Gene Synthesis with H G Khorana

Marvin H Caruthers

"I wish to conclude by hazarding the following rather long range predictions. In the years ahead, genes are going to be synthesized. The next steps would be to learn to manipulate the information content of genes and to learn to insert them into and delete them from the genetic systems. When, in the distant future, all this comes to pass, the temptation to change our biology will be very strong."

H G Khorana, 1968 [1]

Marvin H Caruthers is Distinguished Professor of Chemistry and Biochemistry at the University of Colorado where he teaches and maintains his research program. He is a member of the U.S. National Academy of Sciences and the recipient of the U.S. National Medal of Science (2006). His research interests include the chemical synthesis of the nucleic acids and their analogs as well as the biological and biochemical properties of DNA and RNA.

Gene Synthesis and PCR

I first met Gobind Khorana in October, 1966. I was a graduate student with Robert Letsinger at Northwestern University and had written to Gobind about a postdoctoral position in his laboratory. He asked me to meet him at O'Hare Airport as he was on his way to Cambridge for an extended stay. At our meeting, he presented a plan on how his laboratory had embarked on the chemical and enzymatic synthesis of the gene for yeast alanine transfer RNA and asked me to join the project. During the meeting he told me that they were chemically synthesizing deoxyoligonucleotides (oligomers) twenty deoxynucleotides in length (20mers) and explained how they planned to carry out the total synthesis using these oligomers and DNA polymerase as outlined in *Figure* 1. Two 20mers having a ten base pair overlap would be annealed and repair synthesis used to generate a thirty deoxyoligonucleotide duplex. The deoxyoligonucleotide strands of this duplex would be separated by heating and new chemically synthesized 20mers having ten base pair overlaps at the 3'-ends of these 30mers would be added. Repair synthesis with a new sample of DNA polymerase would then generate a fifty deoxyoligonucleotide duplex. Further repetitions of this cycle with additional 20mers lead to the synthesis of the gene for yeast

Keywords
Chemical synthesis of genes for yeast alanine tRNA and *E. coli* supressor tRNA, Khorana's philosophy on science.

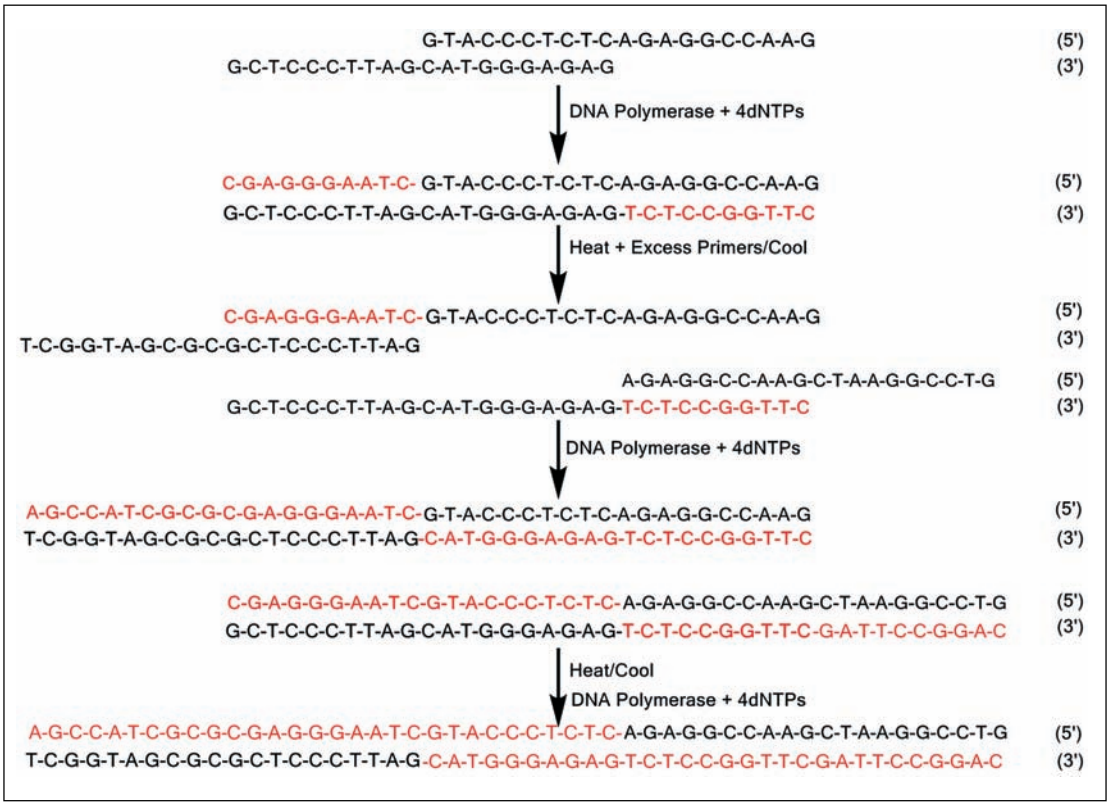

Figure 1. Gene synthesis plan using DNA polymerase.

alanine tRNA. As the final synthetic step, 20mers complementary to the 3'- and 5'-ends would be annealed to the heat-denatured gene and repair synthesis with DNA polymerase carried out. Multiple repetitions of the last step amplify the gene.

I accepted Gobind's offer of a postdoctoral appointment and submitted this plan as a National Institutes of Health Postdoctoral application (which was funded). By the time I joined his group, February 1968, T4 DNA ligase and kinase had been discovered and these enzymes were being explored as an alternative approach, which we eventually used, for enzymatically joining synthetic deoxyoligonucleotides in order to form the yeast alanine tRNA gene.

The enzymatic approach using DNA polymerase was extensively investigated in Gobind's laboratory and published in several manuscripts.

Nevertheless, the enzymatic approach using DNA polymerase, as outlined in *Figure* 1, was extensively investigated in Gobind's laboratory and published in several manuscripts including one

by K Kleppe *et al* [2]. If one examines this approach carefully both theoretically and experimentally, it is the same procedure called 'polymerase chain reaction' (PCR) as developed by Kary Mullis many years later and for which he received the Nobel Prize in 1993. Many of us were extremely disappointed that Gobind did not share this Nobel, which would have been his second.

Many of us were extremely disappointed that Gobind did not share the Nobel for the PCR method, which would have been his second.

Yeast Alanine tRNA Gene Synthesis

Why synthesize genes? Because of the naiveté of biologists and molecular biologists in 1965 when Gobind Khorana initiated synthesis of the yeast alanine tRNA gene, this was a valid question. At that time, we could deduce the sequence of only one gene – the yeast alanine tRNA gene as the corresponding tRNA sequence (the first known tRNA sequence) had just been elucidated by Robert Holley. The synthesis of protein genes was impossible as we did not know how to sequence DNA and the amino acid codons were still being determined by the Nirenberg and Khorana laboratories. As was always typical of Gobind, he chose to begin this project long before others even realized its importance and how the work would eventually lead to the recent advances in the biological sciences and biotechnology.

In 1968 Gobind wrote [3], "We would like to know, for example, what the initiation and termination signals for RNA polymerase are, what kind of sequences are recognized by repressors, by host modification and host restrictive enzymes and by enzymes involved in genetic recombination and so on. For these studies eventually what is required is the ability to synthesize long chains of DNA with specific non-repeating sequences. With this should come the ability to 'manipulate' DNA for different types of studies. We therefore concluded about three years ago that, in continuing our interest in polynucleotide synthesis, the next long-range aim must be the development of methods for the total synthesis of biologically-specific DNA duplexes."

As was always typical of Gobind, he chose to begin the project on synthesis of a tRNA gene long before others even realized its importance.

And so we began our quest to synthesize a gene, but before describing the research, one must still answer an important question. Why choose a tRNA gene – other than the obvious answer

that it was the only one available at the time. Perhaps the most succinct answer can be found from Gobind's work as well [4]: "The general functions of tRNA are clearly established. These molecules have to be recognized by a large number of components of the protein synthesizing machinery such as, by the pyrophosphorylase which repairs the CCA end, by the aminoacyl-t-RNA synthetases, by ribosomes and by messenger RNA. Furthermore, transfer RNAs are a unique class of molecules, possessing attributes of both nucleic acids and proteins. There is a good deal of evidence now to suggest that in addition to a common secondary structure these molecules possess a tertiary structure. Then, all NAs abound in minor bases which are largely found in the loop-out nonhydrogen bonded regions. It is entirely possible that a good part of the evolution of the genetic code is synonymous with the evolution of tRNA molecules. The total area of structure function relationships in these molecules is an open field despite the great current research activity in this field. It is clear that chemical synthesis, provided it could be developed to the point where we can manipulate different parts of the tRNA structures, would open up a definitive approach of wide scope. For example, one could have deletions in different parts, one could take an anti-codon loop from a tRNA specific for one amino acid and replace it with the anti-codon loop for another tRNA." At the time these concepts were proposed by Gobind, the fields of molecular biology and biochemistry were years, if not decades, away from exploiting this approach for understanding the relation of structure to function in biological systems.

At the time these concepts were proposed by Gobind, the fields of molecular biology and biochemistry were years, if not decades, away from exploiting this approach for understanding the relation of structure to function in biological systems.

By the time I joined the Khorana Group, Naba Gupta had determined that we could use T4 DNA ligase and kinase to assemble the yeast alanine tRNA gene. Based upon his results, the research plan outlined in *Figure* 2 was derived. Because T4 DNA ligase could join deoxyoligonucleotides as small as five or six when annealed to a complementary DNA, our synthesis plan was based upon the use of oligomers 10–12 in length. However two oligomers (numbers 6 and 8), which were synthesized very early in the project and at the time when we planned to use repair synthesis,

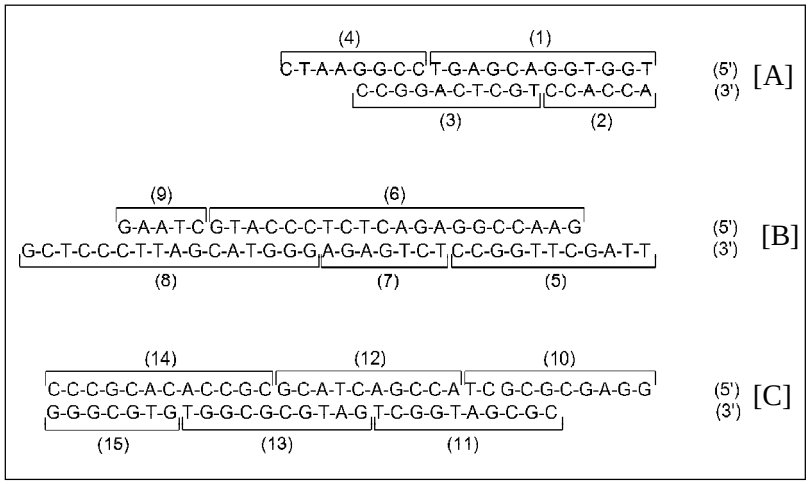

Figure 2. Enzymatic synthesis plan for the yeast alanine tRNA.

were 20 and 16 in length respectively. As can be seen in *Figure* 2, each deoxyoligonucleotide aligned with its complement and also had at least a 4–5 single-stranded or 'sticky-end' which was used to guide the next oligomer into position for ligation to the growing gene duplex. Another advantage relative to using shorter deoxyoligonucleotides was that the synthesis of 10–12mers required far less time and materials than preparation of 20mers.

The deoxyoligonucleotides used to synthesize this gene ranged from five to twenty deoxynucleotides in length. We started with deoxynucleotides as there were no supply houses for ordering protected monomers and certainly not the completed DNA segments as is possible today. The first step was preparing deoxynucleotides in protected form so that various reactive functional groups on the deoxyribose sugar, the purine and pyrimidine bases, and even phosphate were unavailable for generating side-products during synthesis and purification. The protecting groups that were used are shown in *Figure* 3. These were derived from studies carried out in the Khorana Laboratory over a period of ten years (1955–1965) and represent, not the first, but the final, most desirable set relative to the synthetic approach available to us at that time. Our strategy dictated that we build deoxyoligonucleotides from the 5'- to the 3'-end which found us using monomethoxytrityl protection on the 5'-hydroxyl, a group that had to survive many condensation and purification steps.

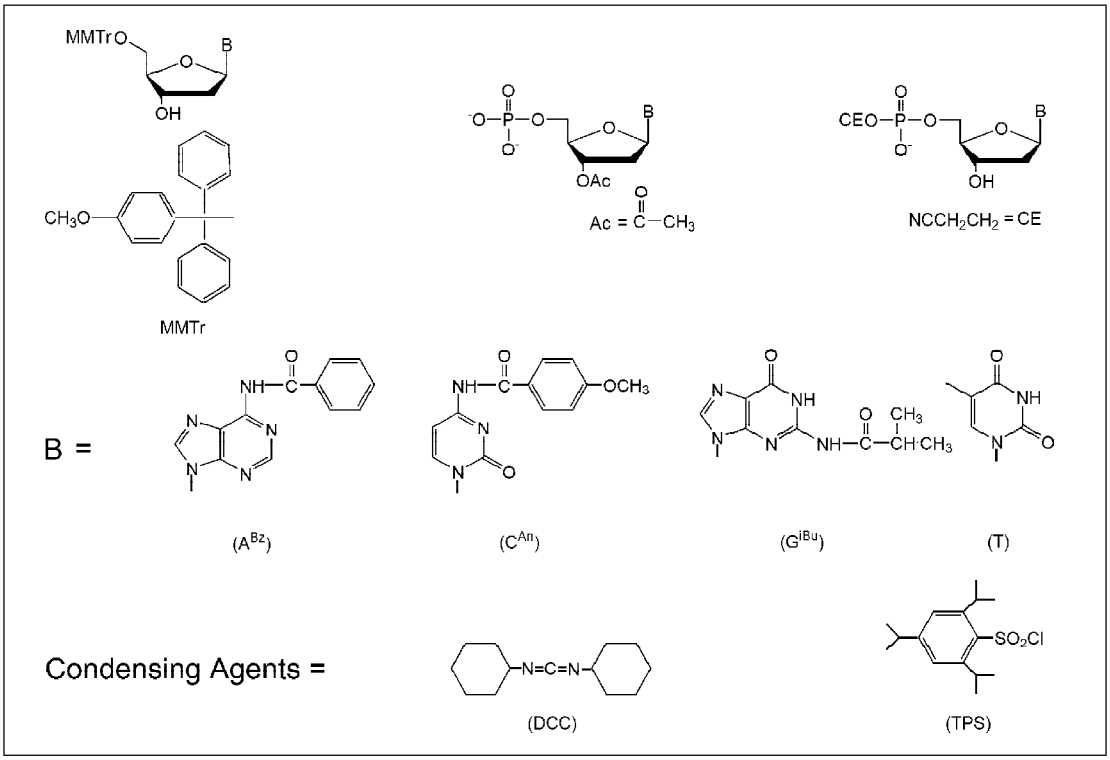

Figure 3. Protecting groups and condensing reagents used for chemically synthesizing DNA.

Similarly, the amino functional groups on cytosine and the purines were protected as amides. Thus cytosine was converted to the N-anisoylderivative, adenine to N-benzoyl, and guanine to N-isobutyroyl. The rationale for this particular set is too extensive for a review here, but can be found in Gobind's early papers. All of our deoxyoligonucleotides were constructed using intermediate mono-, di-, tri-, and tetradeoxyoligonucleotides, which had to be prepared prior to joining with the final, growing 5'-monomethoxytrityl deoxyoligonucleotide. Their preparation required protection of 5'-phosphate with a β-cyanoethyl group and the 3'-sugar hydroxyl with an acetyl ester. These were transient protecting groups that were removed at various steps during preparation of the di-, tri-, and tetradeoxyoligonucleotides.

Once the mono-, di-, tri-, and tetradeoxyoligonucleotides had been prepared, joining of these blocks in the proper order could proceed. These reactions required powerful condensing agents such as dicyclohexylcarbodiimide (DCC) and triisopropylbenzene

sulfonylchloride (TPS). *Figure* 4 contains the outline for synthesis of the 20mer (segment 6, *Figure* 2) as prepared by Hans Weber. The first step was formation of a dinucleotide using an appropriately protected 5'-monomethoxytrityl derivative of deoxyguanosine and a 3'-O-acetyl and N-benzoyl protected deoxyadenosine 5'-phosphate. The condensing reagent was DCC. Upon removal of the 3'-O-acetyl group and purification, this dimer was next reacted with N-benzoyl-3'-O-acetyldeoxyadenosine 5'-phosphate using DCC as the condensing reagent. Removal of the 3'-O-acetyl group and purification generated the 5'-monomethoxytrityl containing trimer. Subsequent condensations with an appropriate dimer and trimer yielded an octamer. Further condensations with three tetramers generated the final product, an icosanucleotide. Starting with 2.1 mmoles of the 5'-monomethoxytrityl derivative of 2'-deoxyguanosine, the yield of icosanucleotide was 1.6 micromole (460 OD_{280} units).

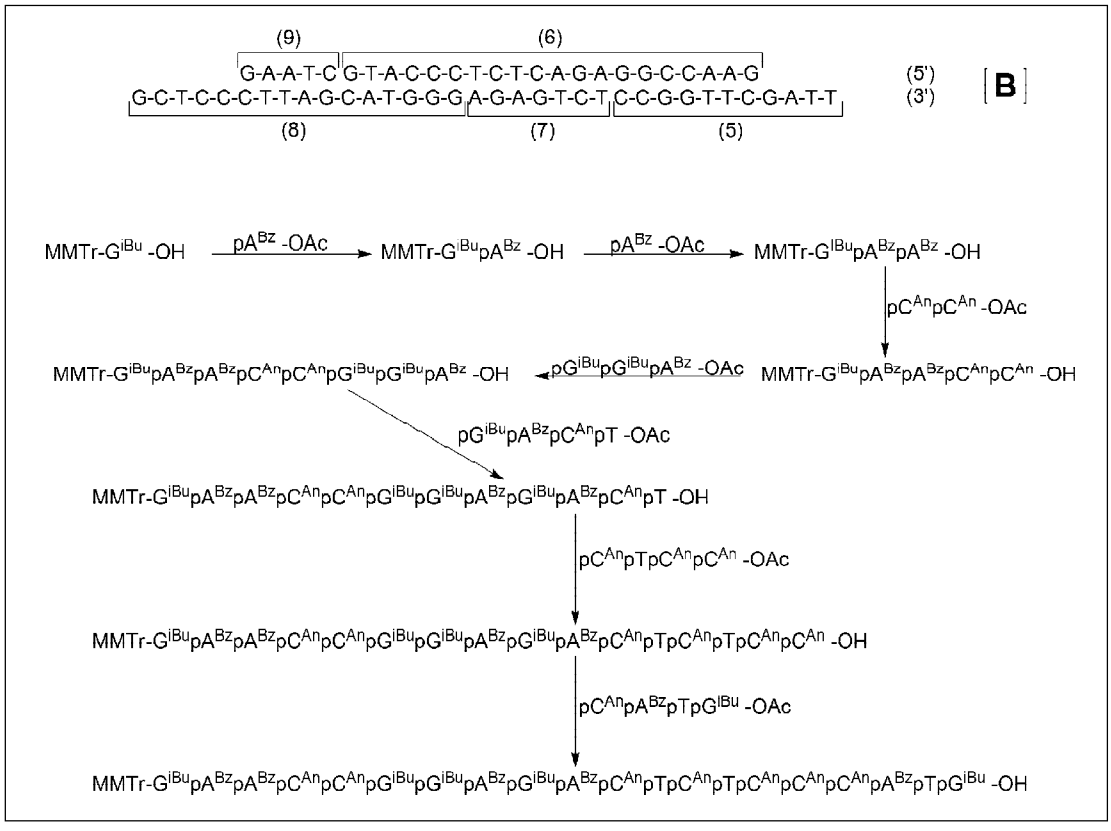

Figure 4. Schematic outline for the chemical synthesis of segment 6.

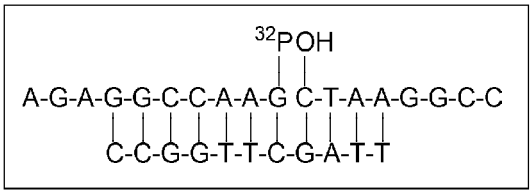

Figure 5. Illustration of a DNA ligase reaction used for synthesizing a tRNA Gene.

Although an extremely low overall yield, the amount of product was far more than needed to synthesize the final gene. Similar strategies were used for the chemical synthesis of all the deoxyoligonucleotides shown in *Figure* 2.

The enzymatic joining of synthetic deoxyoligonucleotides required two steps. The first was converting each oligomer to its 5'-phosphate using ATP and T4 kinase. In order to monitor the phosphorylation and subsequent ligation steps, we used α-^{32}P or α-^{33}P-ATP. The second involved using T4-ligase and ATP to join the 5'-phosphoryl deoxyoligonucleotides. Radiolabels also gave us a powerful method for monitoring ligation reactions by nearest neighbor analysis. In this example (*Figure* 5), which corresponds to nucleotide sequence 13–30 of the yeast alanine tRNA gene, an octamer and a decamer with a 5'-^{32}P-phosphate are joined while forming a partial duplex with a complementary 11mer. The joining reaction was monitored by observing a product having a phosphatase resistant radiolabel, which upon degradation with spleen phosphodiesterase, transfers ^{32}P to deoxycytidine.

Once the basic principles of joining deoxyoligonucleotides using T4 ligase had been determined, the overall approach for gene synthesis was established. We divided the gene into three sections ([A], [B], and [C]; *Figure* 2) and prepared each by enzymatic ligation. Two of these were first joined and the product of this reaction was used in conjunction with the third section to generate the final yeast alanine tRNA gene. The column chromatography profile as shown in *Figure* 6 illustrates the final step in this series where we joined preformed [B-C] with [A]. As can be seen, the final, complete gene, [A-B-C], readily separates on an agarose gel-filtration column from intermediates [A] and [B-C]. Nearest neighbor analysis confirmed the total synthesis.

Once the basic principles of joining deoxyoligonucleotides using T4 ligase had been determined, the overall approach for gene synthesis was established.

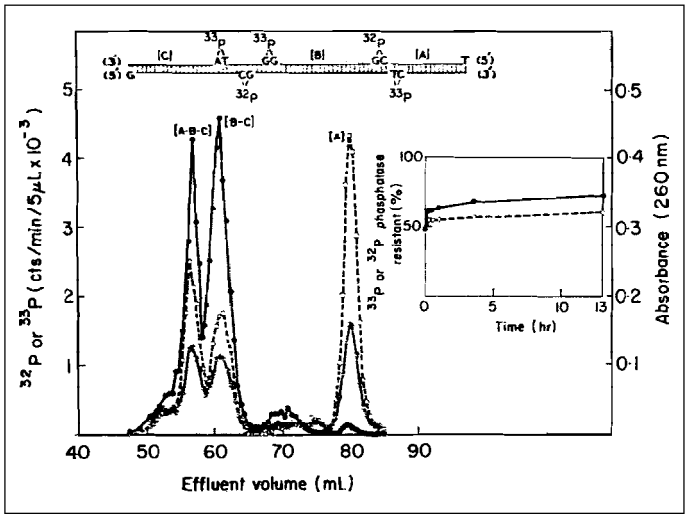

Figure 6. Agarose column purification of the yeast alanine tRNA gene.

The enzymatic joining of the synthetic deoxyoligonucleotides proved to be more difficult than we anticipated. Different sections of the gene, [A], [B] and [C] as shown in *Figure* 2, required different joining strategies. For example, section [A] was best prepared free of unwanted ligation products by joining the individual deoxyoligonucleotides to one another via one ligation step per reaction. Conversely, section [C] could not be prepared using this strategy as either no reaction occurred or incorrect joining was observed when only three segments were mixed. Instead this section was obtained successfully by a one pot reaction. All deoxyoligonucleotides of section [C] were mixed, annealed by heating and slow cooling, and then T4 ligase added.

Synthesis and purification of the final product, [A-B-C], required heating and cooling steps, not only before joining [A] with [B-C], but also prior to chromatography of the final product mixture. Otherwise during this final chromatography on Agarose, no major peaks ([A], [B-C], or [A-B-C]) were observed. Instead only small peaks throughout the effluent were present. These were presumably a mixture of partially ligated products forming incomplete duplexes with one another and with each of the final duplex strands of [A-B-C]. Additionally, recall that the yeast tRNA forms the classical cloverleaf structure. Thus without annealing, each strand of [A], [B-C], or [A-B-C] could form

Synthesis and purification of the final product, [A-B-C], required heating and cooling steps, not only before joining [A] with [B-C], but also prior to chromatography of the final product mixture.

The synthesis of this gene was at the extreme limits of what was technologically possible between 1965 (the project start date) and 1970, when the total synthesis was reported in a *Nature* paper [5].

partial clover leaf structures with one another. When the final synthesis reaction was followed with an annealing step, then the product peak, [A-B-C], was clearly separated on Agarose from unreacted [A] and [B-C]. Perhaps the small peaks observed between [B-C] and [A] intermediates correspond to the single strands of [A-B-C] in the classical cloverleaf tRNA structure.

The synthesis of this gene was at the extreme limits of what was technologically possible between 1965 (the project start date) and 1970, when we reported the total synthesis in a *Nature* paper [5]. It required an enormous investment in time (e.g., two years for Hans Weber to synthesize the 20mer as described in *Figure* 4). By one estimate, the total project required at least 20 man-years to complete [6]. The chemistry was extremely inefficient as coupling yields for preparing the internucleotide linkages were very low (10%–50%) and purification of products (dimer, trimer and tetramer blocks and also the intermediate 5'-monomethoxytrityl deoxyoligonucleotides) required time consuming DEAE-cellulose columns (usually 1–2 weeks per purification). Neither reverse phase nor ion exchange HPLC were available. Slab gel electrophoresis only became possible in 1970–1971. Characterization of products was by techniques (nearest neighbor analysis) that were barely adequate and very time consuming. Mass spectral analysis of deoxyoligonucleotides was still decades away and cloning, coupled with DNA sequencing, was 8–10 years in the future. Starting materials for synthesis required appropriate protection of the deoxynucleotide purine and pyrimidine bases, the sugar hydroxyl groups, and the 5'-phosphate – usually in 20–30 gram scales. There were no chemical supply houses for ordering these appropriately protected monomers. Solvents had to be extensively purified and the hundreds of fractions from each DEAE-cellulose gel-filtration column required individual monitoring by hand using a spectrophotometer. The enzymes T4 ligase, T4 kinase, T4 RNA polymerase, and T4 DNA polymerase were not available commercially. Instead, our isolation of these enzymes started with T4 phage infection of *E. coli* on a scale that would generate 2–3 kg of bacteria paste. Four postdocs set aside

about three weeks to purify these four enzymes on a scale that would last the entire group perhaps six months. All of these procedures required skilled postdoctoral students and were not for laboratory technicians.

In December 1972 the total synthesis of this gene was published as a complete issue (13 manuscripts) in the *Journal of Molecular Biology* [7]. This work was a major milestone in biology as it demonstrated what could be done by one of the great leaders in science. One who had the courage to work at the very edge of the possible and point the way for others to follow in his footsteps.

The total synthesis of yeast alanine gene was a major milestone in biology as it demonstrated what could be done by one of the great leaders in science. One who had the courage to work at the very edge of the possible and point the way for others to follow in his footsteps.

As the historical record demonstrates approximately forty years later, Gobind's foresight was provocative, accurate, and very conservative relative to what synthetic DNA now allows us to do. For example, much of biology, biochemistry, and molecular biology depends upon synthetic DNA as an indispensable component of basic research. These applications include massive DNA sequencing of genes and organismal DNA; PCR; functional studies on genes, proteins, and chromosomes; DNA diagnostics; the use of DNA chips containing 10^5–10^6 unique sequences/chip in order to profile gene expression and to generate modified genes or oligomers useful for monitoring chromosome epigenetic activities; and controlling gene expression using antisense DNA, interfering RNA, and microRNA antimers. Among the many additional applications for synthetic DNA, the cloning of heterologous DNAs in prokaryotic and eukaryotic cells is also noteworthy.

As the historical record demonstrates approximately forty years later, Gobind's foresight was provocative, accurate, and very conservative relative to what synthetic DNA now allows us to do.

E. coli Tyrosine Supressor tRNA Gene Synthesis

Neither Gobind nor his research group were completely satisfied with the synthesis of a yeast alanine transfer tRNA gene. This was because we could not demonstrate biological activity with the gene or easily use it for studies to assess structure–function relationships *in vivo*. In 1968, and well before completion of the alanine gene, we therefore embarked upon the synthesis of the *E. coli* tyrosine suppressor tRNA gene. By the time we started synthesis of this gene, the tRNA sequence had been determined

initially by Sanger's laboratory and, a second time using a different approach, by Tom RajBhandary in collaboration with Susumu Nishimura. We were confident that the sequence was correct. Because of its suppressor activity, we considered it a gene that could be modified and thus used to address a large number of structure–function problems in transcription and translation.

Upon completion of the synthesis plan, the discussion turned to how we were going to express the gene. None of us could propose a viable plan.

And now for some additional insights into the mind and philosophy of a great scientific leader. For two reasons, in October 1968, we decided to develop the synthetic plan for this suppressor gene. Hans van de Sande had just arrived and we wanted to embark on its synthesis. Also Gobind annually attended a small meeting organized by Francis Crick at the Salk Institute and he wanted to present his synthetic ideas to this group. And so we met one afternoon and developed a research plan. Our focus was to minimize the number of synthetic dimers and trimers we had to synthesize and to maximize the possible pyrimidine–pyrimidine chemical joining reactions in the growing 5'-monomethoxytrityl deoxyoligonucleotides. We knew that this type of reaction scheme reduced the number of side-products and increased the overall yields. We also derived a plan whereby repeat and inverted repeat sequences were isolated in separate deoxyoligonucleotides as their presence in the same oligomers would create significant synthesis problems at the enzymatic level. Upon completion of the synthesis plan, the discussion turned to how we were going to express the gene. None of us could propose a viable plan. In 1968, we knew nothing about promoter and terminator sequences and we had no realistic plan on how to incorporate this gene into the *E. coli* chromosome, a phage, or a plasmid. The ability to clone heterologous genes into various organisms would not be available for at least 7 or 8 years. Our only idea, not accepted with any enthusiasm, was to attach the gene to phage λ sticky ends (Ray Wu had determined these sequences by 1968) and hope for some type of viable transfection. After this discussion on expression had continued for 20–30 minutes with Gobind as a silent observer, he stood up (i.e., the meeting was over) and said "Let's first synthesize the gene. By that time we will know how to express it." This is exactly what happened.

After this discussion on expression had continued for 20–30 minutes with Gobind as a silent observer, he stood up (i.e., the meeting was over) and said, "Let's first synthesize the gene. By that time we will know how to express it." This is exactly what happened.

express it." This is exactly what happened. As we proceeded with the synthesis, first the pre-tRNA sequence was discovered by Sid Altman (1971) and later both promoter and terminator motifs were characterized using early DNA sequencing methods. These sequences were added to the overall plan. Meanwhile methods for cloning heterologous DNAs into plasmids and *E. coli* were developed (1975–1976) and the sequences for various restriction enzymes became known. These observations led to the addition of restriction cleavage sites at the ends of the gene. Thus the final *E. coli* tyrosine suppressor construct (207 base pairs) contained the structural gene flanked by a terminator sequence, a promoter, the pre-tRNA region, and restriction sites. It was then cloned into a λ phage at a restriction endonuclease R1 site located near the middle of the total genome. This phage contained two amber mutations. Transfection into *E. coli* generated viable phage which demonstrated suppressor activity for this synthetic gene [8].

Gobind often said, "If you want to break new ground in science, you have to walk the path alone."

Thus Gobind's foresight proved ingenious. By the time we had synthesized the gene, we did know how to express it. Over the years, all of us who knew Gobind well would not consider his solution to this daunting challenge anything but the way he had always chosen to carry out his science. In every endeavor that he accepted, Gobind always led the way before others even knew that the opportunity existed for opening a new era in science. He often said, "If you want to break new ground in science, you have to walk the path alone."

During Nobel Ceremonies in December 1968, while we (Group III) were at the Enzyme Institute in Madison not only preparing for Christmas but also writing Federation Proceedings Abstracts for our first unveiling of the partial synthesis of the yeast tRNA gene, Gobind sent the following telegram from Stockholm. "Group III: You too can win Nobel Prizes. Never cease to discover your potential. Waste no time on your presumed shortcomings. Do not forget Federation Abstracts. Merry Christmas. Gobind." This was Gobind's teachings – never cease to discover your potential and waste no time on your presumed shortcomings. Gobind was

This was Gobind's teachings – never cease to discover your potential and waste no time on your presumed shortcomings.

our lifelong friend and mentor who taught us how to do good science and to enjoy the path it created for each of us.

Suggested Reading

[1] H G Khorana, Nucleic Acid Synthesis, *Pure Appl. Chem*., Vol.17, pp.349–382, 1968.

[2] K Kleppe, E Ohtsuka, R Kleppe, I Molineux and H G Khorana, Studies on Polynucleotides, XCVI, Repair Replication of Short Synthetic DNA's as Catalyzed by DNA Polymerase, *J. Mol. Biol*., Vol.56, pp.341–361, 1971.

[3] H G Khorana, Synthesis in the Study of Nucleic Acids; The Fourth Jubilee Lecture, *Biochemical Journal*, Vol.109, pp.709–725, 1968.

[4] H G Khorana, The Robert A Welch Foundation Conferences on Chemical Research, XII. *Organic Syntheses*, pp.173–211, 1968.

[5] K L Agarwal, H Buchi, M H Caruthers, N Gupta, H G Khorana, K Kleppe, A Kumar, E Ohtsuka, U L Rajbhandary, J H van de Sande, V Sgaramella, H Weber and T Yamada, Total synthesis of the gene for an alanine transfer ribonucleic acid from yeast, *Nature*, Vol.227, pp.27–34, 1970.

[6] G J Powers, R L Jones, G A Randall, M H Caruthers, J H van de Sande and H G Khorana, Optimal strategies for the chemical and enzymatic synthesis of bihelical deoxyribonucleic acids, *J. Am. Chem. Soc*., Vol.97, pp.875–884, 1975.

[7] H G Khorana, K L Agarwal, H Buchi, M H Caruthers, N K Gupta, K Kleppe, A Kumar, E Otsuka U L RajBhandary, J H van de Sande, V Sgaramella, T Terao, H Weber and T Yamada, Studies on polynucleotides, 103. Total synthesis of the structural gene for an alanine transfer ribonucleic acid from yeast, *J. Mol. Biol*., Vol.72, pp.209–217, 1972 and the following twelve manuscripts in this volume.

[8] H G Khorana, Total Synthesis of a Gene, *Science*, Vol.203, pp.614–625, 1979.

Address for Correspondence
Marvin Caruthers
Department of Chemistry and Biochemistry,
University of Colorado at Boulder, Boulder, Colorado 80309, USA
Email:
Marvin.Caruthers@colorado.edu