

~~~ What The Heck is PCR? ~~~

Polymerase chain reaction (PCR) is a technique which is used to amplify the number of copies of a specific region of DNA, in order to produce enough DNA to be adequately tested. This technique can be used to identify with a very high-probability, disease-causing viruses and/or bacteria, a deceased person, or a criminal suspect.

In order to use PCR, one must already know the exact sequences which flank (lie on either side of) both ends of a given region of interest in DNA (may be a gene or any sequence). One need *not* know the DNA sequence in-between. The building-block sequences (nucleotide sequences) of many of the genes and flanking regions of genes of many different organisms are known. We also know that *the DNA of different organisms is different* (while some genes may be the same, or very similar among organisms, there will *always* be genes whose DNA sequences differ among different organisms - otherwise, would be the *same* organism (e.g., same virus, same bacterium, an identical twin; therefore, **by identifying the genes which are different, and therefore unique**, one can use this information to identify an organism).

A gene's building-block sequence is the *precise* order of appearance, one after the other, of 4 different components (deoxyribonucleotides) within a stretch of DNA (deoxyribonucleic acid). The 4 components are: Adenine, Thymidine, Cytosine and Guanine, abbreviated as: A, T, C and G, respectively (a 4-letter alphabet). The arrangement of the letters (one after the other) of this 4-letter alphabet generates a "sentence" (a gene sequence). The number of letters in the sentence may be relatively few, or relatively many, depending on the gene. If the sentence is 1000 letters-long, the sequence would be said to be 1 kilobase (1000 bases).

As an example:

ATATCGGGTTAACCCGGTATGTACGCTA would represent part of one gene. DNA is double-stranded (except in some viruses), and the two strands pair with one another in a very precise way. **EACH** letter in a strand will pair with only one kind of letter across from it in the opposing strand: **A ALWAYS pairs with T; and, C ALWAYS pairs with G across the two strands.**

So:

TTAACGGGGCCCTTTAAA.....TTTAAACCCGGGTTT

Would pair with:

AATTGCCCGGGAAATTT.....AAATTTGGGCCCAA

Now, let's say that the above sequences "flank" (are on either end of..) the gene, which includes a long stretch of letters designated as:

These are known, absolutely identified to be, the sequence of letters which **ONLY** flank a particular region of a particular organism's DNA, and **NO OTHER ORGANISM'S** DNA. This region would be a target sequence for PCR.

The first step for PCR would be to synthesize "primers" of about 20 letters-long, using each of the 4 letters, and a machine which can link the letters together in the order desired - this step is easily done, by adding one letter-at-a-time to the machine (DNA synthesizer). In this example, the primers we wish to make will be exactly the same as the flanking sequences shown above. We make **ONE** primer exactly like the lower left-hand sequence, and **ONE** primer exactly like the upper right-hand sequence, to generate:

TTAACGGGGCCCTTTAAA.....TTTAAACCCGGGTTT

AATTGCCCGGGAAATTT.....>

and:

<.....TTTAAACCCGGGTTT

AATTGCCCGGGAAATTT.....AAATTTGGGCCCAA

Now, the may be a very long set of letters in-between; doesn't matter. If you look at this arrangement, you can see that if the lower left-hand primer sequence (*italics*) paired to the upper strand could be extended to the right in the direction of the arrow, and the upper right-hand sequence paired to the lower strand could be extended to the left in the direction of the arrow (remembering that the also represent letters, and opposite pairing will ALWAYS be A to T and C to G), one could successfully exactly duplicate the original gene's **entire sequence**. Now there would be four strands, where originally there were only two. If one leaves everything in there, and repeats the procedure, now there will be eight strands, do again - now 16, etc.. therefore, about 20 cycles will theoretically produce approximately one-million copies of the original sequences (2 raised to the 20th power).

Thus, with this amplification potential, **there is enough DNA in one-tenth of one-millionth of a liter (0.1 microliter) of human saliva (contains a small number of shed epithelial cells), to use the PCR system to identify a genetic sequence as having come from a human being!** Consequently, only a **very** tiny amount of an organism's DNA need be available originally. Enough DNA is present in an insect trapped within 80 million year-old amber (fossilized pine resin) to amplify by this technique! Scientists have used primers which represent present-day insect's DNA, to do these amplifications.

Here is how PCR is performed:

First step: unknown DNA is heated, which causes the paired strands to separate (single strands now accessible to primers).

Second step: add large excess of primers relative to the amount of DNA being amplified, and cool the reaction mixture to allow double-strands to form again (because of the large excess of primers, the two strands will always bind to the primers, instead of with each other).

Third step: to a mixture of all 4 individual letters (deoxyribonucleotides), add an enzyme which can "read" the opposing strand's "sentence" and extend the primer's "sentence" by "hooking" letters together in the order in which they pair across from one another - A:T and C:G. This particular enzyme is called a DNA polymerase (because makes DNA polymers). One such enzyme used in PCR is called *Taq* polymerase (originally isolated from a bacterium that can live in hot springs - therefore, can withstand the high temperature necessary for DNA-strand separation, and can be left in the reaction). Now, we have the enzyme synthesizing new DNA in opposite directions - BUT ONLY THIS PARTICULAR REGION OF DNA.

After one cycle, add more primers, add 4-letter mixture, and repeat the cycle. The primers will bind to the "old" sequences as well as to the newly-synthesized sequences. The enzyme will again extend primer sentences ... Finally, there will be PLENTY of DNA - and ALL OF IT will be copies of just this particular region. Therefore, by using different primers which represent flanking regions of different genes of various organisms in SEPARATE experiments, one can determine if in fact, any DNA has been amplified. If it has not, then the primers did not bind to the DNA of the sample, and it is therefore highly unlikely that the DNA of an organism which a given set of primers represents, is present. On the other hand, appearance of DNA by PCR will allow precise identification of the source of the amplified material.

Copyright John C. Brown, 1995