Rapid Publications
Sequence of a cDNA Encoding Syrian Hamster Preproinsulin

GRAEME I. BELL AND RAY SANCHEZ-PESCADOR

SUMMARY

A complementary DNA (cDNA) library was prepared from messenger RNA (mRNA) isolated from Syrian hamster islets. Bacterial colonies containing hamster preproinsulin cDNA were identified by cross-hybridization with the human preproinsulin gene. The sequences of two of these established the complete sequence of Syrian hamster preproinsulin mRNA and predicted the sequence of the protein. Hamster preproinsulin is 110 amino acids and possesses 90.0% and 82.7% identity with the corresponding proteins of rats (either I or II) and human beings, respectively. Analysis of the hybridization of hamster preproinsulin cDNA to restriction endonuclease digests of hamster DNA suggests that there is only a single preproinsulin gene in this rodent, in contrast to rats and mice, which possess two nonallelic genes. DIABETES 33:297–300, March 1984.

Santerre et al.[1] have recently established and characterized a beta cell line (HIT) obtained by simian virus 40 transformation of Syrian hamster islets. These cells synthesize insulin, and in one subclone (T15) the insulin content is 0.26% of the total cell protein (the insulin content of normal hamster islets is 5.6%). Moreover, glucose and glucagon stimulate insulin release from these cells. Since HIT cells possess most of the differentiated functions of the beta cell, they represent a model for studying the regulation of insulin biosynthesis and secretion. Knowledge of the structure of Syrian hamster preproinsulin and its gene would be useful for these analyses.

We report here the cloning and complete nucleotide sequence of mRNA coding for Syrian hamster preproinsulin and the predicted amino acid sequence of this protein. In addition, we have examined the organization of the insulin gene by hybridization to restriction endonuclease digests of hamster DNA.

From the Chiron Corporation, 4560 Horton Street, Emeryville, California 94608. Address reprint requests to Dr. Graeme I. Bell at the above address.
Received for publication 14 December 1983.

MATERIALS AND METHODS

The construction of the Syrian hamster (*Mesocricetus auratus*) islet cDNA library has been described previously.[2] Five hundred colonies were grown in arrays on Whatman 541 paper (Whatman, Clifton, New Jersey) and, after amplification of the plasmid DNA in situ,[3] were screened for those encoding insulin by cross-hybridization with a ^{32}P-labeled DNA segment containing the human insulin gene and flanking sequences as previously described.[4]

The sequences of the inserts were determined by the procedures of Maxam and Gilbert[5] and Sanger et al.[6] from the unique Hae II and Pst I sites in the insulin cDNA clones. The majority of the sequence was determined from only one strand; however, two independently isolated clones were sequenced. No potential sequencing artifacts were observed.

Genomic DNA was prepared from HIT-T15 cells,[1] digested with restriction endonucleases, blotted, and hybridized with ^{32}P-labeled pshi1 as described previously.[7]

RESULTS AND DISCUSSION

Twenty of the 500 colonies analyzed hybridized with the human insulin gene probe. DNA was prepared from two of these (pshi1 and 2), and the inserts of both were completely sequenced (Figure 1). The sequence of pshi1 extended from nucleotide 1 to the polyadenylation site. The sequence of pshi2 was identical to pshi1 and extended from nucleotides 1 to 440; this clone lacked the poly A tract and the four preceding bases. The predicted amino acid sequence of hamster preproinsulin is also indicated in Figure 1, as well as the corresponding amino acid of the human protein, if different. Hamster and human preproinsulin are both 110 amino acids and possess 82.7% identity. The sequences of the B- and A-chains each possess one conservative substitution; five of 24 residues of the signal peptide and 11 of 31 in the C-peptide are different. The most interesting difference in

AGCCCUAGGUGACCAGCUAUAAUCAGAAGCCAUCAGCAAGCAGGUUAUUGUUCUACC **57**

−24	Ala			−20							Ala		−11	Ala	
Met	Thr	Leu	Trp	Met	Arg	Leu	Leu	Pro	Leu	Leu	Thr	Leu	Leu	Val	
AUG	ACC	CUG	UGG	AUG	CGC	CUU	CUG	CCC	CUG	UUG	ACC	CUG	CUG	GUC	102
		Gly		Asp			Ala		1						
Leu	Trp	Glu	Pro	Asn	Pro	Ala	Gln	Ala	Phe	Val	Asn	Gln	His	Leu	
CUG	UGG	GAG	CCC	AAC	CCU	GCC	CAG	GCU	UUU	GUC	AAC	CAG	CAC	CUU	147
			10										20		
Cys	Gly	Ser	His	Leu	Val	Glu	Ala	Leu	Tyr	Leu	Val	Cys	Gly	Glu	
UGU	GGC	UCC	CAC	CUU	GUG	GAG	GCA	CUC	UAC	UUG	GUG	UGU	GGG	GAG	192
							29	Thr			Glu	Ala			
Arg	Gly	Phe	Phe	Tyr	Thr	Pro	Lys	Ser	Arg	Arg	Gly	Val	Glu	Asp	
CGU	GGC	UUC	UUC	UAC	ACA	CCC	AAG	UCC	CGU	CGU	GGA	GUG	GAG	GAC	237
Leu		39	Gly		Val								50	Gly	
Pro	Gln	Val	Ala	Gln	Leu	Glu	Leu	Gly	Gly	Gly	Pro	Gly	Ala	Asp	
CCA	CAA	GUG	GCA	CAA	CUG	GAG	CUG	GGU	GGU	GGC	CCC	GGA	GCA	GAU	282
Ser			Pro				59	Gly	Ser	Leu					
Asp	Leu	Gln	Thr	Leu	Ala	Leu	Glu	Val	Ala	Gln	Gln	Lys	Arg	Gly	
GAC	CUU	CAG	ACC	UUG	GCG	CUG	GAG	GUG	GCC	CAG	CAG	AAG	CGC	GGC	327
		Glu	70										80		
Ile	Val	Asp	Gln	Cys	Cys	Thr	Ser	Ile	Cys	Ser	Leu	Tyr	Gln	Leu	
AUU	GUG	GAU	CAG	UGC	UGC	ACC	AGC	AUC	UGC	UCG	CUC	UAC	CAG	CUA	372
				86											
Glu	Asn	Tyr	Cys	Asn	AM										
GAG	AAC	UAC	UGC	AAC	UAG	UCCUCACCACUACCCAGCCCACCUUCCUGCAGCGAA									426

UAAAACCUUUGAAGGAGCAAAAAA

FIGURE 1. Sequence of hamster preproinsulin mRNA and protein and comparison with the human protein. The predicted amino acid sequence of hamster preproinsulin is indicated above the nucleotide sequence. The corresponding amino acid residue in human preproinsulin is indicated above the hamster protein if different. The pairs of basic amino acids on either side of the C-peptide are boxed. The nucleotide at the end of each line is numbered and only a portion of the poly A tract that extends from cytosine–444 is indicated.

the hamster sequence is the Gly (residue C-1) following the Arg Arg at the B-chain C-peptide boundary. (Although only one of the DNA strands was sequenced in this region, the sequence was clear and identical in two independently isolated recombinants, pshi1 and 2.) In 15 other vertebrate C-peptides that have been sequenced,[8] this residue is either Glu or Asp. The significance of this difference, if any, is unknown.

Since the sequence of the 5′-end of the hamster cDNA clone is similar to the 5′-terminal sequences of other mammalian insulin mRNAs[8–12] (Figure 2), it probably contains the entire 5′-untranslated region of the mRNA and is 57 bases. This comparison also indicates that sections of the 5′-untranslated region are remarkedly conserved between rodents, primates, and dogs, and furthermore, that this region is evolving by a process of small deletions/insertions and point mutations. Moreover, many of the substitutions are transitions, i.e., purine → purine or pyrimidine → pyrimidine changes. The homology in the 5′-untranslated region based on this alignment is also indicated. It is tempting to speculate that the sequence conservation in the 5′-untranslated region of insulin mRNA may indicate a role for this region in glucose-mediated regulation of translation.[13–15] The 3′-untranslated region of hamster insulin mRNA is 57 bases and possesses little identity with the corresponding 76-base human segment, except in the region between the AAUAAA and the

Human	**AGCCCUCCAGGAC:AGGCUGCA:UCAGAAGA:GGCCAU:::CAAGCAGAUCACUGUCCUUCUGCC**
Monkey	**AACCCUCCGGGAC:AGGCUGCA:UCAGAAGA:GGUCAG:::CAAGCAGGUCACUGUCCUUCGGCU (88.1%)**
Dog	**AGCCCCC:GGAACCA:UCUGCACCCCGA:CACGGCCGG:::CAAACAGGUC:::::::::::GCC (54.8%)**
Rat I	**AACCCUAAGUGACCAG:CUACAAUCAUAGA::::CCAUCAGCAAGCAGGUCAUUG:::UUCCAAC (59.4%)**
Rat II	**AGCCCUAAGUGACCAG:CUACAGUCGGAAA::::CCAUCAGCAAGCAGGUCAUUG:::UUCCAAC (62.5%)**
Hamster	**AGCCCUAGGUGACCAG:CUAUAAUCAGAAG::::CCAUCAGCAAGCAGGUUAUUG:::UUCUACC (65.6%)**

FIGURE 2. Comparison of the 5′-untranslated regions of mammalian preproinsulin mRNAs. Spaces, indicated by colons, have been introduced to maximize the homology between all the sequences. The identity relative to the human sequence is indicated in parentheses (each colon is counted as a substitution). Conserved nucleotides are underlined in the hamster sequence.

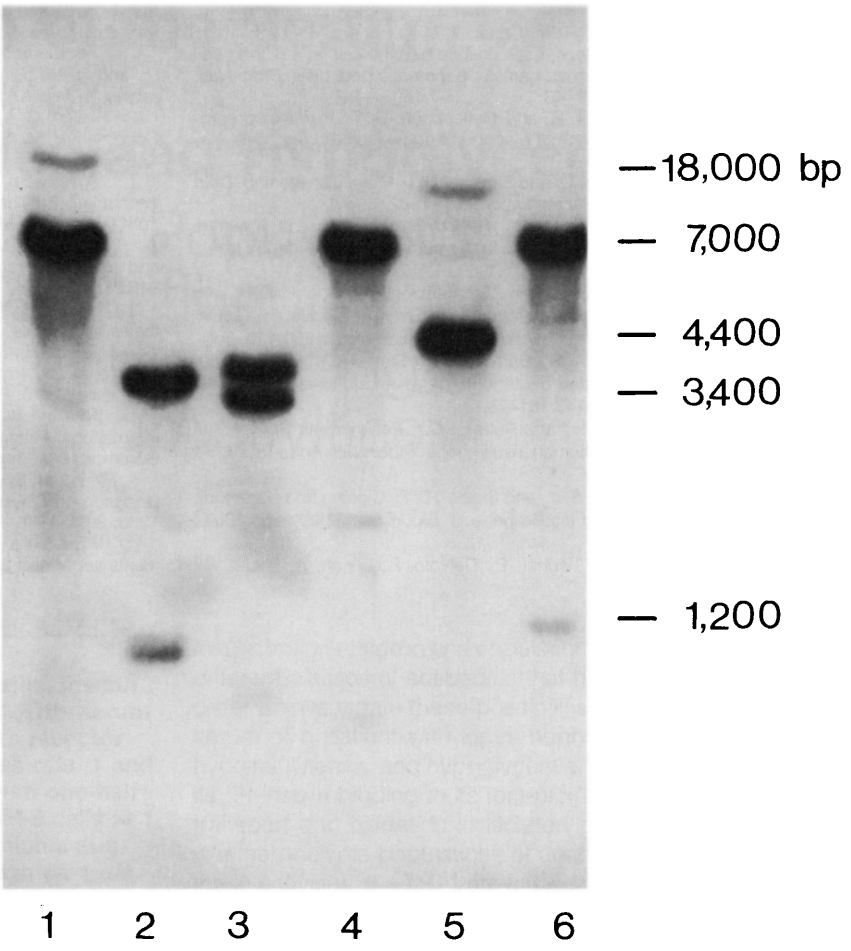

FIGURE 3. Hybridization of hamster insulin cDNA to restriction endonuclease digests of hamster DNA. The sizes of some of the hybridizing fragments are indicated on the right. The digests and the sizes of the hybridizing DNA fragments are: lane 1, EcoRI-(18,000 and 7000 bp); lane 2, Pst I-(3600 and 1000 bp); lane 3, Pvu II-(3900, 3400, and 800 bp); lane 4, Sst I-(7000, 2000, and 700 bp); lane 5, Hind III-(12,000 and 4400 bp); lane 6, Bam HI-(7000 and 1200 bp).

poly A tract. The overall identity between human and hamster insulin mRNA is 73.1%; however, the identity is greater (81.5%) if the 5′- and 3′-untranslated regions are excluded from the comparison. There is 90.0% amino acid sequence identity between Syrian hamster and rat II preproinsulin, and 88.3% nucleotide sequence identity between their mRNAs.

Laboratory rats and mice have at least two nonallelic insulin genes and synthesize two preproinsulin molecules that differ in sequence.[9,10,16] In contrast, molecular cloning has shown that human beings,[11] dogs,[8] and chickens[17] have a single insulin gene. The organization of the Syrian hamster insulin gene was examined by the blotting procedure of Southern.[18] DNA from HIT-T15 cells was digested with EcoRI, Bam HI, Hind III, Pst I, Pvu II, and Sst I and the sizes of the hybridizing fragments determined. Except for Pst I, which is expected to cleave in the 3′-untranslated region of the gene, the recognition sites of these enzymes are not present in the coding (i.e., mRNA) portion of the gene. Each enzyme generated at least two hybridizing bands (Figure 3). However, in most cases, one of the hybridizing fragments was only observed upon prolonged exposure. This could result if there are sites for these enzymes within the intervening sequence (Intron A), which interrupts the 5′-untranslated region of all insulin genes examined to date. Since the coding region present in a restriction fragment containing only exon 1 is probably less than 50 base pairs (bp), its hybridization intensity would be substantially reduced when compared with a restriction fragment containing the remainder of the gene. The pattern observed with Pvu II digestion (Figure 3, lane 3) could result from the presence of a Pvu II site in each of the two introns that interrupt most insulin genes. The simplicity of the hybridization patterns compared with those observed on digestion of rat or mouse DNA[9,10,16] suggests that there is only a single insulin gene in HIT-T15 cells and presumably in the Syrian hamster. However, its cloning will be necessary to unambiguously establish this.

HIT cells provide a model for studying the regulation of insulin biosynthesis. The cloning and characterization of Syrian hamster preproinsulin mRNA and protein provide valuable structural information about the mRNA and protein. In addition, the cDNA clone can be used as a probe in future studies. Since there is probably only a single insulin gene in the Syrian hamster, the regulation of insulin biosynthesis in HIT cells may be similar to that in human beta cells, in contrast to isolated islets or cultured beta cells obtained from rats.

ACKNOWLEDGMENTS

We thank Noel Fong and Margarita Quiroga for their technical assistance, and Dr. Leslie Rall and Maureen Appling for their comments and help in preparing the manuscript.

REFERENCES

[1] Santerre, R. F., Cook, R. A., Crisel, R. M. D., Sharp, J. D., Schmidt, R. J., Williams, D. C., and Wilson, C. P.: Insulin biosynthesis in a cloned cell line of simian virus 40-transformed hamster pancreatic beta cells. Proc. Natl. Acad. Sci. USA 1981; 78:4339–43.

[2] Bell, G. I., Santerre, R. F., and Mullenbach, G. T.: Hamster preproglucagon contains the sequence of glucagon and two related peptides. Nature 1983; 302:716–18.

[3] Maniatis, T., Fritsch, E. F., and Sambrook, J.: Molecular Cloning. Cold Spring Harbor, Cold Spring Harbor Laboratory, 1982.

[4] Bell, G. I., Sanchez-Pescador, R., Laybourn, P. J., and Najarian, R. C.: Exon duplication and divergence in the human preproglucagon gene. Nature 1983; 304:368–71.

[5] Maxam, A. M., and Gilbert, W.: Sequencing end-labeled DNA with base-specific chemical cleavage. *In* Methods in Enzymology. Vol 65. Grossman, L., and Moldave, K., Eds. New York, Academic Press, 1980:499–560.

[6] Sanger, F., Coulson, A. R., Barrell, B. G., Smith, A. J. H., and Roe, B. A.: Cloning in single-stranded bacteriophage as an aid to rapid DNA sequencing. J. Mol. Biol. 1980; 143:161–78.

[7] Bell, G. I., Karam, J. H., and Rutter, W. J.: Polymorphic DNA region adjacent to the 5′ end of the human insulin gene. Proc. Natl. Acad. Sci. USA 1981; 78:5759–63.

[8] Kwok, S. C. M., Chan, S. J., and Steiner, D. F.: Cloning and nucleotide sequence analysis of the dog insulin gene. J. Biol. Chem. 1983; 258:2357–63.

[9] Cordell, B., Bell, G., Tischer, E., DeNoto, F., Ullrich, A., Pictet, R., Rutter, W. J., and Goodman, H. M.: Isolation and characterization of a cloned rat insulin gene. Cell 1979; 18:533–43.

[10] Lomedico, P., Rosenthal, N., Efstratiadis, A., Gilbert, W., Kolodner, R., and Tizard, R.: The structure and evolution of the two nonallelic rat preproinsulin genes. Cell 1979; 18:545–58.

[11] Bell, G. I., Pictet, R. L., Rutter, W. J., Cordell, B., Tischer, E., and Goodman, H. M.: Sequence of the human insulin gene. Nature 1980; 284:26–32.

[12] Wetekam, W., Groneberg, J., Leineweber, M., Wengenmayer, F., and Winnacker, E.: The nucleotide sequence of cDNA coding for preproinsulin from the primate *Macaca fascicularis*. Gene 1982; 19:179–83.

[13] Permutt, M. A.: Insulin biosynthesis IV. Effect of glucose on initiation and elongation rates in isolated rat pancreatic islets. J. Biol. Chem. 1972; 248:2738–42.

[14] Itoh, N., and Okamoto, H.: Translational control of proinsulin biosynthesis by glucose. Nature 1980; 283:100–102.

[15] Jahr, H., Schroder, D., Ziegler, B., Ziegler, M., and Zuhlke, H.: Transcriptional and translational control of glucose-stimulated (pro) insulin biosynthesis. Eur. J. Biochem. 1980; 110:499–505.

[16] Lalley, P. A., and Chirgwin, J.: Mapping of mouse insulin genes. International Human Gene Mapping Workshop VII, Cytogenetics and Cell Genetics 1984. In press.

[17] Perler, F., Efstratiadis, A., Lomedico, P., Gilbert, W., Kolodner, R., and Dodgson, J.: The evolution of genes: the chicken preproinsulin gene. Cell 1980; 20:555–66.

[18] Southern, E. M.: Detection of specific sequences among DNA fragments separated by gel electrophoresis. J. Mol. Biol. 1975; 98:503–17.