

# Génération automatique et exhaustive de structures moléculaires bi- et tri-dimensionnelles à l'aide d'outils issus de l'intelligence artificielle

## 1 Contexte

Le projet de thèse proposé vise à développer des outils de design moléculaire en tirant parti de techniques issues de l'intelligence artificielle, et notamment de la programmation par contraintes, pour générer automatiquement des structures moléculaires bi- et tri-dimensionnelles. Ce projet interdisciplinaire est à la croisée de la chimie, de l'informatique et de la science des matériaux, et permettra de lever des verrous importants en matière de conception moléculaire, d'exploration de structures et de découverte de matériaux.

Ce projet s'inscrit dans la continuité d'une collaboration existante entre une équipe d'informaticiens du LIS<sup>1</sup> (N. Prcovic, C. Terrioux, P. Jégou) et une équipe de chimistes théoriciens de l'iSm2<sup>2</sup> (Y. Carissan et D. Hagebaum-Reignier). Cette collaboration fructueuse a démarré en 2019 grâce à une thèse inter-ED [1] portant sur l'étude de l'aromaticité des benzénoïdes et a donné lieu à 5 publications, 4 en informatique [2, 3, 4, 5] et 1 en chimie [6] ainsi qu'à l'élaboration d'un logiciel libre (BenzAI [7]).

## 2 Sujet de thèse

Le projet s'articule autour de deux axes principaux :

- (1) Le premier concerne la génération exhaustive de structures 2D, de type nanographène, mais avec un nombre arbitraire de cycles composés de 4 à 7 atomes de carbone et la caractérisation de leur aromaticité/antiaromaticité. Les premiers nanographènes ont été synthétisés à partir de fragments de graphène, ce matériau bi-dimensionnel découvert en 2004, aux propriétés de conduction électrique et thermique, de résistance mécanique, de transparence optique remarquables [8]. On appelle aujourd'hui nanographène toute molécule constituée uniquement de cycles d'atomes de carbone fusionnés, saturés par des atomes d'hydrogène. Les benzénoïdes sont des nanographènes formés seulement de cycles à 6 carbones. Grâce aux progrès spectaculaires réalisées en chimie de synthèse dans les dix dernières années [9], il est possible de fabriquer des benzénoïdes dont la taille et la topologie sont totalement maîtrisées. Ils possèdent des propriétés opto-électroniques et magnétiques remarquables et sont en général particulièrement stables chimiquement. Les benzénoïdes possèdent une structure électronique caractérisée essentiellement par les électrons les plus mobiles, appelés électrons  $\pi$ . La répartition de ces électrons  $\pi$  est dite à « couche fermée » lorsqu'ils sont appariés deux à deux. La délocalisation de ces paires d'électrons  $\pi$  dans les cycles benzéniques d'un bout à l'autre de la molécule est à l'origine d'un des concepts les plus fondamentaux en chimie appelé aromaticité [10]. D'une manière simplifiée, un composé aromatique possède des circuits à  $4n+2$  électrons  $\pi$ , ce qui est réalisable à partir de la fusion de noyaux benzéniques ( $4n+2 = 6$  électrons  $\pi$ , avec  $n=1$ ). Plus un composé est aromatique, plus il est stable. Des modèles simples reposant sur la théorie des graphes [11, 12] permettent de prévoir qualitativement l'aromaticité d'un benzénoïde à partir de critères purement topologiques : il s'agit de dénombrer les sextets d'électrons  $\pi$  (modèle de Clar) ou les plus petits circuits conjugués à  $4n+2$  électrons (modèle de Randic) du benzénoïde. De tels modèles basés sur la connectivité des atomes sont particulièrement faciles à mettre en œuvre par des approches de programmation par contraintes (PPC [13]). En effet, chaque benzénoïde peut être modélisé par un graphe connecté, biparti et plan, dont les sommets sont les atomes de carbones et les arêtes sont les liaisons entre carbones. Pour un nombre de cycles benzéniques donné, il est possible de générer facilement à l'aide de la PPC un ensemble exhaustif de benzénoïdes respectant des contraintes données comme leur symétrie, leur forme, la présence de certains motifs, etc. L'aromaticité de ces composés peut être alors obtenue à partir du dénombrement des sextets et la détermination de leur répartition au sein du composé, tout comme les circuits conjugués à  $4n+2$  électrons. Dans le cadre d'une première thèse inter-doctorale, nous avons réglé ce problème de la génération systématique et exhaustive de tous les benzénoïdes et avons montré que les estimations d'aromaticité basées sur ces modèles simples sont en très bon accord avec des calculs de chimie quantique, beaucoup plus complexes à mettre en œuvre et très coûteux en temps de calcul [1, 2, 3, 4, 5, 6, 7]. La modification d'un nanographène par incorporation de cycles à 4, 5 ou 7 atomes de carbones dans un nanographène induit une profonde modification de ses propriétés d'aromaticité. Ces nanographènes modifiés sont qualifiés de « non-benzénoïdes » [14]. Les cycles à 4 électrons  $\pi$  constituent les plus petits cycles à  $4n$  électrons, connus pour être antiaromatiques, synonyme de moindre stabilité et d'un caractère radicalaire (c'est-à-dire à

---

1. Laboratoire d'Informatique et des Systèmes

2. Institut des Sciences Moléculaires de Marseille

« couche ouverte », les électrons  $\pi$  n'étant plus appariés) potentiellement élevé. Les cycles à 5 et 7 électrons  $\pi$  sont par essence radicalaires et seule une combinaison paire de tels cycles peut conduire à des matériaux à couches fermées. Il n'existe pas de modèle général permettant de prévoir les propriétés d'aromaticité ou d'antiaromaticité de ces composés non-benzénoïdes. En effet, la règle de Clar ne peut plus s'appliquer à des cycles à 4, 5 ou 7 électrons  $\pi$ . En revanche, la notion de circuits à  $4n$  ou  $4n+2$  électrons reste valable et leur dénombrement et répartition au sein du système d'électrons  $\pi$  permettrait d'estimer le caractère préférentiellement aromatique/anti-aromatique de certaines régions du composé. Par ailleurs, on peut également se demander quelles sont les combinaisons de cycles à 4, 5, 6 et 7 électrons  $\pi$  permettant d'obtenir des non-benzénoïdes les plus localement ou globalement aromatiques/antiaromatiques. C'est précisément à ces deux questions de génération exhaustive de non-benzénoïdes et de caractérisation de leur aromaticité que le premier axe de ce projet tentera d'apporter des réponses et des solutions. Du point de vue de la programmation par contraintes, cela nécessite une approche radicalement différente de celle envisagée pour les benzénoïdes, les graphes n'étant plus constitués uniquement d'hexagones, mais, d'un nombre arbitraire de cycles à 4, 5, 6 ou 7 carbones.

- (2) Le second axe de recherche concerne le développement d'un outil de génération automatique de structures moléculaires à 3 dimensions, par des approches issues de l'intelligence artificielle. Un tel outil permettrait au chimiste théoricien de sélectionner des géométries chimiquement pertinentes en amont de calculs de chimie quantiques complexes et coûteux en temps de calcul. Les oxydes d'iode et leurs combinaisons avec des molécules d'eau représentent des systèmes de choix pour le développement d'un tel outil : ces systèmes moléculaires complexes sont reconnus pour être des précurseurs d'aérosols, impliqués dans la formation de nuages à la surface des océans [15]. La dynamique de croissance des aérosols est un phénomène essentiel mais mal compris qui a des implications pour la science de l'atmosphère et la modélisation du climat [16]. Les oxydes d'iode présentent une grande diversité d'agencement tridimensionnel d'atomes d'iode et d'oxygène chimiquement liés, et les agrégats formés par ces oxydes d'iode et des molécules d'eau en interaction par des liaisons hydrogène peuvent avoir des géométries très diverses. Il existe en chimie théorique des outils d'exploration automatique des mécanismes réactionnels de réactions chimiques [17, 18], ou de génération automatique d'agrégats moléculaires obtenus comme états d'énergie minimum d'un agencement tridimensionnel de molécules reliées entre elles par des interactions faibles et dont la géométrie est connue au préalable [19]. Celles-ci, bien qu'efficaces, peuvent requérir des temps de calcul importants et nécessitent une connaissance au préalable des géométries tridimensionnelles des molécules mises en jeu dans les réactions ou dans les agrégats. L'avantage qu'offre une approche basée sur la programmation par contraintes est l'ajout progressif de contraintes spécifiques qui permettrait de sélectionner efficacement des formes tridimensionnelles pertinentes parmi l'ensemble des structures possibles. Une telle approche nécessite toutefois l'utilisation de variables continues encadrées par des valeurs minimales et maximales, ce qui représente un défi de taille pour la recherche efficace de solutions [20]. Aussi, au niveau de la PPC, ce second axe pourrait nécessiter de développer de nouvelles approches de résolution afin de mieux répondre aux besoins exprimés par les chimistes.

## Références

- [1] Adrien Varet. *Programmation par contraintes et chimie théorique : utilisation du formalisme CSP pour résoudre des problématiques liées aux benzénoïdes*. PhD thesis, 2022.
- [2] Yannick Carissan, Denis Hagebaum-Reignier, Nicolas Prcovic, Cyril Terrioux, and Adrien Varet. Exhaustive generation of benzenoid structures sharing common patterns. In *Proceedings of the 27th International Conference on Principles and Practice of Constraint Programming (CP)*, volume 210 of *LIPIcs*, pages 19 :1–19 :18, 2021.
- [3] Yannick Carissan, Denis Hagebaum-Reignier, Nicolas Prcovic, Cyril Terrioux, and Adrien Varet. Using constraint programming to generate benzenoid structures in theoretical chemistry. In *Proceedings of the 26th International Conference Principles and Practice of Constraint Programming (CP)*, volume 12333 of *Lecture Notes in Computer Science*, pages 690–706, 2020.
- [4] Yannick Carissan, Chisom-Adaobi Dim, Denis Hagebaum-Reignier, Nicolas Prcovic, Cyril Terrioux, and Adrien Varet. Computing the local aromaticity of benzenoids thanks to constraint programming. In *Proceedings of the 26th International Conference Principles and Practice of Constraint Programming (CP)*, volume 12333 of *Lecture Notes in Computer Science*, pages 673–689, 2020.
- [5] Yannick Carissan, Denis Hagebaum-Reignier, Nicolas Prcovic, Cyril Terrioux, and Adrien Varet. How constraint programming can help chemists to generate benzenoid structures and assess the local aromaticity of benzenoids. *Constraints*, 27(3) :192–248, 2022.
- [6] Adrien Varet, Nicolas Prcovic, Cyril Terrioux, Denis Hagebaum-Reignier, and Yannick Carissan. Benzai : A program to design benzenoids with defined properties using constraint programming. *J. Chem. Inf. Model.*, 62(11) :2811–2820, 2022.
- [7] BenzAI : an open-source software designed for the study of benzenoids using artificial intelligence techniques. <https://benzai-team.github.io/BenzAI>.
- [8] Jishan Wu, Wojciech Pisula, and Klaus Müllen. Graphenes as Potential Material for Electronics. *Chemical Reviews*, 107(3) :718–747, 2007.
- [9] Akimitsu Narita, Xiao-Ye Wang, Xinliang Feng, and Klaus Müllen. New Advances in Nanographene Chemistry. *Chem. Soc. Rev.*, 44(18) :6616–6643, 2015.

- [10] Gabriel Merino, Miquel Solà, Israel Fernández, Cina Foroutan-Nejad, Paolo Lazzeretti, Gernot Frenking, Harry L. Anderson, Dage Sundholm, Fernando P. Cossío, Marina A. Petrukhina, Jishan Wu, Judy I. Wu, and Albeiro Restrepo. Quo Vadis. *Chem. Sci.*, 2023.
- [11] Miquel Solà. Forty Years of Clar’s Aromatic  $\pi$ -Sextet Rule. *Front Chem*, page 1 :22, 2013.
- [12] Milan Randić. Conjugated Circuits and Resonance Energies of Benzenoid Hydrocarbons. *Chemical Physics Letters*, 38 (1) :68–70, 1976.
- [13] F. Rossi, P. van Beek, and T. Walsh. *Handbook of Constraint Programming*. Foundations of Artificial Intelligence. Elsevier, 2006.
- [14] Yoshito Tobe. Quinodimethanes Incorporated in Non-Benzenoid Aromatic or Antiaromatic Frameworks. *Physical Organic Chemistry of Quinodimethanes*, page 107–168, 2018.
- [15] Alfonso Saiz-Lopez, John M. C. Plane, Alex R. Baker, Lucy J. Carpenter, Roland von Glasow, Juan C. Gómez Martín, Gordon McFiggans, and Russell W. Saunders. Atmospheric Chemistry of Iodine. *Chem. Rev.*, 112 (3) :1773–1804, 2012.
- [16] Juan Carlos Gómez Martín, Thomas R. Lewis, Alexander D. James, Alfonso Saiz-Lopez, and John M. C. Plane. Insights into the Chemistry of Iodine New Particle Formation : The Role of Iodine Oxides and the Source of Iodic Acid. *J. Am. Chem. Soc.*, 144 (21) :9240–9253, 2022.
- [17] Saulo A. Vázquez, Xose L. Otero, and Emilio Martínez-Núñez. A Trajectory-Based Method to Explore Reaction Mechanisms. *Molecules*, 23 (12) :3156, 2018.
- [18] Satoshi Maeda, Koichi Ohno, and Keiji Morokuma. Systematic Exploration of the Mechanism of Chemical Reactions : The Global Reaction Route Mapping (GRRM) Strategy Using the ADDF and AFIR Methods. *Physical Chemistry Chemical Physics*, 15 (11) :3683–3701, 2013.
- [19] Christoph Plett and Stefan Grimme. Automated and Efficient Generation of General Molecular Aggregate Structures. *Angewandte Chemie International Edition*, 62 (4), 2023.
- [20] Frédéric Benhamou and Laurent Granvilliers. Continuous and interval constraints. In Francesca Rossi, Peter van Beek, and Toby Walsh, editors, *Handbook of Constraint Programming*, pages 571–603. Elsevier, 2006.

### 3 Informations générales et candidature

**Localisation :** Université d’Aix-Marseille, Faculté des Sciences, Laboratoire d’Informatique et des Systèmes (équipe COALA), Marseille

**Encadrement :**

- Cyril Terrioux (LIS) [cyril.terrioux@univ-amu.fr](mailto:cyril.terrioux@univ-amu.fr)
- Denis Hagebaum-Reignier (iSm2) [denis.hagebaum-reignier@univ-amu.fr](mailto:denis.hagebaum-reignier@univ-amu.fr)

**Date de début :** Automne 2024

**Profil recherché :** La personne recherchée devra être titulaire d’un Master 2 ou d’un diplôme d’ingénieur avec de solides compétences en informatique et en intelligence artificielle, notamment en programmation par contraintes, en algorithmique et en programmation. Des connaissances en optimisation combinatoire ou recherche opérationnelle seront un plus apprécié. Aucune connaissance en chimie n’est requise. L’inscription en thèse se fera dans la discipline informatique de l’école doctorale en Mathématiques et Informatique (ED 184).

**Dépôt de candidature :**

Les pièces demandées pour la candidature sont les suivantes :

- CV (au plus 3 pages),
- Relevés de notes, résultats et classements du diplôme de Master ou équivalent (première et deuxième années),
- Lettre de motivation,
- Lettres de recommandation.

Les candidatures doivent être déposées avant le 30 avril 2024 sous la forme d’un unique fichier pdf envoyé par courriel à Cyril Terrioux ([cyril.terrioux@univ-amu.fr](mailto:cyril.terrioux@univ-amu.fr)).