

# Utiliser la PPC pour générer des structures de benzénoïdes en chimie théorique\*

Yannick Carissan<sup>1</sup> Denis Hagebaum-Reignier<sup>1</sup>  
Nicolas Prcovic<sup>2</sup> Cyril Terrioux<sup>2</sup> Adrien Varet<sup>2</sup>

<sup>1</sup> Aix Marseille Univ, CNRS, Centrale Marseille, ISM2, Marseille, France

<sup>2</sup> Aix Marseille Univ, Université de Toulon, CNRS, LIS, Marseille, France  
{prénom.nom}@univ-amu.fr

## Résumé

Les benzénoïdes sont une sous-famille d'hydrocarbures (molécules composées uniquement d'atomes d'hydrogène et de carbone) dont les atomes de carbone forment des hexagones. Ces molécules ont fait l'objet de nombreuses études en chimie théorique et possèdent de nombreuses applications concrètes. La génération de benzénoïdes ayant certaines propriétés structurales (par exemple, ayant un nombre donné d'hexagones ou ayant une structure particulière du point de vue graphique) est un problème intéressant et important. Il constitue une étape préliminaire à l'étude de leurs propriétés chimiques. Dans cet article, nous montrons que modéliser ce problème dans Choco Solver et laisser son moteur de recherche générer les solutions constitue une approche rapide et très flexible. Elle permet notamment de générer des structures répondant aux besoins des chimistes simplement en ajoutant de nouvelles variables et/ou contraintes tout en évitant d'avoir à développer des méthodes algorithmiques ad-hoc.

Ce papier est un résumé de [2].

## 1 Introduction

Les *hydrocarbures aromatiques polycycliques (HAP)* sont des hydrocarbures dont les atomes de carbone forment des cycles de différentes tailles (de taille 6 dans le cas des *benzénoïdes*). Ils ont été très étudiés dans divers domaines (nanoélectronique moléculaire, synthèse organique, chimie interstellaire, ...) en raison de leur stabilité énergétique, de leurs structures moléculaires ou de leurs spectres optiques. Ils présentent une grande variété de propriétés physico-chimiques en fonction de leur taille et de leur structure. Ainsi, générer des structures de benzénoïdes respectant certaines

propriétés est d'un grand intérêt du point de vue de la chimie théorique.

Dans la littérature, des approches ont été proposées pour générer des structures quelconques ou satisfaisant des propriétés particulières [1]. Il s'agit d'approches ad-hoc qui s'avèrent très efficaces en pratique mais qui sont difficiles à adapter aux besoins des chimistes et qui requièrent un travail de développement significatif. Dans cet article, nous exploitons la PPC pour proposer une nouvelle approche qui soit plus flexible tout en étant relativement compétitive par rapport aux méthodes ad-hoc. En effet, d'une part, une fois le problème modélisé dans toute sa généralité, nous pouvons aisément le spécialiser en ajoutant des variables et/ou des contraintes afin de répondre aux besoins exprimés par les chimistes. D'autre part, la PPC offre des outils de résolution très efficaces.

## 2 Génération de benzénoïdes

Le *benzène*, représenté à la figure 1(a), est une molécule dont les six atomes de carbone forme un hexagone (appelé *cycle benzénique*). Les benzénoïdes sont les molécules qui peuvent être obtenues en fusionnant des cycles benzéniques. Par exemple, l'anthracène (voir figure 1(b)) contient trois cycles benzéniques fusionnés. En exploitant cette caractéristique, nous représentons la structure d'un benzénoïde  $B$  par le graphe non orienté  $B_h$  (appelé *graphe d'hexagones*) dont chaque sommet correspond à un hexagone (cycle benzénique) de  $B$  et tel que deux sommets sont reliés par une arête si les hexagones correspondants partagent une arête. La figure 1(d) présente ce graphe pour le coronène. Nous définissons le problème de génération ainsi : Étant donné un ensemble de propriétés structurales  $\mathcal{P}$ , générer toutes les structures satisfaisant les propriétés

\*Ce travail est soutenu par l'Agence Nationale de la Recherche dans le cadre du projet DEMOGRAPH (ANR-16-CE40-0028).

de  $\mathcal{P}$ . Ces propriétés peuvent porter sur le nombre de carbone ou d'hexagone ou sur des formes particulières du graphe d'hexagones (arbre, rectangle, ...).

Nous présentons maintenant une modélisation CSP permettant de générer toutes les structures ayant  $n$  hexagones. Elle repose sur la propriété que tout benzénoïde de  $n$  hexagones peut être placé dans un coronénoïde de taille au plus  $k(n) = \lfloor \frac{n}{2} + 1 \rfloor$ . Un coronénoïde de taille  $k$  est une molécule de benzène à laquelle on a ajouté successivement  $k - 1$  couronnes d'hexagones (le coronène est le coronénoïde de taille 2). Par la suite, on note  $B_h^{c(k(n))}$  le graphe d'hexagones du coronénoïde de taille  $k(n)$ ,  $n_c$  son nombre d'hexagones et  $m_c$  celui d'arêtes. Nous numérotions arbitrairement les hexagones de  $B_h^{c(k(n))}$  à partir de 1. Tout d'abord, nous considérons une variable de graphe  $x_G$  pour représenter le graphe d'hexagones de la structure souhaitée. Son domaine est l'ensemble de tous les sous-graphes entre le graphe vide et  $B_h^{c(k(n))}$ . Nous exploitons également un ensemble de  $n_c$  variables booléennes  $\{x_1, \dots, x_{n_c}\}$ . La variable  $x_i$  vaut 1 si le  $i$ -ème hexagone de  $B_h^{c(k(n))}$  est utilisé dans  $x_G$ , 0 sinon. De même, nous considérons un ensemble de  $m_c$  variables booléennes  $y_{i,j}$ . La variable  $y_{i,j}$  vaut 1 si l'arête  $\{i, j\}$  de  $B_h^{c(k(n))}$  est utilisée dans  $x_G$ , 0 sinon. Ensuite, nous modélisons les propriétés suivantes à l'aide de contraintes :

- *Lien entre  $x_G$  et  $x_i$  (resp.  $y_{i,j}$ )* : nous utilisons une contrainte **channeling** qui impose  $x_i = 1 \iff x_G$  contient le sommet  $i$  (resp.  $y_{i,j} = 1 \iff x_G$  contient l'arête  $\{i, j\}$ ).
- *$x_G$  est un sous-graphe induit de  $B_h^{c(k(n))}$*  : Toute valeur de  $x_G$  n'est pas nécessairement un graphe d'hexagones valide. Pour garantir sa validité, il doit correspondre à un sous-graphe de  $B_h^{c(k(n))}$  induit par les sommets appartenant à  $x_G$ . Ainsi, pour chaque arête  $\{i, j\}$  de  $B_h^{c(k(n))}$ , on ajoute une contrainte  $x_i = 1 \wedge x_j = 1 \iff y_{i,j} = 1$ . En d'autres termes, l'arête  $\{i, j\}$  existe dans  $x_G$  si et seulement si les sommets  $i$  et  $j$  apparaissent dans  $x_G$ .
- *La structure a  $n$  hexagones* :  $\sum_{i \in \{1, \dots, n_c\}} x_i = n$ .
- *Le graphe d'hexagones est connexe* : nous appli-

quons la contrainte de graphe **connected** sur  $x_G$ .

- *Six hexagones formant un cycle génèrent un hexagone (et non un trou)* : Pour chaque hexagone  $u$ , nous considérons l'ensemble  $N(u)$  de voisins de  $u$  dans le graphe d'hexagones. Pour chaque sommet  $u$ , on pose la contrainte  $\sum_{v \in N(u)} x_v = 6 \implies x_u = 1$ .

Enfin, nous ajoutons plusieurs contraintes pour éviter les redondances. D'abord  $x_G$  doit avoir au moins un sommet sur le bord supérieur (resp. gauche) de  $B_h^{c(k(n))}$  afin d'éviter les symétries par translation. Nous posons donc une contrainte qui spécifie que la somme des variables binaires  $x_i$  associées au bord supérieur (resp. gauche) est strictement positive. Ensuite, il faut s'assurer que le graphe décrit par  $x_G$  est le seul représentant de sa classe de symétrie. Il existe jusqu'à douze solutions symétriques : six symétries de rotation de 60 degrés combinées à une éventuelle symétrie axiale. Ces symétries sont cassées grâce à la contrainte **lex-lead**. Pour chacune des douze symétries, il faut ajouter  $n_c$  variables booléennes (chacune associée à une variable  $x_i$ ) et un total de  $3 \cdot n_c$  clauses ternaires.

Ce modèle peut facilement être mis en œuvre avec le solveur *Choco* [3]. Il peut également être spécialisé pour prendre en compte les besoins des chimistes en ajoutant des variables et/ou des contraintes. Par exemple, générer des structures ayant une forme arborescente (appelées benzénoïdes *catacondensés*) nécessite simplement d'ajouter au modèle général la contrainte de graphe **tree** appliquée à  $x_G$ . D'autres propriétés ont été modélisées afin de générer des structures ayant une forme rectangulaire, possédant un trou ou symétriques.

### 3 Conclusion

Nous avons présenté une approche basée sur la PPC permettant de générer des structures de benzénoïdes ayant des propriétés structurales. Elle s'avère flexible et permet de répondre aux besoins des chimistes tout en se révélant compétitive vis-à-vis des méthodes ad-hoc existantes.

### Références

- [1] G. BRINKMANN, G. CAPOROSSI et P. HANSEN : A Constructive Enumeration of Fusenes and Benzenoids. *Journal of Algorithms*, 45(2), 2002.
- [2] Y. CARISSAN, D. HAGEBAUM-REIGNIER, N. PROCOVIC, C. TERRIOUX et A. VARET : Using Constraint Programming to Generate Benzenoid Structures in Theoretical Chemistry. *In CP*, pages 690–706, 2020.
- [3] J.-G. FAGES, X. LORCA et C. PRUD'HOMME : Choco solver user guide documentation.

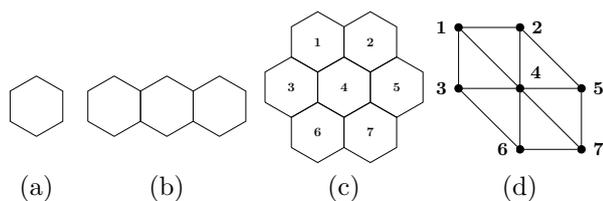


FIGURE 1 – Le benzène (a), l'anthracène (b), le coronène (c) (les atomes d'hydrogène sont omis) et son graphe d'hexagones (d).