

BiolSys Metabolomika - raport

Piotr Kupidura

11 czerwca 2025

1 Wstęp

Spektroskopia NMR jest techniką eksperymentalną opartą na właściwościach magnetycznych atomów. Umożliwia ona m.in. identyfikację danego związku chemicznego na podstawie jego widma. Z tego powodu znajduje ona zastosowanie w metabolomice - dziedzinie zajmującej się badaniem związków niskocząsteczkowych zaangażowanych w przemiany metaboliczne. Może ją bowiem wykorzystać w celu wyznaczenia składu chemicznego mieszaniny, chociażby próbki pobranej od pacjenta. Umożliwia ona więc wykrycie zmian w stężeniach metabolitów, a co za tym idzie diagnozowanie chorób związanych z metabolizmem.

2 Metody

2.1 Widma eksperymentalne

Zbiór widm eksperymentalnych pochodzi z badania oceniającego zmiany poziomu metabolitów w moczu sześciu wolontariuszy przebywających w standardyzowanym i stabilnym środowisku. Do analizy wybrane zostały próbki od jednego z wolontariuszy, pobierane co około 10 dni. Ponadto jest też próbka kontrolna, powstała poprzez zmieszanie pozostałych próbek. Pliki csv z widmami zawierają dane o ich położeniu (ppm) i natężeniu. W ramach preprocessingu z widm usunięto ujemne natężenia (oznaczają one błąd eksperymentalny) i je znormalizowano (zniwelowanie różnic między widmami z różnych źródeł).

2.2 Widma referencyjne

Widma referencyjne pochodzą z HMDB - bazy danych o metabolitach ludzkiego organizmu zawierającej m.in. ich widma NMR. Są one następnie przetwarzane poprzez: grupowanie peaków (niektóre widma zawierają kilka peaków w tym samym miejscu co może zaburzać analizę), usunięcie ujemnych natężeń (oznaczają one błąd eksperymentalny) i ich normalizację (zniwelowanie różnic między widmami z różnych źródeł). Spośród wybranych metabolitów, 26 posiada widma eksperymentalne, W celu dalszego zmniejszenia różnic pomiędzy widmami, do analizy wybrano widma symulowane o częstotliwości 600 MHz (z wyjątkiem 3-HIBA, dla którego dostępne było jedynie 500 MHz).

2.3 Parametry i solvery

W celu znalezienia optymalnych parametrów porównano wyniki estymacji proporcji dla trzech widm eksperymentalnych i widma kontrolnego. Sprawdzono działanie dwóch solverów: HiGHS i CBC_CMD, jednak dały one identyczne wyniki (tab. 1), więc w dalszych analizach wykorzystano CBC_CMD. Model zawiera również parametry \mathcal{K} , odpowiadające karom za usunięcie szumu z badanego widma ($\mathcal{K}_{mixture}$) i widm referencyjnych ($\mathcal{K}_{components}$). W celu znalezienia optymalnych wartości zastosowano grid-search dla $\mathcal{K}_{mixture} \in \{0.2, 0.5, 0.3\}$ i $\mathcal{K}_{components} \in \{0.18, 0.22, 0.26\}$ (tab. 2). Najlepsze wyniki osiągnięto dla domyślnych wartości: $\mathcal{K}_{mixture} = 0.25$ i $\mathcal{K}_{components} = 0.22$.

3 Wyniki

solver	HiGHS	CBC_CMD
mean signal proportion	0.9713	0.9713
mean Wasserstein distance	0.0523	0.0523

Tabela 1: Porównanie wyników dla trzech widm eksperymentalnych i widma kontrolnego przy użyciu dwóch różnych solverów.

parametry	signal proportion	Wasserstein distance
$\mathcal{K}_{mixture} = 0.2, \mathcal{K}_{components} = 0.18$	0.9771	0.1444
$\mathcal{K}_{mixture} = 0.2, \mathcal{K}_{components} = 0.22$	0.9515	0.1494
$\mathcal{K}_{mixture} = 0.2, \mathcal{K}_{components} = 0.26$	0.9544	0.0495
$\mathcal{K}_{mixture} = 0.25, \mathcal{K}_{components} = 0.18$	0.9506	0.0514
$\mathcal{K}_{mixture} = 0.25, \mathcal{K}_{components} = 0.22$	0.9695	0.0514
$\mathcal{K}_{mixture} = 0.25, \mathcal{K}_{components} = 0.26$	0.9693	0.0515
$\mathcal{K}_{mixture} = 0.3, \mathcal{K}_{components} = 0.18$	0.9702	0.0531
$\mathcal{K}_{mixture} = 0.3, \mathcal{K}_{components} = 0.22$	0.9701	0.0532
$\mathcal{K}_{mixture} = 0.3, \mathcal{K}_{components} = 0.26$	0.9701	0.0532

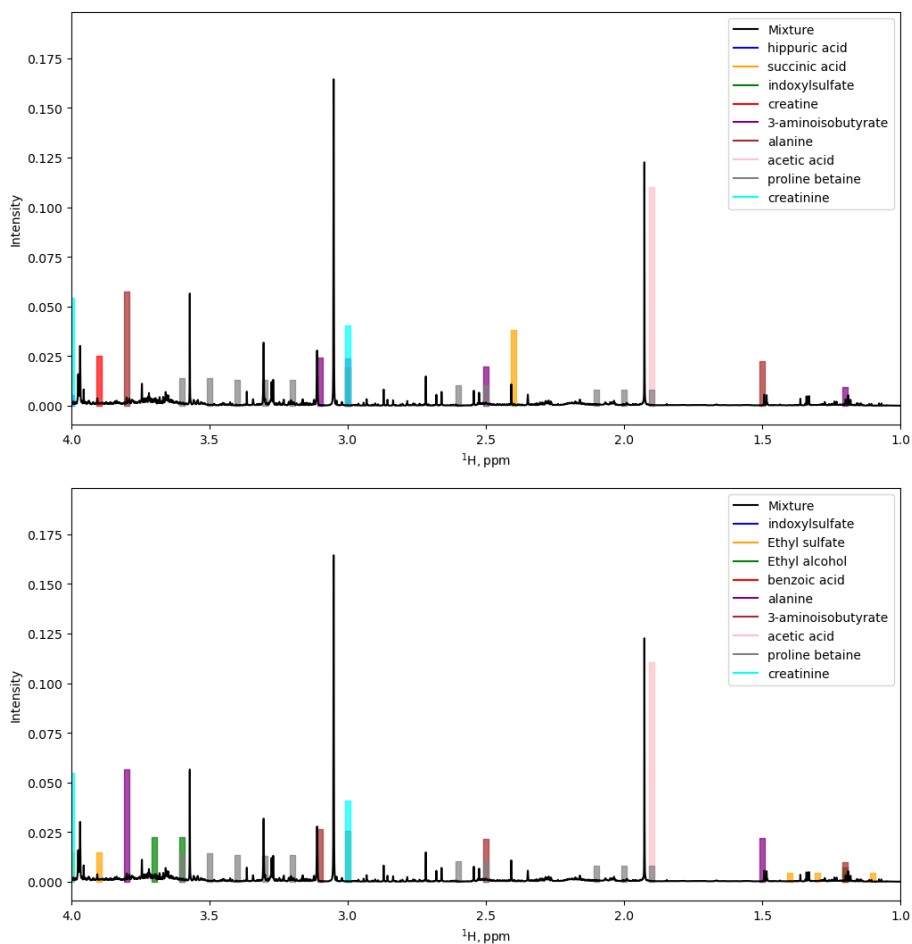
Tabela 2: Porównanie wyników dla widma kontrolnego przy różnych wartościach parametrów, dla proporcji sygnału podawana jest średnia po chunkach.

Na przygotowanych danych przeprowadzono dwie analizy. Najpierw sprawdzono, jak zmieniają się w czasie stężenia metabolitów (rys. 1). Widać, że niektóre metabolity (np. kwas cytrynowy czy dimetyloamina) wcale nie występują w badanych próbkach. Inne (np. 3-hydroksyzowalerianian czy mioinozytol) pojawiają się w pojedynczych krokach czasowych, choć w przypadku np. mioinozytolu proporcje są na tyle niskie, że mogłyby to być również błąd modelu. Ponadto, dla wszystkich obecnych metabolitów ich proporcje wykazują wysoką zmienność w czasie.

Następna analiza miała na celu porównanie wyników estymacji z inną biblioteką referencyjną. W tym celu, użytą w poprzedniej analizie bibliotekę rozszerzono o widma (wybrane według tych samych kryteriów co wcześniej) 4 związków, związanych z metabolizmem alkoholu etylowego (aldehyd octowy, alkohol etylowy, glukuronian etylu i siarczan etylu). Porównanie obu bibliotek dla przykładowego widma przedstawia rys. 2. Widać, że wśród dziewięciu metabolitów o najwyższych proporcjach, znajdują się dwa spoza początkowej biblioteki (etanol i siarczan etylu). Zwiększył się też udział pierwszych dziewięciu komponentów - przy rozszerzonej bibliotece stanowią one 73% całości, a przy oryginalnej 69%.



Rysunek 1: Zmiany stężeń metabolitów w czasie (na wykresach pominięto kroki czasowe 713-716 dla lepszej widoczności).



Rysunek 2: Porównanie proporcji metabolitów w widmie uzyskanym po 50 dniach dla bibliotek bazowej (góra) i rozszerzonej o produkty przemiany etanolu (dół). Wyniki zostały ograniczone do 9 metabolitów o najwyższych proporcjach.

4 Dyskusja

Stosując powyższe podejście udało się przeprowadzić prostą analizę składu chemicznego próbek moczu. Otrzymane wyniki są jednak dalekie od ideału. Zasadniczym problemem wydaje się przygotowanie biblioteki referencyjnej - skoro rozszerzenie jej o 4 metabolity polepszyło wyniki, to możemy spodziewać się, że dalsza jej poprawa (nie tylko na drodze doboru metabolitów, ale też doboru widm i ich preprocessingu) doprowadziłaby do jeszcze lepszych wyników. Problematyczna okazała się też wizualizacja widm (suma komponentów nie od-

powiada widmu z próbki, problem ten próbowałem obchodzić ręcznie skalując widmo), prawdopodobnie ze względu na ich normalizację.

Literatura

- [1] Barbara Domżał, Ewa Klaudia Nawrocka, Dariusz Gołowicz, Michał Aleksander Ciach, Błażej Miasojedow, Krzysztof Kazimierczuk, and Anna Gambin. Magnetstein: An open-source tool for quantitative nmr mixture analysis robust to low resolution, distorted lineshapes, and peak shifts. *Analytical Chemistry*, 96(1):188–196, 2024. PMID: 38117933.
- [2] Daniel Raftery G A Nagana Gowda. NMR based metabolomics. *Adv Exp Med Biol*, 2021.