

利用量子化學協助代謝物在串聯式質譜法的鑑定

鍾興翔、徐丞志

摘要

近年來，代謝體學 (metabolomics) 的研究逐漸蓬勃發展，並且被廣泛應用於臨床診斷、藥物開發、以及天然物研究等領域。由於質譜儀具有很好的靈敏度、特異性、以及通量，因此在代謝體的分析中扮演了重要的角色。在質譜分析中，為了鑑定代謝物結構，會以串聯質譜法 (MS/MS) 的方式取得分子的結構資訊。然而，當前可用的串聯質譜圖數據庫相當小，代謝物鑑定已經成為代謝體研究中的一大瓶頸。為了協助鑑定不在資料庫中的代謝物，目前已有許多團隊開發出電腦輔助代謝物串聯質譜圖鑑定的程式。例如，Neumann 教授團隊開發的 MetFrag 程式，便是藉由預測結構資料庫中化合物的二次質譜圖，並和實驗圖譜比對，進而協助未知物的鑑定。然而，當前方法較少考慮化合物的結構對分子碎裂的影響，因而高估了許多實際上不會產生的二次圖譜碎片。為了提升鑑定的準確度，在此計畫中，我們目標利用量子化學計算來了解化合物的結構與二次圖譜之間的關係，進而協助代謝物在串聯質譜法中的鑑定。

一、 研究動機

質譜儀是分析代謝體 (metabolomics) 非常重要的工具。然而，目前利用質譜進行代謝體的分析中，碰到最大的瓶頸是對小分子化合物的鑑定。在 Silva 等人的研究中發現，在液相層析串聯質譜 (Liquid chromatography tandem mass spectrometry) 的分析中，平均僅有 1.8% 的代謝物能夠被鑑定[1]。而大部分的代謝物即使能夠被偵測到，卻無法得知其結構為何，大大限制了非標靶式 (untargeted) 代謝體的研究速度。因此，在本計畫中，我們預計利用量子化學的計算，來輔助二次質譜對於未知小分子的鑑定。

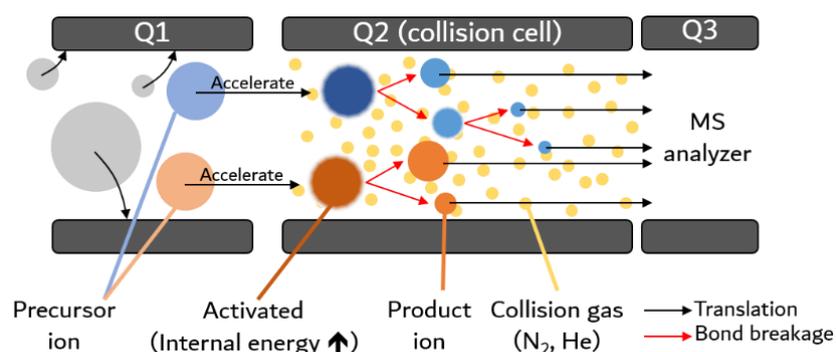
二、 文獻回顧

1. 代謝體質譜法 (mass spectrometry-based metabolomics)

代謝物是指生物代謝過程中的中間物以及最終產物，其在生物體中扮演相當重要的角色，包含訊號傳遞、能量產生、以及和環境溝通等。這些小分子是在分子生物學中心法則 (central dogma) 的最下游產物，因此對於了解生物的表型 (phenotype) 是相當重要的[2, 3]。針對生物樣品中的代謝物進行全面的定性、定量分析的研究稱為代謝體學，目前已被廣泛應用於疾病診斷[4, 5]、藥物開發[6, 7]、天然產物研究[8]等領域中。

質譜儀具有靈敏度 (sensitivity) 高、特異性 (specificity) 高、動態範圍 (dynamic range) 廣等優勢，因此常被用於代謝體學的分析中。質譜儀利用測量待測分子的質荷比 (m/z)，來提供待測分子的準確質量以及分子組成。然而，要從準確質量推出化合物的結構仍然很具挑戰性。即使是利用高解析質譜儀，質荷比的誤差小於 1 ppm，可能的異構物仍然太多[9]。因此，為了取得待測物的結構資訊，會利用串聯質譜法 (tandem mass spectrometry) 收集待測物的二次質譜圖 (MS/MS)。

串聯質譜法 (tandem mass spectrometry) 是利用將兩個質譜分析器串聯，藉以提供分析物的結構資訊 (圖一)。具體來說，串聯質譜法是測量分析物在撞碎前及撞碎後的離子質荷比。由於不同結構的分析物其斷裂方式不同，因此會產生不同的二次質譜圖。目前最常用的串聯質譜法為碰撞誘發解離法 (collision-induced dissociation, CID)，便是將分析物與中性惰性氣體，如氬氣、氮氣，進行碰撞，進而收集撞碎後的碎片離子。



圖一、三段四極柱式質譜類的碰撞誘發解離法示意圖

串聯式質譜法已經被廣泛應用於鑑定序列式的化合物，如蛋白質、多肽、核糖核酸等，且目前已有許多程式能夠利用二次質譜圖自動進行序列鑑定[10]。然而，這樣的方法並不能夠應用在小分子代謝物中，因為代謝物的結構多樣性太高，沒有單一的規則能夠由二次質譜圖中回推代謝物的結構。因此，目前的代謝物鑑定，多是以比對線上資料庫為主。然而，相較於結構化學資料庫如 ChemSpider，收集了超過 1 億種化合物，二次質譜圖的資料庫目前僅有約 2 萬種化合物的資料[11, 12]。因此，據統計，平均僅有 1.8% 的代謝物能夠被鑑定，而大部分的代謝物即使能夠被偵測到，卻無法得知其結構為何[1]。

2. 電腦輔助二次質譜圖鑑定方法

為了提升代謝物鑑定的數量，目前已經有許多團隊提出利用電腦演算法的方式來彌補二次圖譜資料庫的不足。根據演算法的原理，主要可以分成三種類別：規則式 (rule-based) 方法、機器學習式 (machine learning-based) 方法、以及組合式 (combinatorial-based) 方法。以下會分別介紹此三種方式的原理以及代表性的程式 (表一)。

表一、目前電腦輔助二次圖譜建定之方法

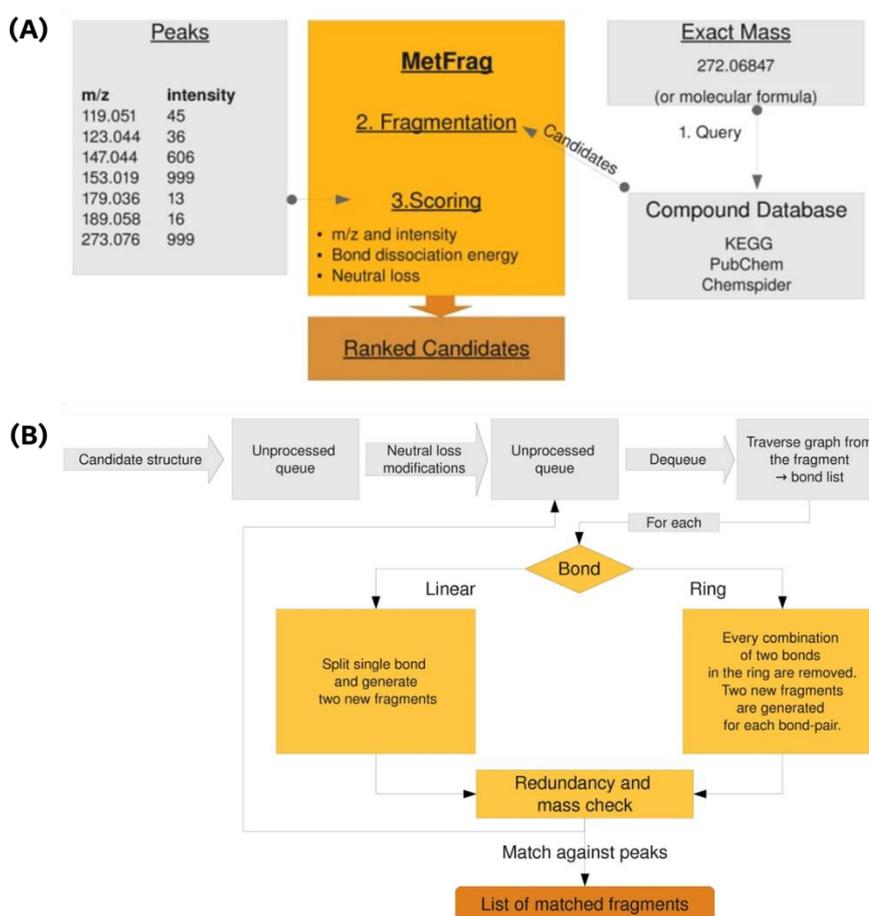
	規則式	機器學習式	組合式
策略	特定類別化合物	實驗二次質譜圖	提取可能結構
	↓	↓	↓
	利用規則產生二次	預測化學指紋	預測二次質譜圖
		↓	↓

	質譜圖	比對結構資料庫	比對實驗二次圖譜
代表軟體	LipidBlast[13] Mass Frontier	SIRIUS 4.0[14] MetFID[15]	MetFrag[16] CFM-ID[17]
目前瓶頸及可能原因	<ul style="list-style-type: none"> • 侷限於特定種類化合物，如酯質 	<ul style="list-style-type: none"> • 準確度待提升，且侷限於結構資料庫 • 缺乏足夠大小及多樣性的訓練資料庫 	<ul style="list-style-type: none"> • 預測偽陽性高，且侷限於結構資料庫 • 未考慮分子結構特性

規則式方法是直接將化合物的斷裂規則寫進程式中，進而達到預測化合物二次質譜圖的方法，此種方法和傳統蛋白質序列鑑定的方式相似。然而，在代謝體中，由於代謝物的結構過於多樣化，因此此種方法只能套用在特定類別的代謝物中，例如：酯質 (lipid)。酯質類的結構相當固定，且其在串聯質譜的斷裂模式也相當規律。在 Kind 等人的研究中，便利用酯質結構固定的特性，將酯質斷裂的規則寫入程式 LipidBlast 中，並利用此規則預測了 119,200 個酯質分子的二次質譜圖，用以鑑定酯質類化合物[13]。然而，此種方式僅能夠套用到特定種類的代謝物中。

第二類方法，是利用機器學習的演算法來輔助二次圖譜圖的鑑定，其主要的原理是以二次質譜圖為資料輸入，並訓練機器學習模型預測化合物的化學指紋 (Chemical fingerprint)。在預測出化學指紋後，便會以此為根據，搜尋結構資料庫，比對出較符合此化學指紋描述的化學分子。由 Dührkop 等人所開發出的 SIRIUS 4.0 軟體，便是使用此方法來進行二次圖譜的結構鑑定 [14]。SIRIUS 4.0 軟體在 2017 年的小分子結構鑑定競賽 (Critical Assessment of Small Molecule Identification (CASMI) contest) 中獲得了第一名。然而，在該次比賽的 45 張二次圖譜中，SIRIUS 4.0 也僅能將其中 28 張二次圖譜的正確結構排在前 10 當中，準確度仍待改進。

第三類方法為組合式方法，其原理是針對候選化合物進行二次圖譜預測，再與實驗圖譜進行比較，進而鑑定不在二次質譜資料庫中的化合物。例如由 Wolf 等人開發的 MetFrag 平臺，就是利用組合式方法來協助鑑定未知化合物，其運算流程如圖二所示[16, 18]。首先，會根據未知物的準確質量及同位素比例等資訊計算化合物的分子組成，並搜尋結構資料庫，如 KEGG、PubChem、ChempSpider 等，從中獲取可能的代謝物結構。接著，程式會以完全碎裂的方式預測所有候選分子的二次質譜圖。最後，將實驗的二次質譜圖與預測的二次質譜圖進行比較，利用相似程度進行排序。然而，此種方法未考量分子結構的特性，因此會預測出許多實驗上不會產生的碎片離子 (product ion)，造成偽陽性較高[19]。



圖二、(A) MetFrag 運算流程圖 (B) 二次圖譜計算流程[20]

3. 量子化學計算在二次圖譜中的應用

量子化學計算目前已經被廣泛應用於化合物的性質計算中，包含構型的優化、能量計算等。量子化學計算也已被用於二次質譜圖的解釋中(表二)，例如 Koopman 等人利用量子化學進行化學動力學模擬，用以解釋代謝物在進行碰撞誘發解離的過程[21]；Schüler 等人則是藉由計算斷鍵反應的反應熱來推估斷裂位置[22]；Wright 等人發現代謝物在離子化後，鍵長伸長 (>0.04 Å) 的化學鍵會在碰撞誘發解離時斷裂[19]。然而，目前利用量子化學計算進行二次質譜圖解釋的研究中，大部分都僅針對特定化合物進行解釋，截至 2020 年為止，尚未有量子化學計算的方法能夠準確的預測各式代謝物的二次質譜圖[23]。

表二、近年利用量子化學計算解釋二次質譜圖的研究整理

文獻	化合物選擇	計算方法	解釋方法
Koopman <i>et al.</i> , (2021)[21]	3-Methylhistidine, phenylephrine, acetylsulfamethoxazole	GFN2-xTB	化學動力學模擬
Bortňák <i>et al.</i> , (2020)[24]	21 pentafluorophenylhydrazones	B3LYP / cc-pVDZ	質子化後鍵長改變
Macaluso <i>et al.</i> , (2019)[25]	S-磺基-L-半胱氨酸	PM6	化學動力學模擬
Chiu <i>et al.</i> , (2019)[26]	葡萄糖胺、半乳糖胺	MP2	活化能計算
Schüler <i>et al.</i> , (2018)[22]	7個禁藥	PM7	鍵數、反應活化能
Janesko <i>et al.</i> , (2017)[27]	鳥嘌呤、脯氨酸、苯氧威	B3LYP / 6-31 + G(d,p),	熱力學可行性 (Thermodynamic

		PM6	feasibility)
Krishnan <i>et al.</i> , (2017)[28]	色氨酸	B3LYP / 6-31 G(d)	化學動力學模擬
Martínez-Cifuentes <i>et al.</i> , (2016)[29]	4 <i>o</i> -carbonyl quinone derivatives	B3LYP / 6-31 G(d,p)	熱力學可行性 (Thermodynamic feasibility)
Lee <i>et al.</i> , (2016)[30]	羰酮	PM3	化學動力學模擬
Wright <i>et al.</i> , (2014)[19]	15個藥物分子	AM1	質子化後鍵長改變

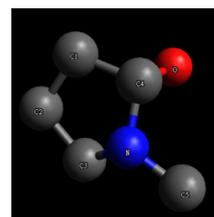
三、 研究計畫

1. 收集代謝物之二次質譜圖
2. 利用 Gaussian 16 軟體進行代謝物結構優化，並根據 Wright 等人的研究，了解質子化後構形改變與二次質譜圖之間的關係
3. 利用理論計算結果輔助預測代謝物之二次質譜圖
4. 驗證模型預測準確度
5. 建立代謝物二次質譜圖鑑定軟體

四、 初期數據

為了初步了解化合物結構與二次圖譜之間的關聯性，我們先重複 Wright 等人於 2014 年發表的結果[19]。我們利用 Gaussian 16 軟體針對 N-甲基吡咯烷醇 (1-methyl-2-pyrrolidinol) 化合物進行研究，分別計算其在分子狀態、以及質子化後的構形。藉由比較質子化前與後的鍵長變化，我們同樣發現鍵長伸長 (>0.04 Å) 的鍵結在碰撞誘發解離會斷鍵，此結果與 Wright 等人發表的結果相同。

Bond	Bond elongation (N+)	Bond elongation (O+)	Predicted fragment	Experimental fragment
C1,C2	-0.0060	-0.0018		
C1,C4	-0.0112	-0.0504		
C2,C3	0.0080	0.0032		
C3,N	0.0470	0.0271	O	O
N,C4	0.1117	-0.1568	O	O
N,C5	0.0429	0.0161	O	O
C4,O	-0.0637	1.3323	O	O



(1-methyl-2-pyrrolidinol)

圖三、N-甲基吡咯烷醇質子化後之鍵長變化，並依據鍵長伸長 $> 0.04 \text{ \AA}$ 做為標準判斷是否在碰撞誘發解離時會斷裂

五、 參考文獻

1. da Silva, R.R., P.C. Dorrestein, and R.A. Quinn, *Illuminating the dark matter in metabolomics*. Proceedings of the National Academy of Sciences, 2015. **112**(41): p. 12549-12550.
2. Fiehn, O., *Metabolomics—the link between genotypes and phenotypes*. Functional genomics, 2002: p. 155-171.
3. Guijas, C., et al., *Metabolomics activity screening for identifying metabolites that modulate phenotype*. Nature biotechnology, 2018. **36**(4): p. 316-320.
4. Huang, Y.-C., et al., *Predicting breast cancer by paper spray ion mobility spectrometry mass spectrometry and machine learning*. Analytical chemistry, 2019. **92**(2): p. 1653-1657.
5. Zhang, J., et al., *Mass spectrometry technologies to advance care for cancer patients IN clinical and intraoperative use*. Mass Spectrometry Reviews, 2020.
6. Wishart, D.S., *Emerging applications of metabolomics in drug discovery and precision medicine*. Nature Reviews Drug Discovery, 2016. **15**(7): p. 473-484.
7. Jacob, M., et al., *Metabolomics toward personalized medicine*. Mass spectrometry reviews, 2019. **38**(3): p. 221-238.
8. Zhao, Q., J.-L. Zhang, and F. Li, *Application of metabolomics in the study of natural products*. Natural products and bioprospecting, 2018. **8**(4): p. 321-334.
9. Kind, T. and O. Fiehn, *Metabolomic database annotations via query of elemental compositions: Mass accuracy is insufficient even at less than 1 ppm*. BMC Bioinformatics, 2006. **7**(1): p. 234.
10. Medzihradzky, K.F. and R.J. Chalkley, *Lessons in de novo peptide sequencing by tandem mass spectrometry*. Mass Spectrom Rev, 2015. **34**(1): p. 43-63.
11. Johnson, S.R. and B.M. Lange, *Open-access metabolomics databases for*

- natural product research: present capabilities and future potential*. Front Bioeng Biotechnol, 2015. **3**: p. 22.
12. Vinaixa, M., et al., *Mass spectral databases for LC/MS-and GC/MS-based metabolomics: State of the field and future prospects*. TrAC Trends in Analytical Chemistry, 2016. **78**: p. 23-35.
 13. Kind, T., et al., *LipidBlast in silico tandem mass spectrometry database for lipid identification*. Nature Methods, 2013. **10**(8): p. 755-758.
 14. Dührkop, K., et al., *SIRIUS 4: a rapid tool for turning tandem mass spectra into metabolite structure information*. Nature methods, 2019. **16**(4): p. 299-302.
 15. Fan, Z., et al., *MetFID: artificial neural network-based compound fingerprint prediction for metabolite annotation*. Metabolomics, 2020. **16**(10): p. 1-11.
 16. Ruttkies, C., S. Neumann, and S. Posch, *Improving MetFrag with statistical learning of fragment annotations*. BMC bioinformatics, 2019. **20**(1): p. 1-14.
 17. Djoumbou-Feunang, Y., et al., *CFM-ID 3.0: Significantly improved ESI-MS/MS prediction and compound identification*. Metabolites, 2019. **9**(4): p. 72.
 18. Ruttkies, C., et al., *MetFrag relaunched: incorporating strategies beyond in silico fragmentation*. Journal of cheminformatics, 2016. **8**(1): p. 1-16.
 19. Wright, P., A. Alex, and F. Pullen, *Predicting collision-induced dissociation spectra: Semi-empirical calculations as a rapid and effective tool in software-aided mass spectral interpretation*. Rapid Communications in Mass Spectrometry, 2014. **28**(10): p. 1127-1143.
 20. Wolf, S., et al., *In silico fragmentation for computer assisted identification of metabolite mass spectra*. BMC bioinformatics, 2010. **11**(1): p. 1-12.
 21. Koopman, J. and S. Grimme, *From QCEIMS to QCxMS: A Tool to Routinely Calculate CID Mass Spectra Using Molecular Dynamics*. Journal of the American Society for Mass Spectrometry, 2021.
 22. Schöler, J.-A., et al., *ChemFrag: Chemically meaningful annotation of fragment ion mass spectra*. Journal of Mass Spectrometry, 2018. **53**(11): p. 1104-1115.
 23. Borges, R.M., et al., *Quantum Chemistry Calculations for Metabolomics: Focus Review*. Chemical Reviews, 2021.
 24. Bortňák, D., et al., *Mass spectrometrical and quantum-chemical study of pentafluorophenylhydrazones*. Journal of Mass Spectrometry, 2020. **55**(10): p. e4540.
 25. Macaluso, V., et al., *l-Cysteine Modified by S-Sulfation: Consequence on Fragmentation Processes Elucidated by Tandem Mass Spectrometry and Chemical Dynamics Simulations*. The Journal of Physical Chemistry A, 2019. **123**(17): p. 3685-3696.
 26. Chiu, C.-c., et al., *Toward closing the gap between hexoses and N-*

- Acetylhexosamines: experimental and computational studies on the collision-induced dissociation of Hexosamines.* The Journal of Physical Chemistry A, 2019. **123**(31): p. 6683-6700.
27. Janesko, B.G., L. Li, and R. Mensing, *Quantum Chemical Fragment Precursor Tests: Accelerating de novo annotation of tandem mass spectra.* Analytica chimica acta, 2017. **995**: p. 52-64.
28. Krishnan, Y., et al., *Classical Dynamics Simulations of Dissociation of Protonated Tryptophan in the Gas Phase.* The Journal of Physical Chemistry A, 2017. **121**(23): p. 4389-4396.
29. Martínez-Cifuentes, M., et al., *Protonation sites, tandem mass spectrometry and computational calculations of o-carbonyl carbazolequinone derivatives.* International journal of molecular sciences, 2016. **17**(7): p. 1071.
30. Lee, G., et al., *Gas phase fragmentation mechanisms of protonated testosterone as revealed by chemical dynamics simulations.* International Journal of Mass Spectrometry, 2016. **407**: p. 40-50.