



Helix instability and self-pairing prevent unnatural base pairs from expanding the genetic alphabet

Thomas P. Hettinger^{a,1}

Several papers have described the use of hydrophobic unnatural base pairs (UBPs) to study mechanisms of DNA replication and transcription (1–7). In PNAS, the study by Zhang et al. (8) claims that UBPs can form stable complementary structures in the absence of hydrogen bonds between base pairs, and that UBPs can be replicated in a bacterial plasmid in virtually any sequence context without loss of the UBPs over many DNA doublings. However, Hirao and Kimoto (4) described two serious flaws with UBPs: (i) hydrophobic UBPs destabilize DNA so that only single UBPs sandwiched between long stretches of natural base pairs will form typical helical double-stranded DNA, and (ii) hydrophobic UBPs are prone to self-pairing during replication that results in inversion of the configuration of the UBPs in the double helix. When an unnatural base pair dX^*dY undergoes self-pairing during DNA replication, dX^*dX or dY^*dY will be formed. Subsequent cognate replication will generate dY^*dX , where the configuration of the UBP is inverted from its original structure. The UBP inversion is a covert mutation invisible to the sequencing methods that rely only on identification of natural bases. The inversion appears to be counted as a faithful replication (8), even though it is a common type of mutation peculiar to hydrophobic UBPs (4). Similarly, the biotin-shift assay (8) recognizes the presence of the UBP in replicated plasmids, but does not distinguish whether the UBP is in its original or inverted configuration. The problem of self-pairing seems to have been ignored because it was assumed that primer extension stops after formation

of the self-pair (1, 4), but this has not been shown to be general or absolute. Even a small amount of self-pairing and extension would eventually equilibrate the UBP orientations. Furthermore, none of the studies used consecutive sequences of UBPs, so the results may not apply to all sequence contexts.

The studies of Hirao's group (2–4) and Romesberg's group (1, 5–8) show that single hydrophobic UBPs can be replicated as long as they are embedded in a sequence of natural base pairs. Hydrophobic UBP replication is a default condition, where a variety of base pairs, including self-pairs, can be incorporated into DNA because natural bases are excluded (9). However, consecutive sequences cannot be replicated, suggesting that single hydrophobic UBPs are stabilized by the presence of long stretches of natural base pairs and are in effect parasitic to the natural base pairs instead of promoting DNA stability. The artificial base pair dP^*dZ of Benner and colleagues (10) does not suffer from this problem. The base pair dP^*dZ has three hydrogen bonds and consecutive sequences of as many as four dP^*dZ pairs can be replicated. Although dP^*dZ causes changes in DNA geometry, the effects are small compared with those of hydrophobic base pairs.

The Zhang et al. (8) study challenges the classic concept of a four-letter alphabet of a DNA double helix with hydrogen-bonded base pairs. However, hydrophobic unnatural bases destabilize DNA and are prone to self-pairing, making them poor candidates for expanding the genetic alphabet.

- 1 Leconte AM, et al. (2008) Discovery, characterization, and optimization of an unnatural base pair for expansion of the genetic alphabet. *J Am Chem Soc* 130:2336–2343.
- 2 Kimoto M, Kawai R, Mitsui T, Yokoyama S, Hirao I (2009) An unnatural base pair system for efficient PCR amplification and functionalization of DNA molecules. *Nucleic Acids Res* 37:e14.
- 3 Yamashige R, et al. (2012) Highly specific unnatural base pair systems as a third base pair for PCR amplification. *Nucleic Acids Res* 40:2793–2806.
- 4 Hirao I, Kimoto M (2012) Unnatural base pair systems toward the expansion of the genetic alphabet in the central dogma. *Proc Jpn Acad, Ser B, Phys Biol Sci* 88:345–367.
- 5 Malyshev DA, et al. (2012) Efficient and sequence-independent replication of DNA containing a third base pair establishes a functional six-letter genetic alphabet. *Proc Natl Acad Sci USA* 109:12005–12010.
- 6 Malyshev DA, et al. (2014) A semi-synthetic organism with an expanded genetic alphabet. *Nature* 509:385–388.
- 7 Dhami K, et al. (2014) Systematic exploration of a class of hydrophobic unnatural base pairs yields multiple new candidates for the expansion of the genetic alphabet. *Nucleic Acids Res* 42:10235–10244.

^aDepartment of Oral Health and Diagnostic Sciences, UConn Health, Farmington, CT 06030

Author contributions: T.P.H. wrote the paper.

The author declares no conflict of interest.

¹Email: thomas.p.hettinger@gmail.com.

- 8 Zhang Y, et al. (2017) A semisynthetic organism engineered for the stable expansion of the genetic alphabet. *Proc Natl Acad Sci USA* 114:1317–1322.
- 9 Khakshoor O, Wheeler SE, Houk KN, Kool ET (2012) Measurement and theory of hydrogen bonding contribution to isosteric DNA base pairs. *J Am Chem Soc* 134:3154–3163.
- 10 Georgiadis MM, et al. (2015) Structural basis for a six nucleotide genetic alphabet. *J Am Chem Soc* 137:6947–6955.