WESTERN
WASHINGTON UNIVERSITY

CSCI 474/597K
Bioinformatics

Filip Jagodzinski

Announcement

Article 3 Summaries/Critiques

- Engelman A, Englund G, Orenstein JM, Martin MA, and Craigie R. Multiple effects of mutations in human immunodeficiency virus type 1 integrase on viral replication. Journal of Virology, 1995 May; Volume 69, pgs 2729–2736.
- Edgar RC. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Research, 2004 Mar; Volume 19, pgs 1792-1797.
- Reva B, Antipin Y, and Sander C. Predicting the functional impact of protein mutations: application to cancer genomics. Nucleic Acids Research, 2011 Sep; Volume 39, pgs 1-14.
- Thompson JD, Higgins DG, and Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Research, 1994 Nov; Volume 11, pgs 4673-80.

Pick one of these publications, and write a summary and critique.

Write a summary and critique that is easily digestible by somebody who has NOT read the article. Still focusing on the WHAT, WHY, and HOW

You'll be asked to assess your peer's summaries and critiques

Course "Summary"

DNA sequence alignment, scoring, etc.
Protein Sequence and Structure
Analysis of Structure (mutations)
Simulation of Motion

Course Design : learn by doing

Course Project

Course Project

- Proposal Presentation
- Proof of Concept
- Initial Results
- Final Presentation
- Peer Assessment
- Final Report

Course Project

- Proposal Presentation
- Proof of Concept
- Initial Results
- Final Presentation
- Peer Assessment
- Final Report

Submitted via Canvas
Due 15 May

- Slide 1 : Title
- Slide 2 : Introduction (WHAT)
- Slide 3 : Motivation (WHY)
- Slide 4 : Methods (HOW)
- Slide 5 : Expected Obstacles (these most likely will change ONCE you begin)
- You must mention at least 2 references

This step will require you to do a quick literature search to better understand what has been done previously and/or the tool(s) or method(s) that you'll be using

Course Project

- Proposal Presentation
- Proof of Concept
- Initial Results
- Final Presentation
- Peer Assessment
- Final Report

Submission Via Canvas
Due 22 May
short report (1 page)

- Mention data found/used
- Method(s) being tested/used … should have been able to implement

Course Project

- Proposal Presentation
- Proof of Concept
- Initial Results
- Final Presentation
- Peer Assessment
- Final Report

Submission Via Canvas
Due 22 May
short report (1 page)

- Mention data found/used
- Method(s) being tested/used … should have been able to implement

This step will require you to do a quick literature search to better understand what has been done previously and/or the tool(s) or method(s) that you'll be using

Course Project

- Proposal Presentation
- Proof of Concept
- Initial Results
- Final Presentation
- Peer Assessment
- Final Report

Submission Via Canvas
Due 22 May
short report (1 page)

- Mention data found/used
- Method(s) being tested/used … should have been able to implement

This step will require you to do a quick literature search to better understand what has been done previously and/or the tool(s) or method(s) that you'll be using

Highly recommended : different people take on different roles … implement, assess, write, etc. … everybody should NOT be doing everything, but likewise don't do things in isolation

Course Project

- Proposal Presentation
- Proof of Concept
- Initial Results
- Final Presentation
- Peer Assessment
- Final Report

Submission Via Canvas
Due 29 May
short report (1 page)

- Discuss initial results
- Expected Obstacles / Obstacles Encountered
- Next steps

Course Project

- Proposal Presentation
- Proof of Concept
- Initial Results
- Final Presentation
- Peer Assessment
- Final Report

Due last week of classes
Submission .pptx or .pdf via canvas

- Zoom presentations : 5 minutes
- Introduction
- Motivation
- Methods
- Results
- Discussion

Course Project

- Proposal Presentation
- Proof of Concept
- Initial Results
- Final Presentation
- Peer Assessment
- Final Report

You will be asked to assess (via Canvas) the contribution of your group members

Course Project

- Proposal Presentation
- Proof of Concept
- Initial Results
- Final Presentation
- Peer Assessment
- Final Report

Due on 12 June
Submission via Canvas

- Introduction
- Motivation
- Methods
- Results
- Conclusions
- Next Steps

Don’t make the final report your magnum opus

If you’ve done the group work leading up to the final presentation, just stich together all parts, do a few read throughs, etc.

Course Project

- Proposal Presentation
- Proof of Concept
- Initial Results
- Final Presentation
- Peer Assessment
- Final Report

Due on 12 June
Submission via Canvas

- Introduction
- Motivation
- Methods
- Results
- Conclusions
- Next Steps

Any team with one or more graduate students

The final report must use the ACM sigconf LaTeX template
(http://www.acm.org/publications/proceedings-template)

Course Project

Group Selection

- Each group (ideally) must have at LEAST 1 CS, and 1 bio/chem non-CS person
- 1 Person groups are not allowed

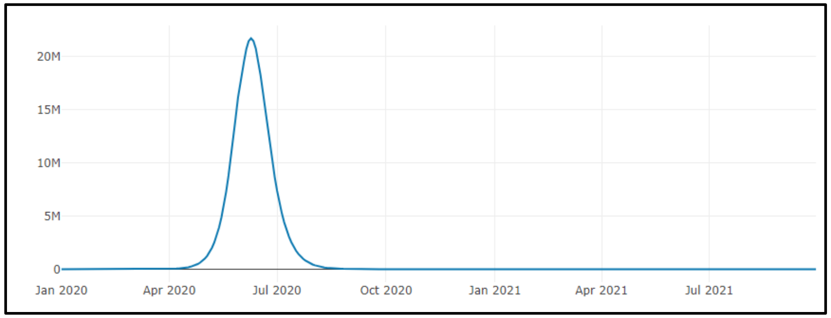
- Project 1 : SEIR Model sensitivity analysis
- Project 2 : SEIR Model extension; at risk and not (as) at risk populations, and social distancing measures
- Project 3 : SEIR Model Threshold scenarios
- Project 4 : Energy Profiles of Mutants (single AA substitution) of Covid-19 proteins
- Project 5 : Energy Profiles of Mutant(s) (double AA substitutions) of Covid-19 proteins
- Project 6 : Data Visualization / web server, for Projects 4 and 5 data
- Project 7 : PDZ domain, mutation analysis

Course Project

Project 1 : SEIR Model sensitivity analysis

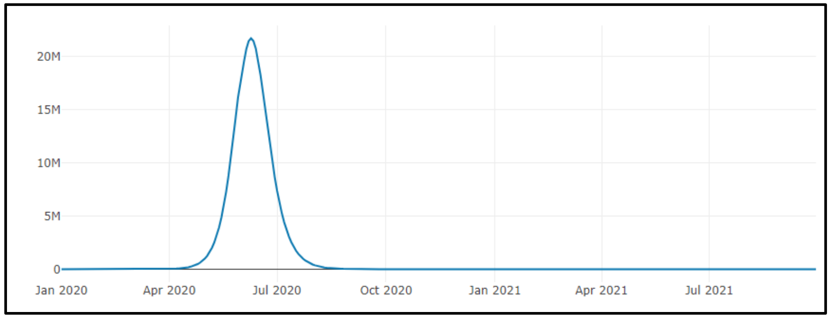
Course Project

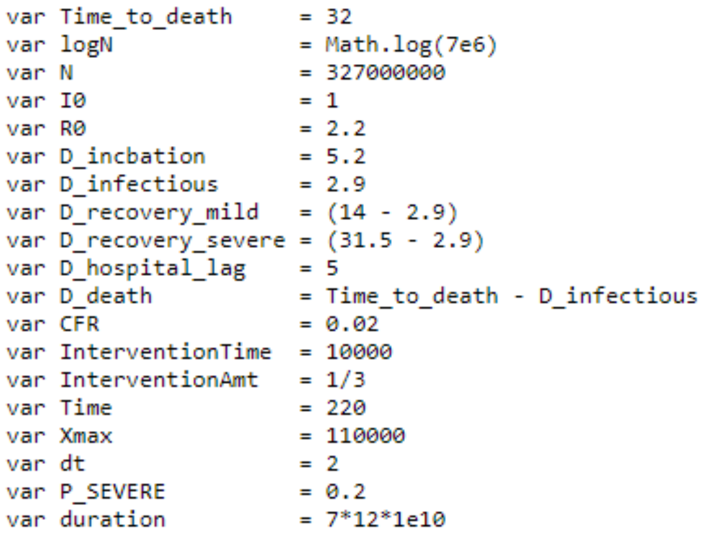
Project 1 : SEIR Model sensitivity analysis



Course Project

Project 1 : SEIR Model sensitivity analysis

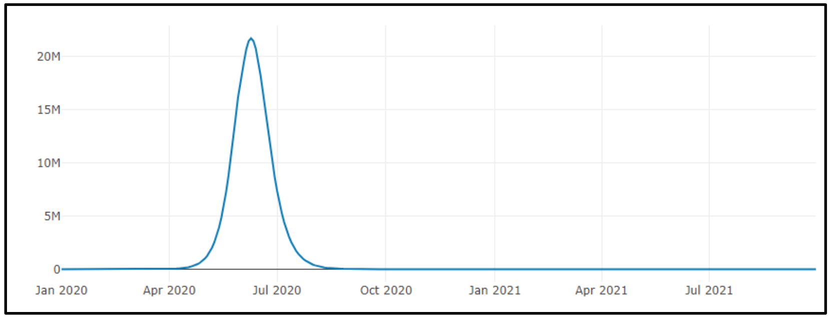


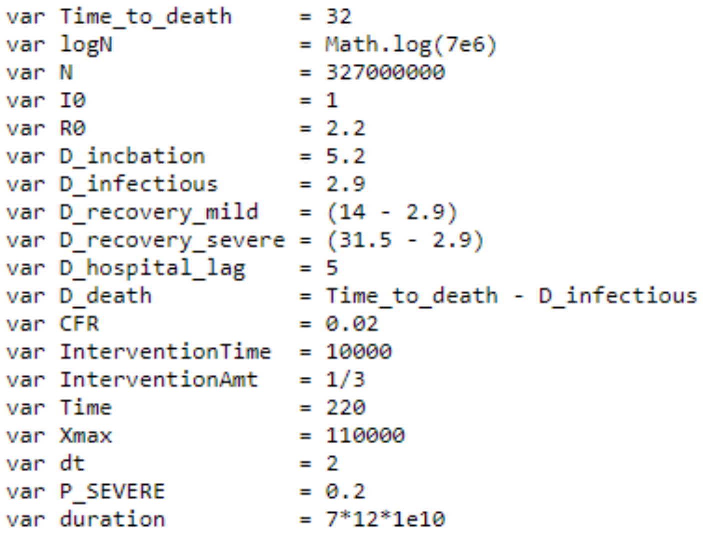


```
var Time_to_death     = 32
var logN              = Math.log(7e6)
var N                 = 327000000
var I0                = 1
var R0                = 2.2
var D_incbation       = 5.2
var D_infectious      = 2.9
var D_recovery_mild   = (14 - 2.9)
var D_recovery_severe = (31.5 - 2.9)
var D_hospital_lag    = 5
var D_death           = Time_to_death - D_infectious
var CFR               = 0.02
var InterventionTime  = 10000
var InterventionAmt   = 1/3
var Time              = 220
var Xmax              = 110000
var dt                = 2
var P_SEVERE          = 0.2
var duration          = 7*12*1e10
```

Course Project

Project 1 : SEIR Model sensitivity analysis





```
var Time_to_death      = 32
var logN               = Math.log(7e6)
var N                  = 327000000
var I0                 = 1
var R0                 = 2.2
var D_incbation        = 5.2
var D_infectious       = 2.9
var D_recovery_mild    = (14 - 2.9)
var D_recovery_severe  = (31.5 - 2.9)
var D_hospital_lag     = 5
var D_death            = Time_to_death - D_infectious
var CFR                = 0.02
var InterventionTime   = 10000
var InterventionAmt    = 1/3
var Time               = 220
var Xmax               = 110000
var dt                 = 2
var P_SEVERE           = 0.2
var duration           = 7*12*1e10
```

- Q: How sensitive is this model to each of these variables?
- Q: What are the ranges of “valid” values for R0, I0, hospital Lag?
- Q: What does the S, E, I and R in SEIR refer to?

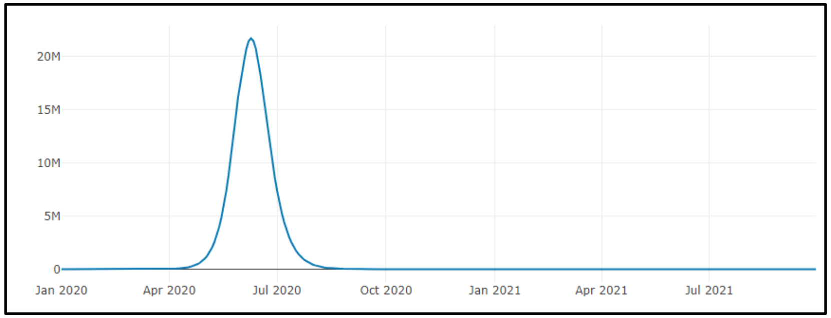
- Code: https://facultyweb.cs.wwu.edu/~jagodzf/covid-19/
- Rewrite code (java, python, C) … your choice
- For a subset of these variables, do exhaustive runs of the model, and report on outliers

Course Project

Project 2 : SEIR Model extension; at risk and not (as) at risk populations, and social distancing measures

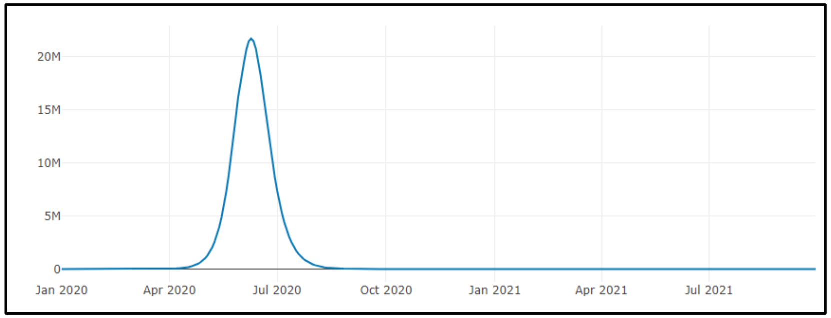
Course Project

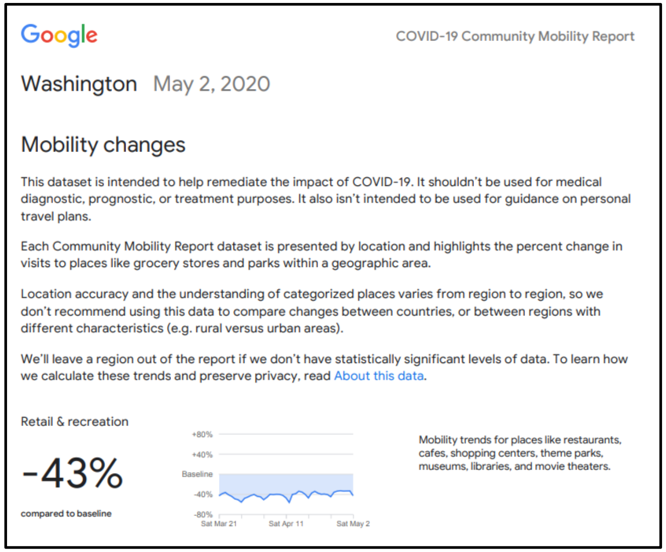
Project 2 : SEIR Model extension; at risk and not (as) at risk populations, and social distancing measures



Course Project

Project 2 : SEIR Model extension; at risk and not (as) at risk populations, and social distancing measures





COVID-19 Community Mobility Report

Washington May 2, 2020

Mobility changes

This dataset is intended to help remediate the impact of COVID-19. It shouldn't be used for medical diagnostic, prognostic, or treatment purposes. It also isn't intended to be used for guidance on personal travel plans.

Each Community Mobility Report dataset is presented by location and highlights the percent change in visits to places like grocery stores and parks within a geographic area.

Location accuracy and the understanding of categorized places varies from region to region, so we don't recommend using this data to compare changes between countries, or between regions with different characteristics (e.g. rural versus urban areas).

We'll leave a region out of the report if we don't have statistically significant levels of data. To learn how we calculate these trends and preserve privacy, read About this data.

Retail & recreation

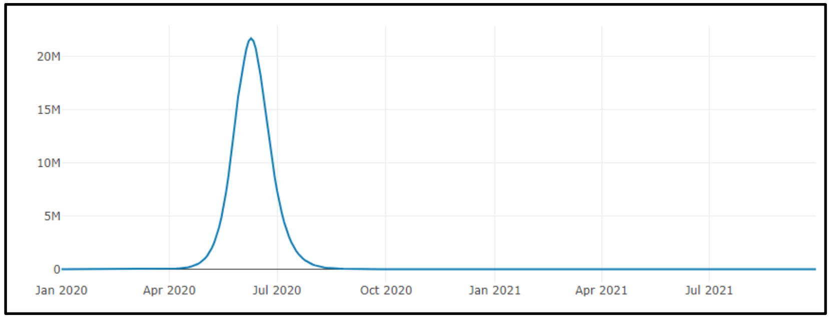
-43%

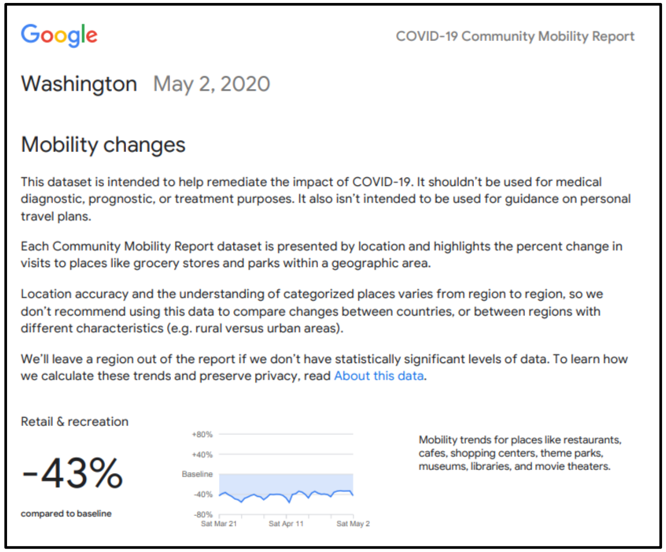
compared to baseline

Mobility trends for places like restaurants, cafes, shopping centers, theme parks, museums, libraries, and movie theaters.

Course Project

Project 2 : SEIR Model extension; at risk and not (as) at risk populations, and social distancing measures





Google COVID-19 Community Mobility Report

Washington May 2, 2020

Mobility changes

This dataset is intended to help remediate the impact of COVID-19. It shouldn't be used for medical diagnostic, prognostic, or treatment purposes. It also isn't intended to be used for guidance on personal travel plans.

Each Community Mobility Report dataset is presented by location and highlights the percent change in visits to places like grocery stores and parks within a geographic area.

Location accuracy and the understanding of categorized places varies from region to region, so we don't recommend using this data to compare changes between countries, or between regions with different characteristics (e.g. rural versus urban areas).

We'll leave a region out of the report if we don't have statistically significant levels of data. To learn how we calculate these trends and preserve privacy, read About this data.

Retail & recreation

-43%

compared to baseline

Mobility trends for places like restaurants, cafes, shopping centers, theme parks, museums, libraries, and movie theaters.

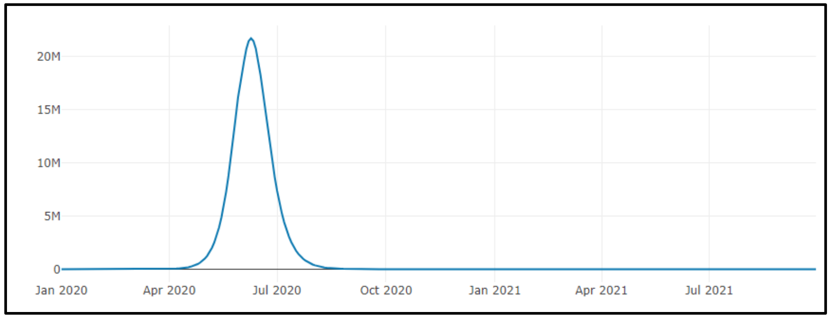
- Q: How can the "R" of SEIR be extended to take into account at risk versus (not as much) at risk populations?
- Code: https://facultyweb.cs.wwu.edu/~jagodzf/covid-19/
- Rewrite code (java, python, C) … your choice (can work with Project 1 team)

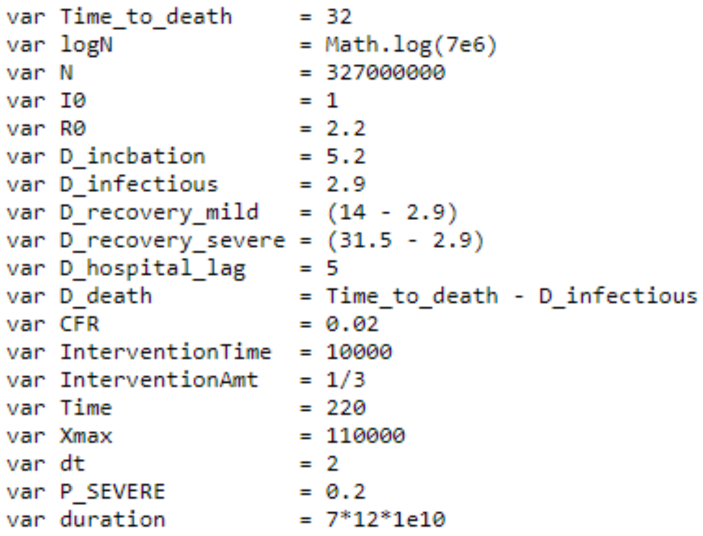
Course Project

Project 3 : SEIR Model Threshold scenarios

Course Project

Project 3 : SEIR Model Threshold scenarios





```
var Time_to_death       = 32
var logN                = Math.log(7e6)
var N                   = 327000000
var I0                  = 1
var R0                  = 2.2
var D_incbation         = 5.2
var D_infectious        = 2.9
var D_recovery_mild     = (14 - 2.9)
var D_recovery_severe = (31.5 - 2.9)
var D_hospital_lag      = 5
var D_death             = Time_to_death - D_infectious
var CFR                 = 0.02
var InterventionTime    = 10000
var InterventionAmt     = 1/3
var Time                = 220
var Xmax                = 110000
var dt                  = 2
var P_SEVERE            = 0.2
var duration            = 7*12*1e10
```

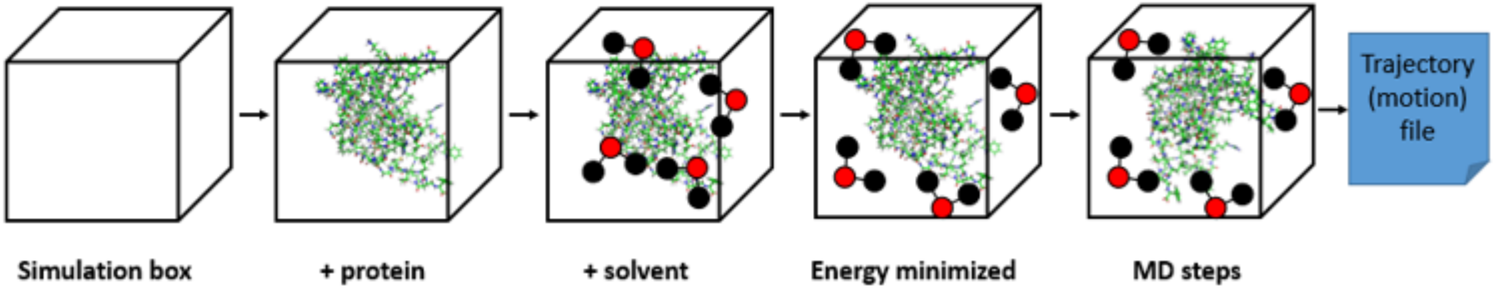
- Q: What are the range(s) of variable values such that a threshold of number of infected people is NOT reached by time t_1, t_2, t_3, etc.
- Note that between t_1 and t_2, for example, there may be a different R0 than between t_2 and t_3, due to social distancing measures
- Code: https://facultyweb.cs.wwu.edu/~jagodzf/covid-19/
- Rewrite code (java, python, C) … your choice (can work with Project 1 and 2 teams)

Course Project

Project 4 : Energy Profiles of Mutants (single AA substitution) of Covid-19 proteins

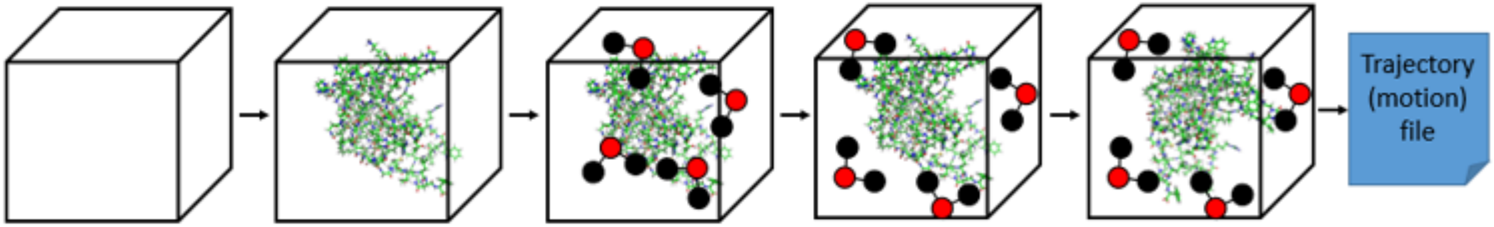
Course Project

Project 4 : Energy Profiles of Mutants (single AA substitution) of Covid-19 proteins



Course Project

Project 4 : Energy Profiles of Mutants (single AA substitution) of Covid-19 proteins



Trajectory (motion) file

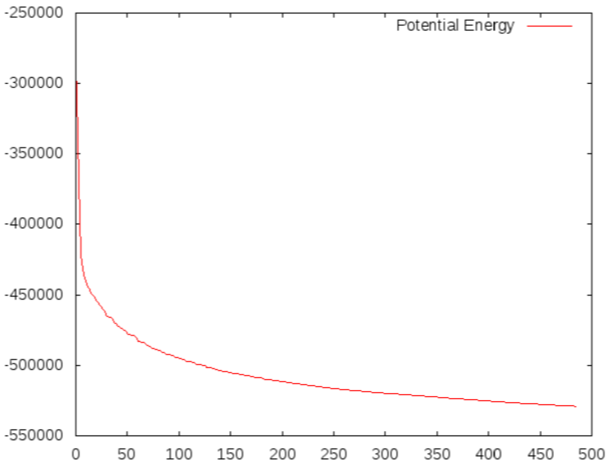
Simulation box

+ protein

+ solvent

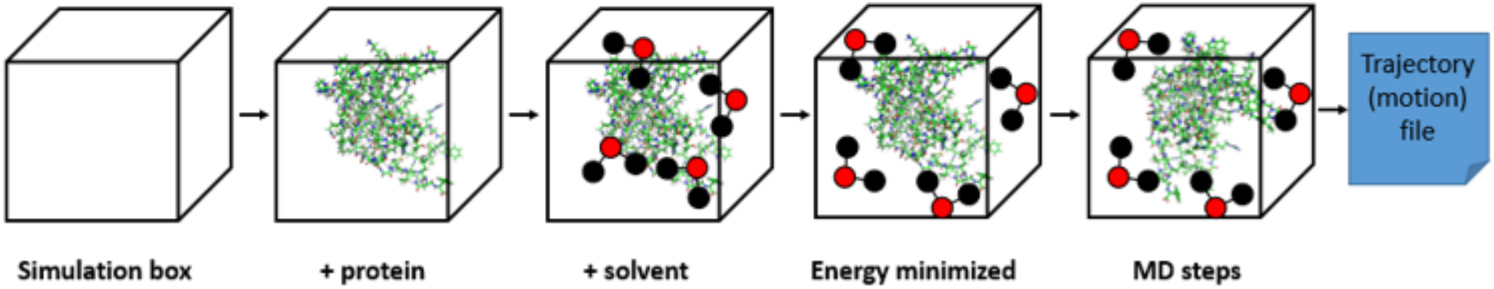
Energy minimized

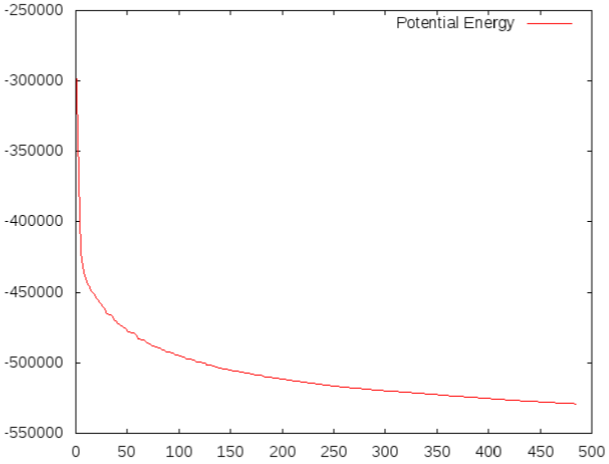
MD steps

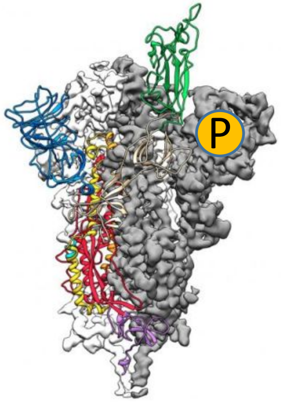


Course Project

Project 4 : Energy Profiles of Mutants (single AA substitution) of Covid-19 proteins

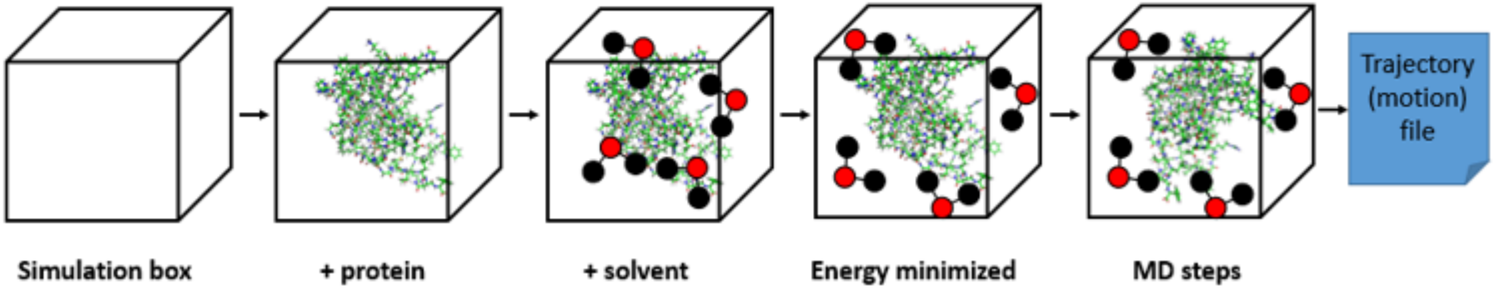




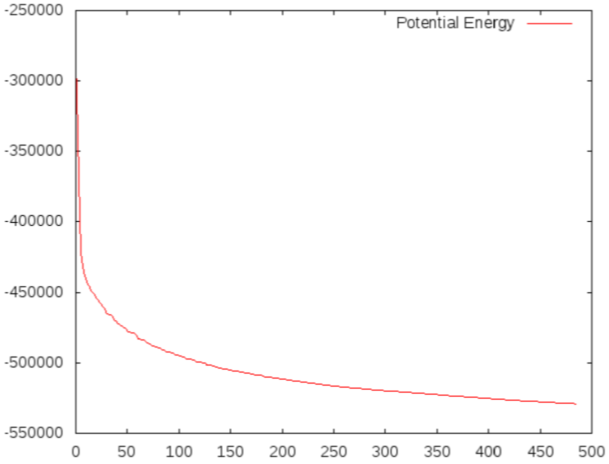


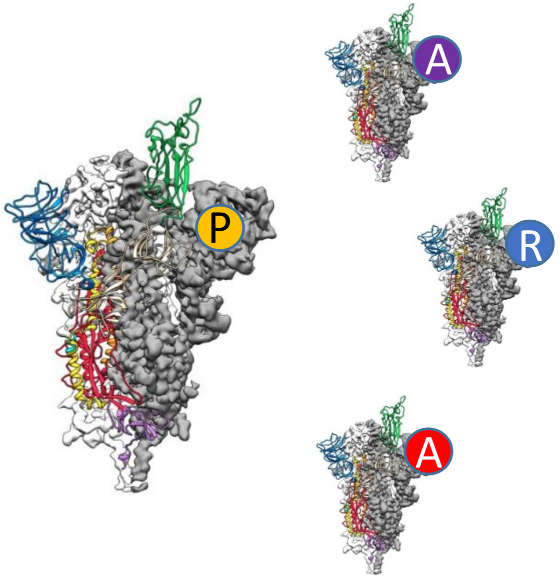
Course Project

Project 4 : Energy Profiles of Mutants (single AA substitution) of Covid-19 proteins



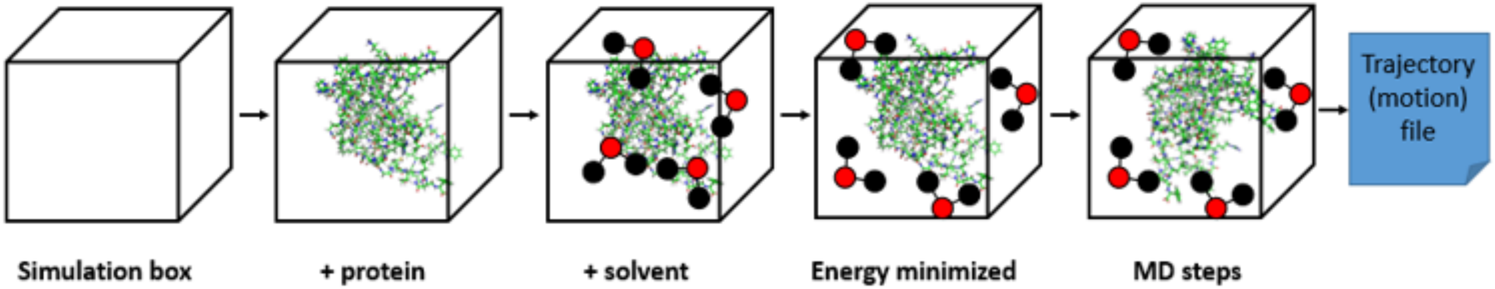
Simulation box → + protein → + solvent → Energy minimized → MD steps → Trajectory (motion) file

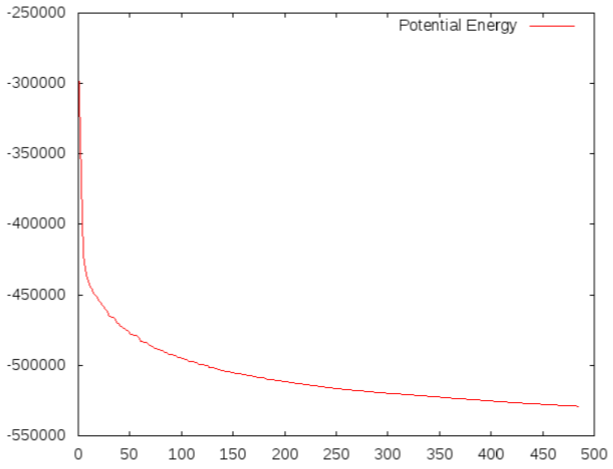


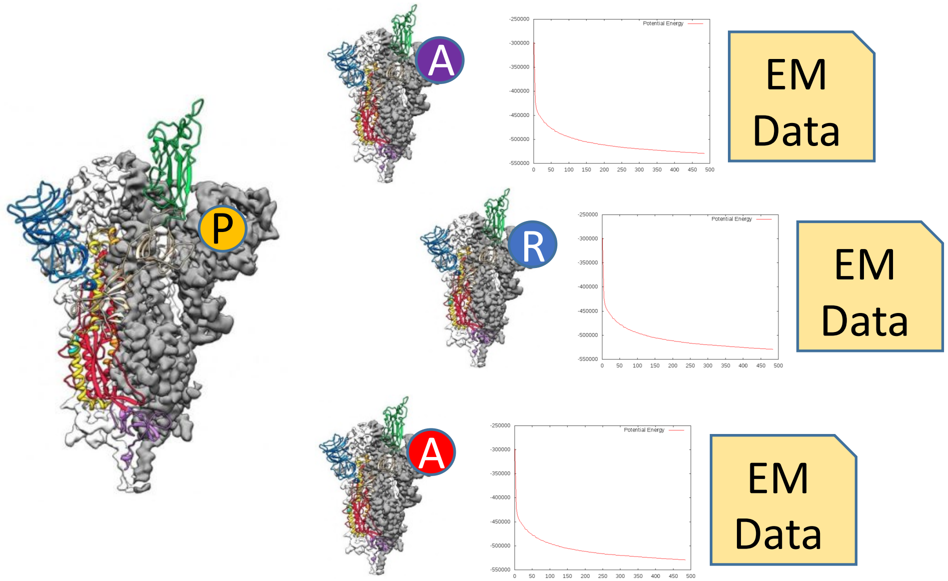


Course Project

Project 4 : Energy Profiles of Mutants (single AA substitution) of Covid-19 proteins







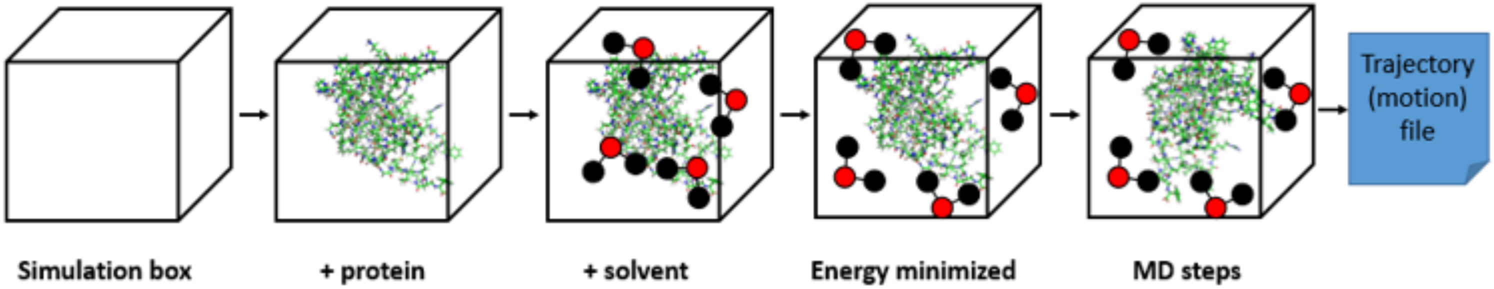
EM Data

EM Data

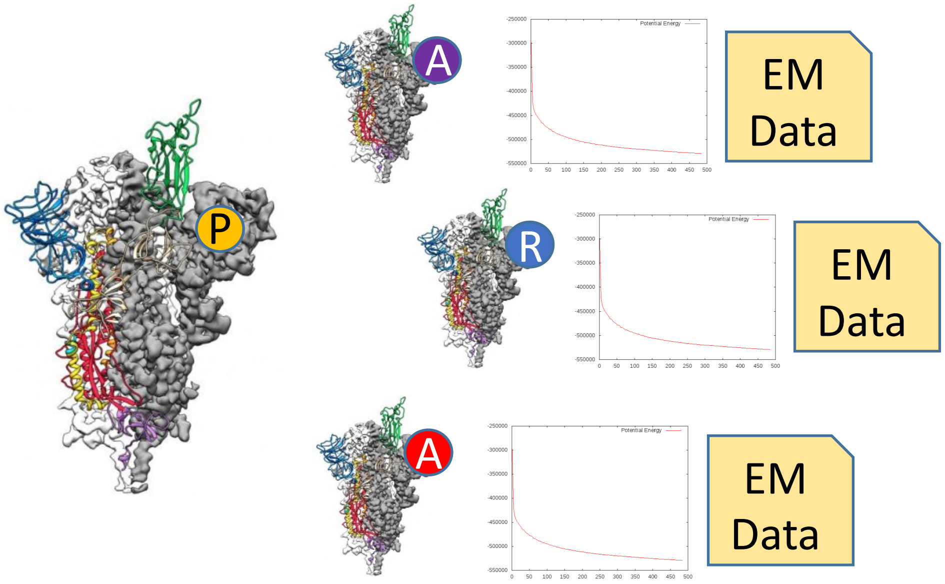
EM Data

Course Project

Project 4 : Energy Profiles of Mutants (single AA substitution) of Covid-19 proteins



- Generate datasets of EM data for ALL single-point mutations in Covid-19 proteins
- Aggregate the results
- **Q: Which residue(s) when mutated are the most impactful in affecting the structural stability of a protein?**

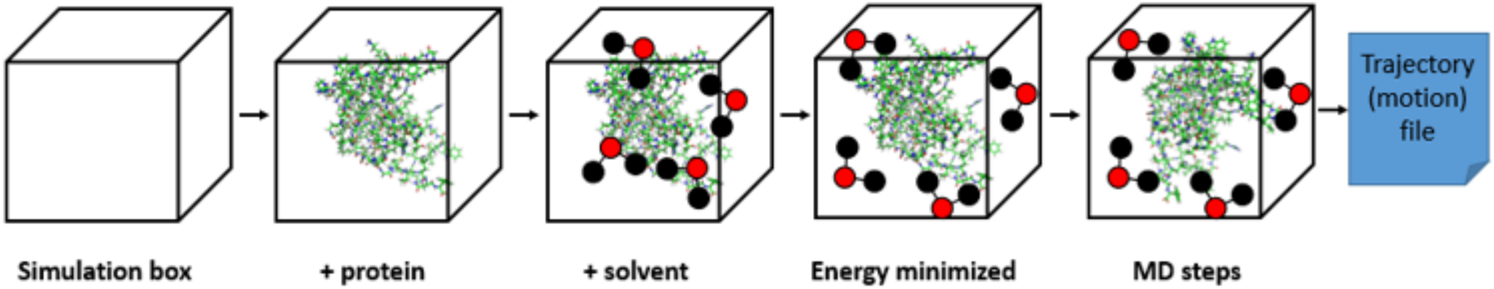


Course Project

Project 5 : Energy Profiles of Mutant(s) (double AA substitutions) of Covid-19 proteins

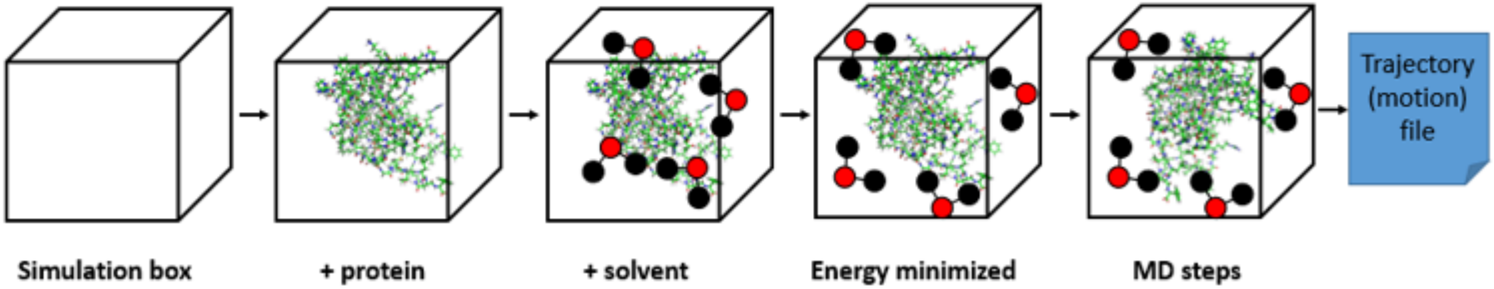
Course Project

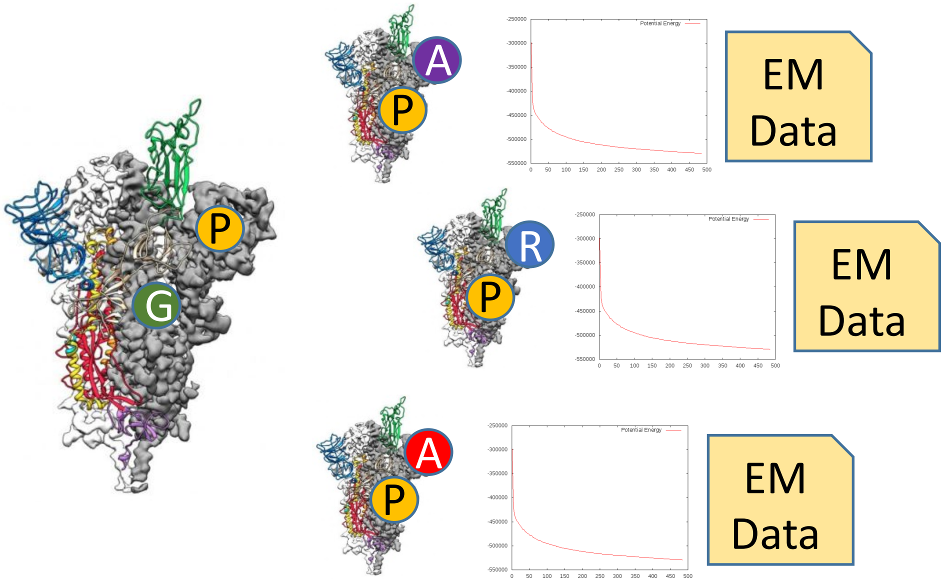
Project 5 : Energy Profiles of Mutant(s) (double AA substitutions) of Covid-19 proteins



Course Project

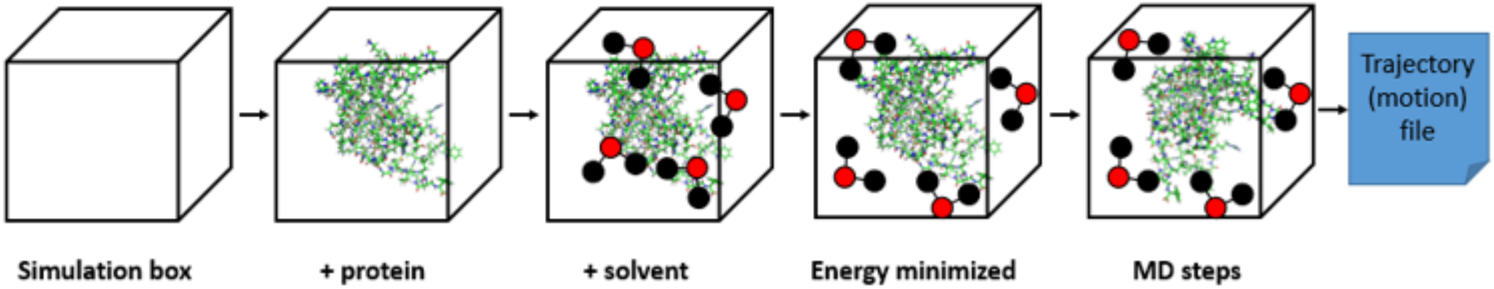
Project 5 : Energy Profiles of Mutant(s) (double AA substitutions) of Covid-19 proteins



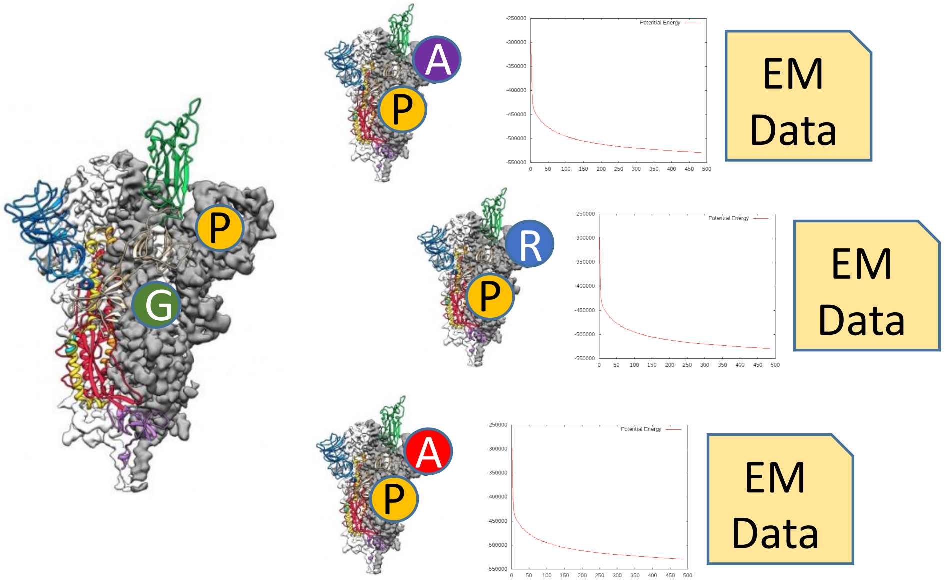


Course Project

Project 5 : Energy Profiles of Mutant(s) (double AA substitutions) of Covid-19 proteins



- Generate dataset(s) of EM data for ALL 2-point mutations in Covid-19 protein(s)
- Aggregate the results
- **Q: Which pairs of residue(s) when mutated are the most impactful in affecting the structural stability of a protein?**

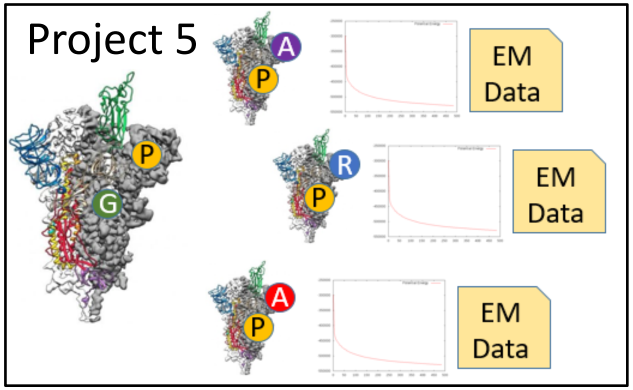


Course Project

Project 6 : Data Visualization / web server, for Projects 4 and 5 data

Course Project

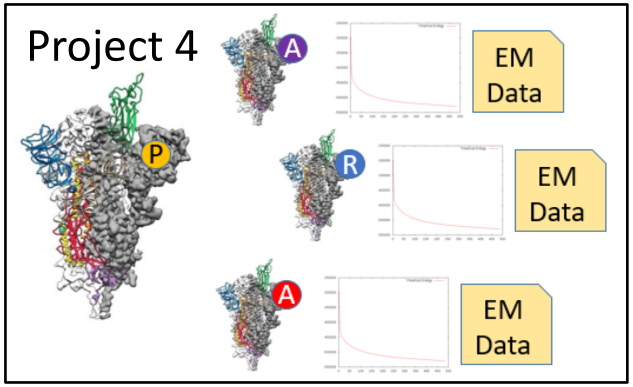
Project 6 : Data Visualization / web server, for Projects 4 and 5 data


Project 5

EM Data

EM Data

EM Data


Project 4

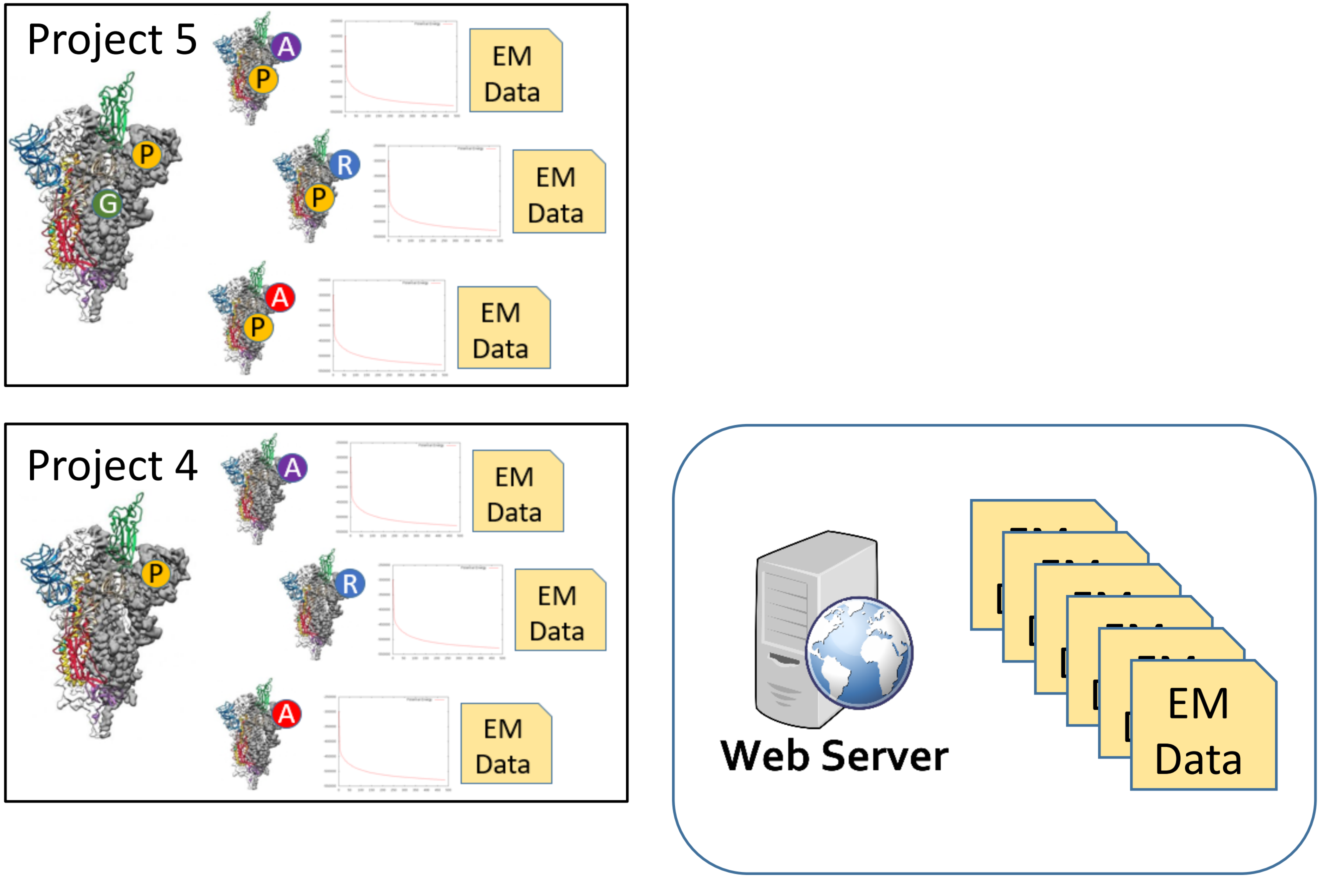
EM Data

EM Data

EM Data

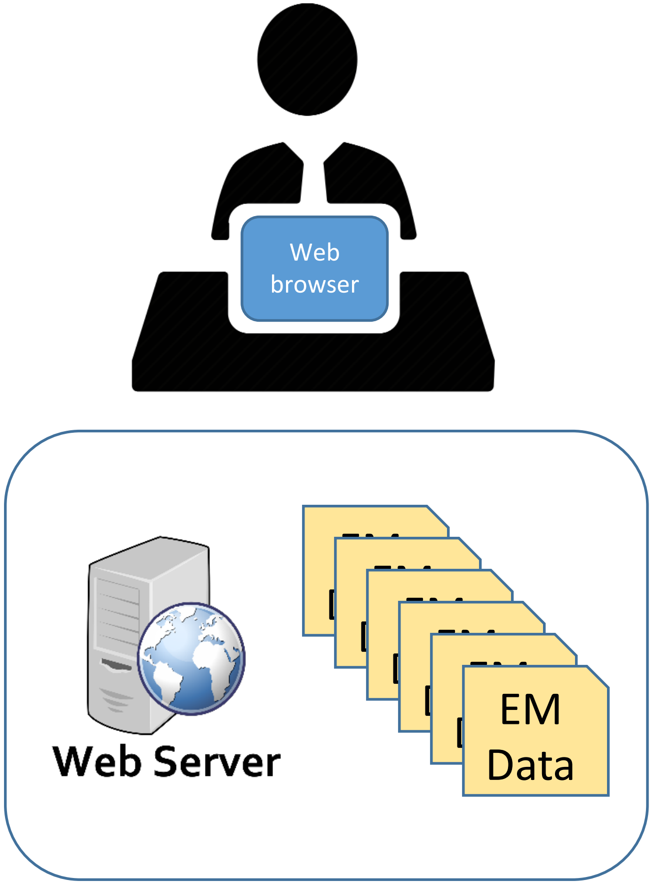
Course Project

Project 6 : Data Visualization / web server, for Projects 4 and 5 data



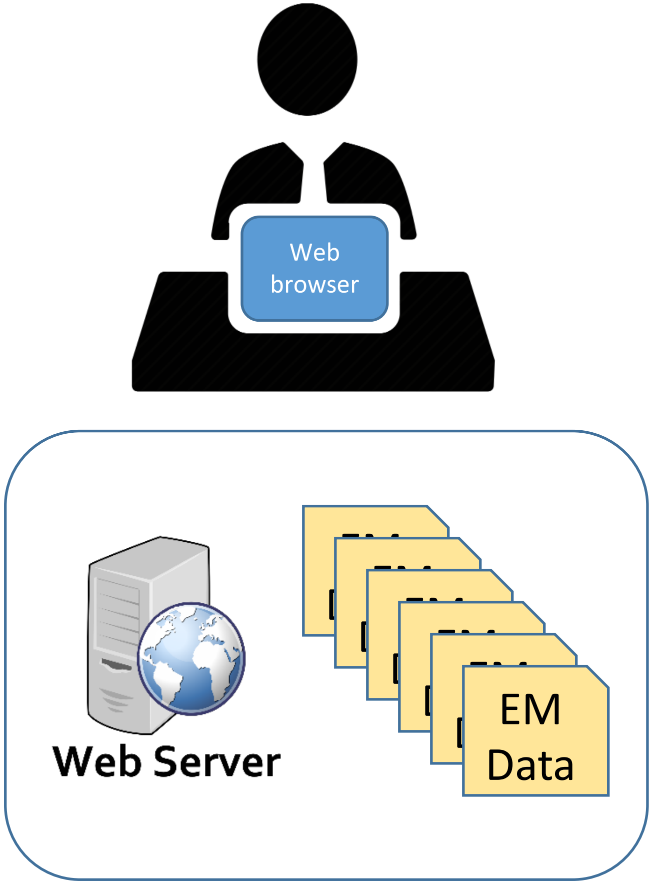
Course Project

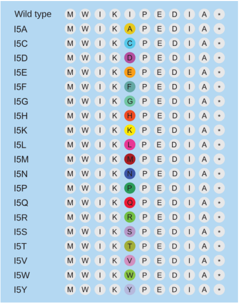
Project 6 : Data Visualization / web server, for Projects 4 and 5 data



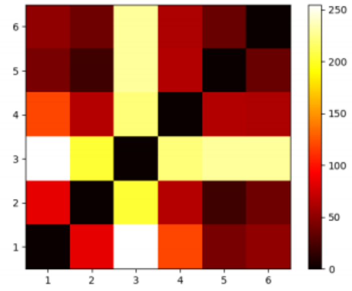
Course Project

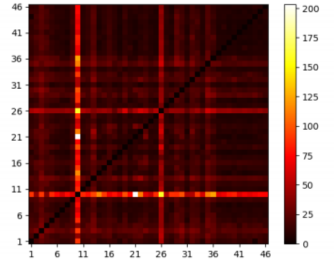
Project 6 : Data Visualization / web server, for Projects 4 and 5 data





Wild type	M	W	I	K	I	P	E	D	I	A	*
I5A	M	W	I	K	A	P	E	D	I	A	*
I5C	M	W	I	K	C	P	E	D	I	A	*
I5D	M	W	I	K	D	P	E	D	I	A	*
I5E	M	W	I	K	E	P	E	D	I	A	*
I5F	M	W	I	K	F	P	E	D	I	A	*
I5G	M	W	I	K	G	P	E	D	I	A	*
I5H	M	W	I	K	H	P	E	D	I	A	*
I5K	M	W	I	K	K	P	E	D	I	A	*
I5L	M	W	I	K	L	P	E	D	I	A	*
I5M	M	W	I	K	M	P	E	D	I	A	*
I5N	M	W	I	K	N	P	E	D	I	A	*
I5P	M	W	I	K	P	P	E	D	I	A	*
I5Q	M	W	I	K	Q	P	E	D	I	A	*
I5R	M	W	I	K	R	P	E	D	I	A	*
I5S	M	W	I	K	S	P	E	D	I	A	*
I5T	M	W	I	K	T	P	E	D	I	A	*
I5V	M	W	I	K	V	P	E	D	I	A	*
I5W	M	W	I	K	W	P	E	D	I	A	*
I5Y	M	W	I	K	Y	P	E	D	I	A	*

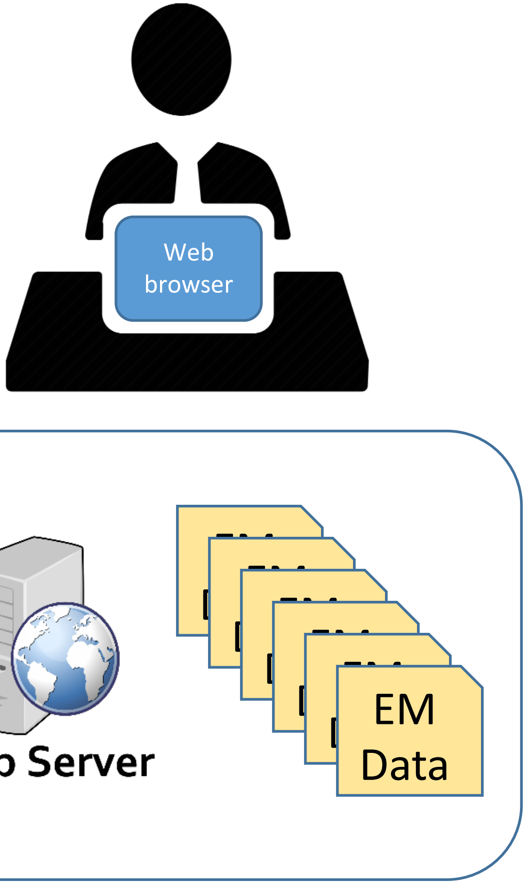

Mutant LRC Outliers 1 Standard Deviation from Global LRC Mean in 5E5X

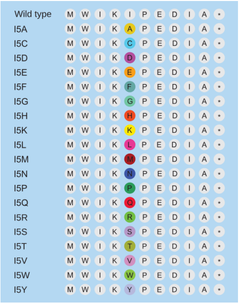

Average Distance Between Wildtype and Mutant LRC in 1CRN

Course Project

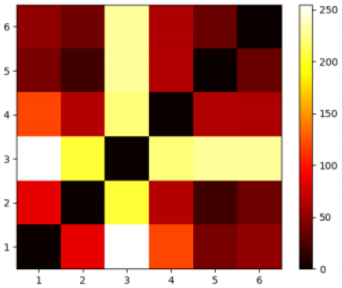
Project 6 : Data Visualization / web server, for Projects 4 and 5 data

- Basic Web server infrastructure
- “simple” GUI
- Interfaces with raw data
- Presents visualizations of the raw data
- **Task: discern which residue(s) are most impactful**

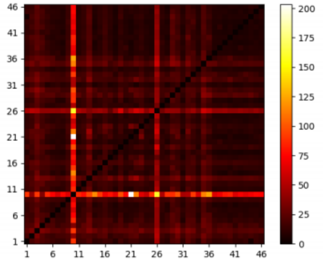




Mutant LRC Outliers 1 Standard Deviation from Global LRC Mean in 5E5X



Average Distance Between Wildtype and Mutant LRC in 1CRN



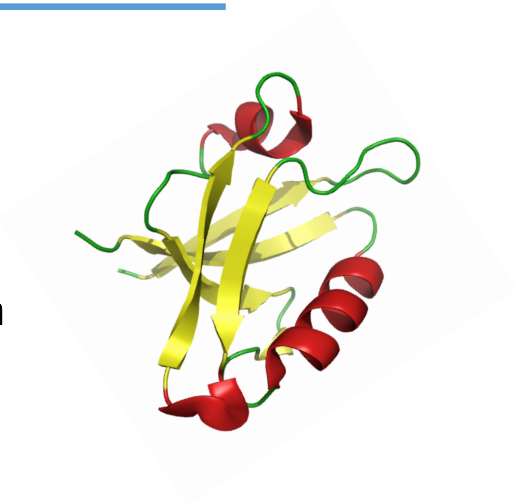
Course Project

Project 7 : PDZ domain, effects of mutations

Course Project

Project 7 : PDZ domain, effects of mutations

PDZ Domains are commonly occurring sequences of upwards of 100 amino acids in signaling proteins. These PDZ domains play a role in anchoring a receptor protein to another protein, to form a complex, so that signaling can occur.

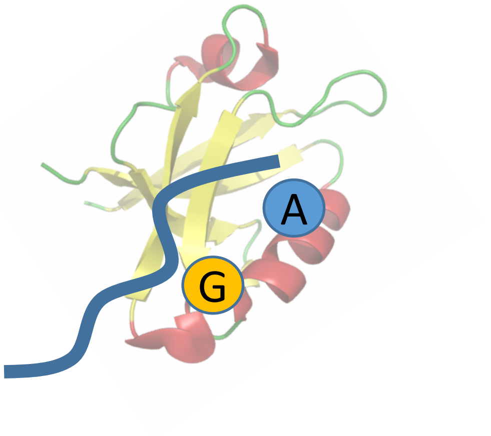


Course Project

Project 7 : PDZ domain, effects of mutations

PDZ Domains are commonly occurring sequences of upwards of 100 amino acids in signaling proteins. These PDZ domains play a role in anchoring a receptor protein to another protein, to form a complex, so that signaling can occur.

Most often it is the “last” 10-or-so amino acids of these PDZ domains that are responsible for securing in place the complex.

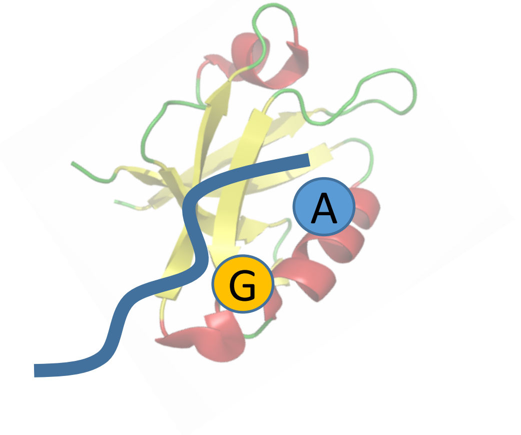


Course Project

Project 7 : PDZ domain, effects of mutations

PDZ Domains are commonly occurring sequences of upwards of 100 amino acids in signaling proteins. These PDZ domains play a role in anchoring a receptor protein to another protein, to form a complex, so that signaling can occur.

Most often it is the “last” 10-or-so amino acids of these PDZ domains that are responsible for securing in place the complex.

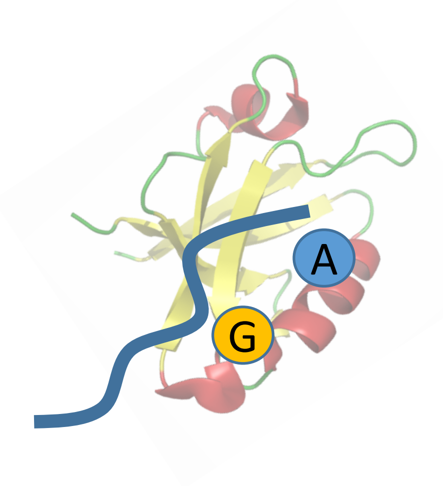


- On-going work with Jeanine Amacher (chemistry)
- We have the crystal structure of several PDZ domains bound to a larger protein
- **Q: What mutations to the 10 amino acids disrupt the binding and hence formation of the complex?**

Course Project

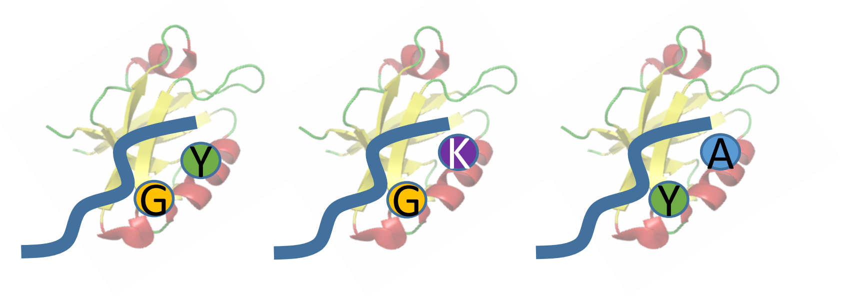
Project 7 : PDZ domain, effects of mutations

PDZ Domains are commonly occurring sequences of upwards of 100 amino acids in signaling proteins. These PDZ domains play a role in anchoring a receptor protein to another protein, to form a complex, so that signaling can occur.



Most often it is the "last" 10-or-so amino acids of these PDZ domains that are responsible for securing in place the complex.

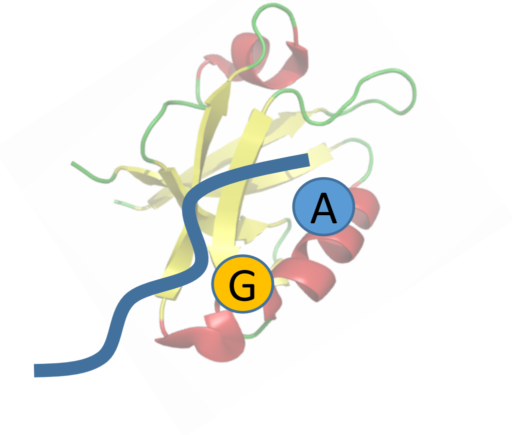
- On-going work with Jeanine Amacher (chemistry)
- We have the crystal structure of several PDZ domains bound to a larger protein
- **Q: What mutations to the 10 amino acids disrupt the binding and hence formation of the complex?**



Course Project

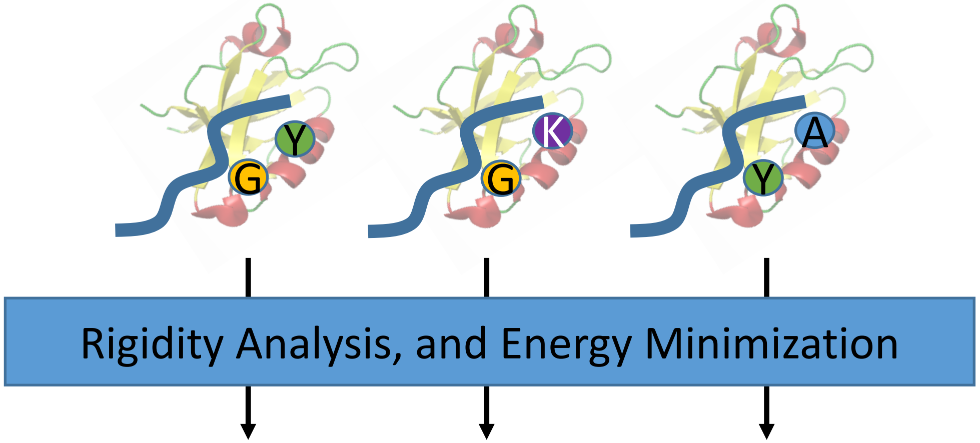
Project 7 : PDZ domain, effects of mutations

PDZ Domains are commonly occurring sequences of upwards of 100 amino acids in signaling proteins. These PDZ domains play a role in anchoring a receptor protein to another protein, to form a complex, so that signaling can occur.



Most often it is the "last" 10-or-so amino acids of these PDZ domains that are responsible for securing in place the complex.

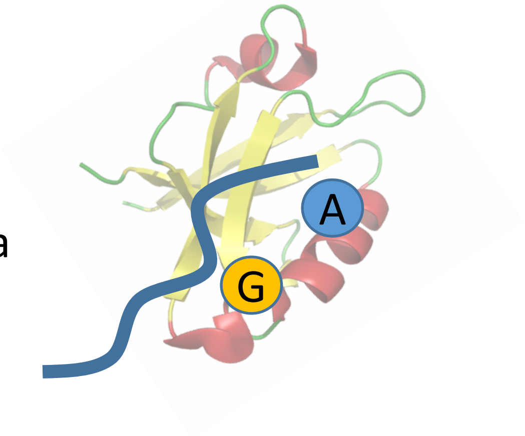
- On-going work with Jeanine Amacher (chemistry)
- We have the crystal structure of several PDZ domains bound to a larger protein
- **Q: What mutations to the 10 amino acids disrupt the binding and hence formation of the complex?**



Course Project

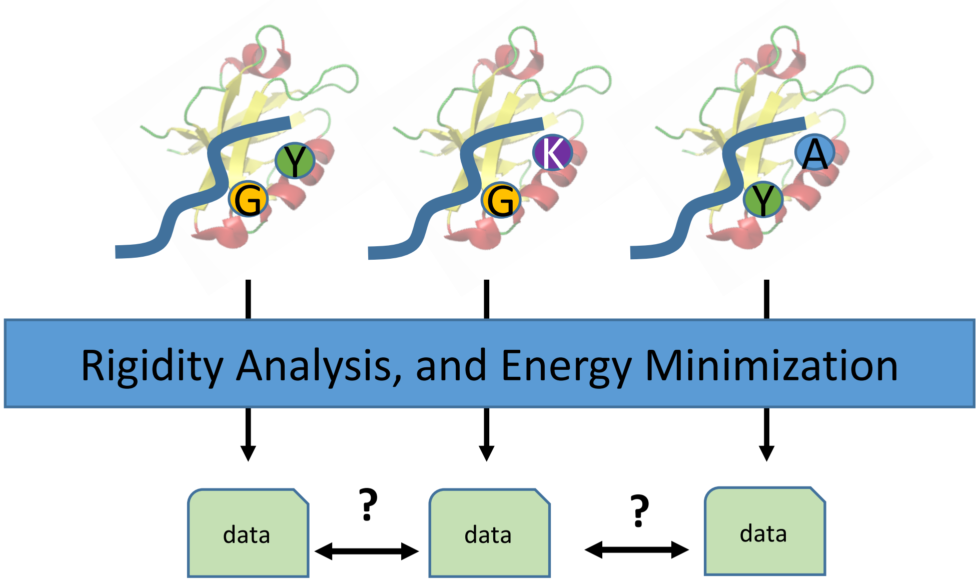
Project 7 : PDZ domain, effects of mutations

PDZ Domains are commonly occurring sequences of upwards of 100 amino acids in signaling proteins. These PDZ domains play a role in anchoring a receptor protein to another protein, to form a complex, so that signaling can occur.



Most often it is the "last" 10-or-so amino acids of these PDZ domains that are responsible for securing in place the complex.

- On-going work with Jeanine Amacher (chemistry)
- We have the crystal structure of several PDZ domains bound to a larger protein
- **Q: What mutations to the 10 amino acids disrupt the binding and hence formation of the complex?**



Rigidity Analysis, and Energy Minimization

data ? data ? data

Course Project

Use Canvas's Discussion feature to discuss group members
Self-select group membership (canvas)

Use Canvas's Discussion feature to discuss group members
Self-select group membership (canvas)

- Proposal Presentation
- Proof of Concept
- Initial Results
- Final Presentation
- Peer Assessment
- Final Report

Submitted via Canvas
Due 15 May

- Slide 1 : Title
- Slide 2 : Introduction (WHAT)
- Slide 3 : Motivation (WHY)
- Slide 4 : Methods (HOW)
- Slide 5 : Expected Obstacles (these most likely will change ONCE you begin)
- You must mention at least 2 references

This Week's schedule

Friday : Remaining labs due