Nucleotide Sequence of Human Preproinsulin Complementary DNA

Abstract. *Recombinant bacterial plasmids that contain DNA complementary to human preproinsulin messenger RNA have been constructed. One clone contains the entire preproinsulin coding region, as well as the 3′ untranslated region of the messenger RNA and eight nucleotides of the 5′ untranslated region. Additional sequence information for the 5′ untranslated region was obtained with the use of insulinoma messenger RNA in conjunction with specific primers from the cloned DNA for enzymatic chain termination sequence analysis. The results confirm the amino acid sequence of human proinsulin previously determined, and predict the amino acid sequence of the human preproinsulin signal peptide.*

Active insulin consists of two polypeptide chains, the A and B chains, which are linked by disulfide bonds. The protein is synthesized as a single precursor polypeptide, in which the A and B chains are joined to each other via a connecting peptide, the C peptide (*1*). This proinsulin precursor is then processed by proteolytic cleavage to yield the mature insulin protein.

Proinsulin is contained within an even larger polypeptide (preproinsulin) which was first identified by in vitro translation of messenger RNA (mRNA) derived from pancreatic B cells (*2, 3*). The higher molecular weight of preproinsulin is consistent with the presence of an amino terminal signal sequence (*4*) involved in the transfer of newly synthesized secretory proteins into the lumen of the endoplasmic reticulum (*4*). The primary structures of insulin prepeptides for several species have previously been determined by microsequencing techniques with radioactively labeled in vitro translation products (*2*), and by DNA sequence analysis in conjunction with molecular cloning techniques (*5–7*).

We now report the nucleotide sequence of human preproinsulin mRNA. The nucleotide sequence confirms the amino acid sequence for human proinsulin (*8*). In addition, the nucleotide sequence permits determination of the amino acid sequence of the preproinsulin signal peptide. We also compare the human preproinsulin coding and untranslated sequences of the mRNA with the corresponding mRNA sequences of rat preproinsulin I and II.

To isolate insulin-specific nucleotide sequences, we prepared complementary DNA (cDNA) from total human insulinoma mRNA. Total mRNA was isolated from 100 mg of human insulinoma tissue (*9*) by the guanidinium thiocyanate procedure (*5*), and double-stranded cDNA was prepared (*5, 10*) and cloned into the Pst I site of pBR322 (*6*).

Escherichia coli χ1776 was transformed with the recombinant plasmids and transformants were selected for their resistance to tetracycline. Approximately 4000 transformants were obtained from 10 ng of cDNA; of these, 50 were screened for insulin sequences by means of a radioactively labeled rat insulin double-stranded cDNA probe (*5*) in conjunction with the Grunstein-Hogness colony screening procedure (*11*). Twelve clones which hybridized to the probe were further characterized by restriction endonuclease cleavage analysis. The clone containing the largest insert (pHI3) was sequenced by both chemical and enzymatic methods (*12, 13*) (Fig. 1). The nucleotide sequence of pHI3 contained eight nucleotides from the 5′ untranslated region, the entire preproinsulin coding region, and the entire 3′ end of the untranslated region, including a polyadenylate [poly(A)] tract of 40 nucleotides. The sequence allowed us to predict the as yet undetermined signal peptide amino acid sequence for human preproinsulin (Fig. 1).

To obtain more complete sequence information from the 5′ untranslated region of preproinsulin mRNA, we used a technique that adapts the chain termination DNA sequencing method (*13*) to the sequencing of RNA (*14*). Synthesis of DNA complementary to insulinoma mRNA was primed with a 39-base-pair DNA fragment derived from the 5′ terminal end of the preproinsulin coding region, generated by the restriction endonuclease Hae III. The additional sequence information obtained by this procedure is shown in Fig. 1.

The nucleotide sequence of human preproinsulin mRNA shows striking similarities to the mRNA sequences of rat preproinsulin I and II. Amino acid sequence analyses have shown that insulin A and B chains are highly conserved, whereas the C peptide is variable, in respect to both its amino acid composition and length. The degree of homology between the rat I, II, and human amino acid sequences is matched by the overall nucleotide sequence homology (*15*) (Fig. 1). Close examination of the nucleotide sequence, however, reveals regions of the prepeptide B chain and A chain, which are absolutely conserved. Some of these regions have been identified to be important for insulin activity. For example, amino acid residues 24 to 28 of the B chain are responsible for the negative cooperativity observed when insulin binds to its cell surface receptor (*16*). The nucleotide sequence in this region is completely conserved even though the genetic code would tolerate nucleotide changes in the third position of translation codons without altering the specified amino acid. The A chain, which contains most of the amino acids implicated in receptor binding (*16*), displays the highest degree of both nucleotide and amino acid sequence homology (*15*).

The insulin prepeptide is thought to serve as the signal for transfer of this protein into the endoplasmic reticulum (*4*); however, nothing is known about which specific amino acids are essential for possible membrane recognition or traversal (or both). One might expect that amino acid residues proximal to the prepeptide cleavage site would bear some common feature important for prepeptide cleavage enzyme activity, and thus the amino acid differences in this region (glutamic acid compared to glycine, aspartic acid compared to arginine, and alanine compared to glutamine at positions −7, −5, and −2, respectively) are rather surprising. In contrast, the amino terminal region of the prepeptide is much more highly conserved; hence, the amino acid sequence in this region must be important for insulin signal sequence activity. The C peptide helps the proinsulin molecule to assume its correct three-dimensional conformation. The greater amount of amino acid sequence variability in the C peptide does not interfere with its ability to function in this manner. The nucleotide sequence of the C peptide is as variable as the amino acid sequence; first and second position codon changes are more prevalent than in the A and B chains. The function of the 3′ untranslated region is not known. The human preproinsulin mRNA extends 73 nucleotides beyond the translation termination codon, whereas both rat I and rat II mRNA 3′ untranslated regions are about 20 nucleotides shorter in length [52 and 53 nucleotides, respectively (*5, 7*)]. The nucleotides following the termination codons bear little resemblance to one another; however, the sequences proximal to the site of mRNA polyadenylation display a high degree of homology. In this region all three mRNA's

0036/8075/80/0404-0057$00.50/0

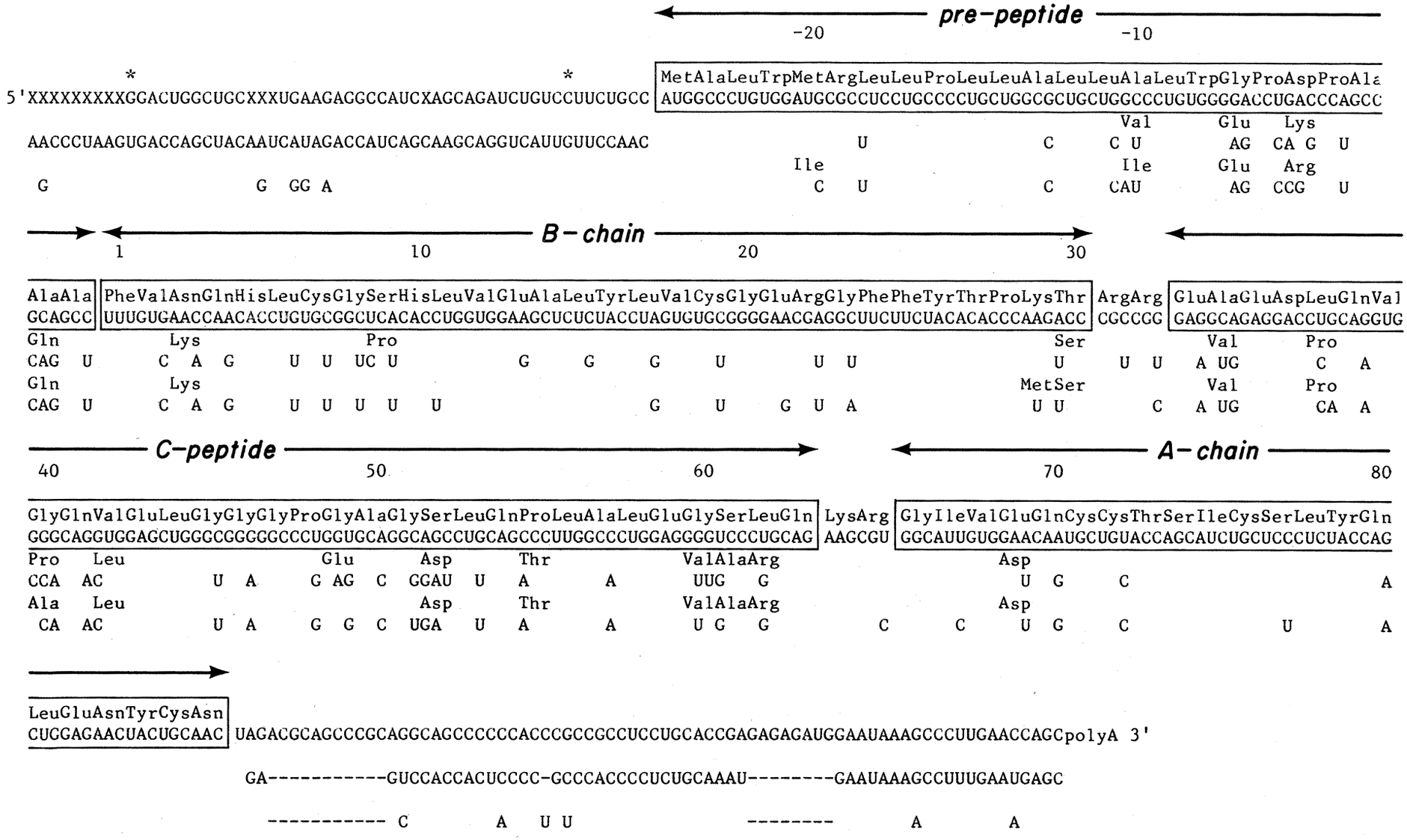

Fig. 1. Nucleotide sequence of human preproinsulin mRNA and comparison with corresponding rat I and rat II sequences. Only coding region differences in rat I and rat II mRNA sequences are presented. In noncoding regions the complete rat I and human sequences are compared; differences in rat II as compared with rat I sequences are also shown. The character "X" denotes uncertainties in the sequence; dashed lines indicate gaps which, when introduced into the sequence, increase apparent homologies between the rat and human mRNA sequences. These gaps may reflect deletions or insertions (or both) that may have occurred during the evolution of preproinsulin. Nucleotides at positions 8 and 48 in the 5′ untranslated region are marked by asterisks.

contain the sequence: AAUAAA (A, adenylate; U, uridylate), a feature common to most eukaryotic mRNA's. The 5′ untranslated region is thought to be important for the translation of mRNA molecules. Structural features in this region may be important for ribosome recognition, binding, and translation efficiency. The rat and human preproinsulin mRNA 5′ untranslated regions contain a significant level of homology, when occasional small insertions or deletions (or both) are considered. The predicted secondary structure of the rat mRNA 5′ untranslated region includes a stable hairpin structure that exposes the sequence CCAUCUAGGA (C, cytidylate; G, guanylate) for potential base-pairing with a complementary sequence present in 18*S* ribosomal RNA (*6*, *17*). The corresponding human sequence differs by only one nucleotide and is also contained within a strong potential hairpin structure between the nucleotides at positions 8 and 48 (denoted by asterisks in Fig. 1).

This cloned human preproinsulin cDNA will facilitate the isolation of the human insulin gene, and the information gained from the complete mRNA sequence will permit the distinction between mRNA sequences, intervening sequences, and nucleotide sequences that flank the insulin gene (*18*).

IRMI SURES*
Stanford University, Veterans Administration Medical Center, Palo Alto, California 94304
DAVID V. GOEDDEL
ALANE GRAY
AXEL ULLRICH†
Genentech, Inc., 460 Point San Bruno Boulevard, South San Francisco, California 94080

References and Notes

1. D. F. Steiner and P. E. Oyer, *Proc. Natl. Acad. Sci. U.S.A.* **57**, 473 (1967); R. E. Chance, R. M. Ellis, W. W. Bromer, *Science* **161**, 165 (1968).
2. S. J. Chan, P. Keim, D. F. Steiner, *Proc. Natl. Acad. Sci. U.S.A.* **73**, 1964 (1976); P. T. Lomedico, S. J. Chan, D. F. Steiner, J. F. Saunders, *J. Biol. Chem.* **252**, 7971 (1977); M. A. Permutt and A. Routman, *Biochem. Biophys. Res. Commun.* **78**, 855 (1977); D. Shields and G. Blobel, *Proc. Natl. Acad. Sci. U.S.A.* **74**, 2059 (1977).
3. C. Yip, C. L. Hew, H. Hsu, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 477 (1975); P. T. Lomedico and G. F. Saunders, *Nucleic Acids Res.* **3**, 381 (1976).
4. C. Milstein, C. G. Brownlee, T. M. Harrison, M. B. Mathews, *Nature (London) New Biol.* **239**, 117 (1972); G. Blobel and B. Dobberstein, *J. Cell Biol.* **67**, 835 (1975).
5. A. Ullrich, J. Shine, J. Chirgwin, R. Pictet, E. Tischer, W. J. Rutter, H. M. Goodman, *Science* **196**, 1313 (1977).
6. L. Villa-Komaroff *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **75**, 3727 (1978).
7. P. Lomedico, N. Rosenthal, A. Efstratiadis, W. Gilbert, R. Kolodner, R. Tizard, *Cell* **18**, 545 (1979).
8. P. E. Oyer, S. Cho, J. D. Peterson, D. F. Steiner, *J. Biol. Chem.* **246**, 1375 (1971); A. S. C. Ko, D. G. Smyth, J. Markussen, F. Sundby, *Eur. J. Biochem.* **20**, 190 (1971).
9. Human insulinoma tissue (B cell adenoma that produces insulin) was provided by Dr. John Baxter and Dr. Wolfgang Kemmler.
10. M. P. Wickens, G. N. Buell, R. T. Schimke, *J. Biol. Chem.* **253**, 2483 (1978); D. V. Goeddel *et al.*, *Nature (London)* **281**, 544 (1979).
11. M. Grunstein and D. S. Hogness, *Proc. Natl. Acad. Sci. U.S.A.* **72**, 3961 (1975).
12. A. Maxam and W. Gilbert, *ibid.* **74**, 560 (1977).
13. F. Sanger, S. Nicklen, A. R. Coulson, *ibid.*, p. 5463.
14. S. Levy, I. Sures, L. H. Kedes, *Nature (London)* **279**, 737 (1979).
15. Sequence homology with human preproinsulin for rat I and rat II amino acid sequences, respectively, were: prepeptide, 79 and 75 percent; B chain, 90 and 90 percent; C peptide, 79 and 74 percent; A chain 95 and 95 percent. Nucleotide sequences: prepeptide, 81 and 78 percent; B chain, 83 and 83 percent; C peptide, 72 and 75 percent; A chain, 94 and 90 percent.
16. P. DeMeyts, E. Van Obberghen, J. Roth, A. Wollmer, D. Brandenburg, *Nature (London)* **273**, 504 (1978).
17. O. Hagenbuchle, M. Santer, J. Steitz, R. J. Mans, *Cell* **13**, 551 (1978).
18. *Note added in proof*: While this report was being reviewed, part of the human preproinsulin mRNA sequence was reported elsewhere [G. Bell *et al.*, *Nature (London)* **282**, 525 (1979)].
19. We thank D. G. Kleid, P. H. Seeburg, J. Brosius, H. Hendrickson, S. Levy, L. H. Kedes, and R. Swanson for support and discussions. We thank Dr. S. Schmus, without whom this work could not have been completed. Supported by a contract from Eli Lilly & Co. and carried out under P2 EK2 conditions in compliance with the NIH guidelines for recombinant DNA research.
* Present address: Département de Biologie Animale, Université de Genève, Genève, Switzerland.
† Reprint requests should be sent to A.U.

8 November 1979; revised 5 February 1980

Role of the Spleen in the Growth of a Murine B Cell Leukemia

Abstract. *A spontaneous B cell leukemia (BCL_1) grew progressively in normal BALB/c mice after injection of tumor cells but did not grow in splenectomized recipients. Despite the absence of progressive tumor growth, residual tumor cells with malignant potential were found in the peripheral blood of the splenectomized animals. Splenectomy performed after injection of tumor cells but before the development of marked leukocytosis also prevented progressive tumor growth and death of the host. Thus the spleen appears to be necessary for progressive proliferation of this lymphocytic leukemia early after passage in vivo.*

We recently described a spontaneous B cell leukemia (BCL_1) of BALB/c mice that expresses B cell surface markers including immunoglobulins M and D, the receptors of the Fc fragment, and I-region antigens (*1*, *2*). The tumor cell line appears unique in its response to stimulation in vitro with the B cell mitogen lipopolysaccharide, since this mitogen stimulates proliferation of and immunoglobulin secretion by the malignant cells (*3*). The tumor cells grow progessively when transferred to normal BALB/c mice and result in their death. Splenomegaly is a consistent manifestation of the disease, and other organs including bone marrow and peripheral blood appear to be involved secondarily (*2*, *4*). Warnke *et al.* (*4*) used autoradiography to demonstrate that there is a pronounced localization of the tumor cells in the spleen early in the course of the disease (*4*). A kinetic study showed that the tumor is manifest in the spleen before large numbers of malignant cells appear in the peripheral blood (*2*). The present study supports and extends these findings by showing that the spleen is required for the progressive growth of the BCL_1 leukemia.

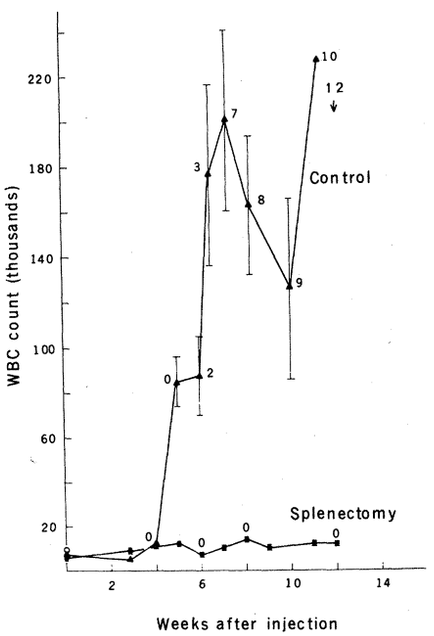

Fig. 1. Growth of BCL_1 tumor cells in splenectomized (N = 24) and normal (N = 12) 4-month-old BALB/c mice. The data are expressed as mean WBC counts ± standard errors. Numbers adjacent to symbols indicate the number of mice dead at that time.

We investigated the role of the spleen in the proliferation of the BCL_1 cells by injecting them into splenectomized mice. The tumor cell line is maintained in our laboratory by passage in unirradiated 2- to 4-month-old BALB/c mice (*5*). To obtain tumor cells for injection, blood was withdrawn from the retro-orbital veins and then diluted in phosphate-buffered saline. The nucleated cells were counted in a hemocytometer. Recipients were splenectomized through an incision in the left upper quadrant under pentobarbital anesthesia. Splenic vessels were tied off with silk, and the incision was closed with metal clips. Twenty-four 4-month-old female BALB/c mice were

0036/8075/80/0404-0059$00.50/0