4. Chang, F., Lee, J.T., Navolanic, P.M., Steelman, L.S., Shelton, J.G., Blalock, W.L., Franklin, R.A., and McCubrey, J.A.. (2003). Leukemia *17*, 590–603.
5. Eng, C. (2003). Hum. Mutat. *22*, 183–198.
6. Wu, H., Goel, V., and Haluska, F.G. (2003). Oncogene *22*, 3113–3122.
7. Stambolic, V., Suzuki, A., de la Pompa, J.L., Brothers, G.M., Mirtsos, C., Sasaki, T., Ruland, J., Penninger, J.M., Siderovski, D.P., Mak, T.W. (1998). Cell *95*, 29–39.
8. Di Cristofano, A., Pesce, B., Cordon-Cardo, C., and Pandolfi, P.P. (1998). Nat. Genet. *19*, 348–355.
9. Nakamura, N., Ramaswamy, S., Vazquez, F., Signoretti, S., Loda, M., and Sellers, W.R. (2000). Mol. Cell. Biol. *20*, 8969–8982.
10. Paez, J., and Sellers, W.R. (2003). Cancer Treat. Res. *115*, 145–167.
11. Brunet, A., Bonni, A., Zigmond, M.J., Lin, M.Z., Juo, P., Hu, L.S., Anderson, M.J., Arden, K.C., Blenis, J., and Greenberg, M.E. (1999). Cell *96*, 857–868.
12. Medema, R.H., Kops, G.J., Bos, J.L., and Burgering, B.M. (2000). Nature *404*, 782–787.
13. Ramaswamy, S., Nakamura, N., Sansal, I., Bergeron, L., and Sellers, W.R. (2002). Cancer Cell *2*, 81–91.
14. Xu, X., Sakon, M., Nagano, H., Hiraoka, N., Yamamoto, H., Hayashi, N., Dono, K., Nakamori, S., Umeshita, K., Ito, Y., Matsuura, N., and Monden, M. (2004). Oncol. Rep. *11*, 25–32.
15. Kau, T.R., Schroeder, F., Ramaswamy, S., Wojciechowski, C.L., Zhao, J.J., Roberts, T.M., Clardy, J., Sellers, W.R., and Silver, P.A. (2003). Cancer Cell *4*, 463–476.
16. Fornerod, M., Ohno, M., Yoshida, M., and Mattaj, I.W. (1997). Cell *90*, 1051–1060.
17. Mayo, L.D., and Donner, D.B. (2002). Trends Biochem. Sci. *27*, 462–467.
18. Wang, W., Rastinejad, F., and El-Deiry, W.S. (2003). Cancer Biol. Ther. *2*, S55–S63.
19. Wang, W., Takimoto, R., Rastinejad, F., and El-Deiry, W.S. (2003). Mol. Cell. Biol. *23*, 2171–2181.
20. El-Deiry, W.S. (2001). Nat. Cell Biol. *3*, E71–E73.
21. Sager, J.A., and Lengauer, C. (2003). Cancer Biol. Ther. *2*, 452–455.

Chemistry & Biology, Vol. 11, January, 2004, ©2004 Elsevier Science Ltd. All rights reserved. DOI 10.1016/j.chembiol.2004.01.007

Orthogonal Base Pairs Continue to Evolve

Recent developments in the design and construction of unusual analogs of the natural nucleic acid bases have reached a milestone with the report (in this issue of *Chemistry & Biology* [1]) of a new orthogonal base pair that allows site-specific introduction of a photo-crosslinkable modified base into an RNA molecule by T7 RNA polymerase-mediated transcription of DNA containing the base-pairing partner.

One of the more interesting challenges in modified base design has been the creation of new orthogonal base pairs that would function as substrates for enzymes involved in the synthesis and processing of nucleic acids and would thereby provide a means for the expansion of the genetic code [2]. The earliest concerted effort to expand the genetic code and create new base pairs came from the laboratory of Benner and coworkers. In studies initiated during the 1980s, a series of putative base pairs containing alternative orthogonal hydrogen bonding patterns were designed and synthesized [3, 4]. Although the initial results with the isoC-isoG base pair looked promising, subsequent work showed that isoC was unstable in aqueous solution [5]. Furthermore, different polymerases show a broad range of specificity and recognition characteristics with these bases as template and as triphosphate substrates. In subsequent years the Benner laboratory continued to explore new variations on alternative H-bonding pattern modified bases. These studies included C-linked pyrimidine analogs [4, 6], which although optimal from the standpoint of alternative H-bonding patterns, proved to be unstable toward epimerization at C1′.

In a recent paper, Hutter and Benner now appear to have solved this problem [7]. 6-Amino-2-oxo-(1H) pyridine linked C3 to C1′ of 2′-deoxyribose is prone to epimerization through a mechanism initiated by protonation of the furanose oxygen followed by furanose ring opening coordinated with loss of the proton on N1. Reclosure leads to both the α and β epimers about C1′. Hutter and Benner rationalized that the introduction of a strong electron withdrawing group at C5 would prevent the electron release required for ring opening. Indeed they have now demonstrated that the iso-C analog, 6-amino-3-(2′-deoxy-β-D-ribofuranosyl)-5-nitro-(1H)-pyridine-2-one is stable to epimerization under conditions that one typically encounters during synthesis and manipulation of nucleoside triphosphates and oligonucleotides. One can expect that this work will soon be extended to polymerase-mediated nucleic acid replication.

As an alternative to orthogonal hydrogen bonding nucleoside analogs, Kool and coworkers have been exploring nonhydrogen-bonding isosteres of the natural bases [8]. A critical finding from this laboratory was that the hydrophobic shape-complementary purine-pyrimidine analogs preferentially pair with one another [9, 10]. Others variations on this theme have uncovered a number of interesting new base pairs. Most noteworthy has been the effort by Romesberg, Schultz, and coworkers, who have focused on the theme of creating unique hydrophobic base analogs with significant preference for self-pairing in nucleic acid duplexes [11].

Recently Hirao and coworkers reported a strategy for the construction of orthogonal base pairs that depend more on molding complementary shapes rather than hydrogen bonding pattern [12, 13]. They have found pairing specificity between a purine bearing a nonhydrogen-bonding substituent on C6 of a purine and a pyridine C-nucleoside that mimics a pyrimidine nucleoside minus the C4 amino or oxo group. This pairing is illustrated in

Figure 1 for a purine containing C2 amino and C6 thienyl groups and the pyridine a C5 iodo group [1]. This pair appears to be able to occupy space without deviating significantly from the natural parameters (shown for a G•C base pair in the figure). Comparison to a natural G•C base pair indicates that the space defined by this pair could have a slightly shorter C1′ to C1′ distance and modestly higher angles for λ. The size and geometry of the thienyl group suggest that there is not much difference between the two possible conformations about the bond between the purine C6 and thiophene C2. Indeed a simple AM1 calculation predicts less than 0.3 kcal difference between the two conformations.

In the design of new base pairs, there are certain structural properties that appear to be universally required if a base pair is to be incorporated and extended by polymerases. These include the presence of a planar aromatic moiety, a shape that matches an A•U(T) or G•C base pair and hydrogen bond acceptor sites on the minor groove side of each base positioned in space equivalent to O2 of C or U(T) and N3 of A or G. Protruding groups extending beyond the dimensions of a natural base pair are not tolerated, with one important exception. Substituents that protrude into the major groove from C5 of the pyrimidine nucleosides appear to be accommodated. This exception is important because it allows one to place modifications into nucleic acids by enzyme-mediated reactions. Kimoto et al. [1] exploit polymerase tolerance for substituents in this region. Both of the modified bases in the **s•Iy** base pair contain modifications that protrude into the major groove. In the case of the thienylpurine analog, **s**, the presence of the thienyl sterically excludes the four natural nucleosides but accommodates an analog, 2-oxo-(1H)pyridine (**Iy**), that lacks a substituent at the site occupied by O4 in U and the external NH_2 in C. There is the prospect that significantly more elaborate substituents could be tolerated.

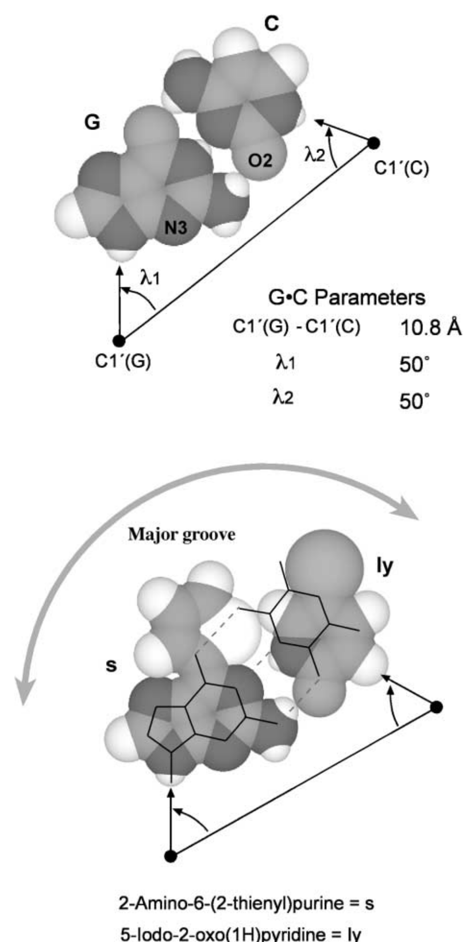

Figure 1. Comparison of the Structures of a G•C Base Pair and the Orthogonal **s•Iy** Base Pair

Eaton and coworkers were the first to show that T7 RNA polymerase could accept a UTP modified at the 5 position with a relatively large substituent [14]. Vaish et al. later demonstrated that other C5 side chains were tolerated as well and introduced amino and thiol functional groups into RNA via T7 RNA polymerase-mediated transcription [15]. The ability to be able to incorporate substituents at this position has important implications for the development of molecular tools in the RNA arena. The demonstration by Kimoto et al. [1] that the 2-oxopyridine base surrogate could also accommodate a substituent at this position bodes well for future applications. For example, one could imagine adding C5 alkynyl spacers terminating in reactive functional groups that provide a pathway for the introduction of other crosslinking reagents, reporter groups, ligands, or other functionalities useful as tools for probing RNA interactions. For example, the introduction of other types of photo-crosslinking reagents would be useful [16].

One of the benefits of the technology reported by Hirao and coworkers is that one can introduce modified bases into virtually any site within a long RNA, since the RNA is obtained by transcription off of a DNA template. The modified base **s** can be incorporated into oligodeoxyribonucleotide primers by means of its phosphoramidite, and the primers extended by conventional means to obtain full-length DNA complementary to the desired RNA. Of course this strategy still requires that one has both a modified phosphoramidite and a complementary nucleoside triphosphate. It will be of interest to see if this technology can be translated into higher throughput synthesis of sets of RNA molecules modified at different sites. Complete mapping of interactions will likely require libraries of RNA molecules containing members each substituted at a different site. The advantage of the Kimoto strategy is that automated synthesis of oligodeoxyribonucleotides is relatively inexpensive, and potentially only a single modified nucleoside phosphoramidite would be required. The economics of synthesizing the complementary modified nucleoside triphosphates could be more of an issue.

Other alternatives for site-specific introduction of modified bases into RNA exist, and it will be interesting to see how these competing strategies are used in the next few years. Earnshaw and Gait reviewed methodology for site-specific introductions of modified nucleosides into RNA 5 yr ago [17]. With recent improvements in RNA synthesis and ligation technologies, it is likely to become more routine to build oligoribonucleotides containing photoaffinity labels at specific positions and

incorporate these into full-length RNA by ligation. RNA ligations are typically done using T4 DNA ligase, as originally reported by Moore and Sharp [18]. Such ligations require a DNA template and yields based on total RNA can appear low because of the presence of N+1 segments when transcribed RNA is used. This approach offers a significant advantage over nontemplated RNA ligation by T4 RNA ligase, which suffers from problems of extraneous ligations. But even the T4 RNA ligase method continues to improve. Kurata et al. [19] recently described a rapid ligation RNA strategy that overcomes many of the problems associated with the use of T4 RNA ligase. The key step is incorporation of a periodate oxidation after ligation of the modified base as pNp in order to convert unreacted 3′-terminal ribosyl groups to dialdehydes, which are incapable of further ligation to a second RNA fragment.

Although Kimoto et al. [1] used the modified base **Iy** to demonstrate crosslinking between RNA molecules, it is likely that this technology will find its greatest application in exploring RNA protein interactions. Methodology for the identification of crosslinking sites has been widely published [20], but in relatively few instances have techniques for precise placement of crosslinkable moieties within an RNA sequence been defined [21]. The biological landscape is rich with a diversity of RNAs and proteins rapidly interacting in ways that spectroscopic methods of analysis may not be suited. Given the complexity, diversity, and dynamics of these interactions, methods that allow recording events (e.g., by laser-mediated photo-crosslinking) are likely to be of considerable interest.

Kimoto et al. [1] have clearly demonstrated that a new orthogonal base pair can be put to valuable use. But it is even clearer that the race to discover new orthogonal base pairs has just begun. The current generation of candidates does not yet have the specificity required for many applications. Hirao and collaborators continue to explore new base pairing partners [22] and most recently have reported a pair based on 4-propynylpyrrole-2-carbaldehyde and 9-methylimidazol[(4.5)-*b*]pyridine, a base introduced by Kool [23]. Romesberg and his collaborators have a significant stable of candidates that show promising properties [24, 25]. Based on the same theme of pairing a five-membered heterocycle with larger polycyclic aromatic heterocycles, Romesberg and coworkers have reported base pairs containing either a thiophene or furan derivative paired with isocarbostyryl derivatives [26]. It is likely that other candidates will soon appear, along with a significant number of novel applications.

Donald E. Bergstrom
Department of Medicinal Chemistry and Molecular Pharmacology
Purdue University
Hansen Life Science Research Building
201 South University Street
West Lafayette, Indiana 47907

Selected Reading

1. Kimoto, M., Endo, M., Mitsui, T., Okuni, T., Hirao, I., and Yokoyama, S. (2004). Chem. Biol. *11*, this issue, 47–55.
2. Wang, L., and Schultz, P.G. (2002). Chem. Comm. 1–11.
3. Switzer, C., Moroney, S.E., and Benner, S.A. (1989). J. Am. Chem. Soc. *111*, 8322–8323.
4. Piccirilli, J.A., Krauch, T., Moroney, S.E., and Benner, S.A. (1990). Nature *343*, 33–37.
5. Switzer, C.Y., Moroney, S.E., and Benner, S.A. (1993). Biochemistry *32*, 10489–10496.
6. Voegel, J.J., and Benner, S.A. (1994). J. Am. Chem. Soc. *116*, 6929–6930.
7. Hutter, D., and Benner, S.A. (2003). J. Org. Chem. *68*, 9839–9842.
8. Kool, E.T. (2002). Acc. Chem. Res. *35*, 936–943.
9. Guckian, K.M., Morales, J.C., and Kool, E.T. (1998). J. Org. Chem. *63*, 9652–9656.
10. Morales, J.C., and Kool, E.T. (1998). Nat. Structure Biol. *5*, 950–954.
11. Ogawa, A.K., Wu, Y., McMinn, D.L., Liu, J., Schultz, P.G., and Romesberg, F.E. (2000). J. Am. Chem. Soc. *122*, 3274–3287.
12. Hirao, I., Ohtsuki, T., Fujiwara, T., Mitsui, T., Yokogawa, T., Okuni, T., Nakayama, H., Takio, K., Yabuki, T., Kigawa, T., et al. (2002). Nat. Biotech. *20*, 177–182.
13. Ohtsuki, T., Kimoto, M., Ishikawa, M., Mitsui, T., Hirao, I., and Yokoyama, S. (2001). Proc. Natl. Acad. Sci. USA *98*, 4922–4925.
14. Dewey, T.M., Mundt, A.A., Crouch, G.J., Zyzniewski, M.C., and Eaton, B.E. (1995). J. Am. Chem. Soc. *117*, 8474–8475.
15. Vaish, N.K., Fraley, A.W., Szotak, J.W., and McLaughlin, L.W. (2000). Nucleic Acids Res. *28*, 3316–3322.
16. Tate, J.J., Persinger, J., and Bartholomew, B. (1998). Nucleic Acids Res. *26*, 1421–1426.
17. Earnshaw, D.J., and Gait, M.J. (1998). Biopolymers *48*, 39–55.
18. Moore, M.J., and Sharp, P.A. (1992). Science *256*, 992–997.
19. Kurata, S., Ohtsuki, T., Suzuki, T., and Watanabe, K. (2003). Nucleic Acids Res. *31*, e145.
20. Urlaub, H., Hartmuth, K., and Lührmann, R. (2002). Methods *26*, 170–181.
21. Harris, M.E., and Christian, E.L. (1999). Methods *18*, 51–59.
22. Mitsui, T., Kimoto, M., Sata, A., Yokoyama, S., and Hirao, I. (2003). Bioorg. Med. Chem. Letts. *13*, 4515–4518.
23. Morales, J.C., and Kool, E.T. (1999). J. Am. Chem. Soc. *121*, 2323–2324.
24. Matsuda, S., Henry, A.A., Schultz, P.G., and Romesberg, F.E. (2003). J. Am. Chem. Soc. *125*, 6134–6139.
25. Henry, A.A., Yu, C., and Romesberg, F.E. (2003). J. Am. Chem. Soc. *125*, 9638–9646.
26. Berger, M., Luzzi, S.D., Henry, A.A., and Romesberg, F.E. (2002). J. Am. Chem. Soc. *124*, 1222–1226.