**Piwi Matters**

By Laurel Lorenz1 and Shuchismita Dutta2\*

1Department of Molecular Biology, Princeton University, Princeton NJ 08544

2Institute of Quantitative Biomedicine, Rutgers University, Piscataway NJ 08854

\*contact author: sdutta@rcsb.rutgers.edu

**Part 3: Insights into Piwi’s Structure (Learning from Homologs)**

This part of the case focuses on learning about Piwi’s structure in order to understand how it mediates its functions. To find an experimentally determined structure of Piwi, we will explore the Protein Data Bank (PDB) at [www.rcsb.org](http://www.rcsb.org).

*Box 5: Resource*

RCSB Protein Data Bank (**RCSB PDB**, [www.rcsb.org](http://www.rcsb.org)) provides access to 3D structural data of biological macromolecules (proteins, nucleic acids, carbohydrates and their various complexes). In addition, it provides information about the experiment used to derive the data, details about the molecules included in the experiment, and links to various bioinformatics resources that can provide additional information about the protein/molecule of interest. Each structure in the PDB is identified by a unique identifier (called PDB ID). Atomic coordinates form the PDB can be visualized and analyzed using various visualization software (some available from RCSB PDB).

1. Search for *Drosophila* Piwi structure(s) in the PDB ([www.rcsb.org](http://www.rcsb.org)) to see what structure(s) is/are available.

Type the protein name “Piwi” in the top search box and hit on the search button. Once the results are returned, click on Drosophila melanogaster in the left hand menu to view only the structures that contain the fruit fly Piwi protein

Q1. List the PDB identifiers and titles for 2-5 structures that contain Drosophila Piwi.

1. The PDB structures may represent a single protein or a complex structure – i.e., the file may contain multiple polymer chains (protein, DNA, and RNA) that form a complex structure and are co-crystalized.

Select one of the PDB entries you listed above and open the structure summary page for this entry to explore the Piwi protein residues, peptides, and chains that are present in these structures.

*Box 6: Vocab*

**Residues**: Building blocks of biological macromolecules are sometimes referred to as residues. Depending on the context, this may refer to amino acids (frequent use) or a nucleotide (less common use).

**Protein** vs **Peptide**: while both proteins and peptides are composed of covalently linked amino acids, peptides are short (composed of 2-25 amino acids)

**Chains**: A chain (such as protein/ peptide chain) refers to covalently linked amino acids. To help locate amino acids in a protein structure or complex a chain is given an identifier (called Chain ID) and each amino acid in the chain is assigned a number. Note that large proteins with multiple domains may be cut into smaller pieces – a short peptide or a single domain and included in the experiment. For this experiment the polymer chain formed by the peptide or domain is assigned a chain ID. Nucleic acids or polymers of nucleotide residues (DNA/RNA) may also be referred to as chains and assigned chains IDs.

Q2. In the list of Piwi containing structures, you should find a structure with Piwi and Papi proteins. List the PDB ID you are exploring and answer the following questions.

To identify the polymer chains, present in this structure entry, examine the section titled Macromolecules. What is the Chain ID of Papi? What is the Chain ID of Piwi? What is the region of the Piwi protein that is present in this structure? (Refer back to your diagrams of Piwi as needed)

1. We will begin our Piwi structure and function exploration using our knowledge of the relationship between protein sequence, structure, and functions.

Q3. Do you know the relationship between protein sequence, structure, and functions? Describe it in 1-2 sentences.

*Box 7: Concept*

The foundation for **bioinformatics** is that protein **sequence** can be used to predict protein **structure**, which in turn can be used to predict protein **functions**. Thus, proteins with similar amino acid **sequences** tend to fold and have the same overall protein **structure** and usually perform similar **functions**. An application of this knowledge is that when experimentally determined structures of a protein (or part of a protein) is not available, the structure of another protein with similar sequences (i.e., a homolog) can be used to provide insights about its structure and function.

1. It appears that experimentally determined structures of the complete *Drosophila* Piwi protein are not available in the PDB (as of October 2019). Therefore, we will need to study similar proteins in other species (orthologues). We will query the PDB for structures of proteins with sequences that are similar to the *Drosophila* Piwi amino acid sequence.

* Go back to the UniProt page for Piwi. Click on the left hand menu tab called “Sequence”. Click on the tab on top labeled “FASTA” to download the protein sequence in FASTA format and copy the sequence and save it in a text file.
* Go to the RCSB PDB home page (<https://www.rcsb.org/>) and click on the Search option in the left hand menu >> Sequences. In the “Option B: Paste Sequence”. Paste the FASTA sequence saved above and click on the Run Sequence Search button.

Once the search results are returned look at the matches to answer the following.

*Box 8: Concept*

*How similar do the protein sequences have to be for such analysis?*

Analysis of large numbers of protein structures and sequences show that for proteins with ~100 amino acids, a sequence identity >25% is indicative of similar folds and functions (Sander and Schneider 1991\*). Finding structures of proteins/domains with >25% identity to the protein of interest, can be used to learn about the structure and function of the protein.

Note that the relationship between protein sequence, structure, and function is complex, and not fully understood yet. For example, there are known protein structures with less than 25% sequence identity that have highly conserved structures, and also there are proteins with high sequence identity that have different structures. Therefore, any predictions or models based on the structures of sequence neighbors should be experimentally tested.

\*Sander, C. and Schneider, R. (1991) Proteins: Structure, Function, and Genetics 9, 56-68

Q4. Fill in the following table for the top (best match) structure.

|  |  |
| --- | --- |
| PDB ID |  |
| Structure title |  |
| length (# of aa) that matches query |  |
| percentage of sequence identity |  |

Q5. Do you think that this structure can be used to learn about the structure and function of Piwi? Why or why not?

Open and explore the structure summary page for the top entry in the query results list.

Q6. Complete the following table for the PDB entry reported in Ans4.

|  |  |
| --- | --- |
| PDB ID |  |
| Author(s) of entry |  |
| Year when the structure was published/ released |  |
| Structure determination method |  |
| # of macromolecular chains  (# Protein + # nucleic acid) |  |
| Names of proteins in these chains (chain ID) |  |
| Names of genes of proteins included here |  |
| Name of any other macromolecular chain |  |

1. In the Macromolecules section of the structure summary page for the PDB entry you are exploring, click on the UniProt ID (orange box) for the protein chain in the structure.

Q7. What is the UniProt Identifier for this chain?

Q8: Examine the “Family & Domains” section of the protein chain listed in Q7 on the UniProt page. Draw a linear diagram of the protein and mark all the motifs, regions, and domains.

1. Click on the 3D view tab on the top of the structure summary page on the RCSB PDB website to visualize the molecule. Color the structure by Chain, using the various pull down menu options.

Q9. Take a screen-shot of the graphical window and insert the image here to show the overall shape of the protein complex. Describe in 1-2 sentence what you see.

*Box 9: Concept*

**Missing residues**: Due to specific conditions in the structure determination methods, most PDB entries do not have the coordinates of all atoms present in the structure. For example, flexible regions of the structure cannot be “seen” in the experiment, so these coordinates are missing. Also, all H atoms in most structures determined by X-ray crystallography are not included in the coordinate files. To learn more read <http://pdb101.rcsb.org/learn/guide-to-understanding-pdb-data/missing-coordinates-and-biological-assemblies>.

1. In order to visualize and analyze the protein structure(s) in greater detail we will use an online molecular structure visualization tool, called iCn3D.

*Box 10: Resource*

**iCn3D** (<https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html>) is a web-based visualization tool that allows users to directly open any structure from the Protein Data Bank (PDB) and visualize the structure. Users can interactively rotate the molecule/complex, select specific regions and represent them in different ways, compare structures, analyze interactions and make simple distance measurements.

* Go to iCn3D (<https://www.ncbi.nlm.nih.gov/Structure/icn3d/full.html>)
* Click on the button called File >> Retrieve by ID >> PDB ID so that a new window opens. Input the PDB ID of the structure you wish to visualize and click on Load.
* The structure opens in a new tab – rotate the molecule and examine the overall structure.

Q10. How many polymer chains do you see? Take and screenshot of the structure and include it below.

* Spend a few minutes playing around with the different pull-down menus to see some of the different ways this protein view can be adjusted.
* Figure out where the Piwi and PAZ domains are located in the 3D structure.
  + Click on the button called Windows >> View Sequences & Annotations.
  + Click on the Details button to see the one letter code sequence of all the protein chains in the structure.
  + Use the residue number limits identified for the Piwi and Paz domains of Siwi (from your answer to Q8 above). Click and drag on these residues in the sequence and annotation window. The residues will be highlighted in yellow, both in the sequence and graphics windows.
  + Click on Color >> Unicolor >> Green (for the PAZ domain). Now click on Select >> Clear Selection.
  + Similarly color the Piwi domain Yellow.

Q11. Save a picture of the Siwi structure after the domain coloring and include it here.

Q12. In relation to the Piwi and PAZ domains, where is the piRNA located in the structure?

1. We know that the sequences of Piwi and Siwi are similar. Aligning these sequences will help highlight where the conserved regions are located.

* Go to the Uniprot home page at <https://www.uniprot.org/> and click on the Align button in the top left corner. Alternatively, directly go to the alignment page at <https://www.uniprot.org/align/>. Paste the UniProt identifiers of Piwi (see Part 2, Ans 1) and Siwi (see Part 2, Ans 7), separated by a comma (or in two separate lines) and click on the Align button.
* Examine the sequence alignment of these two sequences. Turn on specific annotations on this sequence alignment by checking on boxes denoting specific criteria.

Q13. Based on the annotations seen on the sequence alignment complete the table:

|  |  |
| --- | --- |
| Annotation | Residue number in Fruit fly (Silk worm) |
| Active site conserved in both proteins |  |
| Mutagenesis studies done in both proteins |  |
| Protein domain that shows sequence conservation |  |

1. Referring back to the mutational studies discussed in Part -1 of the case the maximum impact on renewal of germline cells was seen in Mutation C or the Piwi-YK mutant had mutations in two amino acids - Y551L and K555E. Locate the corresponding amino acid residues in the Siwi structure to visualize their interactions.

* Locate the corresponding Siwi residues in the aligned sequence window and note down the residue numbers.

Q14. What do the residues Y551 and K555 correspond to in the Siwi structure?

*Box 11: Concept*

Biomolecular structural stability, interactions and functions are dependent on various non-covalent interactions. Some key interactions in molecular structures are

**Hydrogen bonds** - formed between two partially negatively charged atoms with a hydrogen atom between and covalently linked to one of them. e.g. in structures look for examples of O/N … H\_\_O/N, where … denotes hydrogen bond and \_\_ denotes a covalent bond

**Salt bridges** or **ionic interactions** - formed between oppositely charged amino acid side chains and/or charged ligands/ions. e.g. in structures look for interactions between Lys/Arg/His and Glu/Asp. These interactions may also involve phosphate groups and ions such as K+, Na+, Cl- etc.

**Hydrophobic interactions** - formed between hydrophobic amino acid side chains positioned away from the aqueous environment. e.g. look for regions with large numbers of carbon and hydrogen atoms in close proximity. Aliphatic amino acids such as Ala, Leu, Val, Ile participate in hydrophobic interactions.

**Pi stacking** - seen between amino acids with aromatic side chains. Pi clouds of aromatic rings interact with each other in staggered stacks, face to edge interactions. Alternatively they may also interact with positively charged amino acid side chains (pi-cation interaction).

Learn more at <https://earth.callutheran.edu/Academic_Programs/Departments/BioDev/omm/jsmolnew/bonding/chymo.html#Topic2>

* Go to the iCn3D window and identify these residues in the Sequences and Annotations window by clicking and dragging on them. When you release the mouse button these residues are highlighted in yellow in the graphics window.
* Click on the button Style >> Sidechains >> Stick. This should display the side-chains of the selected amino acids. Color >> Atom (to show the sidechains in CPK coloring scheme)
* Display H-bonding interactions between the selected amino acids and non-selected ones by clicking on View >> H-bonds & interactions. A new window pops-up. Keep the default options (i.e. first set selected and second set non-selected) and click on Display do see green dashed lines show up representing H-bonds.
* Clear all selections and take a screen shot or save images of the displayed structure

Q14. Include the screenshot of the interactions below.

Q15. Examine the structure in iCn3D and describe at least 2 key interactions that stabilize the selected Tyr and Lys residues. (Hint: when you display the interactions using the View >> H-bond and interactions options, remember to select the salt bridges box too!)

Save the image and label relevant interactions.

Note: If you have trouble identifying an amino acid using the mouse over option, hold down option on Mac or alt on PC and simultaneously select the amino acid. The amino residue and number will appear below the 3D structure (in the grey window).

1. In the Piwi mutant studies the Y551L and K555E mutant could not rescue the morphological changes caused by the piwi-null mutation.

Q16. Based on the structure explorations above, explain in 2-3 sentences the intermolecular forces that explain why the YK mutant could not perform piwi’s function. (Hint: Review the amino acid properties at <https://cdn.rcsb.org/pdb101/learn/resources/what-is-a-protein/what-is-a-protein-pres.pdf>. What amino acid properties are changed when Tyrosine (Y) is mutated to Leucine (L) and Lysine (K) is mutated to Glutamic Acid (E)?)

**Part 4:** **Structure Based Hypothesis Development**

Exploring the structure of Siwi (Piwi ortholog) allows us to predict additional amino acids that enable Siwi and Piwi to bind to piRNAs.

a. piRNA binding is required for asymmetric stem cell division because when Piwi Y551 and K555 are mutated, piRNA binding is disrupted and the ovaries are rudimentary in size. While these are results for one piRNA binding mutant, it is ideal to test a hypothesis in multiple ways – such as using multiple mutants. Using the Siwi structure, we will identify additional amino acids that are critical for piRNA binding and further hypothesize how to mutate them.

Q1. Where do you think the residues critical to Piwi’s self-renewal function are located in the piwi domain? Explain why.

Q2. Using the rationale, described in the above answer, identify 2 amino acids in Siwi (besides the Y and K discussed in part 3 of the case) that may be critical for its function. List the amino acid residue number and type and explain why you think they are important. Support your answers with figures (using the siwi structure).

(Hint: to make the analysis a little easier, select part of the Siwi or RNA chain and figure out which amino acids interact within 4 angstroms of the selected chain. Use the same steps for selection as used in Part 3: section i)

Q3. What would you mutate these amino acid residues to in order to test if they are functionally important? Predict what phenotype you would observe if you used these to make mutant flies.

Q4. For each of the mutants that you have proposed above, could you mutate the corresponding residues in piwi to affect piwi’s function? Explain your answer. (Hint: remember to consult the sequence alignment that you generated earlier in the case.)

b. To test whether a mutation identified in Q3 and Q4 affects piRNA germline stem cell asymetric division and piRNA binding, the following steps would need to be completed.

1. Generate a Piwi mutant DNA sequence by changing the codon for the original amino acid to a codon that will encode the new the new amino acid. This process will lead to an amino acid substitution in the Piwi protein.
2. Using the Piwi mutant sequence from step one, use CRISPR technology to create a transgenic fly. The new Piwi mutant fly will have the wild-type Piwi replaced with the mutant Piwi.
3. Test whether your Piwi mutant fly has well developed ovaries and egg chambers.
4. Test whether the Piwi mutant in the transgenic fly is able to bind to piRNA. To learn more about piRNA binding assays read details at: <http://www.genesdev.org/cgi/doi/10.1101/gad.209841.112>

Q5. For any 2 mutations that you have proposed in Q3 and Q4, identify a codon sequence for the original and substituted amino acid. You may use a genetic code table as necessary.

Original Amino Acid:

Codon sequence for Original Amino Acid:

Substituted Amino Acid:

Codon sequence for Substituted Amino Acid: