

mgr Jan Grzegorz Stanek  
Pracownia Oddziaływań Międzycząsteczkowych  
Wydział Chemii UW

Warszawa, 8 grudnia 2014 r.

Autoreferat rozprawy doktorskiej pt.:  
**„Novel data processing methods for randomly sampled  
multi-dimensional NMR experiments”**

(tytuł w jęz. pol: “Nowe metody przetwarzania losowo próbkowanych wielowymiarowych  
eksperymentów NMR”

Promotor: prof. dr hab. Wiktor Koźmiński  
Wydział Chemii Uniwersytetu Warszawskiego

Spektroskopia jądrowego rezonansu magnetycznego (NMR) jest, obok krytalografii rentgenowskiej, wiodącą metodą analizy strukturalnej biomolekuł takich jak białka i kwasy rybo- i deoksyrybonukleinowe. Różnorodne metody NMR pozwalają na poznanie z *rozdzielczością atomową* nie tylko cech strukturalnych, ale również dynamiki w skali od sekund do pikosekund, charakterystyki oddziaływań międzydomenowych lub białko-RNA czy stałych dysocjacji kompleksów z ligandami.

Obecnie można wskazać na dwie główne bariery zastosowań NMR w odniesieniu do makromolekuł o znaczeniu biologicznym: niską czułość metody oraz nałożenie sygnałów widmowych. Intensywny postęp m.in. w konstrukcji magnesów NMR oraz zastosowanie sond kriogenicznych doprowadziły do znacznego wzrostu czułości, pozwalającego na badanie biomolekuł w stężeniach od ok. 100 nmol/l oraz o coraz większych rozmiarach (standardowo do ok. 25 kDa). Drugą z wymienionych trudności, tj. zatłoczenie sygnałów, rozwiązuje się poprzez selektywne bądź nieselektywne znakowanie izotopami  $^{13}\text{C}$ ,  $^{15}\text{N}$  i  $^2\text{H}$ , w połączeniu z wielowymiarową heterojądrową spektroskopią korelacyjną. W metodach tych dzięki przeniesieniu magnetyzacji na heterojądro ( $^{13}\text{C}$ ,  $^{15}\text{N}$  bądź  $^{31}\text{P}$ ) oraz ewolucji jego przesunięcia chemicznego, dokonuje się rozdziału nałożonych rezonansów bez zwiększenia ich ogólnej liczby.

Wielowymiarowe techniki NMR napotykać jednak zasadniczą trudność z uwagi na teoremat o próbkowaniu (Nyquista-Shannona), odnoszący się do pośredniego próbkowania ewolucji wzbudzanych koherencji. Długie czasy relaksacji podłużnej dla makromolekuł (~1 s) ograniczają szybkość detekcji pośredniej do kilku próbek/min. Z obu wymienionych powodów standardowa metodologia NMR ogranicza się do metod co najwyżej trójwymiarowych, w których – z uwagi na ograniczony czas pomiarowy – często konieczne jest poświęcenie

rozdzielczości w wymiarach pośrednich. Niemożliwe jest także wykorzystanie wysokiego pola  $B_0$  do poprawy rozdzielczości widmowej.

Rozwiązaniem problemu próbkowania wielowymiarowych eksperymentów NMR jest losowe próbkowanie oszczędne. Pozwala ono na zwężenie linii aż do ich naturalnej szerokości oraz optymalizację czułości przy próbkowaniu. Jednak pominięcie akwizycji większości sygnału niesie konsekwencje w postaci *artefaktów próbkowania*, tj. dodatkowego deterministycznego szumu. Zakłócenia te obniżają czułość wysokowymiarowych eksperymentów NMR i w konsekwencji niepotrzebnie wydłużają pomiary.

W mojej pracy przyjąłem za cel stworzenie nowego algorytmu przetwarzania oszczędnie próbkowanych sygnałów, skutecznie eliminującego *artefakty próbkowania*. Metodę oparto na transformacji Fouriera uzupełnionej zerami, a także na znanych od lat 60-tych iteracyjnych algorytmach CLEAN oraz *iterative soft thresholding*.

Opracowany *algorytm rozdziału sygnałów* (SSA) wprowadza szereg udoskonaleń w stosunku do CLEAN, m. in. możliwość automatycznego dopasowania sygnału funkcjami analitycznymi, iteracyjne udokładnianie modeli sygnału i spekulacyjne poszukiwanie sygnałów blisko krytycznego stosunku sygnału do szumu. Przeprowadzone symulacje komputerowe jak również modelowe dane eksperymentalne potwierdziły, iż nowy algorytm skuteczniej usuwa artefakty niż CLEAN, oraz daje konkurencyjne wyniki do innych popularnych metod przetwarzania, m.in. maksymalizacji entropii (MaxEnt) czy rozkładu wielowymiarowego (MDD). Prostota stworzonego algorytmu nie pozwala w ogólności na przewyższenie dobrze ugruntowanych matematycznie metod optymalizacyjnych, np. *compressed sensing* (CS). Jednak jest ona konieczna do szybkiego przetwarzania wielowymiarowych widm o wysokiej rozdzielczości, w których metody CS czy MaxEnt nie mogą w praktyce być zastosowane.

Szczególne uwagi zwrócono na wydajność obliczeń i możliwości ich zrównoleżenia. Implementacja algorytmu w języku C++ wykorzystuje wystandaryzowane biblioteki algebry liniowej (BLAS) oraz OpenMP (współbieżność). Zużycie pamięci RAM przy przetwarzaniu widm 4D i 5D ograniczono przez podział widm na bloki o stałej wielkości, ich selekcję i kompresję. Zoptymalizowano obliczanie sum Fouriera dla rzadkich wielowymiarowych danych, wykorzystując zmienność gęstości danych na poszczególnych etapach transformacji. Aby umożliwić rekonstrukcję wybranych fragmentów widma wielowymiarowego lub też dowolnych dwuwymiarowych przecięć zastosowano sekwencję transformacji w przód/wstecz, w której unika się dużych rozmiarów danych pośrednich. Stworzony program został udostępniony w 2010 r. na stronie internetowej grupy prof. Koźmińskiego (bezpłatnie do celów akademickich) i jest wykorzystywany w kilku zagranicznych ośrodkach naukowych.

Istotnym aspektem prezentowanej pracy są nowe zastosowania próbkowania losowego w połączeniu z efektywnym usuwaniem artefaktów. Pokazano szereg aplikacji do czterowymiarowych podwójnie edytowanych ( $^{13}\text{C}$  lub/i  $^{15}\text{N}$ ) widm jądrowego efektu Overhausera (NOESY), jak również widm TOCSY do przypisań sygnałów w łańcuchach bocznych białek i pierścieniach rybozy RNA. Szczególnym zastosowaniem okazały się ilościowe pomiary szybkości relaksacji wzajemnie skorelowanej (CCR) dla koherencji wielokwantowych w łańcuchu głównym białek. Dzięki ultra wysokiej rozdzielczości widmowej, wysokiej wymiarowości (4D) i tłumieniu artefaktów możliwe było po raz pierwszy wyznaczenie tych wielkości dla białek częściowo nieustrukturyzowanych (ang. *intrinsically disordered proteins*, *IDP*), stanowiących wyzwanie dla NMR z powodu szczególnie dużego zatłoczenia widm. Pomiary CCR stanowią nowe narzędzie do charakteryzacji lokalnej geometrii łańcucha polipeptydowego dla dominujących konformerów IDP w roztworze.

Ponadto wykazano użyteczność cztero i pięciowymiarowych widm TOCSY, uzyskanych metodą SSA, do przypisania sygnałów w małych i dużych białkach zwiniętych. Pomiary przeprowadzono dla standardowych próbek białek: ubikwityny (86 aa) i MBP (370 aa) oraz – we współpracy z ośrodkami z Polski, Holandii, Słowenii, Szwajcarii i Francji – dla dopiero badanych białek LipaseA (181 aa), HDN (114 aa), EMAP-II (169 aa) i S100A1 (2x93 aa).

Wykazano, iż w przypadku trój- i czterowymiarowych widm NOESY efektywne tłumienie artefaktów jest konieczne dla zachowania zawartej w nich informacji strukturalnej. W tym przypadku osiągnięto redukcję efektywnego szumu o max. 98% (tj. max. 50-krotną). W zależności od czułości eksperymentu (więc także wielkości i stężenia białka), pozwoliło to na całkowitą lub niemal zupełną eliminację artefaktów z widm. Pokazano również możliwości rejestracji bezdiagonalnych asymetrycznych widm NOESY, w których brakuje dominującego źródła artefaktów. Techniki te znajdują zastosowanie w badaniach strukturalnych białek ( $^{15}\text{N}$ ,  $^{13}\text{C}$ -, czy  $^{13}\text{C}(\text{aro})$ ,  $^{13}\text{C}(\text{ali})$ - edytowane NOESY) oraz do przypisania sekwencyjnego RNA w oparciu o kontakty H1'-H6/H8 ( $^{13}\text{C}(\text{ribo})$ ,  $^{13}\text{C}(\text{aro})$ -edytowane NOESY).

Podsumowując, stworzono wszechstronny algorytm przetwarzania sygnałów z losowo próbkowanych eksperymentów NMR. Pozwala on na zachowanie czułości w wysokorozdzielczych wymiarowych technikach NMR stosowanych dla białek i kwasów nukleinowych. Wyniki przeprowadzonych badań opublikowano w czasopismach *Journal of Biomolecular NMR* (4), *Journal of Magnetic Resonance* (1), *Journal of Biomolecular NMR Assignments* (2) i *Angewandte Chemie-International Edition* (1).