

Selective nucleobase pairing extends plausibly prebiotic conditions to the solid state

Tomislav Stolar,¹ Stipe Lukin,¹ Maša Rajić Linarić,² Martin Etter,³
Krunoslav Užarević,¹ Ernest Meštrović,^{2,4*} Ivan Halasz^{1*}

¹Division of Physical Chemistry, Rudjer Bošković Institute, Bijenička 54, 10000 Zagreb, Croatia

²PLIVA, Teva Active Pharmaceutical Ingredients, Prilaz baruna Filipovića 29, 10000 Zagreb, Croatia

³Deutsches Elektronen-Synchrotron (DESY), 22607 Hamburg, Germany

⁴Xellia Pharmaceuticals, Slavonska avenija 24/6, 10000 Zagreb, Croatia

*To whom correspondence should be addressed;

E-mail: ivan.halasz@irb.hr or ernest.mestrovic@xellia.com

Despite its immense relevance in the context of the emergence of life, the pairing of nucleobases has only been observed for regioselectively methylated adenine and thymine. This is surprising as it raises the question why would nucleobases be incorporated into DNA if they were unable to self-assemble beforehand. Here, we have discovered that elusive pairing of methylated guanine and cytosine is readily available in the solid state by heating where the two nucleobases self-assemble via Watson-Crick hydrogen-bonding. Dry heating preserves DNA-specificity as a four-component mixture of nucleobases provides self-assembly only of complementary pairs. We thus emphasize the importance of the solid state as the reaction medium, even for the supramolecular chemistry of life, which was thus far underexplored in the prebiotic context.

The emergence of life is among the most complex and intriguing questions, and the majority of the scientific community agrees that we are far from being able to provide an answer to it (1). Lack of historical evidence forces us to "re-invent" life and chemical systems that may undergo the transition from non-living to living (2) and recognize chemical pathways that could have led to nucleotides (3), lipids (4), and amino acids (5, 6) under plausibly prebiotic conditions. Prebiotic plausibility is flexible, but generally assumes minimal researcher's intervention or avoiding the use of elaborate laboratory designs that would not have been feasible on early Earth (7, 8).

While life itself successfully continues to evade definition, one of its features is undoubtedly replication with imminent copying of genetic information from the parent to an offspring. At the core of this is a pairing of nucleobases in DNA via specific hydrogen-bonding interactions, which thus constitutes the most famous example of molecular recognition (9). DNA replication in living organisms is enzyme-catalyzed and extremely precise. However, no enzymes for this purpose were likely present at early Earth and before the emergence of the DNA double-strand (7).

We have thus decided to explore the chemical origin of the specificity in nucleobase pairing under constraints of prebiotic plausibility (10). Canonical nucleobases are not a proper model for this task since the presence of the N–H group at the glycosidic nitrogen atom, the nitrogen linking to the sugar ring in nucleotides, contributes to hydrogen-bonding interactions and high stability of solid canonical nucleobases (11–13). We thus focus on regioselectively methylated nucleobases where hydrogen-bonding of glycosidic N–H group is hindered by a methyl group substituent (Figure 1A). Methylated nucleobases have previously been used in studies of nucleobase recognition and, except 9-methyladenine (9-mA) and 1-methylthymine (1-mT) (Figure 1B), no other complementary or non-complementary nucleobase pairs were observed to exhibit specific supramolecular recognition upon crystallization from water (14) or in the solid

state (15). In the context of the emergence of life, it, however, seems unlikely that specific pairs of nucleobases would have been coupled into DNA if they were unable to selectively and specifically self-assemble beforehand (16). Indeed, given the numerous modes the four nucleobases could potentially self assemble in pairs via hydrogen bonds (17), it is even more surprising that no other nucleobase pairs were found under plausibly prebiotic conditions.

Since solution chemistry, in general, fails to yield nucleobase pairing due to competition with solvation (14, 18), and since solid-state chemistry could be more relevant in the prebiotic context (10) than thus far anticipated (19–21), we have decided to avoid solvation issues and explore self-assembly in binary, ternary and quaternary combinations of nucleobases in the solid state by using mechanochemical ball-milling and dry heating (for details see SM). We used milling since it overcomes inherently slow solid-state diffusion through continuous comminution and growth of solid milled particles, resulting in every part of bulk solids becoming exposed to the surface and available for chemical reactivity (22). Such mechanochemical agitation has recently been considered in a prebiotic impact scenario for the synthesis of α -amino acid derivatives and sugars (23, 24), and is further supported by the fact that likely sources providing mechanochemical energy on early Earth, albeit on a larger time scale, were erosion, meteorite impacts, and earthquakes (25). Also, elevated temperatures and dry conditions, or cycles of wet and dry conditions, were likely on early Earth (26–28). Heating comes hand-in-hand with reactions in the solid state in the prebiotic context since elevated temperature could have caused water evaporation and subsequent heating of the solid precipitate.

First, we applied mechanochemical milling at room temperature to solid nucleobase mixtures. In our hands, milling was not successful in preparing any pair of complementary or non-complementary canonical or methylated nucleobases, except for the known 9-mA and 1-mT Hoogsteen pair (15) (fig. S1, for details see SM). All other milled reaction mixtures remained physical mixtures of reactants (figs. S2-S5). Next, we were wondering if heating would

have an effect on nucleobase pairing and previously milled solid mixtures of nucleobases were used for differential scanning calorimetry (DSC) experiments. The DSC thermogram of the 9-methylguanine (9-mG) and 1-methylcytosine (1-mC) physical mixture exhibited an interesting feature (Figure 2A). Before a large endotherm corresponding to melting at 287 °C, there was a subtle endotherm at 242 °C. In a repeated experiment, we have stopped heating after this small endotherm and analyzed the mixture by powder X-ray diffraction (PXRD). From the obtained diffraction pattern, it was evident that the sample was no longer a physical mixture of nucleobases. We have undertaken crystal structure determination from PXRD data to reveal a 1:1 nucleobase pair where 9-mG and 1-mC have self-assembled via Watson-Crick hydrogen-bonding, employing three hydrogen bonds between 9-mG and 1-mC molecules, as is characteristic for their pairing in DNA (Figures 2B and 2C). In the crystal, 9-mG:1-mC pairs further connect through N–H···O and N–H···N hydrogen bonds (fig. S6) to form hydrogen-bonded tapes (fig. S7).

To better understand structural changes occurring during heating, mixtures of nucleobases were subjected to controlled heating and in situ monitoring by PXRD. Canonical nucleobases were again stable in their respective pure solid forms (figs. S8-S17). Continuing with methylated nucleobases, initial heating of the 1:1 mixture of 9-mG and 1-mC exhibited only peak shifts in PXRD patterns due to unit cell expansion until the mixture was heated to 200 °C, when the diffraction pattern changed significantly indicating the formation of a new crystalline phase corresponding to 9-mG and 1-mC Watson-Crick nucleobase pair (Figure 2D).

Similarly, DSC thermogram of the 9-mA and 1-mT physical mixture exhibited a major endotherm corresponding to melting at 245 °C as well as a subtle endotherm at 180 °C (Figure 3A). During temperature-resolved PXRD in situ monitoring, the diffraction pattern of the 9-mA and 1-mT physical mixture changed at 100 °C and indicated the formation of a new crystalline phase (Figure 3B). Rietveld refinement showed that this crystalline phase corresponds to the

known 9-mA:1-mT nucleobase pair (fig. S18) previously obtained by crystallization from water (14) as well as by milling in the solid state (15). We thus confirm that Hoogsteen-type hydrogen-bonding is the preferred mode of interaction for the 9-mA:1-mT nucleobase pair in the solid state and is persistent even at elevated temperatures. This is in agreement with calculations in the water where it was shown that for 9-mA and 1-mT Hoogsteen pairing is around 1 kcal/mol more stable than the Watson-Crick pairing (29), and in vacuo where Hoogsteen pairing is preferred to stacking interactions (30). While Hoogsteen pairing usually is not present in the DNA double-strand, it is biologically relevant as it appears transiently in the DNA where it may serve in the binding of the transcription factors (31).

Noteworthy, 9-mA and 1-methyluracil (1-mU) did not form a nucleobase pair during heating in the solid state (fig. S19), and there are no endotherms other than the one attributed to melting of the physical mixture (fig. S20). All other binary mixtures of solid non-complementary methylated nucleobases failed to give any nucleobase pairs during heating, and these have remained physical mixtures (figs. S21-S27).

Since specific supramolecular recognition is at the core of selectivity in DNA replication, we explored selectivity in the formation of complementary nucleobase pairs in the solid state where physical mixtures containing more than two nucleobases in equimolar amounts were heated (figs. S28 and S29). Temperature-resolved PXRD patterns of the ternary mixture of 9-mA, 1-mT, and 1-mU (fig. S30) indicated the formation of only the 9-mA:1-mT nucleobase pair at 100 °C (Figure 4A), similarly to dry heating of the physical mixture of 9-mA and 1-mT. More interesting, dry heating of the quaternary mixture of 9-mA, 1-mT, 9-mG, and 1-mC at 250 °C, resulted in separate self-assembly and the formation of the above described complementary 9-mA:1-mT and 9-mG:1-mC nucleobase pairs (Figure 4B).

We thus show that nucleobases methylated at the glycosidic nitrogen atom achieve supramolecular self-assembly upon heating in the solid state and maintain DNA-specific complementarity

during this process. These results firmly position the solid state as the reaction medium for studying the self-assembly of other plausible purine and pyrimidine derivatives. The solid state represents an alternative pathway, absent of water, where chemical selection based on molecular recognition could have occurred as part of the prebiotic chemistry.

REFERENCES AND NOTES

1. R. Krishnamurthy, Experimentally investigating the origin of DNA / RNA on early Earth. *Nat. Commun.* **9**, 5175 (2018).
2. A. Eschenmoser, The search for the chemistry of life' s origin. *Tetrahedron* **63**, 12821–12844 (2007).
3. M. W. Powner, B. Gerland, J. D. Sutherland, Synthesis of activated pyrimidine ribonucleotides in prebiotically plausible conditions. *Nature* **459**, 239–242 (2009).
4. D. Segre, D. Ben-Eli, D. W. Deamer, D. Lancet, The lipid world. *Orig Life Evol Biosph* **31**, 119–145 (2001).
5. S. L. Miller, A Production of Amino Acids Under Possible Primitive Earth Conditions. *Science* **117**, 528–530 (1953).
6. B. H. Patel, C. Percivalle, D. J. Ritson, C. D. Duffy, J. D. Sutherland, Common origins of RNA, protein and lipid precursors in a cyanosulfidic protometabolism. *Nat. Chem.* **7**, 301–307 (2015).
7. C. Richert, Prebiotic chemistry and human intervention. *Nat. Commun.* **9**, 5177 (2018).
8. D. Whitaker, M. W. Powner, Prebiotic nucleic acids need space to grow. *Nat. Commun.* **9**, 5172 (2018).

9. J. D. Watson, F. H. C. Crick, Genetical Implications of the Structure of Deoxyribonucleic Acid. *Nature* **171**, 964–967 (1953).
10. L. E. Orgel, Prebiotic Chemistry and the Origin of the RNA World. *Crit. Rev. Biochem. Mol. Biol.* **39**, 99–123 (2004).
11. R. E. A. Kelly, Y. J. Lee, L. N. Kantorovich, Homopairing Possibilities of the DNA Base Adenine. *J. Phys. Chem. B* **109**, 11933–11939 (2005).
12. R. E. A. Kelly, Y. J. Lee, L. N. Kantorovich, Homopairing Possibilities of the DNA Bases Cytosine and Guanine : An ab Initio DFT Study. *J. Phys. Chem. B* **109**, 22045–22052 (2005).
13. R. E. A. Kelly, L. N. Kantorovich, Homopairing Possibilities of the DNA Base Thymine and the RNA Base Uracil : An ab Initio Density Functional Theory Study. *J. Phys. Chem. B* **110**, 2249–2255 (2006).
14. K. Hoogsteen, The Crystal and Molecular Structure of a Hydrogen-Bonded Complex Between 1-Methylthymine and 9-Methyladenine * (I). *Acta Cryst.* **16**, 907–916 (1963).
15. M. C. Etter, S. M. Reutzler, C. G. Choo, Self-organization of Adenine and Thymine in the Solid State. *J. Am. Chem. Soc.* **115**, 4411–4412 (1993).
16. N. V. Hud, S. S. Jain, X. Li, D. G. Lynn, Addressing the Problems of Base Pairing and Strand Cyclization in Template- Directed Synthesis. *Chem. Biodivers.* **4**, 768–783 (2007).
17. J. Donohue, K. N. Trueblood, Base pairing in DNA. *J. Mol. Biol.* **2**, 363–371 (1960).
18. P. Cieplak, P. A. Kollman, Calculation of the Free Energy of Association of Nucleic Acid Bases in Vacuo and Water Solution. *J. Am. Chem. Soc.* **110**, 3734–3739 (1988).

19. S. Becker, I. Thoma, A. Deutsch, T. Gehrke, P. Mayer, H. Zipse, T. Carell, A high-yielding, strictly regioselective prebiotic purine nucleoside formation pathway. *Science* **352**, 833–836 (2016).
20. S. Becker, C. Schneider, M. Dejmek, T. Carell, H. Okamura, A. Crisp, Wet-dry cycles enable the parallel origin of canonical and non-canonical nucleosides by continuous synthesis. *Nat. Commun.* **9**, 163 (2018).
21. N. V. Hud, Searching for lost nucleotides of the pre-RNA World with a self-refining model of early Earth. *Nat. Commun.* **9**, 5171 (2018).
22. S. Lukin, M. Tireli, T. Stolar, D. Barišić, M. V. Blanco, M. di Michiel, K. Užarević, I. Halasz, Isotope labeling reveals fast atomic and molecular exchange in mechanochemical milling reactions. *J. Am. Chem. Soc.* **141**, 1212-1216 (2019).
23. C. Bolm, R. Mocci, C. Schumacher, M. Turberg, F. Puccetti, J. G. Hernandez, Mechanochemical Activation of Iron Cyano Complexes: A Prebiotic Impact Scenario for the Synthesis of α -Amino Acid Derivatives. *Angew. Chem. Int. Ed.* **57**, 2423–2426 (2018).
24. S. Lamour, S. Pallmann, M. Haas, O. Trapp, Prebiotic Sugar Formation Under Nonaqueous Conditions and Mechanochemical Acceleration. *Life* **9**, 52 (2019).
25. S. Lherminier, R. Planet, V. L. Vehel, G. Simon, L. Vanel, K. J. Måløy, O. Ramos, Continuously Sheared Granular Matter Reproduces in Detail Seismicity Laws. *Phys. Rev. Lett.* **122**, 218501 (2019).
26. H. J. Cleaves, Prebiotic Chemistry: Geochemical Context and Reaction Screening. *Life* **3**, 331–345 (2013).

27. B. Damer, D. Deamer, Coupled Phases and Combinatorial Selection in Fluctuating Hydrothermal Pools: A Scenario to Guide Experimental Approaches to the Origin of Cellular Life. *Life* **5**, 872–887 (2015).
28. S. Becker, C. Schneider, T. Carell, Non-canonical nucleosides and chemistry of the emergence of life. *Nat. Commun.* **9**, 5174 (2018).
29. I. R. Gould, P. A. Kollman, Theoretical Investigation of the Hydrogen Bond Strengths in Guanine-Cytosine and Adenine-Thymine Base Pairs. *J. Am. Chem. Soc.* **116**, 2493–2499 (1994).
30. P. Jurec, P. Hobza, True Stabilization Energies for the Optimal Planar Hydrogen-Bonded and Stacked Structures of Guanine...Cytosine, Adenine...Thymine, and Their 9- and 1-Methyl Derivatives: Complete Basis Set Calculations at the MP2 and CCSD(T) Levels and Comparison with Experiment. *J. Am. Chem. Soc.* **125**, 15608–15613 (2003).
31. E. N. Nikolova, E. Kim, A. A. Wise, P. J. O. Brien, I. Andricioaei, H. M. Al-Hashimi, Transient Hoogsteen base pairs in canonical duplex DNA. *Nature* **470**, 498–502 (2011).

ACKNOWLEDGEMENTS

We are grateful to Lidija Čubra, Marina Guja, Karmen Čolakić, Helena Cerić and Mirta Rubčić for their assistance in the course of this study. We are grateful to DESY (Hamburg) for beam-time. **Funding:** We are thankful to the Rudjer Bošković Institute (Zagreb) for financial support. S.L. is supported by the Croatian Science Foundation. **Author contributions:** E.M. conceived the idea of nucleobase pairing in the solid state. I.H., T.S., and E.M. designed the study. T.S. performed the experiments. M.R.L. performed DSC experiments. T.S. and M.E. performed synchrotron measurements. T.S., S.L., I.H. and K.U. analyzed the data. I.H. and T.S. solved

the crystal structure. S.L., T.S. and I.H. prepared figures. I.H. and E.M. supervised the study. T.S. and I.H. wrote the manuscript with input from all authors. **Competing interests:** Authors declare no competing interest. **Data and materials availability:** All data needed to evaluate the conclusions in the paper is available in the main text or the supplementary materials.

Supplementary materials

Materials and Methods

List of figures

Figs. S1 to S35

Table S1

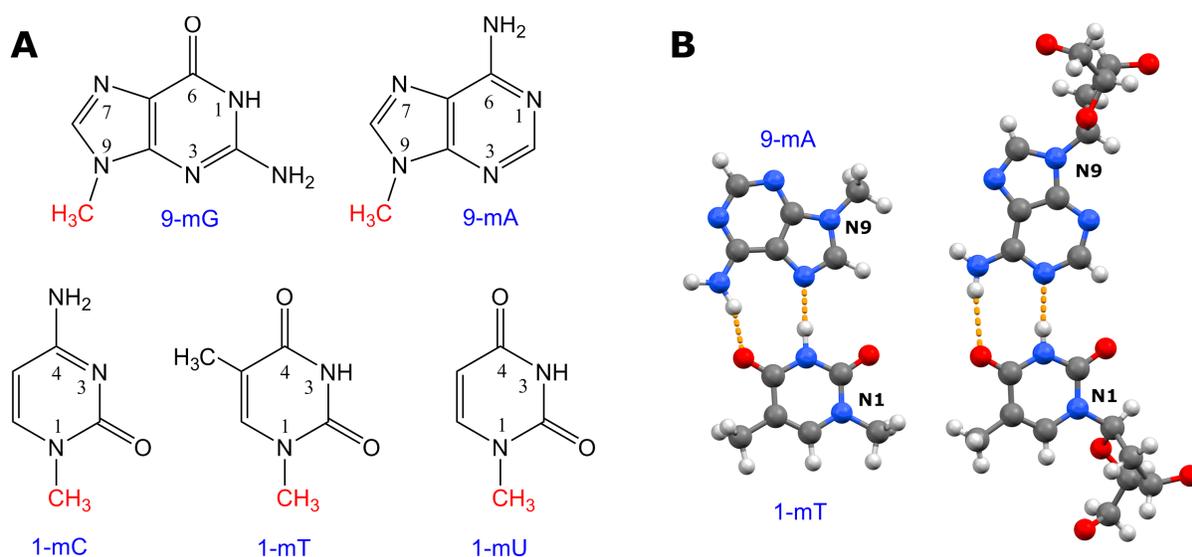


Figure 1. Studying chemical origins of specificity in nucleobase pairing. (A) We used nucleobases in which nitrogen atom, that bonds to sugar moiety in nucleotides, is regioselectively methylated on N9 position in purines (adenine and guanine) and N1 position in pyrimidines (thymine, uracil, and cytosine). (B) On the left is known 9-mA:1-mT nucleobase pair (CSD code: MTHMAD13), obtained upon crystallization from water and milling in the solid state, enabled via Hoogsteen hydrogen-bonding. On the right is Watson-Crick base pairing of adenine and thymine from the DNA (PDB code: 6CQ3).

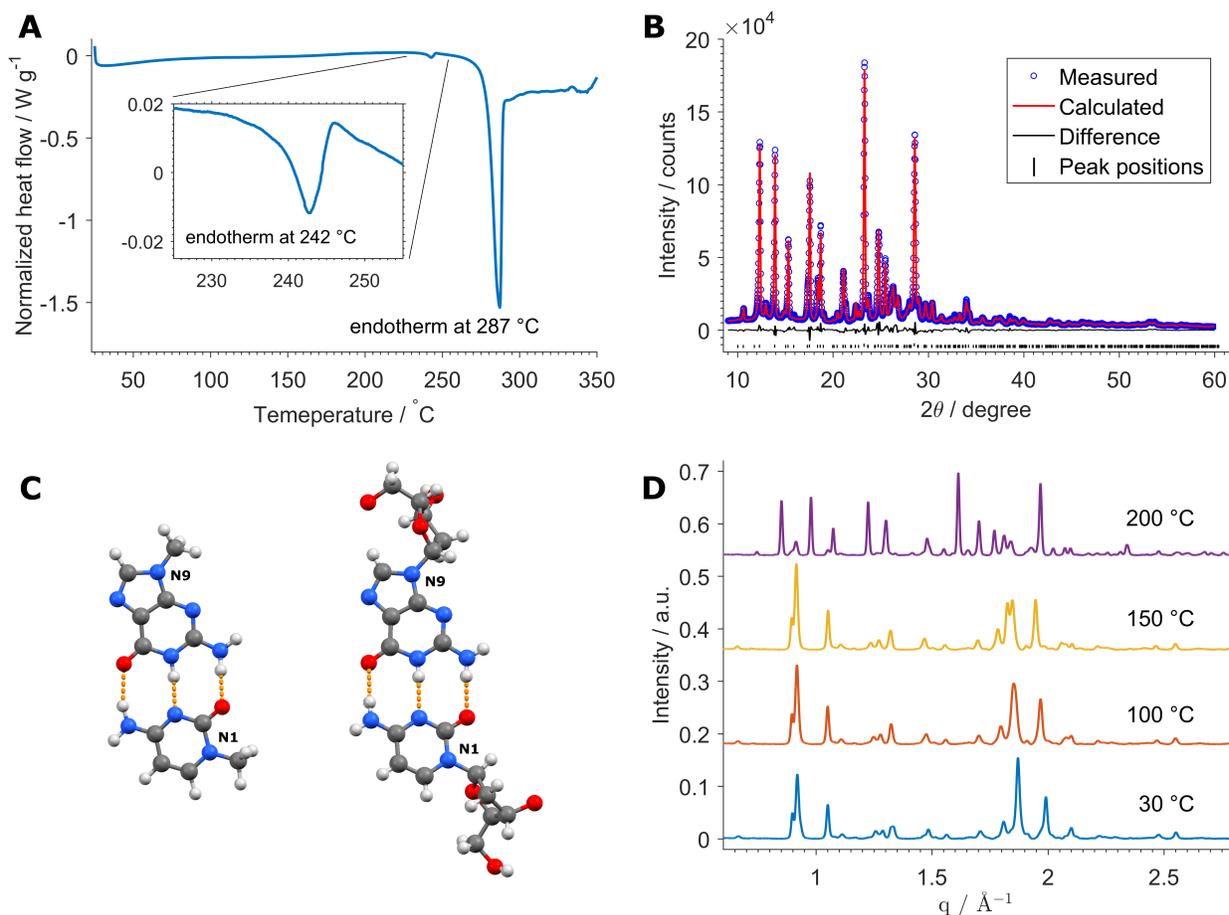


Figure 2. Structural characterization of 9-mG and 1-mC nucleobase pair. (A) DSC thermogram of the pre-milled physical mixture of 9-mG and 1-mC collected with $5^{\circ}\text{C}/\text{min}$ heating rate. (B) Rietveld fit for crystal structure determination of 9-mG and 1-mC nucleobase pair ($\lambda=1.54 \text{ \AA}$). (C) Comparison of Watson-Crick hydrogen-bonding between 9-mG and 1-mC nucleobase pair (on the left) and guanine and cytosine from the DNA (PDB code: 6CQ3, on the right). (D) In situ monitoring by synchrotron PXRD ($\lambda=0.207 \text{ \AA}$) of dry heating pre-milled physical mixture of 9-mG and 1-mC with $5^{\circ}\text{C}/\text{min}$ heating rate. The inset in (A) is zoomed region showing subtle endotherm at 242°C .

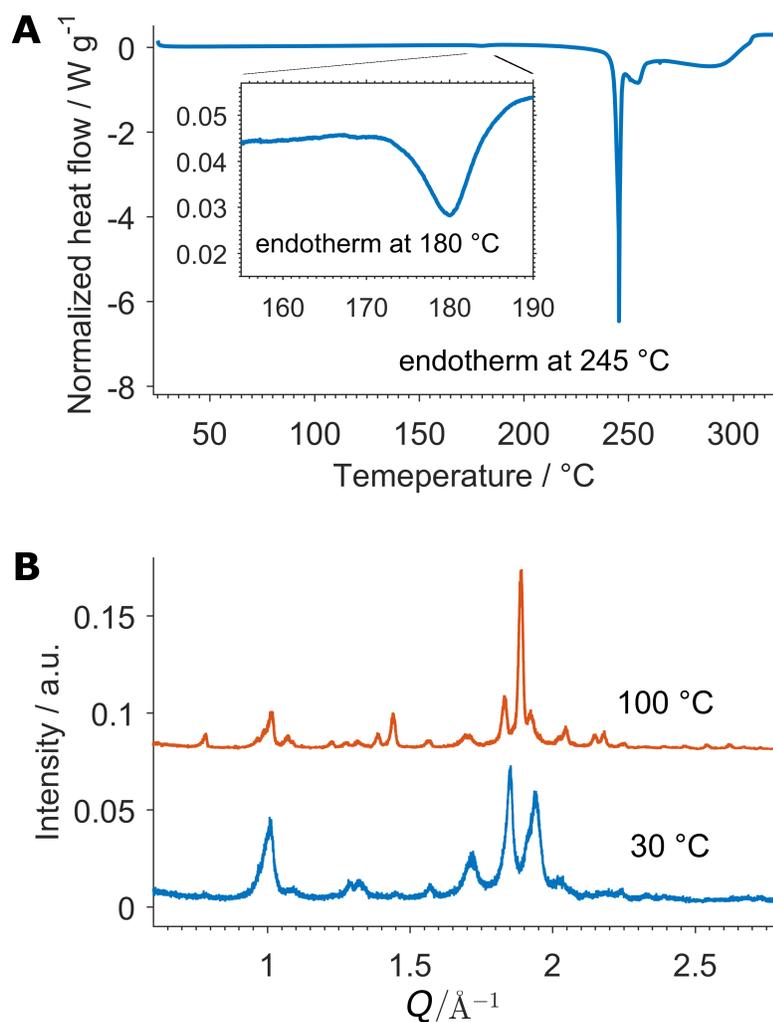


Figure 3. Structural characterization of 9-mA and 1-mT nucleobase pair. (A) DSC thermogram of the pre-milled physical mixture of 9-mA and 1-mT collected with 5 $^{\circ}\text{C}/\text{min}$ heating rate. (B) In situ monitoring by laboratory PXRD ($\lambda=1.54 \text{ \AA}$) of dry heating pre-milled physical mixture of 9-mA and 1-mT with 5 $^{\circ}\text{C}/\text{min}$ heating rate. The inset in (A) is zoomed region showing subtle endotherm at 180 $^{\circ}\text{C}$.

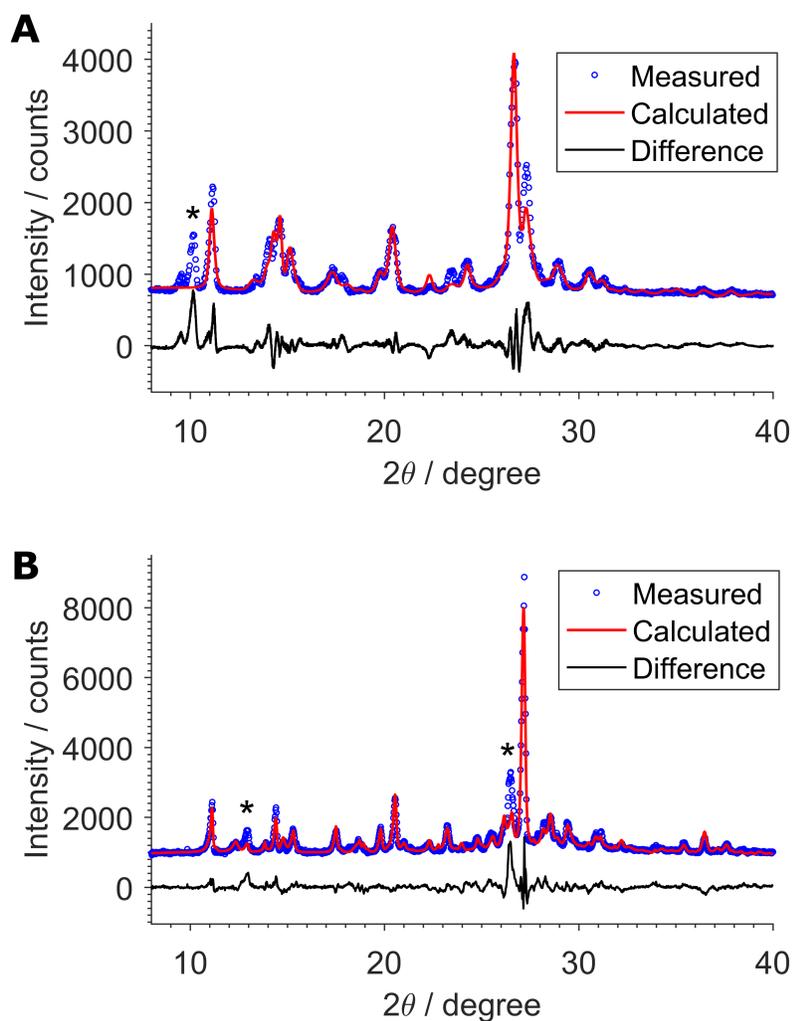


Figure 4. Rietveld plots for selectivity experiments of dry heating heterogeneous mixtures of *N*-methylated nucleobases. (A) The pre-milled ternary mixture of 9-mA, 1-mT, and 1-mU after dry heating at 100 °C. Calculated using crystal structures of the 9-mA:1-mT nucleobase pair (CSD code: MTHMAD13), 9-mA (CSD code: MEADEN01) and 1-mT (CSD code: METHYM04). 1-mU was not evident in the diffraction pattern, and an asterisk marks an unattributed diffraction peak. (B) The pre-milled quaternary mixture of 9-mG, 1-mC, 9-mA, and 1-mT after dry heating at 250 °C. Calculated using crystal structures of the 9-mG:1-mC nucleobase pair, 9-mA:1-mT nucleobase pair (CSD code: MTHMAD13), 9-mA (CSD code: MEADEN01), 1-mT (CSD code: METHYM04) and 1-mC (CSD code: METCYT01). Crystal structure of 9-mG is unknown, and asterisks mark two strongest diffraction peaks corresponding to 9-mG.