



## UNIVERSITÉ CHEIKH ANTA DIOP DE DAKAR

Faculté des Sciences et Techniques

Master Modélisation de Systèmes et Phénomènes Physiques (MSPP)

### Analyse de propriétés croisées d'hydrocarbures aromatiques polycycliques à l'aide d'outils de la théorie des graphes

Soutenu par :

Monsieur Mouhamet TOURE - 07 juillet 2025, Marseille

Composition du jury :

Jury	Prénom(s) et Nom	Grade	Établissement
Président	Serigne Bira Gueye	Professeur Titulaire	UCAD
Membre	Frédéric Zolla	Professeur Titulaire	AMU
Membre	Stéphane Humbel	Professeur Titulaire	AMU
Membre	Mor Ndiaye	Maître de Conférences	UCAD
Membre	Guillaume Demesy	Maître de Conférences	AMU
Encadrant	Denis Hagebaum-Reignier	Maître de Conférences	AMU

Stage effectué à :

Institut des Sciences Moléculaires de Marseille (iSm2)

Aix-Marseille Université Mars - Juin 2025

## Résumé

Ce travail s'inscrit dans l'étude des hydrocarbures aromatiques polycycliques (HAP), des composés d'intérêt majeur en astrochimie et en science des matériaux, notamment pour leur rôle dans l'émission des bandes infrarouges observées dans le milieu interstellaire. Le but principal de ce stage est de réaliser une analyse de propriétés croisées de la base de données BenzDB (Benzenoids Data Base) récemment créée à Aix-Marseille Université. Pour cela, des outils de la théorie des graphes sont combinés à des calculs de chimie quantique avec les méthodes DFT et TD-DFT (B3LYP (CAM-B3LYP)/6-31+G(d) (6-31G)) et à la programmation python afin d'étudier des corrélations entre des descripteurs topologiques (nombre de ronds de Clar, irrégularité), des propriétés optoélectroniques (gap optique, gap HOMO-LUMO) et spectres infrarouges (IR) simulés. Ce stage de recherche vise à mieux comprendre l'impact de la structure des HAP, particulièrement les benzenoïdes sur leurs signatures spectrales, avec des perspectives en modélisation des spectres astrophysiques.

## Abstract

This work is part of the study of polycyclic aromatic hydrocarbons (PAHs), compounds of major interest in astrochemistry and materials science, particularly for their role in the emission of infrared bands observed in the interstellar medium. The main goal of this internship is to perform a cross-property analysis of the BenzDB database recently created at Aix-Marseille University. To this end, graph theory tools are combined with quantum chemistry calculations using DFT and TD-DFT methods (B3LYP (CAM-B3LYP)/6-31+G(d) (6-31G)) and Python programming to study correlations between topological descriptors (Clar ring number, irregularity), optoelectronic properties (optical gap, HOMO-LUMO gap), and simulated infrared spectra. This research internship aims to better understand the impact of the structure of PAHs, particularly benzenoids, on their spectral signatures, with perspectives in modeling astrophysical spectra.

## Dédicaces

Je dédie ce travail à mes parents **Katime TOURE** et **Ndeye Penda DIOP** pour leur amour, leur soutien constant et leurs sacrifices inestimables tout au long de mon parcours.

Je pense également avec émotion à mon oncle **Youssoupha SY**, disparu neuf jours après mon arrivée à Marseille, qui restera à jamais dans mon cœur. Que ce travail lui soit un humble hommage.

## Remerciements

Je souhaite profiter de la rédaction de ce rapport afin de remercier toutes les personnes y ayant contribué, de près comme de loin.

Les travaux présentés dans ce document sont avant tout issus d'un travail d'équipe important. Ainsi, je souhaiterais remercier tout particulièrement Dr Denis HAGEBAUM-REIGNIER, directeur de stage à l'iSm2 (équipe CTOM), de m'avoir accueilli au sein de son équipe et pour l'encadrement de qualité qu'il m'a offert tout au long de ce stage. Sa disponibilité, ses conseils éclairés et ses encouragements ont grandement facilité mon travail et enrichi mes réflexions.

Je remercie également tout le personnel des équipes Chimie Théorique et Modèles (CTOM/iSm2) et Contraintes, Algorithmes et Applications (COALA/LiS) pour leur accueil chaleureux, contribution active au projet et leur aide précieuse au quotidien.

Je suis reconnaissant envers l'ensemble du corps enseignant du master MSPP de l'UCAD et de l'AMU, pour la qualité de la formation dispensée et pour m'avoir permis d'acquérir les compétences nécessaires à la réalisation de ce stage.

Je tiens à adresser mes remerciements les plus affectueux à ma famille, et en particulier mes parents, mon grand frère BABACAR et ses sœurs AWA, FATOUMATA, MARIAMA, NAR et NDEYE MARIÈME, pour leur soutien constant, leurs prières, leurs encouragements et leur confiance en moi tout au long de ce parcours.

Je souhaite aussi exprimer toute ma gratitude à M. Ousmane SAGNANE, qui m'a grandement aidé durant ce parcours.

Enfin je remercie mes camarades de promotion du master MSPP, ainsi que mes amis proches, parmi lesquels Modou SY, Khadime NIASSE, Abdoulaye NGOM, Fallou TOURE et Ibrahima MBOW, pour leur soutien moral, leurs encouragements et les moments de convivialité partagés ensemble.

# Table des matières

<b>I</b>	<b>Introduction générale</b>	<b>9</b>
<b>II</b>	<b>Propriétés et descripteurs des HAP</b>	<b>11</b>
II.1	Introduction . . . . .	11
II.2	Aromaticité locale et globale . . . . .	11
II.3	Paramètre d'irrégularité . . . . .	12
II.4	Structures de Kekulé et délocalisation des électrons . . . . .	13
II.5	Structure de Clar et couverture de Clar . . . . .	14
II.6	Gap HOMO-LUMO et gap optique . . . . .	14
II.7	La topologie des HAP . . . . .	15
II.8	Spectroscopie IR . . . . .	16
II.9	Corrélation entre structure-propriétés des HAP . . . . .	17
II.10	Théorie des graphes pour la modélisation moléculaire . . . . .	17
II.11	Conclusion . . . . .	17
<b>III</b>	<b>Méthodologie</b>	<b>18</b>
III.1	Introduction . . . . .	18
III.2	Bases de données moléculaires étudiées . . . . .	18
III.2.1	La base de données BenzDB . . . . .	18
III.2.2	Présentation des autres bases étudiées . . . . .	19
III.2.2.1	La base de données COMPAS-3 . . . . .	19
III.2.2.2	La base de données de la NASA PAHdb . . . . .	20
III.3	Outils d'analyse et de traitements des données des bases . . . . .	20
III.3.1	Étude comparative entre BenzDB et PAHdb . . . . .	20
III.3.2	Outils pour sélectionner des structures benzenoïdes dans PAHdb . . . . .	21
III.4	Calculs quantiques réalisés . . . . .	22
III.4.1	Méthode de la théorie de la densité fonctionnelle DFT . . . . .	22
III.4.2	Méthode de la théorie de la densité fonctionnelle dépendante du temps TD-DFT . . . . .	23
III.5	Logiciels de Chimie quantique . . . . .	23
III.5.1	Logiciel Gaussian . . . . .	24
III.5.2	Logiciel Agui . . . . .	24
III.6	Conclusion . . . . .	24

<b>IV Résultats et discussion</b>	<b>25</b>
IV.1 Introduction . . . . .	25
IV.2 Analyse Comparative des bases de données étudiées . . . . .	25
IV.2.1 Statistiques des bases de données étudiées . . . . .	25
IV.2.2 Visualisation des composés des bases étudiées . . . . .	26
IV.3 Répartition des benzenoïdes de BenzDB selon leur symétrie . . . . .	27
IV.3.1 Classification des composés dans BenzDB selon les critères structuraux et symétries . . . . .	27
IV.3.2 Visualisation de la distribution des composés en fonction des groupes de symétrie . . . . .	28
IV.4 Étude des benzenoïdes par intervalle de Gap (eV) . . . . .	29
IV.4.1 Répartition des composés dans BenzDB selon leur gap (eV) . . . . .	29
IV.4.2 visualisation des benzenoïdes en fonction du gap homo-lumo . . . . .	30
IV.5 Étude des benzenoides par intervalle d'irrégularité . . . . .	30
IV.5.1 Répartition des compoés dans BenzDB selon le paramètre d'irrégularité . . . . .	30
IV.5.2 Visualisation des benzenoïdes en fonction des paramètres irrégularités . . . . .	31
IV.6 Corrélation entre propriétés structurales, spectrales et d'aromaticité . . . . .	32
IV.6.1 corrélation entre nombre de ronds de Clar, topologie, délocalisation $\pi$ et gap optique . . . . .	32
IV.6.2 Corrélation entre structures moléculaires et spectres IR d'un HAP . . . . .	33
IV.6.3 Relation entre le nombre de ronds de Clar et le HOMA . . . . .	34
IV.7 Conclusion . . . . .	35
<b>V Conclusion</b>	<b>36</b>
<b>A Script python pour l'analyse de PAHdb</b>	<b>41</b>
<b>B Exemple de fichier pour des calculs quantiques avec DFT et TD-DFT</b>	<b>44</b>
<b>C Formules mathématiques de la méthode de DFT</b>	<b>46</b>
C.1 Fonctionnelle énergétique . . . . .	46
C.2 Nombre de particules . . . . .	46
C.3 Comparaison des énergies . . . . .	46
C.4 Approximation Locale de la Densité électronique proposée par Kohn et Sham . . . . .	46
C.5 équation de Kohn et Sham . . . . .	46

# Table des figures

II.1	Illustration des hydrogènes périphériques ( $N_{solo} = 3$ , $N_{duo} = 4$ , $N_{trio} = 3$ , $N_{quartet} = 3$ . . . . .	13
II.2	Le benzène et ses deux structures Kekulé . . . . .	14
II.3	(a) Une structure de Kekulé du pérylène et (b) une des structures de Clar possibles . . . . .	14
II.4	Illustration des familles de bordures des HAP [1] . . . . .	16
IV.1	Histogramme comparatif entre les bases de données BenzDB, CompasDB et PAHdb (échelle logarithmique) . . . . .	26
IV.2	Benzenoïdes par symétrie dans BenzDB . . . . .	28
IV.3	Benzenoïdes en fonction du gap homo-lumo dans BenzDB . . . . .	30
IV.4	Benzenoïdes en fonction des paramètres d'irrégularités dans BenzDB . . . . .	31
IV.5	Illustration de la distribution des gaps HOMO-LUMO et optique en fonction du nombre de ronds de Clar . . . . .	32
IV.6	Gap optique et HOMO-LUMO en fonction du nombre de ronds de Clar . . . . .	33
IV.7	L'indice HOMA en fonction du nombre de ronds de Clar . . . . .	35

# Liste des tableaux

III.1 Répartition des composés benzenoïdes dans BenzDB en fonction du nombre d'hexagones et de leur symétrie . . . . .	19
IV.1 Nombre de composés selon le nombre d'hexagones dans chaque base de données .	26
IV.2 Nombre de composés selon le type de critères et symétries dans la base de données BenzDB . . . . .	28
IV.3 Nombre de composés selon l'intervalle du Gap HOMO-LUMO dans BenzDB . . .	29
IV.4 Nombres de composés selon l'intervalle du nombre d'irrégularité dans BenzDB .	31
IV.5 Correspondance entre longueurs d'onde et nombres d'onde pour les vibrations C-H	33

## Abbreviations

- **PAHs** : Polycyclic Aromatic Hydrocarbons
- **BenzDB** : Benzenoid DataBase
- **PAHdb** : Polycyclic Aromatic Hydrocarbon database
- **DFT** : Density Functional Theory
- **TD-DFT** : Time-Dependent Density Functional Theory
- **HOMO** : Highest Occupied Molecular Orbital
- **LUMO** : Lowest Unoccupied Molecular Orbital
- **B3LYP** : Becke, 3-parameter, Lee–Yang–Parr functional
- **CML** : Chemical Markup Language



# Chapitre I

## Introduction générale

Les hydrocarbures aromatiques polycycliques (HAP) sont des composés principalement constitués d'atomes de carbone et d'hydrogène. Chimiquement, les HAP de grande taille contiennent au moins deux cycles fusionnés de même taille ou de taille différente et sont disposés selon un arrangement linéaire, en amas ou angulaire.

Du fait de leur diversité structurale, les HAP ont suscités de nombreuses approches de modélisation pour une meilleure compréhension des ces composés. Pour approfondir la compréhension des propriétés structurales et électroniques des HAP, nous nous concentrerons dans ce travail, sur une famille particulière largement étudiée : les benzenoïdes.

Les benzenoïdes sont des hydrocarbures aromatiques polycycliques (HAP) constitués uniquement de cycles à six chaînons fusionnés. Ils sont étudiés dans divers domaines tels que la chimie, la physique, l'astrophysique et l'astrochimie en raison de leurs diverses propriétés structurales, optoélectroniques [2] et spectroscopiques. Ils sont également largement étudiés en chimie interstellaire car on soupçonne qu'ils sont présents dans les nuages interstellaires et qu'ils agissent comme catalyseurs pour les réactions chimiques se déroulant dans l'espace [3] [4].

L'étude du caractère aromatique des composés benzenoïdes est centrée principalement dans le cadre de ce stage sur plusieurs règles et critères telles que la règle de sextets de Clar (structure de Clar) pour les états fondamentaux [5], le paramètre d'irrégularité (hydrogènes périphériques) et les gaps HOMO-LUMO (Highest Occupied Molecular Orbital and Lowest Unoccupied Molecular Orbital) ou gaps optiques.

Concernant les benzenoïdes de tailles plus grandes, la topologie c'est-à-dire la forme des bordures a une grande influence sur l'étude des propriétés. Ils constituent également un défi majeur pour les chimistes de synthèse, ce qui conduit souvent à la synthèse de molécules remarquables sur des surfaces [6].

Leur grande stabilité, leur réactivité, leur propriétés optiques et électroniques, sont directement liés à leur structure électronique.

Les HAP, soupçonnés d'être à l'origine des bandes infrarouges (IR) aromatiques diffuses observées dans de nombreuses régions interstellaires, font l'objet d'étude approfondies afin de mieux interpréter ces observations et comprendre la chimie du milieu interstellaire.

## Objectifs du stage

Ce stage s'inscrit dans le cadre d'une collaboration entre l'équipe CTOM de l'Institut des Sciences Moléculaires de Marseille (iSm2) et l'équipe COALA du Laboratoire Informatique et Système (LiS), avec comme ambition d'approfondir l'étude sur différentes familles d'hydrocarbures aromatiques polycycliques (HAP). Ces composés font l'objet de nombreuses études de par leur grande stabilité et leurs propriétés spectrales bien caractéristiques. Ils jouent un rôle important en astrochimie, notamment dans l'interprétation des spectres d'infrarouge (IR), visible et ultraviolet (UV-vis) des milieux interstellaires (ISM).

L'objectif général de ce stage est d'analyser en détail la base de données BenzDB, qui contient 8392 composés benzenoïdes allant du benzène jusqu'à neuf hexagones fusionnés (HAP formés uniquement de cycle benzéniques fusionnés).

Cette analyse vise à :

1. Simuler les spectres IR de familles de composés définies selon certaines propriétés d'irrégularité, de symétrie, de compacité et de les comparer à ceux d'autres études.
2. étudier les corrélations possibles entre les propriétés structurales, électroniques, spectrales et d'aromaticité (locale et globale) des structures benzenoïdes étudiées.

Ce travail combine plusieurs approches telles que la chimie quantique pour les simulations de spectres, la théorie des graphes pour la caractérisation structurale et aromatique des molécules et la programmation en python pour l'analyse de la base à l'aide de scripts.

## Plan du rapport

Afin de mieux répondre aux objectifs de ce stage, la suite du rapport est organisé de la manière suivante :

- Une partie présentant de manière générale l'aromaticité, la théorie des graphes appliquées aux HAP, certains descripteurs topologiques (structures de Kekulé, sextets de Clar, le paramètre d'irrégularité, la topologie des bordures, etc ...), les gaps optiques et HOMO-LUMO, et le spectre infrarouge (IR),
- Une partie méthodologie, détaillant les techniques d'analyse de BenzDB, une base de données de 8392 HAP [7] et d'autres bases de HAP comparables dans la littérature [8, 9] (article en cours de rédaction), les calculs en chimie quantique réalisés avec les méthodes DFT et TD-DFT (géométries optimisées, spectre IR simulés et extractions des indices structuraux et topologiques),
- Une partie résultats et discussions, où seront exposés en détail les résultats issus d'analyse de différentes propriétés croisées dans BenzDB, les corrélations entre propriétés et aromaticité observées et leur interprétation,
- Une conclusion récapitulant les contributions de ce travail et dégageant des perspectives.

## Chapitre II

# Propriétés et descripteurs des hydrocarbures aromatiques polycycliques (HAP)

### II.1 Introduction

Cette partie présente les concepts fondamentaux mobilisés au cours de ce travail. Nous y aborderons notamment les notions d’aromaticité (locale et globale), la théorie des graphes appliquée à la modélisation moléculaire, ainsi que quelques descripteurs topologiques, électroniques et spectroscopiques qui permettent de caractériser les composés benzenoïdes dans BenzDB.

### II.2 Aromaticité locale et globale

La notion d’aromaticité est un concept fondamental en chimie organique, introduit par Michael Faraday en 1825 suite à la découverte du benzène. Ce concept très intéressant a été formaliser quelques années plus tard par August Kekulé, qui proposa la structure hexagonale du benzène avec une alternance des liaisons simples et doubles au sein du cycle benzénique. L’aromaticité désigne des systèmes conjugués ayant une grande stabilité particulière liée à la délocalisation des électrons  $\pi$ . Dans l’article de Schleyer et al. [10], les auteurs ont rappelé que l’aromaticité d’un composé reste difficile à définir de façons précise et rigoureuse, car ne se réduit pas à une seule propriété physique mesurable. Il s’agit plutôt d’un concept regroupant un ensemble de critères structurels, énergétiques, magnétiques et électroniques. Dans ce travail, nous nous intéressons à la fois de l’aromaticité locale et globale en utilisant différents descripteurs. L’aromaticité locale, est caractérisée par des descripteurs de types topologiques tels que le nombre de ronds de Clar représentant les sextets  $\pi$  localisés les plus stables au sein du cycle [11] [12] et les structures de Kekulé, qui montre les différentes manières d’agencer les doubles liaisons dans le squelette moléculaire.

L’aromaticité globale, peut, quant à elle, être caractérisées par un indice HOMA (Harmonic Oscillator Model of Aromaticity) [11]. Cet indice est calculé à partir de distances en Angströms des liaisons carbone-carbone (C-C) au sein du cycle selon l’équation suivante :

$$\text{HOMA} = 1 - \frac{\alpha}{n} \sum_i (R_{\text{opt}} - R_i)^2 \quad (\text{II.1})$$

où  $\alpha$  est une constante de normalisation ( $257.7 \text{ \AA}^{-2}$ ),  $n$  est le nombre de liaisons C-C dans le cycle et  $R_{\text{opt}}$  est une distance référence pour une liaison aromatique parfaite obtenue à partir des mesures de diffraction électronique en phase gazeuse du 1-3 butadiène comme moyenne géométrique idéale ( $1.388 \text{ \AA}$ ) entre les liaisons simples ( $C_s = 1.467 \text{ \AA}$ ) et double ( $C_d = 1.349 \text{ \AA}$ ) dans les systèmes conjugués [11]

Cet indice indique dans quelle mesure la structure est proche d'un système aromatique ou pas. Un HOMA proche de 1 révèle une aromaticité forte, alors qu'une valeur faible indique des distorsions hors plan ou une faible délocalisation électronique. L'analyse conjointe de l'aromaticité locale et globale permet une évaluation approfondie des caractéristiques aromatiques des composés benzenoïdes présents dans BenzDB.

## II.3 Paramètre d'irrégularité

Le paramètre d'irrégularité, noté  $\xi$ , est un paramètre topologique défini selon la formule suivante :

$$\xi = \frac{N_{\text{trio}} + N_{\text{quartet}}}{N_{\text{solo}} + N_{\text{duo}} + N_{\text{trio}} + N_{\text{quartet}}} \quad (\text{II.2})$$

Ce paramètre permet de quantifier la distribution des atomes d'hydrogènes périphériques (II.1) (solo, duo, trio, quartet,  $N_{\text{solo}}$  étant le nombre d'hydrogènes "solo",  $N_{\text{duo}}$  le nombre d'hydrogènes "duo" etc...) dans une molécule benzenoïde. Un hydrogène "solo" est un hydrogène en périphérie du HAP lié à un atome de carbone n'ayant aucun voisin possédant un atome d'hydrogène. Lorsque deux carbones voisins sont liés chacun à un atome d'hydrogène, on parle d'hydrogènes "duo", etc... Ce paramètre permet, notamment dans le domaine de l'astrochimie, de caractériser la régularité de la périphérie des HAP [4]. L'outil BenzAI permet de calculer directement cette irrégularité pour un ensemble de molécule dans BenzDB, ou de générer automatiquement des benzenoïdes selon les contraintes fixées sur le paramètre d'irrégularité  $\xi$ , ou sur les quantités de chaque type d'atome d'hydrogène périphérique. Dans certains travaux sur l'étude des HAP tels que Bouwman et al. [13], les auteurs ont montré que la base de la NASA (National Aeronautics and Space Administration) PAHdb manquait de grandes molécules, car elle contient relativement peu de grandes molécules avec plus de 34 atomes de carbone environ ayant une irrégularité élevée. Cette description est affirmée par une analyse en détail de la distribution de  $\xi$  que nous avons effectuée dans tous les benzenoïdes de taille inférieure ou égale à neuf hexagones dans BenzDB. Cette analyse a révélé une prédominance de structures à faible irrégularité, surtout pour les composés de grands tailles. Ce manque dans PAHdb est significatif, car dans plusieurs études consultées [14] [15] [16], les auteurs ont montré que le type d'hydrogène périphérique (solo, duo, trio, quarto) a une grande influence directe sur la position et l'intensité des bandes d'absorption dans les spectres infrarouges. Ainsi, la diversité topologique des hydrocarbures aromatiques polycycliques (HAP) et en particulier leur irrégularité est un facteur déterminant dans la modélisation et l'interprétation de leur signatures spectroscopiques.

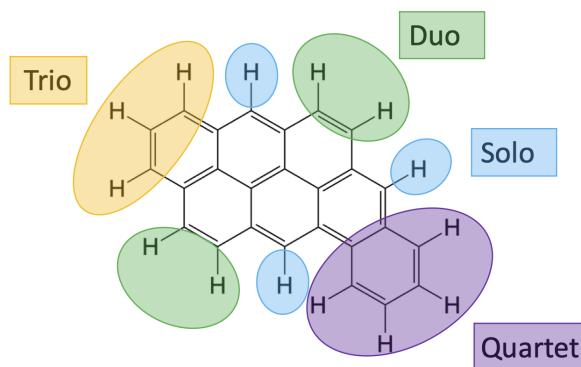


FIGURE II.1 – Illustration des hydrogènes périphériques ( $N_{solo} = 3$ ,  $N_{duo} = 4$ ,  $N_{trio} = 3$ ,  $N_{quartet} = 3$

)

## II.4 Structures de Kekulé et délocalisation des électrons

Les notions de structures de Kekulé et de délocalisation des électrons  $\pi$  ont été abordées lors de ce stage pour caractériser les propriétés structurales et d'aromaticité des molécules étudiées dans la base de données BenzDB. L'étude des structures aromatiques a connue une grande avancée vers les années 1865, Auguste Kekulé [17] a proposé la première structure définie pour le benzène. Il suggère que le benzène est une structure cyclique de six atomes de carbone avec une alternance de liaisons simples et doubles. Chaque atome de carbone est attaché à un atome d'hydrogène [18]. Cette vision de Kekulé (1865) introduit ce que l'on appelle aujourd'hui structure de Kekulé pour mieux cerner les propriétés moléculaires et l'aromaticité. Quelques dizaines d'années plus tard, J.J. Thomson a fait son entrée à l'institution royale le 30 avril 1897 pour annoncer sa découverte de l'électron, un constituant fondamental de tout atome. Il reçut le prix Nobel 1906 pour avoir montré que l'électron était une particule [19]. Cette découverte a permis de réinterpréter ces structures. On comprend alors que les doubles liaisons dans les structures de Kekulé des hydrocarbures aromatiques polycycliques ne sont pas fixes mais résultent d'une délocalisation des électrons  $\pi$ , des électrons provenant des orbitales de type  $p$ , perpendiculaires au plan moléculaire, de chaque atome de carbone au sein du cycle [20]. Ces électrons ne sont pas localisés spécifiquement entre deux atomes de carbone mais se déplacent dans l'ensemble du système plan conjugué, produisant des liaisons supplémentaires dynamiques voir figure (II.2). Ce phénomène de délocalisation des électrons rend compte de la stabilité des composés aromatiques tels que les benzenoïdes, et justifie la représentation alternative sous forme de résonance entre plusieurs structures Kekulé, chacun montrant une configuration possible des doubles liaisons.

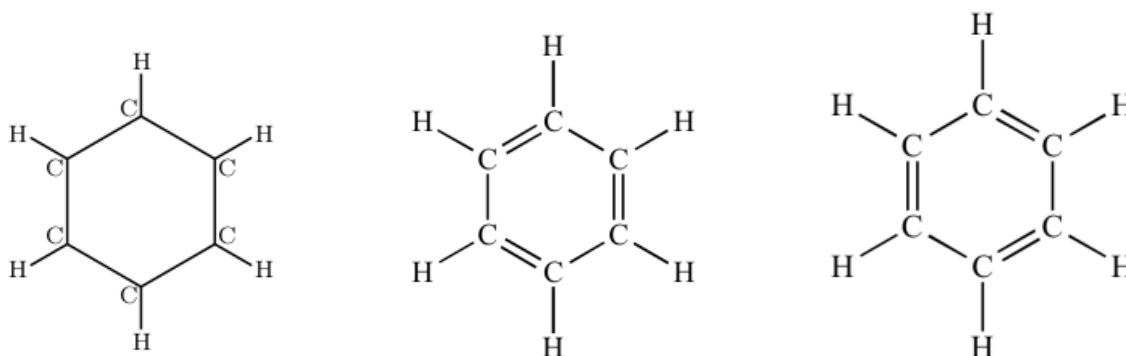


FIGURE II.2 – Le benzène et ses deux structures Kekulé

## II.5 Structure de Clar et couverture de Clar

La structure de Clar, introduit par F. Clar en 1972, constitue pour les chimistes un outil important pour représenter la délocalisation des électrons au sein des HAP. Elle est définie comme la structure possédant le maximum de sextets, ou ronds, correspondant à 6 électrons délocalisés au sein d'un noyau benzénique, les autres électrons étant répartis en doubles liaisons. En d'autres termes, elle peut être vue comme une somme de deux structures de Kekulé (voir figure II.3). L'ensemble des structures de Clar équivalentes est appelée couverture de Clar.



FIGURE II.3 – (a) Une structure de Kekulé du pérylène et (b) une des structures de Clar possibles

## II.6 Gap HOMO-LUMO et gap optique

Le gap HOMO-LUMO (**HOMO** (Highest Occupied Molecular Orbital) : orbitale moléculaire la plus haute occupée et **LUMO** (Lowest Unoccupied Molecular Orbital) : orbitale moléculaire la plus basse inoccupée) est un paramètre fondamental qui peut nous renseigner sur la réactivité chimique des composés, la stabilité (aromaticité) et les propriétés d'absorption et d'émission de tous les HAP, particulièrement les benzenoïdes. Dans plusieurs études sur le gap HOMO-LUMO, les auteurs ont montré que plusieurs corrélations entre le gap et d'autres propriétés structurales

et spectrales. Dans le cadre de la modélisation des propriétés électroniques des composés étudiés, une évaluation précise des niveaux d'énergies des orbitales moléculaires frontières (HOMO et LUMO) est d'une importance capitale. Dans l'article Ruiz-Morales et al., (2002), les auteurs ont montré en général, plus une molécule possède des cycles aromatiques fusionnés, plus son gap HOMO-LUMO diminue, traduisant ainsi une grande délocalisation des électrons et une transition vers un comportement semi-conducteur [21]. Cette tendance est non seulement influencée par la taille des molécules, mais aussi par la topologie des structures aromatiques des composés benzenoïdes (krygowski et al., 2001) [22]. Dans l'article Kateris et al., (2023), les auteurs se sont intéressés à la variation du gap HOMO-LUMO dans des HAP de grandes tailles et à leur corrélation avec les propriétés électroniques et optiques des nanoparticules de carbone formées par flamme [23]. Dans cet article, il est bien décrit que ces nanoparticules présentent un comportement de confinement quantique, similaire à celui des points quantiques semi-conducteurs c'est-à-dire que les petites particules ont des gaps plus élevés, alors que les plus grandes tendent souvent vers une valeur de gap nul, comme l'exemple du graphène. Pour mieux cerner cette transition énergétique, ils ont effectué des calculs de gap des HAP de grandes tailles  $C_{384}H_{48}$  (environ 168 hexagones fusionnés) en utilisant une optimisation géométrique par champ de force empirique suivi d'un calcul en DFT avec la fonctionnelle et base B3LYP/6-31G(d). Ils montrent, à l'issue de ces calculs quantiques, que les HAP hexagonaux voient leur gap HOMO-LUMO diminuer avec la taille de la molécule, tendant vers celui du graphène, tandis que les polyacènes présentent un gap qui se stabilise à une valeur finie. En plus, ils démontrent que le gap d'un cluster de HAP dépend fortement des molécules les plus grandes qu'il contient. Enfin, cette étude apporte une compréhension importante de la corrélation entre la taille et la composition moléculaire dans les suies et nanoparticules de carbones, et met en évidence que le gap optique observé est principalement déterminé par les grands HAP, plutôt que par une moyenne sur l'ensemble du cluster. Elle offre ainsi un cadre quantitatif pour interpréter des propriétés électroniques, en reliant les tendances des gaps aux mécanismes de croissances des HAP. En nous basant sur ces études précédentes, notre travail sur la notion de gap HOMO-LUMO constitue un outil capital pour mieux comprendre les propriétés électroniques des HAP, et particulièrement des benzenoïdes.

Le gap optique, selon Adkins et Miller, 2017 [11], correspond à l'énergie de transition entre l'état fondamental et le premier état excité présentant une force d'oscillation significative. Les auteurs affirment également que ce gap est fortement influencé par la topologie moléculaire (délocalisation des électrons  $\pi$ ). Il traduit ainsi la facilité avec laquelle une molécule peut absorber la lumière et passer à un état excité, ce qui en fait un descripteur clé des propriétés électroniques et optiques des HAP.

## II.7 La topologie des HAP

L'étude sur la topologie de la bordure des HAP constitue selon les chimistes un outil essentiel pour caractériser ces composés. Dans plusieurs travaux, les auteurs ont montré que la nature géométrique des contours moléculaires influence fortement les propriétés électroniques, optiques, spectroscopiques et voir même magnétiques [24] [25]. Ils distinguent de façon générale deux grandes familles de bordures : les baies, qui incluent les configurations géométriques de types

deep bay (ou cove), shallow-armchair bay (ou armchair), zigzag bay et ultra-deep bay (ou fjord) et les protusions, qui sont constituées essentiellement les bordures de types c3-h3 protusion, c4-h4 protusion, zigzag edge et armchair edge. La figure (II.4) illustre de façon détaillée les différentes bordures des HAP définies plus haut. Elles jouent un rôle crucial dans la réactivité locale et la stabilité (aromaticité) des structures benzenoïdes. L'identification précise de ces bordures permet une meilleure compréhension des relations entre structures des benzenoïdes étudiés et leurs propriétés électroniques, optiques et spectroscopiques.

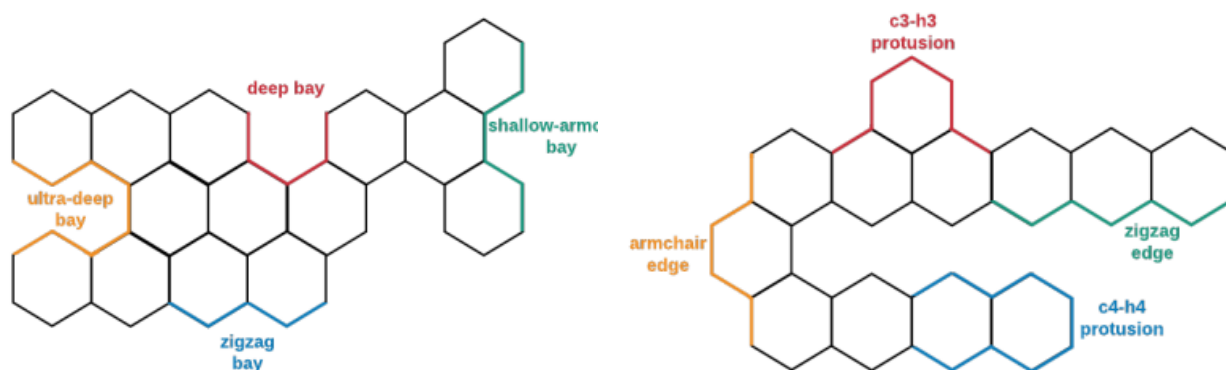


FIGURE II.4 – Illustration des familles de bordures des HAP [1]

## II.8 Spectroscopie IR

La spectroscopie infrarouge (IR) permet d'identifier, de tester la présence d'une molécule dans un échantillon. Le spectre infrarouge (IR) est une technique analytique qui permet d'explorer la structure moléculaire en étudiant les vibrations internes. Lorsqu'une molécule est soumise à un rayonnement infrarouge, certaines longueurs d'onde sont absorbées si elles correspondent aux fréquences de vibrations des groupes fonctionnels présents. Le spectre IR obtenu constitue une zone de vibration caractéristique de la molécule, révélant ainsi des informations précises sur la nature des liaisons chimiques, l'irrégularité, la symétrie et parfois même l'aromaticité du système étudié. Dans l'étude des HAP, l'analyse des spectres IR permet d'identifier les modes de vibrations caractéristiques des cycles aromatiques fusionnés, notamment les vibrations C-H hors du plan et les vibrations C=C dans le plan, dont les positions et intensités dépendent fortement de la structure moléculaire (Hudgins et al., 1999) [14]. Ainsi, la spectroscopie IR constitue un outil essentiel non seulement pour l'identification des HAP dans les mélanges complexes tels que les suies ou les asphaltènes, mais aussi pour la validation théorique de leurs propriétés structurales et électroniques simulées par des méthodes de chimie quantique. Dans d'autres études sur le spectre infrarouge (IR) des HAP, les auteurs ont montré que la bande d'émission des hydrocarbures aromatiques polycycliques (HAP) domine le spectre IR moyen d'une grande variété de sources astronomiques, englobant presque tous les stades de l'évolution stellaire. Malgré leurs similitudes, les détails de la position et de la forme des bandes ont permis d'élaborer une classification de l'émission des HAP [13]. Les bandes interstellaires diffuses sont des phénomènes couvrant le domaine spectral du proche ultraviolet (UV) au proche infrarouge (IR), observés dans le spectre des étoiles masquées par des nuages interstellaires diffus. Depuis leur découverte



en 1922 par Heger [26], l'identification des porteurs de ces bandes constitue un défi permanent pour les scientifiques du XXe siècle [27].

## II.9 Corrélation entre structure-propriétés des HAP

Des études précédentes ont montré une corrélation entre structures, aromaticité et spectroscopie IR [28] [29].

L'étude de Bauschlicher et al., 2009 [29] explore la relation entre les structures géométriques des HAP et leurs spectres IR, en se focalisant sur des composés de grandes tailles et irréguliers. En comparant les spectres des HAP compacts et symétriques par rapport à ceux présentant des irrégularités, tels que des régions déformées ou des hydrogènes en quartet, les auteurs ont montré que ces caractéristiques structurales influencent significativement la position et l'intensité des bandes spectrales IR.

## II.10 Théorie des graphes pour la modélisation moléculaire

La théorie des graphes est aujourd'hui un domaine largement utilisée pour la modélisation moléculaire. Il offre un cadre mathématique important pour la représentation et l'analyse de structures moléculaires. En chimie théorique, une molécule organique peut être modélisée en graphe moléculaire en représentant l'ensemble des atomes par des sommets de graphe (ou noeuds) et les liaisons entre atomes par des arêtes. Cette représentation permet d'étudier les propriétés structurelles et topologiques des molécules sans se référer directement aux coordonnées spatiales des atomes de la molécule. Les graphes moléculaires sont particulièrement utiles pour comprendre la connectivité des atomes, la symétrie et la planarité au sein des molécules. Ils sont aussi utiles pour la classification des hydrocarbures aromatiques polycycliques (HAP), la recherche des structures de Clar (sextets aromatiques) et de Kekulé, ou bien pour l'identification des isomères et des motifs structuraux spécifiques. La théorie des graphes est aussi un outil essentiel pour calculer des indices de topologies, tels que le paramètre d'irrégularité, qui peut être corrélé avec des propriétés physico-chimiques et spectroscopiques des molécules. En modélisation moléculaire, ces représentations permettent une génération exhaustive et automatique de famille entière de molécule, par exemple tous les benzenoïdes dans BenzDB avec un nombre d'hexagones donnés et de prédire éventuellement leurs propriétés à l'aide des méthodes combinatoires, statistiques ou d'apprentissage automatique [30] [31] [32] [33]. Cette approche joue un rôle capital dans le développement de bases de données moléculaires, par exemple BenzDB, dans la modélisation et la simulation des propriétés électroniques, optiques et vibratoires des molécules.

## II.11 Conclusion

Cette partie a permis d'établir un cadre théorique solide autour des descripteurs structuraux, topologiques et spectroscopiques des HAP. Ces outils sont essentiels pour relier forme moléculaire, stabilité et propriétés électroniques et optiques, en particulier à travers l'aromaticité.

## Chapitre III

# Méthodologie

### III.1 Introduction

Dans cette section, nous présentons les différentes approches méthodologiques et les outils computationnels mobilisés pour l'analyse structurale et électronique des hydrocarbures aromatiques polycycliques (HAP). Notre objectif principal est d'établir des corrélations entre les propriétés structurales, spectrales et d'aromaticité des composés benzenoïdes étudiés. Pour ce faire, nous avons adopté une double approche combinant des outils issus de la théorie des graphes afin de modéliser les structures sous forme de graphe moléculaire planaire et des méthodes de chimie quantique basées sur la théorie de la fonctionnelle de la densité (DFT) et sa version dépendante du temps (TD-DFT) afin de simuler les propriétés électroniques et spectroscopiques.

### III.2 Bases de données moléculaires étudiées

#### III.2.1 La base de données BenzDB

BenzDB est une base de données computationnelle développée par des Professeurs d'Aix-Marseille Université (équipe Chimie Théorique et Modèles (CTOM/iSm2) et Contraintes, Algorithmes et Applications (COALA/LiS)). Cette base est spécialisée dans les hydrocarbures aromatiques polycyclique (HAP), plus précisément dans les composés benzenoïdes d'où le nom Benzenoïdes Data Base. Cette base contient des molécules aromatiques formées d'un ou de plusieurs cycles benzéniques fusionnés en fournissant leurs propriétés structurales et électroniques. Elle renferme toutes les structures de benzenoïdes possibles jusqu'à neuf cycles fusionnés, couvrant une large gamme de diversité des motifs tels que les benzenoïdes Kekuléens (possédant au moins une structure de résonance de Kekulé, donc tous les électrons  $\pi$  sont appariés en double liaisons), non-Kekuléens (contraire aux benzenoïdes Kekuléens), catacondensés (au moins un cycle partage deux sommets), péricondensés (au moins un cycle partage trois sommets), et coronoides (structures contenant un trou) avec un ensemble de groupe de symétrie moléculaire. Pour chaque molécule de cette base de données, on peut avoir les informations détaillées telles que les valeurs des orbitales frontières (HOMO-LUMO), le Gap, le paramètre d'irrégularité, des indices d'aromaticité comme le NICS, RBO, couverture de clar etc., la géométrie optimisée ainsi que des spectres théoriques IR par des méthodes de calculs de types DFT et TD-DFT. L'un des avantages majeur de BenzDB se repose dans l'intégration de l'outil BenzAI [7]. BenzAI

est un logiciel open source destiné aux chimistes qui aborde plusieurs questions relatives aux benzénoïdes grâce à des techniques d’intelligence artificielle. Il permet notamment de générer des structures de benzénoïdes selon plusieurs critères (nombre d’hexagones/atomes de carbone/atomes d’hydrogène, propriétés structurales, présence ou absence d’un ou plusieurs motifs, etc.), d’analyser leur structure électronique, et fournir des spectres infrarouges stockés dans la base de données externe BenzDB [4]. Ces problèmes sont modélisés et résolus grâce à la programmation par contraintes, et notamment à l’utilisation de Choco Solver dont on pose des contraintes (carbones dans des hexagones, chaque carbone a trois voisins etc...) que le solveur (Choco) doit satisfaire. Cela se fait par raisonnement automatique, c’est-à-dire qu’une fois les contraintes bien définies, le solveur cherche toutes les solutions qui satisfont ces contraintes générales sans avoir à traiter différents cas spécifiques. Grâce à cet outil, il est possible de visualiser les structures des composés sous forme de graphe moléculaire, d’extraire des données personnalisées telles que la géométrie, et d’effectuer des analyses statistiques ou de similarités sur l’ensemble des composés benzénoïdes disponibles dans la base BenzDB. Ce couple BenzDB/BenzAI offre ainsi un environnement complet et convivial pour l’étude des hydrocarbures aromatiques polycycliques et les relations entre propriétés et structures. Leur utilisation gagne en temps et évite la redondance des calculs quantiques coûteux pour la communauté chimiste en accédant directement à des résultats fiables, sans avoir à effectuer des calculs lourds.

Nbre d’hexa- gones	$C_s$ ou $C_1$	$C_{2v}$	$C_{2h}$	$D_{2h}$	$C_{3h}$	$D_{3h}$	$C_{6h}$	$D_{6h}$
1	0	0	0	0	0	0	0	1
2	0	0	0	1	0	0	0	0
3	0	1	0	1	0	1	0	0
4	2	1	1	2	0	1	0	0
5	10	9	1	2	0	0	0	0
6	57	12	7	3	1	1	0	0
7	279	39	7	3	1	1	0	1
8	1333	61	35	7	0	0	0	0
9	6281	179	36	7	5	2	0	0

TABLE III.1 – Répartition des composés benzenoïdes dans BenzDB en fonction du nombre d’hexagones et de leur symétrie

### III.2.2 Présentation des autres bases étudiées

#### III.2.2.1 La base de données COMPAS-3

COMPAS-3 est une base de données computationnelle récente de systèmes aromatiques et polycycliques, orientée vers les hydrocarbures polybenzenoïdes péricondensés et des structures aromatiques compactes [8]. Elle contient environ 39k structures optimisées par la méthode semi-empirique GFN2-xTB dont 8844 structures benzenoïdes optimisées par la méthode de DFT avec CAM-B3LYP-D3BJ/cc-pvdz//CAM-B3LYP-D3BJ/def2-SVP jusqu’à dix cycles benzeniques. Les données incluent les propriétés électroniques tel que le Gap HOMO-LUMO et

structurales. Cette base de donnée permet d'étudier les relations entre propriétés et structures pour la conception de matériaux conducteur tels que les semi-conducteurs et permet aussi un entraînement de modèle de Machine Learning et Deep Learning en chimie computationnelle.

### III.2.2.2 La base de données de la NASA PAHdb

PAHdb est une base de données de spectres IR expérimentaux et théoriques développée par la NASA (National Aeronautics and Space Administration) en accès libre sur internet [9]. Elle se concentre principalement sur les hydrocarbures aromatiques polycycliques (HAP) dans l'espace interstellaire. Contrairement aux bases de données BenzDB et COMPAS-3, PAHdb contient aussi des espèces ionisés, des hétéroatomes et des radicaux, composés cruciaux pour l'astrophysique. L'étude de cette base dans le cadre de ce stage porte sur les deux version à savoir PAHdb *v3.20* et PAHdb *v4.00* au format CML( Chemical Markup Language).

## III.3 Outils d'analyse et de traitements des données des bases

### III.3.1 Étude comparative entre BenzDB et PAHdb

La première étape de cette étude a consisté à sélectionner tous les composés de ces deux bases de données présentant des caractéristiques différentes en termes de tailles des molécules, types et structures. La base de données BenzDB contient environ 8392 benzenoïdes, tandis que les deux versions de PAHdb — pahdbv3.20 et pahdb v4.00 comprennent respectivement 4233 et 10890 composés, incluant des espèces chargées et des hétéroatomes. Nous avons essayé de faire un tri de cette dernière pour sélectionner uniquement les composés neutres ne contenant que du carbone et de l'hydrogène. Et pour ce faire, nous avons utilisé plusieurs méthodes pour pouvoir faire cette comparaison. Pour rendre plus simple ce travail, nous avons adopté une modélisation des molécules contenues dans la base de données PAHdb de la NASA en utilisant les outils de la théorie des graphes. Ainsi, chaque molécule a été modélisée sous forme d'un graphe, où les sommets représentent les atomes de carbone et les arêtes représentent les liaisons chimiques entre les atomes. Cette représentation nous a permis de passer d'une approche chimique à une approche purement mathématique et algorithmique, facilitant le calcul de paramètres topologiques des composés.

Plusieurs paramètres ont été calculés afin de caractériser les propriétés structurales de chaque composé :

$n$  : nombre de sommets

$h$  : nombre d'hexagones

$m$  : nombre d'arêtes

$n_i$  : nombre de sommets internes

$H$  : nombre de sommets externes de degrés deux

Ces paramètres sont liés via quelques relations ci-dessous établies par Gutman [34] :

$$n = 4h + 2 - n_i \quad (\text{III.1})$$

$$m = 5h + 1 - n_i \quad (\text{III.2})$$

$$m = n + h - 1 \quad (\text{III.3})$$

La base de données BenzDB contient uniquement des benzenoïdes avec un nombre d'hexagones inférieur ou égal à neuf et celle de PAHdb contient uniquement des formules brutes. A partir du logiciel BenzAI, on n'a récupéré le nombre de formules brutes possibles des benzenoïdes classés par nombre de cycles(hexagones). Ainsi, pour un nombre d'hexagones donné, on compte le nombre de benzenoïdes dans BenzDB et pour PAHdb comme il n'est pas possible de filtrer directement les résultats par le nombre d'hexagones, nous avons utilisé les formules brutes pour repérer les formules identiques à celles de BenzDB. Une fois les formules brutes obtenues, j'ai établi des correspondances afin d'identifier les benzenoïdes dans PAHdb identiques à ceux de BenzDB. Parmi les relations nous avons :

$$H = 2h + 4 - n_i \quad (\text{III.4})$$

$$2h + 1 + \left\{ \sqrt{12h - 3} \right\} \leq n \leq 4h + 2 \quad (\text{III.5})$$

Soit  $\{x\}$  le plus petit entier supérieur ou égale à  $x$ , par exemple :  $\{3\} = 3$  et  $\{3.4\} = 4$ . Le nombre de sommets externes de degré deux pour chaque composé nous indique directement le nombre d'hydrogènes d'où la relation (III.4). Et la relation (III.5) nous permet d'obtenir la plage du nombre de sommets  $n$ , correspondant à un nombre donné de cycles hexagonaux. Ainsi, la combinaison de ces équations nous a permis de récupérer les benzenoïdes de la base de données PAHdb et de comparer les composés avec ceux de BenzDB, est la combinaison des équations (III.3) et (III.4) :

$$h = \frac{n - H + 2}{2} \quad (\text{III.6})$$

Enfin, les équations (III.5) et (III.6) ont été organisées et compilées à travers un script python que j'ai établi pour générer les benzenoïdes dans PAHdb identiques à ceux de BenzDB. Ce qui nous permet d'obtenir l'histogramme comparatif voire figure ci-dessous(IV.1) entre BenzDB, pahdb\_v3.20 et pahdb\_v4.00.

### III.3.2 Outils pour sélectionner des structures benzenoïdes dans PAHdb

A partir d'un script shell en python, nous traitons les fichiers xml des bases de données de la NASA pahdb-v3.20 et pahdb-v4.00 pour trier les composés. Après ce tri, les fichiers générés contiennent plusieurs types de composés. Nous sélectionnons uniquement ceux qui nous intéressent c'est-à-dire uniquement les fichiers contenant que des composés neutres avec un nombre défini d'atomes de carbone (C) et d'hydrogène (H). En suite ce fichier est trié pour isoler les potentiels benzenoïdes en extrayant les géométries au format XYZ à l'aide d'un script en annexe que j'ai écrit (A), exécuté comme suit : `python3 convert_xml_to_xyz.py fichier.xml fichier.xyz`. Ce programme nous permet d'éliminer toutes les formules brutes ne correspondant pas à des structures benzenoïdes potentielles. Le résultat issu de ce programme est ensuite converti en format CML via Open Babel, puis traité par un script python (cml.py) développé par un collègue informaticien du laboratoire LIS (AMU). Ce dernier programme permet d'identifier les véritables composés benzenoïdes (hexagones à six atomes de carbone) dans les deux bases.

## III.4 Calculs quantiques réalisés

Dans ce stage, nous avons utilisé des méthodes de chimie computationnelle afin de modéliser avec précision les propriétés électroniques et spectrales des structures benzenoïdes de BenzDB. Ces méthodes reposent essentiellement sur la théorie de la fonctionnelle de la densité (DFT), ainsi sur sa version dépendante du temps (TD-DFT), permettant l'étude des états excités tels que le gap optique.

### III.4.1 Méthode de la théorie de la densité fonctionnelle DFT

La théorie de la fonctionnelle de la densité (DFT) a été introduit en 1964 par Hohenberg et Kohn [35]. Elle repose sur le théorème suivant : si un nombre arbitraire d'électrons dans une boîte est soumis à un potentiel externe  $V(\vec{r})$  et à la répulsion coulombienne électron-électron, ce potentiel est une unique fonctionnelle de la densité électronique  $\rho(\vec{r})$ , à une constante additive près. L'Hamiltonien du système s'écrit comme suit(III.7) :

$$\hat{H} = \hat{T} + \hat{U} + \hat{V} \quad (\text{III.7})$$

ou  $\hat{T}$  et  $\hat{U}$  sont respectivement les opérateurs d'énergie cinétique et d'interaction électron-électron, avec le potentiel  $V$  qui peut s'écrit comme suit :

$$V = \int v(\vec{r})\psi^*(\vec{r})\psi(\vec{r})dr \quad (\text{III.8})$$

La théorie de la fonctionnelle de la densité (DFT), est de nos jours largement utilisée par la communauté des chimistes pour les calculs de structures électroniques, basés sur deux théorèmes essentiels décrit par Hohenberge et Kohn en 1964 [35].

Le premier stipule que la densité électronique d'un système à l'état fondamental détermine de manière unique toutes les propriétés du système, notamment son énergie totale. Le second théorème, connu aussi sous le nom de principe variationnel de la DFT, affirme que la densité de charge  $\rho(\vec{r})$  qui minimise la fonctionnelle d'énergie  $E[\rho]$  est exactement celle de l'état fondamental. Ainsi, tout autre densité  $\rho'(\vec{r})$  associée à un potentiel externe différent, conduit à une énergie supérieure à celle de l'état fondamental (voir annexe C.3). Ces principes ont été formalisés à travers une inégalité fondamentale qui constitue une partie importante du second théorème de Hohenberg et Kohn, mettant en exergue que la fonctionnelle de l'énergie est minimale pour la densité exacte du système. Malgré tout, la méthode de la DFT ne fournit pas exactement de techniques directe pour obtenir cette densité.

Pour combler cette difficulté, Kohn et Sham [36] ont proposé en 1965 une approche pratique basée sur la réécriture de la fonctionnelle de l'énergie dont l'objectif n'est pas de la détailler pour ce stage, produisant la même densité électronique (voir annexe C.4). Dans le cas de densité électronique variant lentement, Kohn et Sham ont proposé l'approximation de la densité locale (LDA)de la DFT, dans laquelle, la fonctionnelle de l'énergie  $E_{xc}(\rho)$  est exprimée en fonction de l'énergie d'échange-corrélation d'un gaz électronique homogène. Cette simplification permet de dériver les équations de Kohn et Sham, qui décrivent les électrons non interactifs évoluant dans un potentiel effectif, réunissant à la fois le potentiel externe, le potentiel de Hartree et le potentiel échange-corrélation (équation de Kohn et Sham voir annexeC.6). Il est essentiel de

noter que toutes les applications de la méthode de la DFT repose sur l'hypothèse que le nombre d'électrons  $N$  est conservé. De plus, pour que les résultats obtenus à travers les équations de Kohn et Sham (KS) soient physiquement pertinents, la densité électronique obtenue doit correspondre à un minimum global de la fonctionnelle d'énergie (voir annexe C.1). Cela signifie que la solution convergée représente bien l'état fondamental du système considéré [37].

Dans le cadre de ce travail, consacré à l'analyse de propriétés croisées des hydrocarbures polycycliques aromatiques (HAP), la théorie de la densité fonctionnelle a été choisie comme la principale méthode pour les calculs de structures électroniques et l'analyse des propriétés associées. Cette méthode, grâce à son excellent compromis entre précision des calculs et coût computationnel, reste particulièrement adaptée pour la modélisation des molécules aromatiques tels que les benzenoïdes. Elle nous permet de prédire des grandeurs essentielles comme le gap, les densités de charges ou des spectres d'absorption obtenus expérimentalement. Les calculs dans ce travail ont été effectués en utilisant les fonctionnelles B3LYP et CAM-B3LYP et des bases de fonctions gaussiennes 6-31+g(d).

### III.4.2 Méthode de la théorie de la densité fonctionnelle dépendante du temps TD-DFT

La théorie de la densité fonctionnelle dépendante du temps (TD-DFT) est de nos jours une approche théorique en chimie quantique très utilisée pour la modélisation des propriétés des états excités électroniques et la simulation de propriétés optiques des molécules organique et inorganique. Cette notion a été introduite pour la première fois par Runge et Gross [38] [39]. Dans l'article [40], les auteurs ont présenté une analyse approfondie des performances de la méthode de la fonctionnelle de densité dépendante du temps (TD-DFT). Elle constitue une extension de la DFT au domaine dynamique, formellement proposée par Runge et Gross, puis dix ans plus tard, Casida [41] l'a rendue opérationnelle en développant un formalisme de réponse linéaire (LR) efficace pour la TD-DFT appelée approximation de phase aléatoire ou équation de Casida [42] permettant de déterminer rapidement et efficacement la solution des équations TD-DFT pour les molécules. Grâce à ces travaux intéressants, la TD-DFT est devenue une approche extrêmement populaire pour la modélisation des énergies, des structures et des propriétés des états électroniquement excités (ES).

Dans ce travail, j'ai utilisé la TD-DFT pour calculer les 10 premiers états excités des benzenoïdes pour déterminer leur gap optique. Les simulations ont été menées en utilisant la fonctionnelle d'échange-corrélation CAM-B3LYP et la base de fonctions gaussiennes 6-31+g(d), garantissant une description adéquate des effets de corrélation et de délocalisation électronique caractéristiques des composés étudiés.

## III.5 Logiciels de Chimie quantique

Dans le cadre de ce stage de recherche centré sur l'analyse croisée des propriétés des hydrocarbures aromatiques polycycliques (HAP), plusieurs outils logiciels ont été utilisés pour la modélisation moléculaire, la simulation quantique et l'analyse de propriétés structurales et spectrales. Dans cette partie, nous décrivons brièvement les principaux logiciels utilisés.

### III.5.1 Logiciel Gaussian

Gaussian [43] est un logiciel de chimie computationnelle largement utilisé pour la modélisation des propriétés moléculaires. Il permet d’effectuer, entre autres, des calculs DFT et TD-DFT pour déterminer des géométries optimisées et leurs énergies associées, des fréquences de vibration, des états excités, spectres électroniques. ... Dans le cadre de ce travail, Gaussian 16 a été utilisé pour réaliser des calculs d’optimisation de géométrie et de calcul de spectres IR des HAP étudiés. Les calculs ont été effectués sur le Cluster de calcul de haute performance de l’équipe CTOM de l’iSm2, permettant de traiter avec efficacité des systèmes de grande taille avec un niveau de théorie approprié tels que B3LYP ou CAM-B3LYP et des bases comme 6-31g(d).

### III.5.2 Logiciel Agui

Agui est une interface graphique conçue pour faciliter la construction, la visualisation et l’édition de structures moléculaire en trois dimensions, ainsi que la visualisation d’orbitales moléculaires et de spectres d’absorption. Dans le cadre de ce stage, ce logiciel a permis la sélection de géométries issues des bases de données étudiées, et la préparation des fichiers d’entrée pour des calculs DFT et TD-DFT avec Gaussian 16.

## III.6 Conclusion

Nous avons mis en place une méthodologie combinant modélisation structurale, scripts Python, et calculs quantiques avec DFT/TD-DFT. Cette approche nous a permis d’extraire efficacement les propriétés des HAP à partir des bases de données telles que BenzDB, PAHdb et COMPAS-3.



# Chapitre IV

## Résultats et discussion

### IV.1 Introduction

Cette partie présente et analyse en détail les résultats obtenus de l'analyse croisée des propriétés des benzenoïdes. En s'appuyant sur une modélisation moléculaire basée sur les outils issus de la théorie des graphes et une analyse détaillée de la base de données BenzDB avec le logiciel BenzAI, nous avons extrait un ensemble de descripteurs topologiques tels que le nombre d'hexagones, les groupes de symétrie, les structures de Kekulé et non Kekulé, le paramètre d'irrégularité et le gap HOMO-LUMO..., permettant de mieux comprendre les propriétés électroniques et spectroscopiques de ces composés.

Dans un premier temps, les différentes bases de données considérées (BenzDB, PAHdb et COMPAS-3), sont étudiées en détail et comparées en fonction de la richesse et de la diversité des structures qu'elles contiennent.

Nous étudions ensuite comment certaines propriétés structurales, spectrales influencent les caractéristiques moléculaires fondamentales telles que l'aromaticité et tentons d'établir une corrélation de ces propriétés pour pouvoir faire une comparaison avec la base de données de la NASA, référence actuelle pour simuler des HAP.

Enfin, nous discutons la pertinence de ces résultats au regard des données issues de la littérature.

### IV.2 Analyse Comparative des bases de données étudiées

#### IV.2.1 Statistiques des bases de données étudiées

Le tableau ci-dessus ([IV.1](#)) met en évidence les différences notables entre les bases de données étudiées. BenzDB couvre de manière systématique l'ensemble des benzenoïdes jusqu'à neuf cycles fusionnés, avec une croissance quasi exponentielle du nombre de composés à mesure que le nombre d'hexagones augmente. Elle atteint un maximum avec 6510 composés à neuf cycles, ce qui en fait la base la plus exhaustive pour cette classe de molécules. À l'inverse, PAHdb, dans ses versions v3.20 et v4.00, présente une sous-représentation importante des benzenoïdes, surtout à partir de 7 hexagones. Bien que la version v4.00 introduise quelques composés à 10 cycles (35), elle reste très inférieure en couverture par rapport à BenzDB. Cela reflète une absence de génération systématique des structures benzenoïdes dans PAHdb. Enfin, la base COMPAS-3 se distingue par sa concentration sur les grands benzenoïdes. Elle contient très peu de composés en

dessous de 6 cycles, mais devient extrêmement fournie à partir de 9 hexagones, atteignant 1531 composés à 9 cycles et 6891 à 10 cycles. Elle cible donc des structures compactes et d'intérêt pour la science des matériaux.

Nb of hexagons	BenzDB	PAHdb_v3.20	PAHdb_v4.00	CompasDB
1	1	0	0	0
2	1	1	1	0
3	3	3	3	0
4	7	7	7	1
5	22	21	20	3
6	81	76	76	15
7	331	288	288	68
8	1436	1157	1157	329
9	6510	4	8	1531
10	0	8	35	6891

TABLE IV.1 – Nombre de composés selon le nombre d'hexagones dans chaque base de données

## IV.2.2 Visualisation des composés des bases étudiées

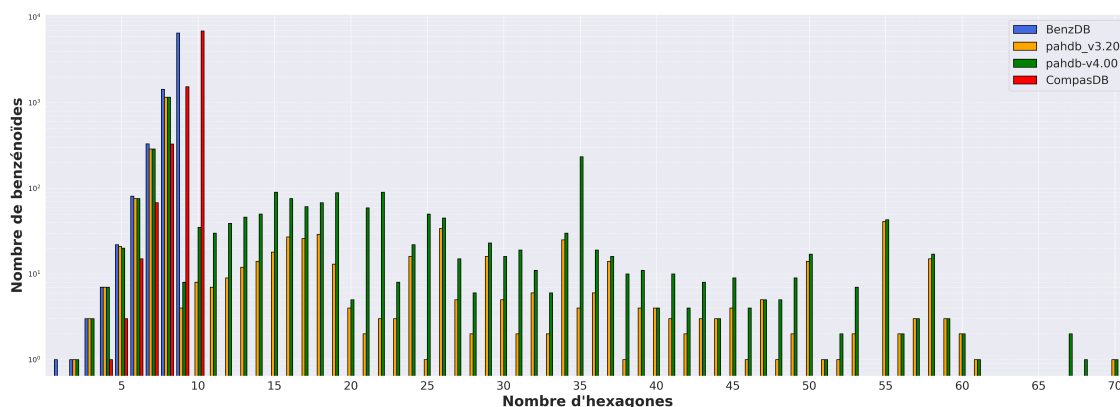


FIGURE IV.1 – Histogramme comparatif entre les bases de données BenzDB, CompasDB et PAHdb (échelle logarithmique)

Dans le cadre de cette étude, les bases de données BenzDB, COMPAS-3, et les deux versions de la NASA (PAHdb\_v3.20 et PAHdb\_v4.00) ont été comparées en termes de distribution du nombre d'hexagones par molécule. L'historgramme présenté ci-dessus (Figure IV.1) permet de mettre en lumière les différentes statistiques majeures entre ces bases.

On observe tout d'abord que nombre de benzenoïdes croît de façon exponentielle avec le nombre de cycles, et lorsqu'on ajoute un cycle supplémentaire à partir des benzenoïdes de taille 5, le nombre de benzenoïdes est multiplié par un facteur  $\simeq 4$ . On voit également que BenzDB possède le plus grand nombre de benzenoïdes jusqu'à 9 cycles fusionnés, comparativement aux deux autres bases de données. Ceci n'est pas surprenant car les benzenoïdes qu'elles contient ont été générés de façon systématique à l'aide de la programmation par contrainte, assurant

l'exhaustivité des types de benzenoïdes générés, à l'exception de ceux possédant des motifs hélicéniques à 6 cycles ou plus (spirale de benzènes fusionnés). Ainsi, cette base présente un atout majeur pour l'étude systématique des corrélations entre structures et propriétés sur une grande diversité de tailles et topologies des benzenoïdes. La base PAHdb avec ses deux versions v3.20 et v4.00, contient un grand nombre de HAP allant jusqu'à 168 hexagones fusionnés. Mais le nombre de benzenoïdes de taille supérieure à 70 est faible (quelques dizaines). Pour des raisons de lisibilité de la figure (IV.1), nous n'avons affiché que les benzenoïdes de taille inférieure ou égale à 70 cycles. La répartition des benzenoïdes de taille supérieure à 10 apparaît relativement homogène : jusqu'à trente benzenoïdes par taille fixée pour la version v3.20, une centaine pour la version v4.00. À noter toutefois, le nombre extrêmement faible de benzenoïdes de taille 9 et 10, comparés autres bases BenzDB et COMPAS-3. La base COMPAS-3 (CompassDB sur l'histogramme) contient essentiellement, quant à elle, des composés de taille 9 et 10.

### IV.3 Répartition des benzenoïdes de BenzDB selon leur symétrie

#### IV.3.1 Classification des composés dans BenzDB selon les critères structuraux et symétries

Le tableau ci-dessous (IV.2) présente une classification détaillée, par symétrie, des composés de BenzDB, selon plusieurs critères structuraux :

- structures à couches fermées (Kekuléens) ou possédant des électrons célibataires (non-Kekuléens)
- topologie ramifiée (catacondensés) ou plutôt compacte (périmcondensés)
- existence de trous (coronoïdes)

Sans surprise, les benzenoïdes sans symétrie ( $C_1$ ) ou ne possédant qu'un seul plan de symétrie, souvent le plan moléculaire, ( $C_s$ ) sont les plus nombreux. Il n'était a priori pas attendu que les composés à couches fermées et les composés radicalaires sont en proportion 50/50 environ. Il existe également environ trois fois plus de périmcondensés ( $\simeq 6000$ ) que de catacondensés ( $\simeq 2000$ ) alors que la répartition entre ces deux types parmi les composés de plus haute symétrie est beaucoup plus homogène.

Symétries	Kekuléens	Non-Kekuléens	Catacondensés	Péricondensés	Coronoïdes
$C_s$ ou $C_1$	3913	4049	1897	6065	3
$C_{2va}$	175	0	84	91	1
$C_{2vb}$	50	77	36	91	0
$C_{2hi}$	44	0	22	22	0
$C_{2hii}$	43	0	22	21	0
$D_{2hi}$	13	0	6	7	0
$D_{2hii}$	13	0	7	6	1
$C_{3hi}$	1	0	1	0	0
$C_{3hii}$	0	6	0	6	0
$D_{3hia}$	2	0	2	0	0
$D_{3hii}$	1	3	1	3	1
$C_{6h}$	0	0	0	0	0
$D_{6h}$	2	0	1	1	0

TABLE IV.2 – Nombre de composés selon le type de critères et symétries dans la base de données BenzDB

### IV.3.2 Visualisation de la distribution des composés en fonction des groupes de symétrie

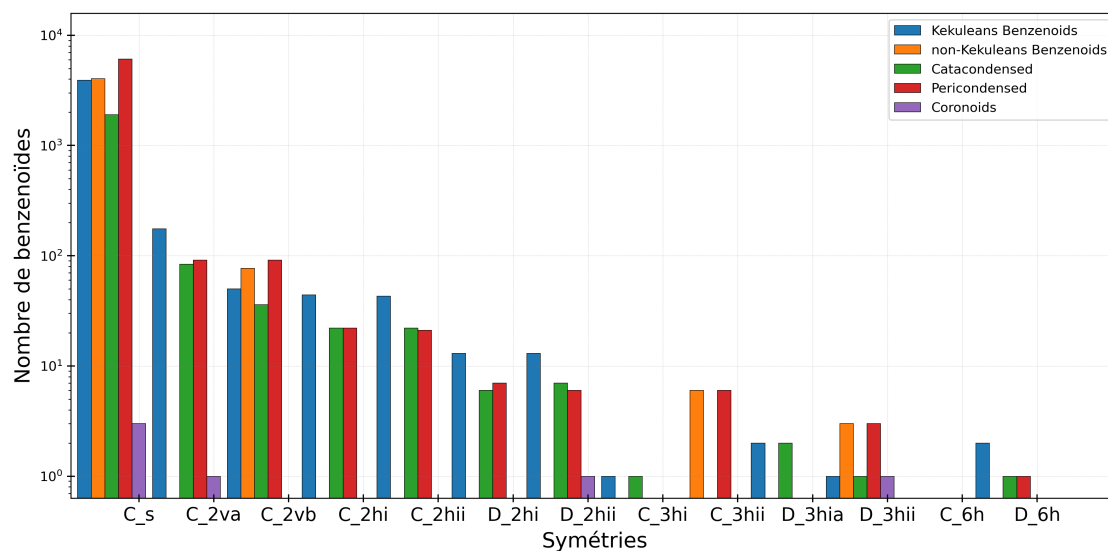


FIGURE IV.2 – Benzenoïdes par symétrie dans BenzDB

La figure (IV.2) présente une distribution des composés selon leur symétrie, en distinguant plusieurs familles telles que les kekuléens, non-kekuléens, catacondensés, péricondensés et les coronoïdes. On observe une forte présence de la symétrie  $C_s$  ou  $C_1$ , qui regroupe la majorité des composés pour toutes les familles. Cela montre la prédominance des composés à faible symétrie dans BenzDB, notamment pour les kekuléens et péricondensés. A partir de  $C_{2va}$ , on note une

diminution progressive des composés pour toutes les symétries élevées, affirmant la rareté des composés plus la symétrie est très élevée dans la base BenzDB. Elles correspondent à des molécules plus régulières et hautement symétriques, qui sont minoritaire dans certaines familles. Les coronoides apparaissent faiblement dans certaines symétries comme  $C_{2va}$ ,  $D_{2hii}$  et  $D_{3hii}$ , ce qui reflète leur forme cyclique particulière. Cette diversité est plus visible chez les kekuléens et péricondensés, alors que les catacondensés présentent une distribution focalisée sur des symétries plus simples.

Cette distribution dans BenzDB traduit du fait que la plus part des composés générés ou répertoriés sont des structures à faibles symétries, en corrélation avec la complexité topologique et la variété des isomères possibles à nombre d'hexagones donnés. La symétrie est donc un paramètre clé qui peut être corrélé à d'autres propriétés électroniques ou spectroscopiques.

## IV.4 Étude des benzenoïdes par intervalle de Gap (eV)

### IV.4.1 Répartition des composés dans BenzDB selon leur gap (eV)

Le tableau (IV.3) représente la répartition du nombre de composés benzenoïdes en fonction de leur gap énergétique (HOMO-LUMO) exprimé en électronvolts (eV). Cette grandeur est importante pour évaluer les propriétés électroniques des molécules, à savoir leur aromaticité, stabilité, réactivité. La classification par intervalle permet de mettre en exergue les tendances de distribution des structures moléculaires de benzenoïdes étudiées au cours de ce stage.

Intervalle du Gap (eV)	Nombre de Benzénoïdes
[0.13 :0.81[	376
[0.81 :1.49[	116
[1.49 :2.18[	385
[2.18 :2.86[	3687
[2.86 :3.54[	2790
[3.54 :4.22[	1024
[4.22 :4.91[	11
[4.91 :5.59[	2
[5.59 :6.27[	0
[6.27 :6.95[	1

TABLE IV.3 – Nombre de composés selon l'intervalle du Gap HOMO-LUMO dans BenzDB

On observe qu'environ 50 % (3687) des composés de BenzDB ont un gap HOMO-LUMO entre 2,2 et 2,9 eV. Si l'on ajoute ceux entre 2,9 et 3,5 eV, cela représente près de 80 % (6477) des composés. Les composés avec des gaps très faibles ou très élevés sont peu nombreux.

## IV.4.2 visualisation des benzenoïdes en fonction du gap homo-lumo

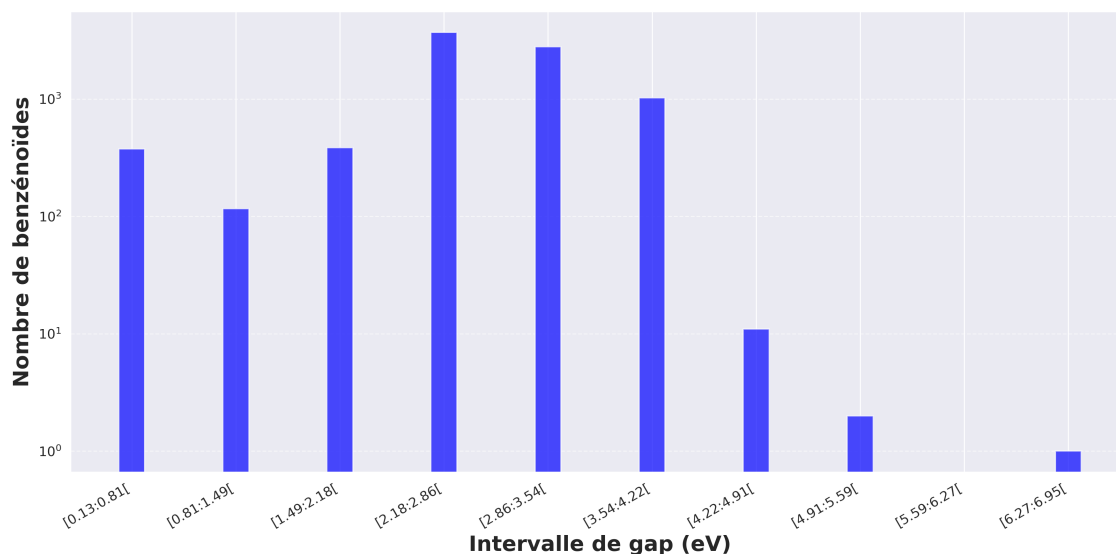


FIGURE IV.3 – Benzenoïdes en fonction du gap homo-lumo dans BenzDB

La figure (IV.3) visualise la distribution des composés dans BenzDB par intervalles de valeurs de gaps données calculées par la méthode de la DFT (B3LYP/6-31G). On observe que la majorité des composés ont un gaps situés entre 2.18eV et 4.22eV. Cette plage dominante indique que la plus part des molécules de BenzDB présentent une stabilité électronique importante, caractéristiques des systèmes peu conjugués ou saturés. Pour le nombre de composés au-delà de 4.22 eV, la distribution devient moins régulière. Ainsi, on peut dire que la base de données BenzDB est principalement constituée de molécules stables, avec des gaps significatifs cohérents avec des structures peu réactives. Ce sont généralement des molécules aromatiques stables présentant des propriétés électroniques intermédiaires.

## IV.5 Étude des benzenoides par intervalle d'irrégularité

### IV.5.1 Répartition des compoés dans BenzDB selon le paramètre d'irrégularité

Le tableau (IV.4) nous donne la répartition des composés dans la base de données BenzDB en fonction de leur paramètre d'irrégularité. Ce paramètre issue de l'application de la théorie des graphe dans la modélisation moléculaire, permet d'évaluer la répartition asymétrique des liaisons ou cycles au sein de la molécule.

Intervalle d'irrégularité	Nombre de Benzénoïdes
[0.0 :0.1[	3
[0.1 :0.2[	2
[0.2 :0.3[	8
[0.3 :0.4[	1026
[0.4 :0.5[	635
[0.5 :0.6[	3284
[0.6 :0.7[	1600
[0.7 :0.8[	1455
[0.8 :0.9[	322
[0.9 :1.0[	56

TABLE IV.4 – Nombres de composés selon l'intervalle du nombre d'irrégularité dans BenzDB

## IV.5.2 Visualisation des benzenoïdes en fonction des paramètres irrégularités

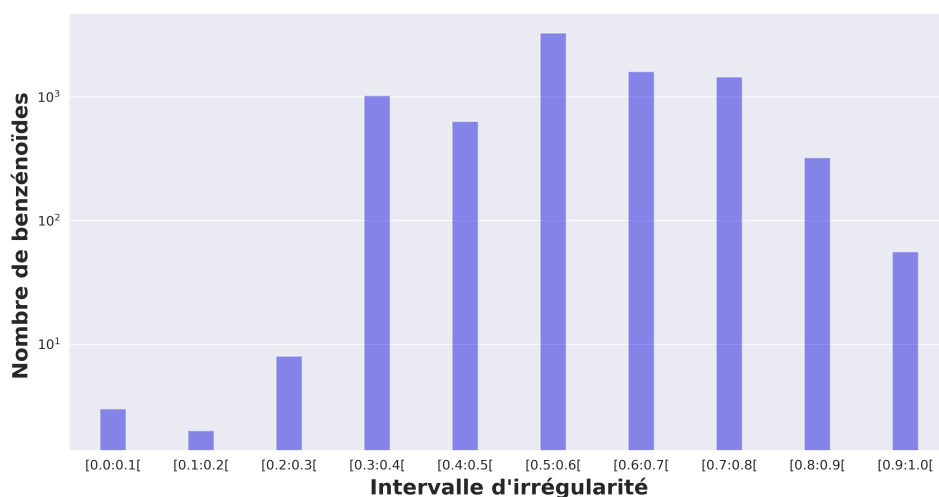


FIGURE IV.4 – Benzenoïdes en fonction des paramètres d'irrégularités dans BenzDB

La figure (IV.4) illustre la répartition des composés benzenoïdes dans BenzDB selon leur paramètre d'irrégularité  $\xi$ . On observe que la totalité des structures étudiées présentent une irrégularité intermédiaire, principalement située dans l'intervalle  $0.3 \leq \xi < 0.8$ . L'intervalle  $0.5 \leq \xi < 0.6$  concentre le plus grand nombre de benzenoïdes. Par contre, les benzenoïdes très réguliers ( $\xi < 0.3$ ) sont nettement moins fréquents. Cette distribution observée peut s'expliquer par la nature des motifs structuraux des benzenoïdes. Ainsi, les composés de valeur d'irrégularité moyenne regroupent des proportions équilibrées d'hydrogènes solo, duo, trio et quartet, qui sont courants dans les structures observées.

Les benzenoïdes très réguliers correspondent à des HAP avec une grande symétrie, peu fréquents dans BenzDB car nécessitant des géométries particulières. De même, les benzenoïdes très irréguliers, riche en hydrogène trio et quartet ne sont pas nombreux dans BenzDB du fait leur

géométrie et structure complexes.

## IV.6 Corrélation entre propriétés structurales, spectrales et d'aromaticité

Dans ce travail, nous nous intéressons en premièrement à une corrélation entre le nombre de ronds de Clar et le gap optique pour estimer l'aromaticité locale et en second lieu, une corrélation entre structures moléculaires (par exemple, le paramètre d'irrégularité et topologie des HAP) et la position, l'intensité des bandes de spectres IR observées. Nous rapportons une étude théorique des spectres infrarouges (IR) des HAP avec uniquement des benzenoïdes, des composés à six cycles fusionnés. Les calculs théoriques sont principalement effectués en utilisant la théorie hybride de la fonctionnelle de la densité (DFT) B3LYP avec un ensemble de base gaussienne 6-31G(d).

### IV.6.1 corrélation entre nombre de ronds de Clar, topologie, délocalisation $\pi$ et gap optique

La figures ci-dessous (IV.5), démontre une corrélation croissante entre le nombre de structure de Clar (sextets de Clar) et le gap optique (gap HOMO-LUMO) dans les HAP à neuf hexagones sans symétrie  $C_s$  dans BenzDB.

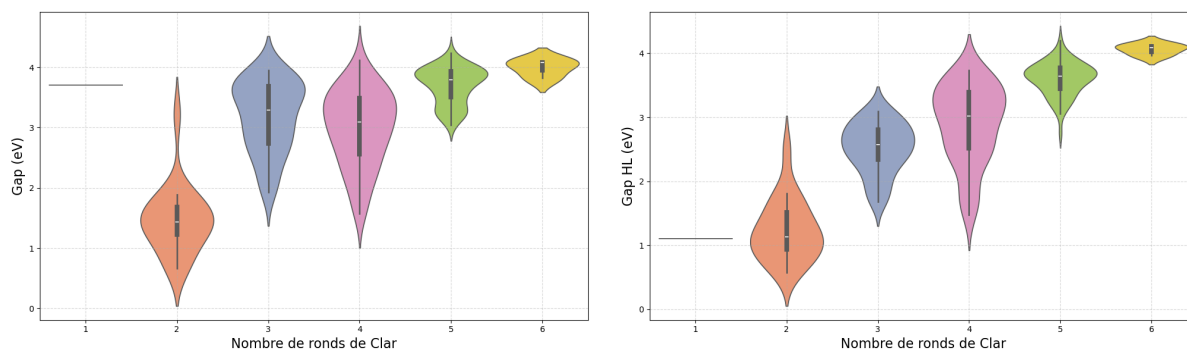


FIGURE IV.5 – Illustration de la distribution des gaps HOMO-LUMO et optique en fonction du nombre de ronds de Clar

On observe une tendance globale croissante du gap optique avec le nombre de structures de Clar (ronds de Clar), plus une molécule possède de sextets de Clar, plus son gap optique est élevé. Cette tendance s'interprète en s'appuyant sur les résultats de Adkins & et Miller (2017) [11], qui montre que les sextets de Clar représentent des zones de forte délocalisation électronique locale. Lorsque le nombre de structures de Clar augmente, la délocalisation des électrons  $\pi$  devient plus localisée autour de ces cycles stables, au détriment d'une délocalisation globale à travers tout le squelette moléculaire. Cette réduction de la conjugaison étendue conduit à une augmentation du gap HOMO-LUMO, et par conséquent, du gap optique.

Ainsi, cette étude illustre clairement que le nombre de ronds de Clar des composés benzenoïdes influencent fortement leurs propriétés électroniques. Plus une molécule possède un nombre donné



de structure de Clar, plus son comportement s'approche de celui du benzène (aromatique), et plus son gap optique est élevé.

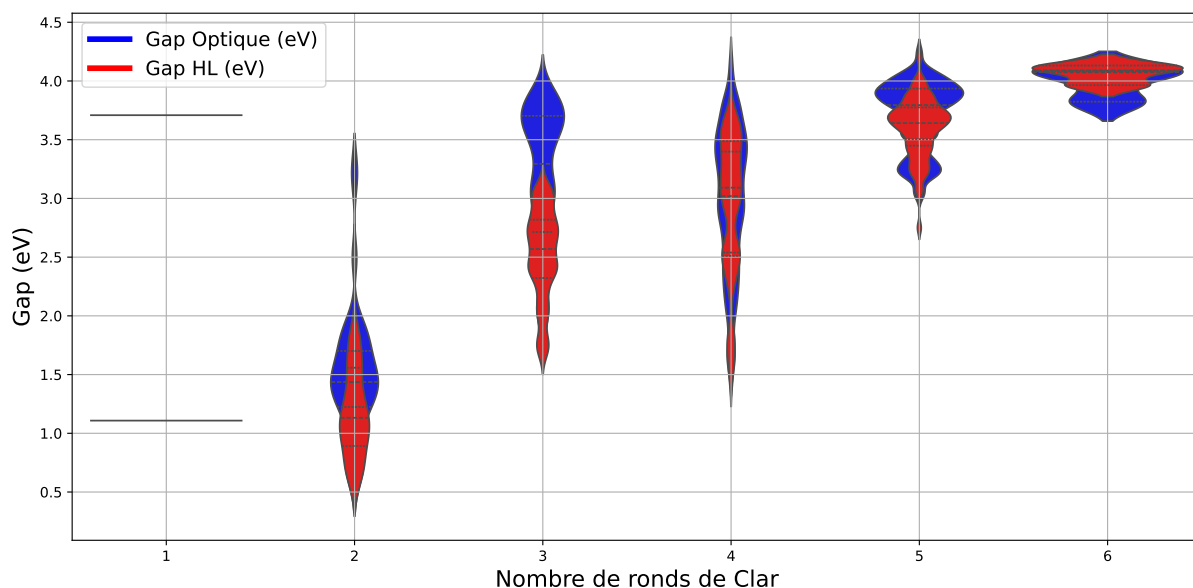


FIGURE IV.6 – Gap optique et HOMO-LUMO en fonction du nombre de ronds de Clar

La figure (IV.6) regroupe les gaps (eV) optiques (0,6630 - 4,2396) et HOMO-LUMO (0,5772 - 4,1979) par nombre de ronds de Clar (1 à 6). On constat que le gap optique varie largement, reflétant les transitions électroniques des états excités, tandis que, le gaps HOMO-LUMO, un peu plus bas, mesure la différence entre les orbitales moléculaires HOMO-LUMO. La dispersion des violons observée indique une hétérogénéité selon les ronds de Clar.

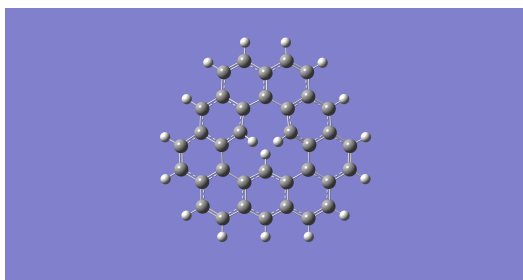
Généralement on note une tendance ascendente des gaps optiques (3.0 à 4.0 eV) avec le nombre de ronds de Clar, ce qui suggère une stabilisation des orbitales  $\pi$ . Alors que les gap HOMO-LUMO varient moins, entrainant une stabilité reduite des orbitales  $\pi$ . Ainsi, une corrélation positive semble exister pour les gaps optiques, moins clair pour les gaps HOMO-LUMO.

#### IV.6.2 Corrélation entre structures moléculaires et spectres IR d'un HAP

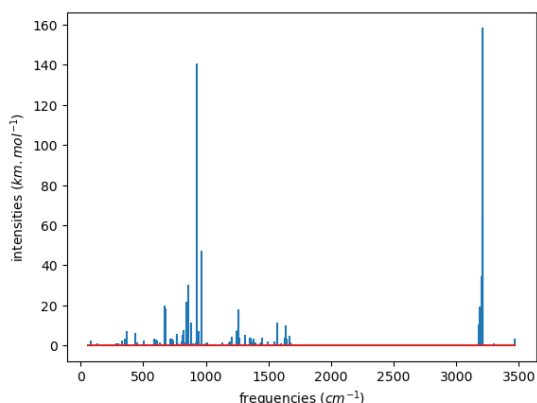
L'étude de Bauschlicher et al. (2009) a mis en évidence un lien clair entre la distribution des types d'hydrogènes périphériques dans les HAP (solo, duo, trio, quartet) et les caractéristiques du spectres infrarouge (IR). En particulier, les modes de vibration C-H responsables des bandes IR intenses sont fortement influencés par le type de voisinage des atomes d'hydrogènes. Pour mieux illustrer cette relation, le tableau suivant présente la correspondance entre les domaines spectraux (en  $cm^{-1}$ ) et les longueurs d'onde (en  $\mu m$ ) des principales vibrations observées dans les HAP.

Types de vibration	Domaine de vibration $cm^{-1}$	Équivalence en $\mu m : \frac{10^4}{V(cm)}$
C-H hors du plan	700–900	14–11.1
C-H élongation (stretch)	300–3200	3.3–3.1

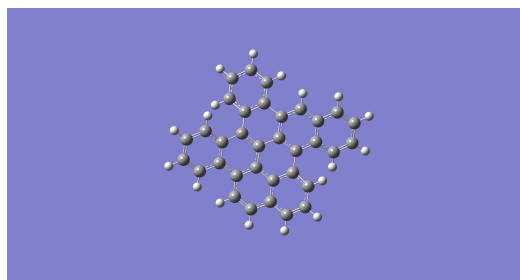
TABLE IV.5 – Correspondance entre longueurs d'onde et nombres d'onde pour les vibrations C-H



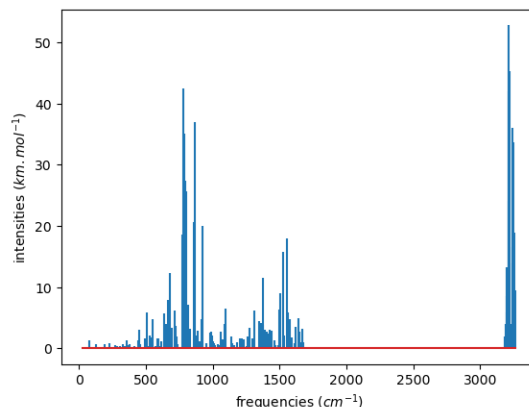
Molécule : A



Spectre IR : A



Molécule : B



Spectre IR : B

A partir de cette base spectroscopique, nous avons analysé les spectres IR de plusieurs composés issus de BenzDB (deux molécules illustrées dans ce rapport) afin d'explorer plus en détail comment la distribution des hydrogènes périphériques influence la position et l'intensité des bandes observées.

La molécule A (ID :5570) dominée par des hydrogènes solo et duo ( $N_{solo} = 6$ ,  $N_{duo} = 12$ ,  $N_{trio} = 0$ ,  $N_{quartet} = 0$ ) avec un paramètre d'irrégularité  $\xi = 0$ , son spectre IR montre des bandes forte entre  $800 - 900 \text{ cm}^{-1}$  et autour de  $3100 \text{ cm}^{-1}$ , caractéristiques des vibrations solo et duo C-H hors du plan et des vibrations d'élongation C-H dans le plan. Ces modes sont bien décrits et localisés, produisant des pics nets et marqués.

La molécule B (ID :464) riche en hydrogènes quartet ( $N_{solo} = 1$ ,  $N_{duo} = 2$ ,  $N_{trio} = 3$ ,  $N_{quartet} = 12$ ), son spectre est caractérisé par des bandes plus diffuses et un profil plus complexe autour de  $700 - 850 \text{ cm}^{-1}$  caractéristique des vibrations trio et quartet C-H hors du plan et dans la zone  $3100 - 3200 \text{ cm}^{-1}$ . Cela reflète la nature plus collective et couplée des vibrations C-H dans des environnements quartet, moins bien définis que les solos.

En conclusion, on peut dire que les hydrogènes solos et duos sont associés à des bandes IR plus intenses et plus localisées, tandis que les trios et quartes induisent des vibrations plus diffuses. Cette corrélation permet d'utiliser la topologie (hydrogènes périphériques) comme indicateur pour l'identification des HAP dans des environnements astrophysiques.

#### IV.6.3 Relation entre le nombre de ronds de Clar et le HOMA

Dans le but d'illustrer la corrélation entre le nombre de ronds de Clar (critère topologique) et l'indice HOMA (critère géométrique) nous avons mené une première analyse sur un échantillon représentatif de 6 composés à 9 hexagones extraits de la base de données BenzDB, couvrant des

structures ayant entre 1 et 6 ronds de Clar.

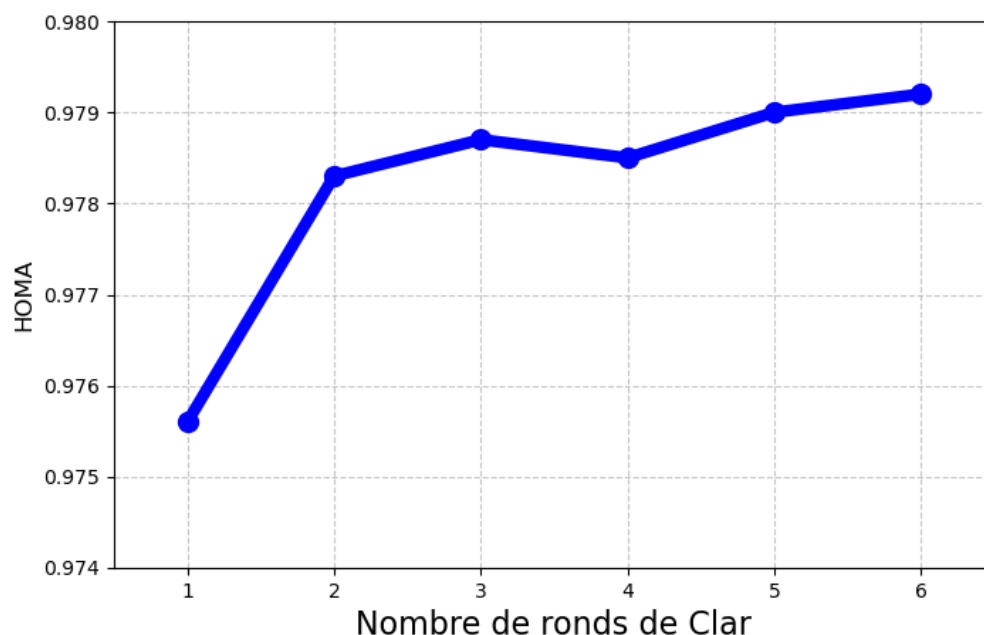


FIGURE IV.7 – L'indice HOMA en fonction du nombre de ronds de Clar

Cette étude montre une tendance globale croissante de l'indice HOMA avec le nombre de ronds de Clar. Ainsi, les structures présentant un plus grand nombre de ronds de Clar tendent à avoir une valeur HOMA plus proche de 1, traduisant une meilleure délocalisation des électrons  $\pi$ , donc une aromaticité globale plus marquée.

Il convient toutefois de souligner que cette tendance est ici illustrée à titre d'exemple et que sa validité statistique est à vérifier sur l'ensemble de la base BenzDB, qui contient 8392 composés benzenoïdes. Cette corrélation met en lumière la complémentarité entre les critères topologiques (Clar) et structuraux (HOMA) pour une caractérisation fine de l'aromaticité des HAP.

## IV.7 Conclusion

L'analyse a révélé des corrélations entre les propriétés topologiques, optoélectroniques et l'aromaticité des HAP. Les descripteurs comme le nombre de sextets Clar et l'irrégularité s'avèrent pertinents pour prédire leurs signatures spectrales et aromaticité, ouvrant des perspectives pour la modélisation en astrophysique et en science des matériaux.

## Chapitre V

# Conclusion

Au cours de ce stage porté sur l'analyse de propriétés croisées d'hydrocarbures aromatiques polycycliques (HAP) à l'aide d'outils issus de la théorie des graphes, plusieurs problèmes ont été abordés relatives aux HAP, particulièrement aux benzenoïdes en s'appuyant sur la base de données BenzDB. L'objectif était de faire une analyse complète de cette dernière et de mettre en relation des descripteurs topologiques tels que le nombre de ronds de Clar et le paramètre d'irrégularité avec des propriétés optoélectronique (gap optique et gap HOMO-LUMO) et les signatures spectrales infrarouges des composés étudiés.

Nous avons tout d'abord donné un cadre théorique combinant des outils issus de la théorie des graphes pour la modélisation des structures moléculaires et des calculs de chimie quantique pour les propriétés électroniques et spectroscopiques, en utilisant notamment le logiciel Gaussian 16. L'aromaticité, locale (ronds de Clar) et globale (HOMA), a été étudiée comme un excellent indicateur reliant la structure à la stabilité et aux propriétés optiques des molécules.

Pour la partie méthodologie, nous avons utilisé principalement les méthodes DFT et DT-DFT avec comme fonctionnelles et bases gaussiennes B3LYP (CAM-B3LYP)/6-31G(6-31+G(d)) via le logiciel Gaussian, le logiciel Agui pour préparer certains fichiers d'optimisation géométriques et des scripts python pour l'analyse des composés (BenzDB, PAHdb et COMPAS-3) et la simulation de spectre issus de calculs DFT.

Sur le plan des résultats obtenus, nous avons constaté que la majorité des composés de BenzDB présentent une irrégularité intermédiaire  $\xi$  compris entre 0.3 et 0.7, les composés beaucoup plus réguliers étant moins nombreux dans la base.

On a établi une corrélation entre le nombre de Clar et les gap optiques et HOMO-LUMO, les composés avec un plus grand nombre de Clar tendent à avoir un gap plus élevé traduisant une plus grande stabilité aromatique.

L'analyse des spectres IR a montré que la distribution des hydrogènes périphériques (solo, duo, trio, quartet) influencent directement la position et l'intensité des bandes IR, en accord avec les résultats de Bauschlicher et al., 2009. Nous avons illustré comment le paramètre d'irrégularité  $\xi$  peut nous servir de prédicteur des signatures spectroscopiques des HAP.

Nous avons essayé aussi de mettre en lumière la relation qui existe entre le nombre de ronds de Clar et l'indice de HOMA pour avoir une vision globale de l'aromaticité des structures dans BenzDB.

L'ensemble de ces résultats contribue à une meilleure compréhension des relations structures et

propriétés des HAP et donne des outils importants pour la modélisation de leurs spectres dans un contexte astrophysique où ces molécules sont envisagées comme des candidats pour expliquer certains certaines bandes observées.

En définitive, ce travail ouvre plusieurs perspectives, notamment la génération de nouvelles familles de benzénoïdes avec BenzAI pour vérifier ces corrélations sur des composés de plus de 9 hexagones dans BenzDB. Il serait aussi intéressant de développer des outils automatisés d'analyse spectroscopique intégrant des techniques d'intelligence artificielle (IA) pour prédire les spectres à partir des descripteurs topologiques comme le nombre de ronds de Clar. Mais aussi de valider les corrélations établies ici par comparaison directe avec des spectres expérimentaux. Ainsi, ce stage nous a permis d'avoir les compétences d'une méthodologie combinant modélisation structurale, développement de scripts pour l'analyse des données et calculs quantiques pour l'étude des HAP.

# Bibliographie

- [1] Adrien Varet. Soutenue à Aix-Marseille Université le 13 décembre 2022 par. *Aix-Marseille Université*, 2022.
- [2] Ramachandran Rakhi and Cherumuttathu H Suresh. Optoelectronic properties of polycyclic benzenoid hydrocarbons of various sizes and shapes for donor- $\pi$ -acceptor systems : a dft study. *ChemistrySelect*, 6(11) :2760–2769, 2021.
- [3] J Bouwman, P Castellanos, M Bulak, J Terwisscha Van Scheltinga, J Cami, H Linnartz, and AGGM Tielens. Effect of molecular structure on the infrared signatures of astronomically relevant pahs. *Astronomy & Astrophysics*, 621 :A80, 2019.
- [4] Adrien Varet, Nicolas Prcovic, Cyril Terrioux, Denis Hagebaum-Reignier, and Yannick Carrissan. BenzAI : A Program to Design Benzenoids with Defined Properties Using Constraint Programming. *Journal of Chemical Information and Modeling*, 62(11) :2811–2820, 2022.
- [5] Yajun Wang. Quantitative resonance theory based on the clar sextet model. *The Journal of Physical Chemistry A*, 126(1) :164–176, 2022. Epub 2022 Jan 2.
- [6] Tanguy Jousselin-Oba, Masashi Mamada, Karen Wright, Jérôme Marrot, Chihaya Adachi, Abderrahim Yassar, and Michel Frigoli. Synthesis, aromaticity, and application of peri-pentacenopentacene : Localized representation of benzenoid aromatic compounds. *Angewandte Chemie International Edition*, 61(1) :e202112794, 2022.
- [7] <https://benzai-team.github.io/BenzAI/> ; consulté en mars 2025’.
- [8] Alexandra Wahab and Renana Gershoni-Poranne. COMPAS-3 : A dataset of *peri* -condensed polybenzenoid hydrocarbons. *Physical Chemistry Chemical Physics*, 26(21) :15344–15357, 2024.
- [9] <https://www.astrochem.org/pahdb/> ; consulté en mars 2025’.
- [10] Paul Von Ragué Schleyer. Introduction : Aromaticity. *Chemical Reviews*, 101(5) :1115–1118, 2001.
- [11] Erin M. Adkins and J. Houston Miller. Towards a taxonomy of topology for polynuclear aromatic hydrocarbons : Linking electronic and molecular structure. *Physical Chemistry Chemical Physics*, 19(41) :28458–28469, 2017.
- [12] Miquel Solà. Forty years of Clar’s aromatic  $\pi$ -sextet rule. *Frontiers in Chemistry*, 1, 2013.
- [13] J. Bouwman, P. Castellanos, M. Bulak, J. Terwisscha Van Scheltinga, J. Cami, H. Linnartz, and A. G. G. M. Tielens. Effect of molecular structure on the infrared signatures of astronomically relevant PAHs. *Astronomy & Astrophysics*, 621 :A80, 2019.

- [14] Douglas M Hudgins and LJ Allamandola. Interstellar pah emission in the 11-14 micron region : New insights from laboratory data and a tracer of ionized pahas. *The Astrophysical Journal*, 516(1) :L41, 1999.
- [15] Alessandra Ricca, Joseph E. Roser, Els Peeters, and Christiaan Boersma. Polycyclic Aromatic Hydrocarbons with Armchair Edges : Potential Emitters in Class B Sources. *The Astrophysical Journal*, 882(1) :56, 2019.
- [16] S. Hony, C. Van Kerckhoven, E. Peeters, A. G. G. M. Tielens, D. M. Hudgins, and L. J. Allamandola. The CH out-of-plane bending modes of PAH molecules in astrophysical environments. *Astronomy & Astrophysics*, 370(3) :1030–1043, 2001.
- [17] Friedrich August Kekulé. Ueber die constitution und die metamorphosen der chemischen verbindungen und über die chemische natur des kohlenstoffs. *Berichte der Deutschen Chemischen Gesellschaft*, 8(2) :182–196, 1865.
- [18] <https://new.societechimiquedefrance.fr/produits/friedrich-august-kekule> ; consulté en mai 2025’.
- [19] Frank Close. Personne n’aurait même pu imaginer les transformations que connaîtrait la science au cours des cent années qui allaient suivre. *Cern*, 1900.
- [20] J. J. Thomson. Cathode rays. *Philosophical Magazine*, 44(269) :293–316, 1897.
- [21] Yosadara Ruiz-Morales. Homo- lumo gap as an index of molecular size and structure for polycyclic aromatic hydrocarbons (pahs) and asphaltenes : A theoretical study. i. *The Journal of Physical Chemistry A*, 106(46) :11283–11308, 2002.
- [22] Tadeusz Marek Krygowski and Michał Ksawery Cyrański. Structural aspects of aromaticity. *Chemical Reviews*, 101(5) :1385–1420, 2001.
- [23] Nikolaos Kateris, Amitesh S Jayaraman, and Hai Wang. Homo-lumo gaps of large polycyclic aromatic hydrocarbons and their implication on the quantum confinement behavior of flame-formed carbon nanoparticles. *Proceedings of the Combustion Institute*, 39(1) :1069–1077, 2023.
- [24] Sihan Zhao, Gabriela Borin Barin, Ting Cao, Jan Overbeck, Rimah Darawish, Tairu Lyu, Steve Drapcho, Sheng Wang, Tim Dumlaff, Akimitsu Narita, et al. Optical imaging and spectroscopy of atomically precise armchair graphene nanoribbons. *Nano letters*, 20(2) :1124–1130, 2020.
- [25] Tim Dumlaff, Yanwei Gu, Giuseppe M Paternò, Zijie Qiu, Ali Maghsoumi, Matteo Tomasini, Xinliang Feng, Francesco Scotognella, Akimitsu Narita, and Klaus Müllen. Hexaperi-benzocoronene with two extra k-regions in an ortho-configuration. *Chemical Science*, 11(47) :12816–12821, 2020.
- [26] ML Heger. Lick obs. *Bull*, 10(337) :146–147, 1922.
- [27] Xiaofeng Tan. Towards a comprehensive electronic database of polycyclic aromatic hydrocarbons and its application in constraining the identities of possible carriers of the diffuse interstellar bands. *Spectrochimica Acta Part A : Molecular and Biomolecular Spectroscopy*, 71(5) :2005–2011, 2009.

- [28] Masiyappan Karuppusamy, Shyam Vinod Kumar Panneer, Abigail Jennifer G, Elumalai Varathan, Mahesh Kumar Ravva, and Venkatesan Subramanian. Structure-aromaticity-spectroscopy relationship in conjugated polymers. *Theoretical Chemistry Accounts*, 142(5) :51, 2023.
- [29] Charles W. Bauschlicher, Els Peeters, and Louis J. Allamandola. the infrared spectra of very large irregular polycyclic aromatic hydrocarbons (PAHs) : observational probes of astronomical PAH geometry, size, and charge. *The Astrophysical Journal*, 697(1) :311–327, 2009.
- [30] Nenad Trinajstić. *Chemical graph theory*. CRC press, 2018.
- [31] Gilles Caporossi and Pierre Hansen. Enumeration of polyhex hydrocarbons to  $h=21$ . *Journal of chemical information and computer sciences*, 38(4) :610–619, 1998.
- [32] Ramón García-Domenech, Jorge Gálvez, Jesus V. De Julián-Ortiz, and Lionello Pogliani. Some New Trends in Chemical Graph Theory. *Chemical Reviews*, 108(3) :1127–1169, 2008.
- [33] Alexandru T. Balaban. Applications of graph theory in chemistry. *Journal of Chemical Information and Computer Sciences*, 25(3) :334–343, 1985.
- [34] Ivan Gutman and Sven Josef Cyvin. *Introduction to the Theory of Benzenoid Hydrocarbons*. Springer Berlin Heidelberg, Berlin, Heidelberg, 1989.
- [35] Pierre Hohenberg and Walter Kohn. Inhomogeneous electron gas. *Physical review*, 136(3B) :B864, 1964.
- [36] Walter Kohn and Lu Jeu Sham. Self-consistent equations including exchange and correlation effects. *Physical review*, 140(4A) :A1133, 1965.
- [37] Diola Bagayoko. Understanding density functional theory (dft) and completing it in practice. *AIP Advances*, 4(12), 2014.
- [38] Erich Runge and Eberhard KU Gross. Density-functional theory for time-dependent systems. *Physical review letters*, 52(12) :997, 1984.
- [39] E.K.U. Gross, J.F. Dobson, and M. Petersilka. Density functional theory of time-dependent phenomena. *Density Functional Theory II : Relativistic and Time Dependent Extensions*, pages 81–172, 2005.
- [40] Adele D Laurent and Denis Jacquemin. Td-dft benchmarks : a review. *International Journal of Quantum Chemistry*, 113(17) :2019–2039, 2013.
- [41] M.E. Casida. Recent advances in density functional methods part i, world scientific. *World Scientific*, 1995.
- [42] M.E. Casida. Recent advances in density functional methods, ed. d. p. *World Scientific*, 1995.
- [43] <https://gaussian.com/gaussian16/>, consulté en mars 2025.



## Annexe A

# Script python pour l'analyse de PAHdb

```
1 import xml.etree.ElementTree as ET
2 import math
3 import re
4 import sys
5
6
7 if len(sys.argv) != 3:
8     print("Usage: python3 convert_xml_to_xyz.py fichier_entree.xml
9         fichier_sortie.xyz")
10     sys.exit(1)
11
12 xml_file = sys.argv[1]
13 output_file = sys.argv[2]
14
15 ns = {"pah": "http://www.astrochemistry.org/pahdb/theoretical"}
16
17 def parse_formula(formula):
18     match = re.match(r'C(\d+)H(\d+)', formula)
19     if match:
20         n = int(match.group(1))
21         m = int(match.group(2))
22         return n, m
23     return None, None
24
25 def calculate_hexagons(n, m):
26     return (n - m + 2) / 2
27
28 def is_benzenoid(n, m, h):
29     if 12 * h - 3 >= 0:
30         if 2 * h + 1 + math.ceil(math.sqrt(12 * h - 3)) <= n <= 4 * h +
31             2:
32             return True
33     return False
```

```

32
33 tree = ET.parse(xml_file)
34 root = tree.getroot()
35
36 with open(output_file, "w") as f_out:
37     count = 0
38     for specie in root.findall("pah:species/pah:specie", ns):
39         formula_elem = specie.find("pah:formula", ns)
40         charge_elem = specie.find("pah:charge", ns)
41
42         if formula_elem is None or charge_elem is None:
43             continue
44
45         formula = formula_elem.text.strip()
46         charge = charge_elem.text.strip()
47
48         if charge != "0":
49             continue
50         if any(c not in "CH0123456789" for c in formula):
51             continue
52
53         n, m = parse_formula(formula)
54         if n is None or m is None:
55             continue
56         h = calculate_hexagons(n, m)
57         if not h.is_integer():
58             continue
59         h = int(h)
60         if not is_benzenoid(n, m, h):
61             continue
62
63         geometry = specie.find("pah:geometry", ns)
64         if geometry is None:
65             continue
66
67         atoms = geometry.findall("pah:atom", ns)
68         if not atoms:
69             continue
70
71         xyz_lines = []
72         for atom in atoms:
73             element_type = atom.find("pah:type", ns).text.strip()
74             x = atom.find("pah:x", ns).text.strip()
75             y = atom.find("pah:y", ns).text.strip()
76             z = atom.find("pah:z", ns).text.strip()
77
78             symbol = {"1": "H", "6": "C"}.get(element_type)
79             if symbol is None:
80                 break

```

```

81         xyz_lines.append(f"{symbol} {x} {y} {z}")
82     else:
83         f_out.write(f"{len(xyz_lines)}\n")
84         f_out.write(f"UID: {specie.get('uid')} Formula: {formula}
85                     Hexagones: {h} Group: {specie.get('group')}\n")
86         for line in xyz_lines:
87             f_out.write(line + "\n")
88         count += 1
89 print(f {output_file})

```

## Annexe B

# Exemple de fichier pour des calculs quantiques avec DFT et TD-DFT

```
1 %nprocshared=8
2 %mem=8GB
3 # TD(NStates=10) cam-b3lyp/6-31+g(d) EmpiricalDispersion=GD3BJ ---> TD-
   DFT
4 # opt+freq b3lyp/6-31g -----> DFT
5
6 2-11-18-19-20-30-40-41-49
7
8 0 1
9 C 5.087752 -3.613930 -0.000000
10 C 3.866142 -2.877420 -0.000000
11 C 5.077612 -4.992570 -0.000000
12 C 3.829042 -1.466070 -0.000000
13 C 2.623212 -3.593040 -0.000000
14 C 2.640642 -0.738960 -0.000000
15 C 3.844252 -5.702680 -0.000000
16 C 2.646872 -5.019300 -0.000000
17 C 1.421412 -2.852090 -0.000000
18 C 3.829042 1.466070 -0.000000
19 C 3.866142 2.877420 -0.000000
20 C 2.640642 0.738960 -0.000000
21 C 5.087752 3.613930 -0.000000
22 C 2.623212 3.593040 -0.000000
23 C 5.077612 4.992570 -0.000000
24 C 1.389952 -1.459090 -0.000000
25 C 1.389952 1.459090 -0.000000
26 C 0.116162 -0.712870 -0.000000
27 C 2.646872 5.019300 -0.000000
28 C 1.421412 2.852090 -0.000000
29 C 3.844252 5.702680 -0.000000
30 C 0.116162 0.712870 -0.000000
31 C -1.119058 -1.371750 0.000000
32 C -1.119058 1.371750 0.000000
```

33	C	-2.358438	-0.712490	0.000000
34	C	-2.358438	0.712490	0.000000
35	C	-3.629128	-1.449620	0.000000
36	C	-3.629128	1.449620	0.000000
37	C	-3.667818	-2.864480	0.000000
38	C	-4.865468	-3.564730	0.000000
39	C	-4.862588	-0.735300	0.000000
40	C	-6.080178	-2.861600	0.000000
41	C	-3.667818	2.864480	0.000000
42	C	-4.862588	0.735300	0.000000
43	C	-4.865468	3.564730	0.000000
44	C	-6.069818	-1.474120	0.000000
45	C	-6.069818	1.474120	0.000000
46	C	-6.080178	2.861600	0.000000
47	H	6.027462	-3.068720	-0.000000
48	H	6.011682	-5.545140	-0.000000
49	H	4.783282	-0.953280	-0.000000
50	H	3.853172	-6.787930	-0.000000
51	H	1.703612	-5.558390	-0.000000
52	H	0.499352	-3.420750	-0.000000
53	H	4.783282	0.953280	-0.000000

## Annexe C

# Formules mathématiques de la méthode de DFT

### C.1 Fonctionnelle énergétique

$$E_v[\rho] = \int v(\vec{r})\rho(\vec{r})d\vec{r} + F[\rho] \quad (\text{C.1})$$

### C.2 Nombre de particules

$$N[\rho] = \int \rho(\vec{r})d\vec{r} = N \quad (\text{C.2})$$

### C.3 Comparaison des énergies

$$\int v(\vec{r})\rho'(\vec{r})d\vec{r} + F[\rho'] > \int v(\vec{r})\rho(\vec{r})d\vec{r} + F[\rho] \quad (\text{C.3})$$

### C.4 Approximation Locale de la Densité électronique proposée par Kohn et Sham

$$E = \int v(\vec{r})\rho(\vec{r})d\vec{r} + \frac{1}{2} \int \int \frac{\rho(\vec{r})\rho(\vec{r}')}{|\vec{r} - \vec{r}'|} d\vec{r}d\vec{r}' + T_s(\rho) + E_{xc}(\rho), \quad (\text{C.4})$$

où  $T_s(\rho)$  représente l'énergie cinétique d'un système d'électrons non interactifs dont la densité est  $\rho(\vec{r})$  et  $E_{xc}(\rho)$  [C.5](#) est définie ici comme l'énergie de corrélation d'échange d'un système d'électrons en interaction avec la densité.

### C.5 équation de Kohn et Sham

$$E[\rho] = \int \rho(\vec{r}) \epsilon_{xc}(\rho(\vec{r}))d\vec{r} \quad (\text{C.5})$$

Avec  $\epsilon_{xc}(\rho(\vec{r}))$  est l'énergie d'échange et de corrélation par électron d'un gaz électronique uniforme de densité  $\rho$ .

$$\epsilon = -\frac{1}{2}\nabla^2 + [v(\vec{r}) + \int \frac{\rho(\vec{r})}{(\vec{r}) - (\vec{r}')} + \mu_{xc}(\rho(\vec{r}))] \quad (\text{C.6})$$

avec la densité de charge calculée en utilisant uniquement les fonctions d'ondes des états excités donnée comme suit :

$$\rho(\vec{r}) = \sum_{i=1}^N |\psi_i(\vec{r})|^2 \quad (\text{C.7})$$