

*Electronic Supplementary Information for*

**Beyond *DAD*: Proposing a one-letter code for nucleobase-mediated molecular recognition**

Aiden J. Ward and Benjamin E. Partridge\*

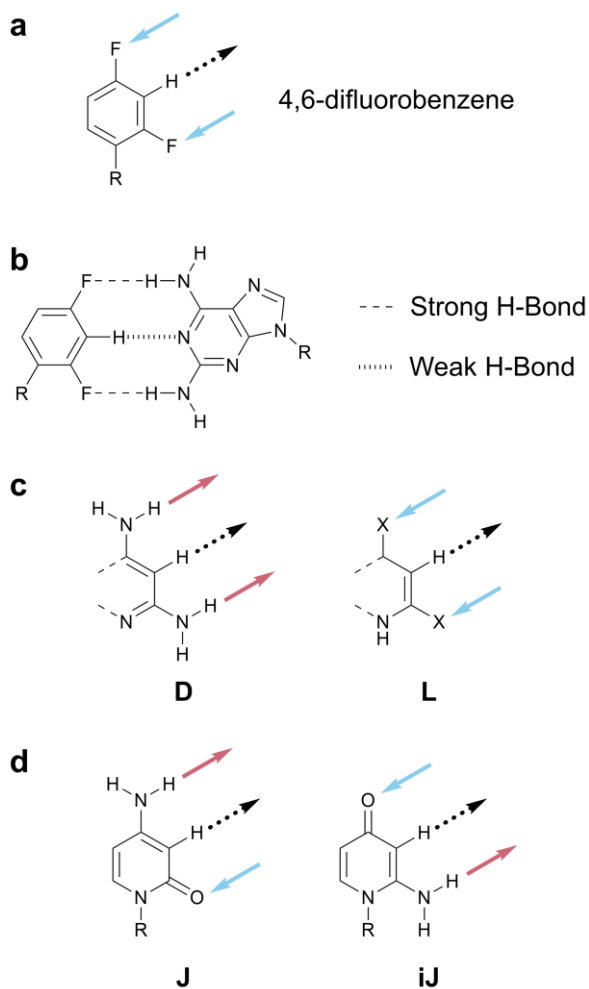
*Department of Chemistry, University of Rochester, Rochester, NY 14627-0216, United States.*

\*Email: [benjamin.partridge@rochester.edu](mailto:benjamin.partridge@rochester.edu)

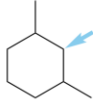
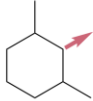
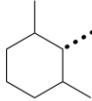
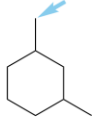












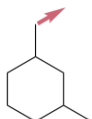












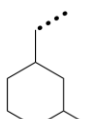












**Table of Contents**

Scheme S1 .....	S2
Scheme S2 .....	S3
Scheme S3 .....	S4
References for the Electronic Supplementary Information .....	S5

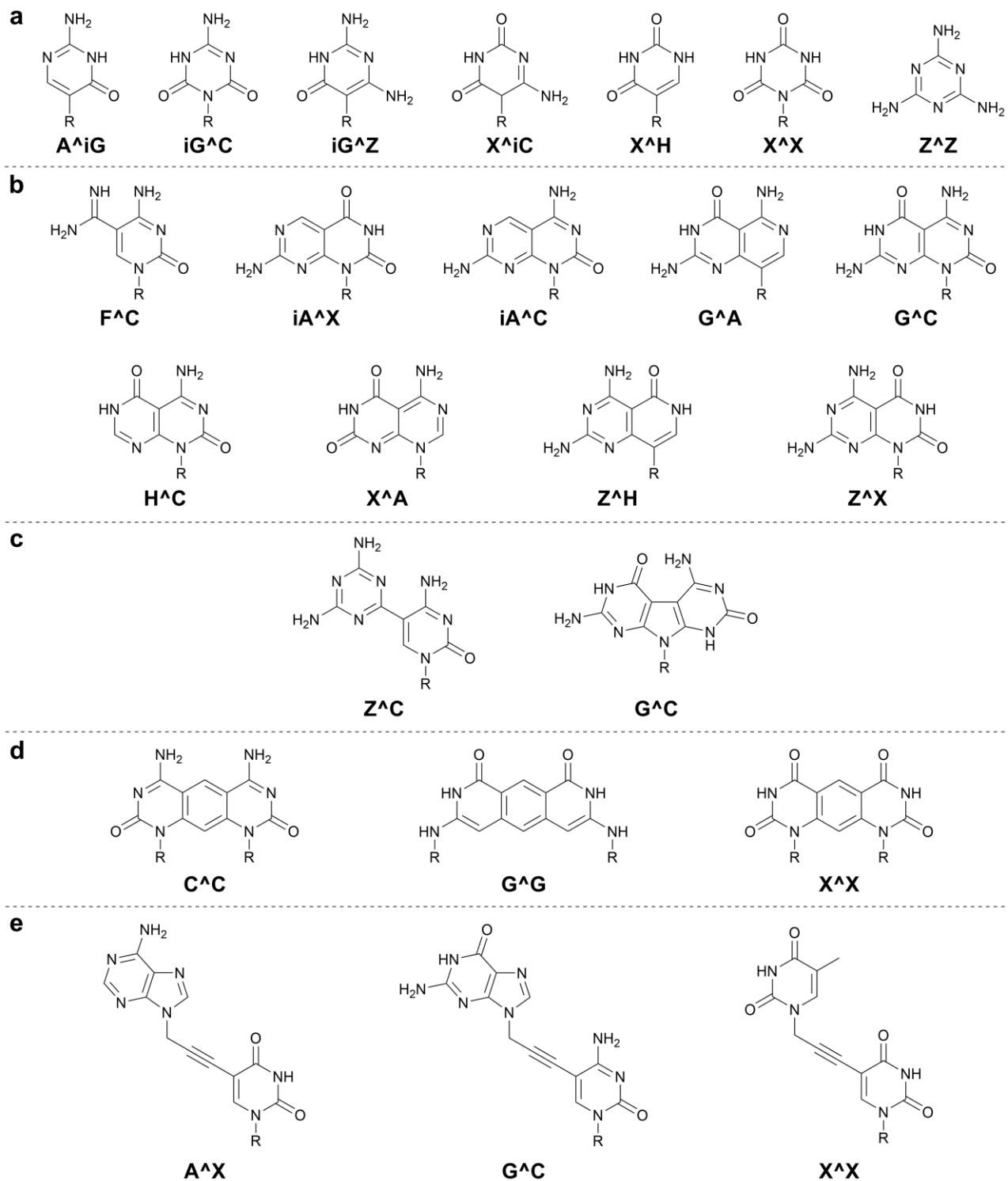
**Scheme S1. HB patterns with a null site at the 2-position.** (a) Difluorobenzyl derivatives have been employed as nucleobase motifs that rely on two fluorine atoms as HB acceptors. (b) Rather than being considered as a null site, a C–H bond may act as a weak HB donor when paired opposite a HB-accepting heteroatom. (c) Although *DND* and *ANA* sites are practically difficult to realize, they are assigned the one-letter codes **D** (for *diamino*) and **L** (for *halogen*), for the sake of completeness. (d) HB patterns that present *DNA* and *AND* are impractical as the C–H in the 2-position will act like a HB donor. Nevertheless, they are named here as **J** and **isoJ** (or **iJ** for short).



**Scheme S2. All 27 combinations of HB acceptors, donors, and null sites.** Although there are 27 combinations of HB acceptors (*A*, blue arrows), HB donors (*D*, maroon arrows), and null sites (*N*, black broken lines), six of these combinations result in only one HB and one combination results in no HBs (*NNN*). Of the remaining 20, four include a null site in the 2-position that is impractical to realize (see Scheme S1). The remaining 16 combinations are the focus of the main text and are summarized in Scheme 1.

1	2			3
				
	 <b>O</b>	 <b>X</b>	 <b>L</b>	
	 <b>iC</b>	 <b>G</b>	 <b>iJ</b>	
	 <b>P</b>	 <b>H</b>	 <b>--</b>	
	 <b>C</b>	 <b>iG</b>	 <b>J</b>	
	 <b>Z</b>	 <b>E</b>	 <b>D</b>	
	 <b>A</b>	 <b>F</b>	 <b>--</b>	
	 <b>iP</b>	 <b>iH</b>	 <b>--</b>	
	 <b>iA</b>	 <b>iF</b>	 <b>--</b>	
	 <b>--</b>	 <b>--</b>	 <b>--</b>	

**Scheme S3. HB patterns of reported bifacial (Janus) nucleobases described using our proposed one-letter codes.** (a) Monocyclic Janus wedges: **A<sup>^</sup>iG**,<sup>1</sup> **iG<sup>^</sup>A<sup>C</sup>**,<sup>2</sup> **iG<sup>^</sup>A<sup>Z</sup>**,<sup>3</sup> **X<sup>^</sup>iC**,<sup>4</sup> **X<sup>^</sup>A<sup>H</sup>**,<sup>5</sup> **X<sup>^</sup>A<sup>X</sup>**,<sup>6</sup> and **Z<sup>^</sup>Z**.<sup>6</sup> (b) Bicyclic Janus nucleobases: **F<sup>^</sup>A<sup>C</sup>**,<sup>5</sup> **iA<sup>^</sup>A<sup>X</sup>**,<sup>7</sup> **iA<sup>^</sup>A<sup>C</sup>**,<sup>8</sup> **G<sup>^</sup>A**,<sup>5</sup> **G<sup>^</sup>A<sup>C</sup>**,<sup>9</sup> **H<sup>^</sup>A<sup>C</sup>**,<sup>9</sup> **X<sup>^</sup>A**,<sup>10</sup> **Z<sup>^</sup>A<sup>H</sup>**,<sup>11</sup> and **Z<sup>^</sup>A<sup>X</sup>**.<sup>12</sup> (c) Nonsymmetric bifacial nucleobases not containing two fused heterocycles: **Z<sup>^</sup>A<sup>C</sup>**<sup>8</sup> and **G<sup>^</sup>A<sup>C</sup>**.<sup>13</sup> (d) Extended, tricyclic symmetric bifacial nucleobases: **C<sup>^</sup>A<sup>C</sup>**, **G<sup>^</sup>A<sup>G</sup>**, and **X<sup>^</sup>A<sup>X</sup>**.<sup>14</sup> (e) Double-headed alkynyl nucleosides: **A<sup>^</sup>A<sup>X</sup>**, **G<sup>^</sup>A<sup>C</sup>**, and **X<sup>^</sup>A<sup>X</sup>**.<sup>15</sup>



## References for the Electronic Supplementary Information

1. M. Egholm, L. Christensen, K. L. Deuholm, O. Buchardt, J. Coull and P. E. Nielsen, Efficient pH-independent sequence-specific DNA binding by pseudoisocytosine-containing bis-PNA, *Nucleic Acids Res.*, 1995, **23**, 217–222.
2. C. L. Meena, D. Singh, B. Kizhakeetil, M. Prasad, M. George, S. Tothadi and G. J. Sanjayan, Triazine-Based Janus G–C Nucleobase as a Building Block for Self-Assembly, Peptide Nucleic Acids, and Smart Polymers, *J. Org. Chem.*, 2021, **86**, 3186–3195.
3. D. Chen, Meena, S. K. Sharma and L. W. McLaughlin, Formation and Stability of a Janus-Wedge Type of DNA Triplex, *J. Am. Chem. Soc.*, 2004, **126**, 70–71.
4. H. Chen, Meena and L. W. McLaughlin, A Janus-Wedge DNA Triplex with A-W<sub>1</sub>-T and G-W<sub>2</sub>-C Base Triplets, *J. Am. Chem. Soc.*, 2008, **130**, 13190–13191.
5. S. A. Thadke, J. D. R. Perera, V. M. Hridya, K. Bhatt, A. Y. Shaikh, W.-C. Hsieh, M. Chen, C. Gayathri, R. R. Gil, G. S. Rule, A. Mukherjee, C. A. Thornton and D. H. Ly, Design of Bivalent Nucleic Acid Ligands for Recognition of RNA-Repeated Expansion Associated with Huntington's Disease, *Biochemistry*, 2018, **57**, 2094–2108.
6. J. A. Zerkowski, C. T. Seto and G. M. Whitesides, Solid-state structures of rosette and crinkled tape motifs derived from the cyanuric acid melamine lattice, *J. Am. Chem. Soc.*, 1992, **114**, 5473–5475.
7. D. Shin, R. W. Sinkeldam and Y. Tor, Emissive RNA Alphabet, *J. Am. Chem. Soc.*, 2011, **133**, 14912–14915.
8. G. Artigas and V. Marchán, Synthesis of Janus Compounds for the Recognition of G-U Mismatched Nucleobase Pairs, *J. Org. Chem.*, 2013, **78**, 10666–10677.
9. H.-Z. Yang, M.-Y. Pan, D.-W. Jiang and Y. He, Synthesis of Janus type nucleoside analogues and their preliminary bioactivity, *Org. Biomol. Chem.*, 2011, **9**, 1516.
10. H. Zhao, W. Huang, X. Wu, Z. Xing, Y. He and Q. Chen, Different superstructures formed by Janus-type nucleosides, *Chem. Commun.*, 2012, **48**, 6097.
11. S. A. Thadke, V. M. Hridya, J. D. R. Perera, R. R. Gil, A. Mukherjee and D. H. Ly, Shape selective bifacial recognition of double helical DNA, *Commun. Chem.*, 2018, **1**, 79.
12. A. Asadi, B. O. Patrick and D. M. Perrin, Janus-AT Bases: Synthesis, Self-Assembly, and Solid State Structures, *J. Org. Chem.*, 2007, **72**, 466–475.
13. A. Asadi, B. O. Patrick and D. M. Perrin, G<sup>+</sup>C Quartet — A DNA-Inspired Janus-GC Heterocycle: Synthesis, Structural Analysis, and Self-Organization, *J. Am. Chem. Soc.*, 2008, **130**, 12860–12861.
14. A. Marsh, E. G. Nolen, K. M. Gardinier and J.-M. Lehn, Janus molecules: Synthesis of double-headed heterocycles containing two identical hydrogen bonding arrays, *Tetrahedron Lett.*, 1994, **35**, 397–400.
15. P. Kumar, A. F. Sorinas, L. J. Nielsen, M. Slot, K. Skytte, A. S. Nielsen, M. D. Jensen, P. K. Sharma, B. Vester, M. Petersen and P. Nielsen, Double-Coding Nucleic Acids: Introduction of a Nucleobase Sequence in the Major Groove of the DNA Duplex Using Double-Headed Nucleotides, *J. Org. Chem.*, 2014, **79**, 8020–8030.