

Streszczenie

Chromatyna - skomplikowany układ białek DNA, białek i innych czynników regulacyjnych odgrywa fundamentalną rolę w regulacji ekspresji genów i zachowaniu strukturalnym genomu. Niniejsza praca doktorska stanowi holistyczny przewodnik po komputerowych metodach trójwymiarowego modelowania struktury chromatyny. Autor we wstępie rozpoczyna od omówienia biofizycznych aspektów budowy tej cząsteczki, następnie omawia dostępne typy danych biologicznych o zastosowaniach modelarskich. Na końcu dokonuje krytycznego przeglądu istniejących metod modelarskich.

W zasadniczej części pracy autor prezentuje własne metody konstruowania modeli, a wśród nich GD-MDS (ang. *Graph Distance Multi Dimensional Scaling*) z metryką opartą o dystans grafowy, spring-model – metodę opartą o mechanikę molekularną – zaimplementowaną i udostępnioną w formie usługi WWW. W dalszej części omawia stworzone przez siebie oprogramowanie md-soft stanowiące uogólnienie metod użytych w spring-model i tworzący złożony ekosystem modelarski umożliwiający integrację komplementarnych typów danych oraz symulację ich ewolucji w czasie. Przedstawione są rozmaite scenariusze przeprowadzania symulacji w różnych skalach czasu i przestrzeni i dla różnych typów danych. Zaproponowana jest systematyczna strategia projektowania tego typu eksperymentów.

W ostatniej części przedstawione są dwa autorskie algorytmy modelowania na podstawie obrazów mikroskopowych: metoda siłowa - z wykorzystaniem techniki MDFF (ang. *Molecular Dynamics Flexible Fitting*) ora metoda korelacyjna polegająca na porównywaniu obrazu mikroskopowego z obrazem symulowanym z bieżącej konformacji struktury.

Na końcu autor omawia techniczne (w tym infrastrukturalne) aspekty modelowania i prezentuje zasady jakimi kierował się przy budowie i wdrażaniu klastra Eden^N, oraz organizacji ośrodka obliczeniowego MiNI PW - którym kieruje. Ośrodek zapewnia dostęp do największej mocy obliczeniowej w Politechnice Warszawskiej, która między innymi posłużyła do powstania niniejszej pracy, a obecnie służy ponad setce innych badaczy.

Słowa kluczowe

chromatyna, modelowanie, MDS, multi-dimensional scaling, skalowanie wielowymiarowe, MDFF, HPC, modelowanie sterowane obrazem

Abstract

Chromatin - a complex system of DNA, proteins, and other regulatory factors—plays a fundamental role in gene expression regulation and maintaining the structural integrity of the genome. This doctoral dissertation serves as a holistic guide to computational methods for three-dimensional chromatin structure modeling. The author begins with an introduction discussing the biophysical aspects of chromatin structure, followed by an overview of available types of biological data applicable to modeling. Finally, a critical review of existing modeling methods is presented.

In the main part of the dissertation, the author introduces their own modeling methods, including GD-MDS (Graph Distance Multi-Dimensional Scaling) using a graph distance-based metric and the spring-model – a molecular mechanics-based method – implemented and made available as a web service. The author further discusses their developed software, md-soft, which generalizes the methods used in the spring-model and creates a comprehensive modeling ecosystem that integrates complementary data types and simulates their evolution over time. Various simulation scenarios are described, covering different time and spatial scales and data types, along with a systematic strategy for designing such experiments.

The final section presents two original modeling algorithms based on microscopic images: a force-based method utilizing the MDFF (Molecular Dynamics Flexible Fitting) technique, and a correlation method comparing microscopic images with simulated images derived from the current structure conformation.

Lastly, the author discusses the technical (including infrastructural) aspects of modeling and outlines the principles behind the development and implementation of the EdenN cluster, as well as the organization of the MiNI PW computing center, which the author manages. The center provides access to the highest computational power at Warsaw University of Technology, which was used for this dissertation and now serves over a hundred other researchers.

Keywords

chromatin, modeling, MDS, multi-dimensional scaling, multi-dimensional scaling, MDFF, HPC, image-driven modeling.