nature reviews physics

https://doi.org/10.1038/s42254-022-00550-3

Perspective

Check for updates

Physical non-equilibria for prebiotic nucleic acid chemistry

In the format provided by the authors and unedited

SUPPLEMENTARY INFORMATION

Physical non-equilibria for prebiotic nucleic acid chemistry

Alan Ianeselli[1,2,*], Annalena Salditt[1,*], Christof Mast[1,*], Corinna L. Kufner[3], Bettina Scheu[4] and Dieter Braun[1,#]

[1]Systems Biophysics, Center for Nanoscience, Ludwig-Maximilians-Universität München, Geschwister-Scholl-Platz 1, 80539 Munich, Germany

[2]Faculty of Computer Science, Free University of Bozen-Bolzano, NOI Techpark A.-Volta-Straße 13/A 39100 Bolzano, Italy

[3]Harvard-Smithsonian Center for Astrophysics Department of Astronomy, Harvard University, 60 Garden Street, Cambridge, MA 02138 USA

[4]Earth and Environmental Sciences, Ludwig-Maximilians-Universität München, Theresienstrasse 41, 80333 Munich, Germany

[#]Corresponding author: dieter.braun@lmu.de

Content:

1. PCA of oligonucleotides' stability

A technique to reduce dimensionality was necessary to visualize how the oligonucleotides' stability determinants (temperature, mono- and bi-valent ions, pH) synergistically interact. Data points have been generated using an oligonucleotide hybridization model based on the van 't Hoff equation[1]. It has been used to calculate the fraction of double stranded oligonucleotides as a function of every combination of temperature, salts and pH in the following ranges (as indicated in Figure 3a-d): $0 < T\ (°C) < 90$, $0 < [Na^+]\ (mM) < 300$, $0 < [Mg^{2+}]\ (mM) < 3$, $3 < pH < 7$. Results are shown in the next Table 1.1.

T (°C)	pH	Na^+ (mM)	Mg^{2+} (mM)	Duplex fraction
36.0	4.6	0	2.7	1.00
85.5	5.2	165	0.6	0.00
63.0	6.0	135	0.0	0.48
58.5	4.8	210	1.2	0.49
…	…	…	…	…

Table 1: Features used for the kPCA analysis. The last feature (duplex fraction) has been used as a label of the duplex fraction in the plot of Figure 3e.

Prior to further processing, the features have been scaled between 0 and 1. At this point, we applied a kernel Principal Component Analysis (kPCA) on the first 4 features. We used a linear kernel and reduced the dimensionality of the dataset down to 2 principal components. The last feature (the duplex fraction) has been used as a colorbar to label the features according to their duplex fraction in the plot at reduced-dimensionality (Figure 3e). For this analysis, we have used the Kernel PCA machine learning package offered by the scikit-learn Python library[2].

2. Sequence dependence of the UV damage[3]

We determined the sequence-dependent damage rates necessary for calculating the influence of UV radiation on oligonucleotide pools from available literature[4–8] . This gives approximately for the cyclobutane pyrimidine (CPD) lesions TT: 20e-3 dmg/photon, for TC/CT: 10e-3 dmg/photon and for AA: 2e-3 dmg/photon. To calculate the damage rate per photon and strand as a function of their melting temperature shown in Figure 3d, we considered all possible 7-mer sequences and summed the damage rates of the contained dimers in each respective strands. We then grouped all sequences with the same GC content corresponding to a common melting temperature[9,10] and averaged the damage rates for each of these groups. The dose of 10 photons per base used corresponds approximately to that on the surface of the early Earth after about 12h[11]. This suggests that superficial oligonucleotides would suffer greatly from UV light, so sequence-dependent selection pressure should be considered.

3. References

1. Ianeselli, A. *et al.* Water cycles in a Hadean CO2 atmosphere drive the evolution of long DNA. *Nat. Phys. 2022* 1–7 (2022). doi:10.1038/s41567-022-01516-z
2. sklearn.decomposition.KernelPCA — scikit-learn 1.1.1 documentation. Available at: https://scikit-learn.org/stable/modules/generated/sklearn.decomposition.KernelPCA.html. (Accessed: 29th June 2022)
3. Kufner, C. L. *et al.* Sequence Dependent UV Damage of Complete Pools of Oligonucleotides. *bioRxiv* 2022.08.01.502267 (2022). doi:10.1101/2022.08.01.502267
4. Johns, H. E., Pearson, M. L., LeBlanc, J. C. & Helleiner, C. W. The ultraviolet photochemistry of thymidylyl-(3′→5′)-thymidine. *J. Mol. Biol.* **9**, 503-IN1 (1964).
5. Sugiyama, T., Keinard, B., Best, G. & Sanyal, M. R. Biochemical and photochemical mechanisms that produce different UV-induced mutation spectra. *Mutat. Res. Mol. Mech. Mutagen.* **823**, 111762 (2021).
6. Law, Y. K., Azadi, J., Crespo-Hernández, C. E., Olmon, E. & Kohler, B. Predicting Thymine Dimerization Yields from Molecular Dynamics Simulations. *Biophys. J.* **94**, 3590–3600 (2008).
7. Lemaire, D. G. E. & Ruzsicska, B. P. Quantum Yields and Secondary Photoreactions of the Photoproducts of dTpdT, dTpdC and dTpdU. *Photochem. Photobiol.* **57**, 755–757 (1993).
8. Kumar, S. *et al.* Adenine photodimerization in deoxyadenylate sequences: elucidation of the mechanism through structural studies of a major d(ApA) photoproduct. *Nucleic Acids*

Res. **19**, 2841–2847 (1991).

9. Marmur, J. & Doty, P. Determination of the base composition of deoxyribonucleic acid from its thermal denaturation temperature. *J. Mol. Biol.* **5**, 109–118 (1962).

10. Von Ahsen, N., Wittwer, C. T. & Schütz, E. Oligonucleotide Melting Temperatures under PCR Conditions: Nearest-Neighbor Corrections for Mg2+, Deoxynucleotide Triphosphate, and Dimethyl Sulfoxide Concentrations with Comparison to Alternative Empirical Formulas. *Clin. Chem.* **47**, 1956–1961 (2001).

11. Ranjan, S. & Sasselov, D. D. Constraints on the Early Terrestrial Surface UV Environment Relevant to Prebiotic Chemistry. *Astrobiology* **17**, 169–204 (2017).