

КИЇВСЬКИЙ НАЦІОНАЛЬНИЙ УНІВЕРСИТЕТ

ІМЕНІ ТАРАСА ШЕВЧЕНКА

ІНСТИТУТ ВИСОКИХ ТЕХНОЛОГІЙ

Завідувач кафедри супрамолекулярної хімії
д.х.н., професор Ігор Володимирович Комаров
Протокол № ____ засідання кафедри
Від «__» _____ 2021 року

**РОЗРОБКА ЛІНІЙНОГО ПРЕДСТАВЛЕННЯ
КООРДИНАЦІЙНИХ СПОЛУК**

Кваліфікаційна робота магістра
студента 2 року навчання спеціальності
102 «Хімія» (хемоінформатика)»
Березовського Олександра Натанієльовича

Науковий керівник: доктор хімічних
наук, професор **Волочнюк Д. М.**
Куратор: **Кротко Д. Г.**

Оцінка захисту роботи

Київ-2021

Анотація

Тема даної дипломної роботи – «Розробка лінійного представлення координаційних сполук» – у рамках якої було розглянуто лінійне представлення SMILES та стратегії побудови розширень до нього. Також проаналізовано труднощі які виникають при спробі лінійного представлення координаційних сполук та методи їх подолання які були запропоновані на момент написання цієї роботи. Запропонований новий метод запису, який зможе вирішити недосконалості попередніх спроб.

Метою цієї роботи була розробка лінійного представлення та методу структурного пошуку, які б вирішили такі проблеми запису та пошуку координаційних сполук такі як: гіпервалентність, неунікальність запису делокалізованих структур та відсутність стандарту задання структур координаційних сполук.

У практичній частині роботи описані структура та правила утворення нового розробленого мною лінійного представлення, правила стандартного запису структур координаційних сполук та сполук з багатоцентровими зв'язками, також описано алгоритм структурного пошуку, що у поєднанні вирішує наведені вище труднощі лінійного запису координаційних сполук.

Зміст

Вступ.....	4
Літературний огляд.....	6
SMILES	6
Стратегії побудови розширень SMILES	6
Приклади розширень SMILES.....	7
Проблеми представлення координаційних сполук.....	8
Практична частина.....	11
Загальна структура та метод утворення Mu_SMILES (моє лінійне представлення).....	11
Структура	11
Метод утворення	11
Об'єднання атомів у систему та системний блок	11
Блок координаційних зв'язків.....	14
«Ковалентні» багатоцентрові зв'язки	14
Еталонізація структур.....	15
Координаційні системи	15
Ковалентні багатоцентрові системи.....	17
Концепція структурного пошуку.....	18
Візуалізація	18
Правила пошуку	19
Приклади Mu_SMILES для різних структур	19
Висновки.....	23
Використана література та посилання.....	24

Вступ

Виокремлення хемоінформатики як науки пов'язано з накопиченням великої кількості хімічної інформації, та необхідністю нею оперувати: зберігати, робити пошук, та обробляти.

Спершу це робилось у паперових хімічних каталогах, але з накопиченням дуже великих кількостей цієї інформації, паперовий варіант ставав ще більш та більш незручним, особливо якщо потрібно було робити пошук за структурою, хоча це як говориться, пів біди у порівнянні з підструктурним пошуком.

Інформація згодом все більше стала переходити у електронний варіант та пошук у паперових каталогах ставав історією, та засобом знущання над студентами.

Інформація про структуру є дуже важливою для речовини, та потрібно було її якось заносити у комп'ютер. Настала потреба розробити комп'ютерні представлення для структурної інформації, бо таке класичне хімічне представлення структури як структурна формула, виявилось дуже незручним для комп'ютерного зберігання, а особливо для пошуку та обробки.

Розроблялися різні способи представлення. Дуже добре показала себе концепція лінійного представлення, тобто запис структури у рядок. Можна провести аналогію з номенклатурними назвами речовин, але у більш зручному для комп'ютера вигляді.

Одним з перших найбільш вдалих були представлення Вісвессера (WLN) [1], які були розроблені для ручного кодування хімічних структур та зберігання на перфокартах. Саме на їх основі були розроблені перші методи пошуку структур, та створені перші комп'ютерні бази даних хімічних структур у 1960-х роках.[2-4]

Але з'явилась проблема з автоматичним переведенням цих структур у машинно-читабельний код, та складним алгоритмом канонізації, що і повело за собою пошук більш зручного представлення.

Таким представленням стали графи. Для них були розроблені дуже зручні алгоритми пошуку за структурою і підструктурою, та інші варіанти пошуку, що складно переоцінити. Графове представлення є актуальним і зараз, хоча воно не є найбільш ємним, і бажання представляти структури у лінійній формі залишилось.

У 1986 році Девідом Вейнінгером була випущена стаття з правилами нового лінійного представлення SMILES, яке було їм розроблено.[5, 7] Та згодом вийшла стаття з правилами канонізації.[6]

Це представлення було настільки простим та зручним, що швидко набрало популярність. Зараз хімічний софт, який не розуміє SMILES, виглядає якось неповноцінно.

Згодом на основі SMILES були розроблені ще два дуже потужних інструмента: SMARTS (лінійне представлення паттернів, для запитів пошуку за підструктурою) та SMIRKS (лінійне представлення хімічних перетворень).

А також на основі SMILES були розроблені представлення для кодування полімерних систем CHUCKLES[8] та SHORTLES[9].

Також варте згадки представлення InChI, яке є міжнародним хімічним ідентифікатором IUPAC, розробленим на початку 2000-х років.[10] Для нього була розроблена дуже корисна можливість. На основі InChI за допомогою хеш-функцій був розроблений InChI-Key, ідентифікатор, який містить фіксоване число символів.

Представлення SMILES було дуже зручним, але й воно не виявилось ідеальним, та не змогло задовольнити усі потреби користувачів. Це спричинило розробку різних розширень SMILES направлених на вирішення таких недосконалостей.

Метою даної роботи було розробка такого розширення для деяких проблемних класів сполук.

Ключові слова: SMILES, InChI-Key, координаційні сполуки.

Літературний огляд

1. SMILES

Simplified Molecular Input Line Entry Specification

Перед тим як говорити про недосконалість, трохи розглянемо принципи за якими будується SMILES.

- Обирається початок (атом з якого починають), та записують послідовно символи атомів «проходяться» молекулою, при чому неорганогенні елементи позначаються символами хімічних елементів у квадратних дужках з укаванням числа воднів з якими вони зв'язані. Для органогенних елементів дужки можна опустити та не вказувати число атомів водню, якщо їх кількість відповідає стандартній валентності.

До органогенних входять такі елементи: C, O, N, S, P, B, F, Cl, Br, I.

- Також вказуються зв'язки між атомами. У SMILES є такі типи зв'язків як простий("-"), подвійний("="), потрійний("#"), ароматичний(":") та відсутність зв'язку(".") (якщо два послідовних атома не зв'язані ковалентним зв'язком), а також є орієнтовані зв'язки ("/", "\"), що несуть інформацію про орієнтацію навколо подвійного зв'язку.
- Розгалуження вказуються круглими дужками (як у напівструктурних формулах).
- Цикли вказуються числами біля атомів, які позначають що було «розірвано» зв'язок, щоб записати структуру.
- Заряди чи радикали позначаються у квадратних дужках після атому з укаванням воднів зв'язаних з ним.
- Ароматичні атоми позначаються маленькими літерами, та зв'язок ароматичний може не вказуватись, а от простий зв'язок між ароматичними атомами вказувати потрібно у явному вигляді.

2. Стратегії побудови розширень SMILES

- 1) Втручання у «тіло» SMILES (введення додаткових позначок, чи щось подібне)

Перевагою такого розширення є можливість більш "тонко" задавати структурні особливості хімічних сполук. Але тоді потрібен новий

алгоритм канонізації та нові інструменти візуалізації, що можна віднести до деяких незручностей такого способу.

- 2) Використання допустимих позначок у синтаксисі SMILES для кодування додаткових речей

При такому способі розширення не потрібен новий алгоритм канонізації та є можливість описувати структури не руйнуючи синтаксис SMILES. Це можна віднести до переваг, проте наші дії обмежені синтаксисом SMILES та канонізацією.

- 3) Запис додаткової інформації «у коментарі» (після пробілу)

Інтерпретатори SMILES, все що записано після пробілу вважають коментарем. Якщо грамотно підійти до розробки, то не потрібен новий алгоритм канонізації. Якщо візуалізувати інтерпретатором для класичних SMILES, то ми побачимо хоч щось, та чим ближче воно до справжньої структури, тим краще.

3. Приклади розширень SMILES

Згодом можливостей SMILES стало недостатньо, документація практично не обновлювалась, та стали створюватись різні розширення для SMILES, які могли б описати такі системи, при описі яких можливостей SMILES бракувало.

Одним з таких проектів був OpenSMILES який підтримувався деякий час, та в який вносились зміни, наприклад у кодування різних типів хіральності.**[11]** Зараз він перетворився у IUPAC SMILES +, представлення яке підтримується IUPAC.**[12]**

Також було розроблено ще одне розширення для OpenSMILES – Jmol SMILES.**[13]** В ньому були внесені деякі корисні зміни, наприклад:

- З'явилась можливість позначати стереохімію атропо-ізомерів.
- Був доданий фіктивний атом.
- Стала можливою безлімітна кількість одночасних «відкривань» циклів (у класичних SMILES не можливо «розірвати» більше 100 зв'язків одночасно).

- З'явилась можливість кекулізації зі збереженням позначок ароматичних атомів.

Також на його основі розроблен Jmol SMARTS, який є розширенням до OpenSMARTS.

Одним із самих потужних з точки зору кількості інформації, яку можливо вказати є CXSMILES розроблений у ChemAxon. [14]

Він складається з 2 частин, у першій пишеться класичний SMILES, а потім через пробіл у коментарі записуються різні додаткові характеристики, які не можливо записати у SMILES. За рахунок того, що вся додаткова інформація у коментарі, інтерпретатори SMILES не мають проблем з візуалізацією. Додаткова інформація йде блоками, та викликається за потреби.

Але нажаль в них нема такого поняття як канонізація.

4. Проблеми представлення координаційних сполук
 - 1) Труднощі у записі координаційних сполук у SMILES починаються з того, що у SMILES нема координаційного зв'язку.
 - 2) Другою проблемою стали багатоцентрові зв'язки, які зустрічаються досить часто у координаційних сполуках.
 - 3) Ще одним підступним моментом виявилась делокалізація.
 - 4) Та найважчим у подоланні є відсутність правил зображення (рисування) структур координаційних сполук.

Є декілька різних підходів до цієї задачі:

- А) Зв'яжемо всі центри простими зв'язками

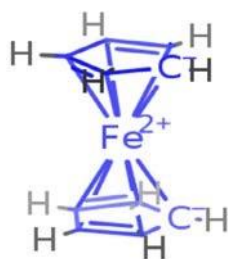


Рис. 1 Представлення фероцену у Reaxys [15]

Такий підхід використовується у Reaxys, але у цієї моделі дуже великі проблеми з валентністю. (рис. 1)

Хоча це не заважає у Reaxys робити структурний пошук, але якщо спробувати пошукати фероцен за структурою без вказання усіх зв'язків із залізом, то він не знаходиться.

Б) Вирішити проблеми з валентністю у координаційних сполуках спробували за рахунок введення додатково дативного типу зв'язку.

При цьому насичення валентності врахувалось тільки з одної сторони зв'язку. (рис. 2) Таке рішення реалізовано у RDKit. [16]

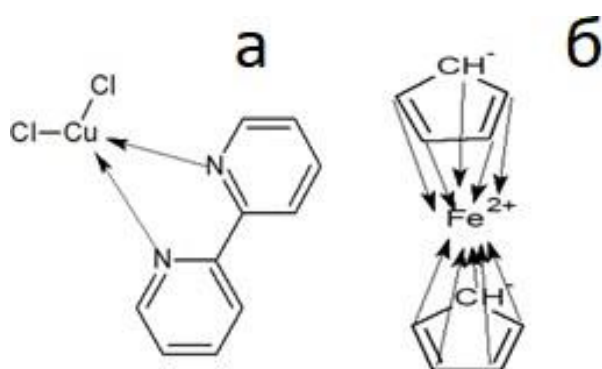


Рис. 2 Запис структур з дативним зв'язком

Якщо розглянемо на структуру 2а, то можна побачити, що проблема з гіпервалентністю для азотів вирішена, так як насичення валентності йде тільки на «гострому» кінці зв'язка.

А от для структури 2б валентність заліза все ще дорівнює 10.

В) Ще однією спробою вирішення даної проблеми було введення деякого суб-атома як посередника зв'язку. (рис. 3)

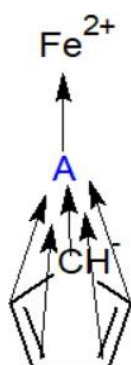


Рис. 3 Ілюстрація концепції суб-атомів

При цьому виходить що залізо вступає в утворення одного зв'язку з усіма атомами карбону у циклопентадиєніл-аніоні одночасно.

Така спроба використана у CXSMILES для опису багатоцентрових зв'язків.

Хоча підхід «зв'язання всіх центрів простими зв'язками» не заважає робити структурний пошук фероцену (хоч рисувати ці зв'язки не дуже зручно та структура потім «засмічена» цими зв'язками, та на неї незручно дивитись), але все ж пошук деяких координаційних систем у Reaxys викликає труднощі. (рис. 4)

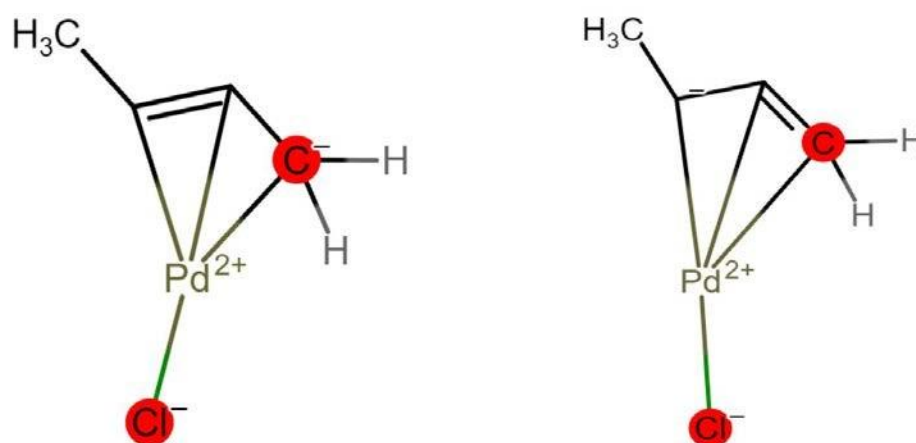


Рис. 4 Приклад проблемних координаційних систем для пошуку у Reaxys

Обидва запити шукають одні й ті самі сполуки, та є резонансними структурами один для одного, але Reaxys зрозуміє ці два запити по різному і знайде різні результати. Таким чином, шукаючи тільки один з запитів, не отримуємо повну відповідь.

Практична частина

Метою цієї роботи було розробити таке лінійне представлення, яке могло б записувати координаційні сполуки найбільш правильно, та за яким було би зручно проводити пошук.

1. Загальна структура та метод утворення `My_SMILES` (моє лінійне представлення)

1) Структура

`SMILES_"system"_1"coord"_2"multicenter"`

Спочатку пишеться деякий класичний SMILES («класична частина»), а потім через пробіл (щоб була можливість візуалізувати звичайним інтерпретатором SMILES) пишеться друга частина («некласична частина»), яка може складатися з різних блоків, таких як: системний блок, координаційний блок та блок багаточентрових (ковалентних) зв'язків .

2) Метод утворення

Спочатку беруться класичні SMILES, які найбільш відповідають системі (Основоположні), та переводяться у такий класичний SMILES на котрому будуть робитись надбудови (фундамент, який буде відрізнятися від основоположних у сполуках з неароматичною делокалізацією тим, що делокалізовані зв'язки будуть простими). Він і буде являтися класичною частиною `My_SMILES`.

2. Об'єднання атомів у систему та системний блок

Об'єднання атомів у систему грає роль суб-атому (рис. 3), який є посередником зв'язку між елементами (атомами чи системами), що входять до системи, та елементом з яким пов'язана система. Також система отримує номер після атомів та систем які були до неї.

Системи можна поділити на такі типи:

1) Не зв'язані

2) Зв'язані

3) Делокалізовані (не ароматичні)

4) Ароматичні

Реалізація 1, 2 та 4 записуються однаково за рахунок того, що не потрібно вказувати делокалізацію (у ароматичних системах делокалізація вказується у класичному SMILES тим, що атоми записуються маленькими літерами).

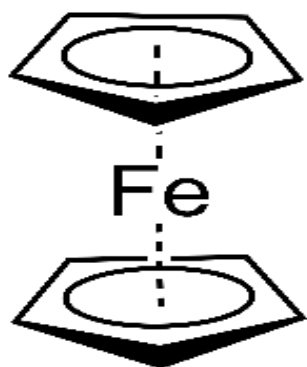


Рис. 5 Фероцен

Розберемо на прикладі фероцену(рис. 5). Його системний блок буде виглядати так:

`_[2,3,4,5,6][7,8,9,10,11]`

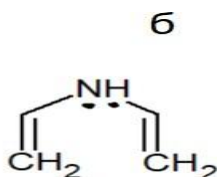
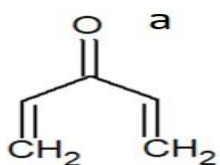
У квадратних дужках записані номери атомів (порядок їх розташування у SMILES) які входять у систему, а це в даному випадку атоми одного циклопентадієнілу у перших дужках та іншого у других.

Тепер розберемо більш складний випадок, коли потрібно вказати делокалізацію у системному блоці.

Тоді описання кожної системи може складатися з 4 підблоків розділених нулями:

`[atom,0,cross,0,charge,0,double]`

- 1 підблок – це номери атомів у порядку зв'язності у системі, та якщо є розгалуження (не плутати з кросс-точками (дивимось далі)), тоді позначати їх дужками.
- Перший 0 – означає, що в системі є делокалізація, та він може бути присутній навіть коли нема підблоків далі.
- 2 підблок – номери кросс-точок (фрагментів які вступають у кросс-спряження(рис. 6а) з системою чи р,л-спряження за рахунок неподіленої електронної пари (рис. 6б)).



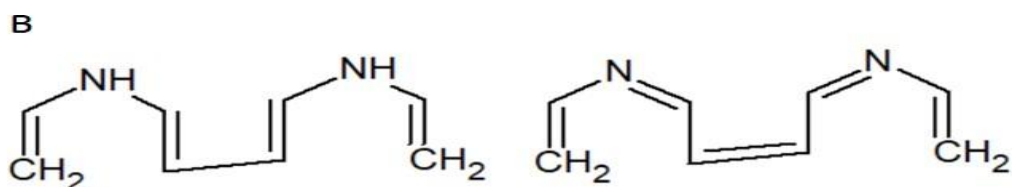


Рис. 6 Приклади кросс-точок (а, б) та «вироджених» систем (в)

Важливо вказувати кросс-точки, тому що інакше може бути не зрозуміло про яку з систем іде мова, як на рис. 6в.

- 3 підблок описує делокалізацію заряду чи радикалу в системі. Перше число позначає заряд чи радикал (записується у дробовій частині числа після 0, наприклад 0.1 чи 0.2) делокалізований у системі, а далі номери атомів на яких він фокусується (у резонансних структурах). Якщо делокалізація заряду чи радикалу відбувається по усій системі, то не потрібно писати номери атомів.
- 4 підблок описує орієнтацію замісників навколо подвійного зв'язку, який ми зробили простим коли робили фундамент. Перше число 1 чи -1. Номери всіх атомів з якими є орієнтований зв'язок у SMILES ("/") пишуться після 1, а тих з якими ("\") – після -1, та група після одиниці пишеться раніше. Таким чином ми уникаємо неточностей, тому що 1 – це може бути і номер атома, а -1 не може.
- Якщо наступні підблоки відсутні, то ми їх упускаємо. Проте якщо наступні підблоки наявні, то треба записати підблок, навіть якщо він порожній.

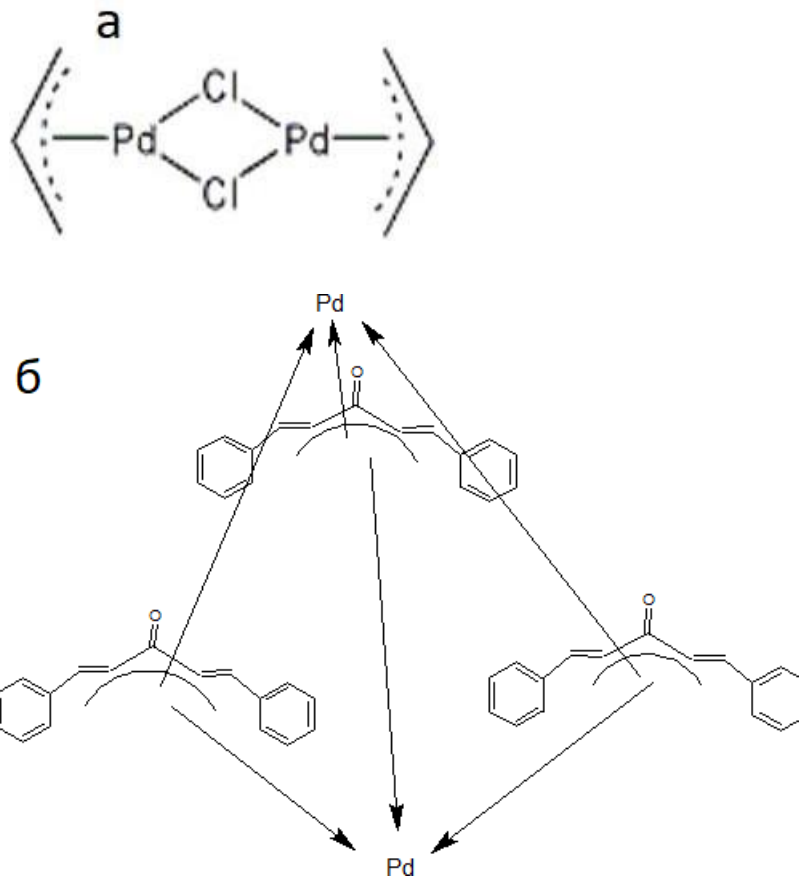


Рис. 7 Делокалізовані координаційні системи (у 7б не прибрані подвійні зв'язки навмисно, для кращої візуалізації)

Приклад системи з рис. 7а:

[1,2,3,0,0,-1,1,3]

Приклад системи з рис. 7б:

[4,3,2,11,12,0,2,0,0,1,3,5,11,13]

3. Блок координаційних зв'язків

Блок координаційних зв'язків починається з _1 та далі у квадратних дужках пишуться номери елементів які утворюють координаційний зв'язок, для кожного зв'язку.

Приклад:

_1[7,9][7,10]

7 елемент утворює координаційні зв'язки з 9 та з 10

4. «Ковалентні» багатоцентрові зв'язки

За аналогією до координаційних багатоцентрових зв'язків. Утворюється система, яка записується у системному блоці, та у блоці

багатоцентрових зв'язків записуються номери елементів які приймають участь у зв'язку. Цей блок починається з $_2$.

Приклад:

$_2[5,6]$

Зв'язок між елементами 5 та 6, хоча б один з яких є системою.

Якщо потрібно вказати багатоцентрові зв'язки не «прибираючи» їх з класичного SMILES, тоді перед описанням зв'язку ставимо 0:

$_2[0,5,6]$

5. Еталонізація структур

I. Координаційні системи

• Проблема

У представленні координаційних сполук велика перешкода – це відсутність чітких правил запису структур, що сильно ускладнює їх лінійне кодування, тому мною були розроблені такі правила.

Щоб було більш зрозуміло розглянемо мостикову паладієво-хлорну систему (рис. 8).

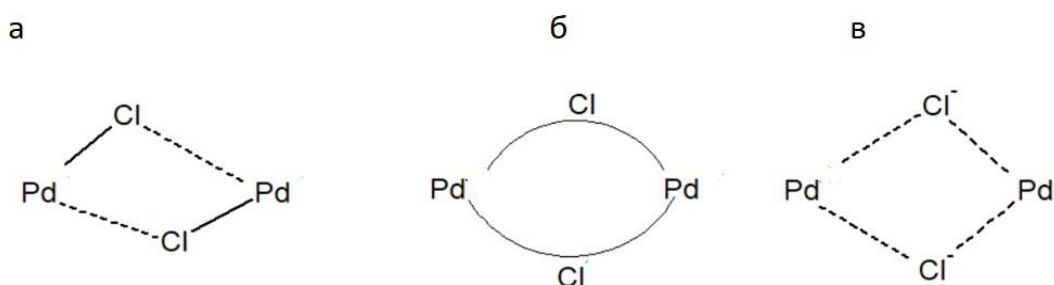


Рис. 8 Паладієво-хлорна система аналогічна хлориду паладію(II) (інші зв'язки атомів паладію упушені для кращого сприйняття)

У таких системах можливі різні форми запису, такі як 8а (один зв'язок ковалентний, а другий координаційний), 8б (ковалентний багатоцентровий зв'язок) та 8в (два координаційних зв'язки з хлорид-аніоном), та у різних речовин більш доречним може бути будь-який з цих видів запису, але навряд чи є багато речовин які відрізняються один від одного тільки вибором однієї з таких «форм».

Хоча в конкретних випадках може бути більш доречною одна з цих форм, а в інших – інша, сумнівно що той хто шукає буде це

знати, та захоче знайти сполуки які відносяться до цих форм разом, не шукаючи кожна окремо.

Потрібно вибрати одну еталонну форму, тоді канонічний Mu_SMILES еталонної структури буде унікальним для речовини.

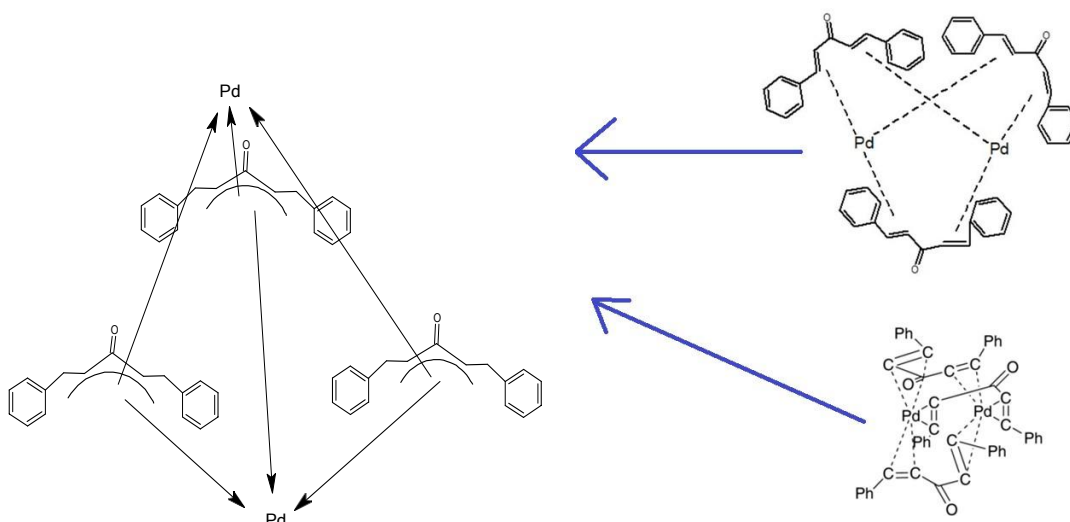
- **Правила еталонізації**

- 1) Мостикові ліганди, які зазвичай є однодентатними, записувати з «звичним» для них зарядом, та координаційними зв'язками.
- 2) Водневі зв'язки можна не зображати (хоча з цим ще не остаточно вирішено).
- 3) Спряження та делокалізацію не потрібно зображати, якщо нема координаційних зв'язків з системою та якщо делокалізація не приводить до зміни структури чи зміни еквівалентності атомів у структурі.
- 4) Якщо атом чи система пов'язується координаційним багатоцентровим зв'язком з частиною спряженої системи, тоді еталонна структура буде записуватися зі зв'язком цього елемента зі всією системою, якщо вона не більше 10 атомів, а якщо більше, то тільки з 10 найближчими, а якщо система циклічна, то це обмеження росте до 20.

При усьому цьому не «заходячи» у кільця і не виходячи з них.

Також, якщо елемент пов'язаний зі всіма елементами делокалізованої системи координаційними зв'язками – то ці зв'язки перетворюються в один багатоцентровий.

Приклади на рис. 9



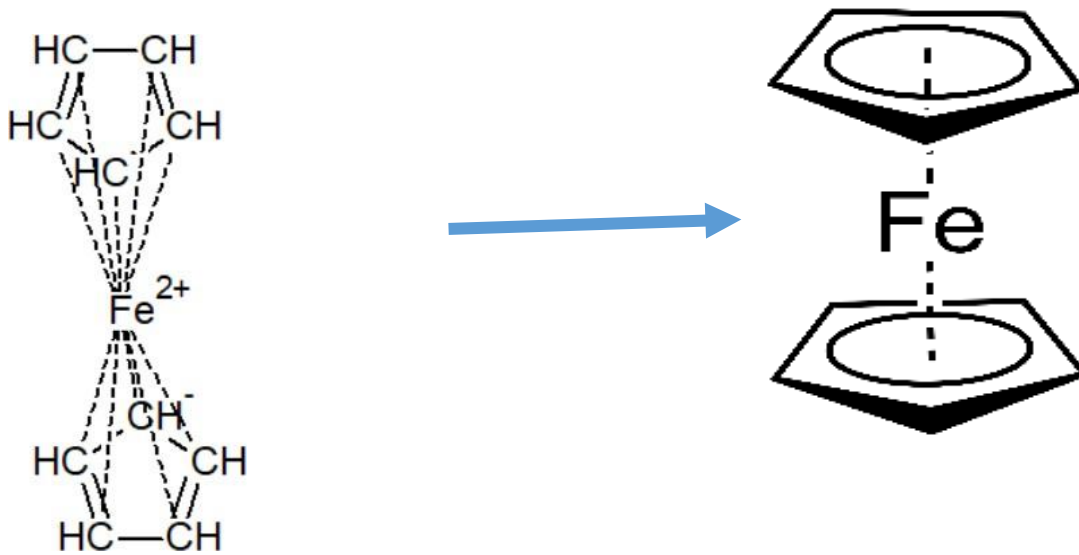


Рис. 9 Приклади еталонізації

II. Ковалентні багатоцентрові системи

1) 2-електронні 3-центрові системи

При описі таких систем, наприклад диборану (рис.10), виникає деяка невизначеність їх запису пов'язана з наявністю альтернативи вибору системи – виродженністю систем.

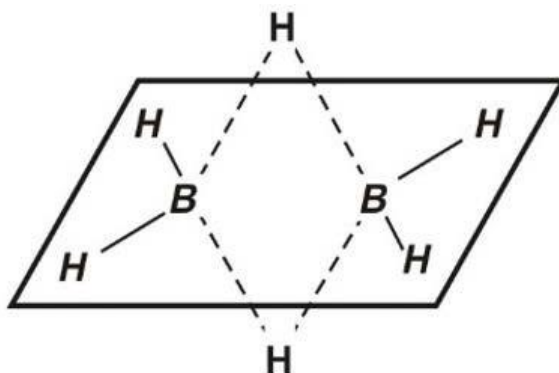
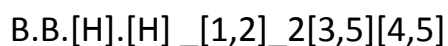
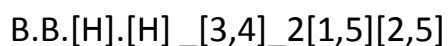


Рис. 10 Диборан

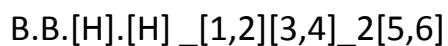
Наприклад можемо «об'єднати» у систему 2 Бори:



А можемо водні:



Така альтернатива з'являється коли усі елементи однієї системи пов'язані з усіма елементами другої. Тоді еталонною має стати структура з багатоцентровим зв'язком між першою системою та другою:



2) «Радикально-димерні» системи

Ще одним розповсюдженим видом хімічних систем в яких зустрічається багатоцентровий зв'язок є димери компонентів з відкритою електронною оболонкою (радикалів) із π -спряженням з радикальним центром (рис.11). [17-30]

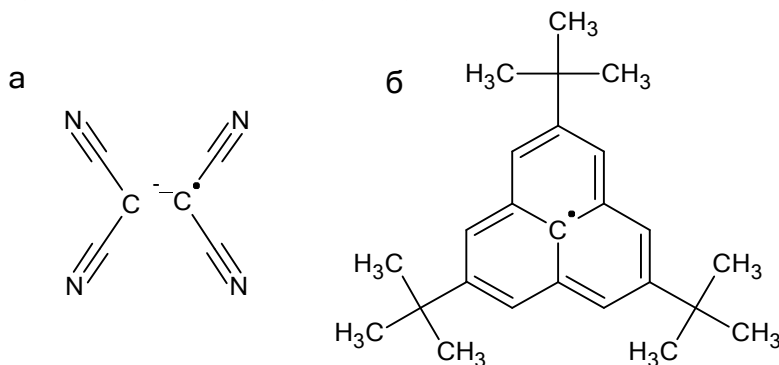


Рис. 11 Радикально-спряженні системи

а – Тетраціаноетилен-аніон-радикал [TCNE] [17-22]

б – Тритретбутилфеналеніл-радикал [23]

У еталонних структурах зв'язки будуть утворені між двома делокалізованими системами.

6. Концепція структурного пошуку

1) Візуалізація

Якщо канонічні Mu_SMILES візуалізувати софтом призначеним для звичайних SMILES, тоді структура може вводити в оману. Наприклад алліл-аніон буде виглядати як просто пропан (рис. 12).

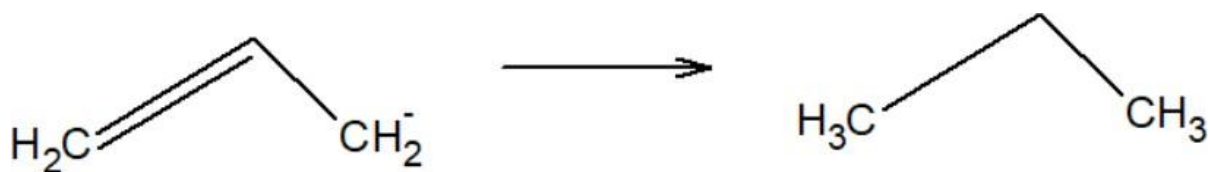


Рис. 12 Неправильне відображення канонічних Mu_SMILES візуалізаторами для класичних SMILES

Щоб уникнути цього:

- У базах буде зберігатись 2 колонки: одна з канонічними Mu_SMILES , а друга з експортним не канонічним варіантом, «класична» частина якого найбільш відповідає структурі (основоположні),

при цьому дані про делокалізацію зникають, але системи та атоми в них помічені як «делокалізовані».

- За канонічними буде проводитись пошук, а експортні будуть вивантажуватись, хоча залишиться можливість вивантажити й канонічні.

При цьому системний блок скоротиться.

Наприклад:

Канонічний

[4,3,2,11,12,0,2,0,0,1,3,5,11,13]

Експортний:

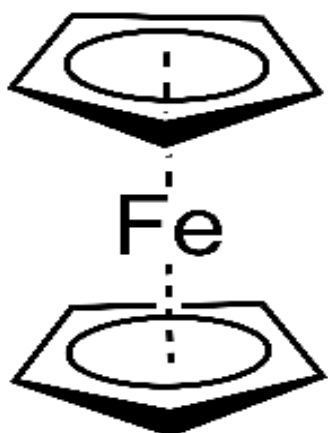
[0,4,3,2,11,12]

При цьому 0 на першому місці у експортному варіанті, говорить що в системі є делокалізація, яка не «розмазана».

2) Правила пошуку

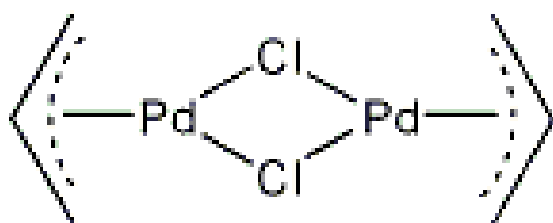
- Результатом структурного пошуку є структури з точним співпадінням класичної частини та наявністю вказаної надбудови.
- Якщо «забити» у пошук структуру без надбудови, то результатом буде структура з довільною надбудовою, чи без неї. А якщо «забити» структуру з 1 координаційним зв'язком, то у результаті він повинен бути, а інша надбудова – довільна.
- Це звільнить у більшості випадків від побудови надбудови взагалі, при заданні структури для пошуку.

7. Приклади Mu_SMILES для різних структур



Фероцен

[Fe+2].c1cc[cH-]c1.c1cc[cH-]c1_[2,3,4,5,6][7,8,9,10,11]_1[1,12][1,13]

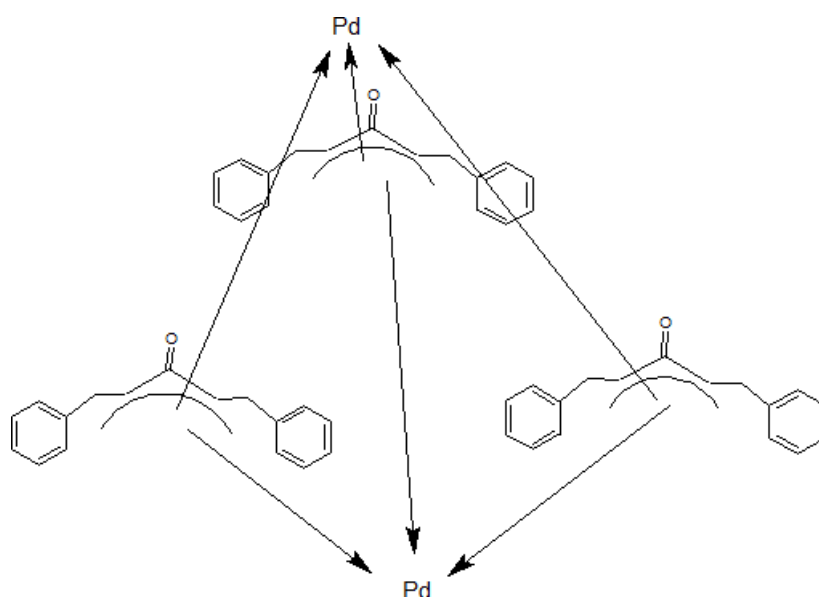


Канонічний:

CCC.CCC.[Cl-].[Cl-].[Pd+2].[Pd+2] _[1,2,3,0,0,-1,1,3][4,5,6,0,0,-1,4,6]
_1[7,9][7,10][8,9][8,10][9,11][10,12]

Експортний:

[CH2-]C=C. [CH2-]C=C.[Cl-].[Cl-].[Pd+2].[Pd+2] _[0,1,2,3][0,4,5,6]
_1[7,9][7,10][8,9][8,10][9,11][10,12]

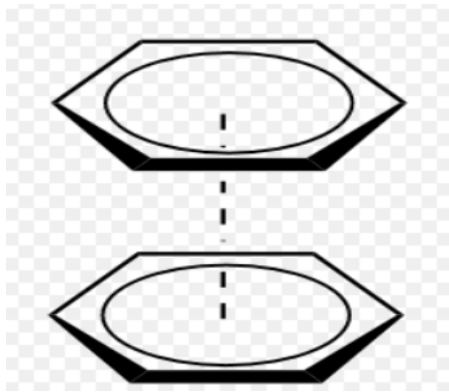


Канонічний:

O=C(CCc1ccccc1)C=C/c1ccccc1.O=C(CCc1ccccc1)C=C/c1ccccc1.O=C(CCc1ccccc1)
CCc1ccccc1.[Pd].[Pd]
_[4,3,2,11,12,0,2,0,0,1,3,5,11,13][22,21,20,29,30,0,20,0,0,21,23,29,31][40,3
9,38,47,48,0,38,0,0,39,41,47,49]_1[55,57][55,58][55,59][56,57][56,58][56,5
9]

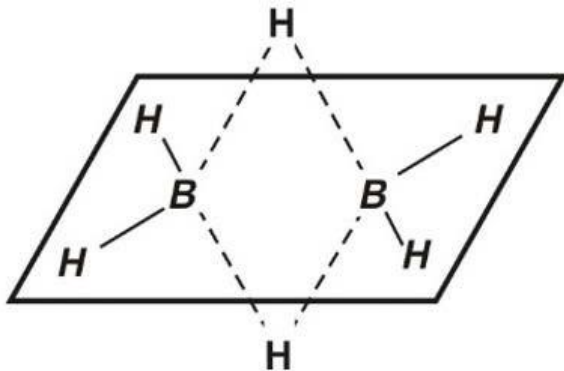
Експортний:

O=C(/C=C/c1ccccc1)/C=C/c1ccccc1.O=C(/C=C/c1ccccc1)/C=C/c1ccccc1.O=C(
/C=C/c1ccccc1)/C=C/c1ccccc1.[Pd].[Pd]
_[0,4,3,2,11,12][0,22,21,20,29,30][0,40,39,38,47,48]_1[55,57][55,58][55,59
][56,57][56,58][56,59]



Димер бензолу

c1ccccc1.c1ccccc1 _[1,2,3,4,5,6]_[7,8,9,10,11,12]₁_[13,14]



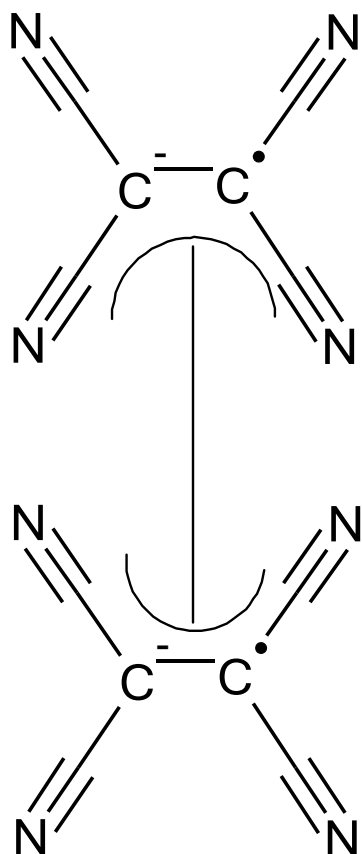
Диборан

Канонічний:

B.B _[H]_[H] _[1,2]_[3,4]₂_[5,6]

Експортний:

[BH2][H].[BH2][H] _[1,3]_[2,4]₂_[0,5,6]



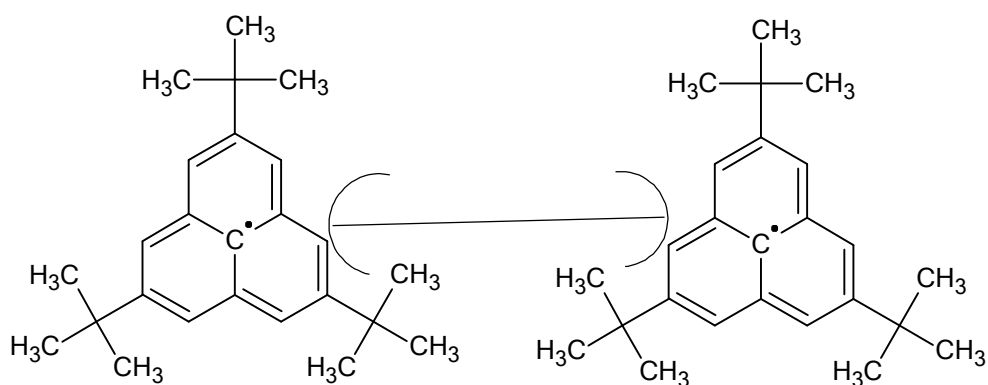
Димер тетраціаноетилен-аніон-радикалу [TCNE]²⁻

Канонічний:

N=CC(C=N)C(C=N)C=N. N=CC(C=N)C(C=N)C=N _{[1,2,3,(4,5),6,(7,8),9,10,0,0,-1,3,6,0.1,6,3][11,12,13,(14,15),16,(17,18),19,20,0,0,-1,13,16,0.1,16,13]} _{2[21,22]}

Експортний:

N#[C][C](C#N)[C-](C#N)C#N.N#[C][C](C#N)[C-](C#N)C#N _{[0,1,2,3,(4,5),6,(7,8),9,10][0,11,12,13,(14,15),16,(17,18),19,20]} _{2[21,22]}



Димер тритретбутилфеналеніл-радикалу

Канонічний:

CC(C)(C)C1=CC2=CC(C(C)(C)C)=CC3=CC(C(C)(C)C)=CC(=C1)C23.CC(C)(C)C1=CC2=CC(C(C)(C)C)=CC3=CC(C(C)(C)C)=CC(=C1)C23
_[5,6,7,8,9,14,15,16,17,22,23,24,25,0,0,0.1][30,31,32,33,34,39,40,41,42,47,48,49,50,0,0,0.1]_2[51,52]

Експортний:

CC(C)(C)C1=CC2=CC(C(C)(C)C)=CC3=CC(C(C)(C)C)=CC(=C1)[C]23.CC(C)(C)C1=CC2=CC(C(C)(C)C)=CC3=CC(C(C)(C)C)=CC(=C1)[C]23
_[0,5,6,7,8,9,14,15,16,17,22,23,24,25][0,30,31,32,33,34,39,40,41,42,47,48,49,50]_2[51,52]

Висновки

1. Розроблено лінійне представлення для координаційних сполук та деяких інших типів структур (структур з довгими (багатоцентровими-ковалентними) зв'язками та структур з неароматичною делокалізацією)
2. Розроблено формат запису структур координаційних сполук
3. Розроблено дієвий підхід до структурного пошуку
4. Подолані такі проблеми як:
 - Гіпервалентність
 - Невизначеність структур за рахунок делокалізації
 - Загромадженість зв'язками
 - Незручність рисування структур для пошуку координаційних сполук
 - Невизначеність запису структур координаційних сполук
 - Канонізація
5. Розроблене представлення має такі переваги:
 - Можливість задавати координаційні зв'язки
 - Можливість задавати багатоцентрові зв'язки
 - Можливість задавати делокалізацію
 - Не потрібний новий алгоритм канонізації
 - Можливість візуалізувати структури у звичайних інтерпретаторах SMILES (мова йде про експортний варіант, але ніяких надбудов ми не побачимо)

Використана література та посилання

1. *Wiswesser W.J.* A line-formula chemical notation / W.J. Wiswesser. – New York: Crowell, 1954. – 149 p.
2. *Benson F.R.* Recording and Recovering Chemical Information with Standard Tabulating Equipment / F.R. Benson // 124-th National Meeting of the American Chemical Society. – Chicago: Remington Rand Inc, 1953.
3. *Smith E.G.* Machine Searching for Chemical Structures / E.G. Smith // Science. – 1960. – V. 131, Is. 3394. – P. 142-146.
4. *Granito C.E.* Chemical Substructure Index (CSI)-A New Research Tool / C.E. Granito, M.D. Rosenberg // Journal of Chemical Documentation. – 1971. – V. 11, Is. 4. – P. 251-256.
5. *Weininger D.* SMILES. A Chemical Language and Information System. 1. Introduction to Methodology and Encoding Rules. / D. Weininger // Journal of Chemical Information and Computer Sciences. – 1988. – V. 28, Is. 1. – P. 31-36.
6. *Weininger D.* SMILES. 2. Algorithm for generation of unique SMILES notation / D. Weininger, A. Weininger, J.L. Weininger // Journal of Chemical Information and Computer Sciences. – 1989. – V. 29, Is. 2. – P. 97-101.
7. <https://www.daylight.com/dayhtml/doc/theory/theory.smiles.html>
8. Siani MA, Weininger D, Blaney JM (1994) CHUCKLES: a method for representing and searching peptide and peptoid sequences on both monomer and atomic levels. *J Chem Inf Comput Sci* 34:588–593
9. Siani MA, Weininger D, James CA, Blaney JM (1995) CHORTLES: a method for representing oligomeric and template-based mixtures. *J Chem Inf Comput Sci* 35:1026–1033
10. https://www.inchi-trust.org/wp/wp-content/uploads/2014/06/InChI_Source_Code_Documentation_v1.0.pdf
11. <http://opensmiles.org/opensmiles.html>
12. https://github.com/IUPAC/IUPAC_SMILES_plus/blob/master/IUPAC_SMILES%2B.asciidoc#what-is-a-molecule-the-valence-model-of-chemistry
13. Hanson, R.M. Jmol SMILES and Jmol SMARTS: specifications and applications. *J Cheminform* 8, 50 (2016). <https://doi.org/10.1186/s13321-016-0160-4>
14. <https://docs.chemaxon.com/display/docs/chemaxon-extended-smiles-and-smarts-cxsmiles-and-cxsmarts.md>
15. <https://www.reaxys.com/>
16. https://www.rdkit.org/docs/RDKit_Book.html

17. Yujie Cui , Longjiu Cheng *RSC Adv.*, 2017,**7**, 49526-49531
- 18.J. J. Novoa, P. Lafuente, R. E. D. Sesto and J. S. Miller, *Angew. Chem., Int. Ed.*, 2001, **40**, 2540–2545
- 19.R. E. D. Sesto, J. S. Miller, P. Lafuente and J. J. Novoa, *Chem.–Eur. J.*, 2002, **8**, 4894–4908
- 20.A. G. Graham, F. Mota, E. Shurdha, A. L. Rheingold, J. J. Novoa and J. S. Miller, *Chem.–Eur. J.*, 2015, **21**, 13240–13245
21. J. Jakowski and J. Simons, *J. Am. Chem. Soc.*, 2003, **125**, 16089–16096
- 22.J. S. Miller, *Acc. Chem. Res.*, 2007, **40**, 189–196
- 23.Fernando Mota, Joel S. Miller, Juan J. Novoa *J. Am. Chem. Soc.* 2009, **131**, 22, 7699–7707
- 24.S. Z. Goldberg, B. Spivack, G. Stanley, E. Richard, D. M. Braitsch, J. S. Miller and M. Abkowitz, *J. Am. Chem. Soc.*, 1977, **99**, 110–117
- 25.C. Willi, A. H. Reis, E. Gebert and J. S. Miller, *Inorg. Chem.*, 1981, **20**, 313–318
- 26.J. D. Bagnato, W. W. Shum, M. Strohmeier, D. M. Grant, A. M. Arif and J. S. Miller, *Angew. Chem., Int. Ed.*, 2006, **45**, 5322–5326
- 27.E. B. Vickers, T. D. Selby and J. S. Miller, *J. Am. Chem. Soc.*, 2004, **126**, 3716–3717
- 28.B. A. Scott, S. J. L. Placa, J. B. Torrance, B. D. Silverman and B. Welber, *J. Am. Chem. Soc.*, 1977, **99**, 6631–6639
- 29.M. J. Hove, B. M. Hoffman and J. A. Ibers, *J. Chem. Phys.*, 1972, **56**, 3490–3502
- 30.J. K. Kochi, R. Rathore and P. L. Maguères, *J. Org. Chem.*, 2000, **65**, 6826–6836