INTEGRATED DNA TECHNOLOGIES

Chemical Synthesis and Purification of Oligonucleotides

Contents

1. Introduction

Nearly every molecular biology technique in use today employs chemically synthesized DNAs or RNAs. This includes PCR, Real-Time PCR, DNA sequencing, site directed mutagenesis, single-nucleotide polymorphism (SNP) assays, microarrays, and the rapidly expanding world of small RNAs. Unlike other reagents, however, each oligonucleotide is custom made according to the specific needs of the individual researcher and purified according to the application for which it is intended.

In recent years, advances in oligonucleotide synthesis chemistries as well as in purification and quality control technologies have led to substantial increases in both quality and yield and to substantial decreases in cost. At the same time, developments in oligonucleotide applications have spurred a dramatic expansion of the range of available modifications. As technological advances in applications have ushered in an era of high throughput analyses, so, too, have advances in oligonucleotide synthesis, purification, and QC led to the establishment of high throughput manufacturing of the full range of possible syntheses such as those available at Integrated DNA Technologies.

Here, we give a brief history of chemical synthesis, present the fundamentals of chemical synthesis of oligonucleotides as it is practiced today, and present some of the more basic issues arising from syntheses including those affecting scale and yield.

---

2. A History of Chemical Synthesis

The elucidation of the Genetic Code was accomplished by 1966 and is an example of brute force applied to a delicate problem. One of the main figures was Professor H. Gobind Khorana**,** a biochemist at the University of Wisconsin. It was understood then that, in nature, the formation of the phosphate linkages in DNA was catalyzed enzymatically in a reversible reaction:

$$dNTP + (dNMP)_n \leftrightarrow (dNMP)_{n+1} + pp_i$$

That is, a deoxynucleotide triphosphate was added to a growing deoxynucleotide monophosphate polymer by a polymerase in the presence of Mg^{++} and the forward reaction resulted in the n+1 polymer and a pyrophosphate. Based on this understanding, Khorana, and other investigators, began to try ways to accomplish the reaction synthetically. Many of these investigators were already well versed in the chemistry involved in the synthesis of polypeptides. This very successful chemistry, shown graphically in Figure 1, utilized a foundation of a single amino acid bound to a solid support (in this case a resin bead), a protected amino acid, and a condensing agent, dicyclohexylcarbodiimide (DCC) which catalyzed

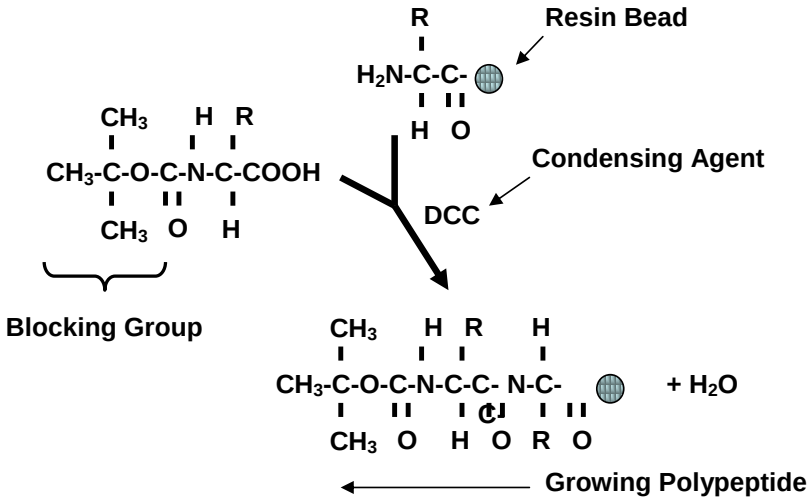


Figure 1. The basic scheme of polypeptide synthesis from a solid support (resin bead). The final deprotection step removed the blocking group from the coupled amino acid and permitted the next coupling reaction to proceed in the presence of DCC.

the peptide linkage. The reaction would link the protected amino acid to the solid support in only one way and the resulting di-peptide was then deprotected for the addition of the next amino acid. In this way, the synthetic polypeptide grew at the direction of the synthesizing chemist one link at a time through a repeated cycle of steps.

Khorana and his group at the Institute for Enzyme Research at the University of Wisconsin began experimenting with DCC on various solid supports and protecting groups on both ribonucleotides and deoxyribonucleotides. The scheme they developed involved three different protecting groups surrounding the nucleotide (Figure 2). The integrity of the ring structure of the base was protected by a benzoyl or an isobutyryl group, the oxygen on the 3' carbon of the sugar was prevented from forming a reactive hydroxyl until needed by the addition of an acetyl group, and the 5' carbon was blocked by a trityl group. In the synthesis of the trityl group Khorana chose to use a soluble polymeric support. In the case of polypeptide synthesis the support was a resin bead but Khorana settled on a polystyrene bead for nucleic acid synthesis. Synthesis of the trityl began with the attachment of a benzene ring to the polystyrene bead. Next, a second ring was attached via a Friedel-Crafts reaction which is the acetylation (or alkylation) of aromatic compounds by aluminum chloride. This structure was then exposed to a Grignard reagent, an extremely reactive class of compounds used in the synthesis of hydrocarbons. Finally, the nearly completed trityl group was treated with acetyl chloride resulting in a polystyrene methoxytritylchloride species that could be easily attached to the 5' carbon of the pentose sugar (Figure 2).

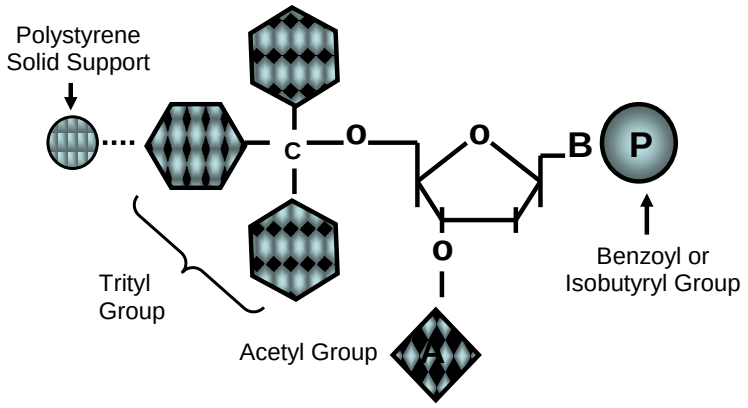


Figure 2. Starting point structure of an oligonucleotide synthesis by the Khorana group. Subsequent couplings would proceed through the 3' carbon after removal of the acetyl group. All nucleotides added from this point would have a reactive hydroxyl in place of the trityl group but would retain the other two protecting groups.

By placing the solid support on the trityl group and using the DCC condensing reagent, the synthesis of an oligodeoxynucleotide proceeded in the same 5'→3' direction as in nature. The nucleotide to which the trityl was attached became the anchor of the synthesis. The oxygen on the 3' carbon of the sugar was deprotected leaving the reactive hydroxyl group ready for coupling. The protected nucleotide, without a trityl, was added via the condensing reaction of DCC. This method was tedious and only modestly efficient but it led to two major breakthroughs. First, Khorana and his group were able to synthesize oligoribonucleotides that were used to confirm the Genetic Code. Second, in 1967, the Khorana-led team at Wisconsin announced their intention to use this chemistry to synthesize a gene. Their attempt to do this

was catalyzed by the discovery in that year of the enzyme, DNA ligase. Khorana realized that they could synthesize overlapping, complementary oligodeoxynucleotides and assemble the gene using the ligase. In 1968, Khorana received the Nobel Prize in Physiology or Medicine for their interpretation of the genetic code and its function in protein synthesis. In 1970 they published the first completely synthetic gene, the 77bp yeast $tRNA_{Ala}$ gene [1]. In this Nature article, they noted, "Unpublished experiments by two of us have given encouraging results on the use of DNA polymerase for replication of the gene in the presence of suitable primers." [1]. Some have speculated that this statement may have been the beginning of the invention of PCR.

---

3. Contemporary Oligonucleotide Synthesis

In spite of the successes of researchers like Khorana, chemical synthesis of oligonucleotides remained labor intensive and inefficient throughout the 1970's and into the early 1980's. In 1983, a breakthrough was achieved in synthesis chemistry that made it possible to make longer oligonucleotides and to make them much more efficiently. The new synthesis process was based on the use of phosphoramidite monomers as building blocks and the use of tetrazole catalysis [2]. A phosphoramidite monomer is a very different synthesis unit compared to its predecessors. First, a phorphoramidite is a normal nucleotide but with protection groups, such as a trityl group, added to its reactive amine, hydroxyl, and phosphate groups. These protection groups prevent unwanted side reactions and force the formation of the desired product during synthesis (Figure 3). The protection groups are removed after the completion of the synthesis process. Second, the link to the solid support is made through the 3' carbon and synthesis proceeds 3' to 5' rather than the 5' to 3' synthesis used previously (Figure 4A). The solid support is a 5 micron controlled pore glass bead (CPG) with holes and channels where the protected nucleotide is attached (Figure 4B). The advances in oligonucleotide synthesis chemistries have resulted in substantial increases in quality and yield with the added advantage of decreasing cost. This is particularly important since each oligonucleotide has to be custom made dependent on the needs of the individual researcher.

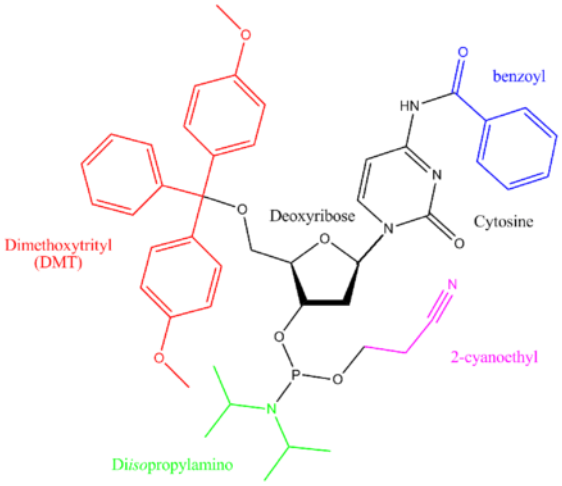


Figure 3. Protection Groups

Phophoramidite synthesis begins with the 3'-most nucleotide and proceeds through a series of cycles composed of fours steps that are repeated until the 5'-most nucleotide is attached. These steps are **deprotection**, **coupling**, **capping**, and **stabilization**.

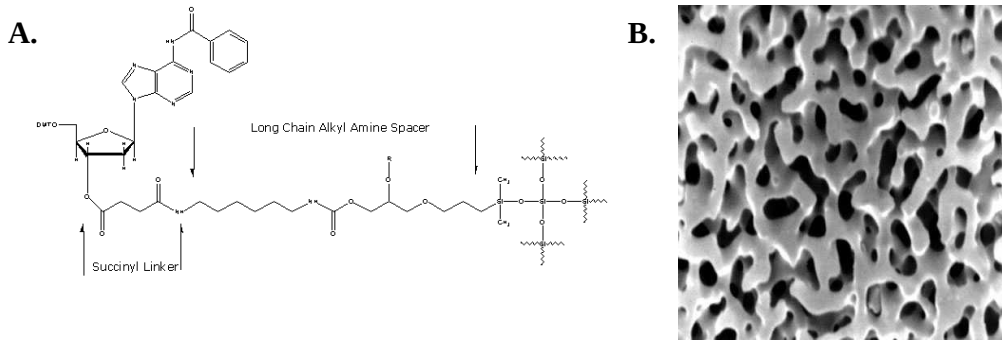


Figure 4. A. A protected nucleoside attached to a CPG. B. An electron photomicrograph of the surface of a CPG bead. The scale of this picture is 10 millionths of an inch square.

3.1 Deprotection

In the classic deprotection step, the trityl group, which is attached to the 5' carbon of the pentose sugar of the recipient nucleotide, is removed by trichloroacetic acid (TCA) leaving a reactive hydroxyl group to which the next base is added.

3.2 Coupling

It is here that the advent of tetrazole activation replaces the use of condensing agents like DCC. Berner et al. showed that tetrazole, a weak acid, attacks the coupling phosphoramidite nucleoside forming a tetrazolyl phosphoramidite intermediate [3]. This structure then reacts with the hydroxyl group of the nucleoside attached to the CPG bead and the 5' to 3' linkage is formed (Figure 5). The unbound base and by-products are washed out, the tetrazole is reconstituted, and the process continues. The use of tetrazole increased coupling efficiency to greater than 99% which allowed longer and longer oligonucleotides to be synthesized.

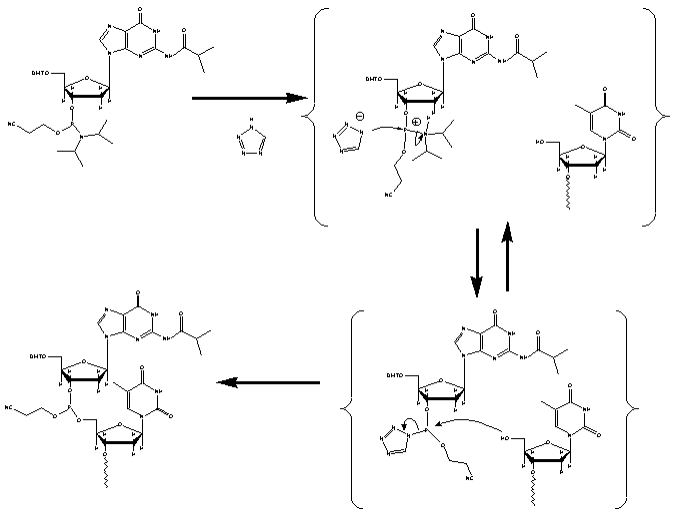


Figure 5. The pathway of tetrazole phosphoramidite-intermediate coupling. The phosphoramidite is introduced in the presence of tetrazole which protonates diisopropylamine leading to the formation of the tetrazole phosphoramidite intermediate. Coupling of the intermediate to the growing oligonucleotide is the final step which returns the tetrazole to its original state.

3.3 Capping

While the increased efficiency afforded by the advent of tetrazole phosphoramidite-intermediate coupling was a major advance in oligonucleotide synthesis, it was still a chemical process and so had a finite failure rate. A coupling failure results when an oligonucleotide retains a reactive hydroxyl group on its 5'-most end. If this were to remain freely reactive, it would be able to couple in the next round and would result in a missing base in the synthesis. Thus, coupling failures must be removed from further participation in the synthesis. This is accomplished by adding an acetylating reagent composed of acetic anhydride and N-methyl imidazole. This reagent reacts only with free hydroxyl groups to irreversibly cap the oligonucleotides in which coupling failed.

3.4 Stabilization

Once the capping step is accomplished, the last step in the cycle is oxidation which stabilizes the phosphate linkage between the growing oligonucleotide chain and the most recently added base. The phosphate linkage between the first and second base must be stabilized by making the phosphate group pentavalent. This is achieved by adding iodine and water which leads to the oxidation of the phosphite into phosphate leaving the phospho-triester bond stabilized.

3.5 Cleavage, Detritylation, Deprotection

This cycle is repeated for each nucleotide in the sequence. At the end of the synthesis the oligonucleotide exists as, for example, a 25-mer with the 3' end still attached to the CPG and the 5' end protected with a trityl group. In addition, protecting groups remain on three of the four bases to maintain the integrity of the ring structures of the bases. The protecting groups are benzoyl on A and C and N-2-isobutyryl on G (figure 6). Thymidine needs no protecting group. The completed synthesis is detritylated and then cleaved off the CPG leaving a hydroxyl on both the 3' and 5' ends. At this point the oligo (base and phosphate) is deprotected by base hydrolysis using ammonium hydroxide at high temperature. The final product is a functional single-stranded DNA molecule.

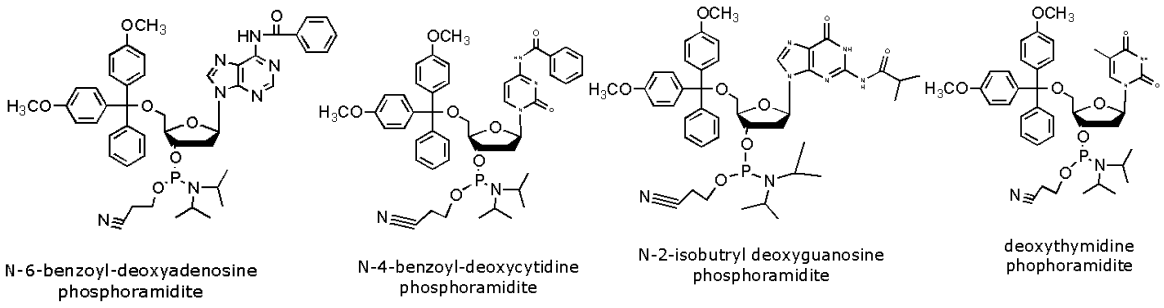


Figure 6. Structures of the four nucleoside phosphoramidite monomers. The benzoyl and isobutyryl protecting groups on the A, C, and G monomers are shown.

3.6 Desalting

Although deprotection removes the protecting groups, they remain with the oligonucleotide as organic salts. The process of removing these contaminants is called **desalting**.

4. Synthesis Scale and Yield

Scale refers to the amount of starting material which is composed solely of the 3'-most nucleotide of a sequence attached to a solid support used to make the oligonucleotide. **Yield** refers to the amount of final product recovered after all the synthesis, processing, and purification steps associated with the oligonucleotide have been completed.

As noted above, because of chemical and physical restraints, coupling efficiency is less than 100% at each step in the synthesis. In addition, coupling efficiency varies for each base both by type and position in the growing oligonucleotide [4, 5]. IDT's experience gained from the synthesis of millions of oligonucleotides shows that some sequences will result in better yields than others. **Monitoring efforts at IDT confirm that our average coupling efficiency for oligonucleotides exceeds 99%.** This means that the failure rates should be 1% or less per coupling event. However, this results in a cumulative population of failures. For a 25-mer with a 99.5% coupling efficiency, the percentage of full-length product can be estimated as $(.995)^{24}$ which is equal to 88.6% full-length product (FLP). For a 25-mer with a 99% coupling efficiency, the estimated FLP would be 78.5%. (Figure 7). The exponent is (n-1) because the very first base on the 3' end is given as a result of the CPG. Thus, while there are 25 nucleotides there are only 24 couplings.

Based on the relationship between expected yield and the actual scale of the synthesis, IDT routinely offers yield guarantees based upon a combination of oligonucleotide length and synthesis scale. Synthesis of very long oligonucleotides (those greater than 60 bases) is problematic. Thus, IDT does not offer any yield guarantees for oligonucleotides greater than 60 bases at our highest synthesis scale and greater than 100 bases on any scale.

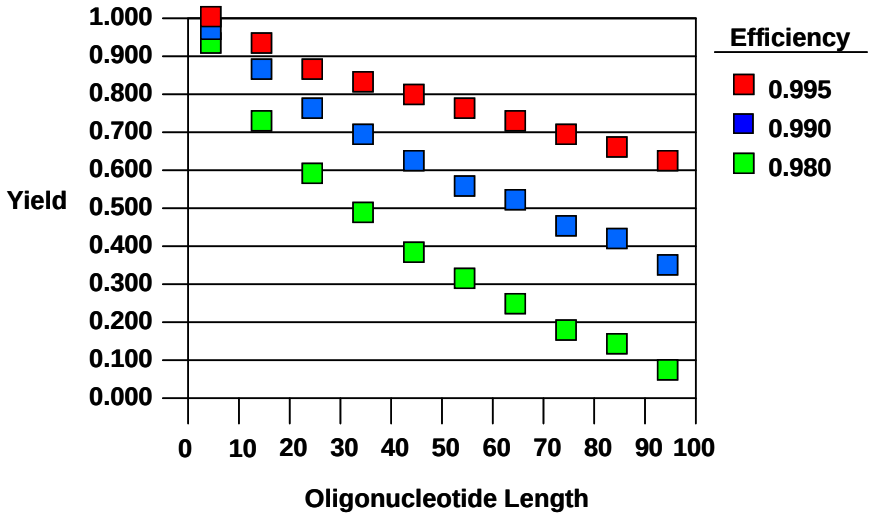


Figure 7. Relationship between average coupling efficiency and synthesis yield over a range of oligonucleotide lengths.

Most PCR primers are in the 20- to 28-mer range which means that a standard synthesis will contain between 85% and 90% full-length product. For PCR applications, the overwhelming mass of full-length product will be sufficient to produce the amplicon of interest in an equally overwhelming mass. Thus, for PCR, there is little need for purification beyond that of desalting particularly with the very high quality of oligos that IDT maintains. IDT also offers standard desalting for products with a number of the more common modifications appropriate for many applications (see the IDT website for more information).

As a general rule, IDT recommends that any oligonucleotide longer than 40 bases should receive further purification. In addition, for demanding applications such as site-directed mutagenesis, cloning, and gel-shift protein-binding assays, additional purification is recommended even for oligonucleotides shorter than 40 bases. Please note that additional purification will result in a decrease in final oligonucleotide yield.

---

5. Purification

Purification of an oligonucleotide is application-dependent. For longer oligonucleotides, such as those that are typically used for cloning and hybridization applications, purification is important because the non-full-length mass might interfere with the applications. Similarly, oligonucleotides modified with non-standard bases, fluorescent dyes, linkers, etc. should be subjected to additional purifications.

IDT uses two methods for post-production purification:

1. Polyacrylamide gel electrophoresis (PAGE)
2. High performance liquid chromatography (HPLC)

5.1 PAGE

High percentage acrylamide gels are used to separate and elute the full-length product from all shorter species with great efficiency. PAGE is the most efficient means of purification for oligonucleotides that are **unmodified** and only need to have the capped products removed. PAGE purification does result in an unavoidable loss of mass because it is physically impossible to recover every bit of full-length product from a gel slice. Thus, a PAGE-purified 50-mer will yield a substantially lower mass relative to the starting synthesis scale. However, the loss of mass is typically an acceptable trade for the increase in purity. PAGE purification is strongly recommended for oligos over 60 bases in length. For unmodified oligos, purity of >90% is routinely achieved.

5.2 HPLC

HPLC utilizes the concept that a 25-mer oligonucleotide will have a specific charge and a specific affinity, or lack of affinity for particular solvents. This method is usually reserved for oligonucleotides that have been modified through the addition of a linker or spacer, a non-

standard base or bases, or hydrophobic molecules. HPLC is a form of column chromatography that utilizes a column to hold a stationary phase while the sample is applied within a mobile phase. The analyte's motion through the column is slowed by specific chemical or physical interactions with the stationary phase as it passes through the length of the column. The amount of retardation depends on the nature of the analyte, stationary phase and mobile phase composition and, thus, retention time is unique for each analyte.
HPLC comes in two basic forms.

1) **Reverse-phase HPLC** - separation is on the basis of the difference in solvent affinity between modified and unmodified oligonucleotide.
2) **Ion-exchange HPLC** - separation is on the basis of the differences in the net charge between modified and unmodified oligonucleotide.

The type of HPLC used is dependant on the particular sample. Again, there will be an unavoidable loss of mass due to purification but this will be offset by the gain in purity.

5.3 Purification Quality Control

Synthesizing and purifying an oligonucleotide does not provide information about the quality of the oligonucleotide. The oligonucleotide can be tested for quality prior to its use in an experiment through two methods:

1) Mass spectrometry
2) Capillary electrophoresis (CE)

5.3.1 Mass spectrometry

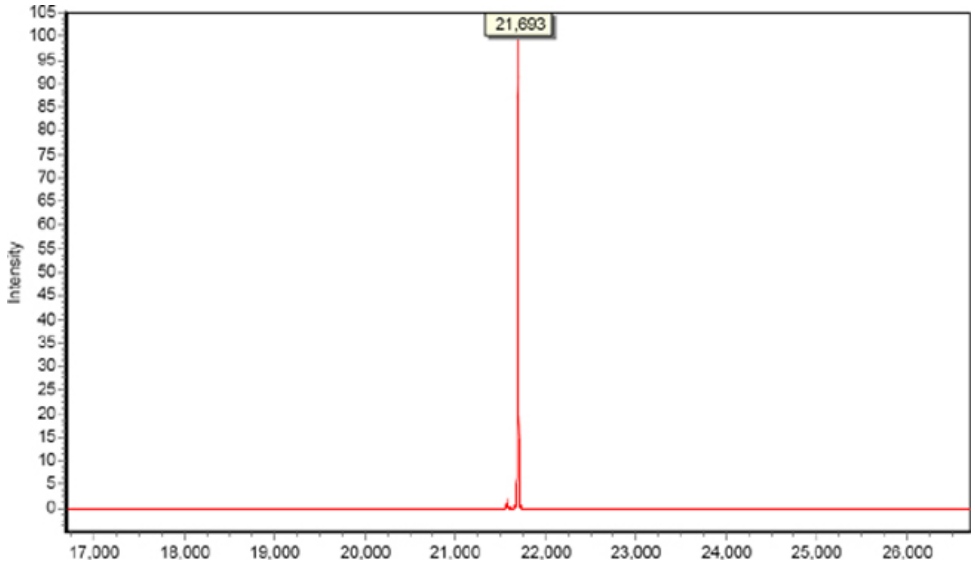
Two types of Mass spectrometry may be employed:

1) MALDI-TOF: Matrix Assisted Laser Desorption Ionization – Time Of Flight
2) ESI: Electrospray Ionization

MALDI-TOF can resolve oligonucleotides up to 50 bases in length (~15,000 Daltons) and is based on the principle that the speed at which an ion moves is inversely proportional to its mass. Thus, in the ion chamber of a MALDI instrument, materials are ionized and given the same potential energy, eV, where V is the potential and e refers to the number of charges on the ion. As the ions emerge from the ion source, the potential is converted to kinetic energy in the moving ion. Ions in motion obey the rule $E = 1/2mv^2$, or $m = 2E/v^2$, where m is the mass of the ion, v is velocity, and E is energy. Since the amount of energy is a constant in the MALDI, the mass of the ion can be determined by velocity alone. Velocity is simply time over distance such that the time of arrival of the ion in the detector of the MALDI (time of flight) is directly converted to velocity because the distance is a constant as well. In oligonucleotide synthesis QC, MALDI-TOF has the added advantage that the mass of an oligonucleotide can be estimated with precision since the masses of the individual nucleotides are fixed (Figure 8). For any given sequence, the expected arrival time will be given by the expected mass and any deviation from that time of arrival will indicate a deviation from ideal size and/or purity.

Electrospray Ionization (ESI) mass spectrometry can resolve longer oligonucleotides (up to ~150 bases in length). The material to be analyzed by ESI begins in liquid form and is dispersed into a fine aerosol by electrospray. The aerosol enters a heated capillary where any remaining solvent is evaporated from the charged droplets. Smaller droplets will evaporate more quickly than larger droplets. Once the solvent is evaporated, only the ions remain in the gas phase. The ions will have a variety of charges and so will hit the detector at different times based on the charge. Based on the peak from when an ion hits the detector, the charge state is determined and then, from this, the mass.

Figure 8. Example of a Mass spectrometry trace. The main peak is the main product at the indicated mass in Daltons.



5.3.2 Capillary electrophoresis

Capillary electrophoresis uses a small amount of the final synthesis product that is subjected to a constant electrical field in a hair-thin capillary. As the product migrates into the capillary it is separated into component sizes in a manner exactly like gel electrophoresis. The fragments will migrate past an optical window and an ultraviolet beam detector will assess the density of the fragments. The “trace” that is produced is composed of a series of peaks corresponding to material densities flowing past the detector (Figure 9). This density profile is made quantitative by establishing a base line and integrating the area under the individual peaks. The purity of the product is the ratio of the main peak to the total area under all peaks. In practice, the main peak should be the last peak since no species should be longer than the full-length product. Peaks that appear to the right of the main peak indicate residual impurities and other potential contaminants.

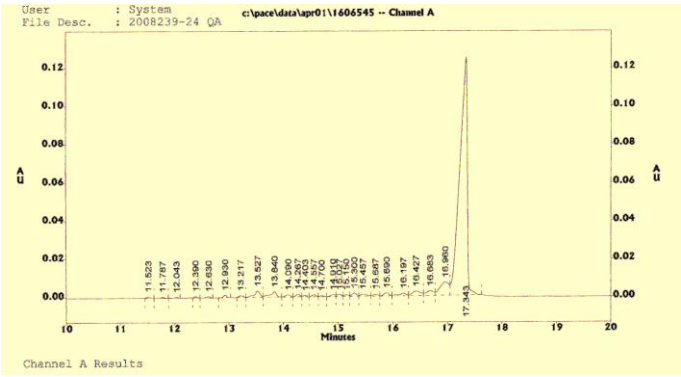


Figure 9. A capillary electrophoresis of an oligonucleotide synthesis. The axes are uV absorbance units (AU) versus time on the column in minutes. The main peak is clearly the dominant peak and the "shoulder" to the left is the (n-1)-mer peak. Note, there is no material detected to the right. The ratio of the main peak to the total indicates a purity

6. Summary

The history of chemical synthesis of oligonucleotides goes back to the discovery of DNA itself. The model of contemporary oligonucleotide synthesis including protected synthesis units, solid support anchors, and end-to-end synthesis, dates to the first attempts to synthesize polypeptides in the 1930's. The advent of phosphoramidite chemistry involving tetrazole intermediates in the 1980's made it possible to make longer and longer oligonucleotides and to make standard lengths faster and far more efficiently. Coupled with this, the transition of techniques like PAGE, HPLC, Mass spectrometry, and CE and MALDI to high throughput analytical platforms, has made the synthesis of high quality, high purity oligonucleotides on an industrial scale possible at Integrated DNA Technologies.

7. References

1. Agarwal KL, Buchi H, et al. (1970) Total synthesis of the gene for an alanine transfer ribonucleic acid from yeast. Nature, 227(5253): 27–34.
2. McBride LJ and Caruthers MH. (1983) An investigation of several deoxynucleoside phosphoramidites useful for synthesizing deoxyoligonucleotides. Tetrahedron Letters, 24(3): 245–248.
3. Berner S, Muhlegger K, and Seliger H. (1989) Studies on the role of tetrazole in the activation of phosphoramidites. Nucleic Acids Res, 17(3): 853–864.

4. Temsamani J, Kubert M, and Agrawal S. (1995) Sequence identity of the n-1 product of a synthetic oligonucleotide. Nucleic Acids Res, 23(11): 1841–1844.
5. Hecker KH and Rill RL. (1998) Error analysis of chemically synthesized polynucleotides. Biotechniques, 24(2): 256–60.

