Volume 13 Number 6 1985 Nucleic Acids Research

Optimizing the expression in *E. coli* of a synthetic gene encoding somatomedin-C (IGF-I)

Gary Buell, Marie-Francoise Schulz, Gerald Selzer, Andre Chollet, N.Rao Movva, Dominique Semon, Sonia Escanez and Eric Kawashima

Departments of Molecular Biology and Chemistry, BIOGEN SA, 3 Route de Troinex, 1227 Carouge/Geneva, Switzerland

Received 21 December 1984; Accepted 5 March 1985

ABSTRACT

Double-stranded DNA encoding the human hormone somatomedin-C (SMC) has been synthesized. This synthetic gene has been inserted into a plasmid bearing the strong leftward promoter (P_L) of bacteriophage λ and expressed in E. coli. Codons for the N-terminal region of SMC which maximized the hormone's synthesis were chosen in an SMC-lac z fusion assay. The amounts of SMC accumulated in E. coli were influenced by mutations at two chromosomal loci, lon and htpR.

INTRODUCTION

Somatomedin-C, a 70 amino acid protein found in human serum, is believed to mediate many of the effects of growth hormone (1). In particular it can stimulate growth in rats made growth hormone-deficient by hypophysectomization (2). SMC, also known as insulin-like growth factor I (IGF-I) (3,4), has extensive amino acid sequence homology with other members of the insulin gene family (5). However, unlike insulin, SMC consists of a single polypeptide chain and thus probably has a structure more analogous to that of proinsulin (5).

We have synthesized a gene that encodes SMC, and studied its expression in E. coli under control of the bacteriophage λ P_L promoter. Use of this promoter and the temperature-sensitive repressor encoded by the λ cI_{857} gene allowed us to control expression of the gene by varying the temperature at which the host cells were grown. We have also fused the gene to the E. coli lac z gene to create a hybrid protein, the initiation of whose synthesis is governed by the SMC gene. Similar gene fusions with lac z have been used extensively to study the expression of both prokaryotic and eukaryotic genes (6,8). Because the hybrid protein retains β-galactosidase activity, we could rapidly screen a large number of variant SMC genes for their effect on the level of expression by using a simple indicator plate technique.

Higher levels of some eukaryotic proteins can accumulate in E. coli

strains deficient in lon protease (9). Recent work has shown that the intracellular level of the lon gene product increases following the same kind of heat shock treatment that we used to trigger SMC synthesis (10,11). Thus, we examined the effect on SMC accumulation of mutations at lon and htpR, a gene whose product is necessary for the heat shock response (12).

MATERIALS AND METHODS

Bacterial strains

strain	genotype	reference
C600 λ	thi, thr, leu, lacY, tonA, supE, λ	13
HB101	hsd, recA, ara, proA, lacY, galK, rpsL, xyl, mtl, supE	13
JM101	Δ(lac-pro), supE, thi, F′[proAB, $lacI^Q$, ZΔM15, traD]	14
JM83	ara, Δ(lac-pro), rpsL, thi, Ø80d[ZΔM15]	15
SG931	lac_{am}, trp_{am}, pho_{am}, $supC^{ts}$, rpsL, mal, tsx::TN10	11
SG933	as SG931 plus lon	11
SG934	as SG931 plus htpR	11
SG936	as SG931 plus lon, htpR	11

Oligodeoxynucleotide synthesis

Oligodeoxynucleotides were synthesized using the solid-phase phosphotriester method (16). After deprotection the crude products were desalted by gel filtration on Sephadex G-50 and purified by electrophoresis on slab gels of polyacrylamide containing urea (17). The oligodeoxynucleotides were localized by UV shadowing and isolated by electroelution from gel slices. The gel purified material was phosphorylated using T4 polynucleotide kinase (13) and repurified as above.

Cloning the synthetic DNA in M13

Double-stranded M13mp8 vector DNA was prepared by endonuclease digestion of the replicative form (RF) of M13mp8 phage using the BamH I and/or Hind III enzymes followed by treatment with E. coli DNA polymerase I (Klenow fragment) to give a blunt-ended fragment. The DNA was treated with calf intestinal phosphatase before use (18). Twenty ng of the blunt-ended vector was mixed with 0.1 pmole each of the synthetic DNA fragments (see below) and T4 DNA ligase and incubated overnight at 16°C. The ligation mixture was used to transfect JM101 which was then plated on LB agar containing 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside (X-GAL). Phage with inserted DNA were identified by their white plaque phenotype. The sequences of the inserted DNAs were determined by the dideoxy-chain termination method (19).

Induction of SMC synthesis

Five ml of L-broth (20) containing 50 ug/ml ampicillin and 40 ug/ml kanamycin was inoculated from plates and incubated overnight at 28°C. Two ml of the overnight culture was added to 10 ml L-broth, at 42°C in a 100 ml erlenmeyer flask which was then agitated vigorously at 42°C for 2 hours. A 1 ml aliquot was centrifuged and the pelleted cells were resuspended at a density equivalent to A_{550}= 20 in a lysis buffer containing SDS and β-mercaptoethanol (21). Cells were lysed by incubating for 10 min at 100°C.

Assays

β-galactosidase activity was assayed using o-nitrophenyl-β-D-galactoside (ONPG) as described by J. Miller (20). Identical β-galactosidase specific activities were found in cells grown to saturation in L-broth or growing exponentially in M9 minimal media (20). SMC was measured by radioimmuno-assay (RIA) with a commercial kit (Nichols Institute Diagnostics, San Juan Capistrano, Ca). Kit standards were verified with purified IGF-I, kindly provided by Dr. R. Humbel, Univ. of Zurich. SMC containing lysates were prepared as for gel electrophoresis (see above). Aliquots of the lysates were diluted at least 20-fold in the assay buffer prior to assaying in duplicate. We found that human SMC, denatured under our standard lysis conditions was as reactive in the RIA as the native hormone (data not shown).

Other Techniques

Restriction endonuclease digestion (13), ligation (13), electroelution of DNA (22), transformation (23) and SDS polyacyrlamide gel electrophoresis (21) were carried out essentially as described. Chemical sequencing of DNA was performed as described (24) using fragments end-labelled with the Klenow fragment of DNA polymerase I (13).

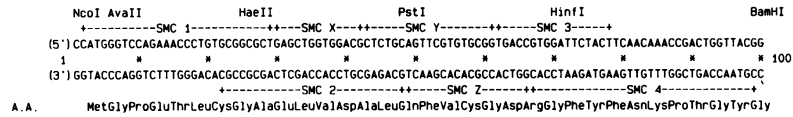
RESULTS

Synthesis of a gene encoding SMC

A synthetic gene which encodes human SMC (IGF-I) was designed on the basis of the known amino acid sequence of the protein (3,4). The nucleotide sequence of the synthetic gene (fig. 1) was designed with two constraints. First, we chose codons that are most often found in highly expressed genes of E. coli (25) or correspond to the most abundant tRNAs in E. coli (26). Second, we included endonuclease recognition sites at several positions to facilitate manipulation of the DNA.

To construct the gene, we first synthesized fourteen

Fragment A



Fragment B

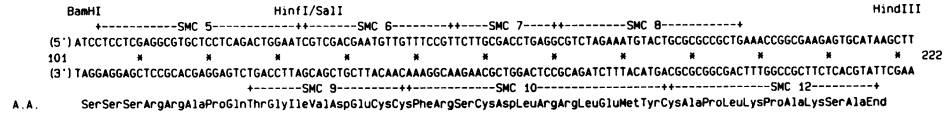


Figure 1. Synthesis of the synthetic gene. 200 pmole each of fragments SMC-1,2,3,4,X,Y and Z were heated to 95°C in 20 ul reannealing buffer (50mM Tris HCl pH 7.6, 10mM $MgCl_2$) and slowly cooled to 4°C. Dithiothreitol, ATP and T4 DNA ligase were added, respectively, to final concentations of 5mM, 70uM and 20 u/ml and the reaction was incubated at 4°C for 10 hours. The reaction mixture was ethanol precipitated, separated on a 8% polyacrylamide/7M urea gel and the 77 and 78 base strands were eluted. 25 pmole of each strand in 5 ul of reannealing buffer was heated to 95°C and slowly cooled to 15°C. Dithiothreitol, dNTPs and the Klenow fragment of DNA polymerase were added to 5mM, 250 uM and 2 units final concentration, respectively and the reaction allowed to proceed for 30 minutes at room temperature. Reaction products were gel purified as above and 7 pmole of the 98 base pair Fragment A was eluted. 20 pmole of the 114 base pair Fragment B was isolated in the same manner, commencing with 200 pmole each of fragments SMC-5,6,7,8,9,10 and 11.

oligodeoxynucleotides that varied in size from 13 to 37 bases. These were phosphorylated and assembled into two double-stranded DNA fragments using T4 DNA ligase and DNA polymerase I as indicated in figure 1. The 98 bp fragment A was ligated to M13mp8 DNA (see methods) to give a recombinant, mp8SMCA, in which Nco I and BamH I recognition sites flank the inserted DNA. Similarly, ligation of fragment B to M13mp8 vector gave a recombinant, mp8SMCB in which Hind III and BamH I sites flank the inserted DNA.

In the final step of the construction, fragments A and B were reisolated from the recombinant phage and joined to give the intact gene. Purified mp8SMCA and mp8SMCB RF DNAs were digested with Nco I plus BamH I and BamH I plus Hind III, respectively. The 98 bp SMC A fragment and 138 bp SMC B fragment were isolated from the digestion products by polyacyrlamide gel electrophoresis and were then mixed and ligated with a third fragment of 67 bp which includes the ribosome-binding site (27) from the ner gene of bacteriophage mu. The ner gene fragment, described previously (28), was

modified from the natural mu sequence by B. Allet (unpublished results). It consists of nucleotides 1043-1096 in the mu sequence (fig. 4 in ref. 29), preceded by a site for EcoR I and followed by a site for Nco I (CCATGG). The internal ATG of the Nco I site forms the codon for methionine which initiates translation.

The 303 bp EcoR I-Hind III ligation product containing all three fragments was isolated and introduced into pPLc24 (30), prepared with the same restriction enzymes (bottom, fig. 2). The resulting plasmid, $pPLmuSMC_{ori}$ thus contains in clockwise order the λP_L promoter (contributed by pPLc24), the ner gene ribosome binding site and the synthetic SMC gene.

To examine the expression of the SMC gene, this plasmid was introduced into HB101 by cotransformation with pcI_{857}, a plasmid compatible with pPLc24, which encodes a temperature-sensitive repressor of P_L (31). When $HB101/pcI_{857}/pPLmuSMC_{ori}$ was grown at 42°C, the cells contained little SMC. The SMC was barely detectable by RIA (Table I) and undetectable by coomassie blue staining of lysates subjected to SDS polyacylamide gel electrophoresis (fig. 5A1). The RIA data indicate that only a few hundred molecules of SMC per cell accumulate.

Construction and expression of SMC-lac z fusions in pUC8

To investigate the reasons for poor expression of the SMC gene, we constructed a gene fusion in which the beginning of the SMC gene was fused to a portion of the lac z gene present on plasmid pUC8 (15,18). This allowed use of convenient assays for β-galactosidase to measure the synthesis of the fusion protein. We made the fusion by inserting the SMC DNA into the promoter proximal region of the lac z gene; thus expression of the hybrid gene was actually under control of the lac promoter, present on pUC8. Ribosomes were expected to initiate translation both at the beginning of the lac z gene, in which case they would quickly encounter an in-frame stop codon in the region of inserted DNA, or, more distal, at the beginning of the SMC gene in which case they would continue through both the SMC and β-galactosidase sequences to produce the fusion protein.

To construct the fusion (fig. 3) $pPLmuSMC_{ori}$ was digested with EcoR I and BamH I to release a 165 bp fragment containing the ribosome binding site and SMC A region. This was ligated to pUC8, digested with the same two enzymes. The resulting plasmid $pUCmuSMC_{ori}$, was introduced into strain JM83 and plated on LB-agar containing the chromogenic substrate, X-GAL to give white colonies after 16 hours at 37°C. This indicated little or no β-galactosidase activity in the cells, a result consistent with the low

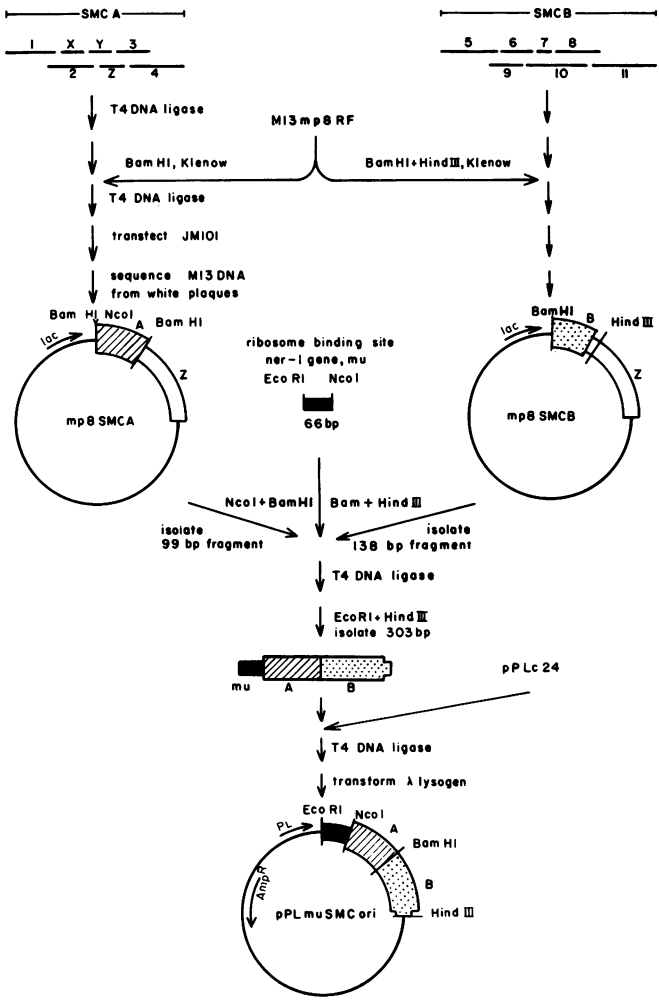


Figure 2. Construction of pPLmuSMC$_{ori}$. The M13mp8 vector was prepared by treating 2 ug RF DNA with BamH I followed by reaction with 1 unit of Klenow polymerase, conditions as described in fig. 1. After ethanol precipitation 20 units of calf intestinal phosphatase were aded in 10mM Tris HCl pH 9.2, 0.2 mM EDTA for 30 minutes. The vector was extracted twice with phenol, precipitated and dissolved in 10mM Tris HCl pH 7.2, 1mM EDTA. 0.1 pmole of Fragment A was ligated with 20 ng linear M13mp8 vector in 10 ul ligation buffer (13) with 40 units of T4 DNA polymerase at 15°C for 24 hours. JM101 was transfected with the ligation mixture and plated onto L-broth plates with X-GAL. Phage DNA was purified from 24 white plaques and sequenced by the chain termination method (18). Fragment B was incorporated into the blunt-end Hind III-BamH I vector under similiar conditions. BamH I and Hind III sites were reformed as well as the Nco I site, GGATCCATGG in Fragment A by the ligations. Intracellular RF DNA was prepared for mp8SMCA

and mp8SMCB and the Nco I-BamH I and BamH I-Hind III fragments were isolated, respectively by gel electrophoresis (see METHODS). They were ligated with the 67 bp EcoR I-Nco I fragment (29) in a 3 fragment ligation. The desired ligation product, the 303 bp EcoR I-Hind III fragment was isolated as above and ligated with pLc24 (30) vector, prepared with the same enzymes.

level of SMC observed in the previous experiment. In contrast, analogous plasmid constructions in which genes encoding human glucagon or bovine growth hormone were fused to lac z led to formation of blue colonies which contained high levels of β-galactosidase (G. Buell, unpublished results). However, the growth hormone and glucagon DNA fragments, used, were excised from P_L plasmids that promoted high-level expression of the respective proteins.

These results suggested to us that the low level of SMC expression was due to inefficient translation of the SMC N-terminal region. They also suggested that we might be able to find changes in the SMC sequence that permitted more efficient expression by screening for fusion plasmids that conferred a blue-colony phenotype. To avoid searching among spontaneous mutants of $pUCmuSMC_{ori}$, we generated a large number of specific variants using synthetic DNA. We synthesized a mixture of short oligodeoxynucleotides that included the 256 possible sequences encoding amino acids 2-6 of SMC (fig. 3, center). These were introduced into the SMC-lac z fusion by ligation to a fragment encoding the mu ribosome binding site plus amino acid 1 (70 bp EcoR I-Ava II) and a fragment encoding amino acids 7-32 of SMC (71 bp Ava II-BamH I). The resulting EcoR I-BamH I ligation products were inserted between the EcoR I and BamH I sites of pUC8 (fig. 3). The mixture was used to transform JM83, and 5000 colonies were obtained on X-GAL plus ampicillin plates. Approximately 10% of the colonies were pale blue.

We selected a number of these colonies, both blue and white, for more extensive analysis. The analysis included a more quantitative measurement of β-galactosidase activity, determination of the nucleotide sequence at the beginning of the SMC gene, and substitution of the variant sequence for the original in $pPLmuSMC_{ori}$ followed by measurement of the effect on SMC expression.

The results obtained with 10 blue and 4 white colonies are shown in Table I. Cells containing plasmids conferring the blue colony phenotype, pUCmuSMC 1-10 contain 2.5-8x more units of β-galactosidase than

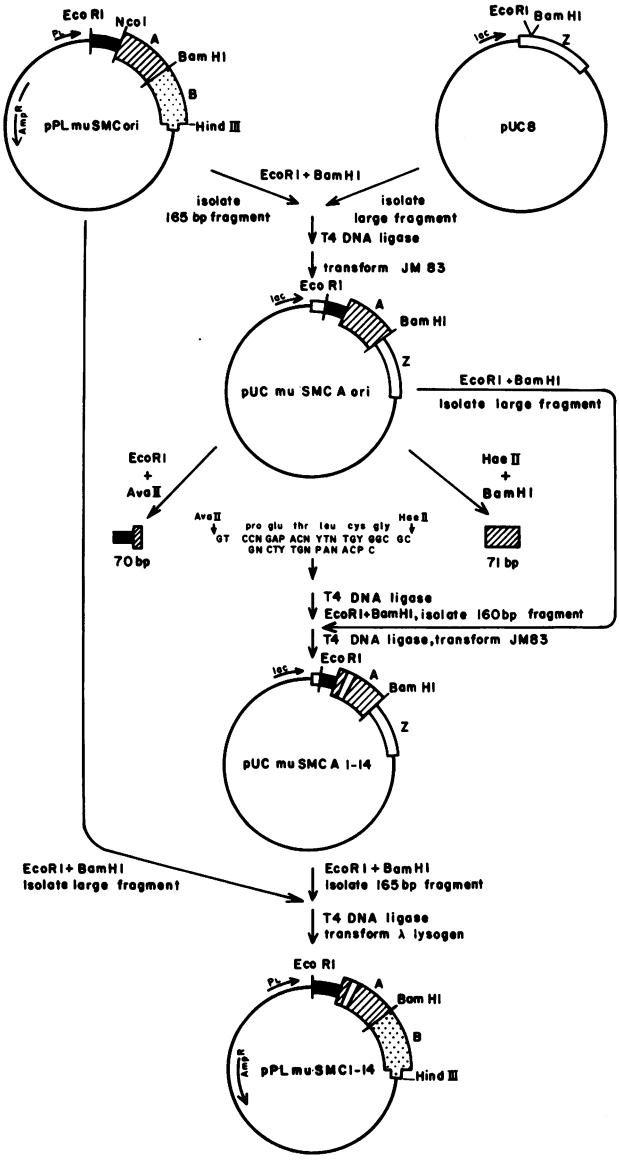


Figure 3. Construction of SMC-lac z fusion plasmids. Each of the DNA fragment isolations was done by restriction digestion, gel electrophoresis on 10% polyacrylamide (except "large fragments" from pPLmuSMC$_{ori}$, and pUCmuSMC$_{ori}$, which were isolated from 1% agarose) and electroelution as stated in METHODS. The 256x degenerate synthetic DNA linker shown at the center of the figure was ligated without prior phosphorylation to avoid linker concatemers. N = all 4 base possibilities, P = purines and Y = pyrimidines. " λ lysogen" refers to C600 λ^+.

TABLE I

plasmid	2 3 4 5 6 pro.glu.thr.leu.cys	β-gal activity units/cell (JM83)	SMC ng/10^7 cells (HB101)
original seq	x x x x x x		
pUCmuSMCA$_{ori}$	CCA GAA ACC CTG TGC	0.4	1.4
blue colonies			
pUCmuSMCA 1	CCC GAA ACT CTG TGT	3.1	33
2	CCT GAA ACT TTG TGC	2.6	45
3	CCA GAG ACG TTG TGC	0.9	35
4	CCA GAG ACG TTG TGT	0.9	43
5	CCT GAA ACT TTG TGT	2.9	33
6	CCT GAG ACG TTG TGT	1.2	58
7	CCG GAA ACG TTA TGT	1.9	50
8	CCG GAA ACA TTG TGT	1.2	65
9	CCA GAA ACG TTG TGT	1.1	32
10	CCT GAG ACT CTA TGT	2.3	42
white colonies			
pUCmuSMCA11	CCC GAA ACC CTC TGT	<0.1	0.10
12	CCT GAA ACC CTC TGT	<0.1	0.11
13	CCG GAA ACC CTC TGT	<0.1	0.10
14	CCA GAA ACC CTC TGT	<0.1	0.09

JM83/pUCmuSMCA$_{ori}$ (the original sequence). These plasmids contain base changes, derived from the synthetic oligodeoxynucleotide, which increase SMC expression 23-46x upon substitution into pPLmuSMC$_{ori}$. The cells carrying pUCmuSMCA 11-14 (white colony phenotype) made no detectable β-galactosidase and contain base changes that reduce SMC expression 10x upon introduction into pPLmuSMC$_{ori}$. Although, there is no direct proportionality between units of β-galactosidase for a particular fusion and SMC yield for the P_L plasmid bearing the same base changes, these two parameters show a positive correlation (Table I). More importantly, these data verify that the color reaction on X-GAL indicator plates can distinquish between base changes that increase or decrease SMC synthesis.

Throughout this work, we monitored the likely secondary structures of the RNA, encoding the N-terminal SMC residues in the plasmids shown in Table I. The minimum free energy conformation for a portion of an RNA molecule was determined with a published computer program (32). We suspect that the various codon choices shown for pUCmuSMCA 1-14 manifest their effects on β-galactosidase activity by forming different RNA "stem and loop" structures that effect translation initiation (33). Figure 4 shows one potential RNA folding for the mu ribosome-binding site region and the first

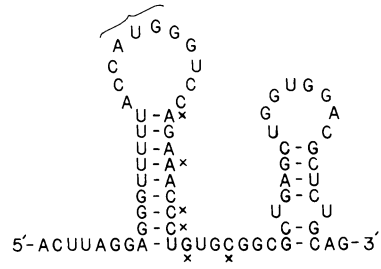


Figure 4. Possible RNA secondary structure for pPLmuSMC$_{ori}$. The model was derived by computer aided analysis as described (32). Bases marked with an x either to their right or below, were varied in the synthetic linker population (fig. 3). Initiating AUG is identified at the top of the left loop.

16 codons in pUCmuSMC$_{ori}$. The six bases which varied in the synthetic linker population are marked. Four of the variant positions are hydrogen-bonded in a 9 bp stem that includes the last three bases of the ribosome binding site, GGAGG . According to this model, of the 6 degenerate positions only those in the 9 bp stem should effect translation. The data in Table I are consistent with this prediction (see DISCUSSION).

Expression of SMC in lon and htpR mutants of E. coli.

The procedure we used to examine SMC expression required growth at 42°C in order to inactivate the temperature-sensitive repressor and thus induce expression. Recent work has shown that transfer of E. coli to 42°C induces expression of a series of genes, "heat shock genes" (34), at least one of which, lon encodes a protease (10,11). For this reason we decided to examine the effect on SMC expression of mutations in lon and in a gene, htpR, whose product is required for induction of the heat shock response (12). We introduced pPLmuSMC2 plus pcI$_{857}$ into strains, isogenic except for the alleles in question, and then performed temperature inductions in parallel. Figure 5B shows SMC accumulation after a 2 hour induction of five different strains, each bearing pPLmuSMC2 and pcI$_{857}$. By RIA analysis the lysates of HB101 (lane 1), lon$^+$htpR$^+$ (lane 2), lon htpR$^+$ (lane 3), lon$^+$htpR (lane 4) and lon htpR (lane 5) contained SMC in the ratios of 1 : 1 : 2.6 : 2.3 : 4.1, respectively. The double mutant, lon htpR, accumulated the most SMC and the two mutations are additive in their effects. pPLmuSMC$_{ori}$ was also tested in the same five strains of E. coli (fig. 5A). Similar effects of the lon and htpR alleles and their additive nature were also seen with this plasmid. But strain comparisons with pPLmuSMC$_{ori}$ demonstrate,

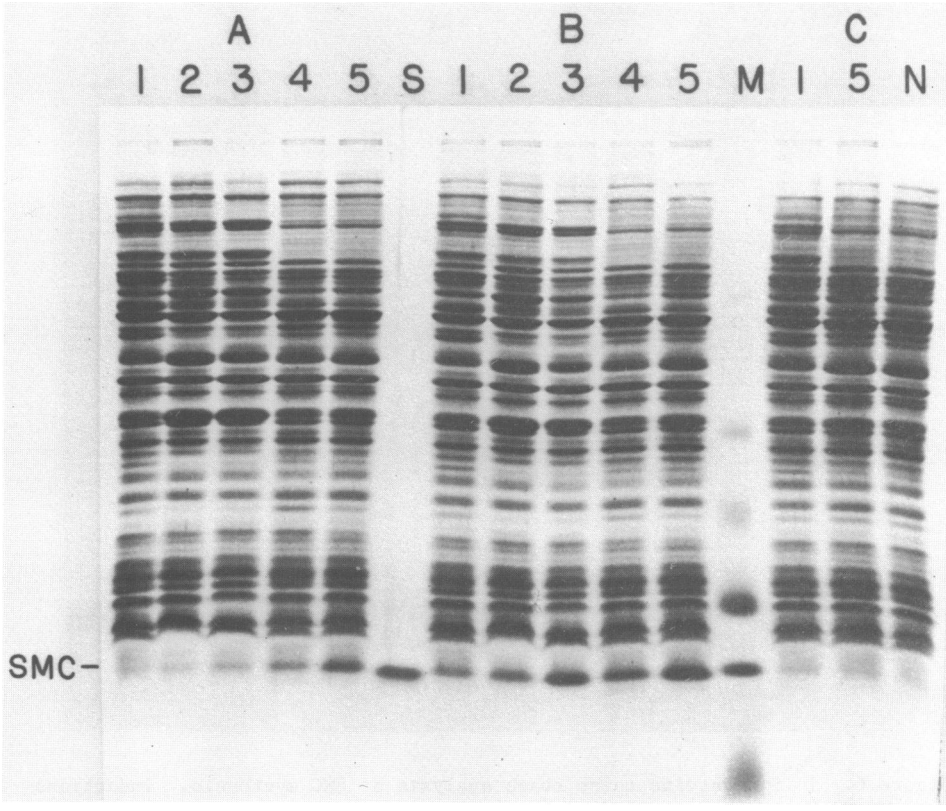


Figure 5. Coomassie blue stained SDS-polyacyrlamide gel of E. coli extracts containing SMC. All cells were transformed with pcI_{857} plus either $pPLmuSMC_{ori}$ (A), pPLmuSMC2 (B) or pPLmuSMC11 (C). In each panel lane 1 shows HB101; lane 2, SG931; lane 3, SG933 (lon); lane 4, SG934 (htpR) and lane 5, SG936 (lon htpR). Lane S is purified SMC from E. coli, lane M contains molecular weight markers: 43Kd, 25.7Kd, 18.4Kd, 14.3Kd, 12.3Kd, 6.2Kd and 2.3+3.4Kd. Lane N (no SMC-containing plasmid) shows control $SG936/pcI_{857}$.

quantitatively, a larger enhancement of SMC accumulation by the lon htpR host. RIA analysis showed that the relative amounts of SMC in the lysates used for figure 5A were 1 : 1 : 3 : 15 : 33, respectively, for the strains in lanes 1-5. We also examined one of the white colony plasmids, pUCmuSMC 11, in HB101 and in the double mutant. As shown in figure 5C no SMC was detectable in either strain. This conslusion was confirmed by RIA (data not shown).

Whereas HB101 bearing pPLmuSMC2 or $pPLmuSMC_{ori}$ differ by 32x in their

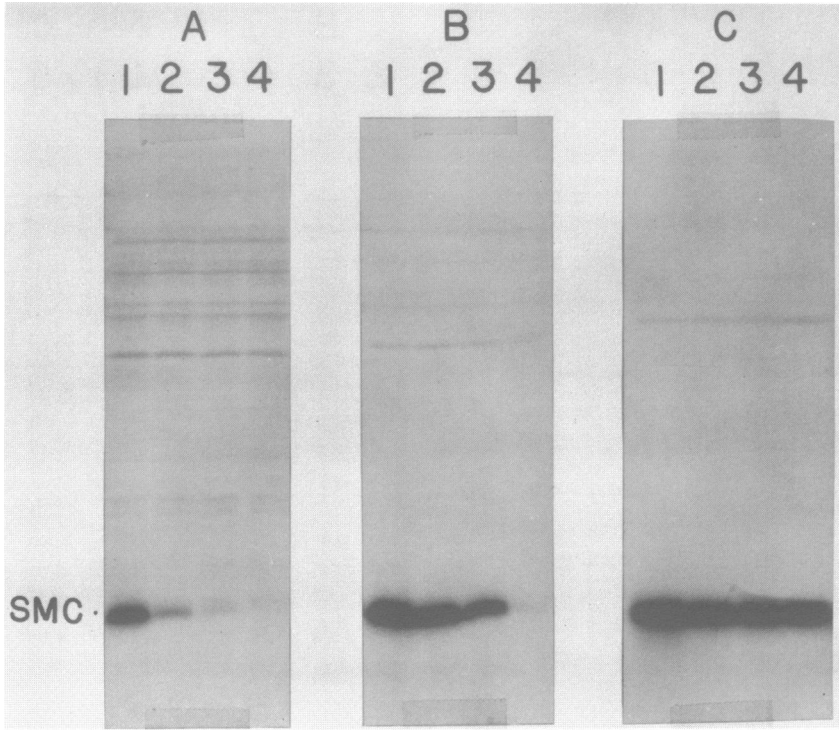


Figure 6. [^{35}S]-cysteine pulse chase analysis of SMC synthesis. Inductions were carried out as described in METHODS except in M9 minimal media (20) lacking cysteine. Cells were induced at 42°C for 1 hour, pulsed for 30 seconds with [^{35}S]-cysteine (3uCi/ml) ,lanes 1 and chased with 100 ug/ml cysteine for 2 minutes, lanes 2; 5 minutes, lanes 3 and 60 minutes, lanes 4. Strains are: panel A, pPLmuSMC$_{ori}$/HB101; panel B, pPLmuSMC2/HB101 and panel C, pPLmuSMC2/SG936. All cells also contained pcI$_{857}$. Samples were precipitated with 10% trichloroacetic acid, dissolved in gel sample buffer (21) and analyzed by electrophoresis and autoradiography.

accumulation of SMC (Table I), lon htpR cells differ by only 4-5x when compared with the same two plasmids (lane 5A versus lane 5B). We conclude that pPLmuSMC2 and pPLmuSMC$_{ori}$ support different rates of SMC synthesis and that these rates are more accurately reflected by the levels of SMC found in the relatively protease deficient environment of lon htpR E. coli.

This conclusion is supported by the pulse-chase study shown in figure 6. Cells were grown overnight at 28°C in M9 minimal medium with all amino acids, except cysteine, followed by induction at 42°C in the same media. The growing cultures were labelled for 30 seconds with [^{35}S]-cysteine (lanes 1) then chased by addition of an excess of unlabeled

cysteine to the culture. Inductions of cells carrying pPLmuSMC$_{ori}$ (panel A) or pPLmuSMC2 (panel B) are compared with HB101. As seen in the gel autoradiogram and verified by scanning densitometry (not shown), HB101 bearing pPLmuSMC2 incorporates about 4x more [^{35}S]-cysteine into SMC than the parallel culture of HB101/pPLmuSMC$_{ori}$. The half-life of SMC appears to be 5 minutes or less in this host and SMC is almost undetectable after a 60 minutes chase. In contrast SMC in the lon htpR host (panel C) has a half-life of approximately 60 minutes.

DISCUSSION

Our initial strategy for the synthesis of SMC in bacteria was based on an SMC synthetic gene which employed codons that are found most frequently in genes of E. coli (25). Our codon choices were also based on the relative abundancies of iso-acceptor tRNAs (26) and, where possible, the recognition sites of several restriction nucleases we thought might be of use in subsequent manipulation of the gene. We then assembled the gene from a series of overlapping, partially complementary oligodeoxynucleotides, a technique previously used for the construction of synthetic genes (35,36). The gene was joined to a fragment specifying a ribosome binding site from bacteriophage mu and inserted into an expression vector carrying the leftward promoter of bacteriophage λ.

Despite the theoretical soundness of this approach, only a small amount of SMC accumulated following induction of the gene in cells carrying the plasmid. Experiments not described here showed that this poor expression was if anything, worse when several other combinations of promoter and ribosome binding sites were tested. These results suggested that some feature of the structural gene itself was at fault. Thus, we constructed a series of variant SMC genes all of which still encoded the normal human amino acid sequence. These were first screened indirectly, using gene fusion techniques. The two methods employed, site-directed mutagenesis with "polylinkers" and gene fusion have both been used before to maximize gene expression (6-8,37). We believe this to be the first time that the two have been combined to examine the role of codon choice in gene expression.

In other SMC-lac z fusion experiments, we tested different ribsome-binding sites and varied the length of the portion of SMC fused to lac z. We found cases in which the starting β-galactosidase activity was too high to permit detection of favorable variants in the polylinker preparation and others in which only white colonies were generated when the synthetic

linkers were substituted into the starting plasmid (data not shown). The exact choice of these elements will probably be important in employing this combined method to optimize expression of other genes.

To investigate possible sources of the effect of codon choice on expression, we determined possible RNA secondary structures at the beginning of the coding region of the SMC transcript. Figure 4 shows a probable folding for transcripts of the original gene generated by the computer program of Zuker and Stiegler (32). A prominent feature is a stem and loop structure with a double-stranded region of 9 bp. Since this includes three bases of the presumed ribosome binding site (GGAGG), its formation might be expected to interfere with the initiation of translation (33,37). When we examined the effect of base changes in the variant genes (marked x in fig. 4), we saw two striking features. First was the loss of the sequence CCC, formed by the adjacent codons for threonine and leucine (ACN.CTN), in all genes that were high expressors (blue colony phenotype). Second was the presence of the base C at the third position for leucine in all low expressors, but not in the original gene. The effect of these changes is to either decrease (for high expressors) or increase (for low expressors) the contribution of the ribosome binding site to the double stranded structure. This supports the idea that the original gene is expressed poorly because of limited ribosome binding to the message.

Other positions altered in the SMC variants appear less significant in this experiment. If plasmid sequences are compared in pairs we note that pUCmuSMCA 4 and 6 differ only at the third position in the codon for proline, A and T, respectively. Likewise, pairs pUCmuSMCA 3 and 4 and pairs pUCmuSMCA 2 and 5 vary only at the third position for cysteine, C or T. None of these variations have more than a 30% effect on SMC expression. The variable nucleotide in the codon for proline is found at the top of the 9 bp stem, where there would be a small stabilizing effect. The variable nucleotide for cysteine, the most distal position tested, is in a region of single-stranded RNA. Thus, the proposed structure in figure 4 agrees well with the data in Table I.

Our results with mutant host strains indicate that SMC acumulation is not simply determined by the level of gene expression. We have shown that the *lon htpR* genotype is particularly well suited for accumulation of SMC and that this effect is manifested by SMC's longer half-life in this strain (fig. 6). The *htpR* gene product is required for the heat shock response of *E. coli* (12,34). At least thirteen proteins, normally heat inducible by

28°C to 42°C temperature shift, fail to be induced in the htpR mutant (34). The lon gene product, a protease, is one of these proteins, present in increased amounts in heat-shocked cells (10,11). Since the effects of lon and htpR on SMC accumulation are additive (fig. 5) we suspect that at least one other heat shock gene encodes a protease which can degrade SMC.

The lon htpR strain discussed here, SG936, is a derivative of K165, the original nitrosoguanidine-mutagenized strain in which the htpR mutation was found (38). The properties which we have described appear to result from the htpR allele, rather than an unrecognized cryptic mutation, since a lon htpR strain constructed by P1 transducing the htpR locus into the parental strain of K165, SC122 (38) behaves in an identical fashion (data not shown).

ACKNOWLEDGEMENTS

We thank Dietrich Hoerlein for providing pure E. coli derived SMC, and Rene Humbel for providing human IGF-I, Carol Gross for the P1 backcross of htpR and Fred Goldberg and Steve Goff for all SG strains and David Lane for proofreading the manuscript. This work was made possible by the technical assistance of Patricia and Jean-Pierre Voegeli, Elizabeth Schmid, Elizabeth Bader and Mohamed El Alama.

REFERENCES

1. Zapf, J., Froesch, E.R. and Humbel, R.E. (1981) Current Topics in Cell. Reg., 19, 257-309.
2. Schoenle, E., Zapf, J., Humbel, R.E. and Froesch, E.R. (1982) Nature 296 252-253.
3. Klapper, D.G., Svoboda, M.E., Van Wyk, J.J. (1983) Endocrinol. 112 2215-2217.
4. Rinderknecht, E. and Humbel, R.E. (1978) J. Biol. Chem. 253 2769-2776.
5. Blundell,T.L. and Humbel, R.E. (1980) Nature 287 781-787.
6. Guarente, L., Lauer, G., Roberts, T.M. and Ptashne, M. (1980) Cell 20 543-553.
7. Casadaban, M.J., Chou, J. and Cohen, S.N. (1980) J. Bacteriol. 143 971-980.
8. Castilho, B.A., Olfson, P. and Casadaban, M.J. (1984) J. Bacteriol. 158 488-495.
9. Young, R.A. and Davis, R.W. (1983) Proc. Natl. Acad. Sci. USA 80 1194-1198.
10. Phillips, T.A., VanBogelen, R.A., Neidhardt, F.C. (1984) J. Bacteriol. 159 283-287.
11. Goff, S.A., Casson, L.P. and Goldberg, A.L. (1984) Proc. Natl. Acad. Sci. USA in press.
12. Yamamori, T. and Yura, T. (1982) Proc. Natl. Acad. Sci. USA 79 860-864.
13. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning (Cold Spring Harbor Laboratory).
14. Messing, J. and Vieira, J. (1982) Gene 19 269-276.
15. Vieira, J. and Messing, J. (1982) Gene 19 259-268.

16. Ito, H., Ike, Y., Ikuta, S. and Itakura, K. (1982) Nucl. Acids Res. 10 1755-1769.
17. Maniatis, T., Jeffery, A. and van de Sande, H. (1975) Biochem. 14 3787-3794.
18. Messing, J. (1983) Methods in Enzymology 101 20-78.
19. Smith, A.J.H. (1980) Methods in Enzymology 65 560-580.
20. Miller, J.H. (1972) Experiments in Molecular Genetics (Cold Spring Harbor).
21. Laemmli, U.K. (1970) Nature (London) 227 680-685.
22. Buell, G.N., Wickens, M.P., Carbon, J. and Schimke, R.T. (1979) J. Biol. Chem. 254 9277-9283.
23. Mandel, M. and Higa, A. (1970) J. Mol. Biol. 53 154.
24. Maxam, A.M. and Gilbert, W. (1980) Methods in Enzymology 65 499-560.
25. Grantham, R., Gautier, C. and Gouy, M. (1980) Nucl. Acids Res. 8 1983-1992.
26. Ikemura, T. (1981) J. Mol. Biol. 151 389-409.
27. Shine, J. and Dalgarno, L. (1975) Nature 254 34-38.
28. Gray, G., Selzer, G., Buell, G. Escanez, S., Hofer, S., Voegeli, P. and Thompson, C. (1984) Gene 32 in press.
29. Priess, H., Kamp, D., Kahmann, R., Brauer, B. and Delius, H. (1982) Mol. Gen. Genet. 186 315-321.
30. Remaut, E., Stanssens, P. and Fiers, W. (1981) Gene 15 81-93.
31. Remaut, E., Tsao, H. and Fiers, W. (1983) Gene 22 103-113.
32. Zuker, M. and Stiegler, P. (1981) Nucl. Acids Res. 9 133-148.
33. Iserentant, D. and Fiers, W. (1980) Gene 9 1-12.
34. Neidhardt, F.C. and VanBogelen, R.A. (1981) Biochem. Biophy. Res. Comm. 100 894-900.
35. Crea, R., Kraszewski, A., Hirose, T. and Itakura, K. (1978) Proc. Natl. Acad. Sci. USA 75 5765-5769.
36. Smith, J., Cook, E., Fotheringham, I., Pheby, S., Derbyshire, R., Eaton, M.A.W., Doel, M., Lilley, D.M.J., Pardon, J.F., Patel, T., Lewis, H. and Bell, L.D. (1982) Nucl. Acids Res. 10 4467-4482.
37. Tessier, L., Sondermeyer, P., Faure, T., Dreyer, D., Benavente, A., Villeval, D., Courtney, M. and Lecocq, J. (1984) Nucl. Acids Res. 12 7663-7676.
38. Cooper, S. and Ruettinger, T. (1975) Mol. Gen. Genet. 139 167-176.