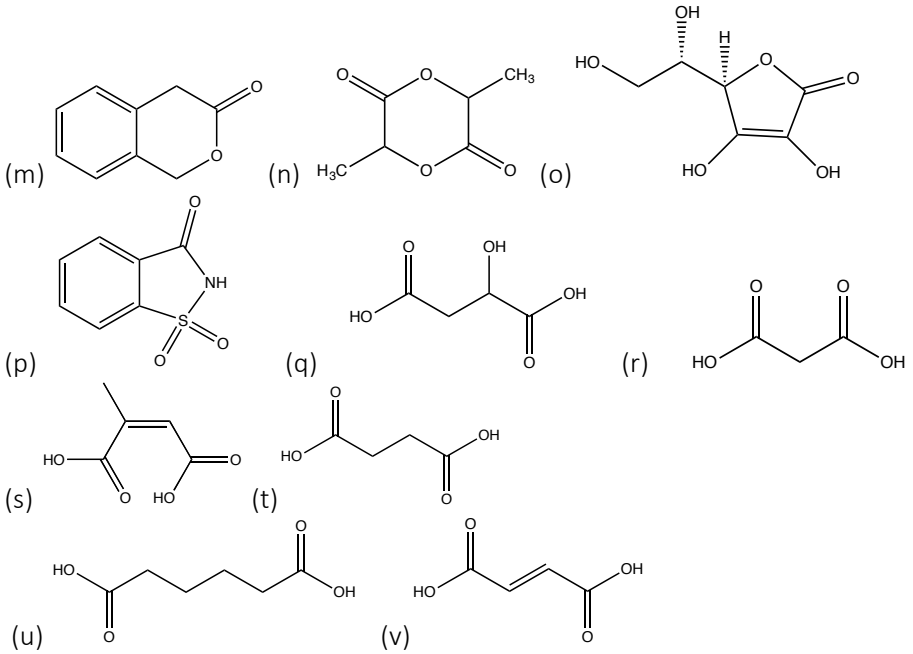
Moreover, it seems that coformers are more prone to accept the hydroxyl donor than the NH group. This may be due to the fact that the phenol group is the best donor and, as Etter stated,[42] prefers to H-bond to the best acceptor, which often comes from the coformer, as the presence of a stronger heterointeraction is part of the driving force leading to cocrystallization.

Motifs searches are thus less helpful *a priori* for Paracetamol cocrystal screening. Indeed, unlike the case of Levetiracetam in which the motif search suggests only one functional group, three different functional groups (carboxylic acids, carbamoyls and aromatic nitrogen atoms) are identified as competitors for at least one of the Paracetamol functional groups. Besides,

other functionalities (aliphatic cyclic amine/ether) are shown to form viable cocrystals and thus represent false negatives in this case. However, this is coherent with these groups being better acceptors than the hydroxy group for the H-bond donated by the OH group of another Paracetamol molecule.

In fact, a similar motifs search was conducted by Srirambhatla *et al.* for the potential interactions of the phenolic group of Paracetamol and they chose coformer candidates accordingly (3 cocrystallizing and 17 non-cocrystallizing compounds included in Scheme 5).[17] Those coformers that failed to cocrystallize with Paracetamol therefore display the same functional groups than those found on the successful coformers (amines, carboxylic acids, ethers and amide). However, further analyses could have helped to reduce the set of coformers to test (see section 4.4.B.4).





Scheme 5. Molecular diagrams of coformers non-cocrystallizing with paracetamol: (a) 4,4-trimethylene-dipyridine, (b) Pyrazine, (c) Melamine, (d) Imidazole, (e) Theobromin, (f) Caffeine, (g) nicotinamide, (h) isonicotinamide, (i) Benzoic acid, (j) 2,5-dihydroxybenzoic acid, (k) Resorcinol, (l) 1-naphthol, (m) 3-isochromanone, (n) Lactide, (o) Ascorbic acid, (p) Saccharin, (q) malic acid, (r) malonic acid, (s) maleic acid, (t) succinic acid, (u) adipic acid, (v) fumaric acid.

Note that certain cocrystallizing coformers (dioxane, dimethylpiperazine) are liquid at room temperature. This is in contradiction with one commonly accepted definition of cocrystals stating that all components must be solid at RT.[43] But this in accordance with the classification of pharmaceutical cocrystals by the FDA, which rather insists on their neutrality and on the nonionic character of their interactions.[44]

Finally, it is worth mentioning that a motif search comparing the potential homo-interactions in Paracetamol to the hetero-interaction formed with common solvents is able to predict the occurrence of solvates of cocrystals of Paracetamol (Table 18).

Table 18. CSD frequency of occurrence statistics for intermolecular interactions involving the functional groups of paracetamol and common solvents. Bold numbers correspond to competitive hetero-interactions.

D	A	Freq	# struct	Total # struct
NH1 of acetylamino_1	O of acetylamino_1	37%	621	1665
	OH ar_hydroxy	25%	27	110
	OH2 of water	**44,1%**	93	211
	OH1 of methanol	**40,0%**	10	25
	O of ethanol	**50,0%**	7	14
	N of acetonitrile	15,4%	2	13
	O of acetone	0,0%	0	6

D	A	Freq	# struct	Total # struct
NH amide II	O of acetylamino_1	37%	621	1665
OH ar_hydroxy		53%	58	110
OH2 of water		**67,3%**	142	211
OH1 of methanol		40%	10	25
OH of ethanol		42,9%	6	14

D	A	Freq	# struct	Total # struct
OH1 of ar_hydroxy	O ar_hydroxy	13%	2054	16121
	O of acetylamino_1	53%	58	110
	O of water	44,2%	1165	2633
	O of methanol	40,2%	264	657
	O of ethanol	44,0%	102	232
	N of acetonitrile	22.8%	57	250
	O of acetone	48,8%	79	162

D	A	Freq	# struct	Total # struct
NH amide II	O ar_hydroxy	25%	27	110
OH ar_hydroxy		13%	2054	16121
OH2 of water		**35,5%**	935	2633
OH1 of methanol		18,0%	118	657
OH of ethanol		23,3%	54	232

We indeed observe that, in each of the five solvates of cocrystals, at least one interaction between Paracetamol and solvent molecules (Table 19) is

characterized by a larger probability than the one associated to the potential homo-interactions.[i]

Table 19. Interactions involving solvent molecules in Para cocrystals

Coformers	Cocrystals	D	A
Piperazine	COKCEL	NH1 of acetylamino_1	O of ethanol
		OH of ethanol	O ar_hydroxy
1,4-di-4-pyridyl-ethylene	KETZAM	NH1 of acetylamino_1	OH2 of water
		OH2 of water	O of acetylamino_1
		OH2 of water	O of water
	KETYIUF	NH1 of acetylamino_1	O of ethanol
		OH of ethanol	O of acetylamino_1
	WIGCAS	NH1 of acetylamino_1	O of methanol
		OH1 of methanol	O of acetylamino_1
	KETZEQ	NH1 of acetylamino_1	O of water
		OH1 of ar_hydroxy	O of water
		OH2 of water	O of acetylamino_1
		OH2 of water	O of ar_hydroxy
		OH2 of water	N coformer

4.4.B.2 Molecular complementarity tests

Using the conclusions of Fabian,[31] we expect that Paracetamol, which is planar (S<5 and S/L <0,6), will preferably cocrystallize with (almost) planar coformers so as to form layered structures, as observed for cocrystals of Diclofenac and Niclosamide.[45,46] All the coformers forming a cocrystal with Paracetamol look planar, except for citric acid.

However, the molecular complementarity test indicates disparity of shape between paracetamol and 3 of the cocrystallizing coformers while citric acid passes all the tests. In fact, while being planar, 1,4-dioxane (DIOX), morpholine (MORPH) and piperazine (PIP) are globally classified as cubic while Paracetamol has a rod-like resemblance (Table 20). Hence they would

[i] Concerning the presence of hydrated forms for certain cocrystals, a common explanation is related to an imbalance between the number of D and A atoms in the potential unsolvated cocrystals:[15,40,45,62] hydration could even out it by increasing the number of donor atoms while ensuring packing cohesion. This seems reasonable here as incorporation of water molecules occurs with 1,4-di-4-pyridyl-ethylene owning only acceptor atoms such that there are only two donors for four acceptors in the 1:1 adduct.

have been rejected, along with oxalic acid (OXA) which is judged too polar in comparison with Paracetamol (Table 21). However, these differences do not prevent close packing of the corresponding cocrystals; suggesting that another factor is more important in this case.

Table 20. Axis lengths of the molecular enclosing box, corresponding ratios and associated shapes for Paracetamol and the three cocrystallizing coformers rejected for shape dissimilarity.

API/coformers	L (Å)	M (Å)	S (Å)	Ratio	Shape
PARA	11,516	6,967	4,238	2:1:1	Rod
DIOX	5,815	5,815	5,023	1:1:1	Cube
PIP	6,621	6,621	5,033	1:1:1	Cube
MORPH	6,475	6,475	5,034	1:1:1	Cube

Table 21. Results of the molecular complementarity analysis for Paracetamol and the five cocrystallizing coformers failing to at least one of the tests. See section 4.2.D for definitions of descriptor abbreviations.

Coformers	M/L	Results [a]	S axis (Å)	Results [b]	S/L	Results [c]
PARA	0,605		4,238		0,368	
DIOX	0,881	v	5,023	v	0,761	**x**
PIP	0,934	**x**	5,033	v	0,71	**x**
MORPH	0,984	**x**	5,034	v	0,765	**x**
OXA	0,764	v	3,406	v	0,476	v

Coformers	Dipole (Debye)	Results [d]	FNO	Results [e]
PARA	3,744		0,273	
DIOX	0	v	0,333	v
PIP	0,002	v	0,333	v
MORPH	2,006	v	0,333	v
OXA	0,002	v	0,667	**x**

a: pass if delta < 0.31, b: pass if delta < 3.23; c: pass < 0.275; d: pass if delta < 5.94; e: pass if delta < 0.294. Passing mark v, failing mark x.

Moreover, the proportion of false positive results is very high (Table 22), so that this tool does not help to significantly reduce the number of experiments to carry out. The fact that all tested coformers are planar (S <6 angstrom) may reveal an a priori awareness of the importance of this

criterion when coformers were selected and this thus undermines the utility of the corresponding test performed here.

Table 22. Confusion matrix showing the results of the tests of molecular complementarity performed with the 5 usual descriptors for Para and 33 potential coformers (all likely conformations included)

Molecular complementarity tests		Predicted Outcomes	
		CC	No CC
Actual Outcomes	CC	8	4
	No CC	18	3

4.4.B.3 Hydrogen-bond propensities

This analysis has been previously realized by Wood *et al.* for cocrystals of Paracetamol (Scheme 4) and non-cocrystallization coformers (Scheme 5 and Table 24).[8] They concluded that the proportion of cocrystallizing coformers with high MC Scores was higher than the one with low MC Scores (Figure 15).

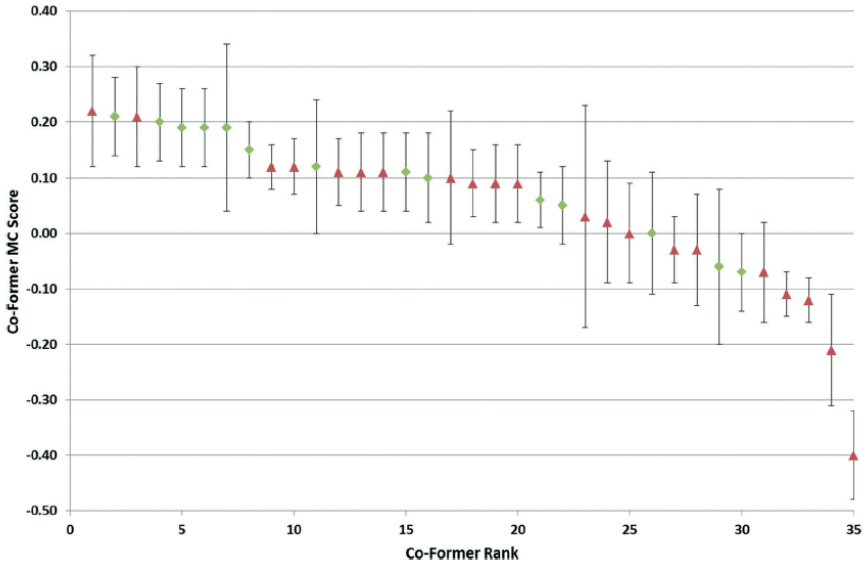


Figure 15. MC Score computed for cocrystallizing coformers (green diamonds) and non-cocrystallizing coformers (red triangles) ranked by their MC Scores.[8]

But there are still 14 non-cocrystallizing coformers characterized by a positive MC Score (False positives), which is more than the number of true positives (Table 23). Note in passing that they classify the coformer 4,4-trimethylene-dipyridine (Scheme 5 a; rank 2 on Figure 14) as successful as it formed an amorphous phase with Para, while we considered it as non-cocrystallizing due to the absence of corresponding single crystal.

Table 23. Confusion matrix showing the results of *a posteriori* HBP analyses performed on the Para-coformer system to reveal cocrystallization ability.

HBP analyses		Predicted Outcomes	
		CC	No CC
Actual Outcomes	CC	12	2
	No CC	14	7

The percentage of successful experiments is thus around 46% (12/(12+14)). This is quite high in comparison with the results of random screening of coformers.[19,47] But coformers tested here were not chosen randomly either: some were selected based on the knowledge of other cocrystals of paracetamol[13,40] and/or following a CSD search.[17,40] Besides, as for the Leviteracetam screening, coformers with similar functional groups led to MC Scores of comparable magnitude, regardless of their ability to cocrystallize (Table 24), such that HBP calculations do not give any additional information in comparison with motif searches. We would also like to highlight that the functional group ranking in Table 24 is in line with the results obtained when performing a motif search (Table 14).

Table 24. Ranking of Paracetamol coformers according to their MC Scores, along with their functional groups.

Rank	**Coformer**	**Functional groups**
1	non cocr f	aromatic N, amide and amine III
2	non cocr a	aromatic N
3	non cocr b	aromatic N
4	cocr a	aromatic N
5	cocr c	aromatic N
6	cocr b	aromatic N
7	cocr d	aromatic N
8	cocr e	carboxylic acid & OH
9	non cocr	carboxylic acid
10	non cocr p	carboxylic acid

11	cocr k	amine III & ether
12	non cocr s	carboxylic acid
13	non cocr j	arom carboxylic acid & OH
14	non cocr i	arom carboxylic acid
15	cocr l	amines II
16	cocr h	amines I
17	non cocr e	amine III, amide and N arom
18	non cocr r	carboxylic acid
19	non cocr t	carboxylic acid
20	non cocr q	carboxylic acid
21	cocr j	amine II & ether
22	cocr m	amines III

4.4.B.4 Packing Feature Searches

Packing features searches were performed, starting from known cocrystals of Paracetamol, in order to investigate if some aspects of molecular packing could help to virtually screen coformers.

A posteriori, we observe different types of packing/coordination in the various cocrystals: [13,15,17,18,40]
- The coformer possesses two acceptors that interact with Paracetamol hydroxyl groups, while chains of Paracetamol molecules are formed through their secondary amides (Figure 16);
- The coformer has two acceptors, one that interacts with the hydroxyl group and another one with a secondary amide hydrogen (Figure 17);
- The coformer accepts two hydroxyl groups but also contains donors that interact with the hydroxyl group on other Paracetamol molecules (Figure 18).
...

However, one can notice that all coformers cocrystallizing with Paracetamol possess at least two acceptor atoms with their lone pairs pointing in opposite direction and interacting with two different Paracetamol molecules. This is in agreement with the hypothesis of Oswald that 1,3,5-trioxane did not cocrystallize due to its lack of inversion symmetry[13] and the results of Srirambhatla that emphasis the need for "*difunctional coformers, able to form chain-like structural motifs through distinct HB for cocrystal formation with paracetamol*".[17] This feature could thus be selected for subsequent packing searches (with nitrogen and oxygen atoms allowed in position of the acceptors).

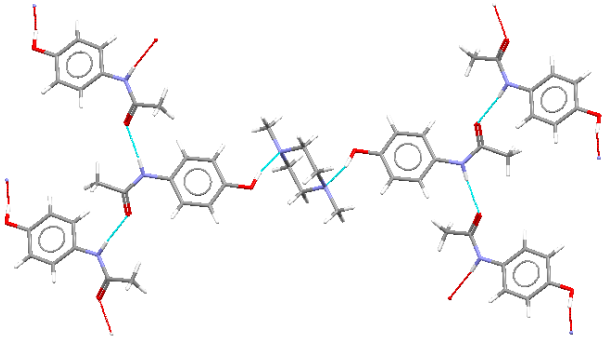


Figure 16. H-bonding in Para - N,N-dimethyl-piperazine (MUPPIW)

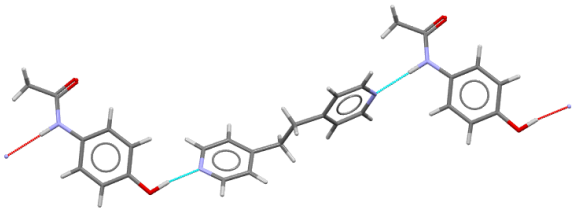


Figure 17. H-bonding in Para - 4,4'-ethane-1,2-diyldipyridine cocrystal (WIGBUL)

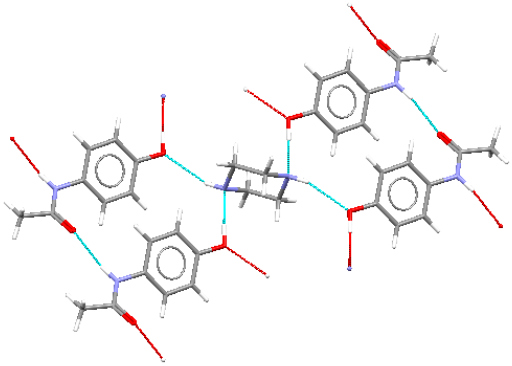
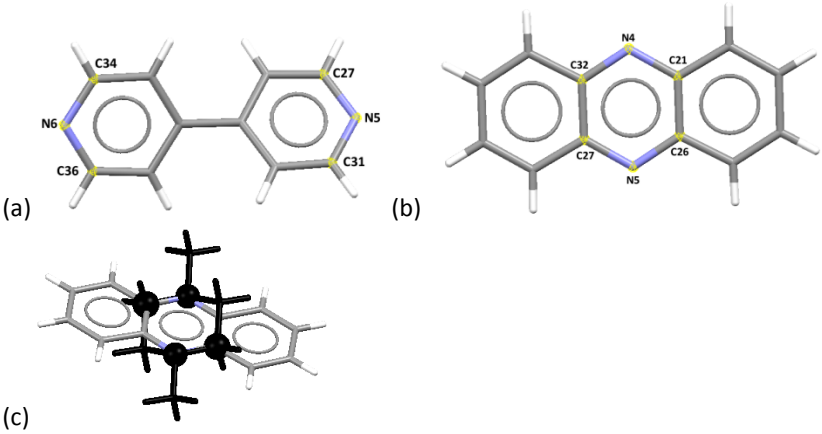


Figure 18. H-bonding in Para -piperazine cocrystal (MUPPUI)

In practice, care was paid to suitably select packing features and search criteria in order to embrace all the coformers that have the matching features but display different packing types. This was achieved by selecting in addition to the two acceptors, their adjacent carbon atoms to ensure that the mutual direction of acceptor lone pairs is adequate (Figure 19, a). Note that in the case of phenazine, this implies the selection of the whole inner cycle such that no angular variation is allowed between the positions of the two acceptors (Figure 19, b). 33 coformers used by Wood for HBP analysis (see section 0) were selected as potential candidates. One compound was

excluded because it is only known in a disordered structure and there is no structure in the CSD corresponding to the other dismissed coformer.



(a) (b)

(c)

Figure 19. Atoms selected (labeled and highlighted in yellow) for the packing feature search in the Para- 4,4-bipyridine cocrystal (MUPQAP) and in the Para-phenazine cocrystal (LUJSOZ) (c) Superimposition of phenazine (from its Para cocrystal, LUSJOZ) and N,N-dimethyl-piperazine (from its pure form, UGUHAF) molecules showing the matching positions of their acceptors.

Doing so, it appears that most coformers could be sorted in two groups according to the distances separating the 2 acceptors atoms and that performing a packing feature search on one coformer of each group (phenazine for short distances and 4,4-bipyridine for longer distances, Figure 18 b and a respectively) allows to identify all the other group members (Table 25 and example Figure 18 c).

Table 25. CSD refcodes of matching structures (pure forms and corresponding cocrystals) for the packing feature search starting from the Para- 4,4-bipyridine cocrystal (MUPQAP) and from the Para- phenazine cocrystal (LUJSOZ). Underline refcodes correspond to the pure forms of the coformers used for the search in the mother structures.

Coformers	Pure forms	RMSD	(Solvates of) Cocrystals
phenazine	WOQBIN	0.013	LUJSOZ
pyrazine	QAMQUR	0.026	-
1,4-dioxane	CUKCIU	0.228	MUPPES02
N,N-dimethyl-piperazine	UGUHAF	0.237	MUPPIW
morpholine	ITIZUG	0.243	MUPQET, AHEPUY
Lactide	BICVIS	0.248	-
piperazine	ITIZOA01	0.253	MUPPUI, COKCEL

Table 25. Continued

Coformers	Pure forms	RMSD	(Solvates of) Cocrystals
4,4-bipyridine	HIQWEJ02	0.029	MUPQAP
1,2-bis-4-pyridyl-ethane	ZEXKIW	1.103	WIGBUL
1,4-di-4-pyridyl-ethylene	AZSTBB	1.173	KETZAM, KETZEQ, KETYUF, WIGCAS

In fact, the distances are quasi identical in the short-distance group and applying medium geometric tolerance worked fine. But for the longer-distance group, distances are more heterogeneous and we needed to authorize a 99% distance tolerance to find the two corresponding coformers as it was not possible to directly remove the distance constrain. Yet it is the mutual orientation of acceptors that prevails in Paracetamol coformers. In fact, similarities of different scales are often encountered and the possibility of performing packing searches without the distance match criterion could constitute a further improvement of the wizard, as noticed by Galek.[48]

We note in passing that the pure forms of the two coformers used for the selection of packing features were found to match the query started from their corresponding cocrystal (Table 25), contrary to what was observed when performing this analysis for Levetiracetam (section 4.4.A.4). This is expected, as both coformers are rigid such that their conformation does not change from one structure to another.

These results emphasize that, for a packing feature search to work, only a section of the crystal structure/coordination shell around a molecule has to be similar between structures and not the entire packing. Indeed, in this case, some coformers were found to match another coformer features even if they possess donor atoms not present on the query molecule and that form additional interactions with Paracetamol (example: piperazine, Figure 18).

This technique also enables to find compatible coformers from the point of view of the localization of polar/apolar zones and close-packing, even when their H-bonding partners do not form any directional interaction (Figure 20).

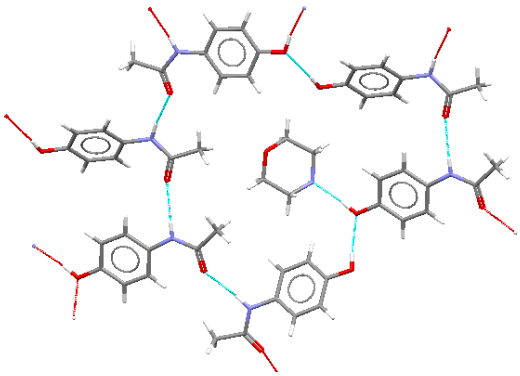


Figure 20. Para- morpholine cocrystal (MUPQET) packing. The ether group of the morpholine molecule is not involved in any H-bond but it is embedded in a hydrophobic cavity to allow close-packing of the coformers.

However, four cocrystals were missed by these two searches. Among them are the two cocrystals of carboxylic acids (AMUBAM & LUJTAM) mentioned in the motifs search section. Their specific packings are due to the unusual concentration of functionalities on the coformer molecules: citric acid owns a lot of functional groups oriented in various ways, while oxalic acid has two adjacent carboxylic acids. This allows them to cocrystallize with very different partners (for instance oxalic acid with caffeine[49] and citric acid with nitrofurantoin[50]), but it seems unlikely that other molecules will adopt a similar packing to theirs.

The third exception to the packing searches is realized by the theophylline cocrystal (KIGLUI). Interactions with paracetamol are in this case also due to two acceptors except that these are originally not in opposite directions. To align them in such directions, theophylline molecules have first to form a centrosymmetric carboxamide dimer (Figure 21). This explains why caffeine (Scheme 5 e) does not form a cocrystal with Paracetamol. This behaviour cannot be predicted by the packing search method and so the existence of this cocrystal could not be identified. However an analysis of theophylline in other cocrystals could have shown the same $R^2_2(10)$ ring motif in different instances (DUCROJ,[51] HEBFEB,[46] WOCHIH and WOCHON[52]).

The last unidentified coformer is trans-1,4-diaminocyclohexane (Scheme 4 h). In this cocrystal (WIGCEW), the lone pairs of the two acceptors also point in opposed directions but acceptors atoms are not part of a cycle and are bound to a single carbon atom. Besides, the two acceptor-carbon segments are collinear while they are only parallel in the other successful coformers. Hence this coformer does not possess the packing patterns defined in the

two queries even if it forms a similar packing. This confirms the importance of the selection step and the need to sometimes perform multiple searches in parallel and with different criteria to obtain complementary results.

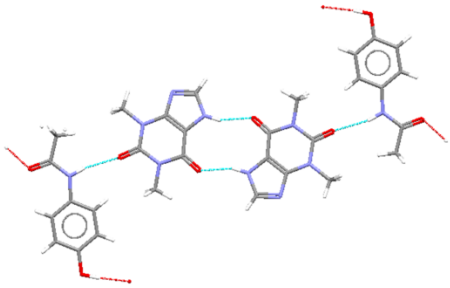


Figure 21. Theophyllin molecules forming a homo-synthon in Para cocrystal (KIGLUI) in order to present acceptor atoms in adequate position for packing.

To sum up, packing features searches started from two known cocrystals of Paracetamol allowed to identify six other cocrystals by performing only 8 experiments (Table 26), for a 75% (6/8) efficiency. And among the two false positive results, it seems understandable that lactide (Schem 5 n) did not cocrystallize with Paracetamol as it possesses two additional acceptors not included as packing features, that could prevent the corresponding interactions of forming. However, the fact that pyrazine (Scheme 5 b) failed to cocrystallize seems doubtful as it displays the required features and nothing more. As liquid-assisted grinding does not provide a 100% success rate,[53] it might be interesting to try new crystallization experiments with this coformer.

Table 26. Confusion matrix showing the results of packing features searches started from two known cocrsystals of Para (MUPQAP & LUJSOZ) to identify coformer candidates for it.

Packing searches started from MUPQAP & LUJSOZ		Predicted Outcomes	
		CC	No CC
Actual Outcomes	CC	6	4
	No CC	2	21

4.5 Discussion

This section aims to identify the limits and discuss optimal conditions for the use of the knowledge-based tools investigated here to predict cocrystallization occurrence. Each tool will first be discussed separately and then, an overall methodology will be suggested.

4.5.A Motifs searches

For some APIs, motifs searches may highlight a lot of different functional groups that could promisingly interact with its moieties, as for the Paracetamol screening. In this case, cocrystallization is thus statistically more likely to occur. But as the number of potential candidates also increases, the need for other selection criteria becomes necessary. On the contrary, when only a few functionalities are competitive, there are less candidates to test but this also means that the API functional groups are already strong H-bond donors/acceptors and that it will be harder to substitute them.

For some pairwise interactions, there is a limited amount of structures in the CSD that contain the two functional groups simultaneously. In such cases, the calculated frequencies may vary significantly according to the size of the dataset and are thus not reliable. This is often encountered for more specific functional groups. Hence, a minimum of 20-30 occurrences is recommended for each motif frequency, for a valid analysis.

Such a low occurrence may reveal either that very few crystallization experiments were tested or, more interestingly, that these moieties do not have a strong incentive to crystallize together, irrespective of their frequency in the dataset. Once more, this highlights the shortcoming of not recording, in the CSD or elsewhere, all the unsuccessful cocrystallization experiments. In fact, it drastically limits the usefulness of the database for validation and prediction, as contrast between successful and unsuccessful trials cannot easily be revealed.

On top, among all the interactions recorded, there might be some identified close contacts that are in fact artifacts of stronger interactions and that introduce a bias in the analysis, as no angle criterion is defined for searches involving newly created motifs.

Contrary to H-bond propensity models, motifs searches do not directly take into account competition between H-bond donors and acceptors, as there is

no indication concerning their number and nature. It would be interesting, albeit time-consuming, to calculate these frequencies in presence of competition (traditional search) and without it, by analyzing a filtered set of structures. This was done for the first time by Shattock *et al.* in their study of carboxylic acid and alcohols homo-/hetero-synthons hierarchy in presence of competing functionalities.[54]They qualify the traditional search as "raw" and the second as "refined". Srirambhatla *et al.* then applied their methodology to the analysis of competing interactions involving the phenolic group of Paracetamol.[17] As expected, refining the data often causes an increase in the percentage of heterosynthons. And for almost all functional groups studied, the conclusions were identical using the two sets of data. But in case of alcohol in presence of primary amide, the frequency of the amide homosynthon (55%) was bigger than the one of related heterosynthons (32 and 24%) in the raw data set while it was the contrary with the refined dataset (32% vs 76% and 70%).; justifying the additional treatment. Of course, this is doable only if there is a sufficient amount of structures in the unfiltered dataset.

Similarly, when the corresponding dataset is large enough, it is worth performing a given motif search on a structural subset containing, in addition to the involved functionalities, the moiety that has to be out-competed, to judge the relative strength of the competing groups. This 3-members structures set could be generated prior the motif search by formulating the appropriate query in Conquest.[37]

4.5.B Molecular complementarity tests

Shape descriptors depend on the conformation of the molecules (API and coformer) in the structures selected for analysis. To limit this dependence, one can a priori use the conformer generator included in Mercury, to isolate the most likely conformations of each flexible compound.

Application of the molecular complementarity test to the Leviteracetam and Paracetamol screenings revealed the importance of not using the molecular complementarity module blindly. In particular, one should keep in mind the particularities of the API and the conditions and limitations of the different tests. This allows to discard a test if required.

These restrictions are in part due to the fact that the descriptors emanate from empiric observations that represent only the majority of cases and that we are still unable to fully rationalize the underlying behaviour.[33]

For example and contrary to the main trend, dissimilarity of coformer shapes may favor cocrystallization of compounds with non-self complementary shapes. In that case, the shape tests are rather counterproductive. The difficulty here lies in being able to tell when one test is helpful and when not.

According to the results of the two cocrystal screenings investigated here, it seems wiser to limit any analysis to the S-test for "absolute" planarity (in opposition to the "relative" planarity depicted by the S/L ratio, which does not look that pertinent) in case of planar API and to the M/L test for API with plate or cubic shape.

Tests about polarity are indeed not recommended for several reasons. The first one is related to the likely explanation for the likeness of polarity descriptors of cocrystallizing molecules; namely the inability of less dipolar molecules to compete with the strong homo-interactions taking place in strong dipolar molecules.[31,33] But in applying polarity tests, one risks to miss molecules with large local but small global dipoles, as it was the case for oxalic acid in both the Leviteracetam and Paracetamol cocrystal screenings. Besides, motifs searches are more efficient and especially dedicated to discard uncompetitive functional groups. And one can easily remove charged species with strong dipoles when retrieving promising coformers from the CSD.

The second reason is that molecules with very distinct values at the polarity descriptor may still generate a favourable segregation of hydrophobic and hydrophilic regions in their cocrystal. Indeed, the two polarity tests compare overall polarity scores rather than local values that could better depict the actual topology of the molecules. And for that matter, there exists another module, the packing features searches, that is more appropriate to identify matching topologies when some related cocrystals are known.

Finally, note that macthing of coformer solubilities does not seem to be related to the preferred polarity similarity as there exist descriptors directly associated with it (Log P, polar surface area) that show no correlation.[31]

Concerning the importance of shape in cocrystal design, Fabian & Friscic showed that similarity of shape cannot counterbalance the effect of good homomolecular synthons and that, reciprocally, heteromolecular synthons cannot offset the negative impact of shape mismatch, except when involving very strong acceptors/donors or ionizable groups. In these latter case, “the stronger the heterosynthon, the larger the shape difference that can be overcomed”.[33]

This explains why shape did not prove influential in Leviteracetam cocrystals and why the opposite was true for Paracetamol. Indeed, all the coformers cocrystallizing with Levi have carboxylic acid moieties (prone to form strong heteromolecular synthons) while often being of different shape. On the contrary, motifs searches did not identify outstanding interactions for Paracetamol functional groups, but almost all its cocrystallizing coformers are planar.

4.5.C HBP & CL models

Some remarks can be expressed about the ability of these models to characterize the probability of a given arrangement of H-bonds to occur.

First, HBP and CN models are said to be usable for both observed structures and predictions but there is one major limitation to their use in the latter case, as both models depend on the conformation of the target molecule.

Hence, for a given system, the results of a model built starting from the existing structure and the results of one model obtained from the sketched 2D representation can be quite different as the conformation automatically assigned by the software does not necessarily correspond to the actual conformation (and there is no way to influence it). This explains why the coordination likelihoods are slightly different for the Levi H-bonding atoms in its pure form and in its cocrystals: its conformation varies between these structures.

Moreover, the coordination likelihoods computed for the target molecule cannot be used for all the putative structures resulting from different H-bonding patterns, as it is the case in the HBP-CN chart, as the conformation of the molecule may vary.

Similarly, the algorithm used to build the training set for H-bond propensities takes into account the conformation of the molecules so that molecules chosen in the dataset are as similar as possible to the target. Consequently, the dataset used for the model built on a structure and the model built on a 2D diagram for the same system will also differ and give distinct predictions. In other words, the utility of HBP-CN models for prediction is questionable. This is the reason why we did not take into account the information gathered from the HBP-NC models generated for the Levi-non cocrystallizing coformer systems in the analysis of section 4.4.A.3. Hopefully, this may not

be a problem anymore in the future as the CCDC confirmed their awareness of this limitation.

Besides, it is worth mentioning that these problems should affect rigid/planar molecules in a limited way and that such models can thus still be useful for prediction of this family of molecules. This has been verified by generating the HBP-NC model from the 2D diagram of the 2,4-dihydroxybenzoic acid planar molecule : the outcomes are very similar to those obtained for the corresponding model built on the observed structure.

From its application to Levetiracetam screening, it appears that HBP models do not give any additional information about coformer cocrystallization ability in comparison with motifs searches. They indeed determine all the potential interactions in a structure while the knowledge of one competitive interaction driving molecular recognition may be enough. These models may be used in replacement of the motifs searches to identify promising functional groups but motifs searches are more efficient as they concern whole categories of functional groups instead of isolated compounds. HBP models may however be useful to evaluate the likelihood of interactions under-represented in the CSD (between less frequent functional groups for example) and evaluate what is not (yet) observed.

The setback of motif searches is that they do not indicate how many moieties from each sort is required to reach optimal coordination in the structure. For this matter, coordination models are highly valuable as they go beyond simply counting donors and acceptors. In particular, they take into account the availability of each functional group, through the computation of 'accessible surfaces', to determine feasible interactions. The accessible surfaces parameter in this model allows to explain why strong H-bonding groups deviate from Etter's first rule[j] and do not form any hydrogen bond in certain structures due to steric hindrance, which account for 2/3 of H-bond-absent structures.[3] This is of great importance as the number of structures possessing a strong donor not involved in H-bonding amounts to 2,5% of the total CSD and up to 25% in certain pharmaceutical cocrystals families.[55]

Finally, note in passing that this example shows the limit of such models to evaluate polymorphism likelihood. Indeed, the fact that there is no equally stable putative structure on the solid-form landscape of Levetiracetam is a priori consistent with its absence of polymorphism. But a similar chart is

[j] « All good proton donors and acceptors are used in hydrogen bonding »[63]

obtained for its racemic equivalent, Etiracetam, which shows two polymorphic forms (refcodes OFIQUR & OFIQUR01). The less stable morph (OFIQUR01) displays a poor coordination score, which is in agreement with its lower stability (Figure 22). But, contrary to what is suggested by Galek *et al.*, it does not figure in the lower right corner of the chart, such that its existence could not have been predicted. And, as Cruz-Cabeza points out, this tool cannot predict either if packing polymorphism will occur.[56]

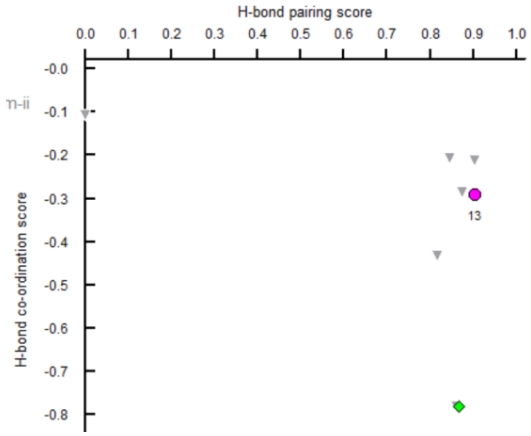


Figure 22. Etiracetam solid-state landscape. The less stable form is denoted by the pink circle while the green diamond corresponds to the thermodynamic form.

Overall, if motifs and coordination analyses help to rationally select coformer candidates according to their H-bonding ability, they give no indication concerning the packing viability of the potential cocrystal, while this latter is at least as important. Indeed, for some compounds, such as large flat or unusually shaped molecules, space filling is the primary concern and D-H…pi interactions and pi-stacking dominate over H-bonding (eg. indomethacin[57]),[58] and this explains the remaining third of the structures with unsatisfied H-bonding partners.[3] In such extreme cases, motifs and coordination analyses are less effective but molecular complementarity and packing feature searches may still be very useful.

4.5.D Packing feature searches

The presence of matching fragments on a coformer does not guarantee that it will actually cocrystallize with the API, as only a fraction of the structural features has been searched for and the rest of the molecule (steric

hindrance by adjacent functional groups, incompatible position of the polar/apolar regions…) may prevent it from happening.

Similarly the absence of a compound on the list of matching compounds does not imply that it won't form a cocrystal with the API under investigation. Indeed, the knowledge of some cocrystal features does not give information on the presence or absence of other feasible packing patterns.

For some APIs, their cocrystals display a lot of alternative efficient packing patterns. This is expected in particular for APIs showing conformational flexibility, as they are likely to adapt their conformation to improve their interaction with coformers, as it is the case for Levetiracetam. The scope of this kind of research is thus highly variable and cannot be predicted. However some conditions seem to favour success.

The first one logically concerns the structural flexibility of the host and guest molecules. Packing feature searches are more conclusive when the host molecule is planar. Indeed, the position and relative orientation of H-bond acceptors/donors are rarely similar for non-planar molecules as they have more degrees of freedom, and the probability of a match is thus smaller. This explains why the packing feature tool was more successful for Paracetamol, which is a planar and rigid API, in comparison with Levetiracetam, which is not. This tool is thus unfortunately less appropriate for chiral APIs, as chirality often implies flexibility.

Similarly, if a potential guest molecule is highly flexible, the conformation used in the structure included in the training dataset may be not representative and the coformer could be a false negative. To limit this dependence, one can a priori use the conformer generator included in Mercury to isolate the most likely conformations of each flexible compound.

Finally, it seems that the probability of effective hits is also increased when the API owns two donor atoms that are so close that they need to be satisfied by acceptors on the same molecule. This is the case when both donors are on the same cycle (see lamivudine case, Figure 23)[59]. Otherwise, the molecule will often adopt a conformation that places them in shifted positions. Having 2 acceptor atoms in strategic positions (at the two ends of the molecule for example) can be sufficient in some cases as well (cf Paracetamol screening).

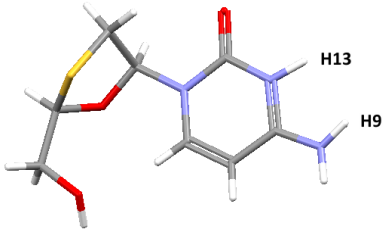


Figure 23. Lamivudine chemical structure showing the proximity of two hydrogen donors (H9 & H13).

Some points may also be discussed concerning the implementation steps. In the context of cocrystal screening, it is recommended to select H-bond donor/acceptor atoms without looking at the actual donated hydrogen atom. For each hit, it is also necessary to check that the matching atoms are not participating in any intramolecular contact and are indeed available for intermolecular interactions.

Another question to address concerns the choice of packing features to be selected as query. It is indeed not necessary to select all the functional groups available for H-bonding on the guest molecule. In fact, the optimal search settings are very dependent on the system studied and may be found by progressive variations.[48] However, it seems reasonable to think that more likely outcomes will be obtained using best donor/acceptor atoms or functional groups identified as best competitors by a previous motifs search. It is also often required to select atoms on the carbon backbone of the mother molecule to ensure proper orientations of the donor/acceptor groups. Ideally, one should however be able to define their preferential orientations without having to be too specific about the rest of the molecule, to increase the scope of the search.

The choice of geometric tolerance is also left to the discretion of the user and will depend on the time and budget allowed for the screening, and on the number of tolerated false negative and false positive results. The “medium” tolerance (chosen as the default value) seems to be a good starting point for an exploratory analysis, even though higher search criteria may be more appropriate for flexible compounds.

Note that theoretically, the same procedure could be started from the API structure itself by considering that the API plays the role of its own coformer in its pure form. One could also perform a packing search on compounds similar to the API. Indeed, structures of such compounds are frequent in the CSD - whether there are single-component crystals, cocrystals or solvates-

and knowledge about packing or proclivity for solvation can be drawn from them.[10]

As we emphasized the importance of shape when discussing the utility of molecular complementarity analyses, it seems natural to also envisage packing prediction based on shape considerations rather than focusing on interactions. The work of Motherwell and Pidcock from the CCDC follow this line.[34,58,60] In particular, they proved that compounds with similar molecular shapes may display similar crystal packing despite having no formal chemical similarity or space group. And for a particular shape, there are generally only a few close-packed arrangements so that it is conceivable to design packing templates for main molecular shapes (cubes, rods or discs). But such results remain to be implemented in the CCDC softwares or elsewhere, to prove their full validity.

4.5.E Recommended methodology

1. Coordination analysis of the API to detect sterically hindered donor groups.

Molecules should be considered via their CIF files (one for each conformation if the coformer is flexible) to avoid conformation issues found by drawing 2D diagrams (see section 4.5.C). Up to now, coordination analyses are inseparable from the HBP analyses. But coordination likelihoods can be generated quickly as coordination models have been generated once and for all by Galek[6] and are thus independent of the functional groups and other parameters that one has to carefully choose in order to create valuable HBP models.

Note that for API without donor or acceptor atoms, such analysis cannot be performed.

2. Motifs searches to determine H-bond donors and acceptors able to compete with the API functional groups; targeting in priority the weakest homo-interaction or functional groups not involved in hydrogen bonding in the original API. The absence of strongly competitive heteromolecular interactions informs on the importance of the next steps.

In case of positive results in the previous step, selection of a list of coformers with promising functional groups. This can be directly performed within the GRAS (Generally Recognized As Safe) list.

3. Packing feature searches starting from the API pure form or known cocrystal structures of the API to identify coformers that could substitute the API or another coformer in its pure form/cocrystal respectively.

These searches should be performed on the list of coformers that pass the previous steps; all their plausible conformations being included in the training set.

4. In case where packing searches are unfeasible or did not bring any results, molecular complementarity tests to exclude partners with incompatible shape, restricting to S-test in case of planar APIs and to the M/L test for APIs with plate or cubic shape.

Before performing the tests, it is required to search for (in existing structures) or generate (via the Conformer Generator module) all the potential conformations of the API and coformers to encompass all possible cases. One should also have in mind the conditions of use and limitations of the different tests.

Overall, there will be a certain number of experiments to perform and in all likelihood, false positive results.

Indeed, as crystal structure prediction (CSP) calculations, CCDC tools cannot predict if an HB pairing is feasible experimentally in usual conditions or not, as they totally ignore kinetic factors.[61] Matching of coformer solubilities, for example, is important for solution crystallization, which is the technique most frequently used to produce single crystals. However, some information can be gathered by including common solvent in the initial motifs searches to assess the risk of cocrystal solvation or the potential preference for solvent molecules over coformers as complexation partners.

4.6 Conclusion and perspectives

Through the analysis of two active pharmaceutical ingredients, Levetiracetam and Paracetamol, we have demonstrated how knowledge-based informatics can help take decisions concerning coformer selection and experimental prioritization, and how one can optimally apply them to a given system.

In particular, we showed that the presence of one hetero-molecular interaction competitive with the homo-molecular interactions existing in the pure coformers may be enough to drive their cocrystallization, but that complexes won't be stable unless they display high coordination score. The prime importance of close-packing optimization through shape considerations and functional groups orientations was also illustrated here. Hence, combining methods is often profitable as, contrary to CSP calculations, the plausible compromises between the requirements of the chemical and geometrical factor is not directly encompassed and apparent.

However, promoting an automated multi-stage screening process still appears questionable as the pertinence and complementarity of the different analyses varies from one system to another, depending on the relative strength of the interactions/influences in play, and may require careful assessment of each step implications.

Moreover, further progress remains to be made to fully adapt the Solid-Form modules to the cocrystallization issue and to uncover the exact mechanisms occurring during supramolecular recognition, so that these techniques could reveal their maximal predictive potential.

In the future, one would ideally be able to:
1) compute the best combinations of hydrogen and halogen bonds from the point of view of interaction strength and donor/acceptor coordinations, using effective 3-D descriptions of the API-coformer systems, for coformers with compatible shape and orientation of the donor/acceptor groups;
2) generate, through *ab initio* calculations, likely packing patterns using pre-defined templates relative to the molecular shape of the coformers or their H-bonded synthons.

4.7 Bibliography

(1) Galek, P. T. a.; Fábián, L.; Motherwell, W. D. S.; Allen, F. H.; Feeder, N. *Acta Crystallogr. Sect. B Struct. Sci.* **2007**, *63*, 768–782.

(2) Galek, P. T. A.; Fábián, L.; Allen, F. H. *Acta Crystallogr. Sect. B Struct. Sci.* **2009**, *65*, 68–85.

(3) Wood, P. A.; Galek, P. T. A. *CrystEngComm* **2010**, *12*, 2485.

(4) Galek, P. T. a.; Fábián, L.; Allen, F. H. *Acta Crystallogr. Sect. B Struct. Sci.* **2010**, *66*, 237–252.

(5) Galek, P. T. a.; Fábián, L.; Allen, F. H. *CrystEngComm* **2010**, *12*, 2091.

(6) Galek, P. T. a; Chisholm, J. a; Pidcock, E.; Wood, P. a. *Acta Crystallogr. B. Struct. Sci. Cryst. Eng. Mater.* **2014**, *70*, 91–105.

(7) Delori, A.; Galek, P. T. a.; Pidcock, E.; Patni, M.; Jones, W. *Cryst. Eng. Comm.* **2013**, *15*, 2916.

(8) Wood, P. a.; Feeder, N.; Furlow, M.; Galek, P. T. a.; Groom, C. R.; Pidcock, E. *CrystEngComm* **2014**, *16*, 5839.

(9) Macrae, C. F.; Bruno, I. J.; Chisholm, J. A.; Edgington, P. R.; McCabe, P.; Pidcock, E.; Rodriguez-Monge, L.; Taylor, R.; van de Streek, J.; Wood, P. A. *J. Appl. Crystallogr.* **2008**, *41*, 466–470.

(10) Galek, P. T. A.; Pidcock, E.; Wood, P. A.; Bruno, I. J.; Groom, C. R. *CrystEngComm* **2012**, *14*, 2391–2403.

(11) Wood, P. A.; Olsson, T. S. G.; Cole, J. C.; Cottrell, S. J.; Feeder, N.; Galek, P. T. A.; Groom, C. R.; Pidcock, E. *CrystEngComm* **2013**, *15*, 65–72.

(12) Feeder, N.; Pidcock, E.; Reilly, A. M.; Sadiq, G.; Doherty, C. L.; Back, K. R.; Meenan, P.; Docherty, R. *J. Pharm. Pharmacol.* **2015**, *67*, 857–868.

(13) Oswald, I. D. H.; David, R.; Mcgregor, P. A.; Motherwell, W. D. S.; Parsons, S.; Colin, R. *Acta Crystallogr. Sect. B Struct. Sci.* **2002**, 1057–1066.

(14) Childs, S. L.; Stahly, G. P.; Park, A. *Mol. Pharm.* **2007**, *4*, 323–338.

(15) Karki, S.; Tomislav, F.; Fabian, L.; Laity, P. R.; Day, G. M.; Jones, W. *Adv. Mater.* **2009**, *21*, 3905–3909.

(16) Elbagerma, M. A.; Edwards, H. G. M.; Munshi, T.; Scowen, I. J. *CrystEngComm* **2011**, *13*, 1877–1884.

(17) Srirambhatla, V. K.; Kraft, A.; Watt, S.; Powell, A. V. *Cryst. Growth Des.* **2012**, *12*, 4870–4879.

(18) Sander, J. R. G.; Bučar, D.-K.; Henry, R. F.; Giangiorgi, B. N.; Zhang, G. G. Z.; MacGillivray, L. R. *CrystEngComm* **2013**, *15*, 4816.

(19) George, F.; Tumanov, N.; Norberg, B.; Robeyns, K.; Filinchuk, Y.; Wouters, J.; Leyssens, T. *Cryst. Growth Des.* **2014**, *14*, 2880–2892.

(20) Musumeci, D.; Hunter, C. a.; Prohens, R.; Scuderi, S.; McCabe, J. F.

Chem. Sci. **2011**, *2*, 883.
(21) Issa, N.; Karamertzanis, P. G.; Welch, G. W. A.; Price, S. L. *Cryst. Growth Des.* **2009**, *9*, 442–453.
(22) Karamertzanis, P. G.; Kazantsev, A. V.; Issa, N.; Welch, G. W. A.; Adjiman, C. S.; Pantelides, C. C.; Price, S. L. *J. Chem. Theory Comput.* **2009**, *5*, 1432–1448.
(23) Habgood, M.; Price, S. L. *Cryst. Growth Des.* **2010**, *10*, 3263–3272.
(24) Habgood, M. *Cryst. Growth Des.* **2013**, *13*, 4549–4558.
(25) Galek, P. T. a.; Allen, F. H.; Fábián, L.; Feeder, N. *CrystEngComm* **2009**, *11*, 2634.
(26) Majumder, M.; Buckton, G.; Rawlinson-Malone, C. F.; Williams, A. C.; Spillman, M. J.; Pidcock, E.; Shankland, K. *CrystEngComm* **2013**, *15*, 4041.
(27) Wood, P. a.; Allen, F. H.; Pidcock, E. *CrystEngComm* **2009**, *11*, 1563–1571.
(28) Hosmer, D. W.; Lemeshow, S.; Sturdivant, R. X. *Applied Logistic Regression,3rd Edition*; Wiley.; 2013.
(29) Gasteiger, J.; Marsili, M. *Tetrahedron* **1980**, *36*, 3219–3228.
(30) Bilton, C.; Allen, F. H.; Shields, G. P.; Howard, J. A. K. *Acta Crystallogr. Sect. B Struct. Sci.* **2000**, *56*, 849–856.
(31) Fábián, L. *Cryst. Growth Des.* **2009**, *9*, 1436–1443.
(32) Karki, S.; Friscic, T.; Fábián, L.; Jones, W. *CrystEngComm* **2010**, *12*, 4038.
(33) Fábián, L.; Frišcic, T. In *Pharmaceutical Salts and Co-crystals*; 2011; pp. 89–109.
(34) Pidcock, E.; Motherwell, W. D. S. *Cryst. Growth Des.* **2004**, *4*, 611–620.
(35) Desiraju, G. R. *Angew. Chem. Int. Ed. Engl.* **1995**, *34*, 2311–2327.
(36) Chisholm, J. A.; Motherwell, S. *J. Appl. Crystallogr.* **2005**, *38*, 228–231.
(37) Bruno, I. J.; Cole, J. C.; Edgington, P. R.; Macrae, C. F.; Pearson, J.; Taylor, R. *Acta Crystallogr. Sect. B* **2002**, 389–397.
(38) Van de Streek, J.; Motherwell, S. *Acta Cryst . B* **2005**, *61*, 504–510.
(39) Karki, S.; Friscic, T.; Jones, W.; Motherwell, W. D. S. *Mol. Pharm.* **2007**, *4*, 347–354.
(40) André, V.; M. da Piedade, M. F.; Duarte, M. T. *CrystEngComm* **2012**, *14*, 5005.
(41) Cheney, M. L.; Weyna, D. R.; Shan, N.; Hanna, M.; Wojtas, L.; Zaworotko, M. J. *J. Pharm. Sci.* **2011**, *100*, 2172–2181.
(42) Etter, M. C. *J. Phys. Chem.* **1991**, *95*, 4601–4610.
(43) Aitipamula, S.; Banerjee, R.; Bansal, A. K.; Biradha, K.; Cheney, M. L.; Choudhury, A. R.; Desiraju, G. R.; Dikundwar, A. G.; Dubey, R.;

Duggirala, N.; Ghogale, P. P.; Ghosh, S.; Goswami, P. K.; Goud, N. R.; Jetti, R. R. K. R.; Karpinski, P.; Kaushik, P.; Kumar, D.; Kumar, V.; Moulton, B.; Mukherjee, A.; Mukherjee, G.; Myerson, A. S.; Puri, V.; Ramanan, A.; Rajamannar, T.; Reddy, C. M.; Rodriguez-Hornedo, N.; Rogers, R. D.; Row, T. N. G.; Sanphui, P.; Shan, N.; Shete, G.; Singh, A.; Sun, C. C.; Swift, J. A.; Thaimattam, R.; Thakur, T. S.; Kumar Thaper, R.; Thomas, S. P.; Tothadi, S.; Vangala, V. R.; Variankaval, N.; Vishweshwar, P.; Weyna, D. R.; Zaworotko, M. J. *Cryst. Growth Des.* **2012**, *12*, 2147–2152.

(44) FDA. *Fda* **2013**, 1–5.

(45) Aakeröy, C. B.; Grommet, A. B.; Desper, J. *Pharmaceutics* **2011**, *3*, 601–614.

(46) Sanphui, P.; Kumar, S. S.; Nangia, A. *Cryst. Growth Des.* **2012**, *12*, 4588–4599.

(47) Blagden, N.; Berry, D. J.; Parkin, A.; Javed, H.; Ibrahim, A.; Gavan, P. T.; De Matos, L. L.; Seaton, C. C. *New J. Chem.* **2008**, *32*, 1659.

(48) Galek, P. T. a. *CrystEngComm* **2011**, *13*, 841.

(49) Trask, A. V.; Motherwell, W. D. S.; Jones, W.; Samuel Motherwell, W. D.; Jones, W. *Cryst. Growth Des.* **2005**, *5*, 1013–1021.

(50) Alhalaweh, A.; George, S.; Basavoju, S.; Childs, S. L.; Rizvi, S. A. A.; Velaga, S. P. *CrystEngComm* **2012**, *14*, 5078.

(51) Aitipamula, S.; Chow, P. S.; Tan, R. B. H. *Acta Crystallogr. Sect. E Struct. Reports Online* **2009**, *65*, o2126–o2127.

(52) Sarma, B.; Saikia, B. *CrystEngComm* **2014**, *16*, 4753.

(53) Rahim, S. A.; Hammond, R. B.; Sheikh, A. Y.; Roberts, K. J. *CrystEngComm* **2013**, *15*, 3862.

(54) Shattock, T. R.; Arora, K. K.; Vishweshwar, P.; Zaworotko, M. J. *Cryst. Growth Des.* **2008**, *8*, 4533–4545.

(55) Childs, S. L.; Rodríguez-Hornedo, N.; Reddy, L. S.; Jayasankar, A.; Maheshwari, C.; McCausland, L.; Shipplett, R.; Stahly, B. C. *CrystEngComm* **2008**, *10*, 856.

(56) Cruz-Cabeza, A. J.; Reutzel-Edens, S. M.; Bernstein, J. *Chem. Soc. Rev.* **2015**, *44*, 8619–8635.

(57) Dudenko, D. V.; Yates, J. R.; Harris, K. D. M.; Brown, S. P. *CrystEngComm* **2013**, *15*, 8797.

(58) Motherwell, W. D. S. *CrystEngComm* **2010**, *12*, 3554.

(59) Martins, F. T.; Paparidis, N.; Doriguetto, A. C.; Ellena, J. *Cryst. Growth Des.* **2009**, *9*, 5283–5292.

(60) Pidcock, E.; Motherwell, W. D. S. *Cryst. Growth Des.* **2005**, *5*, 2322–2330.

(61) Dunitz, J. D. *Chem. Commun. (Camb).* **2003**, 545–548.

(62) Clarke, H. D.; Arora, K. K.; Bass, H.; Kavuru, P.; Ong, T. T.; Pujari, T.;

Wojtas, L.; Zaworotko, M. J. *Cryst. Growth Des.* **2010**, *10*, 2152–2167.
(63) Etter, M. C.; MacDonald, J. C.; Bernstein, J. *Acta Crystallogr. Sect. B Struct. Sci.* **1990**, *46*, 256–262.

Part IV

Conclusion and Perspectives

1. Conclusion

In this thesis, we looked for ways of improving cocrystal screening procedures, by studying overall trends in cocrystal formation and by considering new screening methodologies.

By comparing the results of their respective cocrystal screening, we first showed a tendency for enantiopure and racemic versions of a selected API to form cocrystals with identical non-chiral partners. Accordingly, we suggested a new procedure to identify cocrystals more efficiently in early stages of drug research, when a racemic compound is often more readily available than its enantiopure counterpart: one should first perform an extended screen using the racemic compound, and then a focused screen with the enantiopure one, using only the coformers that led to positive hits in the first screen.

Second, we reported three novel cocrystal structures involving AKGA and (S)-/(R)-Leviteracetam or (RS)-Etiracetam, which show two very interesting features. Namely, the Levetiracetam cocrystals are formed with the lactol tautomer of AKGA, which had never been isolated in the solid-state up to now; and the existence of a stable cocrystal conglomerate in the Etiracetam-AKGA system, which is only the second report of a cocrystal conglomerate, to the best of our knowledge.

The existence of a stable conglomerate in this system was dissected and finally put in relation with the enantiospecificity of the Levetiracetam cocrystals, which is likely related to the ability of the Etiracetam enantiomers to stabilize one lactol tautomer at a time in solution, or to promote its formation by H-bonding.

More generally, by comparing the peculiarities of the system in hand to the general behavior of cocrystallizing chiral systems with and without zwitterionic coformers, it was suggested that for a pseudoquaternary cocrystal (*i.e.* cocrystal made up of two racemate compounds) to exist, the pseudoternary combinations (*i.e.* cocrystal made up of one racemate and an enantiomer of the second compound) should exist and the enantiomers of the two compounds should form a diastereomeric pair at the binary level, rather than behave enantiospecifically.

We also evidenced that a tautomeric equilibrium may be induced by grinding, without the requirement of any amount of solvent.

Third, we examined the possibility of using Isothermal Titration Calorimetry (ITC) to measure interactions between an API and complexing agents in solution, to determine whether these are indicative of successful cocrystal formation. We showed that interactions in solution between non-charged compounds, despite being quite small, can be detected by ITC, and that the recurrence of heteromeric interactions in the solid-state can be correlated with their superior strength in solution with respect to the corresponding homo-interactions.

We however demonstrated that these interactions cannot be used to identify cocrystal formers of a given API, as one needs to also consider the feasibility of an efficient tridimensional packing involving the two molecular partners. Besides, this kind of study further confirmed the importance of studying non-cocrystallizing systems in parallel to the successful ones.

We also directed attention on the limited accuracy of such results and in particular to their dependence on the API/complexing agent concentration ratio achievable with respect to the species solubilities and the desire to avoid uncontrolled occurrence of higher level aggregates.

In a final contribution, we have demonstrated how CSD knowledge-based informatics can help take decisions concerning coformer selection and experimental prioritization, and how one can optimally apply them to a given system.

We highlighted in particular that pertinence and complementarity of the different analyses varies from one system to another, depending on the relative strength of the interactions/influences in play; which goes against the idea of a one-fits-all automated multi-stage screening process.

We also pointed their limits and some potential sources of improvement in this context, as some of these tools were not initially designed to help cocrystal screenings. Taking these into account, we suggested an overall methodology involving some Mercury modules to predict cocrystallization occurrence, keeping in mind that it is currently not possible to avoid a certain number of false positive results, due in part to the ignorance, in the procedure, of the kinetic cocrystallization factors.

2. Perspectives

The Levetiracetam cocrystal screening studied throughout this thesis perfectly illustrates that cocrystallization may be driven by strong heteromolecular interactions but that these are not sufficient to generate a viable structure, and that packing considerations should not be neglected when designing a multi-component material.

It thus seems of primary importance to now pay much attention on theories, experimental techniques or methodologies that intend to attest crystal packing subtleties or to jointly take the chemical and geometrical factors into account. This is the route adopted by Pidcock and Motherwell from the CCDC on one hand, and by Desiraju and Mukherjee on the other hand, whose inspiring work has been already mentioned in this document. Some of their latest contributions are briefly discussed here along with some of their potential developments.

In this context, the work of Pidcock and Motherwell consisted in the derivation of an optimal (simple and efficient) set of parameters to describe shape and molecular packing. They introduced in particular the enclosing box model[1] mentioned previously in this thesis, which is based on the principal axes of inertia of the molecule and later refined to include some description of the void included in the box.[2] Using this model, they were able to sort structures by similarity indexes. They however noticed the presence of substructures in sets of similar structures, which still require some parameters to be fully described and discriminated.

They also investigated the relationship between the packing pattern, the faces of the molecules within the unit cell, and the strongest interactions present in a given structure.[3] They notably detected two likely packing patterns: the ones with low surface area on one hand (Figure 1) and those in which the largest faces of the molecules are related by symmetry operators rather than by unit cell translation on the other hand. Interactions between the large molecular faces are indeed often energetically strong. They however emphasized the importance of considering the motif direction with respect to the direction of the large faces, when there are strong symmetry-demanding motifs in the structure as in that case, the resulting packing pattern must accommodate both (Figure 2).

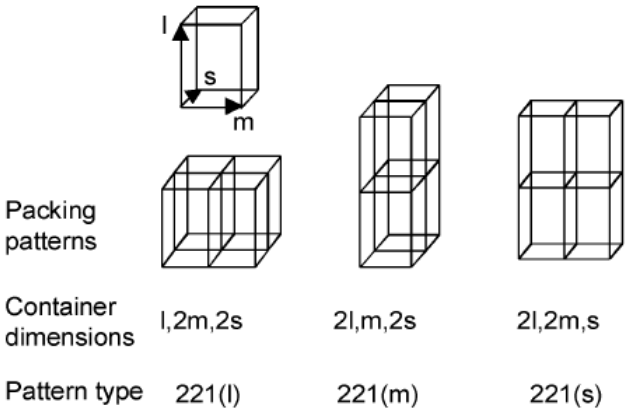


Figure 1. Three ways of stacking 4 boxes; the enclosing containers being characterized by the same volume but different surface area. The arrangement at the left shows the least surface area and is thus the more likely pattern in the 221 family.[1]

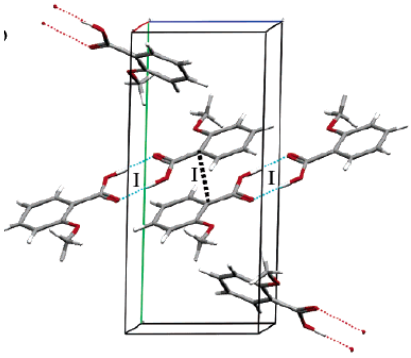


Figure 2. Example of a structure (FUFBOX) in which both the strong symmetry-demanding motif (the carboxylic acid dimer) and the interaction of the large faces of the molecules are mediated through inversion.[3]

Such information seems of great interest but one still has to figure how it could be used in practice for crystal engineering, for example as a selection criteria to rank putative structures in CSP. Fayos did one step in this direction but conceded that more extensive and diversified training sets than the one he used should be employed for more accurate predictions.[4] Besides, one could imagine performing such a CSD investigation focusing solely on cocrystal datasets as the conclusions that hold for them may be different from the ones drawn by studying single-component crystals. The population rank by symmetry operators may differ, for example.

The route explored by Desiraju, Mukherjee and coworkers is slightly different and aims to account for the fact that primary synthons are now quite predictable but often have a poor predictive capacity in terms of packing due to their lack of representativeness in this regard. They proposed a model to allow the description of the crystallization path from 1D synthon to the final 3D crystal, through the formation of Long-range Synthons Aufbau Modules (LSAM).[5–7] According to this theory, molecular recognition in solution is a systematic process that follows the hierarchy of interactions to create assemblies of increasing complexity.

LSAM are composite synthons formed by combining primary synthons to secondary ones and thus uniting the chemical (Desiraju's supramolecular synthons) and the geometrical (Kitaigorodskii's close-packing principle) approaches. Indeed, early synthons rely solely on chemical compatibility whereas the interactions formed at later stages of crystal packing are weaker but responsible for the structural symmetry or space-group. LSAM may involve conventional H-bonds but also pi-pi interactions, halogen bonds and other anisotropic interactions. By taking into account the late and weaker interactions, the aim is not to consider their probability of formation or their individual contribution in deciding the packing, which are both inherently low, but their combined effect and their compatibility in terms of overall molecular topology to form a higher aggregate.

LSAM may allow prediction and tailor-made engineering in case where primary synthons responsible for growth in the three unit cell directions have been identified and show modularity (*i.e.* independent existence in solution, Figure 3).[8,9] This reasoning was successfully used for the rational design of ternary and quaternary cocrystals without interference (*i.e.* no polymorphism or appearance of lower level cocrystals),[10] which are challenging operations as they require the knowledge of all steps in the nucleation process. In some cases, LSAM were also proven to persist to the detriment of very robust heterosynthons such as the carboxylic acid-amide one,[9] highlighting their broad utility.

Figure 3. LSAM formed in 3,4,5-trichlorophenol as the combination of the separate synthons formed in 3,5-dichlorophenol and 4-chlorophenol crystals; showing their "blindness" with respect to each other in solution.[8]

It thus seems coherent to assume that potential for larger synthons could be used as a new criterion to evaluate coformer propensity to cocrystallize. In fact, consideration of LSAM could have been the missing characteristic to sort coformers tested in the Levetiracetam cocrystal screening. Indeed, it was early recognized that cocrystallizing coformers display functional groups oriented in almost opposite directions. But it was not possible to take this feature into account in the packing feature search module, due to the flexibility of the coformers preventing any standardized search. Similarly, one can rationalize the formation of the Paracetamol cocrystals and the success of the corresponding packing searches by noticing that the three types of interactions responsible for cocrystal cohesion are structurally insulated and spread in the three directions (*i.e.* the amide interactions in the first, the phenol one in the second and the aromatic ring interaction in the last direction), explaining their modularity and their assembly in various LSAMs, even though they were not recognized as such.

Hence, one could suggest new potential coformers by considering how each interaction could be altered (*i.e.* which moieties could be used as substitute for each interaction on the target), without necessarily requiring common structural features, which was the limiting condition in CSD packing searches. Work thus remains to assess interactions equivalence and hierarchy and to suggest a systematic procedure that could lead to robust predictions. For that matter, studying the incremental formation of interactions in solution by increasing concentration using various NMR

methodologies,[9] as mentioned in the third chapter of this thesis, appears very promising. This is especially interesting as LSAM are less deformable than short-range synthons and their presence in solution likely attest to the formation of thermodynamic products less prone to polymorphism.

In order to do so, the optimal way seems to take a combinatorial approach and vary systematically different parameters, such as substituent position or nature, before comparing the structural features of the generated family of compounds. Doing so, information on secondary synthons is easily gathered and one can spot in particular the modulable interactions and regions that may be further elaborated. Similarly variations may be applied to the experimental methodology (e.g. grinding time or nature of the solvent in LAG)[11] to assess nucleation paths, kinetics and thermodynamics competition and their consequence on the packing pattern.

Such a systematic approach led to the early successes of Aakeröy and coworkers[12,13] and more recently to some interesting conclusions by Desiraju et *al.*;[14] which may be opposed to a large majority of contributions that rather focus on features of selected compounds without considering the bigger picture. In this context, using the existing structures recorded in the CSD may help to considerably reduce the experimental work. For that matter, one can however not insist enough on how prejudicial the absence is of a database recording all the unsuccessful crystallization experiments.

Hence, crystal engineering remains a fascinating subject full of subtleties and potential developments, but important breakthroughs may be expected in a near future thanks to all the experienced teams of passionate researchers working in this field.

3. References

(1) Pidcock, E.; Motherwell, W. D. S. *Cryst. Growth Des.* **2004**, *4*, 611–620.
(2) Motherwell, W. D. S. *CrystEngComm* **2010**, *12*, 3554.
(3) Pidcock, E.; Motherwell, W. D. S. *Cryst. Growth Des.* **2005**, *5*, 2322–2330.
(4) Fayos, J. *Cryst. Growth Des.* **2009**, *9*, 3142–3153.
(5) Ganguly, P.; Desiraju, G. R. *CrystEngComm* **2010**, *12*, 817–833.
(6) Desiraju, G. R. *J. Am. Chem. Soc.* **2013**, *135*, 9952–9967.
(7) Mukherjee, A. *Cryst. Growth Des.* **2015**, *15*, 3076–3085.
(8) Mukherjee, A.; Desiraju, G. R. *Cryst. Growth Des.* **2011**, *11*, 3735–

3739.
(9) Mukherjee, A.; Dixit, K.; Sarma, S. P.; Desiraju, G. R. *IUCrJ* **2014**, *1*, 228–239.
(10) Dubey, R.; Mir, N. A.; Desiraju, G. R. *IUCrJ* **2016**, *3*, 102–107.
(11) Cinčić, D.; Friščić, T.; Jones, W. *J. Am. Chem. Soc.* **2008**, *130*, 7524–7525.
(12) Aakeröy, C. B.; Salmon, D. J. *CrystEngComm* **2005**, *7*, 439.
(13) Aakeröy, C. B.; Desper, J.; Smith, M. M. *Chem. Commun.* **2007**, 3936.
(14) Mukherjee, A.; Desiraju, G. R. *Cryst. Growth Des.* **2014**, *14*, 1375–1385.

Part V

Appendices

Structural investigation of substituent effect on hydrogen bonding in (S)-phenylglycine amide benzaldimines

George, F.; Norberg, B.; Wouters, J.; Leyssens, T. *Cryst. Growth Des.* **2015**, *15*, 4005–4019.

Abstract: A detailed structural analysis of twenty-three new crystal structures of (S)-phenylglycine amide benzaldimines with various substituents (CH_3, Ph, OCH_3, F, Cl, Br, NO_2) on the benzylidene is performed in this contribution. These compounds belong to the highly studied family of Schiff bases. Etter's nomenclature and Hirshfeld surfaces are used to describe respectively the strong hydrogen bonds and the secondary interactions existing in these compounds. Surprisingly, all 23 obtained structures can be sorted in five types according to their hydrogen bonding motifs. The potential interplay of steric and electronic effects of the substituents on the resulting bonding patterns, conformational features and packing was investigated. Our analysis revealed that neither mesomeric/inductive factors of halogens nor π-π stacking, C-H…π and other hydrophobic interactions affect the structural outcome. The type affiliation is rather due to the interplay of three parameters: (1) the number of strong hydrogen bonds forming the motif (thermodynamic factor), (2) the ease with which the motif is formed (kinetic factor) and (3) the capacity of the motif to accommodate substituents on the different positions (steric factor). It was thus possible to suggest a stability ranking of the five structural types and to identify stable forms when polymorphism was encountered.

1. Introduction

During the past years, Schiff bases have received much attention due to their wide range of biological activities[1,2] and industrial applications. Among their pharmacological properties, they show antibacterial,[3] anticancer,[4] antifungal,[5] and radical scavenging[6] activities. They can also be used as enzymatic intermediates or inhibitor.[7] Stable and easily synthesized, chiral Schiff bases are widely used in organic chemistry as intermediates in the formation of chiral amines and various carbonyl compounds. Due to the π-acceptor properties of the imine nitrogen, they are commonly encountered ligands in coordination chemistry.[8,9] Furthermore, they have shown their use in asymmetric catalysis.[10]

Among Schiff bases, N-(2-Methylbenzylidene)phenylglycine amide has recently been used as a model compound for deracemization through abrasive grinding [11–15] while 2-(benzylideneamino)-2-(2-chlorophenyl)acetamide helped to demonstrate the possibility of using attrition-enhanced deracemization in an up-scaled process.[16] Both these compounds fulfill the requirements for the deracemization technique to work; they form racemic conglomerates in the solid phase (i.e. R and S molecules crystallize in different crystals) and they are easily racemizable in solution.

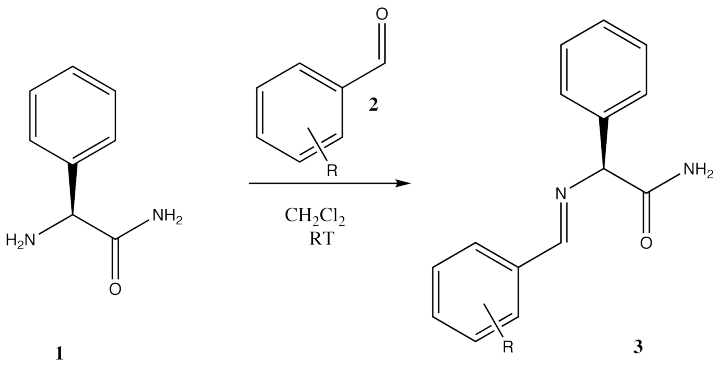
Schiff bases have extensively been structurally characterized[17–22] but only a limited amount of studies investigate the relationship between supramolecular motifs and nature/position of different substituents on a molecular framework.[23,24]

In the current contribution, we analyze the crystal structures of twenty (S)-phenylglycine amide benzaldimines having various substituents located on different positions on the benzylidene. This study will help to understand the solid state behavior of this type of imines, and yield insight into how the nature, size and position of the substituent impact the hydrogen bonding patterns.

All compounds were synthesized by condensation of (S)-phenylglycine amide ((S)-PGA) and the corresponding monosubstituted benzaldehyde (Scheme 1). Etter's nomenclature[25] was used to describe the strong hydrogen bonds existing in the twenty-three crystal structures presented here, while Hirshfeld surfaces[26] served to identify their secondary interactions. The potential interplay of steric and electronic effects of the substituents on the

resulting bonding patterns, conformational features and packing, was investigated in detail.

Scheme 1. Synthesis of (S)-phenylgycine derivatives (3) by condensation of (S)-PGA (1) and a monosubstituted benzaldehyde (2), in dichloromethane at room temperature. R = CH_3, Ph, OCH_3, F, Cl, Br, NO_2.



2. Experimental Section

Starting Materials. (S)-Phenylglycine amide, 2-anisaldehyde, 3-anisaldehyde, 3-nitrobenzaldehyde and 4-nitrobenzaldehyde were purchased from Acros Organics. 2-Tolualdehyde, 3-tolualdehyde, 4-tolualdehyde, 2-bromobenzaldehyde, 3-bromobenzaldehyde, 2-chlorobenzaldehyde, 4-chlorobenzaldehyde, 3-fluorobenzaldehyde, 4-chlorobenzaldehyde and biphenyl- 2-carboxaldehyde were purchased from Sigma-Aldrich. 4-Anisaldehyde and 2- fluorobenzaldehyde were purchased from Alfa Aesar. 4-Bromobenzaldehyde and 2- nitrobenzaldehyde were purchased from Maybridge. 3- Chlorobenzaldehyde and biphenyl- 4-carboxaldehyde were purchased from TCI.

Synthesis. (S)-PGA-aldimines were prepared by addition of the substituted benzaldehyde to a suspension of (S)-PGA in dichloromethane and left to stirr overnight at room temperature, as described by Dalmolen *et al.*[27]

Single Crystals. Most single crystals were grown by slow evaporation of the corresponding solution or by cooling crystallization to 3°C. Different solvents (methanol, acetonitrile, ethyl acetate, dichloromethane and acetone) were used as polymorphism was suspected for some compounds (see below). For the 2-chlorobenzaldehyde derivative, different polymorphs were obtained when using acetonitrile, acetone or methanol as crystallization solvent.

Those were named FI, FII and FIII respectively. Similarly, two polymorphs were isolated for the 2-anisaldehyde product when using methanol and dichloromethane, and named FI and FII respectively.

Single Crystal X-ray Diffraction. Single crystal X-ray diffraction was performed on a Gemini Ultra R system (4-circle kappa platform, Ruby CCD detector) using Cu Kα radiation (λ = 1.54056 Å). Cell parameters were estimated from a pre-experiment run and full data sets collected at room temperature. The structures were solved by direct methods with the SHELXS-97 program and then refined on $|F|^2$ using SHELXL-97 software[28]. The final reported R_1 value is calculated on |F| for the observed reflections (I > 2 sigma(I)). Non-hydrogen atoms were anisotropically refined, and the hydrogen atoms in riding mode with isotropic temperature factors were fixed at 1.2 times U(eq) of the parent atoms (1.5 times for methyl groups). Hydrogen atoms implicated in H-bonds were located in the Fourier difference maps and freely refined.

Hirshfeld surfaces. Hirshfeld surfaces are among other techniques[29] that allow the visualisation of intermolecular interactions formed by a molecule in a given crystal structure. The Hirshfeld surface of a molecule in a crystal is the surface delimiting « the region where the electron distribution of a sum of spherical atoms for the molecule dominates the corresponding sum over the crystal ».[26] This surface can be mapped with different functions. Here, we used only the Hirshfeld surface mapped with d_e (distance external to the surface), the distance from the surface to the nearest nucleus in another molecule, which give information of close intermolecular contacts. The surface colour reflects the proximity of the neighbours : 0.55 angstrom (red) – 1.5 angstrom (green) – 2.4 angstrom (blue). Hydrogen bonds are visible on the d_e surface as large red regions adjacent to the H bond acceptor and as smaller orange-red dot adjacent to the H bond donor.

A 2D fingerprint plot is a plot of d_e in function of d_i, the distance from the surface to the nearest atom in the molecule itself (distance internal to the surface). It summarizes all the intermolecular interactions in a given crystal and provides the relative area of the surface corresponding to each such interaction. Points are coloured from blue, corresponding to the smallest non-zero contribution to the total surface, to red, for contribution of 0.1% or greater to the total surface.

Hirshfeld surfaces and 2D fingerprint plots were generated using the licenced free-of-charge CrystalExplorer software.[30]

3. Results

All twenty synthesized aldimines are labelled according to the nature (CH_3, Ph, OCH_3, F, Cl, Br, NO_2) and position (ortho, meta, para) of the substituent on the benzylydene. They were structurally characterized through single crystal analysis. Crystallographic parameters of all compounds are displayed in Table 1. The crystal structure of o-Me at 208K has already been reported in the CSD[31,32] and shows similar parameters.

Overall, in every structure type, the imine adopts a trans configuration with respect to the C=N bond. Moreover, except for what we will define later on as type IV structures, the amide hydrogen H2B always faces the imine nitrogen N1 (Figure 1). In type IV structures, the carbonyl occupies this position. Although this might seem unfavorable due to the proximity of the lone pairs of the carbonyl and imine groups, this orientation allows H2B to form a hydrogen bond with the imine nitrogen in an intermolecular way. This conformation and the overall hydrogen bonding pattern of type IV structures also occurs in 2-(Benzylideneamino)-2-(2-chlorophenyl)acetamide, which is the only related structure reported in the CSD.[33]

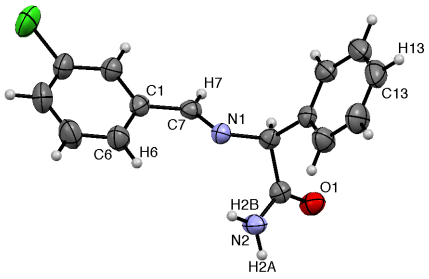


Figure 1. ORTEP plot (Mercury software 3.0) of m-Cl showing crystallographic numbering scheme on the atoms potentially involved in inter- and intramolecular interactions in the various structures.

A further general feature is the presence of the substituents in ortho and meta positions on the H7 side and not on the imine nitrogen side. This conformation is expected to be favored, as most of the hydrogen bonding partners are located on the nitrogen side, and substituents on this side would prevent strong hydrogen bonding interaction due to steric effects. Furthermore, ortho-substituents on the nitrogen side would lead to steric hindrance between the substituent and the nitrogen lone pair.

The only exception to the above observation is m-F for which some rotational disorder around the C1-C7 bond can be found, with about 75% of all molecules having the fluorine on the C3 atom (H7 side) and 25% on the C5 (nitrogen side). This can be explained by the small size of the fluorine atom and the reduced steric effect.

Table 1. Crystallographic parameters of the 23 new structures sorted by type

Type I Compounds	**o-Ph**	**p-OMe**	**p-F**	**p-Cl**	**p-Br**
Structural formula	$C_{21}H_{18}N_2O$	$C_{16}H_{16}N_2O_2$	$C_{15}H_{13}N_2OF$	$C_{15}H_{13}N_2OCl$	$C_{15}H_{13}N_2OBr$
Formula weight (g/mol)	314.37	268.31	256.27	272.72	317.17
Space system	monoclinic	monoclinic	monoclinic	monoclinic	monoclinic
Space group	$P2_1$	$P2_1$	$P2_1$	$P2_1$	$P2_1$
a (Å)	10.3943(8)	6.7394(3)	6.8172(8)	6.7343(4)	6.7098(4)
b (Å)	7.7222(5)	7.9886(3)	7.9216(9)	8.0362(4)	8.0495(4)
c (Å)	11.8976(10)	13.3084(10)	12.594(2)	12.9186(7)	13.1003(8)
α (°)	90	90	90	90	90
β (°)	112.922(10)	94.308(5)	103.040(16)	100.314(6)	98.867(6)
γ (°)	90	90	90	90	90
V ($Å^3$)	879.574	714.478	662.577	687.834	699.098
Z	2	2	2	2	2
R1-factor (%)	3.48	5.05	4.27	3.07	4.46
d ($g.cm^{-3}$)	1.187	1.247	1.285	1.317	1.507

Type I Compounds	**p-NO2**	**m-F**	**m-Cl**	**m-Br**	**m-NO2**
Structural formula	$C_{15}H_{13}N_3O_3$	$C_{15}H_{13}N_2OF$	$C_{15}H_{13}N_2OCl$	$C_{15}H_{13}N_2OBr$	$C_{15}H_{13}N_3O_3$
Formula weight (g/mol)	283.28	256.27	272.72	317.17	283.28
Space system	monoclinic	monoclinic	orthorhombic	orthorhombic	orthorhombic
Space group	$P2_1$	$P2_1$	$P2_1\,2_1\,2_1$	$P2_1\,2_1\,2_1$	$P2_1\,2_1\,2_1$
a (Å)	6.7108(4)	6.9740(3)	6.9705(3)	7.0276(4)	7.1806(6)
b (Å)	7.9363(4)	7.8282(3)	7.7622(4)	7.445(5)	7.7128(7)
c (Å)	13.2410(8)	12.5581(6)	25.4745(12)	25.6078(13)	25.267(2)
α (°)	90	90	90	90	90
β (°)	98.269(6)	105.511(5)	90	90	90
γ (°)	90	90	90	90	90
V ($Å^3$)	697.87	660.625	1378.33	1393.71	1399.35
Z	2	2	4	4	4
R1-factor (%)	3.97	3.03	3.5	3.6	4.22
d ($g.cm^{-3}$)	1.348	1.288	1.314	1.512	1.345

Table 1. Continued

Type II Compounds	m-Me [a]	o-Cl FI [a]	o-Cl FII [a]	o-NO2 [a]
Structural formula	$C_{16}H_{16}N_2O$	$C_{15}H_{13}N_2OCl$	$C_{15}H_{13}N_2OCl$	$C_{15}H_{13}N_3O_3$
Formula weight (g/mol)	252.31	272.72	272.72	283.28
Space system	orthorhombic	orthorhombic	orthorhombic	orthorhombic
Space group	$P2_1\,2_1\,2_1$	$P\,n\,a\,2_1$	$P2_1\,2_1\,2_1$	$P2_1\,2_1\,2_1$
a (Å)	5.1707(5)	18.549(3)	5.4131(2)	5.2320(4)
b (Å)	13.2714(12)	13.827(2)	13.2515(6)	13.7900(13)
c (Å)	20.434(2)	5.3162(7)	18.5144(9)	18.6379(17)
α (°)	90	90	90	90
β (°)	90	90	90	90
γ (°)	90	90	90	90
V ($Å^3$)	1402.23	1363.48	1328.07	1344.71
Z	4	4	4	4
R1-factor (%)	4.74	4.48	3.58	5.09
d ($g.cm^{-3}$)	1.195	1.329	1.364	1.399

Type II Compounds	o-Me [b]	o-Ome FI [b]	o-Cl FIII [b]	o-Br [b]	o-F [c]
Structural formula	$C_{16}H_{16}N_2O$	$C_{16}H_{16}N_2O_2$	$C_{15}H_{13}N_2OCl$	$C_{15}H_{13}N_2OBr$	$C_{15}H_{13}N_2OF$
Formula weight (g/mol)	252.31	268.31	272.72	317.17	256.27
Space system	orthorhombic	orthorhombic	orthorhombic	orthorhombic	monoclinic
Space group	$P2_1\,2_1\,2_1$	$P2_1\,2_1\,2_1$	$P2_1\,2_1\,2_1$	$P2_1\,2_1\,2_1$	C2
a (Å)	5.2563(5)	5.1883(6)	5.20340(10)	5.2600(3)	20.3160(18)
b (Å)	9.7698(12)	9.8853(10)	9.7982(2)	9.7112(5)	5.2050(3)
c (Å)	26.617(3)	27.403(3)	26.4764(5)	26.707(2)	13.5364(14)
α (°)	90	90	90	90	90
β (°)	90	90	90	90	114.387(12)
γ (°)	90	90	90	90	90
V ($Å^3$)	1366.86	1405.44	1349.87	1364.22	1303.69
Z	4	4	4	4	4
R1-factor (%)	6.53	4.51	3.17	3.79	3.51
d ($g.cm^{-3}$)	1.226	1.268	1.342	1.554	1.306

[a] First sub-group; [b] second sub-group; [c] third sub-group.

Table 1. Continued

Type III Compounds	p-Me	p-Ph
Structural formula	$C_{16}H_{16}N_2O$	$C_{21}H_{18}N_2O$
Formula weight (g/mol)	252.31	314.37
Space system	monoclinic	monoclinic
Space group	$P2_1$	$P2_1$
a (Å)	8.2225(4)	8.3299(3)
b (Å)	5.7932(3)	5.7214(2)
c (Å)	14.5687(6)	17.5391(5)
α (°)	90	90
β (°)	90.377(4)	91.143(3)
γ (°)	90	90
V ($Å^3$)	639.959	835.724
Z	2	2
R1-factor (%)	3.09	3.29
d ($g.cm^{-3}$)	1.207	1.249

Type IV Compounds	m-OMe
Structural formula	$C_{16}H_{16}N_2O_2$
Formula weight (g/mol)	268.31
Space system	monoclinic
Space group	$P2_1$
a (Å)	8.6581(7)
b (Å)	5.1375(3)
c (Å)	15.7993(10)
α (°)	90
β (°)	100.982(7)
γ (°)	90
V ($Å^3$)	689.899
Z	2
R1-factor (%)	4.27
d ($g.cm^{-3}$)	1.292

Type V Compounds	o-OMe FII
Structural formula	$C_{16}H_{16}N_2O_2$
Formula weight (g/mol)	268,31
Space system	monoclinic
Space group	C 2/c
a (Å)	25.096(2)
b (Å)	5.6957(4)
c (Å)	20.3318(15)
α (°)	90
β (°)	107.199(6)
γ (°)	90
V ($Å^3$)	2776.26
Z	8
R1-factor (%)	3.62
d ($g.cm^{-3}$)	1.284

Surprisingly, we were able to categorize all 23 obtained structures (20 different compounds and respective polymorphs) in five structurally based types, according to their main hydrogen bonding motifs. The structural analysis below uses graph sets [25] to describe each type of main bonding pattern encountered and Hirshfeld surfaces to consider all secondary contacts which, given their number, play a key role in the structure building and packing efficiency. Indeed, according to Desiraju,[34] the presence or absence of those weaker interactions could even be determinant for the patterns formed by the stronger hydrogen bonds in the crystal.

3.1 Type I : p-OMe, o-Ph, m-F, p-F, m-Cl, p-Cl, m-Br, p-Br, m-NO_2, p-NO_2

Type I structures are characterized by repeating [R^2_2 (9)] ring motifs (Figure 2): the imine lone pair of a first molecule (A) accepts a H bond from an amide hydrogen of a second molecule (B), while the amide hydrogen of the first molecule donates a hydrogen bond to the carbonyl of the second molecule. Each molecule is therefore involved in four different hydrogen bonds, with every potential H bond former being used.

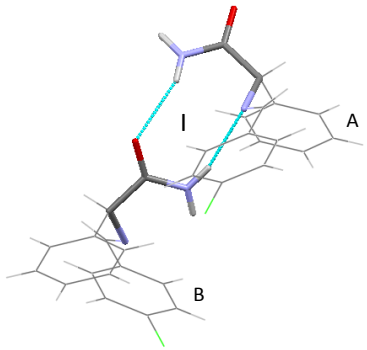


Figure 2. Type I motif displaying a [R^2_2 (9)] ring between two molecules (A and B) of p-Cl.

Consecutive ring motifs produce zigzag ladders, containing two parallel infinite hydrogen-bond chains, described in graph set notation as [C^2_2(7)] and directed along the *b*-axis. From one sport of the ladder to the other, molecules are rotated by 180° around the *b*-axis; all phenyl groups pointing outwards. The one-dimensional ladders are stacked periodically along the *a* and *c* axes (Figure 3).

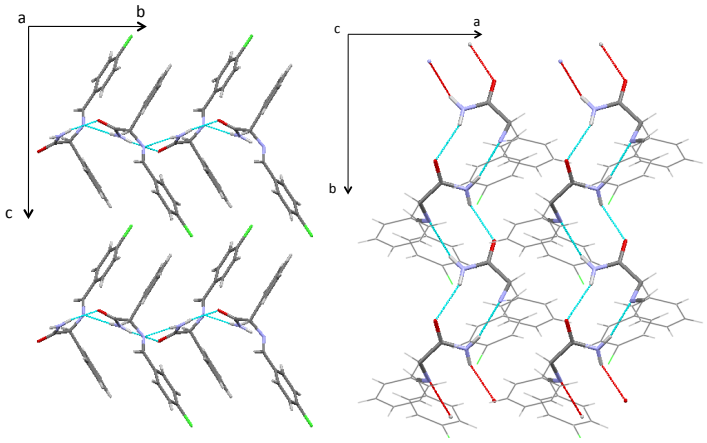


Figure 3. Stacking of p-Cl in the *bc* (left) and *ab* (right) planes, displaying zigzag ladders with [$C^2_2(7)$] graph set notation, forming undulating one dimensional ribbons.

Type I structures crystallize either in the monoclinic $P2_1$ or in the orthorhombic $P2_12_12_1$ space groups.

Concerning the weaker hydrogen bonds, a particularly strong C—H···O interaction is found between C6—H6 and O1 (Figure 4 and Table 4). This is in accordance with the results of Lo Presti *et al.* (2006) stating that C--H···O bonds of comparable strength to O--H···O bonds can exist in organic molecules.[35]

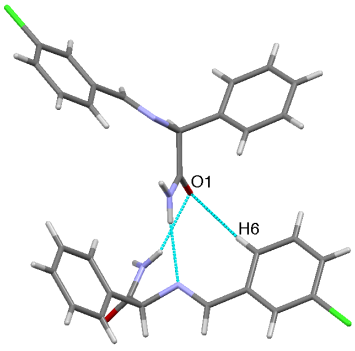


Figure 4. C6—H6···O1 intermolecular bond in type I structures (m-Cl)

One can also note the presence of a weaker C-H···O bond (C7-H7···O1) directed along the *a*-axis connecting the carbonyl of one ladder to the imine hydrogen H7 of another ladder and influencing the tridimensional

arrangement. However, this additional interaction does not occur in o-Ph because the biphenyl group prevents a sufficient proximity between adjacent ladders; the packing stability being due supposedly to hydrophobic interactions in this case (see below). This inter-ladder interaction is not present either in m-NO_2, which rather shows a C5-H5¨O3 bond, involving the nitro group holding ladders together (Figure 5).

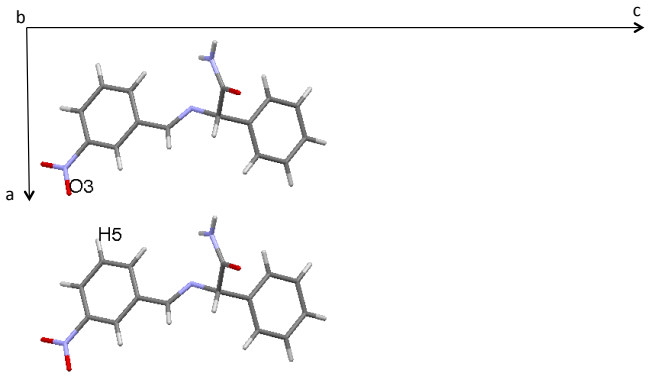


Figure 5. C5-H5¨O3 bond in m-NO_2

In addition, comparison of π- π stacking, C-H…π and other hydrophobic interactions in the different structures can easily be performed by analyzing the 2D fingerprint plot and the corresponding Hirshfeld surfaces of the molecules in the different structures. Indeed, they gives us a complete view of intermolecular interactions, focusing not solely on 'assumed important interactions'.[26]

Among type I structures, disparities are found within the secondary interactions.

For example, the 2D plot of m-F (Figure 6, left and middle) displays so-called « horns » between the spikes while other meta-substituted type I structures do not. Those account for the presence of the short C-F¨H interaction present in m-F.

In addition, comparing it with the 2D fingerprint plot of m-Br (Figure 6, right), one observes that the latter structure is less efficiently packed, as illustrated by the presence of a diffuse blue region at high distances.

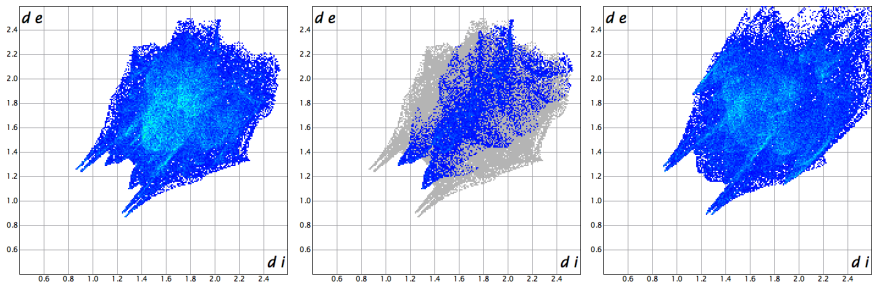


Figure 6. 2D fingerprint plots of m-F (left and central figure) displaying the "horns" corresponding to the short C-H···F interaction (highlighted on the central figure). On the right, the 2D fingerprint plots of m-Br with no horns and less efficient packing.

Similarly, the 2D fingerprint plot of o-Ph differs significantly from the others by the presence of a bump between the spikes and of a pair of wings on the other side of the spikes (Figure 7). The central bump corresponds to all the hydrophobic H···H contacts while the external wings represent the C-H···π interactions which are highly represented in this particular structure, visible as orange zones above the rings on the de Hirshfeld surface (Figure 8).

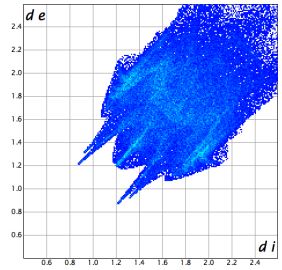


Figure 7. 2D fingerprint plot of o-Ph displaying a central bump, corresponding to H···H contacts, and a pair of external wings corresponding to the C-H···π interactions.

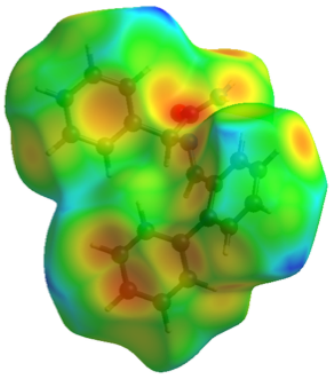


Figure 8. Hirshfeld surface of o-Ph mapped with d_e and displaying bright orange regions above the rings, corresponding to various C-H...π interactions.

3.2 Type II: m-Me, o-Me, o-OMe FI, o-F, o-Cl (FI, FII, FIII), o-Br, o-NO_2

As for type I structures, type II structures are also organized in ladders but this time running along the *a*-axis. The ladders are constituted by the succession of inverted [$R^2_3(8)$] ring motifs involving three molecules (two adjacent ones, A & C, and one on the other side of the ladder, B). The carbonyl of the first molecule (A) forms a bifurcated H bond with one amide hydrogen of the second (B) and the third (C) molecule. In addition, the second amide hydrogen of the third molecule is linked to the carbonyl of the second molecule (Figure 9).

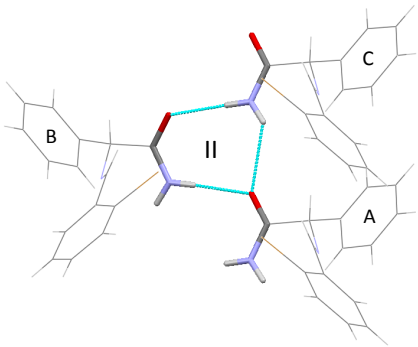


Figure 9. Type II motif displaying a [$R^2_3(8)$] ring formed by three molecules (A,B & C) of o-Br.

As for type I structures, each molecule takes part in four H bonds, but this time two donating hydrogen bonds using the amide hydrogens and two accepting bonds through the carbonyl group. Contrary to the type I structures, the imine is not included in any H bonding pattern in type II structures.

Contrary to type I, no distinctive inter-ladder interaction is reported. However, when the substituent is a halogen/nitro group, an extra intramolecular hydrogen bond is formed between the substituent and the imine hydrogen H7.

The successive ring motifs form two parallel hydrogen-bonded infinite chains C(4) directed along the *a*-axis (Figure 10). Ladders are almost planar, except in o-F structure, in which ladders are slightly undulating.

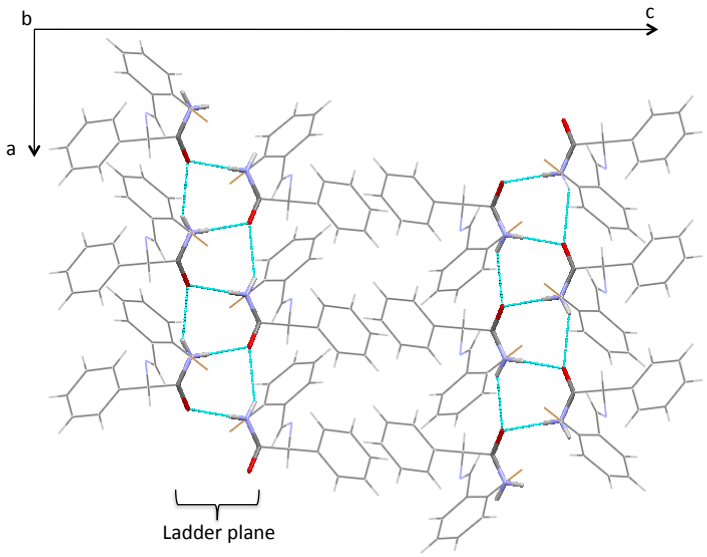


Figure 10. o-Br stacking in the *ac* plane, displaying infinite chains C(4) creating almost planar ladders.

Although all motif II structures display the same hydrogen bonding patterns, their overall molecular packing is less homogenous. Therefore the type II structures can be sorted in three different sub-groups according to the stacking of the ladders along the *b* and *c* axes.

A first sub-group includes structures m-Me, o-Cl (FI, FII) and o-NO_2. Ladders are stacked in alternating rows along the longest axis (*a* or *c*). The ladders

planes are parallel within a row but form an angle with ladders planes situated in the next row. On top, consecutive ladders rows are interpenetrating (Fig. 11).

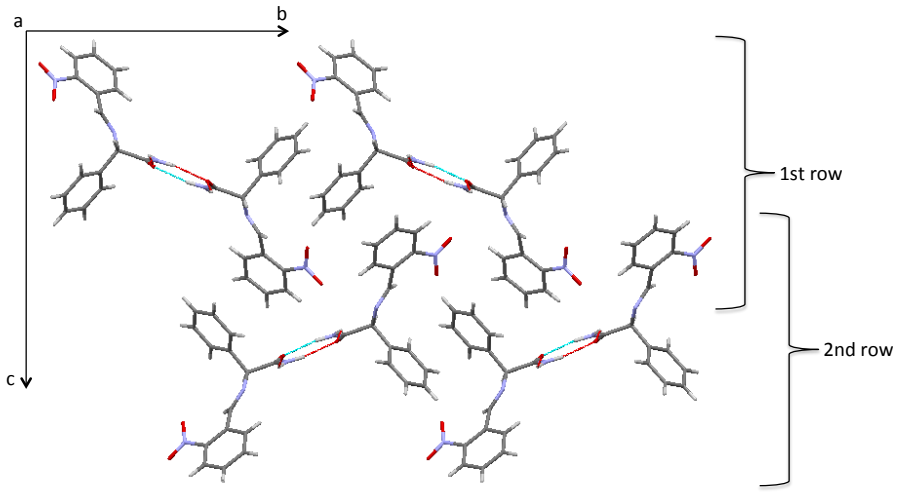


Figure 11. o-NO_2 rows stacking in the *bc* plane.

In the second sub-group (o-Me, o-OMe FI, o-Cl FIII and o-Br), one observes the same packing as in the first sub-group, except that subsequent rows do not interpenetrate, but form herringbone arrangements (Figure 12).

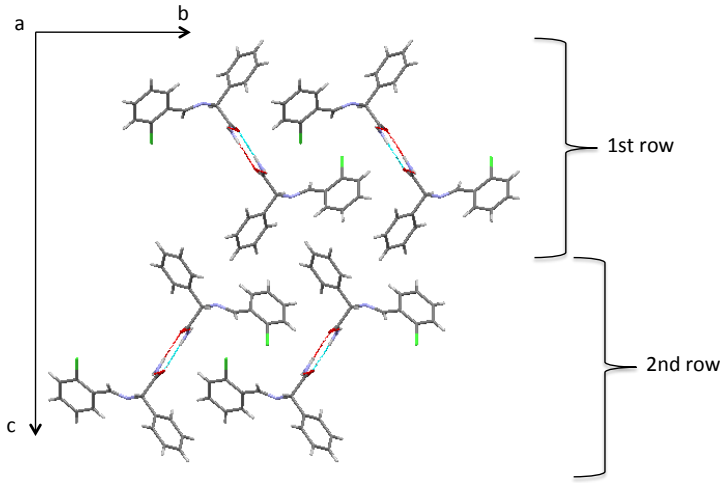


Figure 12. o-Cl (FIII) stacking in the *bc* plane exhibiting herringbone arrangement.

This is due to the fact that the inclination angle of the ladders planes with respect to the *b*-axis is more pronounced in this last sub-group (Figure 13).

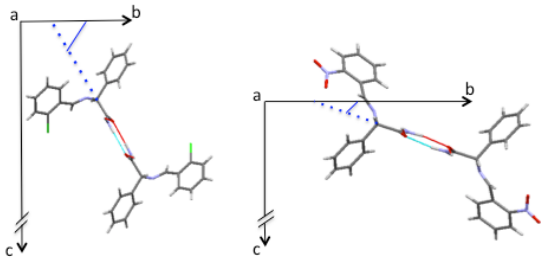


Figure 13. The inclination angle of the ladders planes in the *bc* plane with respect to the *b*-axis is smaller in the first sub-group (right, o-NO_2) than in the second (left, o-Cl (FIII)).

In the last sub-group (o-F), ladders planes are parallel within and between rows (Figure 14). This is therefore the only compound that crystallizes in the monoclinic space group *C2*; all the other compounds in type II belonging to orthorhombic space groups.

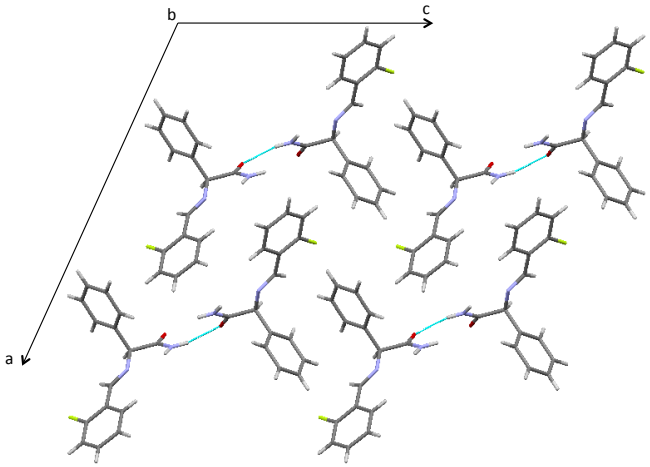


Figure 14. o-F stacking in the *ac* plane.

The analysis of the secondary interactions also reveals disparities among type II structures. For most ortho substituted type II structures, there is a spike on the diagonal of the 2D fingerprint plot. This accounts for the directional H···H interactions found in those structures. Those interactions are depicted by orange areas on the bottom left and right of the Hirshfled surface of o-Cl FII, together with the interacting molecules (Figure 15). The

strength of those interactions varies tremendously according to the nature of the substituent as well as the nature of the polymorph as shown by the three o-Cl structures (Figure 16).

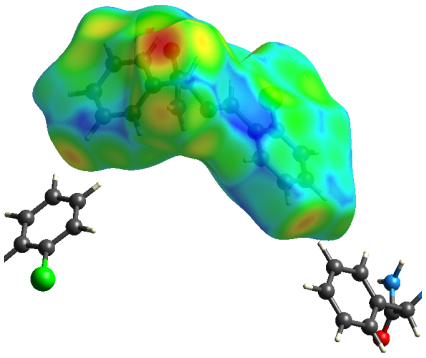


Figure 15. Hirshfeld surface of o-Cl (FII) mapped with d_e function. Directional H···H contacts are represented.

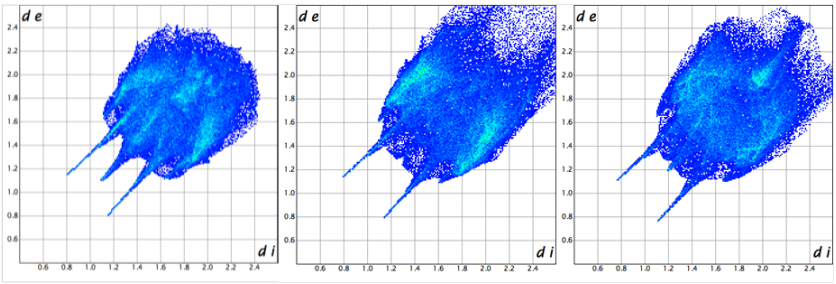


Figure 16. 2D fingerprint plot of o-Cl FII (left), FI (middle) and FIII (right).

3.3 Type III: p-Me and p-Ph

These structures are characterized by head-to-tail catemers, propagating along the b-axis and characterized by a $[C^2_2(8)]$ infinite hydrogen-bonded chain motif: a H bond connects the carbonyl of a molecule to one amide hydrogen of a second molecule, related to the first one by the twofold screw axis in $P2_1$ (Figure 17). The imine is not participating in any supramolecular motif and only one hydrogen on the amide nitrogen atom is involved in H bond formation. Hence, unlike type I and II structures, only two hydrogen bonds are formed per molecule in this type.

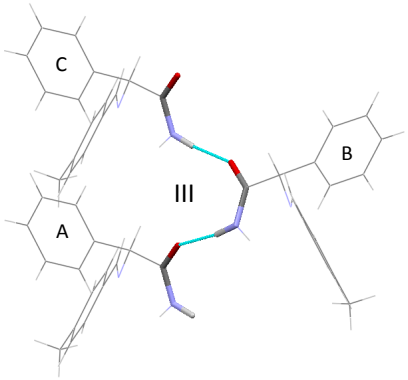


Figure 17. Type III motif displaying [$C^2_2(8)$] hydrogen-bonded infinite chains formed by three molecules (A, B & C) of p-Me.

Type III structures exhibit an inter-chain link joining the carbonyl and the para hydrogen on the phenyl group, providing chain cohesion along the *a*-axis (Figure 18). The same connection is present in type IV albeit with greater strength (see Table 4 below).

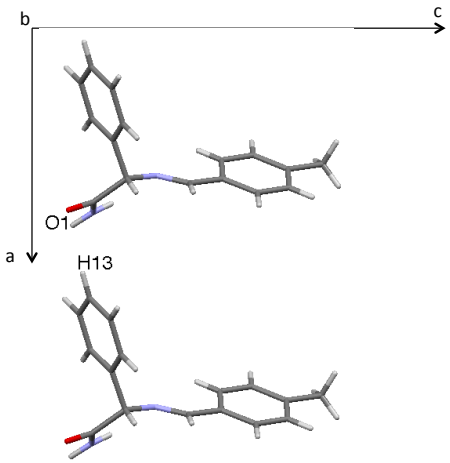


Figure 18. C13-H13···O1 bond in p-Me.

Besides, an additional C8—H8···O1 interaction is present in p-Ph (but not in p-Me). Accordingly, the catemer formed by p-Me molecules is planar (Figure 17) while it is angular (113° between successive hydrogen bonds in the catemer) in the p-Ph structure (Figure 19).

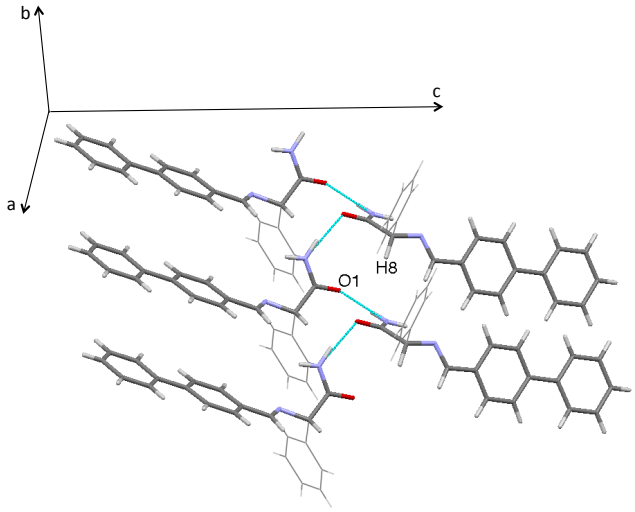
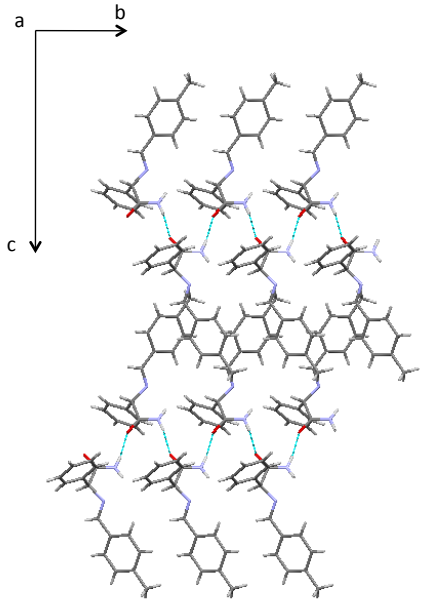


Figure 19. C8—H8···O1 intermolecular interaction and angular catemer in p-Ph (type III).

The network can be described by the stacking of non-interacting chains along the *a* and *c* axes. As in o-F (type II), chains are intercalated and chains planes parallel within and between rows in the *bc* plane (Figure 20).



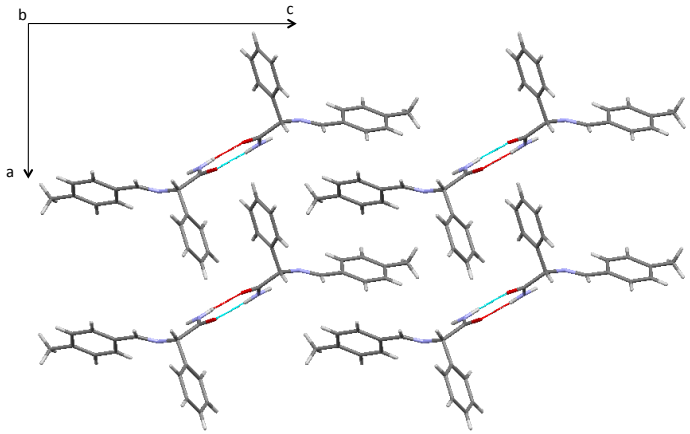


Figure 20. p-Me stacking in *bc* and *ac* planes.

Both p-Me and p-Ph structures belong to the monoclinic space group $P2_1$.

As for the other types, the two type III structures differ significantly from one another with respect to the hydrophobic contacts. The p-Me 2D fingerprint plot is characterized by a central bump and external wings, corresponding to the presence of H···H contacts and C-H···π interactions respectively, while p-Ph does not have any of these features (Figure 21).

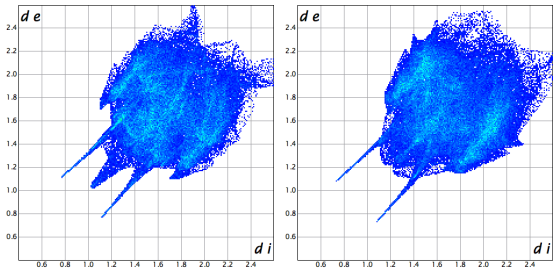


Figure 21. 2D fingerprint plot of p-Me (left) and p-Ph (right). p-Me plot displays a central bump and a pair of external wings corresponding to H···H contacts and C-H···π interactions respectively.

3.4 Type IV: m-OMe

This type of structures exhibit succession of $[R^3_3(11)]$ ring motifs involving three molecules (Figure 22). A H bond joins one amide hydrogen of a first molecule (A) to the imine of a second molecule (C). Another H bond links the second molecule carbonyl to the third molecule's amide hydrogen (B). A last

H bond exists between the third molecule's carbonyl (B) and the first molecule's amide hydrogen (A).

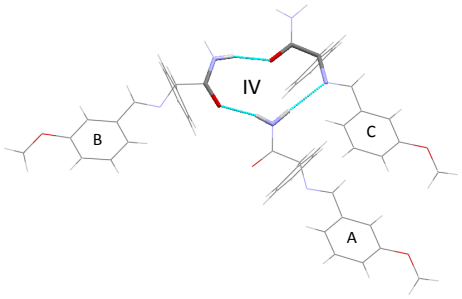


Figure 22. Type IV motif displaying a $[R^3_3(11)]$ ring formed by three molecules (A, B & C) of m-OMe.

Those successive motifs form one inner $[C^2_2(8)]$ and two outer C(5) hydrogen-bonded infinite chains directed along the *b*-axis (Figure 23).

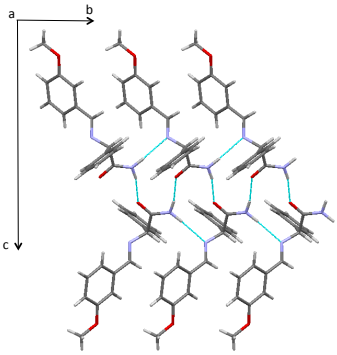


Figure 23. Inner $[C^2_2(8)]$ and outer C(5) hydrogen-bonded infinite chains constituted by successive motifs.

Contrary to the type II structures, the imine takes part in the hydrogen bonding patterns and each carbonyl accepts only one H bond. Furthermore, unlike the type III structures, both amide hydrogens are involved in such interactions. Thus, as in type I, each molecule is involved in 4 different hydrogen bonds.

Those motifs generate one dimensional hydrogen bonded twisted ladders stacked along the *a* and *c* directions. Once again, those twisted ladders are intercalated and ladders planes are parallel within and between rows in the *ac* plane (Figure 24).

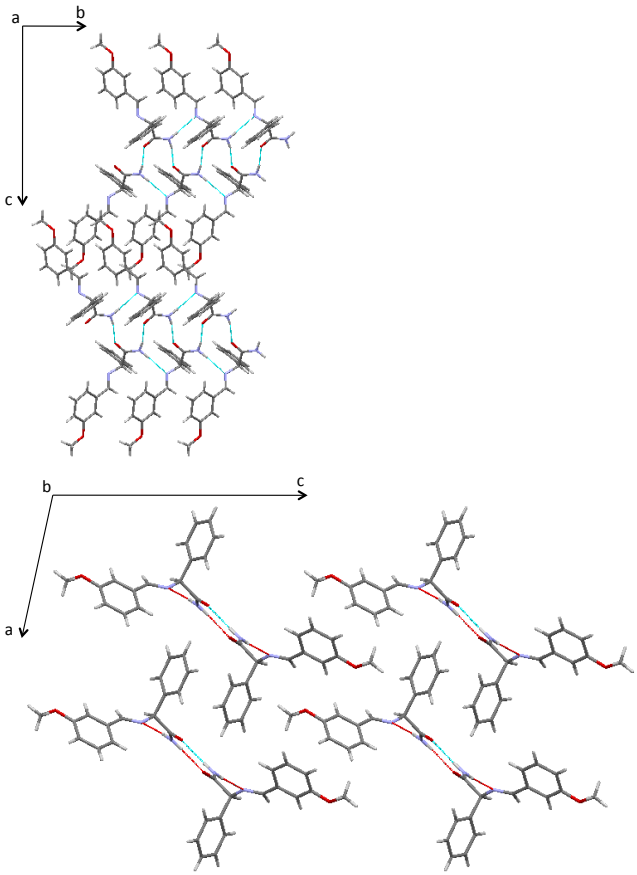


Figure 24. Stacking of m-OMe in the *bc* and *ac* planes exhibiting one dimensional hydrogen bonded tapes.

Concerning other interactions, one can denote the presence of particularly strong C13—H13···O1 (Figure 25 and Table 4) interaction.

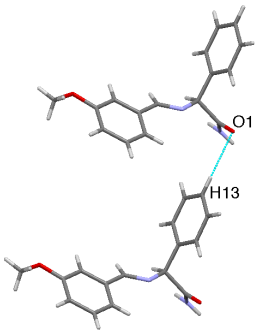


Figure 25. C13—H13···O1 interchain bond in type IV structure (m-OMe).

The type IV structure also crystallizes in the monoclinic space-group $P2_1$.

On the 2D fingerprint plot of type IV, the only significant feature concerning the hydrophobic contacts is the presence of horns near the diagonal, which are due to the fact that in the $[R^3_3(11)]$ hydrogen bonding pattern of this structure, the amide hydrogens of two interacting molecules are very close and hence provide a non-directional H···H contact (Figure 26).

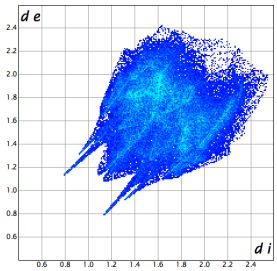


Figure 26. 2D fingerprint plot of m-OMe displaying a pair of horns near the diagonal, corresponding to close non-directional H···H contact in the hydrogen bonding pattern.

3.5 Type V: o-OMe FII

The type V structure has a dimer motif as main pattern. The dimers are formed by amide-amide homosynthons $[R^2_2(8)]$ joining the amides of two molecules (A & B, Figure 27). It is the only type in which no infinite motifs (*i.e.* chains) are present. As in type II and III, the imine lone pair does not take part in any intermolecular hydrogen bonding and as in type III, only one amide hydrogen gives a hydrogen bond. Hence, as in type III, each molecule forms only two distinct intermolecular hydrogen bonds. However, a weak intra-molecular hydrogen bond is formed between the imine hydrogen H7 and the oxygen O2 on the substituent.

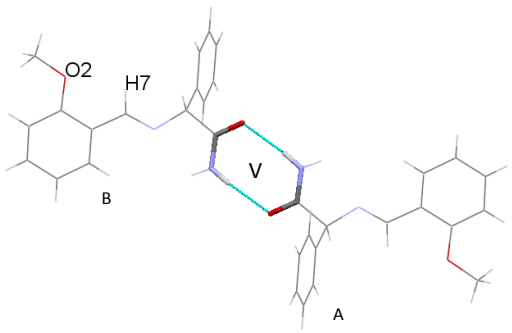


Figure 27. Type V motif displaying a [$R^2_2(8)$] ring formed by two molecules (A & B) of o-OMe.

Overall, the packing is as in type II first subgroup: dimers are stacked in alternating and intercalated rows along the *c*-axis. Type V structure belongs to the monoclinic space group *C* 2/c.

As main feature, the 2D fingerprint of type V structure shows a central bump corresponding to the hydrophobic H···H contacts. Furthermore, the structure seems to be less tightly packed according to the presence of a sparse region at the high distances on Figure 28.

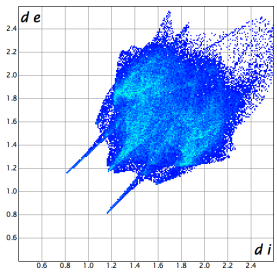


Figure 28. 2D fingerprint plot of o-OMe displaying a central bump and a diffuse region at the high distances.

4. Discussion

The allocation of structures in different types (displayed in Table 2) reveals that:

- All halogen/nitro meta/para substituted compounds show type I structures;

- All halogen/nitro ortho substituted compounds show type II structures;

- Structures with an alkyl/methoxy substituent are encountered in various types.

- Type I and II motifs are by far the most frequently encountered, with 19 out of the 23 structures belonging to these types.

This leads to the question whether the attribution to a given type is mostly due to electronic or rather steric effects specific to the substituents or their position on the benzylidene.

Table 2. Graph sets and bond nature of the most prominent features in each type, along with all structures belonging to those types.

Types	Hydrogen Bonding Patterns	Main interactions [a]	Interactions localisation
I	$[R^2_2(9)]$, $[C^2_2(7)]$	N2 -- H2A .. N1	intermolecular
		N2 -- H2B .. O1	intermolecular
		N2 -- H2B .. N1	intramolecular
II	$[R^2_3(8)]$, C(4)	N2 -- H2A .. O1	intermolecular
		N2 -- H2B .. O1	intermolecular
		N2 -- H2B .. N1	intramolecular
III	C(8)	N2 -- H2A .. O1	intermolecular
		N2 -- H2B .. N1	intramolecular
IV	$[R^3_3(11)]$, $[C^2_2(7)]$, C(5)	N2 -- H2B .. N1	intermolecular
		N2 -- H2A .. O1	intermolecular
		N2 -- H2B .. O1	intermolecular
V	$[R^2_2(8)]$	N2 -- H2A .. O1	intermolecular
		N2 -- H2B .. N1	intramolecular

[a] C--H¨O interactions not taken into account

Concerning the electronic effects, we note that both meta and para halogen substituted moieties exhibit type I structures. However, theoretical charges generated by mesomeric and inductive effects of halogens are located at different positions on the benzylidene for meta and para-substituted compounds. In other words, these charges do not impact the formation of hydrogen bonds involving the benzylidene hydrogens or the imine nitrogen. This result is in agreement with DFT calculations, showing the charges present on the benzylidene atoms are very similar regardless of the nature and position of the substituent on the ring. One can thus conclude that mesomeric and inductive factors of halogens do not significantly affect the structural outcome and that their position (and subsequent steric occupation) on the ring may play a more important role.

Since π-π stacking, C-H¨π and other hydrophobic interactions vary tremendously within a type of structures, they should not be determining for type affiliation either and the discussion below will therefore focus on the stronger interactions present in the various structures (Table 3). Their analysis reveals that only four different main interactions occur, no matter the overall hydrogen motifs formed. Indeed, there are two potential strong donors (both amide hydrogens) and two potential strong acceptors (the imine nitrogen and the carbonyl oxygen) common to all compounds, so the combinations are limited.

Among those interactions, the N2--H2···N1 bond is present in each type, following Etter's rule of the best H-bond acceptor and donor associating with each other.[36] In type II, III and V, this interaction is intramolecular while being intermolecular in type IV. In type I, an intramolecular N2--H2B···N1 is found in addition to an intermolecular N2--H2A···N1.

However, comparing the intermolecular bonds lengths and angles, it appears that N2--H2···O1, which is also present in every type, is the shortest and more linear interaction. In fact, the lone pair of the imine nitrogen is more basic than the carbonyl oxygen but also more sterically hindered (by the phenyl groups situated on both sides). The carbonyl oxygen is thus more accessible and prompt to form short hydrogen bonds with neighboring molecules.

Table 3. Bond lengths (angstrom) and angles (°) of the main intermolecular hydrogen bonds in the 20 compounds sorted by types.

Type I Compounds	D-H···A	D-H (Å)	H···A (Å)	D···A (Å)	D-H···A (Å)	Symmetry
o-Ph	N2 -- H2A .. N1	0.92(2)	2.32(2)	3.203(2)	161(2)	1-x,-1/2+y,1-z
	N2 -- H2B .. O1	0.95(3)	2.14(3)	2.998(2)	150.2(19)	1-x,1/2+y,1-z
	N2 -- H2B .. N1	0.95(3)	2.34(2)	2.736(2)	104.5(16)	
p-OMe	N2 -- H2A .. N1	0.86(3)	2.28(3)	3.099(3)	159(2)	2-x,-1/2+y,1-z
	N2 -- H2B .. O1	0.86(3)	2.30(3)	3.088(3)	153(2)	2-x,1/2+y,1-z
	N2 -- H2B .. N1	0.86(3)	2.33(3)	2.692(3)	105(2)	
m-F	N2 -- H2A .. N1	0.85(2)	2.28(2)	3.1064(19)	164(2)	1-x,-1/2+y,1-z
	N2 -- H2B .. O1	0.85(2)	2.29(2)	3.0808(18)	154.9(14)	1-x,1/2+y,1-z
	N2 -- H2B .. N1	0.85(2)	2.343(16)	2.6919(19)	105.0(12)	
p-F	N2 -- H2A .. N1	0.86	2.28	3.103(3)	161	-x,-1/2+y,-z
	N2 -- H2B .. O1	0.86	2.26	3.064(3)	156	-x,1/2+y,-z
	N2 -- H2B .. N1	0.86	2.34	2.702(3)	106	
m-Cl	N2 -- H2A .. N1	0.88(2)	2.24(2)	3.107(2)	169.2(19)	1-x,1/2+y,1/2-z
	N2 -- H2B .. O1	0.84(2)	2.32(2)	3.087(2)	153(2)	1-x,-1/2+y,1/2-z
	N2 -- H2B .. N1	0.84(2)	2.33(2)	2.691(2)	106.4(17)	
p-Cl	N2 -- H2A .. N1	0.90(2)	2.26(2)	3.146(2)	165.0(19)	1-x,-1/2+y,1-z
	N2 -- H2B .. O1	0.86(2)	2.27(2)	3.069(2)	154.8(18)	1-x,1/2+y,1-z
	N2 -- H2B .. N1	0.86(2)	2.33(2)	2.700(2)	106.3(16)	
m-Br	N2 -- H2A .. N1	0.89	2.28	3.115(3)	156	-x,-1/2+y,3/2-z
	N2 -- H2B .. O1	0.93	2.21	3.089(3)	157	-x,1/2+y,3/2-z
	N2 -- H2B .. N1	0.9300	2.34	2.680(3)	101	
p-Br	N2 -- H2A .. N1	0.84(7)	2.32(8)	3.139(7)	167(6)	-x,-1/2+y,1-z
	N2 -- H2B .. O1	0.87(7)	2.26(7)	3.083(6)	159(4)	-x,1/2+y,1-z
	N2 -- H2B .. N1	0.87(7)	2.36(5)	2.700(6)	103(4)	
m-NO_2	N2 -- H2A .. N1	0.87(2)	2.29(2)	3.134(3)	164(2)	1-x,1/2+y,3/2-z
	N2 -- H2B .. O1	0.85(3)	2.34(3)	3.090(3)	147.8(19)	1-x,-1/2+y,3/2-z
	N2 -- H2B .. N1	0.85(3)	2.27(2)	2.687(3)	110.5(17)	
p-NO_2	N2 -- H2A .. N1	0.91(3)	2.26(2)	3.127(2)	160(2)	2-x,-1/2+y,-z
	N2 -- H2B .. O1	0.90(2)	2.21(2)	3.048(2)	154.0(18)	2-x,1/2+y,-z
	N2 -- H2B .. N1	0.90(2)	2.33(2)	2.694(2)	104.2(15)	

Table 3. Continued

Type II Compounds	D-H···A	D-H (Å)	H···A (Å)	D···A (Å)	D-H···A (Å)	Symmetry
m-Me	N2 -- H2A .. O1	0.81(3)	2.15(3)	2.944(3)	169(3)	-1/2+x,-1/2-y,2-z
	N2 -- H2B .. O1	0.92(3)	2.19(3)	2.962(3)	142(2)	-1+x,y,z
	N2 -- H2B .. N1	0.92(3)	2.33(3)	2.726(3)	106(2)	
o-Me	N2 -- H2A .. O1	0.93(3)	1.97(3)	2.889(3)	175(2)	-1/2+x,-1/2-y,1-z
	N2 -- H2B .. O1	0.84(3)	2.50(3)	3.058(3)	124(2)	-1+x,y,z
	N2 -- H2B .. N1	0.84(3)	2.36(3)	2.736(4)	108(2)	
o-OMe FI	N2 -- H2A .. O1	0.94(2)	1.98(2)	2.915(3)	174(2)	1/2+x,-1/2-y,2-z
	N2 -- H2B .. O1	0.82(3)	2.39(3)	2.978(3)	130(2)	1+x,y,z
	N2 -- H2B .. N1	0.82(3)	2.40(3)	2.742(3)	107(2)	
o-F	N2 -- H2A .. O1	0.86	2.14	2.959(2)	160	1/2-x,-1/2+y,1-z
	N2 -- H2B .. O1	0.86	2.58	3.180(3)	128	x,-1+y,z
	N2 -- H2B .. N1	0.86	2.34	2.705(2)	106	
o-Cl FI	N2 -- H2A .. O1	0.86	2.08	2.934(4)	170	1-x,-y,-1/2+z
	N2 -- H2B .. O1	0.86	2.43	3.107(5)	136	x,y,-1+z
	N2 -- H2B .. N1	0.86	2.35	2.705(4)	105	
o-Cl FII	N2 -- H2A .. O1	0.89(4)	2.08(4)	2.942(3)	165(3)	-1/2+x,1/2-y,2-z
	N2 -- H2B .. O1	0.83(3)	2.58(3)	3.233(3)	138(3)	-1+x,y,z
	N2 -- H2B .. N1	0.83(3)	2.33(3)	2.714(3)	109(2)	
o-Cl FIII	N2 -- H2A .. O1	0.84(2)	2.04(2)	2.8819(19)	179(2)	1/2+x,3/2-y,2-z
	N2 -- H2B .. O1	0.81(2)	2.45(3)	3.0248(19)	128.1(19)	1+x,y,z
	N2 -- H2B .. N1	0.81(2)	2.43(2)	2.748(2)	104.3(19)	
o-Br	N2 -- H2A .. O1	0.86	2.04	2.892(4)	172	1/2+x,3/2-y,-z
	N2 -- H2B .. O1	0.86	2.51	3.074(4)	124	1+x,y,z
	N2 -- H2B .. N1	0.86	2.37	2.743(4)	107	
o-NO_2	N2 -- H2A .. O1	0.93(4)	1.97(4)	2.887(3)	168(3)	1/2+x,3/2-y,1-z
	N2 -- H2B .. O1	0.91(4)	2.31(4)	3.015(3)	134(3)	1+x,y,z
	N2 -- H2B .. N1	0.91(4)	2.38(4)	2.737(3)	104(3)	

Type III Compounds	D-H···A	D-H (Å)	H···A (Å)	D···A (Å)	D-H···A (Å)	Symmetry
p-Me	N2 -- H2A .. O1	0.94(2)	1.96(2)	2.886(2)	172(2)	1-x,1/2+y,2-z
	N2 -- H2B .. N1	0.86(2)	2.27(2)	2.704(2)	111.6(18)	
p-Ph	N2 -- H2A .. O1	0.95(3)	1.88(3)	2.803(2)	164(2)	1-x,1/2+y,-z
	N2 -- H2B .. N1	0.90(2)	2.20(3)	2.646(2)	110(2)	

Type IV Compounds	D-H···A	D-H (Å)	H···A (Å)	D···A (Å)	D-H···A (Å)	Symmetry
m-OMe	N2 -- H2B .. N1	0.98(3)	2.26(3)	3.241(3)	176(2)	x,1+y,z
	N2 -- H2A .. O1	0.89(3)	2.04(3)	2.920(3)	167(3)	1-x,1/2+y,2-z
	N2 -- H2B .. O1	0.98(3)	2.55(3)	3.054(3)	111.6(19)	x,1+y,z

Type V Compounds	D-H···A	D-H (Å)	H···A (Å)	D···A (Å)	D-H···A (Å)	Symmetry
o-OMe FII	N2 -- H2A .. O1	0.883(19)	2.099(19)	2.9773(17)	172.7(16)	-x,2-y,-z
	N2 -- H2B .. N1	0.877(17)	2.276(18)	2.6881(17)	108.7(14)	

Furthermore, according to Galek *et al.*,[37] structures where all the good donors are satisfied (*i.e.* forming their preferential number of coordination) are favored, even if some acceptors are left unemployed. This is in

accordance with our results: in every structure type, both amide hydrogens take part in H bonds while for some types, the imine accepts no hydrogen bond. Similarly, the carbonyl can accept two hydrogen bonds but this happens only in type II structures. In the other types, the carbonyl does not form bifurcated hydrogen bonds.

However, depending on the type, H2B forms an intra- or an intermolecular bond. This latter is notably stronger than the former, as the corresponding bond lengths and angles testify.

Moreover, as Bilton emphasizes,[38] the most probable intramolecular H-bonding motifs are planar conjugated 6-membered rings. Hence, in structures where they are observable, certain 6-membered ring motifs are almost 100% likely to form.[39] Yet the present intramolecular H-bond (N2--H2B$^{\cdot\cdot}$N1) rather forms a 5-membred ring without π-electron delocalization. It has thus a reduced probability to appear in a given structure.

This is supported by the fact that, usually, when a donor hydrogen is involved in a intramolecular H bond, it is less likely to participate in an additional intermolecular contact.[38] Yet, in type I and II, H2B forms an intra- and an intermolecular H bond. This proves that this particular intramolecular bond is so weak that it does not prevent additional interactions.

Given that, one can even question if the intramolecular (N2--H2B$^{\cdot\cdot}$N1) bond identified by Platon software[40] is really a true hydrogen bond or rather an « artefact of other stronger interactions » as Taylor called them.[41] This is supported by Wood *et al.* analysis, showing that contacts with D-H$^{\cdot\cdot\cdot}$A angles below 120° are not significant interactions *per se*.[42]

In an attempt to classify the different types according to their respective stability, we notice that type III and V structures display only one strong hydrogen bond (angle > 120°), while the other types possess two, and that H2B is involved only in the aforementioned very weak intramolecular bond. Hence those two structure types are less stable than the other structure types, explaining why they are among the less encountered. Accordingly, among the two polymorphs identified for o-OMe structure, form I is expected to be the thermodynamically stable one.

In the three other types (I, II & IV) structures, the two strong hydrogen bonds formed seem to be of comparable magnitude and are not expected to be the main cause for type affiliation.

Nonetheless, the types can clearly be distinguished when looking at the connectivity of each molecule. In type II and IV, each molecule is connected to four other ones; each intermolecular bond being formed with a different partner. While in type I, one molecule is linked to only two other molecules; two bonds being formed with the same partner. Consequently, types II and IV structures may be more difficult to form since it requires the concomitant approach of fives molecules, constrained by their mutual steric hindrance. In other words, type I structures seems kinetically favored in comparison with type II and IV structures.

On top, type IV is presumably less represented than type I and II because the H bonding motif present in this type does not seem to suit ortho- and para-derivatives. Indeed, a substituent in those positions would sterically hinder the approach of the adjacent molecules on the ladder and other neighboring ladders molecules.

Finally, type I structures present a particularly strong C-H···O intermolecular interaction (Table 4) between molecules of one ladder that may be evoked to justify its preponderance toward type II structures, which do not display this additional contact.

Table 4. Bond lengths (angstrom) and angles (°) of the C—H···O intermolecular interactions in compounds of type I, III & IV.

Type I Compounds	D-H···A	D-H (Å)	H···A (Å)	D···A (Å)	D-H···A (Å)	Symmetry
o-Ph	C6 -- H6 .. O1	0.93	2.54	3.466(2)	173	1-x,1/2+y,1-z
p-OMe	C6 -- H6 .. O1	0.93	2.43	3.357(3)	175	2-x,1/2+y,1-z
p-F	C6 -- H6 .. O1	0.93	2.49	3.415(3)	172	-x,1/2+y,-z
p-Cl	C6 -- H6 .. O1	0.93	2.47	3.396(2)	174	1-x,1/2+y,1-z
p-Br	C6 -- H6 .. O1	0.93	2.46	3.382(6)	173	-x,1/2+y,1-z
p-NO_2	C6 -- H6 .. O1	0.93	2.45	3.373(2)	171	2-x,1/2+y,-z
m-F	C6 -- H6 .. O1	0.93	2.56	3.470(2)	167	1-x,1/2+y,1-z
m-Cl	C6 -- H6 .. O1	0.93	2.49	3.414(2)	173	1-x,-1/2+y,1/2-z
m-Br	C6 -- H6 .. O1	0.93	2.48	3.403(4)	174	-x,1/2+y,3/2-z
m-NO_2	C6 -- H6 .. O1	0.93	2.54	3.468(3)	174	1-x,-1/2+y,3/2-z
Type III Compounds	**D-H···A**	**D-H (Å)**	**H···A (Å)**	**D···A (Å)**	**D-H···A (Å)**	**Symmetry**
p-Ph	C8 -- H8 .. O1	1.00(2)	2.58(2)	3.396(2)	139.1(16)	1-x,-1/2+y,-z
Type IV Compounds	**D-H···A**	**D-H (Å)**	**H···A (Å)**	**D···A (Å)**	**D-H···A (Å)**	**Symmetry**
m-OMe	C13 -- H13 .. O1	0.93	2.6	3.501(4)	164	-1+x,y,z

Hence it is reasonable to expect that crystallization of another derivative from this family of compounds would preferentially lead to a structure belonging to type I or II, with a slight preference for type I.

Other steric considerations can also be taken into account to further differentiate between type I and II. For example, one can easily understand that most structures with an ortho substituent do not belong to type I in which the substituent would be too close to the carbonyl of an adjacent ladder. Conversely, the o-Ph structure is part of type I instead of II because, in this case, the cavity occupied by the other ortho substituents in type II would be too small to accommodate the phenyl group.

One should however keep in mind, that for some compounds described here, it seems that different types would still be sterically allowed and it is therefore not a straightforward task to predict the resulting type. Polymorphism of these compounds seems likely, especially for those having an alkyl/methoxy substituent on the benzylidene moeity. For example, one can easily imagine m-Me (type II), p-Me and p-Ph (type III) in type I. Similarly, m-oMe could form an intramolecular N2--H2B···N1 bond rather than an intermolecular one and belong to another structure type. Unfortunately, the polymorphism investigation carried out so far to confirm this hypothesis did not lead to any of these alternative forms (Table 5).

Table 5. Solvents[a] in which a single crystal was successfully grown and polymorphs identified for the 20 compounds sorted by types.

Types	Compounds	Solvents of crystallization	Polymorphs found in specified solvent
I	o-Ph	MeOH, ACN, EtAc	-
	p-OMe	ACN, EtAc	-
	p-F	MeOH, ACN, EtAc	-
	p-Cl	MeOH, ACN, EtAc, DCM	-
	p-Br	MeOH, EtAc	-
	p-NO2	ACN, EtAc	-
	m-F	MeOH, EtAc, DCM	-
	m-Cl	MeOH, ACN, EtAc	-
	m-Br	MeOH, ACN	-
	m-NO2	MeOH, ACN	-
II	m-Me	MeOH, DCM	-
	o-Me	MeOH, ACN, EtAc	-
	o-OMe	MeOH,ACN,EtAc,DCM	DCM: FII in type V
	o-F	MeOH, EtAc	-
	o-Cl	MeOH, ACN, Ac	ACN: FI, Ac: FII, MeOH: FIII
	o-Br	MeOH, ACN, EtAc	-
	o-NO2	ACN, EtAc	-
III	p-Me	MeOH,ACN,EtAc,DCM	-
	p-Ph	MeOH, ACN, EtAc, DCM	-
IV	m-OMe	MeOH, ACN	-
V	o-OMe	MeOH,ACN,EtAc,DCM	MeOH: FI in type II

[a]MeOH: methanol, ACN: acetonitrile, EtAc: ethyl acetate, DCM: dichloromethane, Ac: acetone.

5. Conclusion

In this contribution, the structural analysis of twenty compounds from the family of (S)-phenylglycine amide benzaldimines has been performed.

Paying attention only to strong hydrogen bonds (*i.e.* bonds in which the hydrogen is linked to a highly electronegative atom), it was possible to sort the twenty compounds into five types according to the hydrogen bonding pattern formed. In most structure types, the nature of the hydrogen bonds

is similar and the difference resides in their number and position (inter- or intramolecular) in the crystal.

We then performed a more thorough investigation of each type by considering secondary interactions. Some interactions inter-or intramotif (such as C-H···O ones) were found to be specific to certain types. But as far as C-H···π, π-π stacking and other hydrophobic interactions are concerned, we noticed they vary considerably within a type and therefore do not seem to be responsible for type affiliation.

Our analysis reveals that there are 3 other factors that guide the formation of a specific motif and its preponderance over the other motifs:

1. the number of strong hydrogen bonds formed in the motif, which can include C-H···O contacts (thermodynamic considerations),

2. the ease with which the motif is formed, which is related to the coordination number of each molecule in the structure type (kinetic considerations),

3. the capacity of the motif to accommodate substituents on the different positions (ortho, meta, para) of the benzylidene, which is linked to the proximity of the molecules in the structure type (steric considerations).

By evoking those differences and some steric considerations, we were thus able to suggest a rationalization of the type allocation. According to our analysis, another derivative from this family of compounds would preferentially crystallize in type I or II, with a slight preference for type I.

However, it seems that for some compounds, especially the alkyl/methoxy derivatives, crystallization could reasonably lead to different outcomes. Polymorphism seems thus highly likely in this family of compounds.

Hence, despite many research ongoing in this area and with new analytical tools available, it appears that the rationalization and prediction of structures based on hydrogen-bonding patterns remains very much a challenge.

6. References

(1) Shapiro, H. K. *Am. J. Ther.* 1998, *5*.

(2) Jin, X.-D.; Jin, Y.-H.; Zou, Z.-Y.; Cui, Z.-G.; Wang, H.-B.; Kang, P.-L.; Ge, C.-H.; Li, K. *J. Coord. Chem.* 2011, *64*, 1533–1543.

(3) Shi, L.; Fang, R.-Q.; Zhu, Z.-W.; Yang, Y.; Cheng, K.; Zhong, W.-Q.; Zhu, H.-L. *Eur. J. Med. Chem.* 2010, *45*, 4358–4364.

(4) Villar, R.; Encio, I.; Migliaccio, M.; Gil, M. J.; Martinez-Merino, V. *Bioorg. Med. Chem.* 2004, *12*, 963–968.

(5) Abdel Aziz, A. A.; Salem, A. N. M.; Sayed, M. A.; Aboaly, M. M. *J. Mol. Struct.* 2012, *1010*, 130–138.

(6) Lu, J.; Li, C.; Chai, Y.-F.; Yang, D.-Y.; Sun, C.-R. *Bioorg. Med. Chem. Lett.* 2012, *22*, 5744–5747.

(7) Schmidt, M. F.; El-Dahshan, A.; Keller, S.; Rademann, J. *Angew. Chemie Int. Ed.* 2009, *48*, 6346–6349.

(8) Kargar, H.; Jamshidvand, A.; Fun, H.-K.; Kia, R. *Acta Crystallogr. Sect. E* 2009, *65*, m403–m404.

(9) Yeap, C. S.; Kia, R.; Kargar, H.; Fun, H.-K. *Acta Crystallogr. Sect. E* 2009, *65*, m570–m571.

(10) Nozaki, H.; Takaya, H.; Moriuti, S.; Noyori, R. *Tetrahedron* 1968, *24*, 3655–3669.

(11) Noorduin, W. L.; Izumi, T.; Millemaggi, A.; Leeman, M.; Meekes, H.; Van Enckevort, W. J. P.; Kellogg, R. M.; Kaptein, B.; Vlieg, E.; Blackmond, D. G. *J. Am. Chem. Soc.* 2008, *130*, 1158–1159.

(12) Noorduin, W. L.; Meekes, H.; van Enckevort, W. J. P.; Kaptein, B.; Kellogg, R. M.; Vlieg, E. *Angew. Chem. Int. Ed. Engl.* 2010, *49*, 2539–2541.

(13) Noorduin, W. L.; Meekes, H.; van Enckevort, W. J. P.; Millemaggi, A.; Leeman, M.; Kaptein, B.; Kellogg, R. M.; Vlieg, E. *Angew. Chem. Int. Ed. Engl.* 2008, *47*, 6445–6447.

(14) Noorduin, W. L.; Meekes, H.; Bode, A. a. C.; van Enckevort, W. J. P.; Kaptein, B.; Kellogg, R. M.; Vlieg, E. *Cryst. Growth Des.* 2008, *8*, 1675–1681.

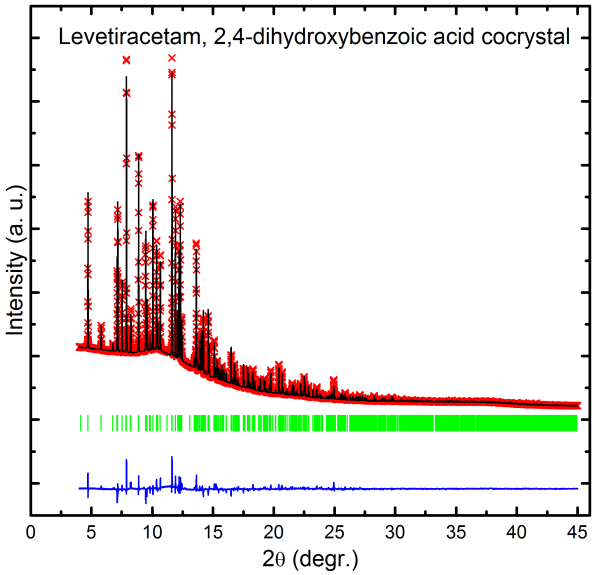
(15) Noorduin, W. L.; van Enckevort, W. J. P.; Meekes, H.; Kaptein, B.; Kellogg, R. M.; Tully, J. C.; McBride, J. M.; Vlieg, E. *Angew. Chem. Int. Ed. Engl.* 2010, *49*, 8435–8438.

(16) Noorduin, W. L.; Van Der Asdonk, P.; Bode, A. A. C.; Meekes, H.; Van Enckevort, W. J. P.; Vlieg, E.; Kaptein, B.; Van Der Meijden, M. W.; Kellogg, R. M.; Deroover, G. *Org. Process Res. Dev.* 2010, *14*, 908–911.

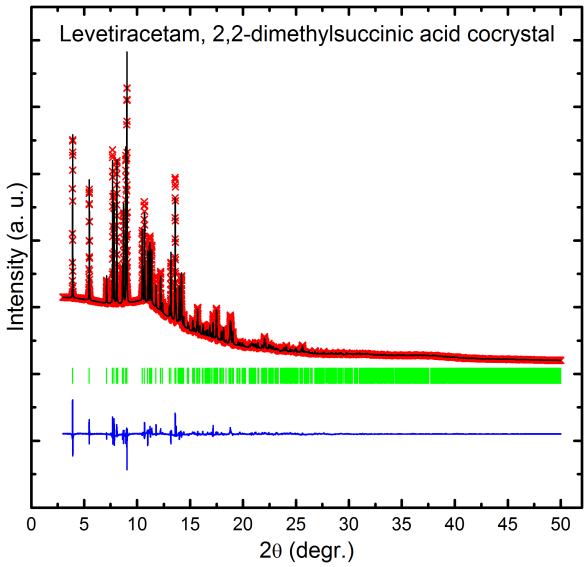
(17) Guo, H.-F.; Pan, Y.; Ma, D.-Y.; Lu, K.; Qin, L. *Transit. Met. Chem.*

2012, *37*, 661–669.
(18) Gül, Z. S.; Ersahin, F.; Agar, E.; Isik, S. *Acta Crystallogr. Sect. E* 2007, *63*, o2902.
(19) Kantar, E. N.; Köysal, Y.; Gümüs, S.; Agar, E.; Soylu, M. S. *Acta Crystallogr. Sect. E* 2012, *68*, o1587.
(20) Kargili, H.; Macit, M.; Alpaslan, G.; Kazak, C.; Erdönmez, A. *Acta Crystallogr. Sect. E* 2012, *68*, o3176.
(21) Pekdemir, M.; Isik, S.; Alaman Agar, A. *Acta Crystallogr. Sect. E* 2012, *68*, o2148.
(22) Vesek, H.; Kazak, C.; Alaman A\ugar, A.; Macit, M.; Soylu, M. S. *Acta Crystallogr. Sect. E* 2012, *68*, o2518.
(23) Kaur, G.; Panini, P.; Chopra, D.; Roy Choudhury, A. *Cryst. Growth Des.* 2012, *12*, 5096–5110.
(24) Cruz-Cabeza, A. J.; Schwalbe, C. H. *New J. Chem.* 2012, *36*, 1347.
(25) Etter, M. C.; MacDonald, J. C.; Bernstein, J. *Acta Crystallogr. Sect. B Struct. Sci.* 1990, *46*, 256–262.
(26) McKinnon, J. J.; Spackman, M. a; Mitchell, A. S. *Novel tools for visualizing and exploring intermolecular interactions in molecular crystals.*; 2004; Vol. 60.
(27) Dalmolen, J.; van der Sluis, M.; Nieuwenhuijzen, J. W.; Meetsma, A.; de Lange, B.; Kaptein, B.; Kellogg, R. M.; Broxterman, Q. B. *European J. Org. Chem.* 2004, *2004*, 1544–1557.
(28) Sheldrick, G. M. *Acta Crystallogr. A.* 2008, *64*, 112–122.
(29) Wood, P. A.; Olsson, T. S. G.; Cole, J. C.; Cottrell, S. J.; Feeder, N.; Galek, P. T. A.; Groom, C. R.; Pidcock, E. *CrystEngComm* 2013, *15*, 65–72.
(30) Wolff, S. K.; Grimwood, D. J.; McKinnon, J. J.; Turner, M. J.; Jayatilaka, D.; Spackman, M. A. CrystalExplorer, 2012.
(31) Noorduin, W. L.; van der Asdonk, P.; Meekes, H.; van Enckevort, W. J. P.; Kaptein, B.; Leeman, M.; Kellogg, R. M.; Vlieg, E. *Angew. Chem. Int. Ed. Engl.* 2009, *48*, 3278–3280.
(32) Leeman, M.; Noorduin, W. L.; Millemaggi, A.; Vlieg, E.; Meekes, H.; van Enckevort, W. J. P.; Kaptein, B.; Kellogg, R. M. *CrystEngComm* 2010, *12*, 2051.
(33) Van Der Meijden, M. W.; Leeman, M.; Gelens, E.; Noorduin, W. L.; Meekes, H.; Van Enckevort, W. J. P.; Kaptein, B.; Vlieg, E.; Kellogg, R. M. *Org. Process Res. Dev.* 2009, *13*, 1195–1198.
(34) Desiraju, G. R.; Murty, B. N.; Kishan, K. V. R. 1990, 447–449.
(35) Lo Presti, L.; Soave, R.; Destro, R. *J. Phys. Chem. B* 2006, *110*, 6405–6414.
(36) Etter, M. C. *J. Phys. Chem.* 1991, *95*, 4601–4610.
(37) Galek, P. T. a; Chisholm, J. a; Pidcock, E.; Wood, P. a. *Acta Crystallogr.*

B. Struct. Sci. Cryst. Eng. Mater. 2014, *70*, 91–105.
(38) Bilton, C.; Allen, F. H.; Shields, G. P.; Howard, J. A. K. *Acta Crystallogr. Sect. B Struct. Sci.* 2000, *56*, 849–856.
(39) Galek, P. T. a.; Fábián, L.; Allen, F. H. *Acta Crystallogr. Sect. B Struct. Sci.* 2010, *66*, 237–252.
(40) Spek, A. L. *Acta Crystallogr. Sect. D Biol. Crystallogr.* 2009, *65*, 148–155.
(41) Taylor, R. *CrystEngComm* 2014, *16*, 6852–6865.
(42) Wood, P. a.; Allen, F. H.; Pidcock, E. *CrystEngComm* 2009, *11*, 1563–1571.



(a)



(b)

Figure S1. Refinement profiles of (a) Levetiracetam- 2,4-dihydroxybenzoic acid and (b) Levetiracetam- 2,2-dimethylsuccinic acid cocrystals. Experimental and calculated data are shown in red symbols and black lines, respectively. Blue line shows the difference line, while green marks show Bragg positions.

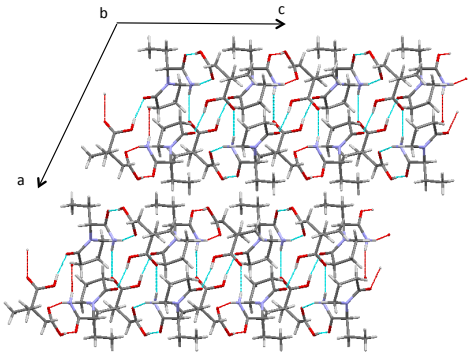


Figure S2. Side view of two layers in ETI-DMSA cocrystal.

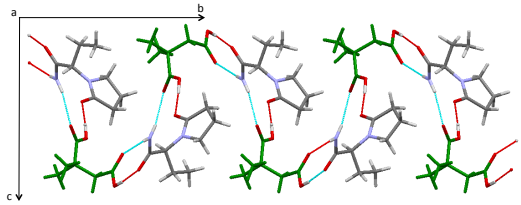


Figure S3. One layer of LEVI-DMSA cocrystal.

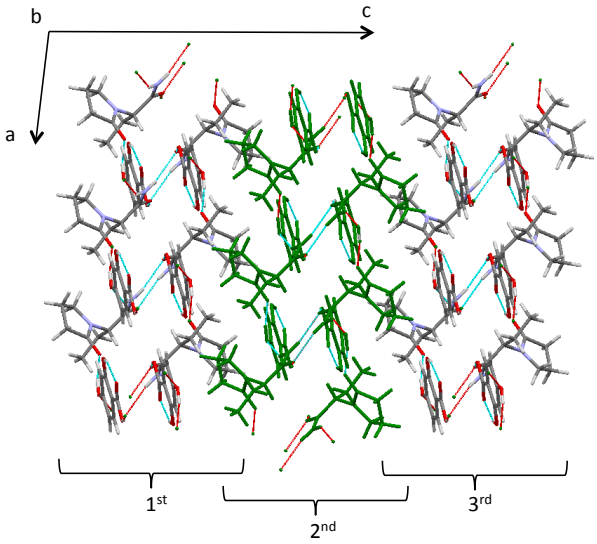


Figure S4. Stacking of three layers in ETI-DHBA cocrystal, showing infinite chain motifs $C^2_2(12)$ along the c axis and π stacking of molecules of 2,4 dihydroxybenzoic acid in the ac plane.

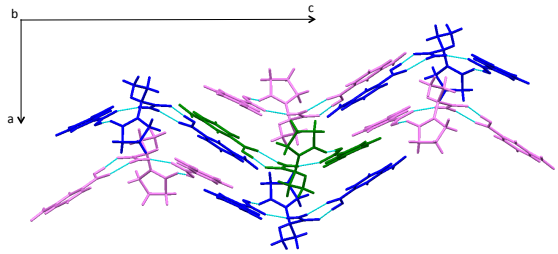


Figure S5. Interlocking of trimers and π stacking of NBZA molecules in LEVI-NBZA cocrystal.

List of achiral coformers used in the Levetiracetam cocrystal screening

Note that from the original list of 152 coformers, 14 coformers are not enumerated here because either they formed an amorphous phase during the screening experiments so we are uncertain about their cocrystallizing ability or they were later found to be duplicates.

1-hydroxy-2-naphtoic acid
2,3-dihydroxybenzoic acid
2,4-dihydroxybenzoic acid
2,5-dihydroxybenzoic acid (gentisic)
3,4-dihydroxybenzoic acid
3,5-dihydroxybenzoic acid
4 hydroxybenzoic acid
salicylic acid
p-coumaric acid
3-hydroxy-2-naphtoic acid
3-hydroxybenzoic acid
3,4,5-trihydroxybenzoic acid
5-hydroxylisophthalic acid
6-hydroxy-2naphtoic acid
Citric acid
4-hydroxy-3-methoxycinnamic acid (Ferulic)
5-chlorosalicylic acid
acetylsalicylic acid
4-aminobenzoic acid
4-dimethylaminobenzoic acid
1H-Pyrazole-3,5-dicarboxylic acid monohydrate

indole-3-acetic acid
pyridine-2,6-dicarboxylic acid
4-nitrobenzoic acid
5-nitroisophtalic acid
3-nitrobenzoic acid
2-ketoglutaric acid
2-hydroxy-1-naphtoic acid
2,2-dimethylsuccinic acid
citraconic acid
fumaric acid
maleic acid
mesaconic acid
oxalic acid
succinic acid
Adipic acid
Benzoic Acid
phtalic acid
Sorbic acid
1,3,5-benzenetricarboxylic acid
isophtalic acid
acide dimethylmalonique
hydroquinone
p-benzoquinone
4,4-bipyridine
terephthalaldehyde
2,2′-Bipyridyl
5,5'-Dimethyl-2,2'-bipyridyl
benzylsulfoxide
phenylsulfoxide
methyl 3,4,5-trihydroxybenzoate
Isobutyl-4hydroxybenzoate
ethyl3,4dihydroxybenzoate
propyl-p-hydroxybenzoate
Gallic acid ethyl ester
trimellitic acid
Methyl 4-hydroxybenzoate
Ethyl-4hydroxybenzoate

n-butyl 4 hydroxybenzoate
Malonamide
Oxamide
Acrylamide
Benzamide
Acetanilide
N-tert-Butylacrylamide
2-Acetamidofluorene
Methacrylamide
Acetamide
Fumaramide
3-methylbutanamide
Propionamide
2-Phenylbutyramide
Glycine anhydre
Piracetam
nefiracetam
isonicotinamide
nicotinamide
4-Aminobenzamide
3-Aminobenzamide
Anthranilamide
3'-Aminoacetanilide
3-Amino-4-methylbenzamide
2-(diethylamino)-N-(2,6-dimethylphenyl)-acetamide
Pyrazinamide
N,N-diethylnicotinamide
phenacetin
2-ethoxybenzamide
4'-ethoxyacetanilide
1-ethoxybenzamide
p-Acetanisidide
Acetaminophen (paracetamol)
Salicylamide
Acetoacetamide
pramiracetam
4-nitrobenzamide

2-Nitrobenzamide
3-Nitrobenzamide
5-fluorouracil
2-Chloroacetamide
4-Fluorobenzamide
2-Fluorobenzamide
4'-Fluoro-2'-nitroacetanilide
2'-Chloroacetanilide
3'-Fluoroacetanilide
3-chloro-4-fluorobenzamide
2-Cyanoacetamide
thioacetamide
4-chlorothiobenzamide
Thiobenzamide
bisurea
carbamazepine
urea
Oxcarbazepine
4-chlorobenzaldéhyde
3-nitrobenzaldehyde
4-nitrobenzaldehyde
4-bromobenzaldehyde
2-nitrobenzaldehyde
Biphenyl-4-carboxaldehyde
Caffeine
succinimide
vanillin
Theophylline
7-(2-hydroxyethyl)theophylline
Theophylline-7-acetic acid
ethyl 7-theophyllineacetate
7-(2-hydroxypropyl)theophylline
o-benzoic sulfuimide
ellagic acid hydrate
xanthine
Sulfathiazole
sulfacetamide

5,5-Dimethylhydantoin
2-(4-hydroxyphenyl)acetamide
N1-[4-bromo-2-(trifluoromethoxy)phenyl]acetamide
N'-Hydroxy-3-methyl-2-pyridinecarboximidamide
Methyl 3,4,5 trimethoxybenzoate
Aniracetam

Supporting information of Chapter 4

1. HBP-CL models

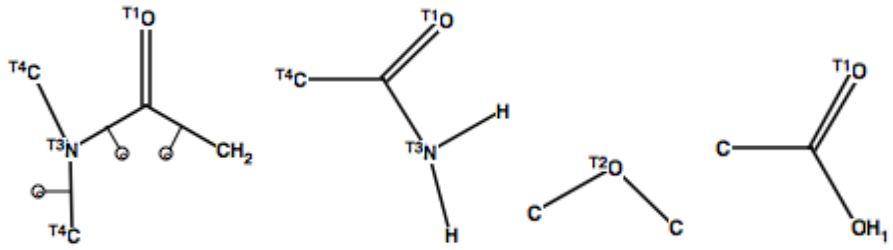
Here are some computational details concerning the HBP-CL models corresponding to the most stable polymorph for all the following compounds. Stability was evaluated taking into account the observed NC Score of each form.

1.1 One molecule models

1.1.A Active pharmaceutical ingredients

Levetiracetam (OMIVUB)

Functional groups used in the model:
- AmIII_cycl_2CT4
- Carbamoyl_C4_NH2T3
- ether
- carboxylic_acid

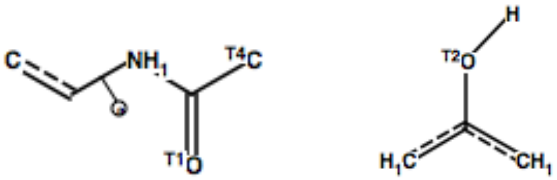


Total number of structures: 1106
AUC: 81%

Paracetamol

– hxacan11 (F1)

Functional groups used in the model:
ar_al_trans_amide
ar_hydroxy_2H_OHT2
carboxylic_acid



Total number of structures: 1057
AUC: 83,6%

– hxacan33 (F2)

Functional groups used in the model:
ar_al_trans_amide
ar_hydroxy_2H_OHT2
carboxylic_acid

Total number of structures: 1057
AUC: 83,7%

– hxacan29 (F3)

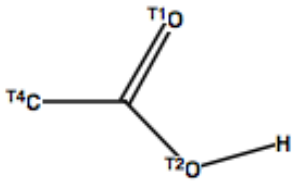
Functional groups used in the model:
trans_amide_CH3
hydroxy_2H_OHT2
carboxylic_acid

Total number of structures: 1056
AUC: 83,7%

1.1.B Cocrystallizing coformers

DMSA (OLENIC)

Functional groups used in the model:
- al_cooh_OHT2
- ether



Total number of structures: 1239
AUC: 86,16%

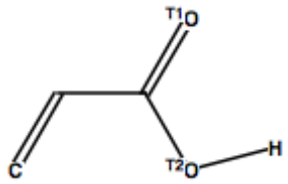
DMMA (MMALAC01)

Functional groups used in the model:
- al_cooh_OHT2
- ether

Total number of structures: 1239
AUC: 82,2%

CCA (NEDNOZ)

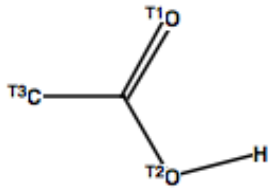
Functional groups used in the model:
- cjg_cooh_OHT2
- ether



Total number of structures: 1213
AUC: 85,5%

OXA (OXALAC07)

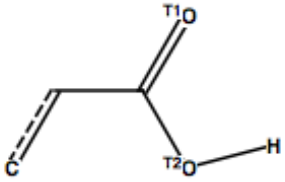
Functional groups used in the model:
- al_cooh_CT3_OHT2
- ether

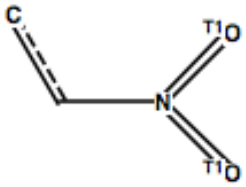


Total number of structures: 1239
AUC: 82,5%

4NBA (NBZOAC02)

Functional groups used in the model:
- ar_cooh_OHT2
- ar_nitro
- ether





Total number of structures: 1289
AUC: 84%

NIA (COFDUW10)

Functional groups used in the model:
- ar_cooh_OHT2
- ar_nitro

Total number of structures: 1160
AUC: 85,5%

3NBA (MNBZAC)

Functional groups used in the model:
- ar_cooh_OHT2
- ar_nitro

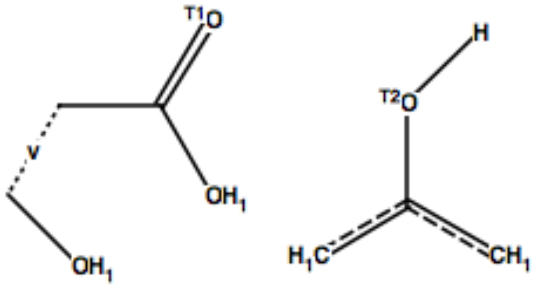
Total number of structures: 1520
AUC: 85%

DHBA (ZZZEEU07)

The two hydroxyl groups present on this molecule are expected to show different behaviours as one is most probably involved in an intramolecular H-bond and not the other one. To take this into account, both the carboxylic acid and the hydroxyl group adjacent to it were included in the same group definition. Indeed it was expected that their combination will have a very specific and robust behavior in the dataset.

Then, to differentiate the two hydroxyl groups, it was specified in the second definition that there are two hydrogen atoms in ortho positions of the non-adjacent hydroxyl group on the aromatic ring.

Functional groups used in the model:
- ar_cjg_cooh_OHadj
- ar_hydroxy_2H_OHT2



Total number of structures: 774
AUC: 92,5% (outstanding discrimination)

1.1.C Non-cocrystallizing coformers

MLA (MALIAC02)

Functional groups used in the model:
- cjg_cooh_OHT2
- ether

Total number of structures: 1212
AUC: 85,5% (excellent discrimination)

PTA (PHTHAC01)

Functional groups used in the model:
- ar_cooh_OHT2
- ether

Total number of structures: 1199
AUC: 83,8%

SLA (SALIAC16)

Functional groups used in the model:
- ar_cjg_cooh_OHadj

Total number of structures: 407
AUC: 94,3%

HBA (BIDLOP02)

Functional groups used in the model:
- ar_cooh_OHT2
- ar_hydroxy_OHT2

Total number of structures: 1232
AUC: 88,2%

ABA (AMBNAC04)

Functional groups used in the model:
- ar_cooh_OHT2
- ar_hydroxy_OHT2
- prim_amine

Total number of structures: 1379
AUC: 85,5%

FRA (GASVOL)

Functional groups used in the model:
- cjg_cooh_OHT2
- ar_hydroxy_OHT2
- ar_methoxy

Total number of structures: 1621
AUC: 84,8%

MSA (MESCON)

Functional groups used in the model:
- cjg_cooh_OHT2
- ether

Total number of structures: 1170
AUC: 85%

1.2 Two molecules models

Pour les modèles des cocristaux/mix, les fonctions utilisées pour décrire les groupes fonctionnels du Levi sont ceux qui ont été choisis pour le modèle du Levi seul.

1.2.A Cocrystals

Levi - DMSA (XOGMOX)

Functional groups used in the model:
- AmIII_cycl_2CT4
- Carbamoyl_C4_NH2T3
- al_cooh_OHT2
- ether

Total number of structures: 1162
AUC: 83,5%

Levi - DMMA

Functional groups used in the model:
- AmIII_cycl_2CT4
- Carbamoyl_C4_NH2T3
- al_cooh_OHT2
- ether

Total number of structures: 1361
AUC: 83,6%

Levi -CCA

Functional groups used in the model:
- AmIII_cycl_2CT4
- Carbamoyl_C4_NH2T3
- cjg_cooh_OHT2
- ether

Total number of structures: 1205
AUC: 84,2%

Levi - OXA (XOGPEQ)

Functional groups used in the model:
- AmIII_cycl_2CT4
- Carbamoyl_C4_NH2T3
- al_cooh_CT3_OHT2
- ether

Total number of structures: 1380
AUC: 83,6%

Levi -OXA (sketched)

Functional groups used in the model:
- AmIII_cycl_2CT4
- Carbamoyl_C4_NH2T3
- al_cooh_CT3_OHT2
- ether

Total number of structures: 1378
AUC: 83,1%

Levi -4NBA_1:2 (XOGNUE)

Functional groups used in the model:
- AmIII_cycl_2CT4
- Carbamoyl_C4_NH2T3
- ar_cooh_OHT2
- ar_nitro

Total number of structures: 1453 ; AUC: 83,9%

Levi -NIA

Functional groups used in the model:
- AmIII_cycl_2CT4
- Carbamoyl_C4_NH2T3
- ar_cooh_OHT2
- ar_nitro

Total number of structures: 1391
AUC: 83,9%

Levi -3NBA_1:2 (XOGNIS)

Functional groups used in the model:
- AmIII_cycl_2CT4
- Carbamoyl_C4_NH2T3
- ar_cooh_OHT2
- ar_nitro
- ether

Total number of structures: 1520
AUC: 83,9%

Levi -DHBA (YASGAC01)

Functional groups used in the model:
- AmIII_cycl_2CT4
- Carbamoyl_C4_NH2T3
- ar_cjg_cooh_OHadj
- ar_hydroxy_2H_OHT2

Total number of structures: 1512
AUC: 90,8%

1.2.B Levi-non cocrystallizing coformers mix (sketched)

Levi -MLA

Functional groups used in the model:
- AmIII_cycl_2CT4
- Carbamoyl_C4_NH2T3

- cjg_cooh_OHT2
- ether

Total number of structures: 1462
AUC: 83,4%

Levi -PTA

Functional groups used in the model:
- AmIII_cycl_2CT4
- Carbamoyl_C4_NH2T3
- ar_cooh_OHT2
- ether

Total number of structures: 1453
AUC: 84,1%

Levi -SLA

Functional groups used in the model:
- AmIII_cycl_2CT4
- Carbamoyl_C4_NH2T3
- ar_cjg_cooh_OHadj

Total number of structures: 1315
AUC: 89,76%

Levi -HBA

Functional groups used in the model:
- AmIII_cycl_2CT4
- Carbamoyl_C4_NH2T3
- ar_cooh_OHT2
- ar_hydroxy_OHT2

Total number of structures: 1334
AUC: 84,8%

Levi -ABA

Functional groups used in the model:
- AmIII_cycl_2CT4

- Carbamoyl_C4_NH2T3
- ar_cooh_OHT2
- prim_amine

Total number of structures: 1379
AUC: 84,9%

Levi - FRA

Functional groups used in the model:
- AmIII_cycl_2CT4
- Carbamoyl_C4_NH2T3
- cjg_cooh_OHT2
- ar_hydroxy_OHT2
- ar_methoxy

Total number of structures: 1623
AUC: 84,3%

Levi - MSA

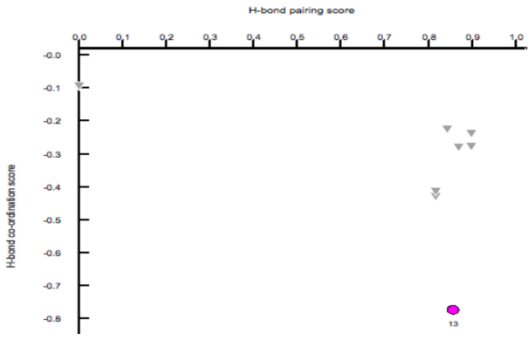
Functional groups used in the model:
- AmIII_cycl_2CT4
- Carbamoyl_C4_NH2T3
- cjg_cooh_OHT2
- ether

Total number of structures: 1456
AUC: 82,8%

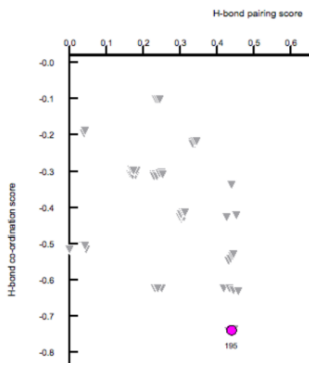
2. Solid-form landscapes

Unfortunately, as the number of feasible H-bonding sets is related to the upper coordination limits for each donor/acceptor atom, the software is unable to generate this chart for some crystal structures with more than one symmetry-irreducible molecule.

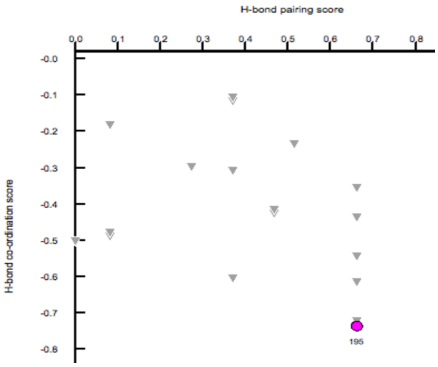
2.1 Solid-form landscapes of Levetiracetam and its cocrystallizing coformers



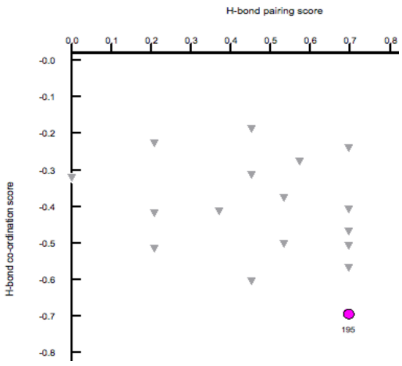
Levetiracetam



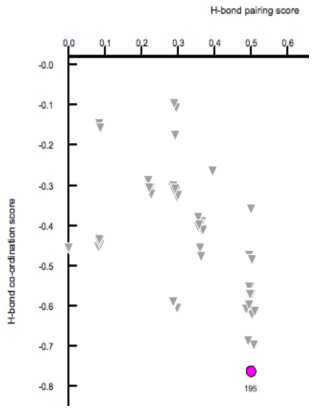
DMSA



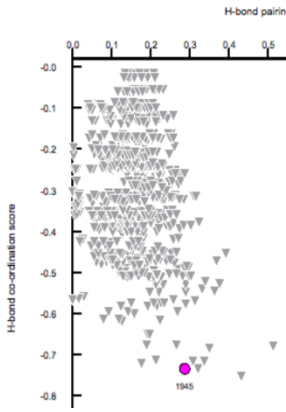
DMMA



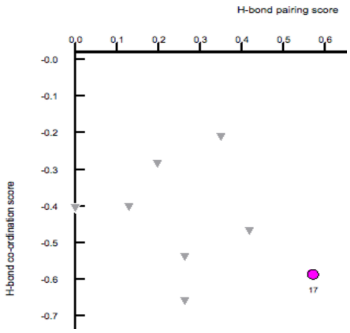
OXA



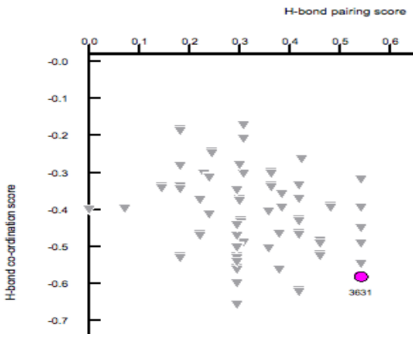
CCA



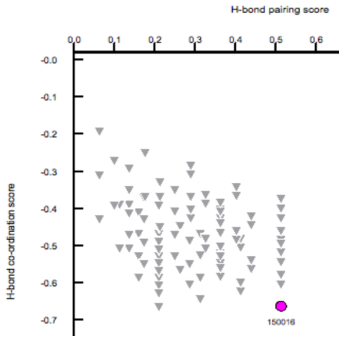
DHBA



4NBA

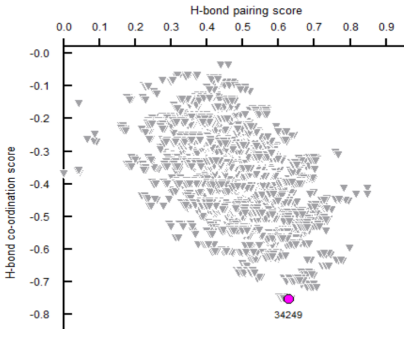


NIA

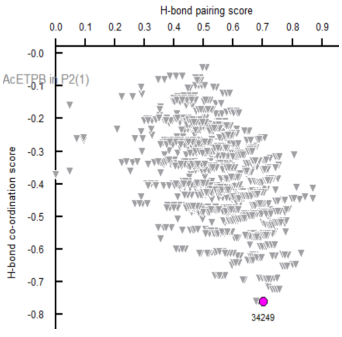


3NBA

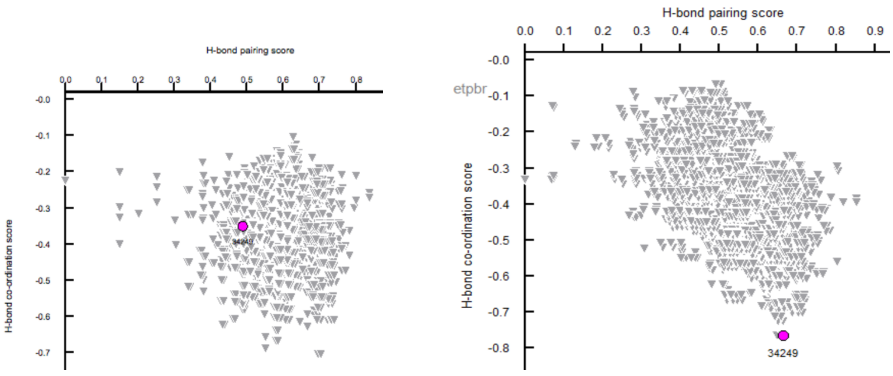
2.2 Solid-form landscapes of the Levetiracetam cocrystals



Levi - DMSA

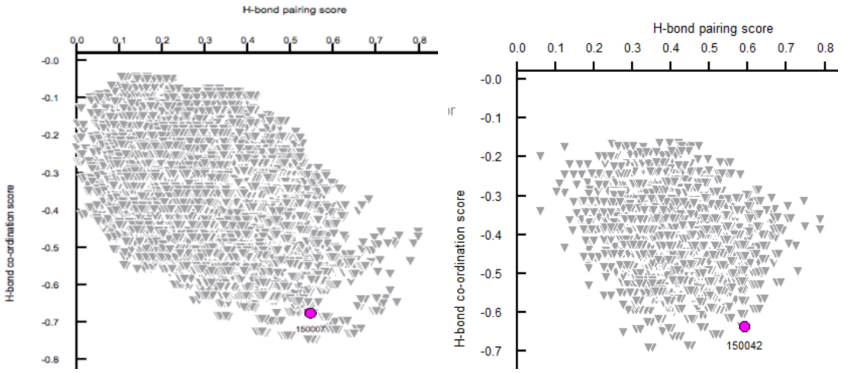


Levi -DMMA



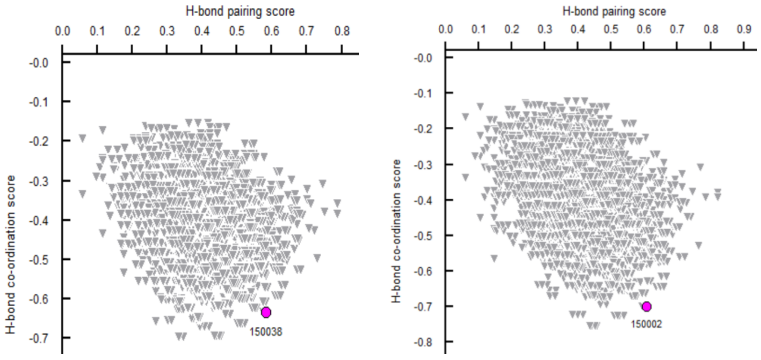
Levi - OXA

Levi - CCA



Levi - DHBA

Levi - 4NBA

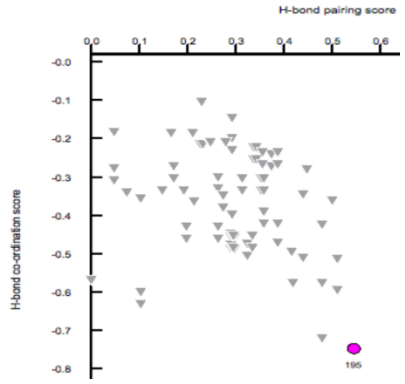


Levi - 3NBA

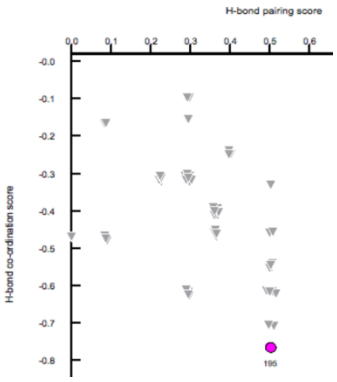
Levi -NIA

Note that the pink point on the Levi - OXA landscape does not correspond to the actual cocrystal structure as the chart was obtained from a model based on a 2D representation rather than on the imported structure (as this latter contains more than 2 independent molecules in the unit cell).

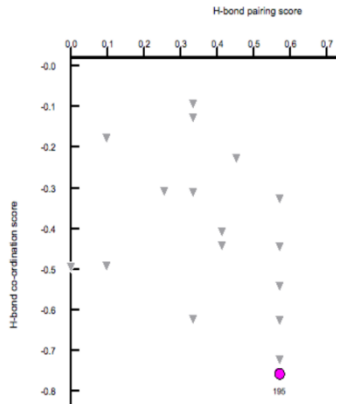
2.3 Solid-form landscapes of some non-cocrystallizing coformers of Levetiracetam



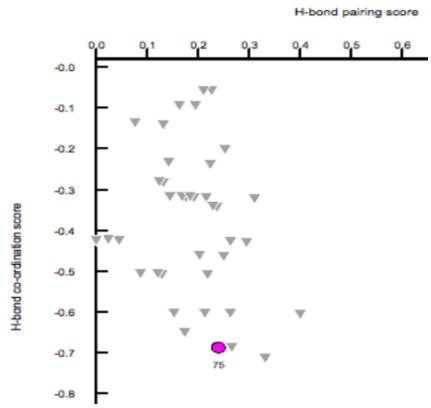
MLA



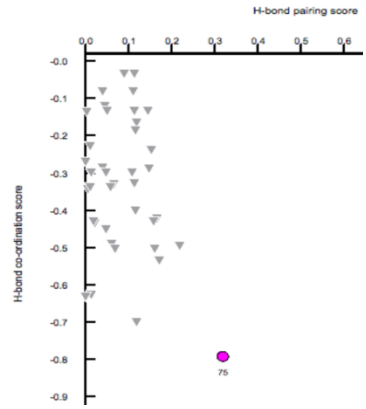
MSA



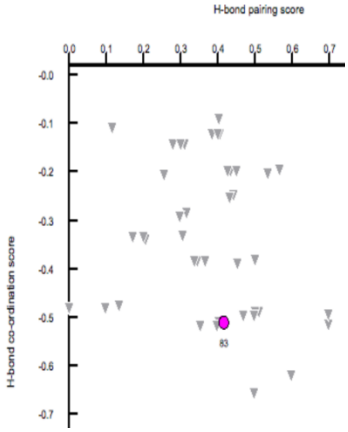
PTA



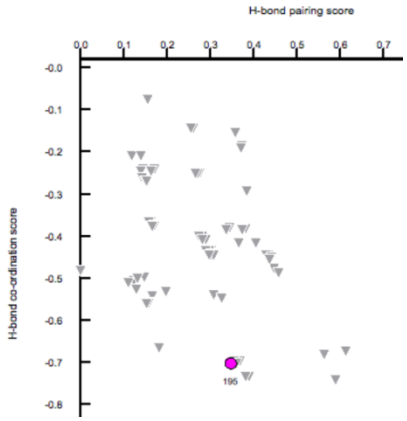
HBA



SLA



ABA


H-bond pairing score
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
H-bond co-ordination score
-0.0
-0.1
-0.2
-0.3
-0.4
-0.5
-0.6
-0.7
-0.8
195
FRA