BBT204/MIC209: Molecular Biology 1
Lecture note 1

Replication
Part 1:

THE BASICS OF THE DNA STRUCTURE:

- **A PAIRS WITH T, G PAIRS WITH C, OR VICE VERSA.**
- **AND THE FUNDAMENTAL REASON TO HAVE DNA REPLICATION OF THE CELLS TO DUPLICATE THE CELLS' GENE.**
  THAT'S THE SINGLE MOST IMPORTANT REASON FOR HAVING DNA REPLICATION.

But certainly, all fundamental events of a cell are controlled by the cell's genome. So obviously, you want this process to work well and to be very accurate.

DNA replication:
Synthesising a DNA from another DNA.

Does Replication process makes mistakes?

replication process is if you do make mistakes, there's often consequences.

There are potentially ***good consequences***: You can develop a new attribute that is good for your survival. And this happens with some frequency.

But more often than not, there are ***bad consequences***. And that can be anything from straight up lethality-- the organism can no longer grow-- to things like cancer or a new attribute that actually makes an organism less able to grow.

And because more generally the outcome is the latter than the former, again, you'd like this to be as accurate as possible.

And we'll talk about how it becomes accurate as we go along.

Properties of DNA replication:

1. RAPID: 100-1000 bp/sec

2. COMPLETE: ESSENTIALLY EVERY BP MUST BE REPLICATED IN THE GENOME

We'll talk about telomere replication at some point. That's one example where not every base pair is replicated. And the final thing is it's extremely accurate.

3. ACCURATE: 1 MISTAKE IN 10^{10} BP SYNTHESIS

There's two ways to think about this:
E. coli has a genome of about $5X\ 10^6$ base pairs.
And that means it's going to make a mistake about one out of every roughly 1,000 organisms will have a mistake in its genome.

It's a little more scary if you're human, right?
So we have a genome of $3X\ 10^9$.
So that means we make a mistake about one every three cell divisions.
But it's worth keeping in mind, most of those 10^{10} th base pairs are not coding and are not that important to be correct.
But if it's in the middle of a coding region, it changes the amino acid and the protein, it can be a problem.

So we're pretty much full of errors is one way to think about it.
Even under the best of circumstances, we have thousands and thousands of errors.

Who knows how much DNA is in each one of your cells?

How long is it? You think it's 2 inches? 2 meters.

How much in your whole body?
It's enough to go to the Sun. It's about 90 million miles.

How much do you synthesize in your lifetime?
It's a light year.

So pretty impressive machinery that can do that and not with us being sort of a bubbling mass of mutant cells that just sort of sits around like a pile of jello.

So this is an incredibly important, very accurate process that makes sure that we stay who we are throughout our lifetime.

The studies of replication led to some very important things that all of you probably know about.

1. molecular cloning, the fact that we can splice genes together or make all sorts of different things directly came from studies of DNA replication.
2. PCR directly came from that.
3. all of this massive sequencing. The fact that we can sequence entire genomes, including our own, is a direct result of studying DNA replication.
4. And then the other side of it is that DNA replication is very frequently the target of drugs.
So most frontline chemotherapies target DNA replication.

There's lots of fancy new ones that lots of people are developing, and hopefully will work much better than the DNA replication ones because, unfortunately, they typically not only make your cancer cells sick, but they make all your other cells sick.
But if you get cancer, the odds of you not being treated with a DNA replication inhibitor is extremely small, because most of these fancy new things are used either in collaboration with those or are not used for the particular type of tumor you have, because they mostly are targeting very small numbers of tumors that can be gotten.
And the other thing is most antivirals target DNA replications.
So for example, there is one of the best anti-AIDS drug targets the DNA replication of its genome or RNA slash DNA replication of its genome. So these are very important targets. And so understanding replication helps us to hopefully develop new targets. And if nothing else, maybe make some that won't make us quite so sick when we treat them.

Answer 2nd Part

BBT204/MIC209: Molecular Biology 1
Lecture note 1

The chemistry of DNA replication.

SLIDES

So this just shows the substrate for DNA replication, which I will call a primer: template junction.

OK, the primer has a free 3 prime OH hydroxyl group.

And the catalysis works simply by having that OH attack the alpha phosphate of an incoming deoxynucleotide triphosphate, releasing a pyrophosphate.

And this only occurs when this base correctly base pairs with the base on the template.

I want to start with one, which is, why does it go forward?

OK, so why do we use triphosphates instead of, for example, diphosphates to do this reaction?
OK, so all of the precursors that we'll talk about will be triphosphates.

And there's an important reason for that, which is that this second reaction of cleaving this pyrophosphate-- which is two phosphates coupled to one another-- into two phosphates, drives the reaction such that it's essentially irreversible.

THE REACTION WITH THE ACTION OF PYROPHOSPHOTASE HAS A ΔG = - 7 KCALS. (IRREVERSIBLE)
BUT WITHOUT, PYROPHOSPHOTASE ΔG = -3.5 KCAL.

And while this would still be a forward reaction, this reaction is essentially irreversible.

So the cell makes sure that this reaction goes forward by not only having it release the pyrophosphate, but very, very rapidly taking that pyrophosphate and degrading it into two phosphates.

DNA POLYMERASE : CATALYZES THE DNA SYNTHESIS

primer template junction (PTJ)

1. REQUIRE 3' OH PRIMER

2. ANNEALED TO A LONGER SINGLE STRANDED (SS) DNA (w/ss DNA ADJ TO 3' OH PRIMER)

3. ALL 4 dNTP

4.ADDED dNTP MUST BP WITH TEMPLATE FOR CATALYSIS

5. 3' OH PRIMER IS EXTENDED BY DNA SYNTHESIS

Where is 5 prime and 3 prime up here?
Q3

PART 3
DNA POLYMERASE ACTIVE SITE

How many reactions can a DNA polymerase catalyze?
SLIDE
How many active sites do you think DNA polymerase has?
You think it has four?
You think it has one.

HOW DOES ONE ACTIVE SITE CATALYZE ADDITION OF MULTIPLE dNTPS?

- **BASE PAIRS HAVE SAME DIMENTIONS**
- **CORRECT BASE PAIRING PLACES 3'OH PROXIMAL TO ALPHA-PHOSPHATE OF THE dNTP**

PART 4
Structure of DNA Polymerase:

Slide

So what does this thing look like?

So there are lots of DNA polymerases. For the most part, they all share this same basic architecture, which the crystallographers-- who are very creative people-- have called a "hand."

And if you see this, you think, "Wow. They're right.

It looks just like a hand."
But this isn't what a real structure looks like. This is actually one of the simpler DNA polymerase structures, and it looks like that.

And now, you see why they're very creative people.

What they call the "fingers" is over here.The thumb here. The palm here.
So it's sort of hand-like. You could say it's a "claw." They say it's right-handed, not left-handed.

There's an argument for why they say that, and that's because this-- as I'll tell you in a moment-- moves and this stays still during the catalysis reaction.
And so that's sort of why they call it the "fingers"--because our fingers move a lot more than our thumb does. And in that case, it's definitely right-handed.

WHERE ARE THE VARIOUS PARTS OF THE DNA BINDING?
- **PLAM BINDS PTH AND NEWLY SYNTHESIZED DNA**

slide

So over here, you have the newly-synthesized DNA, the primer: template junction.
One interesting attribute of this structure and the structure of all polymerasesis you can see this is the template immediately after the active site.

So there's one base unmatched here. And then the rest of the single-stranded DNA, it has a 45-degree bend in the DNA after that. And that's important because that means the only nucleotide that's actually around the active site to be base-paired with is the next nucleotide after the last one added.

But the palm is interacting with this. And again, there's all this other stuffthat we're not going to go into detail, but the palm is considered this region right down here.

So this sort of creates an interesting question, which is--
HOW DOES A PROTEIN BIND ALL DNA?

Because we want to bind that sequence. Whether it's GATC, whether it's AGAA, whether it's GGGG, it needs to bind to that.

SO WHAT PART OF THE DNA WOULD IT BE BINDING?

THE BACKBONE - THE PHOSPHATE BACKBONE, WHICH DOES NOT CHANGE, REGARDLESS OF SEQUENCE.

But what if I told you its primary interactions are actually not with the backbone?

What would it be binding then? If you could be bind-- what's in the major groove?
What kind of information? If you want a sequence-specific DNA binding protein to recognize DNA, do you know what part of the DNA it binds?

SO IT BINDS THE MAJOR GROOVE.

SLIDE:

The major groove has these very interesting patterns of hydrogen bond acceptors and donors, and they're very different.
So here it's AADH, and H is sort of a non-polar hydrogen that can't really be involved.
Here, it's HDAA. Here, it's ADA, and then it has a methyl group, which can have a very different kind of interaction. And then this one is the opposite-- MADA.

So depending on which base pair you make, you can read out all this information about which base pair it is.

You can tell if it's a TA or an AT residue. You can tell if it's a CG or a GC residue by looking at the major groove.

But that's not what we want. Right?

We like this minor groove, and the reason is, well, the middle of the nucleotides-- sorry, the hydrogen bond-- is different.

The outside ones are the same in every base pair, and they're actually positioned in identical positions.
So whether you have an AT, CG, GC, or a TA base pair, you'll get the same pair of acceptors for hydrogen bonding in the minor groove.
And in fact, that's what happens.

THERE ARE INTERACTIONS----THAT ARE NON-SEQUENCE-SPECIFIC IN THE MAJOR GROOVE, BUT WATSON AND CRICK BASE PAIR SPECIFIC IN THE MINOR GROOVE.

And we're going to come back to this because it turns out many proteins that recognize DNA non-specifically use these interactions to do so.
And that is important because what that allows the polymerase to do is, not only to recognize no matter what base pair is there, but it can also say, yes, it's a base pair.

So if you have an AG here, these things are all screwed up.

They're not in the right place, and it can detect that.
Or if you have a CC or a CT or any number of other mismatched base pairs, that distorts where those two acceptors.
And it can say, that new DNA you just made, that's not right.
And it can do things in response to that.
So it's detecting, not only all different sequences, but it can detect whether there is a base pair that's correct or non-correct, non-Watson and Crick pair of nucleotides.

PART 5

SO WHAT PART OF THE DNA WOULD IT BE BINDING?

-PALM BINDS PTJ AND NEWLY SYNTH DNA
- BIND THE PHOSPHATE BACKBONE, WHICH DOES NOT CHANGE, REGARDLESS OF SEQUENCE.D INTERACTIONS----THAT ARE NON-SEQUENCE-SPECIFIC IN THE MAJOR GROOVE, BUT WATSON AND CRICK BASE PAIR SPECIFIC IN THE MINOR GROOVE.
-DNTP BINDING IS MEDIATED BY BASE PAIRING AND BINDING TO DIVALENT METALIC CATION (Mg++), the function of DNA polymerase can be inhibit by other chelating agent like EDTA.

EDTA, you can kill virtually any DNA polymerase's activity.

So if we just go to the next **slide**-- so this just shows, here, this is a nucleotide that's coming in.

It's going to hydrogen bond. That's going to be one mechanism for it to bind to this,

but the other is this triphosphate interacts with, actually, a pair of divalent metal cations that are holding this in addition to the base pairing in the correct position for catalysis.

Once that happens, there's another event that occurs, which is that this so-called O-helix-- and when

I'm talking about helix here, I'm talking about an alpha helix, not the DNA helix-- and it's part of the finger domain.

And this is what I was mentioning earlier, that

THE FINGERS OPEN AND CLOSE AROUND THE BOUND DNTP.

And this involves this O-helix, which is an alpha helix, and it forms both there's a tyrosine that forms the pi-pi interactions with the ring structure of the incoming nucleotide.

And it's going to be doing so with the hexameric part of the ring, whether it's a purine, which will have a hexamer and the pentamer next to it, or if it's a pyrimidine, it's just going to have the hexamer.

But it's going to interact with that, and it's going to facilitate closing that.

But so are these lysine and arginines, which interact with the beta and gamma phosphates.

So those interactions cause this to clamp down, and sit on top of the base pair.

And what's important about that is that actually prevents hydrolysis from occurring.
So water can't get into that active site at that point, and so the only reaction that can occur is for this hydroxyl group to attack the phosphate.

And just to give you a little bit of a sense of what that looks like, this is actually the real crystal structure of an intermediate in the process of adding a new nucleotide.

Here's the O-helix. This pink guy right here is that tyrosine. This is the lysine and the arginine interacting with the diphosphates.

This is the alpha, beta, and gamma phosphates.
And these are the two metal ions that are involved in holding those triphosphates together.

And the primer is going to be right here.

And you can see how tightly clamped down that is.

So now it's clamped down.

Catalysis occurs. What happens next?

-Finger Domain Closes on Bound dNTP
- **O-helix (alpha helix)**
- **-Tyr- Forms Pi-Pi bonds with base**
- **Lys-Arg Bind triphosphates**

-Release of the pyrophoshate Releases the O-helix

We'll just call this pi-pi bonds with the base and a lysine and arginine bind the triphosphate.

So it's held in place for catalysis.
Catalysis occurs.
And basically, the last step is release
of the pyrophosphate releases the O-helix.

So when this pyrophosphate-- these two phosphates-- are
released, these lysine and arginines
have much lower affinity for the base that's now located there.
And it just pops back open.
So if we go back one, you go back
from this situation, where you have these interactions,
to this situation, where you do not.

And then the last thing that happens
is that when we're finished and we've
added the nucleotide to here, we don't look like this anymore.
We have actually a base pair here.
But that's actually not a high-affinity interaction
for the polymerase.
It wants to have a 3-prime OH here, not at the next position
up.
So how does it get to that point?
Well, it basically slides the next single-stranded DNA
template in here by moving the whole DNA down.
So it has a higher infinity for this substrate
than it does for this substrate.
And it can get there by simply moving one position.
And because all the interactions with the double-stranded DNA
are base pair nonspecific, that's
essentially an energetically equivalent thing
to slide the DNA by one base in the palm portion.

So sliding this by one base pair doesn't change the interactions
with the polymerase, but having this 3-prime OH here as opposed
to here is a much lower affinity interaction.
You get from here to here by moving one base pair.
And that all occurs incredibly rapidly.

We'll get to more about this later,
but about a millisecond for each round of addition.
So it can do this very rapidly.

Any questions about how the catalysis works?

Yeah.

Why don't they just stay where they
are when the reaction occurs?

So partly because they're no longer
attached to the nucleotide.
So the nucleotide is holding them there before,
but once you make this bond here,
you cannot maintain the remaining bonds.
So that's the largest reason they're released.
And then it's basically entropy at that point.
If they can get away, even for a little tiny second,
they're going to go away for good,
because they're no longer covalently bound.