DSAP ANALYSIS

Yes, Comic Sans.

Note: This is a complete transcript to the powerpoint. It is good to read through this once to understand everything. If you ever need help and just need a quick a quick glance at something, there is a separate, much shorter help guide. It has less explanation and more directions.

Some things before we start:

What you need:

Cn3D: A program used to view 3-D protein structures. You can download this from NCBI. The download link will be seen below. It's quick and painless to download.

Important Terms:

Function: What the protein actually does.

Structure: What the protein looks like. These two definitions are simple but hard to actually determine when you get down to it; structure and function can get very complex!

Domain: A specific, sometimes repeated section of a protein. Domains can be thought of as ***"areas" of a protein with a specific property or function****. It is usually conserved among related organisms: in other words, it would be found in duckweed, corn, and mustard plants regardless of what organism it is. Domains are important in determining function. Here's an analogy: Your arm is responsible for catching, throwing, waving, etc. That can be considered a domain of the entire human body.*

Alpha Helix and Beta Pleated Sheets:
Alpha helices *are amino acids wound up in a* ***spiral****, and* ***beta pleated*** *sheets are amino acids* ***layered in sheets****. They interact with each other to form the complete structure of the protein.*

So you've got your sequence, cropped out the vector backbone, edited the waveform and ran it through the BLASTs. You've got the protein your DNA codes for. Now what?

Review:

1. *Cropping. Remember that your DNA sequence starts right after CGGCCGGGGG(etc.) and ends at the last base of the poly-A tail.*
    a. *Please EDIT your waveform! Scan the top of your sequence for any "Ns"-that means the sequencing machine messed up and it doesn't know what's going on. You should easily be able to tell what base it is by the peak.*
    b. *Here's an easy code: A=Green, T=Red, C=Blue, G=Black.*
    c. *Remember to go over cases of unreadables in the Cropping guide.*
2. *BLASTs. As a quick overview:*
    a. *Blastn matches your DNA nucleotide sequence to other DNA sequences.*
    b. *Blastx translates your DNA nucleotide sequence to a protein sequence and matches that to other protein sequences.*
    c. *Blastp translates your newly found protein sequence and matches that to other protein sequences.*
    d. *Good matches have very low E values (it's an exponent, remember, so a very large negative number is actually very low). Good matches will also have green bands in the graphic organizer. Good matches should always be from other plants.*

Now, Analysis!

But why analyze?

*Well, **that's the point of Waksman**. The whole cropping and BLASTing deal seems really hard and complicated, but it gets easier with practice. **Analysis is our end result of all that work and is what we dedicate most of our time to.** To analyze proteins, we determine its **function** and **structure** using outside sources and 3-D modeling programs, and then **hypothesize applications** of these proteins (that's what adult scientists do all the time-finding vaccines, cures, or new drugs are all almost exactly the same as this process).*

This sounds hard and intimidating, but don't worry! We'll walk you through it.

(**Note: Look on the ppt provided for extra visuals. The ones provided here are the more intricate ones that were unable to fit on the ppt.)

First: Review

The "Review" portion of DSAP forces us to take into account the entire sequence. Remember that throughout Blastx and Blastp, we've been working with our ***ORF****: the open reading frame we found from toolbox.*

However, ***that's not the entire sequence we cropped from the first step****. The part* ***before*** *the ORF is called the 5' end, and the part* ***after*** *the ORF is called the 3' end. (See Slide 10). These are* ***non-coding regions*** *and tend to have* ***regulatory and promoting functions****: they help the DNA replicate. It's good to account for your whole sequence instead of just the ORF, even though only the ORF is the coding region.*

DSAP will have the ORF bases for you. To answer this question, simply use the DNA box above and click on the base (guesstimate where it is). When you find the base, enter it in the boxes. See Slide 9 for reference. Remember that 5' ends at the base right before the ORF, and the 3' starts right after the ORF ends.

Next: Analysis Begins

Step One

The first step is researching your protein (a "literature search"). You're really just googling your protein. It's always good to use WIKIPEDIA as a GENERAL SOURCE to get a BASIC IDEA of your protein. You can cite it as a source; however:

This year, we are requiring you to use two sources: One Wikipedia, and a second, legitimate source.

Popular sources we use are NCBI and PubMed. Reading research abstracts may be very difficult, but it will soon become easy after practice. That's why we let you use wiki to help you understand it better! :) yw

When you find the general function of your protein, simply type IN YOUR OWN WORDS the info you found in the box, and copy/paste the URLs. See Slide 11 for an example.

Step Two
Is this protein also found in humans?

Now, we've reiterated a couple times that we're looking for proteins in plants–we're interested in duckweed, after all, and function in humans is pretty irrelevant (excluding the fact that many plant enzymes aren't even found in humans). However, it's always good to check, in order to see the ***evolutionary importance*** *of the protein.*

To find related proteins in humans (Slide 12)

1. *Go back to Blastp*
2. *Click on your first result.*
3. *Click on "Identical Proteins"*
4. *On the results page, click on "Related sequences" under the first result.*
5. *On the side bar, change the filter settings to search in humans only*
    a. *Click on "More"*
    b. *Start typing "Homo sapiens" to filter for human beings.*

The cop-out method: Just search it up on Google. However, this is definitely not 100% reliable.

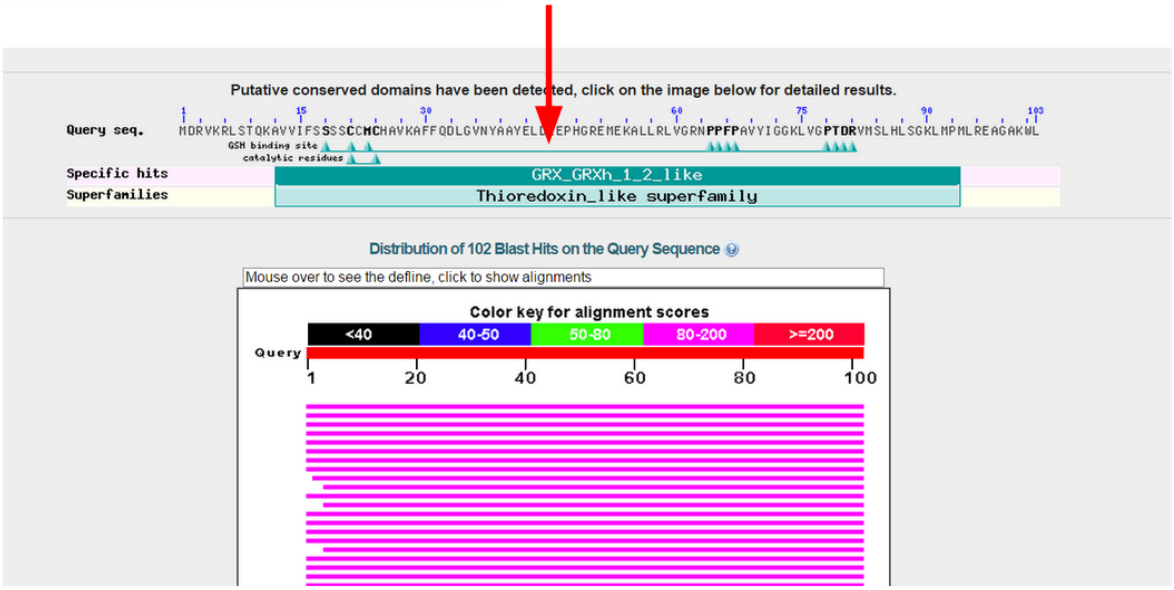
**BONUS SLIDE: Slide 13 is an extra "cheat sheet" of common "model organisms". Model organisms are those used by scientists in experiments. You might recognize fruit flies and rats on that list. You'll also see a number of plants…you don't need to remember it, but it's good to know. You'll be seeing zea mays a lot.*

Step Three

Step One covered ***function*** *and step two was an "extra" step on evolutionary importance. Here, we research* ***structure.***

Structure is very heavily linked to function*. Function can mean a large, extracellular function (regulating cell cycles, reproduction, etc.) However,* ***very tiny interactions*** *between proteins also compose the "function" of proteins. These very tiny interactions are* ***governed by structure****: hydrophobicity, electric charges, etc. This will become more evident as you continue your Waksman journey.*

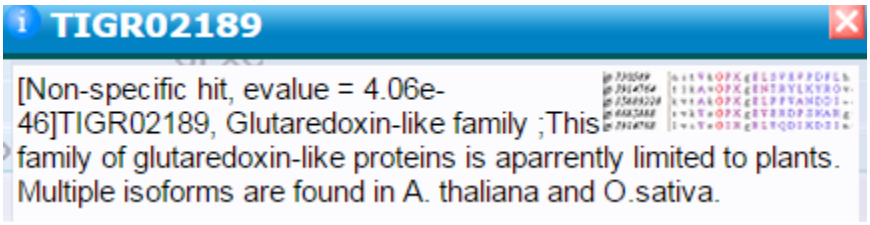
To find the protein structure, first go back to ***Blastp*** *and click at that banner at the top:*



Then you'll arrive at the screenshot on Slide 16. This screenshot is a list of ***domains****: conserved regions in your sequence. When you mouse over these results, some will have thumbnails with a 3-D blob on it, generally green and gold but sometimes purple and yellow. If this thumbnail appears, then your protein has a 3-D structure available in the database! Here's a pic:*

If there is no thumbnail, that does not mean your protein doesn't exist. It just means that scientists haven't got a clear image of the protein, and cannot give a clear diagram of it.

#Jelisa15-16



No Structure.....

NOTE: *This answers Question Three in analysis. The results you've been mousing over are your domains. If you cannot find domains, you can always google "(Protein Name) Domains". However, it is usually difficult to read these papers, so if you are confused feel free to ask for help!*

When you click on a result with a 3-D structure, you will arrive at Slide 17. Here, when you maximize the "Structure" menu, you can download Cn3D (if you have not already done so) and open up the structure by clicking "structure view".

Using Cn3D
*(*Disclaimer: Click "Dismiss" on that first pop-up box, if you get one. You don't need that).*

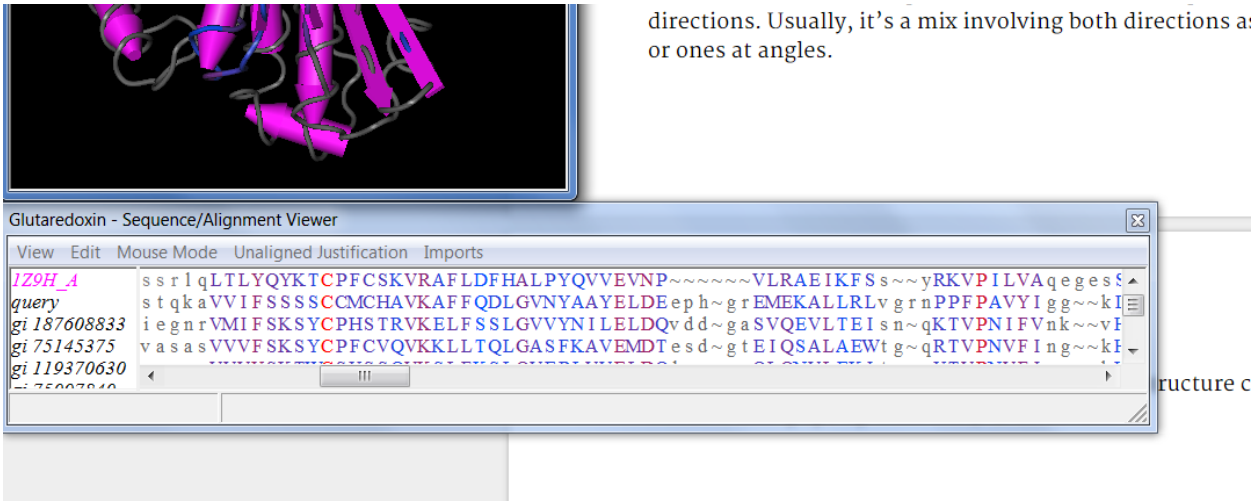
When you first open your 3-D structure on Cn3D, it will look like Slide 18. This looks really disorienting! With more experience, you'll learn how to analyze this. However, we'll skip this for now and go into an easier mode.

1. *At the top menu, click on **"Style"**, and in the drop down menu, click on **"Rendering Shortcuts"**. From the menu, select **"Worms"**. This makes it easy to see the alpha helices and beta pleated sheets.*
2. *Once again, click on **"Style"** and this time, click on **"Coloring Shortcuts"**. Then click on **"Secondary Structure"**.*
3. *You can see this on slide 19. The **purple protein** is in worms. The **green** and **yellow** is in worms AND secondary structure highlighted.*

Now, you can answer how many alpha helices and beta pleated sheets there are.

Antiparallel v.s. parallel: *Parallel beta pleated sheets are just two sheets pointing in the same direction. Antiparallel means two sheets pointing in exact opposite directions. Usually, it's a mix involving both directions as well as perpendicular ones or ones at angles.*

*To answer question four, or the **organism this structure comes from**, find this window that pops up with the 3-D structure.*



See that purply-pink word on the left? That's the accession number-you remember that from BLAST-of the protein. If you click on it, you can find the data sheet, including the organism it comes from. Easy!

(NOTE: WE WILL SKIP QUESTION SIX FOR NOW. THIS WILL BE RELEVANT LONG FROM NOW, AFTER LABS. FOR NOW, PUT N/A).

Step Four: Applications

People love to BS application questions (it's true). We tend to think it's easy to "think outside the box" and "apply what you've learned" to real life. However, this question is much harder and more serious than real life.

***Point 1**: Real scientists are using this research.*

*For those who don't know, WSSP (Waksman Student Scholars Program) is run by Rutgers professors. These professors are ALSO studying proteins in duckweed, with applications in biofuels and bioremediation. Our research DIRECTLY affects theirs! For new members, the enormity of our efforts is hard to understand/realize, but as you do this more (especially if you go to Rutgers meetings) you will see the direct consequences of our research. So, this is the part that we should BS the **least**. Why am I saying this? Because it's very easy to brush off. I know that I was prone to doing it when I was a sophomore (Jeff's Confessions #1). However, **we cannot do this**.*

Point 2: *Scientific applications are hard.*

After a couple of times, even in your first year you can easily find "BS" applications for proteins. (Example one: cancer research. Not everything helps with cancer research). In addition, many things aren't plausible simply because ***it's not scientifically possible*** *(or economically possible, too). We are looking for real-life examples, so this means we need to follow the rules of science. We can't just make things up.*

So, how do I answer this question?

First, ***understand the science****. This involves as much chemistry as biology. Understand the bonds between amino acids in your protein. Understand the properties of each domain in your protein, understand the interactions between them. Understand the interactions between your protein and other proteins. This can immensely help you in your thinking. Need help? Totally confused by your protein? This is NATURAL! Ask for help from any experienced member!*

Second, ***try designing an experiment****. It doesn't need to be a fool-proof, totally awesome experiment. It could have a couple of holes in it. WSSP staff is trying to see that you can extrapolate the information you found and apply it somewhere. So start easy in the lab-try "finding how the deletion of my DNA sequence affects the production of so-and-so".*

Finally, ***remember biofuels and bioremediation****. Those are the two major things scientists are working on right now regarding duckweed, especially in Rutgers. However, you can also look at other biology-related industries: medicine/pharmacy, life sciences, etc.*

Here's an example of one good analysis response:

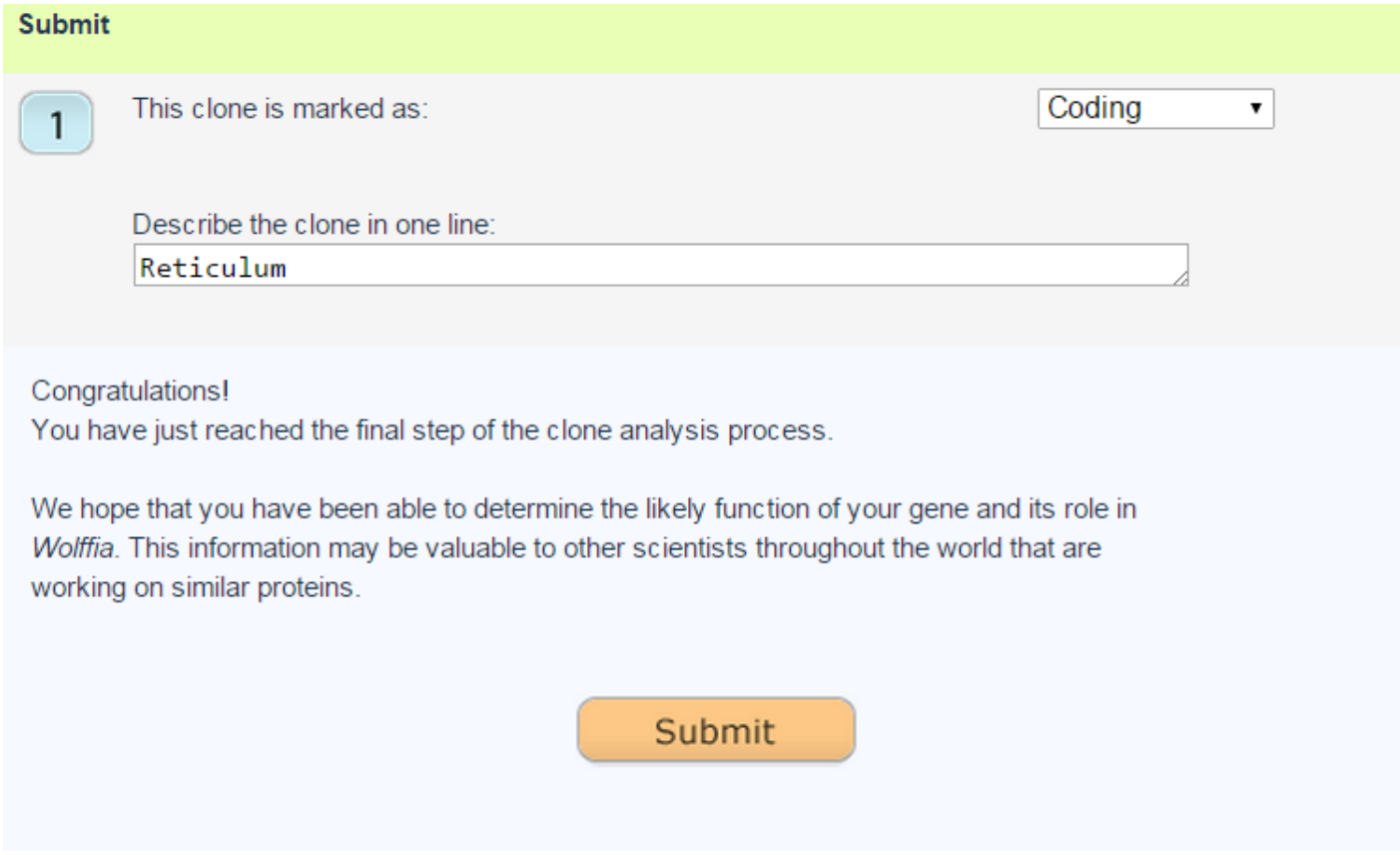
Tetraspanin-18 is a tetraspanin that does not yet have a clear function; it is known to be a scaffolding protein like other tetraspanins. One option is to build a protein assay by setting up a multitude environments that favor certain signaling pathways. To build this we can create environments that stimulate the release of signaling molecules whose receptor proteins are not yet known. Then, we can collect all the samples and run a Western blot to see which one has the most expressed gene for Tetraspanin-18. Another possibility is setting up a radioactive marker on the cysteine active sites. Whenever they react, they will release radiation, and we can trace that back to the original signaling molecule. Then we can find what signaling pathway tetraspanin-18 is involved in.

In this example the function of Tetraspanin-18 was not clear, so I couldn't link it to biology-related industries. However, I crudely outlined a couple of experiments I could do to

determine this function. It's also clear that I understand the interactions between this protein and others.

This might be totally confusing and super hard for you. DO NOT WORRY! If you are stuck, please do not hesitate to ask and we will help you!

Once that is over, you will reach the last page. Click "Submit" to finish!



NOTE: PLEASE PRESS THIS SUBMIT BUTTON. SOME PEOPLE LITERALLY FORGOT TO PRESS IT AND THEN THEY GOT IN TROUBLE BY ACCIDENT!! SMH

CONGRATULATIONS UR DONE.

#YEEZYFORPRES2020