Published in final edited form as:

Curr Opin Chem Biol. 2007 December ; 11(6): 588–594.

Model Systems for Understanding DNA Base Pairing

Andrew T. Krueger and Eric T. Kool*

Department of Chemistry, Stanford University, Stanford CA 94305-5080

Abstract

The fact that nucleic acid bases recognize each other to form pairs is a canonical part of the dogma of biology. However, they do not recognize each other well enough in water to account for the selectivity and efficiency that is needed in the transmission of biological information through a cell. Thus proteins assist in this recognition in multiple ways, and recent data suggests that these mechanisms of recognition can vary widely with context. To probe how the chemical differences of the four nucleobases are defined in various biological contexts, chemists and biochemists have developed modified versions that differ in their polarity, shape, size, and functional groups. This brief review covers recent advances in this field of research.

Keywords

base pair; stacking; minor groove; hydrogen bond; DNA polymerase; replication

Introduction

The information that is stored and transferred in a cell is carried in the structure and sequence of nucleic acid bases. These heterocycles must be distinguished from one another by the cellular machinery, and this recognition and discrimination is carried out both by nucleobases themselves, and by proteins as well. For successful cellular function, the genetic information as a base sequence must be recognized accurately during many processes, including DNA replication, transcription, DNA repair, recombination, translation, and RNA interference. All of these processes involve the pairing of one base with another in a selective way. Errors in this pairing can lead to medically serious outcomes such as cancer and drug resistance; but at the same time, such errors also make the process of evolution, and thus life on earth, possible.

A central and basic question is what chemically defines the nucleobases as different from one another, and how does the answer vary in the different biological processes mentioned above. The four bases adenine, guanine, thymine and cytosine vary from one another in their size, shape, polarity, local electrostatics, and functional groups (Fig. 1). In principle, all of these differences can be used in the identification of a base, and yet which of these are most important is often not clear. We know from many solution studies that the bases have some ability to recognize each other in nonpolar solvents by complementary hydrogen bonding, and of course we also know that short molecules of DNA and RNA can self-assemble into double helices in water. However, studies have also shown that simple base-base recognition is not strong or selective enough on its own to account for the selectivities observed in many biological

*to whom correspondence should be addressed: kool@stanford.edu.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

processes, which suggests that proteins act to enhance the specificity and efficiency of nucleobase recognition.

Understanding the mechanisms of accurate genetic pairing matters in both basic and practical ways. Since this transfer of information is central to many diseases, a full understanding of its origin and propagation in living systems will require a better grasp of this recognition processes. In addition, design of therapeutic agents aimed at disrupting or enhancing this recognition also can benefit from such knowledge. Finally, the development of improved tools for postgenomic analysis can result from better understanding of the recognition of genetic sequences as well.

This brief review describes recent work in which nonnatural nucleobases are used as tools to study base recognition and pairing mechanisms. It covers only very recent work, mainly carried out in the last 2–3 years, since multiple reviews have addressed related topics recently. In addition, because of the large increase in work in this field recently, it focuses mainly on studies aimed at basic science, and only briefly notes some designed systems, which have been reviewed recently in this journal.

Testing electrostatic effects in pairing

Electrostatic effects can be important in DNA since the structure is highly polar. This is of course tempered by the fact that the solvent is water, which is not only highly polar itself (providing a high dielectric that suppresses electrostatic attractions) but also competes directly for electrostatic interactions such as with hydrogen bonding groups. One of the main electrostatic interactions to be studied recently is base-base hydrogen bonding, and the question of the role of Watson-Crick hydrogen bonds in pairing has received a good deal of attention. Continuing studies of nonpolar nucleoside isosteres have shown that they pair poorly and nonselectively with natural bases in DNA alone, suggesting that Watson-Crick hydrogen bonds play a substantial role in stabilizing the helix [1]. However, the results with DNA polymerases can be markedly different. Nonpolar shape analogs of thymine were shown to be replicated with high efficiency and fidelity, and led to the conclusion that for the Klenow enzyme and other replicative polymerases such as T7 DNA polymerase, hydrogen bonds play only a small positive role in the efficiency and selectivity of pairing, while steric effects play a substantially larger role [2,3]. Several laboratories have demonstrated efficient and selective polymerase synthesis of base pairs that completely lack Watson-Crick hydrogen bonds [2–8]. Most of these pairs lack hydrogen bonding groups in both partners, thus avoiding the cost of desolvating a polar base inside the helix. The Hirao laboratory [4,5] and the Romesberg laboratory [6–8] have recently developed some especially effective nonpolar pairs, and Engels has studied fluoro-aromatic base analogues for their effect on pairing in RNA [9].

Recent studies from the Benner laboratory have demonstrated new examples of hydrogen-bonded pairs in which the H-bonding arrangement is different from the canonical A-T, G-C pairs [10,11]. The studies showed that some non-Watson-Crick hydrogen bonding arrangements can lead to mismatching interference as a result of tautomerism. By making steric adjustments to avoid mismatches and by careful design, pairs were identified can function well in polymerase amplification. This may lead to “synthetic biological systems” [12] that can evolve *in vitro*.

Importantly, recent studies have shown that the mechanism by which polymerases choose partners may vary greatly with different classes of enzymes. For example, Kuchta and Engels found that nonpolar DNA bases were processed poorly by human DNA primase, suggesting the importance of Watson-Crick hydrogen bonds [13]. Several studies with Y-family polymerases (including pols kappa, eta, Dpo4, and Dbh), which have the role as repair enzymes and operate with low fidelity, have also shown that non-H-bonding nucleobases are processed very poorly, which has led to the suggestion that this class of enzyme may rely on Watson-

Crick hydrogen bonds as a central mechanism for pairing efficiency and fidelity [14–16]. It seems somewhat ironic that low-fidelity enzymes may require these bonds, while many high-fidelity polymerases may benefit considerably less from such electrostatic interactions [2,3].

Base-base hydrogen bonding effects were also studied recently in the mechanism of RNA interference, by use of nonpolar uridine analogs as replacements in guide strand RNAs. Both Manoharan and Kool simultaneously carried out similar studies [17,18], and both found that the dependence on Watson-Crick hydrogen bonding was position-dependent in the siRNA. Interestingly, one central position (position 7) retained nearly full bioactivity in gene suppression (in both studies) without hydrogen bonds, whereas positions nearby showed strong dependence on these polar interactions, suggesting that interactions with protein(s) near this position are important.

Probes for minor groove effects

Studies in several laboratories have shown that polar interactions in the minor groove of DNA are exceedingly important in pairing stability and structure. All natural DNA bases have minor groove hydrogen bond acceptors in similar positions, and these are typically well solvated. McLaughlin has developed a number of base analogs lacking minor groove hydrogen bond acceptors, and has shown a strong destabilizing effect on DNA [19]. Minor groove carbonyl groups of thymine have been implicated also in the curvature of DNA containing several consecutive A-T pairs; McLaughlin has localized electrostatic interactions related to curvature to a single thymine out of several, implicating interactions with cations in solution [20].

Minor groove interactions are also important in many polymerases, where hydrogen bonds between the protein and DNA have been suggested to be even more important than Watson-Crick hydrogen bonds between the bases [21]. Spratt has synthesized 3-deaza-variants of guanine and adenine, and has studied their effects on polymerases, particularly in bypassing pairs already made [22]. Related studies with nonpolar nucleoside analogs from several laboratories have shown that having minor groove hydrogen bond acceptors can be crucial to successful polymerase replication of pairs; Romesberg has made this point with multiple designed bases [23–25]. Hirao's most successful pairs also take advantage of this effect [26]. The effect arises presumably because moving a nonpolar base placement into position opposite a polar amino acid sidechain in the minor groove would invoke a costly desolvation penalty. The specific positions and influences of these sidechain interactions appear to vary from enzyme to enzyme.

Testing shape and size effects

The canonical pairs adopt very similar geometries, occupying space that overlaps to a large extent (Fig. 2). Since both correct and incorrectly matched base pairs in DNA are hydrogen-bonded, it is clear that the nonstandard geometry of mismatches plays a substantial role in pairing selectivity. An important question is how much do steric effects alone (in the absence of hydrogen bonds) affect pairing selectivity in the double helix? And secondly, it is clear that pairing selectivity in DNA alone is not nearly high enough (especially at the ends of duplexes) to explain the remarkably high pairing selectivity of polymerase enzymes. As a result, how do polymerase enzymes magnify the importance of these steric effects? Fortunately, there now exists a wide variety of DNA base analogs and designed DNA bases of greatly varied shapes and sizes, and these have been quite useful recently in evaluating steric effects in pairing.

We initiated a systematic study of steric effects in pairing by development of a series of base analogs in which size and shape were gradually changed. In this light, Kim reported a series of nonpolar thymidine analogs as variants of the earlier difluorotoluene isostere; this set has H, F, Cl, Br, and I at the 2,4 positions, analogous to thymine carbonyls. In DNA alone these

compounds had very little difference among them for pairing; all of them were destabilizing against natural bases and displayed very small selectivity for adenine over other partners [1]. However, as polymerase substrates the series showed widely varying effects, with a rapid rise in activity as size increased from the smallest, followed by a rapid decrease beyond an optimum [2,3]. Selectivity followed the same trend. The series varied by several orders in magnitude across a subtle 1 Å change in size, which established a strong response to steric effects. This series was also studied in a bacterial replication assay, showing similar results (Fig. 3), and establishing that such steric effects are biologically relevant [2]. In a similar way, shapes of thymidine analogs were varied by Sintim as well, and again large kinetic effects were observed in replication *in vitro* [27].

In addition to the above work, there have been more polymerase studies with nonpolar and polar DNA base analogs having varied shapes, and the studies have reached sometimes similar, and in some cases divergent, conclusions about the influence of steric effects. For example, Berdis, using various indole derivatives, found roles for size and shape in translesion synthesis past an abasic site [28], but found that pi-surface area and stacking were greater factors in other studies [29,30]. Engels and Kuchta found that base shape was “essentially irrelevant” to nucleotide selection by polymerase alpha and the Klenow enzyme [31,32]; this conclusion with the latter enzyme is in contrast to the subsequent results of Kim *et al.* and Sintim and Kool [2,27]. Bergstrom and Davisson studied some small, polar oxazole base analogues, and found both electrostatic and steric effects to be relevant [33]. Some of the seeming conflicts can be explained by the fact that some base analogs (such as indoles, benzimidazoles and oxazoles) can change conformation, leading to ambiguity in their steric effects. In some cases, different enzymes were studied, which may also explain varied conclusions.

Importantly, different enzymes clearly behave differently in response to steric changes. The Kim series of compounds was recently studied with the low-fidelity repair polymerase Dpo4, where it was found that steric effects were quite small [15]. This suggested that this low-fidelity enzyme is quite flexible in adjusting to different steric changes; this may make sense in the biological role of such repair enzymes, which is to replicate past damaged or mismatched pairs of noncanonical shape and size.

Probes for recognition of damaged bases

DNA analogues have also been recently implemented as probes of damage in DNA. Sturla has demonstrated selective and stable base pairing between the biologically relevant damage adduct *O*⁶-benzyldeoxyguanosine and a designed diaminonaphthyl-derived nucleoside [34]. In addition, two laboratories have constructed analogs of oxidized purines as probes for how such damage is recognized by repair and polymerase enzymes. Hamm used 8-haloguanines to measure the steric effect of the 8-substituent on base pairing stability [35]. Taniguchi and Kool reported the base pairing and polymerase substrate capabilities of nonpolar isosteres for 8-oxo-G and 8-oxoA, and found that they recapitulated the mutagenic properties of the polar naturally damaged purines [36]. This suggested the importance of the shapes of the *syn*-oriented oxopurines in their mutagenic biological activity.

Structural studies

Structural studies of nucleobase analogs in DNA or RNA yield the benefit of helping to understand how the helix adjusts to changes in structure or interactions. Manoharan, Frank-Kamenetskii and Egli (as part of their siRNA studies) published the x-ray crystal structure of an RNA-RNA duplex containing a non-H-bonded pair between adenine and difluorotoluene. The majority of the RNA and local base pair structure was unaffected by this pair, although small local adjustments to pair geometry were noted [17]. McLaughlin and Williams published a crystal structure of a duplex missing a specific minor groove hydrogen bond acceptor, and

analyzed its effect on hydration [37]. Romesberg recently published an impressive array of structural studies of more than one modified base pair in DNA, using both NMR and x-ray crystallography [38]. The studies were useful in identifying which geometry was preferred in a base that can freely rotate around a glycosidic bond, and giving some insight into the efficient replication of this pair. Loakes published the NMR-derived structure of a DNA duplex containing a larger nonpolar “universal” base (nitroindole), which revealed that if a base is too large to fit opposite a natural partner, it can simply intercalate next to it instead [39]. Finally, Lynch et al. published the NMR-derived structure of DNA in which all pairs were modified by benzo-extension (“xDNA”) [40]. The duplex showed eight different modified pairs, all of which were 2.4 Å larger than natural pairs. Interestingly, this large base pair expansion only mildly affected the DNA backbone conformation, which resembled B-DNA; presumably this is because all the pairs adopted a similar geometry, thus allowing for a regular helical conformation.

Alternative principles for base pairing

Some of the insights gained from studies of pairing have led to the development and testing of other, non-natural principles for design of new base pairs. For example, Sekine recently tested the interesting idea of using “halogen bonds” to help pairs associate [41]. Saito examined the use of Schiff base linkage to connect opposing bases [42]. Purine-purine pairs were examined by Battersby [43]. A number of laboratories have recently described metal-bridged base pairs; in this light, Carell recently demonstrated duplexes impressively containing two different metals in several contiguous pairs [44]. The notion of hydrophobic pairs led recently to the testing of super-hydrophobic “fluorous” pairs [45,46], which paired selectively and some of which are active as polymerase substrates. Finally, principles of hydrogen bonding in Watson-Crick pairing led recently to Matsuda’s development of elegant base pairs containing four hydrogen bonds rather than Nature’s three or two [47,48].

Conclusions

There are multiple points in the transmission of cellular genetic information where accurate base pairing is crucial, and the growing complement of DNA base analogs developed by researchers (Fig. 4) will be essential in evaluating how this recognition occurs. It is no doubt the case that both steric and electrostatic effects have important influences on base pairing, but the relative importance of these factors will vary depending on the context, whether it be DNA or RNA. In addition, it is becoming increasingly clear that proteins often assist in base-base recognition by modulating or amplifying their affinity or specificity; once again, the evidence is growing that different proteins use different mechanisms to achieve this. Overall, the wide span of cellular functions that transmit genetic information means that there are a large variety of problems at the chemistry-biology interface for researchers to pursue, and an unlimited set of new designs that can be developed.

Acknowledgements

This work was supported by the U.S. National Institutes of Health (GM072705 and GM63587).

References and Recommended Reading

- = of special interest
 - = of outstanding interest
1. Kim TW, Kool ET. A series of nonpolar thymine analogues of increasing size: DNA base pairing and stacking properties. *J Org Chem* 2005;70:2048–2053. [PubMed: 15760186]

- 2. Kim TW, Delaney JC, Essigmann JM, Kool ET. Probing the active site tightness of DNA polymerase in sub-angstrom increments. *Proc Natl Acad Sci USA* 2005;102:15803–15808. [PubMed: 16249340]A systematic study of steric effects on replication, carried out both in vitro and in living bacteria.
3. Kim TW, Briebe LG, Ellenberger T, Kool ET. Functional evidence for a small and rigid active site in a high fidelity DNA polymerase: probing T7 DNA polymerase with variably sized base pairs. *J Biol Chem* 2006;281:2289–2295. [PubMed: 16311403]
4. Hirao I, Kimoto M, Mitsui T, Fujiwara T, Kawai R, Sato A, Harada Y, Yokoyama S. An unnatural hydrophobic base pair system: site-specific incorporation of nucleotide analogs into DNA and RNA. *Nat Methods* 2006;9:729–735. [PubMed: 16929319]
- 5. Mitsui T, Kimoto M, Harada Y, Yokoyama S, Hirao I. An efficient unnatural base pair for a base-pair-expanded transcription system. *J Am Chem Soc* 2005;127:8652–8658. [PubMed: 15954770] A particularly effective example of a nonnatural, non-hydrogen-bonded base pair that is a substrate for polymerases. It is has been shown to have useful applications in labeling RNAs.
6. Leconte AM, Matsuda S, Romesberg FE. An efficiently extended class of unnatural base pairs. *J Am Chem Soc* 2006;128:6780–6781. [PubMed: 16719445]
7. Matsuda S, Henry AA, Romesberg FE. Optimization of unnatural base pair packing for polymerase recognition. *J Am Chem Soc* 2006;128:6369–6375. [PubMed: 16683801]
8. Hwang GT, Romesberg FE. Substituent effects on the pairing and polymerase recognition of simple unnatural base pairs. *Nucleic Acids Res* 2006;34:2037–2045. [PubMed: 16617144]
9. Zivkovic A, Engels JW. RNA recognition by fluor-aromatic substituted Nucleosides. *Nucleotides Nucleic Acids* 2005;24:1023–1027.
- 10. Yang Z, Sismour AM, Sheng P, Puskar NL, Benner SA. Enzymatic incorporation of a third nucleobase pair. *Nucleic Acids Res* 2007;35:4238–4249. [PubMed: 17576683]This study describes one of the best existing hydrogen-bonded alternative pairs, and demonstrates its replication alongside the natural pairs.
11. Sismour AM, Benner SA. The use of thymidine analogs to improve the replication of an extra DNA base pair: a synthetic biological system. *Nucleic Acids Res* 2005;33:5640–5646. [PubMed: 16192575]
12. Benner SA, Sismour AM. Synthetic biology. *Nat Rev Genet* 2005;7:533–543. [PubMed: 15995697]
13. Moore CL, Zivkovic A, Engels JW, Kuchta RD. Human DNA primase uses Watson-Crick hydrogen bonds to distinguish between correct and incorrect nucleoside triphosphates. *Biochemistry* 2004;43:12367–12374. [PubMed: 15379576]
14. Wolffe WT, Washington MT, Kool ET, Spratt TE, Helquist SA, Prakash L, Prakash S. Evidence for a Watson-Crick hydrogen bonding requirement in DNA synthesis by human DNA polymerase kappa. *Mol Cell Biol* 2005;16:7137–7143. [PubMed: 16055723]
15. Mizukami S, Kim TW, Helquist SA, Kool ET. Varying DNA base-pair size in subangstrom increments: evidence for a loose, not large, active site in low-fidelity Dpo4 polymerase. *Biochemistry* 2006;45:2772–2778. [PubMed: 16503632]
16. Potapova O, Chan C, DeLucia AM, Helquist SA, Kool ET, Grindley ND, Joyce CM. DNA polymerase catalysis in the absence of Watson-Crick hydrogen bonds: analysis by single-turnover kinetics. *Biochemistry* 2006;45:890–898. [PubMed: 16411765]
- 17. Xia J, Noronha A, Toudjarska I, Li F, Akinc A, Braich R, Frank-Kamenetskii M, Rajeev KG, Egli M, Manoharan M. Gene silencing activity of siRNAs with a ribo-difluorotoluy nucleotide. *ACS Chem Biol* 2006;1:176–183. [PubMed: 17163665]This report (combined with ref. 18) showed that hydrogen bonding is not essential at all positions for pairing and biological activity in cellular RNA interference.
18. Somoza A, Chelliserrykattil J, Kool ET. The roles of hydrogen bonding and sterics in RNA interference. *Angew Chem Int Ed Engl* 2006;45:4994–4997. [PubMed: 16802393]
19. Sun Z, McLaughlin LW. Effects of the minor groove pyrimidine nucleobase functional groups on the stability of duplex DNA: The impact of uncompensated minor groove amino groups. *Biopolymers* 2007;87:183–195. [PubMed: 17657710]
20. Meena Sun Z, Mulligan C, McLaughlin LW. Removal of a single minor-groove functional group eliminates A-tract curvature. *J Am Chem Soc* 2006;128:11756–11757. [PubMed: 16953605]

21. Morales JC, Kool ET. Minor groove interactions between polymerase and DNA: More essential than Watson-Crick bonds. *J Am Chem Soc* 1999;121:2723–2724.
22. McCain MD, Meyer AS, Schultz SS, Glekas A, Spratt TE. Fidelity of mispair formation and mispair extension is dependent on the interaction between the minor groove of the primer terminus and Arg668 of DNA polymerase I of *Escherichia coli*. *Biochemistry* 2005;44:5647–5659. [PubMed: 15823023]
23. Matsuda S, Leconte AM, Romesberg FE. Minor groove hydrogen bonds and the replication of unnatural base pairs. *J Am Chem Soc* 2007;129:5551–5557. [PubMed: 17411040]
24. Kim Y, Leconte AM, Hari Y, Romesberg FE. Stability and polymerase recognition of pyridine nucleobase analogues: role of minor-groove H-bond acceptors. *Angew Chem Int Ed* 2006;45:7809–7812.
25. Leconte AM, Matsuda S, Hwang GT, Romesberg FE. Efforts towards expansion of the genetic alphabet: pyridone and methyl pyridone nucleobases. *Angew Chem Int Ed* 2006;45:4326–4329.
26. Hirao I. Unnatural base pair systems for DNA/RNA-based biotechnology. *Curr Opin Chem Biol* 2006;10:622–627. [PubMed: 17035074]
- 27. Sintim HO, Kool ET. Remarkable sensitivity to DNA base shape in the DNA polymerase active site. *Angew Chem Int Ed Engl* 2006;45:1974–1979. [PubMed: 16506248] This report describes systematic shape variations in DNA base analogs, resulting in strong effects in DNA replication.
28. Zhang X, Lee I, Zhou X, Berdis AJ. Hydrophobicity, shape, and π -electron contributions during translesion DNA synthesis. *J Am Chem Soc* 2006;128:143–149. [PubMed: 16390141]
29. Zhang X, Lee I, Berdis AJ. The use of nonnatural nucleotides to probe the contributions of shape complementarity and π -electron surface area during DNA polymerization. *Biochemistry* 2005;44:13101–13110. [PubMed: 16185078]
30. Reineks EZ, Berdis AJ. Evaluating the contribution of base stacking during translesion DNA replication. *Biochemistry* 2004;43:393–404. [PubMed: 14717593]
31. Beckman J, Kincaid K, Hock M, Spratt T, Engels J, Cosstick R, Kuchta RD. Human DNA polymerase α uses a combination of positive and negative selectivity to polymerize purine dNTPs with high fidelity. *Biochemistry* 2007;46:448–460. [PubMed: 17209555]
32. Kincaid K, Beckman J, Zivkovic A, Halcomb RL, Engels JW, Kuchta RD. Exploration of factors driving incorporation of unnatural dNTPs into DNA by Klenow fragment (DNA polymerase I) and DNA polymerase α . *Nucleic Acids Res* 2005;33:2620–2628. [PubMed: 15879351]
33. Paul N, Nashine VC, Hoops G, Zhang P, Zhou J, Bergstrom DE, Davisson VJ. DNA polymerase template interactions probed by degenerate isosteric nucleobase analogs. *Chem Biol* 2003;10:815–825. [PubMed: 14522052]
34. Gong J, Sturla SJ. A synthetic nucleoside probe that discerns a DNA adduct from unmodified DNA. *J Am Chem Soc* 2007;129:4882–4883. [PubMed: 17402738]
35. Hamm ML, Rajguru S, Downs AM, Cholera R. Base pair stability of 8-chloro- and 8-iodo-2'-deoxyguanosine opposite 2'-deoxycytidine: Implications regarding the bioactivity of 8-oxo-2'-deoxyguanosine. *J Am Chem Soc* 2005;127:12220–12221. [PubMed: 16131189]
36. Taniguchi Y, Kool ET. Nonpolar isosteres of damaged DNA bases: effective mimicry of mutagenic properties of 8-oxopurines. *J Am Chem Soc* 2007;129:8836–8844. [PubMed: 17592846]
37. Woods KK, Lan T, McLaughlin LW, Williams LD. The role of minor groove functional groups in DNA hydration. *Nucleic Acids Res* 2003;31:1536–1544. [PubMed: 12595562]
38. Matsuda S, Fillo JD, Henry AA, Rai P, Wilkens SJ, Dwyer TJ, Geierstanger BH, Wemmer DE, Schultz PG, Spraggon G, Romesberg FE. Efforts toward expansion of the genetic alphabet: Structure and replication of unnatural base pairs. *J Am Chem Soc* 2007;129:10466–10473. [PubMed: 17685517]
39. Gallego J, Loakes D. Solution structure and dynamics of DNA duplexes containing the universal base analogues 5-nitroindole and 5-nitroindole 3-carboxamide. *Nucleic Acids Res* 2007;35:2904–2912. [PubMed: 17438041]
40. Lynch SR, Liu H, Gao J, Kool ET. Toward a designed, functioning genetic system with expanded-size base pairs: solution structure of the eight-base xDNA double helix. *J Am Chem Soc* 2006;128:14704–14711. [PubMed: 17090058]
41. Tawarada R, Seio K, Sekine M. Synthesis and properties of artificial base pairs by use of halogen bonds. *Nucleic Acids Symp Ser (Oxf)* 2006;50:121–122.

42. Dohno C, Okamoto A, Saito I. Stable, specific, and reversible base pairing via Schiff base. *J Am Chem Soc* 2005;127:16681–16684. [PubMed: 16305258]
43. Battersby TR, Albalos Maria, Friesenhahn Michel J. An unusual mode of DNA duplex association: Watson-Crick interaction of all-purine deoxyribonucleic acids. *Chemistry & Biology* 2007;14:525–531. [PubMed: 17524983]
44. Clever GH, Kaul C, Carell T. DNA-metal base pairs. *Angew Chem Int Ed Engl* 2007;46:6226–6236. [PubMed: 17640011]
45. Lai JS, Kool ET. Selective pairing of polyfluorinated DNA bases. *J Am Chem Soc* 2004;126:3040–3041. [PubMed: 15012120]
46. Zahn A, Brotschi C, Leumann CJ. Pentafluorophenyl-phenyl interactions in biphenyl-DNA. *Chemistry* 2005;11:2125–2129. [PubMed: 15714531]
- 47. Hikishima S, Minakawa N, Kuramoto K, Fujisawa Y, Ogawa M, Matsuda A. Synthesis of 1,8-naphthyridine C-nucleosides and their base-pairing properties in oligodeoxynucleotides: Thermally stable naphthyridine:imidazopyridopyrimidine base-pairing motifs. *Angew Chem Int Ed Engl* 2005;44:596–598. [PubMed: 15612071] This study shows the results of elegant design of new base pairs that form four hydrogen bonds.
48. Minakawa N, Kojima N, Hikishima S, Sasaki T, Kiyosue A, Atsumi N, Ueno Y, Matsuda A. New base pairing motifs. The synthesis and thermal stability of oligodeoxynucleotides containing imidazopyridopyrimidine nucleosides with the ability to form four hydrogen bonds. *J Am Chem Soc* 2003;125:9970–9982. [PubMed: 12914460]

Figure 1. Varied depictions of the four natural DNA bases, showing differences in size, shape, electrostatics, and functional groups. The chemical information of the cell is defined by these differences.

Figure 2.

The consensus base pair shape. Space-filling models of the four common pairs (with sugars attached) are overlaid. Darkest green color indicates space that is occupied by all four, while lightest areas show variations that project outward.

Figure 3.

The strong effect of systematic changes in base size on the fidelity of replication in *Escherichia coli*. Thymine analogs varied in size over a 1 Å range, and were found to be replaced by the bases shown after replication. Data from Kim *et al.* [2].

Figure 4. Representative examples from the menagerie of modified DNA bases in the literature, showing the wide range of structures, sizes, shapes, and polarities.