Total Synthesis of a Gene

H G Khorana*

Summary. The method developed for the total synthesis of a given DNA containing biologically specific sequences consists of the following. The DNA in the double-stranded form is carefully divided into short single-stranded segments with suitable overlaps in the complementary strands. All the segments are chemically synthesized starting with protected nucleosides and mono-nucleotides. The 5' -OH ends of the appropriate oligonucleotides are then phosphorylated with the use of [γ-32P]ATP and polynucleotide kinase. A few to several neighboring oligonucleotides are then allowed to form bilhedral complexes in aqueous solution, and the latter are joined end to end by polynucleotide ligase to form covalently linked duplexes. Subsequent head-to-tail joining of the short duplexes leads to the total DNA. The methods are described for the construction of a biologically functional suppressor transfer RNA gene. The total work involved (i) the synthesis of a 126-nucleotide-long bilhedral DNA corresponding to a known precursor to the tyrosine suppressor transfer RNA, (ii) the sequencing of the promoter region and the distal region adjoining the C-C-A end, which contained a signal for the processing of the RNA transcript, (iii) total synthesis of the 207 base-pair-long DNA, which included the control elements, as well as the Eco R1 restriction endonuclease specific sequences at the two ends, and (iv) full characterization by transcription in vitro and amber suppressor activity in vivo of the synthetic gene.

Organo-chemical methods for the synthesis of oligonucleotides began to be developed (1) soon after the elucidation of the structures of the nucleic acids (2, 3). While considerable advances were made in the 1950’s and 1960’s in constructing polydeoxyribonucleotides of defined nucleotide sequences (4), there continued to be severe practical limits on the size of the polynucleotide chains that could be assembled

* The author is Alfred P. Sloan, Professor of Biology and Chemistry, Massachusetts Institute of Technology, Cambridge 02139.

unambiguously by purely chemical methods. However, for biological studies, completely defined polynucleotides in the size range often much higher than those accessible by chemical methods are required. The formidable tasks confronting organic chemistry were recognized early (5) and, therefore, attempts were made to couple chemical methods, which alone offer oligonucleotides of controlled sequences, to other concepts, which together would afford high-molecular-weight nucleic acids of defined structures. Thus, in the 1960’s, it was possible to use short synthetic deoxyribo-oligonucleotides with repeating sequences as templates for the nucleic acid-polymerizing enzymes, and this approach enabled the preparation of a variety of double-stranded DNA-like polymers of high molecular weight and messenger RNA’s (mRNA) of defined sequences (4, 6, 7). The latter proved to be very useful in studies of the genetic code. However, the large objective of the synthesis of macromolecular DNA’s having nonrepeating and biologically specific sequences (5) required a different approach. Toward this goal, the central concept was the inherent ability of polynucleotide chains to form ordered birehelical complexes by virtue of base pairing. Thus, the goal was to join, end to end, chemically synthesized polydeoxyribonucleotides while these were held together in properly aligned birehelical complexes. The discoveries of the enzymes polynucleotide ligase (8) and polynucleotide kinase (9) proved crucial in these studies and, fortunately, the average size of oligonucleotides, which was demanded of chemical synthesis, proved to be quite short, being well within the range of chemical techniques (10). Synthesis of a double-stranded DNA corresponding to the major yeast alanine transfer RNA (tRNA), which was the first one to be sequenced, became the first objective, and by 1970 this objective had been accomplished (11). From various considerations, which are described later, an Escherichia coli tRNA gene offered much greater opportunities for biochemical and biological studies; and, already in 1968, the E. coli tyrosine suppressor tRNA gene had been chosen as the target for synthesis. The total project involved (i) the synthesis of a 126-nucleotide-long birehelical DNA corresponding to a precursor to the tRNA, (ii) the sequencing of the adjacent promoter region and the distal region, adjoining the C-C-A end, which contained a signal for processing the RNA transcript, (iii) syntheses of the DNA’s corresponding to these control regions, and (iv) biochemical and biological studies of the totally synthetic gene. This article presents a brief overview of the different phases of the above work, from its start in the late 1960’s until its completion recently.

Chemical Synthesis of Deoxyribo-oligonucleotides

While a detailed account of the chemical methodology is beyond the scope of this article,