

Undergraduate Bioinformatics

Conceptualizing Form and Function on a Molecular Scale

By Paige Kramer* and JF Trembl†

Over the past 64 years, since the first actual, three-dimensional structure of a protein was determined by X-Ray diffraction in 1958,¹ the library of solved structures has grown to more than 170,000 (as of 2020).² With the expansion of our knowledge of structures and the advent of applications to visualize them, the tools to aid in our understanding of these as physical objects have greatly improved.

Nevertheless, the concept of proteins as the real, 3-dimensional objects that they are can be difficult for undergraduate biology students to grasp. In part, this may be due to the duality of these biomolecules' existence as both carriers of information and as physical actors. The understanding of the former lends itself to a common representation of proteins as mere strings of amino acids (AAs) written out as text, whereas the latter is often left to the imagination or reduced to abstract representations (Figure 1) that have no relationship to their actual structures.

We know that the sizes and shapes of AAs take up space in a protein and that their positions within this space allow for interactions between atoms which maintains a given shape and positions certain atoms in ways that they can interact with their substrates. One challenge of communicating this in foundational biology courses is the difficulty of rendering such a shape in three dimensions using a whiteboard or PowerPoint presentation. Animations, such as those maintained by the Howard Hughes Medical Institute's BioInteractive website, provide great insight into the molecular-scale world, however, there are only a small number of these available, and student interaction with them is entirely passive.

There exists a significant body of research describing how students utilize multiple learning styles to grasp new concepts (cite). Often, these learning styles are shorthand into the four categories of the 'VARCK' model as Visual, Auditory, Reading & Writing, and Kinesthetic styles of learning. While the reality of these discrete styles is debatable,³ students do have preferences for how they are presented with new information. When these preferences are given, students are more invested in the material and spend more time engaging actively with it.^{4,5} Therefore, the development of active, engaging instructional materials that utilizes all of the approaches of the VARCK model remains a valid objective. Specific data for the efficacy of this approach in regards to teaching protein structures demonstrated that students' grasp of these structures was improved when they were told to draw how proteins were structured versus those who were only asked to describe those structures in words.⁶ The authors acknowledge that the positive effects of their active lessons could be attributable to the increased investment made by students rather than the visual style being more or less superior.

Incorporating these ideas, the following is a walkthrough of a project designed to overcome the lack of sense for proteins as real objects. In doing this, students actively engage with these proteins as structures subject to mutations that may, or may not, alter their function. It uses publicly available tools to manipulate a primary sequence of AAs, solving for structure and function. These structures are then tested against mutations to determine how they will impact the stability of the protein. As a final deliverable, this project requires the production of a 3D model of the protein or some portion of it. All resources are called by name and the web addresses for each are found in the Resources section of the assignment, presented after the article.

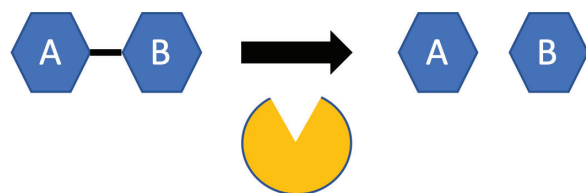


Figure 1 | Pac-man snipping a covalent bond.

*University of Kansas, Edwards Campus, Biotechnology

†Corresponding Author jtrembl@ku.edu

The following is a walkthrough of this assignment with the Programmed cell Death protein 1 (PD-1)¹¹ discussed at length in the previous review articles.

Student Assignment

Background: Protein Identification

Use the following resources to identify your protein and provide some background information

I. NCBI Gene Database⁷

1. Common Name
2. Host Species (*Genus species*)
3. NCBI Gene ID/Official Symbol
4. Short Description
 - i. This should be distilled from the NCBI Gene Summary and/or Wikipedia.
5. Expression Bar Chart
 - i. Click "See more" to view a bar chart containing data on where in the body's tissues the gene is expressed (as determined by RNA sequencing). Save and include this bar chart as the deliverable for this step.

II. Universal Protein Research Knowledgebase (UniProtKB)⁸

6. UniProt Entry Number
 - i. Follow the UniProt link in the Resources then search for the protein using the NCBI Gene ID
 - ii. Carefully select the result that best matches the gene and organism of interest by clicking on the blue entry number.
 - iii. This page will be used later to gather further details about the protein.

III. RCSB Protein Data Bank (PDB)⁹

7. RCSB PDB Solved Structure Identifier
 - i. Follow the RCSB PDB link in the Resources and search for the protein by either the common name or the NCBI Gene ID, making sure to select the organism of interest on the left.
 - ii. You must ensure that your chosen protein has an existing solved structure in this data bank in order to do a mutational analysis in later parts of this exercise.

IV. NCBI GenBank¹⁰

8. AA Protein Sequence
 - i. From the NCBI Gene page, go to the "Genomic regions, transcripts, and products" section and then click "GenBank" on the right. Scroll down to the first Coding Sequence "CDS" section and look directly after "/translation=" for the full protein sequence.
 - ii. Sequence needs to be in FASTA Format consisting of '>' followed by a simple name, a return, and then the sequence in one continuous line of text. See "FASTA Formatting" link in Resources.

Example Assignment

1. Common Name: PD-1
2. Species: *Homo sapien*
3. NCBI Gene ID: PDCD1
4. Description: An immune-inhibitory receptor expressed on activated T cells
5. Expression Chart: Figure 2
6. UniProt Number: Q15116
7. PDB Identifier: 2M2D
8. AA Protein Sequence:

```
>HumanPD1
MQIPQAPWPVVWAVLQLGWRPGWFL
DSPDRPWNPPTFSPALLVVTEGDNA
TFTCSFSNTSESFVLNWRMSPSNQ
TDKLAAFPEDRSQPGQDCRFVTLQL
PNGRDFHMSVVRARRNDSGYLCSGA
ISLAPKAQIKESLRAELRVTERRAE
VPTAHPSPSPRPAQGFQTLVVGVVG
GLLGLSLVLLVWVLAVICSRARGTI
GARRTGQPLKEDPSAVPVFSVDYGE
LDFQWREKTPEPPVPCVPEQTEYAT
IVFPSGMGTSSPARRGSADGPRSAQ
PLRPEDGHCSWPL
```

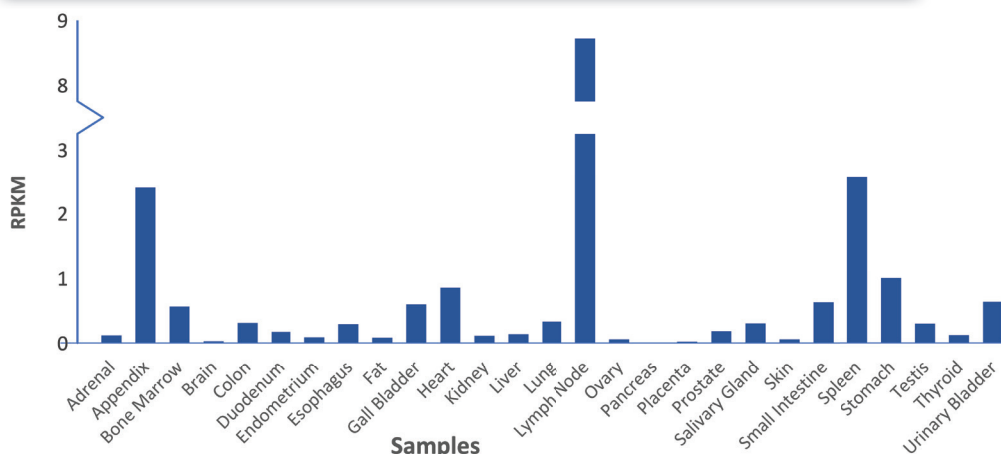


Figure 2 | Protein Expression of PD-1 as determined by RNA sequencing in various tissue types. This chart reports that expression is highest in the lymph nodes and spleen tissues. This is consistent with current knowledge that PD-1 is expressed on T cells and both of these secondary lymphoid organs are enriched for T cells of both CD4+ and CD8+ types.

Student Assignment

Signal Sequence and Transmembrane Regions

Use the following resources to identify various regions of your protein

V. SignalP 6.0 Server¹²

9. Predicted Signal Sequence

- Following the link in the Resources for SignalP 6.0, paste the AA sequence copied from step 8, and select the following options: Organism: Other; Output format: Long Output; and Model Mode: Slow.
- The results here will predict if a signal sequence is present on the protein, directing that ribosome towards the ER for synthesis.
- If this signal sequence is present, copy and paste the FASTA sequence here, making the [predicted signal sequence blue](#). If no signal sequence is present, indicate so.

10. Predicted Cut Site

- If a signal peptide is predicted, the cut site will be reported by SignalP as well.

11. Signal Sequence Probability Plot

- SignalP will also create a graphic representation showing the AAs predicted to be in this signal peptide region. Attach this plot here.

VI. Expsy ProtScale Web Tool¹³

12. Kyte-Doolittle Hydropathy Plot

- Follow the ProtScale link in the Resources to open the web tool. Making sure to **exclude** the predicted signal sequence (if present), copy and paste the remaining chain of AAs from Step 9 into the indicated box.
- Select the "Hydropath./Kyte & Doolittle" plot and leave all other options as the default then hit submit. Attach the resulting hydropathy plot as the deliverable for this step.

VII. UniProt Knowledgebase

13. Comparison of Predicted vs. Experimental Results

- Copy and paste the FASTA sequence from Step 8 here with the [predicted signal sequence in blue](#) for further editing.
- Return to the UniProt protein page and select the "Subcellular Location" section on the left and scroll down to the "Features" subheading.
- Here the experimentally determined extracellular, transmembrane, and intracellular regions of the protein can be found.
- On the FASTA sequence here, underline the extracellular region, make the [transmembrane region red](#), and the [intracellular region green](#).
- Note any discrepancies between the predicted and the actual regions of the protein.

Example Assignment

9. Predicted Signal Sequence:

```
>HumanPD1
MQIPQAPWPVVWAVLQLGWRPGWFL
DSPDRPWNPPTFSPALLVVTEGDNA
TFTCSFSNTSESFVLNWRMSPSNQ
TDKLAAFPEDRSQPGQDCRFRVTQL
PNGRDFHMSVVRARRNDSGTYLCGA
ISLAPKAQIKESLRAELRVTERRAE
VPTAHPSPSPRPAGQFQTLVVGVVG
GLLGSLLVLLVWLAVICRAARGTI
GARRTGQPLKEDPSAVPVFVSDYGE
LDFQWREKTPEPPVPCVPEQTEYAT
IVFPSGMTSSPARRGSADGPRSAQ
PLRPEDGHCSWPL
```

10. Predicted Cut Site: 97% probability for a cut site between AAs 24 and 25

11. Signal Sequence Plot: Figure 3

12. Hydropathy Plot: Figure 4

13. Predicted vs. Experimental:

```
>HumanPD1
MQIPQAPWPVVWAVLQLGWRPGWFL
DSPDRPWNPPTFSPALLVVTEGDNA
TFTCSFSNTSESFVLNWRMSPSNQ
TDKLAAFPEDRSQPGQDCRFRVTQL
PNGRDFHMSVVRARRNDSGTYLCGA
ISLAPKAQIKESLRAELRVTERRAE
VPTAHPSPSPRPAGQFQTLVVGVVG
GLLGSLLVLLVWLAVICRAARGTI
GARRTGQPLKEDPSAVPVFVSDYGE
LDFQWREKTPEPPVPCVPEQTEYAT
IVFPSGMTSSPARRGSADGPRSAQ
PLRPEDGHCSWPL
```

For PD-1, there is a discrepancy in the location of the cut site for the signal sequence between the predicted and the experimental results. The extracellular sequence identified by UniProt includes the final phenylalanine residue that SignalP predicted would be cut off as part of the signal sequence. This residue is marked by overlapping underlined and blue text above.

Student Assignment

3-Dimensional Modeling and Printing

Use the following resources to go from an AA sequence to a 3-D printed model

VIII. Swiss-Model Protein Structure Homology-Modeling Server^{14,15}

Swiss-Model is a fully automated server that references a database of proteins whose structure has been solved via X-Ray Crystallography. It then uses these solved structures as templates and combines them to generate a predictive model for new proteins.

Either the entire protein or just the extracellular or intracellular portions of the protein may be modeled, whichever is of greater interest.

14. 3-D Model of the Protein (.pdb file)

- i. To generate this model, follow the Swiss-Model link provided in the resources, paste the desired portion of the AA sequence from Step 17 into the first box, give the model a project title, and click "Build Model".
- ii. Once the model is done rendering, change the view on the right from "Cartoon" to "Surface".
- iii. Attach a screenshot of this entire page here. This view should show the template Swiss-Model used to construct the model (left) and the final image of the 3-D rendered protein (right).
- iv. To download the .pdb file, click the "Model 01" dropdown menu and select the first "PDB Format" option.

IX. UCSF ChimeraX Molecular Visualization Program^{16,17}

Swiss-Model creates protein database (.pdb) files, however, most 3-D printers need the file in a stereolithography (.stl) format.

15. Printable 3-D Model of the Protein (.stl file)

- i. Follow the ChimeraX link in the Resources and follow the instructions to download and install the "Latest Production Release" of this software.
- ii. This program is only available for mac iOS however alternative software such as Jmol or PyMol can be used if using a Windows- or Linux-based machine (links in Resources).
- iii. Once this program is installed and open, go to "File" then "Open" and open the .pdb file from Swiss-Model for the protein.
- iv. Once the file is open, select "Molecule Display" from the top banner then select "Electrostatic" in the "Coloring" section at the top.
- v. Attach a clear screenshot of this protein here. Go to the "Right Mouse" section in the top banner and use the "Movement" tools to gain a better view of the protein.
- vi. Next, go to "File" then "Save..." then select "STL (3D Printing)" for the type of file. Save this file for the next step.

X. 3-D Printer

The Johnson County Central Resource Library's MakerSpace program allows all members to print one object a week free of charge.

16. Printed 3-D Model

- i. Using a 3-D printing software that is paired with the available 3-D printer, upload the .stl file and then print the protein.

Example Assignment

14. 3-D Model (.pdb file): Figure 5 and Figure 6a
15. 3-D Model (.stl file): Figure 6b
16. Printed 3-D Model: Figure 6c



If you find our students' work as inspiring as we do

DONATE

Consider sponsoring a student or our program by contacting Jack Treml at jtreml@ku.edu

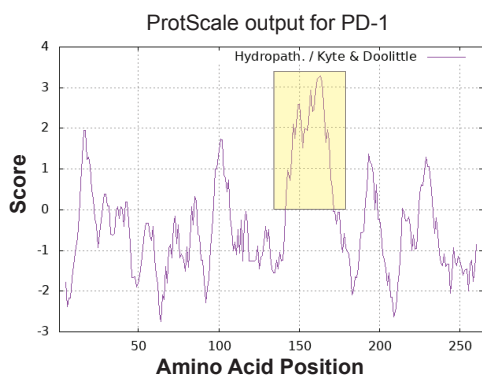
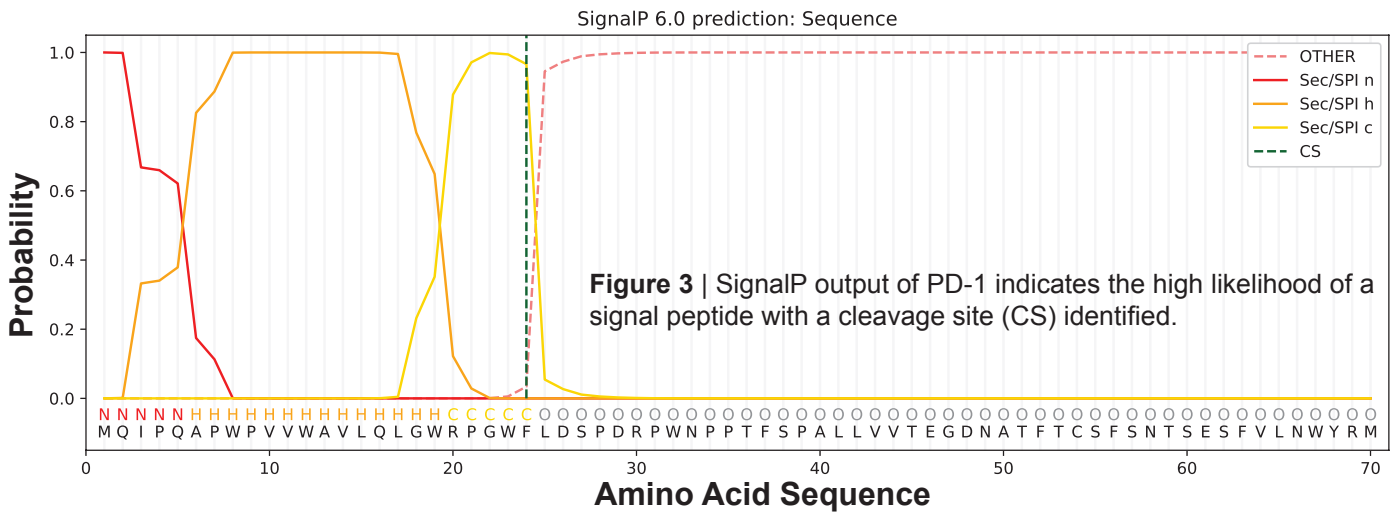


Figure 4 | Kyte-Doolittle hydropathy analysis of the mature PD-1 protein. As this analysis is done, a clear picture emerges of a trans-membrane, hydrophobic expanse of 21 AAs (highlighted above in yellow).

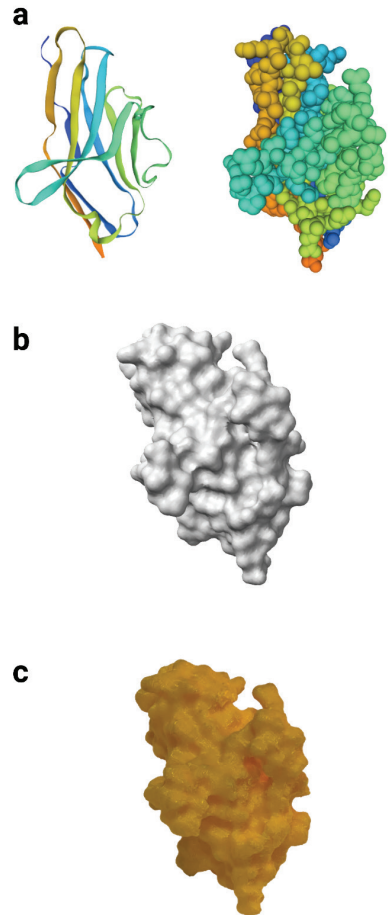


Figure 6 | Renderings and 3D Printed model of extracellular portion of PD-1. (a) Ribbon and Surface renderings, PDB File. (b) ChimeraX STL File. (c) Photograph of 3D printed protein.

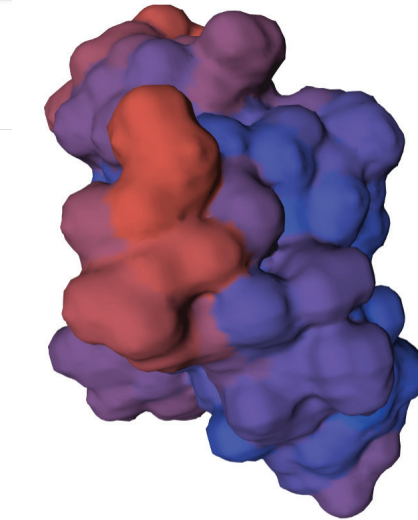


Figure 5 | Screenshot of output from swiss-Model of extracellular region of PD-1 analysis.

Student Assignment

Experimental Analysis of Mutations

The goal for this portion of the assignment is to better visualize and understand how different mutations can affect proteins in different ways and it will require critical thinking along with some trial and error.

For these next deliverables, use what is known of the properties of AAs (size, charge, hydrophobicity) to ascertain one tolerated and one non-tolerated mutation in the protein.

XI. RCSB PDB

The tools being used can only perform mutational analysis on a portion of a protein if that portion has a solved structure in the RCSB Protein Data Bank. Therefore for Steps 17-XX, use the FASTA format given by this data bank.

17. Solved RCSB FASTA Sequence

- Return to the RCSB PDB page for the protein, click the "Display Files" button on the right of the "Structure Summary" page, and then click "FASTA Sequence".
- It is of note that the crystallized structure may be a fraction of the total protein. This sequence may also be expressed with an artificial start codon or other tags of AAs used for isolation of the recombinant protein. These AAs will not be found within the original sequence and should be ignored.
- Paste the AA sequence from Step 13 here then **bold** the portion of the sequence that corresponds to the FASTA sequence from the RCSB PDB.

XII. Sorting Intolerant From Tolerant (SIFT) Sequence Tool¹⁸

The SIFT Sequence tool is an AA substitution tool that predicts whether any given AA change will alter the structure and function of a particular protein based on sequence homology and the physical and electrochemical properties of the AAs themselves.

18. Tolerated Mutation Example

- Follow the SIFT link in the Resources. Under "Single Protein Tools" on the right, select the "SIFT Sequence" tool.
- Paste the AA sequence **excluding the signal sequence** in FASTA format (including the name) copied from Step 17 into the first box.
- In the second box, enter an AA mutation that is predicted to be tolerated. The format is to have X#Y where X is the original AA, # is the position of the mutation, and Y is the new AA. When determining the AA position, for this resource specifically, start counting from the first AA in the *submitted* sequence.
- It is essential that the AA residue picked is within the **bold** portion of the sequence.
- Leave all other parameters on this page the same and click "Submit".
- If this AA substitution was predicted to be tolerated, copy and paste the resulting text here.
- If this AA substitution was reported as "affecting the protein's function" and is therefore non-tolerated, evaluate why that might be and try again.

19. Reason For Being Tolerated

- Propose an explanation for why this mutation would be tolerated based on the structural and electrochemical properties of the AAs involved.

Example Assignment

17. Solved FASTA Sequence:

```
>HumanPD1
MQIQAPWPVVWAVLQLGWRPGWFL
DSPDRPWNPPTFSPALLVVTEGDNA
TFTCSFSNTSESFVLNWRMSPSNQ
TDKLAAFPEDRSQPGQDCRFVTLQ
PNGRDFHMSVVRARRNDSGYLCGA
ISLAPKAOIKESLRAELRVTERRAE
VPTAHPSPSPRPAGQFQTLVGVVVG
GLLGSVLVLLVWVLAVICSRARGTI
GARRTGQPLKEDPSAVPVFSVDYGE
LDFQWREKTPEPPVPCVPEQTEYAT
IVFPSGMGTSSPARRGSADGPRSAQ
PLRPEDGHCSWPL
```

18. **Tolerated Mutation:** L18I "Substitution at pos. 18 from Leucine-to-Isoleucine is predicted to be TOLERATED with a score of 0.51. Median sequence conservation: 3.46. Sequences represented at this position: 5"

19. **Explanation:** This result is reasonable as Leucine and Isoleucine are both hydrophobic amino acids with similar structures (Figure 7).

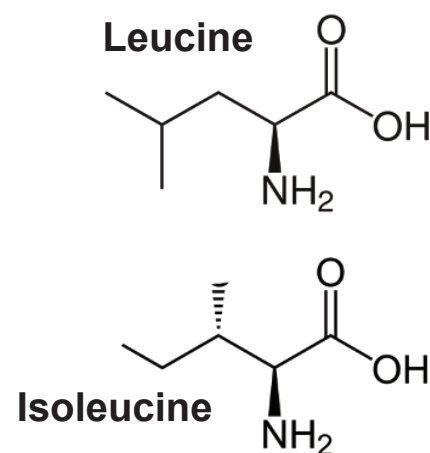


Figure 7 | Chemical structures of Leucine and Isoleucine showing the similarity in their structures.

Student Assignment

Experimental Analysis of Mutations Cont'd

20. Non-Tolerated Mutation Example
 - i. Repeat Step 18, but with the intent of finding an AA substitution that would be non-tolerated and would affect the protein's function.
21. Reason For Being Non-Tolerated
 - i. Propose an explanation for why this mutation would be non-tolerated based on the structural and electrochemical properties of the AAs involved.

XIII. Site-Directed Mutator (SDM) Server¹⁹

SDM is a web server used to predict the effects of various mutations on a specific protein's structure and function. A wild-type AA structure along with a proposed AA substitution are input and a stability score is calculated.

22. Tolerated Mutation Comparison

- i. Follow the SDM link in the Resources.
- ii. In the "Single Mutation" section enter the following:
 - First text box: the PDB Identifier from Step 7
 - Second: Enter the AA mutation that was predicted to be **Tolerated** by SIFT. The format is the same, however, when determining the AA position, for this resource specifically, start counting from the first AA in the Step 17 sequence.
 - Third: For most cases, entering "A" here will work
- iii. Click the green "Run SDM" button
- iv. The results of this prediction will be on the top left under "Predicted pseudo $\Delta\Delta G$:" This will reflect how this specific mutation will the form and possibly function of the protein.
- v. Use the interactive prediction tool on the right to examine how the substituted AA physically fits into the folded protein.
- vi. Capture a screenshot of this entire page and attach it here along with a statement comparing these results to those previously obtained from SIFT.

23. Non-Tolerated Mutation Comparison

- i. Repeat Step 22, but with the mutation that was predicted by SIFT to be **Non-Tolerated**.

Example Assignment

20. Non-Tolerated Mutation: V20W

"Substitution at pos. 20 from Valine-to-Tryptophan is predicted to AFFECT PROTEIN FUNCTION with a score of 0.02. Median sequence conservation: 3.46. Sequences represented at this position: 5"

- #### 21. Explanation:
- This result is reasonable as Valine and Tryptophan have very different structures where Valine is a branched-chain hydrophobic amino acid and histidine is a charged ring structure (Figure 8).

22. Tolerated Mutation Comparison:

Consistent with the SIFT results above, a Leucine-to-Isoleucine mutation does not destabilize the protein. Instead, it is predicted to increase the stability of the structure. Also see Figure 9.

23. Non-Tolerated Mutation Comparison:

Consistent with the SIFT results above, a Valine-to-Tryptophan mutation results in a destabilized protein. Also see Figure 10.

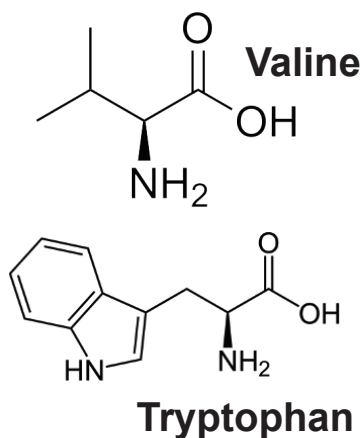


Figure 8 | Chemical structures of Valine and Tryptophan showing the differences in their structures.

Predicted pseudo $\Delta\Delta G$:
0.5 (Increased stability)

Mutation:
PDB name: 2M2D.pdb
Chain: A
Wild-type: LEU
Position: 42
Mutant-type: ILE

Environment	Wild type	Mutant
Secondary structure	E	E
Solvent accessibility (%)	20.1	24.6
DEPTH (Å)	4.4	4.4
OSP	0.40	0.43
HBOND_SS	False	False
HBOND_SN	False	False
HBOND_SO	False	False



Figure 9 | Screenshot of the SDM output for the tolerated Leucine-to-Isoleucine mutation of PD-1 that was predicted to be tolerated by SIFT.

Student Assignment

Experimental Analysis of Mutations Cont'd

XIV. PinSnps Tool for Human Protein Networks and SNPs²⁰

While the two previous mutational analysis were done using random, user-defined mutations, in the final portion of this assignment, use the PinSnps tool to investigate what actual, clinical mutations exist and appreciate their corresponding health consequences. Note that Single Nucleotide Polymorphisms (SNPs) are commonly observed mutations that may result in a change to the amino acid called for at a specific location. Many known SNPs are associated with pathologies.

24. Pathological Mutation ID with Reference

- Follow the PinSnps link in the Resources and clear out the default entries in the first three boxes.
- Copy and paste the UniProt Entry Number from Step 6 into the first box and click "Query".
- Shown below are any published SNPs or SNVs (Single Nucleotide Variants) that are known, sorted into five sections. Expand any of these sections containing identified mutations by clicking the plus button on the right.
- If possible, identify one of these known mutations that is associated with a predicted non-benign/"possibly damaging" outcome.
- Follow the link for that mutation by clicking on the "Mutation ID"
- Record the Mutation ID here.
- Scroll down to the References sections and also include the Reference Title, Author, year, and Journal of any one of the sources where this mutation was described.

Example Assignment

24. Mutation ID and Reference:

COSV57690725. "Integrated Molecular analysis of clear-cell renal cell carcinoma. Sato Y et al. 2013. Nature Genetics."

- The UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 49, D480–D489 (2021).
- Berman, H. M. et al. The Protein Data Bank. *Nucleic Acids Res.* 28, 235–242 (2000).
- GenBank [Internet]. Bethesda (MD): National Library of Medicine (US), National Center for Biotechnology Information; [1982] - [cited 2022 Sept 21]. Available from: <https://www.ncbi.nlm.nih.gov/nucleotide/>
- Hillier, L. W. et al. Generation and annotation of the DNA sequences of human chromosomes 2 and 4. *Nature* 434, 724–731 (2005).
- Nielsen, H., Tsirigos, K. D., Brunak, S. & von Heijne, G. A Brief History of Protein Sorting Prediction. *Protein J.* 38, 200–216 (2019).
- Kyte, J. & Doolittle, R. F. A simple method for displaying the hydropathic character of a protein. *J. Mol. Biol.* 157, 105–132 (1982).
- Bienert, S. et al. The SWISS-MODEL Repository—new features and functionality. *Nucleic Acids Res.* 45, D313–D319 (2017).
- Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 46, W296–W303 (2018).
- Goddard, T. D. et al. UCSF ChimeraX: Meeting modern challenges in visualization and analysis. *Protein Sci. Publ. Protein Soc.* 27, 14–25 (2018).
- Pettersen, E. F. et al. UCSF ChimeraX: Structure visualization for researchers, educators, and developers. *Protein Sci. Publ. Protein Soc.* 30, 70–82 (2021).
- Ng, P. C. & Henikoff, S. SIFT: predicting amino acid changes that affect protein function. *Nucleic Acids Res.* 31, 3812–3814 (2003).
- Worth, C. L., Preissner, R. & Blundell, T. L. SDM—a server for predicting effects of mutations on protein stability and malfunction. *Nucleic Acids Res.* 39, W215–W222 (2011).
- Lu, H.-C., Herrera Braga, J. & Fraternali, F. PinSnps: structural and functional analysis of SNPs in the context of protein interaction networks. *Bioinforma. Oxf. Engl.* 32, 2534–2536 (2016).

Predicted pseudo $\Delta\Delta G$:
-1.82 (Reduced stability)

Mutation:
PDB name: 2M2D.pdb
Chain: A
Wild-type: VAL
Position: 44
Mutant-type: TRP

Environment	Wild type	Mutant
Secondary structure	E	E
Solvent accessibility (%)	14.1	53.8
DEPTH (Å)	4.6	3.8
OSP	0.44	0.26
HBOND_SS	False	False
HBOND_SN	False	False
HBOND_SO	False	False

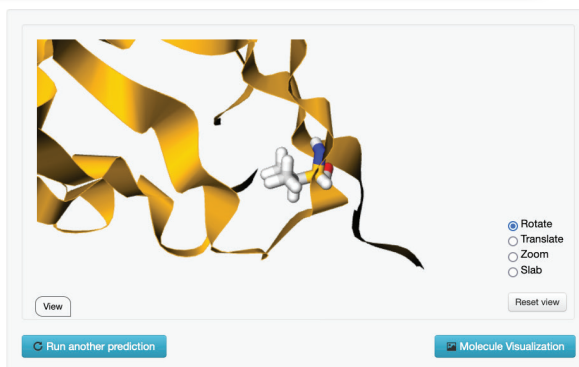


Figure 10 | Screenshot of the SDM output for the non-tolerated Valine-to-Tryptophan mutation of PD-1 that was predicted to affect protein function by SIFT.

References

- Kendrew, J. C. et al. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. *Nature* 181, 662–666 (1958).
- 'The game has changed.' AI triumphs at solving protein structures. <https://www.science.org/content/article/game-has-changed-ai-triumphs-solving-protein-structures>.
- Learning styles: where's the evidence? - Rohrer - 2012 - Medical Education - Wiley Online Library. <https://onlinelibrary.wiley.com/doi/10.1111/j.1365-2923.2012.04273.x>.
- Wang, Y. & Ji, Y. How do they learn: types and characteristics of medical and healthcare student engagement in a simulation-based learning environment. *BMC Med. Educ.* 21, 420 (2021).
- Chet Meyers & Jones, T. Promoting Active Learning: Strategies for the College Classroom. (Jossey-Bass Inc.).
- Leopold, C. & Leutner, D. Science text comprehension: Drawing, main idea selection, and summarizing as learning strategies. *Learn. Instr.* 22, 16–26 (2012).
- National Center for Biotechnology Information. (1988).

ChimeraX Acknowledgement

Molecular graphics and analyses performed with UCSF ChimeraX, developed by the Resource for Biocomputing, Visualization, and Informatics at the University of California, San Francisco, with support from National Institutes of Health R01-GM129325 and the Office of Cyber Infrastructure and Computational Biology, National Institute of Allergy and Infectious Diseases.