Proc. Natl. Acad. Sci. USA
Vol. 83, pp. 599–603, February 1986
Biochemistry

Total synthesis of a gene for bovine rhodopsin

(oligonucleotide synthesis/polynucleotide kinase/DNA ligase/site-specific mutagenesis/restriction sites)

LUCA FERRETTI, SADASHIVA S. KARNIK, H. GOBIND KHORANA, MICHAEL NASSAL, AND DANIEL D. OPRIAN

Departments of Biology and Chemistry, Massachusetts Institute of Technology, 77 Massachusetts Avenue, Cambridge, MA 02139

Contributed by H. Gobind Khorana, September 18, 1985

ABSTRACT **To carry out systematic structure–function studies of bovine rhodopsin by specific amino acid replacements, we have accomplished the total synthesis of its gene, which is 1057 base pairs long. The synthetic gene contains 28 unique restriction sites that are on the average 60 base pairs apart. Replacement of specific restriction fragments by synthetic counterparts containing the desired nucleotide changes permits specific mutagenesis in all parts of the gene. The synthesis of the gene involved enzymatic joining of a total of 72 synthetic oligonucleotides, 15–40 nucleotides long, to form DNA duplexes. The total gene was assembled from three synthetic fragments that were cloned. All synthetic oligonucleotides were characterized by 5′-end analysis, and the accuracy of the joining reactions was confirmed by sequencing the three fragments as well as the complete gene.**

Rhodopsin is the photoreceptor protein of vertebrate retinal rod cells. The amino acid sequence for bovine rhodopsin, a polypeptide chain of 348 amino acids, has been determined (1–3). 11-*cis*-Retinal linked as a Schiff base to the ε-amino group of a lysine residue serves as the chromophore. Absorption of light results in isomerization of 11-*cis*-retinal to the all-*trans* form. This is believed to cause a conformational change in the protein that results in the activation of retinal GTPase and, ultimately, in the hyperpolarization of the rod cell. In addition, bleached rhodopsin undergoes phosphorylation by rhodopsin kinase and interacts with other proteins—interactions that must be involved in light and dark adaptation.

We wish to investigate the structure–function relationships of rhodopsin by site-specific mutagenesis. This will be accomplished by replacement of restriction fragments with synthetic duplexes containing the desired nucleotide changes (4). To facilitate this approach, we have carried out the total synthesis of a gene for bovine rhodopsin that has been designed to contain frequently occurring and suitably placed restriction sites.

Design of the Synthetic Gene

The synthetic gene contains 28 unique restriction sites distributed throughout the sequence that define target fragments for mutagenesis that are, in general, 60 base pairs (bp) in length (Fig. 1). These restriction sites originated as follows: (*i*) by retaining a few of the unique sites present in the native gene, (*ii*) in the case of multiple sites, all but one of the repeated sequences were destroyed to make the remaining site unique, and (*iii*) new unique sites were created. Unique sites were placed at each end of the gene to enable easy transfer from one vector to another. An *Eco*RI site was added at the 5′ end of the gene and a *Bam*HI site at the 3′ end, because these two sites are frequently used in plasmid and bacteriophage vectors. If necessary, both sites can be changed easily by the synthesis of only two oligonucleotides, at the 3′ end by replacement of *Nar* I/*Bam*HI or *Bst*XI/*Bam*HI fragments, and at the 5′ end by replacement of an *Eco*RI/*Kpn* I fragment.

In total, 52 nucleotide changes were made in the synthetic gene relative to the natural gene sequence (3). Thirty-six of these were required for the creation of the 26 unique restriction sites. Twelve changes were made to eliminate stretches of four or more guanosines or cytidines to avoid possible difficulties in chemical synthesis of oligonucleotides or DNA sequence analysis of the gene. However, one sequence of four cytidines was retained to create the unique *Sfi* I restriction site (Fig. 1). While *Escherichia coli* is a likely host for expression of the synthetic gene, no attempt was made to optimize the sequence for codon usage except for two codons that are very infrequently used in this organism (5). The UUG codon appearing for both leucine-119 and -128 was changed to CUG, and the AUA codon for isoleucine-219 was changed to AUC. Finally, one change was introduced at position 33 of the nucleotide sequence to eliminate an *Eco*K restriction site near the 5′ end of the gene. This site was removed because it is recognized by the restriction-modification system of *E. coli* K-12 strains.

The gene was divided into three fragments of approximately equal length (Fig. 1). Each fragment had unique restriction sequences at the cohesive ends. Fragment EX, bounded by *Eco*RI and *Xho* I sites, was 335 base pairs in length and was constructed from 32 oligonucleotides. Fragments XP and PB were 364 and 345 base pairs in length, respectively, and each was constructed with 20 oligomers. These fragments were the first synthetic products to be cloned as intermediates in the total synthesis.

For synthesis of single-stranded oligonucleotides and their joining to form the DNA duplexes, the general principles developed earlier (6) were followed. Thus, the length of the protruding cohesive ends in the complementary oligonucleotides was four to five residues so that the remainder of the chains formed duplexes with the complementary oligonucleotides. No cohesive end in any of the single-stranded oligonucleotides was self-complementary, and oligonucleotides with similar or identical cohesive ends were not used in any one ligation mixture. Thus, a critical general feature of the present strategy was that the sites designed for the joining of the single-stranded oligonucleotides and DNA duplexes were different from the restriction sites created in the gene.

In every oligonucleotide synthesized, the 5′-terminal nucleotide was different from that at the penultimate position. The rationale for this is as follows. Our method of purification separates according to size. Therefore, the most likely impurity in the full-length product, n, would be an $n-1$ homolog. Since, in our method of synthesis, chain elongation proceeds in the 3′ to 5′ direction, the $n-1$ homolog would lack the 5′-terminal nucleotide. Determination of the 5′ end would, therefore, detect the presence of contaminating

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.

Abbreviations: DMT, dimethoxytrityl; bp, base pair(s).

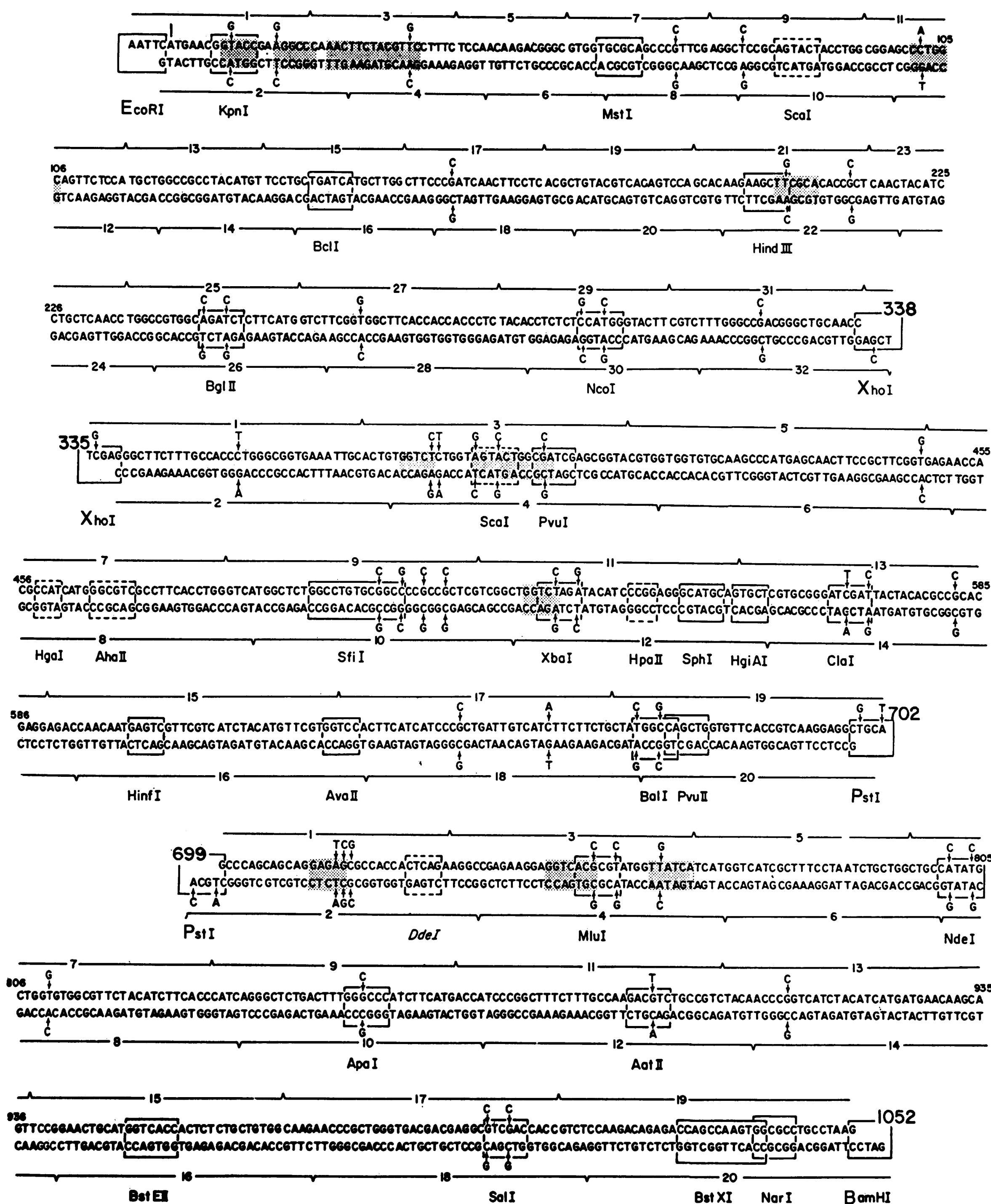


FIG. 1. Plan for the total synthesis of a gene for bovine rhodopsin. The gene is divided into three fragments: fragment EX, nucleotides −5 to 338; fragment XP, nucleotides 335 to 702; and fragment PB, nucleotides 699 to 1052. The oligonucleotides to be synthesized for each fragment are numbered and shown between carets, placed above and below the nucleotide sequences. The vertical arrows indicate nucleotides that were changed from the native gene sequence. The unique restriction sites in the synthetic gene are shown boxed in solid line. The restriction sites unique within a fragment, but not unique in the gene, are shown boxed in dashed line. Restriction sites removed from the native gene are shown in shaded boxes. The *Dde* I site in fragment PB is distinguished by italics, because there is a second *Dde* I site that overlaps the unique *Bam*HI site. However, the second *Dde* I site is not recognized if the gene fragment has been digested with *Bam*HI.

$n - 1$ impurity in the oligomer preparations, provided that the 5′-terminal and -penultimate residues were different.

MATERIALS AND METHODS

Materials. *N,N*-diisopropylphosphoramidites were obtained from Applied Biosystems. Controlled pore glass and long-chain alkylamine controlled pore glass were from Pierce or Electro-Nucleonics (Oak Ridge, TN). Dichloroacetic acid (Fisher) was vacuum distilled. All other reagents were purified according to standard protocols (7) or as described (4). Restriction endonucleases were purchased from New England Biolabs or Boehringer Mannheim and used according to manufacturer's recommendations. T4 polynucleotide kinase was a product of Pharmacia P-L Biochemicals. [γ-^{32}P]ATP ($\approx$5000 Ci/mmol; 1 Ci = 37 GBq) was obtained from Amersham.

Attachment of 5′-*O*-Dimethoxytrityl-Protected Deoxynucle-

osides to Controlled Pore Glass. 5′-DMT-N^6-Bz-dA, 5′-DMT-N^4-Bz-dC, and 5′-DMT-N^2-isobutyryl-dG were prepared via the transient protection method (8, 9) and 5′-DMT-dT according to a standard procedure (10) (DMT, dimethoxyltrityl; Bz, benzoyl). The 3′-OH groups in the protected nucleosides were succinylated with succinic anhydride. The terminal carboxylic acid functions were activated (7) as *p*-nitrophenyl or *N*-hydroxysuccinimide esters and coupled to long-chain alkylamine controlled pore glass (11). For nitrophenyl esters, the coupling reaction was performed in the presence of 1-hydroxybenzotriazole. Unreacted amino groups were blocked by acetylation. The attachment of DMT-nucleosides was in the range of 22 to 30 μmol of nucleoside/g of glass.

A solid support with a shorter spacer was prepared by coupling 6-aminocaproic acid to 500 Å-controlled pore glass derivatized with 3-aminopropyltriethoxysilane (12). The nucleoside bound in this preparation was 35–40 μmol of nucleoside/g of glass.

Oligonucleotide Synthesis. Oligonucleotides were synthesized by the phosphite triester method with an Applied Biosystems Model 380A DNA synthesizer using a 10-fold excess of fully protected 2′-deoxynucleoside *N,N*-diisopropylphosphoramidites (13) and the equivalent of 1 μmol nucleoside bound to either of the above supports. DMT-protecting groups, including that of the terminal 5′-nucleotide, were removed by treatment with 3% (wt/vol) dichloroacetic acid in dichloromethane for varying lengths of time depending on the nucleoside carrying the DMT group. The times given were as follows: 80, 100, 125, and 150 sec at room temperature for adenosine, guanosine, thymidine, and cytidine, respectively. Deprotection and cleavage of the oligomers from the solid support was performed according to standard procedures (13). The repetitive yield, based on the final yield of purified product, was in the range of 94 to 96% per cycle with both types of solid support.

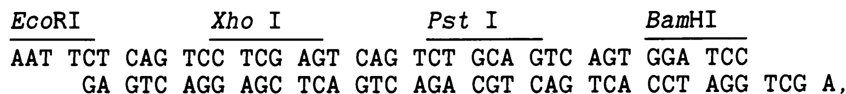
Purification of Oligonucleotides. The crude oligonucleotide mixture (25–40 A_{260} units, 10–15% of the total mixture), was applied directly to a polyacrylamide gel (0.3 × 20 × 40 cm; two 6-cm-wide slots per gel) for electrophoresis and then purified as described (4). The acrylamide concentration in the gels varied from 10–20% depending on the length of the oligomer.

5′-Phosphorylation and Characterization of Oligonucleotides. The oligonucleotides were phosphorylated at the 5′-end groups with T4 polynucleotide kinase using [γ-^{32}P]ATP (specific activity, 2.5 Ci/mmol), and isolated as described (14). The labeled oligonucleotides were checked for homogeneity on polyacrylamide gels under denaturing conditions.

The ^{32}P-labeled oligonucleotides were degraded with snake venom phosphodiesterase in 100 mM Tris·phosphate, pH 8.9, and the 5′ nucleotides were separated by chromatography on PEI-cellulose plates using 1 M LiCl as the solvent.

DNA Ligase-Catalyzed Joining of Oligonucleotides. T4 DNA ligase-catalyzed reactions were performed essentially as described (14). The reaction mixtures (50 μl) containing 4–10 oligonucleotides (2 μM each) at a time were overlaid with a thin layer of mineral oil and then brought to 100°C. The solution was allowed to cool to room temperature (≈2 hr), ligase was added to a concentration of 100 units/ml, and the mixture was incubated for 2 hr at room temperature. The desired joined duplex was purified by preparative polyacrylamide gel electrophoresis under nondenaturing conditions (15). Similar conditions were used for joining of the duplexes.

Preparation of Cloning Vectors M13mpR8 and 9. M13mpR8 was prepared by modifying M13mp8 (16) so as to delete several restriction sites that occur also in the synthetic rhodopsin gene. This was done by replacing the polylinker region between *Eco*RI and *Hin*dIII sites in M13mp8 with the synthetic duplex shown below,



which contained the following features. (*i*) The multiple cloning segment contained only the four restriction sites that mark the boundaries of the three fragments of the gene. (*ii*) The order of the sites in M13mpR8 is the same as in the synthetic gene. (*iii*) The reading frame of the β-galactosidase α-peptide is retained. (*iv*) Finally, the distance between any two cleavage sites is at least 5 bp. Insertion of the duplex to form M13mpR8 also eliminated the *Hin*dIII site. In the preparation of M13mpR9, where the duplex was inserted in the opposite orientation, the adenosine marked with an asterisk was changed to guanosine to destroy the chain-terminating codon.

Cloning of Fragments EX, XP, and PB. Procedures used for cloning of DNA were as described (4).

DNA Sequence Analysis. The dideoxy method (17) was used to sequence each fragment and the entire gene after cloning into M13mpR8 and 9 vectors. Sequencing of specific regions of the gene was facilitated by the use of appropriate primers from the pool of synthetic oligonucleotides. The universal primer is a 24-nucleotide long oligomer that hybridizes to a sequence near the polylinker region in M13mpR8 and 9 (the gift of K.-M. Lo).

RESULTS

Synthesis, Purification, and Characterization of Oligonucleotides. Oligodeoxynucleotides were synthesized. Following deprotection, the crude oligomer mixtures were subjected directly to electrophoresis on preparative denaturing polyacrylamide gels. The major products were the expected full length oligonucleotides, lower homologs being present only as minor bands. The products were extracted from excised gel slices, purified, and phosphorylated using T4 polynucleotide kinase.

Electrophoresis under denaturing conditions uniformly separated the full length oligonucleotides from those shorter by one nucleotide. Further characterization was carried out by analysis of the 5′-end ^{32}P-labeled nucleotides. Such analysis showed only the expected 5′-end residues (data not shown). On the basis of electrophoretic homogeneity and 5′-end analysis, the oligonucleotides were judged to be at least 95% pure.

T4 Ligase-Catalyzed Joining of 5′-Phosphorylated Oligonucleotides to Form Fragment PB. Initial experiments were carried out to determine the combinations of oligonucleotides for optimal joining reactions. Using four adjacent oligonucleotides at a time, reactions went in very high yields (>90%); however, in a few cases, e.g., EX 9–12 (oligonucleotides 9 through 12 of fragment EX) the yields were low (<50%). In the case of PB 13–16, a higher molecular weight product, probably hexamer, was formed. Based on these and further tests using combinations of six or more oligonucleotides, two alternative procedures were used for the synthesis of the fragment PB. In the first (Fig. 2 *a* and *b*), oligonucleotides PB 1–6, PB 7–14 and PB 15–20 were separately joined and the three resulting duplexes were purified by preparative gel electrophoresis. The full length fragment PB was then prepared by ligation of the three

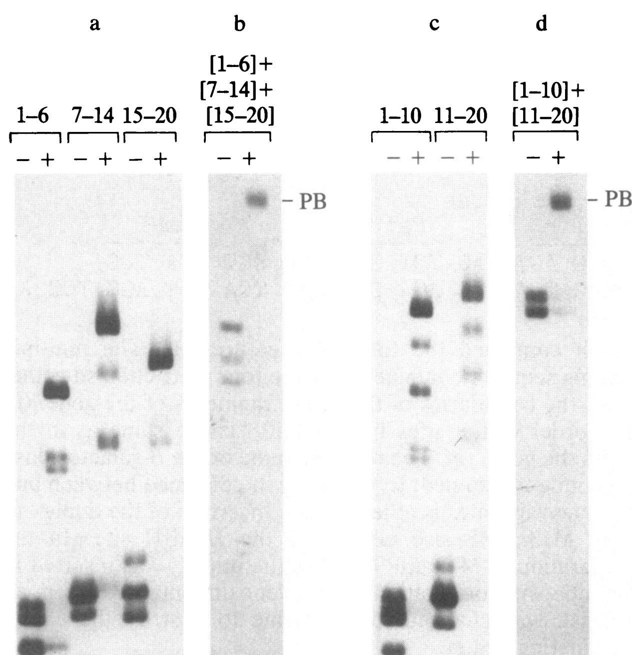


FIG. 2. Synthesis of fragment PB by T4 ligase-catalyzed reactions. (*a*) Ligation reactions of oligonucleotides PB 1–6, PB 7–14, and PB 15–20, without (−) and with (+) ligase. The products were purified and joined in the three component reaction (*b*) to form PB. (*c*) The joining of oligonucleotides PB 1–10 and PB 11–20 in one step reactions. (*d*) Joining of purified products, PB 1–10 and PB 11–20, to form fragment PB. The ligation products were separated by electrophoresis in nondenaturing 8% polyacrylamide gels and visualized by autoradiography.

duplexes. In the second method (Fig. 2 *c* and *d*) two duplexes were prepared from PB 1–10 and PB 11–20 in one step. The purified duplexes were then joined in a second step. The overall yield in the two different strategies was comparable. Yields from joining reactions in the more complex reaction mixtures, PB 1–10 and PB 11–20, were slightly lower. However, this was compensated for in the second step that involved only a two component ligation. The average yield at each ligation step was about 20% after gel purification of the product. Thus, fragment PB, which was assembled in two successive steps, was obtained in overall yield of about 4%. Fragments EX and XP, which were assembled in three ligation steps, were obtained in an overall yield of about 1%.

Cloning of Fragments EX, XP, and PB and Sequencing of the Cloned Fragments. Each fragment was cloned into the bacteriophage vectors M13mpR8 and 9 for amplification of the DNA. The nucleotide sequence of each fragment was determined at this stage by the dideoxy method using the single-stranded M13mpR8 and 9 recombinant phage DNA as templates and selected synthetic oligonucleotides as primers. The sequence analysis for fragment PB is shown in Fig. 3. The sequence of this fragment as well as those of EX and XP corresponded exactly to that expected from the synthesis.

Construction of the Rhodopsin Gene. The fragments were excised at the flanking restriction sites from the cloning vectors, and then purified by preparative agarose gel electrophoresis. The entire gene was synthesized in a four component ligation mixture, containing each of the three fragments, EX, XP, and PB, and *Eco*RI/*Bam*HI doubly digested M13mpR8, and cloned into *E. coli* strain JM103. The entire gene EB, and the three fragments EX, XP, and PB were purified from M13mpR8 vector DNA. Their mobilities on an agarose gel are shown in Fig. 4.

The nucleotide sequence of the entire gene was determined by using the universal primer for sequencing from single-stranded M13 templates, as well as many of the synthetic oligomers used in the synthesis of the gene fragments (Fig. 1).

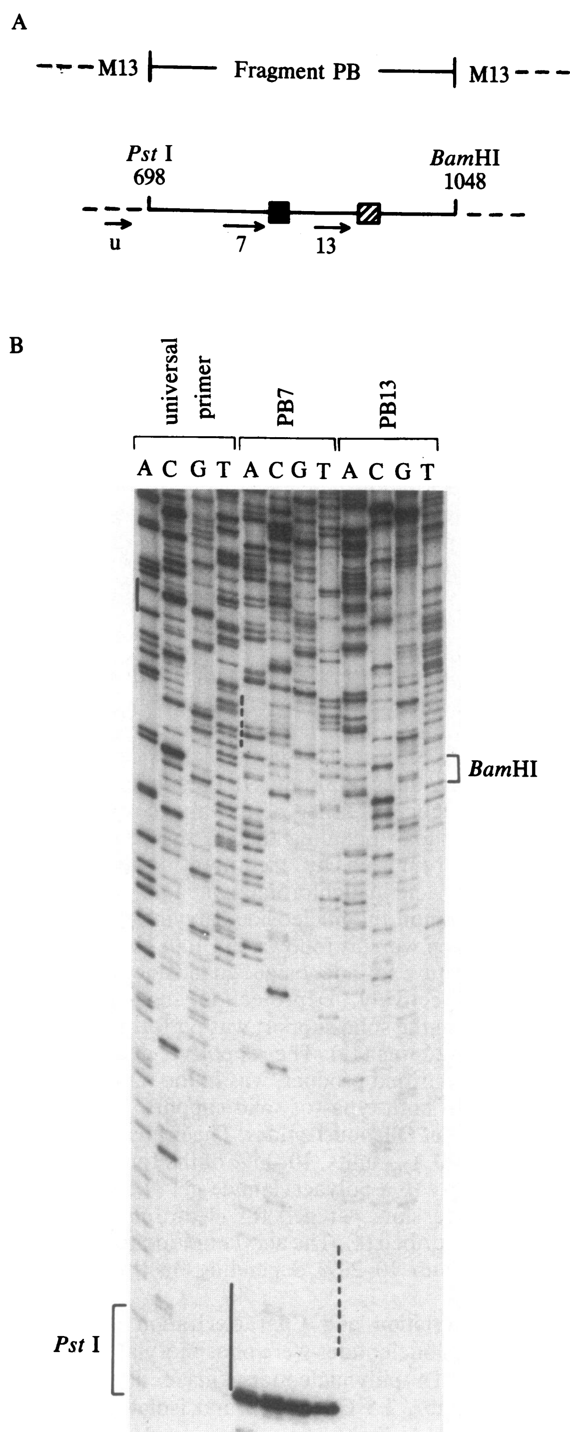


FIG. 3. DNA sequence analysis of fragment PB. (*A*) Schematic diagram of the recombinant M13mpR9 clone containing fragment PB. The three primers used for dideoxy sequencing, shown by the arrows, are as follows: u, universal primer; 7, PB 7; 13, PB 13. The two boxes indicate the regions of sequence overlap. (*B*) Autoradiogram of a 5% acrylamide/8.3 M urea gel used to determine the sequence of fragment PB as shown in *A*. The nucleotide sequences of the *Pst* I site at position 698 and *Bam*HI site at position 1048 have been indicated. The two regions of overlap shown in *A* are indicated by the solid and dashed vertical lines placed to the right of each sequencing lane. The solid block in *A* corresponds to the solid vertical line in *B*, and the hatched box in *A* corresponds to the dotted line in *B*.

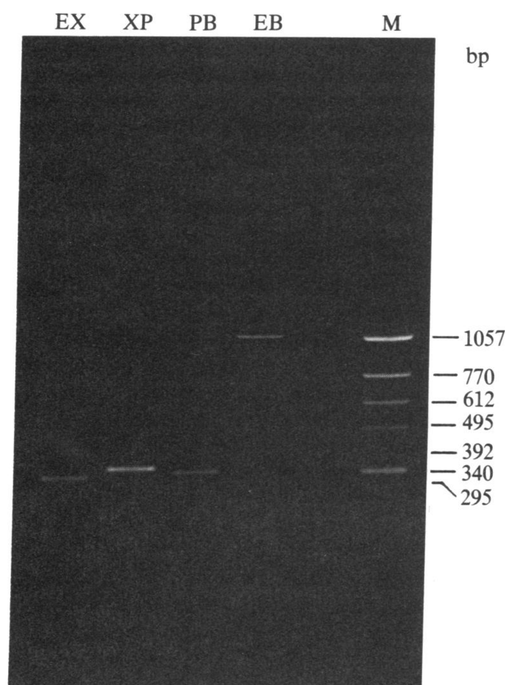


FIG. 4. Gel electrophoretic analysis of the cloned fragments EX, XP, PB, and the total synthetic gene EB. The DNA bands were visualized with ethidium bromide. Molecular size markers are *Hin*cII-digestion products of ϕX174 DNA. The expected size of the fragments is 335, 364, 345, and 1049 bp for EX, XP, PB, and EB, respectively.

The 1 kbp sequence corresponded exactly to that shown in Fig. 1.

DISCUSSION

A variety of methods has been introduced for site-specific mutagenesis. One method involves replacement of restriction fragments by synthetic counterparts containing the desired codon changes (4). While this method allows rapid and extensive mutagenesis, it is limited by the frequency and distribution of unique restriction sites in naturally occurring gene sequences. For this reason, we have designed and carried out the total chemical synthesis of a gene for rhodopsin. Unique restriction sites have been evenly placed throughout the gene such that mutations may be made by replacement of restriction fragments that average 60 bp or less in length. Such fragments would require the synthesis of no more than four oligonucleotides.

Gel electrophoresis was uniformly satisfactory for purification of oligonucleotides from contaminants that differed in size even by one nucleotide. Analysis of the 5′-end groups of the purified oligonucleotides confirmed that the syntheses had gone to completion in all cases. However, in one case we encountered an anomalous side product, an oligonucleotide with electrophoretic mobility of $n - 1$ that had the expected 5′ end but lacked the 3′-terminal nucleotide. While possible mechanisms for the formation of this side product will be discussed elsewhere (18), the example emphasizes the need for careful separation and characterization of the synthetic oligonucleotides.

Since the early work on tRNA genes (6), the total synthesis of genes has become increasingly facile and rapid. A major factor has been the rapidity and efficiency in the synthesis of deoxyribopolynucleotides of size much larger than those synthesized earlier. The size range of genes that fall within the scope of synthetic methodology will continue to increase and, consequently, the ability to carry out structure–function studies.

We thank Dr. U. L. Raj Bhandary for invaluable discussions and encouragement throughout the course of this work. This work was supported by Grants GM28289 and AI11479 from the National Institutes of Health, Grant PCM78-13713 from the National Science Foundation, and contract number N00014-82-K-0668 from the Office of Naval Research, Arlington, VA. D.D.O. was supported by National Institutes of Health Fellowship 5 F32EY05616-02. M.N. is a recipient of a European Molecular Biology Organization long-term fellowship.

1. Ovchinnikov, Y. A., Abdulaev, N. G., Feigina, M. Y., Artamonov, I. D., Zolotarev, A. S., Kostina, M. B., Bogachuk, A. S., Miroshnikov, A. I., Martinov, V. I. & Kudelin, A. B. (1982) *Bioorg. Khim.* **8,** 1011–1014.
2. Hargrave, P. A., McDowell, J. H., Curtis, D. R., Wang, J. K., Juszczak, E., Fong, S.-L., Mohana Rao, J. K. & Argos, P. (1983) *Biophys. Struct. Mech.* **9,** 235–244.
3. Nathans, J. & Hogness, D. S. (1983) *Cell* **34,** 807–814.
4. Lo, K.-M., Jones, S. S., Hackett, N. R. & Khorana, H. G. (1984) *Proc. Natl. Acad. Sci. USA* **81,** 2285–2289.
5. Grantham, R., Gautier, C., Gouy, M., Jacobzone, M. & Mercier, R. (1981) *Nucleic Acids Res.* **9,** r43–r74.
6. Khorana, H. G. (1979) *Science* **203,** 614–625.
7. Caruthers, M. H. (1982) in *Chemical and Enzymatic Synthesis of Gene Fragments*, eds. Gasser, H. G. & Lang, A. (Verlag Chemie, Weinheim, FRG), pp. 71–79.
8. Ti, G. S., Gaffney, B. L. & Jones, R. A. (1982) *J. Am. Chem. Soc.* **104,** 1316–1319.
9. McGee, D. P. C., Martin, J. C. & Webb, A. S. (1983) *Synthesis*, 540–541.
10. Schaller, H., Weimann, G., Lerch, B. & Khorana, H. G. (1963) *J. Am. Chem. Soc.* **85,** 3821–3827.
11. Adams, S. P., Kavka, K. S., Wykes, E. J., Holder, S. B. & Galluppi, G. R. (1983) *J. Am. Chem. Soc.* **105,** 661–663.
12. Matteucci, M. D. & Caruthers, M. H. (1981) *J. Am. Chem. Soc.* **103,** 3185–3191.
13. McBride, L. J. & Caruthers, M. H. (1983) *Tetrahedron Lett.* **24,** 245–248.
14. Sekiya, T., Bresmer, P., Takeya, T. & Khorana, H. G. (1976) *J. Biol. Chem.* **251,** 634–641.
15. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65,** 499–560.
16. Messing, J. & Vieira, J. (1982) *Gene* **19,** 269–276.
17. Sanger, F., Nicklen, S. & Coulson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74,** 5463–5467.
18. Oprian, D. D., Nassal, M., Ferretti, L., Karnik, S. S. & Khorana, H. G. (1986) in *Protein Engineering: Applications in Science, Medicine, and Industry*, eds. Inouye, M. & Sarma, R. (Academic, New York), in press.