



INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 4 : C12P 21/00, C07K 13/00 C07H 15/12	A1	(11) International Publication Number: WO 87/ 05331 (43) International Publication Date: 11 September 1987 (11.09.87)
(21) International Application Number: PCT/US87/00396 (22) International Filing Date: 26 February 1987 (26.02.87) (31) Priority Application Number: 833,920 (32) Priority Date: 26 February 1986 (26.02.86) (33) Priority Country: US (71) Applicant: MASSACHUSETTS INSTITUTE OF TECHNOLOGY [US/US]; 77 Massachusetts Avenue, Cambridge, MA 02139 (US). (72) Inventors: TOTH, Matthew, J. ; Apartment 6H, 60 Wadsworth Street, Cambridge, MA 02142 (US). SCHIMMEL, Paul, R. ; 44 Sanderson Road, Lexington, MA 02173 (US). (74) Agents: PABST, Patrea, L. et al.; Hamilton, Brook, Smith & Reynolds, Two Militia Drive, Lexington, MA 02173 (US).		(81) Designated States: AT (European patent), BE (European patent), CH (European patent), DE (European patent), FR (European patent), GB (European patent), IT (European patent), JP, LU (European patent), NL (European patent), SE (European patent). Published <i>With international search report.</i>
(54) Title: CONSTRUCTION AND APPLICATIONS OF POLYPROTEINS		
(57) Abstract A method for constructing polyproteins which can perform multiple sequential activities. A DNA sequences is constructed using genetic engineering techniques to insert sequences encoding the desired proteins into a plasmid, in the correct order, following a single promoter element and before a single stop codon. The reading frames of the mRNA sequences are phased so that a polyprotein with the desired activities, in the required order, is produced. Modified polyproteins can be produced by inserting or substituting amino acids into the mRNA sequence to create spaces between the individual proteins, to increase the stability of the total polyprotein, to change the spatial orientation of the individual proteins relative to each other and their substrates, and to modify the activity of the individual proteins.		

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	ML	Mali
AU	Australia	GA	Gabon	MR	Mauritania
BB	Barbados	GB	United Kingdom	MW	Malawi
BE	Belgium	HU	Hungary	NL	Netherlands
BG	Bulgaria	IT	Italy	NO	Norway
BJ	Benin	JP	Japan	RO	Romania
BR	Brazil	KP	Democratic People's Republic of Korea	SD	Sudan
CF	Central African Republic	KR	Republic of Korea	SE	Sweden
CG	Congo	LI	Liechtenstein	SN	Senegal
CH	Switzerland	LK	Sri Lanka	SU	Soviet Union
CM	Cameroon	LU	Luxembourg	TD	Chad
DE	Germany, Federal Republic of	MC	Monaco	TG	Togo
DK	Denmark	MG	Madagascar	US	United States of America
FI	Finland				

-1-

CONSTRUCTION AND APPLICATIONS OF POLYPROTEINSBackground of the Invention

Many proteins have specialized biological activities. These activities include catalyzing specific reactions and
05 interacting with specific molecules, including ligands, cell surface receptors and other macromolecules. For many applications, it is necessary to combine the activities of two or more proteins to achieve the desired result. For example, a reaction may proceed in multiple steps where the sequential
10 action of two or more enzymes is required to catalyze the overall reaction. In nature, cells which do not produce all of the proteins with the required activity are unable to exist without supplementation with either the protein with the desired activity or the end product of the sequential reaction. Supplementation
15 with the missing or deficient protein first requires isolating the protein with the required activity and, secondly, combining the isolated protein with the substrate at the proper time in the reaction so that the entire sequential reaction can proceed. This is time consuming, laborious and costly. It would be more
20 desirable to have a single protein which could perform the entire reaction.

Polypeptide fusions have been reported in earlier work. The most common example is that of joining an enzyme, such as beta-galactosidase, to a protein to "tag" the protein. The
25 enzyme element is relatively easy to identify using, for example, a spectrophotometric assay, while the protein piece is generally more difficult to measure. There is no connection between the activity, if any, of the protein and the enzyme label and, in

-2-

general, there is no attempt made to determine whether the protein retains its original activity after it is joined to the enzyme.

5 There are examples in nature where, in a particular organism, enzymes which sequentially operate in a reaction pathway are joined together. One example is the fatty acid synthetase system, in E. coli, consisting of a single polypeptide with multiple activities. The comparable synthetase system in yeast consists of individual proteins.

10 There are many reasons why proteins usually exist in nature as separate molecules rather than being joined together. The necessity for the protein retaining the proper spatial relationship for binding to a receptor or substrate is one consideration. Another is the inherent instability of large
15 proteins. In one group of proteins, the twenty E. coli aminoacyl tRNA synthetases, only two exist as other than a single polypeptide chain. The other two, glycine and phenylalanine, consist of two polypeptide chains. In glycine tRNA synthetase, one chain is responsible for the activation of glycine with ATP
20 to form glycyl-adenylate. The other chain reacts the glycyl-adenylate with tRNA^{Gly} to yield Gly-tRNA^{Gly}. Neither reaction occurs in the absence of the other chain, suggesting that the two chains must interact with each other and the tRNA^{Gly}.

25 Heretofore, no one has joined active proteins in such a manner that they not only retain their original activity but can be used to perform sequential reactions. One problem with fusing the proteins together involved changes in the way they can interact together in solution. For example, in glycine tRNA synthetase, joining the two chains, alpha and beta, could prevent
30 the required interaction between the chains unless a means for

-3-

controlling the spacing and spatial relationship between the two chains was provided. Further, a means for stabilizing the protein in the presence of proteolytic enzymes is required. In their native configuration, areas susceptible to enzymatic degradation are protected by the folding of the protein into a specific three-dimensional structure. Another problem is to provide a means for controlling the spacing along the length of the individual proteins to provide for "channeling" or passing of the reaction products to the appropriate points along the protein chains.

It is therefore an object of the present invention to provide a method for constructing a polyprotein which can perform multiple, sequential activities, in vivo or in vitro.

It is a still further object of the invention to provide a method whereby a protein which can perform multiple, sequential activities can be reproducibly replicated in vivo or in vitro.

It is still another object of the invention to provide a method for modifying polyproteins with multiple biological activities, with respect to their biological activities, their spacial configuration, and the interrelationship of the portions with biological activities.

Summary of the Invention

A method for constructing and modifying polyproteins which can perform multiple, sequential activities. A polyprotein encoding DNA sequence is constructed using genetic engineering techniques to insert into a plasmid, sequences encoding the desired proteins, in the correct order, following a single promoter element and before a single stop codon. The reading

-4-

frames of the mRNA sequences are phased to produce a polyprotein with all of the biological activities of the individual proteins. The polyproteins can be modified by insertion or substitution of amino acids to create spaces between the proteins, to increase
5 the stability of the polyprotein, to change the spatial orientation of the individual proteins relative to each other and their substrates, and to modify the biological activities of the individual proteins.

In a specific example, a polyprotein is made from a
10 DNA sequence encoding both a first polypeptide which activates glycine with ATP to form glycyl-adenylate and a second polypeptide which reacts glycyl-adenylate with tRNA^{Gly}. The fused DNA sequences are produced with both the DNA encoding the Met initiation site and the DNA sequence encoding a ribosome
15 binding site at the beginning of the second polypeptide deleted so that only the complete polyprotein is produced. The polyprotein is stable with respect to both catalytic activity and molecular weight over time at 37°C.

Detailed Description of the Invention

20 The present invention is a method for constructing "polyproteins", proteins made up of individual proteins that have been joined together in a sequence whereby they retain their original biological activities and ability to interact with each other to perform multi-step reactions, in the proper sequential
25 order.

"Biological activity" includes enzymatic activity, interaction with a receptor, ligand, antigen, or cofactor, phosphorylation, acylation or formation of other modified forms of a compound, and other biological activities known to those
30 skilled in the art.

-5-

The polyprotein-encoding DNA sequence is formed by linking or "fusing" sequences encoding each of the individual proteins. The first sequence in the polyprotein DNA construction has a promoter element and a ribosome binding site. These elements assure that transcription of the polyprotein DNA into mRNA begins at a defined site and that the signal, the ribosome binding site, needed for translation of mRNA into protein is present. Synthesis of the polyprotein is made continuous from one protein component to the next by removing or altering any initiation or binding signals and stop codons from the subsequent protein-encoding sequences. The stop codon, normally a signal for the ribosome to stop translation and to end the polypeptide, is not altered or removed from the last DNA sequence. The individual protein encoding sequences are joined such that a proper phasing is made of the mRNA reading frames for translation of the sequence into the desired amino acids.

The primary advantage of the present invention is that a single polyprotein with multiple activities is produced. Another advantage is that the method affords the opportunity to build polyproteins which facilitate "channeling", or the direct passage of the product of one protein, such as an enzyme, to another protein, such as a second enzyme, without passage of the intermediate compound into the solution. Channeling provides a means to sequester and act quickly on highly unstable intermediate compounds which might decay rapidly in free solution.

Once a DNA sequence encoding a polyprotein or a "polyprotein gene" is made, it is necessary to demonstrate that the construction leads to production of a stable polyprotein. If the resulting protein is not stable, for example because the junctions between the proteins are vulnerable to proteolytic

-6-

digestion, then the junction regions are modified. This can be done by inserting different amino acids at or near the junction or by building spacers of amino acids between the individual proteins. Linkers or spacers can also be introduced to modify
5 the overall activity of the polyprotein. By adjusting the space between and orientation of the individual proteins, it is possible to modify the total activity of the polyprotein.

In one example of the present invention, a single polyprotein was formed by the fusion of the alpha and beta chains
10 of the E. coli glycine tRNA synthetase. The objective was to make a polyprotein which could catalyze a two step reaction. Step one of the reaction is to activate glycine with ATP to form glycyl-adenylate. The second step is to react glycyl-adenylate with tRNA^{Gly} to form Gly-tRNA^{Gly}. The entire aminoacylation
15 reaction is written as glycine + ATP + tRNA^{Gly} --> Gly-tRNA^{Gly} + AMP + PPi.

The presence and interaction of both the alpha and beta chains of the glycine tRNA synthetase are required for catalysis of the overall reaction. Neither glycine dependent ATP-PPi
20 exchange nor aminoacylation can be detected with either of the chains alone. Both reactions are detected when the chains are mixed together to generate a complex comprised of two alpha and two beta chains.

In the E. coli genome, the alpha and beta subunits are
25 encoded by a single gene. There is a promoter region, a ribosome binding site, the alpha subunit polypeptide coding region, a stop codon followed by nine nucleotides, the beta subunit coding region, and a stop codon. There is also a separate ribosome binding site for the beta chain, so that synthesis of this chain
30 can proceed independently of synthesis of the alpha chain.

-7-

Using genetic engineering techniques known to those skilled in the art, the alpha and beta protein coding regions were fused together to construct one continuous coding sequence within a plasmid. The translation termination signal at the end of the sequence encoding the alpha subunit was altered, as was the translation initiation site at the beginning of the sequence encoding the beta subunit. A linker of six amino acids was inserted to join the end of the coding region of the alpha chain to what is the normal beginning of the beta chain. In the linker region, several different amino acid sequences were tested for their effect on catalytic activity and stability. These sequences generally differed from each other by a single amino acid.

A two-protein polyprotein was synthesized in E. coli and found to be stable both in vivo and in vitro. Most significantly, the fused protein executes the overall two-step reaction with an efficiency that is comparable to that found for the individual proteins, establishing that fusing the two proteins as described does not significantly attenuate the biological activities while affording all of the advantages of a single polypeptide chain. The fused polyprotein functions in vivo as the sole source of glycine tRNA synthetase activity. This is a rigorous test of the efficacy of the fused polyprotein since glycine tRNA synthetase activity is essential for cell growth and any defects or loss of function would cause the cells to die.

The details of the construction of the fused polyproteins and method for analysis are as follows.

In the first step, the gene for the protein or proteins of interest is deleted from the chromosome. In this example, the

-8-

gene for glycine tRNA synthetase, the glyS locus, was deleted from the E. coli chromosome by a variation of the method in U.S. patent application Serial No. 688,612 filed January 3, 1985 by Jasin and Schimmel, entitled "A Method for Deletion of An

5 Essential Gene from a Bacteria" for deleting an essential gene from E. coli. The procedure is based upon site specific recombination with a linear DNA fragment that replaces the target gene, in this case glyS, with a Kan^r marker that is bounded by glyS 5' and 3' adjoining sequences. This recombination is done

10 in E. coli JC7623(recBC sbcB) where recBC sbcB mutations inactivate exonucleases that degrade linear DNA fragments. When the chromosomal copy of glyS is deleted to give the delta glyS strain TM101, cell viability is maintained with plasmid pMT 901. Plasmid pMT 901 encodes a HindIII fragment that contains glyS

15 cloned into the HindIII site of plasmid pMT 901. Plasmid pMT 901, encoding Cm^R and Tc^R and a temperature-sensitive replicon such that plasmid replication is blocked at the restrictive temperature (42°), is constructed by inserting the Cm^R locus from pBR 325 into pPM 103, which encodes Tc^R and a

20 temperature-sensitive replicon.

Cell viability is maintained by a plasmid (pMT 901) that encodes glyS and a temperature sensitive replicon. By virtue of the temperature sensitive replicon, plasmid replication is blocked at the restrictive temperature of 42°C. Cell survival

25 therefore depends entirely on the ability of a second plasmid to complement the glyS null allele. This second plasmid is introduced by transformation at the permissive temperature and retained at the temperature that is restrictive for the temperature-sensitive plasmid. The second plasmid contains the

30 specific glyS sequences to be tested.

Functional recA activity is required for the site specific recombination of linear DNA fragments. In the example

-9-

described by Jasin and Schimmel in U.S. Serial No. 688,612 filed January 3, 1985, also described in "Deletion of an Essential Gene in Escherichia coli by Site-Specific Recombination with Linear DNA Fragments" in J. of Bacteriology, 159(2), 783-786 (1984), for deletion of alaS, the close proximity of recA to alaS was taken advantage of. A linear DNA fragment was selected that simultaneously deleted alaS and a segment of recA. The resulting recA phenotype thus prevented subsequent recombination of the alaS-maintaining plasmid with the chromosome. In general, recA is required to prevent recombination between plasmid-borne sequences and the chromosome.

In the present example, it was not possible to inactivate recA simultaneously with the deletion of glyS. A two-step procedure was required in which the glyS deletion in TM101/MT901 was first created and the recA 56 allele in E. coli GW554 moved into this strain by P1 transduction. The resulting temperature-sensitive delta glyS recA strain is designated TM 102/pMT 901. For reasons of convenience, the delta glyS and recA alleles were then moved into a wild-type E. coli C600 strain to produce the temperature-sensitive TM 202/pMT 901 strain.

The glyS deletion in TM 101/pMT 901 was created as follows. Plasmid pTK 101 contains a 10 kbp PstI fragment which encompasses glyS and several kbp of 5'- and 3'-adjoining sequences. The central 5 kbp HindIII fragment of the PstI segment encoding glyS was replaced with a 3.3 kbp HindIII fragment encoding Kan^r from bacterial transposon Tn5. One microgram of the resulting plasmid, pMT300, was digested with EcoRI and NdeI at unique sites in the pBR322 portion of pMT300, and transformed into E. coli JC7623/pMT901. Selection for Kan^r yielded approximately 70 transformants. These transformants were presumed to have the delta glyS/Kan^r allele and one of them was designated as TM101/pMT901.

-10-

The recA56 allele was moved by P1 transduction into TM101/pMT901 to produce TM102/pMT901. Subsequently, both delta glyS and recA56 alleles were moved by P1 transduction from TM102/pM901 to a wild-type C600 strain to produce the delta glyS/Kan^r recA strain TM202/pM901. Linkage of Kan^r to xyl-5 in this strain was shown to be similar (90%) to that reported for linkage of glyS to xyl(80%). The genomic arrangement predicted for delta glyS Kan^r was also verified by Southern blot analysis.

Oligonucleotides for insertion into the fixed polypeptide DNA sequence were formed by oligonucleotide-directed mutagenesis using the procedure of M. Zoller and M. Smith in Methods in Enzymology 100, 468-500 (1983), as modified by L. Marsh in a personal communication. In this procedure, an oligonucleotide hybridizes to a specific complementary region of a single-stranded DNA template. The oligonucleotide has one or more mismatches with the region of single-stranded DNA to which it hybridizes. As a result, extension of the primer by DNA synthesis with the Klenow fragment of DNA polymerase results in a heteroduplex product. Replication and segregation of the heteroduplex results in mutant and wild-type double-stranded molecules. If the mutant clone does not produce a recognizable phenotype, it may be distinguished from the wild-type clone by its selective hybridization to [³²P]-labeled "mutagenic" oligonucleotide. The single-stranded DNA of the desired mutant clone hybridizes more strongly to the mutagenic oligonucleotide because of the perfect sequence complementation.

For oligonucleotide-directed mutagenesis, the 5 kbp HindIII fragment that contains the glyS gene was cloned into the single-stranded DNA phage mp8. The initial oligonucleotide-directed mutagenesis used the permuted heptadecamer COGCCTCTCCCTTATCT. Except for the permuted positions, this

-11-

oligonucleotide is complementary to the coding strand of glyS that encodes the alpha-subunit stop codon and its 5'- and 3'-flanking sequences, AGATAAGTAAGAGGCGG. The permuted positions alter the TTA that is complementary to the TAA stop, whose position is underlined above.

This oligonucleotide and others were made by phosphoramidite chemistry on a Systec Microsyn 1450A Automatic DNA Synthesizer. The oligonucleotides were purified by denaturing polyacrylamide gel electrophoresis in 8 M urea on 20% polyacrylamide gels or by HPLC purification on a reverse phase C-18 column using triethylammonium acetate buffer and an acetonitrile gradient, as recommended by the Systec manual.

Mutant mp8 plaques were detected at a frequency of 16% by hybridization with the 5'-[³²P]-labeled mutagenic oligonucleotide. Dideoxy sequencing, as modified for alpha-[³⁵S] dATP, with an oligonucleotide that primes close to the site of modification confirmed the putative mutant sequence. Two mutants referred to as Fusion 1 and Fusion 2, were recovered where the TAA stop codon of the alpha subunit was replaced by glutamic acid and glutamine, respectively.

The double-stranded DNA from these mutant mp8 clones was subcloned into the HindIII site of pBR322. The resulting plasmids are glyS fusion-encoding pFN101 (containing Fusion 1) and pFN201 (containing Fusion 2).

A second round of oligonucleotide-directed mutagenesis was performed essentially as described above on the pFN101 and pFN201 glyS fusions. In this procedure, two heneicosanomers, TTTCTCTTCTGCGAGCGCCTC and TTTCTCAGCTGCGAGCGC, that are complementary to the coding region were used to eliminate the ATG

-12-

(complementary to CAT) codon of the beta subunit. (The underline designates the site that was changed.) This was accomplished by a CAT to TGC change which results in a coding strand change of Met to Ala. An AGA to AGC or AGA to TTC change was introduced adjacent to the Met to Ala change. Mutant yields were similar to those previously obtained, and the putative fusion mutants were sequenced and subcloned. These "second generation" fusion mutants, on plasmid pBR322, are designated as pFN102, pFN202 and pFN203, containing Fusions 3, 4, and 5, respectively.

The sequence of the five Fusions are shown in Table 1.

Table 1: Fusion Sequences in the Region
Between the Alpha and Beta Subunits

Wild Type	GAT	AAG	TAA	GAG	GCG	GCT	ATG	TCT	GAG	AAA
encodes	Asp	Lys	Stop	Glu	Ala	Ala	F-met	Ser	Glu	Lys
15 Fusion 1 (pFN101)			GAA Glu							
Fusion 2 (pFN201)			CAA Gln							
20 Fusion 3 (pFN102)			GAA Glu				GCA Ala	GCT Ala		
Fusion 4 (pFN202)			CAA Gln				GCA Ala	GCT Ala		
Fusion 5 (pFN203)			CAA Gln				GCA Ala	GAA Glu		

-13-

All of the fusion sequences complement the deletion strain of glyS (TM202). To visualize the plasmid encoded proteins, the maxicell technique described by Sancar et al in J. Bact. 137, 692-693 (1979) was used to produce [³⁵S] fusion and
5 wild-type proteins. The labeled proteins were separated on a Laemmli gel (3% stacking gel, 10% running gel), soaked in Autofluor (National Diagnostics), dried onto filter paper and exposed to x-ray film. Plasmid encoded proteins were identified at the anticipated molecular weights.

10 Crude extracts of maxicell preparations from the mutant fusion strains were incubated with extracts of wild-type E. coli C600 cells to demonstrate resistance of the fusion proteins to degradation.

For biochemical analysis of the fusion proteins,
15 deletion strains containing fusion mutant or wild-type glyS plasmids were grown to late log phase in 100 ml of LB media. Cells were collected by centrifugation, washed with 0.7% saline, centrifuged, scraped into a chilled mortar and pestle, ground with approximately two packed cell volumes of alumina in 1 ml of
20 grinding buffer (50 mM potassium phosphate pH 7.5, 15 mM 2-mercaptoethanol, 200 microliters PMSF-saturated ethanol per liter). The extract was centrifuged at 12000 x g in an Eppendorf centrifuge and the supernatant used for analysis.

Aminoacylation assays were performed at 37°C as
25 described by A.A. Schreier and P.R. Schimmel in Biochem. 11, 1582-1589 (1972). Pyrophosphate exchange assays were performed as described by R. Calendar and P. Berg in Procedures in Nucleic Acid Research, G.L. Catoni and D.R. Davies, eds., pp. 384-399 (1966) as modified by Putney et al. in J. Biol. Chem. 256,

-14-

198-204 (1981). Beta-lactamase activity of crude extracts was determined spectrophotometrically according to J.A.T. Jansson in Biochem. Biophys. Acta 99, 171-172 (1965).

Crude extracts were partially purified by application
5 to an S-300 Sephacryl™ (Pharmacia Fine Chemical Co.) gel
filtration column. A 100 cm x 1.1 cm column was packed with
superfine Sephacryl™ S-300 in column buffer, 100 mM NaCl, 1 mM
DTT, 10 mM potassium phosphate pH 7.5. The column was calibrated
with standard proteins supplied by Boehringer-Mannheim
10 Biochemicals. Flow rates with column buffers were about 7 ml per
hour and 1.5 ml fractions were collected. Fractions were assayed
for aminoacylation or pyrophosphate exchange.

E. coli strains were grown on LB plates or in LB media
until stationary phase. M9 minimal media was used in the P1
15 transduction manipulations. Concentrations of antibiotics,
supplied by Sigma Chemical Co., were: Kanamycin, 50
micrograms/ml; chloramphenicol 25 micrograms/ml; ampicillin, 50
micrograms/ml; and tetracycline, 12 micrograms/ml.

All restriction endonucleases were used according to
20 manufacturers specifications. DNA fragments were separated using
polyacrylamide gel or agarose gel electrophoresis in a
Tris-borate-EDTA buffer using methods known to those skilled in
the art.

The efficacy of specific fusion polyproteins was
25 established in vivo in a strain where the chromosomal glyS locus
has been deleted as previously discussed.

Table 1 shows the glyS intersubunit region as
originally formed by the fusion of the alpha and beta subunit.

-15-

The translation stop of the alpha-subunit coding region is separated by nine nucleotides from the translation start of the 689 codons of the beta-subunit open reading frame. The intersubunit region encodes a ribosome binding site which enables independent translational initiation at the beginning of the beta-chain. Edman degradation of purified beta-subunit has shown that the mature protein starts with Glu and lacks the Met and Ser that are at the start of the coding region. Oligonucleotide-directed mutagenesis was used initially to construct two simple in-frame fusions of the alpha and beta polypeptides, Fusion 1 and 2. The alpha-chain TAA stop codon was changed to GAA (Glu) in Fusion 1 and to CAA (Gln) in Fusion 2. Each of these fusion constructions on plasmids was transformed into TM202/pM901. There was stable colony growth of the transformants at temperature to 42°C despite loss of the temperature sensitive plasmid, establishing that the polypeptide per se supports viable cell growth.

To establish the effect of the fusion on the synthesis of the gene product, maxicell analysis of protein products was carried out as described previously. With this procedure, synthesis of plasmid-encoded proteins was visualized by autoradiography of [³⁵S]-labeled products that were separated by SDS-gel electrophoresis. The results for Fusion 1 and Fusion 2 show that the fusion proteins are the predominant glyS species in the system. This suggests that the fusion proteins are stable, even though six amino acids have been inserted between the end of the native alpha and the beginning (Glu) of the mature beta chain. Little or no alpha-chain is evident. However, significant amounts of the beta-subunit, in the absence of significant amounts of the alpha chain, suggests that independent initiation of beta-chain synthesis occurred in Fusion 1 and Fusion 2. The independent synthesis of the beta chain was

-16-

eliminated in Fusions 3, 4, and 5, made by oligonucleotide directed mutagenesis, which were constructed so as to remove the potential for independent ribosome binding at the beginning of the beta subunit coding region.

5 In Fusions 3, 4, and 5, the beta-chain's methionine-initiator codon was replaced with GCA (Ala). This change alone eliminates internal initiation of translation at the start of the beta-piece of the fusion construct. This change was combined with an alteration of the second codon of the beta chain and of
10 the alpha-subunit stop codon TAA.

 Fusions 3, 4 and 5 complement the glyS null strain TM202. Maxicell analysis of the translation products clearly shows that stable polyproteins are produced from all three sequences and that neither alpha nor beta chain is produced in
15 significant amounts. This confirms that internal initiation was responsible for the presence of the free beta chain in synthesis from Fusion 1 and Fusion 2. More significantly, complementation of the glyS null strain by the fusion protein, in the absence of significant amounts of either of the free subunits, establishes
20 that the fusion protein itself is active.

 The enzymatic activities of extracts of strains bearing Fusions 1-5 were compared with the activity of the wild-type protein. Two activities were measured: the glycine dependent ATP-PPi exchange that monitors aminoacyl adenylate synthesis and
25 the glycine-specific aminoacylation of tRNA. The activities were normalized to the beta-lactamase activity that is also encoded by the multi-copy plasmid containing the fusion sequence to be assayed.

 Data given in Table 2 show that, for each of the
30 fusions polyproteins, the glycine-dependent ATP-PPi exchange

-17-

activity is within 1 to 2 fold of that of the wild-type protein. The aminoacylation activities, while relatively lower, are nonetheless at least 35 to 65% of that of the wild-type enzyme.

Table 2: Activity of Wild-Type and Fused
Polyprotein Glycine tRNA Synthetase

5

Strain	Units of [¹⁴ C]-glycine amino- acylation	Units of ATP-Pi	Units of Beta- lactamase	(Amino- acylation) (Beta- lactamase)	(ATP-Pi) (Beta- lactamase)
	Activity 1	Activity 2	Activity 3		
Wild type	6,250	3,318	0.170	1.0	1.0
Fusion 1	1,150	1,515	0.065	0.48	1.19
Fusion 2	430	878	0.036	0.32	1.25
Fusion 3	900	2,303	0.055	0.45	2.15
Fusion 4	850	1,702	0.041	0.56	2.13
Fusion 5	750	1,352	0.057	0.36	1.21

15

- Expressed as the TCA precipitable CPM per minute per microliter of extract
- Expressed as the charcoal adsorbed [³²P] CPM per 15 minute assay per microliter of extract
- Expressed as the absolute change of absorbance at 240 nm per minute per microliter of extract

20

The stability of the Fusion 2 polyprotein over time was also investigated. Maxicell extracts were allowed to incubate for one hour at 37° before electrophoresis. This had no

25

-18-

discernible effect on the yield of fusion protein or on the appearance of alpha- or beta-chain products resulting from breakdown of the fusion protein. In another experiment, the activities of Fusion 3 protein were also unchanged at room
5 temperature for at least three hours.

In summary, this example demonstrates that the method of the present invention is useful in constructing polyproteins, and single DNA sequences encoding such polyproteins, which retain the individual biological activities and which can function in a
10 sequential fashion. The example further demonstrates the usefulness of the method of the present invention in modifying these DNA sequences to produce polyproteins with altered spatial relationships, stability, and biological activities.

The present invention may be embodied in other specific
15 forms without departing from the spirit and scope thereof. These and other modifications of the invention will occur to those skilled in the art. Such other embodiments and modifications are intended to fall within the scope of the appended claims.

We claim:

-19-

1. A method for constructing polyproteins with multiple biological activities comprising:

(a) isolating nucleotide sequences encoding proteins with biological activities;

5 (b) selecting a vector suitable for the insertion of the nucleotide sequences encoding the proteins with biological activities;

(c) inserting the nucleotide sequences isolated in step (a) into the vector of step (b); and

10 (d) removing from the inserted sequences any stop codons other than at the end of the last inserted sequence, while phasing the reading frames as required to produce the desired polyprotein.

2. The method of claim 1 further comprising inserting
15 amino acids in addition to or in place of amino acids in the inserted sequences.

3. The method of claim 1 further comprising removing
any initiator codons other than the initiator codon at the
beginning of the first inserted sequence and any ribosome binding
20 sites other than at the beginning of the first inserted sequence.

4. A method for modifying the biological activities of a polyprotein with multiple activities comprising:

(a) isolating nucleotide sequences encoding proteins with biological activities;

25 (b) selecting a vector suitable for insertion of the isolated nucleotide sequences;

(c) inserting the nucleotide sequences into the vector, while maintaining the phasing of the reading frames required to produce the proteins with biological activities, to
30 form a polyprotein encoding sequence; and

-20-

(d) modifying the polyprotein encoding sequence by insertion or removal of codons encoding selected amino acids.

5 5. The method of claim 4 wherein steps (c) and (d) comprise inserting codons encoding selected amino acids into or adjacent to the polyprotein encoding sequence.

6. The method of claim 4 wherein steps (c) and (d) comprise selecting the amino acids to alter the spatial relationship between the proteins with biological activities.

10 7. The method of claim 6 wherein the spatial relationship is altered to provide for "channeling" between the proteins with biological activities.

15 8. The method of claim 4 wherein steps (c) and (d) comprise selecting the amino acids to allow production of portions of the polyprotein in addition to the complete polyprotein.

9. The method of claim 4 wherein steps (c) and (d) comprise selecting the amino acids from the amino acids in and adjacent the junctions between the proteins with biological activities.

20 10. The method of claim 9 wherein the selected amino acids are replaced or removed to produce a junction resistant to proteolytic attack.

11. The method of claim 4 further comprising selecting for polyproteins with enhanced biological activities.

25 12. A polyprotein comprising multiple functional amino acid sequences wherein said polyprotein is produced from a

-21-

single nucleotide sequence sequentially encoding a promoter region, proteins with biological activities, and a stop codon.

13. The polyprotein of claim 12 further comprising non-functional amino acid sequences separating the amino acid sequences with biological activities.

14. A nucleotide sequence comprising codons sequentially encoding a promoter region, proteins with biological activities, and a stop codon.

INTERNATIONAL SEARCH REPORT

International Application No PCT/US87/00396

I. CLASSIFICATION OF SUBJECT MATTER (if several classification symbols apply, indicate all) ³		
According to International Patent Classification (IPC) or to both National Classification and IPC		
INT. CL ⁴ : C12P 21/00; C07K 13/00; C07H 15/12		
U.S. CL : 435/68; 530/350; 536/27		
II. FIELDS SEARCHED		
Minimum Documentation Searched ⁴		
Classification System	Classification Symbols	
U.S.	435/68, 70, 172.3 530/350 536/27; 935/11, 47, 48	
Documentation Searched other than Minimum Documentation to the Extent that such Documents are Included in the Fields Searched ⁵		
CHEMICAL ABSTRACTS DATA BASE (CAS) 1967-1987, BIOLOGICAL ABSTRACTS DATA BASE (BIOSIS) 1967-1987; AUTOMATED PATENT SYSTEM (US) 1975-1987; KEYWORDS: POLYPROTEINS, PLASMID, RECOMBINANT DNA.		
III. DOCUMENTS CONSIDERED TO BE RELEVANT ¹⁴		
Category [*]	Citation of Document, ¹⁶ with indication, where appropriate, of the relevant passages ¹⁷	Relevant to Claim No. ¹⁸
X, P	US, A, 4,624,926 (INOUE ET AL) 25 November 1986. See columns 10-14.	1-8 and 11-14
X	US, A, 4,559,300 (KOVACEVIC ET AL) 17 December 1985. See columns 2, 5 and 9-12.	1-8 and 11-14
X	US, A, 4,411,994 (GILBERT ET AL) 25 October 1983. See columns 4 and 5.	1-8 and 11-14
X Y	US, A, 4,356,270 (ITAKURA) 26 October 1982. See columns 8, 15 and 16.	<u>1-8, 12-14</u> <u>9-11</u>
[*] Special categories of cited documents: ¹³ "A" document defining the general state of the art which is not considered to be of particular relevance "E" earlier document but published on or after the international filing date "L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified) "O" document referring to an oral disclosure, use, exhibition or other means "P" document published prior to the international filing date but later than the priority date claimed "T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention "X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step "Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art. "&" document member of the same patent family		
IV. CERTIFICATION		
Date of the Actual Completion of the International Search ²		Date of Mailing of this International Search Report ²
19 April 1987		04 MAY 1987
International Searching Authority ¹		Signature of Authorized Officer ²⁰
ISA/US		Thomas D. Mays

III. DOCUMENTS CONSIDERED TO BE RELEVANT (CONTINUED FROM THE SECOND SHEET)

Category *	Citation of Document, ¹⁶ with indication, where appropriate, of the relevant passages ¹⁷	Relevant to Claim No ¹⁸
<u>X</u> <u>Y</u>	CELL (CAMBRIDGE, MASSACHUSETTS, USA) Volume 37, issued July 1984 (HANECAK ET AL) "Expression of a cloned gene segment of poliovirus in <u>E. coli</u> : Evidence for autocatalytic production of the viral proteinase". See pages 1063, 1064, 1071 and 1072.	1-8 and 12-14 <u>9-11</u>
<u>X</u> <u>Y</u>	SCIENCE (WASHINGTON, D.C.) Volume 209, issued 19 September 1980 (GUARENTE ET AL) "A technique for expressing eukaryotic genes in bacteria". See entire document.	1-8 and 12-14 <u>9-11</u>

FURTHER INFORMATION CONTINUED FROM THE SECOND SHEET

X	NUCLEIC ACIDS RESEARCH (OXFORD, UK) Volume 13, issued March 1985 (SKERN ET AL) "Human rhino- virus 2: complete nucleotide sequence and proteolytic processing signals in the capsid protein region". See pages 2112-2114 and figure 2.	1-8 and 12-14
---	---	------------------

V. ☐ OBSERVATIONS WHERE CERTAIN CLAIMS WERE FOUND UNSEARCHABLE ¹⁰

This international search report has not been established in respect of certain claims under Article 17(2) (a) for the following reasons:

1. ☐ Claim numbers , because they relate to subject matter ¹² not required to be searched by this Authority, namely:

2. ☐ Claim numbers , because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out ¹³, specifically:

VI. ☐ OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING ¹¹

This International Searching Authority found multiple inventions in this international application as follows:

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims of the international application.
2. ☐ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims of the international application for which fees were paid, specifically claims:

3. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claim numbers:

4. ☐ As all searchable claims could be searched without effort justifying an additional fee, the International Searching Authority did not invite payment of any additional fee.

Remark on Protest

- ☐ The additional search fees were accompanied by applicant's protest.
☐ No protest accompanied the payment of additional search fees.