Proc. Natl. Acad. Sci. USA
Vol. 81, pp. 5046–5050, August 1984
Biochemistry

Guinea pig preproinsulin gene: An evolutionary compromise?

(insulin genes/hystricomorph rodents/molecular evolution/neutral theory)

SHU JIN CHAN*, VASSO EPISKOPOU†, SCOTT ZEITLIN†, SOTIRIOS K. KARATHANASIS‡, ALBERT MACKRELL*, DONALD F. STEINER*, AND ARGIRIS EFSTRATIADIS†

*Department of Biochemistry, University of Chicago, Chicago, IL 60637; †Department of Human Genetics and Development, Columbia University, New York, NY 10032; and ‡Department of Cardiology, Harvard Medical School, Children's Hospital Medical Center, Boston, MA 02115

Contributed by Donald F. Steiner, April 26, 1984

ABSTRACT **We characterized a clone carrying the guinea pig preproinsulin gene, which, in contrast to other mammalian preproinsulin genes, is highly divergent in its regions encoding the B and A chains of mature insulin. Blot hybridization analysis indicates that this gene is present in only one copy in the guinea pig genome and that other normal or mutated preproinsulin genes do not exist in this animal. Moreover, the position of introns in this gene and the homology of its 3′ flanking region to the corresponding regions of other sequenced mammalian genes show that it has been derived from the common mammalian stock. The rapid evolution of the region encoding the B and A chains can be interpreted, according to our sequence-divergence analysis, as due to the fixation of both neutral and adaptive mutations.**

The genes for preproinsulin provide an interesting system for addressing questions related to molecular evolution. Insights into the evolutionary process can be gained by comparing the gene product or the gene sequences themselves in different species. The primary structures of insulins isolated from over 25 species are known (1), and preproinsulin gene and/or cDNA sequences are available for mammals [human (2–4), dog (5), rat (6), Chinese hamster (7), and the Old World monkey *Macaca fascicularis* (8)], birds [chicken (9)], fishes [carp (10), anglerfish (11), and salmon (12)], and a cyclostome [hagfish (13)].

The interest in preproinsulin gene evolution stems from the fact that (negative) selection operating at the preproinsulin protein level assigns distinct regions to the corresponding gene (encoding signal peptide–B chain–C peptide–A chain) which are under very different constraints. Thus, the B and A chains, which appear in the mature hormone, are highly conserved. Also conserved is the hydrophobic character and certain other features of the signal peptide (preregion), which is involved in transmembrane segregation of the nascent peptide. The C peptide is more divergent, perhaps reflecting its role as a molecular spacer whose functions in the transport, folding, and cleavage of the proinsulin molecule are less sequence dependent (14).

A major discrepancy in this picture has been observed in the insulins of certain hystricomorph rodents, such as the guinea pig. Though the C peptide divergence of guinea pig proinsulin seems to conform to the general pattern (15), the A and B chains (total of 51 residues) differ from human insulin, for example, at 18 positions (16), while most other mammalian insulins differ from each other at only 1–3 sites. This intriguing phenomenon in the hystricomorphs has been explained in the past as being due either to Darwinian selection (17) or to neutral drift (18) (see *Discussion*). More recently, Roth and collaborators (19, 20) have proposed on the basis of immunological assays that the guinea pig has two preproinsulin genes, one closer to the human/porcine type, which is expressed in extrapancreatic tissues, and a second mutated gene, which is expressed in the pancreas instead of the normal gene.

To examine these issues, we cloned and sequenced the guinea pig preproinsulin gene and compared it to the other sequenced mammalian insulin genes. As we show here, the guinea pig has a single preproinsulin gene, which has been derived from the common mammalian stock. Moreover, our analysis indicates that the evolution of the region of this gene encoding the B and A chains is most compatible with a model in which both adaptive and neutral changes accompany the acquisition of an alternative function.

MATERIALS AND METHODS

Materials. Bacteriophage vectors Charon 28 and λgt10 and plasmid vector pUC9 were provided by F. Blattner, T. Huynh, and J. Messing, respectively. Restriction enzymes, T4 DNA ligase, S1 nuclease, polynucleotide kinase, and *Eco*RI linkers were from New England Biolabs or Bethesda Research Laboratories; Klenow fragment of *Escherichia coli* DNA polymerase I and proteinase K were from Boehringer Mannheim; and reverse transcriptase was from Life Sciences (St. Petersburg, FL). [α-^{32}P]dNTPs (700 Ci/mmol; 1 Ci = 37 GBq) were from New England Nuclear; [γ-^{32}P]ATP (7000 Ci/mmol) was from ICN.

Recombinant DNA Procedures. RNA was prepared from isolated guinea pig pancreatic islets by the guanidine thiocyanate/CsCl procedure (21), and further purified by oligo-(dT)-cellulose chromatography (22). Double-stranded cDNA was synthesized as described (23). The ends of these molecules were made flush by treatment with DNA polymerase I. They were then cloned into the vector λgt10, following attachment of *Eco*RI DNA linkers.

Guinea pig chromosomal DNA was isolated from the liver of a single animal as described (24). After partial digestion with *Mbo* I, DNA fragments in the range of 16–24 kilobases were isolated by preparative agarose gel electrophoresis and cloned into the *Bam*HI vector Charon 28 as described (25). Screening of the cDNA and chromosomal DNA libraries was performed by the method of Benton and Davis (26). DNA blotting was performed by the method of Southern (27).

DNA Sequence Determination. DNA sequence analysis was performed either by the chemical method (28) or by the enzymatic method (29) after subcloning into phage M13mp9.

Calculation of Sequence Divergence. A method of calculating sequence divergence between two homologous sequences has been described by Perler *et al.* (9). These authors introduced corrections for multiple events because they were comparing sequences between species that had diverged at different evolutionary times. However, such calculations are complicated by the fact that the correction for-

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "*advertisement*" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

Abbreviations: kb, kilobase(s); PDGF, platelet-derived growth factor.

mulas are valid only under the assumption that transitions and transversions are equally probable, which is not necessarily correct (see, for example, ref. 30). Thus, we used the method only to assign silent and replacement sites and the corresponding nucleotide substitutions. We then calculated divergences [100 × (substitutions/sites)] without introducing corrections for back mutations. This is because we compared the preproinsulin sequences only among mammalian species that diverged at about the same time. These uncorrected percent divergences (percent substitution values) can be treated directly as nonlinear substitution rates (31). They constitute a relative index of evolutionary rate of fixation at the nucleotide level and allow us to derive from the analysis conclusions that do not rely on methodological assumptions. We emphasize, however, that we use these percent substitution values only as indicative values. Their statistical significance is low because the comparisons involve sequences from only six species, and the gene domains we compare are small.

RESULTS

Isolation of Guinea Pig Preproinsulin cDNA and Chromosomal DNA Clones. Poly(A)-enriched RNA was prepared from isolated guinea pig pancreatic islets, enzymatically converted into double-stranded cDNA and cloned into the λgt10 vector. When the recombinants were screened with ^{32}P-labeled guinea pig islet cDNA, a number of strongly positive plaques were obtained. One of these clones was characterized by DNA sequencing, which showed that the insert of this recombinant phage corresponded to most of the guinea pig preproinsulin mRNA, extending from the preregion (minus the codons for the first four amino acids) to the poly(A) tail. This insert was subcloned into pUC9, and the derived clone (pGPin-1) was used as a probe to screen an amplified guinea pig chromosomal DNA library in Charon 28.

We screened approximately 800,000 plaques and obtained 29 positive clones. Analysis of these recombinants by restriction enzyme mapping and cross-hybridization showed that they are all overlapping clones carrying the same guinea pig preproinsulin gene. A restriction map of the guinea pig preproinsulin gene contained in two overlapping clones (λGPin-8 and λGPin-22), which we further characterized by DNA sequencing, is shown in Fig. 1.

To examine whether a second preproinsulin gene is present in the genome, we hybridized a pGPin-1 probe to a chromosomal DNA blot (Fig. 2). This analysis revealed that the isolated gene is unique since the restriction fragments that hybridized corresponded to those expected from the map of the cloned gene.

We reasoned that if the evolutionary scheme proposed by Roth and collaborators (19, 20) is correct, the putative second gene might not hybridize to pGPin-1 sequences under stringent conditions, but it should hybridize to a rat insulin cDNA probe because it was thought to be similar to the typical mammalian preproinsulin genes. However, repeat of the DNA blot analysis with a rat probe failed to reveal any hybridizing bands, even under less stringent conditions (not shown).

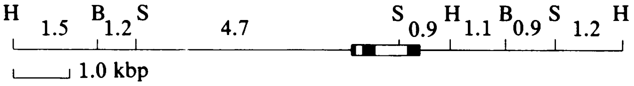

FIG. 1. Structure of the guinea pig preproinsulin gene. Size of fragments between restriction sites for *Hin*dIII (H), *Bam*HI (B), and *Sst* I (S), given in kilobase pairs (kbp), were derived from restriction mapping of two overlapping clones, λGPin-8 and λGPin-22. Shaded boxes represent positions of the three expressed sequences, and open boxes denote the two intervening sequences found in guinea pig preproinsulin gene.

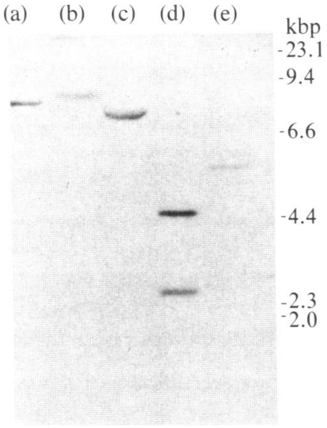

FIG. 2. Southern blot hybridization of guinea pig DNA. High molecular weight DNA isolated from guinea pig spleen (24) was digested with *Bam*HI (lane a), *Hin*dIII (lane b), *Bam*HI/*Hin*dIII (lane c), *Sst* I (lane d), or *Stu* I (lane e), electrophoresed on a 0.6% agarose gel transferred onto nitrocellulose, and hybridized with nick-translated pGPin-1 labeled with [α-^{32}P]dCTP to approximately 10^8 cpm/μg. Standard size markers are from *Hin*dIII-digested λDNA and are shown in kilobase pairs (kbp).

Analysis of the Cloned Guinea Pig Preproinsulin Gene. The nucleotide sequence of guinea pig preproinsulin gene in clone λGPin-22 is shown in Fig. 3. The poly(A) addition site was positioned by comparison to the sequence of the cDNA clone pGPin-1, while the capping site was positioned by comparison to the known capping sites of the rat I and II genes (7).

The structure of this gene is similar to that of the other preproinsulin genes that have been characterized (with the exception of the rat preproinsulin I gene). Thus, the gene contains two introns—a small intron (119 base pairs), which interrupts the region corresponding to the 5′ noncoding region of the mRNA, and a second larger intron (613 bp), which is located between the codons for the sixth and seventh amino acid residues of the C peptide. While the position of the large intron is unequivocal, the small intron was positioned by comparison to other insulin genes (our cDNA clones do not cover this area).

The 5′ flanking region of the guinea pig gene, which includes the "TATA" box, is very homologous to the corresponding region of the other sequenced mammalian genes (Fig. 4*A*), and no lesions are apparent in the primary structure of this promoter-related area. Moreover, the presumably functionless 3′ flanking region of this gene is also homologous to the corresponding region of the other mammalian genes (Fig. 4*B*). These homologies (especially the latter) strongly indicate that the divergent guinea pig preproinsulin gene has the same evolutionary origin as the other mammalian genes because it is embedded in DNA derived from the common ancestor. Thus, guinea pig insulin is not the product of a paralogous gene, as proposed by Rosenzweig *et al.* (20).

Translation of the sequence reveals that the preproinsulin signal peptide is 24 residues long and contains a core rich in amino acids with hydrophobic side chains, as expected. The sequence for the B and A chains is in agreement with the previously determined protein sequence (16). However, the translated sequence for the C peptide is longer by two amino acids from the protein sequence reported earlier (32), which fills the previously proposed gaps in the alignment of mammalian C peptides. Two other residues (positions 27 and 28) are different from those reported. The differences are most likely due to a loss of a COOH-terminal tripeptide during cyanogen bromide cleavage of the methionine residue at position 28. Alternatively, the COOH-terminal part might have been lost prior to the analysis, because C peptides are occasionally truncated during isolation (33), or *in vivo* after the cleavages of the processing pathway (5, 34).

DISCUSSION

Our data indicate that the guinea pig genome contains a unique preproinsulin gene that, despite its common evolutionary origin with preproinsulin genes of other mammals,

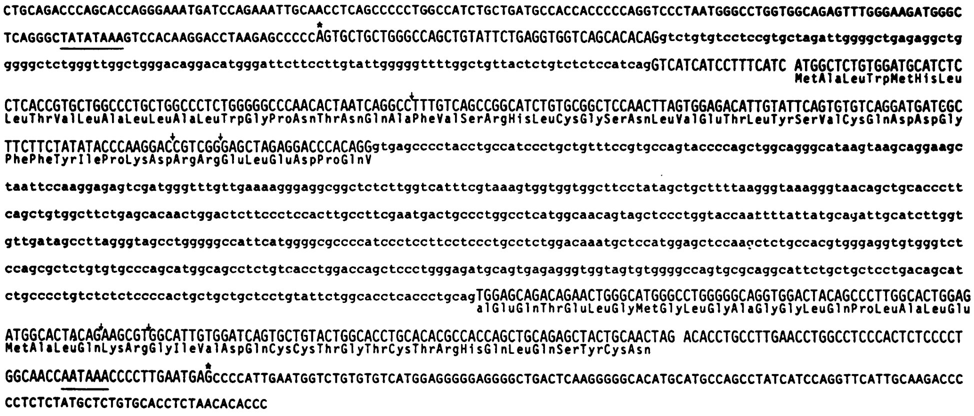

FIG. 3. Nucleotide sequence of the guinea pig preproinsulin gene. The nucleotide sequence of the mRNA strand is displayed from the 5′ to the 3′ direction. Small and large capital letters represent the flanking and mature mRNA sequences, respectively, while the intron sequences are in lower case letters. The amino acid sequence is displayed on a line below the coding sequence. The TATA box and the polyadenylylation signal are underlined. Asterisks indicate the capping and poly(A) addition sites. The arrows indicate, in order, the boundaries of the preregion and the peptides B, C, and A.

has diverged disproportionately in the A and B chain-encoding segments.

In the past, both of the competing hypotheses concerning molecular evolution—i.e., selectionism and neutralism—have been invoked to explain the hystricomorph insulin changes, using different arguments. The point that the neutralists debate is the extension of the neo-Darwinian views on organismal evolution to the molecular level. According to these views, most mutant genes are either more or less adaptive than their ancestors. Less-adaptive genes are eliminated from the population by negative selection, while more-adaptive genes are fixed in the population by positive selection. Negative selection is accepted by the neutralists. They also accept that favorable mutations do occur. However, according to the neutral theory, such mutations are so rare as to be neglected in calculating rates of molecular evolution. Thus, with the exception of the disadvantageous mutants, which will be eliminated, most of the mutant genes are selectively neutral—that is, they are adaptively neither more nor less advantageous than the genes they replace. Most evolutionary changes are then the result of fixation in the population of randomly drifting, selectively equivalent mutant genes.

Although several lines of evidence support the neutralist view (see refs. 35 and 36 for reviews), the crucial question remains as to the exact magnitude of the neutral element in molecular evolution. It seems plausible that in addition to random drift, a significant number of adaptive mutations also may be fixed.

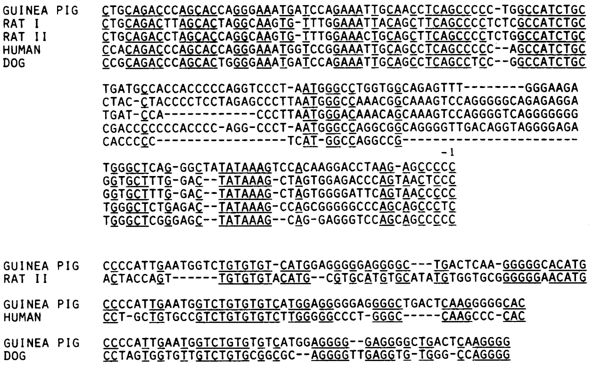

FIG. 4. (*A*) Group alignment of the 5′ flanking regions of sequenced mammalian preproinsulin genes. The sequences begin at the first nucleotide (−1) to the 3′ side of the capping site and extend upstream. Nucleotides identical to all of the sequences are underlined. (*B*) Alignment of the guinea pig 3′ flanking preproinsulin gene sequence to the corresponding sequences of other mammalian preproinsulin genes. The sequences begin with the first nucleotide following the poly(A) addition site and extend downstream. Homologous nucleotides are underlined. Because of extensive sequence divergence in this region, it is difficult to align these sequences as a group. With the exception of the rat sequences (unpublished results), the data are from refs. 4 and 5.

In guinea pig insulin, the usually conserved Zn-coordinating B10 histidine residue (37) has been replaced by aspartate, rendering the molecule unable to form Zn-insulin hexamers (the storage form in most species). The neutralists contend (18) that in the process of speciation, the loss of a selective constraint related to Zn-binding freed the gene from negative selection and allowed the accumulation of additional neutral mutations at a more rapid rate. In contrast, the selectionist view (17) asserts that the observed changes are adaptive. By assuming that storage of the hormone is as important to the fitness of the organism as is receptor binding and biological activity, the following hypothesis can be constructed. A (local) shortage of Zn may have precluded hexamerization during storage of insulin in the beta cells of the guinea pig ancestors. Thus, selectively advantageous mutations were fixed. For example, hydrophobic residues at surface contact points (involved in aggregation to hexamers) were replaced with more hydrophilic residues, making the insulin molecule more stable in the monomeric form in aqueous environments. Specifically the B chain residues at positions B14 (alanine), B17 (leucine), and B20 (glycine) have been replaced by threonine, serine, and glutamine, respectively (Fig. 3).

An environmental factor leading to insulin divergence and presumably to convergent evolution (38) was implicated because, in addition to the hystricomorphs, the insulins of two other New World species (the New World monkeys *Cebus appella* and *Saimiri sciurea*) do not cross-react with porcine insulin antibodies (39). However, environmental Zn shortage would seem unlikely in view of its essentially ubiquitous geographic distribution, and Zn is an essential trace metal (component of many enzymes). Moreover, the insulin of the porcupine, which is an Old World hystricomorph, is also monomeric and of low metabolic activity, despite the fact that it retains the histidine at position B10 (40). Thus, we conclude that Zn shortage as a basis for adaptation is untenable.

What then can we learn from the DNA sequence about the evolutionary events that led to the appearance of the divergent guinea pig insulin gene? Our results clearly show that this gene is a typical mammalian insulin gene in all respects except for increased rates of substitution in the A and B chain replacement sites.

The percent substitution values, derived from comparisons of the guinea pig gene to the other sequenced mammalian genes (Table 1) seem to support the neutral theory. The substitution values in silent sites do not differ from any other mammal and are approximately the same (average of 30%) for all of the gene segments (encoding A and B chains, preregion, and C peptide). This is compatible with the prediction of the neutral theory that, aside from any mRNA structural requirements, silent sites in codons are not usually under selective constraint. However, comparison of the substitution values (Table 1) in the replacement sites of the gene regions leads to interesting conclusions. The values for the C peptide and the preregion are equal and approximately the same for

Table 1. Percent substitution values

	Silent sites			Replacement sites		
	A + B	Pre	C	A + B	Pre	C
Human/dog	32	21	13	1	11	12
Monkey/dog	21	25	16	1	9	14
Human/rat	29	27	39	3	16	15
Monkey/rat	24	24	41	3	14	14
Dog/rat	29	22	38	3	16	19
Human/hamster	31	34	36	2	14	16
Monkey/hamster	24	30	34	2	12	15
Dog/hamster	29	28	38	2	17	20
Rat/hamster	21	15	19	2	11	6
Guinea pig/human	29	36	29	17	17	18
Guinea pig/monkey	27	32	29	17	16	17
Guinea pig/dog	38	30	41	16	19	22
Guinea pig/rat	33	24	40	18	22	18
Guinea pig/hamster	22	21	42	17	19	19
Mammal/mammal (avg)	27	25	30	2	13	15
Guinea pig/mammal (avg)	30	31	36	17	19	19
Chinchilla/mammal (avg)				6		
Porcupine/mammal (avg)				11		
Coypu/mammal (avg)				17		
Casiragua/mammal (avg)				19		
Guinea pig/chinchilla				15		
Guinea pig/porcupine				15		
Guinea pig/coypu				14		
Guinea pig/casiragua				17		
Chinchilla/porcupine				12		
Chinchilla/coypu				15		
Chinchilla/casiragua				18		
Porcupine/coypu				18		
Porcupine/casiragua				19		
Coypu/casiragua				5		
Hystricomorph/ hystricomorph (avg)				16		

The percent substitution values of each pair of sequences was calculated separately for the A and B chains (A + B), the preregion (Pre) and the C peptide as described. The values that involve rat sequences are averages of separate calculations with rat gene *I* and gene *II* preproinsulin encoding segments. The data are from refs. 2–8. The values for the replacement sites of the A and B chains of hystricomorph insulins were calculated by converting the known amino acid sequences (40, 41) to nucleotide sequences. For amino acids with six codons (leucine, serine, and arginine), we chose the codon by homology to the other mammalian sequences.

the guinea pig and the other mammals. They are, however, lower than the corresponding values for silent sites, which suggests that some negative selective pressure is operating in this case, even for the C peptide with its presumably less stringent functions. Actually, 12 of 23 residues in the preregion and 9 of 31 residues in the C peptide are invariant among mammals, including the guinea pig.

An interesting argument can be made from this analysis as follows. The preregion is necessary for the transmembrane segregation of preproinsulin, and this function is related to its hydrophobicity, especially in the central region of the prepeptide (42). Since all of the replacements in this lipophilic core are conservative, involving only hydrophobic (or polar but uncharged) amino acids, they can be considered as neutral (i.e., not harmful). Thus, the replacements in the C peptide (in which the functional constraints may be even less demanding) are definitely neutral. Actually, they are in their majority also conservative.

In the guinea pig, however, the replacement rate in the A and B chains is also at the level of the neutral value. In the absence of other information, this result would lead to the conclusion that the divergence of these regions is also explicable in terms of neutral evolution. We believe, however, that a more complex picture is more realistic in view of the strong negative selection in other mammalian A and B chains in which the replacement rate is approximately one-seventh of the neutral rate.

Further evidence against neutral evolution is also evident from the comparisons in Table 1, which show the range of divergence of various hystricomorph insulins from the typical mammalian sequences. Thus, chinchilla insulin is very close to the typical mammalian molecules (43), while porcupine insulin is in-between the two extremes in terms of divergence but is similar to guinea pig insulin in terms of its biological activity (40). It appears that most of the hystricomorph sequences have diverged between themselves to the same extent that some of them differ from the typical mammalian insulins.[§]

Are we then forced to conclude that all of the changes in hystricomorph insulins are adaptive and therefore positively selected? This is unlikely because changes in certain residues are neutral. For example, certain replacements have occurred in positions where typical mammalian insulins differ among themselves—i.e., positions A8, A9, and A10, and B30. It is known that deletion of the B30 residue does not impair biological activity (45). On the other hand, neutrality cannot easily explain why the drastic change of the Zn-binding histidine at position B10 to glutamine or asparagine in the coypu (or casiragua) and guinea pig insulins, respectively, could be maintained by negative selection as a conservative replacement. These animals have also in common a threonine at position B14, a serine at position B17, and an arginine at position A13, while the A12 or A18 residues are either serine or threonine.

If positive selection has indeed operated, what was its basis? Although an explanation in molecular detail is not feasible for the moment, we will discuss, as examples, two possible evolutionary schemes which are not mutually exclusive.

In the first case, we suggest that one or more slightly deleterious mutations were fixed in the molecule by random drift. Despite a substantial loss in biological activity, the molecule could still function as insulin, as long as negative selection was able to maintain some residual biological activity. This would generate a need for selection of compensatory mutations. The guinea pig has approximately 10 times more insulin in the serum than other mammals, and its insulin is degraded approximately 2-fold slower than is porcine insulin (46). It is known that mutant insulins in humans that have lowered receptor binding potency also accumulate in the serum and turn over more slowly, presumably due to decreased receptor-mediated endocytosis of the hormone (47). But in addition, the insulin receptor level is elevated in guinea pig tissues, although the receptor does not exhibit enhanced binding of guinea pig insulin relative to others (41). Thus, it could be argued that compensatory changes allow the insulin of the guinea pig to remain effective as a metabolic hormone.

The second scheme suggests that hystricomorph insulins might have acquired an alternative functionality, as indicated by a recent, rather remarkable observation. Though the metabolic biological activity of hystricomorph insulins is low in comparison to that of their typical mammalian counter-

[§] To explain this observation with the neutral theory, we would have to conclude that substitutions in replacement sites reached saturation levels very rapidly because the divergence time for these animals is definitely more recent than for the mammalian radiation. In this regard the casiragua/coypu divergence, which is only 5%, remains unexplained unless the taxonomic relationship between these two animals is much closer than has been generally stated. On the other hand, we note that the cuis which belongs to the same subfamily as the guinea pig appears to have a bovine-type insulin (44).

parts, their growth activity is substantially higher than that of any other insulin (or somatomedin) tested in parallel (48). This activity appears to be exerted through a receptor that does not normally recognize other insulins or somatomedins but may recognize platelet-derived growth factor (PDGF) instead (49). Thus, it seems possible that the hystricomorph insulins have compromised their metabolic activity in order to adaptively acquire the ability to recognize the PDGF receptor, possibly because of a deleterious mutation elsewhere (e.g., in the hystricomorph PDGF or somatomedin genes or their respective receptors). The reduction in insulin receptor binding leading to increases in circulating hormone may then be viewed as representing further adaptations that enhance this new function of guinea pig insulin, inasmuch as relatively high levels are required to stimulate the PDGF receptor (49). The foregoing scheme envisions primarily the operation of positive selection, even if we accept that the replacements in the guinea pig insulin are neutral in origin, as the percent substitution values in Table 1 imply. This would be an example of the generation of a new function through the fortuitous assortment of neutral or slightly deleterious mutations fixed by random drift.

It is of interest that the guinea pig has been reported also to have a highly mutated glucagon molecule (50). Its biological potency is not known as yet, but it is tempting to speculate that this otherwise highly invariant metabolic hormone (51, 52) also may have undergone compensatory changes in these rodents. The divergence of glucagon, if confirmed by cloning and DNA sequence analysis of the gene, would demonstrate that adaptive evolution in the guinea pig gastroenteropancreatic endocrine system is extensive.

We thank Ed Fritsch for invaluable help during the construction of the guinea pig chromosomal DNA library and Leigh Van Valen for helpful discussions. We also thank Cathy Christopherson for expert assistance in the preparation of this manuscript. This work was supported by grants from the National Institutes of Health to D.F.S. (AM 13914 and AM 20595) and A.E. (AM 32085). A.E. was supported by a Career Development Award from the Juvenile Diabetes Foundation.

1. Steiner, D. F. (1976) in *Handbook of Biochemistry and Molecular Biology: Proteins*, ed. Fasman, G. D. (CRC, Cleveland, OH), Vol. 3, pp. 378–381.
2. Bell, G. I., Pictet, R. L., Rutter, W. J., Cordell, B., Tischer, E. & Goodman, H. M. (1980) *Nature (London)* **284**, 26–32.
3. Ullrich, A., Dull, T. J., Gray, A., Brosius, J. & Sures, I. (1980) *Science* **209**, 612–615.
4. Ullrich, A., Dull, T. J., Gray, A., Philips, J. A. & Peter, S. (1982) *Nucleic Acids Res.* **10**, 2225–2240.
5. Kwok, S. C. M., Chan, S. J. & Steiner, D. F. (1983) *J. Biol. Chem.* **258**, 2357–2363.
6. Lomedico, P., Rosenthal, N., Efstratiadis, A., Gilbert, W., Kolodner, R. & Tizard, R. (1979) *Cell* **18**, 545–558.
7. Bell, G. I. & Sanchez-Pescador, R. (1984) *Diabetes* **33**, 297–300.
8. Wetekam, W., Groneberg, J., Leineweber, M., Wengenmayer, F. & Winnaker, E.-L. (1982) *Gene* **19**, 179–183.
9. Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R. & Dodgson, J. (1980) *Cell* **20**, 555–566.
10. Hahn, V., Winkler, J., Rapoport, T. A., Liebscher, D.-H., Coutelle, C. & Rosenthal, S. (1983) *Nucleic Acids Res.* **11**, 4541–4552.
11. Hobart, P. M., Shen, L.-P., Crawford, R., Pictet, R. L. & Rutter, W. J. (1980) *Science* **210**, 1360–1363.
12. Sorokin, A. V., Petrenko, O. I., Kavsan, V. M., Kozlov, Y. I., Debabov, V. G. & Zlochevskij, M. L. (1982) *Gene* **20**, 367–376.
13. Chan, S. J., Emdin, S. O., Kwok, S. C. M., Kramer, J. M., Falkmer, S. & Steiner, D. F. (1981) *J. Biol. Chem.* **256**, 7595–7602.
14. Steiner, D. F. (1978) *Diabetes* **27**, Suppl. 1, 145–148.
15. Markussen, J. & Vølund, A. (1974) *Int. J. Pept. Protein Res.* **6**, 79–86.
16. Smith, L. F. (1966) *Am. J. Med.* **40**, 662–666.
17. Blundell, T. L. & Wood, S. P. (1975) *Nature (London)* **257**, 197–203.
18. Kimura, M. & Ohta, T. (1974) *Proc. Natl. Acad. Sci. USA* **71**, 2848–2852.
19. Rosenzweig, J. L., Lesniak, M. A., Samuels, B. E., Yip, C. C., Zimmerman, A. E. & Roth, J. (1980) *Trans. Assoc. Am. Physicians* **93**, 263–278.
20. Rosenzweig, J. L., LeRoith, D., Lesniak, M. A., MacIntyre, I., Sawyer, W. B. & Roth, J. (1983) *Fed. Proc. Fed. Am. Soc. Exp. Biol.* **42**, 2608–2614.
21. Chirgwin, J. M., Przybyla, A. E., MacDonald, R. J. & Rutter, W. J. (1979) *Biochemistry* **18**, 5294–5299.
22. Aviv, H. & Leder, P. (1972) *Proc. Natl. Acad. Sci. USA* **69**, 1408–1412.
23. Chan, S. J., Noyes, B. E., Agarwal, K. L. & Steiner, D. F. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 5036–5040.
24. Blin, N. & Stafford, D. W. (1976) *Nucleic Acids Res.* **3**, 2303–2308.
25. Maniatis, T., Fritsch, E. F. & Sambrook, J. (1982) *Molecular Cloning: A Laboratory Manual* (Cold Spring Harbor Laboratory, Cold Spring Harbor, NY), p. 270.
26. Benton, W. D. & Davis, R. W. (1977) *Science* **196**, 180–182.
27. Southern, E. M. (1975) *J. Mol. Biol.* **98**, 503–517.
28. Maxam, A. M. & Gilbert, W. (1980) *Methods Enzymol.* **65**, 499–560.
29. Sanger, F., Nicklen, S. & Carlson, A. R. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5463–5467.
30. Brown, W. M., Prager, E. M., Wang, A. & Wilson, A. C. (1982) *J. Mol. Evol.* **18**, 225–239.
31. Kafatos, F. C., Efstratiadis, A., Forget, B. G. & Weissman, S. M. (1977) *Proc. Natl. Acad. Sci. USA* **74**, 5618–5622.
32. Smyth, D. G., Markussen, J. & Sundby, F. (1974) *Nature (London)* **248**, 151–152.
33. Kuzuya, H., Blix, P. M., Horwitz, D. L., Rubenstein, A. H., Steiner, D. F., Faber, O. K. & Binder, C. (1978) *Diabetes* **27**, Suppl. 1, 184–191.
34. Tager, H. S., Emdin, S. O., Clark, J. L. & Steiner, D. F. (1973) *J. Biol. Chem.* **248**, 3476–3482.
35. Kimura, M. (1982) in *Molecular Evolution, Protein Polymorphism and the Neutral Theory*, ed. Kimura, M. (Japan Scientific Societies, Tokyo), pp. 3–56.
36. Kimura, M. (1983) in *Evolution of Genes and Proteins*, eds. Nei, M. & Koehn, R. K. (Sinauer Associates, Sunderland, MA), pp. 208–233.
37. Blundell, T. L., Dodson, G. G., Hodgkin, D. C. & Mercola, D. A. (1971) *Adv. Protein Chem.* **26**, 279–402.
38. Wriston, J. C. (1981) *J. Mol. Evol.* **17**, 1–9.
39. Mann, G. V. & Crofford, O. B. (1970) *Science* **169**, 1312–1313.
40. Horuk, R., Blundell, T. L., Lazarus, N. R., Neville, R. W. J., Stone, D. & Wollmer, A. (1980) *Nature (London)* **286**, 822–824.
41. Horuk, R., Goodwin, P., O'Connor, K., Neville, R. W. J., Lazarus, N. R. & Stone, D. (1979) *Nature (London)* **279**, 439–440.
42. Steiner, D. F., Quinn, P. S., Patzelt, C., Chan, S. J., Marsh, J. & Tager, H. S. (1980) in *Cell Biology: A Comprehensive Treatise*, eds. Goldstein, L. & Prescott, D. M. (Academic, New York), Vol. 4, pp. 175–201.
43. Wood, S. P., Blundell, T. L., Wollmer, A., Lazarus, N. R. & Neville, R. W. J. (1975) *Eur. J. Biochem.* **55**, 531–542.
44. Neville, R. W. J., Weir, B. J. & Lazarus, N. R. (1974) *Symp. Zool. Soc. London* **34**, 417–435.
45. Katsoyannis, P. G. (1967) *Adv. Exp. Med. Biol.* **2**, 278–296.
46. Zimmerman, A. E., Moule, M. L. & Yip, C. C. (1974) *J. Biol. Chem.* **249**, 4026–4029.
47. Hanada, M., Polonsky, K. S., Bergenstal, R. M., Jaspan, J. B., Schoelson, S. E., Blix, P. M., Chan, S. J., Kwok, S. C. M., Wishner, W. B., Zeidler, A., Olefsky, J. M., Freidenberg, G., Tager, H. S., Steiner, D. F. & Rubenstein, A. H. (1984) *N. Engl. J. Med.* **310**, 1288–1294.
48. King, G. L. & Kahn, C. R. (1981) *Nature (London)* **292**, 644–646.
49. King, G. L., Kahn, C. R. & Heldin, C.-H. (1983) *Proc. Natl. Acad. Sci. USA* **80**, 1308–1312.
50. Sundby, F. (1976) *Metabolism* **25**, Suppl. 1, 1319–1321.
51. Unger, R. H. & Orci, L. (1981) *N. Engl. J. Med.* **304**, 1518–1524.
52. Unger, R. H. & Orci, L. (1981) *N. Engl. J. Med.* **304**, 1575–1580.