Mutations in the guinea pig preproglucagon gene are restricted to a specific portion of the prohormone sequence

S. Seino, M. Welsh, G.I. Bell*, S.J. Chan+ and D.F. Steiner+°

*Department of Biochemistry & Molecular Biology, The University of Chicago, Chicago, IL, *Chiron Corporation, Emeryville, CA 94608* and *+The Howard Hughes Medical Institute, The University of Chicago, 920 E. 58th Street, Chicago, IL 60637, USA*

Reveived 27 May 1986

A cDNA clone encoding guinea pig preproglucagon has been isolated from a pancreatic cDNA library. The predicted amino acid sequence of proglucagon is highly conserved in all regions, in comparison to other mammals, except for the C-terminal portion of the 29-residue glucagon region, in which 5 amino acid substitutions have occurred. These changes may serve to offset the reduced receptor-binding potency of the highly mutated insulin in this New World species.

Glucagon *Evolution* *Hystricomorph* *cDNA* *Glycogenolytic hormone*

1. INTRODUCTION

Glucagon, the 29-residue pancreatic hormone that stimulates hepatic glycogenolysis and gluconeogenesis, is highly conserved in mammals [1,2]. It is derived from a 180-amino-acid precursor or preproglucagon, which bears an NH_2-terminal signal sequence that is removed after membrane translocation [3,4]. Proglucagon in mammals is an 18 kDa (160 amino acid) protein consisting of an NH_2-terminal propeptide (also called glicentin-related pancreatic peptide, GRPP), glucagon and two COOH-terminal glucagon-like peptides (GLP-1 and 2) separated by a short linker segment. Glucagon is synthesized in the islets of Langerhans, the stomach and intestine and possibly also in the brain [5–8]. Different modes of processing of proglucagon have been observed [3]. Pancreatic islets release glucagon, GRPP and an intact COOH-terminal fragment containing both GLP-1 and GLP-2 [9], whereas the intestine releases a 69-amino-acid glucagon-containing polypeptide called glicentin which contains GRPP, glucagon and spacer peptide 1 [10]. Mammals have a single preproglucagon gene. By contrast, the anglerfish has two non-allelic preproglucagon genes, the products of which are homologous to each other, but they lack the GLP-2 sequence present in the mammalian precursor [11]. In addition, the NH_2-terminal propeptide of the anglerfish protein shows very low homology to its mammalian counterpart [11], however both glucagon and GLP-1 show a certain degree of homology to mammalian sequences.

Hystricomorph rodents, such as the guinea pig, have evolved a number of divergent proteins under conditions of relative isolation in South America following the tertiary migration of the continent [12,13]. One such protein is insulin, which has accumulated mutations in the A and B chains at the same frequency as in the signal sequence and in the C-peptide [14] suggesting that neutral mutations fixed by random drift may have allowed guinea pig insulin to acquire a new function which was fixed by positive selection. Guinea pig insulin has less metabolic activity than other mammalian insulins [15,16], but on the other hand has more growth-stimulating activity [17]. Such an evolutionary

° To whom correspondence should be addressed

Published by Elsevier Science Publishers B.V. (Biomedical Division)

change in insulin's metabolic potency would seem likely to require compensatory changes in insulin counter-regulatory hormones, such as glucagon, as well as other hormones whose actions must be integrated with that of insulin in metabolic regulation [14].

The first indication of the existence of an altered glucagon molecule in guinea pigs was a brief report by Sundby [18] indicating that its amino acid composition was at variance with that of other mammalian glucagons. On this basis, he postulated that guinea pig glucagon might be a larger peptide of approx. 40 residues. During the past year guinea pig glucagon has been isolated and sequenced by Conlon et al. [19] who observed that the COOH-terminal nonapeptide portion of glucagon has 5 changes, compared to other mammals, while the sequence of the NH_2-terminal portion was unchanged. No difference in the overall length of the molecule was found. They concluded that sequence differences in the COOH-terminal part of guinea pig glucagon were an adaptive response to the lower metabolic activity of its divergent insulin. The COOH-terminal portion of glucagon is thought to be involved mainly in receptor binding while the NH_2-terminal part is more directly involved in stimulation of adenylate cyclase [20]. More recent studies of Huang et al. [21] confirm these sequence changes in guinea pig glucagon and demonstrate that this molecule indeed has markedly reduced binding potency in both rat and guinea pig liver membranes.

We have isolated a near full-length cDNA encoding guinea pig preproglucagon. Its nucleotide sequence and predicted amino acid sequence are as similar to those of human as are the other mammalian preproglucagons, in all regions except the COOH-terminal portion of glucagon. These observations agree with the notion that selective pressure to maintain the structure and function of glucagon has been selectively suspended on the COOH-terminal portion of the molecule in order to restore a balanced metabolic hormonal status.

2. MATERIALS AND METHODS

2.1. *Isolation and sequence of cDNA encoding guinea pig preproglucagon*

Poly(A)$^+$ RNA was isolated from adult guinea pig (*Cavia porcellus*) pancreas as well as pancreatic islets using standard procedures [22]. Double-stranded cDNA was prepared [23] and after methylation of internal *Eco*RI sites and the addition of *Eco*RI linkers ligated into the *Eco*RI site of λgt10 [24]. Phage were packaged and recombinants selected by plating on *E. coli* strain BNN102 [24]. Recombinants containing inserts encoding guinea pig preproglucagon were identified in the islet cDNA library [11] by cross-hybridization with a nick-translated insert from a Syrian hamster preproglucagon cDNA [3] under conditions of low stringency [washing in 0.2 × SSC (SSC = 0.15 M NaCl, 0.015 M sodium citrate), 0.1% SDS, 42°C]. A partial fragment of the guinea pig preproglucagon cDNA was identified and then used to screen the library prepared from pancreatic RNA; cDNA inserts were cloned into M13 mp19 and sequenced on both strands [25].

2.2. *RNA blot analysis of guinea pig preproglucagon mRNA*

10 μg poly(A)$^+$ RNA from guinea pig pancreas were denatured with glyoxal [26] and, after electrophoresis through a 1.2% agarose gel, transferred to a nitrocellulose filter [27]. ^{32}P-labeled and glyoxal-denatured fragments of a *Hin*dIII digest of λDNA and a *Hae*III digest of φX174 DNA were included as size standards. Nitrocellulose filter strips were hybridized with nick-translated guinea pig glucagon cDNA insert.

3. RESULTS AND DISCUSSION

cDNAs encoding guinea pig preproglucagon were isolated from libraries prepared with pancreatic islet-enriched as well as pancreatic poly(A)$^+$ RNA. A partial clone was isolated from the former and used to screen the pancreas cDNA library. The frequency of glucagon cDNA clones in this library was about one in 15000 phage. The nucleotide sequence of one of these, gpGCG-2, contained an open reading frame of 540 base pairs (bp) which predicted the sequence of the 180-amino-acid guinea pig preproglucagon (fig.1). The 5′- and 3′-untranslated regions of this clone were 46 and 467 bp, respectively. Guinea pig pancreatic preproglucagon mRNA is about 1350 bases (fig.2) suggesting that the 5′-untranslated region of the mRNA may be 50–100 bases longer than indicated in fig.1. Guinea pig preproglucagon has the typical

```
                                          -20 Signal Peptide                  -10
                                          Met Lys Ser Val Tyr Phe Val Ala Gly Leu Phe Ile Met Leu Ala Gln Gly Ser
GGTGCACCCGTTGCTAGCCACAGCCTACGAGCAGAAGGTAGCAAAA  ATG AAG AGC GTT TAC TTT GTG GCT GGA TTG TTC ATA ATG CTA GCA CAA GGC AGC

     -1 Amino Terminal Peptide (GRPP)       10                                  20
Trp Gln Arg Ser Leu Gln Asp Thr Glu Glu Lys Pro Arg Ser Val Ser Ala Ser Gln Thr Asp Met Leu Asp Asp Pro Asp Gln Met Asn
TGG CAA CGT TCC CTT CAA GAC ACA GAA GAG AAA CCC AGA TCT GTC TCA GCC TCC CAA ACA GAC ATG CTT GAT GAT CCA GAT CAG ATG AAC

     30         Glucagon                  40                                  50
Glu Asp Lys Arg His Ser Gln Gly Thr Phe Thr Ser Asp Tyr Ser Lys Tyr Leu Asp Ser Arg Arg Ala Gln Gln Phe Leu Lys Trp Leu
GAA GAC AAG CGC CAT TCA CAG GGC ACA TTC ACC AGC GAC TAC AGC AAG TAC TTG GAT TCC AGG CGT GCT CAA CAA TTT TTG AAA TGG CTG

     60            Spacer Peptide 1        70     Glucagon-Like Peptide 1      80
Leu Asn Val Lys Arg Asn Arg Asn Asn Ile Ala Lys Arg His Asp Glu Phe Glu Arg His Ala Glu Gly Thr Phe Thr Ser Asp Val Ser
TTG AAT GTC AAG AGG AAC AGG AAC AAC ATT GCC AAA CGT CAT GAT GAA TTT GAG AGA CAT GCT GAA GGG ACC TTT ACT AGT GAT GTA AGT

     90                                  100                                 110 Spacer Peptide 2
Ser Tyr Leu Glu Gly Gln Ala Ala Lys Glu Phe Ile Ala Trp Leu Val Lys Gly Arg Gly Arg Arg Asp Phe Pro Glu Glu Val Ala Ile
TCT TAC TTG GAA GGC CAA GCT GCC AAG GAA TTC ATT GCT TGG CTG GTG AAA GGC CGA GGG AGG CGA GAC TTC CCA GAA GAA GTC GCC ATT

    120                   Glucagon-Like Peptide 2                           140
Val Glu Glu Leu Gly Arg Arg His Ala Asp Gly Ser Phe Ser Asp Glu Met Asn Thr Ile Leu Asp Asn Leu Ala Thr Arg Asp Phe Ile
GTG GAA GAA CTC GGC CGC AGA CAT GCC GAT GGC TCA TTC TCA GAT GAG ATG AAC ACT ATT CTT GAC AAT CTT GCC ACC AGA GAC TTT ATC

    150                                 160
Asn Trp Leu Ile Gln Thr Lys Ile Thr Asp Arg Lys OC
AAC TGG CTT ATT CAG ACC AAA ATC ACT GAC AGG AAG TAA GTATGTCACTCTTCAAGACCATCTTCACATCACCTGCCGTCCACTTGGAATGTTTGAAATTTTACAGTT

CTGTAATTTTACAGAGTTGTACTCTCGAGTATTTCTTTGCAGGCTATTAAACATTTTTTAGCATTGTATAGCCAAATGATTATAAATGGAATAAACTATCGCCAGAATGTTGCTAAAATA

TCAACTTTACAGTATAAAAGTCCTGTCTCTTGTTTTTATCTTATTTTGGTTGAAGTACCCCAACTTGTTTAAATTTAGCAGTGAAATTATTTTTCTATTATATACTTTGTAGATGTAAAT

TAATCCAATCTGAAAATATCTGCATGCAATATCAGGAAAATGAAAACCTTGTAGCCACAGCAGTGAAACTGAAAAGAGAACTTCTTAAAGCCTTTTTCATAAAAATGCTCAGCTTTCAAT

GTCTCAAAGATTGAATTAATAAATTTTCAAGCTTCAAAAAAAAAAAAAAAAAAAA (poly A)
```

Fig.1. Nucleotide sequence of guinea pig preproglucagon cDNA clone (gpGCG2) and the predicted amino acid sequence of protein.

features of the mammalian precursor (fig.3), including signal peptide, NH_2-terminal propeptide, glucagon, GLP-1 and -2, and the two shorter spacer peptides [10,28].

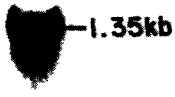

Fig.2. Northern blot of guinea pig pancreatic preproglucagon mRNA. The size of mRNA is indicated.

Comparison of the guinea pig and other mammalian preproglucagon sequences, at both the protein and nucleotide level, with the human sequence reveals that the various segments of the precursor have evolved at relative rates that roughly approximate evolutionary distances (table 1), except for the COOH-terminal nine amino acids of glucagon (proglucagon 53–61, fig.3). Five amino acid substitutions have occurred in this region; two are relatively conservative and three represent significant changes in the properties of the amino acid side chains. These amino acid replacements predicted from the cDNA sequence agree with the sequences reported for guinea pig glucagon by Conlon et al. [19] and Huang et al. [21]. In contrast, only 2 amino acid changes have been found in the duck [29] and alligator [30] glucagons. The other regions of the guinea pig proglucagon molecule show a very high degree of homology to the corresponding regions of other mammalian proglucagons. The striking sequence conservation in the region of GLP-1 and -2 implies an important but as yet undetermined physiological role for

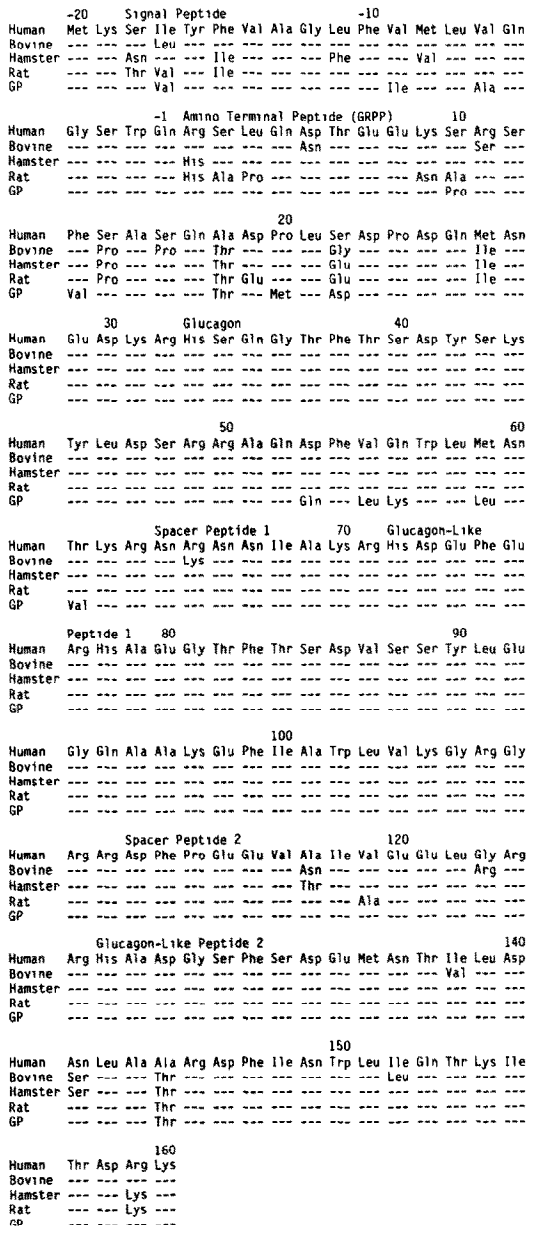

Fig.3. Comparison of the amino acid sequences of human, bovine, Syrian hamster, rat and guinea pig preproglucans. The rat and bovine sequences are from [32] and [33], respectively. The corrected hamster [3] and human [2] sequences are from [10].

these peptides. In contrast, the relatively rapid rate of mutation acceptance in the NH_2-terminal propeptide suggests that the structural requirements for its function (possibly in some aspect of gastrointestinal physiology, since it is part of glicentin, the major glucagon-containing molecule produced in the gut) are less stringent. Alternatively, the relative conservation in this region could reflect special structural requirements for folding or intracellular transport, as in the case of the C-peptide of proinsulin [31]. The homology of the two anglerfish preproglucagons to the human precursor are included for comparison (table 1); the numerous substitutions occur relatively uniformly through both the glucagon and GLP-1 sequences. The sequence of the 3′-untranslated region of guinea pig preproglucagon mRNA is homologous (30–60%) with this region of the other mammalian mRNAs although, except for the region following the polyadenylation signal, there is no large conserved region common to all mammalian preproglucagon mRNAs.

Our data demonstrate that mutations have accumulated more rapidly in the region of the gene encoding the COOH-terminal region of guinea pig glucagon than in the rest of the preproglucagon molecule, which has been evolutionarily stable. These changes may have been required to maintain glucose homeostasis in response to the mutations in the insulin gene resulting in a metabolically less active insulin in this species [15]. On the other hand, the lack of any changes within either GLP-1 or -2 suggests that these peptides do not play a prominent role in glucose homeostasis, in keeping with evidence that these are not processed and released as such from islets [9]. These alterations in glucagon have probably occurred in order to compensate for the lower biological activity of the guinea pig insulin. Their clustering in the COOH-terminal portion of the molecule provides strong support for the hypothesis that this region functions to enhance receptor-binding affinity [18,21]. Although it is possible that mutagenesis in the glucagon molecule may have preceded changes in the guinea pig insulin molecule this sequence of events seems less likely in view of the subordinate role of glucagon relative to that of insulin in regulating glucose homeostasis. Likewise it seems unlikely that the recently described substitutions in vasoactive intestinal polypeptide (VIP) in the guinea pig [34] are related in any direct functional sense to those in the insulin/glucagon axis described here.

Table 1

Comparison of nucleotide and amino acid sequence homology in domains of preproglucagon

	Signal peptide		NH_2-propeptide		Glucagon 1–20		Glucagon 21–29		GLP-1		GLP-2	
	Nucleotide (% of 60)	Amino acid (% of 20)	Nucleotide (% of 90)	Amino acid (% of 30)	Nucleotide (% of 60)	Amino acid (% of 20)	Nucleotide (% of 27)	Amino acid (% of 9)	Nucleotide (% of 111)	Amino acid (% of 37)	Nucleotide (% of 99)	Amino acid (% of 33)
Human/ guinea pig	88	85	83	83	92	100	67	44	96	100	89	97
Human/ hamster	75	80	86	83	88	100	93	100	90	100	88	94
Human/ rat	88	85	82	67	87	100	96	100	93	100	88	99
Human/ bovine	97	95	83	77	93	100	93	100	99	100	87	88
Human/ angler-fish I	48[a]	29[a]	21[a]	17[a]	68	70	78	67	62	61	–	–
Human/ angler-fish II	52[a]	37[a]	14[a]	13[a]	65	75	78	78	67	71	–	–

[a] Gaps due to variations in length in these regions were scored as sequence differences

All dibasic residues are excluded from the sequence comparisons, and the anglerfish I and II GLP-1 were compared to the COOH-terminal part of human GLP-1. The rat and bovine sequences were taken from [32] and [33], respectively

ACKNOWLEDGEMENTS

This work has been supported by USPHS grants AM 13914 and AM 20595, and the Howard Hughes Medical Institute. M.W. is a recipient of a Juvenile Diabetes Foundation Postdoctoral Fellowship. We thank Ms Cathy Christopherson for her expert assistance in preparing this manuscript.

REFERENCES

[1] Unger, R.H. and Orci, L. (1981) N. Engl. J. Med. 304, 1518–1524.
[2] Bell, G.I., Sanchez-Pescador, R., Laybourn, P.J. and Najarian, R.C. (1983) Nature 304, 368–371.
[3] Bell, G.I., Santerre, R.F. and Mullenbach, G.T. (1983) Nature 302, 716–718.
[4] Patzelt, C., Tager, H.S., Carroll, R.J. and Steiner, D.F. (1979) Nature 282, 260–266.
[5] Foa, P., Bajaj, J. and Foa, N. (1977) Glucagon: Its Role in Physiology and Clinical Medicine, Springer, Berlin.
[6] Unger, R.H. and Orci, L. (eds) (1981) Glucagon, Physiology, Pathophysiology and Morphology of the Pancreatic A-cell, Elsevier, Amsterdam, New York.
[7] Conlon, J.M. (1980) Diabetologia 18, 85–88.
[8] Tager, H.S. and Markese, J. (1979) J. Biol. Chem. 254, 2229–2233.
[9] Patzelt, C. and Schiltz, E. (1984) Proc. Natl. Acad. Sci. USA 81, 5007–5011.
[10] Bell, G.I. (1986) Peptides, suppl.7, in press.
[11] Lund, P.K., Goodman, R.H., Montminy, M.R., Dee, P.C. and Habener, J.F. (1983) J. Biol. Chem. 258, 3280–3284.
[12] Beintema, J.J. (1985) in: Evolutionary Relationships Among Rodents (Luckett, W.O. and Hartenberger, J.L. eds) pp.549–565, Plenum, New York.

[13] Wriston, J.C. (1984) Comp. Biochem. Physiol. 77B, 253–278.
[14] Chan, S.J., Episkopou, V., Zeitlin, S., Karathanasis, S.K., MacKrell, A., Steiner, D.F. and Efstratiadis, A. (1984) Proc. Natl. Acad. Sci. USA 81, 5046–5050.
[15] Houck, R., Blundell, T.L., Lazarus, N.R., Neville, R.W.J., Stone, D. and Wollmer, A. (1980) Nature 286, 822–824.
[16] Zimmermann, A.E., Moule, M.L. and Yip, C.C. (1974) J. Biol. Chem. 249, 4026–4029.
[17] King, G.L. and Kahn, C.R. (1981) Nature 292, 644–646.
[18] Sundby, F. (1976) Metabolism 25, Suppl.1, 1319–1321.
[19] Conlon, J.M., Hansen, H.F. and Schwartz, T.W. (1985) Regul. Peptides 11, 309–320.
[20] Frandsen, E.K., Gronvald, F.C., Heding, L.G., Johansen, N.L., Lundt, B.F., Moody, A.J., Markussen, J. and Volund, A. (1981) Hoppe-Seyler's Z. Physiol. Chem. 362, 665–677.
[21] Huang, C.-G., Eng, J., Pan, Y.-C., Hulmes, J.D. and Yalow, R.S. (1986) Diabetes 35, 508–512.
[22] Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning: A Laboratory Manual, Cold Spring Harbor Laboratory, Cold Spring Harbor, NY.
[23] Gubler, U. and Hoffman, B.J. (1983) Gene 25, 263–269.
[24] Huynk, T.D., Young, R.A. and Davis, R.W. (1985) in: DNA Cloning: A Practical Approach (Glover, D.M. ed.) vol.1, pp.49–78, IRL Press, Oxford.
[25] Sanger, F., Coulson, A.R., Barrell, B.G., Smith, A.J.H. and Roe, B.A. (1980) J. Mol. Biol. 143, 161–178.
[26] Carmichael, G.C. and McMaster, G.K. (1980) Methods Enzymol. 65, 380–391.
[27] Thomas, P.S. (1983) Methods Enzymol. 100, 255–260.
[28] Lopez, L.C., Li, W.-H., Frazier, M.L., Luo, C.-C. and Saunders, G.F. (1984) Mol. Biol. Evol. 1, 335–344.
[29] Sundby, F., Frandsen, E.K., Thomsen, J., Kristiansen, K. and Brunfeldt, K. (1972) FEBS Lett. 26, 289–293.
[30] Lance, V., Hamilton, J.W., Rouse, J.B., Kimmel, J.R. and Pollock, H.G. (1984) Gen. Comp. Endocrinol. 55, 112–124.
[31] Steiner, D.F. (1984) in: The Harvey Lectures, Series 78, pp.191–228, Academic Press, New York.
[32] Heinrich, G., Gros, P. and Habener, J.F. (1984) J. Biol. Chem. 259, 14082–14087.
[33] Lopez, L.C., Frazier, M.L., Su, C., Kumar, A. and Saunders, G.F. (1983) Proc. Natl. Acad. Sci. USA 80, 5485–5489.
[34] Du, B.-H., Eng, J., Hulmes, J.D., Chang, M., Pan, Y.-C.E. and Yalow, R.S. (1985) Biochem. Biophys. Res. Commun. 128, 1093–1098.