Vol. 155, No. 1, 1988 BIOCHEMICAL AND BIOPHYSICAL RESEARCH COMMUNICATIONS
August 30, 1988 Pages 518-523

MISTRANSLATION IN IGF-1 DURING OVER-EXPRESSION OF THE PROTEIN IN ESCHERICHIA COLI USING A SYNTHETIC GENE CONTAINING LOW FREQUENCY CODONS

Ramnath Seetharam[1], Robert A. Heeren, Edith Y. Wong, Sarah R. Braford, Barbara K. Klein, Serdar Aykent, Claire E. Kotts, Karl J. Mathis, Bruce F. Bishop, Michael J. Jennings, Christine E. Smith and Ned R. Siegel

Monsanto Company Biol. Sciences, 700 Chesterfield Village Pkwy., St. Louis 63198

Received July 11, 1988

SUMMARY: Partial misincorporation of Lys for Arg has been observed for the Arg residues of IGF-1 when the molecule is expressed in *Escherichia coli* using a synthetic gene with the low frequency AGA codon encoding all six Arg residues and yeast preferred codons encoding the remaining residues. The Lys for Arg substitution at these residues could not be detected when a gene containing *E. coli* preferred codons, with the codon CGT coding for all Arg residues, was used for the expression of the protein. Similarly, no misincorporation of Lys for Arg could be detected when a gene containing *Escherchia coli* preferred codons at all positions, except for an AGA codon at Arg (36), was utilized.

Insulin-like growth factor-1 (IGF-1), a polypeptide hormone of 70 amino acids and three disulfide bonds (1, 2, 3) has potent growth promoting activities (4) and is believed to mediate many of the effects of growth hormone (5, 6). We have produced IGF-1 in *E. coli* by expressing a synthetic gene encoding this protein using a periplasmic expression system (7, 8). The synthetic gene used for the expression of IGF-1 was constructed using codons found either in highly expressed genes of yeast (9) or those found in highly expressed genes of *E. coli* (10, 11). There are a number of differences in codon usage between the two organisms. Consequently, the two synthetic genes have very different nucleotide sequences, although they are designed to code for the same amino acid sequence. One such difference, that is relevant to the present discussion, is the codon used for Arg in the two cases. The preferred codon for Arg is AGA in the case of yeast and CGT in the case of *E. coli*.. Furthermore, the AGA codon for Arg is almost never used in *E. coli.* (10, 11). The data presented in this report indicates that expression of IGF-1 in *E. coli,* using a synthetic

[1]Present address: Dupont Company, Pharmaceutical R & D, Glasgow Site Bldg. 300, Wilmington, DE 19898

0006-291X/88 $1.50

gene with the low frequency codon AGA coding for the six Arg residues of the protein and yeast preferred codons encoding the remaining residues, results in partial misincorporation of Lys at the Arg residues of the protein. No mistranslation can be detected when the low frequency codons for Arg (AGA) are replaced by the more frequently used Arg codons of *E.coli* (CGT).

MATERIALS AND METHODS

An analog of human IGF-1 (Met-59 -> Thr) was obtained from Amgen (Thousand Oaks, CA) and used as a standard. The amino acid substitution at position 59 has been shown to have no effect on the biological activity of the molecule (10).

Expression of IGF-1:
Synthesis of the different IGF-1 genes as well as the expression of the protein in *E. coli* was carried out as described earlier (7, 8). Three versions of the gene were synthesized. The first one had the low frequency codon AGA encoding all the six Arg residues of the protein and yeast preferred codons at all the other positions (Gene I). The second version of the gene had the more frequently used codon CGT at all the six Arg residues of the protein and *E. coli* preferred codons at the other positions (Gene II). The third version was the same as the second except for the presence of a low frequency AGA codon coding for the Arg at position 36 (Gene III)

Purification of IGF-1:
The IGF-1 was purified from *E. coli* extracellular media using a modification of the method described previously (11, 12). The extracellular media (about 2 liters) were concentrated about 6 fold in a stirred cell using a YM-2 membrane (Amicon, Danvers, MA) and subjected to ammonium sulfate precipitation. The precipitate obtained between 25-65% ammonium sulfate was collected, dissolved in 50 mM sodium formate buffer, pH 4.0 and desalted on a G-25 column equilibriated with the same buffer. The desalted sample was loaded onto a HR 10/10 Pharmacia Mono S cation exchange column (Pharmacia, Piscataway, NJ), equilibriated with 50 mM sodium formate, pH 4.0; at a flow rate of 3.0 ml/min. The column was eluted at the same flow rate, using a linear NaCl gradient in 50 mM sodium formate. The absorption of the effluent was monitored at 280 nm. The fractions containing IGF-1 (approx 600 mM NaCl) were pooled and subjected to reversed-phase HPLC chromatography using a 7.0 X 250 mm Aquapore RP 300 Brownlee C8 column (Brownlee labs, CA). The column was equilibrated with 10% (v/v) acetonitrile in water, containing 0.1% (v/v) trifluoroacetic acid (TFA); and eluted with a gradient of10-50 % acetonitrile in water containing 0.1% (v/v) TFA. The IGF-1 was found to elute at approximately 33% acetonitrile under these conditions. Microbore reversed-phase HPLC, using an Applied Biosystems Model 130A separation system, was employed to monitor the IGF-1 during purification as well as to asses the purity of the final product A 2.1 X 30 mm Aquapore RP300 Brownlee C8 column was used. The column was equilibriated in 0.1% TFA in water and the samples were applied to the column in the same solvent. The column was eluted using an acetonitrile/water/TFA gradient as described in Fig-1. The Amgen IGF-1 standard was used to establish the retention time of IGF-1 under the conditions of assay. IGF-1 was quantitated by comparing the area under the peak to the area obtained with a known amount of standard IGF-1.

Biological Acitivity:
Biological activity was determined using a muscle cell proliferation assay (12, 13).

Amino Acid Sequencing :
Amino acid sequence was obtained using automated Edman degradation chemistry. An Applied Biosystem, Inc. Model 470A gas phase sequencer (Foster City, CA) was employed for the degradations (14). The respective PTH amino acid derivatives were identified by reversed-phase HPLC analysis, in an on-line fashion, utilizing an Applied Biosystem Inc., Model 120A PTH analyzer fitted with a Brownlee 2.1 X 150 mm PTH-C18 column. The amino acid sequence through 68 places was determined by

sequential Edman degradation using 13 nmoles of protein. The amino acid at positions 69 and 70, which were difficult to detect in the extended Edman degradation run were determined by carboxypeptidase digestion. 3 nmoles of IGF-1 was taken up in about 160 uls of 100 mM NH4HCO3 and 5.87 units of Carboxypeptidase A (Cooper Biomedical) was added. The reaction was terminated at various time points by the addition of trifluoroacetic acid (TFA). Samples were lyophilized, reconstituted in sodium citrate buffer, and injected onto a Beckman 6300 High Performance Analyzer (Palo Alto, CA). The equivalent of about 750 pmoles was used for the amino acid analyses. A plot of the picomolar amino acids released vs time was used to determine the sequence of cleavage.

RESULTS AND DISCUSSION

The chromatographic profiles obtained during the cation exchange step as well as the reversed-phase HPLC step were virtually identical, irrespective of the synthetic gene used for the expression of IGF-1. About 6 mgs of purified IGF-1 was obtained from approximately 2 liters of 6 fold concentrated media, in all three cases, providing an overall purification yield of about 30%. The purified IGF-1 obtained from all three versions of the synthetic gene gave a single symmetric peak with the same retention time as the IGF-1 standard (Fig 1). The purity of the preparations was estimated to be about 98% by this criterion. In addition, the IGF-1 obtained using different versions of the gene had the same biological activity as standard IGF-1 (data not shown).

The entire amino acid sequence of the IGF-1, obtained using the version of the synthetic gene containing yeast preferred codons (Gene I), was determined and was found to agree with that expected for IGF-1 (Fig 2). However, a significant amount of

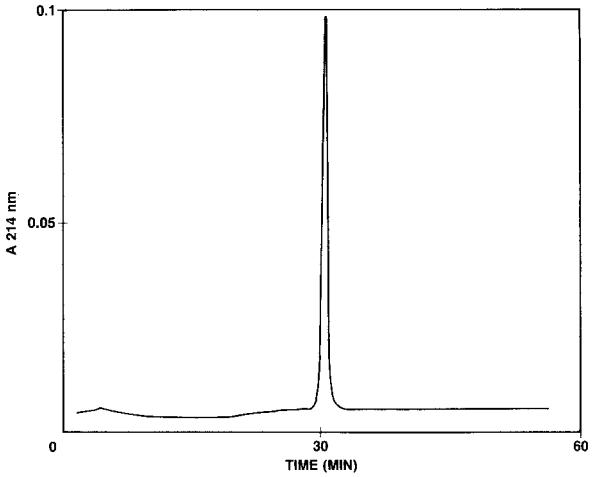

Fig. 1. Microbore reversed-phase HPLC of purified IGF-1. Microbore HPLC was performed on a 2.1 mm X 150 mm C 8 reversed-phase column (Brownlee Aquapore RP300) using an Applied Biosystems Model 130 A Separations System. The purified IGF-1 was applied to the column in 0.1% TFA/water. The following gradient was used: 0% B to 30% B in 7.1 min; 30% B to 70% B in 56 min; 70% B to 100% B in 3 min; 100% B to 0 % B in 5 min. The flow rate was 200 ul/min. Buffer A was 0.1% TFA/water; Buffer was 70% acetonitrile/water containing 0.085% TFA.

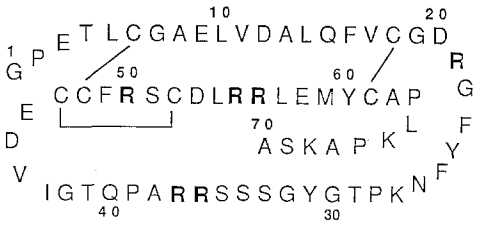

Fig. 2. The amino acid sequence of IGF-1. The amino acid sequence of IGF-1 using the single letter code, along with the disulfide linkages as deduced from its homology to insulin. The positions of the Arg residues in the protein have been highlighted.

lysine could also be detected at the arginine positions 21, 36 and 37. The percentage of lysine detected at each of these positions is shown in Table I. It is quite evident from data subsequently obtained that the Lys for Arg substitution cannot be detected when the gene with *E. coli* preferred codons Gene II) is used for the expression of IGF-1 (Table I). Interestingly, the same results were obtained when the IGF-1 was expressed using the synthetic gene containing *E. coli* preferred codons at all positions, except for an AGA codon at Arg (36) (GeneIII). This suggests that a single

TABLE I

Lys for Arg Misincorporation in IGF-1 encoded by different gene sequences

	% LYSINE		
	Gene I[a]	Gene II[b]	Gene III[c]
ARG (21)	12	0	0
ARG (36)	6	0	0
ARG (37)	5	0	0
ARG(50,55,56)	d	d	d

a - Lys for Arg misincorporation observed when the IGF-I gene containing low frequency AGA codons at all six Arg positions and yeast preferred codons at all other positions is used.

b - Lys for Arg misincorporation observed when the IGF-I gene containing the more frequently used Arg codon CGT is used at all the six Arg positions and E. coli preferred codons at all other positions.

c - Lys for Arg misincorporation observed when the IGF-1 gene containing E.coli preferrred codons at all positions except for the low frequency AGA codon at Arg 36.

d - The low primary sequencing yields obtained at these positions made it difficult to obtain accurate misincorporation data.

low frequency codon is not capable of causing Lys for Arg substitution. It was difficult to quantitate the amount of Lys for Arg substitution at residues 50, 55 and 56 since the main Arg signal itself was quite weak at these residues. This observation is most likely the effect of technologica limitations incurred by the inefficiencies in the Edman degradation chemistry, however, and it is reasonable to think that the mistranslation is probably in the same range as that observed at Args 36 and 37.

The observed Lys for Arg misincorporation implies that the protein, obtained by expressing the gene utilizing yeast preferred codons, is actually a heterogenous mixture of IGF-1 species containing different combinations of Lys for Arg substituted molecules. This indicates that the different IGF-1 species are not separated under the conditions used for purification. This is not surprising considering the fact that the IGF-1 species generated by Lys for Arg substitution would have very similar properties because of the close similarity between these two amino acids. This fact suggests that it would be quite difficult to devise a purification protocol to obtain the authentic IGF-1, free from the mistranslated species, in good yields. The use of the synthetic gene with *E. coli* preferred codons appears to be the easier method to generate IGF-1 with the authentic sequence. In general, this would probably be the preferred approach if the mistranslated products have properties very similar to the authentic protein. Purifying the mistranslated species away from the authentic product would be a viable alternative if the two proteins have quite different properties. However, if that is the case it would be difficult to detect any mistranslation at all by analyzing the purified translation product alone. It is quite conceivable that some of the mistranslated IGF-1 species, generated due to possible mistranslation at some of the other low frequency codons present in the gene utilizing yeast preferred codons, are separated by the protocol used to purify the IGF-1 and are therefore not detected. However, the fact that the chromatographic profiles and purification yields obtained are almost identical with all three versions of the IGF-1 gene suggests such mistranslation products are not present at the same level as the Lys for Arg substituted IGF-1.

The Lys for Arg substitution probably arises due to the misincorporation of the tRNA ^{Lys} (with the anticodon sequence UUU) at the AGA codon for Arg. If so, it would suggest the involvement of a second position wobble pairing of the U in the anticodon with the G in the codon (15). It is interesting to note that the Lys for Arg substitution does not occur at the AGA codon coding for Arg (36), when it is the only low frequency codon present in the synthetic gene, indicating that a single low frequency codon at that position alone is not sufficient enough to cause Lys for Arg misincorporation. This raises the interesting question of the minimum number of low frequency codons required to cause significant mistranslation. It is conceivablethat this number would be affected by the nature of the low frequency codons itself as well as its position in the gene sequence, by the nature of the codons

encoding other amino acids in the protein and by the strain of *E. coli* used for expressing the protein. It appears that the IGF-1 system described here would provide a good model system to study the effects of some these parameters, at least in case of the Lys for Arg substitution. Such information would lead to a better understanding of the mistranslation effects in *E. coli*. It would also be useful in analyzing the mistranslation that can occur during high-level expression of natural genes in a heterologous system, especially when the genes contain codons infrequently used by organism employed for high-level heterologous expression.

ACKNOWLEDGEMENTS

We wish to thank Drs. Gwen G. Krivi, Dr. Gerald R. Galluppi and Dr. Gary Bild for their interest and support.

REFERENCES

1. Rinderknecht, E., and Humbel, R. E. (1976) Proc. Natl. Acad. Sci. USA **73,** 2365.
2. Rinderknecht, E., and Humbel, R. E. (1977) J. Biol. Chem. **253,** 2769.
3. Rinderknecht, E., and Humbel, R. E. (1978) FEBS Lett. **89,** 283.
4. Schoenle, E., Zapf, J. and Froesch E. R.(1982) Nature (London) **296,** 252
5. Zapf, J., Froesch, E. R. and Humbel, R. E. (1981) Curr. Topics Cell Reg. **19,** 257-309
6. Clemmons, D. R. and Van Wyck J. J. (1981) Hand book of Experimental Pharmacology (Springer-Verlag, Berlin) **vol 57,** 161-208.
7. Wong, E. Y., Braford, S. R., Klein, B. K., Heeren, R. A., Seetharam, R., Siegel, N. R., and Tacon, W. C. (1988a) Advances in Gene Technology: Protein Engineering and Production (Proc. of the 1988 Miami Winter Symposium) (Brew, K., Ahmad, F., Baily, H., Black, S., Fenna, R. E., Puett, D., Scott, W. A., Brunt, J. V., Voellmy,R.W., Whelan, W. J., and Woessner, J. F. eds) ICSU Short Reports 8, 104.
8. Wong, E. Y., Seetharam, R., Kotts, C. E., Heeren, R. A., Klein, B. K., Braford, S. R, Mathis, K. J., Bishop, B. F., Siegel, N. R., Smith, C. E. and Tacon, W. J. (1988b) Gene (In Press)
9. Sharp, P. M., Tuohy, P. M. S. and Mosurski, K. R. (1986) Nucl. Acids Res. **14,** 5125-5143.
10. Grantham, R., Gautier C., Gouy, M., Jacobzone, M., and Mercier, R. (1981) Nucl. Acids Res. 9, r43-r74.
11. Gouy, M. and Gautier, C. (1982) Nucleic Acid Res. **10,** 7056-7074.
12. Peters, M. A., Lau, E. P., Snitman, D. L., Van Wyk, J. J., Underwood, L. E., Russell, W. E., and Svodoba, M. E. (1985) Gene **35,** 83.
13. Kotts, C. E., White, M. E., Allen, C. E., Martin, F., and Dayton, W. R. (1987) J. Animal Sci. **64,** 615.
14. Hunkapiller, M. W., Hewick, R. M., Dreyer, W. J., and Hood, L. E. (1982) Methods Enzymol. **91,** 39-413.
15. Mcpherson, D.T. (1988) Nucleic Acid Res. **16,** 4111-4120.