Nucleic acid building blocks: The role of tautomerism

Lukas Eberlein, Nicolas Tielker, Stefan M. Kast

Fakultät für Chemie und Chemische Biologie, TU Dortmund, Germany

Knowledge of the structure and thermodynamic properties of biologically active molecules, in particular their conformational, tautomer, and protonation state preferences is essential to understand their behavior in solution and mode of action in the organism. An accurate pH-, temperature- and pressure-dependent characterization of these properties in solution is of vital importance but poses a challenge to both experiment and theory even for well-studied compounds such as nucleic acid building blocks. Experimentally, rapid conformational changes and the fast proton transfer between multiple tautomeric forms make elucidating these equilibria cumbersome especially under extreme conditions, while the theoretical task is complicated due to the environmental effect on both electronic structure and solvent distributions.

Conceptually footed in the methodology employed within the SAMPL blind prediction challenges for tautomer equilibria, distribution coefficients and acidity constants, [1-3] we here combine tautomer and conformational sampling for natural and non-natural nucleic acid building blocks. Solvent effects on energetics and spectroscopic parameters in quantum-chemical calculations are considered using the "embedded cluster reference interaction site model" (EC-RISM) developed by us, which has been demonstrated to provide accurate estimates of thermodynamic quantities and spectroscopic features in solution even for high pressure solvents. [4,5] The EC-RISM workflow is refined by coupled-cluster extrapolation, which allows treatment of electron correlation effects with high accuracy, and by the incorporation of explicit solvent molecules to further improve the predictive quality for tautomer preferences and spectroscopic parameters. We obtain the contributions of all accessible states to the molecular ensemble as a function of pH, pressure and temperature, even for conformationally flexible species, [6] as well as their respective relevance for understanding experimental NMR spectra. This can not only lead to a further understanding of the implications of tautomerism for the base pairing in nucleic acids under extreme conditions and therefore for the universality of our genetic code and the extended Hachimoji code [7] in the universe, but also to the identification of variations in the purine scaffold of adenine and guanine species to develop new, tautomer-stable, building blocks for DNA-encoded libraries or synthetic biology.

[1] S. M. Kast, J. Heil, S. Güssregen, K. F. Schmidt, *J. Comput.-Aided Mol. Des*. **2010**, *24*, 343-353.
[2] N. Tielker, D. Tomazic, J. Heil, T. Kloss, S. Erhart, S. Güssregen, K. F. Schmidt, S. M. Kast, *J. Comput.-Aided Mol. Des*. **2016**, *30*, 1035-1044.
[3] N. Tielker, L. Eberlein, S. Güssregen, S. M. Kast, *J. Comput.-Aided Mol. Des*. **2018**, *32*, 1151-1163.
[4] R. Frach, P. Kibies, S. Böttcher, T. Pongratz, S. Strohfeldt, S. Kurrmann, J. Koehler, M. Hofmann, W. Kremer, H. R. Kalbitzer, O. Reiser, D. Horinek, S. M. Kast, *Angew. Chemie Int. Ed*. **2016**, *55*, 8757-8760.
[5] T. Pongratz, P. Kibies, L. Eberlein, N. Tielker, C. Hölzl, S. Imoto, M. Beck Erlach, S. Kurrmann, P. H. Schummel, M. Hofmann, O. Reiser, R. Winter, W. Kremer, H. R. Kalbitzer, D. Marx, D. Horinek, S. M. Kast, *Biophys. Chem*. **2020**, *257*, 106258.
[6] N. Tielker, L. Eberlein, C. Chodun, S. Güssregen, S. M. Kast, *J. Mol. Model*. **2019**, *25*, 139.
[7] S. Hoshika, N. A. Leal, M.-J. Kim, M.-S. Kim, N. B. Karalkar, H.-J. Kim, A. M. Bates, N. E. Watkins Jr., H. A. SantaLucia, A. J. Meyer, S. DasGupta, J. A. Piccirilli, A. D. Ellington, J. SantaLucia Jr., M. M. Georgiadis, S. A. Benner, *Science* **2019**, *363*, 884-887.