RecA Protein Promotes Strand Exchange with DNA Substrates Containing Isoguanine and 5-Methyl Isocytosine†

Kevin P. Rice,‡ John C. Chaput,§ Michael M. Cox,‡ and Christopher Switzer*,§

Department of Biochemistry, University of Wisconsin−Madison, Madison, Wisconsin 53706, and Department of Chemistry, University of California−Riverside, Riverside, California 92521

Received February 11, 2000; Revised Manuscript Received May 12, 2000

ABSTRACT: The *Escherichia coli* RecA protein pairs homologous DNA molecules and promotes DNA strand exchange in vitro. We have examined DNA strand exchange between a 70 nucleotide ssDNA fragment and a 40 bp duplex, in which all G and C residues (at 18 positions distributed throughout the 40 bp exchanged region) were replaced with the nonstandard nucleosides 2′-deoxyisoguanosine (iG) and 2′-deoxy-5-methylisocytidine (MiC), respectively. We demonstrate that the nonstandard oligonucleotides are substrates for the RecA protein, permitting DNA strand exchange in vitro at a rate and efficiency comparable to exchange with normal DNA substrates. This observation provides an expanded experimental basis for discussions of potential roles for iG and MiC in a genetic code. Experiments of this type also provide another avenue for exploring RecA-facilitated DNA pairing mechanisms.

There are a number of nonstandard nucleobases capable of forming Watson−Crick base pairs compatible with DNA helices (*1*−*3*). Among the possible nonstandard base pairs, the one formed by isoguanine (iG)[1] and isocytosine (iC) is distinct in that it differs from one of the two natural base pairs only by the relative placement of functional groups (Figure 1) and retains a carbon−nitrogen glycosidic bond. As a consequence, a case can be made for its prebiotic plausibility (*4*, *5*). Indeed, the possible existence of these bases during the early history of Earth or other planets has led to speculation about the origin of the extant genetic system utilizing the bases ATCG(U) (*6*−*10*).

In vitro experimental systems have begun to define the structural and functional utility of the iC:iG base pair. Thus, the iC:iG base pair has been determined to be at least as stable as a G:C pair in double helices (*10*−*12*), DNA and RNA polymerases have been shown to accept iG and iC as substrates (*2*, *7*, *13*, *14*), and ribosomal translation can be mediated by the iG:iC pair (*15*). While in each of the aforementioned systems iG:iC recognition succeeds, they all suffer to some extent from a competing recognition process involving iG and T. In particular, when a competing adenosine-bearing substrate is not supplied in either polymerase (*2*, *7*, *13*) or translation experiments (*15*), iG:T pairs are also functional, although to differing extents. In addition to the N1−H tautomer shown in Figure 1, there is evidence that iG adopts other tautomers including an enol (*16*−*18*), N3−H (*19*), and imino oxo tautomers (*10*). Further, it has been demonstrated both crystallographically and in solution that more than one tautomer of iG can pair with T in a DNA duplex, where one tautomer is the N1−H form and the other is most probably the enol form (*18*). The enol form of iG is also known to be favored by nonpolar environments and increases in temperature (*16*, *18*). From a structural viewpoint, the enol form of iG is the most likely cause for infidelity during polymerase copying and translation experiments since the T:iG(N1−H) pair leads to a wobble geometry in a DNA duplex, and the enol form of iG should retain a Watson−Crick geometry. Here we expand in vitro structure and function studies of iC and iG by examining RecA-mediated strand exchange of DNA bearing these bases, and approach the question of whether a genetic system incorporating iC and iG might be capable of homologous recombination.

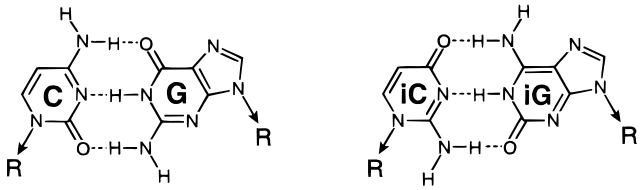


FIGURE 1: The nonstandard nucleotide bases iG and iC differ from G and C in that each has its carbonyl and exocyclic amine groups transposed. This permits Watson−Crick-type pairing between these bases in a DNA double helix.

Protein-mediated homologous recombination provides an essential set of DNA repair pathways in all cells, and contributes to a range of processes including proper chromosome segregation at meiosis in eukaryotes and conjugation

† This work was supported by grants from the NASA Astrobiology Institute (to C.S.) and the National Institutes of Health (GM32335, to M.M.C.). The authors thank Patrick Sung for providing the Rad51 protein as well as help and advice in its use, and Aimee Eggler for work with the Rad51 protein experiments.

* To whom correspondence should be addressed. Phone: (909) 787-7266. Fax: (909) 787-4713. E-mail: switzer@mail.ucr.edu.

‡ University of Wisconsin−Madison.

§ University of California−Riverside.

[1] Abbreviations: UV, ultraviolet light; ssDNA, single-stranded DNA; dsDNA, double-stranded DNA; bp, base pair(s); C, cytosine; G, guanosine; iC, isocytosine; MiC, 5-methylisocytosine; iG, isoguanosine; ATP[_S], adenosine 5′-*O*-(3-thio)triphosphate; PAGE, polyacrylamide gel electrophoresis; Hepes, *N*-(2-hydroxyethyl)piperazine-*N*′-(2-ethanesulfonic acid); DTT, dithiothreitol; EDTA, ethylenediamine tetraacetic acid; SDS, sodium dodecyl sulfate; $Mg(OAc)_2$, magnesium acetate; SSB, the single-stranded DNA binding protein of *E. coli*.

10.1021/bi0003339 CCC: $19.00
Published on Web 07/29/2000

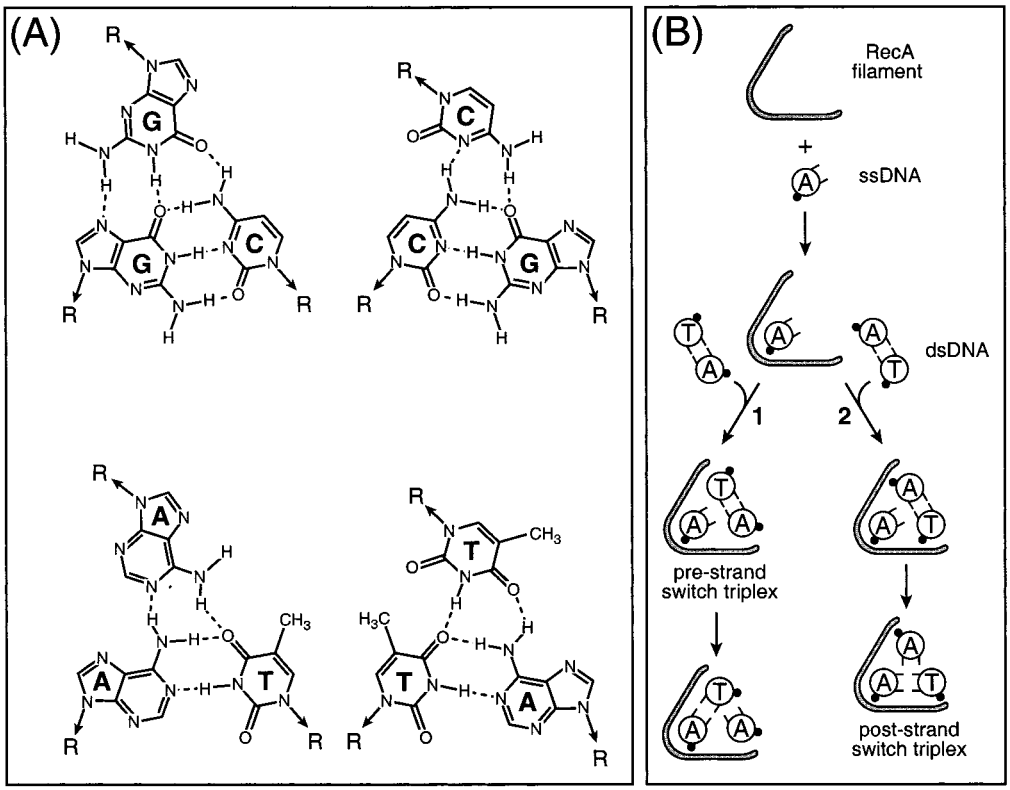


FIGURE 2: (A) These base triples predicted by the structure of R-DNA could serve as the recognition mechanism for homologous alignment during RecA protein mediated strand exchange. Non-Watson−Crick hydrogen bond donors and acceptors in the major groove provide a potentially unique readout such that each base would only recognize one base pair. (B) RecA protein binds to ssDNA forming a nucleoprotein filament. This filament then samples dsDNA for homology. The incoming dsDNA substrate can enter the filament via its major or minor groove. In this figure, the small closed circles on the outsides of the DNA bases represents attachment to the deoxyribose phosphate backbone. Thus, pathway 1 depicts a major groove approach by the dsDNA substrate. A possible intermediate in this pathway is the "pre-strand switch" R-form triplex, which could serve as the molecular basis for homology recognition. Pathway 2 depicts the minor groove approach. Local unwinding of the duplex and Watson−Crick base pairing would be the most likely recognition mechanism in this pathway. However, if the displaced strand persisted in the major groove of the hybrid duplex, a "post strand-switch" triplex could result.

in bacteria. In bacteria, the primary function of the homologous recombination systems is the reactivation of arrested replication forks (*20*, *21*). The central steps of most recombination pathways are promoted by the RecA family of proteins, including the RecA protein of bacteria and the Rad51 and Dmc1 proteins of eukaryotes. In vitro, RecA protein pairs two homologous DNA molecules and promotes an exchange involving up to four DNA strands. A RecA-mediated three-strand exchange reaction (involving a single strand and a homologous duplex) occurs in several stages. The first stage involves ATP-dependent filamentation of RecA protein onto the ssDNA. This nucleoprotein filament represents the active species of RecA protein. The DNA within this filament is extended and underwound by nearly 40%. The filament itself contains a deep helical groove in which the nucleotide bases are presented. DNA pairing follows filamentation. The filament recruits duplex DNA and samples for homology. Once homologous regions are aligned, a DNA strand switch occurs, resulting in a hybrid duplex DNA species and a displaced single strand. Under some circumstances, extensive DNA strand exchange is coupled to ATP hydrolysis as detailed elsewhere (*20*, *22*).

DNA strand exchange has been extensively studied in vitro. However, one of the fundamental aspects of this process, alignment of DNA homology, remains poorly understood. During the course of RecA protein-mediated three-strand DNA exchange, several potential DNA interactions and intermediates have been proposed. In 1966, Lacks proposed a novel three-stranded pairing interaction as the recombination intermediate for homology recognition between dsDNA and ssDNA substrates (*23*). More detailed models for DNA pairing have been offered over the past 15 years (*24*−*27*). Much of the thinking has focused on a hypothetical DNA triplex structure, much like that originally proposed by Lacks and now called R-form DNA (*27*, *28*).

The proposed recombination triplex has broad implications not only as a recombination intermediate but also as a novel DNA structure. Molecular modeling studies have provided much of the structural basis for R-DNA (*27*, *29*). These models include unique hydrogen bond readouts for every base with its homologous base pair (Figure 2A). The third strand is aligned in the major groove of the Watson−Crick duplex, with like DNA strands in parallel. Specific hydrogen bonds are proposed to form between the third strand and the functional groups from both bases of the base pair within the major groove.

To make use of the R form triplex as a mechanism for homologous alignment, the RecA-bound single strand must interact with the incoming duplex DNA in the latter's major groove. While some evidence has favored an initial interaction in the major groove (*26*, *30*−*33*), this is not the only plausible pathway for DNA pairing. Several recent reports have supported a pathway in which the minor groove is presented first (Figure 2B) (*34*−*38*). The novel R-form DNA

triplex could be required for homologous alignment (major groove first path) or could exist as a possibly inconsequential product of the strand switch (minor groove first path). However, there has been no direct observation of an R-form triplex, and no conditions have been found in which such a structure is stable in the absence of RecA or some similar protein. The logical alternative to alignment via an R-form triplex is simple Watson−Crick base pairing between the strands of the hybrid duplex. Homology recognition via base pairing might require base flipping, which could occur within the context of either a major or minor groove first pathway. A number of predictions of a base-flipping pathway for homology recognition have been borne out by experiment (*39*−*41*).

We demonstrate that DNAs containing MiC and iG are excellent substrates for RecA-mediated DNA strand exchange. The same result is obtained for the yeast Rad51 protein. We also present modeling studies of a hypothetical R-form triplex containing iG and MiC. The modeling allows for structures consistent with the proposed R-form structure, but predicts a loss of fidelity when MiC/iG substrate DNAs are employed.

MATERIALS AND METHODS

General Methods. Proton (^{1}H) and phosphorus (^{31}P) NMR spectra were recorded on a Varian Inova-300 (300 and 125 MHz, respectively) spectrometer. Mass spectra (MS) and exact mass measurements were obtained on a VG-ZAB-2FHF mass spectrometer using FAB ionization obtained by the staff of the Southern California Regional Mass Spectrometry Facility (University of California, Riverside, CA). Ultraviolet spectra were obtained using a Hewlett-Packard 8452A diode array spectrophotometer. All experiments involving air and/or moisture sensitive materials were carried out under an argon (Ar) atmosphere.

Wild-type *Escherichia coli* RecA protein was purified using a procedure created for the purification of the RecA mutant K72R (*42*). Storage and handling of RecA protein were as previously described (*43*). The concentration of RecA protein was determined by UV absorption at 280 nm using the extinction coefficient of 0.59 A_{280} mg^{-1} mL (*44*). The yeast Rad51 protein was purified as previously described (*45*). Rad51 protein concentration was calculated by UV absorption at 280 nm using the extinction coefficient of 0.30 A_{280} mg^{-1} mL (*46*). T4 polynucleotide kinase and its accompanying 10× buffer were purchased from Promega. ATPγS was purchased as a 100 mM solution from Boehringer Mannheim. [γ^{32}P]ATP was purchased from Amersham-Pharmacia. Tris base, boric acid, EDTA, magnesium acetate, sodium dodecyl sulfate, acrylamide, bis-acrylamide, urea, glycerol, 1 M hydrochloric acid, and 1 M acetic acid were purchased from Fisher Chemical. Formamide was purchased from Sigma. Dithiothreitol (DTT) was purchased from Research Organics Inc. Solvents were purchased from Fisher Scientific. DNA synthesis reagents, phosphoramidites, and controlled pore glass support was purchased from Glen Research. Pyridine and CH_2Cl_2 were freshly distilled over CaH_2 as needed.

Nucleoside Synthesis. The iG phosphoramidite was synthesized as previously described (*10*). The MiC phosphoramidite was synthesized in steps as follows:

2′-Deoxy-5-methylisocytidine. 2,5′-Anhydrothymidine (*47*) (255 mg, 1.1 mmol) was added to methanol (15 mL) presaturated at 0 °C with ammonia. The mixture was heated for 5 h in a sealed tube at 100 °C. The reaction mixture was cooled to room temperature and concentrated to ∼1 mL and then precipitated with acetonitrile (15 mL). Product was collected by centrifugation, washed and dried to afford 216 mg as a white powder (85% yield) which was used without further purification. ^{1}H NMR (DMSO): δ 1.70 (s, 3H), 2.18 (m, 2H), 3.58 (m, 2H), 3.76 (m, 1H), 4.25 (m, 1H), 5.83 (t, 1H, J = 6.7 Hz), 7.50 (s, 1H).

N-(N,N-Dimethylformamidine)-2′-deoxy-5-methylisocytidine. To a solution of 2′-deoxy-5-methylisocytidine (125 mg, 0.51 mmol) in pyridine (5 mL) was added dropwise *N*,*N*-dimethylformamidine dimethyl acetal (0.12 mL, 0.77 mmol) at room temperature. The reaction was left stirring for 12 h, then concentrated under reduced pressure. Flash chromatography [SiO_2, MeOH (15%)/CH_2Cl_2] afforded 120 mg of the desired product in 80% yield. ^{1}H NMR (DMSO/D_2O): δ 1.79 (s, 3H), 2.15 (m, 2H), 3.04 (s, 3H) 3.18 (s, 3H), 3.60 (m, 2H), 3.79 (m, 1H), 4.22 (m, 1H), 6.63 (t, 1H, J = 6.4 Hz), 7.75 (s, 1H), 8.58 (s, 1H).

5′-O-(4,4′-Dimethoxytrityl)-N-(N,N-dimethylformamidine)-2′-deoxy-5-methylisocytidine. To a stirred solution of *N*-(*N*,*N*-dimethylformamidine)-2′-deoxy-5-methylisocytidine (240 mg, 0.81 mmol) in pyridine (6 mL) was added 4,4′-dimethoxyltrityl chloride (165 mg, 0.5 mmol) at room temperature. After 2 h, a second portion of 4,4′-dimethoxyltrityl chloride (165 mg, 0.5 mmol) was added and the reaction was left stirring for an additional 2 h, at which time CH_2Cl_2 (150 mL) was added. The reaction mixture was extracted with saturated aqueous $NaHCO_3$ and saturated aqueous NaCl, dried over Na_2SO_4, and concentrated under reduced pressure. Flash chromatography (SiO_2, MeOH (6%)/pyridine (1%)/CH_2Cl_2) afforded 406 mg of the desired product in 84% yield. ^{1}H NMR ($CDCl_3$): δ 1.68 (s, 3H), 2.31 (m, 2H), 3.08 (s, 3H), 3.13 (s, 3H), 3.45 (dd, 2H, J = 3.2 Hz, 10.6 Hz), 3.79 (s, 6H), 4.05 (m, 1H), 4.54 (m, 1H), 6.83 (m, 5H), 6.84 (m, 4H), 7.22−7.41 (m, 9H), 7.59 (s, 1H), 8.82 (s, 1H).

5′-O-(4,4′-Dimethoxytrityl)-N-(N,N-dimethylformamidine)-2′-deoxy-5-methylisocytidine-2-cyanoethyl-N,N′-diisopropylphosphoramidite. To a stirred solution of 5′-*O*-(4,4′-dimethoxytrityl)-*N*-(*N*,*N*-dimethylformamidine)-2′-deoxy-5-methylisocytidine (406 mg, 0.68 mmol) in CH_2Cl_2 (5 mL) was added diisopropylethylamine (0.33 mL, 2.04 mmol) followed by 2-cyanoethoxy (*N*,*N*′-diisopropylamino)chlorophosphine (0.18 mL, 0.82 mmol). After 2 h at room temperature, CH_2Cl_2 (20 mL) was added and the reaction mixture was then extracted with saturated $NaHCO_3$, dried over Na_2SO_4, filtered, and concentrated under reduced pressure. Flash chromatography (SiO_2, MeOH (4%)/pyridine (1%)/CH_2Cl_2) afforded 437 mg of the desired product in 81% yield. ^{1}H NMR ($CDCl_3$): δ 1.05 (d, 3H), 1.16 (m, 9H), 1.65 (s, 3H), 2.28 (m, 2H), 2.42 (t, 1H), 2.60 (t, 1H), 3.08 (d, 3H), 3.14 (s, 3H), 3.30 (dd, 1H), 3.46−3.65 (m, 5H), 3.79 (s, 6H), 4.18 (m, 1H), 4.58 (m, 1H), 6.80 (m, 5H), 7.2−7.4 (m, 9H), 7.66 (d, 1H), 8.84 (s, 1H). ^{31}P NMR ($CDCl_3$): δ 150.23, 149.47. MS (FAB$^+$) *m*/*z* 799. Exact mass (FAB$^+$) calcd for $C_{43}H_{56}N_6O_7P$: 799.394812 (MH$^+$). Found: 799.3971.

Oligonucleotide Synthesis. The following oligonucleo tide substrates containing only standard nucleotide bases were purchased from Operon Technologies in PAGE-

purified form: A, AGTAGACTCAGCGAACTCACTGATCCAGTCTTAGCATCAGTCACGATACCTCGAGATACATACGGACGTA; B, TGATCCAGTCTTAGCATCAGTCACGATACCTCGAGATACA; C, TGTATCTCGAGGTCTCGTGACTGATGCTAAGACTGGATCA; B_f, TAATTTAATTTTAATATTAATTATAATATTTCAAAATATA; C_f, TATATTTTAAAATTTTATAATTAATATTAAAATTAAATTA).

Lyophylized oligonucleotides were resuspended in TE [10 mM Tris-Cl (80% cation) and 1 mM EDTA (pH 8.0)] and stored at −20 °C.

The following isoguanine and 5-methyl isocytosine-containing oligonucleotide substrates were synthesized on a 391EP DNA synthesizer (Applied Biosystems) using commercially available reagents, and MiC and iG phosphoramidites prepared as described above: A′, AGTAGACTCAGCGAACTCACTiGATMiCMiCAiGTMiCTTAiGMiCATMiCAiGTMiCAMiCiGATAMiCMiCTMiCiGAiGATAMiCATACGGACGTA; B′, TiGATMiCMiCAiGTMiCTTAiGMiCATMiCAiGTMiCAMiCiGATAMiCMiCTMiCiGAiGATAMiCA; C′, TiGTATMiCTMiCiGAiGiGTATMiCiGTiGAMiCTiGATiGMiCTAAiGAMiCTiGiGATMiCA; A'_{fA}, AGTAGACTCAGCGAACTCACTAATMiCMiCAATMiCTTAAMiCATMiCAATMiCAMiCAATAMiCMiCTMiCAAAA TAMiCATACGGACGTA; A'_{fT}, AGTAGACTCAGCGAACTCACTiGATTTAiGTTTTAiGTATTAiGTTATiGATATTTTiGAiGATATATACGGACGTA.

Oligonucleotides were synthesized trityl-off. Cleavage from the solid support and deprotection was accomplished in concentrated NH_4OH for 16 h at 55 °C. Oligonucleotides were purified by preparative gel electrophoresis and quantified by UV absorbance at 260 nm at 70 °C.

Oligonucleotides were 5′-^{32}P-labeled by incubating with T4 polynucleotide kinase (5−10 units) and [γ^{32}P]ATP (0.5 μM) for 30 min at 37 °C in the buffer recommended by the enzyme supplier [final concentrations: 70 mM Tris-Cl (pH 7.6), 10 mM $MgCl_2$, and 5 mM DTT] with a final reaction volume of 50 μL. The single-stranded oligonucleotides were then purified by first mixing with an equal volume of formamide and then incubating at 55 °C for 5 min, followed by electrophoresis on a denaturing gel containing 12% acrylamide and 7 M urea in 1× TBE buffer [45 mm Tris-borate (50% cation), 1 mM EDTA (pH 8)]. Radiolabeled bands were visualized by a 20 s exposure to X-ray film. Processed film was placed under the gel to facilitate band identification; the bands were then excised with a razor blade. Unlabeled bands were identified by placing the gel over a thin-layer chromatography (TLC) plate containing F-254 indicator (Whatman catalog no. 4861-820) followed by a brief exposure to short-wave UV during which the DNA band appeared due to quenched fluorescence of the TLC plate. DNA was extracted from excised intact gel slices by three successive incubations of at least 1 h in 1 mL of TE at 37 °C. After each incubation, the liquid was removed and the DNA concentrated in Microcon-3 concentrators (Amicon) for 90 min at 13600*g* with the next extraction liquid being added to the previously concentrated material, then reconcentrated. The dsDNA 40mer substrates were formed by annealing the two complementary ssDNA 40mers, heating to 85 °C for 5 min followed by slow cooling to 25 °C. Double-stranded oligonucleotides were isolated from single-strands by electrophoresis in 20% acrylamide/1× TBE native gels and extracted as described above. Oligonucleotide concentration was determined by UV absorption at 260 nm using the extinction coefficients provided by the manufacturer. All DNA concentrations are reported here in nucleotides.

Thermodynamic Calculations. UV absorbance versus temperature profiles were measured on a Hewlett-Packard 8452A diode-array spectrophotometer in a temperature-controlled cell holder with an HP 89090A peltier temperature controller. The absorbance was recorded in the reverse and forward directions at temperatures of 25−90 °C at a rate of 1 °C/min. Experiments were performed in duplicate with different samples and the T_ms were averaged. All melts were done in 1 M NaCl, 10 mM Na_2HPO_4, pH 7, 0.1 mM EDTA, pH 7, and 5 μM total oligonucleotide concentration. Duplex free-energy values were determined from melting curves by nonlinear regression using a two-state model according to the method reported by Turner (*48*).

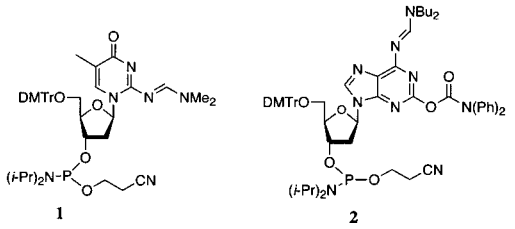
Modeling Methods. Computations on the nucleobase triples were performed using the PM3 semiempirical method implemented by Gaussian 98 (*49*), and the Hartree−Fock ab initio method implemented by Jaguar (Schrödinger, Portland, OR). Triple-stranded R-form DNA models were created through the use of Macromodel (Schrödinger, Portland, OR) and the PM3 optimized geometries of the base triples.

DNA Pairing Reactions. Standard DNA pairing reactions were carried out in 20 mM Tris-OAc buffer (80% cation) with added 1 mM DTT and 1% glycerol at 37 °C. The 70mer ssDNA substrate (56 μM) was preincubated with RecA protein (18.7 μM), ATPγS (3 mM), and 1 mM $Mg(OAc)_2$ in a reaction volume of 9 μL for 10 min. The reaction was initiated by the addition of the ^{32}P-labeled 40mer dsDNA substrate (to 12.8 μM) and $Mg(OAc)_2$ (to 10 mM) yielding a final reaction volume of 12 μL. The final pH after addition of all reaction components was 7.55. Reactions were stopped at the indicated times with the addition of EDTA and SDS to final concentrations of 20 mM and 1%, respectively, and a final volume of 16 μL. Stopped reactions were dialyzed into at least 100 mL of TE with a 1000 MWCO disposable dialyzers (The Nest Group) at 25 °C for 2 h. A gel loading buffer [2.5% ficoll (type 400), 0.08% bromophenol blue, 0.08% xylene cyanol FF, all final concentrations] was then added to each dialyzed reaction, and this was followed by electrophoresis on 20% acrylamide/1× TBE gels. Bands were visualized by exposure to phosporimager screens (Molecular Dynamics) for 30 min and scanned on a Molecular Dynamics PhosphorImager (model 425E). Band intensities were calculated using the ImageQuant software (Molecular Dynamics). The percent reaction was calculated as the combined signal from both product bands divided by the total labeled DNA signal for a given gel lane.

The kinetic analysis reaction conditions differed from the standard reaction conditions as follows. All reaction incubations were carried out at 30 °C. The ssDNA substrate was at 44.8 μM and the RecA protein concentration was at 14.9 μM. In all cases, each reaction time point represents a separate reaction incubated for the indicated time.

DNA pairing reactions with Rad51 protein were carried out in 35 mM K-MOPS buffer (pH 7.2) with added 2.4 mM $MgCl_2$, and 60 mM KCl at 38 °C. The 70mer ssDNA substrate (56 μM) was preincubated with Rad51 protein

70mer ssDNA substrate
5'- AGTAGACTCAGCGAACTCACTiGATMiCMiCAiGTMiCTTAiGMiCATMiCAiGT-
MiCAMiCiGATAMiCMiCTMiCiGAiGATAMiCATACGGACGTA-3'
40mer dsDNA substrate
3'-AMiCTAiGiGTMiCAiGAATMiCiGTAiGTMiCAiGTiGMiCTAT-
iGiGAiGMiCTMiCTATiGT-5'
5'-TiGATMiCMiCAiGTMiCTTAiGMiCATMiCAiGTMiCA-
MiCiGATAMiCMiCTMiCiGAiGATAMiCA-3'



1 2

FIGURE 3: Structures of the nonstandard phosphoramidites used to synthesize MiC and iG and bearing oligonucleotides.

(17 μM) and ATP (2 mM) in a reaction volume of 10 μL for 10 min. The reaction was initiated by the addition of the ^{32}P-labeled 39mer dsDNA substrate (12.8 μM) and 4 mM spermidine-HCl yielding a final reaction volume of 12.5 μL. Reactions were stopped, dialyzed, and analyzed as described for reactions with RecA protein.

RESULTS

Experimental Rationale. To test the iG and MiC nucleotides in substrates for RecA-mediated DNA strand exchange, we made use of a relatively simple reaction system. The length of the DNA substrates was limited by the need to create them via chemical synthesis. We designed a ssDNA, 70 nucleotides in length (Figure 3, oligomer A and A′ in Materials and Methods). This is short enough to be synthesized by chemical means, but long enough to allow efficient binding by RecA protein (*50*, *51*). A′ differs from A only in that iG and MiC has replaced G and C in the sequence. In its central section, this DNA is homologous to a synthetic dsDNA 40mer (made up of the oligos B and C, or B′ and C′ in the case of the iG and MiC-substituted versions). Strand exchange is carried out between the single-stranded 70mer and the 40mer duplex DNAs. The length difference between substrates also allows for differentiation of substrate and product duplexes after electrophoresis. Sequences were designed to minimize the opportunities for the formation of secondary structure in the individual oligonucleotides. Several versions of these DNAs were constructed, containing C/G, MiC/iG, or mixtures of the two, for purposes of experiment and control as described below.

Choice of iC and iG Monomers for Oligonucleotide Synthesis. Protected MiC phosphoramidite **1** (*10*, *12*, *52*) and iG phosphoramidite derivative **2** (*10*, *12*, *53*) were used in the preparation of modified oligonucleotides (Figure 3). 5-Methylisocytosine rather than isocytosine was used owing to the enhanced stability of the 5-methyl derivative to basic conditions (e.g., concentrated NH_4OH) encountered during oligonucleotide deprotection (*10*, *13*). Although the 5-methyl group is known to increase the sensitivity of isocytosine to acid depyrimidination (*10*), 2-*N*,*N*-dialkylformamidine protected MiC phosphoramidite derivatives are known to be sufficiently stable to the acidic detritylation conditions used during oligonucleotide synthesis to enable efficient coupling (*11*, *12*, *52*). The composition of all modified oligonucleotides was verified by digestion with snake venom phosphodiesterase in the presence of bacterial alkaline phosphatase followed by HPLC analysis of the resultant nucleosides (*10*). Implications of adding the 5-methyl group to iC are that the MiC:iG and C:G pairs are no longer isomeric and the major groove functional groups of the MiC:iG pair are identical with those found in the T:A pair.

Substituting iG/MiC for G/C Has Only Modest Effects on RecA-Mediated DNA Strand Exchange. RecA protein-mediated strand exchange was monitored for a set of DNA substrates with standard nucleotide bases, and compared to reactions with another set in which all of the G and C residues (45% of the total) have been replaced by iG and MiC residues (oligos A′, B′, and C′). Conditions for DNA strand exchange were optimized empirically for these short DNA substrates and are described in the Materials and Methods and in the Figure 4 legend. Sufficient RecA protein was added to saturate the ssDNA, and ATPγS was used as the nucleotide cofactor to minimize hydrolysis. RecA protein is allowed to bind to the ssDNA substrate in the presence of a magnesium concentration (1 mM), low enough to minimize internal secondary structure within the DNA and allow for easier filament formation (*40*, *54*). The higher magnesium concentration (10 mM), needed for optimal RecA function, is added with the dsDNA substrate upon initiation of the reaction.

Typical reactions are shown in Figure 4, panels b and c. Strand exchange reactions using oligonucleotide substrates containing only standard nucleotide bases (A and B + C) averaged 89.3 $\pm$ 3.9% (n = 3) product formation after 20 min. Similar DNA strand exchange reactions using the oligonucleotide substrates containing iG and MiC (A′ and B′ + C′) averaged 72.6 $\pm$ 1.9% (n = 3) product formation after 20 min. In separate control experiments with both normal and iG/MiC DNA substrates, strand exchange was completely dependent on both RecA protein and ATPγS, as it failed in the absence of either Rec A protein or ATPγS (data not shown).

Both reactions were relatively fast and appeared to be nearly completed within 5 min under these reaction conditions. Hence, it was possible that the iG/MiC-containing oligonucleotides affected the reaction kinetically, while permitting the reaction to proceed to a similar extent. To better examine and compare the kinetics of the two reactions, the reaction was slowed by decreasing the ssDNA substrate concentration to 44.8 μM and lowering the reaction temperature from 37 to 30 °C. To correct for small differences in extent from one experiment to the next, each reaction was normalized so that data is expressed as percent of maximum reaction (calculated as percent product formation at each time point divided by the percent product formation at 20 min). Both standard and substituted substrates reacted with very similar kinetics under these altered conditions (Figure 5). Both reactions occur rapidly in the first minute reaching approximately 80% of maximum, increasing thereafter to reach about 90% of maximum after 5 min. Under these altered conditions, the actual final extent of reaction was 85.8 $\pm$ 1.7 and 64.7 $\pm$ 4.3% product formation (n = 3) for the normal and altered DNA substrates, respectively.

A set of similar experiments was carried out using the yeast Rad51 protein in place of the *E. coli* RecA protein. The reaction conditions used were based on those reported

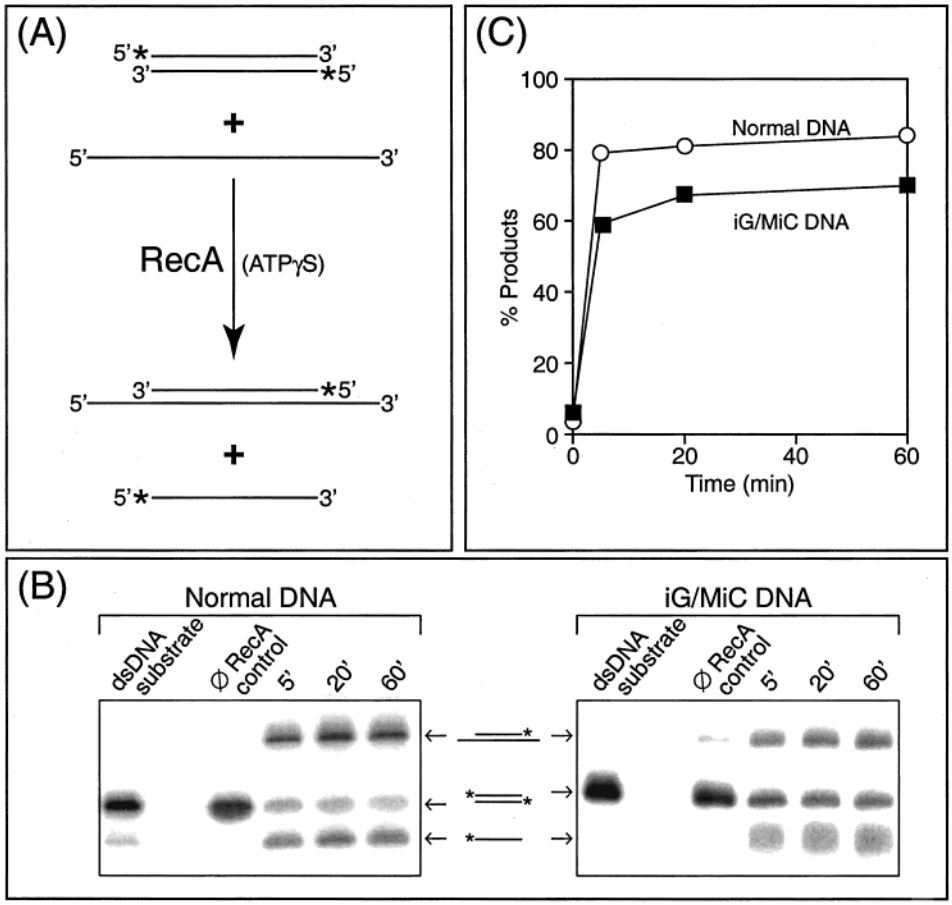


FIGURE 4: (A) The 70mer ssDNA substrate is preincubated with RecA protein and ATPγS. The reaction is then initiated with the dsDNA substrate (5′ end labeled on both strands with ^{32}P). Labeled reaction products are a hybrid dsDNA of a 70mer annealed to a 40mer and a displaced ssDNA 40mer. (B) Reactions using substrates containing both normal DNA bases and iG and MiC were nearly complete after 5 min with only small additional progress after 1 h. (C) Quantitation of the reaction progress from the gels shown in panel B was calculated as the combined signal from both product bands divided by total signal from all three bands multiplied by 100%.

in the literature, optimized for strand exchange of large (ϕX174) DNA substrates (*45*, *55*), but without the addition of the yeast replication protein A (RPA). With normal DNA substrates (using oligos A, B, and C) Rad51 protein promoted strand exchange to a final extent of 73.5%. When MiC/iG substituted substrates were used in an otherwise identical experiment, strand exchange proceeded to a final extent of 61.6% (data not shown). Rad51 pairing reactions were somewhat slower than their RecA counterparts, but the relative difference in total reaction efficiency was essentially the same. The kinetics of Rad51-mediated DNA strand exchange were the same for normal and MiC/iG-substituted substrates.

Oligonucleotides Containing Multiple iG and MiC Form Stable Watson−Crick Duplexes. Since our present strand exchange studies require multiple MiC:iG pairings (18 base pairs out of the 40-mer duplex used in the exchange reaction), we felt it was important to confirm that the occurrence of multiple MiC:iG pairs would not be deleterious to duplex stability. Several studies of (M)iG:iC pairing in duplexes have been reported (*10*−*12*), including in two cases multiple such pairs (*11*, *12*). The issue is especially important due to iG's ability to assume different tautomeric forms. Initially thermal denaturation of the 40-mer duplexes B/C and B′/C′ were investigated, but they gave duplexes that were too stable to yield complete denaturation profiles. We next turned to duplexes based on 10-mer sequences derived from bases 12−21 in oligomers B and B′ and their complements in oligomers C and C′. It may be seen from Figure 6 that the 10-mer duplex bearing four MiC:iG base pairs denatures cooperatively and has a free energy of duplex formation 2.63 kcal/mol greater than the all natural duplex (Table 1). Some measure of this enhanced stablility is expected to derive from the 5-methyl group present on MiC but absent from cytosine (*11*, *56*). On the basis of these data we presume that the sum of the pairing interactions that occur in the MiC/iG containing 40-mer duplexes B′/C′ are similarly favorable.

Modeling: MiC and iG Can Form Triples Consistent with R-Form Triplexes. Modeling studies were undertaken to assess the viability of an R-form triplex intermediate involving MiC and iG. Figure 7A shows putative R-form DNA triples for the nonstandard bases. The success of the iG:iG:MiC triple as represented in this figure relies on the ability of iG to adopt enol and keto tautomers within the third and second strand bases, respectively, and switch between them. Although this might at first appear an unusual arrangement, two tautomeric forms of iG have been observed in a duplex in the crystalline state and in solution at approximately the same temperature at which the exchange experiments are performed (37 °C) (*18*).

Nonstandard triples were evaluated by optimizing their geometries using semiempirical and ab initio methods, and then comparing the results with the geometries of the standard triples determined by the same methods. The only

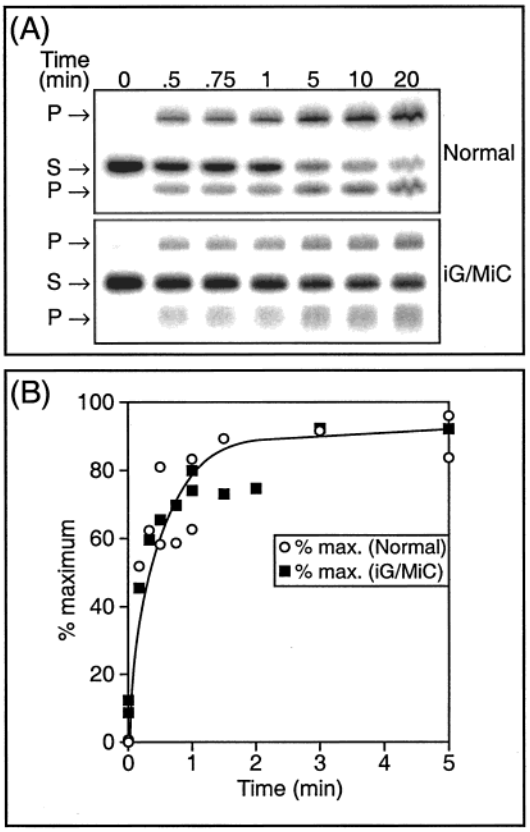
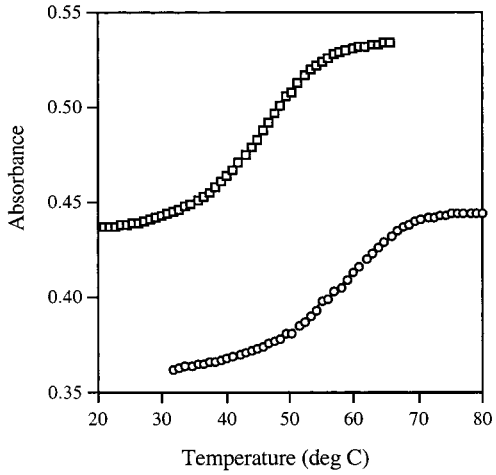


FIGURE 5: (A) Slowed reactions allow for kinetic analysis of the early stages of RecA protein-facilitated strand exchange. (B) Compiled normalized data from several experiments reveals that reactions with both normal and iG/MiC-substituted substrates occur rapidly in the first minute at which point the reaction slows.



❒ = 5′-TAGCATCAGT/3′-ATCGTAGTCA
❍ = 5′-TAiGMiCATMiCAiGT/3′-ATMiCiGTAiGTMiCA

FIGURE 6: Thermal denaturation profiles for two 10-mer duplexes bearing either four iG:MiC base pairs [(○) 5′-TAiGMiCATMiC-AiGT/3′-ATMiCiGTAiGTMiCA] or four G:C base pairs at corresponding positions [(□) 5′-TAGCATCAGT/3′-ATCGTAGTCA].

restriction placed on the bases during the calculations was for all atoms to be coplanar. In the earlier work of Zhurkin et al. (*27*), triple computations were performed by molecular mechanics methods using the same coplanar restriction, and a further restriction that the Watson−Crick pair remain fixed in an Arnott B-DNA geometry. In the present work, no restriction was imposed on the Watson−Crick pair of the triples. Similar to earlier findings (*27*), we observe the carbon-1′ atoms of all third strand bases of all standard R-form DNA triples to fall within a circle of 1 Å diameter, with the center of this circle defined as the center of mass of these four atoms (Figure 7, panels B and C, left). With this correlation in methodology established for the standard bases, we then sought to evaluate the nonstandard triples.

PM3 and Hartree−Fock/6-31G* optimized geometries for MiC:MiC:iG and iG:iG:MiC are shown in Figure 7, panels B and C (at center and right), and the center-of-mass distances for their respective carbon-1′ atoms are depicted in the same figure (at left). Geometrical results for nonstandard triples from the two methods generally conform to the R-form triple model constraints proposed by Zhurkin et al. (*27*), namely that third strand carbon-1′ atoms fall inside (or provide favorable interaction energies when constrained to be within) a 1.5 Å diameter circle defined by arbitrarily expanding the 1 Å diameter circle found for the carbon-1′ atoms for the four standard triples by 50%. The spirit of this geometric test is that the DNA backbone has finite flexibility to accommodate deviations in strand position while at the same time preserving noncovalent interactions in successive triples within DNA strands. Center of mass distances for two nonstandard triples fall slightly outside the 1.5 Å radius—PM3 optimized MiC:MiC:iG (1.59 Å) and HF/6-31G* optimized iG:iG:MiC (1.51 Å). Given the origin of the carbon-1′ test radius, violations of 0.09 and 0.01 Å, respectively, for the latter triples cannot be considered fatal.

Geometric information for PM3 optimized R-form triples, including key mismatches, are summarized in Table 2. As anticipated from the degeneracy of MiC:iG and T:A major groove functional groups, the third strand carbon-1′ geometric information in this table predicts T will compete with MiC

Table 1: Duplex Stabilities. Duplex Parameters Determined from the Denaturation Profiles Shown in Figure 6

(10-mer)
d-5′-TAXYATYAXT
d-3′-ATYXTAXTYA

entry	X	Y	T_m (°C)	$-\Delta G^\circ_{37}$ (kcal/mol)	$-\Delta\Delta G^\circ_{37}$ (kcal/mol)	$-\Delta G^\circ_{37}$[a] (kcal/mol)
1	G	C	45.3	9.77	0	9.84
2	iG	iC	59.2	12.4	2.63	

[a] Calculated from nearest-neighbor free energy parameters using the unified model of Santa Lucia [Santa Lucia, J. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 1460−1465].

Table 2: Distances of the Third Strand Carbon-1′ from the Center of Mass of the Four Third-Strand Carbon-1′s in the Four Natural R-form Triples AAT, TTA, GGC, and CCG[a]

WC strands	third strand					
	MiC	iG	A	C	G	T
MiCiG	**1.59**		2.63	**1.42**	2.77	**0.52**
iGMiC		**1.13**	**0.72**			
AT	2.79		**0.85**			
CG	4.11			**0.83** (1.27)[b]		
GC	1.83				0.56	
TA	1.47					**0.45**

[a] Values are derived from PM3 optimized geometries of Cs symmetric structures [b] Value derives from alternative bifurcated CCG geometry.

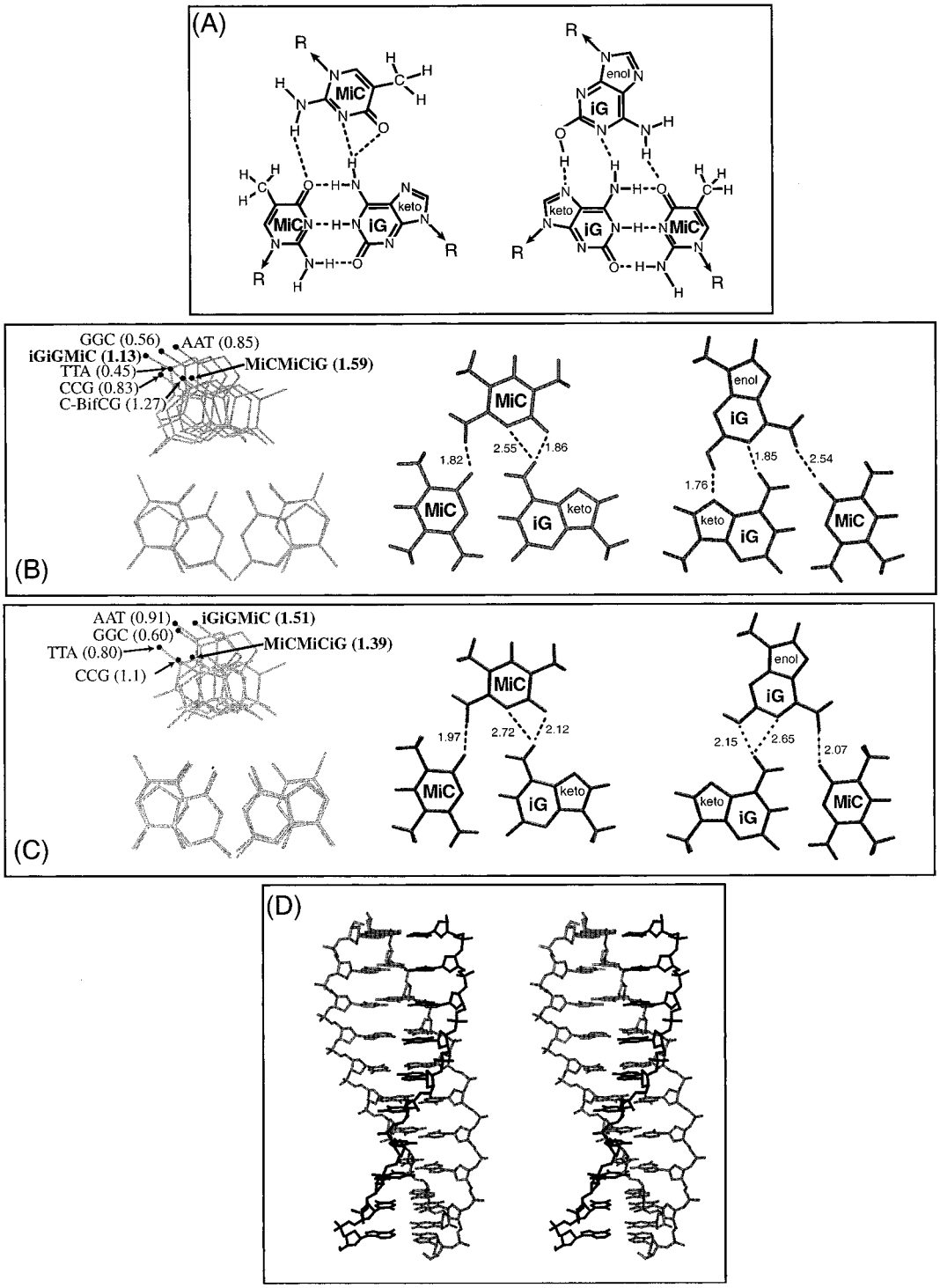


FIGURE 7: (A) Base triples incorporating MiC and iG that conform to the proposed R-form triplex intermediate. (B) Left: superposition of PM3 optimized geometries of R-form triples. The distance in angstroms from the carbon-1′ of each third strand base to the center of mass of the carbon-1′ atoms of the four standard triples is highlighted for each case. C-BifCG refers to an alternative minimum of CCG incorporating apparently bifurcated hydrogen bonds that was also found by Zhurkin et al. (*27*). Center: the PM3 optimized geometry of MiCMiC:iG. The distance traversed by the dashed line is given in angstroms. Right: the PM3 optimized geometry of iGiGMiC. (C) Left: superposition of HF/6-31G* optimized geometries of R-form triples. The HF/6-31G* optimization only found the apparently bifurcated geometry of CCG irrespective of which of the two PM3 optimized geometries was used as the starting geometry. Center: the HF/6-31G* optimized geometry of MiCMiC:iG. Right: the HF/6-31G* optimized geometry of iGiGMiC. (D) R-form triplex model of sequence 5′-dATGCMiCATGCMiC with 5.1 Å rise and 20° twist. The third strand is in bold.

for MiC:iG (0.52 vs 1.59 Å) and A will compete with iG for iG:MiC (0.72 vs 1.13 Å). All of aforementioned competing geometries involve at least two hydrogen-bonding interactions between the second and third strands. Although on the basis of geometry alone Table 2 also predicts C will compete with MiC for MiC:iG, in practice this is unlikely since the C-MiC:iG interaction is mediated by a single hydrogen bond. While all important MiC third strand geometries were determined (Table 2, row 1 and column 1), a similar array of computations was not performed in

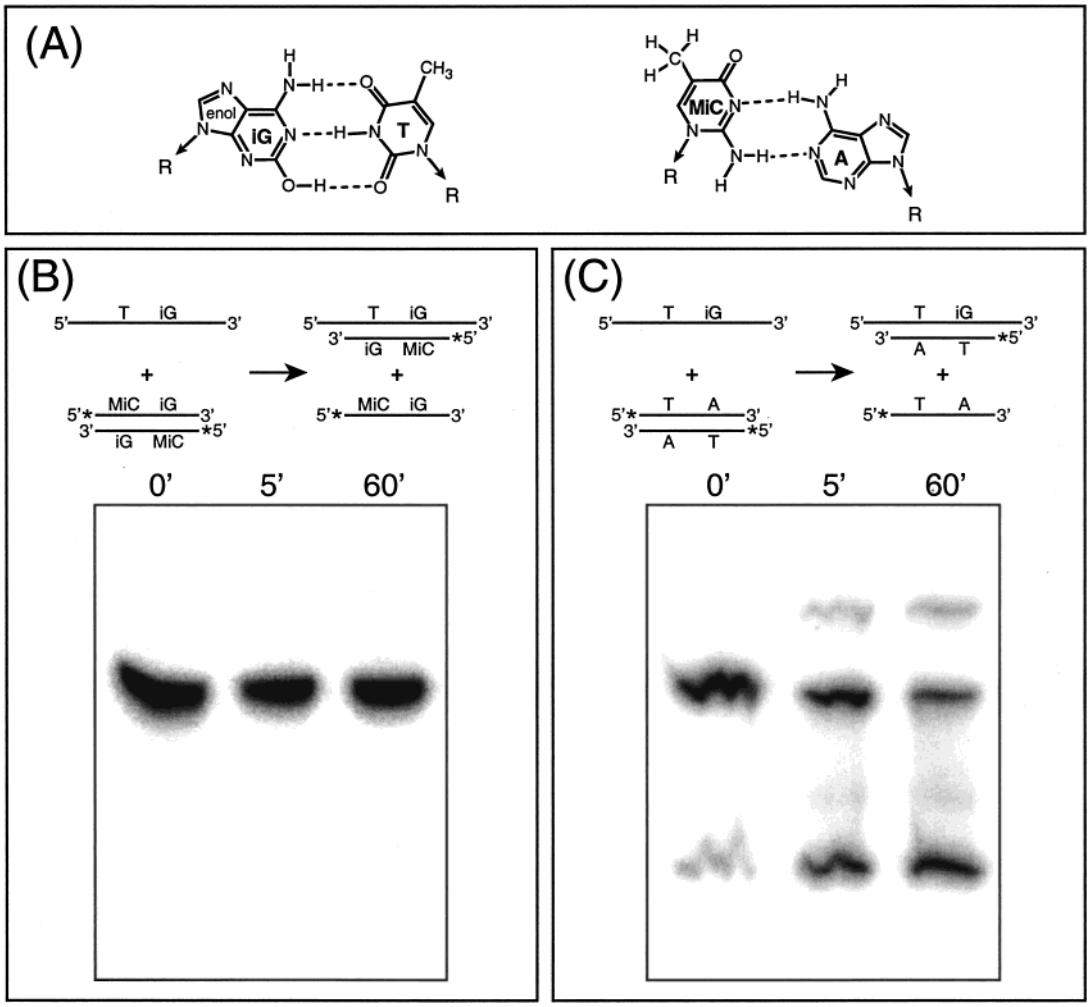


FIGURE 8: (A) These other potential base pairs involving iG or MiC were explored in pairing reactions to examine the fidelity of iG and MiC in homologous recombination. (B) RecA protein-facilitated DNA pairing that would result in a T:iG mismatch resulted in no detectable product formation. (C) When the substrate duplex was entirely A:T pairs, RecA was able to pair DNA substrates that result in products containing iG:T pairs. This particular reaction has a high background because of the exclusive A:T content of the substrate duplex. The ssDNA signal seen in the zero time point (in the absence of RecA protein) remains even after gel purification of the duplex.

the iG series due to the indeterminate nature of its tautomerism. Nevertheless, we confirmed, in keeping with the fact that the iG:MiC and A:T Watson−Crick pairs display identical major groove functionality, that the third strands of A:iG:MiC and AAT triples have very similar geometries (center of mass distances 0.72 and 0.85 Å, respectively).

As a further test of consistency, the MiC:MiC:iG triple in its PM3 optimized geometry was inserted along with optimized forms of the four natural triples into a triplex defined by a 5.1 Å rise and 20° twist consistent with electron microscopic structural data on RecA/DNA complexes (*57*, *58*) (Figure 7D). To arrive at the final structure, atoms in the PM3 optimized triples were restrained while the backbones of the DNA strands were left unrestrained, and the system optimized using molecular mechanics. No aberrations were evident in the resulting R-form triplex structure.

Fidelity of DNA Strand Exchange Is Decreased when DNAs Contain MiC/iG Nucleotides. The hydrogen-bonding groups presented in the major groove of an iG:MiC base pair are identical to those presented in an A:T base pair. In principle, any triple formed by iG with iG:MiC could be formed between iG and A:T, and any triple formed by T with A:T could be formed between T and MiC:iG. In addition, thermodynamic and structural data have shown that iG will stably pair with T (*10*, *18*). The solution structure of the iG:T pair has been shown to exist as an equilibrium between two pairing geometries, where the predominant low temperature form is wobble (∼95% at 0 °C) and the second form, presumed to be Watson−Crick, grows in at higher temperatures (∼40% at 40 °C) (Figure 8A). While no structural data is available for the iC:A pair, thermodynamic data supports pairing of these bases via a wobble geometry (Figure 8A) (*10*, *59*). To explore the effect of MiC and iG on the fidelity of DNA strand exchange, we prepared DNA substrates in which each iG in the duplex was paired with T in the ssDNA, and vice versa. A new 70mer ssDNA (A'_{fT}) was constructed for this purpose, with all the iC residues replaced with T. This ssDNA did not undergo exchange with the MiC/iG-substituted 40mer duplex ($B' + C'$, Figure 8B). However, an exchange was observed with a new 40mer duplex ($B_f + C_f$), consisting of entirely A:T base pairs, in which each iG in the ssDNA would become paired with a T in the product (Figure 8C). SSB is often included in RecA protein-mediated DNA strand exchange reactions, and plays a role in binding the displaced DNA strand. We had not included it in our experiments, since our trials with normal

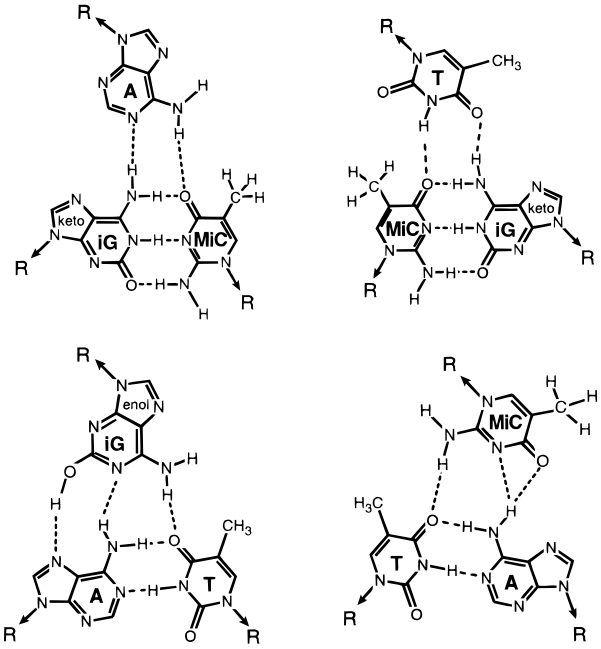


FIGURE 9: These theoretical base triples resemble those drawn in association with the R-DNA triplex and those modeled in this study. Even with these additional interactions possible in the major groove, which would result in a loss of fidelity according to the R-DNA recognition model, RecA promotes DNA pairing efficiently with substrates containing MiC and iG.

DNA substrates had indicated it had no effect on the reaction under our reaction conditions. We added SSB to the experiments in Figure 8B to determine if it might enhance exchange reactions involving mismatched bases, but did not observe any change in the results (data not shown).

In the former reaction, the T:iG pair replaces the very stable MiC:iG pairs, and thus the product lacks the stability of the substrate. We do not know to what extent this might affect the results. Radding and co-workers have demonstrated that pairing efficiencies with Rad51 protein are directly related to differential stability between the substrate and product duplexes (*40*). The observation of exchange in the latter experiment shows that the iG:T pair participates in a product of RecA-mediated DNA strand exchange. In this case, the substrate duplex consists entirely of A:T base pairs, and the iG:T pairing may provide comparable stability in the product DNA. Both of these reactions should allow for the formation of triples compatible with R-form DNA (Figure 9), and thus the capacity to form such triples should not be a factor in the different results obtained.

We also carried out reactions in which the ssDNA 70mer had all of its iG residues replaced with A′s (A'_{fA}), with the potential to generate either A:MiC or MiC:A pairings in the products. These reactions did not work (data not shown). However, control experiments in which the strands of the potential products were directly annealed to each other and subjected to electrophoresis under conditions used to monitor DNA strand exchange demonstrated that the products were inherently unstable. Very little duplex DNA was formed in these experiments. When equal amounts of A'_{fA} and the potentially complementary 40mer (C′) were annealed, less than 20% of the resulting DNA was successfully paired. When equal amounts of the substrate duplex (B′ + C′) and A'_{fA} were heated to 90 °C then allowed to cool, no hybrid duplex (A'_{fA} + C′) was detected after electrophoresis.

DISCUSSION

Our primary conclusion is both the *E. coli* RecA protein and the yeast Rad51 protein promote an efficient DNA strand exchange reaction with DNA substrates containing MiC and iG nucleotides in place of the usual C and G residues. Even when nearly half of the nucleotides to be exchanged were altered, the effects on the rate and efficiency of DNA strand exchange were minor. The results expand the list of enzymes and reaction types for which MiC and iG can be utilized as substrates. The results also clearly apply to the major DNA strand exchange activities of both bacteria and eukaryotes.

General Comments. The MiC and iG tautomers generally listed as most prevalent at first appear to display hydrogen-bonding functional groups in the major groove that are incompatible with the R-form triplex (*9*). However, the modeling carried out here, using in particular the important enol tautomer of iG, provides for triples that generally conform to the parameters established for the R-form in previous modeling exercises (*27*). From that perspective, the present results do not challenge the model for RecA or Rad51-mediated DNA pairing that relies on the R-form for homologous alignment. However, we note that the fidelity inherent in the R-form triples when the standard nucleotides are utilized is compromised when the MiC and iG substitutions are made. An iG in the ssDNA should align as readily with an A:T base pair as with iG:MiC, and this substitution actually permits a substantial level of DNA strand exchange. Similarly, the possible MiC:T:A, T:MiC:iG, and A:iG:MiC triples are essentially identical to the modeled MiC:MiC:iG, T:T:A, and A:A:T triples, respectively, although stability of the ultimate product may be compromised in some of these pairings (Figure 9). Thus, in the experiments of Figures 4 and 5, each nucleotide in the ssDNA has at least two potential types of pairing partners if the R-form triplex plays an important role in homologous alignment. This is in contrast to the unique pairing readout provided by the standard nucleotides in the putative R-form triples (*27*). The potential ambiguity in an alignment step via the hypothetical triplex clearly has very little effect on kinetics or the final product yield in DNA strand exchange reactions promoted by the RecA and Rad51 proteins. In addition, there is no observed kinetic effect of the tautomerization from the enol to N1−H form of iG by the expanded R-form triplex model. In previous work, the rate of exchange between nonstandard nucleobase tautomers in duplexes in the absence of protein has been reported to be slow on the NMR time scale (*18*, *60*−*62*), but has been observed to be fast in the gas phase where only bases were involved (*63*). If one posits a 500 Hz difference in chemical shift between tautomeric protons in an NMR experiment, it follows that the upper bound for the slow exchange regime is defined by $k_{exchange} < \sim 1 \times 10^3\ s^{-1}$ [$k_{exchange} < \pi\Delta\nu/\sqrt{2}$ (*64*)]. The latter rate constant for tautomerization exceeds the rate constant reported for the displacement of an 82-mer DNA strand [$k_{obs} = 0.06\ s^{-1}$ (*65*)] by 4 orders of magnitude, and is not inconsistent with the data of Figure 5B. In the absence of a more refined value for k_{taut}, we cannot ascertain whether this physical step is incompatible with an R-form intermediate. Nonetheless, we acknowledge that it could well be the case.

Nucleotide substitutions that might affect the formation of an R-form triplex were examined for RecA-mediated DNA strand exchange in only one previous study, in which G residues were replaced by 7-deazaguanine (*66*). This substitution had no effect on the efficiency of either 3 or 4 strand exchange reactions, and subsequent work showed that DNA substrates containing both 7-deazaadenine and 7-deazaguanine worked as well or better than normal DNA as strand exchange substrates (*67*). These substitutions disrupt only a subset of the interactions upon which the R-form structure is based and thus do not rule out such a structure. However, we note that both these and our current results are at least as consistent with recent reports favoring an alignment of the RecA-bound ssDNA in the minor groove of the duplex (*34*−*38*) via a mechanism involving base flipping (*39*, *41*).

Compatibility of iG and (M)iC with Genetic Systems That Are Competent at Both Recombination and Replication. Recombination mechanisms have the potential of imposing serious limitations on the fitness of nucleobases beyond those presently found in terrestrial genomes. Selective advantages for organisms that utilize recombination competent genetic materials include an additional means for DNA repair and the capacity for sexual, as opposed to asexual, reproduction. If the R-form triplex should prove to be a ubiquitous intermediate in homologous recombination, then compatibility with this structure becomes pivotal for nucleobases to function within extant biochemistries. Beyond this, to the extent that the R-form triplex intermediate is required for recombination irrespective of how a particular biochemistry might have evolved, compatibility with it defines those nucleobases that support sexual reproduction and those that do not.

Modeling studies reported here indicate MiC and iG are most compatible with the R-form triplex when iG adopts the enol and N1−H tautomeric forms in the third and second strands, respectively. This circumstance results in a paradox: the enol form of iG required for compatibility with the R-form triplex virtually guarantees infidelity during replication via Watson−Crick pairing in any genetic system that includes A:T pairs. In addition, the identical major groove functionality shared by iC:iG and T:A base pairs precludes high fidelity homologous recombination via an R-form triplex intermediate in a genetic system that includes both of these pairs. Nevertheless, a genetic code based on MiC:iG and C:G pairs may be viable (*10*) since in this system the enol form of iG should only assist recombination and not detract from replication fidelity.

If recombination can be achieved by a mechanism relying on Watson−Crick base pairing, then there is no longer a need for the N1−H and enol forms of iG to coexist. Within this mechanistic scenario, it would appear formally possible to evolve a biochemistry in which the iG:MiC pair can recombine and replicate in conjunction with the full set of genomic nucleobases provided the iG enol form could be suppressed. This might be accomplished by controlling the microenvironment surrounding the base, or selecting an alternative iG structure that at once achieves the desired tautomer suppression and fulfills the constraint of prebiotic plausibility.

REFERENCES

1. Leach, A. R., and Kollman, P. A. (1992) *J. Am. Chem. Soc. 114*, 3675−3683.
2. Switzer, C., Moroney, S. E., and Benner, S. A. (1989) *J. Am. Chem. Soc. 111*, 8322−8323.
3. Piccirilli, J. A., Krauch, T., Moroney, S. E., and Benner, S. A. (1990) *Nature 343*, 33−37.
4. Eschenmoser, A., and Loewenthal, B. (1992) *Chem. Soc. Rev. 21*, 1−16.
5. Robertson, M. P., Levy, M., and Miller, S. L. (1996) *J. Mol. Evol. 43*, 543−550.
6. Rich, A. (1962) in *Horizons in Biochemistry* (Kasha, M., and Pullman, B., Eds.) pp 103−126, Academic Press, New York.
7. Switzer, C. Y., Moroney, S. E., and Benner, S. A. (1993) *Biochemistry 32*, 10489−10496.
8. Horlacher, J., Hottiger, M., Podust, V. N., Hubscher, U., and Benner, S. A. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 6329−6333.
9. Cox, M. M. (1997) *Mutat. Res. 384*, 15−22.
10. Roberts, C., Bandaru, R., and Switzer, C. (1997) *J. Am. Chem. Soc. 119*, 4640−4649.
11. Horn, T., Chang, C. A., and Collins, M. L. (1995) *Tetrahedron Lett. 36*, 2033−2036.
12. Seela, F., He, Y., and Wei, C. (1999) *Tetrahedron* 55, 9481−9500.
13. Tor, Y., and Dervan, P. B. (1993) *J. Am. Chem. Soc. 115*, 4461−4467.
14. Lutz, M. J., Horlacher, J., and Benner, S. A. (1998) *Bioorg. Med. Chem. Lett. 8*, 499−504.
15. Bain, J. D., Chamberlin, A. R., Switzer, C., and Benner, S. A. (1992) *Nature 356*, 537−539.
16. Sepiol, J., Kazimierczuk, Z., and Shugar, D. (1976) *Z. Naturforsch. 31*, 361−370.
17. Seela, F., Wei, C., and Kazimierczuk, Z. (1995) *Helv. Chim. Acta 78*, 1843−1854.
18. Robinson, H., Gao, Y. G., Bauer, C., Roberts, C., Switzer, C., and Wang, A. H. J. (1998) *Biochemistry 37*, 10897−10905.
19. Pitsch, S., Krishnamurthy, R., Bolli, M., Wendeborn, S., Holzener, A., Minton, M., Lesueur, C., Schlonvogt, I., Jaun, B., and Eschenmoser, A. (1995) *Helv. Chim. Acta 78*, 1621−35.
20. Cox, M. M. (1999) *Progr. Nucleic Acids Res. Mol. Biol. 63*, 310−366.
21. Cox, M. M., Goodman, M. F., Kreuzer, K. N., Sherratt, D. J., Sandler, S. J., and Marians, K. J., (2000) *Nature 404*, 37−41.
22. Bedale, W. A., and Cox, M. (1996) *J. Biol. Chem. 271*, 5725−5732.
23. Lacks, S. (1966) *Genetics 53*, 207−235.
24. Umlauf, S. (1990) *Molecular Biology*, University of Wisconsin−Madison.
25. Howard-Flanders, P., West, S. C., and Stasiak, A. (1984) *Nature (London) 309*, 215−219.
26. Stasiak, A. (1992) *Mol. Microbiol. 6*, 3267−3276.
27. Zhurkin, V. B., Raghunathan, G., Ulyanov, N. B., Camerini, O. R., and Jernigan, R. L. (1994) *J. Mol. Biol. 239*, 181−200.
28. Kim, M. G., Zhurkin, V. B., Jernigan, R. L., and Camerini-Otero, R. D. (1995) *J. Mol. Biol. 247*, 874−889.
29. Kiran, M. R., and Bansal, M. (1997) *J. Biomol. Struct. Dyn. 15*, 333−345.
30. Hsieh, P., Camerini, O. C., and Camerini, O. R. (1990) *Genes Dev. 4*, 1951−1963.
31. Rao, B. J., Chiu, S. K., and Radding, C. M. (1993) *J. Mol. Biol. 229*, 328−343.
32. Rao, B. J., Dutreix, M., and Radding, C. M. (1991) *Proc. Natl. Acad. Sci. U.S.A. 88*, 2984−2988.
33. Chiu, S. K., Rao, B. J., Story, R. M., and Radding, C. M. (1993) *Biochemistry 32*, 13146−13155.
34. Baliga, R., Singleton, J. W., and Dervan, P. B. (1995) *Proc. Natl. Acad. Sci. U.S.A. 92*, 10393−10397.
35. Podyminogin, M. A., Meyer, R. B., and Gamper, H. B. (1995) *Biochemistry 34*, 13098−13108.
36. Podyminogin, M. A., Meyer, R. B., and Gamper, H. B. (1996) *Biochemistry 35*, 7267−7274.

37. Kumar, K. A., and Muniyappa, K. (1992) *J. Biol. Chem. 267*, 24824−24832.
38. Zhou, X., and Adzuma, K. (1997) *Biochemistry 36*, 4650−4661.
39. Nishinaka, T., Shinohara, A., Ito, Y., Yokoyama, S., and Shibata, T. (1998) *Proc. Natl. Acad. Sci. U.S.A. 95*, 11071−11076.
40. Gupta, R. C., Folta-Stogniew, E., and Radding, C. M. (1999) *J. Biol. Chem.. 274*, 1248−1256.
41. Gupta, R. C., Folta-Stogniew, E., O'Malley, S., Takahashi, M., and Radding, C. M. (1999) *Mol. Cell 4*, 705−714.
42. Shan, Q., Cox, M. M., and Inman, R. B. (1996) *J. Biol. Chem. 271*, 5712−5724.
43. Shan, Q., and Cox, M. M. (1998) *Mol. Cell 1*, 309−317.
44. Craig, N. L., and Roberts, J. W. (1981) *J. Biol. Chem. 256*, 8039−44.
45. Sung, P., and Stratton, S. A. (1996) *J. Biol. Chem. 271*, 27983−6.
46. Sugiyama, T., Zaitseva, E. M., and Kowalczykowski, S. C. (1997) *J. Biol. Chem. 272*, 7940−5.
47. Watanabe, K. A.; Reichman, U.; Chu, C. K.; Fox, J. J. (1978) in *Nucleic Acid Chemistry* (Tipson, R. S., and Townsend, C. B., Eds.) Part 1, pp 273−277, John Wiley and Sons, New York.
48. Longfellow, C. E., Kierzek, R., and Turner, D. H. (1990) *Biochemistry 29*, 278.
49. M. J. Frisch et al. (1998) Gaussian 98, Revision A.7, Gaussian, Inc., Pittsburgh, PA.
50. Brenner, S. L., Mitchell, R. S., Morrical, S. W., Neuendorf, S. K., Schutte, B. C., and Cox, M. M. (1987) *J. Biol. Chem. 262*, 4011−4016.
51. Bianco, P. R., and Weinstock, G. M. (1996) *Nucleic Acids Res. 24*, 4933−4939.
52. Jurczyk, S. C., Kodra, J. T., Rozzell, J. D., Benner, S. A., and Battersby, T. R. (1998) *Helv. Chim. Acta 81*, 793−811.
53. Sugiyama, H., Ikeda, S., and Saito, I. (1996) *J. Am. Chem. Soc. 118*, 9994−9995.
54. Mazin, A. V., and Kowalczykowski, S. C. (1999) *Genes Dev. 13*, 2005−2016.
55. Sung, P. (1994) *Science 265*, 1241−1243.
56. Wang, S., and Kool, E. T. (1995) *Biochemistry 34*, 4125−4132.
57. Stasiak, A., DiCapua, E., and Koller, T. (1981) *J. Mol. Biol. 151*, 557−564.
58. Stasiak A., and DiCapua, E. (1982) *Nature 299*, 185−186.
59. Strobel, S. A., Cech, T. R., Usman, N., and Beigelman, L. (1994) *Biochemistry 33*, 13824−13835.
60. Fazakerley, G. V., Gdaniec, Z., and Sowers, L. (1993) *J. Mol. Biol. 230*, 6−10.
61. Gdaniec, Z., Ban, B., Sowers, L., and Fazakerley, G. V. (1996) *Eur. J. Biochem. 242*, 271−279.
62. Nedderman, A., Stone, M. J., Williams, D. H., Thoo, K., Lin, P., and Brown, D. M. (1993) *J. Mol. Biol. 230*, 1068−1076.
63. Douhal, A., Kim, S. K., and Zewail, A. H. (1995) *Nature 378*, 260−263.
64. Sanders, J. K. M., and Hunter, B. K. (1989) *Modern NMR Spectroscopy*, p 210, Oxford University Press.
65. Bazemore, L. R., Takahashi, M., and Radding, C. M. (1997) *J. Biol. Chem. 272*, 14672−14682.
66. Jain, S. K., Inman, R. B., and Cox, M. M. (1992) *J. Biol. Chem. 267*, 4215−4222.
67. Jain, S. K. (1994) *Department of Biochemistry*, pp 171, University of Wisconsin−Madison, Madison, WI.

BI0003339