Review

Making, Cloning, and the Expression of Human Insulin Genes in Bacteria: The Path to Humulin

Arthur D. Riggs

1Diabetes and Metabolism Research Institute, City of Hope National Medical Center, Duarte, California 91010, USA

ORCID number: 0000-0002-8184-5288 (A. D. Riggs).

Received: 11 July 2020; Editorial Decision: 4 November 2020; First Published Online: 19 December 2020; Corrected and Typeset: 7 January 2021.

Abstract

In the mid- to late 1970s, recombinant deoxyribonucleic acid methods for cloning and expressing genes in *E. coli* were under intense development. The important question had become: Can humans design and chemically synthesize novel genes that function in bacteria? This question was answered in 1978 and in 1979 with the successful expression in *E. coli* of 2 mammalian hormones, first somatostatin and then human insulin. The successful production of human insulin in bacteria provided, for the first time, a practical, scalable source of human insulin and resulted in the approval, in 1982, of human insulin for the treatment of diabetics. In this short review, I give my personal view of how the making, cloning, and expressing of human insulin genes was accomplished by a team of scientists led by Keiichi Itakura, Herbert W. Boyer, and myself.

Graphical Abstract

Key Words: recombinant DNA, biotechnology, genentech, chemical DNA synthesis
ESSENTIAL POINTS
• The desire to understand bacterial and human gene regulation and to improve synthetic deoxyribonucleic acid (DNA) chemistry stimulated the efforts to engineer bacteria to produce human proteins.
• This was first accomplished in \textit{E. coli}, which expressed a synthetic gene for somatostatin and was reported in 1977.
• A similar approach was employed in \textit{E. coli} carrying synthetic genes for the human insulin A and B chains.
• Support from industry and academic collaborators resulted in Food and Drug Administration approval of Humulin in 1982, the first human insulin made by recombinant DNA technology.

Until 1982, when Humulin was approved by the US Federal Drug Administration (FDA), human insulin was not available for the treatment of diabetics. Instead, cow and pig insulin were used. The successful production of human insulin from synthetic genes was first reported in January of 1979 (1). Though the initial yields were low, subsequent work done, first at the start-up company Genentech and then at Eli Lilly, increased yields and led to the commercial production of human insulin by Eli Lilly. In October of 1982 Humulin became the first protein therapeutic product based on recombinant deoxyribonucleic acid (DNA) technology to be approved for use in humans by the FDA. Diabetics could now be treated with human insulin. In addition, success with insulin jump-started the biotechnology industry, which currently provides hundreds of previously unavailable therapeutic agents.

It is commonly assumed that human insulin was first made from the human insulin gene cloned and introduced into \textit{E. coli}, but this was not the case. Neither the messenger RNA (mRNA) sequence nor the gene sequence of human insulin was known at the time. All that was known was the protein sequence and structure of insulin. For the work leading to Humulin, the genes for the A and B chains of insulin were designed using the sequence of amino acids for the insulin A and B peptide chains and then the genetic code, with a selection of codons preferred by \textit{E. coli}. These genes were then chemically synthesized using recently developed organic chemical synthesis methods (2). Thus, the genes for the A and B chains of human insulin were completely human-designed and human-made genes. It was remarkable that these genes functioned in \textit{E. coli}.

Human insulin, which is comprised of a total of 51 amino acids (21 aa A chain and 30 aa B chain), was the first therapeutically useful protein product resulting from recombinant DNA technology. However, the feasibility of the methods had been established just a year earlier, in December 1977, with the production in \textit{E. coli} of somatostatin, a 14 amino acid mammalian hormone (3). The methods used for insulin were essentially the same as for somatostatin (Fig. 1), but with the additional necessity of joining the 2 peptide chains, which were each made and purified separately (Fig. 2).

How the Insulin Project Started
It may be of interest that the initial goals for the principal investigators of this breakthrough in diabetes therapy were not directed towards improving the treatment of diabetes but were directed towards improving chemical DNA synthesis (Keiichi Itakura), improving our understanding of bacterial and mammalian gene regulation (Arthur D. Riggs), and improving restriction enzyme-based recombinant DNA technology (Herbert W. Boyer). The work that led to the insulin project began as a project to crystalize the lac repressor, a protein that controls the transcription of beta-galactosidase mRNA in \textit{E. coli}. The lac repressor was the first repressive transcription factor protein to be purified (4), and I was interested in the details of its sequence-specific interaction with its DNA target, the lac operator, which was known to be a small sequence of DNA, only 21 bp long. To better understand sequence-specific protein DNA interactions, Richard Dickerson, Dickerson’s postdoctoral fellow, John Rosenberg, and I decided to start a project to crystallize the lac repressor bound to the lac operator. To make the lac operator by chemical DNA synthesis, Itakura was recruited to our team and he became a faculty member with me at the City of Hope (COH), a medical research center located near the California Institute of Technology (Caltech), at which Richard Dickerson was a professor and expert in protein crystallography. As a postdoctoral fellow in S. A. Narang’s laboratory in Ottawa, Itakura had worked on improving the phosphotriester method for chemical DNA synthesis (5). After moving to the COH (and Caltech in Dickerson’s lab while waiting for his own labs to be built at the COH), Itakura began making the lac operator as a double-stranded 21 bp DNA fragment. We were never successful in crystallizing the lac repressor bound to the lac operator, although Itakura and Dickerson were successful in crystallizing other DNA molecules made by Itakura (6). The spin-off projects of this failed repressor-operator crystallization project did, however, prove to be quite successful. Figure 3 shows some of the significant advances made in the resulting somatostatin/insulin projects.
After Itakura worked for about a year to make the lac operator DNA using his phosphotriester method, it was found that the final product was not adequately pure as measured by binding to the lac repressor, so we needed a way to further enrich for the correct, functional sequence. At this time, it was not possible to purify DNA according to a base sequence, but, driven by necessity, we realized that cloning, using methods very recently developed by Herbert Boyer (7, 8), and then screening for the correct sequence would function to purify and then produce the desired DNA fragment. This led to a collaboration with Herbert W. Boyer and Herbert Heyneker, a postdoctoral fellow in Boyer's laboratory. Heyneker was quickly successful in cloning the chemically synthesized lac operator and showing, for the first time, that cloned, chemically synthesized DNA could function in vivo as a ligand for its protein target (9).

**Somatostatin Project**

Now knowing that biologically functional DNA could be chemically synthesized and cloned, these results encouraged us to take an important next step—designing a gene that would cause *E. coli* to make a protein product, preferably a medically useful protein product. Insulin was known to be a small protein, totaling only 51 amino acids, but it was composed of 2 polypeptide chains, the A and B chains, and assays for the A chain, including immunoassays, were not very sensitive. I thought we needed a smaller, simpler molecule to demonstrate the feasibility of our approach. Fortunately, a publication on somatostatin (10) came to our attention at the time Itakura and I were thinking about designing a gene for a small, biologically active peptide—probably a hormone. We noted that somatostatin was only 14 amino acids and had a very sensitive radioimmunoassay (11, 12), so sensitive that I calculated that we could detect even 1 somatostatin molecule per bacterial cell. Itakura and I then wrote and submitted to the National Institutes of Health (NIH) (in early 1976) a grant application in which we proposed to chemically synthesize the gene for somatostatin, clone it in *E. coli*, and assay for the production of the somatostatin polypeptide. In this grant, we also stated that if using somatostatin was successful, we would use similar technology to produce insulin. The grant was reviewed moderately well but not funded. The summary statement of the rejection noted: “In conclusion, the goals reflect
extremely complex and time-consuming projects which may not be reasonably accomplished in three years . . .

the only possible outcome of this work would be to confirm that these manipulations can lead to the synthesis of a human peptide in E. coli. Because of the poor choice of the biological system, this appears as an academic exercise.”

In hindsight, our project on somatostatin was indeed an academic exercise, but a novel one that provided strong patents and quickly led to a flourishing new industry.

Fortunately, Herbert Boyer was independently thinking about insulin and contacted Itakura and me in early 1976, saying he had a “businessman friend” who thought he could raise money to start a company based on recombinant DNA technology, with the first product being human insulin. The businessman friend was Robert Swanson, who with Herbert Boyer did start Genentech and did fund our work on somatostatin and then insulin. Boyer had the difficult but ultimately successful task of convincing Swanson to first fund somatostatin, which we all thought would not likely be a commercially important product. Genentech was incorporated in April of 1976, and once funding was in hand from Genentech, Itakura, I, and Boyer began designing the somatostatin gene and its synthesis and cloning, as well as the assay and purification of somatostatin from bacteria. The responsibilities of each laboratory are shown in Fig. 4.

When we began work on somatostatin, the safety of recombinant DNA was being actively discussed (the famous Asilomar Conference was in February of 1975), and safety regulations were being put in place. The main concern was whether it would be safe to produce an active hormone in E. coli, which is an enterobacterium found in the human gut. These concerns proved to be groundless, but at the time we gave them great weight, and for this reason we deliberately designed the hormone product to be inactive until chemically converted to an active hormone after purification. Since a free amino-terminus of somatostatin is required for activity, we designed the chemically synthesized gene to produce in vivo a fused precursor product where somatostatin would be made as a tail on a precursor peptide and thus be inactive. Somatostatin does not contain a methionine, so we planned to use cyanogen bromide to treat crude extracts to cleave off the tail, thereby producing active somatostatin. We were fortunate that this was our strategy, because if we had tried to make somatostatin as a 14 aa peptide, we now know that it would have been destroyed by enzymes in E. coli that are very active on small polypeptides. In fact, our first attempt at producing somatostatin did fail completely, probably for this reason. We had 2 plans for somatostatin, Plan A and Plan B, differing only by where the gene was inserted into the beta-galactosidase gene, which depended on the
presence of EcoR1 sites (GAATTC). For plan A, the chemically synthesized gene was to be inserted very near to the N-terminus of beta-galactosidase. We tried Plan A first and were not able to detect somatostatin, even with a sensitivity of less than 1 molecule per cell. Fortunately, plan B was to use another EcoR1 site near the end of beta-galactosidase, which is a large protein. Plan B was successful, probably because the large beta-galactosidase protein protected the somatostatin tail prior to being released by the cyanogen bromide treatment of cell extracts. The scheme for somatostatin is shown in Fig. 1. Most of the scientists involved in the somatostatin project are shown in Fig. 5.

Insulin Project

Neither of the insulin chains contain a methionine, so given the success with somatostatin, we were confident that we could make and purify—the A and B chains of insulin. Itakura’s team, with Roberto Crea as key member, immediately began working on the chemical synthesis of these 2 genes. Our biggest concern was whether we would be able to efficiently join the A and B chains to make functional insulin. We decided to use a procedure that had been published by Katsoyannis (13), which involves converting cysteines to S-sulfonates, mixing the chains together with excess of A chain over B chain, and forming disulfide bonds by air oxidation (Fig. 2). It may have been fortunate that none of the members of our team were experts in insulin biochemistry, as we did not know that some in the field had difficulty in getting the method to work efficiently. We had good success, however, getting joining yields of up to 20% in preliminary experiments and then good activity of the final insulin product. To help with the cloning, purification, and joining of insulin chains, we were extremely fortunate that David Goeddel and Dennis Kleid, two of the first employees of Genentech, joined the team. With their help, cloning and expression was accomplished quickly and we were able to obtain functional insulin and prepare a paper for submission by October of 1978, with publication in January of 1979 (1). Some of the scientists involved in the insulin project are shown in Fig. 6. After publication of this paper reporting our success, my efforts were then devoted to other projects, including recombinant antibodies (14), but much additional work was needed to bring human insulin to the market.

Genentech and Eli Lilly

The first yields obtained for insulin were enough to establish that the method had promise, but they were much too low for commercial production. In 1978, Genentech leased space in an industrial park in south San Francisco and constructed laboratories for recombinant DNA research and development work for the production and purification of bacterial products. When these laboratories were completed, David Goeddel and Dennis Kleid began working...
there and they were soon joined by Herbert Heyneger, Dan Yansura, Ron Wetzel, and others, including Roberto Crea, who moved from Itakura’s lab at COH to set up a DNA synthesis facility at Genentech. This team focused on improving the yield, in part, by devising improved recombinant bacterial promoters and shortening the precursor protein to which the insulin peptide chains were attached. An agreement was reached between Genentech and Eli Lilly specifying that if Genentech could meet yield milestones, then the technology would be transferred to Eli Lilly for large scale production. The yield milestones were met, and Eli Lilly then proceeded to build large new facilities for the bacterial production of insulin. Eli Lilly conducted preclinical and clinical trials to establish the safety and efficacy of bacterially produced insulin. These results were published in 1981 (15) and were sufficient to obtain FDA approval in 1982. I have little knowledge of this tremendous effort, but it is remarkable that only 4 years after the first demonstration that bacteria could be used to produce human insulin, diabetic patients could now be treated with Humulin, the brand name for human insulin produced by Eli Lilly.

The Race to Produce Human Insulin

At the time, we were trying to establish that bacteria could be used to make human proteins directed by novel, chemically synthesized genes, while others were trying to make human insulin starting with mRNA, converting it to complementary DNA (cDNA), and then cloning the proinsulin gene, which would have the C-peptide joining the A and B chains and not be fully active. However, the plan was to then convert the proinsulin to insulin using proteases. There was strong scientific competition, and we all knew that human insulin had commercial importance. Stephen Hall has written a book titled Invisible Frontiers: The Race to Synthesize a Human Gene (16). I highly recommend this book, which describes in lay terms much of the work that was being done to provide a source of human insulin. At the time we were actively working on our approach, we knew that at least 2 other groups were in the “race.” One group, led by William Rutter, was working in the same building as Herbert Boyer at the University of California at San Francisco. Another group in the race was led by Walter Gilbert at Harvard. The methods for converting mRNA to cDNA were just being developed in 1976 (17). And, it was very difficult to obtain intact insulin mRNA because insulin is only produced in the pancreas, which is also the main source of ribonuclease. The ribonuclease problem was solved by the use of guanidine thiocyanate but was not published about until 1979 (18). It was not practical to work with human pancreases, so the Rutter group developed their methods using rat pancreases, and the Gilbert group used rat insulinoma cells. Both groups then planned to use these methods to clone and produce human proinsulin. Both groups were slowed down by the need to do work with mammalian genes in special P3 containment facilities, and this is well described in Hall’s book. The Rutter group, with Axel Ullrich as the key expert in recombinant DNA, was successful in cloning rat proinsulin cDNA into a plasmid by using reverse transcriptase to convert insulin mRNA into cDNA (19). The Gilbert group was also successful in cloning rat insulin cDNA and getting secretions of preproinsulin (20), but the yields were low. Neither of these groups followed through to produce commercial levels of human proinsulin. Genentech was eventually successful in cloning the human preproinsulin gene (21), and after a few years production was shifted from the separate production of A and B chains to the use of proinsulin, although the details of this, to my knowledge, remain unpublished.

The COH/Genentech team was first to be successful for several reasons. One reason was that NIH guidelines for recombinant DNA made it necessary to work in special containment facilities when cloning mammalian hormone genes. Problems caused by this requirement are well described in Hall’s book (15). Another is that insulin is made as a larger protein, with A and B chains connected by a connecting polypeptide (C-peptide). This C peptide must be removed by proteases during maturation to obtain active insulin. The successful expression and secretion of cloned human proinsulin in mammalian cells was achieved in 1983 (22), but proinsulin is not biologically active. Thus, our approach was actually simpler to perform given the technology available in the late
1970s. Much of the technology used by us and others to obtain human insulin was still in its early stages. It is interesting to think about how much progress has been made since then. For example, in 2020, the genes for insulin can be made in a few hours by automated instruments and then cloned and expressed by a single person in about a week.

Acknowledgments

Financial Support: None.

Additional Information

Correspondence and Reprint Requests: Dr Arthur D. Riggs, Diabetes & Metabolism Research Institute, Gonda North, City of Hope National Medical Center, 1500 East Duarte Road, Duarte, CA 91010, USA. E-mail: ariggs@coh.org.

Disclosure Summary: No disclosures

Data Availability: Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

References

1. Goeddel DV, Kleid DG, Bolivar F, et al. Expression in Escherichia coli of chemically synthesized genes for human insulin. Proc Natl Acad Sci U S A. 1979;76(1):106–110.
2. Crea R, Kraszewski A, Hirose T, Itakura K. Chemical synthesis of genes for human insulin. Proc Natl Acad Sci U S A. 1978;75(12):5765–5769.
3. Itakura K, Hirose T, Crea R, et al. Expression in Escherichia coli of a chemically synthesized gene for the hormone somatostatin. Science. 1977;198(4321):1056–1063.
4. Riggs AD, Bourgeois S. On the assay, isolation and characterization of the lac repressor. J Mol Biol. 1968;34(2):361–364.
5. Itakura K, Katagiri N, Narang SA, Bahl CP, Mariani KJ, Wu R. Chemical synthesis and sequence studies of deoxyribonucleotides which constitute the duplex sequence of the lactose operator of Escherichia coli. J Biol Chem. 1975;250(12):4592–4600.
6. Drew HR, Wing RM, Takano T, et al. Structure of a B-DNA decamer: conformation and dynamics. Proc Natl Acad Sci U S A. 1981;78(4):2179–2183.
7. Cohen SN, Chang AC, Boyer HW, Helling RB. Construction of biologically functional bacterial plasmids in vitro. Proc Natl Acad Sci U S A. 1973;70(11):3240–3244.
8. Morrow JE, Cohen SN, Chang AC, Boyer HW, Goodman HM, Helling RB. Replication and transcription of eukaryotic DNA in Escherichia coli. Proc Natl Acad Sci U S A. 1974;71(5):1743–1747.
9. Heyneker HL, Shine J, Goodman HM, et al. Synthetic lac operator DNA is functional in vivo. Nature. 1976;263(5580):748–752.
10. Schally AV, Dupont A, Arimura A, et al. Isolation and structure of somatostatin from porcine hypothalamus. Biochemistry. 1976;15(3):509–514.
11. Arimura A, Sato H, Dupont A, Nishi N, Schally AV. Somatostatin: abundance of immunoreactive hormone in rat stomach and pancreas. Science. 1975;189(4207):1007–1009.
12. Arimura A, Sato H, Coy DH, Schally AV. Radioimmunoassay for GH-releasing inhibiting hormone. Proc Soc Exp Biol Med. 1975;148(3):784–789.
13. Katsoyannis PG, Tometsko A, Zalut C, Johnson S, Trakatellis AC. Studies on the synthesis of insulin from natural and synthetic A and B chains. I. Splitting of insulin and isolation of the S-sulfonated derivatives of the A and B chains. Biochemistry. 1967;6(9):2635–2642.
14. Cabilly S, Riggs AD, Pande H, et al. Generation of antibody activity from immunoglobulin polypeptide chains produced in Escherichia coli. Proc Natl Acad Sci U S A. 1984;81(11):3273–3277.
15. De Meys P, Halban P, Hepp KD. In vitro studies on biosynthetic human insulin: an overview. Diabetes Care. 1981;4(2):144–146.
16. Hall S. Invisible Frontiers: The Race to Synthesize a Human Gene. New York: Atlantic Monthly Press; 1987.
17. Efstratiadis A, Kafatos FC, Maxam AM, Maniatis T. Enzymatic in vitro synthesis of globin genes. Cell. 1976;7(2):279–288.
18. Chirgwin JM, Przybyla AE, MacDonald RJ, Rutter WJ. Isolation of biologically active ribonucleic acid from sources enriched in ribonuclease. Biochemistry. 1979;18(24):5294–5299.
19. Ullrich A, Shine J, Chirgwin J, et al. Rat insulin genes: construction of plasmids containing the coding sequences. Science. 1977;196(4296):1313–1319.
20. Villa-Komaroff L, Efstratiadis A, Broome S, et al. A bacterial clone synthesizing proinsulin. Proc Natl Acad Sci U S A. 1978;75(8):3727–3731.
21. Sures I, Goeddel DV, Gray A, Ullrich A. Nucleotide sequence of human preproinsulin complementary DNA. Science. 1980;208(4439):57–59.
22. Laub O, Rall L, Bell GJ, Rutter WJ. Expression of the human insulin gene in an alternate mammalian cell and in cell extracts. J Biol Chem. 1983;258(10):6037–6042.