Proc. Natl. Acad. Sci. USA
Vol. 78, No. 4, pp. 2169–2173, April 1981
Biochemistry

Procedure for production of hybrid genes and proteins and its use in assessing significance of amino acid differences in homologous tryptophan synthetase α polypeptides

(intragenic recombination/compatible plasmids/genetic exchanges/protein function)

WILLIAM P. SCHNEIDER, BRIAN P. NICHOLS*, AND CHARLES YANOFSKY

Department of Biological Sciences, Stanford University, Stanford, California 94305

Contributed by Charles Yanofsky, December 29, 1980

ABSTRACT **Hybrid tryptophan synthetase α and β polypeptides were produced by genetic recombination between the *trpB–trpA* regions of *Escherichia coli* and *Salmonella typhimurium* contained on compatible, multicopy plasmids. Intragenic recombination was decreased but still evident in *recA* cells. Genetic exchange occurred at many sites within *trpA*, but every recombinant gene produced a functional α polypeptide despite many amino acid differences from one or the other of the parental polypeptides. The five hybrid tryptophan synthetase α subunits examined resembled the parental polypeptides in catalytic function but differed in thermostability. The stability differences suggest that, as amino acid changes occurred in these proteins during the course of evolution, subsequent changes were limited to those that would allow retention of a desired protein conformation.**

The gene *trpA* of *Escherichia coli* and *Salmonella typhimurium* encodes the α polypeptide subunit of the $\alpha_2\beta_2$ tryptophan synthetase complex. Comparison of the nucleotide sequences of the *trpA*s of the two species reveals that there are 199 sequence differences in the 804 nucleotide pairs in this gene (1). The corresponding α polypeptides have 40 amino acid differences in their 268 residues. Thus, most of the nucleotide differences between the two *trpA*s are in synonymous codons. Despite the 40 amino acid differences between the two α subunits, the polypeptides are functionally interchangeable, both *in vitro* (2–4) and *in vivo* (3, 5). In this and comparable examples, the question arises as to whether the nucleotide and amino acid sequence differences that are seen in homologous genes and proteins are selectively favored in their respective organisms.

The functional significance of the amino acid differences in the two tryptophan synthetase α subunits can be assessed by constructing hybrid *trpA*s containing sequences from both species and examining the properties of the corresponding hybrid polypeptides. By using this approach, one class of hybrid tryptophan synthetase α polypeptides was isolated and studied (6). Unfortunately, such recombinants were extremely rare and were obtained only when *trpA*$^+$ recombinants were selected in crosses between *trpA* mutants of the two species. The single class of hybrid α subunits examined resembled the parental polypeptides in all the functional tests that were performed.

In this investigation, the difficulty of obtaining hybrid α subunits was overcome by incorporating inactive *trpB–trpA* segments of the two species into compatible multicopy plasmids and introducing both plasmids into bacteria with the *trpB–trpA* region deleted. The two plasmids were constructed so that the region of nucleotide sequence homology was confined to the *trpB–trpA* segment and so that genetic recombination in this segment could reconstitute intact *trpB* and *trpA*. Using this approach, we readily recovered hybrid *trpB*s and *trpA*s containing varying segments from the parental genes. It was also possible to modify the selection procedure slightly so that we could determine whether the hybrid *trpA*s obtained formed active or inactive α subunits. We also compared the properties of several of the hybrid α subunits with those of their parental polypeptides.

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.

METHODS

Media. Supplements to L broth or minimal medium were to the following final concentrations: glucose, 0.4%; acid-hydrolyzed casein, 0.5%; indole, 10 mg/liter; chloramphenicol, 20 mg/liter; ampicillin, 40 mg/liter.

DNA Manipulations. Plasmid DNA was prepared as described (7). Restriction endonucleases were either purchased from commercial suppliers (Bethesda Research Laboratories; New England BioLabs) or prepared in this laboratory by published procedures. *Pseudomonas* BAL 31 double-stranded DNA exonuclease was a gift from Horace Gray. BAL 31 reactions were performed as described (8). DNA ligation and bacterial transformation were carried out as described (7). DNA sequences were determined by the method of Maxam and Gilbert (9, 10) using the polyacrylamide/urea gel system of Sanger and Coulson (11).

The 3′ ends of the *Hin*dIII sites of DNA fragments were filled in by using 3 units of Klenow DNA polymerase I (New England BioLabs) in 10 mM Tris·HCl, pH 7.4/10 mM $MgCl_2$/10 mM 2-mercaptoethanol/50 mM NaCl containing dATP, TTP, dCTP, and dGTP at 0.025 mM each. The mixture was incubated at 20°C for 60 min.

Plasmid Construction. Plasmid pWS1 was constructed by mixing pACYC184 (12) and pBN24 (13) DNAs, digesting with *Bam*HI and *Sal* I, partially digesting with *Bgl* II, and ligating the mixture of fragments. The ligated DNA was used to transform *E. coli* W3110 Δ[*tonB trpBA17*] *his* to *trpB*$^+$ and chloramphenicol resistance (Cm^r) (Fig. 1*a*). Transformants contained a 7.8-kilobase (kb) plasmid composed of the 3.7-kb *Sal* I/*Bam*HI fragment of pACYC184 and the 4.1-kb *Bgl* II/*Sal* I fragment of pBN24 (containing *E. coli trp C*$^+$ *B*$^+$ *A*$^+$).

Plasmid pWS2 was constructed by first digesting pWS1 with *Hpa* I. The free ends of the linear plasmid DNA were digested with the double-strand exonuclease BAL 31. The linear DNA molecules were circularized by ligation and used to transform *E. coli* W3110 *trpA33* to Cm^r (Fig. 1*a*). Trp$^-$ colonies were de-

Abbreviations: kb, kilobase(s); Cm^r, chloramphenicol resistance; bp, base pair(s); Ap^r, ampicillin resistance; Tc^r, tetracycline resistance; In→Trp, indole + serine→tryptophan; InGP→Trp, indole glycerol phosphate→tryptophan; Ap^s, ampicillin sensitive.

* Present address: Department of Biological Sciences, University of Illinois, Chicago, IL 60680.

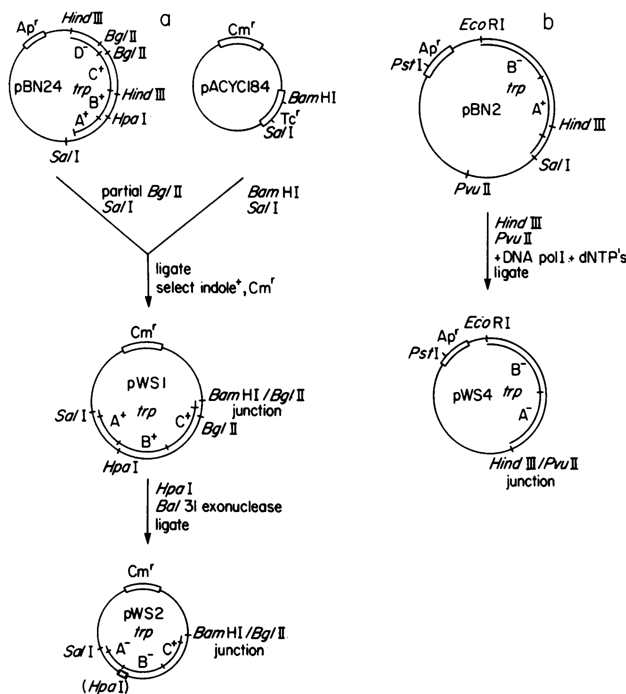

FIG. 1. **Construction of plasmids pWS1, pWS2, and pWS4. (*a*) Segments of pBN24 and pACYC184 were combined to generate pWS1 which was subsequently reduced to generate pWS2. (*b*) The *Hin*dIII/*Pvu* II segment of pBN2 was removed to generate pWS4.**

tected by testing for growth on minimal medium. Plasmid DNA was isolated from many Trp^- Cm^r colonies. The approximate length of the BAL-produced deletion was determined by restriction analysis. The *trpA* plasmid with the shortest deletion was selected. This plasmid, pWS2, was identical to pWS1 except for a deletion of 272 base pairs (bp) from *trpB* codon 335 to *trpA* codon 28, as determined by DNA sequence analysis.

Plasmid pWS4 was constructed by digesting pBN2 (1) with *Pvu* II and *Hin*dIII, filling in the staggered ends, and ligating the mixture of fragments (Fig. 1*b*). The resulting plasmid preparation was used to transform *E. coli* W3110 *trpR* Δ*trpEA2* to ampicillin resistance (Ap^r). Ap^r colonies contained a plasmid, designated pWS4, that was identical to pBN2 except that the *Hin*dIII/*Pvu* II segment was deleted. This deletion removed the last 69 codons of *trpA* of *S. typhimurium*. It also removed the tetracycline resistance (Tc^r) region that would be homologous with a corresponding region of pWS2.

Selection of Intragenic Recombinants. *E. coli* strain W3110 *trpR* Δ*trpEA2* was transformed with pWS2 and pWS4 to Cm^r and Ap^r. Twenty-four single colonies containing both plasmids were grown in separate 4-ml L-broth cultures containing ampicillin and chloramphenicol. A 0.1-ml sample of each culture was plated on minimal agar supplemented with glucose, acid-hydrolyzed casein, indole, and chloramphenicol to select for $trpB^+$ colonies. Colonies collected from each Petri plate were pooled. Pools were grown separately in liquid minimal medium supplemented with glucose, acid-hydrolyzed casein, indole, and chloramphenicol. Plasmid DNA was prepared from these cultures, digested with *Pst* I, and used to transform *E. coli* W3110 Δ[*tonBtrpAC4*] to $trpB^+$ and Cm^r. Transformants were screened for ampicillin sensitivity and for $trpA^{+/-}$ by replica plating to medium containing ampicillin and medium without indole, respectively.

Determination of Frequency of Recombination. Colonies derived from single cells containing both plasmids were grown overnight in L broth containing chloramphenicol and ampicillin. Serial dilutions were used to inoculate tubes containing 5 ml of L broth with chloramphenicol and ampicillin. Overnight cultures that had been inoculated with 20–100 cells were diluted to 50 ml and grown to stationary phase. Each culture was centrifuged at 2500 × *g* for 5 min, resuspended in 10 ml of minimal medium lacking glucose, and centrifuged again. The washed pellet was resuspended in 0.5 ml of minimal medium lacking glucose. The cell suspension was serially diluted and plated on L-broth agar containing chloramphenicol and ampicillin, on minimal medium supplemented with acid-hydrolyzed casein, indole, chloramphenicol, and ampicillin, and on minimal medium supplemented with acid-hydrolyzed casein, chloramphenicol, and ampicillin.

Enzyme Assays, Antibody Inhibition Tests, and Thermal Inactivation Tests. Strains selected for enzyme assays were grown overnight in 50 ml of minimal medium supplemented with glucose, acid-hydrolyzed casein, and chloramphenicol. Cultures were chilled and cells were collected by centrifugation. The cells were suspended in 4 ml of 0.1 M Tris·HCl (pH 7.8) and disrupted by sonication. Cellular debris was removed by centrifugation and the supernatant solutions were assayed for enzyme activity.

Tryptophan synthetase α subunit activity was assayed in the indole + *L*-serine→*L*-tryptophan (In→Trp) and the 3-indole glycerol phosphate + *L*-serine→*L*-tryptophan (InGP→Trp) reactions as described (14). One unit of enzyme corresponds to the synthesis of 0.1 μmol of tryptophan in 20 min at 37°C. An extract of *E. coli* W3110 Δ[*tonB trpA905*] was used as a source of tryptophan synthetase β_2 subunit for the assay of the α subunit. The 3-indole glycerol phosphate was generously supplied by Kasper Kirschner.

Antibody inhibition tests were performed as described (6) with the following modifications. Antibody inhibition of both In→Trp activity and InGP→Trp activity was determined. For the InGP→Trp reaction, 1.2 units of α subunit was incubated for 15 min at 0°C with sufficient antiserum to inhibit 0.5 unit of the α subunit activity of *E. coli*. For In→Trp, 2.5 units of α subunit was incubated for 15 min at 0°C with sufficient antiserum to inhibit 1.0 unit of the α subunit activity of *E. coli*. Antiserum to purified *E. coli* α subunit was kindly provided by Terrence Murphy and Stanley Mills.

Heat-inactivation studies with tryptophan synthetase α subunits were performed as described (6) with the following modifications. The *E. coli* and hybrid α subunits examined were in crude extracts. The *S. typhimurium* α subunit used was partially purified (15). One hundred α units and 100 β_2 units (InGP→Trp) were heated in 2 ml containing bovine serum albumin (4 mg of protein per ml). Remaining α subunit activity was determined in the InGP→Trp reaction by using fresh, unheated β_2 subunit.

RESULTS

Isolation of Recombinants. When cells of strain W3110 *trpR* Δ*trpEA2*/pWS2 + pWS4 were plated on minimal agar containing indole and chloramphenicol, $trpB^+$ Cm^r colonies appeared at a frequency of $8/10^7$ cells plated (Table 1). Because strains with either of these plasmids do not give $trpB^+$ progeny, the $trpB^+$ Cm^r colonies presumably contained a $trpB^+$ recombinant plasmid. A single crossover within the *trpB* segments of the two plasmids would produce an intact *trpB* gene (Fig. 2). The resulting plasmid would be dimeric unless a second crossover occurred in the only additional homologous region, *trpA*. In this case, the $trpB^+$ Cm^r plasmid would be monomeric and would be just slightly larger than pWS2. In effect, in such a double recombinant the deleted *trpB*–*trpA* junction region of *E. coli* would be replaced by the intact segment of *S. typhimurium*. In experiments of this type we could enrich for monomeric pWS2 recombinant plasmids by digesting plasmid DNA

Table 1. Recombination in strains with compatible plasmids

Plasmids	Strain	B^+/total	B^+A^+/total	B^+A^+/B^+
pWS2 + pWS4	w3110 *trpR* Δ*trpEA2*	8×10^{-7}	—	—
pWS2 + pWS4	W3110 *trpR* Δ*trpEA2 recA*	2×10^{-8}	—	—
pWS2 + pWS4	W3110 Δ[*tonB trpAC4*]	4×10^{-6}	3×10^{-7}	9×10^{-2}
pWS2 + pWS4	W3110 Δ[*tonB trpAC4*] *recA*	3×10^{-8}	3×10^{-8}	1.0
pBN18 + pWS2	W3110 *trpR* Δ*trpEA2*	1.0	—	—
pBN18 + pWS2	W3110 *trpR* Δ*trpEA2 recA*	6×10^{-4}	—	—

trpB$^+$ recombinants were detected by growth on indole-supplemented medium; *trpB*$^+$*A*$^+$ recombinants grew on unsupplemented minimal medium. Deletion *tonB trpAC4* removed *trpA*, *trpB*, and part of *trpC*, whereas deletion *trpEA2* deleted the entire operon. Plasmid pWS2 carried *trpC*$^+$. Thus, in bacteria with Δ[*tonB trpAC4*] and pWS2, the cells are *trpE*$^+$, *trpD*$^+$, and *trpC*$^+$. If plasmid recombination produces *trpB*$^+$ and *trpA*$^+$, the bacterium can grow in minimal medium.

from pooled *trpB*$^+$ Cm^r colonies with the restriction endonuclease *Pst* I. pWS2 lacks a *Pst* I site whereas dimeric plasmids would have the *Pst* I site of pWS4 and therefore would be cleaved to linear molecules by the endonuclease treatment. Because circular plasmids transform with greater efficiency than does linear DNA (16), a subsequent transformation with the pooled *Pst* I-digested plasmid DNA should increase the likelihood of obtaining *trpB*$^+$ Cm^r transformants containing monomeric plasmids.

By using this procedure and strain W3110 Δ[*tonB trpAC4*] as recipient, more than 99% of many *trpB*$^+$ Cm^r colonies obtained were ampicillin sensitive (Ap^s). Thus, almost all of the transformed cells probably contained plasmids that were monomeric or, at least, lacked the Ap^r region of pWS4. Approximately 1000 *trpB*$^+$ Cm^r Ap^s colonies in several experiments were replica plated to minimal agar to determine what fraction was *trpA*$^+$. Every colony grew on minimal medium, demonstrating that each *trpB*$^+$ plasmid was *trpA*$^+$ as well. It appears, therefore, that crossovers in the *trpA* regions of the two plasmids invariably generate hybrid *trpA*s that specify functional tryptophan synthetase α polypeptides.

Infrequently, *trpB*$^+$ *trpA* Cm^r Ap^r colonies were recovered. Restriction patterns of plasmid DNA obtained from such colonies did not resemble those of either of the parental plasmids, suggesting that plasmids of this type do not arise by simple genetic exchanges. These plasmids were not examined further.

Frequency of Recombination. When cells of W3110 Δ[*tonB trpAC4*]/pWS2 + pWS4 were plated on indole/chloramphenicol/ampicillin agar, *trpB*$^+$ Cm^r colonies appeared at a frequency of $1/2.5 \times 10^5$ cells plated (Table 1). When *trpB*$^+$ *trpA*$^+$ Cm^r colonies were selected by plating on minimal chloramphenicol/ampicillin agar, colonies appeared at 1/10th the frequency. These findings suggest that the plasmid single recombinant class (*trpB*$^+$) is 10 times more frequent than the double crossover class (*trpB*$^+$*trpA*$^+$). Recall, though, that in the experiment described in the preceding section we observed that every *trpB*$^+$ colony was also *trpA*$^+$. This apparent discrepancy is due to the difference between the procedures used. Undoubtedly, most dimeric plasmids generated by a single crossover in *trpB* subsequently experience a second crossover in *trpA*.

pWS2 and pWS4 were also introduced into *recA* strains to determine whether the *recA*$^+$ system was required for genetic exchange between the two plasmids (Table 1). Recombinant plasmids were still detected, although at 1/100th the frequency. In addition, because the *trpB*$^+$ and *trpB*$^+$*trpA*$^+$ frequencies were identical, it appeared that every *trpB*$^+$ recombinant was also *trpA*$^+$. The recombinant strains were UV sensitive, establishing that the *recA* allele was present. These findings suggest that a *recA*$^+$ independent pathway can generate plasmid recombinants.

We assessed the effect of the 15% nucleotide sequence mismatch between the *trpB*s of *E. coli* and *S. typhimurium* on the frequency of recombination in *trpB* by substituting plasmid pBN18 for pWS4. pBN18 contains the *trpB–trpA* segment of *E. coli* but lacks the initial portion of *trpB* (Fig. 2). Recombination between the *E. coli trpB* segments of pWS2 and pBN18 should give *trpB*$^+$ recombinants. When cells with these two plasmids were plated, every colony was *trpB*$^+$, indicating that recombination is extremely frequent within the identical *trpB* sequences. When a *recA* strain with the two plasmids was examined, 1 in 1600 cells was *trpB*$^+$. Thus, a *recA*$^+$ independent pathway also promotes recombination between identical sequences.

Determination of the Crossover Points in Recombinant Plasmids. The *Sau*3AI and *Taq* I restriction maps of *trpBA* of *E. coli* and *S. typhimurium* differ appreciably (Fig. 3). This allows us to distinguish between segments of *trpB* and *trpA* from either organism. Recombinant plasmids were treated with these two restriction enzymes and, on the basis of the patterns observed, we located the approximate crossover points in *trpB* and in *trpA* of 14 recombinant plasmids. Crossover points in *trpA* were located in all regions along the gene that are discernible by this technique. Six hybrids were chosen as representative of crossover points throughout *trpA*, and the exact sites of genetic exchange were determined by DNA sequence analyses. The sites of exchange in *trpA* are indicated in Figs. 3 and 4; the latter also presents the parental nucleotide and amino acid se-

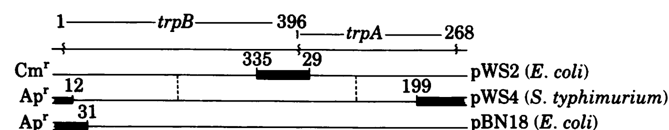

FIG. 2. *trpBA* regions in plasmids pWS2, pWS4, and pBN18. Deletions are indicated by thick bars. Codon numbers marking the ends of deletions are shown. *trpB* and *trpA* are 396 and 268 codons long, respectively. Dashed vertical lines represent possible sites of genetic recombination.

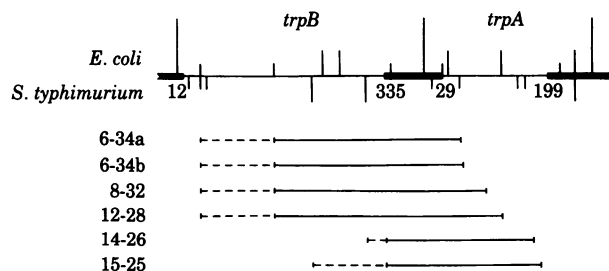

FIG. 3. *Sau*3AI and *Taq* I restriction sites in the *trpBA* regions of *E. coli* and *S. typhimurium* and the locations of crossover sites in the hybrids. Thick bars indicate deletions in pWS2 and pWS4. Short and long vertical lines denote *Sau*3AI and *Taq* I restriction sites, respectively, in the sequences of the two species. The six lines at the bottom indicate the segments of *trpBA* of *S. typhimurium* that are present in the hybrids. The dashed extensions represent the regions within which crossovers occurred in *trpB*. Crossover points in *trpA* of the hybrids are known precisely and are indicated by the right-hand ends of the solid lines.

```
                                    -20                                     -10
S. typhimurium          ASN LEU SER GLY ARG GLY ASP LYS ASP ILE PHE THR VAL HIS ASP ILE LEU LYS ALA ARG GLY GLU ILE
                        AAT CTC TCT GGC CGC GGA GAT AAA GAC ATC TTT ACC GTA CAC GAT ATC CTG AAA GCG CGA GGG GAA ATC TG
E. coli                   C   T   C   T                           C       T           T T         A

                                    10                                      20                                      30
MET GLU ARG TYR GLU ASN LEU PHE ALA GLN LEU ASN ASP ARG ARG GLU GLY ALA PHE VAL PRO PHE VAL THR LEU GLY ASP PRO GLY ILE
ATG GAA CGC TAC GAA AAT TTA TTT GCC CAA CTC AAC GAT CGC CGG GAA GGC GCT TTT GTC CCC TTC GTG ACC CTG GGC GAC CCT GGC ATT
                    TC  C G           G T G   G   G     AAA           A   C   T   T       C   G   C   T   T   G
                    SER                     LYS GLU     LYS
                                                                                                    - - - - - -|
                                    40                                      50                          pWS2        60
GLU GLN SER LEU LYS ILE ILE ASP THR LEU ILE ASP ALA GLY ALA ASP ALA LEU GLU LEU GLY VAL PRO PHE SER ASP PRO LEU ALA ASP
GAA CAG TCA CTG AAA ATT ATT GAC ACA CTG ATT GAT GCC GGC GCC GAC GCT CTG GAA CTG GGG GTT CCC TTC TCC GAT CCG CTG GCC GAT
  G         T             C   T   G   A       A       T   T       G       G T A   T A C               C   A       G
                                            GLU                                 ILE
                                                                                                            |_____|  |__
                                    70                                      80                              6-34a   90
GLY PRO THR ILE GLN ASN ALA ASN LEU ARG ALA PHE ALA ALA GLY VAL THR PRO ALA GLN CYS PHE GLU MET LEU ALA LEU ILE ARG GLU
GGC CCT ACC ATC CAG AAT GCG AAC TTA CGC GCC TTC GCC GCT GGC GTC ACG CCG GCT CAG TGT TTT GAA ATG CTG GCG CTG ATT CGT GAA
      G   G   T   A   C   C  CT C G           T   G   A   T   G   T       A   A                       A           C C G
                            THR                                                                                     GLN
______|
6-34b                               100                                     110                                     120
LYS HIS PRO THR ILE PRO ILE GLY LEU LEU MET TYR ALA ASN LEU VAL PHE ASN ASN GLY ILE ASP ALA PHE TYR ALA ARG CYS GLU GLN
AAA CAC CCG ACC ATT CCG ATT GGC CTG CTA ATG TAC GCG AAT CTG GTG TTC AAT AAC GGC ATA GAT GCG TTC TAT GCC CGT TGT GAA CAG
                      C             T G     T   C               T   C   A       T      A    T          AG   C     A A
                        |__________|                                    LYS             GLU             GLN         LYS
                           8-32     130                                     140                                     150
VAL GLY VAL ASP SER VAL LEU VAL ALA ASP VAL PRO VAL GLU GLU SER ALA PRO PHE ARG GLN ALA ALA LEU ARG HIS ASN ILE ALA PRO
GTT GGC GTA GAT TCC GTG CTG GTC GCA GAT GTC CCG GTT GAA GAA TCG GCC CCC TTC CGC CAG GCA GCG TTA CGG CAT AAT ATC GCG CCG
  C       C       G           T   C       G   A           G   C   G                   C       G   T   C   C G     A   T
            |_____|                                                                                         VAL
            12-28                   160                                     170                                     180
ILE PHE ILE CYS PRO PRO ASN ALA ASP ASP ASP LEU LEU ARG GLN VAL ALA SER TYR GLY ARG GLY TYR THR TYR LEU LEU SER ARG SER
ATC TTC ATC TGC CCG CCA AAT GCG GAT GAC GAT CTT CTG CGC CAG GTC GCA TCT TAC GGC CGC GGT TAC ACC TAC CTG CTT TCG CGT TCG
                              C           C   G             A A   C           T   T               T T     G   A   G A
                                                            ILE                     |____________|                  ALA
                                    190                                     200        14-26                        210
GLY VAL THR GLY ALA GLU ASN ARG GLY ALA LEU PRO LEU HIS HIS LEU ILE GLU LYS LEU LYS GLU TYR HIS ALA ALA PRO ALA LEU GLN
GGT GTC ACC GGC GCG GAA AAC CGT GGC GCA TTG CCG TTG CAT CAT CTC ATT GAG AAG CTT AAA GAG TAC CAT GCC GCG CCT GCG TTA CAG
  C   G           A           C  C    G   A   C C C A         G G    C        G             A C   T   A     C A   G
        |________|              ALA                 ASN         VAL ALA       |_ _ _ _      ASN             PRO
          15-25                     220                                     230 pWS4                                240
GLY PHE GLY ILE SER SER PRO GLU GLN VAL SER ALA ALA VAL ARG ALA GLY ALA ALA GLY ALA ILE SER GLY SER ALA ILE VAL LYS ILE
GGC TTC GGT ATC TCC TCG CCG GAA CAG GTG TCT GCG GCC GTG CGT GCC GGG GCG GCT GGC GCT ATC TCC GGC TCA GCC ATT GTC AAG ATT
  A   T       T     G C       T       A AAA   A   G A T GA    A   A   T   G       G   T   T   T   G           T   A   C
                    ALA     ASP         LYS         ILE ASP

                                    250                                     260                             268
ILE GLU LYS ASN LEU ALA SER PRO LYS GLN MET LEU ALA GLU LEU ARG SER PHE VAL SER ALA MET LYS ALA ALA SER ARG ALA END
ATC GAG AAA AAC CTC GCG TCT CCC AAA CAG ATG TTG GCG GAG CTC AGG TCC TTT GTC TCA GCC ATG AAA GCC GCC AGC CGC GCA TAA
        C   C T A T AAT GAG   A G G A A     C        CA   G  AA GTT       A CA  C G           G   G  CG     AGT
        GLN HIS ILE ASN GLU     GLU LYS             ALA     LYS VAL         GLN PRO             THR         SER
```

FIG. 4. Nucleotide and amino acid sequences of *S. typhimurium* and *E. coli trpA* and regions of recombination in hybrids. The complete sequences of *S. typhimurium* are given. Differences that exist in *E. coli* sequences are listed below. Solid lines mark areas within which crossovers occur. Dashed lines mark the end points of deletions in pWS2 and pWS4.

quences. Each recombinant plasmid is numbered to indicate how many of the 40 amino acid differences between the two parental α polypeptides are from each parental source. Thus, 6-34 indicates that the NH_2-terminal six amino acid differences are from *S. typhimurium* and the remaining 34 differences are those of the *E. coli* polypeptide.

Properties of Hybrid α Polypeptides. To determine whether the hybrid α polypeptides differed from the parental polypeptides in catalytic activity, the hybrid polypeptides were assayed in the In→Trp and InGP→Trp reactions. The former reaction measures the ability of the α chain to complex with and activate the β_2 subunit; the latter reaction measures the catalytic site of the α polypeptide as well as the activation of this site by the β_2 subunit. The ratio of the two activities was about 2 with the α polypeptide of either parent (Table 2). None of the hybrid α polypeptides exhibited greatly increased ratios. Thus, relative catalytic activity and complexing ability appear to be near normal. To establish that the enzymatic activity in each of the reactions was normal per unit of α polypeptide, antiserum to the α polypeptide of *E. coli* was used in activity inhibition tests. The

Table 2. Properties of parental and hybrid α polypeptides

Source of α subunit	Activity ratio, In→Trp / InGP→Trp	Inhibition by antiserum, % compared to *E. coli*	
		InGP→Trp	In→Trp
E. coli	2.2	100	100
Hybrid 6-34a	2.5	83	83
Hybrid 6-34b	2.2	72	87
Hybrid 8-32	2.8	60	74
Hybrid 12-28	3.0	44	70
Hybrid 14-26	2.6	56	66
Hybrid 15-25	3.1	46	70
S. typhimurium	2.0	50	76

FIG. 5. Thermal stability of parental and hybrid α polypeptides. ×, *E. coli* α; ○, hybrid 8-32; ●, *S. typhimurium* α; ■, hybrid 15-25; ▲, hybrids 6-34a, 6-24b, and 12-28; □, hybrid 14-26.

antiserum inhibited the hybrid α polypeptides about as effectively as it did the parental chains or gave intermediate values. The antiserum inhibited the *S. typhimurium* and hybrid polypeptides somewhat less well than the *E. coli* α polypeptide. These comparisons suggest that the hybrid α polypeptides are essentially indistinguishable from the parental polypeptides in catalytic activity.

The thermostability of the hybrid α polypeptides was compared with that of the parental α chains (Fig. 5). One hybrid polypeptide was intermediate in stability; the five others were more labile than either of the parental α polypeptides. The hybrid α polypeptide studied previously (6) had intermediate thermostability.

DISCUSSION

Production of hybrid genes by genetic recombination can be facilitated by using cells containing compatible multicopy plasmids carrying genes parental to the desired hybrids. The participating plasmids can be so constructed that their nucleotide sequence homology is limited to regions of interest, thus preventing undesired exchanges. In addition, the parental genes, by being resident in plasmids, can be easily modified by standard recombinant DNA techniques. And, as done in this investigation, procedures that allow enrichment for and screening of hybrid genes without selecting for the function of the hybrid gene product can be used.

Using compatible multicopy plasmids as described above, we have produced hybrid *trpB*s and hybrid *trpA*s by genetic recombination between modified, inactive forms of the two genes from *E. coli* and *S. typhimurium*. Genetic exchange occurs at reasonable frequencies by both *recA*$^+$-dependent and -independent mechanisms. The 15% mismatch between the *trpB* nucleotide sequences of the two organisms reduced the frequency of recombination at least 5 orders of magnitude compared with the frequency observed under similar conditions between identical *trpB* sequences.

Six different hybrid *trpA*s coding for five different α polypeptides were examined in detail. The crossover junctions in the hybrids were found to be located throughout the *E. coli* and *S. typhimurium* sequences within which genetic exchange could occur (Fig. 4). Although it would be reasonable to expect that the 66-bp region of identity (codons 167–188) in *trpB* of *E. coli* and *S. typhimurium* (13) would serve as a recombinational "hot spot," we note that, in the six hybrids examined here and in eight others examined by restriction analyses, none of the crossovers in *trpB* occurred within this perfectly conserved region. Sequence analysis of the crossover points in *trpA* indicates that exchange involves single phosphodiester switching rather than multiple switching. In addition, the pairing preceding exchange must be precise; otherwise, we would have observed frameshifted, and hence inactive, *trpA*s. Our results further suggest that the site of exchange need not involve extended regions of identical neighboring parental sequences.

Although *trpA* hybrids were not selected on the basis of *trpA*$^+$ function, nevertheless they all produced active α subunits. The five hybrid α polypeptides examined in detail were highly active catalytically and in complementation tests with the β_2 subunit, despite the fact that they differed from the parental α subunits at 6 to 34 amino acid positions. These findings suggest that, with respect to these two α subunit functions, none of the amino acid residues that are exchanged in the hybrid proteins plays a unique and essential role in its respective polypeptide. Nor do the exchanged residues appear to "balance" one another in some crucial manner in the maintenance of a functional active site. If the latter were the case, we would expect the replacement of one member of a balanced pair to impair the function of the protein.

Although the catalytic site appears to have been unaffected by the amino acid changes that occurred in the two α polypeptides since divergence of the two bacterial species, it seems that certain changes in each protein must have been balanced in order to retain a suitable overall conformation. This conclusion is based on the finding that several of the hybrid α polypeptides are significantly more thermolabile than either of the parental α subunits. It appears that, once an amino acid change became established, it placed constraints on what changes subsequently would be acceptable in maintaining a desired protein conformation.

The procedures we have developed probably can be used to generate hybrid genes and proteins by recombination between any two homologous, nonidentical genes. In addition, these methods should make it possible to force recombination between nonhomologous genes. This would allow the production of novel hybrid proteins that do not exist in nature but that have particularly desired characteristics.

These studies were supported by grants from the U.S. Public Health Service (GM09738), the National Science Foundation (PCM77-24333), and the American Heart Association (69C-15). W.P.S. is a Predoctoral Trainee of the U.S. Public Health Service, B.P.N. was a Postdoctoral Fellow of the American Cancer Society, and C.Y. is a Career Investigator of the American Heart Association.

1. Nichols, B. P. & Yanofsky, C. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 5244–5248.
2. Balbinder, E. (1964) *Biochem. Biophys. Res. Commun.* **17**, 770–774.
3. Creighton, T. E., Helinski, D. R., Somerville, R. L. &Yanofsky, C. (1966) *J. Bacteriol.* **91**, 1819–1826.
4. Murphy, T. M. & Mills, S. E. (1969) *J. Bacteriol.* **97**, 1310–1320.
5. Creighton, T. E. (1975) *J. Mol. Evol.* **4**, 121–137.
6. Yanofsky, C., Li, S. S.-L., Horn, V. & Rowe, J. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 286–290.
7. Gunsalus, R. P., Zurawski, G. & Yanofsky, C. (1979) *J. Bacteriol.* **140**, 106–133.
8. Legerski, R. J., Hodnett, J. L. & Gray, H. B., Jr. (1978) *Nucleic Acids Res.* **5**, 1445–1464.
9. Maxam, A. M. & Gilbert, W. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 560–564.
10. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
11. Sanger, R. & Coulson, A. R. (1978) *FEBS Lett.* **87**, 107–110.
12. Chang, A.C.Y. & Cohen, S. N. (1978) *J. Bacteriol.* **134**, 1141–1156.
13. Crawford, I. P., Nichols, B. P. & Yanofsky, C. (1980) *J. Mol. Biol.* **142**, 489–502.
14. Smith, O. H. & Yanofsky, C. (1962) *Methods Enzymol.* **5**, 794–806.
15. Creighton, T. E. & Yanofsky, C. (1970) *Methods Enzymol.* **17**, 365–380.
16. Cohen, S. N., Chang, A.C.Y., Boyer, H. W. & Helling, R. B. (1973) *Proc. Natl. Acad. Sci. USA* **70**, 3240–3244.