DNA-specific selectivity in pairing of model nucleobases in the solid state

Tomislav Stolar,[a] Stipe Lukin,[a] Martin Etter,[b], Maša Rajić Linarić,[c] Krunoslav Užarević[a] Ernest Meštrović,*[d] and Ivan Halasz*[a]

Nucleobases methylated at the glycosidic nitrogen atom achieve DNA-specific self-assembly upon heating in the solid state. We report formation and characterisation of the elusive cocrystal of methylated guanine and methylated cytosine, exhibiting Watson-Crick-type hydrogen bonding, and the crystal structure of 9-methylguanine.

The pairing of nucleotides in the DNA via specific hydrogen-bonding interactions constitutes the most famous example of supramolecular recognition. Following the discovery of the structure of the DNA in 1953,[1] and that complementarity between the two DNA strands is based on supramolecular interactions, numerous studies focused on hydrogen bonding between nucleobases to obtain detailed geometries of their interactions.[2–5] Among canonical nucleobases or their simple derivatives, supramolecular pairing under plausibly prebiotic conditions was observed only for adenine and thymine methylated at the glycosidic nitrogen atom, which was achieved by crystallisation from water,[6] but also in the solid state[7] (Fig. 1). Several other derivatives of adenine and thymine as well as derivatives of guanine and cytosine have been found to pair via Hoogsteen or Watson-Crick hydrogen bonding.[3,5,8] These pairs have, however, usually been obtained from a dimethylsulphoxide (DMSO) solution due to the limited solubility of nucleobases in water. Noteworthy, Etter attempted to achieve pairing of 9-ethylguanine and 1-methylcytosine by grinding in the solid state, a pair previously obtained from a heated DMSO solution,[9] but was not successful.[7]

Here, using the simplest derivatives of canonical nucleobases, we have discovered that pairing of methylated guanine and methylated cytosine is readily achievable in the solid state by heating their physical mixture, where they self-assemble in the Watson-Crick hydrogen-bonded motif. We also explored selectivity in solid-state recognition and found it to be DNA-specific. The four-component mixture of methylated adenine, guanine, cytosine, and thymine provided pairing only of complementary bases.

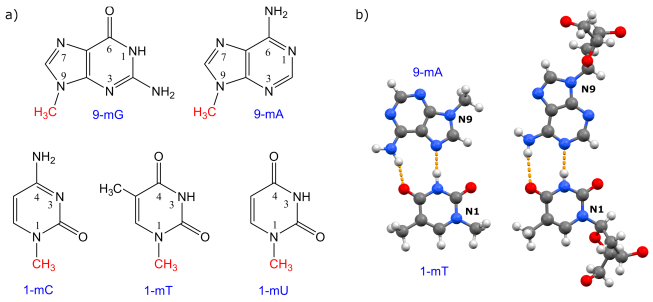

Fig. 1 (a) Nucleobases having the glycosidic nitrogen atom regioselectively methylated (position N9 in purines and position N1 in pyrimidines) were used to study selectivity in nucleobase pairing in the solid state. (b) On the left is the known Hoogsteen-type 9-mA:1-mT nucleobase pair (CSD code: MTHMAD13[10]) obtained upon crystallization from water and by milling in the solid state. On the right is Watson-Crick-type base pairing between adenine and thymine nucleotides in DNA (PDB code: 6CQ3).

Having first unsuccessfully explored pairing of canonical solid nucleobases by ball milling,[11] we realised that these are not a proper model for this task. The presence of an N−H group at the glycosidic nitrogen atom, the nitrogen linking to the sugar ring in nucleotides, contributes to hydrogen-bonding interactions interfering with potential recognition at other faces of the pyrimidine and purine bases. In addition, the presence of this N−H group leads to highly stable solid canonical nucleobases.[12–15] We have thus switched our focus to methylated nucleobases, where hydrogen-bonding involving the glycosidic nitrogen atom is disabled by the methyl group (Fig. 1).

While the methyl substituent limits the usefulness of these model nucleobases in the prebiotic context, these results may still provide an insight into supramolecular selectivity in the origin-of-life context since it should be unlikely that the specific pairs of nucleobases would have been coupled into DNA if they were unwilling to selectively and specifically self-assemble beforehand.[16] In addition, given the numerous modes the five nucleobases could self-assemble via hydrogen bonds,[17] it is even more surprising that no other cocrystals of nucleobases or their derivatives were found under plausibly prebiotic conditions.[18]

[a] *Ruđer Bošković Institute, Bijenička c. 54, 10000 Zagreb, Croatia. E-mail: ivan.halasz@irb.hr*
[b] *Deutsches Elektronen-Synchrotron (DESY), 22607 Hamburg, Germany*
[c] *PLIVA Croatia Ltd., TAPI R&D, Prilaz baruna Filipovića 25, 10000 Zagreb, Croatia*
[d] *Xellia Pharmaceuticals, Slavonska avenija 24/6, 10000 Zagreb, Croatia Email: ernest.mestrovic@xellia.com*

† Electronic Supplementary Information (ESI) available: Additional information about crystal structure determination, PXRD, DSC, FTIR-ATR data, CCDC 1919850 and 1995351.

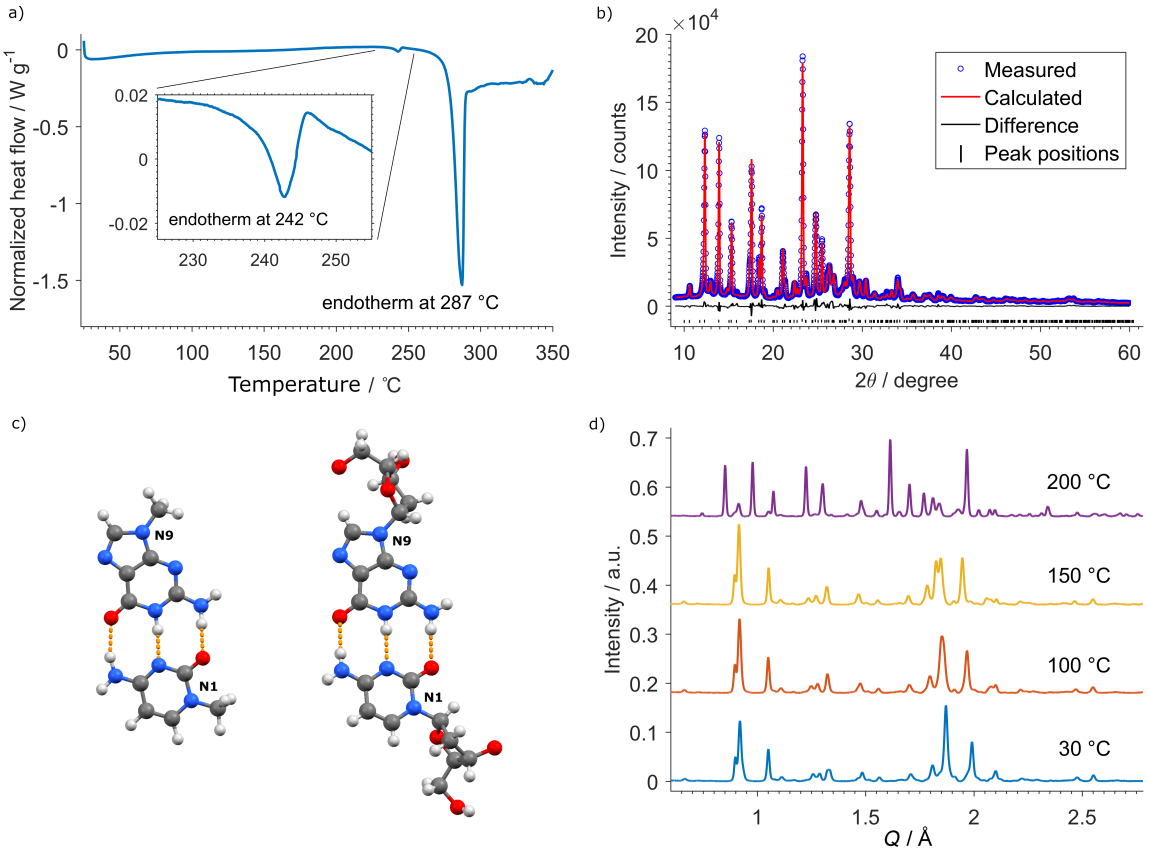

Fig. 2 (a) DSC thermogram of the pre-milled, stoichiometric 1:1 physical mixture of 9-mG and 1-mC. The inset highlights the subtle endotherm at 242 °C. (b) Rietveld fit for the crystal structure of the 1:1 9-mG:1-mC cocrystal (λ=1.54 Å). (c) Comparison of Watson-Crick hydrogen bonding between 9-mG and 1-mC in the cocrystal (to the left) and guanine and cytosine nucleotides from the DNA (PDB code: 6CQ3, to the right). (d) Temperature-resolved monitoring by synchrotron PXRD (λ=0.207 Å) of the pre-milled physical mixture of 9-mG and 1-mC.

In solution, nucleobase recognition competes with solvation.[6,19] Since solid-state chemistry could be more relevant in the prebiotic context then thus far anticipated,[20–25] we have decided to avoid solvation issues and explored self-assembly in binary, ternary and quaternary combinations of methylated nucleobases in the solid state by using ball milling and dry heating (for details see ESI). First, we applied milling at room temperature to 1:1 solid nucleobase mixtures. The resulting samples were analysed by powder X-ray diffraction (PXRD) and, except for reproducing the preparation of the known 9-mA:1-mT Hoogsteen pair[7] (Fig. S6, for details see ESI), we were not successful in preparing any other pair of complementary or non-complementary canonical or methylated nucleobases, and the milled mixtures remained physical mixtures of reactants (Figs. S7-S10, ESI).

Next, we were wondering if heating would have an effect on nucleobase pairing. Previously milled physical mixtures of nucleobases were thus subjected to heating in differential scanning calorimetry (DSC) experiments. The pairing of nucleobases was not observed for any combination of canonical nucleobases and all pairs of non-complementary methylated nucleobases. We did, however, observe pairing of methylated guanine and cytosine. The DSC thermogram of the 9-methylguanine (9-mG) and 1-methylcytosine (1-mC) physical mixture exhibited a subtle endotherm at 242 °C before a large melting endotherm at 287 °C (Fig. 2a).

In a repeated experiment, heating was interrupted after this subtle endotherm and the sample was analysed by PXRD. From the obtained diffraction pattern, it was evident that the sample was highly crystalline and no longer a physical mixture of nucleobases. Crystal structure determination from PXRD data revealed a 1:1 cocrystal 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 (Fig. 2b and 2c). In the crystal, 9-mG:1-mC pairs further connect through N–H···O and N–H···N hydrogen bonds (Fig. S1) to form hydrogen-bonded tapes (Fig. S2).

In the course of this study, we have been able to find in the literature the relevant crystallographic information for all nucleobases and their methylated derivatives, except for the crystal structure of 9-mG. Its crystal structure was solved here from high-resolution powder diffraction data revealing a complex structure in the monoclinic $P2_1/c$ space group with 4 molecules comprising the asymmetric unit (Fig. S3). Molecules of 9-mG form puckered hydrogen-bonded layers in the crystal structure with weaker interactions between layers (Figs. S4 and S5).

To better understand structural changes occurring during heat-

ing, mixtures of nucleobases were subjected to controlled heating and *in situ* monitoring by synchrotron PXRD. Canonical nucleobases were again stable each in its respective pure solid phase (Figs. S11-S20). 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 the 1:1 9-mG:1-mC cocrystal (Fig. 2d).

Similarly, a DSC thermogram of the 1:1 physical mixture of 9-mA and 1-mT exhibited the major endotherm corresponding to melting at 245 °C, but also a subtle endotherm at 180 °C (Fig. 3a) corresponding to the chemical reaction of nucleobase pairing. Rietveld analysis of diffraction patterns collected in a temperature-resolved synchrotron PXRD monitoring experiment confirmed the formation of the known 9-mA:1-mT cocrystal (Figs. 3b and S21). We thus confirm a preference for Hoogsteen-type hydrogen-bonding for the 9-mA:1-mT nucleobase pair that is persistent also at elevated temperatures. This is in agreement with calculations in water where it was shown that for 9-mA and 1-mT the Hoogsteen pairing is around 1 kcal/mol more stable than the Watson-Crick pairing,[26] and *in vacuo* where Hoogsteen pairing is preferred to stacking interactions.[27] While Hoogsteen pairing usually is not present in the DNA double-strand, it is biologically relevant[28–30] and may appear transiently to 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. S22), and there are no endotherms other than the one attributed to melting of the physical mixture (Fig. S23).

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 from physical mixtures containing more than two nucleobases (Figs. S31 and S32). Temperature-resolved PXRD patterns of the ternary mixture of 9-mA, 1-mT, and 1-mU (Fig. S33) indicated the formation of only the 9-mA:1-mT nucleobase pair at 100 °C, similarly to dry heating of the physical mixture of 9-mA and 1-mT. The quaternary mixture of 9-mA, 1-mT, 9-mG, and 1-mC, heated above 200 °C, provided a 6-component mixture of separate cocrystals of complementary nucleobase pairs 9-mA:1-mT and 9-mG:1-mC in addition to unreacted starting nucleobases (Fig. 4). Presence of starting nucleobases in the heated mixture is not surprising since purely heating does not involve mixing of the powder mixture and the low mobility of solid particles has likely prevented the whole of the sample to become engaged.

In summary, pairing of nucleobases into hydrogen-bonded supramolecular complexes is surprisingly difficult to achieve, despite each nucleobase having a number of hydrogen-bond donor and acceptor groups. Having here focused on pairing in the solid state, our attempts to achieve supramolecular recognition among canonical nucleobases, as well as among their derivatives methylated at the glycosidic nitrogen, provided only one new supramolecular complex. The cocrystal of methylated guanine and methylated cytosine exhibiting Watson-Crick hydrogen bonding was obtained upon heating the solid physical mixture of components. We also reproduced solid-state formation of the cocrystal of methylated adenine and methylated thymine and noteworthy, observed DNA-specific selectivity in pairing from ternary and quaternary mixtures of methylated nucleobases. We are currently exploring the potential of the solid-state as a prebiotic reaction medium for emergence of recognition patterns among purine and pyrimidine derivatives that could have been present on early Earth.[32–34]

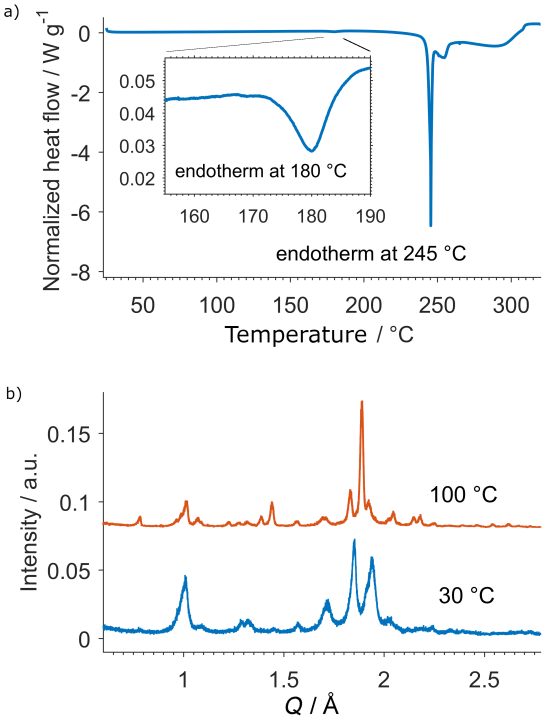

Fig. 3 (a) DSC thermogram of the pre-milled, stoichiometric 1:1 physical mixture of 9-mA and 1-mT. (b) In situ monitoring by laboratory PXRD (λ=1.54 Å) of dry heating pre-milled physical mixture of 9-mA and 1-mT. The inset in (a) highlights the subtle endotherm at 180 °C.

Conflicts of interest

There are no conflicts of interest to declare.

Acknowledgments

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 the Ruđer Bošković Institute (Zagreb) for financial support and DESY (Hamburg) for beamtime at the beamline P02.1. SL is supported by the Croatian Science Foundation.

Notes and references

1 J. D. Watson and F. H. C. Crick, *Nature*, 1953, **171**, 964–967.
2 Y. Kyogoku, R. C. Lord and A. Rich, *Science*, 1966, **154**, 518–520.
3 D. Voet and A. Rich, *A. Rich*, Academic Press, 1970, vol. 10, pp. 183 – 265.
4 J. Donohue, *Proc. Nat. Acad. Sci.*, 1956, **42**, 60–65.
5 F. S. Mathews and A. Rich, *Journal of Molecular Biology*, 1964, **8**, 89 – 95.
6 K. Hoogsteen, *Acta Cryst.*, 1963, **16**, 907–916.

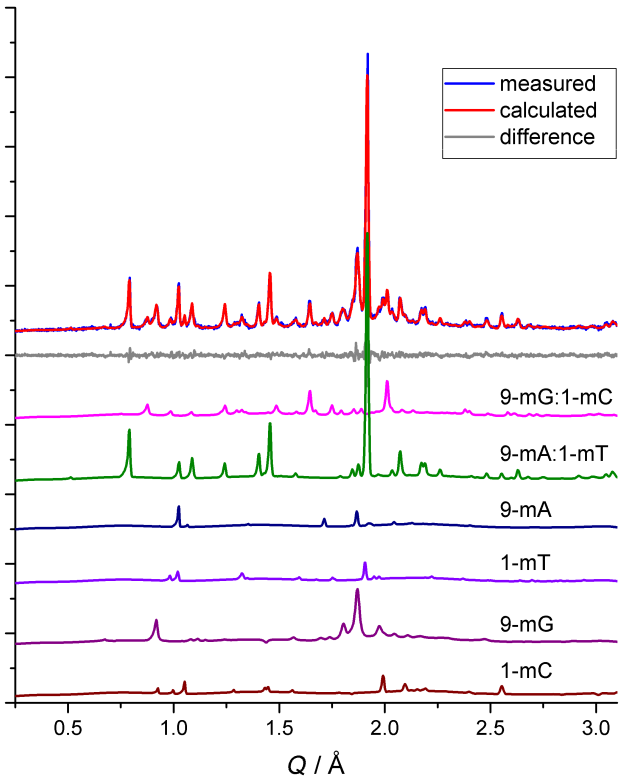

Fig. 4 Rietveld plot for the pre-milled quaternary mixture of 9-mG, 1-mC, 9-mA, and 1-mT after dry heating at 250 °C. The sample contains six phases: the 9-mG:1-mC and 9-mA:1-mT cocrystals, 9-mA, 1-mT, 1-mC and the here solved crystal structure of 9-mG. Data collected using radiation from a copper anode in Bragg-Brentano geometry. Preferred orientation was modelled for some phases.

7 M. C. Etter, S. M. Reutzel and C. G. Choo, *J. Am. Chem. Soc.*, 1993, **115**, 4411–4412.
8 H. M. Sobell, K.-i. Tomita and A. Rich, *Proc. Nat. Acad. Sci.*, 1963, **49**, 885–892.
9 E. J. O'Brien, *Acta Cryst.*, 1967, **23**, 92–106.
10 K. N. Jarzembska, A. M. Goral, R. Gajda and P. M. Dominiak, *Cryst. Growth Des.*, 2013, **13**, 239–254.
11 A. V. Trask and W. Jones, in *Crystal Engineering of Organic Cocrystals by the Solid-State Grinding Approach*, ed. F. Toda, Springer Berlin Heidelberg, Berlin, Heidelberg, 2005, pp. 41–70.
12 R. E. A. Kelly, Y. J. Lee and L. N. Kantorovich, *J. Phys. Chem. B*, 2005, **109**, 11933–11939.
13 R. E. A. Kelly, Y. J. Lee and L. N. Kantorovich, *J. Phys. Chem. B*, 2005, **109**, 22045–22052.
14 R. E. A. Kelly and L. N. Kantorovich, *J. Phys. Chem. B*, 2006, **110**, 2249–2255.
15 T. Stolar, S. Lukin, J. Požar, M. Rubčić, G. M. Day, I. Biljan, D. v. Jung, G. Horvat, K. Užarević, E. Meštrović and I. Halasz, *Cryst. Growth Des.*, 2016, **16**, 3262–3270.
16 D. G. Hud, N. V.; Jain, S. S.; Li, X.; Lynn, *Chem. Biodivers.*, 2007, **4**, 768–783.
17 J. Donohue and K. N. Trueblood, *Journal of Molecular Biology*, 1960, **2**, 363 – 371.
18 L. E. Orgel, *Crit. Rev. Biochem. Mol. Biol.*, 2004, **39**, 99–123.
19 P. Cieplak and P. A. Kollman, *J. Am. Chem. Soc.*, 1988, **110**, 3734–3739.
20 C. Bolm, R. Mocci, C. Schumacher, M. Turberg, F. Puccetti and J. G. Hernandez, *Angew. Chem., Int. Ed.*, 2018, **57**, 2423–2426.
21 S. Lamour, S. Pallmann, M. Haas and O. Trapp, *Life*, 2019, **9**, 52.
22 S. Lherminier, R. Planet, V. L. Vehel, G. Simon, L. Vanel, K. J. Måløy and O. Ramos, *Phys. Rev. Lett.*, 2019, **122**, 218501.
23 J. G. Forsythe, S.-S. Yu, I. Mamajanov, M. A. Grover, R. Krishnamurthy, F. M. Fernandez and N. V. Hud, *Angew. Chem. Int. Ed.*, 2015, **54**, 9871–9875.
24 S. Becker, C. Schneider, M. Dejmek, T. Carell, H. Okamura and A. Crisp, *Nat. Commun.*, 2018, **9**, 163.
25 N. V. Hud, *Nat. Commun.*, 2018, **9**, 5171.
26 I. R. Gould and P. A. Kollman, *J. Am. Chem. Soc.*, 1994, **116**, 2493–2499.
27 P. Jurec and P. Hobza, *J. Am. Chem. Soc.*, 2003, **125**, 15608–15613.
28 N. B. Leontis, J. Stombaugh and E. Westhof, *Nucleic Acids Res.*, 2002, **30**, 3497–3531.
29 E. N. Nikolova, H. Zhou, F. L. Gottardo, H. S. Alvey, I. J. Kimsey and H. M. Al-Hashimi, *Biopolymers*, 2013, **99**, 955–968.
30 H. Zhou, B. J. Hintze, I. J. Kimsey, B. Sathyamoorthy, S. Yang, J. S. Richardson and H. M. Al-Hashimi, *Nucleic Acids Res.*, 2015, **43**, 3420–3433.
31 E. N. Nikolova, E. Kim, A. A. Wise, P. J. O. Brien, I. Andricioaei and H. M. Al-hashimi, *Nature*, 2011, **470**, 498–502.
32 M. P. Callahan, K. E. Smith, H. J. Cleaves, J. Ruzicka, J. C. Stern, D. P. Glavin, C. H. House and J. P. Dworkin, *Proc. Nat. Acad. Sci.*, 2011, **108**, 13995–13998.
33 B. J. Cafferty and N. V. Hud, *Isr. J. Chem.*, 2015, **55**, 891–905.
34 A. C. Rios and Y. Tor, *Isr. J. Chem.*, 2013, **53**, 469–483.