Proc. Natl. Acad. Sci. USA
Vol. 89, pp. 8068–8072, September 1992
Plant Biology

A homolog of *Escherichia coli* RecA protein in plastids of higher plants

(chloroplast/recombination/*recA* gene/DNA damage/DNA repair)

H. CERUTTI†, M. OSMAN, P. GRANDONI, AND A. T. JAGENDORF‡

Section of Plant Biology, Cornell University, Ithaca, NY 14853

Contributed by A. T. Jagendorf, April 24, 1992

ABSTRACT **Studies of chloroplast DNA variations, and several direct experimental observations, indicate the existence of recombination ability in algal and higher plant plastids. However, no studies have been done of the biochemical pathways involved. Using a part of a cyanobacterial *recA* gene as a probe in Southern blots, we have found homologous sequences in total DNA from *Pisum sativum* and *Arabidopsis thaliana* and in a cDNA library from *Arabidopsis*. A cDNA was cloned and sequenced, and its predicted amino acid sequence is 60.7% identical to that of the cyanobacterial RecA protein. This finding is consistent with our other results showing both DNA strand transfer activity and the existence of a protein of the predicted molecular mass crossreactive with antibodies to *Escherichia coli* RecA in the stroma of pea chloroplasts.**

The observation of chloroplast DNA recombinants in somatic hybrids of higher plants (1), genetic studies of the inheritance of chloroplast markers in several crosses of *Chlamydomonas* (2–4), the integration of donor DNA by homologous recombination in chloroplasts of transformed *Chlamydomonas* (5), and extensive comparative analyses of chloroplast genome structure (6–9) indicate that DNA recombination occurs in chloroplasts of both higher plants and green algae. The biochemistry of any recombinational mechanism in chloroplasts is completely unknown, however.

It is generally accepted that plastids originated from cyanobacterial progenitors, acquired by an ancestral eukaryotic cell through an endosymbiotic event (10, 11). Therefore it seemed probable that any chloroplast recombination system should be related to a eubacterial counterpart. In *Escherichia coli* and many other prokaryotes, the RecA protein is essential for homologous recombination and for a variety of SOS responses to DNA damage (12–20). In searching for a possible higher-plant *recA* gene we used a cyanobacterial *recA* as a probe and found homologous sequences in nuclear DNA from pea and *Arabidopsis thaliana*. With the same probe we have cloned an *Arabidopsis thaliana* cDNA that encodes a protein highly homologous to eubacterial RecA, except for a predicted chloroplast transit peptide at its amino terminus.§ The likely expressed protein was detected in chloroplast stromal extracts by crossreaction with polyclonal antibodies to *E. coli* RecA protein.

MATERIALS AND METHODS

Materials. Standard laboratory chemicals were purchased, unless otherwise noted, from Sigma. Purified RecA protein was obtained from United States Biochemical. Cellulysin and Macerase were from Calbiochem. Horseradish peroxidase-conjugated goat antibodies to rabbit IgG, and a chemiluminescent substrate, were from Amersham.

Isolation of Stroma from Intact Chloroplasts and Immunodetection of a RecA Protein. Protoplasts were prepared from leaves of *Arabidopsis* or pea by digestion with 3% (wt/vol) Cellulysin and 0.5% (wt/vol) Macerase. Intact chloroplasts were isolated by passage of washed protoplasts through a nylon mesh (10-μm pore size) for disruption, followed by centrifugation of the homogenate on discontinuous Percoll gradients (21). Chloroplasts were broken by osmotic lysis and thylakoid membranes were removed by centrifugation. The supernatant was concentrated by acetone precipitation to give the chloroplast stromal fraction. Potential bacterial contamination was tested by plating aliquots of isolated protoplasts or chloroplasts on LB medium (22). Proteins from the different fractions were separated by SDS/polyacrylamide gel electrophoresis, transferred to nitrocellulose, and probed with two polyclonal rabbit antisera to RecA protein from *E. coli* strain K-12. Detection was by means of goat anti-rabbit IgG antibodies conjugated to horseradish peroxidase, and a chemiluminescent substrate. In some cases, the enzyme was inactivated by incubation in 15% (wt/vol) H_2O_2 and the blot was reprobed with an antibody against the γ subunit of CF1.

DNA Isolation and Probing with a Cyanobacterial *recA* Gene Fragment. Total DNA was isolated from *Arabidopsis* leaves by a miniprep procedure (R. L. Last, personal communication). Isopycnic CsCl centrifugation (22) was used for the purification of total DNA from pea leaves and of chloroplast DNA from intact pea chloroplasts. Standard procedures were used for digestion, electrophoretic separation, and transfer of the DNA to nylon membranes (22). The filters were probed with a ^{32}P-labeled *Bst*EII fragment comprising the 5′ half of the coding sequence (23). Prehybridization was at 70°C for 6 hr and hybridization was at 60°C overnight (24). Filters were washed three times for 1.5 hr with 2× SSC/0.1% SDS (22) at room temperature, then three times in 0.2× SSC/0.1% SDS at 50°C for another 1.5 hr. The membranes were exposed to Kodak XAR-5 film with an intensifying screen.

cDNA Cloning and Sequencing. A cDNA was isolated by screening ≈300,000 members of an *Arabidopsis* library in the λYES vector (25), using as probe the *Synechococcus recA* gene (23). The bacteriophages were plated on an *E. coli* strain deleted for *recA* (JC 10289, ref. 26). Standard procedures were used for library screening (22). After subcloning into pBluescript (Stratagene), nested deletions were generated and the DNA was sequenced on both strands by the dideoxy chain-termination method using T7 DNA polymerase (22). The sequences were analyzed with the Genetics Computer Group software package (27).

Protoplast Treatments. Protoplasts (10 ml, 10^6 protoplasts per ml) in liquid LP* medium (28) were placed in 9-cm Petri

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.

†Present address: Department of Botany, Duke University, Durham, NC 27706.
‡To whom reprint requests should be addressed.
§The sequence reported in this paper has been deposited in the GenBank data base (accession no. M98039).

dishes. Mitomycin C was added to the desired concentration, and protoplasts were incubated for 12 hr in the dark. Intact protoplasts were reisolated by flotation, and proteins were extracted and analyzed by Western blotting.

RESULTS

Detection of a Protein in Chloroplast Stroma That Is Immunologically Related to *E. coli* RecA. An immunoblotting technique was used to detect *Arabidopsis* proteins related to RecA (Fig. 1*a*). Nonimmune serum did not reveal any bands (data not shown). Two polyclonal antibodies raised against *E. coli* RecA crossreacted with three protoplast proteins. Two of them are soluble chloroplast proteins detected in the stromal fraction (Fig. 1*a*). The apparent molecular mass of the faster moving stromal protein is 40.5 kDa, almost identical with that of *E. coli* RecA (Fig. 1*a*). Similar proteins were also identified in pea chloroplasts (Fig. 1*b*). Consistent with the presence of RecA in chloroplasts, we detected DNA strand exchange, an essential activity of *E. coli* RecA (12, 14), in crude stromal extracts from pea (H.C. and A.T.J., unpublished work).

Induction by DNA Damage. It has been argued that the primary biological role of recombination is the repair of DNA damage (29). Exposure of *E. coli* to DNA-damaging agents induces the SOS response, resulting in derepression of ≈20 genes (including *recA*) (12–20). The RecA protein is involved in multiple aspects of this response: regulation of gene induction by promoting cleavage of the LexA repressor (13, 15, 19, 20), recombinational repair (13, 14, 19, 20), SOS mutagenesis (13, 16, 19), DNA replication (13, 17), and duplication mutagenesis (13). In other organisms, genes involved in DNA repair/recombination are also induced in response to DNA damage (19, 30).

To see whether DNA-damaging agents would affect expression of the proteins immunologically related to RecA, pea protoplasts were incubated with mitomycin C. Mitomycin C is a bifunctional alkylating agent that forms interstrand crosslinks (15, 31), presumably requiring a recombinational pathway for repair (29). The treatments increased the steady-state level of the chloroplast crossreacting protein similar in molecular mass to RecA, suggesting its involvement in DNA repair/recombination (Fig. 1*b*). The same blot was probed with an antibody against the γ subunit of the chloroplast ATP synthetase as a control for the proper loading of the lanes (Fig. 1*c*). Further details of the induction will be described elsewhere (H.C., H.-Z. Ibrahim, and A.T.J., unpublished work).

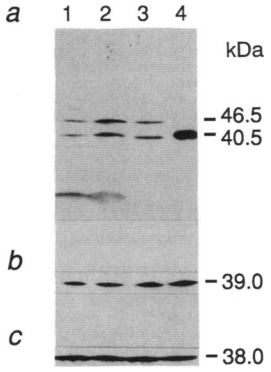

FIG. 1. Detection of proteins immunologically related to *E. coli* RecA in *Arabidopsis thaliana*. (*a*) Lane 1, protoplast fraction; lane 2, chloroplast fraction; lane 3, chloroplast stromal fraction (half the amount of protein loaded in lanes 1 and 2); lane 4, purified *E. coli* RecA protein. (*b*) Mitomycin C induction in the steady-state level of the chloroplast protein similar to *E. coli* RecA. This protein is slightly smaller in pea (39 kDa) than in *Arabidopsis*. Pea protoplasts were incubated for 12 hr in the presence of 0, 6, 15, or 30 μM mitomycin C (lanes 1–4, respectively), before protein isolation. (*c*) The same blot shown in *b* was reprobed with antiserum to the γ subunit of the chloroplast ATP synthetase, a nuclear-encoded chloroplast protein similar in size to RecA. This protein was not induced by DNA-damaging agents and served as a control for the proper loading of the lanes.

Homologous DNA in Higher Plants. Blot hybridization of restriction enzyme-digested genomic DNA revealed sequences homologous to the *Synechococcus recA* gene (23) in pea and *Arabidopsis* (Fig. 2). However, we were unable to detect any hybridization to purified chloroplast DNA (Fig. 2) or mitochondrial DNA (data not shown). Homology to *recA* has not been found in the completely sequenced chloroplast genomes of tobacco (32), *Marchantia* (33), or rice (7). Moreover, induction of the stromal protein was prevented by protein synthesis inhibitors acting on cytosolic (80S) ribosomes, also suggesting a nuclear localization for this gene (data not shown).

A *recA* Gene in *Arabidopsis thaliana*. Using the *Synechococcus recA* gene (23) as a probe, we screened an *Arabidopsis* cDNA library and cloned a gene showing extensive homology to eubacterial *recA*. The cloned cDNA was sequenced by standard techniques. Polymerase chain reaction analysis showed that this was the longest cDNA in the library encoding the RecA protein (data not shown). Although the cDNA is truncated at its 5′ end (Fig. 3*a*), it is long enough to reveal the features of the encoded protein. The amino terminus shows no similarity to bacterial RecA sequences and is probably a chloroplast transit peptide (Fig. 3*b*). Chloroplast transit peptides are not highly homologous, except for a loosely conserved motif at the cleavage site for the stromal processing protease (35, 36). The deduced amino acid sequence contains a perfect match to this consensus cleavage site (Fig. 3*a*). When such a match is found, it is predicted to specify the correct cleavage site with 90% probability (35). The sequence upstream of the putative cleavage site is enriched for serine and threonine (29.5%) and is almost devoid of acidic residues (2.0%). It also lacks predicted secondary structures except for two relatively small regions, one of them an amphiphilic β-strand next to the cleavage site (data not shown). These are typical features of chloroplast

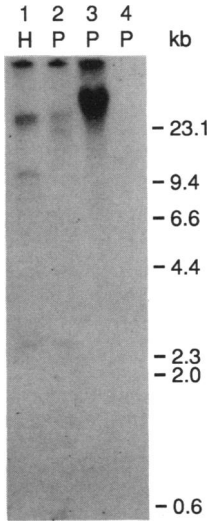

FIG. 2. Southern blot showing sequences related to *Synechococcus recA* in genomic DNA from *Arabidopsis* and pea. Lanes 1 and 2, total *Arabidopsis* DNA (1 μg); lane 3, total pea DNA (8 μg); lane 4, pea chloroplast DNA (1 μg). Restriction enzymes: H, *Hin*dIII; P, *Pst* I. Size markers are in kilobases (kb).

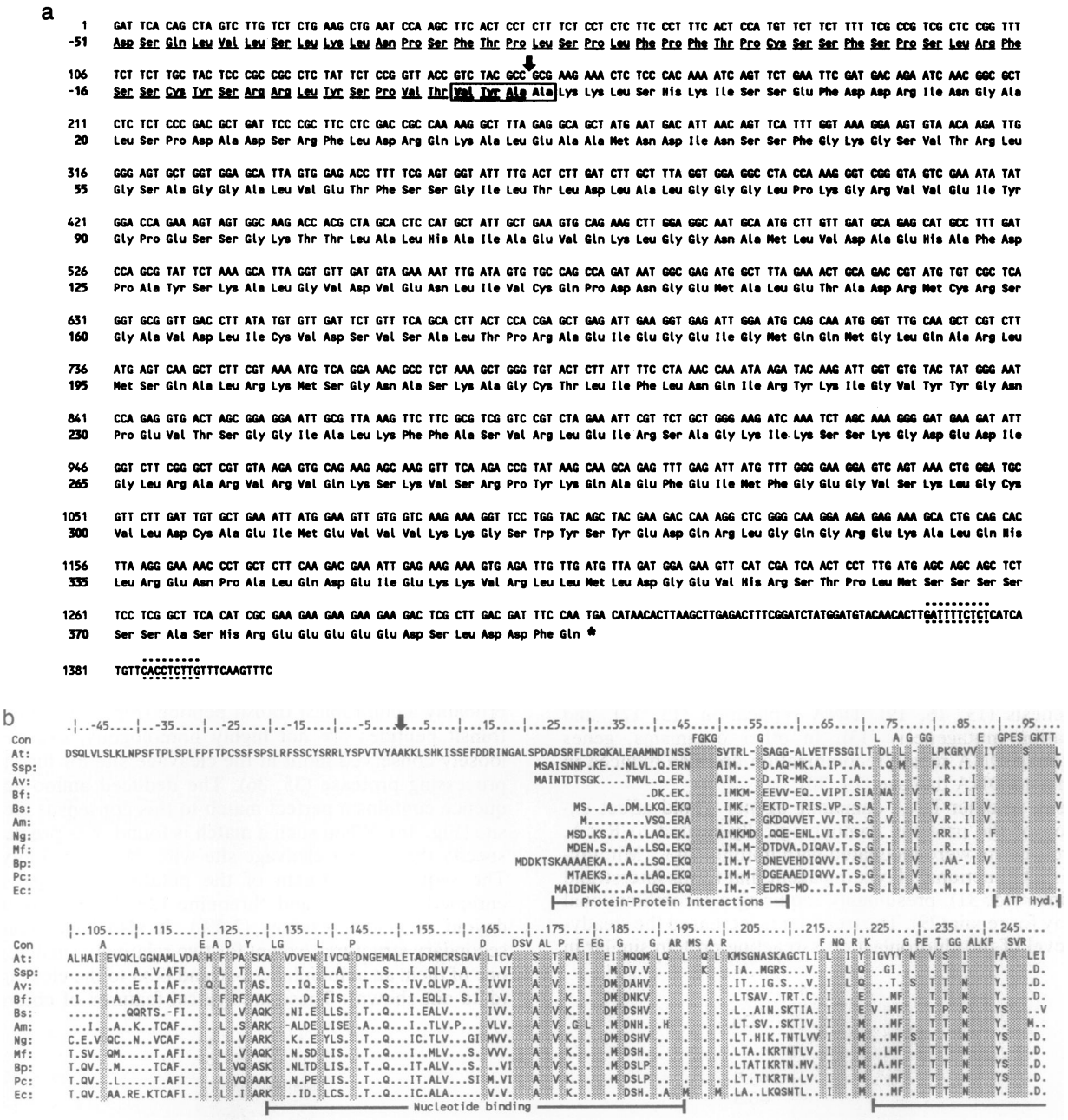

FIG. 3. Nucleotide and deduced amino acid sequences of a cDNA encoding the *A. thaliana* RecA protein. (*a*) The truncated cDNA contains a continuous open reading frame starting at its 5′ end. The stop codon (star) is followed by sequences with 75–80% homology to elements implicated in efficient polyadenylylation of plant mRNAs (34) (dotted lines). The amino acid sequence contains a putative chloroplast transit peptide (underlined residues) with a perfect consensus cleavage site (35) (box). Amino acids are numbered from the predicted start of the mature protein (arrow). (*b*) Comparison of the amino acid sequences of *Arabidopsis* RecA and several eubacterial homologs. Numbering of residues starts at the predicted cleavage site (arrow) for the chloroplast transit peptide in the *Arabidopsis* sequence. Invariant amino acids (row labeled Con) were determined by the alignment of 22 sequences (14, 23) (GenBank release 69 and this work) and are indicated by shaded areas. Dashes represent gaps introduced to optimize alignments. Dots indicate residues conserved with respect to the *Arabidopsis* sequence. Postulated functional domains in *E. coli* RecA (12) are shown below the sequence. Potential caveats on these assignments have been discussed (12, 14). At, *A. thaliana*; Ssp, *Synechococcus* sp.; Av, *Anabaena variabilis*; Bf, *Bacteroides fragilis*; Bs, *Bacillus subtilis*; Am, *Aquaspirillum magnetotacticum*; Ng, *Neisseria gonorrhoeae*; Mf, *Methylobacillus flagellatum*; Bp, *Bordetella pertussis*; Pc, *Pseudomonas cepacia*; Ec, *E. coli*.

transit peptides (35–37) and strongly support that identity for the first 51 amino acids of the truncated sequence.

The mature protein is predicted to be 387 amino acids long (Fig. 3*a*), with a calculated molecular mass of 41.8 kDa. This is close to the apparent molecular mass of the faster moving protein identified in the chloroplast stromal fraction (Fig. 1*a*). The amino acid sequence shows 60.7% overall identity with the *Synechococcus* protein and 52–57% identity with 20 other prokaryotic RecA proteins, 10 of which are shown in Fig. 3*b*. The amino and carboxyl termini are poorly conserved (Fig. 3*b*), although they may have functional significance (12, 14). Interestingly, the carboxyl end is enriched for acidic residues in almost all species analyzed (Fig. 3*b*). Although this sequence is more divergent than any of the eubacterial RecA proteins found to date, predicted functional domains of the *E. coli* protein (12, 14, 16, 18) are largely conserved (Fig. 3*b*). Amino acids known to cause recombination deficiency when altered in *E. coli* RecA (12, 14, 16, 18) are invariant in the *Arabidopsis* sequence. However, residues affecting preferentially co-protease activity (12, 14, 16, 18) and/or causing hyper-recombinogenic phenotypes (12, 14) are not so well conserved in the *Arabidopsis* gene. Since this gene is now located in the nucleus of a eukaryote, it is tempting to speculate that it has acquired a different regulatory system and it is able to evolve independently of LexA.

Table 1. Codon usage in *A. thaliana* nuclear genes encoding chloroplast proteins and several eubacterial *recA* genes

Genes*	CGN/AGR ratio†	Codon usage distance‡
EPSP synthase§	0.62	5.07
Tryptophan synthase	0.47	6.28
Acetolactate synthase	1.36	6.48
Cs gene	0.33	6.42
Rubisco activase	0.91	6.88
Average¶	0.74 ± 0.41	6.22 ± 0.68
Arabidopsis recA	1.56	6.87
Anabaena recA	5.00	10.78
Synechococcus recA	8.50	11.39
Escherichia recA	∞‖	15.82

By Spearman's rank correlation coefficient, the CGN/AGR ratio and codon usage distance are correlated with each other ($r = 0.88$; $P < 0.005$).

*In plant genes (first six listed), sequences corresponding to the mature proteins were used for the analysis (i.e., introns and chloroplast transit peptides were excluded). References for the *Arabidopsis* genes are listed in ref. 41. The Cs gene codes for a chloroplast protein of unknown function. The eubacterial *recA* genes were obtained from GenBank (release 69).

†Arginine coding ratio, where N = A, C, G, or T, and R = A or G (40).

‡Determined by a modification of a published procedure (38). Briefly, synonymous codons differing only in their third nucleotides were grouped. Termination codons and single codon groups (methionine and tryptophan) were excluded, leaving 21 codon groups with a total of 59 codons. The relative frequency of different codons in each group was calculated. The overall difference in codon usage between any gene and the *Arabidopsis* average was computed by a distance algorithm (38):

$$D(A, \overline{X}) = \sum_{i-1}^{i-59} |x(i, A) - x(i, \overline{X})|,$$

where $x(i, A)$ and $x(i, \overline{X})$ are the frequencies of the ith codon in gene A and in the average *Arabidopsis* nuclear gene encoding a chloroplast protein (determined by pooling the coding sequences of the five genes shown).

§5-Enolpyruvylshikimate-3-phosphate synthase.

¶The *Arabidopsis* average was calculated from the five individual genes shown (41) and is expressed as the mean ± SD. To avoid giving excessive weight to fluctuations in codon usage for rare amino acids, a problem with short proteins, only genes of similar length to *recA* were analyzed.

‖*E. coli* RecA lacks arginine residues encoded by AGR.

The plastid genome encodes only a small proportion of the proteins needed for functional chloroplasts (7, 32, 33), and it is thought that most genes have been transferred to the nucleus during evolution (11, 38). The base composition and codon usage of these transferred genes have adjusted to reflect their nuclear localization (38). It has been hypothesized that codon usage is genome-specific and provides a basis for species classification comparable to classical systematics (39). We applied several methods used to compare nonhomologous sequences (38–40), to determine the degree of similarity between the sequenced cDNA and several other *Arabidopsis* nuclear genes encoding chloroplast proteins. The dinucleotide frequency and base composition (TA/AT ratio, and %G + %C) were not significantly different between the genes examined (analyses not shown). However in codon usage distance and the arginine coding ratio, the cloned sequence was indistinguishable from the *Arabidopsis* genes and clearly different from several eubacterial *recA* genes (Table 1). These data suggest that the *Arabidopsis recA* gene has adjusted to reflect its localization in the nuclear genome, clearly diverging from even the more closely related cyanobacterial homologs.

DISCUSSION

Chloroplast DNA recombination has been studied extensively, particularly in *Chlamydomonas*, by genetic analysis (3, 4). There has also been much work suggesting the involvement of certain sequence elements in plastid DNA recombination (4, 6, 7). However, very little is known at the enzymatic level. Our finding of a plastid-localized RecA homolog provides biochemical evidence for a chloroplast recombination system and strongly supports its relationship to the eubacterial counterpart. To our knowledge, this is also the first observation of a *recA* homolog in a eukaryote. While recombination/repair enzymes have been identified in various other eukaryotes (42–46), they have structures and enzymatic characteristics that differ from those of the bacterial RecA protein.

In view of the known roles for RecA in *E. coli*, and induction of the pea enzyme by DNA damage, it is likely that the chloroplast enzyme is also concerned with DNA repair. We have not yet determined whether the *Arabidopsis* gene can complement a *recA*-deficient strain of *E. coli*. However, by using a complementation assay, Pang *et al.* (ref. 47, next paper in this issue) isolated a gene from the same *Arabidopsis* cDNA library, and found that it could increase the survival of a mutant (*phr*$^-$, *uvrB*$^-$, *recA*$^-$) *E. coli* strain exposed to UV light. Its DNA sequence is quite different from the one isolated by hybridization with a *recA* probe, however. In view of the complex pathways for DNA repair in bacteria (18, 19, 20, 48), it is likely that a number of proteins will be needed to interact with RecA in chloroplasts as well. Further work is needed to define fully the enzymology of DNA repair and recombination in chloroplasts.

We thank J. W. Roberts for antibodies against *E. coli* RecA, D. A. Bryant for the *Synechococcus recA* clone, J. T. Mulligan for the *Arabidopsis* cDNA library, and R. L. Last for helpful advice. H.C. was supported by a predoctoral fellowship and P.G. by a postdoctoral fellowship from the Cornell National Science Foundation Plant Science Center. M.O. was supported by a fellowship from the Afgrad Foundation. This work was supported in part by a Hatch grant to A.T.J.

1. Medgyesy, P., Fejes, E. & Maliga, P. (1985) *Proc. Natl. Acad. Sci. USA* **82**, 6960–6964.
2. Gillham, N. W. (1978) *Organelle Heredity* (Raven, New York), pp. 397–422.
3. Lemieux, B., Turmel, M. & Lemieux, C. (1990) *Theor. Appl. Genet.* **79**, 17–27.

4. Boynton, J. E., Gillham, N. W., Newman, S. M., & Harris, E. H. (1991) in *Cell Organelles: Advances in Plant Gene Research*, ed. Herrmann, R. (Springer, Vienna), Vol 6, pp. 1–48.
5. Kindle, K. L., Richards, K. L. & Stern, D. B. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 1721–1725.
6. Howe, C. J., Barker, R. F., Bowman, C. M. & Dyer, T. A. (1988) *Curr. Genet.* **13**, 343–349.
7. Hiratsuka, J., Shimada, H., Whittier, R., Ishibashi, T., Sakamoto, M., Mori, M., Kondo, C., Honji, Y., Sun, C.-R., Meng, B.-Y., Li, Y.-Q., Kanno, A., Nishizawa, Y., Hirai, A., Shinozaki, K. & Sugiura, M. (1989) *Mol. Gen. Genet.* **217**, 185–194.
8. Palmer, J. D. (1985) *Annu. Rev. Genet.* **19**, 325–354.
9. Palmer, J. D. (1990) *Trends Genet.* **6**, 115–120.
10. Cedergren, R., Gray, M. W., Abel, Y. & Sankoff, D. (1988) *J. Mol. Evol.* **28**, 98–112.
11. Gantt, J. S., Baldauf, S. L., Calie, P. J., Weeden, N. F. & Palmer, J. D. (1991) *EMBO J.* **10**, 3073–3078.
12. Kowalczykowski, S. C. (1991) *Biochimie* **73**, 289–304.
13. Echols, H. & Goodman, M. F. (1990) *Mutat. Res.* **236**, 301–311.
14. Roca, A. I. & Cox, M. M. (1990) *CRC Crit. Rev. Biochem. Mol. Biol.* **25**, 415–456.
15. Sassanfar, M. & Roberts, J. W. (1990) *J. Mol. Biol.* **212**, 79–96.
16. Dutreix, M., Moreau, P. L., Bailone, A., Galibert, F., Battista, J. W., Walker, G. C. & Devoret, R. (1989) *J. Bacteriol.* **171**, 2415–2423.
17. Witkin, E. M., Roegner-Manisalco, V., Sweasy, J. B. & McCall, J. O. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 6805–6809.
18. Wang, W. B. & Tessman, E. S. (1986) *J. Bacteriol.* **168**, 901–910.
19. Walker, G. C. (1985) *Annu. Rev. Biochem.* **54**, 425–457.
20. Little, J. W. & Mount, D. W. (1982) *Cell* **29**, 11–22.
21. Nivison, H., Fish, L. E. & Jagendorf, A. T. (1986) *Methods Enzymol.* **118**, 282–295.
22. Sambrook, J., Fritsch, E. F. & Maniatis, T. (1989) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Lab., Cold Spring Harbor, NY).
23. Murphy, R. C., Gasparich, G. E., Bryant, D. R. & Porter, R. D. (1990) *J. Bacteriol.* **172**, 967–976.
24. Church, G. M. & Gilbert, W. (1984) *Proc. Natl. Acad. Sci. USA* **81**, 1991–1995.
25. Elledge, S. J., Mulligan, J. T., Ramer, S. W., Spottswood, M. & Davis, R. W. (1991) *Proc. Natl. Acad. Sci. USA* **88**, 1731–1735.
26. Csonka, L. N. & Clark, A. J. (1979) *Genetics* **93**, 321–343.
27. Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12**, 387–395.
28. Lehminger-Mertens, R. & Jacobsen, J. H. (1989) *In Vitro Cell Dev. Biol.* **25**, 571–574.
29. Bernstein, H., Byerly, H. C., Hopf, F. A. & Michod, R. E. (1985) *Science* **229**, 1277–1281.
30. Friedberg, E. C. (1991) *BioEssays* **13**, 295–302.
31. Tomasz, M., Lipman, R., Chowdary, D., Pawlak, L., Verdine, G. L. & Nakanishi, K. (1987) *Science* **235**, 1204–1208.
32. Shinozaki, K., Ohme, M., Tanaka, M., Wakasugi, T., Hayashida, N., Matsubayashi, T., Zaita, N., Chunwongse, J., Obokata, J., Yamaguchi-Shinozaki, K., Ohto, C., Torazawa, K., Meng, B. Y., Sugita, M., Deno, H., Kamogashira, T., Yamada, K., Kusuda, J., Takaiwa, F., Kato, A., Tohdoh, N., Shimada, H. & Sugiura, M. (1986) *EMBO J.* **5**, 2043–2049.
33. Ohyama, K., Fukuzawa, K., Kohchi, T., Shirai, H., Sano, T., Umesono, K., Shiki, Y., Takeuchi, M., Chang, Z., Aota, S. I., Inokuchi, H. & Ozeki, H. (1986) *Nature (London)* **322**, 572–574.
34. Mogen, B. D., MacDonald, M. H., Graybosch, R. & Hurt, A. G. (1990) *Plant Cell* **2**, 1261–1272.
35. Gavel, Y. & Von Heijne, G. (1990) *FEBS Lett.* **261**, 455–458.
36. Von Heijne, G., Steppuhn, J. & Herrmann, R. G. (1989) *Eur. J. Biochem.* **180**, 535–545.
37. Von Heijne, G. & Nishikawa, K. (1991) *FEBS Lett.* **278**, 1–3.
38. Oliver, J. L., Marín, A. & Martínez-Zapater, J. M. (1990) *Nucleic Acids Res.* **18**, 65–73.
39. Grantham, R., Gautier, C. & Gouy, M. (1980) *Nucleic Acids Res.* **8**, 1893–1912.
40. Grantham, R. (1978) *FEBS Lett.* **95**, 1–11.
41. Von Heijne, G., Hirai, T., Klösgen, R. B., Steppuhn, J., Bruce, B., Keegstra, K. & Herrmann, R. (1991) *Plant Mol. Biol. Reporter* **9**, 104–126.
42. Moore, S. P. & Fishel, R. (1990) *J. Biol. Chem.* **265**, 11108–11117.
43. Kmiec, E. B. & Holloman, W. K. (1982) *Cell* **29**, 367–374.
44. Kolodner, R., Evans, D. H. & Morrison, P. T. (1987) *Proc. Natl. Acad. Sci. USA* **84**, 5560–5564.
45. Sugino, A., Nitiss, J. & Resnick, M. A. (1988) *Proc. Natl. Acad. Sci. USA* **85**, 3683–3687.
46. Lowenhaupt, K., Sander, M., Hauser, C. & Rich, A. (1989) *J. Biol. Chem.* **264**, 20568–20575.
47. Pang, Q., Hays, J. B. & Rajagopal, I. (1992) *Proc. Natl. Acad. Sci. USA* **89**, 8073–8077.
48. Smith, G. R. (1987) *Annu. Rev. Genet.* **21**, 179–201.