

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

Abstract

Post-transcriptional modifications shape RNA structure and function across all domains of life. These chemical alterations influence folding, molecular interactions, and stability, thereby affecting translational efficiency and fidelity. However, their distribution, conservation, and structural impact remain incompletely characterised. In addition to natural modifications, a wide variety of artificial chemical modifications have been introduced into nucleic acids for use in biophysical studies, structural probing, and therapeutic applications. A major part of the work presented in this thesis involves the development of databases that collect and organise diverse types of data related to nucleic acid modifications. First, a substantial expansion of the MODOMICS database was carried out, focusing on RNA modifications and their structural context. Second, a new database, NACDDB, was developed for the collection and analysis of circular dichroism (CD) spectra of nucleic acids, encompassing both naturally occurring and synthetically modified molecules. Beyond the development of these resources, the data contained within them were analysed to investigate how chemical modifications exert position- and context-dependent effects on RNA structure. A comparative analysis of RNA sequence alignments and three-dimensional structures in MODOMICS was performed to identify evolutionary signatures and structural preferences of modified nucleotides. Domain-specific analysis of ribosomal rRNA showed evolutionary conservation of key modifications, especially within the peptidyl transferase centre. The structural contexts of uridine and its modified derivatives — pseudouridine (Ψ) and dihydrouridine (D) — were examined with respect to base stacking interactions. The analysis showed that dihydrouridine residues typically occur outside of stacking interactions, while pseudouridine residues are more frequently involved in stacking, particularly with their 5'-neighbouring nucleotides, compared to uridine. This demonstrates the link between the altered stacking preferences of Ψ and D and their selection at positions where the presence or absence of stacking interactions is important for the RNA structure. Further, circular dichroism data from NACDDB were used to evaluate the structural impact of m¹A and m⁶A modifications. The data indicated that methylated adenosines alter local conformational stability, with thermal destabilisation observed at elevated temperatures. Future research integrating diverse experimental data on the positions of modified residues in RNA sequences, together with structural data for RNAs containing such modifications, will be essential to fully understand how RNA modifications influence RNA structure. Toward this goal, a collaboration has been initiated between the MODOMICS team and the developers of the Sci-ModoM database, which aims to extract modified transcript sequences from high-throughput sequencing datasets and to analyse their structural preferences.