ADAM MICKIEWICZ UNIVERSITY IN POZNAŃ
Faculty of Biology
Institute of Human Biology and Evolution

Izabela Makałowska Ph.D.
Institute of Human Biology and Evolution
Faculty of Biology
Adam Mickiewicz University in Poznań
Poland

Poznań, 20.10.2025

Evaluation of the Doctoral Dissertation prepared by Andrea Cappannini, MSc, a PhD student at the International Institute of Molecular and Cell Biology in Warsaw

Title of the thesis:

"Development of MODOMICS and NACDDB databases and their application to studying the impact of RNA modifications on structure"

The dissertation presented for review was completed under the supervision of Prof. Janusz Bujnicki, an internationally recognized expert in structural biology. The study aimed to investigate the influence of post-transcriptional RNA modifications on RNA structure and function. The topic of the thesis is very timely and relevant given the growing interest in epitranscriptomics. The study integrates database development, data analysis, and structural bioinformatics. Three specific objectives were defined, all of which closely align with the project's primary goal. These objectives encompass two major deliverables: (i) the substantial update and redesign of MODOMICS, RNA modifications database, and (ii) the creation of a new database, NACDDB, dedicated to the collection and analysis of nucleic acid circular dichroism spectra.

The thesis has a classic structure, is clearly organized, and professionally presented. It is written in proper language however, some sentences are long and could be split up for easier reading. I did not find many technical issues, except that the labels on many figures, such as Figure 12B, Figure 13 or Figure 18 for example, were very small and often of low resolution. Therefore, they were difficult to read, even after significant enlargement.

The structure of presented thesis follows a logical progression from background, aims, results, discussion, and conclusions. In the introduction, the PhD candidate provides a detailed description of RNA structure and modifications, as well as the experimental approaches used to determine RNA structure. The candidate also discusses the methods and approaches used to predict the 3D structure of RNA, and selected RNA databases. The introduction is comprehensive and demonstrates a broad understanding of RNA chemistry and structural biology. However, it is overly descriptive at times, resembling a textbook chapter. It could be more concise, focusing on aspects directly relevant to the research questions.

The first of the two chapters that describe the results focuses on the critical evaluation and validation of RNA modification data that has been deposited in MODOMICS, as well as on the structural and evolutionary analyses of RNA modifications. Andrea Cappannini performed a comparative analysis of RNA sequence alignments and three-dimensional structures in order to identify the evolutionary signatures and structural preferences of modified nucleotides. To detect modification patterns that reflect evolutionary constraints within conserved tRNA domains, modification frequencies were quantified and visualized across Eukarya, Bacteria, and Archaea to identify conserved evolutionary signatures. Next, domain-specific and Sarcin-Ricin Loop analyses of ribosomal RNA were performed to investigate the evolutionary conservation of key modifications. Finally, a Ph.D. candidate performed a comparative analysis of experimentally determined RNA structures to investigate the stacking preferences of uridine and its derivatives.

The next chapter describes the technical implementation of the NACDDB (Nucleic Acid Circular Dichroism Database) and MODOMICS updates. The MODOMICS database stores experimentally determined RNA sequences that contain site-specific modifications. Updating this database involved restructuring RNA alignments and restoring access to deprecated data layers. Additionally, the RNA alignment section was reimplemented to enable comparative analyses of sequence conservation and chemical modification across multiple RNA classes. A comprehensive update to the protein section was done to enhance both functionality and accessibility of the data. The NACDDB database includes circular dichroism experiments involving DNA, RNA, and hybrid nucleic acid molecules. The database provides access not only to these experiments, but also to the computational framework necessary to evaluate nucleic acid conformation and make comparisons across diverse experimental conditions and structural classes.

In the discussion, the performed work is summarized. Provided is also evidence from other works supporting main findings of the thesis, which are:

- Identification of evolutionary signatures and structural preferences of modified nucleotides.
- Demonstration that dihydrouridine residues typically occur outside of stacking interactions, while pseudouridine residues are more frequently involved in stacking.
- Evaluation of the structural impact of m^1A and m^6A modifications.
- Implementation of NACDDB database and significant update of MODOMICS database.

The discussion and conclusions are followed by a chapter dedicated to future perspectives, in which the Ph.D. candidate discusses possible further developments of databases and emphasizes the necessity of increasing the interoperability of the developed tools and introducing FAIR principles. In final chapter of the thesis, described are materials and methods used to implement the databases and analyze their content.

Overall, the work presented is original and scientifically significant. The expansion of MODOMICS and the introduction of NACDDB are major contributions to the infrastructure of RNA research. Both resources are freely accessible and have been published in the leading journal Nucleic Acids Research, which underscores their international recognition and enduring value. Importantly, the thesis goes beyond technical implementation by applying these databases to investigate how chemical modifications influence RNA structural stability and evolutionary conservation. Combining database curation, structural analysis, and comparative analysis demonstrates significance and originality of presented research.

The methodological foundation of the thesis is solid and appropriate. The PhD candidate employed a wide array of computational approaches, including sequence and structure alignment tools, clustering and correlation analyses, and database integration using modern software frameworks (Python, Django, SQL, R). Descriptions of algorithms, backend architecture, and analytical pipelines are detailed and reproducible.

The thesis is an impressive and well-executed research, though there are areas that could be improved.

- I already mentioned that the introductory part could be more concise. It contains many schematic illustrations, but in some cases, these illustrations could be more useful in the later sections that describe the results.
- The thesis relies heavily on computational analysis, which is understandable in the context of bioinformatics. However, explicit cross-validation with experimental datasets (e.g. SHAPE, cryo-EM or Nanopore modification data) would lend it more weight. Has this been considered?
- The discussion would benefit from an explicit discussion of how database improvements enable new biological insights. Currently, the focus is more on technical than conceptual advancements.

However, it should be emphasized that the aforementioned weaknesses are minor and do not detract from the work's overall quality and completeness.

In summary, this doctoral dissertation is of very high quality. It represents impressive, well-executed research that combines methodological innovation, technical achievement, and biological insight. The thesis is scientifically rigorous and original and makes substantial and lasting contributions to RNA bioinformatics. The redevelopment of MODOMICS and the creation of NACDDB address significant gaps in RNA informatics infrastructure. Both databases have already attracted use and citations from the wider community, ensuring their continued impact beyond the doctoral work. Furthermore, the analyses of modification-dependent stacking and conservation patterns offer new perspectives on how post-transcriptional modifications impact RNA structure, an area of research that is advancing rapidly. The combination of technical development, scientific insight, and community contribution clearly demonstrates the candidate's capacity for independent postdoctoral-level research.

I, the undersigned, hereby state that the doctoral dissertation of Andrea Cappannini meets the requirements specified in Article 187 of the Act of July 20, 2018 – Law on Higher Education and Science (c.t., Journal of Laws of 2024, item 1571, as amended). I hereby recommend to the Doctoral Committee of the International Institute of Molecular and Cell Biology in Warsaw to admit Andrea Cappannini to the subsequent stages of the procedure for the conferment of the doctoral degree in the field of natural sciences, in the discipline of biological sciences. Considering the quality of the research and the significance of the presented results, I propose awarding the dissertation.