Rabies Virus Glycoprotein Analogs: Biosynthesis in *Escherichia coli*

Elizabeth Yelverton, Shirley Norton
John F. Obijeski, David V. Goeddel

Summary. The surface of rabies virus is composed of an approximately 60,000 dalton glycoprotein, in which most of the antigenic and immunogenic determinants of the virus reside. We have constructed plasmids for the direct expression in *Escherichia coli* of the mature full length rabies glycoprotein gene and also for the expression of a glycoprotein gene which has been truncated to exclude the coding region for a hydrophobic, possibly transmembrane, domain of the protein. *Escherichia coli* harboring the plasmids synthesize analog proteins which conform by several biochemical and antigenic criteria to rabies glycoprotein.

Rabies, the disease, is caused by a promiscuous neurotropic virus whose hosts include most warm-blooded animals and, of course, humans. Virus particles, normally transmitted by the bite of an infected animal, invade the nerve endings and travel via the spinal cord to the brain, where they multiply and infect efferent neurons, including those supplying the salivary glands. The disease has been greatly feared since antiquity because of associated bizarre behaviors in its victims, including convulsions, raging, drooling, and hydrophobia and also because once symptoms develop the disease nearly always proves fatal for humans (*1*). In 1885, Louis Pasteur developed the first rabies vaccine by emulsifying desiccated spinal cords of rabies-infected rabbits. The victim (patient) received a series of painful abdominal injections extracted from cords which had been air-dried for periods of 15 days decreasing to 5 days. Today, largely as a result of the efforts of workers at the Wistar Institute in Philadelphia, rabies virus can be cultured in vitro, concentrated, and inactivated to produce a more potent and more easily tolerated vaccine. The treatment regimen for exposed humans now includes injection with vaccine, the immunoglobulin fraction of antiserum to rabies virus, and sometimes interferon (*2*).

Despite these advances, the incidence of rabies worldwide is still high in hard-to-reach animal reservoirs, and a safe and effective synthetic vaccine is needed. In North America, mandatory vaccination programs have nearly eradicated canine rabies, but the disease is still widespread among the sylvatic population, particularly among skunks. In South America, the death of cattle infected by the bites of rabid vampire bats results in an estimated yearly loss of more than $29 million (*3*). Rabies in foxes is a problem in the northeastern United States and Canada, and an oral vaccine to immunize this species was developed by Baer and colleagues (*4*) in the 1970's. Since the 1940's, the incidence of fox rabies has been increasing in Europe, where a test program to vaccinate foxes orally by means of chicken head baits containing attenuated rabies virus has been instituted (*5*). Because of the possibility that an attenuated virus may revert to virulence, an oral virus-free vaccine would be preferable.

One alternative for virus subunit vaccine design is to use an acquired nucleotide sequence to deduce the amino acid sequence of an immunogenic viral protein, and to select from it short polypeptides which are then chemically synthesized, attached to carrier proteins of known immunogenicity, and injected in an immunopotentiating adjuvant. An important finding, illustrated in the cases of hepatitis virus and foot-and-mouth disease virus (*6*), is that a given linear array of amino acids can be an effective antigenic domain; the native three-dimensional structure of the intact protein is not required to evoke a spectrum of reactive and neutralizing antibodies in vivo.

Another method for producing subunit vaccines is biosynthesis in which the gene for a protein product is isolated and provided with new regulatory signals appropriate for its expression in a new host organism. The biosynthesis of polypeptide hormones such as insulin, growth hormone, and interferons (*7–9*) has been well documented, and many biosynthetically produced polypeptides will become commercially available in the next few years (*10*). Expression of small viral genes has also yielded to techniques of biosynthesis. The gene for the major surface antigen of hepatitis B virus has been directly expressed both in the yeast *Saccharomyces cerevisiae* (*11*) and in simian cells (*12*). Also, Kleid and coworkers have reported the synthesis in *Escherichia coli* of an immunogenic fusion protein between an antigen, VP_3, of foot-and-mouth disease virus and a protein derived from the *E. coli* pathway for tryptophan biosynthesis (*13*).

In this article we describe the direct expression in *E. coli* of the glycoprotein gene of rabies virus.

The rabies virus itself is an RNA virus of the Lyssavirus genus in the Rhabdoviridae family. Some rabies virus strains, such as CVS, HEP, ERA, and PM, have been developed by cultivation in laboratory. Since conventionally prepared, polyspecific antiserum does not differentiate these strains, rabies virus has been considered to consist of a single antigenic species. Correspondingly, all commercially available rabies vaccines now are produced from single strains. However, use of monoclonal antibodies, which provides higher resolution, can reveal some variations among the strains (*14*). Also, variations in the electrophoretic mobilities of some of the gene products have been observed (*15*). Five rabies viral proteins have been identified from rabies virus–infected cells and from purified rabies virions: L (large, the viral polymerase), N (nucleocapsid protein), G (glycoprotein), and M_1 and M_2 (matrix proteins) (*16*). Of these, the glycoprotein (or G protein) forms the external surface of the virus and is responsible for its immunogenicity; that is, antibodies raised against purified rabies virus glycoprotein can neutralize infectious rabies virus. In fact, Dietzschold *et al.* (*17*) showed that 9 nanograms of purified glycoprotein was equivalent to 1630 ng of purified whole virus in a mouse protection test even though the glycoprotein constitutes about 40 percent of the rabies

The authors are staff members in the Departments of Molecular Biology and Vaccine Development, Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, California 94080.

virion protein (*16*). The amino acid sequences of the glycoproteins of the rabies virus CVS and ERA strains have now been deduced from DNA copies of the respective genes (*18, 19*). The homology (about 90 percent at the amino acid level) is in conformity with the closely related antigenic characteristics of the two virus strains. We have undertaken the engineering of a rabies subunit vaccine consisting of the glycoprotein gene product from the CVS rabies virus strain to test against many strains of infectious rabies virus.

Structural Features of Rabies Virus Glycoprotein

From polyadenylated messenger RNA's (mRNA's) of cells infected with the rabies CVS strain, we prepared double-stranded complementary DNA (cDNA) species, introduced single-stranded homopolymeric tails, and annealed them by means of DNA base pairing into the bacterial cloning vehicle pBR322 (*20*). The heterogeneous hybrid plasmids were then transformed into *E. coli* to form a colony library. From the library we identified a plasmid, pRab91, which selectively hybridized to rabies virus glycoprotein mRNA (*18*). Nucleotide sequencing of the cloned DNA, which consisted of > 2000 base pairs (bp) showed that it encoded most of the hydrophobic signal peptide and all of the mature rabies virus glycoprotein, whose NH_2- and COOH-terminus has been established by direct amino acid analysis (*21*). The amino acid sequence predicted from the cDNA is shown in Fig. 1, where some features of the polypeptide are emphasized.

Although, on the whole, the rabies virus glycoprotein is quite hydrophobic, the signal peptide (denoted S_1-S_{19}) and the 22–amino acid (stippled) region between amino acid positions 439 and 462 are particularly hydrophobic. This latter domain has been postulated by Anilionis *et al.* (*19*) to constitute the portion of the protein which would span the viral lipid membrane. In that regard, it is satisfying to note that 9 of these predicted 20 amino acid positions differ between the glycoproteins of the CVS and ERA rabies strains, but none of the changes alter the hydrophobicity. Hypothetically, the native immunogenic domains of the glycoprotein would occur on the NH_2-terminal side of this transmembrane segment; the carboxyl-terminal portion would be the site of interaction with matrix protein. The many charged residues that follow the hydrophobic segment are indicated in Fig. 1 by underlining. Three sites of possible sugar chain attachment, Asn-X-Thr or Asn-X-Ser (*22*) which might occur on the virus exterior, are also shown; these are consistent in number with the most highly glycosylated form of the protein (*15*).

Expression Vectors for the Rabies Glycoprotein Gene

To effect the expression of the mature rabies virus glycoprotein gene in *E. coli*, it was necessary to (i) add a translation initiation signal before the codon for the amino-terminal amino acid, (ii) place the gene downstream from an *E. coli* promoter, or site of initiation of transcription by RNA polymerase, and (iii) make adjustments so that the transcript would be efficiently translated by the host machinery.

The method we used for the first step was similar to that used for the expression of human serum albumin and human fibroblast interferon (*23*). The synthetic deoxyoligonucleotide (*24*) primer 5′-dCATGAAGTTCCCCAT was chosen to incorporate an ATG translation initiation codon before the codon for the first amino acid of the mature rabies glycoprotein gene (AAG and Lys). The primer mismatches with the negative strand template in only two positions, and includes a deoxycytidine residue at the 5′

```
                                S1                                  S10                                 S19
                                Met Val Pro Gln Val Leu Leu Phe Val Leu Leu Leu Gly Phe Ser Leu Cys Phe Gly
Lys Phe Pro Ile Tyr Thr Ile Pro Asp Lys Leu Gly Pro Trp Ser Pro Ile Asp Ile His His Leu Arg Cys Pro
Asn Asn Leu Val Val Glu Asp Glu Gly Cys Thr[Asn Leu Ser]Gly Phe Ser Tyr Met Glu Leu Lys Val Gly Tyr    (50)
Ile Ser Ala Ile Lys Val Asn Gly Phe Thr Cys Thr Gly Val Val Thr Glu Ala Glu Thr Tyr Thr Asn Phe Val
Gly Tyr Val Thr Thr Thr Phe Lys Arg Lys His Phe Arg Pro Thr Pro Asp Ala Cys Arg Ala Ala Tyr Asn Trp   (100)
Lys Met Ala Gly Asp Pro Arg Tyr Glu Glu Ser Leu Gln Asn Pro Tyr Pro Asp Tyr His Trp Val Arg Thr Val
Arg Thr Thr Lys Glu Ser Leu Ile Ile Ile Ser Pro Ser Val Thr Asp Leu Asp Pro Tyr Asp Lys Ser Leu His   (150)
Ser Arg Val Phe Pro Ser Gly Lys Cys Ser Gly Ile Thr Val Ser Ser Thr Tyr Cys Ser Thr Asn His Asp Tyr
Thr Ile Trp Met Pro Glu Asn Pro Arg Pro Gly Thr Pro Cys Asp Ile Phe Thr Asn Ser Arg Gly Lys Arg Ala   (200)
Ser Asn Gly[Asn Lys Thr]Cys Gly Phe Val Asp Glu Arg Gly Leu Tyr Lys Ser Leu Lys Gly Ala Cys Arg Leu
Lys Leu Cys Gly Val Leu Gly Leu Arg Leu Met Asp Gly Thr Trp Val Ala Met Gln Thr Ser Asp Glu Thr Lys   (250)
Trp Cys Ser Pro Asp Gln Leu Val Asn Leu His Asp Phe Arg Ser Asp Glu Ile Glu His Leu Val Val Glu Glu
Leu Val Lys Lys Arg Glu Glu Cys Leu Asp Thr Leu Glu Ser Ile Met Thr Thr Lys Ser Val Ser Phe Arg Arg   (300)
Leu Ser His Leu Arg Lys Leu Val Pro Gly Phe Gly Lys Ala Tyr Thr Ile Phe[Asn Lys Thr]Leu Met Glu Ala
Asp Ala His Tyr Lys Ser Val Arg Thr Trp Asn Glu Ile Ile Pro Ser Lys Gly Cys Leu Lys Val Gly Gly Arg   (350)
Cys His Pro His Val Asn Gly Val Phe Phe Asn Gly Ile Ile Leu Gly Pro Asp Asp Arg Val Leu Ile Pro Glu
Met Gln Ser Ser Leu Leu Arg Gln His Met Glu Leu Leu Glu Ser Ser Val Ile Pro Leu Met His Pro Leu Ala   (400)
Asp Pro Ser Thr Val Phe Lys Glu Gly Asp Glu Ala Glu Asp Phe Val Glu Val His Leu Pro Asp Val Tyr Lys
Gln Ile Ser Gly Val Asp Leu Gly Leu Pro Asn Trp Gly Lys Tyr Val Leu Met Thr Ala Gly Ala Met Ile Gly   (450)
   ↑                                                                          ↑
Bgl II                                                                      Pst I
Leu Val Leu Ile Phe Ser Leu Met Thr Trp Cys Arg Arg Ala Asn Arg Pro Glu Ser Lys Gln Arg Ser Phe Gly
Gly Thr Gly Gly Asn Val Ser Val Thr Ser Gln Ser Gly Lys Val Ile Pro Ser Trp Glu Ser Tyr Lys Ser Gly   (500)
Gly Glu Ile Arg Leu End
```

Fig. 1. The deduced amino acid sequence of a rabies virus glycoprotein from the control virus standard (CVS) strain. The first amino acids of the mature protein are indicated as determined by NH_2-terminal analysis of the glycoprotein from another rabies virus strain (*21*). The stippled portion between amino acids 439 and 462 is a hydrophobic domain which may be the site of membrane attachment. Further on in this sequence, charged residues that may interact with capsid proteins are underlined. The amino acid sequence contains four potential glycosylation sites (Asn-X-Ser or Asn-X-Thr). Those three that would hypothetically occur on the external portion of the glycoprotein are indicated by boxes. Two restriction enzyme recognition sites, which occur in the cDNA and were used in the construction of truncated or fused rabies glycoprotein analogs, are shown.

position for later assembly to an Eco RI recognition site. The assembly for the mature glycoprotein gene is shown in Fig. 2. A Rsa I restriction fragment (136 bp) containing the coding region for the signal peptide cleavage site was denatured in the presence of the direct expression primer. *Escherichia coli* DNA polymerase I (large fragment) was used to synthesize the revised sense strand, remove the 3′ protruding end, and repair mismatches in the minus strand. A second restriction site within the repair template was used to generate a defined Ava II sticky end for ligation with the 1297 bp Ava II–Pst I fragment containing most of the rabies G coding region. The resultant 1337-bp fragment

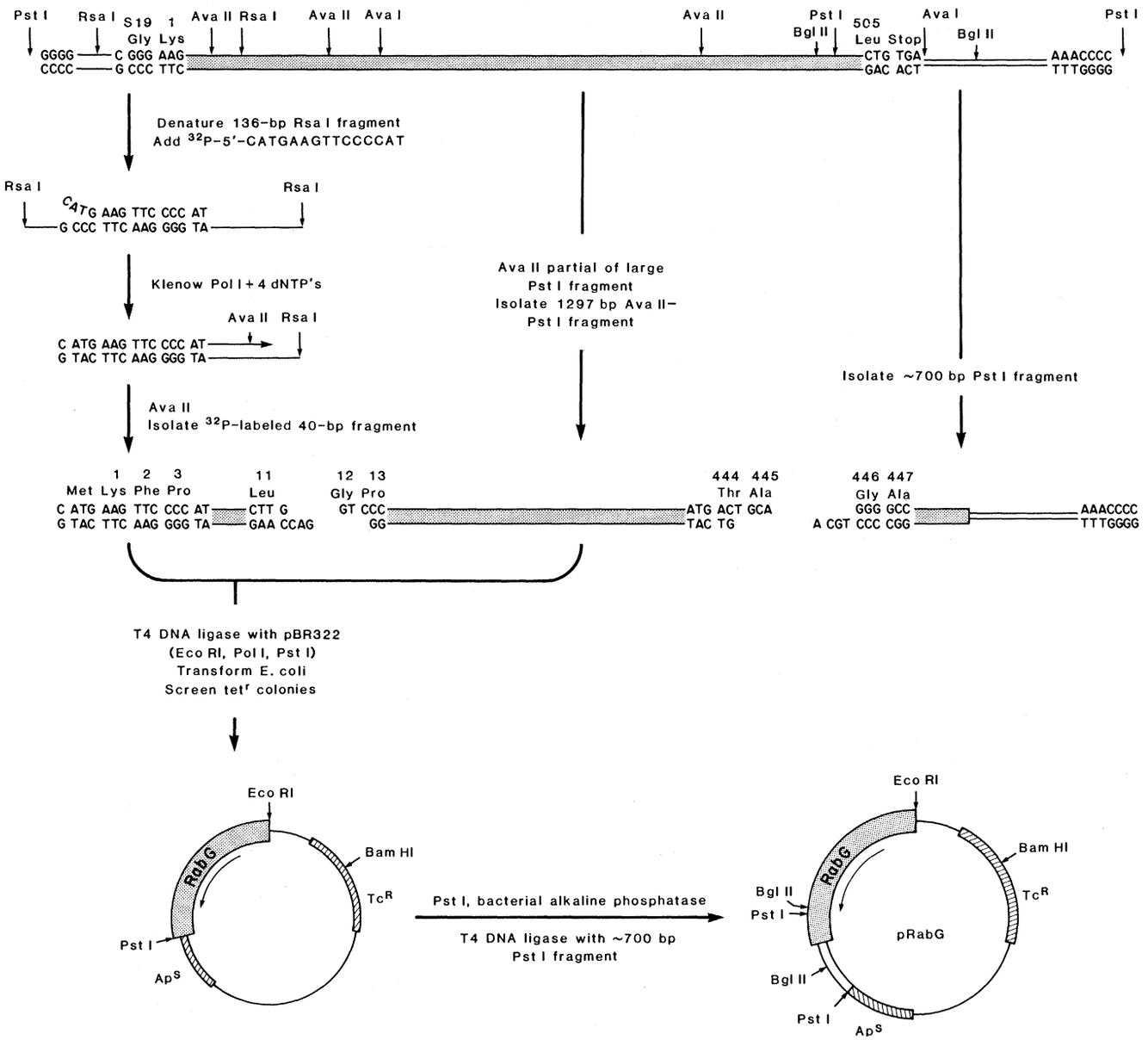

Fig. 2. Construction of a plasmid bearing the gene for mature rabies glycoprotein. The plasmid pRab91, a pBR322 derivative with rabies glycoprotein cDNA introduced by dG/dC pairing at the Pst I site (*18*) was the source of the template for a DNA polymerase I (Klenow fragment) catalyzed repair reaction. The synthetic deoxynucleotide dCATGAAGTTCCCCAT, synthesized by the modified phosphotriester method (*27*) was phosphorylated in a reaction containing 500 pmole of primer, ~ 300 μCi of [γ-^{32}P]ATP (2500 Ci/mmole, New England Nuclear) and 2 units of T4 polynucleotide kinase in 60 m*M* tris-HCl (*p*H 8), 10 m*M* $MgCl_2$, 15 m*M* β-mercaptoethanol. The reaction was allowed to proceed for 20 minutes at 37°C; unlabeled ATP was then added to a final concentration of 200 μ*M*. After 30 minutes the reaction was terminated by extraction with phenol and chloroform, and the aqueous phase was precipitated with ethanol with 3 μg of the 136-bp template Rsa I fragment. The mixture was dissolved in 50 μl of water, boiled for 5 minutes, quenched in dry ice, and diluted at 0°C to 100 μl to contain 100 m*M* tris-HCl (*p*H 7.5), 7 m*M* $MgCl_2$, 60 m*M* NaCl, and 0.5 m*M* of each deoxyribonucleotide triphosphate. Ten units of DNA polymerase I (Klenow fragment) (Boehringer-Mannheim) were added, and the reaction was allowed to proceed at room temperature for 4 hours. After phenol extraction and ethanol precipitation, the mixture was treated with 15 μl of Ava II (New England Biolabs) according to the manufacturer's recommended conditions. The reaction mixture was subjected to electrophoresis through a 10 percent polyacrylamide slab gel; the desired 40-bp fragment was located by autoradiography, excised, and electroeluted. Next, plasmid pRab91 was partially digested with Ava II, then digested to completion with Pst I. A 1297-bp fragment was recovered after gel electrophoresis of the restriction products. This fragment was used in a three-part ligation (by means of T4 DNA ligase) with the ^{32}P-labeled 40-bp repaired fragment and a modified pBR322 vector. The ligation mixture was transformed into *E. coli* 294 with selection on LB plates containing tetracycline at 5 μg/ml. Plasmid DNA of the transformants was screened by hybridization in situ with the labeled synthetic oligonucleotide (*18*) and then by digestion with restriction enzymes. The intermediate plasmid was then cleaved with Pst I, treated with bacterial alkaline phosphatase, and ligated with the ~ 700 bp Pst I fragment containing the remaining complementary DNA.

was propagated in a modified pBR322 plasmid which was prepared by opening at the only Eco RI site, then treating with DNA polymerase I in the presence of deoxyribonucleoside triphosphates to fill in the single-stranded ends, and subsequently treating with Pst I. Tetracycline resistant *E. coli* transformants were first screened by colony hybridization with the same ^{32}P-labeled DNA primer (*18*) to select correct transformation products. The correct intermediate plasmid was then opened at the only Pst I recognition site, treated with bacterial alkaline phosphatase to prevent self-ligation, and ligated in a molar excess of the approximately 700-bp Pst I fragment bearing the remaining coding sequence and nontranslated 3′ end. Asymmetric Ava I sites were used to determine the orientation; approximately one-half had the restriction fragment inserted in the orientation to recreate the entire text. The resultant plasmid, pRabG, is shown at the bottom of Fig. 2.

In pRabG, a single Eco RI recognition site directly precedes the structural gene for mature rabies glycoprotein (Fig. 2). The regulatory elements for expression of the rabies G gene were added by means of a 300 bp Eco RI–bounded promoter fragment isolated from the plasmid pLeIFA25 (*5*) which contained the congruent promoter and repressor binding site (operator), and also the ribosome binding site for the *E. coli* tryptophan operon leader peptide (*25*). The plasmid pRabG1 was opened by restriction with Eco RI, treated with bacterial alkaline phosphatase, and ligated with the 300-bp promoter fragment; the ligation reaction mixture was used to transform competent *E. coli* 294 to tetracycline resistance. An asymmetrically located Xba I site in the Eco RI fragment allowed us to screen for transformant plasmids which had the promoter fragment in the orientation for transcription of the rabies virus G gene. Several plasmid isolates were assembled correctly, and one, pRabG*trp*1, was chosen for study. After the *trp* promoter fragment was added, the nucleotide sequence around the Eco RI junction was determined and showed that incorporation of the new translation initiation codon had occurred as planned (*26–28*). In an alternate order of assembly, the Eco RI promoter fragment was inserted prior to the addition of the ~ 700-bp Pst I fragment bearing the rabies G COOH-terminal coding sequence. In the plasmid pRabGT1, the rabies glycoprotein text is brought into phase at amino acid position 446 with the portion of the β-lactamase gene which extends from the Pst I site of pBR322, an addition of 103 more codons before a translation stop signal is reached.

A preliminary test for expression of the rabies glycoprotein gene in *E. coli* was to examine by gel electrophoresis whole cell extracts of *E. coli* transformed by pRabGT1, pRabG*trp*1, or a control plasmid pIFN-γ*trp*48 (Table 1). A major new protein band, with migration commensurate with a size of about 62,000 daltons, was readily visible by Coomassie-blue staining in the gel lane containing the extract of *E. coli* W3110/pRabGT1 (Fig. 3, lane c). The presence of the protein was correlated with tryptophan depletion in the culture medium; it constituted about 3 percent of the total cell proteins. The protein expressed from pRabG*trp*1 was not as easily detectable. Therefore, one further manipulation was made to the expression vector to increase production of rabies glycoprotein. Studies with *E. coli lacZ* gene fusion proteins (*29*) and with bacterially synthesized human interferons (*30*) show that expression of a gene can be increased by alterations in the ribosome binding site of the transcript. Specifically, modifying by a few base pairs the distance between the Shine-Dalgarno sequence [the region of mRNA complementary to the 3′ end of 16*S* ribosomal RNA (*31*)] and the translation initiation signal, AUG, can result in more than 100-fold difference in the assayed products. The DNA fragment which was introduced to pRabG1 to provide a promoter also supplies a recognition site for Xba I within the presumed ribosome binding site as follows: (5′) GGTATC TAGAATTCATG (the underlined nucleotides constitute the Xba I site, and the barred nucleotides show the Shine-Dalgarno and translation initiation sequences).

The plasmids pRabGT1 and pRabG*trp*1 were altered by restriction with Xba I, treatment with S1 nuclease, and religation with T4 DNA ligase, as described (*30*). Plasmid DNA from transformants which arose on selective medium were screened for the loss of the Xba I recognition site. Plasmids designated pRabdex31 (derived from pRabGT1) and pRabdex2 (from pRabG*trp*1) were selected for further study. Nucleotide se-

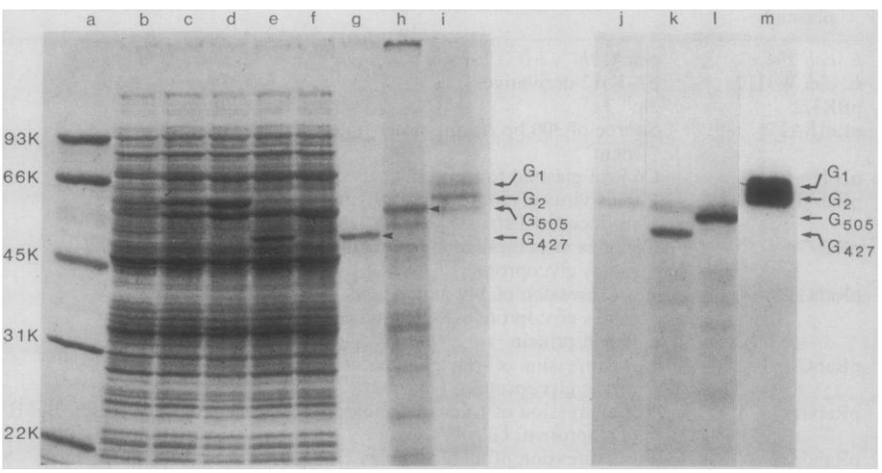

Fig. 3. Discontinuous, 10 percent polyacrylamide slab gel analysis of cellular proteins from *E. coli* bearing rabies virus glycoprotein expression plasmids. (Lane a) Molecular weight standard proteins (Bio-Rad). Lanes b to f are total cell extracts of *E. coli* W3110 transformed with (lane b) a control plasmid, pIFN-γ*trp*48 (*37*); (lane c) pRabGT1, (lane d) pRabdex31, (lane e) pRabdex1, and (lane f) pRabdex2. The inocula were grown overnight in LB plus tetracycline (5 μg/ml), then diluted 1:100 in M9 medium (*40*) containing 0.2 percent glucose, 0.5 percent Casamino acids, and tetracycline. Indole acrylic acid was added to a concentration of 25 μg/ml when A_{550} reached ~ 0.1. Cells were collected by centrifugation when A_{550} reached 1.0. Total cell extracts were prepared by boiling in 2 percent sodium dodecyl sulfate and 0.1*M* β-mercaptoethanol. (Lanes g and h) Enriched preparations of G_{427} and G_{505}, respectively. Cultures (500 ml) of *E. coli* W3110/pRabdex1 and *E. coli* W3110/pRabdex2 were grown in tryptophan-depleted medium as above, harvested by centrifugation at A_{550} = 1.0, and treated with lysozyme, 0.2 percent NP-40, and 1.5*M* NaCl (*13*). The insoluble pellet was slurried into PBS. Samples to be analyzed by gel electrophoresis were dissolved by boiling in buffer containing 20 percent glycerol, 5 percent β-mercaptoethanol, 0.125*M* tris-HCl (*p*H 6.8), and 4 percent SDS. (Lane i) A partially purified rabies virus glycoprotein prepared by NP-40 solubilization of rabies virions. Some faster migrating nucleocapsid (N) protein is also present in this preparation. All proteins in lanes a to i are visualized by staining with Coomassie brilliant blue. Lanes j to m show an autoradiogram of a protein blot (*41*) analysis of *E. coli*–derived or authentic rabies glycoprotein. Total cell extracts of *E. coli* W3110 containing plasmids (lane j) pIFN-γ*trp*48, (lane k) pRabdex1, or (lane l) pRabdex2 were placed on electrophoresis gels next to authentic glycoprotein (lane m) and transferred electrophoretically to nitrocellulose filter paper. The filter paper was treated serially with rabbit serum to rabies G and with ^{125}I-labeled protein A from Amersham (*42*). K represents 1000 daltons.

quencing confirmed that the four nucleotides which form single-stranded ends after Xba I scission had been removed to generate the new ribosome binding site GGTATAATTCATG.

Cultures of *E. coli* W3110/pRabdex31 and *E. coli* W3110/pRabdex2 were grown in the absence of tryptophan (Fig. 3). Gel electrophoresis of the whole cell extracts of those strains are shown in Fig. 3, lanes d and f. The full-length bacterially expressed rabies glycoprotein, termed G_{505}, comigrates with a major class of *E. coli* proteins, but can be observed as a widening of the stained protein band at about 57,000 daltons. Densitometry scanning of the electrophoretically separated *E. coli* W3110/pRabdex2 proteins, compared to control bacterial extract proteins, showed that G_{505} constituted about 2 percent of the cellular proteins, while the amount of fusion protein expressed from pRabdex31 was slightly less than 5 percent.

The synthesis of highly hydrophobic protein moieties can present problems to the *E. coli* host. For example, Rose and Shafferman (*32*) reported that expression of a vesicular stomatitis virus (VSV) glycoprotein-anthranilate synthetase fusion could be obtained only if the coding sequence for the hydrophobic amino terminal signal sequence of the VSV protein were removed. To investigate whether the synthesis of G_{505} was diminished due to the hydrophobic transmembrane amino acids 440 to 461, we undertook a construction in which the translation termination codon (TAG) would be brought into phase immediately preceding the coding sequence for the hydrophobic domain of the rabies virus glycoprotein. In this way analysis of the effect of the hydrophobic domain on expression in *E. coli* would not be confounded by extra amino acids at the carboxyl terminus. Figure 1 illustrates the position of the Pst I site where the rabies glycoprotein gene is joined to the β-lactamase gene to create the fused protein. In addition, Fig. 1 shows the location of a Bgl II restriction site which was used to design a truncated glycoprotein gene. Inspection of the sequence in the 3′-untranslated portion of the rabies G cDNA (not shown) revealed that excision of an internal Bgl II fragment from the plasmid pRabdex2 would result in a TAG triplet immediately after the codon for amino acid 427 of the mature rabies glycoprotein. The positions of the two Bgl II sites are indicated in Fig. 2.

The plasmid pRabdex2 was restricted with Bgl II and religated with T_4 DNA ligase. The ligation mix was used to transform *E. coli* 294 to tetracycline resistance; plasmid DNA from several isolates was screened for the loss of the internal Bgl II fragment. The plasmid pRabdex1 met this criterion and was selected for further characterization. Gel electrophoresis of a whole cell extract of *E. coli* W3110/pRabdex1 is shown in Fig. 3, lane e. The truncated glycoprotein, termed G_{427}, can be observed readily at a position corresponding to about 49,000 daltons; it constitutes about 2 to 3 percent of the cell proteins. In Table 2, the roughly estimated synthetic levels of G_{505} and G_{427} are shown in relation to the number of protein molecules per bacterium. The per cell synthesis of the more hydrophilic polypeptide, G_{427}, is about one and a half times higher than that of G_{505}, or nearly equal to the level of synthesis of the rabies glycoprotein–β-lactamase fusion protein.

Table 1. Bacterial strains and plasmids used.

Strain or plasmid	Relevant characteristics	Reference or source
E. coli 294	*end*A, *thi*$^-$, *hsr*$^-$, *hsm*$^+$	(*7*)
E. coli W3110	F$^-$ K-12 derivative	(*37*)
pBR322	Ap^R Tc^R	(*20*)
pLeIFA25*	Source of 300 bp *trp* promoter fragment	(*8*)
pIFN-γ*trp*48	Control plasmid for protein analysis	(*38*)
pRab91	Rabies virus glycoprotein cDNA introduced at Pst I site of pBR322	(*18*)
pRabG	Contains structural gene for mature rabies glycoprotein	This study; Fig. 2
pRabGT1	For expression of 549 amino acid rabies glycoprotein–β-lactamase fusion protein	This study
pRabG*trp*1	For expression of low level mature rabies glycoprotein, G_{505}	This study
pRabdex1	For expression of truncated rabies glycoprotein, G_{427}	From pRabdex2 by Bgl II restriction
pRabdex2	For expression of mature rabies glycoprotein, G_{505}	From pRabG*trp*1 by Xba I, SI treatment
pRabdex31	For expression of 549 amino acid glycoprotein fusion	From pRabGT1 by Xba I, S1 treatment

*All plasmids confer tetracycline (Tc) but not ampicillin (Ap) resistance (R) unless otherwise noted.

Table 2. Properties of the bacterially derived rabies glycoprotein gene products, G_{427} and G_{505}.

Item	G_{427}	G_{505}
Host or plasmid source	*E. coli* W3110/pRabdex1	*E. coli* W3110/pRabdex2
Molecular weight	48,600	56,900
Percent cell protein*	2.5 to 3.0	1.9 to 2.1
Copies/cell†	112,000 to 135,000	73,000 to 81,000
Percent in enriched preparation‡	24.2	21.6
Activity§ in PBS	1.9	2.8
Activity after solubilization in guanidine + urea	68.2	90.0

*Determined from densitometry tracings of electrophoretically separated proteins from three or more cell extracts (*39*). †Calculation based on the assumption that 1 liter of *E. coli* grown to an absorbance of 10 (A_{550} = 10) yields 2 g of protein from 5×10^{12} cells (*9*). ‡See legend to Fig. 3. §Reactivity with antibody to glycoprotein compared with an equal amount of authentic glycoprotein.

Characterization of Bacterially Synthesized Rabies Glycoprotein

To demonstrate the antigenic identity of the proteins produced by expression from the plasmids pRabdex1 and pRabdex2, cell extracts were subjected to electrophoresis on polyacrylamide slab gels, transferred to nitrocellulose, and reacted in situ with rabbit antiserum to authentic rabies glycoprotein (*33, 34*) (Fig. 3, lanes j–m). The antiserum used reacts somewhat with a normal *E. coli* component (Fig. 3, lane j, a control *E. coli* extract). The more intensely labeled bands in lanes k and l correspond with the rabies virus specific bands easily visible in stained slab gels. The gel from which this transfer was made is not shown, but is typified by extracts shown in lanes b, e, and f. The rabies glycoprotein–β-lactamase fusion gave a similar reaction (not shown). Polyacrylamide gel analysis of proteins from rabies virions of the CVS strain normally discriminates two electrophoretic glycoprotein species, G_1 and G_2, which differ in their

degree of glycosylation (*15*). The ^{125}I-labeled bands corresponding to the more slowly migrating natural glycoproteins, G_1 and G_2, are shown in Fig. 3, lane m. This procedure showed that at least some of the antigenic determinants survived the denaturing and reducing agents present in the gel-loading buffer. Furthermore, it showed that at least some of the antigenic determinants are independent of post-translational modifications, such as carbohydrate side chain attachment, which are made on the authentic glycoprotein in rabies infected cells, but would not be made in *E. coli* cells.

Rabies glycoprotein, when removed from intact virions, is quite insoluble. In fact, Lai and Dietzschold (*21*) reported that in order to dissolve glycoprotein for amino acid sequence determination it was necessary to boil protein preparations in 1 percent sodium dodecyl sulfate (SDS). The bacterially produced glycoprotein derivatives seem to share this characteristic and behave in *E. coli* much like other highly expressed, insoluble fusion proteins (*13, 34*). We obtained an approximately tenfold enrichment of G_{427} and G_{505} from cell extracts of *E. coli* W3110/pRabdex1 and *E. coli* W3110/pRabdex2, respectively, by collecting the precipitate of a lysate resulting from treatment with lysozyme in 0.2 percent NP-40 and 1.5*M* NaCl (Fig. 3, lanes g and h). Enriched preparations were then used to better assess the antigenicity of the proteins by comparison with authentic glycoprotein in an enzyme-linked immunosorbent assay (ELISA).

For the ELISA, dilutions of either purified rabies glycoprotein or the test bacterial samples were bound to the surface of a plastic microtiter tray, and were reacted with high-titer rabbit antiserum directed against authentic rabies glycoprotein. Unbound antibody was washed away, and the samples were treated with goat antiserum to rabbit antibody coupled to the enzyme alkaline phosphatase. Cleavage of the alkaline phosphatase substrate resulted in the development of a yellow color, which was monitored spectrophotometrically; conditions were set so that the absorbance was linear in the concentration range of 50 to 500 nanograms of rabies glycoprotein per milliliter (*35*).

When we assayed the insoluble protein fractions from cultures of *E. coli* W3110/pRabdex1, and *E. coli* W3110/pRabdex2, we found that the proteins G_{427} and G_{505} were only about 2 to 3 percent as effective per weight as purified authentic glycoprotein in binding antibodies to glycoprotein. To recover more activity from the bacterially synthesized proteins, a rather vigorous solubilization procedure was used. The enriched preparations were resuspended in 7*M* guanidine-hydrochloride at room temperature followed by dialysis into 7*M* urea and 5 m*M* tris, *p*H 8.0. (In some experiments, reducing agents were present in the buffers.) Samples were diluted up to 1:2000 in phosphate buffered saline (PBS) just prior to assay. Under these conditions, the proteins dissolved and their apparent activity increased more than 30-fold as indicated by the ELISA. Freshly solubilized preparations of G_{505} reacted with the antibody about 90 percent as effectively as natural glycoprotein, while the activities of solubilized preparations of G_{427} were no higher than 70 percent those of authentic glycoprotein. This reconstitution of activity by solubilization showed that the low reactivities first observed in the ELISA were not inherent properties of the bacterially synthesized glycoprotein analogs, G_{505} and G_{427}. Rather, the amount of activity recovered suggests that most antigenic determinants present on the authentic glycoprotein are conserved in the bacterial counterpart, G_{505}. We have yet to ascertain whether the incomplete recovery of antigenic activity in the preparations of G_{505} is artifactual or signifies the absence of specific immunogenic domains formed only in the virus-derived, glycosylated molecule.

Further purification of the G_{427} and G_{505} proteins for in vitro competition assays with authentic rabies glycoprotein and for animal inoculation is necessary. Although historical precedent (*13*) and the in vitro evidence presented here point toward their usefulness as immunogenic agents, these proteins cannot be dubbed subunit vaccines until after successful protection of an animal against lethal challenge with rabies virus. In 1980, reporting on a newly isolated human alpha interferon cDNA, Weissmann (36) mused: "We do not know whether *E. coli* IF [interferon] has the same specific activity as authentic LeIF [interferon-α] . . . If lack of appropriate glycosylation diminishes the activity of the molecule, we shall have a problem on our hands." In retrospect, glycosylation was not a problem. Today we know that the activities of interferons produced in *E. coli* are comparable to those of authentic interferon in every biological test to which they have been subjected. Further experimentation is required to determine whether bacterially produced rabies glycoprotein analogs will be as effective as virus-derived rabies glycoprotein in eliciting a protective response against rabies infection.

References and Notes

1. See M. M. Kaplan and H. Koprowski [*Sci. Am.* **242**, 120 (January 1980)] for a lively general article on rabies; J. H. Steele [in *The Natural History of Rabies*, G. M. Baer, Ed. (Academic Press, New York, 1975), p. 1] for a survey of the history of rabies and its prevention; J. F. Bell [*ibid.*, p. 331] for evidence that rabies infections are not necessarily fatal.
2. S. A. Plotkin, *Hosp. Pract.* **15** (No. 6), 65 (1980).
3. J. J. Callis, G. C. Poppensiek, D. H. Ferris, in *Virus Diseases of Food Animals*, E. P. J. Gibbs, Ed. (Academic Press, New York, 1981), p. 55.
4. G. M. Baer, M. K. Abelseth, J. G. Debbie, *Am. J. Epidemiol.* **93**, 487 (1971); J. G. Debbie, M. K. Abelseth, G. M. Baer, *ibid.* **96**, 231 (1972).
5. F. Steck, U. Hafliger, Ch. Stocker, A. Wandeler, in Abstracts from the Swiss Society of Microbiology, *Experientia* **34/12**, 1662 (1978).
6. R. A. Lerner, N. Green, H. Alexander, F.-T. Liu, J. G. Sutcliffe, T. M. Shinnick, *Proc. Natl. Acad. Sci. U.S.A.* **78**, 3403 (1981); J. L. Bittle *et al.*, *Nature (London)* **298**, 30 (1982). For a discussion of advances in this field, see J. G. Sutcliffe, T. M. Shinnick, N. Greene, R. A. Lerner, *Science* **219**, 660 (1983).
7. D. V. Goeddel *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 106 (1979); *Nature (London)* **281**, 544 (1979).
8. ———, *Nature (London)* **287**, 411 (1980).
9. M. J. Ross, in *Insulins, Growth Hormone, and Recombinant DNA Technology*, J. L. Gueriguian, Ed. (Raven, New York, 1981), p. 33.
10. See for example the discussion of human insulin (recombinant DNA) [I. S. Johnson, *Science* **219**, 632 (1983)].
11. P. Valenzuela, A. Medina, W. J. Rutter, G. Ammerer, B. D. Hall, *Nature (London)* **298**, 347 (1982); R. Hitzeman, unpublished results.
12. C. C. Liu and A. D. Levinson, in *Eukaryotic Viral Vectors*, Y. Gluzman, Ed. (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1982), p. 55.
13. D. G. Kleid *et al.*, *Science* **214**, 1125 (1981).
14. A. Flamand, T. J. Wiktor, H. Koprowski, *J. Gen. Virol.* **48**, 105 (1980).
15. B. Dietzschold, J. H. Cox, L. G. Schneider, *Virology* **98**, 63 (1979).
16. G. D. Coslett, B. P. Holloway, J. F. Obijeski, *J. Gen. Virol.* **49**, 161 (1980).
17. B. Dietzschold, J. H. Cox, L. G. Schneider, *Dev. Biol. Stand.* **40**, 45 (1978).
18. E. Yelverton, J. F. Obijeski, B. Holloway, D. V. Goeddel, in preparation.
19. A. Anilionis, W. H. Wunner, P. J. Curtis, *Nature (London)* **284**, 275 (1981).
20. F. Bolivar *et al.*, *Gene* **2**, 95 (1977).
21. C. Y. Lai and B. Dietzschold, *Biochem. Biophys. Res. Commun.* **103**, 536 (1981).
22. R. J. Winzler, in *Hormonal Proteins and Peptides*, C. H. Li, Ed. (Academic Press, New York, 1973), p. 1.
23. R. M. Lawn *et al.*, *Nucleic Acids Res.* **9**, 6103 (1981); D. V. Goeddel *et al.*, *ibid.* **8**, 4057 (1980).
24. The abbreviations for the bases are A, adenine; C, cytosine; G, guanine; T, thymine; U, uracil. The abbreviations for the amino acid residues are Ala, alanine; Arg, arginine; Asn, asparagine; Asp, aspartic acid; Cys, cysteine; Gln, glutamine; Glu, glutamic acid; Gly, glycine; His, histidine; Ile, isoleucine; Leu, leucine; Lys, lysine; Met, methionine; Phe, phenylalanine; Pro, proline; Ser, serine; Thr, threonine; Trp, tryptophan; Tyr, tyrosine, and Val, valine.
25. C. Yanofsky *et al.*, *Nucleic Acids Res.* **9**, 6647 (1981).
26. The single-stranded deoxyoligonucleotide dTCTGAAATGAGC, which hybridizes to the antisense strand from −50 to −38 bp 5′ to the transcription initiation site in the tryptophan operon (*25*), was synthesized by modified solid phase phosphotriester method (*27*) and used as primer for dideoxy chain termination sequencing (*28*) of alkali-denatured supercoiled plasmid. Plasmid DNA to be sequenced was denatured in 0.2*M* NaOH containing 0.2 m*M* EDTA for 5 minutes at room temperature, neutralized by addition of 0.13*M* ammonium acetate (*p*H 4.5) and precipitated with 2½ volumes of 95 percent ethanol. Primer was hybridized to the plasmid in tenfold molar excess in 6 m*M* tris-HCl, *p*H 7.6, 6 m*M* $MgCl_2$, 50 m*M* NaCl for 5 minutes at 37°C.
27. R. Crea and T. Horn, *Nucleic Acids Res.* **8**, 2331 (1980).
28. A. J. H. Smith, *Meth. Enzymol.* **65**, 560 (1980).
29. L. Guarente, T. M. Roberts, M. Ptashne, *Science* **209**, 1428 (1980).
30. H. M. Shepard, E. Yelverton, D. V. Goeddel, *DNA* **1**, 125 (1982).
31. J. Shine and L. Dalgarno, *Proc. Natl. Acad. Sci. U.S.A.* **71**, 1342 (1974).

32. J. K. Rose and A. Shafferman, *ibid.* **78**, 6670 (1981).
33. An antibody to rabies glycoprotein was raised against the glycoprotein-containing NP-40 supernatant fraction from purified rabies virions. Rabbits were injected once and then again at 3 weeks; serum titers (all 1:30,000, 3 weeks after the second injection) were determined by immunofluorescent neutralizing antibody assay [J. B. Thomas, in *The Natural History of Rabies*, G. M. Baer, Ed. (Academic Press, New York, 1975), p. 401]. A single crude serum sample from a single rabbit (titer 1:34,800) was diluted 1:300 to 1:1000 for use in ELISA or for filter hybridization. In some experiments, antiserum was first adsorbed to control bacterial extracts to eliminate nonspecific reaction in the ELISA.
34. D. C. Williams, R. M. Van Frank, W. L. Muth, J. P. Burnett, *Science* **215**, 687 (1982).
35. Sample antigens to be assayed were diluted into phosphate-buffered saline (PBS), and 100 μl was pipetted onto the wells of a Falcon flexible assay plate. After a 30-minute incubation period, unbound sample was aspirated away, and 100 μl of PBS plus 0.25 percent gelatin was applied. After 15 minutes, the gelatin solution was removed and the wells were washed three times with PBS plus 0.05 percent Tween. Rabbit antiserum to rabies glycoprotein (100 μl) diluted in PBS-Tween was added and incubated for 30 minutes. After washing three times in PBS-Tween, 100 μl of goat antiserum to rabbit immunoglobulin G–alkaline phosphatase conjugate was added (Zymed Laboratories; diluted 1:1000 in PBS-Tween just prior to use). After 30 to 60 minutes of incubation and washing, 100 μl of substrate (*p*-nitrophenyl phosphate, reconstituted according to Zymed instructions) was added. The absorbance at 405 nm was determined on a Dynatech ELISA reader. The standard plot for each experiment was authentic rabies glycoprotein diluted in duplicate from 5 to 50 ng per well.
36. S. Nagata, H. Taira, A. Hall, L. Johnsrud, M. Streuli, J. Ecsödi, W. Boll, K. Cantell, C. Weissmann, *Nature (London)* **284**, 316 (1980).
37. B. J. Bachmann, *Bacteriol. Rev.* **36**, 525 (1972).
38. P. W. Gray, D. W. Leung, D. Pennica, E. Yelverton, R. Najarian, C. C. Simonsen, R. Derynck, P. J. Sherwood, D. M. Wallace, S. L. Berger, A. D. Levinson, D. V. Goeddel, *Nature (London)* **295**, 503 (1983).
39. Polyacrylamide gel scanning was performed with a Zeineh soft laser scanning densitometer and a Hewlett-Packard 3390A integrator.
40. J. H. Miller, *Experiments in Molecular Genetics* (Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y., 1972), pp. 431–433.
41. H. Towbin, T. Staehelin, J. Gordon, *Proc. Natl. Acad. Sci. U.S.A.* **76**, 4350 (1979).
42. All antibody binding steps and washings were done in 50 m*M* tris, *p*H 7.4, 0.15*M* NaCl, 5 m*M* EDTA, 0.25 percent gelatin, and 0.05 percent NP-40 at room temperature for 1 to 2 hours.
43. We thank George Baer and Brian Holloway for developing the antiserum to glycoprotein, Larry Bock for carrying out the protein blot experiment (*41*), Avima Yaffe and Eric Patzer for help in setting up the ELISA, and Jeanne Arch and Alane Gray for help in manuscript preparation. We thank Art Levinson and Ron Wetzel for many helpful discussions. Supported by Genentech, Inc.

Secretion of Human Interferons by Yeast

Ronald A. Hitzeman, David W. Leung, L. Jeanne Perry
William J. Kohr, Howard L. Levine, David V. Goeddel

Summary. Plasmids were constructed to direct synthesis of the human interferons IFN-α1, IFN-α2, and IFN-γ in the yeast *Saccharomyces cerevisiae.* Expression of IFN genes containing coding sequences for secretion signals resulted in the secretion of IFN activity. A large proportion of the IFN-α1 and IFN-α2 isolated from the yeast cell growth media had the same amino termini as the natural mature interferons, suggesting a removal of the signal sequences identical to that of human cells. These results show that a lower eukaryote, such as yeast, can utilize and process a human signal sequence.

With the advent of recombinant DNA technology, it has been possible to examine the expression and secretion of eukaryotic gene products by the prokaryote *Escherichia coli*. Talmadge *et al.* (*1*) have shown that hybrid gene products containing bacterial–eukaryotic secretion signals (β-lactamase–rat preproinsulin) can be processed to proinsulin, which is then found in the periplasmic space of *E. coli*. They have also shown that for these hybrid proteins the eukaryotic signal (amino-terminal protein sequence) alone is sufficient to obtain secretion and processing in *E. coli* (*2*). Using another *E. coli* expression plasmid, Fraser and Bruce (*3*) have observed expression and secretion of chicken ovalbumin into the periplasmic space of *E. coli*. However, the secretion signal for this protein is thought to be within the protein sequence and there is no cleavage during the process (*4*). These results suggest that secretion signals may be functionally similar for many organisms. In this article we examine the secretion and processing of mammalian interferon (IFN) gene products in the lower eukaryote yeast.

The study of protein secretion from *Saccharomyces cerevisiae* (yeast) has focused primarily on the acid phosphatase and invertase enzymes (*5*), which are secreted into the periplasmic space. These enzymes are expressed as precursors having hydrophobic amino-terminal signal sequences (*6*) that are subsequently removed during the secretion process. Schekman and Novick (*5*) have used these enzymes to examine the secretion pathway in temperature-sensitive secretion mutants (*sec*). Characterization of these mutants has led to the recognition of a pathway in yeast that is organelle-dependent (*5*) and similar to that observed for the mammalian exocrine cells (*7*). The yeast pathway appears to start with translation of the pre-protein at the endoplasmic reticulum where the presequence (amino-terminal signal peptide) is removed after passage through the membrane, and then transported to the Golgi where vesicle intermediates are formed. These vesicles then move to the bud of the plasma membrane where fusion occurs with subsequent release of protein from the cell.

Plasmids that allow expression of heterologous (non-yeast) genes in the lower eukaryote yeast have recently been developed. Hitzeman *et al.* (*8*), using the promoter of the highly expressed yeast alcohol dehydrogenase I gene (*9*) to initiate transcription, demonstrated the synthesis of human IFN-α1 in yeast. However, the IFN-α1 used in these experiments was modified so that the NH_2-terminal signal peptide coding sequence was replaced by an ATG translational start codon. Therefore the polypeptide produced was mature IFN-α1 (the form found after secretion from human cells); it was not secreted from the yeast cell.

We have recently constructed a different portable yeast promoter (*10*) from another highly expressed yeast glycolytic gene encoding 3-phosphoglycerate kinase (PGK). This gene was isolated, characterized, and sequenced (*11*). When the PGK promoter fragment was

The authors are members of the Department of Molecular Biology, Protein Biochemistry, or Process Development of Genentech, Inc., South San Francisco, California 94080.