



Wrocław, 5.03.2026

Recenzja rozprawy doktorskiej Pana mgr Sevastianosa Korsaka pt. "Multi-scale Computational Modeling of Chromatin"

Obecnie wiemy, że aby w pełni zrozumieć zjawiska związane z organizacją i ekspresją informacji genetycznej u człowieka, nie wystarcza już sama znajomość jednowymiarowej sekwencji DNA oraz zestawu białek wiążących kwasy nukleinowe. Kluczem do poznania tych procesów jest wiedza o trójwymiarowej (3D) strukturze chromatyny, która ma decydujące znaczenie dla mechanizmów różnicowania komórek, odpowiedzi na bodźce środowiskowe oraz patogenezy wielu chorób. Co więcej, chromatyna nie jest statyczna – jej architektura jest wysoce dynamiczna i zmienia się w zależności od typu komórki i fazy cyklu komórkowego. Właśnie ta duża złożoność i dynamika sprawiają, że nie jest łatwo badać i zrozumieć te zjawiska. Złożone dane przestrzenne generowane przez nowoczesne techniki eksperymentalne są w dużej mierze niemożliwe do zinterpretowania bez odpowiedniego aparatu matematycznego i informatycznego. Dlatego też konieczne stało się opracowywanie nowych, wydajnych metod komputerowych, które pozwolą na precyzyjne modelowanie, symulowanie i wizualizację trójwymiarowej struktury chromatyny. Dlatego bardzo słusznie ambitnym przedmiotem pracy doktorskiej Pana mgra Sevastianosa Korsaka stało się opracowanie nowatorskich metod komputerowego modelowania trójwymiarowej struktury chromatyny w czasie i przestrzeni.

Rozprawa doktorska została napisana w języku angielskim ze streszczeniem w języku polskim. Zawiera: Abstract, Streszczenie, Podziękowania, Spis treści, Wstęp, Przegląd opublikowanych prac, Wnioski i perspektywy dalszych badań oraz załączone publikacje. Nie jest to typowy układ, jednak przegląd opublikowanych prac można traktować jako Materiały i metody, Wyniki i Dyskusję.

We Wstępie w bardzo przystępny i prosty sposób doktorant scharakteryzował matematyczny opis chromatyny oraz konieczność symulacji komputerowych, ponieważ wyliczenia analityczne nie są praktyczne obliczeniowo. Pan mgr Korsak wyjaśnił też ważne terminy stosowane w modelowaniu, co jest konieczne, aby precyzyjnie rozumieć opisywane zagadnienia. Bardzo pomocne są porównania opisywanych terminów do zjawisk z dziedziny matematyki, astronomii i fizyki. Przy okazji opisu wieloskalowości przedstawiono najnowsze poglądy na organizację chromatyny uwzględniając poziom nukleosomów, pętli i domen (TAD), kompartmentów A i B, regionów chromosomów oraz całego genomu i interakcji z laminami. Uważam, że doktorant bardzo ładnie połączył opis matematyczny i biologiczny chromatyny reprezentowanej jako polimer. Pan Sevastianos Korsak słusznie zauważył, że fraktalny opis chromatyny, mimo zalet obliczeniowych, może nie być odpowiedni ze względu na różnice i złożoność chromatyny na różnych poziomach organizacji.

Interesująca jest część dotycząca losowości, w której doktorant stwierdził, że ze względu na nieliniowy charakter oddziaływań chromatyny oraz dużą wymiarowość układu, to współdziałanie porządku i losowości może prowadzić do wysoce nieprzewidywalnych zachowań. Nie do końca się zgadzam ze stwierdzeniem, że losowe łączenie się i odłączanie białek z chromatyną są głównie zjawiskami losowymi. Może bezpieczniej jest napisać, że wydaje nam się, że są losowe dopóki nie odkryjemy potencjalnych prawidłowości. Dobrze byłoby także wyjaśnić, co to są kompleksy białek Smc przy pierwszym użyciu. Przy opisie entropii Gibbsa można by było wspomnieć dla porównania też o entropii Shannona. Ważne jest poruszenie kwestii stanu początkowego w modelowaniu oraz przejść i rozdziałów fazowych.

Doktorant dobrze wyjaśnił metodologię modelowania chromatyny i wie, że nie jest to proste z powodu wielu lokalnych optimum energii. Należy zwrócić uwagę, że Pan mgr Korsak zdaje sobie także sprawę z niebezpieczeństw i ograniczeń związanych ze stosowanymi algorytmami i metodami obliczeniowymi w tym optymalizacją oraz symulacjami deterministycznymi i stochastycznymi. Dobrze opracowany jest także rozdział dotyczący analiz eksperymentalnych: Chip-Seq, ATAC-Seq i metod wielowymiarowych (Hi-C, ChIA-PET, ChIA-Drop i Hi-ChIP) oraz obrazowania (FISH).

Opisy wyraźnie pokazują, że doktorant wykazuje się dużą wiedzą z różnych dziedzin oraz dogłębnie rozumie różne zagadnienia i ograniczenia metodologiczne. Świadczy o tym także rozdział, w którym przedstawiono wyzwania związane z modelowaniem chromatyny. Jest on także dobrym uzasadnieniem badań opisanych w tej pracy doktorskiej.

Cele pracy doktorskiej, czyli wieloskalowe modelowanie chromatyny, analiza dynamiki i zrozumienie stresu replikacyjnego, zostały jasno sformułowane i uzasadnione w osobnym rozdziale. Wszystkie zostały właściwie zrealizowane. Pozostałe rozdziały charakteryzują trzy oryginalne narzędzia symulacyjne: LoopSage, RepliSage i MultiMM, pozwalające zrozumieć, jak DNA zachowuje się w jądrze komórkowym w czasie i przestrzeni. Zostały one opublikowane w pięciu załączonych pracach opublikowanych w prestiżowych czasopismach.

Narzędzie LoopSage jest stochastycznym modelem fizycznym wykorzystującym algorytm Monte Carlo i funkcję Hamiltona opisującą energię układu, a także uwzględnia dynamikę molekularną. Model opiera się na danych eksperymentalnych dotyczące położenia białek biorących udział w tworzeniu pętli oraz tworzących kompleksy SMC (kohezyny i kondensyny) i CTCF. Głównym zadaniem LoopSage jest symulacja mechanizmu ekstruzji (eksponowania lub wytlaczania) pętli w celu wygenerowania realistycznych struktur chromatyny, które można porównać z wynikami eksperymentów 3C (np. mapami interakcji).

Najważniejszym osiągnięciem w tym modelu jest zmiana sposobu myślenia o symulowaniu procesów biologicznych. Zamiast sztywnych reguł programistycznych doktorant zastosował prawa fizyki. Wcześniejsze modele procesu wytlaczania pętli polegały na ręcznym wpisywaniu do kodu wielu reguł i prawdopodobieństw oddziaływania białek z DNA. Natomiast opracowany model wprowadza fizyczne pojęcie energii (Hamiltonian). Traktuje system biologiczny jak układ fizyczny, który dąży do minimalizacji swojej energii. Funkcja energii w modelu nagradza wydłużanie pętli i zatrzymywanie się na odpowiednio skierowanych białkach granicznych, a karze za "zderzenia" białek ze sobą. Dzięki temu model jest znacznie prostszy i bardziej elastyczny.

Ponieważ model opiera się na fizyce statystycznej i algorytmie Metropolis, pojawia się w nim parametr temperatury oznaczający stopień uporządkowania oraz tendencje białek tworzących pętle do przyłączania i odłączania względem DNA. LoopSage jest modelem dwuetapowym. Najpierw symuluje stochastyczny ruch białek wzdłuż nici DNA. Następnie, używając tych jednowymiarowych informacji, etap dynamiki molekularnej buduje trójwymiarowy model polimeru, w którym utworzone pętle działają jak elastyczne sprężyny. To pozwala w naturalny sposób przejść od mechanizmów molekularnych do pełnej architektury 3D jądra komórkowego. Model ten wyjaśnia także charakterystyczne wzory przypominające pasma w danych z eksperymentów biologicznych (np. Hi-C). Ponieważ był on w stanie je wygenerować, to potwierdził, że powstają one w wyniku tzw. jednostronnego wytlaczania pętli – sytuacji, w której jeden koniec pętli DNA jest zakotwiczony w miejscu, a drugi jest aktywnie przesuwany przez białko motoryczne.

Kolejne narzędzie RepliSage jest bardziej zaawansowane i wykorzystuje dane z eksperymentów 3C, jak poprzedni model, oraz dodatkowo czasy replikacji dla miejsc na chromosomie. Wprowadza modelowanie kompartmentów oraz uwzględnia wiele procesów biologicznych zachodzących w trakcie podziału komórek, jak nierównowagowe zjawiska związane z replikacją DNA i propagacją widełek replikacyjnych wymagających dostępu do chromatyny.

RepliSage składa się z trzech współpracujących ze sobą modułów, które razem tworzą pełen obraz cyklu komórki. Replikator symuluje fazę S (syntezy), rozpoczęcie replikacji DNA oraz prędkość poruszania się widełek replikacyjnych. Moduł stochastyczny jest oparty na algorytmie Metropolis, który łączy tworzenie pętli (z programu LoopSage) z przesuwanymi się widełkami replikacyjnymi oraz, co jest nowością, z propagacją stanów epigenetycznych (opisanych modelem Potts). Dynamika molekularna 3D buduje ostateczną strukturę przestrzenną. W odróżnieniu od pierwszego modelu, tutaj symulowane są dwie siostrzane nici DNA naraz, które ostatecznie muszą zostać od siebie oddzielone, jak to zachodzi przed podziałem komórki.

Dużą zaletą modelu jest to, że jest w stanie zasymulować zjawiska cyklu komórkowego oraz krytyczne zjawisko stresu replikacyjnego, co ma bezpośrednie i kluczowe znaczenie dla badań nad nowotworami, ponieważ podczas zatrzymania replikacji zwiększa się podatność DNA na mutacje. Model ten integrujący trzy kluczowe zjawiska biofizyczne: rozprzestrzenianie się stanu epigenetycznego powodującego kompartmentalizację, ekspozycję pętli oraz replikację DNA. Ocena modelu odbywa się przez porównanie z danymi eksperymentalnymi HiChIP w warunkach bez i ze stresem oraz z danymi Hi-C pojedynczych komórek. Ważne było zastosowanie funkcji Hamiltona jako sumy energii kinetycznej i potencjalnej układu chromatyny w jądrze komórkowym, co umożliwiło opracowanie realistycznego modelu obliczeniowego dynamiki zmian konformacji chromatyny w trakcie cyklu komórkowego i przestrzeni.

Model został opracowany dla podziału mitotycznego. W związku z tym mam pytanie: jak wyglądałby ten model dla podziału mejotycznego, w którym parują się chromosomy homologiczne? W modelu założono, że widełki replikacyjne są barierą dla tworzenia pętli. A czy nie można założyć, że jest odwrotnie? Jak by to wpłynęło na modelowanie? W perspektywach dalszych badań doktorant założył, żeby to w przyszłości zastosować. Dobrze byłoby również jawnie pokazać wyniki walidacji modelu, jeśli jest to możliwe.

MultiMM jest deterministycznym i statycznym algorytmem łączącym wszystkie skale struktury chromatyny, co jest ewidentną nowością. Narzędzie wymaga dane o interakcjach pętli z eksperymentów 3C, przypisania miejsc chromatyny do kompartmentów i położenia nukleosomów z eksperymentów ATAC-Seq. Narzędzie to pozwala na rekonstrukcję chromatyny od poziomu pojedynczych nukleosomów (o wielkości 11 nm), przez pętle, domeny, kompartmenty aż po całe chromosomy wypełniające jądro komórkowe (5 do 10 μm). Nukleosomy zostały zamodelowane jako sztywne obiekty dla przyspieszenia obliczeń, a pętle chromatynowe jako obiekty elastyczne dzięki harmonicznym potencjałom wiązań. Kompartmenty natomiast modelowano dzięki symulacji kopolimerów blokowych uwzględniając interakcje kompartmentu B z błoną jądra komórkowego. Model uwzględnia nawet wypychanie małych chromosomów w stronę jąderka. Aby skrócić czas obliczeń, MultiMM wprowadza inicjację obliczeń od struktury zwanej krzywą Hilberta, która jest fraktalem spełniającym dwa główne warunki DNA: jest ekstremalnie gęsto upakowany, ale jego nić nigdzie się nie przecina. Dzięki temu symulacja nie marnuje czasu na "zwijanie" polimeru, lecz od razu przechodzi do optymalizacji lokalnych interakcji, pozwalając na symulowanie układów składających się z milionów elementów.

MultiMM to wszechstronne narzędzie umożliwiające śledzenie zachowania się organizacji chromatyny w istotnych skalach przestrzennych i czasowych. Użytkownik może wybrać, czy wszystkie pętle mają mieć stałą długość, czy ich siła ma zależeć od intensywności sygnału z eksperymentu, a także skupić się w analizach na otoczeniu wybranych genów. Doktorant stwierdził, że jest to jego największe osiągnięcie. Zagrądam się z nim, chociaż dla mnie najbardziej interesujący jest RepliSage.

Należy zwrócić uwagę na rozbudowany rozdział pod koniec rozprawy zawierający dyskusję, wnioski i przyszłe badania. W nim doktorant przedstawił, w jaki sposób opisane modele realizują cele badawcze niniejszej pracy doktorskiej. Dogłębnie omówił innowacje wprowadzone przez te modele, a także ograniczenia, co świadczy o jego dużym obiektywizmie. Interesujące są perspektywy dotyczące przyszłych kierunków badawczych polepszające dotychczasowe metody na poziomie danych biologicznych, modelowym, algorytmicznym i obliczeniowym. Są one nawet bardzo szczegółowe, co oznacza, że doktorant jest dojrzałym naukowcem z konkretnymi planami. W szczególności omówił obiecującą rolę uczenia maszynowego w rozwiązywaniu aktualnych wyzwań związanych z modelowaniem chromatyny, takich jak optymalizacja doboru hiperparametrów oraz przewidywanie trójwymiarowej struktury genomu bezpośrednio na podstawie sekwencji DNA.

Praca jest napisana jasnym i poprawnym językiem oraz jest dobrze sformatowana. Przy opisie wymagań obliczeniowych dobrze byłoby przedstawić ile czasu, procesorów i pamięci wymagały obliczenia. Nie znalazłem wytłumaczenia Ubk we wzorze 1.28. Chciałbym zwrócić uwagę na oryginalne rysunki, które w przejrzysty sposób obrazują skomplikowane algorytmy.

Podsumowując praca doktorska dostarcza biologom i biofizykom potężnych narzędzi i modeli obliczeniowych, które wyjaśniają, jak struktura chromatyny zmienia się podczas cyklu komórkowego i jak reaguje na stres replikacyjny związany z chorobami nowotworowymi.

Warto podkreślić, że doktorant jest współautorem 9 publikacji w renomowanych czasopiśmie, prezentował wynik na w wielu konferencjach i uczestniczył w grantach. Ponadto nadzorował dwie prace magisterskie i zarządzał klastrem obliczeniowym. Doktorant wyjaśnił również swój udział w pracach, co nie pozostawia wątpliwości o jego dużym wkładzie.

Reasumując mogę stwierdzić, że doktorant włożył dużo trudu w opracowanie nowatorskich metod i przeprowadzone analizy, a przedstawione opisy wyników świadczą o jego dużej dojrzałości naukowej, wnikliwości i głębokim zrozumieniu zagadnienia. Nie mam zastrzeżeń do metodyki przeprowadzonych analiz. Tematyka pracy doktorskiej jest bardzo zasadna, ponieważ istnieje potrzeba tworzenia nowych metod i narzędzi modelujących struktury przestrzenne chromatyny w czasie i przestrzeni, co stało się przedmiotem rozprawy. Warto zaznaczyć, że zastosowane w pracy modele są nowe i zaawansowane, a uzyskane wyniki dowodzą, że wykorzystanie takich metod w biologii molekularnej prowadzi do wartościowych i interesujących odkryć. Uważam, że recenzowana rozprawa stanowi istotne i kompleksowe opracowanie komputerowych metod wieloskalowego trójwymiarowego modelowania struktury i funkcji chromatyny.

Uważam, więc, że przedstawiona do recenzji rozprawa doktorska spełnia wszystkie wymogi Ustawy o Stopniach Naukowych. Zgłaszam, zatem wniosek do Rady Naukowej Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Warszawskiej o uznanie rozprawy Pana mgra Sevastianosa Korsaka za odpowiadającą wymogom stawianym rozprawom doktorskim i o dopuszczenie doktoranta do dalszych etapów przewodu doktorskiego. Ze względu na bardzo dużą jakość uzyskanych wyników i ich bardzo dobre przedstawienie wnioskuję o wyróżnienie rozprawy.

Prof. dr hab. Paweł Mackiewicz

