Articles

Expression of Human Insulin-like Growth Factor I in Bacteria: Use of Optimized Gene Fusion Vectors To Facilitate Protein Purification†

Tomas Moks,‡ Lars Abrahmsén,‡ Erik Holmgren,§ Margareta Bilich,‖ Anders Olsson,‡ Mathias Uhlén,*,‡ Gunnar Pohl,§ Catharina Sterky,§ Hans Hultberg,§ Staffan Josephson,§ Arne Holmgren,‖ Hans Jörnvall,‖ and Björn Nilsson‡

Department of Biochemistry, Royal Institute of Technology, S-100 44 Stockholm, Sweden, Kabigen AB, S-112 87 Stockholm, Sweden, and Department of Chemistry I, Karolinska Institute, S-104 01 Stockholm, Sweden

Received January 30, 1987

ABSTRACT: Several fusions between the gene for human insulin-like growth factor I (IGF-I) and the genes for different IgG-binding fragments of staphylococcal protein A were assembled and compared regarding expression, secretion, and purification of the peptide hormone. After IgG affinity purification of the fusion proteins from the growth medium of *Staphylococcus aureus* or *Escherichia coli*, native IGF-I was released by cleavage of an Asn–Gly peptide bond with hydroxylamine. An optimized expression system based on a modified synthetic IgG-binding domain (z), resistant to hydroxylamine, gave the highest yield of fusion protein. After cleavage, the hormone could be separated from the IgG-binding moiety and from noncleaved fusion protein by a second passage through the IgG affinity column. The biological activity and the purity of the IGF-I obtained were confirmed by a radioreceptor assay, N-terminal sequence analysis, polyacrylamide gel electrophoresis, isoelectric focusing, and high-performance liquid chromatography.

The construction and use of fusion vectors based on the gene for staphylococcal protein A have previously been described (Nilsson et al., 1985a,b). The expression system is based on the "pseudoimmune" interaction between the constant region of immunoglobulins and staphylococcal protein A (Uhlén et al., 1984). The dissociation constant for this interactin has been determined to be 2×10^{-8} M (Langone, 1982), allowing fusion protein products to be rapidly recovered from crude lysates by affinity chromatography. The protein A fusion concept has been used to immobilize enzymes with high efficiency (>95%) (Nilsson et al., 1985b) and to secrete human insulin-like growth factor I (IGF-I) into the culture medium of *Staphylococcus aureus* (Nilsson et al., 1985a). The IGF-I fusion protein was then preceded by the signal peptide and all five IgG-binding regions of protein A. After affinity purification, IGF-I was released by acid cleavage of an Asp–Pro bond. The peptide hormone was biologically active but lacked the N-terminal glycine residue.

Recently the system was improved. It was demonstrated that fusions containing only two (EE or EB) of the five (E, D, A, B, and C) IgG-binding domains of protein A were secreted to the growth medium of *Escherichia coli* (Abrahmsén et al., 1985, 1986). Normally, proteins secreted in other systems are retained in the periplasmic space when expressed in this host (Oliver, 1985). In this study, we compare different systems for bacterial expression of a mammalian peptide hormone and present an optimized version based on a synthetic IgG-binding domain (Z). Native human IGF-I with high specific activity was obtained in a scheme involving hydroxylamine treatment (Bernstein & Balian, 1977) which is a specific chemical cleavage method apparently not previously utilized for fusion proteins.

Materials and Methods

Bacterial Strain and Plasmids. *E. coli* HB101 (Boyer & Roulland-Dussoix, 1969) and *S. aureus* SA113 (Iordanescu, 1975) were used as bacterial hosts. Plasmid and phage vectors were pRIT4 (Nilsson et al., 1985b), pRIT15 (Abrahmsén et al., 1986), pEMBL8, pEMBL9 (Dente et al., 1983), pUC8 (Yanisch-Perron et al., 1985), and M13mp8 (Norrander et al., 1983).

DNA Constructions and Cell Growth. Restriction enzymes (Pharmacia, New England Biolabs, and Boehringer Mannheim) and T4 DNA ligase (Pharmacia) were used according to the recommendation of the supplier. The oligonucleotide dGTGAATTCTAACGGTCCCGAAACT was synthesized with an automated machine (KabiGen AB, Sweden) as described (Chow et al., 1981; Elmblad et al., 1982a). The deprotected oligomer was purified by polyacrylamide gel electrophoresis. *E. coli* and *S. aureus* were transformed and grown as described (Nilsson et al., 1985a; Abrahmsén et al., 1986). After growth, the medium was clarified by cross-flow microfiltration and passed through a column of Fast Flow IgG–Sepharose (Pharmacia, Sweden). The gel was washed with 10 bed volumes of TST [50 mM tris(hydroxymethyl)aminomethane (Tris), pH 7.4, 200 mM NaCl, and 0.05% Tween 20] and eluted with 0.5 M acetic acid (titrated to pH 2.8 with ammonium acetate).

Protein Analysis. Protein A was quantified by ELISA using Fab fragments of rabbit anti-protein A antibodies conjugated to β-galactosidase (a kind gift from Dr. M Inganäs, Pharmacia, Sweden) as described (Abrahmsén et al., 1985; Moks et al., 1986). The biological activity of IGF-I was determined by a radioreceptor assay (RRA) using human placenta mem-

† This investigation was supported by grants from the Swedish National Board for Technical Development, the Swedish Natural Research Council, and the Swedish Cancer Society.

‡ Royal Institute of Technology.

§ Kabigen AB.

‖ Karolinska Institute.

0006-2960/87/0426-5239$01.50/0

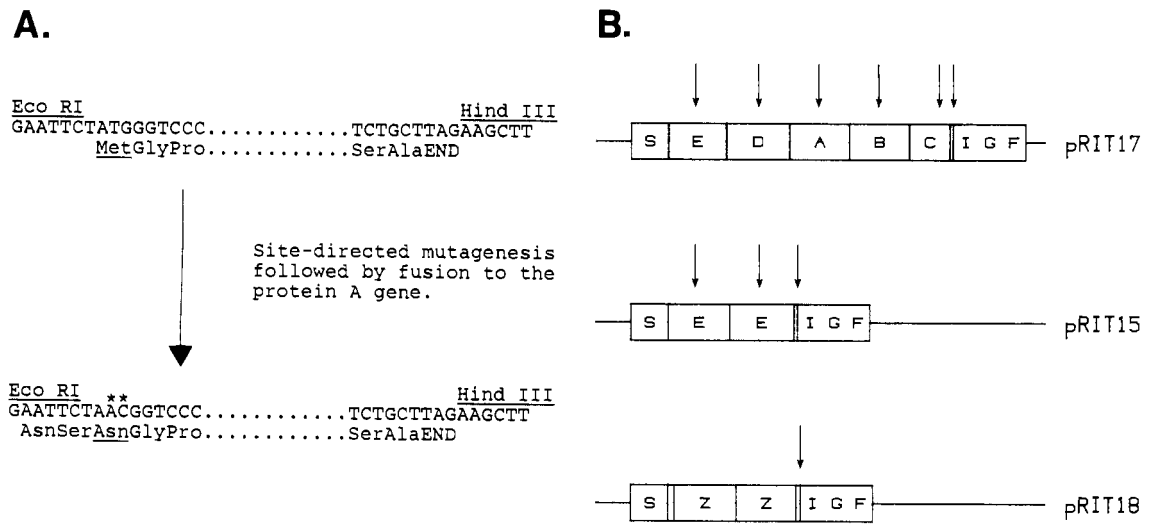

FIGURE 1: Schematic representation of the gene fusion vectors. (A) The strategy for oligonucleotide-mediated in vitro mutagenesis of the synthetic IGF-I gene employed to introduce a site-specific (Asn-Gly) cleavage site at the N-terminus of the IGF-I molecule. Changed nucleotides are indicated with asterisks, and the replaced amino acid residue is underlined. (B) Schematic drawing of relevant parts of the gene fusion vectors used. The arrows indicate possible cleavage sites (Asn-Gly) for hydroxylamine. The signal peptide (S), the IgG-binding domains E, D, A, B, and C of protein A, and the synthetic IgG-binding domain Z are indicated. The nonlabeled box after the signal peptide in pRIT18 represents the six residues of region E that were required for an efficient secretion in *E. coli* (Abrahmsén et al., 1985).

branes and pooled human serum as reference (Hall et al., 1974). Total compositions were obtained by amino acid analysis on a Beckman 121M instrument after acid hydrolysis with 6 M HCl/0.5% phenol in evaculated tubes for 24 h at 110 °C. Amino acid sequence analysis was carried out by liquid-phase sequencer degradations in a Beckman 890D sequencer using glycine-precycled polybrene (Jörnvall & Philipson, 1980). Isoelectric focusing was performed as titration in a polyacrylamide gel (LKB, Sweden). About 100 μg of protein was dissovled in 100 μL of water and applied in the sample trench. After electrophoresis, the pH in the gel was measured along the trench using a surface pH electrode, after which the gel was fixed and stained (Righetti & Chillenni, 1978). Reverse-phase high-performance liquid chromatography (HPLC) was performed by using a two-pump system for gradient elution. The protein (20 μg) was applied to a column (TSK ODS-120T, LKB, Sweden), equilibrated with 0.1% trifluoroacetic acid in distilled water, and eluted at a flow rate of 1 mL/min with a linear gradient of acetonitrile.

Hydroxylamine Cleavage. A solution of 2 M hydroxylamine hydrochloride and 0.2 M Tris base was titrated to pH 9.0 at 45 °C with NaOH. The protein was dissolved in this solution to a concentration of 5 mg/mL. After a 4-h incubation at 45 °C, the reaction was terminated by lowering the temperature and by adjusting the pH with HCl to below 8.0. The buffer was exchanged by gel chromatography on Sephadex G-25 (Pharmacia).

RESULTS

Site-Directed Mutagenesis of the Synthetic IGF-I Gene. Human IGF-I contains an N-terminal glycine residue. Since treatment with hydroxylamine cleaves the Asn–Gly peptide bond (Bornstein & Balian, 1977), we decided to introduce a codon for asparagine before the one for the N-terminal glycine in the synthetic IGF-I gene fragment (Elmblad et al., 1982b). The synthetic gene of 240 base pairs spans from an *Eco*RI to a *Hin*dIII site. A few base pairs downstream from the *Eco*RI site, a potential ATG start codon is followed by the codons for native IGF-I (Figure 1A). Oligonucleotide-mediated mutagenesis (Nilsson et al., 1985a) was performed to change this ATG to AAC which codes for asparagine. The oligonucleotide used in the mutagenesis spans both the mismatches and the *Eco*RI site (Figure 1A). After extension of the mutagenesis primer, the IGF-I gene was cleaved out, using *Eco*RI and *Hin*dIII, followed by recloning into plasmid pUC8. Approximately 25% of the transformants contained the desired mutation as shown by oligonucleotide hybridization followed by DNA sequencing.

Table I: Comparison of Gene Fusion Products Containing Human IGF-I[a]

plasmid	IgG-binding fragments	host	[IGF-I] (mg/L)	localization	Asn-Gly sites
pRIT17	EDABC	*E. coli*	1	periplasm	6
pRIT17	EDABC	*S. aureus*	2	medium	6
pRIT15	EE	*E. coli*	20	medium	3
pRIT18	ZZ	*E. coli*	30	medium	1

[a] Cells were grown in a 10-L fermentor using standard conditions (Abrahmsén et al., 1986). The IgG-binding fragments are defined as described in the legend to Figure 1. The amount of IGF-I is calculated from the protein A content as measured by ELISA (Moks et al., 1986). The localization was as described (Abrahmsén et al., 1985).

Expression of the Gene Fusion EDABC–IGF-I (pRIT17) in S. aureus and E. coli. The modified IGF-I gene was cloned into the plasmid pRIT4 (nilsson et al., 1985b) to yield pRIT17. This expression vector contains the promoter and the protein A gene that code for the signal peptide and all five IgG-binding domains (E, D, A, B, and C) fused to the modified IGF-I (Figure 1B). The bifunctional plasmid pRIT17 was introduced into *E. coli* and *S. aureus* by standard precedures (Nilsson et al., 1985a). The protein A–IGF-I fusion protein was found to be expressed and efficiently transported across the cytoplasmic membrane in both species, to become periplasmic in *E. coli* and extracellular in *S. aureus* (Table I). The level of expression was higher and the proteolytic degradation was lower in *S. aureus*, suggesting that, for production of this gene product, *S. aureus* is a better host than *E. coli*.

The fusion protein was purified from the culture medium of *S. aureus* by IgG affinity chromatography and cleaved with hydroxylamine. The predicted cleavage products of the fusion protein should, in addition to the 7.7-kilodalton (kDa) IGF-I moiety, yield six protein A peptides of 6.8, 6.6, 6.6, 3.0, and 1.5 kDa, respectively (Figure 1B). Analysis by sodium dodecyl sulfate (SDS)–polyacrylamide gel electrophoresis confirms that the full-length fusion protein disappears (Figure 2A, lane 1)

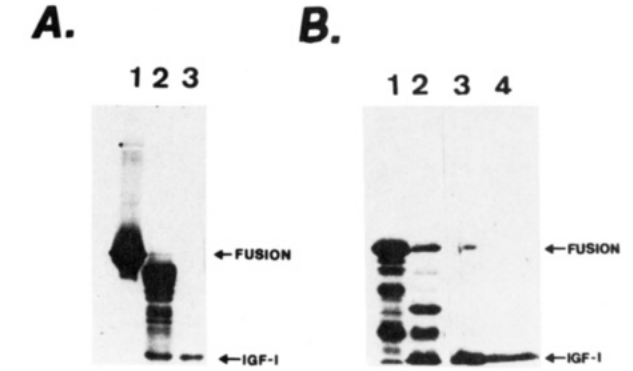

FIGURE 2: Fusion proteins (FUSION) EDABC–IGF-I and EE–IGF-I analyzed by SDS–polyacrylamide gel electrophoresis before and after cleavage for 4 h with 2 M hydroxylamine at 45 °C, pH 9.0 (A) EDABC–IGF-I encoded by pRIT17: lane 1, affinity purified from the culture medium of *S. aureus*; lane 2; after cleavage with hydroxylamine; lane 3, after purification by ion-exchange chromatography and gel filtration. (B) EE–IGF-I encoded by pRIT15: lane 1, affinity purified from the culture medium of *E. coli*; lane 2, after cleavage with hydroxylamine; lane 3, after purification with CM-Sepharose; lane 4, after purification with Mono S FPLC.

and that smaller fragments appear (lane 2). However, except for a distinct band corresponding in size with native IGF-I, most of the bands correspond to partial cleavage products, suggesting that the Asn-Gly structures inside the IgG-binding domains are relatively resistant to hydroxylamine.

The native IGF-I was purified from the cleavage mixture by ion-exchange chromatography on CM-Sepharose and Mono S fast protein liquid chromatography (FPLC) followed by reverse-phase HPLC to yield a 95% pure peptide hormone as determined by SDS-polyacrylamide gel electrophoresis (Figure 2A, lane 3). The biological activity was approximately 7000 units/mg (not shown), i.e., about the same as reported for the native hormone purified from human serum (Hall et al., 1974).

Expression of the Gene Fusion EE–IGF-I (pRIT15) in E. coli. Recently, an expression system allowing secretion of heterologous gene products to the culture medium of *E. coli* was described (Abrahmsén et al., 1986). The fusion protein, encoded by plasmid pRIT15, consisted of the signal peptide and two IgG-binding domains (EE) from protein A fused to human IGF-I. About 85% of the fusion protein EE–IGF-I, schematically shown in Figure 1B, was translocated to the culture medium (Table I).

The clarified medium, from the *E. coli* cells containing pRIT15, was passed through an IgG affinity column. The purity of the fusion protein was analyzed by SDS-polyacrylamide gel electrophoresis. The slowest migrating band, corresponding in size to a full-length fusion protein (Figure 2B, lane 1), was the major band. Smaller sized proteins could also be observed, probably arising from proteolytic degradation of the heterologous fusion protein.

The predicted degradation products from the hydroxylamine cleavage of EE–IGF-I are three peptides of molecular weights 3.0K 6.7K, and 3.9K from the protein A moiety, apart from the IGF-I moiety of M_r 7.7K (Figure 1B). SDS–polyacrylamide gel electrophoresis showed bands of the expected sizes plus bands corresponding to partial cleavage products (Figure 2B, lane 2). The cleaved material was further purified by ion-exchange chromatography on CM-Sepharose and Mono S FPLC to yield a final IGF-I product with a specific activity of around 10000 units/mg. Analysis by SDS–polyacrylamide gel electrophoresis revealed a single band corresponding in size to IGF-I (Figure 2B, lane 4).

Construction of the Gene Fusion Vector pRIT18. We have recently described the design of a synthetic gene fragment

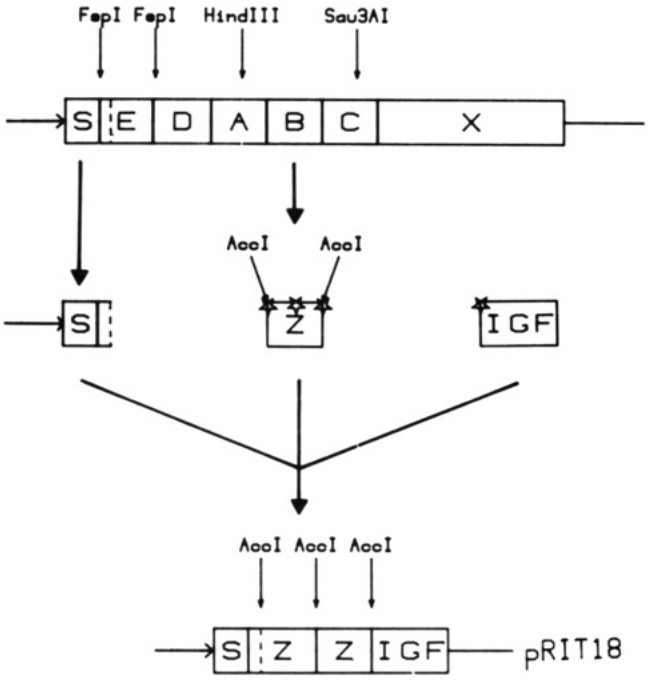

FIGURE 3: Strategy for the construction of the gene fusion vector pRIT18. Stars represent amino acid changes as compared to the domain B sequence of native protein A (Uhlén et al., 1984) or the IGF-I gene fragment. S is the signal sequence, E, D, A, B, and C are the IgG-binding regions, X is the region responsible for the association to the cell wall, Z is the engineered IgG-binding region, and IGF the IGF-I gene.

encoding an IgG-binding domain (Nilsson et al., 1986). On the basis of computer predictions, the Asn-Gly sequence, present in all five IgG-binding domains of native protein A, was replaced by Asn-Ala. Thereby, this fragment will code for a protein resistant to hydroxylamine treatment. It was used in the assembly of a gene fusion consisting of three parts (Figure 3). The first contains the promoter, the signal sequence, and codons for the first six amino acids of domain E, i.e., the N-terminal portion of mature protein A. These six residues separate domain E from the other domains and are necessary for correct processing of the signal peptide (Abrahmsén et al., 1985). The second part is the synthetic Z fragment, and the third is the mutagenized synthetic IGF-I gene (Figure 1A). It has earlier been shown that two IgG binding domains have considerably higher affinity for IgG than one domain (Moks et al., 1986). Therefore, with the use of two flanking nonpalindromic *Acc*I sites of the synthetic Z gene fragment, the plasmid vector pRIR18 with two Z fragments was constructed (Figure 1B and 3).

Expression of the Gene Fusion ZZ–IGF-I (pRIT18) in E. coli. E. coli cells containing plasmid pRIT18 (Figure 1B) were grown in a 10-L fermentor. A production of approximately 80 mg/L of fusion protein, corresponding to 30 mg/L of IGF-I, was obtained. The culture medium was recovered by using cross-flow microfiltration and passed through an IgG–Sepharose column (Nilsson et al., 1985a). The column was eluted with 0.5 M acetic acid, pH 2.8, and the eluted material was lyophilized and analyzed by SDS–polyacrylamide gel electrophoresis. One major band corresponding in size to ZZ–IGF-I (Figure 4, lane 1) and several minor putative degradation products were observed.

The fusion protein was treated with hydroxylamine, and aliquots were taken, desalted, and analyzed by electrophoresis (Figure 4, lanes 2–5). This demonstrated that the full-length fusion protein band disappeared, giving rise to two bands of smaller sizes corresponding to fragment ZZ (14 kDa) and

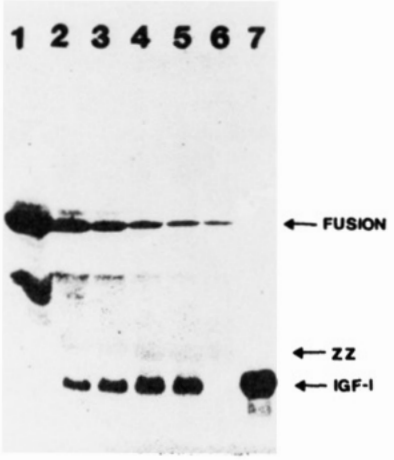

FIGURE 4: Fusion protein ZZ–IGF-I (FUSION) encoded by pRIT18 and the cleavage products after hydroxylamine treatment, analyzed by SDS–polyacrylamide gel electrophoresis. Lane 1, fusion protein after IgG affinity purification; Lanes 2–5, cleavage products after treatment with hydroxylamine for 1, 2, 3, and 4 h, respectively; lane 6, material bound to the IgG affinity column after 4-h cleavage; lane 7, flow-through of the 4-h cleavage mixture after passage through the IgG affinity column. Note that fragment ZZ is poorly stained with the method employed.

Table II: IGF-I Activity before and after Cleavage of Affinity-Purified ZZ–IGF-I Fusion Protein Encoded by pRIT18[a]

protein fraction	act. (units)	protein content (mg)	sp act. (units/mg)	yield (%)
fusion protein	<1000	72		
cleavage and G-25 desalting	53800	69	780	100
flow-through IgG column	48700	15.7	3100	91
eluted from IgG column	3200	53.3	60	
Sephadex G-43	44100	4.5	9800	82

[a] The IGF-I activity was determined by a radioreceptor assay (Jörnvall & Philipson, 1980). The cleavage was performed in 2 M hydroxylamine (pH 9.0) for 4 h at 45 °C. The purification steps were performed as described in the text.

IGF-I (7.7 kDa). More than 80% of the fusion protein was cleaved after 4 h.

The reaction mixture (Figure 4, lane 5) was desalted and passed through an IgG–Sepharose column. Both the flow-through material and the bound material were collected and analyzed by SDS–polyacrylamide gel electrophoresis. This demonstrated that fragment ZZ and noncleaved fusion protein bound efficiently to the column (lane 6) and that the IGF-I molecules were recovered in the flow-through (lane 7). The purity as determined by electrophoresis was higher than 95%.

The biological activity of IGF-I was determined by the human placenta membrane radioreceptor assay (Hall et al., 1974). The intact fusion protein had no detectable biological activity (Table II). In contrast, the desalted cleavage mixture had a specific activity of approximately 780 units/mg, establishing that functional IGF-I molecules were generated by cleavage.

Purification and Characterization of IGF-I. Little activity was lost when the cleavage mixture was passed through the IgG affinity column (Table II). The specific activity increased to 3100 units/mg as a result of the binding of nonactive fragment ZZ and noncleaved fusion protein to the column (Figure 4, lane 6).

The flow-through from the second passage through the IgG column was subjected to a gel filtration step on a column of Sephadex G-43 special grade, to remove traces of IgG and multimeric forms of IGF-I. The major peak contained IGF-I with a biological activity of 9800 units/mg (Table II).

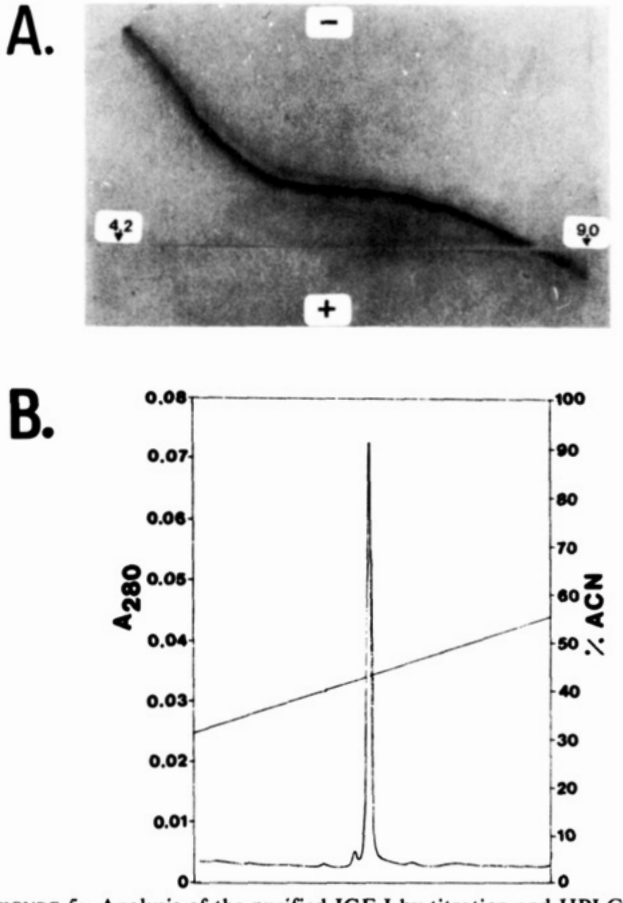

FIGURE 5: Analysis of the purified IGF-I by titration and HPLC. (A) Titration curve of the purified IGF-I. The protein was applied in the sample trench after generation of the pH gradient. Electrophoresis was performed in a second dimension and the pH in the gel along the sample trench measured. The pH values, obtained by a surface electrode at the anode (+) and the cathode (−) sides of the gel, are shown. (B) The whole chromatogram from high-performance liquid chromatography of the purified IGF-I. The sample was eluted from the column with a linear gradient of acetonitrile (%ACN), and the absorbance was monitored at 280 nm (A_{280}).

The purity of the IGF-I recovered was also analyzed by isoelectric focusing, yielding a p*I* of 8.0–8.5 (Figure 5A) which is in agreement with the published value of 8.2 (Humbel, 1984; Froesch et al., 1985). Amino acid analysis after acid hydrolysis showed the expected composition, and the amino acid sequence for 15 cycles of the purified IGF-I revealed a homogeneous N-terminal region identical with the amino acid sequence of native IGF-I. The purified peptide hormone was finally analyzed by reverse-phase HPLC, demonstrating a product of high purity (Figure 5B).

Discussion

Gene fusion systems designed to facilitate the purification of proteins have been developed to enable secretion of heterologous proteins in hosts such as *Bacillus subtilis* (Palva et al., 1983), *Pseudomonas aeruginosa* (Gray et al., 1984), and *Saccharomyces cerevisiae* (Brake et al., 1984). Protein "tags" in the form of polyarginine (Smith et al., 1984), β-galactosidase (Ullmann, 1984; Germino & Bastia, 1984), or protein A (Nilsson et al., 1985a,b) have been used to shift the p*I* or to enable affinity purification of the desired protein. The combination of secretion and affinity purification, as described in this paper, has obvious advantages. *E. coli* contains thousands of intracellular proteins and approximately 100 proteins in the periplasm (Pugsley & Schwartz, 1985). In contrast, less than 10 proteins are normally translocated to the culture

medium. A high level of purification is thus achieved by the secretion of the gene product from the cytoplasm. In addition, the affinity "tail" provides an efficient means for purification with a technically simple procedure.

Human IGF-I with high specific activity was produced with three different gene fusion vectors. All utilized the staphylococcal protein A promoter/signal sequence and a mutagenized IGF-I gene to enable site-specific cleavage with hydroxylamide. However, the IgG-binding part of the fusion protein and the host bacteria varied. This gave a 30-fold difference in recovery but also affected the complexity of the cleavage products (Table I).

The first fusion product encoded by plasmid pRIT17 contains all five IgG-binding domains of protein A. It is efficiently secreted to the culture medium of *S. aureus*, but since this bacterium is potentially pathogenic, the system is less attractive. Another disadvantages is that the hydroxylamine treatment gives a complex mixture of products. It is interesting that the Asn-Gly structure at the junction between protein A and IGF-I seems more susceptible to cleavage than the internal Asn–Gly bonds in each protein A domain (Figure 2A). This may reflect the fact that these structures are stabilized within α-helices (Oliver, 1985).

The second fusion encoded by plasmid pRIT15 contains only two IgG-binding regions of protein A (EE). The main advantage compared with pRIT17 is the efficient secretion to the culture medium. In addition, a 10-fold increase in expression (Table I) and a less complex cleavage pattern after hydroxylamine treatment are obtained. However, also for this fusion protein, the IgG-binding capacity of the affinity tail is destroyed by the cleavage, preventing the optimal purification scheme involving a second affinity chromatography step.

This disadvantage is eliminated by the third fusion encoded by plasmid pRIT18 containing two synthetic IgG-binding domains (ZZ). The lack of Asn-Gly structures in the synthetic fragments makes it possible to utilize the unique processing site at the junction between fragment ZZ and the desired product. The single cleavage site allows recovery of the pure cleavage product by mere collection of the flow-through from a second IgG affinity column.

The nonpalindromic *Acc*I restriction sites flanking the synthesized gene (Figure 3) that encodes the synthetic IgG-binding domain are technically important. They allow construction of expression vectors coding for any number of IgG-binding regions leading to a flexible affinity purification system. By investigation of different multiplicities of the Z fragment (between 1 and 10), a divalent fragment was found to be optimal for the production of human IGF-I (data not shown). Other multiplicities may be used for other products depending on (i) the steric interference of the gene product, (ii) the efficiency of secretion, and (iii) the desired strength of the IgG interaction.

The comparative studies have shown that all three expression systems described are well suited for the production of human insulin-like growth factor I. The highest level of expression was obtained with the synthetic IgG-binding domain, giving 30 mg/L calculated on IGF-I (Table I). The fact that a final product with high specific activity was obtained in all cases suggests that the system allows correct folding of the peptide hormone which contains three disulfide bridges. Obviously, this system is potentially useful for secretion and purification of other gene products. Recently, it has been used to express the *E. coli* human insulin-like growth factor II (B. Hammarberg, unpublished results), human secretin (H. Olson, unpublished results), and several small peptides (B. Löwenadler, unpublished results). The gene products were, in all cases, secreted to the culture medium and could be purified by IgG affinity chromatography.

ACKNOWLEDGMENTS

We thank Dr. Lars-Erik Nyström, Pharmacia AB, Sweden, for providing IgG FF-Sepharose and Sephadex G-43 special grade. We are also grateful to Dr. Lennart Philipson for critical comments and advice throughout the course of this work and to Gerd Benson for skillful secretarial help.

REFERENCES

Abrahmsén, L., Moks, T., Nilsson, B., Hellman, U., & Uhlén, M. (1985) *EMBO J. 4*, 3901–3906.

Abrahmsén, L., Moks, T., Nilsson, B., & Uhlén, M. (1986) *Nucleic Acids Res. 14*, 7487–7500.

Bornstein, P., & Balian, G. (1977) *Methods Enzymol. 67*, 132–145.

Boyer, H. W., & Roulland-Dussoix, D. (1969) *J. Mol. Biol. 41*, 459–472.

Brake, A., Marryweather, J. P., Coit, D. G., Heberlein, V. A., Masiarz, F. R., Mullenbach, G. I., Urclea, M. S., Valenzuela, P., & Barr, P. J. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 4642–4646.

Chow, F., Kempe, T., & Palm, G. (1981) *Nucleic Acids Res. 9*, 2807–2817.

Dente, L., Cesarini, Y., & Cortese, R. (1983) *Nucleic Acids Res. 11*, 1645–1655.

Elmblad, A., Josephson, S., & Palm, G. (1982a) *Nucleic Acids Res. 10*, 3291–3301.

Elmblad, A., Fryklund, L., Heden, L.-O., Holmgren, E., Josephson, S., Lake, M., Löwenadler, B., Palm, G., & Skottner-Lundin, A. (1982b) in *Third European Congress cn Biotechnology III*, pp 287–296, Verlag Chemie, Weinheim, Frg.

Froesch, E. R., Schmid, C., Schwander, J., & Zap, J. (1985) *Annu. Rev. Physiol. 47*, 443–467.

Germino, J., & Bastia, D. (1984) *Proc. Natl. Acad. Sci. U.S.A. 81*, 4692–4696.

Gray, G. L., McKeown, K. A., Jones, A. J. S., Seeburg, P. H., & Heynecker, H. L. (1984) *Bio/Technology 2*, 161–168.

Hall, K., Takano, K., & Fryklund, L. (1974) *J. Clin. Endocrinol. 39*, 973–976.

Humbel, R. E. (1984) in *Hormonal Proteins & Peptides* (Li, C. H., Ed.) Vol. XII, pp 57–79, Academic Press, New York.

Iordanescu, S. (1975) *J. Bacteriol. 12*, 597–601.

Jörnvall, H., & Philipson, L. (1980) *Eur. J. Biochem. 104*, 237–247.

Langone, I. I. (1982) *Adv. Immunol. 32*, 157–252.

Moks, T., Abrahmsén, L., Nilsson, B., Hellman, U., Sjöquist, J., & Uhlén, M. (1986) *Eur. J. Biochem. 156*, 637–643.

Nilsson, B., Holmgren, E., Josephson, S., Gatenbeck, S., Philipson, L., & Uhlén, M. (1985a) *Nucleic Acids Res. 13*, 1151–1162.

Nilsson, B., Abrahmsén, L., & Uhlén, M. (1985b) *EMBO J. 4*, 1075–1080.

Nilsson, B., Moks, T., Jansson, B., Abrahmsén, L., Elmblad, A., Holmgren, E., Henrichson, C., Jones, T. A., & Uhlén, M. (1987) *Protein Eng. 1*, 107–113.

Norrander, J., Kempe, T., & Messing, J. (1983) *Gene 20*, 101–106.

Oliver, D. (1985) *Annu. Rev. Microbiol. 39*, 615–648.

Palva, I., Lehtovaara, P., Kääriäinen, L., Sibakov, M., Cantello, K., Schein, C., Kashiwagi, K., & Weissmann, C. (1983) *Gene 22*, 229–235.

Pugsley, P. P., & Schwartz, M. (1985) *FEMS Microbiol. Rev. 32*, 3-38.
Righetti, P. G., & Chillenni, F. (1978) *J. Chromatogr. 157*, 243-251.
Smith, J. C., Derbyshire, R. B., Cook, E., Dunthorne, L., Viney, J., Biener, S. J., Sassenfeld, H. M., & Bell, L. D. (1984) *Gene 32*, 321-327.
Uhlén, M., Guss, B., Nilsson, B., Gatenbeck, S., Philipson, L., & Lindberg, M. (1984) *J. Biol. Chem. 259*, 1695-1702.
Ullman, A. (1984) *Gene 29*, 27-31.
Yanisch-Perron, C., Vieira, J., & Messing, J. (1985) *Gene 33*, 103-119.

Nucleotide Sequence of the Mouse α_1-Acid Glycoprotein Gene 1†

Robin Cooper,‡ D. Mark Eckley, and John Papaconstantinou*
Department of Human Biological Chemistry and Genetics, Division of Cell Biology, University of Texas Medical Branch, and Division of Cell Biology, Shriners' Burns Institute, Galveston, Texas Unit, Galveston, Texas 77550
Received January 28, 1987; Revised Manuscript Received April 27, 1987

ABSTRACT: In a previous paper we presented evidence for the existence of at least two α_1-acid glycoprotein (AGP) genes in the mouse. One of the cDNA clones characterized in those studies was used to isolate several unique AGP genomic clones. In these studies we present the complete sequence of one of the mouse AGP genes. The sequence analysis includes 595 base pairs (bp) 5′ to the site of initiation of transcription and 135 bp 3′ to the polyadenylation signal. This mouse AGP gene, designated AGP-1, has six exons, a structure similar to those of the AGP genes in rats and humans. Analysis of the sequence has revealed a number of potential regulatory sites. These include a run of alternating purine–pyrimidine bases [$(GT)_n$] at +2890 to +2945, flanked by three potential glucocorticoid receptor binding sites within intron 5. Two of these TGTTCT at +3069 to +3074 and +3082 to +3087 flank the $(GT)_n$ track at its 3′ end, and one, which is oriented in the opposite direction (AGAACA), at +2771 to +2776 flanks the track at its 5′ end. A longer version of the glucocorticoid receptor site, GGGTACAATGTGTCCT, has been located in the 5′ flanking region of the gene (−94 to −79); the sequence AGAACA is another potential glucocorticoid receptor site oriented in the opposite direction and located at −127 to −122. This entire region, from −146 to −42, in the mouse has a strong homology (~85%) to the 5′ flanking region of the rat AGP gene, which contains a 78-bp fragment (−120 to −42) that represents the minimal sequence required for glucocorticoid regulation. A sequence of 38 nucleotides (−22 to +16) that is homologous to similarly located sequences previously observed in three human acute-phase proteins has also been identified. We suggest that this sequence may represent an acute-phase protein regulatory element.

α_1-Acid glycoprotein (AGP), also known as orosomucoid, is a M_r 44 000 component of mammalian serum. In humans, the circulating amounts of AGP, and other acute-phase proteins, increase dramatically following a physiological insult such as acute inflammation, bacterial infection, major surgery, or burns (Koj, 1974). The biological significance of the acute-phase response is not well understood. Nevertheless, the reaction must have an important physiological role as the acute-phase proteins are highly conserved in evolution, and normal levels of these proteins are maintained even during severe malnutrition (Schmid, 1975; Ricca & Taylor, 1981; Ricca et al., 1982; Friedman, 1983).

The synthesis and secretion of the acute-phase proteins, including AGP, can be induced by injection of inflammatory agents such as turpentine. Both dexamethasone and a hepatocyte stimulating factor have also been shown to induce at least some of the acute-phase proteins (Ritchie et al., 1982; Baumann et al., 1983a,b, 1984a,b; Ritchie & Fuller, 1983). Studies with rats have indicated that, at the maximum point of induction by turpentine, the mRNA coding for AGP becomes one of the most abundant mRNA species in the liver (Northemann et al., 1983). Reports have been published in which both transcriptional and posttranscriptional mechanisms have been proposed for the regulation of the AGP gene expression in the rat (Vannice et al., 1984; Kulkarni et al., 1985). Changes in AGP mRNA pool levels, induced by dexamethasone in a rat hepatoma cell line, were reported to be due to stabilization of the primary AGP transcript and the efficient processing of this transcript to form a mature, functional cytoplasmic mRNA. This mechanism may be mediated by a protein factor whose synthesis is regulated by the glucocorticoid (Vannice et al., 1984). A more recent report on the effects of dexamethasone and turpentine in rats indicates that AGP regulation, at least in vivo, occurs at the transcriptional level (Kulkarni et al., 1985).

Previously, we reported the existence of at least two AGP genes in the mouse (Copper & Papaconstantinou, 1986). These conclusions were based on the sequence analysis of cDNA clones from a liver cDNA library. Our more recent work on sequence analysis of genomic clones shows the existence of an additional AGP gene in the mouse (unpublished observations). These data could explain the results of Baumann et al. (1984a,b), who observed multiple forms of AGP on two-dimensional protein gels. We wish, ultimately, to establish whether various mouse AGP genes are coordinately or differentially regulated in response to inducers such as inflammatory agents or hormones. As a first step toward this

† This research was supported by a grant from the Shriners' Burns Institute, Galveston, Texas Unit, to J.P.
* Correspondence should be addressed to this author at the Department of Human Biological Chemistry and Genetics, University of Texas Medical Branch.
‡ Present address: Laboratory of Molecular Parasitology, The Rockefeller University, New York, NY 10021.

0006-2960/87/0426-5244$01.50/0