

Lecture 1

Introduction to Nucleic Acids

Today we're going to cover a few "orientation" issues, discuss some of the course expectations, and even learn a few things about DNA.

*Lets talk
about the
structure of
DNA and
RNA*

Warming up the brain: Nucleic acids are made up of nucleotides, consisting of **bases (purines and pyrimidines)**, as you probably recall from your genetics or cell biology class, **sugars (ribose or deoxyribose)**, and a **phosphate backbone**.

Remember that we have some rules, called "Watson-Crick" base pairing, by which adenylate nucleotides can **hydrogen bond** to thymidylate nucleotides (or uridylate in RNA), while guanylate nucleotides hydrogen bond to cytidylate nucleotides.

C pairs with G
A pairs with T (or U)

Is this all starting to come back to you now? Let's find out.

... about the bases

	Which of these are purine bases?
	A (adenine)
	C (cytosine)
	T (thymidine)
	G (guanine)

Stop me if you've heard this one...

A guy walks into a bar and says "My name's Chargaff, and 22% of my DNA is "A" nucleotides. I'll bet anyone that they can't guess what percentage of my DNA is "C" nucleotides!" You say "I'm thirsty, so I'll take that bet!" and then you think...

Guess right and win the drink.

22% C

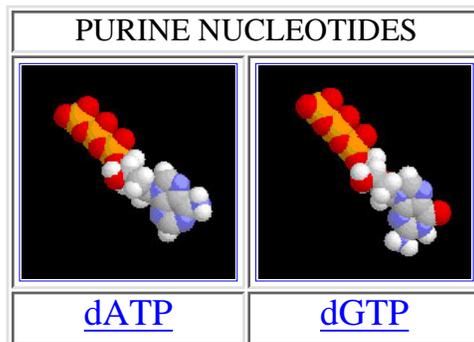
44% C

28% C

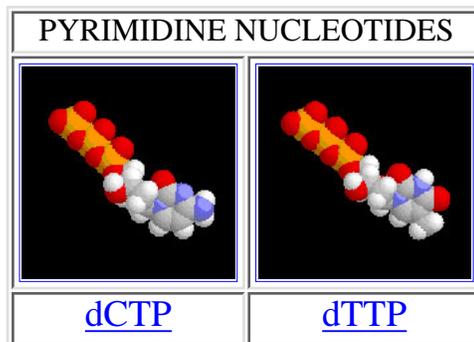
Wait a minute - it can't be done!

Yes, it can be done! Erwin explains, we have double stranded DNA genomes, so if 22% is "A", then there must also be 22% "T", because every "A" base will be paired with a "T" base. You with me? So $22\% + 22\% = 44\%$ is the percentage of the DNA that is either "A" or "T". That implies that the percentage that is "G" or "C" must be whatever is left, or $100\% - 44\% = 56\%$. Every "G" must be base paired with a "C" and every "C" must be base paired with a "G", so exactly half of that 56% must be "C" bases. That is, 28% are "C" bases and 28% are "G" bases.

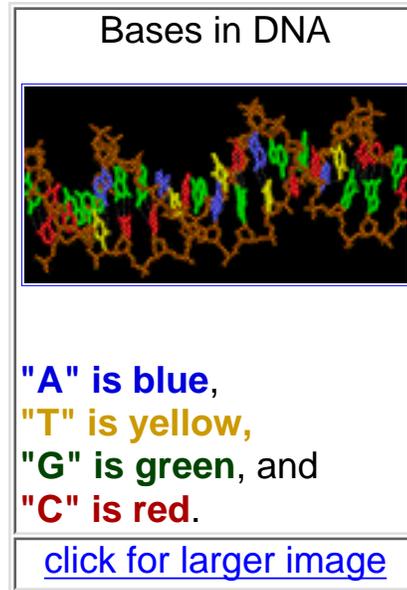
Here's a photo gallery (Click for larger images):



red = oxygen, blue = nitrogen, white = hydrogen, gray = carbon.
What atom does the amber color represent?



The nucleotide bases make up the core of the double helix, as you can see in the picture below.



This snapshot comes from a site you'll probably want to investigate, an "[Interactive Animated Nonlinear Tutorial](#)" by Eric Martz, from the Department of Microbiology at the University of Massachusetts-Amherst.

<http://www.umass.edu/microbio/chime/dna/index.htm>

Here's another good site to visit, to learn about the [Chime plug-in](#), and to study the overall structure of DNA. Chime is pronounced with a hard "K" sound as in "kind", not a "Ch" sound as in "chair."

http://www.umass.edu/microbio/chime/dna/fs_pairs.htm

You can develop a real "feel" for molecules if you familiarize yourself with the shareware [RasMol](#) (RasMac) program. With this program, you can inspect crystallographic structures downloaded from Brookhaven National Labs, turning the molecules on the screen so you can see them from every side and angle. [Downloading instructions](#) are available on the Web, as are instructions for [finding molecules to play with](#).

If you have the Chime plug-in working, you may be able to see the following two examples, generated by GLACTONE (<http://chemistry.gsu.edu/glactone/>). You may also download them directly and use RasMol.

An [AT base pair](#)

A GC base pair

How does hydrogen bonding come to pass?



Well, suppose this is a cherry, and you're going to make chocolate cupcakes with cherries on top. You make the cake mix, fill the little cupcake holders and bake the cupcakes. Then you put a cherry on top of each, and whip up a batch of chocolate icing. Here is one, ready to cover with frosting!



Here's one that was covered well, in fact it was so evenly covered with frosting that you can no longer see the cherry!



Then, an interesting thing happens. On some of the cupcakes, the chocolate icing is very thin. It dribbles down onto the cake, leaving the cherry somewhat visible through the frosting.



It is almost as if the cupcake and the cherry are fighting for the frosting, and the cupcake is winning!

In fact, sometimes the frosting gets so thin, that there's nothing left to hold the cherry in place, so it pops out, leaving the frosting still stuck to the cake.



Hmmm... What does this make us think of? Why polar covalent bonds, of course!

You see, some atoms are more electronegative than others. Oxygen is more electronegative than hydrogen, so in an -OH group, the oxygen takes more than its fair share of electrons. That's just like the cupcake taking more than its fair share of frosting. The electrons get very thinly distributed over the hydrogen and get more thickly distributed over the oxygen.

That gives a partial negative charge to the oxygen and a partial positive charge to the hydrogen. Why? Because the electron is charged, and if more of it is distributed in one place, that place will get a bit of charge.

Nitrogen can play the same trick, because it is also more electronegative than hydrogen. On the other hand, carbon and

hydrogen are about the same in electronegativity, so they share the electrons pretty fairly. There will not be a partial charge on the carbon, because the electrons are distributed evenly in the bond. The carbon-hydrogen bond reminds us of the well-frosted cake - all neutrally distributed:



On the other hand, the oxygen-hydrogen and nitrogen-hydrogen bonds remind us of the thinly-frosted cake, and the thin frosting leads to a "dipole moment", or partial charge:

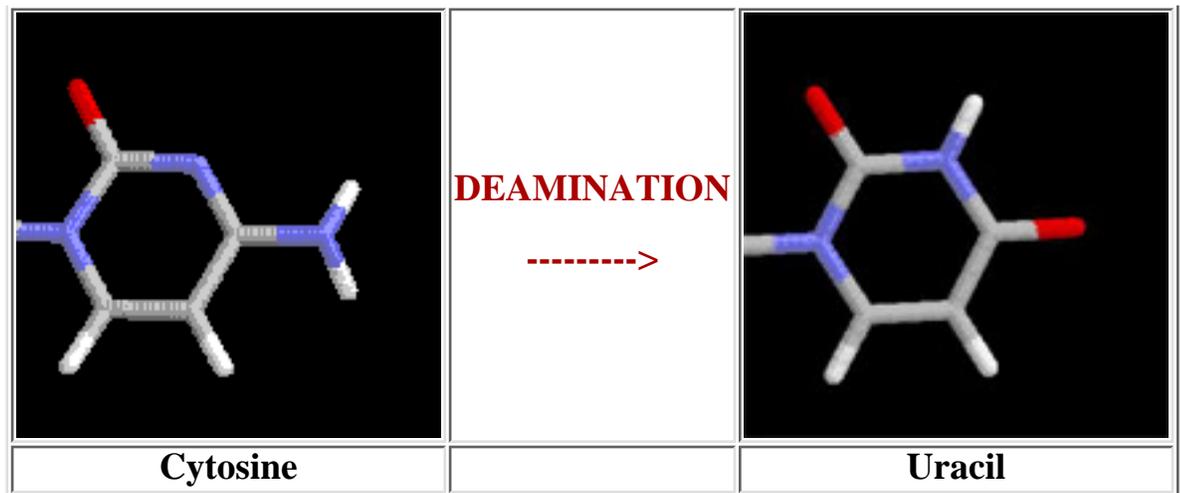


What's the difference between DNA and RNA?

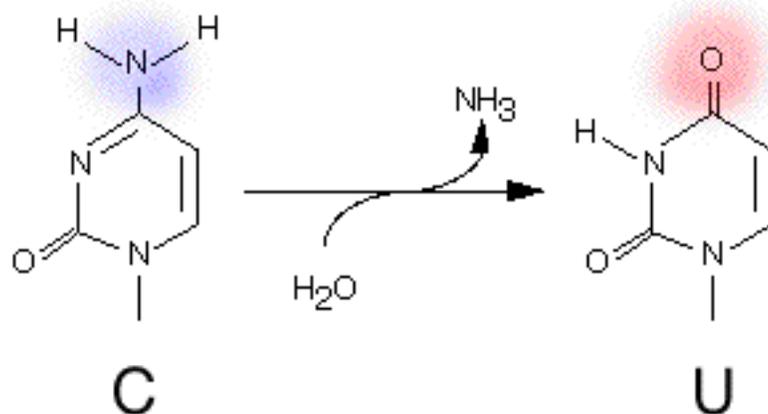
DNA contains the sugar deoxyribose while RNA is made with the sugar ribose. It's just a matter of a single 2' hydroxyl, which deoxyribose doesn't have, and ribose does have. Of course, you all remember that RNA uses the base uracil instead of thymine too.

Cytosine naturally has a high rate of deamination to give uracil

Cytosine deamination (i.e. water attacks!)



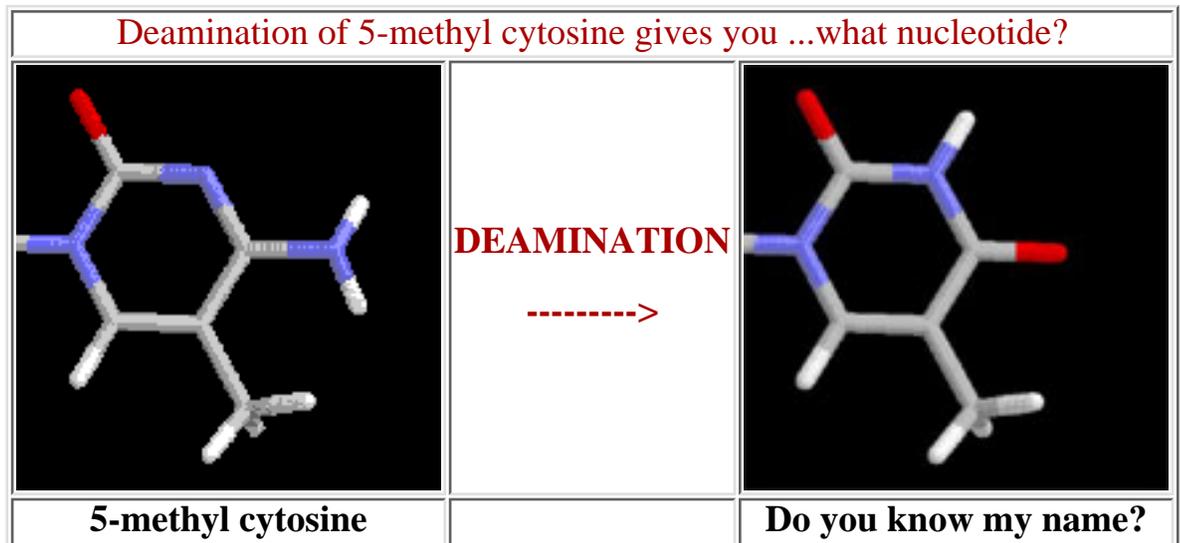
Uracil in the DNA is a big no no, and there are specific enzymes called uracil N-glycosylases (from the gene called *ung*, about which we'll have much more to say in a later lecture) that excises the offending deoxyuridylate nucleotide so that it can be replaced. If the uracil had arisen by deamination, then what will be the nucleotide base across from it? There will be a G nucleotide across from it, if the mutation just occurred. That's because the G was paired with the C that deaminated to a U. On the other hand, if there is a round of DNA replication before the uracil N-glycosylase arrives on the scene, then there will be an A nucleotide across from the U. That's because the U will have had a chance to be a template in DNA replication, and U base pairs to A, right?



If you're an organism that doesn't want to end up looking like a [Teenage Mutant Ninja Turtle](#) (who as you may recall, were suffering from the effects of a "retromutagen" that made them behave like adolescent boys), then you should keep a sharp eye out for deoxyuridylate nucleotides. The dU should be excised rapidly and replaced with a C, so that these deamination events do not become "fixed" as a mutation.

Some types of mutations change a pyrimidine to a different pyrimidine, or a purine to a different purine. We call these **transition mutations**. If a purine is mutated to a pyrimidine, then it is a **transversion mutation**. So, for example, a mutation of A to T or C to A would be what? Right! A transversion, and a mutation of A to G or T to G would be a transition.

Sometimes deoxycytosine is methylated on its "5 position," so what would happen to the coding content of deoxy-5-methyl-cytosine if it were unlucky enough to be naturally deaminated?



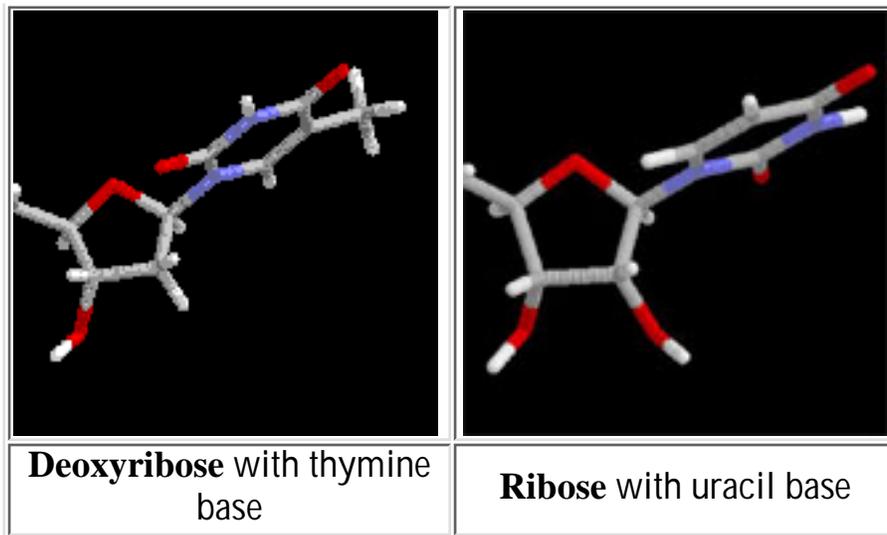
So you see the problem...the 5-methyl cytosine is deaminated to thymidine. The new thymidine looks like any other thymidine - it's a mutation! A transition mutation, because it is a pyrimidine changed to another pyrimidine.

Perhaps that is why there are so few CG dinucleotides in mammalian genomes. CG dinucleotides are frequently methylated on the C base, so CG may frequently mutate to TG, leaving CG "under represented". In fact, CG dinucleotides are sometimes associated with regulatory regions of genes, and we call them "CG islands" because they are so rare.

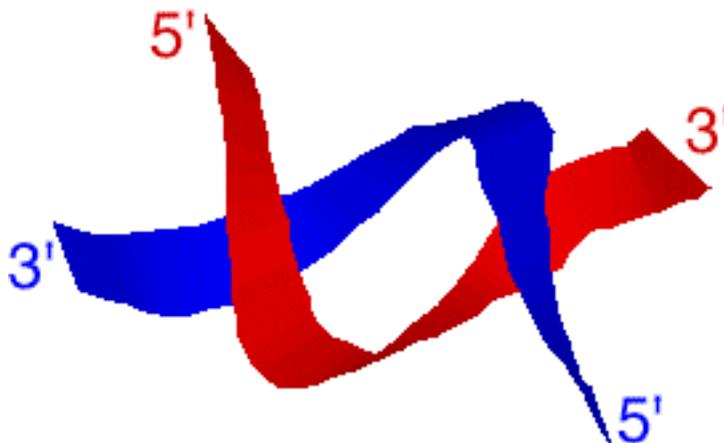
... about the sugars

Now let's look at the sugar component of nucleic acids. Remember that ribose and which is deoxyribose?

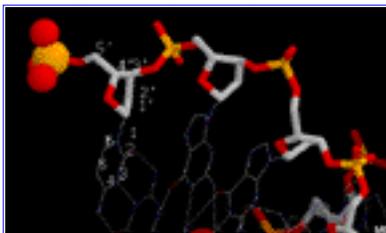
DNA vs. RNA sugars



There is a **5' end** and a **3' end** to a nucleic acid. The 5' end frequently has a phosphate attached, while the 3' end is typically a hydroxyl group. A single strand of DNA has a "*polarity*" or "*directionality*." It isn't like a piece of string, in which you cannot distinguish one end from the other.

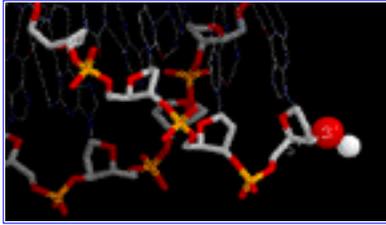


Study the phosphate at the 5' end



[click for larger image](#)

Study the hydroxyl at the 3' end

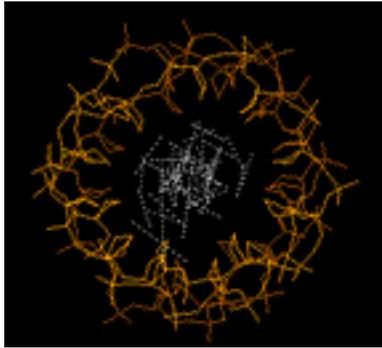


[click for larger image](#)

**Here are
some
exercises:**

1. Go to the site, <http://www.umass.edu/microbio/chime/dna/dnabone.htm> and try erasing each of these components in turn: The sugars, the bases, the H-bonds.

A Puzzle: What is this a picture of, and what is missing?



2. Go to this site, http://www.umass.edu/microbio/chime/dna/fs_ends.htm and click on the 5' end and 3' end keys to study the antiparallel structure

3' Go to this site, <http://www.umass.edu/microbio/chime/dna/codons.htm> and study the coding structure of DNA - the difference between codon and anticodon in RNA.

**We used the
word
"polarity" in
discussing
DNA strands.**

One end of a DNA strand will have a free 5' group on the deoxyribose sugar (perhaps with a phosphate group attached), and the other end will have a free 3' end (probably with a hydroxyl group). If we abbreviate a DNA strand as a sequence of letters GGAGATTCACCAACT, we need to know which end is the 5' end and which is the 3' end. The two possible versions, 5'-GGAGATTCACCAACT-3' and 3'-GGAGATTCACCAACT-5' are completely different!

Here's an important convention we will use: If a sequence is written with no notation of 5' and 3' ends, we will assume that the left end is 5' and the right end is 3'.

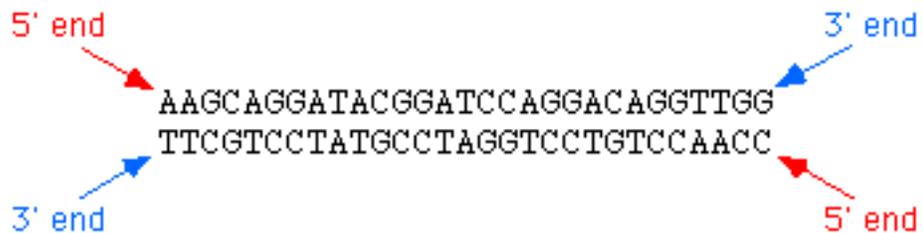
Thus:

AGGTCAGAGGATCAAGAGACCACCTCAGAAGAGCAAC

will mean:

5'-AGGTCAGAGGATCAAGAGACCACCTCAGAAGAGCAAC-3'

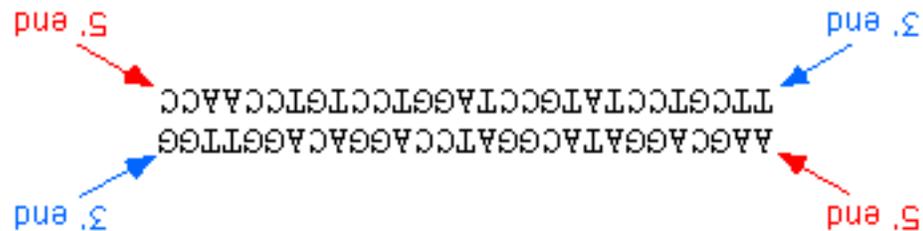
We will make a similar convention for double-stranded molecules, with 5' ends at the upper left and lower right, and 3' ends at the upper right and lower left. Of course, the two 5' ends are opposite each other because the polarities of the strands are opposing in a double helix. We say that double-stranded DNA is made up of two "antiparallel" strands.



These conventions are important because they simplify the language of molecular biology. Remember...In a contiguous sequence of DNA, the 5' end of the top strand will be on the left while the 5' end of the bottom strand will be on the right (unless a notation indicates otherwise). We need to find out how robust this convention is:

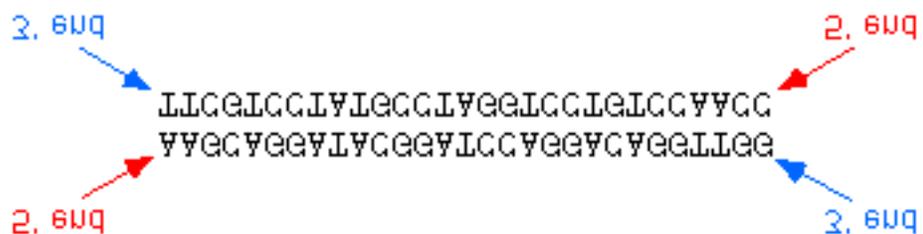
It doesn't matter! The 5' ends are still in the upper left and lower right.

Does it matter if we rotate the molecule?

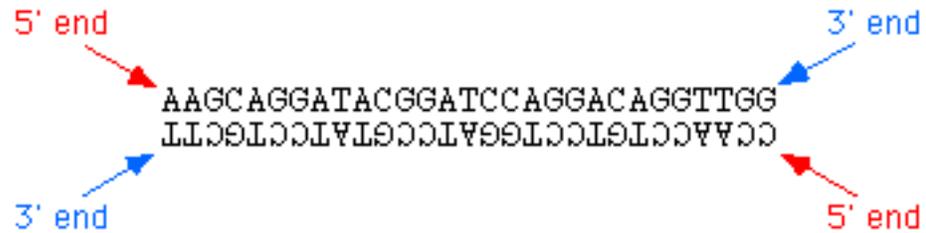


Does it matter if we flip the molecule through a mirror image?

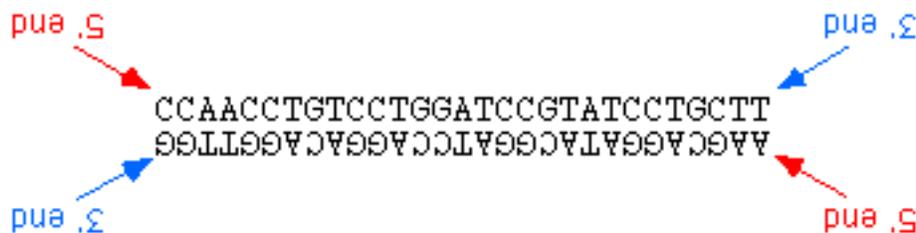
Yes! That ruins the convention because the 3' end is now in the upper left. We couldn't do that unless we make a notation on the paper indicating where the 5' and 3' ends are.



Aida's favorite Here's a slightly different version of the sequence writing convention. My wife will encourage all of you to write your molecules with an upside-down bottom strand, like this...



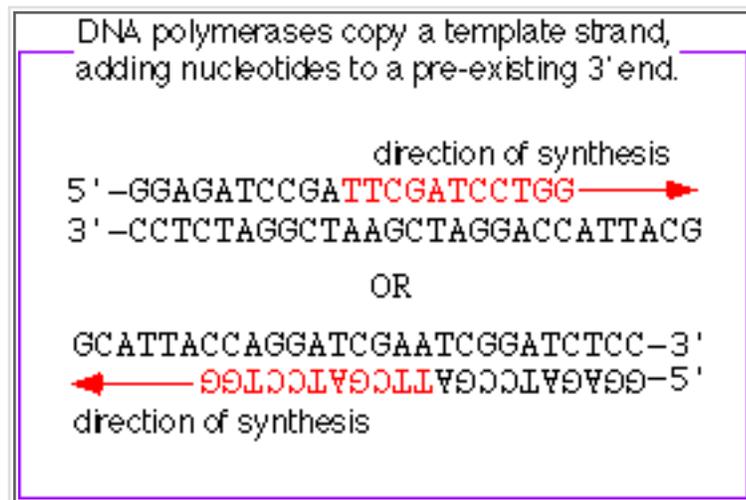
...because then you can rotate the paper and the molecule still looks pretty much the same.



If this upside-down bottom strand makes more sense to you than the other way, feel free to write your sequences this way.

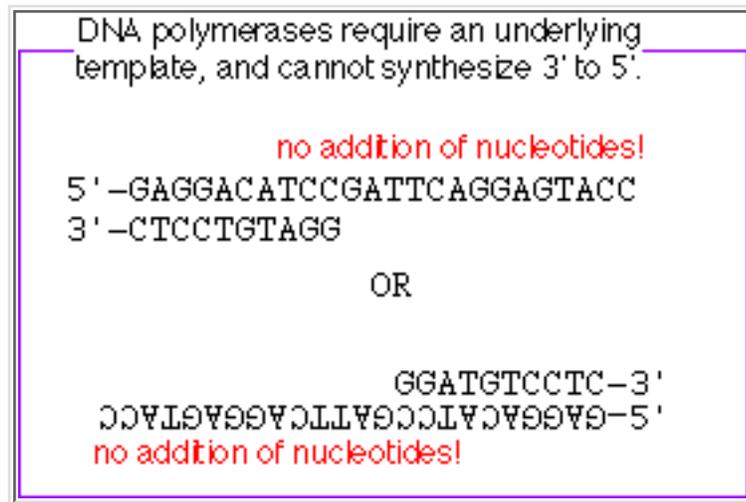
A few general observations about DNA polymerases

DNA polymerases copy a template strand, adding nucleotides to a pre-existing 3' end.



*It all comes
down to three
things:
Primer,
template, and
substrate*

DNA polymerases require an underlying template (and a primer) and cannot synthesize in the direction 3' to 5'. That is, they cannot add nucleotides to a free 5' end.



*Error
correction*

A DNA polymerase enzyme packing a 3' to 5' exonucleolytic activity (in addition to its 5' to 3' synthetic activity) can go back and forth on a template, laying down newly-synthesized DNA (as it moves "to the right" or 5' to 3') and then chewing it again (as it moves "to the left" or 3' to 5'). What a waste you say? Actually this is a good feature. The proof-reading ability of the enzyme is improved if it can easily back up and correct errors.

There are two DNA polymerase enzymes that are commonly used that have only the 5' to 3' synthetic activity and the 3' to 5' exonuclease activity:

- Klenow enzyme (a subfragment of E. coli DNA polymerase)
- T4 DNA polymerase (derived from bacteriophage T4)

These enzymes lack a 5' to 3' exonuclease activity (which we will discuss at a later point in the course). In comparing the two, T4 DNA polymerase enzyme has a 3' to 5' exonuclease activity that is stronger than its counterpart in Klenow enzyme.

Practical application

Suppose we started with a double stranded DNA having an end shown on the right

```
5' ----ATCTCCCGAAAGAATCCGCATACCAGGAAGGGCGC - 3'
3' ----TAGAGGGCTTTCTTAGGCGTATGGTCCTTCC - 5'
```

If we add the enzyme T4 DNA polymerase, along with the four deoxynucleoside triphosphate substrates dGTP, dATP, dTTP, and dCTP, the following sequence of events may happen:

The exonuclease activity removes nucleotides from the 3' end, and may even digest a bit into the double-stranded region.

```
5' ----ATCTCCCGAAAGAATCCGCATACC - 3'
3' ----TAGAGGGCTTTCTTAGGCGTATGGTCCTTCC - 5'
```

Any digestion into the double-stranded region will be repaired by the synthetic activity of the enzyme, using the four dNTP substrates.

```
5' ----ATCTCCCGAAAGAATCCGCATACCAGGAAGG - 3'
3' ----TAGAGGGCTTTCTTAGGCGTATGGTCCTTCC - 5'
```

(new synthesis shown in red)

In the end...

Note that the final product will have a "**blunt end**," provided that we haven't forgotten to add the four dNTP substrates to the reaction. Forward synthesis is favored when the substrates are present, so the exonuclease activity doesn't destroy our fragment. Its destructive work is well compensated by the synthetic activity of the same enzyme.

One more time: If the enzyme has a primer and template, forward synthesis is favored when the substrates are present.

Synthesize? Degrade? Sit and wait? How does an enzyme like DNA polymerase Klenow Fragment know what to do next? Well, there are some general rules of conduct that these enzymes learn in school, and you can learn them too.

General rules of conduct for Klenow and T4 DNA polymerases

1. Remember your base pairing rules: G goes with C and A goes with T.
2. The 5' ends are strictly off limits, unless you have your holoenzyme license (and for your information, you don't!)
3. There will be no synthesis without a free 3' end, unless you have your RNA polymerase license (and for your information, you don't!)
4. There will be no degradation without a free 3' end, unless you have

your endonuclease license (and for your information, you don't!)

5. There will be no synthesis without an underlying template, unless you have your terminal transferase license (and for your information, you don't!). Excess nucleotide substrates is NOT accepted as an excuse for untemplated additions to the 3' end.

6. Under no circumstances may you make a synthetic addition to the 5' end (even holoenzymes are not permitted to do that!). Having a template or substrate available is not an excuse for 3' to 5' synthesis.

7. There will be no reconstruction of a broken phosphodiester bond, unless you have your ligase license (and for your information, you don't!). If you are synthesizing DNA and run into an obstruction on your template, you must stop and leave the nick unrepaired. You may not excise the 5' nucleotide that is obstructing your path (see rule 2).

8. If you have no remaining template, then you must excise the nucleotide at the 3' end (and don't be tempted to break rule 5!). (repeat rule 8 until it does not apply).

9. If you have been provided with a free 3' end, a template, and a substrate molecule that is correct, you must add that nucleotide to the growing end of the strand (i.e. to the 3' end.)

10. If you have a free 3' end and a template, but after waiting for the appropriate number of milliseconds you are still missing the appropriate nucleotide substrate for the next synthetic step, you may go back and remove the one preceding nucleotide. Either of rules 9 or 10 may apply thereafter.

Special note to CSUN Students.

So like if you grew up in the Valley (Shut up! Oh, my god?) and are like totally clueless cuz I make no sense?, go to the [Valley URL site](#). Buffy and Tiffany will translate any site into valleyspeak. Just type in this URL code when you get there, and you'll like TOTALLY understand everything.



So like, here is Tiffany's testimonial from the site...

A testimonial -- Tiffany Johnson, Encino, California

So like I'm sitting in the mall the other day and like my friend Buffy says to me "Hey, Tiff" (my name is Tiffany, but, like, my friends call me Tiff), "have you seen this mega-huge computer thingy called the Internet?" Helloooo, who does she think she's talking to -- Miss Uninformed? It's not like *I* wasn't the one who told her all about the time I caught HER ex-boyfriend Brian making out with that heifer cheerleader last week? Anyway, I told her how I had gone over to Jeff's house (he's, like, a total stud puppy) and asked him to show me this web stuff? I was like "oh my gawd, only a total dweeb would like this" because none of this stuff made ANY sense? Everyone was, like, TOTALLY talking in languages that made no sense... until I found "Valley URL" by those bitchin' '80s Server guys. These guys are like WAY cool!!!

Stan Metzenberg
Department of Biology
California State University Northridge
Northridge CA 91330-8303
stan.metzenberg@csun.edu

© 1996, 1997, 1998, 1999, 2000, 2001, 2002