

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

ІМЕНІ ТАРАСА ШЕВЧЕНКА

ФАКУЛЬТЕТ РАДІОФІЗИКИ ЕЛЕКТРОНІКИ ТА КОМП'ЮТЕРНИХ СИСТЕМ

Кафедра радіотехніки та радіоелектронних систем

До захисту допущено:

«На правах рукопису»

Завідувач кафедри _____ Ігор АНІСІМОВ

18 травня 2023 р.

КВАЛІФІКАЦІЙНА МАГІСТЕРСЬКА РОБОТА

на тему:

**«Алгоритми попереднього оброблення МАС-спектроскопічного сигналу
для автоматизованої ідентифікації хлорорганічних молекул»**

Виконала:

студентка 2-го курсу магістратури

денної форми навчання

спеціальності 172 Телекомунікації та радіотехніка

ОНП «Інформаційна безпека телекомунікаційних систем і мереж»

Гришаєва Юлія Геннадіївна _____

Науковий керівник:

д.т.н., доц. Ольшевський Сергій Валентинович _____

Рецензентка:

к.б.н., с.н.с. Демченко Віолетта Федорівна _____

Засвідчую, що у цій магістерській роботі

немає запозичень з праць інших авторів без

відповідних посилань

Студент _____

Робота допущена до захисту в ЕК рішенням кафедри радіотехніки та радіоелектронних систем від 18 травня 2023 р., протокол № 18.

Завідувач кафедри радіотехніки та радіоелектронних систем,

доктор фіз.-мат. наук, професор

Анісімов Ігор Олексійович _____

ЗМІСТ

| | |
|--|----|
| ВСТУП | 4 |
| РОЗДІЛ 1. ПОСТАНОВКА ЗАДАЧІ | 6 |
| РОЗДІЛ 2. АНАЛІЗ МЕТОДІВ РОЗПІЗНАВАННЯ МОЛЕКУЛЯРНОЇ СТРУКТУРИ | 7 |
| 2.1. Методи виділення схожих фрагментів сигналів | 7 |
| 2.2. Обертально інваріантні представлення зображень молекулярних структур | 10 |
| 2.3. Гіперграфи як спосіб представлення молекулярної структури..... | 13 |
| РОЗДІЛ 3. АЛГОРИТМИ МАШИННОГО АНАЛІЗУ КОМПОНЕНТНОГО СКЛАДУ МОЛЕКУЛИ..... | 18 |
| 3.1. Алгоритм розпізнавання фрагментів хлорорганічних молекул за масспектрометричними паттернами | 18 |
| 3.2. Алгоритм побудови структурного гіперграфа за внутрішніми координатами молекул..... | 22 |
| 3.3. Алгоритм встановлення відповідності між мас-спектрометричним паттерном і структурою молекулярного фрагмента | 37 |
| РОЗДІЛ 4. РЕЗУЛЬТАТИ ЕКСПЕРИМЕНТІВ..... | 39 |
| 4.1. Створення бази даних мас-спектрів та просторових структур хлорорганічних молекул з близьким елементним складом..... | 39 |
| 4.2. Виділення мас-спектроскопічних паттернів поліхлорбіфінілів..... | 40 |
| 4.3. Доведення еквівалентності представлень структури молекул координатною матрицею та зваженим гіперграфом..... | 43 |
| 4.4. Відновлення просторової структури молекулярного фрагмента за мас-спектрометричним паттерном | 49 |
| ВИСНОВОК | 52 |
| СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ | 53 |
| ДОДАТОК А..... | 54 |
| ДОДАТОК Б | 57 |

| | |
|-----------------|----|
| ДОДАТОК В..... | 58 |
| ДОДАТОК Г..... | 59 |
| ДОДАТОК Д..... | 60 |
| ДОДАТОК Е | 63 |
| ДОДАТОК Є..... | 65 |
| | |
| ДОДАТОК Ж..... | 6 |

ВСТУП

Дипломна робота присвячена алгоритмам попереднього оброблення мас-спектроскопічних сигналів для автоматизованої ідентифікації хлороганічних молекул. Мас-спектрометрія являється одним із фундаментальних аналітичних методів сучасності. Це методика що дозволяє вивчати механізми фрагментації та точно визначити: елементний склад молекул та молекулярну масу їх фрагментів, та їх взаємні співвідношення та положення.

Зазвичай, щоб інтерпретувати мас-спектри використовують методи статистики порівняння з еталонними зразками. У даному випадку використовують цілий мас-спектр, в якому кожен іоний пік є дескриптором молекули. Недоліками такого методу є різні умови реєстрації еталонної виборки і великі розміри дескриптора цільового об'єкту.

Таким чином, один із шляхів вирішення задачі інтерпритації мас-спектрів можна спростити до вибору дескриптора більш високого рівня, розмір його вектора може бути значно меншим. Методи виділення мас-спектрів таких молекул, як багатоатомні або змішані, досить складні і погано вивчені. Наразі вирішення цієї проблеми не існує, хоча програмне та апаратне забезпечення вивчаються. Проте в даний час використовуються комплексні рішення аналітичних проблем виділення подібних молекулярних фрагментів. Ці методи є ефективними, але, на жаль, складні і не можуть бути застосовані на практиці, оскільки вони потребують багато часу для обчислення. Рішенням проблеми може бути створення алгоритму, який буде виділяти схожі фрагменти в еталонному та експериментальному мас-спектрах.

Також запропоновано створити метод представлення молекулярної структури зваженим графом, опису побудови матриць графів на основі z – матриці структури молекул та відповідно в зворотньому напрямку – процедуру побудови z – матриці на основі відповідних їм, отриманих раніше

графів, для подальшого використання в алгоритмах для ідентифікації хлорорганічних молекул.

РОЗДІЛ 1

ПОСТАНОВКА ЗАДАЧІ

Створення автоматизованого методу ідентифікації невстановлених молекул за їх масспектрами ґрунтується на алгоритмах відновлення структури молекулярного фрагмента за відповідним фрагментом мас-спектра. Верифікація метода здійснюється на прикладі поліхлорованих біфенілів, оскільки їх ізомери мають однаковий елементний склад та містять фрагменти як з однаковою так і різною просторовою конфігурацією. Початковим кроком має бути накопичення апріорних даних для створення бази мас-спектрів молекул поліхлорованих біфенілів, яка налічує понад 250 ізомерів. Зібрані дані можуть бути використані для подальшої ідентифікації невідомих сполук в тих зразках, яких не має в базах даних еталонних мас-спектрів. Процедура автоматизованої ідентифікації молекул складається з алгоритмів виділення мас-спектрометричних паттернів фрагментів молекул із спільною структурою декмозиції вихідної молекули на окремі структурні елементи та встановлення відповідності між паттерном і структурним елементом.

Таким чином мета роботи: розробити алгоритм автоматизованої ідентифікації хлорорганічних молекул за їх мас-спектром.

Для досягнення мети необхідно вирішити такі задачі:

1. Створення бази даних мас-спектрів та просторових структур хлорорганічних молекул з близьким..елементним..складом.
2. Виділення мас-спектроскопічних паттернів для розпізнавання фрагментів хлорорганічних молекул.
3. Розробити та відлагодити алгоритми попереднього оброблення мас-спектроскопічного сигналу.

РОЗДІЛ 2

АНАЛІЗ МЕТОДІВ РОЗПІЗНАВАННЯ МОЛЕКУЛЯРНОЇ СТРУКТУРИ

2.1. Методи виділення схожих фрагментів сигналів

Щоб знайти рішення проблеми ідентифікації подібних фрагментів у еталонному та експериментальному мас-спектрах, методи виявлення та аналізу подібних фрагментів у дискретизованих сигналах, такі як:

- Класифікація.
- Кластеризація.
- Регресія.
- Прогнозування послідовностей часових рядів.
- Порядок.
- Асоціації;

Асоціацію використовують, коли необхідний аналіз структури, коли кілька подій пов'язані одна з одною.

Класифікація є найвідомішою моделлю інтелектуального аналізу. Використовуючи цю модель, можна виявити ознаки, що характеризують ту чи іншу групу, в яку входять об'єкти. Класифіковані об'єкти використовуються для аналізу та формують певний набір правил. Регресія – використовується, коли зв'язок між змінними є кількісним, наприклад, комбінації однакових змінних. Отримана комбінація використовується для прогнозування того, що припускає обчислена залежна змінна за певного набору значень вхідної змінної. У простих випадках також можна використовувати стандартні статистичні методи, такі як лінійна регресія.

Прогнозування послідовностей. Багато систем прогнозування використовують історичну інформацію, що зберігається у вигляді часових рядів в інформаційних кешах. Створюючи конкретні математичні моделі та знаходячи закономірності, які відображають цю динаміку, ми можемо передбачити, як система поводитиметься в майбутньому при їх використанні. На основі аналізу поведінки цих часових рядів методи прогнозування часових рядів дозволяють оцінити прогнози майбутніх значень. Послідовність – ця модель використовується, коли є низка подій в часі пов'язаних між собою.

Кластеризація схожа на методи класифікації, за винятком того, що при використанні моделі кластеризації групи не визначені заздалегідь, а однорідні класи даних створюються незалежно за допомогою інтелектуальних обчислювальних інструментів. У машинному навчанні для опису об'єкта існує поняття простору ознак (вектор, що визначає певні вимірні ознаки об'єкта). Коли виникає необхідність знайти невеликий і достатньо інформативний опис об'єкта, виникає задача навчання за поданням [1].

Наразі існує низка прямих та ефективних методів оцінки подібності двох кривих, що включають оцінку відстані на основі різних мір відстані, тобто аналіз подібності за допомогою коефіцієнтів кореляції.

Серед них широко використовуються європейська відстань, відстань Хаусдорфа та ін. У 1906 році відомий математик Моріс Рене Фреше запропонував новий метод опису подібності. Цей метод фокусується на розгляді просторових траєкторій, що робить його більш ефективним методом в задачах оцінки подібності кривих з певними просторовими та часовими рядами. На рисунку 2.1.1 показано відстані Хаусдорфа і Фреше між двома кривими.

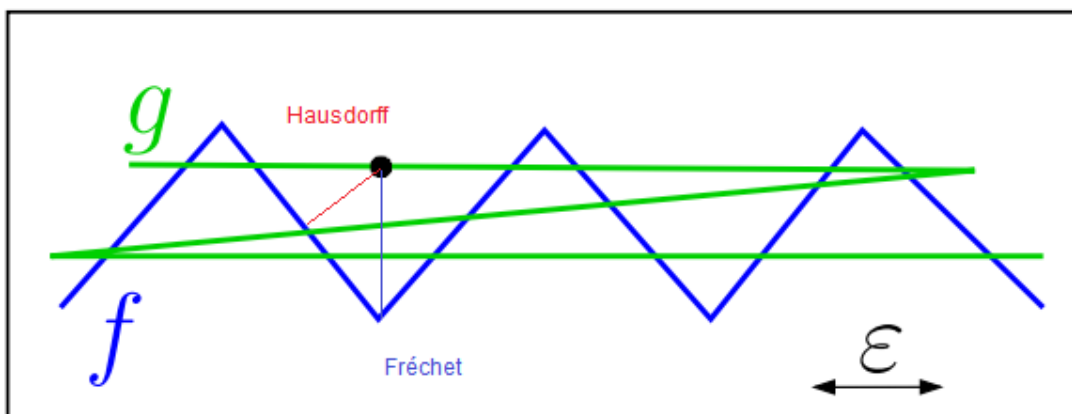


Рис. 2.1.1. Відстані Хаусдорфа та Фреше

Відстань Фреше являє собою міру подібності кривої, що враховує кількість і порядок точок уздовж кривої. Візьмемо метричний простір P . Крива V в цьому просторі S є неперервним відображенням одиничних сегментів S , тобто $V : [0,1] \rightarrow S$. Репараметризація α відрізка $[0,1]$ – це неперервна сюрєкція $\alpha : [0,1] \rightarrow [0,1]$.

Відстань Хаусдорфа – D_H , відстані Фреше – S_f , виходить $d_H \leq \varepsilon$, $s_f \geq 2\varepsilon$, тому у деяких ситуаціях використання відстані Фреше краще, аніж відстані Хаусдорфа.

Отже, нехай V і C 2 криві в просторі P , таким чином відстань Фреше між V та C визначається як точна нижня границя по всім параметризаціям α та β відрізка $[0,1]$ по всім максимумам $t \in [0,1]$ відстаней в P між $A(\alpha(t))$ і $B(\beta(t))$. Математичне представлення відстані Фреше $F(A, B)$

$$F(A, B) = \inf_{\alpha, \beta} \max_{t \in [0,1]} \{d(A(\alpha(t)), B(\beta(t)))\},$$

де d є функцією відстані простору P .

Відстань Фреше приймає до уваги проходження двох кривих, оскільки пари точок, відстань між якими визначає відстань Фреше, пробігають вздовж цих кривих [2].

Провівши аналіз різних методів аналізу даних необхідної для нашої задачі, а саме таких як: класифікація, регресія, асоціація, кластеризація, послідовність, прогнозування часових послідовностей, був зроблений висновок: вище описані методи не достатньо точні для фрагментованих сигналів, таких, як мас-спектри багатоатомних молекул. Отже, для порівняння форми сигналів, більш доцільно використовувати саме міри схожості кривих – відстань Фреше. Даний вибір можна пояснити тим, що алгоритми, орієнтовані на порівняння форм кривих, які представлені дискретним набором відліків. Ці криві не являються рекурсивними і оптимізовані на основі кількості кроків. У задачі виділення мас-спектральних моделей буде використано алгоритм розрахунку відстані Фреше. Його перевагою також є легкість програмування, досить велика швидкодія і він проявив хороші результати в експериментальних тестах.

2.2. Обертально-інваріантні представлення зображень молекулярних структур

Для того щоб виділити спільні фрагмента також можна використати представлення зображень, а саме – моменти, що являють собою розкладання зображень, використовуючи обертально симетричні базисні функції. У літературних джерелах наведено декілька типів обертально-інваріантних представлень зображень, а саме, таких, як- моменти Кравчука, Гана та моменти Церніке і інші.

Окремими випадками подвійних поліномів Гана є Поліноми Кравчука та Чебишова. Подвійні многочлени Гана масштабуються, для того, щоб

забезпечити чисельну стабільність, надалі створюючи набір зважених ортогональних моментів. Та забезпечує мінімальну надлишковість зображення, прибирає необхідність чисельних наближень – така властивість, як дискретна ортогональність подвійних моментів Гана. Подвійні інваріанти моменту Гана виводяться при використанні лінійних комбінацій геометричних моментів [4].

Моменти Кравчука являють собою дискретні ортогональні моменти, що є дуже ефективними для задачі представлення зображення. Інваріанти перекладу і масштабу моментів Кравчука досягаються, шляхом нормалізації зображення, або ж при використанні комбінацій деяких інваріантів геометричних моментів. Цікава властивість даних об'єктів: вони спроможні відокремлювати локальні особливості зображення [3].

Моменти Церніке порядку φ і в степені η зображення $f(r, \theta)$, що представлено в колі одиничного радіусу, визначається по формулі:

$$A_{\varphi\eta} = \frac{p+1}{\eta} \sum_r \sum_{\theta} f(r, \theta) V_{\varphi\eta}(r, \theta),$$

де $V_{\varphi\eta}(r, \theta)$ – Церніке поліном, що створює ортогональну систему в одиничному колі $x^2 + y^2 \leq 1$.

Поліном Церніке $V_{\varphi\eta}(x, y)$ – функція, що визначена в колі одиничного радіусу $x^2 + y^2 \leq 1$.

Отже функція:

$$V_{\varphi\eta}(x, y) = R_{\varphi\eta}(r) e^{-j\eta\theta},$$

$$\text{Де } j = \sqrt{-1}$$

φ – ціле число з знаком +, або число 0.

η – ціле число, $\eta \leq \varphi, \varphi = |\eta|$,

r – є довжиною вектора від початку координат до (x, y) , $r = \sqrt{x^2 + y^2} \leq 1$.

θ – це кут між вектором r і віссю x , $\theta = \arctan(y/x)$, $\theta \in [0, 2\pi]$;

$R_{\varphi_n}(r)$ – радіальний поліном.

Поліноми Церніке необхідні розраховуватися для кожного пікселя зображення із врахуванням того, що моменти Церніке визначаються в полярних координатах. Використовуючи полярну систему координат слід розраховувати нові координати пікселя, так як положення відносно центра кола зміщується [5]., Перетворення в полярну систему координат зображення в декартовій системі координат перетворюється в полярну систему координат представлено на рис. 2.2.1.

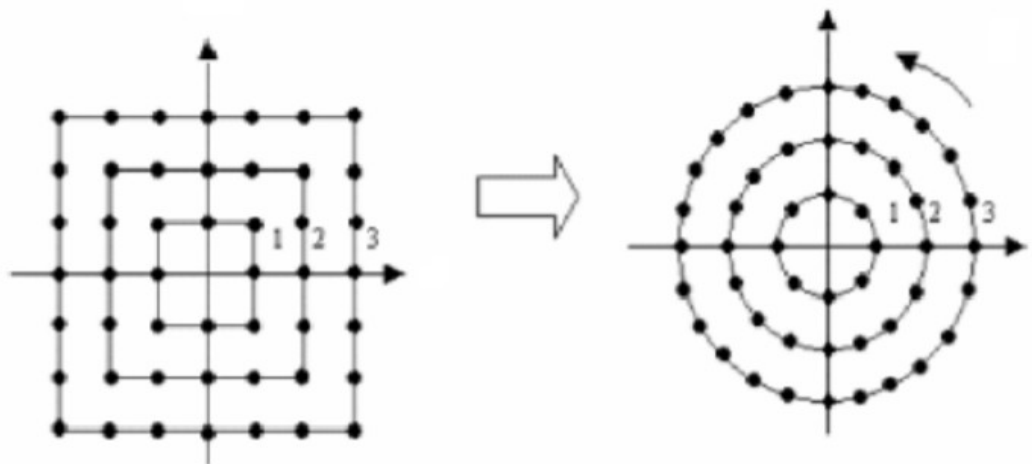


Рис. 2.2.1. Перетворення зображення з декартової системи координат в полярну систему координат

Пікселі у полярній системі координат повинні бути представлені такою формулою:

$$x_i = \frac{2i - N + 1}{N}, y_k = \frac{2k - N + 1}{N}, \quad i, k = 0, 1, 2, \dots, N-1$$

Ще одна добра властивість моментів Церніке це інваріантність відносно повороту і відображення зображення. Завдяки цій властивості моменти Зерніке широко використовуються в розробці систем розпізнавання об'єктів та в інших областях. Двосторонній момент Зерніке перевершує інші за шумочутливістю та дискримінантною здатністю. Добре відомо, що функцію дискретного зображення можна відновити за допомогою моментів Церніке. Відносно невеликий набір моментів Зерніке може досить ефективно описати загальну форму зразка. Моменти нижчого порядку представляють загальну форму зразка, а моменти вищого порядку представляють деталі. [6].

Таким чином, зробивши аналіз безлічі обертально-інваріантних представлень зображень, для подальшого аналізу і виділення схожих фрагментів за символічним зображенням просторової структури молекул, зроблений висновок, що самим підходящим методом є моменти Церніке, так як такий підхід більш поширений при аналізі зображень.

2.3. Гіперграфи як спосіб представлення молекулярної структури

Графи представляють великий розділ математики. Він знайшов використання при рішенні управлінських завдань, економічних, програмуванні, конструюванні теорії електричних ланцюгів, комунікаціях, хімії психології, соціології, лінгвістиці, інших галузях знань. Властивості графів визначає теорія, вони вміщують в себе множини точок та множини ліній, які представляють зв'язки між цими точками. Теорію графів запропонував у 1736 році Леонард Ейлер, він вирішив завдання кенігсберзьких мостів [11].

Використання графів полягає у відображенні різноманітних відносин на безлічі. Прикладом такого застосування є $A = \{a_1, a_2, \dots, a_n\}$ - множина з певних людей, де елементи становлять вигляд точок. У множині $B = \{b_1, b_2, \dots, b_m\}$ – відображені зв'язки. Множина під назвою A відображує зв'язок знайомства між людьми в даній множині множини. Таким чином потрібно побудувати граф із вузлів і зв'язок. Вузлами можна пов'язати знайомих між собою людей. Кількість знайомих в одних людей може відрізнятись знайомих у інших людей, та треба врахувати, що інші можуть бути не знайомі між собою. (дані елементи не з'єднанні), як показано на рис. 2.3.1 [12].

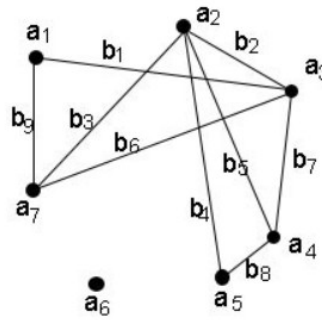


Рис. 2.3.1. Граф.

«Точками» - в графі називаються вузли, а «зв'язками» - ребра графу.

Природу множин не бере до уваги теорія графів. Існує незлічування кількість задач, що при вирішенні їх можливо ненадовгий час не пам'ятати про особливий зміст їх множин, елементів. Дана особливість ніяким чином не відображується при вирішенні задачі, незалежності від її проблеми. Прикладом є, рішення питань де, можливість з точки A дістатися до точки B , проходячи тільки по лініях, з'єднуюючих точок, де не береться до уваги, чи ми маємо задачу з числами, людьми, місцями, і т.д. Вже коли ми маємо рішену задачу, то тоді маємо рішення, яке є правильним для будь-якого

змодельованого змісту в представленні графу. Важливим інструментом при розробці штучного інтелекту є використання теорії графів, тому що штучний інтелект дає можливість говорити з абонентом про кохання, музику або ж спорт, також питання вирішення різноманітних завдань, та може робити це без якого-небудь перемикання, без якого у таких ситуаціях не може обійтись людина [9].

Граф F – представляє множина вершин, де елементи $f \in F$ представляють вершини. Граф $P = P(F)$ де є безлічч вершин $F \in$ деяким сімейством пар: $e = (c, d)$, де $c, d \in F$, що показують, вершини які залишаються зв'язаними. Кожна пара $e = (c, d)$ є ребром графа. Множина E є множиною ребер е графа. Вершини c і d є кінцевими точками ребра e .

Граф, являє собою систему (вершин) та (ребер), які з'єднують певні пари даних об'єктів.

Граф як структура даних. Широке застосування теорії графів в інформатиці та інформаційних технологіях зумовлено додаванням вищезгаданої концепції графів як структур даних. В інформатиці та інформаційних технологіях граф визначається як нелінійна структура даних. Отже, що таке лінійні структури даних і чим відрізняються від них графіки? Лінійні структури даних характеризуються елементами, з'єднаними зв'язками типу «проста сутність». Лінійні структури даних - масив, таблиця, стовпець, черга, потік, рядок. На відміну від них, нелінійні структури даних - це ті, елементи яких знаходяться на різних рівнях ієрархії і поділяються на три типи: необроблені дані, згенеровані дані тощо. Отже, граф є нелінійною структурою даних [14].

Деякі визначення в теорії графів:

Вузол- це точка у графі, окремий об'єкт, для топологічної моделі графа немає значення координати вершини, її розташування, колір, смак, розмір; але при вирішенні деяких задач вершини можуть розфарбовуватися в різні кольори або зберігати числові значення.

Ребром є невпорядкована пара двох вершин, пов'язаних між собою. Вони називаються кінцевими точками чи кінцями ребра. При цьому важливий факт наявності зв'язку, яким саме чином здійснюється цей зв'язок і якою дорогою - не має значення; однак ребрам може бути присвоєно "вагу", це дозволить говорити про "звантажений граф".

Інцидентність – вершина та ребро називаються інцидентними, якщо вершина є для цього ребра кінцевою. Термін “інцидентність” застосовується лише до вершини, ребра.

Суміжність вершин - дві вершини є суміжними, якщо вони інцидентні одному ребру [10].

Суміжність ребер - два ребра є суміжними, якщо вони інцидентні одній вершині.

Говорячи простіше - дві вершини суміжні, якщо вони з'єднані ребром, два ребра суміжні - якщо вони з'єднані вершиною.

Якщо на ребрах (дугах) графа прикріплені деякі ваги (мітки), то такий граф називається зваженим..

Вага дуги чи ребра може означати довжину, пропускну здатність, напругу чи струм і т. д. Ваги можна присвоювати не тільки ребрам, але і вершинам. Наприклад, зважені графіки використовуються в мережевому плануванні. Зважений орієнтований граф називається графом потоку сигналів. Наприклад, зважені орієнтовані графи використовуються в теорії ланцюгів. Зважений граф, який не містить багато ребер, може бути представлений матрицею суміжності. При цьому кожен її ненульовий елемент дорівнює вазі відповідного ребра (дуги) [14].

Тепер розглянемо випадок зважених графів. Елементи матриці суміжності $a[i,j]$, що відповідають такому графу, будуть збігатися зі значенням «ваги» ребра між вершинами i та j (рис. 2.3.2. та рис.2.3.3) [17].

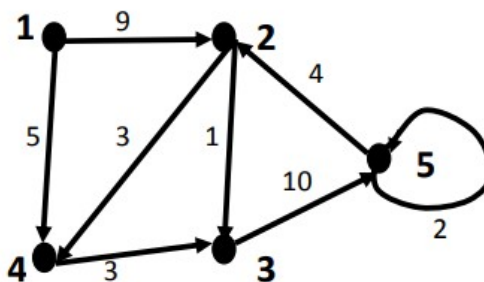


Рис. 2.3.2. Зважений граф

| № вершини | 1 | 2 | 3 | 4 | 5 |
|-----------|---|---|---|---|----|
| 1 | 0 | 9 | 0 | 5 | 0 |
| 2 | 0 | 0 | 1 | 3 | 0 |
| 3 | 0 | 0 | 0 | 0 | 10 |
| 4 | 0 | 0 | 3 | 0 | 0 |
| 5 | 0 | 4 | 0 | 0 | 2 |

Рис. 2.3.3. Матриця суміжності

Способ яким можна представити граф також є список вершин які суміжні між собою. Кожній вершині в даному випадку приписується список суміжних між собою вершин, де можна побачити в якому порядку містять всі суміжні з нею вершини. Представленому графу (рис. 2.3.1) відповідає представлений нижче ряд суміжних між собою вершин:

$$1 \rightarrow 2 \rightarrow 4, \quad 2 \rightarrow 1 \rightarrow 5 \rightarrow 3 \rightarrow 4, \quad 3 \rightarrow 2 \rightarrow 4 \rightarrow 5, \quad 4 \rightarrow 1 \rightarrow 3 \rightarrow 2, \quad 5 \rightarrow 3 \rightarrow 2.$$

Використовуючи теоретичну інформацію, яка наведена вище, буде розроблений алгоритм еквівалентного перетворення з - матриці структури молекул в три взаємопов'язаних зважених графів, в якому вузли та ребра будуть мати певні параметри, які розглянуті та представлені в наступному розділі. Отримані графи будуть описані у вигляді матриці інцидентності, щоб показати як взаємопов'язані між собою вузли та ребра відповідного графа.

РОЗДІЛ 3

АЛГОРИТМИ МАШИННОГО АНАЛІЗУ КОМПОНЕНТНОГО СКЛАДУ МОЛЕКУЛИ

3.1. Алгоритм розпізнавання фрагментів хлорорганічних молекул за маспектроскопічними паттернами

Розроблений метод для розпізнавання та виділення маспектроскопічних паттернів був розроблений з використанням відстані Фреше, блок-схема алгоритму зображена на рис. 3.1.1.

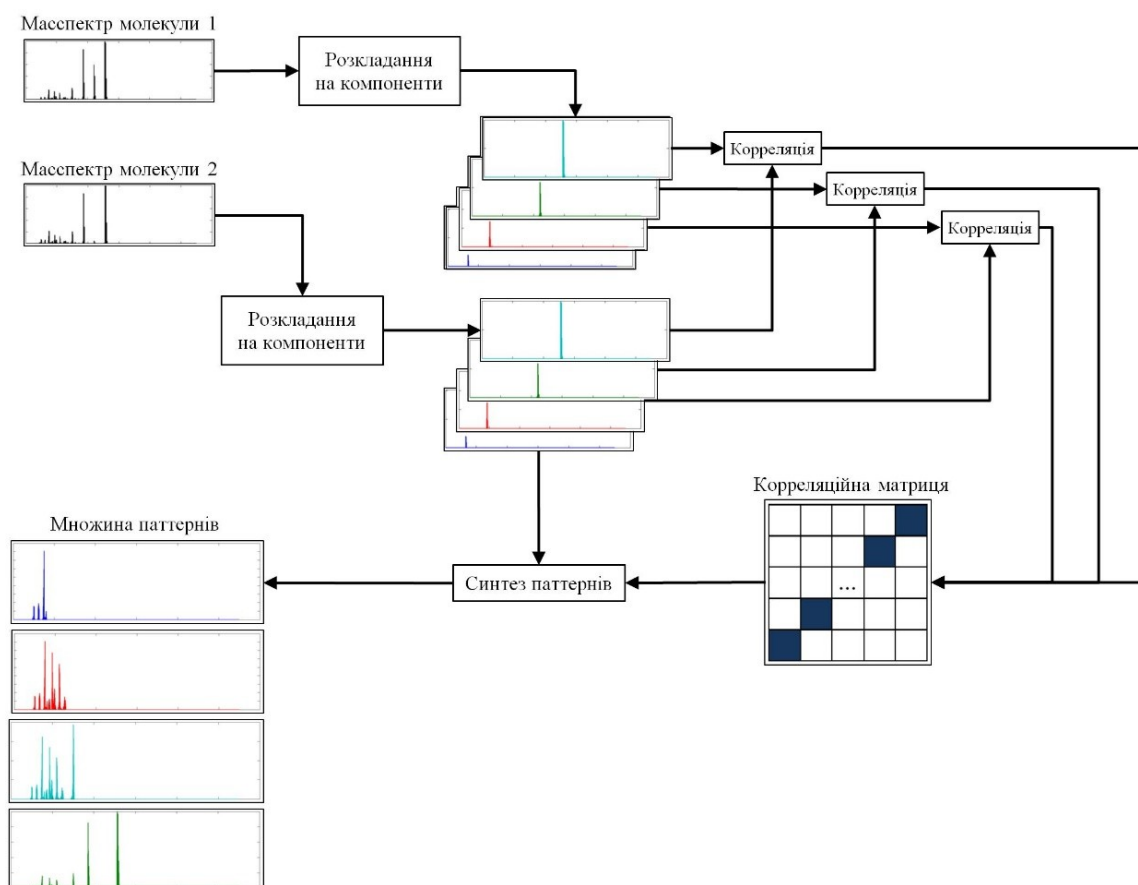


Рис. 3.1.1. Алгоритм методу розпізнавання мас-спектроскопічних паттернів

Представлена на рис. 3.1.1. блок-схема виконання алгоритму методу розпізнавання паттернів. В першу чергу алгоритм перетворює пари N – мірних масивів з значеннями інтенсивності іонних піків порівнювальних молекул у пару $N \times M$ - мірних масивів, M – це кількість окремих груп іонів, що відносяться до одного гомологічного ряду. M -мірний масив зберігає в собі значення інтенсивностей іонних піків, що належать до якого-небудь одного гомологічного ряду, а значення іших піків приводиться до нуля. На наступному етапі між різними елементами двовимірного масиву алгоритм обчислює відстань Фреше та будує вхідну мас-спектральну матрицю на основі цієї відстані. На основі отриманої матриці алгоритм синтезує ідентичні ланцюги іонів по всьому спектру, додає елементи та обчислює відстань Фреше між отриманими сумами. Якщо сума приблизно дорівнює, ці рядки зберігаються в окремому масиві, який є шаблоном. Програмна реалізація цього алгоритму, розроблена в середовищі MATLAB R2014, дуже підходить для реалізації описаного методу, оскільки включає різноманітні математичні операції, а саме для нашої задачі: матричні обчислення та статистику. Нижче описаний програмний код.

Програмний код виконує зчитування даних двох вхідних експериментальних мас-спектрів молекул, що розташовуються у файлі Excel, у змінні Ms_s – мас-спектр основної молекули і Ms_r – в мас-спектр молекули, що досліджується. Розмірність масивів дорівнює діапазону мас, для цього інтенсивність піків, що буде нижче рівня шуму і відсутніх піків примусово приводили в нуль. Розмірність мас-спектрів дорівнює 550. Використовуючи функції $Ms_Splitter$ (..) перетворювали одномірні масиви мас-спектрів основної і досліджуваної молекули у двумірні масиви $Frg_s(N,M)$ та $Frg_s(N,K)$.

В розділі коду- «Порівняння» використана функція Frechet (..), що будує матрицю відповідності $fDst(L \times L)$, де $L = \min\{M, K\}$, функція Frechet (..) рахує відстань Фреше між двома кривими, що задані одномірними масивами.

На наступному кроці в розділі коду «Синтез паттернів» гомологічні послідовності, що належать до різних вихідних мас-спектрів, але близькі одна до одної за відстанями Фреше, накопичуються в масиві $Pttr(N \times J)$. Коли ці послідовності накопичуються, вони порівнюються з відстанню Фреше, якщо відстань не перевищує порогове значення 0,05, тоді послідовність зараховується як зразок і зберігається в таблиці $Prt_d(N \times J')$, де J' — розмірність, що дорівнює числу іонів з однаковою послідовністю в мас-спектрах основної та досліджуваної молекули.

Отже в результаті роботи алгоритму отримано сімейство послідовностей, що асоціюють з відповідними однаковими фрагментами ізомерів та вибирають за мас-спектроскопічні паттерни цих фрагментів.

Також встановлений характерний час обчислення алгоритму для мас-спектрів порядку 500 іонних мас, в ПЗ MatLab складає приблизно 0,94 сек.

3.2. Алгоритм побудови структурного гіперграфа за внутрішніми координатами молекул

Визначення геометричних параметрів молекул.

Для того щоб описати геометричну структуру молекули необхідно знати три геометричні параметри:

1. Довжина зв'язку;
2. Валентний кут – кут, утворений двома напрямками хімічних зв'язків, що виходять з одного атома;
3. Торсіонний кут – кут, утворений двома напівплощинами, що виходять з однієї прямої, а також частина простору, обмежена цими напівплощинами, як показано на рис. 3.2.1.

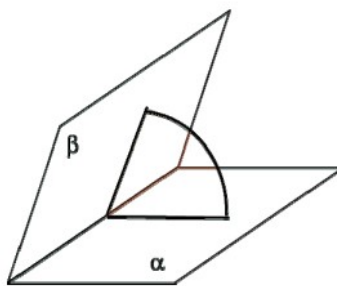


Рис. 3.2.1. Торсіонний кут

Для визначення z- матриці структури молекули будемо використовувати програмне забезпечення «Gaussview 6».

Задання вихідної конфігурації молекули.

Положення ядер електронів α - і β спіна можуть бути визначені як за допомогою Z-матриці, так і за допомогою декартових координатах, та також їх поєднанням.

Перший рядок розділу специфікації молекули визначає чистий електричний заряд (підписане ціле число) та мультиплетність спіна (позитивне ціле число). Мультиплетність розраховується за формулою

$$M = \left(2 \sum_{n=1}^N s + 1 \right)$$

Тут наведено кількість електронів у молекулі та спінове квантове число кожного електрона. Оскільки більшість електронів у молекулах є парними, більшість матеріалів мають характерний нульовий кумулятивний спіновий стан. Н. $M = 1$ (синглетний стан). Коли молекула збуджена, один з її електронів переходить у збуджений стан. Н. Вищі рівні енергії. У цьому випадку кратність змінюється зі зміною взаємної орієнтації спінів або змінюється зі зміною взаємної орієнтації спінів. Наприклад, молекула може переходити з синглетного основного стану в синглетний або триплетний збуджений стан ($M=3$). [7].

Початкова структура молекули у вікні молекулярної специфікації може бути задана 4 способами: за допомогою внутрішньої координат атомів,

декартових координат, комбінованим способом та за допомогою псевдоатомів. Одиницями виміру, які прийняті за замовчуванням, є ангстреми (10^{-10} м) для відстані і градуси для кутів.

Приклад молекулярної специфікації в декартових координатах. При задані конфігурації молекули в декартових координатах у вікні молекулярної специфікації, на першому місці вказується хімічний символ атома, потім роздільник (пробіл або декілька пробілів, або кома), потім через роздільник декартові координати атома x , y , z . Якщо координати атома – цілі числа, то наявність десяткової точки обов'язкова [8].

Координати атома

Кожен рядок містить символ елемента і додатково можуть вказуватися молекулярно-механічні параметри атома. Решта рядка містить інформацію про місцезнаходження атома як у Декартових координатах, так і використовуючи Z -матриці.

Декартові координати є найпростішим способом визначення будови молекули, і полягає у вказівці декартових координат, і для кожного атома молекули. У цьому випадку формат кожного рядка, що описує атом, має вигляд

АТОМ XYZ,

де

АТОМ - символічне позначення даного атома чи атомний номер,

X , Y , Z - набір декартових координат [13]. Наприклад, задання стартової конфігурації структури молекули $C_{12}H_{10}$ за допомогою декартових координат має такий вигляд, таблиця наведена в додатку А.

Вибір початку координат є довільним (це може бути будь-який атом молекули або абсолютно довільна точка простору).

Z -матриця – це формалізований спосіб розміщення атомів, які утворюють молекулу, у тривимірному просторі. Також, таке представлення називають також внутрішніми координатами (internal coordinates). Це представлення визначає кожен атом системи через атомний номер, довжину

зв'язку, валентний кут і двогранний (торсійний) кут. Під зв'язком мається на увазі вектор, який спрямований від одного атома до іншого. Будь-які три атоми, становище яких було визначено раніше, називають базисними або реперними, причому перший базисний атом приймається за центр полярної системи координат, другий базисний атом задає базову вісь, а третій базисний атом базову площину полярної системи координат, як в таблиці в додатку А.

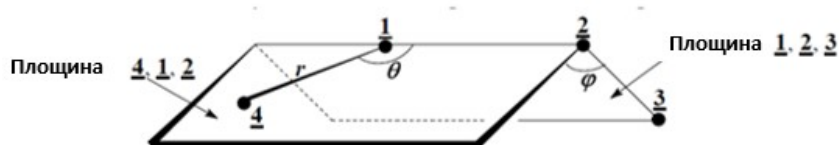


Рис. 3.2.3. Визначення положення атома в просторі.

Для точного визначення положення атома 4 у просторі необхідно визначити сукупність трьох координат: довжину зв'язку r , валентний кут θ та торсійного кута ϕ . Валентні кути повинні бути в діапазоні $0^\circ < \theta < 180^\circ$. Торсійні кути можуть набувати будь-якого значення.

Як видно з рис. 3.2.3, величина знаходиться на відстані між атомом 4, що задається, базисним атомом 1, валентний кут – кут між атомами 4, 1 і 2 з вершиною на базисному атомі 1, торсійний кут – це кут між площинами 4, 1, 2 і 1, 2, 3, тобто. кут між перпендикулярами до базової осі, опущеними з однієї точки обидві площині. Формат кожного рядка, що описує атом, має вигляд

АТОМ [n] atom1 atom2 atom3 [format-code] [16].

Де n - додатковий параметр, пов'язаний із заморожуванням (freezing) атомами під час оптимізації,

$atom1$, $atom2$, $atom3$ - мітки для вказаних раніше атомів, які будуть використовуватися, щоб визначити положення поточних атомів,

Положення даного АТОМ атома визначають, задаючи довжину зв'язку r , як відстань від АТОМ до $atom1$, валентний кут θ є кут утворений лінією з'єднуючої АТОМ і $atom1$, і лінією яка з'єднує $atom1$ і $atom2$, торсійний кут ϕ - кут, утворений лінією з'єднує $atom2$ і $atom3$ з площиною, утвореною АТОМ (даним атомом), $atom1$ і $atom2$.

Додатковий параметр коду формату *format-code* визначає формат входу Z-матриці. Його використання необхідне лише тоді, коли розраховується молекулярна система, розділена двома чи трьома шарами, кожен із яких обробляється різними хімічними моделями (обчислення ONIOM) [8].

Якщо використовується символ елемента для позначення АТОМ, він може супроводжуватися іншими алфавітно-цифровими характеристиками, щоб створити мітку ідентифікації для атома. За назвою елемента слідує ціле число ідентифікації: C1, C2, C3, і т.д.

Розглянемо специфікацію молекули бифенілу C₁₂H₁₀.

Використання внутрішніх координат атомів. Взаємне розташування атомів в молекулі можна описати за допомогою внутрішніх координат атомів, якими є міжатомні відстані (довжина зв'язків), валентний і торсійних кутів. Для цієї цілі прийнято використовувати так звану z-матрицю, яка розташована у вікні молекулярної специфікації. Перед її побудовою всім атомам структури присвоюються номери, наприклад, для молекули бифенілу, рис. 3.2.4.

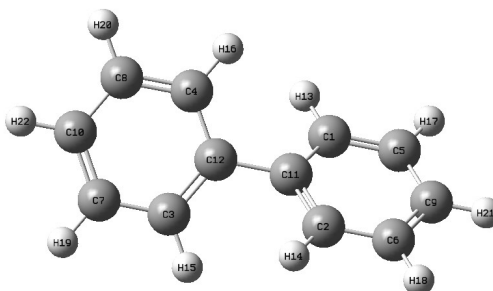


Рис. 3.2.4. Молекула бифенілу C₁₂H₁₀ (атоми пронумеровані).

В першому рядку Z - матриці пишеться тільки символ елемента, якому присвоєний номер 1. Другий рядок починається із символу хімічного елемента, якому було присвоєно номер 2, далі йде роздільник (пробіл або кома), далі цифра 1 (номер елемента, якій стоїть в першому рядку), потім – роздільник, після якого вказується довжина зв'язку атом 2 – атом 1. Наприклад, для молекули бєфєнілу (довжина зв'язку C-C рївна 1.5611000 Å) рис 3.2.5.

Таблиця 3.2.1

Z - матриця структури молекули C₁₂H₁₀.

| Tag | Symbol | NA | NB | NC | Bond | Angle | Dihedral |
|-----|--------|----|----|----|-----------|-------------|--------------|
| 1 | C | | | | | | |
| 2 | C | 1 | | | 1.5611000 | | |
| 3 | C | 1 | 2 | | 1.5600780 | 109.3964352 | |
| 4 | C | 2 | 1 | 3 | 1.5600223 | 109.4200713 | -0.0129610 |
| 5 | C | 2 | 1 | 3 | 1.5250529 | 86.0589537 | -122.1425459 |
| 6 | C | 5 | 2 | 1 | 1.3516002 | 93.9301152 | 0.0275699 |
| 7 | C | 2 | 1 | 6 | 1.5246403 | 86.0703191 | -115.7486101 |
| 8 | C | 7 | 2 | 1 | 1.3516000 | 93.9388686 | -0.0151477 |
| 9 | C | 4 | 2 | 1 | 1.5339657 | 106.3746401 | 56.4460337 |
| 10 | C | 9 | 4 | 2 | 1.3508000 | 113.9869139 | -59.3375106 |
| 11 | C | 4 | 2 | 1 | 1.5338366 | 106.3822497 | -56.4281174 |
| 12 | C | 11 | 4 | 2 | 1.3508000 | 113.9923945 | 59.3289303 |
| 13 | H | 3 | 1 | 6 | 1.0929231 | 112.7099224 | 82.4919109 |
| 14 | H | 4 | 2 | 1 | 1.0928117 | 112.7160534 | -179.9927022 |
| 15 | H | 5 | 2 | 1 | 1.0851450 | 132.6808808 | 179.9537673 |
| 16 | H | 7 | 2 | 1 | 1.0851595 | 132.6732535 | -179.9430544 |
| 17 | H | 6 | 5 | 2 | 1.0852370 | 133.3789070 | 179.9387908 |
| 18 | H | 8 | 7 | 2 | 1.0852140 | 133.3877161 | -179.9346487 |
| 19 | H | 9 | 4 | 2 | 1.0814819 | 122.3965785 | 119.2811654 |
| 20 | H | 11 | 4 | 2 | 1.0815726 | 122.3953212 | -119.2839144 |
| 21 | H | 10 | 9 | 4 | 1.081520 | 123.599728 | 178.6053235 |

| | | | | | | | |
|----|---|----|----|---|---------------|-----------------|--------------|
| | | | | | 4 | 5 | |
| 22 | H | 12 | 11 | 4 | 1.081500 4 | 123.601594 0 | -178.6046770 |

Третій рядок окрім довжину зв'язку між атомом вуглецю C(3) та іншим атомом вуглецю C(1) містить кут 3-1-2 рівний 109.3964352° . Структура всіх рядків починаючи з четвертого, однакова.

Слід зауважити, що перші три атоми в межах z – матриці не використовують повне число параметрів; Тільки четвертому і наступних атомах використовується повне число параметрів. Наприклад останній рядок z – матриці бєфєнілу специфікує атом H(22), після символу елемента і роздільника йде номер будь-якого передуючого елемента, 1.0815004 Å – відстань до атома H(22) до атома 1; 123.6015940 – валентний кут H(22) – атом 12–11; -178.6046770 – торсіонний кут – атом 12 – атом 11 – атом 4.

Отже, використовуючи z – матриці структур молекул, які визначенні в програмному забезпеченні «GaussView 6», в наступному розділі буде представлений розроблений алгоритм побудови еквівалентних графів із z – матриці структур молекул, які були отримані.

В даному розділі представлений розроблений алгоритм побудови графів із z -матриць структур молекул. Побудовані графи мають бути представлені у вигляді трьох взаємопов'язаних графів, де вузлами та ребрами якого будуть:

1. Вузлами першого графа нехай будуть окремі елементи, що складають молекулу. Назва графу: «Граф валентних зв'язків».

Ваговими коефіцієнтами вузлів мають бути порядкові номери елементів з таблиці Менделєєва.

Ваговими коефіцієнтами ребер мають бути довжини валентних зв'язків.

2. Вузлами другого графа нехай будуть пари елементів зі спільним зв'язком, які утворюють валентний кут з відповідним елементом іншої пари.

Ваговим коефіцієнтом ребра має бути величина валентного кута.

Назва графу: «Граф валентних кутів»

3. Вузлами третього графу нехай будуть трійки елементів, що входять в спільний валентний кут.

Ребра цього графу представляють торсіонний кут поточної трійки відповідного елемента іншої трійки, який складає торсіонний кут з даними трьома елементами.

Ваговим коефіцієнтом ребра буде величина торсіонного кута.

Назва графу: «Граф торсіонних кутів».

Отже, побудуємо перший граф, який носить назву «Граф валентних зв'язків». де вузли- окремі елементи, що складають молекулу, вагові коефіцієнти вузлів порядкові номери елементів з періодичної таблиці Менделєєва, вагові коефіцієнти ребер довжини валентних зв'язків.

Для виконання алгоритму побудови графів із z-матриць структур молекул візьмемо молекулярну структуру біфенілу C₁₂H₁₀ рис. 3.2.5., та її z-табл. 3.2.2.

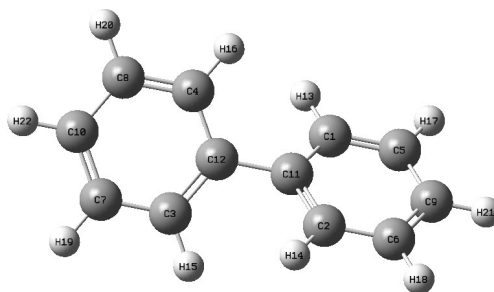


Рис. 3.2.5. Молекула біфенілу C₁₂H₁₀.

Таблиця 3.2.2

Z - матриця структури молекули C₁₂H₁₀.

| Tag | Symbol | NA | NB | NC | Bond | Angle | Dihedral |
|-----|--------|----|----|----|---------------|-----------------|--------------|
| 1 | C | | | | | | |
| 2 | C | 1 | | | 1.561100 0 | | |
| 3 | C | 1 | 2 | | 1.560078 0 | 109.396435 2 | |
| 4 | C | 2 | 1 | 3 | 1.560022 3 | 109.420071 3 | -0.0129610 |
| 5 | C | 2 | 1 | 3 | 1.525052 | 86.0589537 | -122.1425459 |

| | | | | | | | |
|----|---|----|----|---|---------------|-----------------|--------------|
| | | | | | 9 | | |
| 6 | C | 5 | 2 | 1 | 1.351600 2 | 93.9301152 | 0.0275699 |
| 7 | C | 2 | 1 | 6 | 1.524640 3 | 86.0703191 | -115.7486101 |
| 8 | C | 7 | 2 | 1 | 1.351600 0 | 93.9388686 | -0.0151477 |
| 9 | C | 4 | 2 | 1 | 1.533965 7 | 106.374640 1 | 56.4460337 |
| 10 | C | 9 | 4 | 2 | 1.350800 0 | 113.986913 9 | -59.3375106 |
| 11 | C | 4 | 2 | 1 | 1.533836 6 | 106.382249 7 | -56.4281174 |
| 12 | C | 11 | 4 | 2 | 1.350800 0 | 113.992394 5 | 59.3289303 |
| 13 | H | 3 | 1 | 6 | 1.092923 1 | 112.709922 4 | 82.4919109 |
| 14 | H | 4 | 2 | 1 | 1.092811 7 | 112.716053 4 | -179.9927022 |
| 15 | H | 5 | 2 | 1 | 1.085145 0 | 132.680880 8 | 179.9537673 |
| 16 | H | 7 | 2 | 1 | 1.085159 5 | 132.673253 5 | -179.9430544 |
| 17 | H | 6 | 5 | 2 | 1.085237 0 | 133.378907 0 | 179.9387908 |
| 18 | H | 8 | 7 | 2 | 1.085214 0 | 133.387716 1 | -179.9346487 |
| 19 | H | 9 | 4 | 2 | 1.081481 9 | 122.396578 5 | 119.2811654 |
| 20 | H | 11 | 4 | 2 | 1.081572 6 | 122.395321 2 | -119.2839144 |
| 21 | H | 10 | 9 | 4 | 1.081520 4 | 123.599728 5 | 178.6053235 |
| 22 | H | 12 | 11 | 4 | 1.081500 4 | 123.601594 0 | -178.6046770 |

1. «Граф валентних зв'язків».

Сформуємо правила побудови даного графу:

1. Першим кроком, із отриманої в програмному забезпеченні «Gaussview 6» Z – матриці (табл. 3.2.2) визначимо які саме елементи молекули зв'язані між собою валентним зв'язком – номер присвоєний елементу стовпчик «Tag» та номер елементу який зв'язаний з ним стовпчик «NA» і стовпчик таблиці під назвою «Bond», що показує довжину валентного зв'язку між ними.

2. Наступним кроком зробимо 2 таблиці: перша таблиця - це вузли, а друга таблиця - ребра, як показано в табл. 3.2.3.

Таблиця 3.2.3.

Таблиці вузлів та ребер майбутнього графу валентних зв'язків

| ID | LABEL | | 1 | 2 |
|----|-------|--|-----|----|
| 1 | C1 | | C4 | C3 |
| 2 | C2 | | C5 | C3 |
| 3 | C3 | | C6 | C1 |
| 4 | C4 | | C7 | C6 |
| 5 | C5 | | C8 | C1 |
| 6 | C6 | | C9 | C1 |
| 7 | C7 | | C10 | C2 |
| 8 | C8 | | C11 | C1 |
| 9 | C9 | | C12 | C2 |
| 10 | C10 | | H13 | C6 |
| 11 | C11 | | H14 | C1 |
| 12 | C12 | | H15 | C1 |
| 13 | H13 | | H16 | C1 |
| 14 | H14 | | H17 | C2 |
| 15 | H15 | | H18 | C2 |
| 16 | H16 | | H19 | C2 |
| 17 | H17 | | H20 | C2 |
| 18 | H18 | | H21 | C4 |
| 19 | H19 | | H22 | C4 |
| 20 | H20 | | | |
| 21 | H21 | | | |
| 22 | H22 | | | |

3. Імпортуємо отримані таблиці в програму для побудови графів «Gerhi 0.9.7» та пропишемо вагові коефіцієнти ребер у вікні – «таблиця даних» у колонку «вага», таким чином отримаємо граф валентних зв'язків, який представлено на рис. 3.2.6.

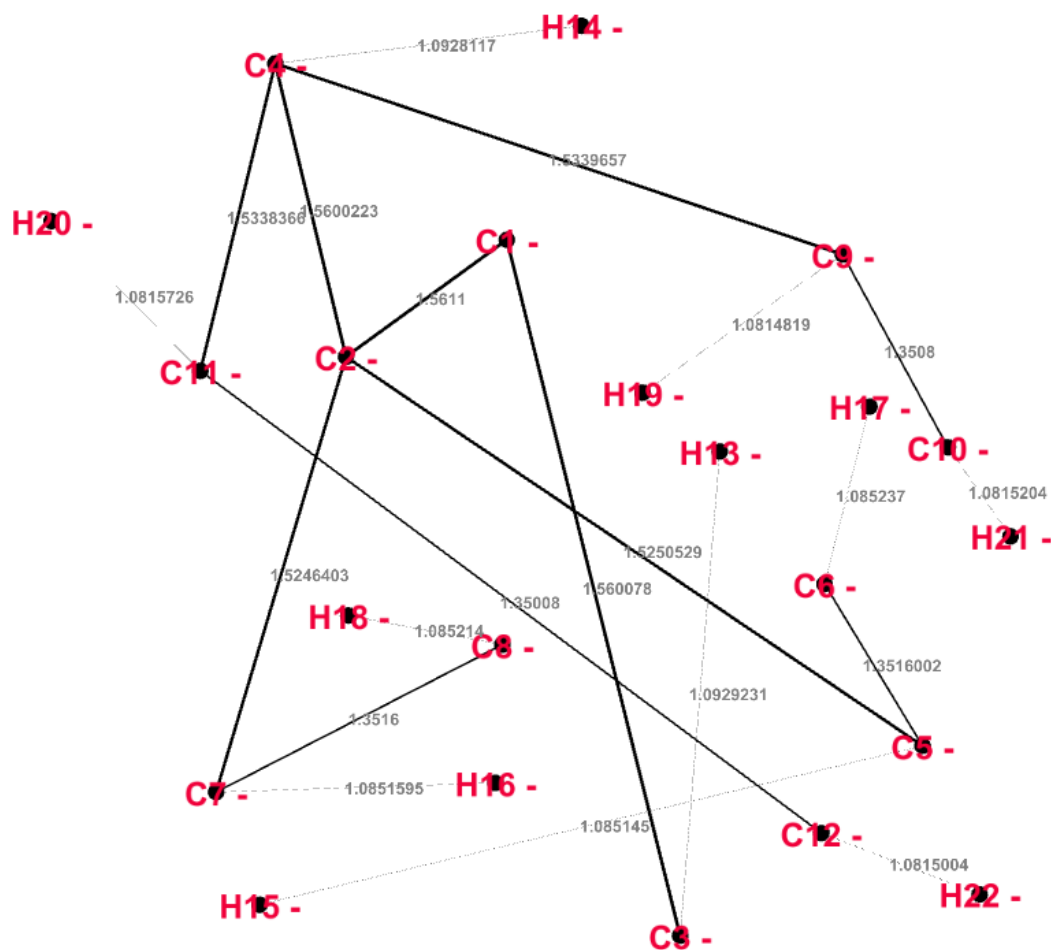


Рис. 3.2.6. Граф валентних зв'язків.

2. Наступним кроком побудуємо матрицю інцидентності. Матриця суміжності будується, щоб показати яким чином суміжні між собою всі елементи, що складають граф, оскільки це зважений граф, то додаються ще вагові коефіцієнти ребер. Таблиця наведена в додатку Б.

3. «Граф валентних кутів»

Сформуємо правила побудови даного графу:

1. Першим кроком, із отриманої в програмному забезпеченні «Gaussview 6» Z – матриці (табл. 3.2.2.) визначимо які саме пари елементів молекули утворюють між собою валентний кут з відповідним елементом іншої пари – пари елементу стовпчик «Tag» і «NB», та номер елементу з яким

утворюють між собою валентний кут ця пара, і стовпчик таблиці під назвою «Angle», що валентний кут, утворений між ними.

2. Наступним кроком зробимо 2 таблиці: перша таблиця - це вузли, а друга таблиця - ребра, як показано в табл. 3.2.4.

Таблиця 3.2.4.

Таблиці вузлів та ребер майбутнього графу валентних кутів

| ID | LABEL | 1 | 2 |
|----|----------|----------|-----|
| 1 | C2, C3 | C2, C3 | C1 |
| 2 | C1, C4 | C1, C4 | C2 |
| 3 | C1, C5 | C1, C5 | C2 |
| 4 | C2, C6 | C2, C6 | C5 |
| 5 | C1, C7 | C1, C7 | C2 |
| 6 | C2, C8 | C2, C8 | C7 |
| 7 | C2, C9 | C2, C9 | C4 |
| 8 | C4, C10 | C4, C10 | C9 |
| 9 | C2, C11 | C2, C11 | C7 |
| 10 | C4, C12 | C4, C12 | C11 |
| 11 | C1, H13 | C1, H13 | C3 |
| 12 | C2, H14 | C2, H14 | C4 |
| 13 | C2, H15 | C2, H15 | C5 |
| 14 | C2, H16 | C2, H16 | C7 |
| 15 | C5, H17 | C5, H17 | C6 |
| 16 | C7, H18 | C7, H18 | C8 |
| 17 | C4, H19 | C4, H19 | C9 |
| 18 | C4, H20 | C4, H20 | C11 |
| 19 | C9, H21 | C9, H21 | C10 |
| 20 | C11, H22 | C11, H22 | C12 |

3. Імпортуємо отримані таблиці в програму для побудови графів «Gerhi 0.9.7» та пропишемо вагові коефіцієнти ребер у вікні – «таблиця даних» у колонку «вага», таким чином отримаємо граф валентних зв'язків, який представлено на рис. 3.2.7.

молекули, що входять в спільний валентний кут, ребра даного графа представляють торсіонний кут поточної трійки і відповідного елемента іншої трійки. Трійки елементів стовпчик «Tag», «NB», «NC», та номер елемента з яким утворюють між собою торсіонний кут ця трійка «NA», і стовпчик таблиці під назвою «Angle» - торсіонний кут, утворений між ними.

2. Наступним кроком зробимо 2 таблиці: перша таблиця - це вузли, а друга таблиця - ребра, як показано на рис. в табл. 3.2.5.

Таблиця 3.2.5.

Таблиці вузлів та ребер майбутнього графу торсіонних кутів

| ID | LABEL | 1 | 2 |
|----|--------------|--------------|-----|
| 1 | C3, C1, C4 | C3, C1, C4 | C2 |
| 2 | C3, C2, C5 | C3, C2, C5 | C2 |
| 3 | C1, C2, C6 | C1, C2, C6 | C5 |
| 4 | C6, C1, C7 | C6, C1, C7 | C2 |
| 5 | C1, C2, C8 | C1, C2, C8 | C7 |
| 6 | C1, C2, C9 | C1, C2, C9 | C4 |
| 7 | C2, C4, C10 | C2, C4, C10 | C9 |
| 8 | C1, C2, C11 | C1, C2, C11 | C4 |
| 9 | C2, C4, C12 | C2, C4, C12 | C11 |
| 10 | C6, C1, H13 | C6, C1, H13 | C3 |
| 11 | C1, C2, H14 | C1, C2, H14 | C4 |
| 12 | C1, C2, H15 | C1, C2, H15 | C5 |
| 13 | C1, C2, H16 | C1, C2, H16 | C7 |
| 14 | C2, C5, H17 | C2, C5, H17 | C6 |
| 15 | C2, C7, H18 | C2, C7, H18 | C8 |
| 16 | C2, C4, H19 | C2, C4, H19 | C9 |
| 17 | C2, C4, H20 | C2, C4, H20 | C11 |
| 18 | C4, C9, H21 | C4, C9, H21 | C10 |
| 19 | C4, C11, H22 | C4, C11, H22 | C12 |

3. Імпортуємо отримані таблиці в програму для побудови графів «Gephi 0.9.7» та пропишемо вагові коефіцієнти ребер у вікні – «таблиця даних» у колонку «вага», таким чином отримаємо граф валентних зв'язків, який представлено на рис. 3.2.8.

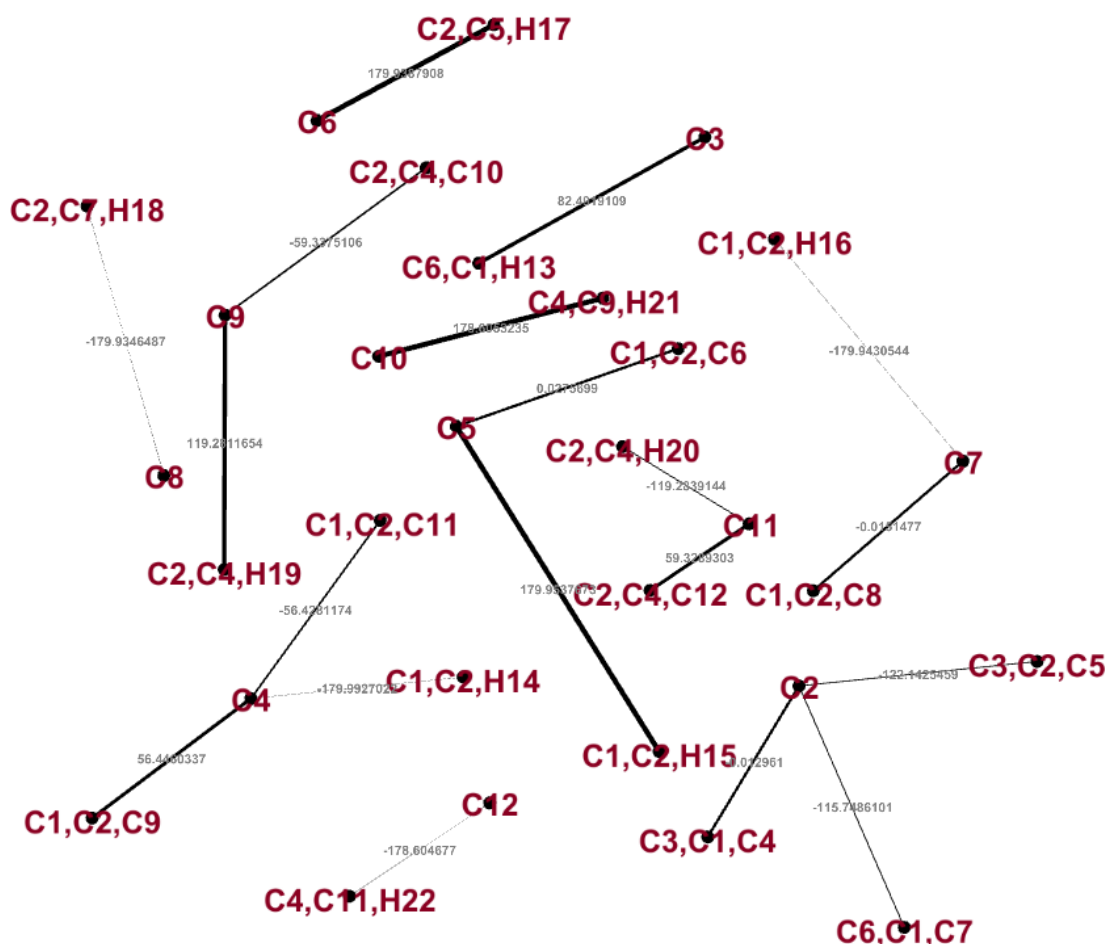


Рис. 3.2.8. Граф торсіонних кутів

4. Наступним кроком побудуємо матрицю інцидентності. Матриця суміжності будується, щоб показати яким чином суміжні між собою всі елементи, що складають граф, оскільки це зважений граф, то додаються ще вагові коефіцієнти ребер. Таблиця наведена в додатку Г.

5. Наступним кроком представимо доведення еквівалентності перетворення зваженого графа в z – матрицю

Сформуємо правила перетворення графу в z -матрицю:

1. Першим кроком для того, щоб побудувати z -матрицю з графу, візьмемо матрицю інцидентності першого графу- «графу валентних зв'язків» (табл. 3.2.2). Треба пам'ятати, що в першому рядку z - матриці

пишеться тільки символ елемента, якому присвоєний номер 1. Далі почнемо заповнювати матрицю, врахувавши, які елементи мають спільний валентний зв'язок між собою, та яка довжина цього валентного зв'язку. Дивлячись на матрицю суміжності, можемо побачити вузли які зв'язані між собою ребрами (табл.), ваговими коефіцієнтами ребер є саме довжини валентних зв'язків, які потрібно записати в Z – матрицю, як показано в табл в доданку Д.

Таблиця 3.2.7

Взаємозв'язок елементів в матриці інцидентності графу валентних зв'язків

| | C1 | C2 | C3 | C4 | C5 | C6 | C7 | C8 | C9 | C10 |
|-----|------------|-------------|-------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| C1 | 0 | 1,5611 | 1.560 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C2 | 1,561 1 | 0 | 0 | 0 | 1.5250 5 | 0 | 1.5246 4 | 0 | 0 | 0 |
| C3 | 1.560 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| C4 | 0 | 1.5600 2 | 0 | 0 | 0 | 0 | 0 | 0 | 1.5339 6 | 0 |
| C5 | 0 | 1.5250 5 | 0 | 0 | 0 | 1.3516 0 | 0 | 0 | 0 | 0 |
| C6 | 0 | 0 | 0 | 0 | 1.3516 0 | 0 | 0 | 0 | 0 | 0 |
| C7 | 0 | 1.5246 4 | 0 | 0 | 0 | 0 | 0 | 1.3516 0 | 0 | 0 |
| C8 | 0 | 0 | 0 | 0 | 0 | 0 | 1.3516 | 0 | 0 | 0 |
| C9 | 0 | 0 | 0 | 1.5339 6 | 0 | 0 | 0 | 0 | 0 | 1.3508 0 |
| C10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1.3508 0 | 0 |

2. Наступним кроком, на основі матриці інцидентності другого графу - «графу валентних кутів», заповнимо другий стовпчик Z - матриці, де ваговим коефіцієнтом ребра є валентний кут між двома елементами молекули.

Валентний кут – кут, утворений між двома напрямками хімічних зв'язків, що виходять з одного атома, тому наприклад взявши зв'язок C2,C3 - C1, потрібно врахувати, що елемент C1, утворює з елементами C2 та C3 валентний кут . Отже в результаті отримуємо другий стовпчик, таблиця наведена в додатку Д.

3. Останнім кроком, на основі матриці інцидентності третього графу - «графу торсіонних кутів», заповнимо третій стовпчик z - матриці, де ваговим коефіцієнтом ребра є торсіонний кут між трьома елементами молекули.

В даному випадку, потрібно врахувати, що торсіонний кут – кут, утворений двома напівплощинами, що виходять з однієї прямої, а також частина простору, обмежена цими напівплощинами, тому наприклад елемент C_2 , утворює з елементами C_3 , C_1 , C_4 торсіонний кут. Таким чином отримуємо останній стовпчик z - матриці, таблиця наведена в додатку Д.

Отже, використовуючи правила, які описані в даному розділі, можна перетворити спорідений граф в матрицю z – структури молекул.

3.3. Алгоритм встановлення відповідності між мас-спектрометричним паттерном і структурою молекулярного фрагмента

В даному розділі описаний розроблений алгоритм виділення підграфа з прив'язкою до мас-спектрів отриманих раніше паттернів рис. 3.3.1.

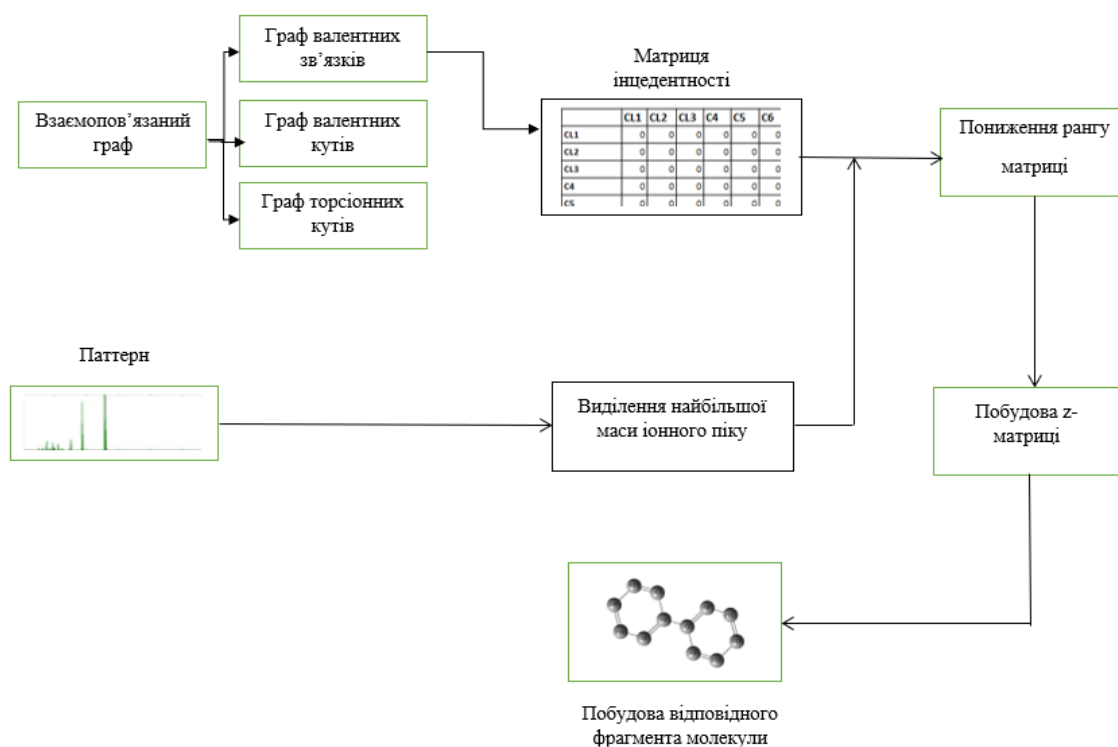


Рис. 3.3.1. Блок-схема алгоритму співвідношення графу до паттерна.

Як можна побачити з приведеної блок-схеми, для того щоб прив'язати граф до паттерна за масою, необхідно взяти за основу матрицю інцидентності.

Важливим моментом є те, що масовий склад масспектрів не несе інформацію про просторову структуру молекули, тому для нашої задачі достатньо лише одного графа, а саме графа валентних зв'язків.

Із паттерну потрібно виділити найбільший пік іонної маси. Далі необхідно перебудувати структуру матриці понизивши її ранг, таким чином щоб при складанні маси вузлів, сума мас елементів дорівнювала тій самій найвищій масі іонного піку.

Наступним кроком в програмному забезпеченні «Gaussian» за отриманою z-матрицею будемо відповідний фрагмент молекули.

РОЗДІЛ 4

РЕЗУЛЬТАТИ ЕКСПЕРИМЕНТІВ

4.1. Створення бази даних мас-спектрів та просторових структур хлороорганічних молекул з близьким елементним складом

Апріорні дані щодо структури молекул були отримані прямими експериментальними мас-спектрометричними вимірюваннями стандартних аналітичних розчинів відповідних речовин. Результати проведених вимірювань описані такими характеристиками, як: назва речовини, формула, характерні мас-спектри для точної кількосної ідентифікації поліхлорованих біфенілів, реєстраційний номер CAS та просторова структура моделі, таблиця наведена в додатку Е.

Представлена таблиця в додатку Е, містить в собі характерні мас-спектри для точної кількісної ідентифікації поліхлорованих біфенілів.

Мас-спектри багатоатомних органічних молекул являють собою багатозв'язний масив молекулярних іонів. На рис. 4.1.1. наведені мас-спектри молекул ізомерів поліхлорованих біфенілів, у яких є однакова формула брутто $C_{12}H_7Cl_3$; PCB_{15} , PCB_{19} , PCB_{20} , PCB_{22} .

З представлених мас-спектрів на рис. 4.1.1 можна побачити, що молекули вміщують в собі однакові елементи, вміщують схожі та відмінні фрагменти. Мас-спектри таких молекул мають однаковий діапазон іонних мас і вміщують в своєму складі однакові та різні послідовності іонів.

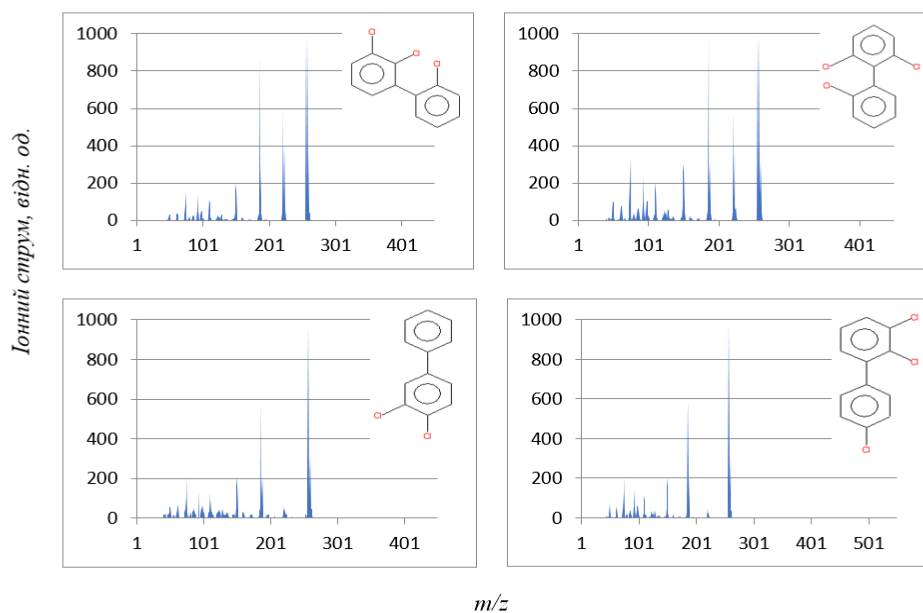


Рис. 4.1.1. Мас-спектри молекул ізомерів поліхлорованих біфенілів із схожою структурою та однаковими елементами.

Окремі фрагменти цього іонного масиву складаються з вершин, розташованих близько одна до одної, що корелює зі стабільними фрагментами багатоатомної молекули. Якщо використовувати класичні аналітичні методи, група іонних піків, що належать до кожної окремої групи, апроксимується неперервною кривою, а параметри цих апроксимацій використовуються як описи відповідних молекул.

Цей метод знижує надійність молекулярної ідентифікації, враховуючи при цьому окремо локалізовані фрагменти мас-спектрометрії, які довели свою ефективність у виявленні молекул зі схожим елементним складом.

4.2. Виділення мас-спектроскопічних паттернів поліхлорбіфенілів

Для того щоб зробити виділення однакових масспектроскопічних паттернів, було розроблено алгоритм. Використовуючи цей алгоритм був

розроблений і налагоджений скрипт у ПЗ MatLab та були отримані маспектрометричні паттерни. Виконавши цей алгоритм були отримані паттерни ізомерів молекул PCB_{16} та PCB_{22} . Тестування алгоритму проводилися на ізомері молекули PCB_{18} . Результати тестування алгоритму представленні на рис. 4.2.1.

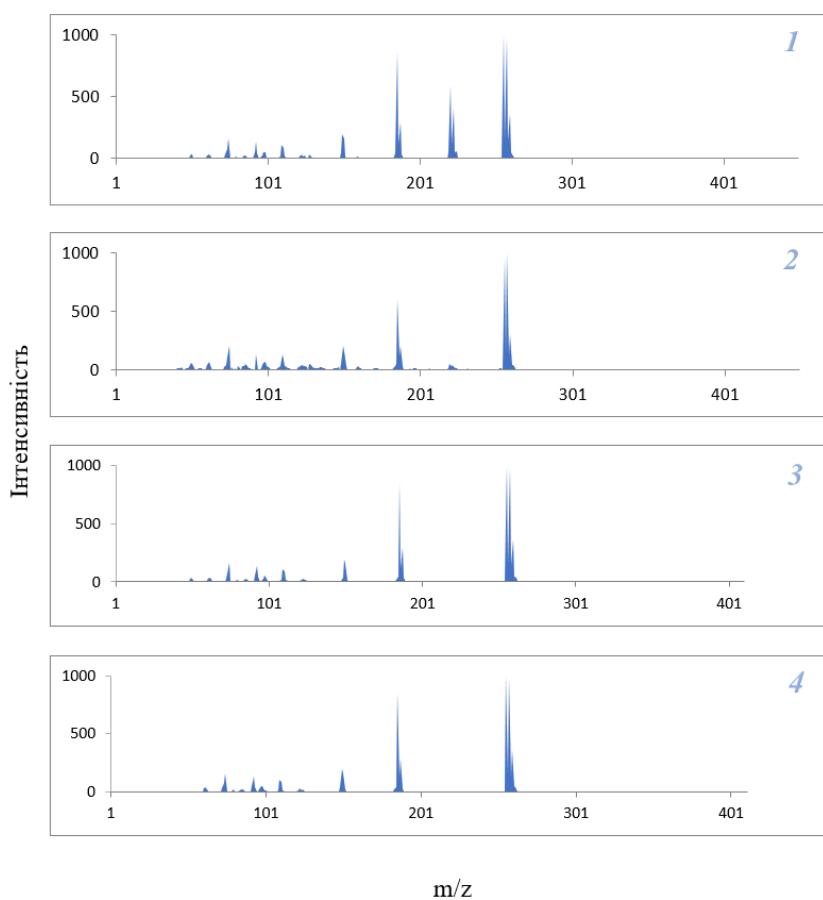


Рис. 4.2.1. Результати тестування алгоритму виділення схожих маспектроскопічних паттернів

Для того щоб протестувати алгоритм був виділений паттерн, що отриманий використовуючи мас-спектри вхідних молекул із мас-спектру експериментальної молекули.

На рис. 4.2.1 представлені діаграми 1, 2 – це мас-спектри вхідних ізомерів і отриманий з цих мас-спектрів паттерн – наведених на діаграмі 3, паттерн виділений з експериментальної молекули наведено на діаграмі 4.

На цих діаграмах зображені мас-спектри молекул, що мають практично однакову форму, але різняться гомологічним рядом. У ізомера молекули - PCB_{16} інтенсивність більше за інтенсивність ізомера молекули PCB_{22} – складає 70% амплітуди. Форма гомологічної послідовності збігається з точністю критерію порівняння. Алгоритм не зберіг цю послідовність подібностей, оскільки форма мас-спектрів відрізнялася лише цією послідовністю. Тому після виконаної роботи розроблено, налагоджено та перевірено на експериментальній мас-спектрометрії методику розпізнавання молекулярних фрагментів хлорорганіки на основі мас-спектрометричних зразків.

Алгоритм методу має досить високу швидкість, що відкриває перспективу його застосування в обробці сигналів реального часу. Цей метод відбирає в мас-спектрометрії багатоатомних молекул ідентичні ланцюги молекулярних іонів для молекул зі схожими структурами. Ці іонні ланцюги можуть служити дескрипторами тих самих фрагментів в ізомерах багатоатомних молекул і представляти мас-спектроскопічні картини цих фрагментів.

Аналіз штучних зображень просторової структури молекул допоміг виявити загальні фрагменти в молекулах поліхлорованого біфенілу. Для аналізу зображень просторової структури молекул було розраховано перші 7 моментів Церніке за допомогою сценарію програмного коду в програмному забезпеченні MatLab для молекулярних ізомерів і отримано моменти Церніке, показано на рис. 4.2.2.

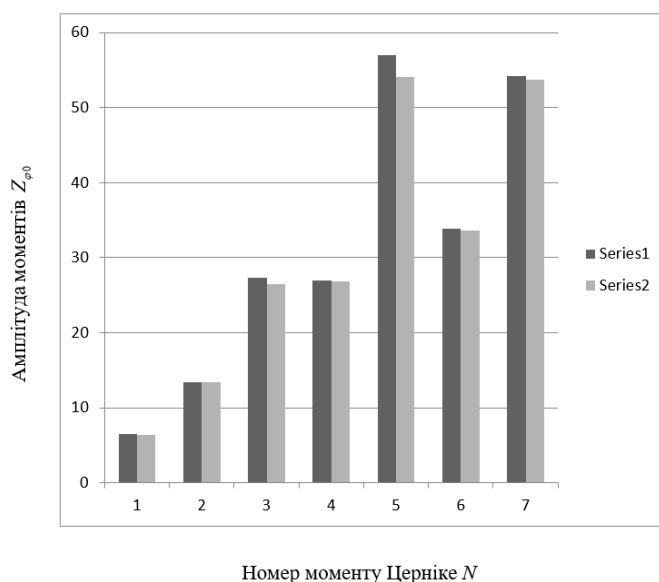


Рис. 4.2.2. Отримані після виконання алгоритму моменти Церніке

На рис. 4.2.2 представлені моменти Церніке, подані в логарифмічній шкалі, з цих спектральних графіків і з отриманих результатів можна зробити висновок, що виділення подібних фрагментів просторової структури молекул поліхлорованих бііфенілів за моментами Церніке не є ефективним, оскільки самі молекули мають дуже близьку симетрію і, відповідно за моментами Церніке, ізомери практично не можна розрізнити.

Тому для подальшої роботи цей метод не слід використовувати і в подальшому пропонується аналізувати просторову структуру молекул за допомогою z-матричного представлення структури молекул у вигляді 'графів.

4.3. Доведення еквівалентності представлень структури молекул координатною матрицею та зваженим гіперграфом

Першим кроком побудуємо взаємопов'язаний граф для ізомерів тих молекул, що використовувались для виділення спільних маспектроскопічних паттернів, а саме $C_{12}H_7Cl_3$; PCB_{16} , PCB_{22} .

Для виконання алгоритму побудови графів із z-матриць структур молекул візьмемо молекулярну структуру ізомеру трихлорбіфенілу $C_{12}H_7Cl_3$ - PCB_{16} рис. 4.3.1, та її z-матрицю таблиця.

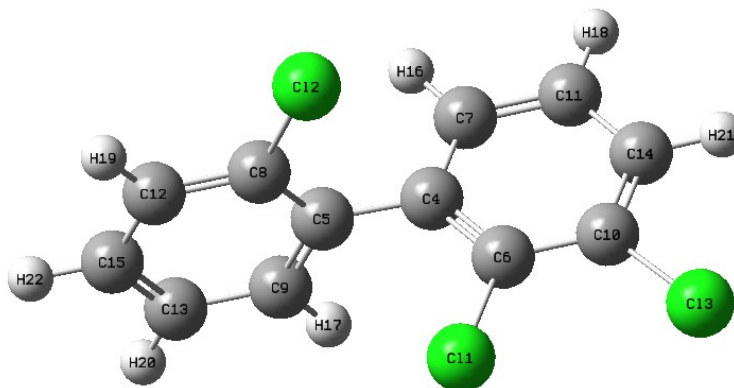


Рис. 4.3.1. Ізомер молекули трихлорбіфенілу $C_{12}H_7Cl_3$ - PCB_{16} .

Таблиця 4.3.1.

Z-матриця ізомеру молекули трихлорбіфенілу $C_{12}H_7Cl_3$ - PCB_{16}

| Tag | Symbol | NA | NB | NC | Bond | Angle | Dihedral |
|-----|--------|----|----|----|-----------|-------------|--------------|
| 1 | Cl | | | | | | |
| 2 | Cl | 1 | | | 4.0869414 | | |
| 3 | Cl | 1 | 2 | | 3.2023746 | 107.1743626 | |
| 4 | C | 1 | 3 | 2 | 2.7167638 | 85.4529091 | 44.5017832 |
| 5 | C | 4 | 1 | 3 | 1.4507446 | 88.4809896 | 179.9889348 |
| 6 | C | 4 | 1 | 5 | 1.3952485 | 33.7633876 | 179.9989784 |
| 7 | C | 4 | 1 | 6 | 1.3953412 | 153.7090256 | 0.0317301 |
| 8 | C | 5 | 4 | 1 | 1.3948091 | 121.8429000 | 90.0433009 |
| 9 | C | 5 | 4 | 1 | 1.3949487 | 118.1568417 | -89.9843889 |
| 10 | C | 6 | 4 | 1 | 1.3949115 | 120.0242892 | 179.9840105 |
| 11 | C | 7 | 4 | 1 | 1.3947947 | 120.0262868 | -0.0247915 |
| 12 | C | 8 | 5 | 4 | 1.3947979 | 120.0017270 | 179.9831044 |
| 13 | C | 9 | 5 | 4 | 1.3948924 | 119.9936998 | -179.9840697 |
| 14 | C | 11 | 7 | 4 | 1.3948655 | 120.0036620 | 0.0040816 |
| 15 | C | 13 | 9 | 5 | 1.3948025 | 120.0076882 | 0.0019413 |
| 16 | H | 7 | 4 | 1 | 1.0877837 | 121.1149398 | -179.9968967 |
| 17 | H | 9 | 5 | 4 | 1.0876353 | 121.1067818 | 0.0452321 |
| 18 | H | 11 | 7 | 4 | 1.0860625 | 119.9990024 | -179.9858768 |
| 19 | H | 12 | 8 | 5 | 1.0871420 | 120.9675748 | 179.9973788 |
| 20 | H | 13 | 9 | 5 | 1.0861401 | 120.0023149 | -179.9876773 |
| 21 | H | 14 | 11 | 7 | 1.0871874 | 118.9822482 | -179.9938561 |
| 22 | H | 15 | 13 | 9 | 1.0860363 | 119.9941417 | -180.0000000 |

Сформуємо правила побудови взаємопов'язаного графу:

1. Першим кроком, із отриманої в програмному забезпеченні «Gaussview 6» z – матриці (табл. 4.3.1) визначимо які саме елементи молекули зв'язані між собою валентним зв'язком – номер присвоєний елементу стовпчик «Tag» та номер елементу який зв'язаний з ним стовпчик «NA» і стовпчик таблиці під назвою «Bond», що показує довжину валентного зв'язку між ними – для графу валентних зв'язків.

Визначимо, які саме пари елементів молекули утворюють між собою валентний кут з відповідним елементом іншої пари – пари елементу стовпчик «Tag» і «NB», та номер елементу з яким утворюють між собою валентний кут ця пара, і стовпчик таблиці під назвою «Angle», що валентний кут, утворений між ними- для графу валентних кутів.

Визначимо, які саме трійки елементів молекули, що входять в спільний валентний кут, ребра даного графа представляють торсіонний кут поточної трійки і відповідного елемента іншої трійки. Трійки елементів стовпчик «Tag», «NB», «NC», та номер елементу з яким утворюють між собою торсіонний кут ця трійка «NA», і стовпчик таблиці під назвою «Angle» - торсіонний кут, утворений між ними – для графу торсіонних кутів.

2. Наступним кроком зробимо 2 таблиці: перша таблиця - це вузли, а друга таблиця - ребра, як показано в табл. 4.3.2. Такі самі таблиці також будуємо для двох інших графів.

Таблиці вузлів та ребер майбутнього графу валентних зв'язків

| ID | LABEL | | 1 | 2 |
|----|-------|--|-----|-----|
| 1 | C12 | | C12 | C11 |
| 2 | C13 | | C13 | C11 |
| 3 | C4 | | C4 | C11 |
| 4 | C5 | | C5 | C4 |
| 5 | C6 | | C6 | C4 |
| 6 | C7 | | C7 | C4 |
| 7 | C8 | | C8 | C5 |
| 8 | C9 | | C9 | C5 |
| 9 | C10 | | C10 | C6 |
| 10 | C11 | | C11 | C7 |
| 11 | C12 | | C12 | C8 |
| 12 | C13 | | C13 | C9 |
| 13 | C14 | | C14 | C11 |
| 14 | C15 | | C15 | C13 |
| 15 | H16 | | H16 | C7 |
| 16 | H17 | | H17 | C9 |
| 17 | H18 | | H18 | C11 |
| 18 | H19 | | H19 | C12 |
| 19 | H20 | | H20 | C13 |
| 20 | H21 | | H21 | C14 |
| 21 | H22 | | H22 | C15 |

3. Наступним етапом імпортуємо отримані таблиці в програму для побудови графів «Gerhi 0.9.7» та пропишемо вагові коефіцієнти ребер у вікні – «таблиця даних» у колонку «вага», таким чином отримаємо три графи, він вміщує в себе: граф валентних зв'язків, валентних кутів та торсіонних кутів. який представлено на рис. 4.3.2.

6. Для доведення еквівалентності представлень структури молекул координатною матрицею та зваженим графом, спочатку використовуючи зв'язок елементів, в отриманих матрицях інцидентності, будуємо z -матрицю, як показано в табл. 4.3.3

Таблиця 4.3.3

Побудова z – матриці з матриць інцидентності графів

| № | Елемент | NA | NB | NC | Довжина валентного зв'язку | Валентий кут | Торсіонний кут |
|----|---------|----|----|----|-------------------------------|--------------|----------------|
| 1 | Cl | | | | | | |
| 2 | Cl | 1 | | | 4.0869414 | | |
| 3 | Cl | 1 | 2 | | 3.2023746 | 107.1743626 | |
| 4 | C | 1 | 3 | 2 | 2.7167638 | 85.4529091 | 44.5017832 |
| 5 | C | 4 | 1 | 3 | 1.4507446 | 88.4809896 | 179.9889348 |
| 6 | C | 4 | 1 | 5 | 1.3952485 | 33.7633876 | 179.9989784 |
| 7 | C | 4 | 1 | 6 | 1.3953412 | 153.7090256 | 0.0317301 |
| 8 | C | 5 | 4 | 1 | 1.3948091 | 121.8429000 | 90.0433009 |
| 9 | C | 5 | 4 | 1 | 1.3949487 | 118.1568417 | -89.9843889 |
| 10 | C | 6 | 4 | 1 | 1.3949115 | 120.0242892 | 179.9840105 |
| 11 | C | 7 | 4 | 1 | 1.3947947 | 120.0262868 | -0.0247915 |
| 12 | C | 8 | 5 | 4 | 1.3947979 | 120.0017270 | 179.9831044 |
| 13 | C | 9 | 5 | 4 | 1.3948924 | 119.9936998 | -179.9840697 |
| 14 | C | 11 | 7 | 4 | 1.3948655 | 120.0036620 | 0.0040816 |
| 15 | C | 13 | 9 | 5 | 1.3948025 | 120.0076882 | 0.0019413 |
| 16 | H | 7 | 4 | 1 | 1.0877837 | 121.1149398 | -179.9968967 |
| 17 | H | 9 | 5 | 4 | 1.0876353 | 121.1067818 | 0.0452321 |
| 18 | H | 11 | 7 | 4 | 1.0860625 | 119.9990024 | -179.9858768 |
| 19 | H | 12 | 8 | 5 | 1.0871420 | 120.9675748 | 179.9973788 |
| 20 | H | 13 | 9 | 5 | 1.0861401 | 120.0023149 | -179.9876773 |
| 21 | H | 14 | 11 | 7 | 1.0871874 | 118.9822482 | -179.9938561 |
| 22 | H | 15 | 13 | 9 | 1.0860363 | 119.9941417 | -180.0000000 |

7. Далі з отриманої z -матриці побудуємо гіперграф, використовуючи програму для побудови графів «Gephi 0.9.7» .

8. На рис. 4.3.3 представлено зіставленні два графи: перший граф, який було побудовано спочатку, а другий - після еквівалентного перетворення. Наступним кроком побудуємо матриці інцидентності та використовуючи зв'язки в них, запишемо координатну матрицю, як показано в табл.4.3.3.

Таблиця 4.3.4

Побудова z – матриці з матриць інцидентності графів

| № | Елемент | NA | NB | NC | Довжина валентного зв'язку | Валентний кут | Торсіонний кут |
|----|---------|----|----|----|-------------------------------|---------------|----------------|
| 2 | Cl | 1 | | | 4.0869414 | | |
| 3 | Cl | 1 | 2 | | 3.2023746 | 107.1743626 | |
| 4 | C | 1 | 3 | 2 | 2.7167638 | 85.4529091 | 44.5017832 |
| 5 | C | 4 | 1 | 3 | 1.4507446 | 88.4809896 | 179.9889348 |
| 6 | C | 4 | 1 | 5 | 1.3952485 | 33.7633876 | 179.9989784 |
| 7 | C | 4 | 1 | 6 | 1.3953412 | 153.7090256 | 0.0317301 |
| 8 | C | 5 | 4 | 1 | 1.3948091 | 121.8429000 | 90.0433009 |
| 9 | C | 5 | 4 | 1 | 1.3949487 | 118.1568417 | -89.9843889 |
| 10 | C | 6 | 4 | 1 | 1.3949115 | 120.0242892 | 179.9840105 |
| 11 | C | 7 | 4 | 1 | 1.3947947 | 120.0262868 | -0.0247915 |
| 12 | C | 8 | 5 | 4 | 1.3947979 | 120.0017270 | 179.9831044 |
| 13 | C | 9 | 5 | 4 | 1.3948924 | 119.9936998 | -179.9840697 |
| 14 | C | 11 | 7 | 4 | 1.3948655 | 120.0036620 | 0.0040816 |
| 15 | C | 13 | 9 | 5 | 1.3948025 | 120.0076882 | 0.0019413 |
| 16 | H | 7 | 4 | 1 | 1.0877837 | 121.1149398 | -179.9968967 |
| 17 | H | 9 | 5 | 4 | 1.0876353 | 121.1067818 | 0.0452321 |
| 18 | H | 11 | 7 | 4 | 1.0860625 | 119.9990024 | -179.9858768 |
| 19 | H | 12 | 8 | 5 | 1.0871420 | 120.9675748 | 179.9973788 |
| 20 | H | 13 | 9 | 5 | 1.0861401 | 120.0023149 | -179.9876773 |
| 21 | H | 14 | 11 | 7 | 1.0871874 | 118.9822482 | -179.9938561 |
| 22 | H | 15 | 13 | 9 | 1.0860363 | 119.9941417 | -180.0000000 |

9. Наступним етапом за отриманою z – матрицею в програмному забезпеченні «Gaussian», побудуємо молекулу та зіставимо з початковою, як показано на рис. 4.3.4.

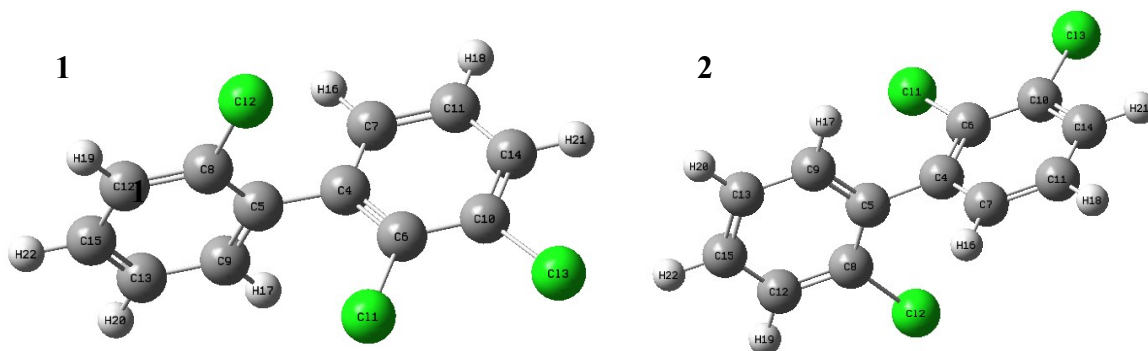


Рис. 4.3.4. Порівняння двох молекул: 1 - початкова, 2 – після перетворення.

Отже, виходячи з наведеного вище рис 4.3.4, можна зробити висновок, що при після проведеного доведення еквівалентності представлень структури молекул координатною матрицею та зваженим графом, молекулярна структура залишається незмінною.

4.4. Відновлення структури молекулярного фрагмента за мас-спектрометричним паттерном

В данному підрозділі розглядається відповідність мас-спектрів до отриманих раніше паттернів до структури графу.

Використовуючи попередньо розроблений алгоритм представимо відповідність графа до виділеного паттерну, для цього потрібно використовувати перший граф – граф валентних зв'язків, де сума мас зв'язаних вершин графа має бути рівною найбільшій масі іонного піку в нашому паттерні. Як можна побачити з рис. 4.4.1, найбільша маса іонного піку приблизно дорівнює 260 m/z .

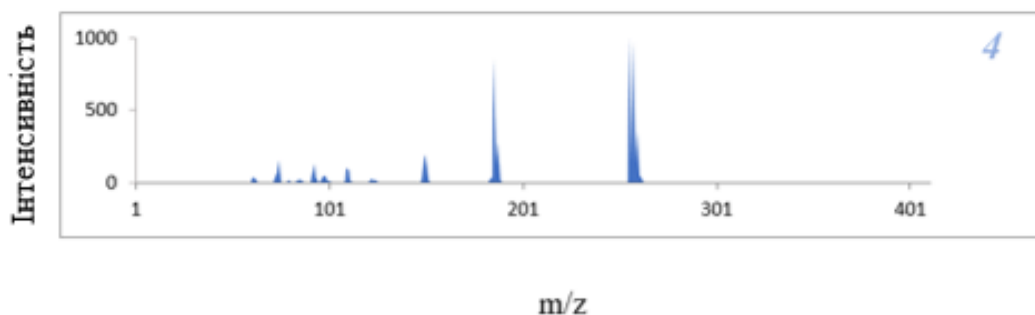


Рис. 4.4.1. Мас-спектрометричний паттерн ізомерів молекул PCB_{16} та PCB_{22}

Використовуючи матрицю інцидентності графу валентних зв'язків складаємо маси вузлів, так щоб сума мас елементів дорівнювала 260, та в результаті отримуємо z-матрицю паттерна молекули трихлорбіфенілу, як представлено в табл. 4.4.1.

Таблиця 4.4.1

Z-матриця паттерна молекули трихлорбіфенілу

| Tag | Symbol | NA | Bond | Angle | Dihedral |
|-----|--------|----|-----------|-------------|--------------|
| 5 | C | 4 | 1.4469702 | 88.5632635 | 179.9883321 |
| 6 | C | 4 | 1.3974017 | 33.8173521 | 179.9961412 |
| 7 | C | 4 | 1.3970540 | 153.4900729 | 0.0255859 |
| 8 | C | 5 | 1.3947278 | 119.9892731 | -89.9789477 |
| 9 | C | 5 | 1.3949365 | 120.0040581 | 90.0575028 |
| 10 | C | 6 | 1.3947776 | 120.1596499 | 179.9820495 |
| 11 | C | 7 | 1.3948603 | 120.1629154 | -0.0163317 |
| 12 | C | 8 | 1.3949534 | 119.9933402 | -179.9477036 |
| 13 | C | 9 | 1.3948783 | 119.9917762 | 179.9576765 |
| 14 | C | 11 | 1.3948775 | 120.0052167 | 0.0047125 |
| 15 | C | 12 | 1.3947692 | 120.0045377 | -0.0205449 |

Наступним кроком в програмному забезпеченні «Gaussian» за отриманою z-матрицею будуємо відповідний фрагмент молекули, як показано на рис. 4. 4. 2.

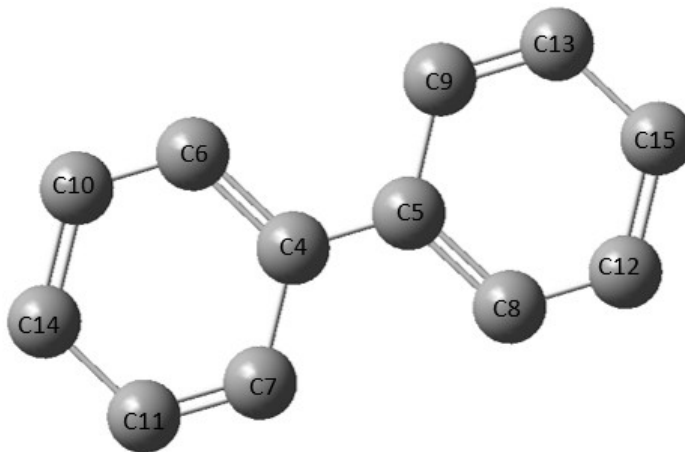


Рис. 4. 4. 2. Побудований фрагмент молекули в ПЗ «Gaussian».

Отже, представити відповідність графа до отриманого паттерна, можна використовуючи таблицю інцидентності графу валентних зв'язків, склавши маси елементів таким чином щоб отримана маса відповідала найвищому іонному піку мас-спектра паттерна. Далі створити z-матрицю з отриманих елементів, та використовуючи їх зв'язок побудувати даний фрагмент молекули.

ВИСНОВОК

В представленій роботі був зроблений аналіз методів виділення схожих фрагментів складних сигналів, а саме: класифікація, регресія, кластеризація, асоціація, послідовність та ін. Визначено, що ці методи є недостатньо точними для даних фрагментованих сигналів, таких як мас-спектри багатоатомних молекул. Тож запропонували новий метод, який використовує відстань Фреше, як міру схожості кривих. Міра схожості кривих являє собою відстань між опорним і фрагментом, який досліджується.

Результатом роботи є розроблений, налагоджений і перевірений на експериментальних мас-спектрах алгоритм розпізнавання фрагментів хлороганічних молекул за маспектрометричними паттернами. За допомогою цього методу можна виділити з мас-спектрів багатоатомних молекул послідовності молекулярних іонів, що є однаковими для молекул з подібною структурою. Такі послідовності молекулярних іонів можуть правити за дескриптори однакових та схожих фрагментів у ізомерах багатоатомних молекул і представляють мас-спектрометричні паттерни тих фрагментів.

Представлений алгоритм аналізу зображення молекул поліхлорованих біфенілів, для виділення схожих фрагментів просторової структури виявився мало придатним, так як ізомери використовуючи моменти Церніке майже не можна розрізнити. Тому для задачі попереднього оброблення мас-спектроскопічного сигналу було запропоновано та розроблено алгоритм побудови графів на основі z-матриць структур молекул і його еквівалентного перетворення, та алгоритм виділення підграфа з прив'язкою до виділеного мас-спектра паттерна.

Результатом дипломної роботи є розроблені алгоритми попереднього оброблення мас-спектроскопічного сигналу для автоматизованої ідентифікації хлороорганічних молекул.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Wang, S. C. (1928-04-01). "The Problem of the Normal Hydrogen Molecule in the New Quantum Mechanics". *Physical Review*. 31 (4): 579–586. doi:10.1103/PhysRev.31.579
2. Лукіянець Б.А. та ін. *Основи квантової фізики* / Львівська політехніка 2009р. – 420 с.
3. McQuarrie, Donald A. (2007). *Quantum Chemistry* (2nd ed.). University Science Books. ISBN 978-1891389504.
4. Pauling, Linus (1939). *The Nature of the Chemical Bond and the Structure of Molecules and Crystals: An Introduction to Modern Structural Chemistry* (1st ed.). Cornell University Press.
5. Hellmann, Hans (1937). *Einführung in die Quantenchemie*. Deuticke, Leipzig und Wien.
6. Lewis, G.N. "The Atom and the Molecule". *Journal of the American Chemical Society*. 38: 762–785.
7. Shaik, S.S.; Hiberty, P.C. (2007). *A Chemist's Guide to Valence Bond Theory*. Wiley-Interscience. ISBN 978-0470037355.
8. Frisch, E. *Gaussian 09 User's Reference* / Frisch E., Frisch M.J., Clemente F.R., Trucks G.W. // Gaussian Inc. - 2009
9. David L. Cooper; Terry P. Cunningham; Joseph Gerratt; Peter B. Karadakov; Mario Raimondi (1994). "Chemical Bonding to Hypercoordinate Second-Row Atoms: d Orbital Participation versus Democracy". *Journal of the American Chemical Society*. 116 (10): 4414–4426. doi:10.1021/ja00089a033

ДОДАТОК А

```

function main()
%-----%
%-----%
clear all;
% Change default axes fonts.
set(0, 'DefaultAxesFontName', 'Times New Roman')
set(0, 'DefaultAxesFontSize', 12)

% Change default text fonts.
set(0, 'DefaultTextFontname', 'Times New Roman')
set(0, 'DefaultTextFontSize', 12)
%=====
=====
Frg=[];
Eps=0;
Tr_h=0.98;
N=552; % Number of MS peaks
Hd='A3';
Tl=':B';
MS=[];
fDst=[];
Pttr=[];
Ptr_d=[];
%-----%
x_area=strcat(Hd,Tl,num2str(N));
filename='PCB[016-039]_C12H7C13.xlsx';
PCP_List = {'PCB 17'; 'PCB 18'};
Sheet_name=strjoin(strcat(PCP_List(1)));
MS_s=xlsread(filename,Sheet_name,x_area);
K=length(MS_s(:,1));
Frg_s=MS_Splitter(MS_s,Eps);
Sheet_name=strjoin(strcat(PCP_List(2)));
MS_r=xlsread(filename,Sheet_name,x_area);
Frg_r=MS_Splitter(MS_r,Eps);
%-----%
%
%
%
%-----%
M_s=length(Frg_s(1,:));
M_r=length(Frg_r(1,:));
for i=2:M_s
    for j=2:M_r
        [c1,q]=corrcoef(Frg_s(:,i),Frg_r(:,j));
fDst(i,j)=c1(1,2);
    end
end
end

```

```

%-----%
%
%                               Синтез паттернів
%
%-----%

for i=1:K
    Pttr(i,1)=0;
    Pttr(i,2)=0;
end

%-----%

J=1;
for i=2:M_s
    for j=2:M_r
        if(fDst(i,j)>Tr_h)
            Pttr(:,1)=Pttr(:,1)+Frg_s(:,i);
            Pttr(:,2)=Pttr(:,2)+Frg_r(:,j);
            [c1,q]=corrcoef(Pttr(:,1),Pttr(:,2));
            if(c1(1,2)>Tr_h)
                Ptr_d(:,J)=Pttr(:,1); J=J+1;
            end
        end
    end
end

%-----%

xlswrite('Patterns.xlsx',Frg_s(:,1),1, 'A3');
xlswrite('Patterns.xlsx',Ptr_d,1, 'B3');

%-----%

subplot(1,1,1);
hold on
%plot(fDst);
imagesc(fDst);
box on
cb1=colorbar('peer',gca);
set(get(cb1,'Ylabel'),'String','Frechet Distance')

%-----%

subplot(2,1,1)
hold on
plot(MS_s(:,2),'b','linewidth',2)
xlabel('m/Z')
ylabel('Intensyti')
box on
subplot(2,1,2)
hold on
plot(Ptr_d(:,12),'b','linewidth',2)
xlabel('m/Z')
ylabel('Intensyti')
box on
%=====

function frgs = MS_Splitter(MS,Eps)
%-----%

Frg=[];
J=length(MS(:,1));
j=1; cntr=0;

```

```
    if (MS(1,2)>Eps) cntr=1; end;
%-----%
for i=1:J Frg(i,1)=MS(i,1); Frg(i,2)=0; end;
for i=1:J
    if (MS(i,2)>Eps)
        cntr=1;
        Frg(i,j)=MS(i,2);
        Frg(i,j+1)=0;
    else
        if(cntr>0)j=j+1; cntr=0;end;
    end
end
%-----%
frgs=Frg;
```

ДОДАТОК Б

**Конфігурація структури молекули C₁₂H₁₀ за допомогою
декартових координат**

| Row | Tag | Symbol | X | Y | Z |
|-----|-----|--------|------------|------------|------------|
| 1 | 1 | C | -0.6922000 | 0.0000000 | 0.7807000 |
| 2 | 2 | C | -0.6921000 | -0.0002000 | -0.7804000 |
| 3 | 3 | C | 0.7793000 | 0.0000000 | 1.2989000 |
| 4 | 4 | C | 0.7792000 | 0.0000000 | -1.2990000 |
| 5 | 5 | C | -1.5015000 | 1.2881000 | -0.6758000 |
| 6 | 6 | C | -1.5009000 | 1.2884000 | 0.6758000 |
| 7 | 7 | C | -1.5012000 | -1.2882000 | -0.6758000 |
| 8 | 8 | C | -1.5011000 | -1.2884000 | 0.6758000 |
| 9 | 9 | C | 1.4577000 | -1.2263000 | -0.6754000 |
| 10 | 10 | C | 1.4577000 | -1.2263000 | 0.6754000 |
| 11 | 11 | C | 1.4575000 | 1.2263000 | -0.6755000 |
| 12 | 12 | C | 1.4575000 | 1.2265000 | 0.6753000 |
| 13 | 13 | H | 0.8424000 | 0.0001000 | 2.3900000 |
| 14 | 14 | H | 0.8421000 | 0.0000000 | -2.3900000 |
| 15 | 15 | H | -1.9205000 | 1.9562000 | -1.4212000 |
| 16 | 16 | H | -1.9199000 | -1.9564000 | -1.4213000 |
| 17 | 17 | H | -1.9195000 | 1.9569000 | 1.4212000 |
| 18 | 18 | H | -1.9198000 | -1.9568000 | 1.4212000 |
| 19 | 19 | H | 1.8744000 | -2.0249000 | -1.2739000 |
| 20 | 20 | H | 1.8741000 | 2.0250000 | -1.2741000 |
| 21 | 21 | H | 1.8745000 | -2.0249000 | 1.2739000 |
| 22 | 22 | H | 1.8742000 | 2.0252000 | 1.2737000 |

ДОДАТОК Е

Побудова z – матриці по першому графу валентих зв'язків

| № | Елемент | NA | NB | NC | Довжина валентного зв'язку | Валентий кут | Торсіонний кут |
|----|---------|----|----|----|----------------------------|--------------|----------------|
| 1 | C | | | | | | |
| 2 | C | 1 | | | 1.5611000 | | |
| 3 | C | 1 | | | 1.5600780 | | |
| 4 | C | 2 | | | 1.5600223 | | |
| 5 | C | 2 | | | 1.5250529 | | |
| 6 | C | 5 | | | 1.3516002 | | |
| 7 | C | 2 | | | 1.5246403 | | |
| 8 | C | 7 | | | 1.3516000 | | |
| 9 | C | 4 | | | 1.5339657 | | |
| 10 | C | 9 | | | 1.3508000 | | |
| 11 | C | 4 | | | 1.5338366 | | |
| 12 | C | 11 | | | 1.3508000 | | |
| 13 | H | 3 | | | 1.0929231 | | |
| 14 | H | 4 | | | 1.0928117 | | |
| 15 | H | 5 | | | 1.0851450 | | |
| 16 | H | 7 | | | 1.0851595 | | |
| 17 | H | 6 | | | 1.0852370 | | |
| 18 | H | 8 | | | 1.0852140 | | |
| 19 | H | 9 | | | 1.0814819 | | |
| 20 | H | 11 | | | 1.0815726 | | |
| 21 | H | 10 | | | 1.0815204 | | |
| 22 | H | 12 | | | 1.0815004 | | |

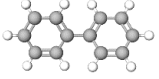
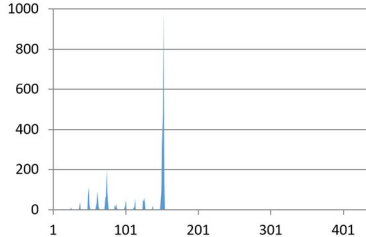
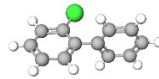
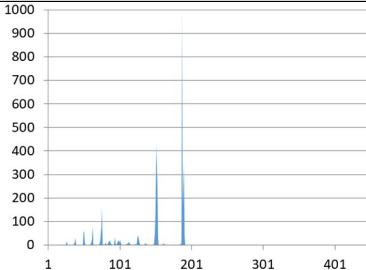
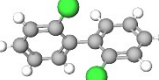
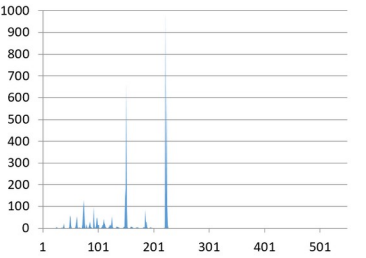
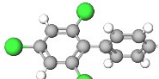
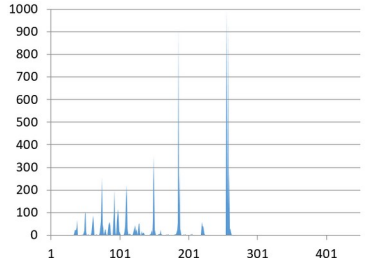
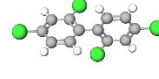
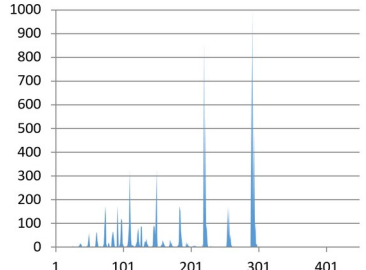
Побудова z – матриці по другому графу валентних кутів

| № | Елемент | NA | NB | NC | Довжина валентного зв'язку | Валентний кут | Торсіонний кут |
|----|---------|----|----|----|-------------------------------|---------------|----------------|
| 1 | C | | | | | | |
| 2 | C | 1 | | | 1.5611000 | | |
| 3 | C | 1 | 2 | | 1.5600780 | 109.3964352 | |
| 4 | C | 2 | 1 | | 1.5600223 | 109.4200713 | |
| 5 | C | 2 | 1 | | 1.5250529 | 86.0589537 | |
| 6 | C | 5 | 2 | | 1.3516002 | 93.9301152 | |
| 7 | C | 2 | 1 | | 1.5246403 | 86.0703191 | |
| 8 | C | 7 | 2 | | 1.3516000 | 93.9388686 | |
| 9 | C | 4 | 2 | | 1.5339657 | 106.3746401 | |
| 10 | C | 9 | 4 | | 1.3508000 | 113.9869139 | |
| 11 | C | 4 | 2 | | 1.5338366 | 106.3822497 | |
| 12 | C | 11 | 4 | | 1.3508000 | 113.9923945 | |
| 13 | H | 3 | 1 | | 1.0929231 | 112.7099224 | |
| 14 | H | 4 | 2 | | 1.0928117 | 112.7160534 | |
| 15 | H | 5 | 2 | | 1.0851450 | 132.6808808 | |
| 16 | H | 7 | 2 | | 1.0851595 | 132.6732535 | |
| 17 | H | 6 | 5 | | 1.0852370 | 133.3789070 | |
| 18 | H | 8 | 7 | | 1.0852140 | 133.3877161 | |
| 19 | H | 9 | 4 | | 1.0814819 | 122.3965785 | |
| 20 | H | 11 | 4 | | 1.0815726 | 122.3953212 | |
| 21 | H | 10 | 9 | | 1.0815204 | 123.5997285 | |
| 22 | H | 12 | 11 | | 1.0815004 | 123.6015940 | |

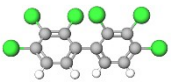
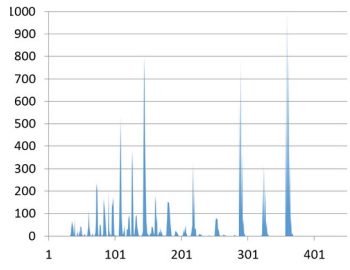
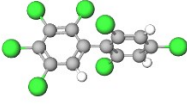
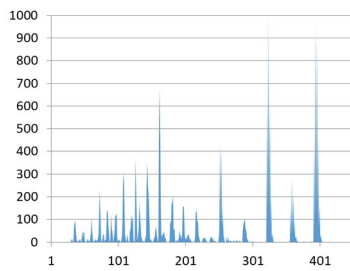
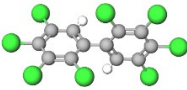
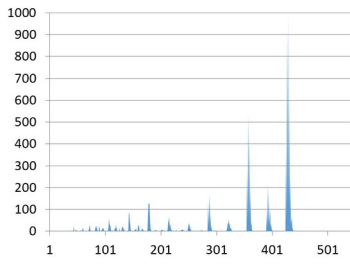
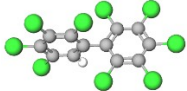
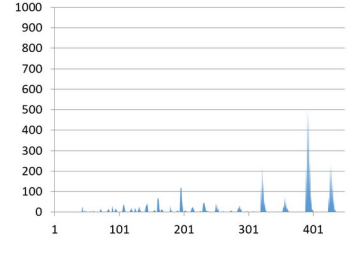
Побудова z – матриці по третьому графу торсіонних кутів

| № | Елемент | NA | NB | NC | Довжина валентного зв'язку | Валентний кут | Торсіонний кут |
|----|---------|----|----|----|-------------------------------|---------------|----------------|
| 1 | C | | | | | | |
| 2 | C | 1 | | | 1.5611000 | | |
| 3 | C | 1 | 2 | | 1.5600780 | 109.3964352 | |
| 4 | C | 2 | 1 | 3 | 1.5600223 | 109.4200713 | -0.0129610 |
| 5 | C | 2 | 1 | 3 | 1.5250529 | 86.0589537 | -122.1425459 |
| 6 | C | 5 | 2 | 1 | 1.3516002 | 93.9301152 | 0.0275699 |
| 7 | C | 2 | 1 | 6 | 1.5246403 | 86.0703191 | -115.7486101 |
| 8 | C | 7 | 2 | 1 | 1.3516000 | 93.9388686 | -0.0151477 |
| 9 | C | 4 | 2 | 1 | 1.5339657 | 106.3746401 | 56.4460337 |
| 10 | C | 9 | 4 | 2 | 1.3508000 | 113.9869139 | -59.3375106 |
| 11 | C | 4 | 2 | 1 | 1.5338366 | 106.3822497 | -56.4281174 |
| 12 | C | 11 | 4 | 2 | 1.3508000 | 113.9923945 | 59.3289303 |
| 13 | H | 3 | 1 | 6 | 1.0929231 | 112.7099224 | 82.4919109 |
| 14 | H | 4 | 2 | 1 | 1.0928117 | 112.7160534 | -179.9927022 |
| 15 | H | 5 | 2 | 1 | 1.0851450 | 132.6808808 | 179.9537673 |
| 16 | H | 7 | 2 | 1 | 1.0851595 | 132.6732535 | -179.9430544 |
| 17 | H | 6 | 5 | 2 | 1.0852370 | 133.3789070 | 179.9387908 |
| 18 | H | 8 | 7 | 2 | 1.0852140 | 133.3877161 | -179.9346487 |
| 19 | H | 9 | 4 | 2 | 1.0814819 | 122.3965785 | 119.2811654 |
| 20 | H | 11 | 4 | 2 | 1.0815726 | 122.3953212 | -119.2839144 |
| 21 | H | 10 | 9 | 4 | 1.0815204 | 123.5997285 | 178.6053235 |
| 22 | H | 12 | 11 | 4 | 1.0815004 | 123.6015940 | -178.6046770 |

Характерні мас-спектри для точної кількісної ідентифікації поліхлор-біфенілів.

| № | Речовина | Формула | Просторова Структура | CAS # | Мас-спектр |
|----|--------------------------------------|-----------------|---|------------|---|
| 1. | Biphenyl | $C_{12}H_{10}$ |  | 92-52-4 |  |
| 2. | 1,1'-Biphenyl,2-chloro- | $C_{12}H_9Cl$ |  | 2051-60-7 |  |
| 3. | 1,1'-Biphenyl, 2,2'-dichloro | $C_{12}H_8Cl_2$ |  | 13029-08-8 |  |
| 4. | 1,1'-Biphenyl,2,2',6trichloro- | $C_{12}H_7Cl_3$ |  | 35693-92-6 |  |
| 5. | 1,1'-Biphenyl, 2,2',4,4'tetra chloro | $C_{12}H_6Cl_4$ |  | 2437-79-8 |  |

**Характерні мас-спектри для точної кількісної
ідентифікації поліхлор-біфенілів.**

| № | Речовина | Формула | Просторова Структура | CAS # | Мас-спектр |
|-----|---|----------------------|---|----------------|---|
| 7. | 1,1'- Biphenyl, 2,2,3,3',4,4'h exachloro- | $C_{12}H_4Cl_6$ |  | 38380- 07-3 |  |
| 8. | 1,1'- Biphenyl, 2,2,3,4,4',5,6 Heptachloro - | $C_{12}H_3Cl_7$ |  | 60145- 23-5 |  |
| 9. | 1,1'- Biphenyl, 2,2',3,3',4,4', 5,5'- octachloro- | $C_{12}H_2Cl_8$ 8 |  | 35694- 08-7 |  |
| 10. | 1,1'- Biphenyl, 2,2',3,3',4,4', 5,5',6- nonachloro- | $C_{12}HCl_9$ |  | 40186- 72-9 |  |

