

A6. Investigating Zika virus with blast

Zika virus is incredibly frightening for any woman who is pregnant or thinking of starting a family. Scientists are working on vaccines to prevent infections with Zika virus. It might also be possible to use existing drugs to treat Zika infections.

Some of the bioinformatics tools we've been studying can be used to identify potential anti-Zika drugs. In this investigation, we're going to see if we can find drugs that might be able to treat people infected with Zika virus.

To carry out this investigation, you will use a variety of resources at the NCBI to learn about Zika virus. Then, you'll use blastp to identify compounds that inhibit viral proteins, last you'll use Molecule World to determine if those compounds might bind to Zika virus proteins and inhibit replication.

About viruses

All viruses need to infect some kind of cell in order to reproduce. To reproduce, they must get into a cell, and hijack the machinery in the cell to make viral proteins and new copies of the viral genome.

Viral genomes all consist of some of type of nucleic acid. Unlike humans, whose genomes are always linear, double-stranded DNA, viral genomes can take many forms.

Viral genomes can be:

- double-stranded DNA,
- single-stranded DNA,
- double-stranded RNA,
- single-stranded, + strand RNA (this is the same sense as mRNA),
- single-stranded, minus (-) strand RNA (this must be copied before translation),
- single-stranded, RNA retroviruses,
- double-stranded, DNA retroviruses.

Viral structures:

1. **Lipid envelope:** Some viruses have a lipid envelope; others do not. For viruses that have a lipid envelope, like influenza virus, the envelope comes from infected cells. Many viruses have proteins in the lipid envelope that help them bind to cells.

2. **Capsid:** The capsid is an outer coat made of protein. The capsid protects the viral genome much like the candy coating of an M & M protects the chocolate.

3. **Polymerase:** This is the enzyme that copies the viral genome. Some viral polymerases are also called L proteins. The name of the polymerase tells you what type of polymer it makes. DNA polymerases make DNA. RNA polymerases make RNA. Sometimes the name of the polymerase tells you something about the template it uses. For example, some viral polymerases can make DNA by using an RNA template. Those polymerases are also called reverse transcriptases.

4. **Proteases:** Many viruses produce an mRNA that gets translated as a single long protein. After translation, this protein is chopped into smaller segments by one or more proteases. When this kind of protein is made, it's called a polyprotein. Proteases are enzymes that cut bonds in proteins.

5. **Non-structural proteins and other viral proteins:** Viruses can have other proteins as well. These can be unique to each type of virus. Some proteins are spikes on the outside of

the envelope that help the virus infect specific types of cells. Some proteins help in translating RNA and making proteins. Other proteins help in copying the viral genome.

Instructions

In this investigation you will:

Part I. Conduct background research to become familiar with the Zika virus and its effects.

Part II. Look at the Zika virus genome and the proteins it makes.

Part III. Use protein blast to find potential drugs in the structure database.

Part IV. Use BLAST to compare the new sequence you found to the protein sequence from the Zika virus.

Part V. Use Molecule World and your sequence alignment from part IV to determine if your drug could bind the Zika virus protein.

Review:

Before you begin, you may want to review using the Find command to locate text on a web page.

Part I. Background research

1. Go to the NCBI Zika virus resource page:

<http://www.ncbi.nlm.nih.gov/genome/viruses/variation/resources/>

2. Click the box that says “Zika virus resources.” This opens a set of links to information about Zika virus. You will be using links in this section to learn more about Zika virus.

3. Click **CDC**.

4. Click “About Zika,” then “What you need to know”

5. Read about the top five things you need to know about Zika virus.

Q1. What is the primary birth defect that’s associated with Zika infections?

6. Return to the Zika virus resources section that you found in step 2.

7. Click the Health Map link and use the map to see where Zika virus has been found.

Health Map works like Google Maps.

You can see the entire map by clicking the expand icon  located in the upper right hand corner.

On an iPad:

- Zoom in by touching the map or spreading your fingers apart.
- Move through the map by using two fingers to drag the image.
- View health reports by touching one of the dots.

On a desktop computer:

- Zoom in with command (or Ctrl) + (or scroll)
- Zoom out with command (or Ctrl) and - (or scroll).
- Click the map and hold the button down to drag it,
- Click a dot to see the reports.

8. Zoom in to see Austin, Texas and click the dot on Austin.

Q2. When was the most recent Zika infection reported in Austin?

9. Return to the Zika virus resources section that you found in step 2.

10. Click the Viral Zone link and read the information on the page. You may want to print a pdf copy of this page and save it since we'll be referring to it later on. Scan and read the information on the page and look at the images to answer the questions below.

Q3. Describe the genome for the Zika virus.

[Hint: Your answer should mention if it's RNA or DNA, if it contains one strand (ss for single stranded) or two (ds for double stranded), and if it's RNA, whether it's the + sense or - sense.]

Q4. How large is the Zika genome in terms of nucleotides or bp?

Q5. How large is the Zika polyprotein? Remember the length of a protein is measured in amino acids.

Q6. What insect transmits the Zika virus?

Q7. What kind of enzyme is NS3?

Q8. What kind of enzyme is NS5?

Enzymes are described in part by their substrates and sometimes by the types of things they make or they kinds of reactions they catalyze. The names of enzymes usually end in "ase." A nuclease, for example, is an enzyme that cuts nucleic acids. A lipase cuts lipids. A protease cuts proteins. A polymerase makes a polymer, RNA or DNA.

Part II. The Zika virus genome

1. Return to the Zika virus resources section that you found in part I, step 2.

2. Click the link to the NCBI Zika virus reference genome.

You'll see a map showing a gene from the Zika virus (green) and directly below it, the polyprotein that gets translated from the Zika virus (red).

Below the polyprotein, you'll see shorter segments that are made by cutting the large polyprotein into smaller pieces. If you hold your mouse over the shorter pieces, the names of those proteins will appear.

3. Click the link above the map that says "[GenBank](#)."

The GenBank record shows the same information as the map, in a text form. This page also contains links to research articles about the virus.

4. Scroll down the page.

The names on the left side of the page, in capital letters, are fields in the nucleotide sequence database.

As you continue scrolling, you will see the FEATURES field name. This field contains information about the Zika genome sequence. Comparing the Zika genome to other viral genomes allowed many of these features to be identified. We call these pieces of information, like features, "annotations." When you hear people talk about annotating genomes, they are referring to the process of finding these kinds of information.

5. One of the first features is the 5' UTR. UTR stands for Untranslated Region. The numbers to the right of the feature name correspond to nucleotide positions in the viral genome.

Q9. Where does the 5' UTR begin and end?

6. One of the next features you'll see is the CDS. CDS stands for Coding Sequence. This part of the genome corresponds to the region of mRNA that gets translated into protein.

Q10. Where does the coding sequence begin and end?

7. You may notice that the coding sequence ends before the end of the genome. Remember, you looked at the genome size in Q5. Scroll down to find the 3'UTR. Remember UTR means Untranslated Region.

Q11. Where does the 3' UTR begin and end?

8. Scroll back up the page to the top of the CDS. In this field, you'll see a section that says "translation." This will show the entire amino acid sequence of the Zika virus polyprotein.

9. Scroll down past the polyprotein.

In this region, you'll see features designated as "mat peptide." This is short for "mature peptide." Mature peptides are shorter protein segments that are produced when the long polyprotein is cut into shorter proteins by proteases.

The position numbers show the nucleotide positions that correspond to the sites where the coding sequence for these peptides begin and end in the full-length genome sequence. These positions are also shown on the map you viewed earlier and in the viral genome map you viewed at the Viral Zone page.

10. Read through the page to find the mature peptides for the protease (nonstructural protein NS3) and the polymerase (NS5).

Q12. Where does the NS3 protease sequence begin and end? What is the protein id for the NS3 protease?

11. Look through the list of mature peptides to find the RNA-dependent RNA polymerase.

Q134. Where does the RNA-dependent RNA polymerase NS5 sequence begin and end? What is the protein id for the NS5 RNA polymerase?

Part III. Use protein blast to find potential drugs in the structure database.

The most straightforward way to find a potential drug is to look at drugs that bind to the active site in similar proteins. In the best cases, these drugs might also prevent replication in related viruses. Two common drug targets in viruses are proteases and polymerases. (Remember, when you see “ase,” you know the protein is an enzyme). Proteases are enzymes that cut up proteins. Polymerases are enzymes that make polymers, like DNA or RNA. HIV infections have become manageable diseases over the past ten years because of protease and polymerase inhibitors. Chronic infections with Hepatitis C are now treatable because of a drug called Solvaldi® (Sofosbuvir) that inhibits the viral polymerase.

In Zika virus, the genome is made of RNA. The Zika virus genome is the + sense, so it can be used as mRNA and translated to make proteins. But when it comes to making new genomes, Zika virus has a problem. It can't use the host cells' RNA polymerase because the host RNA polymerase only makes RNA by copying DNA inside the nucleus.

To overcome that problem, Zika virus makes it's own special RNA polymerase. Unlike the polymerase in the host cell, the Zika RNA polymerase is able to use RNA as template. This allows the Zika RNA polymerase to make a - strand RNA and then copy that - strand RNA to produce new mRNA for making proteins and viral genomes for making new copies of the virus.

1. Go to the NCBI and search for this accession number: NC_012532.1

2. Click the link for the Nucleotide database to view the record for the complete Zika virus genome.

Make sure the GenBank record is showing. This is the record with all the text information.

3. Pick a protein to use as a query. You can either use the protease (NS3) or the RNA polymerase (NS5). You're going to use protein blast to search for related proteins, from other viruses, that are bound to potential drugs. The protein you choose, NS3 or NS5, will be the target of an anti-viral drug.

Q14. Which drug target did you decide to investigate (NS3 or NS5)? What is the protein_id?

4. Scroll through the GenBank record to find your target protein. Click the link to the protein ID for the protein you decide to investigate. This will take you to the GenBank record for that protein.

Q15. How long is your protein? This is in the LOCUS field at the top of the page. Remember to include the units for the length.

5. In the top right hand corner, you'll see a link that says "RUN BLAST." Click it.

This automatically sets up a blast search to compare your protein to other proteins in a database of your choice.

The four steps in a blast search are to:

1. **Enter a query sequence.**
2. **Choose which database to search**
3. **Pick an algorithm to use for the search.**
4. **Start the search:** Click the BLAST button.

You've completed the first step. When you clicked "Run Blast," the NCBI website entered the accession number for your protein in the Query Sequence field. This is similar to entering a query sequence.

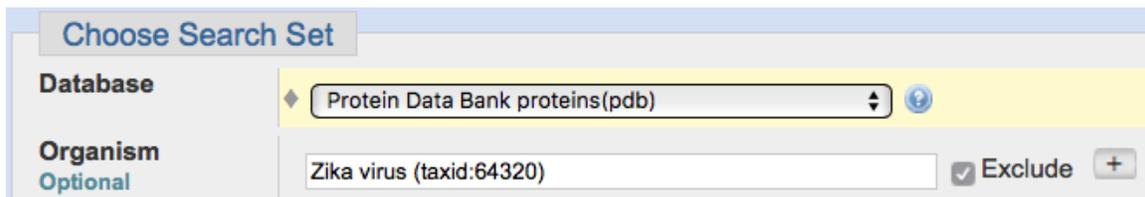
Now, you need to choose a database (step 2). We want to see if we can find structures of related proteins that are bound to inhibitors or drugs. To do this, you'll need to pick a different database.

5. Change the database to "Protein Data Bank proteins (pdb)." All the protein sequences in this database come from molecular structures in the MMDB.

6. In the Organism field, type "zika" and touch it when it appears below the field.

7. Click the Exclude box.

Your Search Set should look like this:



The screenshot shows a web interface titled "Choose Search Set". It has two main sections: "Database" and "Organism". The "Database" section has a dropdown menu currently set to "Protein Data Bank proteins(pdb)". The "Organism" section is labeled "Optional" and has a text input field containing "Zika virus (taxid:64320)". To the right of the input field is a checked checkbox labeled "Exclude" and a plus sign icon.

Excluding structures from Zika viruses will help you find drugs bound to other viruses because Zika virus structures will be excluded from the search.

8. Scroll down and look in the program section to which algorithm is selected by default.

Q16. Which algorithm is selected?

9. If you have an account at the NCBI and you've been using BLAST, the conditions from your last search are often stored at the NCBI. If you see yellow boxes around settings in the General Parameters area, click "Restore default search parameters" to return the algorithm parameters to the default state.

10. Click BLAST to start the search.

11. When the search is complete, look at the top section of your BLAST results.

Q17. How long was your query sequence? Include the units

Q18. Capture an image of the conserved domains graph and circle the part of the graph that corresponds to your protein (either the NS3 protease or NS5 polymerase).

Paste the graph in your notes.

12. Look at the conserved domain graph.

Q19. Where does the conserved domain that corresponds to your sequence begin and end?

You can give approximate numbers based on the graph. This section of the sequence contains either the NS3 protease or the NS5 polymerase based on the choice you made in Q14.

A. Zika protease: Peptidase is another word for protease. If you picked the protease, you can see which part of this protein corresponds to the peptidase (or protease). Look at the numbers above the query sequence part of this graph. These are amino acid positions in your protein sequence.

B. Zika polymerase: If you picked the polymerase, the most important domain in the graph is the Flavi_NS5 superfamily domain. Look at the numbers above the query sequence part of this graph. These are amino acid positions in your protein.

13. Scan the data table and blast graph to find proteins that both align to either the protease or polymerase sections of your protein AND are also bound to a drug.

Identifying structures that are bound to drugs, or inhibitors that could be made into drugs, can be tricky. The best way to find these is to scroll down past the table and look in the alignments themselves.

Some of the key words you want to look for are phrases like "with compound" or words like "bound," "inhibitor," "drug," or words that end in "-vir."

Sometimes proteins are bound to substances that are not drugs, like Sah, GTP, or ATP. (You might remember that ATP and GTP are nucleotides). SAH stands for S-adenosyl methionine. This is not a drug.

Sometimes alignments are hidden because multiple structures contain proteins with identical sequences. If you see an alignment like this:

Chain A, Dengue V
Sequence ID: [pdb|4V0C](#)
[▶ See 2 more title\(s\)](#)

Click the “See X more title(s)” link to see if any of those structures might contain a drug or inhibitor.

Sometimes drug names contain numbers.

You should also make sure that the structure you choose contains the correct part of the sequence (the protease or polymerase). Use the values from your conserved domain graph that you recorded in Q20 and the values in your alignment to check this.

12. Pick a protein structure to investigate. To be a good candidate:

a. The structure must match the right part of the sequence.

b. The structure must contain a drug or inhibitor.

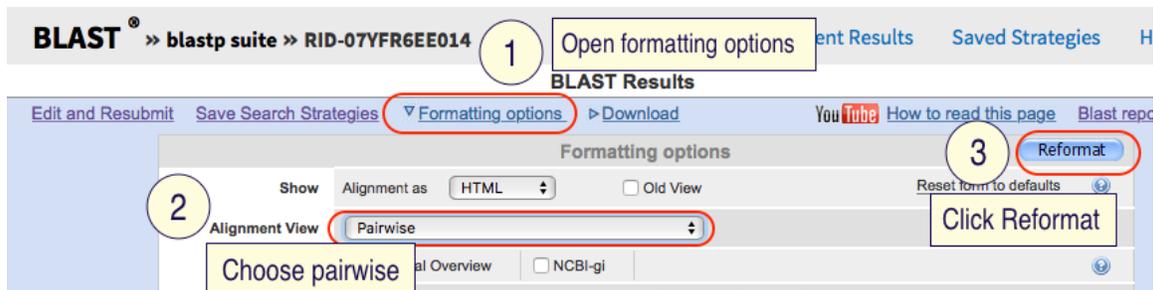
In each case, the sequence that you get from the GenBank record will contain a domain that matches proteins that are neither proteases nor RNA polymerases. Be sure you pick a protein that matches the correct domain in conserved domain graph.

If you don't understand whether the description tells you a structure is bound to a drug or not, you can send me the description of your structure and the PDB ID and I will tell you if it contains a drug or inhibitor.

Q20. Record the description or title of your structure. Underline the part of the description that suggests your protein is bound to a drug.

13. Before going farther, check the formatting option for your BLAST results and make sure that the Alignment view is set to Pairwise.

The formatting settings are located at the top of the BLAST results page as shown below.



14. Scroll down to find the alignment between your query sequence and the protein sequence of the structure you picked.

15. Use your mouse to select and copy the text for the alignment.

Q21. Paste the aligned sequences in your assignment. Format the text so that the font is Courier and the size is 9 point.

```

Query  274  NMKIIGRRIERIRNEHAETWFLDENHPYRTWAYHGSEYAPTQGSASSLVNGVVRLLSKPW  333
        N + I +RI++++ E A TW  D  HPYRTW YHGSYE  GSASSLVNGVV+L+SKPW
Sbjct  9      NQEKIKKRIQKLKEEFATTWHKDPHEPYRTWTYHGSYEVKATGSASSLVNGVVKLMSKPW  68

Query  334  DVVTGVTGIAMTDTTPYGQORVFKEKVDTRVDPQEGTRQVMNIVSSWLWKELGKRKRPR  393
        D + VT +AMTDTTP+GQORVFKEKVD+ P+P  G R+V+N  ++WLW  L + KRPR
Sbjct  69      DAIANVTTMAMTDTTPFGQORVFKEKVDTKAPEPPAGVREVLNETTNLWLAHLSREKRPR  128

Query  394  VCTKEEFINKVRSNAALGAIIEEEKEWKTAVEAVNDPRFWALVDREREHHLRGECHSCVY  453
        +CTKEEFI KV SNAALGA+F E+ +W TA EAVNDPRFW +VD ERE+HLRGECH+C+Y
Sbjct  129     LCTKEEFIKKVNSNAALGAVFAEQNQWSTAREAVNDPRFWEMVDEERENHLRGECHTCIY  188

Query  454  NMMGKREKKQGEFGKAKGSRAIWMWLGARLEFEALGFLNEDHWMGRENSGGGVEGLGL  513
        NMMGKREKK GEFKAKGSRAIW+MWLGAR+LEFEALGFLNEDHW+ RENSGGVEG G+
Sbjct  189     NMMGKREKKPGEFGKAKGSRAIWMWLGARYLEFEALGFLNEDHWSRENSGGGVEGSGV  248

Query  514  QRLGYILEEMNRPAGGKMYADDTAGWDTRISKFDLENEALITNQMEEGHRTLALAVIKYT  573
        Q+LGYIL ++  GGKMYADDTAGWDTRI++ DLENEA +  ++ HR LA A+I+ T
Sbjct  249     QKLGYLRLDIAGKQGGKMYADDTAGWDTRITRDLNEAKVLELLDGEHRMLARAIIELT  308

Query  574  YQNKVVKVLRPAEGGKTVMDIISRQDQSGSQVVVYALNTFTNLVVQLIRNMEAEVLEM  633
        Y++KVVKV+RPA  GKTVM+ISR+DQSGSQVVVYALNTFTN+ VQL+R MEAE V+
Sbjct  309     YRHKVVKVMRPAEAGKTVMDVISREDQSGSQVVVYALNTFTNIAVQLVRLMEAEVIGP  368

Query  634  QDLWLLRKPEKVT--RWLQSNQWDRDKRMAVSGDDCVVKPIDDRFAHALRFLNDMGKVRK  691
        Q L  L + K+  WL  NG +R+ RMA+SGDDCVVKP+DDRFA AL FLN M KVRK
Sbjct  369     QHLEQLPRKNKIAVRTWLFENGEERVTRMAISGDDCVVKPLDDRFATALHFLNAMSKVRK  428

Query  692  DTQEWKPSWGWSNWEVFPFCSHHFNKLYLKDGRSIVVPCRQDELIGRARVSPGAGWSIR  751
        D QEWKPS GW +W++VPFCS+HF ++ +KGRSIVVPCR QDELIGRAR+SPGAGW+++
Sbjct  429     DIQEWKPSHGWHDWQVFPFCSNHQEIIVMKDGRSIVVPCRQDELIGRARISPGAGWNVK  488

Query  752  ETACLAKSYAQMWQLLYFHRRDLRLMANAICSAVPVDWVPTGRTTWSIHKGEMTTEDM  811
        +TACLAK+YAQM W LLYFHRRDLRLMANAICSAVPVDWVPTGRT+WSIH KGEWMTTEDM
Sbjct  489     DTACLAKAYAQM W LLYFHRRDLRLMANAICSAVPVDWVPTGRTSWSIHSKGEWMTTEDM  548

Query  812  LMVWNRVWIEENDHMDKTPVTKWTDIPYLGKREDLWCGSLIGHRPRTTWAENIKDTVNM  871
        L VWNRVWIEEN+ M DKTP+  WTD+PY+GKRED+WCGSLIG R R TWAENI  +N
Sbjct  549     LQVWNRVWIEENEWMMDKTPIASWTDVPPYVGKREDIWCGLIGTRSRATWAENIYAAINQ  608

Query  872  VVRIIGDEEKYMDYLSTQVRY  892
        VR +IG +E Y+DY+++ RY
Sbjct  609     VRAVIG-KENYVDYMTSLRRY  628

```

Your formatted sequence alignment should look like the example above. For my example, I used 4HDG. The 4HDG structure contains an RNA polymerase from the Japanese encephalitis virus bound to GTP. GTP is not a drug. It's a nucleotide that gets

used in making RNA. I'm just using this structure as an example of a case where the virus RNA polymerase is bound to another molecule.

In the example, the Query sequence is from the Zika virus protein and the Subject sequence is from the Japanese encephalitis virus protein. Notice that aligned portion of the Zika sequence starts at nucleotide 274. When I look at the conserved domains graph, I can see that this position is in the right domain and corresponds to the polymerase.

The letters in between the two sequences represent amino acids that are found in those positions in both proteins. When you see a +, it indicates that the amino acid at the position is similar in terms of its chemistry, but it's not identical.

In the last section, you'll use that aligned sequence to determine if the residues in your structure that bind to a drug are also present in the Zika virus protein.

Part IV. Determine if your drug could bind the Zika virus protein.

In this last part, we will identify amino acid residues in our protein that bind to the drug or inhibitor and see if those same residues are present in the Zika virus protease or RNA polymerase. The way we do this is to select the drug, then use the program to select amino acids that are near the drug. Then, we find those residues in our aligned sequences and see if they're present in proteins from both viruses.

1. Use your PDB ID to find your structure in the MMDB.

In my example, I'm using 4HDG.

2. Click the citation title in the MMDB record to view and read the abstract.

Q22. What does the abstract say about the ability of the bound substance (drug/inhibitor/chemical) to inhibit replication of the virus?

A good example, of this is with the structure 4WTG. If I look at that structure in the MMDB and click the title, I get to the abstract for this article:

"Viral replication. Structural basis for RNA replication by the hepatitis C virus polymerase"

The first sentence of the article tells me that nucleotide analog inhibitors have been clinically successful in treating hepatitis C virus. That means that they work. In the last sentence, my abstract says that drug in 4WTG, sofosbuvir, forms a clinically active metabolite. That tells me it has an affect.

3. Download the structure of the protein.
4. Locate the drug / inhibitor in the structure and select it.
5. Select the drug.

If you're using Molecule World, touch the name of the drug in the sequence window to select it. You may need to look at the MMDB record to determine which abbreviation corresponds to the drug.

6. Select amino acids that are near the drug.

Molecule World: Touch the Selection button, touch Select nearby.

The Select nearby option locates and highlights objects that can be found within 5-6 angstroms of a selected drug or residue. It's like finding using Google maps to find pizza places within 5 miles of home.

7. Change the coloring style to residue and the drawing style to space fill.

8. Deselect and hide amino acids that are farther away from the drug so that you end up with only a few amino acids (between 4-12) that are very close to the drug. These residues should almost appear to touch the drug.

Q23. Find 12 or more amino acids that contact the drug and highlight them in your sequence alignment (Q21).

An example is shown below with an alignment between the Zika virus query sequence and a protein sequence from 4HDG. Amino acids that contact the drug are highlighted.

You can use a combination of the residue position and the Find command in Word to find these sequences. Use the formatting command in Word to highlight these residues.

Query	274	NMKIIGRRIERIRNEHAETWFLDENHPYRTWAYHGSYEAPTQGSASSLVNGVVRLLSKPW	333
		N + I +RI++++ E A TW D HPYRTW YHGSYE GSASSLVNGVV+L+SKPW	
Sbjct	9	NQEKIKKRIQKLKEEFATTWHKDPHEPYRTWTYHGSYEVKATGSASSLVNGVVKLMSKPW	68
Query	334	DVVTGVTGIAMTDTPYGGQORVFKEKVDTRVDPQEGTRQVMNIVSSWLWKELGKRKRPR	393
		D + VT +AMTDTP+GQORVFKEKVD+ P+P G R+V+N ++WLW L + KRPR	
Sbjct	69	DAIANVTTMAMTDTPFGQORVFKEKVDTKAPEPPAGVREVLNETTNWLWAHLSREKRPR	128
Query	394	VCTKEEFINKVRSNAALGAIIEEEKEWKTAVEAVNDPRFWALVDREREHHLRGECHSCVY	453
		+CTKEEFI KV SNAALGA+F E+ +W TA EAVNDPRFW +VD ERE+HLRGECH+C+Y	
Sbjct	129	LCTKEEFIKKVNNSNAALGAVFAEQNWSTAREAVNDPRFWEMVDEERENHLRGECHTCIY	188
Query	454	NMMGKREKKQGEFGKAKGSRRAIWMWLGARLFLEFEALGFLNEDHWMGRENSGGGVEGLGL	513
		NMMGKREKK GEFGKAKGSRRAIW+MWLGAR+LEFEALGFLNEDHW+ RENSGGGVEG G+	
Sbjct	189	NMMGKREKKPGEFGKAKGSRRAIWFMWLGARYLEFEALGFLNEDHWSRENSGGGVEGSGV	248
Query	514	QRLGYILEEMNRPAGGKMYADDTAGWDTRISKFDLENEALITNQMEEGHRTLALAVIKYT	573
		Q+LGYIL ++ GGKMYADDTAGWDTRI++ DLENEA + ++ HR LA A+I+ T	
Sbjct	249	QKLGYLRLDIAGKQGGKMYADDTAGWDTRITRDLNEAKVLELLDGEHRMLARAI IELT	308
Query	574	YQNKVVVKVLRPAEGGKTVMDDIISRDQSGSQVVVYALNTFTNLVQLIRNMEAEVLEM	633
		Y++KVVKV+RPA GKTVMDD+ISR+DQSGSQVVVYALNTFTN+ VQL+R MEAE V+	
Sbjct	309	YRHKVVKVMRPAEAGGKTVMDDVISREDQSGSQVVVYALNTFTNIAVQLVRLMEAEVIGP	368
Query	634	QDLWLLRKPEKVT--RWLQSNWDRLRKMAVSGDDCVVKPIDDRFAHALRFLNDMGKVRK	691
		Q L L + K+ WL NG +R+ RMA+SGDDCVVKP+DDRFA AL FLN M KVRK	
Sbjct	369	QHLEQLPRKNKIAVRTWLFENGEERVTRMAISGDDCVVKPLDDRFATALHFLNAMSKVRK	428
Query	692	DTQEWKPKSTGWSNWEVFPFCSHHFNKLYLKDGRSIVVPCRHQDELIGRARVSPGAGWSIR	751
		D QEWKPS GW +W++VPFCS+HF ++ +KDGRSIVVPCR QDELIGRAR+SPGAGW+++	
Sbjct	429	DIQEWKPSHGWHWQQVFPFCSNHFQEIVMKDGRSIVVPCRQDELIGRARISPGAGWNVK	488
Query	752	ETACLAKSYAQMWQLLYFHRRDLRLMANAICSAVPVDWVPTGRTTWSIHKGWEMTTEDM	811
		+TACLAK+YAQM W LLYFHRRDLRLMANAICSAVPVDWVPTGRT+WSIH KGEWMTTEDM	
Sbjct	489	DTACLAKAYAQM W LLYFHRRDLRLMANAICSAVPVDWVPTGRTSWSIHSKGWEMTTEDM	548
Query	812	LMVWNRVWIEENDHMDKTPVTKWTDIPYLGKREDLWCGSLIGHRPRTTWAENIKDVTNM	871
		L VWNRVWIEEN+ M DKTP+ WTD+PY+GKRED+WCGSLIG R R TWAENI +N	
Sbjct	549	LQVWNRVWIEENEWMMDKTPIASWTDVPPYVGKREDIWCGLIGTRSRATWAENIYAAINQ	608
Query	872	VRIIGDEEKYMDYLSTQVRY	892
		VR +IG +E Y+DY+++ RY	
Sbjct	609	VRAVIG-KENYVDYMTSLRRY	628

10. Pick one interaction between one residue and the drug and hide everything else. If there are metals between the drug and the amino acid, keep them.

Q24. (6 pts) Pick one case where an amino acid interacts with the drug.

- a. Paste an image in your answer sheet that clearly shows the interaction between this amino acid and the drug. The image should be shown with the element coloring style.
- b. Record the name of the amino acid and the position in the protein. If there are two location values, specify whether you used the NCBI or PDB position.
- c. Describe the type of interaction or bond between an atom in the drug and the amino acid.
- d. Is this same amino acid present in the same position in the Zika virus sequence?
- e. What evidence supports the likelihood that your drug / inhibitor will bind to the Zika virus protein?

Your answer must include the number of residues from both viral proteins that contact the drug. Use the number of highlighted residues in (Q23).