

**RADA NAUKOWA DYSCYPLINY
INFORMATYKA TECHNICZNA I TELEKOMUNIKACJA POLITECHNIKI WARSZAWSKIEJ**

zaprasza na
OBRONĘ ROZPRAWY DOKTORSKIEJ

mgr Sevastianos Korsak

która odbędzie się w dniu **12 czerwca 2026 roku**, o godzinie **14:00** w trybie stacjonarnym

Temat rozprawy:

“Multi-scale Computational Modeling of Chromatin ”

Promotor: prof. dr hab. Dariusz Plewczyński – Politechnika Warszawska

Recenzenci: dr hab. inż. Aleksandra Gruca – Politechnika Śląska

prof. dr hab. Paweł Mackiewicz – Uniwersytet Wrocławski

dr hab. Witold Rudnicki – Uniwersytet w Białymstoku

Obrona odbędzie się stacjonarnie w Sali 40 (parter) Wydziału Matematyki i Nauk Informatycznych Politechniki Warszawskiej w Warszawie, ul Koszykowej 75.

Z rozprawą doktorską i recenzjami można zapoznać się w Czytelni Biblioteki Głównej Politechniki Warszawskiej, Warszawa, Plac Politechniki 1.

Streszczenie rozprawy doktorskiej i recenzje są zamieszczone na stronie internetowej: <https://www.bip.pw.edu.pl/Postepowania-w-sprawie-nadania-stopnia-naukowego/Doktoraty/Wszczete-po-30-kwietnia-2019-r/Rada-Naukowa-Dyscypliny-Informatyka-Techniczna-i-Telekomunikacja/mgr-Sevastianos-Korsak>

Przewodniczący Rady Naukowej Dyscypliny
Informatyka Techniczna i Telekomunikacja
Politechniki Warszawskiej
prof. dr hab. inż. Jarosław Arabas

Niniejsza praca stanowi wkład w dziedzinę modelowania komputerowego złożonych systemów biologicznych, jak również pełni rolę przystępnego wprowadzenia do symulacji dynamicznej struktury trójwymiarowej chromatyny w komórkach ludzkich. Zaczynając od podstawowych pojęć, czytelnik jest prowadzony przez kluczowe narzędzia matematyczne, na których opiera się modelowanie polimerów, w tym omówienie integratorów numerycznych, konstrukcji pól sił molekularnych oraz uzasadnienia stosowania symulacji stochastycznych w kontekście dynamiki struktury chromatynowej. Krytyczna dyskusja, podkreśla istniejące luki w modelowaniu chromatyny, zwracając uwagę na interdyscyplinarny charakter aktualnie wykorzystywanych modeli komputerowych oraz krytyczną rolę danych pochodzących z doświadczeń biologii molekularnej i genomiki w tworzeniu modeli trójwymiarowych oraz ich walidacji.

Doktorat rozpoczyna się od omówienia modelu LoopSage który jest komputerową implementacją biofizycznych obserwacji procesu ekstruzji pętli. W tej części pracy doktorskiej omówione zostają kluczowe biomolekuły, które biorą udział w formowaniu się pętli, takie jak kohezyna czy białko CTCF. Następnie wprowadzona zostaje metoda Monte Carlo oparta na biofizycznym Hamiltonianie, rygorystycznie zweryfikowana względem eksperymentalnych map interakcji pochodzących z doświadczeń typu 3C (HiC, HiChIP, ChIA-PET). Na tej podstawie opracowana została metoda RepliSage, która uwzględnia szereg procesów biologicznych zachodzących w trakcie podziału komórek ludzkich, tworząc bardziej kompleksowy i obliczeniowo wymagający model integrujący trzy kluczowe zjawiska biofizyczne: propagację stanu epigenetycznego prowadzącą do kompartmentalizacji, ekstruzję pętli oraz replikację DNA.

Istotnym wyzwaniem było opracowanie metody walidacji tak złożonego modelu komputerowego za pomocą obserwacji eksperymentalnych. Osiągnięto to przez porównanie przewidywań modelu komputerowego RepliSage z danymi eksperymentalnymi z doświadczeń HiChIP przeprowadzonych w laboratorium, w warunkach bez i ze stresem. Dodatkowo porównano wyniki symulacji z danymi Hi-C pojedynczych komórek. Porównanie symulacji biofizycznych oraz wyników doświadczeń pokazało jak integrować w rygorystycznym modelu komputerowym zarówno biologiczne spostrzeżenia dotyczące obserwowanego zjawiska, jak również biofizyczny opis oddziaływań molekularnych za pomocą funkcji Hamiltona reprezentującej sumę energii kinetycznej i potencjalnej układu chromatyny w jądrze komórkowym. Tylko realistyczne uwzględnienie zarówno biologii jak i fizyki układu umożliwiło opracowanie realistycznego modelu komputerowego dynamiki zmian konformacji trójwymiarowej chromatyny w trakcie cyklu komórkowego.

Na koniec w pracy przedstawiono nowatorski algorytm MultiMM, który jest wszechstronnym i wydajnym komputerowym modelem wielo-skalowym, który łączy organizację chromatyny od poziomu nukleosomów o rozmiarze 11 nm aż po całe chromosomy i pełną strukturę 3D jądra komórkowego zajmującą w przybliżeniu kulistą przestrzeń o średnicy od 5 do 10 μm . W tym podejściu nukleosomy reprezentowane są za pomocą sztywnych obiektów trójwymiarowych w ramach podejścia geometrycznego w celu zwiększenia szybkości obliczeń, zaś pętle chromatynowe symulowane są dzięki harmonicznym potencjałom wiązań. Kompartymenty natomiast modelowane są dzięki zastosowaniu symulacji kopolimerów blokowych oraz uwzględniając dodatkowe interakcje, np. między kompartmentem B a błoną jądra komórkowego. MultiMM dostarcza w ten sposób wszechstronne i przystępne narzędzie wielo-skalowego modelowania komputerowego dla biologów i biofizyków umożliwiające

śledzenie zachowania się architektury chromatyny we wszystkich istotnych skalach przestrzennych i czasowych.

Podsumowując, praca doktorska omawia wyzwania komputerowej reprezentacji genomu ludzkiego, wprowadza szereg innowacji do aktualnie wykorzystywanych modeli, stanowiąc istotny wkład w genomikę obliczeniową oraz wskazując na nowe możliwości, jakie otwierają się dla symulacji molekularnych złożonych systemów biologicznych, potencjalnie prowadząc do dalszych odkryć w biologii molekularnej. W szczególności zwrócono uwagę na ograniczenia obecnie stosowanych metod. Podkreślono, że podejścia hybrydowe i wielo-skalowe, łączące dynamikę molekularną z technikami uczenia maszynowego wykorzystywanymi w procesie optymalizacji hiperparametrów, mają potencjał, aby sprostać tym wyzwaniom i przyczynić się do dalszego rozwoju dziedziny modelowania komputerowego systemów biologicznych w skali całej komórki.

Gliwice, 04.04.2026

Recenzja rozprawy doktorskiej mgr-a Sevastianosa Korsaka

Multi-scale Computational Modelling of Chromatin

Ukończonej na Wydziale Matematyki i Nauk Informatycznych
Politechniki Warszawskiej

Pod opieką promotora Profesora Dariusza Plewczyńskiego

Tematyka i cel pracy, problem badawczy i jego znaczenie

Przedstawiona mi do recenzji rozprawa doktorska powstała na Wydziale Matematyki i Nauk Informatycznych Politechniki Warszawskiej pod kierunkiem Profesora Dariusza Plewczyńskiego.

Tematyka rozprawy obejmuje wykorzystanie metod komputerowych do wieloskalowego (*multi-scale*) modelowania trójwymiarowej struktury chromatyny. W swoich badaniach Doktorant koncentruje się na stworzeniu zestawu metod, które umożliwią przewidywanie struktury tej cząsteczki ze szczególnym uwzględnieniem interakcji między jej różnymi skalami organizacyjnymi (od poziomu nukleosomów po całe terytoria chromosomowe). Równocześnie część jego badań obejmuje nie tylko samo modelowanie struktury chromatyny, ale także jej dynamiki w czasie całego cyklu komórkowego.

Przestrzenna organizacja chromatyny stanowi obecnie jeden z intensywnie rozwijanych obszarów współczesnej biologii molekularnej. Postęp technologiczny, obejmujący zarówno nowoczesne metody sekwencjonowania, jak i zaawansowane techniki obrazowania, umożliwia coraz dokładniejsze poznawanie trójwymiarowej architektury genomu. Zagadnienie to ma fundamentalne znaczenie dla zrozumienia mechanizmów regulujących funkcjonowanie komórki. Szczególną rolę odgrywają tu pętle chromatynowe, które pośredniczą w kontaktach pomiędzy odległymi regionami DNA, w tym promotorami i enhancerami, wpływając tym samym na kontrolę ekspresji genów. Architektura chromatyny pozostaje również ściśle związana z przebiegiem replikacji materiału genetycznego oraz procesami naprawy DNA, a jej zaburzenia mogą sprzyjać rozwojowi wielu schorzeń, zwłaszcza chorób nowotworowych. Z tego względu badania nad strukturą przestrzenną chromatyny mają istotne znaczenie nie tylko poznawcze, lecz także aplikacyjne, ponieważ mogą przyczynić się do lepszego zrozumienia patogenezы chorób oraz wspierać poszukiwanie nowych strategii terapeutycznych.

W niniejszej rozprawie Doktorant podejmuje tematykę badawczą związaną z wielkoskalowym modelowaniem struktury chromatyny. Autor nie ogranicza swoich analiz do pojedynczego poziomu organizacji, lecz podejmuje próbę ujęcia wzajemnych zależności pomiędzy różnymi skalami jej struktury w ramach spójnego modelu. W tym celu przeprowadzone badania obejmują niemal wszystkie istotne poziomy organizacji chromatyny, w tym symulacje geometryczne nukleosomów, stochastyczne modele ekstruzji pętli, dynamikę kompartmentalizacji oraz oddziaływania chromatyny z laminą jądrową. Celem rozprawy nie jest wyłącznie odtworzenie struktury chromatyny, lecz również poznanie jej dynamiki, przemian fazowych zachodzących pod wpływem kluczowych parametrów, a przede wszystkim związków pomiędzy zmianami strukturalnymi a podstawowym procesami, takimi jak cykl komórkowy oraz związany z nim stres replikacyjny, który jest procesem biologicznym ściśle powiązanego z rozwojem chorób nowotworowych

Oceniając tematykę badawczą podjętą w rozprawie, należy uznać, że dotyczy ona problemu istotnego i aktualnego, wyraźnie osadzonego w aktualnych kierunkach rozwoju bioinformatyki oraz biologii obliczeniowej.

Charakterystyka rozprawy

Przedstawiona praca doktorska składa się z trzech rozdziałów.

Pierwszy rozdział (*Introduction*) stanowi obszernie, interdyscyplinarne wprowadzenie do niezbędnych koncepcji z dziedziny informatyki, biologii i fizyki związanych z tematyką rozprawy. Autor wyjaśnia, jak matematycznie opisać chromatynę oraz szczegółowo omawia podstawowe terminy, takie jak właściwości wieloskalowe, fraktale, stochastyczność czy przejścia fazowe. Dodatkowo opisane są metody pozyskiwania danych eksperymentalnych oraz wyzwania i powszechnie stosowane podejścia w modelowaniu chromatyny. Na końcu tego rozdziału sformułowano motywację badawczą oraz główne cele rozprawy.

Rozdział drugi (*Overview of Published Research*) to przegląd opublikowanych badań. Jest to rozdział, w którym w zwięzły sposób Doktorant zaprezentował trzy kluczowe modele będące wynikiem badań prowadzonych w ramach doktoratu: LoopSage, RepliSage oraz MultiMM. Rozdział ten ma na celu wyjaśnienie oraz opis zastosowanych rozwiązań metodologicznych.

Pierwszą z opisanych metod jest **LoopSage** – metoda modelowania dynamiki procesu ekstruzji pętli chromatynowych. Zaproponowana metoda LoopSage jest dwuetapowa: w pierwszym etapie symulacji Monte Carlo algorytm bada przestrzeń prawdopodobieństwa, symulując trajektorie LEFs (*Loop Extrusion Factors*). Drugi etap metody to faza dynamiki molekularnej z wykorzystaniem narzędzia OpenMM, gdzie wykorzystuje się otrzymane z poprzedniego etapu parametry (pozycje LEFs). Tym, co w głównej mierze charakteryzuje tę metodę, jest odejście od tradycyjnego, sztywnego kodowania reguł biofizycznych w kodzie programu na rzecz stochastycznego modelu opartego na energii układu.

Druga opisana w tym rozdziale metoda, **RepliSage**, stanowi rozszerzenie metody LoopSage i służy do symulacji dynamiki chromatyny w trakcie całego cyklu komórkowego. Metoda ta działa w trzech etapach. Pierwszy moduł systemu RepliSage zbudowany jest na bazie opisanej wcześniej metody LoopSage i odpowiada za symulację replikacji DNA, skupiając się na losowej aktywacji miejsc inicjacji oraz postępującym

ruchu widełek replikacyjnych. Drugi etap obejmuje symulację stochastyczną łączącą ekstruzję pętli ze zmianami epigenetycznymi (z wykorzystaniem modelu Potts'a). Ostatni etap analizy, to moduł dynamiki molekularnej, który na podstawie wygenerowanych stanów tworzy końcowe, trójwymiarowe struktury chromatyny. Metoda RepliSage została zaprojektowana tak, aby móc skutecznie modelować warunki stresu replikacyjnego, który jest zjawiskiem ściśle powiązaniem z niestabilnością genomową i powstawaniem nowotworów.

Ostatnie opisane w rozdziale trzecim narzędzie to **MultiMM**. Framework MultiMM to deterministyczny, wieloskalowy model obliczeniowy służący do rekonstrukcji trójwymiarowej architektury przestrzennej chromatyny w skali genomowej. Stanowi on kompleksowe narzędzie, którego główną zaletą jest zdolność do integracji organizacji chromatyny na wszystkich istotnych biologicznie poziomach – od rozdzielczości pojedynczych nukleosomów, poprzez pętle chromatynowe i domeny TAD, po kompartmenty, terytoria chromosomowe oraz pełną przestrzeń jądra komórkowego. Model ten charakteryzuje się modułowością. Funkcja całkowitej energii układu może być elastycznie modyfikowana poprzez włączanie i wyłączenie poszczególnych komponentów sił, w zależności od zadanego problemu oraz dostępnych danych wejściowych. Składowe modelu obejmują m.in.: reprezentację pętli chromatynowych, modelowanie zjawiska kompartmentalizacji chromatyny przy użyciu potencjałów kopolimerów blokowych oraz wydajną, geometryczną reprezentację nukleosomów (w przeciwieństwie do kosztownych obliczeniowo symulacji atomowych). Dodatkowo, system MultiMM oferuje szereg opcji pozwalających na personalizację dynamiki, modelowanie wybranych fragmentów genomu oraz generowanie statystycznych zespołów struktur w celu zbadania zmienności i stabilności architektury chromatyny.

Ostatni trzeci rozdział (*Conclusions and Future Perspectives*) zwiera wnioski i przyszłe kierunki badawcze, w których można rozwijać zaprezentowane prace. Rozdział ten podsumowuje całość pracy, analizując w jaki sposób opracowane modele zrealizowały założone cele i jak przyczyniają się do rozwoju wiedzy o strukturze 3D genomu i jej powiązaniach z procesami biofizycznymi. Zawiera on również dyskusję na temat ograniczeń i niedoskonałości zaprezentowanych metod. Na koniec Autor zarysowuje perspektywy dla przyszłych kierunków badań, wskazując m.in. na ogromny potencjał wykorzystania uczenia maszynowego do przewidywania struktury chromatyny bezpośrednio z sekwencji DNA, a także podsumowuje swoje osiągnięcia naukowe, publikacje oraz udział w grantach badawczych.

Opinia o rozprawie

Przedstawione w rozprawie wyniki stanowią spójny i logicznie skonstruowany cykl badań poświęconych modelowaniu struktury chromatyny. Pierwszy rozdział pracy stanowi wprowadzenie do tematu, drugi opisuje opracowane metody, w trzecim Doktorant podejmuje dyskusję wskazując problemy i słabości istniejących metod oraz możliwe ich rozwiązania. Analizując opracowane w ramach Doktoratu metody szczególnie godne podkreślenia jest to, że przyjęta strategia badawcza realizowana jest konsekwentnie według zasady przejścia od szczegółu do ogółu – od modelowania pojedynczych elementów strukturalnych, poprzez metody modelowania uwzględniające cykl komórkowy chromatyny, aż po framework umożliwiając odwzorowanie organizacji chromatyny jednocześnie na wielu skalach przestrzennych.

Podejmowana problematyka wpisuje się w aktualne trendy badawcze z pogranicza biofizyki, bioinformatyki obliczeniowej oraz biologii strukturalnej genomu — dziedzin, które w ostatnich latach przeżywają dynamiczny rozwój, czego dowodem jest rosnąca liczba publikacji poświęconych trójwymiarowej organizacji chromatyny. Warto przy tym podkreślić, że postęp w tej dziedzinie ma charakter ściśle sprzężony: dynamiczny rozwój metod eksperymentalnych, takich jak techniki Hi-C, ChIP-seq, ATAC-seq czy obrazowanie FISH, dostarcza coraz bogatszych i bardziej precyzyjnych danych, które z kolei stymulują i umożliwiają tworzenie coraz bardziej zaawansowanych metod obliczeniowych. Rozprawa doskonale wpisuje się w tę dwukierunkową zależność, czerpiąc z najnowszych danych eksperymentalnych jako podstawy dla proponowanych modeli. Dopelnieniem całości jest krytyczna analiza metod walidacji modeli chromatyny, która ujawnia istotne słabości obecnie stosowanych podejść walidacyjnych i stanowi cenny wkład w metodologię dziedziny.

Bibliografia rozprawy jest obszerna i obejmuje 278 pozycji literaturowych, co świadczy o gruntownym zapoznaniu się Doktoranta z dorobkiem naukowym w omawianej dziedzinie. Pewne zdziwienie budzi jednak fakt, że zaledwie około 1/4 przywołanych źródeł stanowią publikacje z ostatnich pięciu lat (tj. z roku 2021 lub nowsze). Biorąc pod uwagę, że pole badawcze objęte rozprawą cechuje się wyjątkowo szybkim tempem rozwoju, wskazane byłoby szersze uwzględnienie najnowszego piśmiennictwa, które mogłoby wzbogacić dyskusję wyników oraz zapewnić pełniejsze osadzenie pracy w bieżącym stanie wiedzy.

Pozytywnie oceniam również fakt, że kod źródłowy wszystkich opracowanych metod i narzędzi obliczeniowych został udostępniony publicznie w repozytoriach serwisu GitHub. Praktyka ta, zgodna z postulatami otwartej nauki, znacząco podnosi wartość uzyskanych w ramach rozprawy wyników, umożliwiając innym badaczom weryfikację uzyskanych wyników oraz dalszy rozwój zaproponowanych rozwiązań.

Podsumowując, w pracy umiejętnie połączono modelowania fizycznego z podejściami obliczeniowymi — formalizmy mechaniki statystycznej i dynamiki molekularnej zostały z powodzeniem zintegrowane z nowoczesnymi technikami symulacyjnymi i analizą danych genomicznych, tworząc spójne ramy badawcze. Zastosowane metody badawcze są adekwatne do postawionych celów naukowych, a przyjęte w pracy podejścia do przetwarzania i analizy danych wskazują na dobrą znajomość przez Doktoranta nowoczesnych i efektywnych narzędzi stosowanych we współczesnej biologii obliczeniowej.

Uwagi krytyczne i dyskusyjne

Na wstępie niniejszej części recenzji chciałabym wyraźnie zaznaczyć, że problematyka podjęta przez Doktoranta należy do zagadnień o istotnym znaczeniu naukowym i znacznym stopniu trudności. Po szczegółowej analizie przedłożonej rozprawy nie stwierdzam obecności błędów merytorycznych o zasadniczym charakterze. Uwagi przedstawione poniżej formułuję w duchu dyskusji naukowej i nie stanowią one podstawy do obniżenia ogólnej, pozytywnej oceny pracy.

Pewną wątpliwość budzi niejednoznaczna forma redakcyjna rozprawy. Czytając odnosiłam wrażenie, że Doktorant nie podjął do końca jednoznacznej decyzji co do przyjętego modelu prezentacji — czy praca ma stanowić monografię w tradycyjnym ujęciu, czy też cykl powiązanych tematycznie artykułów naukowych. W efekcie rozprawa jest z jednej strony stosunkowo obszerna, z drugiej zaś w odniesieniu do szeregu przedstawionych metod brakuje pełnej informacji dotyczącej ich wydajności i skuteczności na rzeczywistych danych, oraz opisu tych danych. Kwestia walidacji modeli sygnalizowana jest jedynie skrótowo, ograniczając się w wielu miejscach do ogólnych komentarzy na temat trudności z nią związanych, bez przedstawienia konkretnych wyników ilościowych. Co więcej, w pracy nie zamieszczono bezpośrednich odnośników do artykułów lub ich konkretnych sekcji, w których wyniki te zostały opublikowane. Recenzent jest wprawdzie w stanie domyślić się, w których publikacjach należy szukać brakujących informacji, niemniej jednak rozprawa doktorska powinna zawierać tego rodzaju informacje podane wprost, a nie pozostawiać je do samodzielnego ustalenia przez czytelnika. Uważam, że jest to istotny minus pracy. Wskazane byłoby uzupełnienie odpowiednich fragmentów o jednoznaczne odesłania do powiązanych publikacji oraz tam, gdzie to możliwe o związku podsumowanie kluczowych wyników ilościowych.

Niedosyt budzi sposób przedstawienia celów badawczych we wstępie rozprawy. Wstęp zawiera jedynie bardzo ogólne stwierdzenie, że przedmiotem pracy jest modelowanie trójwymiarowej struktury chromatyny, bez precyzyjnego określenia szczegółowych celów naukowych i pytań badawczych, które Doktorant zamierzał podjąć. Informacje te pojawiają się dopiero w końcowej części rozdziału pierwszego, co utrudnia czytelnikowi — a także recenzentowi — właściwe osadzenie prezentowanych wyników w zamierzonym kontekście badawczym już na etapie lektury wstępnej. Zgodnie z przyjętymi standardami redakcji rozpraw doktorskich, cele pracy powinny być wyraźnie sformułowane we wstępie, tak aby od samego początku wyznaczały perspektywę dla oceny zawartości wszystkich kolejnych rozdziałów.

W sekcji *Limitations and Open Challenges* rozdziału trzeciego, Doktorant słusznie zauważa, że przedstawione w pracy metody wymagają określenia zestawu hiperparametrów, co w praktyce stanowi niebagatelne wyzwanie. Doktorant wspomina o możliwości zastosowania przeszukiwania siatki jako strategii ich doboru, jednocześnie jednak przyznaje, że brak jednoznacznych danych eksperymentalnych opisujących strukturę 3D chromatyny istotnie utrudnia ocenę jakości uzyskanych rozwiązań. Doceniam fakt, że w rozdziale trzecim Doktorant podejmuje dyskusję nad możliwymi podejściami do wyznaczania hiperparametrów i jest świadomy złożoności tego zagadnienia. Niemniej jednak chciałabym w tym miejscu zadać Doktorantowi pytanie: czy na podstawie dotychczasowych doświadczeń z opracowanymi modelami jest w stanie zaproponować potencjalnym użytkownikom konkretny, praktyczny „przepis” na dobór hiperparametrów? Innymi słowy — czy możliwe jest sformułowanie choćby ogólnych wytycznych lub heurystyk, które pozwoliłyby nowemu użytkownikowi na sensowne zainicjowanie procesu optymalizacji, bez konieczności przeprowadzania kosztownych obliczeń metodą prób i błędów lub implementacji nowych metod optymalizacji?

Podsumowanie

Pan mgr Sevastianos Korsak przedstawił rozprawę doktorską rozwiązującą aktualny problem naukowy, która przyczyni się do rozwoju reprezentowanej dyscypliny naukowej. Nie mam wątpliwości, że rozprawa zawiera oryginalne rozwiązanie problemu naukowego, a Doktorant wykazał, że zarówno posiada ogólną wiedzę teoretyczną w dyscyplinie informatyka techniczna i telekomunikacja jak i umiejętność prowadzenia pracy naukowej.

Doktorant jest współautorem 9 publikacji naukowych, z których jedna jest aktualnie dostępna wyłącznie na platformie *bioRxiv* i pozostaje w trakcie procesu recenzyjnego. Jest on pierwszym autorem w czterech spośród tych prac, trzech opublikowanych oraz jednej recenzowanej. Przy czym należy podkreślić, że w pracach opisujących główne metody obliczeniowe przedstawione w rozprawie, Doktorant figuruje jako pierwszy autor, co świadczy o jego samodzielnym i wiodącym wkładzie w prowadzone badania. W okresie doktoratu brał on ponadto aktywny udział w projektach naukowych, co dodatkowo potwierdza jego zaangażowanie w działalność badawczą.

Biorąc pod uwagę ocenę przedstawioną w niniejszej recenzji, nie mam wątpliwości, że przedstawiona do oceny praca doktorska w pełni odpowiada warunkom określonym w Art. 187 ustawy z dnia 20 lipca 2018 r. Prawo o Szkolnictwie Wyższym i Nauce (tekst jednolity Dz. U. z 2023 r. poz. 742 z późn. zm.) i na tej podstawie wnoszę do Wysokiej Rady Naukowej Dyscypliny Informatyka Techniczna i Telekomunikacja Politechniki Warszawskiej o dopuszczenie mgr-a Sevastianosa Korsaka do dalszych etapów przewodu doktorskiego.

dr inż. hab. Aleksandra Gruca, prof. PŚ.



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 μ m). 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 U_{bk} 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



Recenzja pracy doktorskiej mgra Sevastianosa Korsaka

dr hab. Witold Rudnicki, prof. UwB
Wydział Informatyki, Uniwersytet w Białymstoku,
ul. Konstantego Ciołkowskiego 1M, 15-245 Białystok,
tel. +48 85 738 8276
e-mail: W.Rudnicki@uwb.edu.pl

Ocena pracy doktorskiej pt. „*Multi-scale Computational Modeling of Chromatin*” przygotowanej przez mgra Sevastianosa Korsaka, pod opieką prof. dr hab. Dariusza Plewczyńskiego, w dziedzinie nauk inżynieryjno-technicznych, w dyscyplinie informatyka techniczna i telekomunikacja.

I Omówienie rozprawy

Rozprawa doktorska została napisana w języku angielskim. Została przygotowana na podstawie materiału zawartego w następujących artykułach (punktacja MNiSW w nawiasach):

1. Agarwal, A., Korsak, S., Choudhury, A., & Plewczynski, D. (2023). The dynamic role of cohesin in maintaining human genome architecture. *BioEssays*, 45(10), 2200240. (100)
2. Korsak, S., & Plewczynski, D. (2024). LoopSage: An energy-based Monte Carlo approach for the loop extrusion modeling of chromatin. *Methods*, 223, 106-117. (100)
3. Korsak, S., Banecki, K. H., Buka, K., Górski, P. J., & Plewczynski, D. (2025). Chromatin as a Coevolutionary Graph: Modeling the Interplay of Replication with Chromatin Dynamics. *bioRxiv*, 2025-03. (preprint)
4. Korsak, S., Banecki, K., & Plewczynski, D. (2024). Multiscale molecular modeling of chromatin with MultiMM: from nucleosomes to the whole genome. *Computational and Structural Biotechnology Journal*, 23, 3537-3548. (100)
5. Kubica, J., Korsak, S., Banecki, K. H., Schirman, D., Yadavalli, A. D., Brenner Clerkin, A., ... & Plewczynski, D. (2025). The challenge of chromatin model comparison and validation: A project from the first international 4D Nucleome Hackathon. *PLOS Computational Biology*, 21(8), e1013358. (140)

Rozprawa składa się z podziękowań, streszczeń w języku angielskim i polskim, spisu treści, trzech rozdziałów merytorycznych, spisu osiągnięć autora, spisu literatury oraz dodatku, w którym zamieszczone są reprinty wymienionych wyżej artykułów. Rozprawa stanowi szerokie omówienie badań opisanych w tych artykułach, scalające je we wspólnych ramach biologicznych, fizycznych i algorytmicznych: przywołuje podstawy biologiczne, fizyczne i algorytmiczne, pokazuje kontekst poszczególnych publikacji, a także dyskutuje wyniki i perspektywy dalszych badań nad tym problemem.

Rozdział pierwszy to wstęp, wprowadzający przedmiot badań, fundamentalne pojęcia niezbędne do zrozumienia rozprawy, a także podstawy metodologii badawczej. Podzielony jest na osiem sekcji.

Sekcja 1.1 wprowadza matematyczny opis dynamiki chromatyny w formalizmie równań Newtona.

W sekcji 1.2 doktorant omawia kilka zagadnień związanych z modelowaniem chromatyny: konieczność opisu wieloskalowego, adekwatność opisu chromatyny jako fraktala, kwestię porządku i losowości w opisie, wreszcie bardzo ważny problem: w jaki sposób wybór warunków początkowych może wpłynąć na przebieg symulacji i jej wyniki. Sekcję zamyka krótki opis przejść fazowych i separacji faz. Oba typy przejść są istotne w kontekście chromatyny: z jednej strony zmiana uporządkowania - przejścia między chromatyną skondensowaną i zdecondensowaną na różnych etapach cyklu komórkowego, z drugiej separacja przestrzenna zdecondensowanej chromatyny należącej do różnych chromosomów w stanie interfazy.

Sekcja 1.3 poświęcona jest wprowadzeniu algorytmów stosowanych w modelowaniu chromatyny. Doktorant przedstawia w niej podstawy mechaniki i dynamiki molekularnej, łącznie z opisem podstawowych typów oddziaływań i stosowanych algorytmów całkowania równań ruchu. Duży fragment sekcji poświęcony jest metodom Monte Carlo, w szczególności Markov Chain Monte Carlo i najbardziej znanemu algorytmowi z tej klasy, czyli algorytmowi Metropolisa.

Sekcja 1.4 poświęcona jest omówieniu metod eksperymentalnych, które rzucają światło na strukturę przestrzenną chromatyny. W pierwszej kolejności doktorant omawia metody służące do analizy jednowymiarowej: Chip-Seq dostarczająca informację o rozmieszczeniu białek wzdłuż sekwencji DNA oraz metoda ATAC-Seq dostarczająca informacji o lokalizacji otwartych regionów chromatyny. Następnie doktorant omawia podstawowe metody dostarczające wskazówek co do trójwymiarowej struktury chromatyny, czyli metody 3C (Chromatin Conformation Capture). Należą do nich Hi-C (High-throughput chromosome conformation capture), ChIA-PET (Chromatin Interaction Analysis by Paired-End Tag sequencing), ChIA-Drop (Chromatin Interaction Analysis via Droplet-based and Barcode-linked sequencing) i Hi-Chip (High-throughput chromosome conformation capture combined with chromatin immunoprecipitation). Doktorant omawia pokrótce na czym one polegają, jakiego typu informację można z nich uzyskać i jakie są ich ograniczenia.

Sekcja 1.5 poświęcona jest wyzwaniom związanym z modelowaniem chromatyny. Doktorant przedstawia trzy klasy wyzwań. Pierwsza to wyzwania związane z analizą i przetwarzaniem danych eksperymentalnych ze względu na ich objętość, złożoność i zaszumienie. Druga grupa obejmuje wyzwania związane z samym procesem modelowania. Należą do nich złożoność obliczeniowa oraz konieczność stosowania modeli gruboziarnistych w różnych skalach. Obejmują one również problemy techniczne związane z integracją modeli wieloskalowych. Wreszcie doktorant wskazuje na wyzwania teoretyczne wynikające z faktu, że klasyczne narzędzia modelowania w skalach mikroskopowych, wykorzystujące dobrze znane metody fizyki obliczeniowej, nie dają wglądu w to, co w modelach chromatyny jest najbardziej interesujące, czyli w powiązania dynamiki z realnymi procesami biologicznymi. Dodatkowym wyzwaniem wskazanym w tej sekcji jest konieczność komunikacji wiedzy między badaczami wywodzącymi się z różnych dyscyplin,

operującymi w różnych paradygmatach i używającymi odmiennych pojęć oraz schematów poznawczych. Trzeci rodzaj wyzwań to powiązanie wyników modelowania i wyników eksperymentalnych i walidacja modeli. Największy problem polega na tym, że nie istnieją techniki eksperymentalne pozwalające na bezpośrednie porównanie wyników symulacji z eksperymentem. Pozostają metody pośrednie, ale w tym wypadku weryfikacja jest trudna, bo konieczne jest stosowanie różnych typów danych dla różnych klas modeli. Co więcej dane eksperymentalne są na tyle niedoskonałe, że trzeba używać różnych wstępnych metod ich obróbki tak, aby wyniki były porównywalne. Mimo wszystko weryfikacja eksperymentalna jest możliwa do przeprowadzenia i autor przedstawia klarownie w jaki sposób należy ją przeprowadzać.

W sekcji 1.6 doktorant przedstawia krótki przegląd metod używanych w modelowaniu chromatyny. Sekcja jest krótka ale pokazuje bardzo dobrą znajomość metod używanych w tej dziedzinie badawczej. Autor wymienia różne klasy metod i podaje dobrze dobrane odnośniki do prac oryginalnych.

W sekcji 1.7 doktorant przedstawia motywację do prowadzenia badań na temat organizacji przestrzennej i dynamiki chromatyny i dlaczego modelowanie jest właściwym sposobem prowadzenia badań.

W sekcji 1.8 doktorant przedstawia cel badań opisanych w dysertacji, czyli opracowanie metodologii wielkoskalowego modelowania chromatyny, a następnie przedstawia podstawowe osiągnięcia opisane w pracy:

- opracowanie metodologii modelowania chromatyny w wielu skalach zaimplementowanej w zbudowanej przez autora bibliotece MultiMM,
- opracowanie metodologii modelowania tworzenia pętli w chromatynie zaimplementowanej w bibliotece LoopSage,
- opracowanie metodologii modelowania chromatyny w trakcie cyklu komórkowego, a w szczególności podczas replikacji, zaimplementowanej w bibliotece RepliSage.

Rozdział drugi stanowi omówienie badań opisanych w pięciu artykułach stanowiących podstawę rozprawy, poświęconych modelowaniu formowania pętli w chromatynie, modelowaniu replikacji chromatyny oraz budowie wieloskalowego modelu chromatyny.

Rozdział należy traktować jako rozszerzone omówienie publikacji, nadające im wspólne ramy i porządkujące ich miejsce w logice całej rozprawy. Zawiera on wspólne tło i dodatkowe wyjaśnienia, na które nie ma miejsca w oryginalnych artykułach.

Sekcja 2.1 poświęcona jest omówieniu modelowania pętli w programie LoopSage. Sekcja oparta jest o artykuły [1] (Agarwal, Korsak i inni; 2023) i [2] (Korsak i Plewczyński; 2024). W sekcji doktorant przedstawia najpierw biologiczne mechanizmy tworzenia pętli, następnie omawia dane wejściowe i stochastyczny algorytm symulacji. Artykuł [1] ma charakter przeglądowy i porządkuje wiedzę o roli kohezyny w organizacji architektury genomu, ze szczególnym uwzględnieniem pętli chromatynowych, domen topologicznych chromatyny (TADs) i modeli opisujących mechanizm formowania i powiększania pętli chromatynowych (loop extrusion). Artykuł jest syntetyczną analizą literatury eksperymentalnej i obliczeniowej; obejmuje również omówienie metod 3C oraz pipeline'ów bioinformatycznych służących do analizy danych eksperymentalnych o interakcjach

chromatynowych. Artykuł pozwala na zbudowanie spójnej ramy pojęciowej dla późniejszych prac doktoranta: uzasadnia, dlaczego modelowanie architektury chromatyny musi łączyć dane kontaktowe, mechanikę polimerów i modele stochastyczne, oraz pokazuje, że kohezyna jest naturalnym punktem wyjścia dla obliczeniowych modeli ekstruzji pętli. Artykuł [2] prezentuje LoopSage: obliczeniowy model ewolucji pętli chromatyny, służący do generowania zespołów struktur 3D dla domen topologicznych chromatyny. Wyniki modelu są następnie porównywane z danymi kontaktowymi 3C.

Model składa się z dwóch modułów - jeden to liniowy model Monte Carlo opisujący topologię pętli z funkcją energii opisującą wielkoskalową organizację pętli chromatynowych, drugi to mezoskalowy model dynamiki molekularnej opisujący dynamikę chromatyny jako polimeru, który jest używany do generowania równowagowych zespołów konformacji. Topologia DNA jest modelowana przy użyciu algorytmu Metropolis w stałej temperaturze lub w schemacie symulowanego wygrzewania. Funkcja pseudoenergii opisuje zachowanie chromatyny w następujący sposób: naturalną tendencją do tworzenia pętli zapewnia człon pseudoenergii, którego wartość maleje logarytmicznie wraz ze wzrostem długości pętli, człon dodatni energii to kara za krzyżowanie się pętli. W module mezoskopowym chromatyna jest opisywana jako losowy polimer z wyłączoną objętością plus więzy harmoniczne nałożone na odległości między lokalizacjami początku i końca pętli. Lokalizacje początków i końców pętli zmieniają się w czasie, więc efektywnie potencjał jest zmienny w czasie. Połączenie stochastycznego modelu topologii z dynamicznym modelem mezoskopowym daje w efekcie dynamiczny model przestrzenny chromatyny. Na podstawie takiego modelu można wygenerować rozkłady prawdopodobieństwa połączeń przestrzennych między różnymi fragmentami chromatyny i porównać je z doświadczalnymi danymi typu 3C. Model LoopSage potrafi odtwarzać jakościowo różne stany upakowania chromatyny, w tym przejście między stanem rzadszym i bardziej skondensowanym, i generuje dane zgodne z eksperymentem.

W sekcji 2.2 doktorant przedstawia algorytm RepliSage, służący do modelowania chromatyny w czasie replikacji. Sekcja oparta jest o artykuł [3] (Korsak i in.; 2025). Algorytm jest oparty na wcześniej opisanym algorytmie LoopSage. Algorytm uwzględnia różnice w zachowaniu chromatyny w różnych fazach cyklu komórkowego. W szczególności podstawowy algorytm LoopSage opisuje zachowanie chromatyny w fazie G1. Proces replikacji jest opisywany przez wprowadzenie do opisu procesu inicjacji widełek replikacyjnych i ich propagacji. Oba procesy są opisane jako procesy stochastyczne, z rozkładami prawdopodobieństwa inicjacji i szybkości propagacji zgodnymi z danymi eksperymentalnymi. Zachowanie łańcuchów chromatyny w trakcie replikacji jest opisywane przez algorytm LoopSage w którym uwzględniono interakcje między domenami topologicznymi i procesem replikacji. O ile w fazie G1 interesujące są rozkłady równowagowe i do opisu chromatyny w tej fazie można używać algorytmu symulowanego wygrzewania, to w fazie S proces jest daleki od równowagi i w związku z tym stosowanych jest zwykły algorytm Metropolis. W fazie G2 wprowadzana jest większa aktywność tworzenia pętli przez dodanie drugiej rodziny białek tworzących pętle (kondensyn), o większej szybkości generowania pętli. Faza M jest reprezentowana przez dodanie siły działającej globalnie na dwie nici, każda z nich jest odciągana w przeciwną stronę.

Sekcja 2.3 jest poświęcona omówieniu narzędzia MultiMM opisanego w artykule [4] (Korsak i in.; 2024). Algorytm pozwala na symulację chromatyny w różnych skalach - od

pojedynczych nukleosomów do całych chromosomów, a nawet oddziaływać chromosomów z blaszką jądrową. Dla każdej skali oddziaływań zaproponowana jest inna postać energii oddziaływań, adekwatna dla danej skali i zgodna z wiedzą biologiczną. Model został zaimplementowany z wykorzystaniem powszechnie stosowanej biblioteki symulacyjnej OpenMM. Dzięki temu dostępny jest bogaty wybór metod symulacyjnych zaimplementowanych w OpenMM, a także optymalizacje pozwalające na wykorzystanie akceleratorów CUDA i prowadzenie obliczeń równoległych.

Sekcja 2.4 jest poświęcona krótkiej dyskusji algorytmów i narzędzi opisanych w poprzednich sekcjach - porównanie obszarów zastosowań i omówienie ograniczeń.

W Rozdziale trzecim doktorant przedstawia szerszą syntetyczną dyskusję na temat modelowania chromatyny przy użyciu opracowanych przez siebie narzędzi, ich ograniczeń i dalszego ich rozwoju. Doktorant przedstawia również swoją opinię na temat dalszego kierunku badań w tej dziedzinie, ze szczególnym naciskiem na możliwość zastosowania metod sztucznej inteligencji.

W sekcji 3.1 doktorant w szczególności dyskutuje zastosowanie modularnego formalizmu opartego o funkcję pseudoenergii, wskazuje na adaptowalność opracowanych przez siebie narzędzi do różnych skal symulacji i do różnych kontekstów biologicznych, dyskutuje również zastosowanie metod stochastycznych, a także dynamiki zależnej od czasu.

W sekcji 3.2 doktorant wskazuje na interpretację wyników symulacji i wpływ przeprowadzonych badań na dziedzinę:

- wskazuje na to, że jego praca wprowadza możliwość symulacji różnych procesów zachodzących w chromatynie,
- wskazuje na możliwość zastosowania tych narzędzi do badania realnych procesów biologicznych,
- uzyskanie wyników symulacyjnych zgodnych z danymi doświadczalnymi,
- doktorant podkreśla reprodukowalność wyników i dostępność narzędzi dla innych badaczy.

W tej sekcji doktorant dyskutuje również ograniczenia:

- brak głębszego modelu teoretycznego; modele opracowane przez doktoranta mają charakter modeli ad hoc - stosowalnych do konkretnych zastosowań,
- problem wyboru parametrów modeli,
- wymaganie dużych zasobów obliczeniowych,
- wyzwania dotyczące walidacji wyników.

Sekcja 3.3 to omówienie otwartych problemów, a w szczególności najważniejszego wyzwania, czyli opracowania modelu 3D chromatyny wyłącznie na podstawie sekwencji.

Sekcja 3.4 poświęcona jest dyskusji nad dalszymi kierunkami rozwoju modeli. Obejmuje ona zarówno rozszerzanie istniejących modeli, jak i sposoby łączenia ich z metodami uczenia maszynowego. Doktorant wskazuje tu przykładowo na możliwość wykorzystania metod uczenia maszynowego do doboru parametrów modeli, przewidywania elementów strukturalnych na podstawie sekwencji oraz konstruowania funkcji pseudoenergii z użyciem sieci neuronowych.

Sekcja 3.5, ostatnia, to krótkie podsumowanie wyników uzyskanych w rozprawie.

Oddzielny rozdział rozprawy poświęcony jest przedstawieniu osiągnięć naukowych doktoranta:

- wszystkich publikacji, których był współautorem, za szczególnym uwzględnieniem publikacji włączonych do rozprawy,
- aktywnego udziału w konferencjach i warsztatach naukowych,
- opracowanych przez siebie algorytmów i narzędzi do modelowania chromatyny,
- innych aktywności w trakcie doktoratu.

Zasadniczą część rozprawy kończy bogaty spis literatury cytowanej w głównym tekście rozprawy (artykuły oryginalne mają własne spisy literatury).

Rozprawę zamyka dodatek zawierający reprodukcje wymienionych wcześniej artykułów stanowiących podstawę pracy.

II Ogólna ocena rozprawy

Przedstawiona przez pana Sevastianosa Korsaka rozprawa doktorska dotyczy problematyki fundamentalnej dla zrozumienia podstawowych mechanizmów funkcjonowania komórki eukariotycznej. Tematyka rozprawy leży na przecięciu trzech dyscyplin nauki - biologii, informatyki, a także fizyki.

W rozprawie doktorant przedstawił opracowane przez siebie trzy narzędzia służące do budowy przestrzennych modeli chromatyny w różnych skalach i w różnych procesach. Wyniki przeprowadzonych symulacji wykazują dużą zgodność z danymi eksperymentalnymi. Praca wykonana przez doktoranta jest nowatorska i wartościowa, stanowi istotny wkład w zrozumienie dynamiki struktury przestrzennej chromatyny. Cztery z włączonych do rozprawy artykułów zostały opublikowane w bardzo dobrych międzynarodowych czasopismach, jeden pozostaje na etapie preprintu, jednak jego poziom wskazuje na realną perspektywę publikacji w dobrym czasopiśmie międzynarodowym.

Jednym z ważnych osiągnięć doktoranta jest opracowanie takich reprezentacji obliczeniowych procesów zachodzących w chromatynie, które pozwalają na ich modelowanie, symulację i porównywanie z danymi eksperymentalnymi. Dotyczy to zarówno doboru poziomu opisu, jak i sposobu powiązania modelu z obserwacjami eksperymentalnymi. Na szczególne podkreślenie zasługuje operowanie kilkoma komplementarnymi reprezentacjami tego samego układu: liniową, przestrzenną, dynamiczną, a w części badań również grafową. Szczególnie interesujące jest tu dynamiczne ujęcie grafowe, w którym relacje między elementami układu są traktowane jako obiekt ewoluujący w czasie. Istotnym walorem pracy jest także integracja tych poziomów opisu w ramach spójnych procedur obliczeniowych oraz implementacja własnych bibliotek służących do wieloskalowego modelowania chromatyny.

Opracowane przez doktoranta algorytmy i narzędzia są oparte na metodologii mocno osadzonej w aparacie fizyki klasycznej i statystycznej. Taki wybór metodologii symulacji procesów biologicznych pozwala na bardzo skuteczne uproszczenie problemu i wykorzystanie uniwersalnego aparatu algorytmicznego. Metodologia zastosowana przez

doktoranta wymaga opisanie problemu w języku energii lub pseudoenergii oddziaływań między komponentami. Należy podkreślić, że zarówno energia, jak i używana w modelach ewolucji statystycznej temperatura w większości wypadków nie są prawdziwą energią oddziaływania ani rzeczywistą temperaturą symulowanego układu. Są to formalne konstrukty pozwalające zmapować problem do odpowiedniego formalizmu, w którym znane są efektywne algorytmy, dobrze ugruntowane procedury symulacyjne oraz bogaty repertuar narzędzi opracowanych do modelowania układów fizycznych. Na przykład do optymalizacji można użyć metod opartych na mechanice molekularnej lub na algorytmach Monte Carlo, do symulacji procesów dynamicznych można użyć algorytmów opracowanych na potrzeby dynamiki molekularnej lub implementacji symulacji łańcuchów Markowa. Są to metody doskonale ugruntowane teoretycznie i obliczeniowo, których właściwości są dobrze poznane, co zapewnia wysoką jakość technicznej strony modelowania.

Doktorant nie poprzestaje przy tym na użyciu gotowych algorytmów, lecz buduje z nich spójne procedury obliczeniowe łączące reprezentację problemu, dobór parametrów, symulację i walidację wyników. W tym sensie rozprawa wnosi nie tylko modele biologiczne, ale również dobrze zaprojektowane instrumentarium obliczeniowe.

Ważnym efektem takiego podejścia jest możliwość budowy modeli modułowych, dających się adaptować do różnych skal opisu i różnych kontekstów biologicznych, a zarazem zachowujących spójność obliczeniową. To właśnie ta umiejętność łączenia wieloskalowego opisu, integracji danych, dynamicznej reprezentacji grafowej oraz własnej implementacji bibliotek obliczeniowych stanowi jeden z najmocniejszych elementów dorobku doktoranta.

III Poprawność

Rozprawa ma interesującą kompozycję. Część podstawowa to trzy rozdziały poświęcone odpowiednio

- przedstawieniu podstaw biologicznych i metodologicznych modelowania chromatyny,
- omówieniu treści prac oryginalnych z naciskiem na zrozumienie założeń i podstaw metodologicznych konkretnego modelu
- dyskusji różnic między algorytmami i ich obszarów zastosowań, a także dyskusję na temat dalszego rozwoju modeli i ewentualnego powiązania metod opartych na formalizmie fizyki z metodami opartymi o sztuczną inteligencję.

Rozprawa jest dobrze napisana, zawiera bardzo bogaty przegląd literatury, w przystępny sposób przedstawia podstawy zastosowanych w pracach metod. Bardzo cennym elementem pracy jest przedstawienie szerszej perspektywy zarówno rozwoju narzędzi opisanych w rozprawie jak i możliwości i sposobów wykorzystania w modelowaniu metod sztucznej inteligencji.

Od strony merytorycznej nie mam poważnych zastrzeżeń, opisy modeli i algorytmów są poprawne i wyczerpujące, wyniki są dobrze opisane.

Uchybienia

Doktorant włączył do rozprawy w dodatku pięć publikacji. Jednak szczegółowego omówienia doczekały się tylko cztery z nich. Piąta, która jest wieloautorską pracą omawiającą wyniki Hackathonu poświęconego symulacjom chromatyny, została potraktowana szczątkowo. Jest wspomniana w rozprawie w trzech miejscach w kontekście walidacji symulacji za pomocą danych eksperymentalnych (strona 64), a następnie na stronach 83 i 107 odpowiednio w kontekście problemów z walidacją modelu LoopSage oraz braku i/lub niewystarczającej jakości danych eksperymentalnych dla modeli „bottom-up”. Odnośniki są oczywiście adekwatne, ale w tekście rozprawy zabrakło opisu metodologii przyjętej w tej pracy i opisu wyników dotyczących modeli opracowanych przez doktoranta. Nie jest to zarzut dyskwalifikujący: tematyka poruszana w pracy [5] jest omawiana w rozprawie, a włączenie jej do dodatku pozwala na zapoznanie się z nią w całości. Jednak zabrakło mi szerszego i bardziej wyczerpującego omówienia tej pracy w głównym tekście rozprawy.

Drobniejsze uchybienia.

W rozdziale poświęconym prezentacji swoich osiągnięć doktorant przyjął dziwną konwencję pisania o swojej pracy. O ile cała rozprawa jest napisana w pierwszej osobie, to w tym rozdziale przeszedł do opisu w trzeciej osobie. Jest to dziwny wtręt do spójnego stylu rozprawy i wygląda jak opinia promotora wklejona do tekstu rozprawy. To zupełnie niepotrzebny zgrzyt, nie pasujący do stylu całej rozprawy napisanej w pierwszej osobie.

Opis członu energii oddziaływań kompartmentowych jest mało przejrzysty. Zastosowana notacja typu $\delta(s\pm 1)$ interpretowana jako delta Kroneckera jest niestandardowa i utrudnia odczytanie sensu modelu. Znacznie czytelniejsze byłoby sformułowanie go przy użyciu standardowej delty Kroneckera: δ_{ij} dla stanów $i, j \in \{A, 0, B\}$ z jawnym podaniem jedynie niezerowych elementów diagonalnych, np. E_{AA} , E_{BB} oraz $E_{00}=0$.

Dodatkowo w samym tekście warto byłoby wyraźnie zaznaczyć, że znak parametrów E_{AA} i E_{BB} jest ujemny, ponieważ bez tej informacji nie jest od razu jasne, że rozważany człon odpowiada oddziaływaniu przyciągającemu. Dopiero odwołanie się do publikacji oryginalnej pozwala zrozumieć działanie zaproponowanego modelu. Uwaga ta dotyczy przejrzystości zapisu, a nie wartości merytorycznej samego modelu.

Przyjęta przez doktoranta forma rozprawy tłumaczy brak szczegółowego omówienia wyników prac oryginalnych, a częściowo także brak zwyczajowych spisów tabel i ilustracji. Ponieważ jednak w treści rozprawy zamieszczono 25 ilustracji, spis ilustracji powinien się w niej znaleźć.

Poza spisem ilustracji, brakowało mi w rozprawie słownika pojęć i skrótów. Poruszona w treści rozprawy tematyka jest obszerna i interdyscyplinarna. Z tego względu używany jest w niej bardzo bogaty aparat pojęciowy z dziedziny fizyki, biologii i informatyki. Słownik pojęć i skrótów znacząco ułatwiłby lekturę.

Najpoważniejsza z powyższych uwag dotyczy kompozycji rozprawy, a dokładniej niewystarczającego omówienia jednej z prac w tekście głównym. Nie jest to jednak zarzut odnoszący się do jakości samych badań, ich oryginalności ani wartości uzyskanych wyników. Z tego względu wskazane uchybienia nie zmieniają mojej bardzo wysokiej oceny całości rozprawy i przedstawionego w niej dorobku.

III Wiedza doktoranta

Doktorant wykazał w pracy wiedzę zarówno szeroką, jak i głęboką. Jak wspomniałem wcześniej, rozprawa leży na przecięciu trzech dyscyplin naukowych; doktorant wykazał się głęboką wiedzą w każdej z nich: wiedzą na temat badanych procesów biologicznych, doskonałą znajomością podstaw teoretycznych zastosowanych algorytmów, a także umiejętnością doboru algorytmów i budowania na ich bazie narzędzi do analiz bioinformatycznych chromatyny. Cytowane przez doktoranta pozycje z literatury przedmiotu są dobrze dobrane i wyczerpujące.

IV Podsumowanie

Po przeczytaniu i przeanalizowaniu przedstawionej rozprawy oceniam, że:

1. Rozprawa przedstawia oryginalne rozwiązanie problemu naukowego (zdecydowanie TAK);
2. Kandydat posiada ogólną wiedzę teoretyczną w dyscyplinie informatyka techniczna i telekomunikacja (zdecydowanie TAK);
3. Kandydat posiada umiejętność samodzielnego prowadzenia pracy naukowej (zdecydowanie TAK).

Stwierdzam, że rozprawa doktorska pana magistra Sevastianosa Korsaka spełnia warunki określone w art. 187 Ustawy z dnia 20 lipca 2018 r. Prawo o szkolnictwie wyższym i nauce, tekst jednolity: Dz.U. z 2021 r. poz. 478. W związku z powyższym wnioskuję o dopuszczenie pana Sevastianosa Korsaka do dalszych etapów przewodu doktorskiego.

Ze względu na wysoki poziom naukowy rozprawy, oryginalność uzyskanych wyników oraz wartość przedstawionego dorobku, a także zważywszy, że wskazane wcześniej uchybienia mają charakter kompozycyjny i nie podważają jakości samych badań, wnioskuję o wyróżnienie rozprawy.

dr hab. Witold Rudnicki, prof. UwB

Białystok, 20.04.2026