JANUS WEDGES: a new approach towards nucleobase-pair recognition

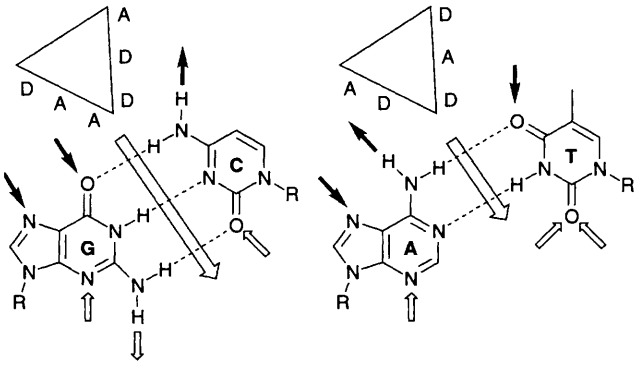
Neil Branda,† Guido Kurz and Jean-Marie Lehn*

Laboratoire de Chimie Supramoléculaire (URA 422), Institut Le Bel, Université Louis Pasteur, 4 rue Blaise Pascal, F-67000 Strasbourg, France

The two-faced JANUS-type heterocycles 1a and 1b are designed to recognize a cytosine–uracil mismatched base-pair by insertion in a wedge-like fashion between the bases to give a triad motif utilizing the maximum number of Watson–Crick interactions, as shown by ^{1}H NMR spectroscopic studies.

Double stranded structures of nucleic acids may be recognized by molecules that either bind in the major or minor groove of the double helix, or intercalate between the base-pairs.[1] In contrast to the rather unspecific aromatic π-stacking interactions of intercalators, major and minor groove binders like oligonucleotides[2] or distamycin and netropsin derivatives,[3] present a more or less pronounced sequence specificity through the formation of specific hydrogen bonds. In this approach, base-pair discrimination relies on subtle differences in number and accessibility of hydrogen bonding sites in the grooves of the double helix. As a consequence, a complete set of DNA binders with specificity for every natural base-pair (AT, TA, GC and CG) has not yet been developed.

We propose here a new approach towards base-pair recognition, the WEDGE concept. It is based on the design of wedge-shaped heterocycles presenting two hydrogen bonding arrays that are complementary to the Watson–Crick faces of a pair of either matched or unmatched nucleobases. By analogy with previously described JANUS heterocycles[4] presenting two hydrogen bonding 'faces', they are referred to as JANUS wedges, after the Roman god depicted with two faces on opposite sites of his head. Instead of targeting hydrogen bonding sides of the intact base-pairs as realized in most of the DNA recognition motifs, the JANUS wedges are designed to bind by inserting *between* the two nucleobases (Scheme 1). The resulting formation of five or six hydrogen bonds should provide high binding affinity and specificity. However, since JANUS heterocycles presenting recognition surfaces complementary to those of the two nucleobases in a natural pair, are necessarily self-complementary, the formation of the derived base–wedge–base triad has to compete both with disruption of the base-pair and with self-aggregation of the wedge molecules.

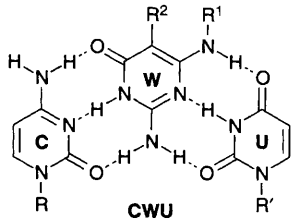


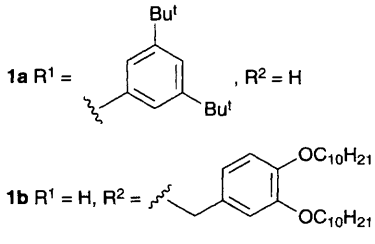
Scheme 1 The WEDGE concept. JANUS-type wedges with two hydrogen bonding faces (A: acceptor, D: donor) are designed to bind by insertion *between* base-pairs forming a triplet with the maximum number of Watson–Crick interactions. In contrast, common DNA binding molecules target hydrogen bonding sites of the *intact* base-pair, as illustrated by black (or small open) arrows for the major (or minor) groove side of the AT base-pair.

In order to avoid these interferences one may target not the canonical base-pairs AT and GC but, instead, pairs of mismatched bases which are present in important structural features such as mutated sites of DNA and unpaired regions of RNA forming bulges or loops. Of the possible purine–purine, purine–pyrimidine and pyrimidine–pyrimidine mismatches, the cytosine–uracil pair is a particularly attractive target. Crystal structures of double helical nucleic acids show that, unlike *e.g.* the guanine–adenine[5] and the guanine–uracil[6] 'wobble' pairs which form two hydrogen bonds, the cytosine–uracil pair exhibits only one direct and one water-bridged hydrogen bond.[6] The JANUS molecules **1a** and **1b**, differing only in the position of the organic solubilizing groups, are CU-wedges designed to insert between cytosine and uracil so as to yield a total of six hydrogen bonds upon formation of the ternary CWU complex. Additional stabilization may be provided by enhanced aromatic π-stacking interactions due to the increased area of the planar triad.

The lipophilic wedges **1a** and **1b** were synthesized as outlined in Scheme 2.‡ Compound **1a** (white solid, mp 290–293 °C, R_f 0.36 in CH_2Cl_2–10%MeOH–1%Et_3N on silica) was obtained in one step from commercially available starting materials by the reaction of 2-amino-6-chloropyrimidin-4-ol with 3,5-di-*tert*-butylaniline in a refluxing mixture of acetic and hydrochloric acid (28% yield after recrystallization from EtOH). Derivative **1b** (white solid, mp 192–195 °C, R_f 0.41 in CH_2Cl_2–10%MeOH–1%Et_3N on silica) was prepared in 40% overall yield through a Knoevenagel condensation of ethyl cyanoacetate and 3,4-didecylbenzaldehyde, followed by selective hydrogenation of the alkene and subsequent cyclocondensation with guanidine.

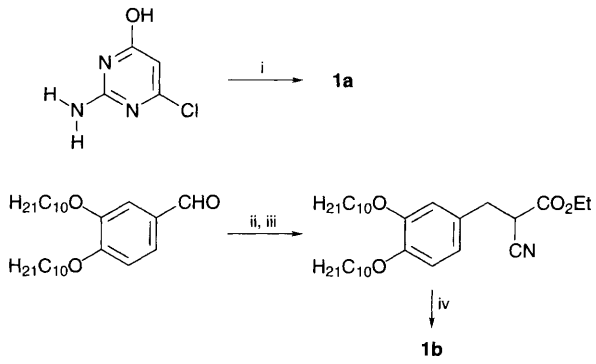
The binding properties of the CU-wedges **1a** and **1b** were assessed by analysing the ^{1}H NMR spectral changes occurring on titration[7] of a $CDCl_3$ solution of **1a** or **1b** (10 mM) with a 1 : 1





mixture of the lipophilic derivatives *N*-propyluracil (PU) and 3,5-di-*tert*-butylbenzylcytosine (BC, 100 mM).§ The results suggest an intimate association through hydrogen bonding. While the signals of all aromatic protons shift to a very small extent (<0.1 ppm) excluding binding contributions from aromatic stacking, those of the heteroatomic protons of **1a** or **1b** move significantly downfield (>1 ppm for NH, NH_2 and NHAr) as would be expected for the formation of strong hydrogen bonds.

Titrations of **1a** and **1b** with PU and BC in separate experiments allows an evaluation of the association constants (K_a) in $CDCl_3$ (Fig. 1). Both data sets show good correlation with calculated curves using a 1:1 binding model. The K_a values of 2.6–4.2 × 10^2 and 6.4–8 ×10^4 dm^3 mol^{-1} obtained are in the range expected for cytosine and uracil binding to their complements (170 dm^3 mol^{-1} for diaminopurine–U; 10^4–10^5 dm^3 mol^{-1} for G–C).[8] The position of the solubilizing groups on the two wedges (W) appears to have little affect on the binding affinity, since both give similar binding constants. Because W and C have such a high affinity, they are fully associated at the concentration used and can be considered as a single species. This is supported by the observation that the W–NH moves insignificantly when a 1:1 mixture of **1a** and BC is titrated with PU. The K_a values translate into a rough estimate of the amount of ternary complex as 50–60% at 10 mM concentration.



Scheme 2 *Reagents and conditions*: i, 3,5-di-*tert*-butylaniline, AcOH, HCl, reflux, 28%; ii, ethyl 2-cyanoacetate, piperidine, C_6H_6, 91%; iii, H_2, PtO_2, EtOH, Et_2O, 88%; iv, guanidinium chloride, NaOEt, EtOH, 50%

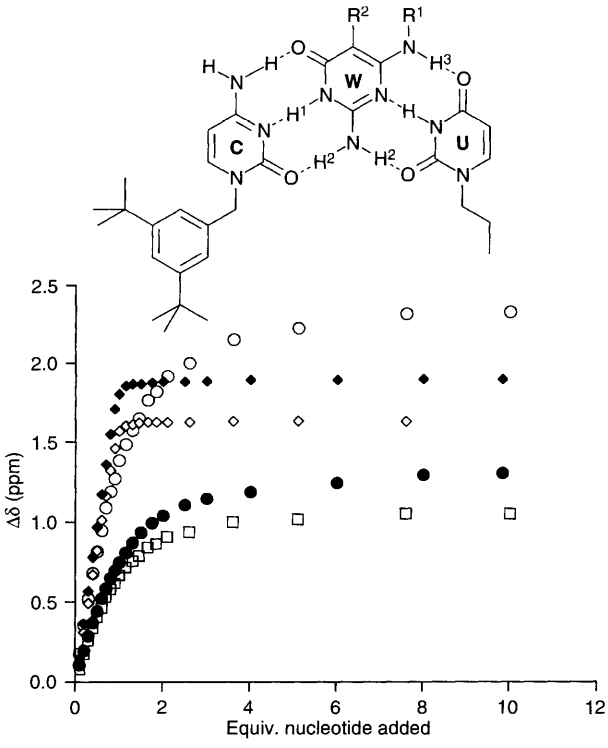


Fig. 1 ^1H NMR titrations of CU-wedges **1a** and **1b** with *N*-propyluracil, [(○) **1a**-H^3; (●) **1b**-H^3; (□) **1a**-H^2] and 1-(3,5-di-*tert*-butylbenzyl)cytosine [(◇) **1a**-H^1; (◆) **1b**-H^1] in $CDCl_3$. In all cases, a 0.01 M solution of the wedge was titrated with 0.10 M solutions of nucleobase derivative, except for wedge **1a**–cytosine, which was titrated with a 0.05 M solution of the cytosine derivative.

In this report, we described the new WEDGE strategy for the recognition of nucleobase-pairs based on the use of JANUS type heterocycles **1a** and **1b** as wedges for insertion between a CU pair. These CU wedges simultaneously bind liphophilic derivatives of C and U in chloroform to form a CWU triad maximising all Watson–Crick hydrogen bonds.

These model studies encourage follow-up investigations with water-soluble wedges targeting mismatch regions in double-stranded DNA (mutations) and RNA (loops, bulges). Wedges for other pairs of nucleobases (for instance derivatives of triaminotriazine for UU or TT pairs) as well as larger JANUS molecules with more distant hydrogen bonding faces[4] may also be considered. Longer stretches of mismatched base-pairs might form triplex structures[9] with oligomeric wedges that contain a DNA or PNA backbone. The concept presented here provides a novel approach directed towards the recognition of double helical nucleic acids by wedging between mismatched and, eventually, canonical base-pairs.

We thank the Natural Sciences and Engineering Research Council of Canada (N. B.) and the French Government (G. K.) for postdoctoral fellowships.

Footnotes

† Current address: Department of Chemistry, University of Alberta, Edmonton, AB, Canada, T6G 2E1.

‡ All new compounds were characterized by ^1H and ^{13}C NMR spectroscopy and by high-resolution mass spectrometry.

§ Chloroform-soluble nucleotides were prepared in one step by alkylating uracil with 3-bromopropane or cytosine with 3,5-di-*tert*-butylbenzyl chloride (NaH, DMF, 70 °C in both cases).

References

1 C. L. Probst and T. J. Perun, *Nucleic Acid Targeted Drug Design*, Marcel Dekker, New York, 1989.

2 H. E. Moser and P. B. Dervan, *Science*, 1987, **238**, 645; T. Le Doan, L. Perrouault, D. Praseuth, N. Habhoub, J. L. Decout, N. T. Thuong, J. Lhomme and C. Hélène, *Nucleic Acids Res.*, 1987, **15**, 7749; M. Cooney, G. Czernuszewicz, E. H. Postel, J. Flint and M. E. Hogan, *Science*, 1988, **241**, 456; P. A. Beal and P. B. Dervan, *Science*, 1991, **251**, 1360. For a review, see N. T. Thuong and C. Hélène, *Angew. Chem., Int. Ed. Engl.*, 1993, **32**, 666.

3 See for instance, M. Mrksich and P. B. Dervan, *J. Am. Chem. Soc.*, 1995, **117**, 3325 and references cited therein.

4 A. Marsh, E. G. Nolen, K. M. Gardinier and J.-M. Lehn, *Tetrahedron Lett.*, 1994, **35**, 397.

5 T. Brown, W. N. Hunter, G. Kneale and O. Kennard, *Proc. Natl. Acad. Sci. USA*, 1986, **83**, 2402; G. G. Privé, U. Heinemann, S. Chandrasegaran, L.-S. Kan, M. L. Kopka and R. E. Dickerson, *Science*, 1987, **238**, 498; G. G. Privé, K. Yanagi and R. E. Dickerson, *J. Mol. Biol.*, 1991, **217**, 177.

6 S. R. Hollbrook, C. Cheong, I. Tinoco and S.-H. Kim, *Nature*, 1991, **353**, 579.

7 ^1H NMR titrations were performed following the procedure of K. Williams, B. Askew, P. Ballester, C. Buhr, K. S. Jeong, S. Jones and J. Rebek Jr., *J. Am. Chem. Soc.*, 1989, **111**, 1090. Values of K_a were estimated using least-squares fit programming.

8 Y. Kyogoku, R. C. Lord and A. Rich, *Proc. Natl. Acad. Sci. USA*, 1967, **57**, 250; Y. Kyogoku, R. C. Lord and A. Rich, *Biochim. Biophys. Acta*, 1969, **179**, 10.

9 For studies of triplex structures, see for instance T. L. Trapane, M. S. Christopherson, C. D. Roby, P. O. P. Ts'o and D. Wang, *J. Am. Chem. Soc.*, 1994, **116**, 8412; R. Bandaru, H. Hashimoto and C. Switzer, *J. Org. Chem.*, 1995, **60**, 786.

Received, 6th August 1996; Com. 6/05494C