

Addressing quality gaps and enabling data interoperability of nucleic acid structures

Bohdan Schneider, Lada Biedermannová, Paulína Božíková, Michal Malý, Terezie Prchalová,
Jakub Svoboda & Jiří Černý

Institute of Biotechnology of the Czech Academy of Sciences, 252 50 Vestec, Czech Republic

bohdan.schneider@ibt.cas.cz

While the Protein Data Bank (PDB) and community standards have fostered a robust structural biology landscape, exemplified by the widespread adoption of mmCIF dictionaries and the revolutionary success of AlphaFold2 in protein (but not NA [1]) structure prediction, the quality of nucleic acid-containing structures presents significant challenges. These issues include (i) inconsistent application of valence geometry restraints, (ii) poorly refined backbone conformations, and (iii) incomplete or incorrect assignment of base pairing. This presentation will demonstrate that addressing these quality concerns is both feasible and straightforward. We introduce web-based tools accessible at dnatco.datmos.org [2] designed to enhance nucleic acid structure annotation and validation. The Annotation TAB provides an accessible overview, while the Validation TAB offers in-depth analysis, including dinucleotide conformation analysis based on our NtC classification [3], novel rmsd/RSCC scatterplots, and valence geometry assessment. A forthcoming integrated tool will analyze base pairs, complementing our current base pair overview at basepairs.datmos.org. The Browse TAB facilitates multi-perspective structure visualization. Our ongoing development at dnatco.datmos.org aims to provide tools for structure modification during refinement.

- [1] Bohdan Schneider, Blake Alexander Sweeney, Alex Bateman, Jiří Černý, Tomasz Zok, & Marta Szachniuk: When will RNA get its AlphaFold moment? **Nucleic Acids Research** 51: 9522-9532 (2023). doi: 10.1093/nar/gkad726.
- [2] Jiří Černý, Paulína Božíková, Michal Malý, Michal Tykač, Lada Biedermannová & Bohdan Schneider: Structural alphabets for conformational analysis of nucleic acids available at dnatco.datmos.org. **Acta Crystallographica D**76: 805-813 (2020). doi: 10.1107/S2059798320009389.
- [3] Jiří Černý, Paulína Božíková, Jakub Svoboda & Bohdan Schneider: A unified dinucleotide alphabet describing both RNA and DNA structures. **Nucleic Acids Research** 48: 6367-6381 (2020). doi: 10.1093/nar/gkaa383.

This research was funded by Czech Academy of Sciences, grant RVO 86652036 and by grant LM2023055 to ELIXIR CZ from MEYS Czech Republic.