Name: _____

Date: _____ Period: _____

DNA, Hot Pockets, & The Longest Word Ever

Crash Course Biology #11

1. What is the nickname and function of the longest known protein on earth?

2. During _____, the DNA instructions are copied gene by gene and taken out of the nucleus.
3. During _____, amino acids are assembled into strings of polypeptides, or proteins.
4. Most of the polypeptides that get made are _____ which go on to make cell material.

Transcription

5. The length of DNA that is going to be transcribed onto an RNA molecule is called our _____.
6. Each transcription unit has a sequence 'upstream' on the strand called a _____, which marks the beginning of the transcription unit.
7. What are the four nitrogenous bases?

8. T-A-T-A-A-A-A is the sequence that makes the nearly universal promoter known as the _____ _____.
9. In this case, upstream means toward the ____ end and downstream means toward the ____ end.
10. The first enzyme in the process is _____ _____, which copies the DNA sequence downstream of the TATA Box.
11. RNA polymerase works toward the 5' end and copies the DNA into _____ _____.
12. Where does the initial RNA polymerase enzyme in a cell come from?

13. RNA has the base _____ and not _____.
14. The RNA polymerase is triggered to pull off when it reaches a downstream sequence called a _____ _____.
15. A special type of guanine is added to the 5' end, called a _____.
16. About 250 adenines are added to the 3' end, called a _____.
17. The caps on either end of the mRNA package make it easier for the mRNA to leave the _____ and protect it from _____ from nearby passing enzymes, while also making it easier to connect with other organelles later on.

RNA Splicing

18. RNA _____ cuts out 'extra information.'
19. The snRNPs (_____ _____ ribonucleoproteins) are a combination of RNA and proteins that recognize the sequences that signal the start and end of the areas to be spliced.
20. The _____ breaks down the bases that are being cut out and joins the ends of the remaining segments.
21. The sections of mRNA that will eventually be expressed are called _____.
22. The sections of mRNA that are spliced out are called _____.

Translation

23. The organelles where proteins are synthesized are _____, which are made up of protein and ribosomal RNA, or rRNA.
24. Ribosomes have _____ _____ that allow the incoming mRNA to interact with transfer RNA, or _____.
25. On one end of the tRNA is an _____ _____. On the other end is a specific sequence of _____ nitrogenous bases.
26. How many amino acids do humans have in their bodies?
27. The mRNA is read three letters at a time - each set called a triplet _____.
28. The complementary end of the tRNA is called the _____.
29. The tRNA also brings an _____ _____.
30. For almost every gene, the first codon on the mRNA is _____. The matching tRNA anticodon UAC has the amino acid _____ on its other end.
31. If the mRNA codon is UUA, then the matching tRNA anticodon is _____. If the codon is AGA, then the anticodon is _____.
32. Connecting amino acids forms a _____ _____, which is the beginning of a protein.
33. Multiple codons may code for the same amino acid. Why might that be considered positive?

Folding and Protein Structure

34. The actual sequence of amino acids in a polypeptide is called its _____ structure.
35. _____ bonds give polypeptides their secondary structure, which may include spirals, called a _____ or pleated _____.
36. R groups define each amino acid and their interactions cause additional bending to the polypeptide chain, giving a protein its _____ structure.
    a. Some R groups are _____, and they will draw together.
    b. Other R groups are _____, and they will tend to form hydrogen bonds together.
37. Several different polypeptide chains coming together form a _____ structure.
38. What are some functions of completed polypeptides?

Video Info:
This video can be accessed via YouTube: https://www.youtube.com/watch?v=itsb2SqR-R0
Video length: 14:07. Using the table of contents provided by CrashCourse with the YouTube video, you can click and watch certain sections as desired.

Transcript

Introduction

Ok, roll it. You know what this is? It is the longest word in the world, like, anywhere, any language, more than 189,000 letters. If you were to write it down, though I don't know why you would, it would fill up more than 100 pages, and if you could actually say it, without, like, breaking your face, it'd take about 5 hours. What the frick is this word?

Well, it is the name of the longest known protein on earth, and it's actually in you right now. Because of its enormous size, it was given the name 'Titin' by scientists (and that's with two i's) and it's a protein that helps give you like the springiness to your muscles.

Today we're going to be talking about DNA, and how it, along with three versions of its cousin RNA, unleash chemical kung-fu to synthesize proteins just like this. So this is going to take a while to explain, so how about, uh, we make ourselves some hot pockets?

(0:54) Mmm, they're my favorite - ham and cheese. Every time I take a bite I wonder, how do they do it? How do they pack exactly the same flavor into every foil cardboard-wrapped food-ish item? Clearly there's gotta be some super-secret instruction manual kept in a location known only to two people. And since I'm talking about biology here, that brings up a related question: how did I get built from the DNA instructions and biological molecules we've been talking about?

Today, that's what I'm going to do - not actually make hot pockets or a person, but I'm going to be talking about DNA transcription and translation, which is how we get made into the delicious things that we are today. Though hopefully none of us know how delicious people are. Animals, plants and also hot pockets are really nothing more than salty water, carbohydrates, fats, and you know proteins combined in precise proportions following very explicit instructions.

Let's say I wanted to make my own hot pocket. I would have to:
1. Break into the lair of the hot pocket company holding the secret manual.
2. Read the instructions on how to make the machinery to produce the hot pocket and the proportions of the ingredients. 3. Quickly write down that information in shorthand before I get caught by the hot pocket police.
4. Go home, follow the instructions, build the machinery, mix the ingredients together until I have a perfect hot pocket.

1) Transcription (2:12)

And that is how we get us. Very simply, inside the cell's nucleus, the DNA instruction manual is copied gene by gene by transcription onto a kind of RNA and then taken out of the lair where the instructions are followed by the process of translation to assemble amino acid strings into polypeptides, or proteins, that make up all kinds of stuff, from this titin down here, to the keratin in my hair.

But most of the polypeptides that get made aren't structural proteins like hair, they're enzymes which go on to act like the assembly machinery, breaking down and building and combining carbohydrates into lipids and proteins that make up variations of cell material. So enzymes are just like whatever ingenious machinery they use at the factory to make this.

Okay, let's start off in the lair, I mean the nucleus. The length of DNA that we're going to transcribe onto an RNA molecule is called our transcription unit.

a)Transcription Unit (3:00)

And let's say, in today's example, that it's going to include the gene that transcribes for our friend titin, which, in humans at least, occurs on Chromosome 2.

b) Promoter (3:10)

Now each transcription unit has a sequence just above it on the strand and that's called 'upstream', biologists call that 'upstream' on the strand, and that sequence sort of defines when the transcription unit is going to begin. This special sequence is the promoter, and it almost always contains a sequence of two of the four nitrogenous bases that we talked about in our last episode: adenine (A), thymine (T), cytosine (C) and guanine (G).

c) TATA Box (3:32)

Specifically, the promoter is a really simple repetition, we got thymine, adenine, thymine, adenine and then A-A-A, and then on the other side: ATATTTT. Because you know how this works, right?!

This is called the TATA Box. It's nearly universal and helps our enzyme figure out where to bind to the strand. Now, you'll remember from our episode about DNA structure that DNA strands run in one of two directions depending on which end of the strand is free and which end has a phosphate bond. One direction is 5'-3' and the other direction is 3'-5'. In this case, upstream means toward the 3' end and downstream means toward the 5' end.

d) RNA Polymerase (4:12) and e) mRNA

So the first enzyme in this process is RNA polymerase, and it copies the DNA sequence downstream of the TATA Box. That's towards the 5' end and copies it into a similar type of language: messenger RNA (mRNA).

Quick aside: So you'll notice that to read the DNA in order to make enzymes we need an enzyme in the first place. So it kinda gets 'chicken/egg' here. We need the enzyme to make the DNA and the DNA to make, make the enzyme. So, where did RNA polymerase come from in the first place if we haven't made it yet?

What an excellent question! It turns out that all of these basic necessities get handed down from your Mom. She packed quite a lot more into her egg than just her DNA so y'know we had a healthy start. So, thanks Mom!

So the RNA polymerase binds to the DNA at that TATA box and begins to unzip the double-helix. Working along the DNA chain, the enzyme reads the nitrogenous bases, those are the letters, and helps the RNA version of the nitrogenous bases floating around in the nucleus to find their match.

Now you might also recall from our previous episodes that nitrogenous bases only have one counterpart that they can bond with. But RNA, which is the pink one here, doesn't have thymine like DNA does, which is the green and the blue, instead it has uracil (U). So, U appears here in T's place as a partner to adenine.

f) Termination Signal (5:21)

As it moves, the RNA polymerase re-zips the DNA behind it and lets our new strand of messenger RNA peel away. Eventually, the RNA polymerase reaches another sequence downstream, called a termination signal, that triggers it to pull off.

g) 5' Cap and Poly-A Tail (5:34)

Now, some finishing touches before this info can safely leave the lair. First, a special type of guanine (G) is added to the 5' end. That's the first part of the mRNA that we copied and that's called the 5' cap. On the other end, it looks like I fell asleep with my finger on the A key of my keyboard, but another enzyme added about 250 adenines on to the 3' end. This is called our poly-A tail.

These caps on either end of the mRNA package make it easier for the mRNA to leave the nucleus. They also help protect it from degradation from nearby passing enzymes, while also making it easier to

connect with other organelles later on.

2) RNA Splicing (6:08)

But that's still not the end of it. As if to try to confuse me to protect the secret Hot Pocket recipe, the original recipe book also contains lots of extra, misleading information. So just before leaving the nucleus, that extra information gets cut out of the RNA in a process called RNA splicing. And it's something like editing this video.

a) snRNPs and Spliceosome (6:26)

This process is really complicated, but I just had to tell you about two of the key players, because they have such cool names.

One, the snRNPs, which are Small Nuclear Ribonucleoprotein. These are a combination of RNA and proteins, and they recognize the sequences that signal the start and end of the areas to be spliced. snRNPs bunch together with a bunch of other proteins to form the spliceosome, which is what does the actual editing, as it were, breaking the junk segments down so their nitrogenous bases can be reused in DNA or RNA, and sticking together the two ends of the good stuff.

b) Exons and Introns (6:56)

That good stuff that gets spliced together, by the way, are called the exons because they will eventually be expressed. The junk that gets cut out are just the intervening segments, or the introns. The material in the introns will stay in the nucleus and get recycled. So, for instance, titin down there is thought to have hundreds of exons when it's all said and done, probably more than 360, which may be more than any other protein. And it also contains the longest intron in humans, some 17, 000 base pairs long. Man, titin! It is just a world record holder!

3) Translation (7:28)

So now that it's been protected and refined, the messenger RNA can now move out of the nucleus.

Okay, so, a quick review of our Hot Pocket Mission Impossible caper so far: We broke into the lair containing the instructions, we copied down those instructions in shorthand, we added some protective coatings, and then we cut out some extra notes that we didn't need, and then we escaped back out of the lair. Now I have to actually read the notes, make the machinery and assemble the ingredients. This process is called translation.

So next, rewind your memory - or just watch that video again - to the episode about animal cells.

a) mRNA and tRNA (8:01)

Do you remember the rough endoplasmic reticulum? I hope you do. Those little dots on the membranes are the ribosomes, and the processed messenger RNA gets fed into a ribosome like a dollar bill into a vending machine. Ribosomes are a mixture of protein and a second kind of RNA, called ribosomal RNA or rRNA. And they act together as a sort of work space. rRNA doesn't contribute any genetic information to the process, instead it has binding sites that allow the incoming rRNA to interact with another special type of RNA, the third in this caper, called transfer RNA, or tRNA. And tRNA really might as well be called 'translation RNA' because that's what it does; it translates from the language of nucleotides into the language of amino acids and proteins.

b) Triplet Codons and Anticodons (8:39)

On one end of the tRNA is an amino acid. On the other end is a specific sequence of three nitrogenous bases. These two ends are kind of matched to each other. Each of the 20 amino acids that we have in our body has its own sequence at the end. So if the tRNA has the amino acid methionine on end end, for instance, it can have UAC as the nucleotide sequence on the other end.

Now, it's just like building a puzzle. The mRNA slides through the ribosome. The ribosome reads the mRNA three letters at a time - each set called a triplet codon. The ribosome then finds the matching piece of the puzzle: a tRNA with three bases that will pair with the codon sequence. That end of the tRNA, by the

way, is called the anticodon. Sorry for all the terminology. YOU NEED TO KNOW IT!

And of course, by bringing in the matching tRNA, the ribosome is also bringing in whatever amino acid is on that tRNA. Okay so, starting at the 5' end of the mRNA that's fed into the ribosome, after the 5' cap, for almost every gene, you find the nucleotide sequence AUG on the mRNA. The ribosome finds a tRNA with the anticodon UAC, and on the other end of that tRNA is methionine. The mRNA, like a mile-long dollar bill, keeps sliding into that ribosome so that the next codon can be read, and another tRNA molecule with the right anticodon binds on.

If the codon is UUA, then the matching tRNA has an AAU on one end and a leucine on the other. And if the mRNA has an AGA, then the matching tRNA has a UCU on one end and an arginine on the other. In each case that new amino acid gets connected on to the previous amino acid - starting a polypeptide chain. Which is the beginning, the very beginning, of a protein.

But it turns out that here are LOTS of different ways to read this code. Because UUA is not only the triplet that codes for leucine, UUG does too! And arginine is coded for by six different triplets! This is actually a good thing. It means that we can make a few errors in copying and transcribing and translating DNA, and we won't necessarily change the end product. This process continues, with the mRNA sliding in a bit more, and the ribosome bringing in another tRNA with another amino acid, that amino acid binding to the existing chain and on and on and on and on, sometimes for THOUSANDS of amino acids to make a single polypeptide chain, for example.

This whole word is basically just the names of the amino acids in the sequence in the order in which they occur in the protein all 34, 350 of them.

4) Folding and Protein Structure (10:51)

But before we can make our own Hot Pockets, and that string of amino acids becomes my muscle tissue, we have some folding to do. That's because proteins, in addition to being hella big, can also contort into very complex and downright lovely formations.

One key to understanding how a protein works is to understand how it folds, and scientists have been working for decades on computer programs to try and figure out protein folding.

a) Primary Structure (11:11)

Now, the actual sequence of amino acids in a polypeptide - what you see scrolling along down there - is called its primary structure. One amino acid covalently bonded to another, and that one to another, in a single file.

b) Secondary Structure (11:23)

But some amino acids don't like to just hold hands with two others, they're a bit more promiscuous than that. The hydrogens on the main backbone of the amino acids like to sometimes form bonds on the side, hydrogen bonds, to the oxygens on amino acids a few doors down. When they do that, depending on the primary structure, they bend and fold and twist in to a chain of spirals, called a helix. We also sometimes find several kinked strands laying parallel to one another, called pleated sheets. All those hydrogen bonds in pleated sheets are what make silk strong, for instance. So in the end, our promiscuous amino acids lead to wrinkled sheets. Ah-ha! These hydrogen bonds are what help give polypeptides their secondary structure.

c) Tertiary Structure (11:58)

But it doesn't end there. Remember the R groups that define each amino acid? Well some of them are hydrophobic. And since the protein is in the cell, which is mostly water, all of those hydrophobic groups try to hide from the water by huddling together, and that can bend up the chain some more.

Other R groups are hydrophilic, which if nothing else means that they like to form hydrogen bonds with other hydrophilic R groups. So we get more bonding, and more bending, and our single-file line has now taken on a massively complex 3-dimensional shape. It also explains why I can fix my bed-head by wetting

my hair with water. The water helps break some of those hydrogen bonds in the keratin which relaxes its structure. That way I can comb it out, and when it dries those bonds reform and voila, perfect hair! All of this shape caused by bonding between R groups gives our polypeptide its tertiary structure.

d) Quaternary Structure (12:44)

So now we have a massively contorted polypeptide chain, and it actually contorts very precisely. Sometimes, just one chain is what makes up the whole enzyme or protein. In other proteins, like hemoglobin, several different chains come together to form a quaternary structure.

So a quick review of structure: the sequence is the primary structure, the backbone of hydrogen bonds forming sheets and spirals are secondary structure, R group bonds are tertiary, and the arrangement of multiple proteins together give the quaternary structure.

These polypeptides are either structural proteins, like this thing at the bottom here that you can find in your muscle or in my hot pocket. They might also be enzymes, and enzymes like, do stuff. They can cut up biological molecules like I so with this chef's knife, they can mix stuff and they can put stuff together. So from that one recipe book we got all of the ingredients and all of the tools necessary to make me, which is better than a hot pocket. Would you all agree?

Conclusion (13:34)

Now take your time with this stuff, feel free to watch the episode a couple times, because next week we're going to talk about how cells swap all of this genetic information through reproduction.

Thank you for watching this episode. By now, you should probably know how this works. You can click on any of the links over there, and it'll take you back to that point in the show as long as you are not watching on your cell phone. It doesn't work on cell phones, I apologize for that.

Thank you to everyone who helped us put this show together, and thank you to you, for watching it today. If you have any questions about this episode please leave them in the comments below, or you can get us on Facebook or Twitter. And that's all! Goodbye.

ANSWERS

DNA, Hot Pockets, & The Longest Word Ever

Crash Course Biology #11

1. What is the nickname and function of the longest known protein on earth? **Titin – gives "springiness" to your muscles** (*also known as connectin, titin is responsible for the passive elasticity of muscle)*

2. During **transcription**, the DNA instructions are copied gene by gene and taken out of the nucleus.

3. During **translation**, amino acids are assembled into strings of polypeptides, or proteins.

4. Most of the polypeptides that get made are **enzymes** which go on to make cell material.

Transcription

5. The length of DNA that is going to be transcribed onto an RNA molecule is called our **transcription unit**.

6. Each transcription unit has a sequence 'upstream' on the strand called a **promoter**, which marks the beginning of the transcription unit.

7. What are the four nitrogenous bases? **Adenine (A), thymine (T), cytosine (C), and guanine (G)**

8. T-A-T-A-A-A-A is the sequence that makes the nearly universal promoter known as the **TATA Box**.

9. In this case, upstream means toward the **3'** end and downstream means toward the **5'** end.

10. The first enzyme in the process is **RNA polymerase**, which copies the DNA sequence downstream of the TATA Box.

11. RNA polymerase works toward the 5' end and copies the DNA into **messenger RNA (mRNA)**.

12. Where does the initial RNA polymerase enzyme in a cell come from? **It is inherited maternally via the egg**

13. RNA has the base **uracil** and not **thymine**.

14. The RNA polymerase is triggered to pull off when it reaches a downstream sequence called a **termination signal**,

15. A special type of guanine is added to the 5' end, called a **5' cap**.

16. About 250 adenines are added to the 3' end, called a **poly-A tail**.

17. The caps on either end of the mRNA package make it easier for the mRNA to leave the **nucleus** and protect it from **degradation** from nearby passing enzymes, while also making it easier to connect with other organelles later on.

RNA Splicing

18. RNA **splicing** cuts out 'extra information.'

19. The snRNPs (**small nuclear** ribonucleoproteins) are a combination of RNA and proteins that recognize the sequences that signal the start and end of the areas to be spliced.

20. The **spliceosome** breaks down the bases that are being cut out and joins the ends of the remaining segments.

21. The sections of mRNA that will eventually be expressed are called **exons**.

22. The sections of mRNA that are spliced out are called **introns**.

Translation

23. The organelles where proteins are synthesized are **ribosomes**, which are made up of protein and ribosomal RNA, or rRNA.

24. Ribosomes have **binding sites** that allow the incoming mRNA to interact with transfer RNA, or **tRNA**.

25. On one end of the tRNA is an **amino acid**. On the other end is a specific sequence of **three** nitrogenous bases.

26. How many amino acids do humans have in their bodies? **20**

27. As the mRNA slides through the ribosome, it is read three letters at a time - each set called a triplet **codon**.

28. The complementary end of the tRNA is called the **anticodon**.

29. The tRNA also brings an amino acid.

30. For almost every gene, the first codon on the mRNA is **AUG**. The matching tRNA anticodon UAC has the amino acid **methionine** on its other end.

31. If the mRNA codon is UUA, then the matching tRNA anticodon is **AAU**. If the codon is AGA, then the anticodon is **UCU**.

32. Connecting amino acids forms a **polypeptide chain**, which is the beginning of a protein.

33. Multiple codons may code for the same amino acid. Why might that be considered positive? **A few errors can be made without necessarily changing the end product.**

Folding and Protein Structure

34. The actual sequence of amino acids in a polypeptide is called its **primary** structure.

35. **Hydrogen** bonds give polypeptides their secondary structure, which may include spirals, called a **helix** or pleated **sheets**.

36. R groups define each amino acid and their interactions cause additional bending to the polypeptide chain, giving a protein its **tertiary** structure.

    a. Some R groups are **hydrophobic**, and they will draw together.

    b. Other R groups are **hydrophilic**, and they will tend to form hydrogen bonds together.

37. Several different polypeptide chains coming together form a **quaternary** structure.

38. What are some functions of completed polypeptides? **Structural proteins, enzymes – break down biological molecules, catalyze chemical reactions**