

## 1. Wprowadzenie

Spektroskopia NMR (rezonansu magnetycznego jądrowego) to technika służąca do poznawania struktury chemicznej cząsteczek. Wykorzystuje ona magnetyczne właściwości jąder atomowych, badając spiny jąder oraz ich reakcje z otoczeniem [1]. Jest to nieinwazyjna metoda ilościowa zdolna do jednoczesnego wykrywania setek różnych związków. Dzięki tym cechom NMR znalazło szerokie zastosowanie w metabolomice- nauce zajmującej się analizą złożonych mieszanin związków molekularnych (metabolitów) obecnych w próbkach biologicznych takich jak komórki, tkanki, płyny ustrojowe [2]. Metabolity dostarczają informacji o stanie fizjologicznym organizmu i przebiegu procesów biochemicznych. Są również potencjalnymi biomarkerami przy wykrywaniu chorób [3]. Analizę metabolomiczną można przeprowadzić na przykład na krwi lub moczu, co jest przydatne przy analizie przebiegu chorób czy badaniu wpływu środowiska na jednostkę. Użycie metody NMR do wykrycia odpowiednich związków jest proste w przygotowaniu, wysoce reprodukowalne i nie niszczy próbki biologicznej [4].

W tej pracy badamy próbki moczu uzyskanych przy użyciu jednowymiarowej spektroskopii NMR  $^1\text{H}$  1D pod kontem obecności wybranych metabolitów. Naszym celem jest ustalenie, czy zmienność składu moczu zależy bardziej od czynników środowiskowych czy od indywidualnych różnic między badanymi osobami.

## 2. Metody

- Dane eksperymentalne

Analizujemy dane z projektu, w którym brało udział 6 wolontariuszy. Każdy z nich przebywał w tym samym środowisku, stosował tę samą dietę i rytm dobowy. Mocz był zbierany codziennie przez 254 dni oraz w dniach 713-716 od początku eksperymentu. Poza tym zmieszano ze sobą wszystkie próbki zebrane podczas eksperymentu i wykonano 32 pomiary stworzonej w ten sposób próby kontrolnej (QC). Te próbki zostały poddane badaniu przy pomocy metody NMR  $^1\text{H}$  1D. W tej pracy skupimy się na 29 widmach NMR próbek pochodzących od jednego wolontariusza oraz jednego widma próbki QC. W ramach preprocessingu usunięto ujemne intensywności i dane zostały znormalizowane.

- Biblioteki referencyjne

Do analizy wykorzystano 2 zestawy metabolitów referencyjnych. Pierwszy z nich obejmuje 35 wybranych wcześniej metabolitów. Drugi składa się z 53 metabolitów i zawiera zarówno część metabolitów z pierwszego zestawu jak i nowe. Związki takie jak np. glicyna, kreatynina, alanina są wykorzystywane w obu bibliotekach, a np. fenol, mannitol czy seryna są nowym rozszerzeniem. Drugi zbiór metabolitów został przygotowany na podstawie Figure 1 z artykułu Bouatra et. al. 2013 [5], gdzie opisany jest typowy metaboliczny skład moczu ludzkiego od zdrowego pacjenta uzyskany za pomocą NMR. Sprawia to, że jest to dobra baza referencyjna do określenia składu próbek. Dokładna lista metabolitów jest w pliku `new_selected_metabolites.csv`.

Dane referencyjne zostały pobrane z HMDB (Human Metabolome Database). Jest to baza danych metabolomicznych skoncentrowana na człowieku. Zawiera ona szczegółowe informacje o metabolitach takie jak: nazwa związku, wzory sumaryczne i cząstkowe, spektra NMR i MS, ścieżki biochemiczne,

interakcje z lekami i inne. Spośród metabolitów z bazy HMDB wykrytych w moczu 1843 posiada eksperymentalne widma 1D 1H NMR, z czego 1607 jest wykrytych i skwantyfikowanych, a 236 wykrytych i nieskwantyfikowanych. Spośród 35 wybranych wcześniej metabolitów 26 posiada spektra eksperymentalne 1H 1D. Dla wszystkich z tych metabolitów najwyższa częstotliwość w przypadku danych przewidywanych to 1000 MHz.

Obydwa zestawy widm referencyjnych zostały wyszukane w bazie HMDB, a następnie odfiltrowane aby pozostały tylko dane uzyskane za pomocą metody NMR 1H 1D. Piki zostały zgrupowane po ich pozycji ppm shift bez zaokrąglania, ponieważ wartości w plikach są podawane z 2 miejscami po przecinku, a według artykułu [6] najlepiej podawać je z dokładnością do przynajmniej 3 miejsc, więc kolejne zaokrąglanie jeszcze bardziej zmniejszyłoby precyzję danych. Zgrupowanym wartościom przypisano nową intensywność równą sumie intensywności w grupie. Po tych działaniach dla większości metabolitów zostało przynajmniej kilka spektrów mogących być potencjalną referencją. Ostatecznie wybrano te z największą liczbą pików dla danego metabolitu (ale nie takie, których wszystkie intensywności miały wartość 0). Większa liczba pików pozwala na lepsze porównywanie i dopasowanie do danych eksperymentalnych. W kolejnym kroku usunięte zostały intensywności ujemne, ponieważ model liniowy takich nie przewiduje i nie mają one biologicznego sensu. Poza tym referencje znormalizowano tak jak dane eksperymentalne, aby były w tej samej skali i dało się je porównywać.

- Funkcje, solvery, parametry

Porównywanie składów wielu próbek moczu odbywało się za pomocą biblioteki Magnetstein [7].

Przy wyliczaniu proporcji odpowiednich metabolitów wybrano opcję `what_to_compare = area`, czyli wynikowe proporcje dotyczą pola pod wykresem. Jest ono proporcjonalne do liczby protonów w związku, które generują piki. Dodatkowo funkcja zwraca wtedy również o szumie, szumie w referencji i jego proporcji w komponentach.

Trzy porównywane solvery to Gurobi, Mosek i HiGHs. Osiągają one podobne wyniki patrząc na procent szumu i stopień wyjaśnienia spektrum mieszaniny. Jednak HiGHs produkuje intensywności ujemne (które co do wartości bezwzględnej zgadzają się z pozostałymi solverami). Istotny jest też czas wykonywania obliczeń, który jest najmniejszy dla Mosek, więc główna analiza została przeprowadzona przy pomocy właśnie tego solvera.

W trakcie badań manipulujemy dwoma parametrami w poszukiwaniu najlepszych rezultatów. Parametr  $\kappa_{\text{mixture}}$  to wartość przedstawiająca maksymalny dystans transportu dla mieszaniny. W trakcie estymowania proporcji sygnał będzie przenoszony jedynie na taki dystans. Jest to też interpretowane jako kara usuwania szumu z mieszaniny.  $\kappa_{\text{compounds}}$  to maksymalny dystans transportu dla poszczególnych komponentów (referencji). Przy estymacji proporcji sygnał ze spektrów komponentów będzie transportowany do tego dystansu. Jest to kara za usuwanie szumu dla komponentów. Sprawdzane były różne wartości tych parametrów, gdzie  $\kappa_{\text{mixture}}$  i  $\kappa_{\text{compounds}}$  wynosiły odpowiednio 0.25, 0.22 (wartość domyślna); 0.25, 0.15; 0.25, 0.30; 0.15, 0.22; 0.30, 0.22. Najlepsze wyniki uzyskano dla  $\kappa_{\text{mixture}} = 0.3$   $\kappa_{\text{compounds}} = 0.22$ , więc te wartości były wykorzystywane we właściwej analizie.

### 3. Wyniki

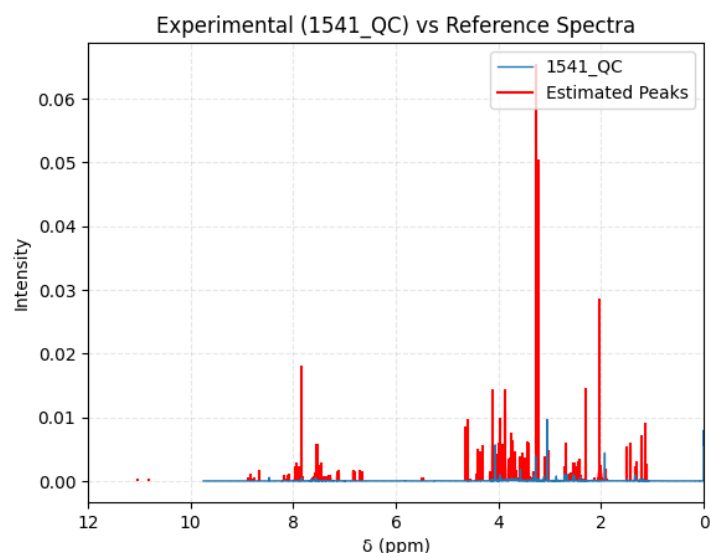
- Porównywanie solverów

Porównywanie solverów zostało przeprowadzone na 35 wybranych wcześniej metabolitach jako referencja i dwóch przykładowych próbkach moczu (z dnia 50 – mix1 i z dnia 160 – mix 2) oraz próbki QC. Parametry  $\kappa$  były tutaj domyślne czyli 0.25 i 0.22. Osiągane wyniki pod względem proporcji sygnału, szumu i całkowitego wytłumaczenia danych są bardzo podobne (tab. 1). Przy solverze HiGHs wartości tych wskaźników oraz proporcje są ujemne, ale co do wartości bezwzględnej są spójne z pozostałymi solverami. Porównane zostały także wartości proporcji dla poszczególnych metabolitów w

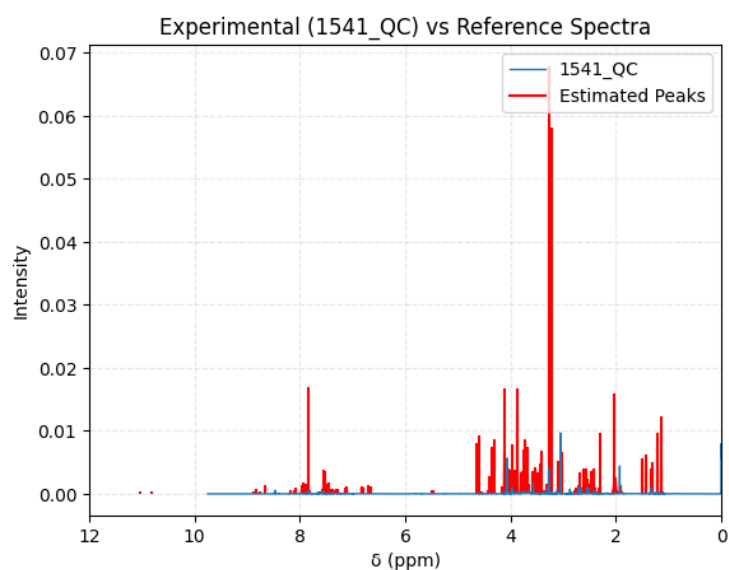
tych 9 testach (3 próbki moczu, 3 solvery). Również uzyskaliśmy bardzo podobne wyniki, a przynajmniej tego samego rzędu. Wykresy widm modelowych i eksperymentalnych QC (fig.1, fig.2, fig. 3) także wyglądają podobnie dla wszystkich solverów (poza tym, że dla HiGHs mamy wartości ujemne). Sugeruje to, że wyniki są wiarygodne i nie ma między solverami znaczących różnic. Z tego powodu do dalszej analizy używano Mosek, gdyż osiągał on rezultaty w znacząco krótszym czasie niż pozostałe solvery.

	Gurobi			Mosek			HiGHs		
	mix1	mix2	QC	mix1	mix2	QC	mix1	mix2	QC
Signal proportion in mixture's spectrum	0.916	0.916	0.916	0.916	0.916	0.919	-0.916	-0.916	-0.019
Noise proportion in mixture's spectrum	0.084	0.084	0.084	0.084	0.084	0.08	-0.08	-0.084	-0.08
Total explanation	1.0	1.0	1.0	1.0	1.0	1.0	-1.0	-1.0	-1.0
Noise proportion in combination of components' spectra	0.025	0.025	0.025	0.025	0.034	0.015	-0.025	-0.034	-0.015

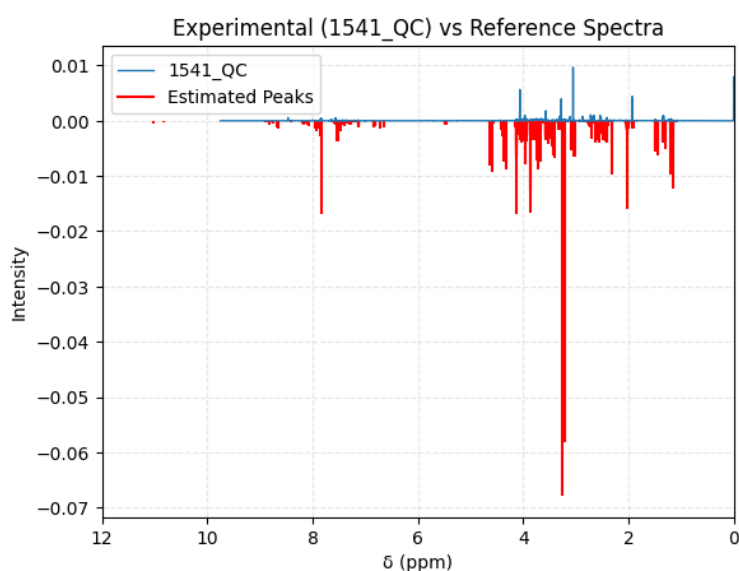
**Tabela 1.** Wyniki proporcji sygnału, szumu i całkowitego wyjaśnienia spektrów eksperymentalnych dla 3 solverów i 3 próbek moczu. Osiągane wyniki są porównywalne.



**Fig. 1** Widmo modelowe i eksperymentalne dla danych Quality Control uzyskane za pomocą solvera Gurobi. Oś X przedstawia wartości ppm shift, oś Y intensywności. Na czerwono zaznaczono widmo modelowe utworzone przez przemnożenie widm referencyjnych przez wynikowe proporcje, a na niebiesko widmo QC.



**Fig. 2** Widmo modelowe i eksperymentalne dla danych Quality Control uzyskane za pomocą solvera Mosek. Oś X przedstawia wartości ppm shift, oś Y intensywności. Na czerwono zaznaczono widmo modelowe utworzone przez przemnożenie widm referencyjnych przez wynikowe proporcje, a na niebiesko widmo QC.



**Fig. 3** Widmo modelowe i eksperymentalne dla danych Quality Control uzyskane za pomocą solvera HiGHs. Oś X przedstawia wartości ppm shift, oś Y intensywności. Na czerwono zaznaczono widmo modelowe utworzone przez przemnożenie widm referencyjnych przez wynikowe proporcje, a na niebiesko widmo QC.

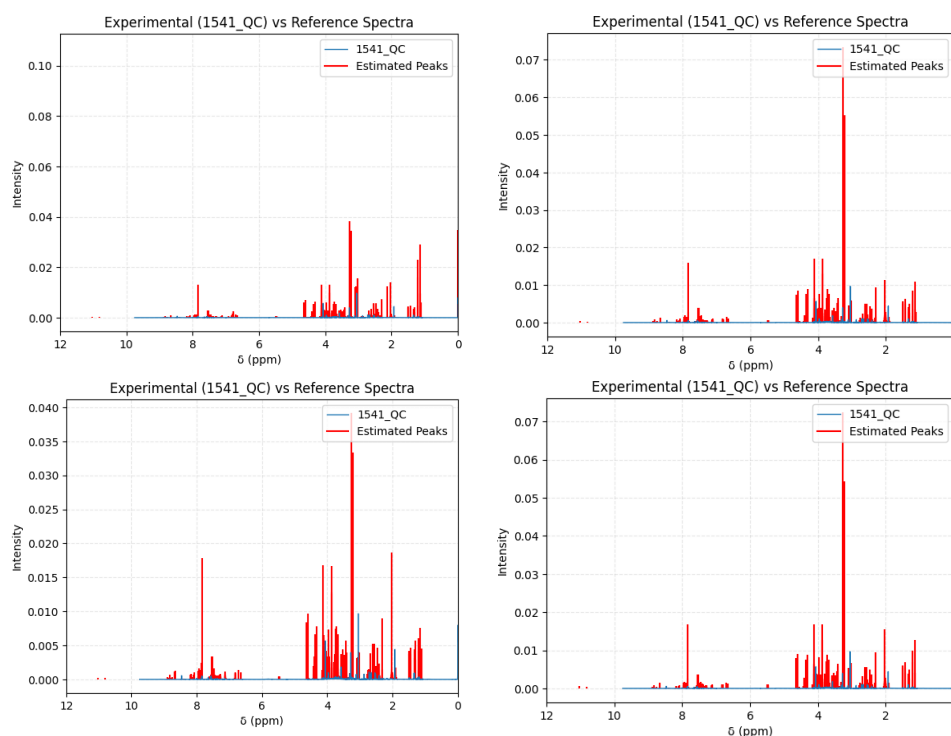
- Porównywanie parametrów  $\kappa$

Do sprawdzenia różnych parametrów używano wartości gdzie  $\kappa_{\text{mixture}}$  i  $\kappa_{\text{compounds}}$  wynosiły odpowiednio 0.25, 0.22 (wartość domyślna); 0.25, 0.15; 0.25, 0.30; 0.15, 0.22; 0.30, 0.22. Takie zróżnicowanie parametrów pozwala na sprawdzenie wartości domyślnych, oraz wartości większych i mniejszych obu parametrów. Wszystkie te analizy przeprowadzono przy użyciu solvera Mosek i próbki QC moczu. Dla wartości (0.25, 0.15) uzyskano największą proporcję sygnału w uzyskanym spektrum, ale procent szumu dla komponentów jest zbyt duży, żeby uznać te parametry za optymalne. Parametry (0.15, 0.22) osiągają słabą proporcję sygnału. Parametry (0.25, 0.30), (0.30, 0.22) i (0.25, 0.22) osiągają podobne

wyniki, jednak procent sygnału jest największy dla (0.30, 0.22) więc to te parametry będą wykorzystywane w dalszej analizie (tab. 2). Patrząc na wykresy widm modelowych i eksperymentalnych dla tych parametrów (fig. 4) widzimy że najlepsze dopasowanie występuje dla parametrów (0.25, 0.15), czyli w przypadku największej proporcji sygnału w mieszaninie, ale dużym szumie w komponentach. Reszta wykresów jest porównywalna z niewielkimi różnicami w skali. Te obserwacje są spójne z analizą dla wartości z tabeli 2. Przy szukaniu optymalnych wartości parametrów także porównano uzyskane proporcje dla poszczególnych metabolitów i są one podobne dla różnych parametrów, jak i dla wyników z poprzedniego punktu przy porównywaniu solverów.

	(0.25, 0.15)	(0.25, 0.30)	(0.15, 0.22)	(0.30, 0.22)	(0.25, 0.22)
Signal proportion in mixture's spectrum	0.97	0.91	0.79	0.94	0.919
Noise proportion in mixture's spectrum	0.027	0.08	0.21	0.06	0.08
Total explanation	1.0	1.0	1.0	1.0	1.0
Noise proportion in combination of components' spectra	0.22	0.01	0.14	0.015	0.015

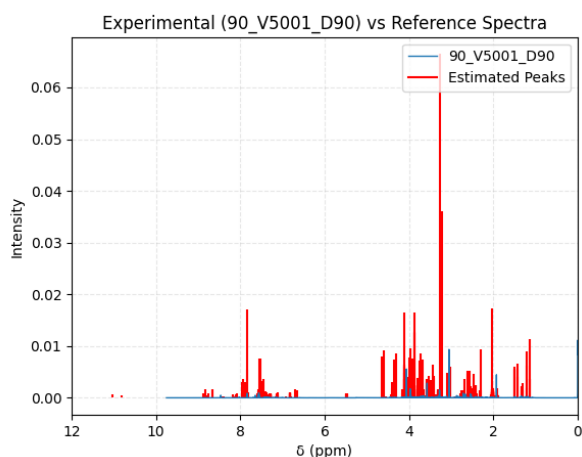
**Tabela 2.** Wyniki proporcji sygnału, szumu i całkowitego wyjaśnienia spektrów eksperymentalnych dla 4 różnych zestawów parametrów  $\kappa$  w porównaniu do domyślnych wartości (na niebiesko). Wartości parametrów są podawane w postaci ( $\kappa_{\text{mixture}}$ ,  $\kappa_{\text{compounds}}$ ).



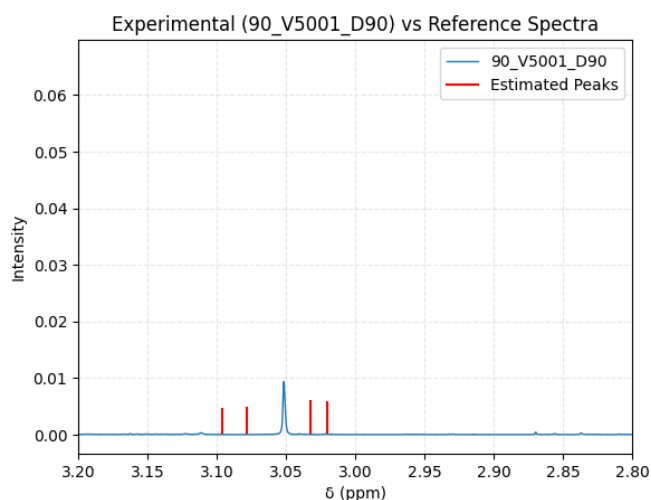
**Fig. 4** Widma modelowe i eksperymentalne dla danych Quality Control uzyskane za pomocą solvera HiGHs i różnych parametrów  $\kappa$ . Oś X przedstawia wartości ppm shift, oś Y intensywności. Na czerwono zaznaczono widmo modelowe utworzone przez przemnożenie widm referencyjnych przez wynikowe proporcje, a na niebiesko widmo QC. W lewym górnym rogu parametry wynoszą  $\kappa_{\text{mixture}}=0.25$ ,  $\kappa_{\text{compounds}}=0.15$ , w prawym górnym  $\kappa_{\text{mixture}}=0.25$ ,  $\kappa_{\text{compounds}}=0.30$ , w lewym dolnym  $\kappa_{\text{mixture}}=0.15$ ,  $\kappa_{\text{compounds}}=0.22$  w prawym dolnym  $\kappa_{\text{mixture}}=0.30$ ,  $\kappa_{\text{compounds}}=0.22$ .

- Szacowanie proporcji dla wszystkich próbek

Szukanie proporcji metabolitów referencyjnych we wszystkich próbkach odbyło się za pomocą solvera Mosek i parametrów  $\kappa_{\text{mixture}}=0.30$ ,  $\kappa_{\text{compounds}}=0.22$ . Patrząc na przykładowy wykres uzyskany w tej analizie (fig. 5), ale też na pojawiające się w tej pracy poprzednie wykresy widać, że dopasowania piki modelowe pomnożone przez uzyskane proporcje nie dopasowują się poprawnie do pików eksperymentalnych. Może to wynikać np. z nieprawidłowego działania funkcji `normalize()`, która jednak nie zwraca widm w zawsze tej samej skali. Żeby być pewnym, trzeba by jednak dokładnie prześledzić mapę transportu. Możemy jednak zwrócić uwagę, że niektóre piki jednak wyglądają rozsądnie. Przybliżając nasz wykres (fig. 6) widzimy pik eksperymentalny, który mógłby być wyjaśniony otaczającymi go pikami estymowanymi, co świadczy, że na podstawie tej analizy można jednak wnioskować o składzie próbek.

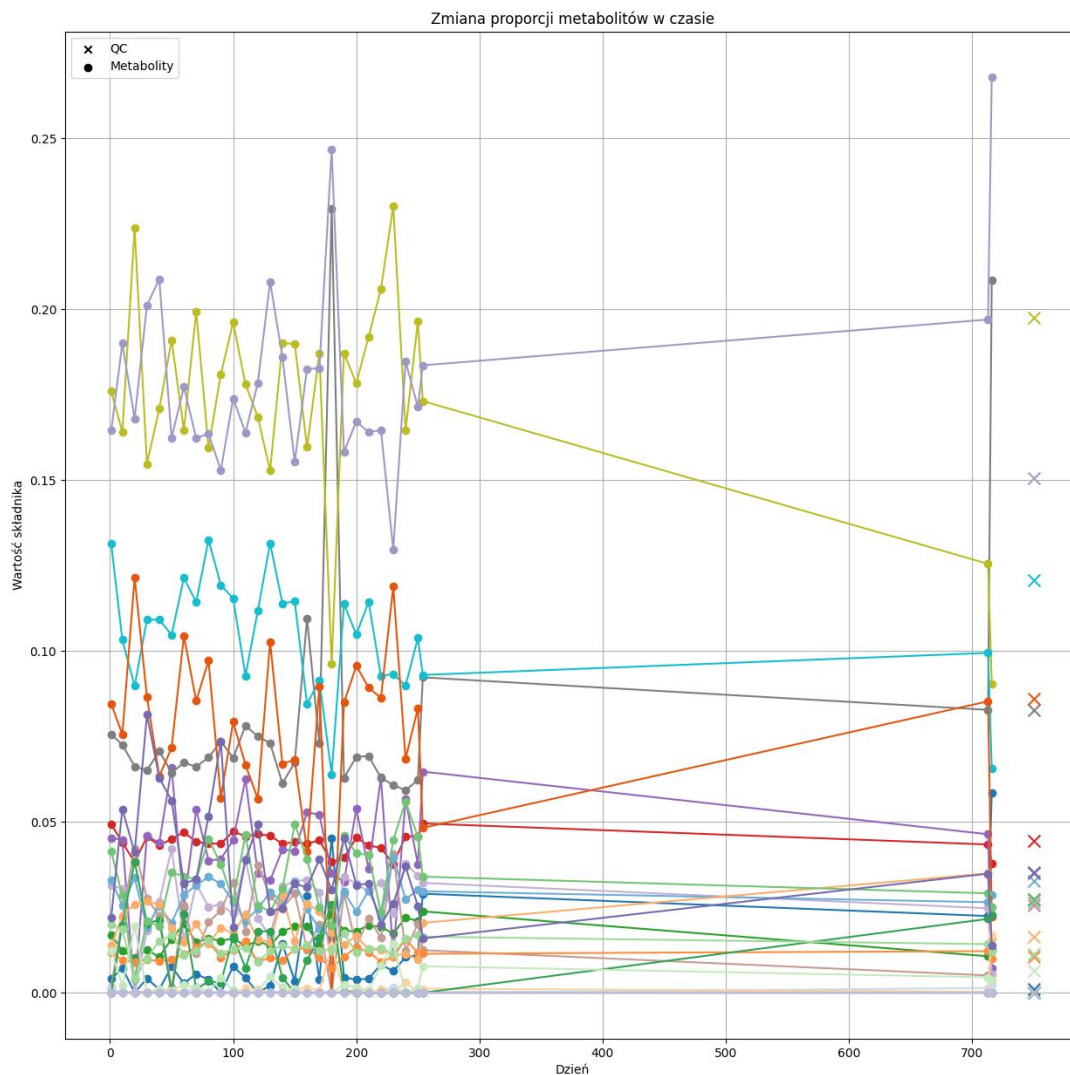


**Fig. 5** Przykładowe widmo modelowe i eksperymentalne dla danych z dnia 90 uzyskane za pomocą solvera Mosek i parametrów  $\kappa_{\text{mixture}}=0.30$ ,  $\kappa_{\text{compounds}}=0.22$ . Oś X przedstawia wartości ppm shift, oś Y intensywności. Na czerwono zaznaczono widmo modelowe utworzone przez przemnożenie widm referencyjnych przez wynikowe proporcje, a na niebiesko widmo QC.



**Fig. 6** Widmo modelowe i eksperymentalne dla danych z dnia 90 przybliżone na osi X do mniejszego zakresu. Dane uzyskane za pomocą solvera Mosek i parametrów  $\kappa_{\text{mixture}}=0.30$ ,  $\kappa_{\text{compounds}}=0.22$ . Oś X przedstawia wartości ppm shift, oś Y intensywności. Na czerwono zaznaczono widmo modelowe utworzone przez przemnożenie widm referencyjnych przez wynikowe proporcje, a na niebiesko widmo QC.

Wykonano także analizę zmiany proporcji poszczególnych metabolitów w czasie (fig. 7). Widać na niej, że substancje zmieniają swój udział w moczu pacjenta. Nie jest to jednoznaczna zmiana w stronę wartości QC, poza pojedynczymi wyjątkami, substancje utrzymują się raczej na stałym poziomie. Może to świadczyć o tym, że na skład ludzkiego moczu największy wpływ mają jednak osobnicze różnice, a nie środowisko.



**Fig. 7** Zmiany proporcji poszczególnych metabolitów w czasie. Kropkami zaznaczone są poszczególne proporcje metabolitów w kolejnych dniach, a symbolem x wartości dla próbki QC. Na osi X są kolejne dni, a na osi Y proporcje.

- Nowa biblioteka referencyjna

Metabolity z drugiej biblioteki referencyjnej zostały przygotowane w taki sam sposób jak przy pierwszej bibliotece. Do porównania dwóch bibliotek użyty przykładowych danych: próbek poprzanych dnia 30, 170 i QC. Wyniki pod względem proporcji sygnału i szumu wyszły lepsze (tab. 3). Zarówno jest mniej szumu jak i więcej sygnału. Patrząc na wykresy widm modelowych (fig. 8) widzimy kilka podobieństw. Na przykład występują np. ‘grupy’ pików w okolicach ok 7 do 9 i 1 do 5 z przerwą około 6 ppm, ale poza tym kształty uzyskanych widm są inne. Zmiana biblioteki zmienia wyniki, więc istotnym jest podjęcie dobrej decyzji podczas gromadzenia danych referencyjnych.

	stara biblioteka			nowa biblioteka		
	dzień 30	dzień 170	QC	dzień 30	dzień 170	QC
Signal proportion in mixture's spectrum	0.96	0.93	0.94	0.99	0.98	0.98
Noise proportion in mixture's spectrum	0.04	0.066	0.063	0.011	0.017	0.017
Total explanation	1.0	1.0	1.0	1.0	1.0	1.0
Noise proportion in combination of components' spectra	0.03	0.03	0.015	0.05	0.08	0.07

**Tabela 3.** Wyniki proporcji sygnału, szumu i całkowitego wyjaśnienia spektrów eksperymentalnych dla przykładowych próbek dla 2 różnych bibliotek referencyjnych.

#### 4. Dyskusja

Uzyskane przez nas wyniki analizy NMR nie są jednoznaczne. Zaobserwowano pewne nieścisłości przy tworzeniu widm modelowych, którym należy się dokładniej przyjrzeć. Wnioskujemy również, że największy wpływ na skład metaboliczny moczu mają różnice osobnicze pomiędzy organizmami. Nie jest to jednak wniosek jednoznaczny, gdyż wyniki mogą nie być dokładne. Należałoby także wykonać testy statystyczne istotności zmian proporcji poszczególnych metabolitów. Poza tym wyniki różnią się w zależności od zastosowanej biblioteki referencyjnej. Z tego względu kluczowe jest odpowiednie dobranie i przygotowanie biblioteki referencyjnej.

Aby móc sformułować bardziej wiarygodne wnioski, należałoby powtórzyć analizę na większej liczbie pacjentów, a najlepiej także kilkakrotnie powtórzyć cały proces badawczy, w tym zbieranie próbek. Przed kolejną analizą warto również zidentyfikować i wyeliminować problemy związane z generowaniem wykresów. Należy także prawidłowo dobrać bibliotekę referencyjną.



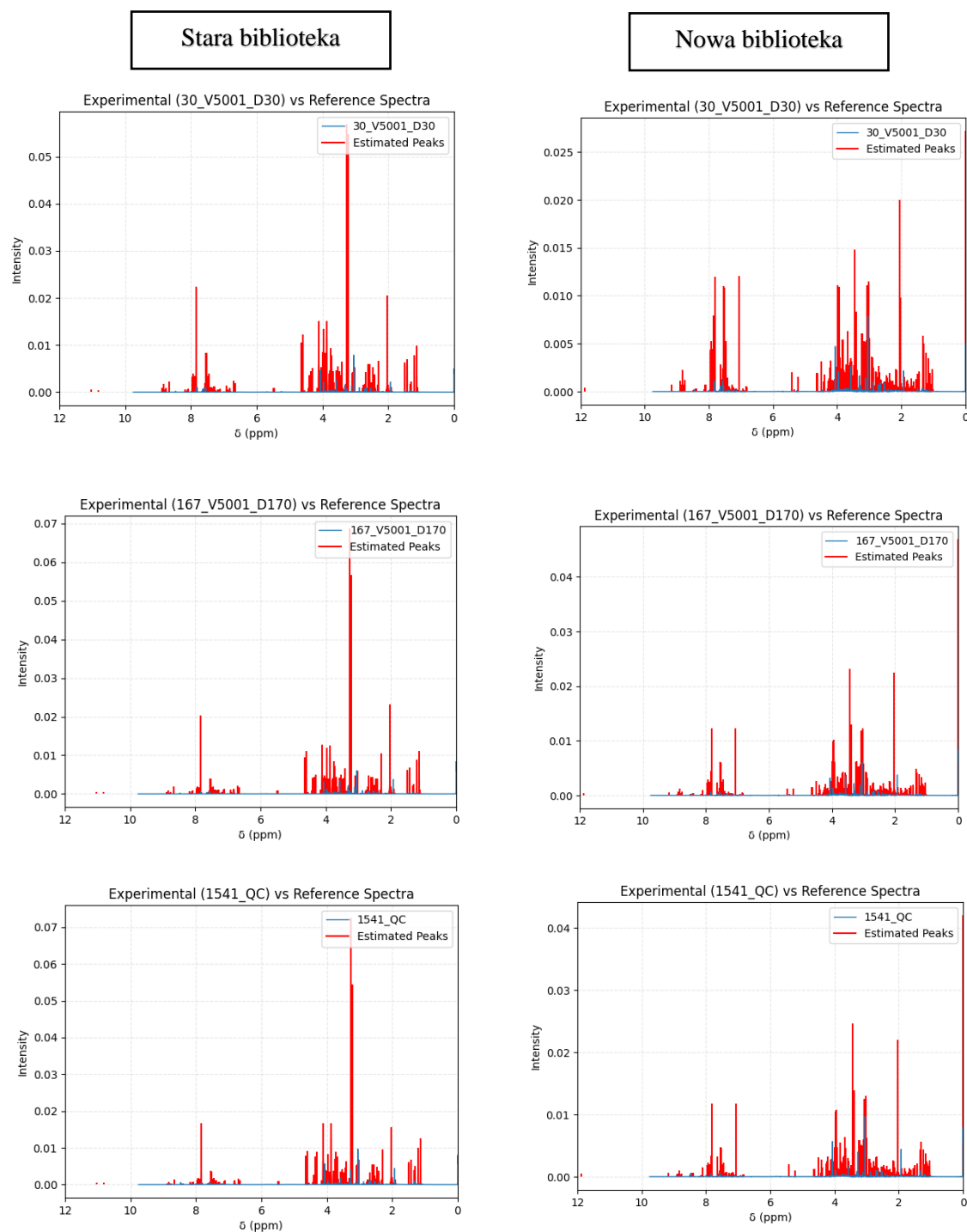


Fig. 8 Porównanie dwóch bibliotek referencyjnych. Po lewej widma uzyskane starą biblioteką, a po prawej nową. Widma są z próbek z dni 30, 170 i QC. Dane uzyskane za pomocą solvera Mosek i parametrów  $\kappa_{\text{mixture}}=0.30$ ,  $\kappa_{\text{compounds}}=0.22$ . Oś X przedstawia wartości ppm shift, oś Y intensywności. Na czerwono zaznaczono widmo modelowe utworzone przez przemnożenie widm referencyjnych przez wynikowe proporcje, a na niebiesko widmo QC.

## 5. Bibliografia

- [1] Eills J, Budker D, Cavagnero S, Chekmenev EY, Elliott SJ, Jannin S, Lesage A, Matysik J, Meersmann T, Prisner T, Reimer JA, Yang H, Koptug IV. Spin Hyperpolarization in Modern Magnetic Resonance. *Chem Rev.* 2023 Feb 22;123(4):1417-1551. doi: 10.1021/acs.chemrev.2c00534. Epub 2023 Jan 26. PMID: 36701528; PMCID: PMC9951229.
- [2] G.A. Nagana Gowda, Wentao Zhu, Daniel Raftery, NMR-based metabolomics: Where are we now and where are we going?, *Progress in Nuclear Magnetic Resonance Spectroscopy*, Volumes 150–151, 2025,101564, ISSN 0079-6565, <https://doi.org/10.1016/j.pnmrs.2025.101564>
- [3] Bo Wang, Aaron M. Goodpaster, Michael A. Kennedy, Coefficient of variation, signal-to-noise ratio, and effects of normalization in validation of biomarkers from NMR-based metabonomics studies, *Chemometrics and Intelligent Laboratory Systems*, Volume 128, 2013, Pages 9-16, ISSN 0169-7439, <https://doi.org/10.1016/j.chemolab.2013.07.007>.
- [4] Emwas AH, Roy R, McKay RT, Tenori L, Saccenti E, Gowda GAN, Raftery D, Alahmari F, Jaremko L, Jaremko M, Wishart DS. NMR Spectroscopy for Metabolomics Research. *Metabolites*. 2019 Jun 27;9(7):123. doi: 10.3390/metabo9070123. PMID: 31252628; PMCID: PMC6680826.
- [5] Bouatra S, Aziat F, Mandal R, Guo AC, Wilson MR, Knox C, Bjorndahl TC, Krishnamurthy R, Saleem F, Liu P, Dame ZT, Poelzer J, Huynh J, Yallou FS, Psychogios N, Dong E, Bogumil R, Roehring C, Wishart DS. The human urine metabolome. *PLoS One*. 2013 Sep 4;8(9):e73076. doi: 10.1371/journal.pone.0073076. PMID: 24023812; PMCID: PMC3762851.
- [6] Pauli GF, Chen SN, Lankin DC, Bisson J, Case RJ, Chadwick LR, Gödecke T, Inui T, Kronic A, Jaki BU, McAlpine JB, Mo S, Napolitano JG, Orjala J, Lehtivarjo J, Korhonen SP, Niemitz M. Essential parameters for structural analysis and dereplication by (1)H NMR spectroscopy. *J Nat Prod*. 2014 Jun 27;77(6):1473-87. doi: 10.1021/np5002384. Epub 2014 Jun 4. PMID: 24895010; PMCID: PMC4076039.
- [7] Domżał, B.; Nawrocka, E. K.; Gołowicz, D.; Ciach, M. A.; Miasojedow, B.; Kazimierczuk, K.; Gambin, A. Magnetstein: An Open-Source Tool for Quantitative NMR Mixture Analysis Robust to Low Resolution, Distorted Lineshapes, and Peak Shifts. *Analytical Chemistry* 2023, 96, 188–196.