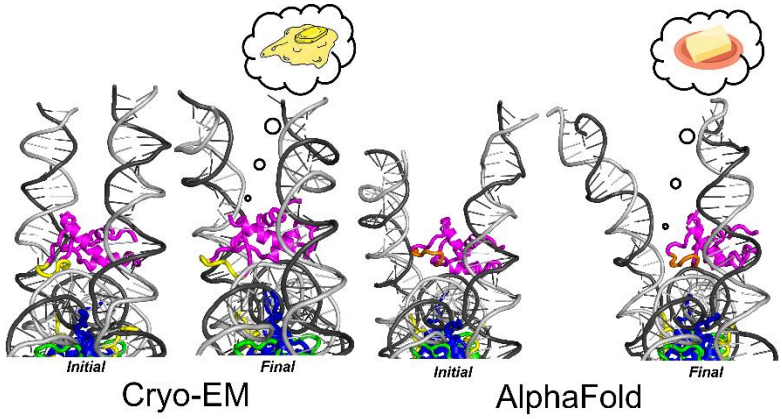
Assessment of machine-learning- versus experimental structural models of sex-specific isoforms of linker histone H1

Merve UÇA[1,2], Stefan DIMITROV[1,3], Seyit KALE[1,*]
[1]Izmir Biomedicine and Genome Center, Dokuz Eylül University Health Campus, Balçova, Izmir, Turkey
[2]Izmir International Biomedicine and Genome Institute, Dokuz Eylül University Health Campus, Balçova, Izmir, Turkey
[3]Université Grenoble Alpes, CNRS UMR 5309, INSERM U1209, Institute for Advanced Biosciences (IAB), Site Santé – Allée des Alpes, 38700 La Tronche, France

*Correspondence: seyit.kale@ibg.edu.tr

Abstract

The nucleosome is the repeating unit of chromatin and it is composed of an octamer histone core that holds together around 145 base pairs of DNA in a disc-shaped configuration. In higher eukaryotes, an additional histone, known as linker histone H1, provides further structural order by stabilizing the nucleosomal DNA termini. Histone H1 family includes seven somatic variants (H1.0 to H1.5 and H1.10) and four variants expressed only in germ cells, three of which are testis-specific (H1T, H1T2m and HILS1), and one oocyte-specific (H1.8). These isoforms are expressed differentially across cell types and stages, suggesting diverse functional roles in various physiological processes including gene expression, differentiation, aging, and cancer. In this work, we employed molecular modeling and molecular dynamics simulations to gain insights into the dynamic behavior of these sex-specific isoforms in free- and chromatin-bound states.



We surveyed the Protein Data Bank (PDB) for available linker histone structures in nucleosome-bound states. For isoforms where an experimental structure is not available or is not sufficiently reliable, we used AlphaFold predicted structures as well as SuperLooper2 structural corrections. We solvated each H1 isoform in a cubic water box to mimic a physiological buffer solution. We used the Charmm36m force field together with the OPC water model. We collected production trajectories sub-microsecond using Gromacs molecular dynamics suite, version 2018.3. We used VMD and in-house Python scripts for analysis.

During molecular dynamics simulation of H1.8 isoform in chromatin-bound state, we observed local structural instabilities such as loss of secondary structure features (graphical abstract, left panel). To resolve this issue, we generated multiple different starting models of H1.8 using conventional molecular modelling approaches as well as the more recent machine-learning based structure prediction tools. Among all the structural models we tested, the AlphaFold predicted initial structure provided the best overall stability (graphical abstract, right panel), excelling not only in comparison to the original, cryo-EM experimental model but also in comparison to the conventional molecular modelling proposals.

Encouraged by this result, we are now in the process of exploring the possible binding modes of the testis-specific H1.T to chromatin, for which only machine-learning predicted atomic models are available, albeit in isolated (not nucleosome bound) state. This endeavor is beneficial not only from a computational structural biological standpoint but also in terms of validation of machine-learning based protein structure predictions which yet lack a consideration of the effect of binding partners.

Acknowledgement and/or disclaimers, if any:
This project is funded by the European Molecular Biology Organization (EMBO) Installation Grant No 5056, Tübitak BİDEB 2232 project no 118C354, Tübitak ULAKBIM Truba High Performance Computing resource, Partnership for Advanced Computing in Europe (PRACE) project DECI-17-EPICENTROMERE, as well as local high performance computing resources of Izmir Biomedicine and Genome Center.