

Modelowanie wywlekania pętli chromatynowej

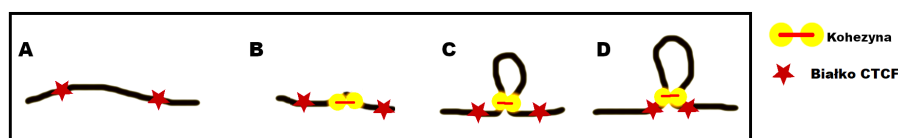
Łukasz Milewski, Maciej Zagrodzki, Elżbieta Karpińska
w Laboratorium Strukturalnej i Funkcjonalnej Genomiki,
Centrum Nowych Technologii
pod kierunkiem prof. Dariusza Plewczyńskiego i Michała Kadlofa

Informacje genetyczne w organizmach eukariotycznych gromadzone są w jądrze komórkowym w postaci włókna chromatynowego: kwasu nukleinowego owiniętego na białka – histony. W chromatynie człowieka wyróżnia się obszary, w których widzimy częste interakcje wewnętrzne TAD (ang. *Topologically associating domain*), których funkcja związana jest z regulacją ekspresji genów. TADy prawdopodobnie formowane są poprzez wywlekanie pętli chromatynowych wywlekane z udziałem kompleksu kohezynowego. Ekstruzja pętli blokowana jest przez białka CTCF. Celem projektu było sprawdzenie tych dwóch hipotez formowania się TADów. W tym celu symulowano dwa rodzaje wywlekania pętli: dwustronną (pętla wywleka się z obu stron) i jednostronną (ekstruzja zablokowana z jednej strony). Pierwszym etapem było upłynnienie działającego skokowo w kodzie prototypowym dwustronnego tworzenia się pętli. Następnie symulowano pętlę wywlekaną jednostronnie. Napisano program generujący mapę kontaktów w polimerze dla obu rodzajów wywlekania. Wyniki porównano wizualnie z przykładową mapą kontaktów w fragmencie genomu ludzkiego bazującą na danych z eksperymentów Hi-C. W wyniku analizy wizualnej uznaliśmy model ekstruzji jednostronnej za bardziej prawdopodobny od wywlekania dwustronnego.

Wstęp

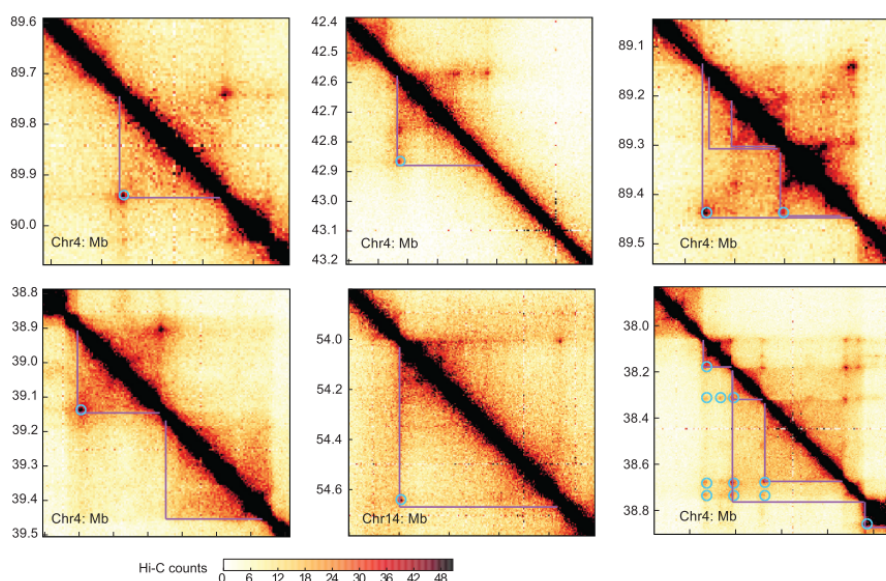
DNA w jądrze komórkowej upakowane jest pod postacią chromatyny. Chromatyna ma wysoce uporządkowaną strukturę na kilku poziomach organizacji podzieloną na zróżnicowane funkcjonalnie domeny¹.

Chromatyna umożliwia zachodzenie ważnych procesów biologicznych takich jak ekspresja genów, replikacja. W celu umożliwienia tych procesów, odpowiednie enzymy muszą uzyskać dostęp do fragmentów genomu, a do tego konieczne są zmiany konformacji chromatyny². Jest ona głównym składnikiem chromosomów. Do analizy procesu zwijania chromatyny można użyć metod dynamiki molekularnej, które polegają na symulowaniu zachowań cząsteczek z wykorzystaniem całkowania równań dynamiki Newtona. Tworzenie się pętli chromatynowej opisywane jest poprzez model wywlekania pętli (ang. *Loop Extrusion Model* - LEM), który przewiduje, że pętle chromatynowe powstają przy udziale pierścienia białkowego (kompleks kohezynowy), a dalsza ekstruzja pętli blokowana jest przez białko CTCF znajdujące się na włóknie chromatynowym³. Przedstawiono to na Rys. 1.



Rysunek 1: Schemat tworzenia się pętli chromatynowej zgodnie z LEM (modelem wywlekania pętli, od ang. *Loop Extrusion Model*). Czarna nić – włókno chromatynowe; żółte okręgi - pierścień kohezynowy; czerwone gwiazdki - kompleksy blokujące zawierające białko CTCF.

Miejsca wiążące kohezynę działają jako czynniki wywołujące ekstruzję pętli – LEFs (ang. *loop extrusion factors*). Pośród wielu metod analizy organizacji chromatyny wyróżniamy między innymi metodę Hi-C, która pozwala na badanie obszarów takich jak np. TADy (ang. *topologically associated domains*)⁴, czyli obszary włókna chromatynowego bogate w kontakty wewnętrzne⁵. W metodzie Hi-C w celu zidentyfikowania TADów używa się algorytmów. Mapy Hi-C (Rys. 2) to wykresy przedstawiające interakcje między dwoma loci na chromosomie (na osiach umieszczone są współrzędne). Im ciemniejszy kolor danej części mapy, tym częściej mamy interakcje między tymi loci. Ciemne kwadraty na diagonalu reprezentują ten sam locus chromosomu, a kwadraty "wokół" diagonalu to TADu widoczne jako obszary bogatsze w interakcje wewnętrzne (mamy więcej interakcji wewnątrz tego regionu niż z pozostałymi obszarami genomu)⁶. Piki na rogach TAD'ów interpretuje się jako pętle chromatynowe.

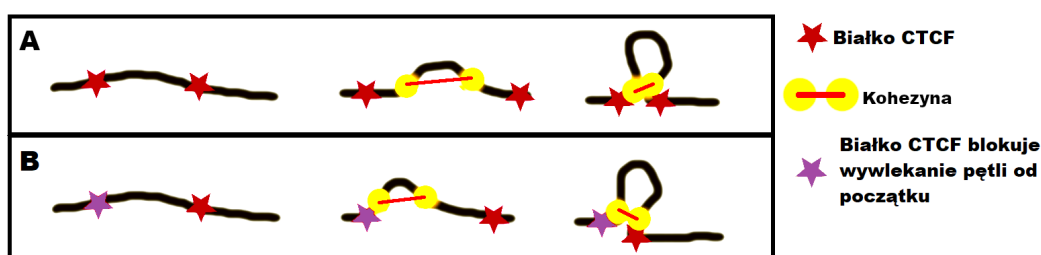


Rysunek 2: Mapy kontaktów z eksperymentów Hi-C dla sześciu regionów chromosomowych. Fioletowymi liniami zaznaczono TADy, niebieskie okręgi to miejsca pików. Rysunek na podstawie Rys. 1 u Fudenberg, et al., 2016³, dane pochodzą z pracy Rao, et al., 2014⁶.

Materiały i metody

Symulacje bazowano na prototypowym kodzie napisanym w języku Python dostarczonym przez opiekuna projektu. Symulacje opierają się na bibliotekach symulacji dynamik molekularnych OpenMM 7.4.2⁷ oraz MDtraj 1.9.3⁸. Mapy kontaktów pochodzącą z danych eksperymentalnych wygenerowano przy użyciu narzędzia Juicebox 1.5.1⁹. Filmy przedstawiające symulację zwizualizowano z użyciem narzędzia MDMovie w programie Chimera 1.14¹⁰. W Dodatku 1 przedstawiono podział prac między uczestników projektu. Kod końcowego programu do obu symulacji i wizualizowania ich za pomocą mapy kontaktów znajduje się w prywatnym repozytorium, ale dostęp do kodu źródłowego może zostać uzyskany za zgodą autorów sprawozdania.

W pracy badano dwa modele wywlekania pętli chromatynowej: model wywlekania dwustronnego (A na Rys. 3) zakładający ekstruzję pętli symetrycznie z obu stron aż do natrafienie na białka CTCF ustawione w orientacji blokującej oraz model wywlekania jednostronnego, w którym pętla jest od początku blokowana z jednej strony i ekstruzja jest jednostronna (B na Rys. 3).



Rysunek 3: Dwie hipotezy wywleknięcia pętli chromatynowej badane w projekcie. A) dwustronne; B) jednostronne. Czarna nić – włókno chromatynowe; żółte okręgi - pierścień kohezynowy; czerwone gwiazdki - kompleksy blokujące zawierające białko CTCF.

W projekcie użyto metod dynamiki molekularnej, do których wykorzystano gruboziarnisty model polimeru. Polega on na zastąpieniu grup atomów kulkami połączonymi ze sobą wiązaniami. Wykorzystanie modelu gruboziarnistego pozwala uprościć opis układu biofizycznego i zmniejszyć ilość oraz czas obliczeń potrzebnych w symulacji. Model taki w naszym przypadku reprezentuje włókno chromatynowe. Wywleknięcie pętli symulowano przez tworzenie się wiązania harmonicznego między kolejnymi parami kulek. Wiązanie harmoniczne polega na zbliżaniu się kulek zgodnie z potencjałem harmonicznym, który określa wzór¹¹:

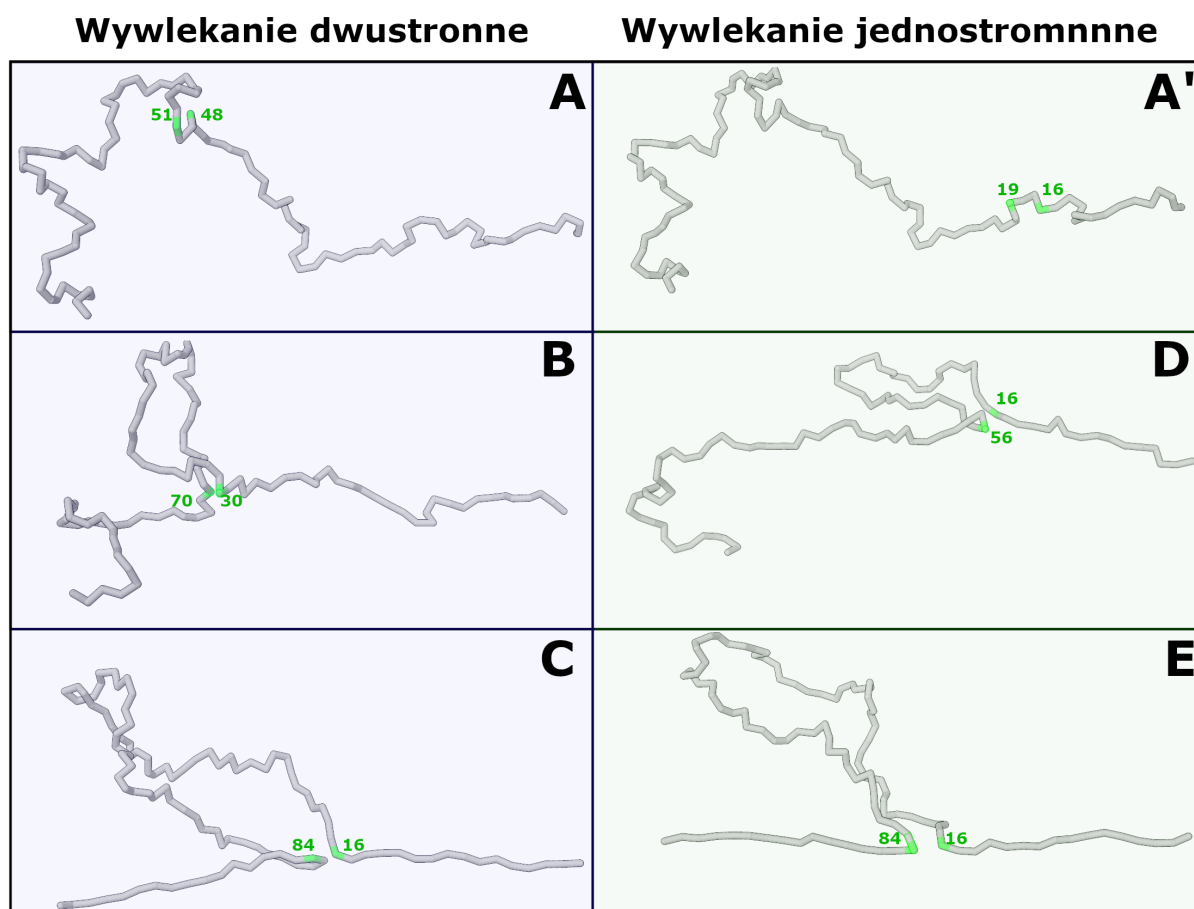
$$E = k \cdot x^2$$

gdzie k to stała siłowa wiązania w jednostkach energii podzielonych przez kwadrat jednostek odległości, a x to odległość między kulkami.

Siła siłowa określa jak mocno kulki oddziałują ze sobą. Tworzenie się pętli symulowane było poprzez płynne zmiany stałej siłowej wiązania między parą kulek i, j od początkowego $k = 3$ do $k = 0$ oraz między parą kulek $i-1, j+1$ (w przypadku wywleknięcia dwustronnego, w wywlekaniu jednostronnym było to $i, j+1$) od początkowego $k = 0$ do $k = 3$. Stała siłowa wiązań zmieniała się zgodnie z wygenerowaną sinusoidalną macierzą zawierającą wartości stałych dla każdego kroku symulacji. Co krok aktualizowane były parametry symulacji. Kod napisano w ten sposób, aby pętłe w obu rodzajach symulacji były na koniec takie same (od kulki nr 16 do nr 84, C i E na Rys. 4). W przypadku ekstruzji dwustronnej, pętla zaczynała się tworzyć od pary kulek 48-51 (A na Rys. 4), zaś dla jednostronnej zaczynała się od pary 16-19 (A' na Rys. 4).

Wyniki

Na Rysunku 4 przedstawiono klatki z początku symulacji, w połowie jej trwania i na koniec symulacji.



Rysunek 4: Dwa badane modele wywleknięcia pętli chromatynowej. A, A' odpowiednio dwustronne i jednostronne wywlekanie, początek symulacji; B, D w połowie symulacji; C, E na końcu symulacji

Dostęp do filmów przedstawiających symulację umożliwiają adresy zamieszczone poniżej.

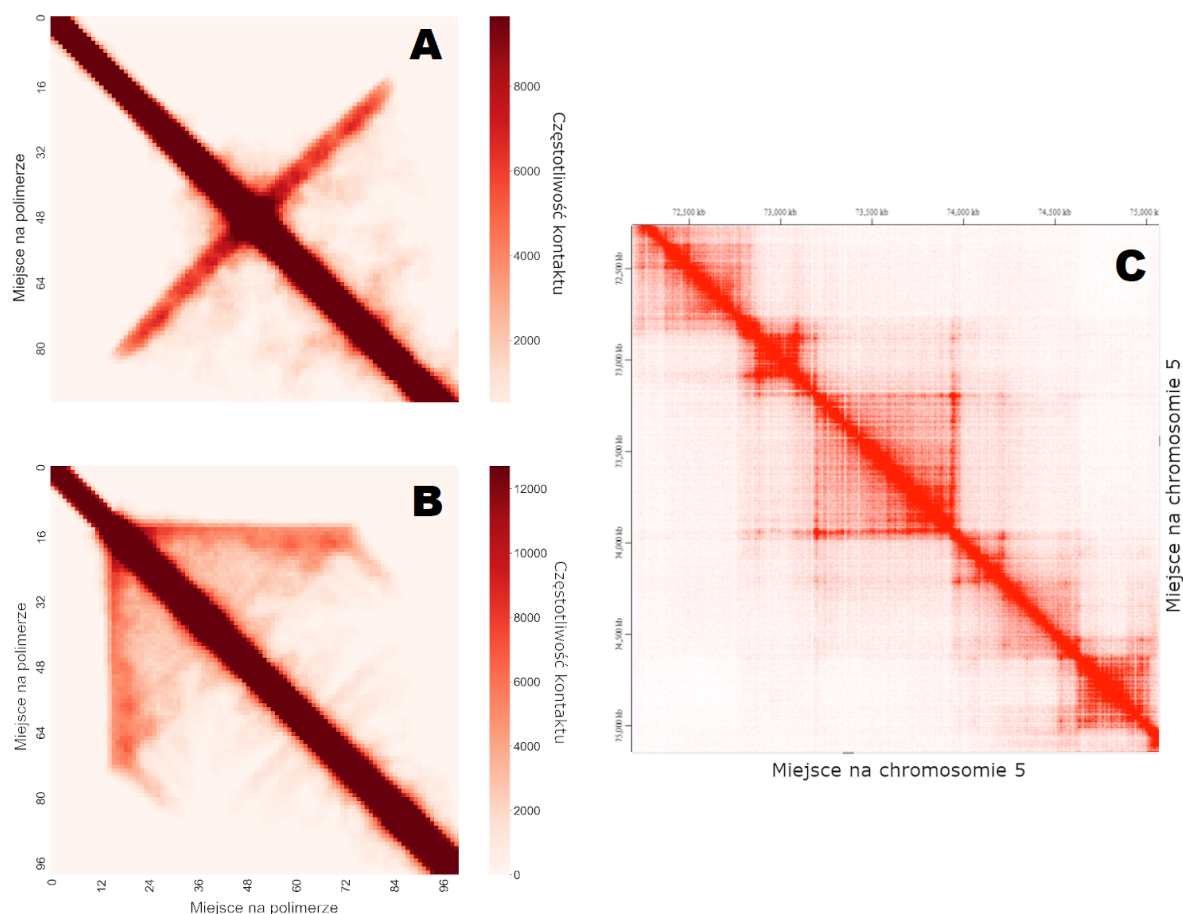
Wywleknięcia jednostronne:

<https://www.youtube.com/watch?v=xQyi1JQZWF8>

Wywleknięcia dwustronne:

<https://www.youtube.com/watch?v=F6jQcELYt8>

Na Rysunku 5 przedstawiono wyniki projektu w formie mapy kontaktów pokazującej kontakty ziaren gruboziarnistego modelu polimerowego. Wizualizuje ona sumy po macierzach kontaktów dla poszczególnego kroku symulacji. Kulki będące blisko siebie (odległość mniejsza lub równa 0,3 dobrana arbitralnie) dają większe wkłady liczbowe do macierzy kontaktów od kulek znajdujących się dalej. Skutkuje to ciemniejszym zabarwieniem mapy.



Rysunek 5: Mapy kontaktów. A) dla symulacji dwustronnego wywleknięcia pętli chromatynowej. B) dla symulacji jednostronnego wywleknięcia pętli chromatynowej. C) dla fragmentu ludzkiego chromosomu piątego. Zwizualizowano przy użyciu narzędzia Juicebox⁹. Dane eksperymentalne pochodzą z pracy Rao et al., 2014⁶. Im bardziej kolor czerwony jest intensywny, tym częściej zachodzą w tym obszarze wewnętrzne interakcje.

Przy symulacji dwustronnego wywleknięcia pętli na mapie kontaktów zauważalne są ciemne prążki odchodzące od diagonalu pod kątem prostym (A na Rys. 5), zaś dla symulacji jednostronnego wywleknięcia pętli widzimy motyw prążków odchodzących od przekątnej pod kątem 45 stopni (B na Rys. 5). Analiza wizualna poprzez porównanie z podobną mapą kontaktów uzyskaną na podstawie danych z pracy Rao et al., 2014⁶ pozwala przypuszczać, iż model jednostronnego wywleknięcia pętli chromatynowej lepiej opisuje rzeczywistość powstawania pętli chromatynowych w genomie ludzkim. Wnioski takie wymagałyby jednak przeprowadzenia symulacji rzeczywistego biofizycznego układu i kolejnych badań.

Bibliografia

- [1] Mondal, T., Rasmussen, M., Pandey, G. K., Isaksson, A. & Kanduri, C. Characterization of the RNA content of chromatin. *Genome Research* **20**, 899–907 (2010).
- [2] Cook, K. B., Hristov, B. H., Le Roch, K. G., Vert, J. P. & Noble, W. S. Measuring significant changes in chromatin conformation with ACCOST. *Nucleic Acids Research* **48**, 2303–2311 (2020).
- [3] Fudenberg, G. *et al.* Formation of Chromosomal Domains by Loop Extrusion. *Cell Reports* **15**, 2038–2049 (2016). URL <http://dx.doi.org/10.1016/j.celrep.2016.04.085>.
- [4] Rusk, N. When ChIA PETs meet Hi-C. *Nature Methods* **6**, 863 (2009).
- [5] Pombo, A. & Dillon, N. Three-dimensional genome architecture: Players and mechanisms. *Nature Reviews Molecular Cell Biology* **16**, 245–257 (2015).
- [6] Rao, S. S. *et al.* A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* **159**, 1665–1680 (2014). URL <http://dx.doi.org/10.1016/j.cell.2014.11.021>.
- [7] Eastman, P. *et al.* OpenMM 7: Rapid development of high performance algorithms for molecular dynamics. *PLoS Computational Biology* **13**, 1–17 (2017).
- [8] McGibbon, R. T. *et al.* MDTraj: A Modern Open Library for the Analysis of Molecular Dynamics Trajectories. *Biophysical Journal* **109**, 1528–1532 (2015).
- [9] Durand, N. C. *et al.* Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom. *Cell Systems* **3**, 99–101 (2016). URL <http://dx.doi.org/10.1016/j.cels.2015.07.012>.
- [10] Pettersen, E. F. *et al.* UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605–12 (2004).
- [11] WebFF. WebFF: Harmonic bond. URL <https://webff-documentation.readthedocs.io/en/latest/Reference/Bond-Harmonic.html>.

Dodatek 1. Podział prac między uczestników projektu

Łukasz Milewski	nadzorowanie prac grupy, komunikacja z opiekunem projektu, utworzenie i utrzymywanie w porządku repozytorium na github; dzielenie zadań między członków zespołu, utworzenie wstępnej macierzy zawierającej czas symulacji i siły wiązań, które się mają zmieniać; utworzenie pętli, która bierze z macierzy kolejne wartości sił wiązań, poprawki kodu; opisanie wyników i wniosków w raporcie formatowanie raportu tworzenie rysunków
Maciej Zagrodzki	uruchamianie symulacji i tworzenie filmów; stworzenie generatora sinusoidalnej oraz liniowej macierzy z wartościami sił wiązania od 0 do zadanej wielkości, optymalizacja energii, napisanie programu generującego macierz kontaktów z pliku trajektorii symulacji, napisanie kodu symulacji jednostronnej utworzenie filmów końcowych symulacji
Elżbieta Karpińska	uruchamianie symulacji, sprawdzanie poprawnego działania programu, optymalizacja programu poprzez uruchamianie symulacji z różną liczbą kroków symulacji, analiza prototypowego kodu, ustalanie strategii działania, napisanie wstępu teoretycznego w sprawozdaniu