GC-rich DNA sequences block homologous recombination in vitro

Alexandra A. Gruss, Véronique Moretto, Stanislav S. Ehrlich, Patrick Duwat, Patrick Dabert

Communication

THE JOURNAL OF BIOLOGICAL CHEMISTRY
Vol. 266, No. 11, Issue of April 15, pp. 6667-6669, 1991

GC-rich DNA Sequences Block Homologous Recombination *in Vitro**

(Received for publication, January 3, 1991)

Alexandra Gruss‡, Véronique Moretto, S. Dusko Ehrlich, Patrick Duwat, and Patrick Dabert

From the Laboratoire de Génétique Microbienne, Institut de Biotechnologie, Institut National de la Recherche Agronomique, Domaine de Vilvert, 78350 Jouy en Josas, France

The capacity of the RecA protein of *Escherichia coli* to promote an essential step in homologous recombination, strand transfer, was tested on DNA substrates varying in percentage GC. GC content was determined by a novel method using the polymerase chain reaction. Strand transfer activity is greatly reduced as a function of increasing GC content of the DNA. Some reduction is observed with substrates having a GC percentage similar to that of *E. coli*. The transfer reaction between sequences adjacent, but not distal, to GC-rich sequences is similarly decreased, suggesting that the structure of RecA-DNA complexes may differ with the GC content of DNA. Our results implicate an important role of DNA sequence in homologous recombination and suggest that many sequences are excluded due to their GC content.

The RecA protein of *Escherichia coli* is necessary for homologous recombination (1, 2), induction of repair processes (3, 4), and normal cell maintenance. Its analogues appear to exist in every organism. RecA activity depends on its binding to ssDNA.[1] Binding to ssDNA, and under certain conditions to dsDNA, is affected by DNA sequence and potential structure (5; reviewed in Ref. 6). For example, RecA has high affinity for adenine-thymine tracts of dsDNA (7), whereas Z-form DNA is more rapidly bound than the B-form (8, 9). Binding of RecA to ssDNA is reduced by secondary structure (10–13). We effected experiments, using percentage GC content as the criterion, to determine whether recombination by RecA is affected by base composition of the substrates. It is reported that homologous recombination is severely reduced if substrates are GC-rich or if substrates are adjacent to GC-rich DNA.

EXPERIMENTAL PROCEDURES

DNA—DNA fragments 2 kb long were isolated from species varying in their average GC content and cloned into vector pHV33ΔPvu (14) at the *Hin*dIII-*Eco*RI sites. The 2-kb *Hin*dIII-*Eco*RI fragments and the 2-kb *Xmn*I fragment of M13 mp11 were purified from agarose gels using a Gene-Clean kit (Bio101). Fragments were subcloned onto M13 mp11 for preparation of ssDNA from phage particles. dsDNA fragments were end-labeled by Klenow polymerase (Boehringer Mannheim) according to the supplier's instructions.

Determination of GC content of a DNA Fragment by PCR—Five sequenced fragments, 1.5–2 kb long, were used as standards, and nine 2-kb unknown fragments were tested. DNAs, which were all cloned onto sequenced plasmids, were amplified by PCR (15) using the buffer and *Taq* DNA polymerase supplied by Promega Corp., appropriate primers (0.5 μM each), and nucleotides (50 μM each). 7′-Deaza-2′-dGTP was used in place of dGTP to improve amplification of GC-rich segments (16). 3 μCi of [^{3}H]dTTP and 1 μCi of [^{32}P]dCTP were added with cold nucleotides. Radiolabeled nucleotides were mixed prior to addition to ensure that a fixed ratio was present in each sample. The PCR cycle was 5 min at 94 °C and then 25 cycles of 0.5 min at 96 °C, 2 min at 48 °C, and 3 min at 72 °C. Amplified DNA samples were purified from agarose gels by freeze-squeeze (Millipore), and labeled DNA samples were counted on two channels for ^{3}H and ^{32}P in a Beckman 3801 scintillation counter. After correction for background and spillover, the ratio of radioactive nucleotide incorporated (inc) to total radioactivity added (total) was determined for [^{3}H]dTTP and [^{32}P]dCTP. The percentage GC of the five known fragments were plotted against the incorporated C (%) (see Fig. 1).

$$\frac{^{32}P_{inc}/^{32}P_{total}}{^{32}P_{inc}/^{32}P_{total} + ^{3}H_{inc}/^{3}H_{total}} \times 100$$

A best fit line was determined by linear regression analysis on standards, with a correlation coefficient greater than 0.99. The percentage GC of unknown fragments was calculated by the equation obtained.

Strand Transfer Assay—Reactions were performed in a 20-μl volume containing dsDNA, ssDNA (about one-fifth of the circular ssDNA molecule is homologous to the dsDNA fragment), RecA, and SSB (Pharmacia LKB Biotechnology Inc) in the amounts indicated in the figure legends, in 20 mM Tris-HCl, pH 7.5, 10 mM $MgCl_2$, 2 mM dithiothreitol, 2 mM ATP, and 100 μg/ml bovine serum albumin, as previously described (10). Reactions were stopped after 30 min (unless otherwise indicated) at 37 °C by addition of 0.1 volume of 4% sodium dodecyl sulfate, 50 mM EDTA, and 0.1 mg/ml proteinase K. Agarose gels (0.7 or 0.8%) were run in Tris borate buffer containing 40 μg/ml ethidium bromide overnight at 1.5 V/cm.

RESULTS AND DISCUSSION

Two-kb segments of DNA originating from prokaryotes with different percentage GC contents were cloned for use in strand transfer assays. Their GC content varied between 28 and 70%. This was determined by a rapid method we have developed, based on PCR (15) in which double label ([^{32}P]dCTP and [^{3}H]dTTP) was incorporated into the amplified DNA (Fig. 1); segments of known GC content served as standards. Unlike previous methods, this determination of GC content is unbiased by sequence organization (17).

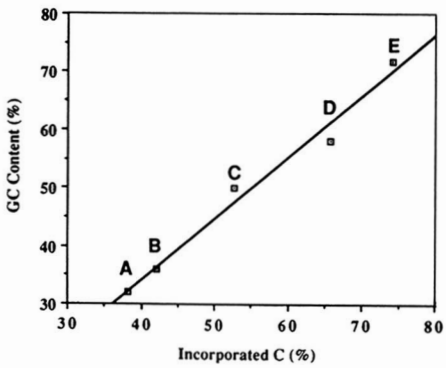
A strand transfer activity test (10; diagrammed in Fig. 2*A*) was performed on purified 2-kb dsDNA fragments and ssDNA of M13 clones with the corresponding inserts (Fig. 2*B*, *top*). A 2-kb *Xmn*I fragment of M13 (45% GC) was also tested against all ssDNAs (Fig. 2*B*, *bottom*). Efficiency of strand transfer in the presence of RecA and SSB diminished with increasing percentage GC of the DNA substrates. The diminution was even observed with substrates having the same GC content as *E. coli* DNA (inserts 3 and 4).

One explanation for inefficient strand transfer is that secondary structure is greater in GC-rich ssDNA and the quantity of SSB used in reactions is insufficient to melt the structures. Sequence analyses of DNA fragments having 36

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

‡ To whom correspondence should be addressed. Tel.: (33-1)-34-65-25-27; Fax: (33-1)-34-65-22-73.

[1] The abbreviations used are: ssDNA, single-stranded DNA; dsDNA, double-stranded DNA; kb, kilobase(s); PCR, polymerase chain reaction; SSB, single-stranded DNA-binding protein.



Standard DNA	%GC	Source
A	32	*Lactococcus lactis*
B	36	*Staphylococcus aureus*
C	50	*Escherichia coli*
D	58	*Escherichia coli*
E	70	*Streptomyces limosus*

Test DNA	Calculated %GC	
1	28	*Staphylococcus aureus*
2	30	*Staphylococcus aureus*
3	52	*Salmonella typhimurium*
4	56	*Salmonella typhimurium*
5	62	*Pseudomonas aeruginosa*
6	64	*Pseudomonas aeruginosa*
7	61	*Pseudomonas aeruginosa*
8	67	*Pseudomonas aeruginosa*
9	70	*Pseudomonas aeruginosa*

FIG. 1. **Determination of GC content of a DNA fragment by double-label incorporation in PCR.** Five sequenced fragments (*A–E*) 1.5–2 kb long, used as standards, and nine unknown fragments were subjected to PCR in the presence of [^{3}H]dTTP and [^{32}P]dCTP. The percentage GC of the five known fragments were plotted against the percentage of [^{32}P]dCTP incorporated (*Incorporated C (%)*; see "Experimental Procedures"). A best fit line was determined by linear regression analysis on standards, with a correlation coefficient greater than 0.99. The percentage GC of unknown fragments was calculated by the equation obtained. The averages of two determinations are listed.

or 71% GC revealed that each contains numerous short inverted repeats, which are more stable in the GC-rich DNA (not shown). A kinetic study of strand transfer of these two fragments in the presence of 6 times the amount of SSB used in the previous assay was performed using labeled dsDNA (Fig. 3). Reactions containing the usual amount, and 12 times the amount of SSB, were included as controls. Compared to results with the amounts usually employed (10, 11), the 6-fold increase in SSB augmented transfer of DNA substrates containing secondary structure, whereas a 12-fold increase did not improve, or even inhibited, the reaction. Complete transfer of the 36% GC substrate was observed, whereas only a minor proportion of the 71% GC DNA was converted to heteroduplex product. These results indicate that SSB may be unable to overcome the effects of high percentage GC or that other factors besides secondary structure account for inefficient strand transfer of high percentage GC substrates. A slowed unwinding of GC-rich DNA by RecA may also be a factor, since the reaction time with the GC-rich substrate is increased (Fig. 3). The heteroduplex product of 36% GC DNA substrates appeared after 5–10 min, as expected from published studies (18), whereas that of the 71% GC DNA appeared after 20–30 min. Similar results were obtained when kinetics were performed with 1 × SSB (not shown).

Another explanation for inefficient strand transfer is that GC-rich DNA binds to RecA in a way that does not permit homologous pairing. Previous studies indicate that each base is not equivalently disposed on DNA (19); guanine is more accessible to modifying agents than adenine on both ssDNA and dsDNA. This difference is greatly enhanced when DNA is complexed with RecA (19, 20), even though DNA recognition by RecA is effected through the phosphate backbone (20). Structural differences are thus expected between a RecA complex with low or high percentage GC DNA, and could be the basis for reduced transfer of high percentage GC molecules. The following experiment suggests that these structural differences exist.

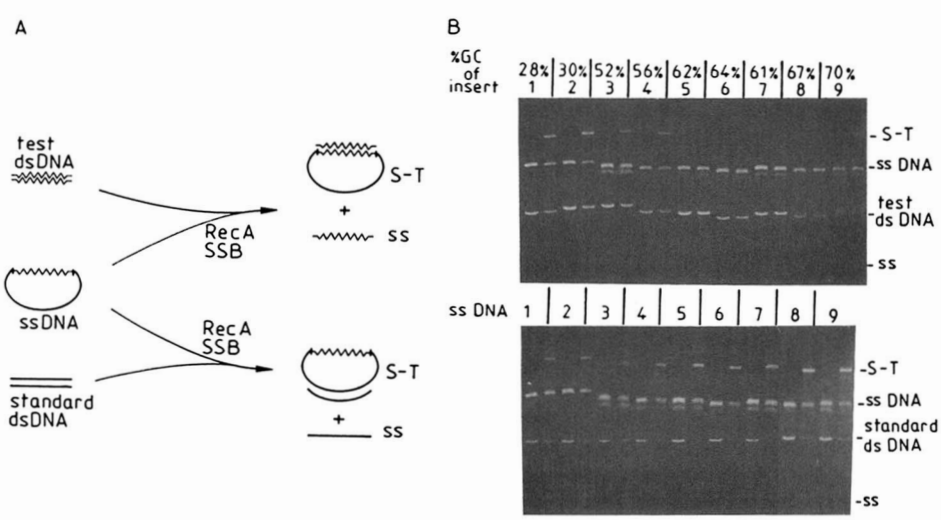


FIG. 2. **Strand transfer is inefficient between high percentage GC substrates.** *Panel A*, schematic representation of DNA substrates and products of strand transfer. The ssDNA substrate is M13 carrying inserts of varying GC content. The dsDNA substrates, approximately 2 kb in length, correspond either to the inserts (*wavy line*) or to a constant region on M13 (*smooth line*). RecA and SSB proteins were added to test strand transfer, which would produce a heteroduplex molecule and a 2-kb ssDNA fragment. *Panel B*, *top*, nine dsDNA substrates (*1–9*), each with a different percentage GC (see Fig. 1), are tested for strand transfer with ssDNA containing the corresponding homology. *Panel B*, *bottom*, each of the nine ssDNAs was also tested with a 2-kb *Xmn*I dsDNA segment of M13 (45% GC) homologous with the M13 part of the ssDNAs. Reaction mixtures contain 2 μM dsDNA, 10 μM ssDNA, 5 μM RecA, and 0.5 μM SSB. *ssDNA*, circular ssDNA substrate; *dsDNA*, dsDNA fragment substrate; *S-T*, strand transfer product; *ss*, free ssDNA product. For each set of substrates, a sample is taken just before (*left*) and 30 min after (*right*) RecA and SSB addition and incubation. The *number* above each pair of wells corresponds to the insert used.

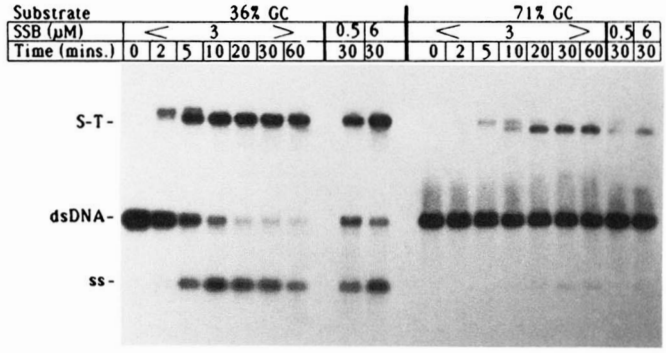


FIG. 3. **Strand transfer of GC-rich substrates in the presence of usual or increased SSB concentrations is inefficient.** One DNA substrate (*Staphylococcus aureus* plasmid pUB110; fragment base pairs 780–2363 on the published map (21)) is 36% GC, and the other (*Streptomyces limosus* DNA fragment; base pairs 270–1820 of a published gene sequence (22)) is 71% GC. These substrates are subfragments of DNA samples *A* and *E* (Fig. 1) and are referred to by their percentage GC. Kinetics were performed during a 1-h period, with 10-μl samples removed at the times indicated above the wells. Each sample contained 1 μM dsDNA labeled by Klenow polymerase at *Bam*HI-*Bgl*II (*36% GC*) or *Eco*RI-*Hin*dIII (*71% GC*) ends, 5 μM of ssDNA, 5 μM RecA, and 3 μM SSB (as compared with 0.5 μM SSB in previous assays). Reactions in the presence of 0.5 and 6 μM SSB are also presented. An autoradiograph of a 0.7% agarose gel is shown. Intermediates of the transfer reaction migrate above the complete strand transfer reaction product. *dsDNA*, dsDNA fragment substrate; *S-T*, strand transfer product; *ss*, free ssDNA product.

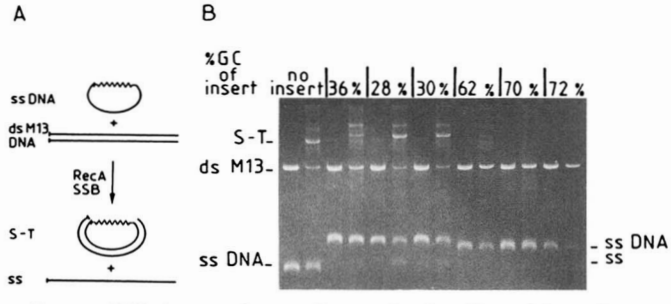


FIG. 4. **Efficiency of strand transfer is affected by adjacent sequences on ssDNA substrates.** In these experiments, the dsDNA substrate was M13-linearized at the site of fragment insertion, *Hin*dIII-*Eco*RI. The different ssDNA substrates tested were M13, M13 containing inserts A, 2, 3, 6, 10 (see Fig. 1), or a 72% GC *S. limosus* fragment that showed poor strand transfer (not shown). *Panel A*, schematic representation of the experiment. Strand transfer initiates within the homologous ssDNA region (*smooth line*) abutting the site of foreign DNA insertions (*wavy line*). *Panel B*, experimental results. Reactions contained 4 μM ssDNA, 4 μM dsDNA, 6 μM RecA, and 0.6 μM SSB. For each set of substrates, samples loaded on the gel are taken just before (*left*) and 30 min after (*right*) RecA and SSB addition and incubation. The percentage GC of the inserts is indicated above each reaction. *dsM13*, linear dsM13 DNA substrate; *ssDNA*, circular ssDNA substrate; *S-T*, strand transfer products; *ss*, free ssDNA product.

The efficiency of strand transfer of sequences neighboring the high percentage GC DNA on the ssDNA substrate was tested (diagrammed in Fig. 4*A*). The ssDNA substrates were M13 or M13 carrying inserts of different percentage GC, and the dsDNA substrate was M13 linearized at the sites used for insert cloning. Transfer initiates from sequences adjacent to the inserts (Fig. 4*B*). With three ssDNA substrates carrying inserts low in percentage GC (28–36%), recombination was as efficient as with the substrate having no insert. In contrast, heteroduplex formation was barely detectable with three substrates carrying inserts rich in GC (62–72%). This interference is provoked from outside the region involved in recombination. It occurs when strand transfer initates adjacent to the site of the insert, but not if the homology is distal to the insert (see Fig. 2*B*, *bottom*). These results suggest that RecA-SSB loading on high percentage GC DNA, or possibly the DNA itself, imposes a steric block to adjacent initiation of strand transfer. Thus, a complex between RecA and high percentage GC ssDNA may not present the DNA in an efficient way for homologous pairing and may sterically block initiation of nearby transfer events. To our knowledge, a complex between RecA and high percentage GC ssDNA has not yet been examined.

The above observations may indicate an exclusion of GC-rich sequences from recombination *in vivo*. Previously published RecA studies often employ either M13 (41% GC) or φX174 (46% GC) DNA as substrates, below the average of 51% for *E. coli* DNA. An intriguing possibility is that RecA (and/or SSB) activity is not optimal for the average base composition of *E. coli* DNA and that GC-rich DNA may consequently remain more highly conserved. We would expect that RecA analogues in organisms with high percentage GC DNA are more efficient than RecA in recombining GC-rich substrates, and we are currently examining this hypothesis.

Acknowledgments—We thank E. Egelman, E. Maguin, and B. Michel for helpful discussion, E. Cassuto for advice and the generous gift of RecA protein, M. J. Virolle for the clone of *S. limosus* DNA, and V. Akueson for art work. We appreciate the critical reading of this manuscript by R. D'Ari, E. Cassuto, E. Egelman, and C. Radding.

REFERENCES

1. Clark, A. J. (1973) *Annu. Rev. Genet.* **7,** 67–86
2. West, S. C., Cassuto, E., Mursalim, J., and Howard-Flanders, P. (1980) *Proc. Natl. Acad. Sci. U. S. A.* **77,** 2569–2573
3. Hanawalt, P. C., Cooper, P. K., Ganesan, A. K., and Smith, C. A. (1979) *Annu. Rev. Biochem.* **48,** 783–836
4. Walker, G. C. (1984) *Microbiol. Rev.* **48,** 60–93
5. Pugh, B. F., and Cox, M. M. (1988) *J. Mol. Biol.* **203,** 479–493
6. Cox, M., and Lehman, I. R. (1987) *Annu. Rev. Biochem.* **56,** 229–262
7. Kowalczykowski, S. C., Clow, J., and Krupp, R. A. (1987) *Proc. Natl. Acad. Sci. U. S. A.* **84,** 3127–3131
8. Blaho, J. A., and Wells, R. D. (1987) *J. Biol. Chem.* **262,** 6082–6088
9. Kim, J.-I., Heuser, J., and Cox, M. M. (1989) *J. Biol. Chem.* **264,** 21848–21856
10. Muniyappa, K., Shaner, S. L., Tsang, S. S., and Radding, C. M. (1984) *Proc. Natl. Acad. Sci. U. S. A.* **81,** 2757–2761
11. Julin, D. A., Riddles, P. W., and Lehman, I. R. (1986) *J. Biol. Chem.* **261,** 1025–1030
12. Kowalczykowski, S. C., Clow, J., Somani, R., and Varghese, A. (1987) *J. Mol. Biol.* **193,** 81–95
13. Kowalczykowski, S. C., and Krupp, R. A. (1987) *J. Mol. Biol.* **193,** 97–113
14. te Riele, H., Michel, B., and Ehrlich, S. D. (1986) *EMBO J.* **5,** 631–637
15. Saiki, R. K., Gelfand, D. H., Stoffel, S., Scharf, S. J., Higuchi, R., Horn, G. T., Mullis, K. B., and Erlich, H. A. (1988) *Science* **239,** 487–491
16. Innis, M. A., Gelfand, D. H., Sninsky, J. J., and White, T. J. (eds) (1990) *PCR Protocols,* pp. 54–59, Academic Press, Inc., San Diego
17. de Ley, J. (1970) *J. Bacteriol.* **101,** 738–754
18. Chow, S. A., Honigberg, S. M., Bainton, R. J., and Radding, C. M. (1986) *J. Biol. Chem.* **261,** 6961–6971
19. Di Capua, A., and Muller, B. (1987) *EMBO J.* **6,** 2493–2498
20. Leahy, M., and Radding, C. M. (1986) *J. Biol. Chem.* **261,** 6954–6960
21. McKenzie, T., Hoshino, T., Tanaka, T., and Sueoka, N. (1986) *Plasmid* **15,** 93–103
22. Long, C. M., Virolle, M.-J., Chang, S.-Y., Chang, S., and Bibb, M. J. (1987) *J. Bacteriol.* **169,** 5745–5754