Proc. Natl. Acad. Sci. USA
Vol. 91, pp. 6161–6165, June 1994
Biochemistry

Formation of base triplets by non-Watson–Crick bonds mediates homologous recognition in RecA recombination filaments

B. J. RAO AND C. M. RADDING

Departments of Genetics and Molecular Biophysics and Biochemistry, Yale University School of Medicine, New Haven, CT 06510

Communicated by Aaron B. Lerner, March 18, 1994

ABSTRACT **Whereas complementary strands of DNA recognize one another by forming Watson–Crick base pairs, the way in which RecA protein enables a single strand to recognize homology in duplex DNA has remained unknown. Recent experiments, however, have shown that a single plus strand in the RecA filament can recognize an identical plus strand via bonds that, by definition, are non-Watson–Crick. In experiments reported here, base substitutions had the same qualitative and quantitative effects on the pairing of two identical strands in the RecA filament as on the recognition of duplex DNA by a third strand, indicating that similar non-Watson–Crick interactions govern both reactions.**

Escherichia coli RecA protein is the prototype of a class of recombination proteins that appear to be universally distributed among prokaryotes and eukaryotes (1–4). RecA protein forms a helical nucleoprotein filament on single-stranded DNA that mediates homologous recognition of duplex DNA and subsequent strand exchange (5, 6). How DNA molecules recognize one another is one of the central riddles of homologous recombination. According to a base-pair model, an interacting duplex melts locally to form a new set of Watson–Crick bonds with a third strand. According to a base-triplet model, a single strand in the RecA filament recognizes homology within duplex DNA without breaking Watson–Crick hydrogen bonds.

We observed (7) that a RecA filament formed on a single Watson or plus strand of DNA forms a homology-dependent complex not only with a complementary minus strand but also with an identical plus strand. This so-called two-strand reaction reveals that homologous recognition governed by RecA protein can occur via non-Watson–Crick bonds. We found further that a RecA filament formed on a plus strand to which a short minus strand had been annealed could also recognize either a plus or a minus strand, which we term a three-strand reaction.

On several counts, self-recognition, the homology-dependent binding of an identical plus strand by the plus strand within the RecA filament, has no exact precedents. The nonenzymic formation of parallel-strand DNA and triplex DNA *in vitro* are limited to special sequences, whereas, consistent with its function in homologous recombination, the RecA reaction appears to work with any two homologous sequences (8–12). In addition, the structure of DNA within the RecA filament is different from that of free DNA, and indeed the complex of two identical parallel strands appears to be unstable when RecA protein is removed (7).

To decipher the determinants of non-Watson–Crick recognition, we systematically substituted bases and assessed the effects on both the two-strand reaction, a model for self-recognition, and the three-strand reaction, a model for homologous pairing in recombination.

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

EXPERIMENTAL PROCEDURES

Materials. RecA protein was purified as described (13). Polynucleotide kinase was from New England Biolabs. Nuclease P1 was from United States Biochemical.

DNA Substrates. Single-stranded oligonucleotides were synthesized, end-labeled, and purified as described (14). Standard reaction conditions and assays are as described in figure legends.

RESULTS

Effects of Base Substitutions on the Formation of Two-Strand and Three-Strand Complexes. We have shown (7) that in the three-strand reaction, when the third strand was a plus strand, a new duplex molecule was formed by stereospecific strand exchange. Formally, this is the reverse of the usual strand exchange involving naked duplex DNA and a single strand in the RecA filament (15, 16). However, this reverse reaction appears particularly suited to the study of the initial recognition steps because although strand exchange ensues, it is about 1000 times slower than in the forward reaction (data not shown). Furthermore, in all the three-strand reactions that are described below, we paired 33-mer oligonucleotides with a region that was located in the middle of a 43-mer tract of duplex DNA in the RecA filament, which provided 5 extra base pairs of duplex DNA on either side of the pairing region; the shorter length of the 33-mer prevents it from displacing its homolog from the filament (data not shown). In addition, probing with nucleases and diethylpyrocarbonate revealed a specific footprint of protection corresponding to the duplex portion of the RecA filament; and competition with unlabeled oligonucleotides showed that the complementary oligonucleotide annealed to the 83-mer within the RecA filament remained stably paired under the conditions of the following experiments (data not shown).

We paired 5′-^{32}P-labeled 33-mer oligonucleotide (plus strand) with RecA nucleoprotein filament containing either single-stranded or double-stranded DNA and analyzed the samples by a gel retardation assay. When a labeled oligonucleotide binds to the RecA nucleoprotein filament, it migrates more slowly upon gel electrophoresis (7). Homologous 33-mer oligonucleotide was efficiently recognized. The signal was specific as it was markedly reduced by competition with a 10-fold excess of the same unlabeled sequence but not with a heterologous one (Fig. 1, lanes 5 and 6). In addition, another heterologous labeled 33-mer oligonucleotide yielded a background signal of only 10% (Fig. 1, lanes 1–3). Two-strand vs. three-strand homologous reactions showed identical pairing efficiencies. In both reactions, recognition was dependent on RecA protein and adenosine 5′[γ-thio]triphosphate and was sensitive to deproteinization (Fig. 1, lanes 43–45).

To determine the rules that govern self-recognition, we systematically substituted all four bases in the homologous 33-mer oligonucleotide. The homologous wild-type 33-mer oligonucleotide had all four bases roughly in equal number (nine A, nine G, seven C, and eight T residues). Mutant

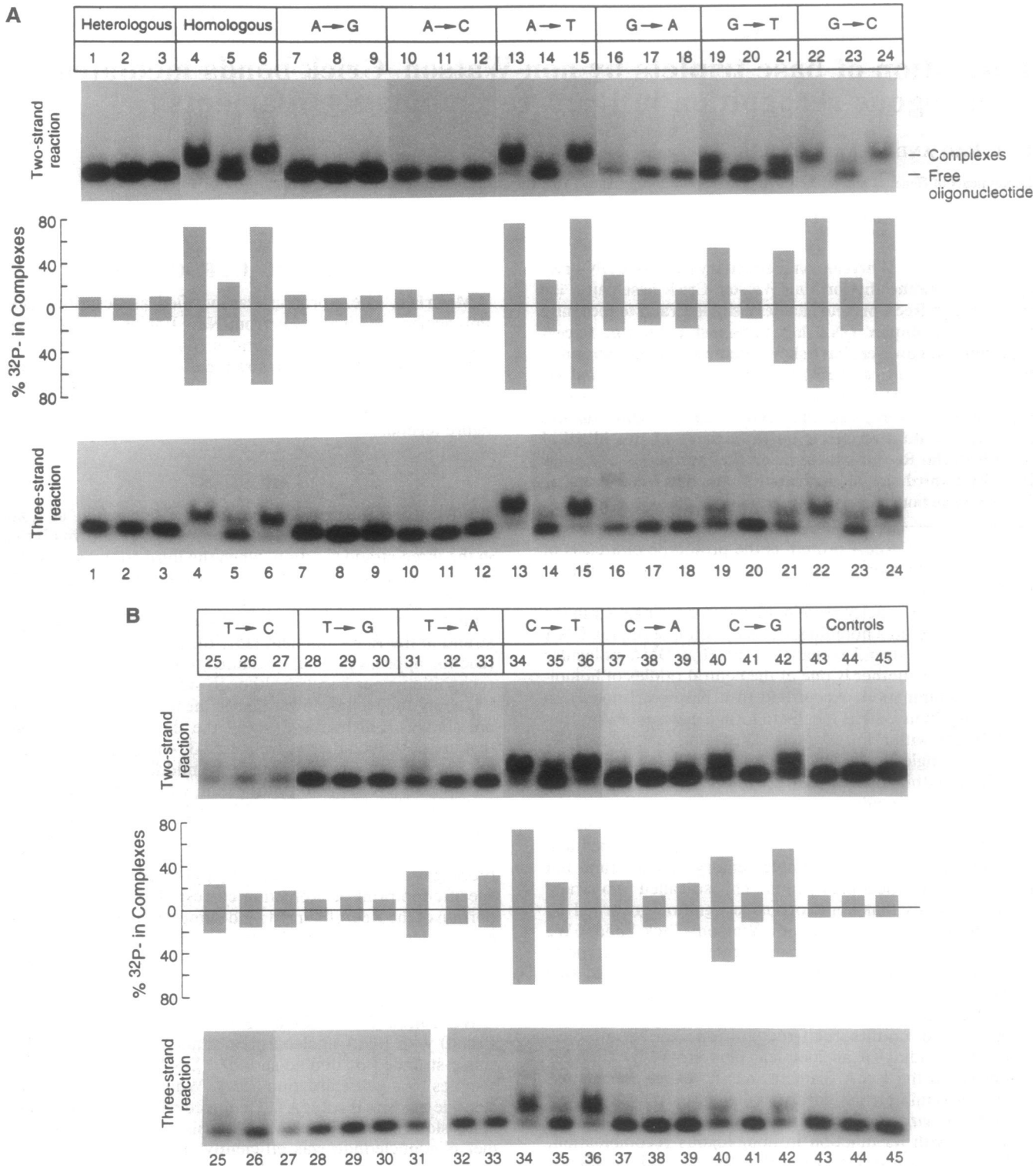

FIG. 1. Effects of base substitutions. The top of *A* and *B* shows the effects of base substitutions on the two-strand reaction; the bottom shows results of the three-strand reaction. A single-stranded 83-mer oligonucleotide corresponded to the (+) sequence in M13 circular single-stranded genome at nt 1082–1165: (5′-TCTGCGCCTCGTTCCGGCTA·AGTAACATGGAGCAGGTCGCGGATTTCGACACAATTTATCAGG-C-GATGATACAAATCTCCGTT-3′). For three-strand reactions, this was annealed with a 43-mer oligonucleotide that was complementary to the sequence between the dots, and purified on a native 10% polyacrylamide gel. The duplex was eluted and subsequently desalted as described (14). Nucleoprotein filaments were formed by incubating RecA protein (7 μM) with either the 83-mer oligonucleotide (12 μM) or the partially duplex oligonucleotide (18 μM) at 37°C for 12 min in 33 mM Hepes, pH 7.0/1.2 mM magnesium acetate/2 mM dithiothreitol/1.0 mM adenosine 5′-[γ-thio]triphosphate/bovine serum albumin (100 μg/ml). The wild-type 33-mer plus-strand oligonucleotide is as shown in boldface type in the sequence above. Pairing was initiated by adding 5′-labeled 33-mer homologous oligonucleotide (5 μM), a homologous oligonucleotide carrying base substitutions (5 μM), or a heterologous 33-mer oligonucleotide (5 μM, 5′-ACACACTGGTGCAGCTAACTGTTCTTCAGCTCG-3′), in the presence of 16 mM magnesium acetate and heterologous unlabeled carrier oligonucleotide (120 μM, 5′-CGATTGATGCGTACCAGCTTAC-CGAAGTTA-3′). Each pairing reaction (in 30 μl) was done in three ways: without any additional competitor (first lane in each set of three, e.g., lane 4), with a 10-fold excess (50 μM) of unlabeled competitor of the same sequence as the labeled 33-mer (second lane in each set of three, e.g., lane 5), or with another heterologous competitor (50 μM, 5′-TTCACAAACGAATGGATCCTCATTAAAGCCAGA-3′)

33-mer oligonucleotides were synthesized in which all A residues were replaced with G residues (designated as A → G), C residues (A → C), or T residues (A → T). Similarly, other substitutions were G → A, G → T, G → C, T → C, T → G, T → A, C → T, C → A, and C → G. The panel of 12 substitutions was studied in the same experiment (Fig. 1). As described above for homologous reactions, the specificity of each reaction was assessed by comparing the relative competition of excess unlabeled mutant oligonucleotide with that of an excess unlabeled heterologous sequence. In addition, by a computer search, we looked for sequences in 83-mer oligonucleotide that were either identical or complementary to any in the entire set of 33-mer oligonucleotides used here (17). No fortuitous matches were found until the stringency of search was lowered to 40% homology or less. Moreover, the distribution and frequency of partial homologies were similar for all 33-mer oligonucleotides, which included 1 homologous (wild-type), 2 heterologous, and 12 mutant 33-mer oligonucleotides (data not shown). The analysis ruled out fortuitous complementarity or identity as the basis for observed differences in pairing among this set of oligonucleotides.

The observations shown in Fig. 1 and summarized in Fig. 2 lead to several conclusions:

(*i*) In every case where a specific base in the naked 33-mer oligonucleotide was replaced with each of the other three, e.g., A → G, A → C, and A → T (see Fig. 2, row 1), there was at least one substitution in the set that abolished the ability to pair and did so in both two-strand and three-strand reactions (Fig. 1).

(*ii*) Not all substitutions in the naked oligonucleotide abolished recognition. All complementary substitutions except T → A (i.e., A → T, G → C, and C → G) were unimpaired. In addition, G → T and C → T substitutions had no effect.

(*iii*) An apparent asymmetry in recognition was revealed: If we assign the first position in a pair to the base in the filament and the second to the base in the oligonucleotide, we see that A·T, G·T, and C·T pairs allowed recognition, whereas the reciprocal ones (namely, T·A, T·G, and T·C) did not, which suggests that T residues in the filament pair with only T residues in the naked oligonucleotide, whereas T residues in the oligonucleotide opposite A, G, or C residues do not disrupt pairing.

(*iv*) In spite of the complexity inherent in the observations just described, the effects of base substitutions on recognition were identical in two-strand vs. three-strand reactions. The quantitative data obtained by gel scanning of two- vs. three-strand reactions are virtually mirror images of one another (Fig. 1). In contrast, seven base substitutions (T → C, T → G, C → A, C → T, G → A, G → T, and A → C) had no effect on the pairing of a 33-mer minus-strand oligonucleotide that was complementary to the plus-strand 83-mer in the RecA filament (data not shown).

Base Substitutions in Another Sequence. We repeated two-strand and three-strand pairings with an 83-mer oligonucleotide that was completely unrelated in sequence to that used in the previous experiments. The substitutions in labeled 33-mer oligonucleotides were of two types: all residues of a specific purine were replaced with another purine (A → G and G → A) or all residues of a specific pyrimidine were replaced with another pyrimidine (C → T and T → C). Homologous reactions showed a specific signal of expected strength while heterologous reactions yielded a background level (Fig. 3*A*).

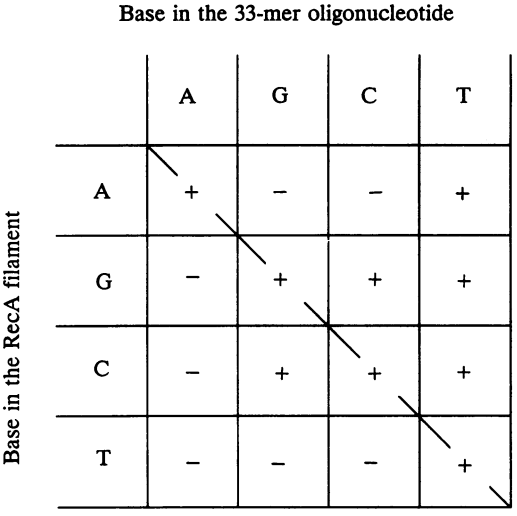

FIG. 2. Summary chart. Effects of base substitutions on recognition of sequence identity by the RecA nucleoprotein filament. +, Pairing of the oligonucleotide occurred; −, pairing disrupted.

As expected from the results diagramed in Fig. 2, in the two-strand and three-strand pairings, the A → G substitution eliminated pairing and the T → C substitution reduced pairing considerably, whereas the C → T mispairing had no effect. However, the outcome of a G → A substitution appeared contrary to expectation. A G → A substitution in the new set had little effect, while the same substitution in the earlier set disrupted pairing. However, in the new set, there were only five G residues in the region of homology as opposed to nine in the earlier one. To explore the effect of the number of substitutions, we made a direct comparison of pairing between either a single-stranded or double-stranded RecA filament with a set of three 33-mer oligonucleotides containing three, six, or nine substitutions (G → A) that were placed as far apart as was possible in the region of homology (for details, see Fig. 3*B*). Nine G → A substitutions effectively reduced the reaction to background, as was seen earlier (Fig. 1*A*). However, pairing reactions involving only six G → A substitutions produced measurable amounts of homologous complexes (Fig. 3*B*), and the yield of homologous complexes increased further as the number of substitutions was lowered to three.

Base Substitutions in the Filament. In the experiments described above, the base substitutions were all made in the naked 33-mer oligonucleotides, whereas the sequence of the nucleoprotein filament was unchanged (Fig. 1). As a reciprocal control, we introduced substitutions in the filament sequence. All T residues in the homologous region of the original 83-mer oligonucleotide (described in Fig. 1) were replaced with C residues. When this new filament was paired with the same set of 33-mer oligonucleotides in which the T residue had been changed to an A, G, or C residue (Fig. 2, row T), the predicted outcome is that shown in Fig. 2, row C. The same 33-mer oligonucleotides that previously produced T·G and T·C pairs that abolished the formation of complexes should now produce active C·G and C·C pairs; the oligonucleotide that produced inactive T·A pairs should produce inactive C·A pairs; and the wild-type oligonucleotide that produced active T·T pairs should produce active C·T pairs. As shown in Fig. 4, the results quantitatively fulfilled these

(third lane in each set of three, e.g., lane 6). Pairing was done for 6 min; the samples were analyzed by a gel retardation assay and quantitated with PhosphorImager (Molecular Dynamics) (7). In the middle, the results of the three-strand reactions are plotted as inverted bars for ease of comparison with the two-strand reactions. Additional controls included: omission of RecA protein (lane 43), omission of adenosine 5′-[γ-thio]triphosphate (lane 44), and deproteinization of wild-type reactions with SDS and proteinase K as described (lane 45) (7).

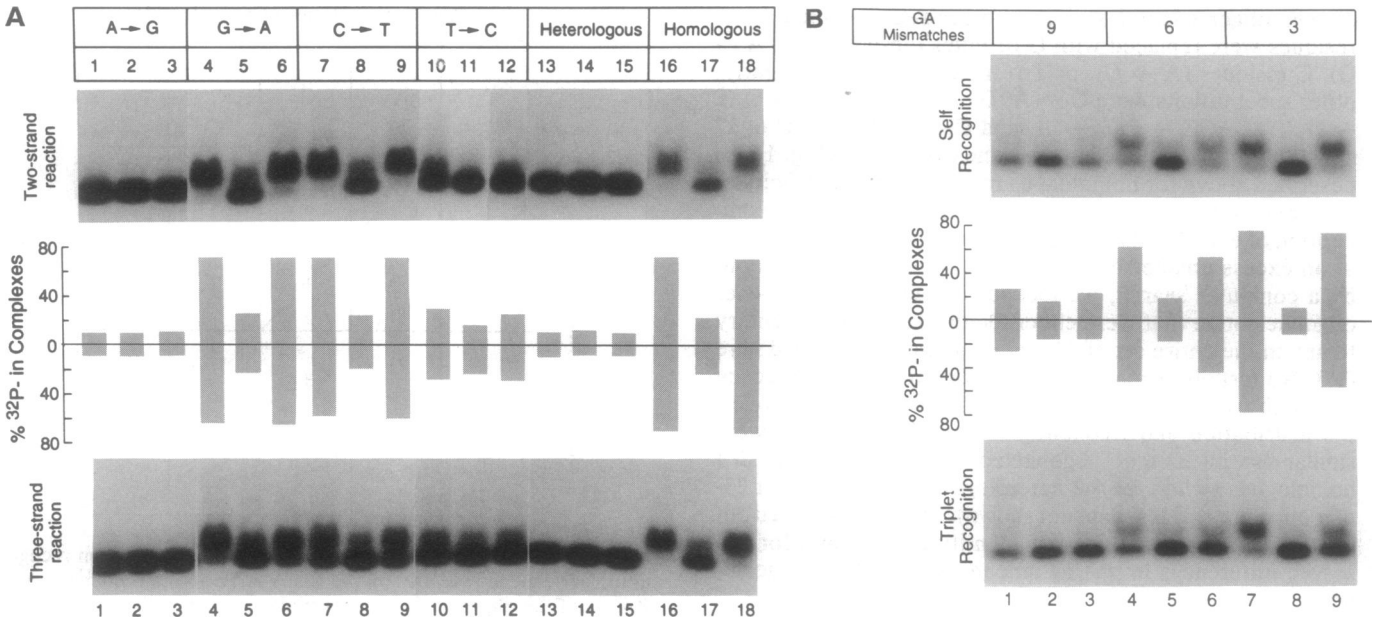

FIG. 3. (*A*) Base substitutions in another sequence. A filament bearing another 83-mer sequence [(+) sequence in M13 circular single-stranded genome at nt 182–265] (5′-TTGCATATTTAAAACATGTTGAGCTACAGCACCAGATTCAGCAATTAAGCTCTAAGCCATCCGCAAAAATGACCTCTTATCAA-3′) was used. The letters in boldface type correspond to homologous sequence. (*B*) Effect of the number of base substitutions on recognition of identity. Three mutant oligonucleotides where different G residues were replaced with A residues in the wild-type 33-mer sequence (shown in boldface type in Fig. 1) were: three G residues at nt 7, 14, and 23; six G residues at nt 4, 7, 11, 14, 17, and 23; all nine G residues. Other details were as described in Fig. 1.

predictions. These observations further validate the rules that are revealed in Fig. 2.

DISCUSSION

When we substituted bases in oligonucleotides that we paired with RecA nucleoprotein filaments, we found that the effects were both qualitatively and quantitatively indistinguishable whether the filament contained a single strand of identical sequence or duplex DNA. Since the mutual recognition of two identical sequences occurs by non-Watson–Crick bonds, it follows that recognition of homology involving three strands must occur via the same or similar bonds and, hence, that base triplets rather than Watson–Crick base pairs mediate the initial recognition in the three-strand reaction.

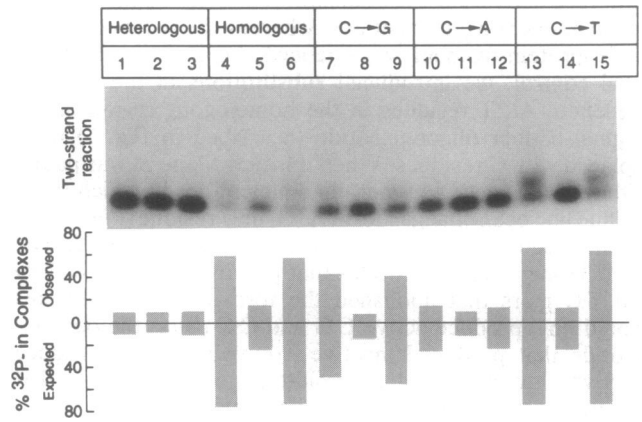

FIG. 4. Base substitutions in the filament. All T residues in the homologous part of the filament described in Fig. 1 were replaced with C residues. Upright bars are the values observed with the new filament, whereas the inverted bars belong to values expected on the basis of previous changes in the 33-mer oligonucleotide (described in Fig. 1). For other details, see Fig. 1.

Controls showed that similar base substitutions had no detectable effect on the pairing of complementary strands.

On the basis of studies of deproteinized products of the RecA reaction, we previously proposed a model for homologous recognition via base triplets (Fig. 5) that was similar to several earlier proposals (14, 21–23). Such models, however, fail to account satisfactorily for self-recognition if the third complementary strand [C(−) in Fig. 5] is simply removed from the triplet. Recently, Zhurkin *et al.* (20) pointed out that in such triplet schemes, there is a complementary pattern of partial charges, including the well-known charges that produce hydrogen bonds and weaker charges as well (Fig. 5); they specifically suggested that this pattern of complementary charges might constitute an electrostatic code for recognition. That complementary pattern, we note, is precisely the same for a two-strand vs. a three-strand reaction, although major rotations are required to bring pairs of complementary charges into proximity. Thus in principle, the electrostatic code rationalizes the observed identity of the two- and three-strand reactions.

According to our observations, every base in the filament contributes to the specificity of recognition in both the two-strand and three-strand reactions: an A residue will not pair with a G or C residue; a G or C residue will not pair with an A residue; and a T residue will pair only with a T residue (Fig. 2). These observations are at odds with an earlier model for homologous recognition that made use of self-recognition by purines (24). However, even when we take into account the possible electrostatic interactions, the model shown in Fig. 5 does not appear to explain all of the complexities revealed by the data, as summarized in Fig. 2. Some base changes had no effect, which might be due to lack of interference with the correct pairing of the other bases or to alternative pairing interactions. In addition, Fig. 2 reveals an unexplained asymmetry: The effects of replacing T residues with any other base differed when the same changes were made in the strand that was in the filament initially vs. the strand that was added.

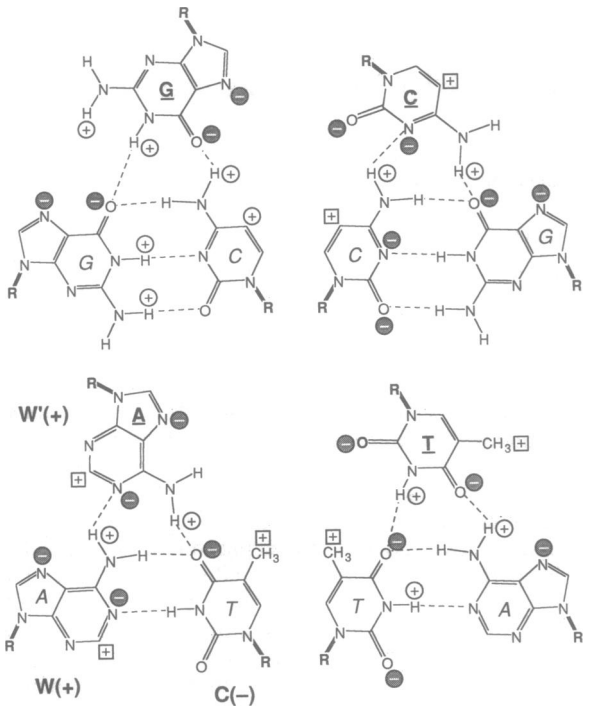

FIG. 5. Base-triplet model for homologous recognition. In the current experiments, W'(+) represents the base from incoming naked 33-mer oligonucleotide that is recognized by either a single strand [W(+)] (self-recognition in two-strand pairing) or a double strand [W(+)·C(−)] (triplet-recognition in three-strand pairing) in the filament. Charges in circles are stronger than those in squares (18, 19). This scheme was originally proposed to explain the effects of methylation of cytosine at N-4 and adenine at N-6 and the lack of effect of methylation of guanine at N-7 (14). The complementarity of partial charges, suggested by Zhurkin *et al.* (20) as an electrostatic recognition code, rationalizes the identity of two-strand and three-strand pairings as reported here. When the strand labeled C(−) is removed, the identical pattern of charge complementarity of two-strand reaction and three-strand reaction may be realized by either clockwise or counter-clockwise rotation of bases in W'(+).

Another remarkable feature of the base substitutions is the relative insensitivity of pairing to the number of mismatches. Of the nine G·G pairs in one 33-mer, more than six had to be changed into G·A mismatches before recognition was substantially abolished (Fig. 3*B*). We observed previously that in both the two-strand and three-strand reactions, homologous recognition was insensitive to the direction of the sugar-phosphate backbone (7). The assays employed in all these experiments may detect the earliest step of recombination—namely, the recognition of homology, which is fast, but apparently not stringent. In contrast, strand exchange is slow and stereospecific (ref. 7 and unpublished observations), which leads to the notion that specificity in recognition may be achieved by a quick imprecise search, followed by a more selective step associated with strand exchange, which provides a further barrier to nonhomologous interactions.

Three sets of observations now support the view that base triplets and non-Watson–Crick bonds mediate homologous recognition: These are the similar requirements and similar stereochemistry of the two-strand vs. three-strand reactions (7), the indistinguishable effects of base substitutions on both reactions as described here, and recent studies of a RecA mutant that promotes Watson–Crick pairing normally but is partially defective in recognition of homology in duplex DNA (25) and in self-recognition (H. Kurumizaka, T. Ogawa, T. Shibata, B.J.R., and C.M.R., unpublished work). The observations on the mutant, moreover, implicate the non-Watson–Crick interactions of self-recognition in recombination. More experiments are required to understand the complexities revealed in Fig. 2, which nonetheless provide the outline of a recognition code that might eventually be useful in gene targeting and gene therapy based on endogenous RecA homologs.

We are grateful to Ning Ye for technical assistance, to Jan Zulkeski for secretarial assistance, and to SungKay Chiu, Rochelle Reves Bazemore, Guruchuran Reddy, Efim Golub, Hitoshi Kurmizaka, and Takehiko Shibata for collegial advice. We are especially grateful to Victor Zhurkin for discussion prior to publication of his ideas on base recognition by electrostatic interactions. This research was sponsored by Grant 2R01GM33504 from the National Institutes of Health.

1. Bishop, D. K., Park, D., Xu, L. & Kleckner, N. (1992) *Cell* **69**, 439–456.
2. Shinohara, A., Ogawa, H. & Ogawa, T. (1992) *Cell* **69**, 457–470.
3. Ogawa, T., Yu, X., Shinohara, A. & Egelman, E. H. (1993) *Science* **259**, 1896–1899.
4. Radding, C. M. (1993) *Curr. Biol.* **3**, 358–360.
5. Flory, J., Tsang, S. S. & Muniyappa, K. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 7026–7030.
6. Stasiak, A. & Egelman, E. H. (1988) in *Genetic Recombination*, eds. Kucherlapati, R. & Smith, G. R. (Am. Soc. Microbiol., Washington, DC), pp. 265–308.
7. Rao, B. J. & Radding, C. M. (1993) *Proc. Natl. Acad. Sci. USA* **90**, 6646–6650.
8. Htun, H. & Dahlberg, J. E. (1988) *Science* **241**, 1792–1796.
9. Hanvey, J. C., Shimizu, M. & Wells, R. D. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 6292–6296.
10. Praseuth, D., Perouault, L., Doan, T. L., Chassignol, M., Thuong, N. & Helene, C. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 1349–1353.
11. Rippe, K., Ramsingh, N. B. & Jovin, T. M. (1989) *Biochemistry* **28**, 9536–9541.
12. Beal, P. A. & Dervan, P. B. (1991) *Science* **251**, 1360–1363.
13. Shibata, T., DasGupta, C., Cunningham, R. P., Williams, J. G. K., Osber, L. & Radding, C. M. (1981) *J. Biol. Chem.* **256**, 7565–7572.
14. Rao, B. J., Chiu, S. K. & Radding, C. M. (1993) *J. Mol. Biol.* **229**, 328–343.
15. Rao, B. J., Jwang, B. & Radding, C. M. (1990) *J. Mol. Biol.* **213**, 789–809.
16. Radding, C. M. (1991) *J. Biol. Chem.* **266**, 5355–5358.
17. Schwindinger, W. F. & Warner, J. R. (1984) *Nucleic Acids Res.* **12**, 601–604.
18. Renugopalakrishnan, V., Lakshminarayanan, A. V. & Sasisekharan, V. (1971) *Biopolymers* **10**, 1159–1167.
19. Saenger, W. (1984) *Principles of Nucleic Acid Structure* (Springer, New York).
20. Zhurkin, V. B., Raghunathan, G., Ulyanov, N. B., Camerini-Otero, R. D. & Jernigan, R. L. (1994) *J. Mol. Biol.* **239**, in press.
21. Lacks, S. (1966) *Genetics* **53**, 207–235.
22. Umlauf, S. (1990) Ph.D. thesis (Univ. Wisconsin, Madison).
23. Chiu, S. K., Rao, B. J. & Radding, C. M. (1993) *Biochemistry* **32**, 13146–13155.
24. Hseih, P., Camerini-Otero, C. S. & Camerini-Otero, R. D. (1990) *Genes Dev.* **4**, 1951–1963.
25. Kurumizaka, H., Ikawa, S., Ikeya, T., Ogawa, T. & Shibata, T. (1994) *J. Biol. Chem.* **269**, 3068–3075.