



Wrocław, 15.05.2025

**Recenzja rozprawy doktorskiej Pana mgr inż. Michała Kadlofa pt. "Komputerowe modelowanie trójwymiarowej struktury chromatyny"**

Organizacja i ekspresja informacji genetycznej, zwłaszcza u organizmów eukariotycznych, są skomplikowanymi zjawiskami wymagającym interakcji materiału genetycznego i białek. Włókniste kompleksy tych związków zwane chromatyną umożliwiają upakowanie bardzo długiej cząsteczki DNA (u człowieka ok. 2 metry) w niewielkiej objętości jądra komórkowego o średnicy kilku mikrometrów. Poza funkcją strukturalną, chromatyna odgrywa kluczową rolę w regulacji ekspresji genów. Jej organizacja na różnych poziomach umożliwia kontrolę dostępu do konkretnych fragmentów materiału genetycznego. Aby zrozumieć te zjawiska nie wystarcza już znajomość sekwencji genomu oraz białek łączących się z kwasami nukleinowymi, ale potrzebna jest wiedza o trójwymiarowej strukturze chromatyny. Badania nad trójwymiarową organizacją chromatyny mają kluczowe znaczenie dla zrozumienia mechanizmów regulacji genów, różnicowania komórek, odpowiedzi na bodźce środowiskowe oraz powstawania chorób. Chromatyna nie jest statyczna – jej przestrzenna organizacja jest dynamiczna i zależy od typu komórki, fazy cyklu komórkowego oraz sygnałów zewnętrznych. Dlatego bardzo słusznie ambitnym przedmiotem pracy doktorskiej Pana mgr inż. Michała Kadlofa stało się opracowanie metod komputerowego modelowania trójwymiarowej struktury chromatyny.

Rozprawa doktorska została napisana w języku polskim ze streszczeniem w języku angielskim. Zawiera ona: Streszczenie, Abstract, Spis treści, Wykaz użytych skrótów, Część teoretyczną i praktyczną oraz Dodatki zawierające wykaz publikacji autora, warianty strukturalne nukleosomów, Spis rysunków i tabel, Bibliografię i Skorowidz. Nie jest to typowy układ, jednak Część teoretyczną można traktować jako wstęp, a praktyczną jako wyniki i dyskusję. Przydałby się jednak wyróżnić cele pracy w osobnym rozdziale oraz dodać rozdział z wnioskami lub podsumowaniem zawierający najważniejsze informacje uzyskane z opracowanych modeli i przeprowadzonych badań.

W Części teoretycznej we Wstępie w bardzo przystępny i prosty sposób doktorant scharakteryzował chromatynę podając łatwy sposób na jej izolację, którą wykonał, a efekt uwiecznił na zdjęciach. Pod koniec Wstępu podał cel pracy, którym było zaproponowanie nowatorskich metod obliczeniowych do określania przestrzennych modeli chromatyny. Cel pracy doktorskiej został jasno sformułowany i właściwie zrealizowany. Pozostałe rozdziały w przejrzysty i wystarczający sposób zbierają najważniejsze fakty o budowie i roli chromatyny w regulacji ekspresji genów i zachowaniu struktury genomu. Ciekawe są wyliczenia związane z długością, masą i gęstością chromatyny. Zwrócono uwagę, że włókno 30 nm nie jest strukturą powszechnie występującą, jak wcześniej sądzono, a głównymi poziomami organizacji chromatyny są włókno 11 nm, pętle, domeny i kompartmenty. Dobrze byłoby przedstawić schematy pokazujące tę hierarchiczną organizację.

Osobny podrozdział dotyczy technik eksperymentalnych służących do określania struktury chromatyny, genomowych (3C, 4C i 5C, Hi-C, ChIA-PET, ChIA-Drop) i mikroskopowych (FISH, EM-ISH, iPALM). Opisano je w zrozumiały sposób uwzględniając ich wady i zalety, jednak idealnie byłoby do opisu dołączyć schematy tych metod. W dodatkowym rozdziale przedstawione zostały modele chromatyny: łańcuch idealny, model Kratky-Poroda oraz model WLC, model Odijk-Skolnick-Fixman oraz globule fraktalne uwzględniające różne cechy, jak sztywność, efekty odpychania elektrostatycznego oraz bliskiego oddziaływania regionów.

W ostatnim rozdziale części teoretycznej scharakteryzowano metody modelowania struktury chromatyny. W rozdziale zwrócono także rozważania natury filozoficznej, a nie tylko suche techniczne fakty. Podkreślono różnice w modelowaniu chromatyny i białek dotyczące natury ich struktur oraz skali i rozdzielczości danych do modelowania. Można byłoby tutaj wspomnieć o słynnym programie AlphaFold. Ciekawe jest porównanie struktury chromatyny do mapy świata. Doktorant przedyskutował ważne parametry związane z modelowaniem, jak: skala, rozdzielczość, typ danych i cechy polimeru oraz przedstawił ważne rozważania dotyczące idei modelowania, co wskazuje, że zdaje sobie sprawę z jego uproszczeń i ograniczeń. Świadczy to o głębokim zrozumieniu zagadnienia przez Pana Michała Kadlofa.

Mam jednak jeszcze inne uwagi do tej części. W przypadku celów modelowania można byłoby je lepiej sprecyzować, zwłaszcza dotyczące zagadnień badawczych obok wizualizacyjnych. W części dotyczącej wyboru regionów do modelowania prosiłbym dokładniej wyjaśnić kwestię macierzy po normalizacji. Podrozdział dotyczący wyboru skali mógłby zawierać wyszczególnienie tych skal i podanie rzędów wielkości, albo można byłoby go połączyć z podrozdziałem o rozdzielczości. Dobrze byłoby wyjaśnić, co to jest punkt

modelowanego układu. Nie wiem czy wyrażenie "Im lepsza (tzn. mniejsza) rozdzielczość" powinno brzmieć "Im lepsza (tzn. większa) rozdzielczość". Dobrze byłoby wyjaśnić promień bezwładności dla struktur modelowanych oraz potencjał FENE.

Główną częścią doktoratu jest część praktyczna, która jest podzielona na kilka rozdziałów opisujących różne narzędzia do modelowania struktury chromatyny. Doktorant przedstawił dwie metody opracowane przez siebie do wyznaczania użytecznych granic domen. Jedna z metod nazwana CCD Caller służy do identyfikacji regionów o ponadprzeciętnym pokryciu łukami interakcji w porównaniu do sygnału oczekiwanego. Jest on użyteczny, ponieważ w obiektywny sposób wyszukuje takie regiony (domeny typu CCD) i ułatwia automatyzację analiz. Druga metoda MetaTAD Caller analizuje wyniki wielu innych programów wyznaczający domeny typu TAD w oparciu o głosowanie wykorzystując funkcję Gaussa opisującą rozmycie położenia granicy domen. Dobrze byłoby przedstawić jaka jest skuteczność tych metod, np. procent fałszywie i poprawnie przewidzianych regionów w porównaniu do danych rzeczywistych oraz potencjalne ograniczenia. Mam kilka pytań związanych z tą częścią. Prosiłbym o wyjaśnienie zdania, że TAD-y tworzą zbiór stycznych do siebie odcinków dzielących cały chromosom na części. Według mnie styczne odcinki do siebie będą się nakładać i będą równoległe. Mam jeszcze dodatkowe pytanie: Czy liczenie indeksu kierunkowego od tyłu, czyli różnica interakcji późniejszych do wcześniejszych da ten sam wynik, co różnica między wcześniejszymi i późniejszymi interakcjami? Uważam, że Rysunek 6.4 powinien być lepiej opisany.

Jasno jest scharakteryzowana metoda skalowania wielowymiarowego i dystanse grafowe oraz sposoby eliminacji szumu, jednak brakuje wyjaśnienia po co jest ten opis. Czy został on zastosowany w jakimś programie? Wypadałoby pokazać skuteczność działania tej metody na przykładzie.

Pan mgr inż. Michał Kadlof przedstawił także algorytm budowania modeli polimerowych nazwany modelem sprężynowym (ang. SM, Spring Model) polegający na znalezieniu takiej konformacji punktów w przestrzeni trójwymiarowej, która minimalizuje funkcję energii potencjalnej. Model jest atrakcyjny i uwzględniający wiele istotnych parametrów, a jego opis jest jasny i zawiera wiele praktycznych wskazówek. Szczególnie dużo miejsca poświęcono strukturom początkowym. Podano także interesujące przykłady zastosowań oraz implementacje. Mam pewne uwagi również do tej części. We wzorze i opisie na energię niedostępnej objętości dobrze byłoby dać ten sam symbol na średnicę kulki. Chciałbym się także dowiedzieć skąd bierze się potęga 12 w tym wzorze? Następujące stwierdzenie jest dla mnie niejasne: Z tego powodu dobrą praktyką jest wybierać takie struktury

początkowe, które już na początku zostały wybrane w taki sposób by ich energia nie odbiegała znacznie od minimum układu. Jest to dla mnie dziwne, ponieważ jak wybrać te struktury, skoro aby dowiedzieć się jakie jest minimum energetyczne układu trzeba przeprowadzić obliczenia już na jakiś strukturach początkowych? W związku z tym, że struktury początkowe mogą mieć wpływ na końcowy wynik chciałbym uzyskać odpowiedź na pytanie: Jak duża jest powtarzalność wyników dla różnych stanów początkowych? Domyślam się, że może to zależeć od złożoności układu. W przypadku opisu zastosowań tej metody w podrozdziale 8.6.1 dobrze byłoby skomentować uzyskane wyniki na Rys. 8.12. Jakie struktury należałoby oczekiwać w podanym przykładzie? Dobrze byłoby również napisać jaka jest skuteczność modelu sprężynowego w porównaniu z innymi. Chciałbym jeszcze uzyskać odpowiedź na pytanie: Do jakich jeszcze innych polimerów może służyć opracowany model sprężynowy?

Interesujące jest podejście opisane w następnym rozdziale uwzględniające zastosowanie wysokorozdzielczych obrazów mikroskopowych w metodzie siłowej i korelacyjnej służących do modelowania struktur polimerów. Pokazano także wynik działania tego podejścia na oczekiwanej strukturze metodą siłową przy dwóch stanach początkowych oraz na danych jądra komórkowego *Arabidopsis thaliana*. Model generalnie dobrze wypadł, ale pokazano, że wpływ na wynik ma stan początkowy. Ładnie przedyskutowano wyniki symulacji na danych układu testowego. Podobny komentarz przydałby się dla danych rzeczywistych. Mam jeszcze dodatkowe pytanie: Jaki byłby wynik przy zastawianiu metody korelacyjnej oraz innych stanów początkowych dla obu metod?

Ostatni rozdział dotyczy aspektów technicznych zaawansowanych obliczeń. Doktorant przedstawił najważniejsze problemy obliczeniowe i sposoby ich rozwiązania. Szczegółowo opisał architekturę i zarządzanie na klastrze Eden, który służy do wykonywania skomplikowanych obliczeń struktury chromatyny. Klaster ten posłużył do uzyskania wyników przedstawionych w pracy doktorskiej. Należy zaznaczyć, że Pan mgr inż. Michał Kadlof administrował tym klastrem. Dobrze byłoby, aby doktorant wyeksponował jaki miał udział w jego budowie i wdrażaniu.

Warto podkreślić, że doktorant jest współautorem 9 publikacji w renomowanych czasopismach, a dwie metody zostały opublikowane w dwóch pracach.

Generalnie praca jest napisana poprawnym językiem, miejscami występuje żargon i swobodny styl, ale dobrze się ją czyta. Znalazłem następujące błędy z propozycją poprawek: 2m -> 2 m; dwóch typów para zasad -> dwóch typów par zasad; ilością wiązań -> liczbą wiązań; które łatwo tworzą jony H<sup>+</sup> -> które łatwo przyciągają jony H<sup>+</sup>; 3mm -> 3 mm, dna -> DNA;

odległości ziemi od słońca -> odległości Ziemi od Słońca; Histogram długości chromosomów -> Wykres długości chromosomów; parametru GC-content -> zawartości par GC; bilateria -> Bilateria; rys. -> Rys; fig. -> Fig., kompartmenty -> kompartmenty, intra i inter chromosomalnych -> intra- i interchromosomalnych; nie wielką -> niewielką; trudne do sekwencjonowania Np, regiony o wysokiej repetytywności, niemapowalne itp. -> trudne do sekwencjonowania, np. regiony o wysokiej repetytywności, niemapowalne itp., odmienne interpretacje danych. Np. wartości z macierzy Chia-PET -> odmienne interpretacje danych, np. wartości z macierzy Chia-PET, ilości ilości uzyskanych odczytów -> liczby uzyskanych odczytów, Prezentuje on ten przypadki -> Prezentuje on przypadki, giętkąść -> giętkość, po rzeczywisty -> po rzeczywistym, nie wpływa na pomiary -> nie wpływa on na pomiary, nie pożądanym -> niepożądanym, płaskie. (rys. 8.4) -> płaskie (Rys. 8.4)., Tek -> Tak, pomiędzy każdym trzem sąsiadującymi punktami -> pomiędzy każdym z trzech sąsiadujących punktów, problemowi *problemowi* -> *problemowi*, nie zbyt -> niezbyt, nie długich -> niedługich, w w -> w, dlatego. Z tego -> dlatego, z tego, przyz -> przy, kolekcję kolekcję -> kolekcję, Na struktura -> Ta struktura, Dokładnie identyczne zastosowano -> Dokładnie identyczne założenia zastosowano, sać się -> stać się, macierzy. Np. serwerów -> macierzy, np. serwerów, są -> są, Obniża bezpieczeństwo -> Obniża to bezpieczeństwo, ilość -> liczba.

Ponadto w wielu miejscach brakuje przecinków przed: aby, zaś, gdy, gdyby, co, czyli, w której, którą, które, w których, ponieważ, jeśli. Brak jest czasownika w kilku zdaniach. Jedno zdanie było urwane. Niejasne jest zdanie: W efekcie pokrycie średnie ma bardzo, krótkie obszary początkowe i końcowe, a większość chromosomu. Niepoprawnie jest sformatowana Tabela 2.2. Niektóre rysunki powinny być większe. Raz jest używane fig. a raz rys. Jako separator dziesiętny raz jest stosowany przecinek, a raz kropka.

Reasumując mogę stwierdzić, że powyższe uwagi nie umniejszają znaczenia pracy doktorskiej. 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, 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 trójwymiarowego modelowania struktury 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 mgr inż. Michał Kadlof za odpowiadającą wymogom stawianym rozprawom doktorskim i o dopuszczenie doktoranta do dalszych etapów przewodu doktorskiego.

Prof. dr hab. Paweł Mackiewicz

