

# A - PRESENTATION DU PROJET DOCTORAL

## TITRE DU PROJET

PRÉDICTION DE L'ASSEMBLAGE MOLÉCULAIRE DANS DES FORMES CRISTALLINES DES PRINCIPES ACTIFS PHARMACEUTIQUES (API) A PARTIR DE MODÈLES D'IA - MACHINE LEARNING

## DESCRIPTION DU PROJET

### Contexte

Au cours des dernières années, l'intelligence artificielle (IA) et le Machine Learning se sont imposés comme des outils majeurs pour l'exploration de l'espace chimique et la prédiction de propriétés moléculaires complexes. En établissant des relations non linéaires entre structure, propriétés électroniques et interactions intermoléculaires, ces approches ouvrent des perspectives nouvelles pour modéliser des phénomènes difficiles à traiter par des méthodes purement expérimentales ou théoriques, en particulier lorsqu'il s'agit de prédire l'autoassemblage et l'organisation tridimensionnelle des molécules à l'état solide.

Les formes galéniques des ingrédients pharmaceutiques actifs (*Active Pharmaceutical Ingredients*, API) sont souvent solides puisqu'elles sont chimiquement stables et faciles à administrer et à produire, tout en offrant un dosage précis.<sup>1</sup> Les médicaments sous forme solide peuvent exister soit à l'état cristallin soit à l'état amorphe, avec une préférence pour le premier vu l'instabilité de nombreux matériaux amorphes.<sup>2</sup> Un API peut être cristallisé seul, en tant qu'un composant unique, ou avec d'autres molécules pour former des cristaux à plusieurs composants, tels que des hydrates, des solvates, des sels ou des co-cristaux. Pour chacune de ces phases, ainsi que pour leurs différentes formes polymorphiques qui peuvent exister, les divers cristaux d'un même API auront des propriétés physico-chimiques différentes de celles des autres formes, ce qui peut affecter la stabilité et la biodisponibilité du médicament.<sup>3-5</sup> Une telle diversité offre la possibilité de contrôler les propriétés physico-chimiques de l'API en conservant l'activité thérapeutique du principe actif.

Le co-cristal est un solide cristallin homogène qui se forme par l'association de deux ou plusieurs molécules électriquement neutres (Fig. 1 gauche), qui sont présentes en quantités stœchiométriques précises à l'état solide et à température ambiante.<sup>6</sup> Les composants d'un co-cristal sont maintenus ensemble dans la phase cristalline par des interactions intermoléculaires (non covalentes) comme les liaisons hydrogène, halogène, chalcogène, ou encore les interactions  $\pi \cdots \pi$ .<sup>7</sup> Il est possible d'obtenir des hydrates ou des solvates de co-cristaux si des molécules de solvants s'introduisent dans leurs structures cristallines (Fig. 1 gauche).

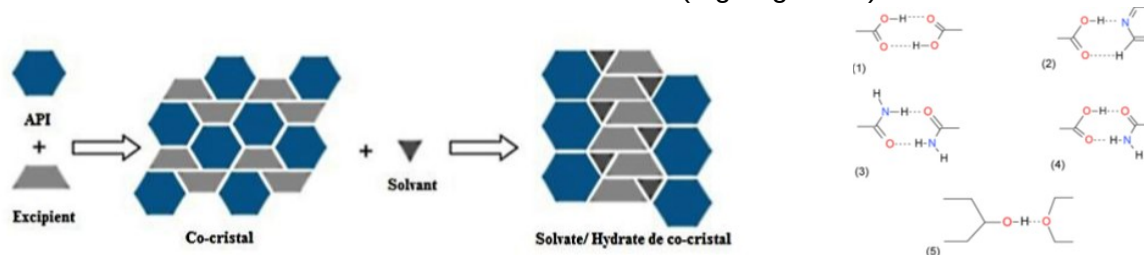


Fig. 1. (gauche) Formation des solvates/hydrates de co-cristaux. (droite) 5 synthons observés dans des co-cristaux organiques.

Les co-cristaux possèdent un champ d'applications très vaste,<sup>8</sup> cependant les applications les plus intéressantes touchent le domaine pharmaceutique.<sup>3,9</sup> La solubilité des médicaments est l'une des propriétés biopharmaceutiques les plus importantes, et son amélioration est actuellement l'un des principaux défis de l'industrie pharmaceutique. A ce titre, la formation de co-cristaux d'APIs représente l'une des méthodes utilisées pour améliorer la biodisponibilité des médicaments possédant des APIs avec de faibles solubilités.<sup>10</sup> Par ailleurs, en plus des améliorations des propriétés de solubilité, de taux de dissolution et de biodisponibilité, la formation des co-cristaux permet aussi l'amélioration d'autres propriétés des API, telles que la stabilité, la compressibilité, l'écoulement, le point de fusion et l'hygroscopicité.<sup>11</sup> Dans les co-cristaux pharmaceutiques l'un des composants est un ingrédient pharmaceutique actif (API) et les autres composants sont des excipients, qui favorisent la prise du médicament mais n'ont pas de réelles fonctions thérapeutiques. Les molécules qui peuvent être utilisées comme excipients sont regroupées dans une liste appelée GRAS (*Generally Recognized As Safe*) qui est publiée par l'administration américaine *Food & Drug Administration* (<https://www.fda.gov/>).

La liaison hydrogène est l'interaction intermoléculaire la plus couramment observée dans le cas des cristaux de composés organiques comportant des atomes d'hydrogène à caractère acide. L'utilisation des règles qui concernent la formation des liaisons hydrogène et des synthons moléculaires, qui sont définis comme des motifs structuraux intermoléculaires récurrents dans des phases cristallines, peut aider à la conception et à l'analyse des co-cristaux des composés organiques (Fig. 1 droite). Des études sur la compétition entre des différents synthons supramoléculaires ont été réalisées pour faciliter la prédiction de la formation des co-cristaux.<sup>12</sup> Les résultats de ces études suggèrent que le synthon énergétiquement le plus stable est celui figurant dans l'arrangement du co-cristal formé.<sup>9</sup> Ces hypothèses ont été vérifiées expérimentalement grâce à des comparaisons entre des structures monocristallines formées.<sup>13</sup> La complémentarité entre groupements fonctionnels moléculaires et leurs synthons supramoléculaires donnent les conditions préalables à la conception des co-cristaux. Elles facilitent la sélection de molécules appropriées participant à la formation des co-cristaux.

Le solide cristallin d'un API, qu'il soit formé par un ou plusieurs composants, peut présenter plusieurs formes qui sont associées au phénomène du polymorphisme (Fig. 2). Les différentes formes polymorphiques possèdent différentes propriétés physico-chimiques, puisque celles-ci sont dépendantes de la structure à l'état solide, laquelle varie selon le polymorphe formé. Ainsi, en agissant sur la biodisponibilité du principe actif, les diverses structures polymorphes d'un API se manifestent avec des différences (parfois notables) dans l'efficacité du médicament.<sup>14</sup> Pour un cristal moléculaire d'un API, il peut exister un grand nombre de polymorphes et il est extrêmement difficile de savoir le nombre de formes polymorphiques qu'un composé chimique peut former. Dans le domaine pharmaceutique, le polymorphisme touche plus de % des molécules de principes actifs<sup>15</sup> et environ 30 à % des composés pharmaceutiques les plus vendus dans le marché sont polymorphiques<sup>16</sup>. De ce fait, les enjeux du polymorphisme sont très importants. Il faut signaler que, malgré les différences significatives dans les interactions intermoléculaires, la conformation, la position des molécules dans l'espace et les différences dans les structures cristallines des polymorphes d'un même composé, leur différence d'énergie est généralement très faible, typiquement 2 à 16 kJ·mol<sup>-1</sup>. Des études statistiques ont montré<sup>17,18</sup> que 85% des polymorphes connus présentent une différence d'énergie inférieure à 10% avec une limite supérieure d'énergie de 20 kJ·mol<sup>-1</sup> ce qui souligne la difficulté à les identifier, d'autant plus que les formes cristallisées ne sont pas nécessairement celles de plus basse énergie ou plus stables.

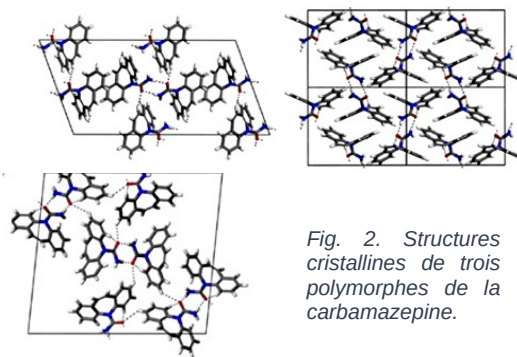


Fig. 2. Structures cristallines de trois polymorphes de la carbamazépine.

80  
50

## Objectifs

L'un des aspects les plus importants des diverses formes cristallines des APIs et de leurs polymorphes repose sur le fait qu'ils peuvent améliorer les propriétés du principe actif sans changer son activité biologique intrinsèque. La reconnaissance moléculaire, puis l'autoassemblage entre les molécules formant le cristal, se fait via des interactions intermoléculaires, souvent réalisées au sein de synthons. Ainsi, les excipients les plus utilisés pour la formation de co-cristaux pharmaceutiques possèdent des groupements fonctionnels capables de construire les synthons avec les APIs. Cependant, malgré l'utilisation de synthons et des règles phénoménologiques guidant la synthèse supramoléculaire des co-cristaux, leur formation ne peut pas être prédite. En effet, la problématique étant liée à l'assemblage 3D qui est nécessaire pour former un cristal, l'utilisation des groupements moléculaires complémentaires pouvant former un synthon *ne permet de prédire qu'une partie de l'assemblage global du cristal moléculaire*. En plus, comme il a été indiqué, *il est très difficile aussi d'identifier l'existence de formes polymorphiques à partir de calculs d'énergie*.

Par ailleurs, très récemment, et en collaboration avec le Dr. Sabeur Aridhi (LORIA, Laboratoire Lorrain de Recherche en Informatique et ses Applications), nous avons développé des modèles de IA – *Machine Learning* pour la prédiction de propriétés moléculaires (charges atomiques et énergie totale moléculaire). Pour ce faire, nous avons généré une base de données SQL avec plus de 455000 entrées issues des calculs quantiques DFT de la fonction d'onde moléculaire (B3LYP/Def2TZVPP) qui ont été obtenues avec plus de 16700 molécules organiques issues de la base de données *Cambridge Structural Database* (<https://www.ccdc.cam.ac.uk/>) possédant les atomes H, C, N, O, F, P, S, Cl, Se et Br et leurs types chimiques étendus. Après l'entraînement,

le modèle est capable de prédire les charges atomiques<sup>19</sup> avec une déviation standard < 0.01e et l'énergie totale des molécules avec une déviation standard de 0.005 Hartree (13 kJ/mol) qui devrait encore s'améliorer avec une exploration conformationnelle des molécules. Ces recherches vont donner lieu à une publication scientifique qui est en phase finale de rédaction.<sup>20</sup>

Dans ce contexte, et à l'aide de méthodes IA - *Machine Learning*, les objectifs de ce projet de thèse de doctorat sont :

- **Apprendre des représentations moléculaires informatives** (descripteurs / *embeddings*) à partir de données quantiques (charges atomiques, densités, énergies), permettant de décrire finement les interactions intermoléculaires
- **Prédire les propriétés moléculaires** qui décrivent les interactions intermoléculaires
- **Prédire l'assemblage moléculaire 3D d'un principe actif (API)** (cristaux, co-cristaux et polymorphes) en utilisant les descripteurs moléculaires prédits précédemment

### Déroulé du projet doctoral

Le déroulement de ce projet implique le développement d'une méthodologie en trois stades : (i) prédiction des propriétés atomiques et moléculaires en phase gaz, (ii) prédiction des assemblages moléculaires (orientations et intensités), notamment dans une ou plusieurs couches 3D autour d'une molécule centrale, et (iii) à partir des assemblages obtenus en (ii), déterminer les conditions de périodicité sur les descripteurs électroniques obtenus pour pouvoir former un cristal.

Ce projet scientifique présente un fort caractère innovateur de par sa méthodologie avec l'utilisation de modèles d'IA-*Machine Learning* pour la prédiction des assemblages moléculaires, en particulier des APIs. De ce fait, des compétences dans le domaine de l'IA – *Machine Learning* sont un **prérequis** important pour candidater, alors que les compétences en calculs de chimie quantique et en modélisation moléculaire pourront être acquises pendant le déroulement des travaux de thèse. Le développement informatique et des modèles d'IA pour la prédiction des assemblages moléculaires sera réalisé en collaboration avec le Pr. Sabeur Aridhi (LORIA). Les moyens de calcul locaux (LORIA, CRM2) et régionaux (mésocentre Explor) seront utilisés.

### Références

- <sup>1</sup> Hilfiker, R. (2006). *Polymorphism: In the Pharmaceutical Industry*. Wiley-VCH, Germany.
- <sup>2</sup> Vippagunta, S. R., Brittain, H. G. & Grant, D. J. W. (2001). *Adv. Drug Delivery Rev.*, 48, 3-26.
- <sup>3</sup> Shan, N. & Zaworotko, M.J. (2008). *Drug Discov. Today*, 13, 440–446.
- <sup>4</sup> Byrn, S. R., Pfeiffer, R., Stephenson, G. A., Grant, D. J. W. & Gleason, W. (1994). *Chem. Mater.* 6, 1148-1158.
- <sup>5</sup> Byrn, S. R., Pfeiffer, R. R., Ganey, M., Hoiberg, C. & Poochikian, G. (1995). *Pharmaceut. Res.* 12, 945-954.
- <sup>6</sup> Aakeroy, C. B. & Salmon, D. J. (2005). *CrystEngComm*, 7, 439–448.
- <sup>7</sup> Aakeroy, C. B. (1997). *Acta Cryst.* B53, 569–586.
- <sup>8</sup> Steed, J. W. (2013). *Trends Pharmacol. Sci.* 34, 185–193.
- <sup>9</sup> Blagden, N., Berry, D. J., Parkin, A., Javed, H., Ibrahim, A., Gavan, P. T., De Matos, L. L. & Seaton, C. C. (2008). *New J. Chem.* 32, 1659–1672
- <sup>10</sup> Qiao, N., Li, M., Schlindwein, W., Malek, N., Davies, A., Trappitt, G. (2011). *Int. J. Pharm.* 419, 1–11.
- <sup>11</sup> Lu, J. & Rohani, S. (2009). *Org. Process Res. Dev.* 13, 1269–1275.
- <sup>12</sup> Etter, M. C. (1991). *J. Phys. Chem.* 95, 4601–4610.
- <sup>13</sup> Vishweshwar, P., McMahon, J. A., Bis, J. A. & Zaworotko, M. J. (2006). *J. Pharm. Sci.* 95, 499–516.
- <sup>14</sup> Brittain, H.G. & Grant, D.J.W. (1999). Effect of polymorphism and solid- state solvation on solubility and dissolution rate, in: H.G. Brittain (Ed.), *Polymorphism in Pharmaceutical Solids*. Marcel Dekker, New York.
- <sup>15</sup> Giron. D. (1995). *Thermochim. Acta.* 248, 1-59.
- <sup>16</sup> Threlfall, T. L. (1995). *Analyst.* 120, 2435-2460.
- <sup>17</sup> Gavezzotti, A. & Filippini, G. (1995). *J. Am. Chem. Soc.* 117, 12299-12305.
- <sup>18</sup> Gavezzotti, A. (2007). *Molecular Aggregation*. Oxford University Press, New York.
- <sup>19</sup> Bader, R. F. W. (1990). *Atoms in Molecules – A Quantum Theory*. Clarendon: Oxford, U.K.
- <sup>20</sup> Autef, L., Aubert, E., Aridhi, S. & Espinosa, E. (2026). *CrystEngComm* (in preparation).

## B – UNITES DE RECHERCHE IMPLIQUEES – EQUIPES D'ACCUEIL

### Unités de recherche impliquées :

- Laboratoire de Cristallographie, Résonance Magnétique et Modélisations (CRM2), UMR UL-CNRS 7036.

- Laboratoire Lorrain de Recherche en Informatique et ses Applications (LORIA), UMR UL-Lorraine INP 7503.

#### Equipes d'accueil :

- Biologie structurale, Modélisation des Interactions Moléculaires, Ingénierie Cristalline (BioMIMIC, CRM2)
- Artificial Intelligence and Big data (LORIA – TELECOM Nancy)

## C – CONDITIONS SCIENTIFIQUES MATERIELLES ET FINANCIERES DU PROJET DE RECHERCHE

Dans les deux laboratoires, nous disposons des moyens de calcul (puissance et stockage dans nos serveurs, ainsi que l'accès aux centres de calcul régional (MESOCENTRE) ou national pour les calculs plus lourds) et d'infrastructure (bureau, ordinateurs, etc) nécessaires pour accueillir et accompagner la personne qui sera recrutée sur ce sujet de recherche.

Le financement du projet de recherche est de l'Université de Lorraine traduisant son choix d'appuyer sa recherche par l'octroi de contrats doctoraux.

## D - ENCADREMENT DU PROJET DOCTORAL

Les travaux de recherche de cette thèse interdisciplinaire sera réalisée au sein des laboratoires CRM2 et LORIA, avec l'encadrement de :

Pr. Enrique Espinosa (CRM2) : [enrique.espinosa@univ-lorraine.fr](mailto:enrique.espinosa@univ-lorraine.fr)

Dr. Emmanuel Aubert (CRM2) : [emmanuel.aubert@univ-lorraine.fr](mailto:emmanuel.aubert@univ-lorraine.fr)

Pr. Sabeur Aridhi (LORIA) : [sabeur.aridhi@loria.fr](mailto:sabeur.aridhi@loria.fr)

## E – PROFIL ET COMPETENCES RECHERCHEES

Titulaire d'un Master (ou équivalent) en IA, Data Science, informatique, Chimie ou Physique, le/la candidat(e) possède de bonnes compétences en *Machine Learning* et en programmation scientifique (Python). Un intérêt pour les approches interdisciplinaires à l'interface entre IA et modélisation moléculaire est essentiel. Des connaissances en chimie quantique ou cristallographie seront appréciées mais non obligatoires.

Profil recherché :

- ✓ Master en IA / Data Science / Informatique / Chimie / Physique
- ✓ Bon niveau en *Machine Learning* / *Deep Learning* (PyTorch, scikit-learn...)
- ✓ Intérêt pour la santé numérique et les données multimodales

Environnement : recherche appliquée, collaboration avec informaticiens – chimistes quantiques - cristallographes, données réelles, publications internationales.

## F – CANDIDATURE

Envoyer les documents suivants aux courriers électroniques des trois encadrants (voir ci-dessus section D) :

- CV
- Lettre de motivation
- Relevés de notes de M1 et M2 (notes de stages M1 et M2 exclues)