A Roadmap to the Assembly of Synthetic DNA from Raw Materials

Yogesh S. Sanghvi

Rasayan Inc., Encinitas, California

Cite as:

Sanghvi Y. 2005. A Roadmap to the Assembly of Synthetic DNA from Raw Materials. In: *Working Papers for Synthetic Genomics: Risks and Benefits for Science and Society*, pp. 17-33. Garfinkel MS, Endy D, Epstein GL, Friedman RM, editors. 2007.

(this page blank)

A Roadmap to the Assembly of Synthetic DNA from Raw Materials

Yogesh S. Sanghvi

Rasayan Inc., Encinitas California

Introduction

Until recently, the synthesis of DNA has been a tedious, time consuming, expensive and experimentally challenging task. But advances in automated instrumentation and improved chemistry have now made it possible to make any moderate-length sequence of DNA in any quantity. The ease of automated chemical synthesis of DNA has triggered a whole new industry of low-cost DNA suppliers around the globe. The convenience of ordering DNA sequence by mail has opened new avenues in research both in academia and in the healthcare products developed by pharmaceutical companies. At the same time, these advances have made it theoretically possible to synthesize DNA that could be used to do harm. This article aims to describe the first stages of DNA synthesis, from readily available raw materials to medium-sized segments with a desired sequence (oligonucleotides), and examines whether there are points at which such activities could be, for example, monitored or controlled. Some academic and commercial applications of DNA synthesis require the construction of very small quantities of the desired sequence; others involve synthesis at the gram scale or larger. I provide comments on possible intervention points for both types of application. Terms shown in bold are defined in the glossary.

1. History and key landmarks

Our current methods of DNA synthesis have evolved over almost 150 years. In 1869, Friedrich Miescher first isolated a new substance from human pus cells, which he named nuclein. Two years later, he found that the same material could also be isolated from salmon caught in the river Rhine. Subsequently, Richard Altman in 1889, further purified

nuclein as a protein-free product that he called **nucleic acid**. In 1900, Albrecht Kossel studied the chemical composition of nucleic acids, and found that they contained adenine, cytosine, guanine and thymine bases (Figure 1). Kossel was awarded the Nobel Prize for his work. In his acceptance speech he noted that "nucleic acids possess a great biological significance". In 1902, Emil Fischer received his Nobel Prize for the first chemical synthesis of a **purine** base.

In 1955, the first chemical synthesis of a dimeric block of DNA was accomplished by M. Michelson and Alexander (Lord) Todd. Subsequently, this contribution was recognized by a Nobel Prize to Todd. Next, Har Gobind Khorana and his colleagues showed how a DNA sequence could be assembled via chemical means, now known as the phosphodiester method. In 1976 Khorana with his 19 co-workers reported on the synthesis of a 126-residue long DNA. This project took 8 years; today the same product can be made in one day using an automated DNA synthesizer.

The pioneering work of Robert Letsinger, Kevin Ogilvie and Colin Reese using the phosphotriester method also helped to pave the road to solution-phase synthesis of DNA. In the mid-1970's, the first solid-phase preparation of DNA was performed in the laboratories of Hubert Köster, Michael Gait and K. Itakura. Solid-phase synthesis is the dominant method used today. The specific chemistry we use today came slightly later, in 1981, when Mark Matteucci and Marv Caruthers reported an efficient automated synthesis of DNA employing the P(III) **amidite** chemistry.

2. Transforming raw materials into the building blocks of DNA

DNA is a long chain polymer that is made up of four repeating units called **nucleotides**. Half of the structure is identical for all four nucleotides, and consists of the sugar and phosphate groups (red boxes, Figure 1). The other half of the structure, the **base** (blue boxes, Figure 1) comes in four varieties, divided into two groups. The **pyrimidines** (thymidine and cytosine) each have a six-membered ring containing nitrogen, while the **purines** (adenine and guanine) have a double ring, a fusion of a six-membered ring with a five-membered ring. In the famous double helix of DNA, these nucleotides line up as

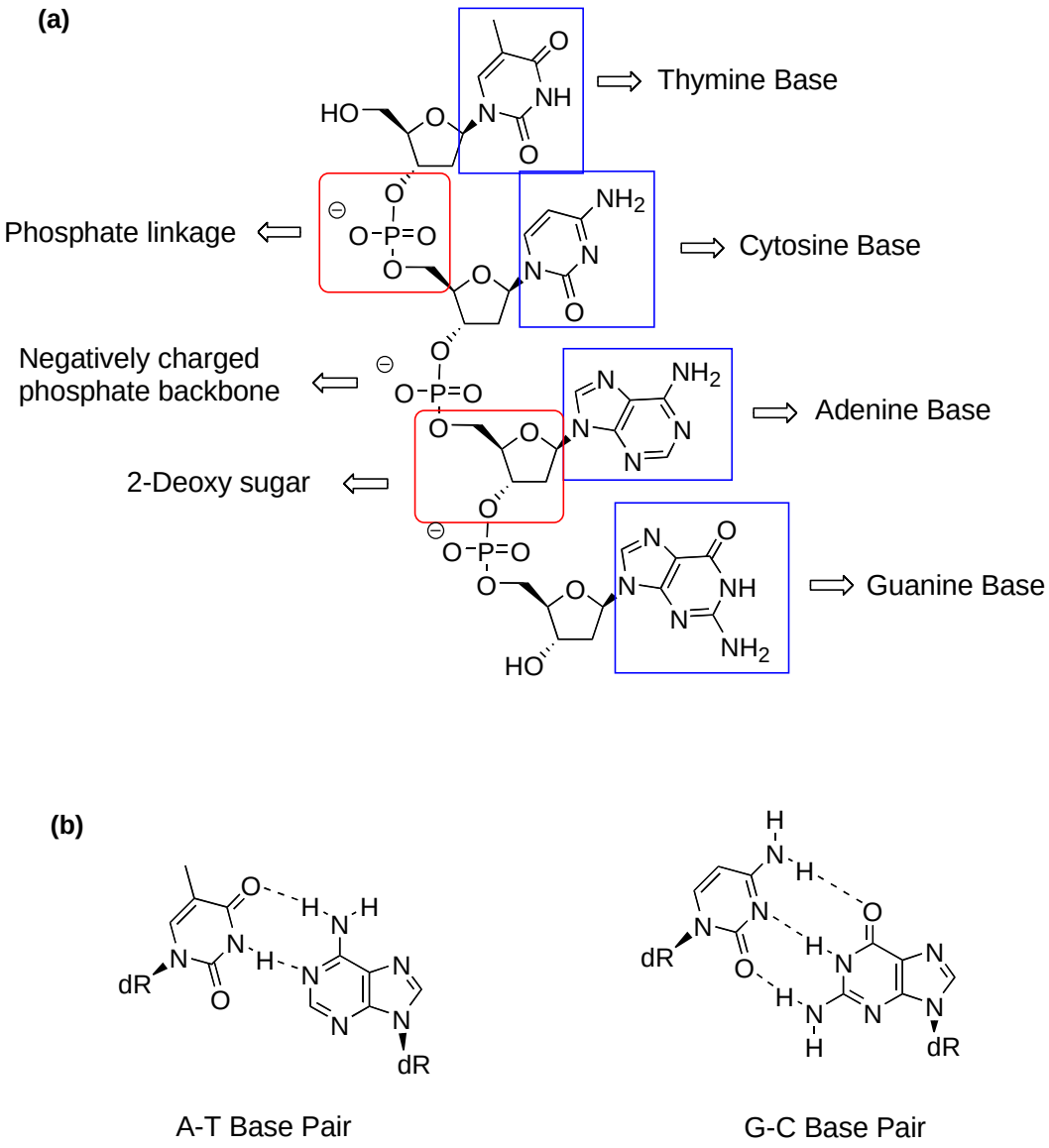

Figure 1: Structural elements of DNA. (a) A single strand of DNA, showing the structure of the bases and the backbone elements. Note that the negative charge on the backbone is balanced by positively charged ions such as Na^+ in a solution of DNA. (b) Structure of the base pairs that form when two matched strands of DNA are allowed to pair to make a double helix.

pairs: as shown in Figure 1, adenine (A) pairs with thymine (T), and cytosine (C) pairs with guanine (G). Because a six-membered-ring base always pairs with a double-ring base, the spacing between the two strands of DNA is maintained, and the overall shape of the molecule is the same no matter what the sequence. The "backbone" of the structure is also always the same, a repeated pattern of sugar-phosphate groups. It is this uniformity of structure that makes it possible to automate the synthesis of DNA. No matter the sequence to be produced, the chemical reaction required is always the same. The problem of making the sequence of DNA needed can thus be reduced to the problem of using the right nucleotide building blocks in the right order.

2.1. Availability of bases

All four bases are available in metric ton quantities from a variety of sources. The cheapest suppliers are in China; they sell their product for under $100/Kg. These products are chemically synthesized and are stable indefinitely when stored appropriately. The chemical synthesis of all four bases is straightforward and it can be carried out almost anywhere with the help of easily accessible reagents in a chemical laboratory. However, easy access to the bases does not lead to easy access to the nucleotides that are essential for the assembly of DNA (see below).

2.2 Availability of nucleotides and nucleosides

Nucleotides are the key reagents used in DNA synthesis. They can readily be made from **nucleosides** by adding phosphate groups. As recently as six years ago, all the nucleotides needed for DNA synthesis were made from nucleosides isolated from natural sources, such as fish milt. A flow chart of this isolation process is shown in Figure 2. Several companies, including Yamasa in Japan, Reliable in the USA and ProBioSint in Italy have used this method to produce nucleosides in metric ton quantities. It is not a rapid process (it can take 1.5 years from beginning to end) and it is very labor intensive. Some years ago, attempts began to develop alternative sources for nucleoside production. Today, at least six Asian companies manufacture the pyrimidine nucleosides at low-cost and in metric ton quantities using a completely chemical process, starting with cane sugar

(Figure 3). Processing of cane sugar furnishes D-glucose, which is transformed into **2-deoxy-D-ribose** in just a few steps. Next, the 2-deoxy-D-ribose is converted into a reactive α-chloro-sugar that is easily converted into the pyrimidine nucleosides (T and C). Mitsui Chemicals has developed a process for producing purine nucleosides at a very large scale, using a phosphate analog of 2-deoxy-D-ribose. The new process is patent protected and currently practiced in Japan for the production of the purine nucleosides (A and G). All four nucleosides are now available in large quantities from chemical synthesis at a significantly lower cost than the nucleosides isolated from fish milt.

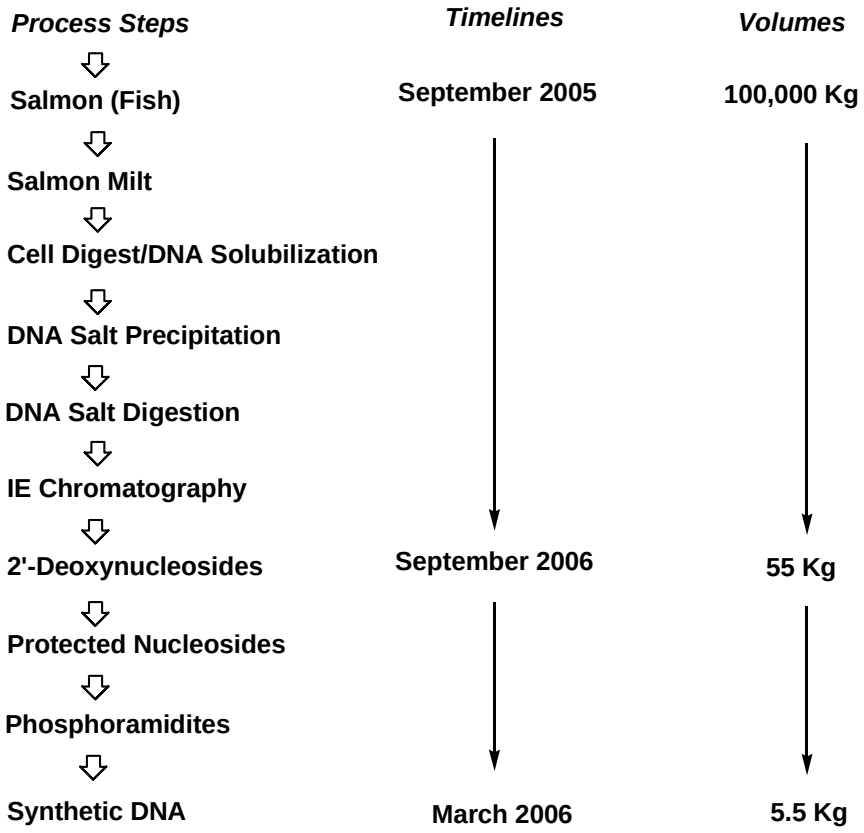

Figure 2: Raw material pipeline from fish to synthetic DNA

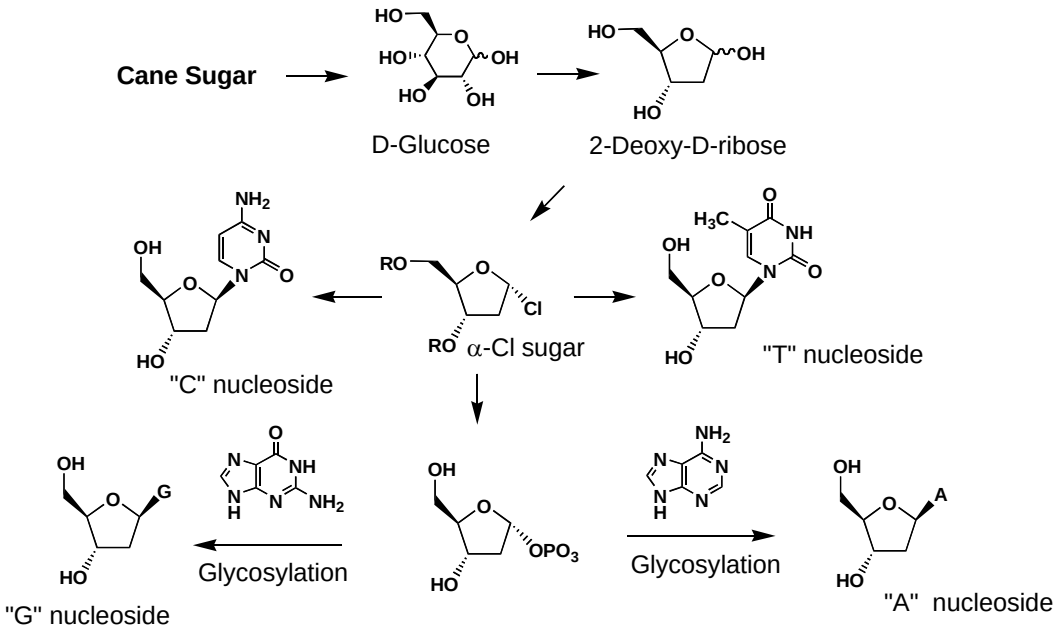

Figure 3: Chemical synthesis pathway for nucleosides from cane sugar

Both of the methods of producing nucleosides require significant skill, especially the chemical synthesis approach. Chemical synthesis of nucleosides requires Ph.D.-level chemistry personnel, specialty chemicals and specialized equipment. The most difficult part in the synthesis of nucleosides is to chemically connect a base to the top face (β-connection) of the sugar. An incorrect linkage from the bottom face will result in the formation of α-nucleosides, which are useless for DNA synthesis. Despite easy access to the bases from China, the synthesis of pure β-nucleosides and high purity amidites is not an easy task for a novice in the field.

In practice, most DNA synthesis today depends on the availability of nucleotide **amidites** (Figure 4), since the amidite chemistry is the dominant chemistry used in automated synthesizers. Good quality amidites are essential for successful synthesis of DNA on an automated machine. The production of good quality amidites is also a skilled task, and large quantities of **anhydrous** solvents and airtight equipment are required.

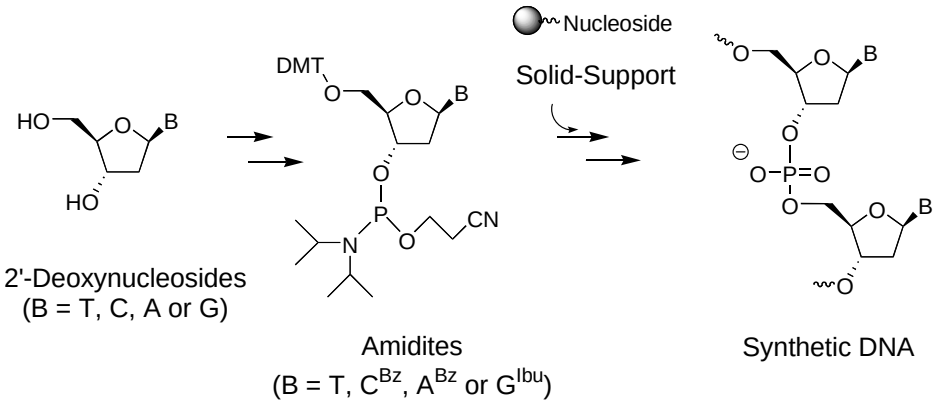

Figure 4: Key raw materials for DNA synthesis. Several steps are required to convert nucleosides to amidites.

2.3 What would it take to make building blocks from scratch?

If a chemist were cut off from all the sources of ready-made nucleotides, nucleosides and bases described above, how would he or she approach the problem of putting together the essential ingredients for DNA synthesis? First, this person would need to be an excellent chemist and have access to a well-equipped chemistry laboratory. The most likely route for such a person to take would be to use the older method of purification from salmon milt. As noted above, isolation of nucleosides from fish is long, tedious and inefficient. To isolate 1Kg of four nucleosides, one would need 1,818 Kg of salmon.

If access to nucleosides is not a problem, the chemist would still need to synthesize nucleotide amidites. This requires the use of special reagents (e.g. phosphitylation reagent), solvents (e.g. anhydrous acetonitrile) and airtight equipment. Isolation, storage and handling of P(III) amidites is an art that is not easily acquired even by an experienced chemist. However, given the tools, training and chemicals, an expert in the field could produce gram quantities of amidites in about six months.

In reality, substantial supplies of nucleosides and all four amidites are already distributed across the globe in large quantities, held by a large number of potential suppliers. It is highly unlikely that even the most concerted international effort would be able to restrict the raw material supplies available to the degree envisioned above.

3. Using the building blocks to make a desired sequence

The assembly of a useful (or harmful) sequence of DNA starts with the assembly of several nucleotides into a medium-length DNA strand, called an **oligonucleotide**. This is done using automated solid-phase synthesis; in other words, the chain of nucleotides is built on a solid bead, one at a time, with washing steps in between. The solid phase is essential to allow multiple steps to be performed with reasonable efficiency. Usually, processes that require a large number of chemical steps give a poor yield; if each step of a six-step synthesis is 95% efficient, the overall yield is only 73%. It would require about 80 individual steps to complete the sequential assembly of a 20-unit long oligonucleotide. The result is that the product is mixed with unreacted starting material and the products of undesired reactions, and can be very hard to purify. The larger the number of steps, the worse the problem gets.

Solid phase synthesis makes multi-step synthesis easier in two ways. First, it is very easy to separate the product (which is attached to the bead) from the unreacted starting material (which is in solution) by simply washing the beads extensively. Second, this ease of separation makes it possible to use large excesses of starting material to drive the reaction very close to completion. In the case of the amidite chemistry that is typically used, each reaction occurs with >99% efficiency. The only impurity left is the product of partial reactions.

Partial reactions are inevitable in any chemical process. For example, if one is trying to construct the sequence ATGCCAA, one would start with an A attached to a bead, then react it with a T. In most cases, the sequence AT will be made, but in a few cases the T will not be added. If the unreacted A is allowed to continue in the elongation reaction, the end result would be the wrong sequence, AGCCAA, which might have a completely different biological effect from the desired sequence. The same problem can occur at any step of the elongation process. Current DNA synthesis technology uses a trick to minimize the problems caused by incomplete reaction, "capping" unreacted sequences with special blocks that prevent their further elongation.

Each cycle of elongation takes place in four main steps, with wash steps in between: **(1)** deprotection, in which a group that prevents premature reaction is removed from the end of a growing nucleotide chain; **(2)** coupling, in which a new nucleotide is added; **(3)** oxidation to stabilize the newly formed linkage; and **(4)** capping of partial products. Finally, the completed chain must be cleaved from the bead. The reagents needed for each step are discussed below. I will focus primarily on the amidite method, because of its widespread use on automated machines that produce thousands of DNA sequences every day around the world. Several other reaction schemes are possible, although they are less efficient.

3.1. Building blocks and reagents required for solid-phase synthesis

The most important reagent required for amidite chemistry is a protected stable amidite derivative (see Figure 4), which provides extremely high (>99%) reaction efficiencies. These amidites are easily synthesized from nucleosides in just a few chemical steps, none of which would be challenging for a reasonably well-trained chemist with access to a sophisticated laboratory. Until recently the manufacturing and sales of these amidites was restricted due to the Köster patent. With the expiration of the patent this year, a number of low-cost Asian suppliers are now producing amidites in commercial quantities. This is one of the key reasons for the recent reduction in the cost of synthetic DNA. The four amidites of interest are wax-like, hygroscopic and easily decomposed upon heating. They must be carefully protected from air, water and heat. For most DNA synthesis applications, the amidites are sold in convenient pre-packed bottles that are simply plugged into a synthesizer without exposing them to air.

The solid support is the second most important raw material needed for DNA synthesis (Figure 5). In essence, the solid support is a small mechanically sturdy polymeric porous bead that is chemically inert during DNA synthesis. The bead must have a reasonable surface area so that each bead can accommodate many growing chains. The most popular solid-support for small-scale synthesis is controlled pore glass (CPG) made from glass or silica (Figure 5). CPG is a special bead custom-made for the synthesis of DNA by a

handful of companies. Synthesis of DNA on ordinary glass is possible but less efficient and would lead to decreased production of the desired DNA strand. Beads made of cross-linked polymers (reminiscent of nylon, but more rigid) can also be used as an alternative support.

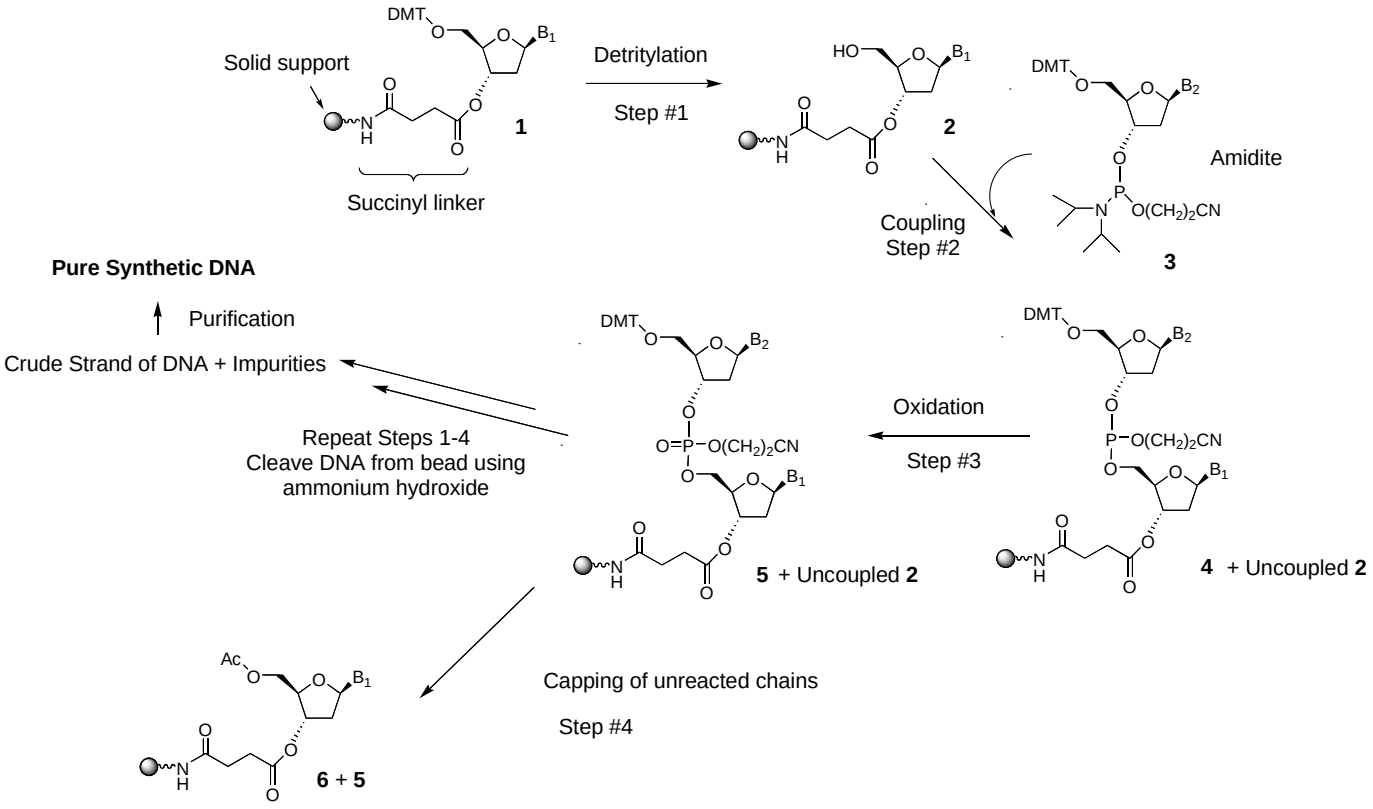

Figure 5: General scheme for automated synthesis of DNA using amidites

Generally solid-supports are sold with the first nucleoside unit already anchored to the surface of the bead via a short cleavable succinyl linker (Figure 4c). The support is placed in a reactor (column) and connected to the automated synthesizer for the chain extension. The first step in oligonucleotide synthesis is the coupling of the first nucleotide to the

nucleoside already attached to the surface of the bead. The addition of each nucleotide unit requires four individual chemical steps and a number of reagents.

These reagents include: (a) a deblocking solution that contains an acid such as dichloroacetic acid (DCA) in dichloromethane (DCM) or toluene; (b) an activator solution such as 1H-tetrazole or 4,5-dicyanoimidazole in acetonitrile; (c) an oxidation solution such as iodine in pyridine, THF and water; (d) two capping solutions, one containing N-methyl imidazole in pyridine and acetonitrile, the other containing acetic anhydride in acetonitrile. These are the only special reagents needed for the four-step repetitive synthesis cycle used in DNA construction (Figure 4). All of them are easily produced from common materials that would be next to impossible to control; alternative reagents have also been described in a variety of publications. In the final step, ammonium hydroxide solution is used to cleave the succinyl linker arm that holds the DNA chain attached to the surface of the solid support and removes the protecting groups that avoid side reactions during synthesis. Ammonium hydroxide is also a very common reagent.

One of the important reagents used for DNA synthesis is the anhydrous acetonitrile required for the washing steps. Because the amidites are very sensitive to water, the grade of acetonitrile needed is higher than for most other applications.

3.2. Chemical steps during assembly

The specifics of the chemical reactions that take place in an average DNA synthesizer are shown in Figure 4. The beads that make up the solid support, with the first nucleoside residue **1** attached**,** are packed into a column to allow solvents to be flowed through them efficiently. The synthesizer is programmed to pump reagents and solvent through the column, and the order of amidite addition is determined by the sequence of DNA needed. The chemical steps shown are: (1) detritylation, consisting of the removal of an acid-labile protecting group from the 5'-hydroxyl group of the nucleoside residue at the end of the growing oligonucleotide chain; (2) coupling of an activated amidite with the 5'-hydroxyl group generated in step 1. (3) oxidation of the labile P(III) intermediate **4** to

stable P(V) product **5** (Figure 4); and (4) capping using a mixture of two solutions, cap A (N-methyl imidazole in pyridine and acetonitrile) and cap B (acetic anhydride in acetonitrile), pumped through the column at the same time. A washing cycle between each step is essential. The four-step protocol is short and very efficient with each cycle completed in just a few minutes.

Note that the "cap" added to the chain that failed to complete the desired reaction is an acetyl (Ac) residue, which is chemically different from the dimethoxytrityl (DMT) blocking group. The DMT group can be removed by gentle acid treatment, freeing it to react in the next cycle. The Ac group withstands this treatment, preventing the chain from elongating. DMT serves two purposes in the cycle; it prevents the amidites from reacting with themselves, and it prevents a chain that has been successfully elongated from being capped. After the capping step is complete the DMT can be removed to allow a new coupling reaction.

3.3. Automated synthesizers

In the early 1980's, the first commercial DNA synthesizers were built and sold by Applied Biosystems. These were single-column 380A and 380B instruments with capabilities to make one DNA sequence at a time on a very small scale (0.2 – 10 μmol). Today, there are a number of instruments on the market with the ability to produce hundreds or thousands of DNA sequences in parallel using both commercial and proprietary instrumentations. For example, Applied Biosystems 3900 DNA synthesizers use 96- or 384-well plates, making a different sequence in each well; specialized companies such as Illumina have adapted this strategy to use synthesizers with large platforms that carry many 384-well plates, again making an individual sequence in each well of each plate. Similarly, high throughput DNA synthesis is available on the MerMade Bioautomation or the Oligator Farm. Although synthesis of DNA without an automated synthesizer is in principle possible, in practice it would be highly inconvenient.

4. Purification

Some uses for oligonucleotides require a purification step to remove the products of incomplete reactions. The most widely used purification technologies are: (1) anion exchange, in which the oligonucleotide (after being cleaved from the solid support used in synthesis) is passed over positively-charged beads that retard the progress of individual oligonucleotides depending on how many negative charges are present on the molecule; and (2) reverse-phase chromatography, which separates molecules based on their degree of hydrophobicity. Both of these purification methods are very widely used for a variety of biochemical applications in academia and industry. New purification methods currently being explored include membrane-based chromatography and simulated moving bed (SMB) chromatography. SMB in particular looks promising, with the potential for >98% purity at the kilogram scale.

5. Points of intervention

The reagents required for oligonucleotide synthesis are almost all so common, or so readily produced, as to defy restriction. The possibilities for restriction differ depending on whether the application to be controlled requires small amounts of material (as is the case for most genetic engineering applications) or large amounts (as for many medical applications). In both cases I focus on the amidite chemistry, since this is by far the most efficient chemistry currently available. Other chemistries can be used, but no sensible chemist would use them unless there was no other option.

5.1 Small-scale synthesis

- *Nucleosides, nucleotides and amidites*. These are the key building blocks of oligonucleotide synthesis. They are used in a range of peaceful industries, including the production of important medicines such as AZT for the treatment of HIV. They are made and sold in very large scale; for example, Proligo (recently acquired by Sigma Aldrich) produces tons of amidites per year. Denial of all ready nucleoside supplies to an oligonucleotide chemist might slow the progress of DNA synthesis for months or years. At the same time, such restrictions would destroy or severely hamper the biotechnology

industry, and the progress of biomedical research. Given the ready availability of amidites from many suppliers across the world, it is hard to imagine an effective program to restrict access to these chemicals.

- *Solid support*. A handful of companies produce the beads that permit efficient DNA synthesis on automated machines. Because of the highly specialized equipment and training required for the preparation of these beads, a skilled individual cannot make these products alone. Restricting access to beads may be worth exploring as a method of controlling DNA synthesis. Note, however, that unless one is willing to destroy the entire DNA synthesis industry, there will be a large number of companies that have legitimate uses for these beads. It would be a significant challenge to track every shipment of beads to every DNA synthesis company and ensure that all the beads are used for legitimate purposes. It is also increasingly possible to make oligonucleotides on derivatized glass slides, which are relatively easy to make by hand.

- *DNA synthesizer*. A number of automated DNA synthesizers are available in the market place. It is possible that tracking the sales of new instruments might allow identification of potential terrorists. However, a very large number of existing instruments have already been sold and would be hard to track in this way; furthermore, an experienced engineer could construct one from spare parts with little difficulty.

5.2 Large-scale synthesis

- *Raw materials*. The amount of raw materials needed for large-scale synthesis provides significant logistical challenges. Only a handful of companies are able to mass-produce these building blocks in the purity required for DNA synthesis. It should therefore be possible to identify and track bulk users of these chemicals. Furthermore, the capital investment in building and running a facility to produce DNA on large-scale is significant. For example, a kilo-scale plant used for the production of DNA would cost $2-5 million in capital investment alone. This does not include the cost of running the facilities. It would be relatively easy to keep track of the construction of such plants world-wide.

- *Solvents and reagents*. Although there are a number of choices for solvents and reagents for DNA synthesis, anhydrous acetonitrile is one solvent that is absolutely essential for the coupling step. A limited number of companies are producing DNA synthesis grade acetonitrile. Monitoring the sales of high purity acetonitrile might allow suspicious organizations to be identified if they are performing large-scale reactions. For small-scale oligonucleotide synthesis, however, anhydrous acetonitrile can be readily produced in the laboratory using a still.

- *Plant permit*. Because a majority of large-scale (e.g. >10kg/year) plants are manufacturing medicines based on DNA they are regulated by the FDA for their GMP compliance. Therefore, it should be possible to monitor any suspicious or non-therapeutic activities and to require careful reagent tracking to minimize the risk that beads or solvents are diverted to other purposes. One possible hurdle could be put in place for such activity is to require a permit of some kind from an official entity before an organization is allowed to produce DNA in kilo quantities.

- *Product registration*. It is possible to envision a system where someone requesting kilo-scale custom synthesis of DNA is required to register with an organization describing its potential use. This system may create a barrier for the synthesis of DNA for harmful applications.

Conclusion

It would not be an easy matter to restrict the supply of the reagents needed for DNA synthesis to such an extent as to prevent a motivated individual from making oligonucleotides at a small scale. As noted above, the least implausible option for tracking and restriction would seem to be solid support beads. However, since these are widely used by the legitimate DNA synthesis industry, the restrictions must also include protocols for monitoring reagent use within a company and reporting their disposition. Several of the companies making and using these reagents reside outside the USA, complicating the task of imposing effective tracking policies.

Glossary

AMIDITE (also known as phosphoramidite): This is a protected version of a **nucleoside** that is easily activated for the coupling reaction. The P atom, which will eventually form part of the phosphate backbone, is protected with β-cyanoethyl and diisopropylamine groups. In the first stage of the coupling reaction, a weak acid protonates the nitrogen atom of the diisopropylamine protecting group, causing it to become positively charged and making it into a good leaving group. This allows nucleophilic attack by the free 5' hydroxyl group of the bead-attached monomer on the phosphorous atom, forming the molecule referred to as **4** in Figure 4.

Different protecting groups are also attached to the amines ($-NH_2$) that are not part of a ring in the bases A, G and C, to prevent them from becoming protonated and causing unwanted reactions. These protecting groups, and the cyanoethyl protecting group on the phosphate, remain on the growing chain until it is finally released from the bead.

ANHYDROUS: Water-free. Because the **amidite** chemistry depends on the hydroxyl group of the bead-attached monomer performing a nucleophilic attack on the positively charged diisopropylamine group, any other nucleophiles in the solution will reduce the efficiency of the coupling reaction. Water can act as a nucleophile, and must be rigorously excluded from the reaction.

BASE: The structures of the bases are shown in Figure 1. The information content of a DNA molecule consists of the linear arrangement of the bases A, T, G and C along a phosphate/sugar backbone (also shown in Figure 1). It is the pairing of the bases, A with T and G with C, that allows DNA to be copied.

2-DEOXY-D-RIBOSE: The particular form of sugar that is used in DNA synthesis. The D in DNA stands for "deoxy"; this refers to the fact that carbon number 2 in the ribose ring does not carry an oxygen in the DNA structure. In RNA, the sugar used is ribose, not deoxyribose.

NUCLEIC ACID: A polymer of nucleotide subunits.

NUCLEOTIDES: A base connected to a sugar (ribose) ring and one or more phosphate groups. In the DNA structure, the nucleotides have one phosphate group each, which forms part of the backbone of the DNA. Each phosphate group is linked to two sugar groups, through two different oxygen atoms (see Figure 1).

NUCLEOSIDE: The structures of the four nucleosides relevant to DNA synthesis are shown in Figure 3. A phosphate group must be added to nucleosides before they can be linked together to form DNA. In amidite chemistry, the phosphate group is formed by oxidation after the coupling step (see Figure 5).

OLIGONUCLEOTIDE: A short stretch of DNA (for example, 20 nucleotide subunits linked together).

PURINE: A base in which the **pyrimidine** ring is fused to a second ring, the imidazole ring. Imidazole is a five-membered aromatic ring with two nitrogens.

PYRIMIDINE: A base consisting of a six-membered aromatic ring with two nitrogen atoms at positions 1 and 3.