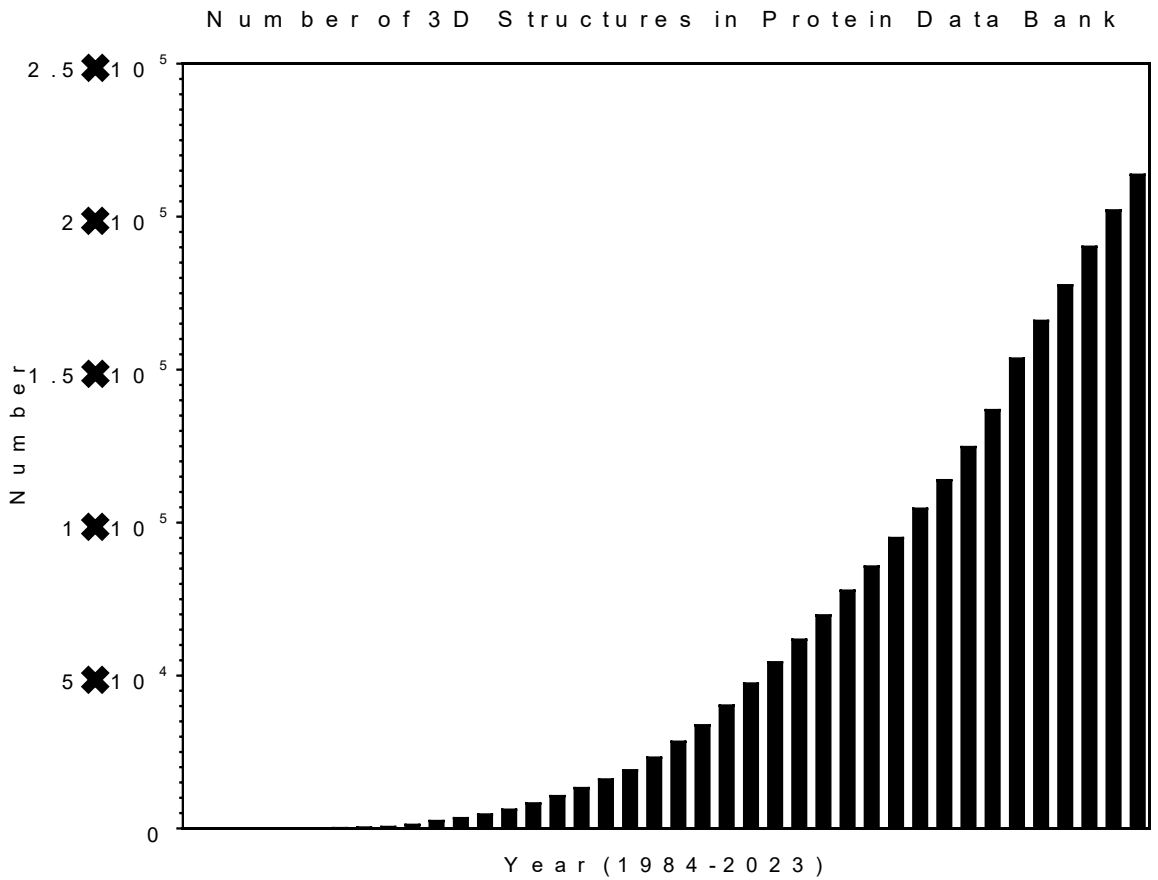
(i) Supplementary figures

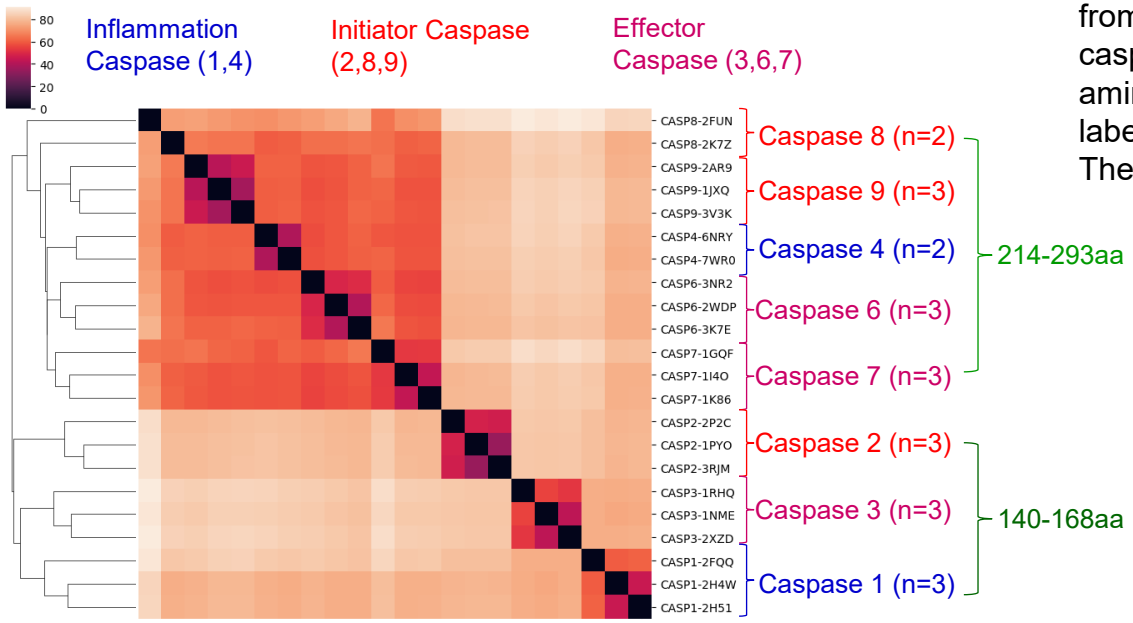
Supplementary Figure 1



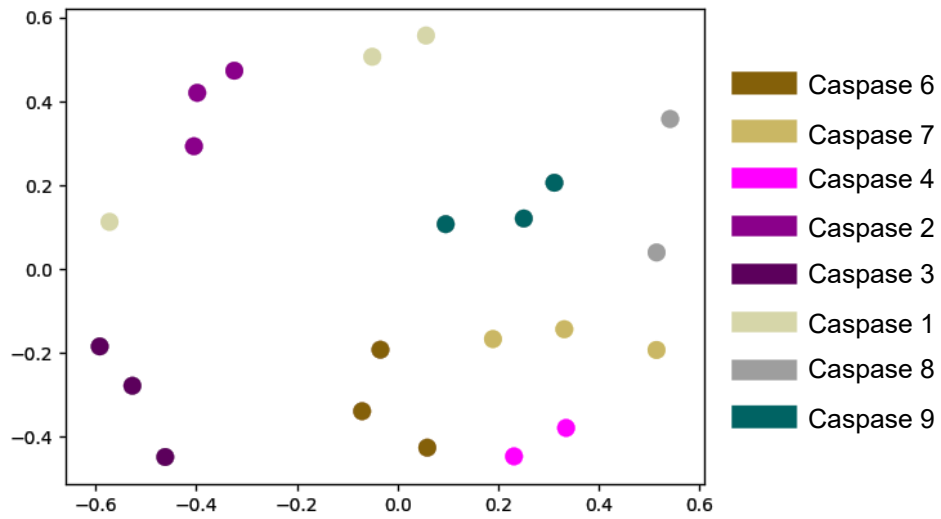
Supplementary Figure 2

The hierarchical cluster and MDS analyses demonstrate the structural characteristic of different types of caspases. a, The result from the hierarchical cluster analysis. The caspases and their numbers, numbers of amino acids and functional classifications are labeled; b, The result from the MDS analysis. The caspases are labeled.

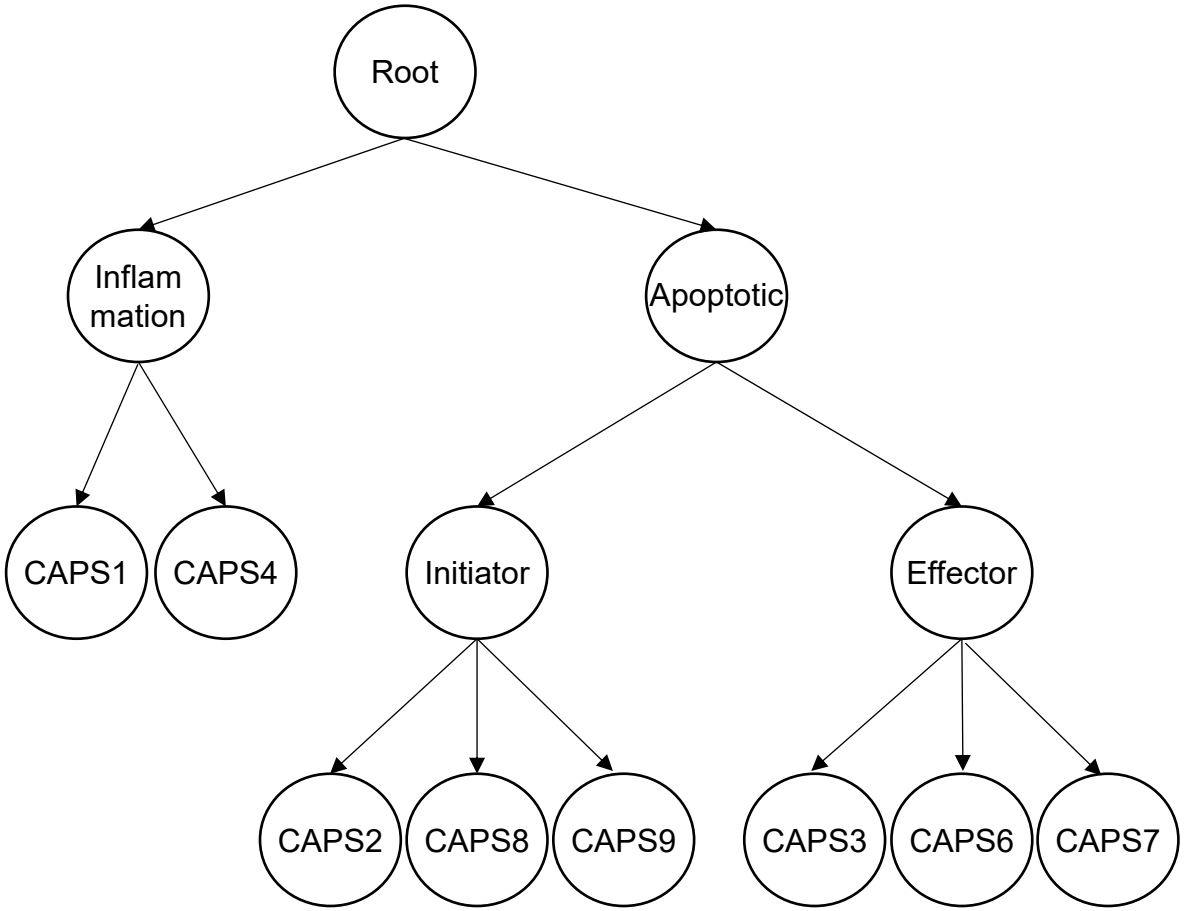
a



b

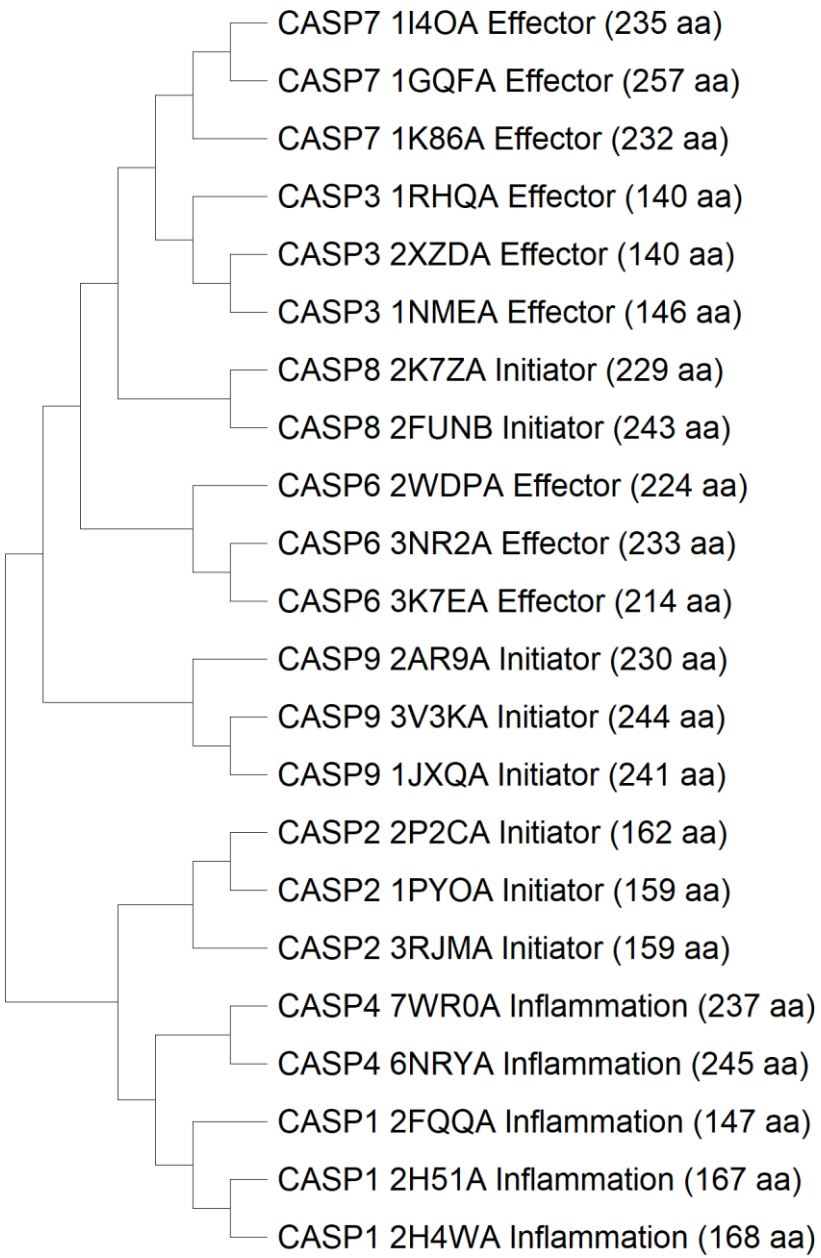


Supplementary Figure 3



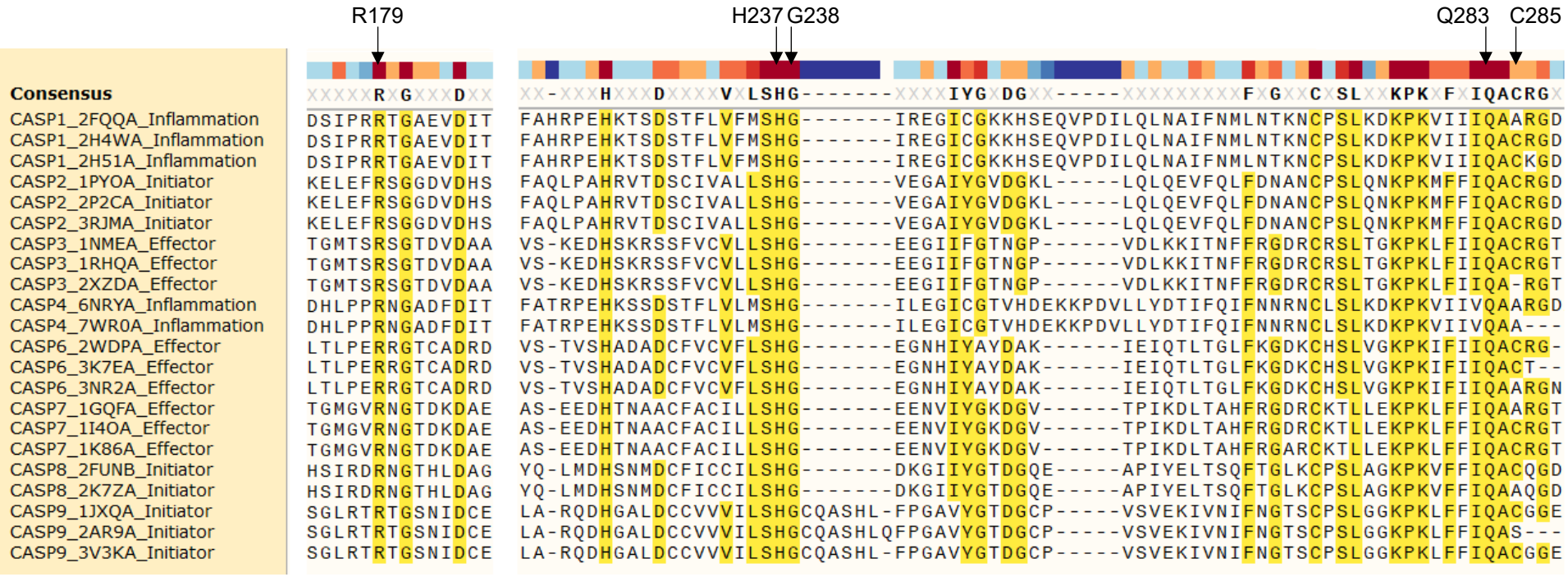
The hierarchical organization of the caspases

Supplementary Figure 4



The phylogenetic analysis shows the caspase sequence relationships. The numbers of amino acids of each individual caspases and their functional classifications are labeled. The caspase sequences were aligned using ClustalW and the phylogenetic tree was constructed using the Neighbor-Joining algorithm built in MEGA.

Supplementary Figure 5



The sequence alignment using SnapGene shows the residues that closely interact with the caspase 3 inhibitor. Part of the aligned sequences is shown.

Supplementary File 1: Instructions on how to run on the LONI cluster (New users can copy and paste each command to the terminal window and run them):

1. Log onto QB3 cluster using the command:

   ssh username@qbc2.loni.org

2. Change to the /work directory:

   cd /work/$USER

3. Clone the current repository:

   git clone https://github.com/dbxmcf/wu_sizegap.git

4. Switch to the repository directory on the cluster:

   cd wu_sizegap

5. Submit the first job:

   sbatch stepall.sh

6. Submit the second job after the first job is complete:

   sbatch common_step.sh

7. Monitor the jobs
   qstat –u username

Supplementary File 2: Instruction of how to use the TSR-based software for the users of Windows or Mac Book computers

1. Find proteins on https://www.rcsb.org/

2. In excel:
   - cell A1- protein: 4 letter code for protein from PDB
   - cell B1- chain: example: A or B…
   - cell C1- group: Gene name from PDB (usually 4-5 letter code), use correct capital letters
   - cell D1 group1: full name of the protein- ex: Dopamine_Receptor

3. Open FileZilla:
   - Host: sftp://qbc2.loni.org
   - Username:
   - Password:
   - Port: 22
   - Quick connect
   - Change /home/ to /work/
   - Open "wu_sizegap", "sample_root"
   - Create new file named "t#" where # is the next chronological number or t1 is first job.
   - Return to "wu_sizegap" and open a previous "stepall.sh" document

4. Stepall.sh steps- all # represent the number you are using for your stepall and t folder, DO NOT PUT # IN ACTUAL JOB. 1st job will have no number only stepall.sh and t1.
   - Open previous stepall.sh from desktop
   - Change run time to 10 hrs
   - change stepall#.sh à to number you are running in #PBS-N and #PBS-O
   - change SAMPLE_NAME to t# you are running
   - rename to correct stepall#.sh and save
   - transfer new stepall.sh to supercomputer in "wu_sizegap" folder

5. .csv file steps
   - Save excel file as "sample_details.csv" – the MS-DOS comma separated option
   - Transfer file to supercomputer under your new t# folder

6. Terminal steps
   - Open "terminal" app
   - Type: ssh username@qb.loni.org. Press enter
   - Type in your LONI password (you wont be able to see it)
   - Type "cd /work/username" and press enter
   - Type "cd wu_sizegap" and press enter
   - Type "sbatch stepall#.sh" and press enter
   - Type "qstat -u username" and press enter to check the job

7. Checking the job

- Open the t# file in the supercomputer and check for results (may not appear right away). Cluster map will be in folder 'theta_29'
- Drag 1st file named "clustermap_generalised_jaccard…." to home computer to view cluster map

To create wu_sizegap folder:

Open terminal/ putty

Log into loni account: ssh username@qbc2.loni.org
- Type in password
- Change /home/ to /work/ and enter

Copy: git clone https://github.com/dbxmcf/wu_sizegap.git
- Refresh filezilla/ putty
- Folder should come up once logged in under /work/

How to change structure sequence to PDB:
1. Open browser: https://zhanggroup.org/pdb2fasta/
2. Download PDB protein as PDB file (second option)
3. Upload PDB file to zhanggroup, or simply drag downloaded file from your computer to the 'choose file' button
4. Select pdb2fasta

How to view sequence in SnapGene:
1. Open snap gene
2. Select "open main collection"
3. Select 'new' on upper left corner
4. Copy and paste structure sequence from zhanggroup to the box
5. Select 'okay'
6. On bottom of page select 'sequence'
7. Save sequence: fileàsave asà name it to know this is the structure sequenceà format should be SnapGene Proteinà save

How to align sequences in SnapGene:
1. Open main collection in SnapGene
2. Select 'tools' à 'align multiple proteins' for 3 or more sequences or 'align two proteins' for only 2
3. Select 'import sequences to align', upload structure sequence
4. Repeat for however many proteins needed
5. Select 'align'

How to find sequence motif in SnapGene:
- Open sequence alignment of interest

- Select the ‘edit’ tab in top left corner of screen
- Scroll down a select ‘find’
- Type sequence motif od interest in search bar on bottom left of screen (ex: NYxxP)
- Select next to view in sequence

Multiple sequence alignment in MEGA link:
https://www.rcsb.org/downloads/fasta

Sequences for phylogenetic tree steps: edited PDB files to MEGA
1. Open pdb2fasta website
    - open correct .pdb file
    - select ‘pdb2fasta’
    - copy correct chain sequence and paste in notepad/sublime text (include >pdb:)

2. Save as PDB ID and .fasta

3. Open MEGA new alignment
    - insert sequence from file
    - select .fasta file and insert
4. Repeat for all desired sequences ^

5. MUSCLE align

6. Data -> phylogenetic analysis

7. On main MEGA page select ‘phylogeny’
    - select construct/test neighbor-joining tree

8. Select use current data

9. Open tree on main MEGA page

Supplementary File 3: Assessment Rubric

Table 1. Overall Assessment

Assessment Category	Percentage
Basic Supercomputer Skills	15
Preparation of Dataset	15
Structural Analysis	20
Report	20
Presentation	20
Structural Visualization	5
Sequence Analyses	5
Total	**100**

Table 2. Assessment Rubric for Basic Supercomputer Skills

Category	Criteria	Score
Needs work	• Need help to transfer files between local computer and supercomputer. • Need help to submit jobs to supercomputer. • Need help to create a new job script. • Need help to do trouble shooting.	< 10
Competent	• Be able to independently transfer files between local computer and supercomputer. • Be able to independently submit jobs to supercomputer. • Need help to create a new job script. • Need help to do trouble shooting.	10-12
Exemplary	• Be able to independently transfer files between local computer and supercomputer. • Be able to independently submit jobs to supercomputer. • Be able to independently create a new job script. • Be able to independently do trouble shooting.	13-15

Table 3. Assessment Rubric for Dataset Preparation

Category	Criteria	Score
Needs work	• Need help to identify proteins of interest from the PDB. • Need help to generate the input file including correct PDB IDs and chains. • Need help to correctly label the protein group names. • Need help to design the dataset by including a testing group and a control group.	< 10

	• Need help to prepare the dataset for common and specific TSR key calculations.	
Competent	•Be able to independently identify proteins of interest from the PDB. • Be able to independently generate the input file including correct PDB IDs and chains. • Be able to independently and correctly label the protein group names. • Be able to independently design the dataset by including a testing group and a control group. • Be able to independently prepare the dataset for common and specific TSR key calculations.	10-12
Exemplary	• Be able to independently identify proteins of interest from the PDB. • Be able to independently generate the input file including correct PDB IDs and chains. • Be able to independently and correctly label the protein group names. • Be able to independently design the dataset by including a testing group and a control group. • Be able to independently prepare the dataset for common and specific TSR key calculations.	13-15

Table 4. Assessment Rubric for Structural Analysis

Category	Criteria	Score
Needs work	• Need help to perform cluster analysis. • Need help to perform common TSR key calculations. • Need help to perform specific TSR key calculations. • Need help to do trouble shooting, maintain a good record of all the jobs and organize all the raw data and all the files. • Need help to interpret clustering results. • Need help to interpret results of common TSR key calculations. • Need help to interpret results of specific TSR key calculations.	<12
Competent	• Be able to independently perform cluster analysis. • Be able to independently perform common TSR key calculations. • Be able to independently perform specific TSR key calculations. • Be able to independently perform trouble shooting, maintain a good record of all the jobs and organize all the raw data and all the files. • Be able to independently interpret clustering results. • Need help to interpret results of common TSR key calculations. • Need help to interpret results of specific TSR key calculations.	12-15
Exemplary	• Be able to independently perform cluster analysis. • Be able to independently perform common TSR key calculations. • Be able to independently perform specific TSR key calculations.	16-20

	• Be able to independently perform trouble shooting, maintain a good record of all the jobs and organize all the raw data and all the files. • Be able to independently interpret clustering results. • Be able to independently interpret results of common TSR key calculations. • Be able to independently interpret results of specific TSR key calculations.	

Table 5. Assessment Rubric for Report

Category/Criteria	Beginning	Developing	Accomplished	Exemplary
Organization	Details and examples are not organized, are hard to follow and understand.	Information is scattered and needs further development.	Information is logically ordered with paragraphs and transitions.	Information is presented in effective order. Excellent structure of paragraphs and transitions enhances readability and comprehension.
Quality of Information	Unable to find specific details.	Details are somewhat sketchy.	Some details do not support the report topic.	Supporting details are specific to topic and provide the necessary information.
Introduction	Introductory paragraph is not apparent.	Introductory paragraph is vague.	Introductory paragraph is clearly stated with a focus.	Introductory paragraph is clearly stated, has a sharp, distinct focus and enhances the impact of the report.
Methods	Not all the methods are stated.	All the methods stated, but lack of details.	All the methods clearly stated with sufficient details, but this section needs to re-organized.	All the methods are organized well and clearly stated with sufficient details.
Results	Presentation of the data needs to be improved. Not	Data are logically presented. Not	Data are logically presented.	Results section is organized well. Data are

	all the results are based on the actual data. This section needs to be significantly re-organized.	all the results are based on the actual data. This section needs to be re-organized.	Results are based on the actual data. This section needs to be re-organized.	logically presented. Results are based on the actual data.
Conclusions	Concluding paragraph is not apparent.	Concluding paragraph is only remotely related to the report topic.	Concluding paragraph follows and summarizes the report discussion and draws a conclusion.	Concluding paragraph summarizes and draws a clear, effective conclusion and enhances the impact of the report.
References	Resources not cited in paper or proper format not used.	Some resources are cited but not all. Not formatted correctly.	All resources are cited, but formatting is not correct or inconsistent.	All resources are cited and appear with correct formatting.
Scores	< 7	7-11	12-15	16-20

Table 6. Assessment Rubric for Presentation

Category/Criteria	Beginning	Developing	Accomplished	Exemplary
Structure	• More than two of the points for exemplary need to be improved.	• Two of the points for exemplary need to be improved.	• One of the areas for exemplary needs to be improved.	• The presentation has a concise and clearly focus that is relevant to the audience. • The presentation is well-structured with a clear storyline. • Ideas are logically arranged. They strongly support the focus of the talk. • Sections are well-connected with smooth transition.

Content	• More than two of the points for exemplary need to be improved.	• Two of the points for exemplary need to be improved.	• One of the points for exemplary needs to be improved.	•The materials are coherently organized, demonstrating the presenter's mastery of the subject knowledge. • Figures are clear and well designed. • All materials presented are relevant and lead naturally to the conclusions. • Ideas are supported by evidence with appropriate use of facts, examples, statistics and references.
Communication	• More than two of the points for exemplary need to be improved.	• Two of the points for exemplary need to be improved.	• One of the points for exemplary needs to be improved.	• The presenter is fluent and articulate. The use and variation of tone and pace is effective. • Demonstrates good grammar and choice of words. • Maintains proper eye contact with audience. Posture and gesture show a good level of confidence and enthusiasm. • Answer questions well.

Time management	Presentation finished in the expected time ± 4 min	Presentation finished in the expected time ± 3 min	Presentation finished in the expected time ± 2 min	Presentation finished in the expected time ± 1 min
Scores	< 7	7-11	12-15	16-20

Table 7. Assessment Rubric for Structural Visualization

Category	Criteria	Score
Needs work	• Two or more of the points for exemplary need to be improved.	< 2
Competent	• One of the points for exemplary needs to be improved.	3-4
Exemplary	• Be able to independently show global and local structures. • Be able to independently show specific interactions between molecules. • The colors and atoms are well designed and clearly labeled. • The visualization is effective.	5

Table 8. Assessment Rubric for Sequence Analysis

Category	Criteria	Score
Needs work	• Two or more of the points for exemplary need to be improved.	< 2
Competent	• One of the points for exemplary needs to be improved.	3-4
Exemplary	• Be able to independently import sequence data into MEGA. • Be able to independently and correctly label each sequence. • Be able to independently perform alignment analyses. • Be able to independently create phylogenetic trees. • Be able to independently interpret the results.	5

Table 8. Assessment Rubric for Sequence Analysis

Category	Criteria	Score
Needs work	• Two or more of the points for exemplary need to be improved.	< 2
Competent	• One of the points for exemplary needs to be improved.	3-4
Exemplary	• Be able to independently import sequence data into MEGA. • Be able to independently and correctly label each sequence. • Be able to independently perform alignment analyses. • Be able to independently create phylogenetic trees. • Be able to independently interpret the results.	5

Supplementary File 4: Students' feedback

"I recently earned my Bachelor of Science degree in Psychology with a minor in Chemistry from the University of Louisiana at Lafayette. It was during my undergraduate studies that the bulk of this research was conducted. Research was a central aspect of my time as an undergrad, and it is partly because of my efforts in multiple research studies, including this project, that I will be starting medical school this summer. I worked closely with other undergraduates who also plan to pursue careers in the medical field. This research is novel and provides a new perspective of protein analyzation, but this research also allowed my coauthors and I to learn how to conduct sound research, learn many accepts of the TSR-based method, and learn how to translate our results into an academic article. This research was led and conducted by undergraduate students and then supervised by our professor, Dr. Wu Xu."

"I recently graduated from the University of Louisiana at Lafayette where I earned my Bachelor of Science degree in Biology with a minor in Chemistry. For this manuscript, the main role I played was identifying neurotransmitter receptors with their respective chains and sequences. From there, using MEGA software, I aligned these sequences based on different criteria, some of which were entire classes, families, or groups. I generated phylogenetic trees which I could then analyze to observe relationships upon these receptors which were then compared to respective cluster maps to further ensure that relationships were valid. Further contributions included helping to write and edit the paper to ensure that grammatical accuracy and general flow were present and upheld throughout the draft. With my fellows, data was formulated into presentations which were presented to other students. By being a part of this project, I have learned a lot about research in general, working with others, and the long process that encompasses being a part of a research project. These skills that I have acquired throughout this experience have helped me in my current occupational research role where I am involved in multiple clinical research trials. The attention to detail, data entry, and teamwork skills that were required for this project have aided me in the workplace, and I plan to continue to use them while furthering my education toward a career in healthcare."

"I am a recent graduate of the University of Louisiana at Lafayette, where I earned my Bachelor of Science degree in Biology with minors in Chemistry and Psychology. My research endeavors began when I was advised by one of my chemistry professors to join Dr. Wu Xu's research cohort. Dr. Xu's passion for research motivated me to join my fellows in conducting our research on neurotransmitter receptors. My contribution to our research consisted of using Dr. Xu's protein analyzation software and demonstrating how to use it to the undergraduate research students, analyzing our data and translating the data into written form, and assisting with proofreading and editing our manuscript drafts. After my experience working with my coauthors, I hope to continue my research career as I obtain my Master of Science Degree in Medical Sciences, as well as a medical student in the future."

"I am a senior at the University of Louisiana at Lafayette majoring in Biology with minors in Chemistry and Psychology. I am graduating in May of 2023 and attending medical school at LSU Health Shreveport starting in July of 2023. I started research with Dr. Wu and other students about two years ago with the intention of studying neurotransmitter receptors and diseases because of our common interest in these topics. During our first semester, we familiarized ourselves with the comparison software and learned to analyze the results. In our second semester, we created platforms and resources to educate future research students to

become familiar with the programs. In our third semester, we began drafting this manuscript. The ultimate goal of our manuscript is to analyze and compare various substructures of ionotropic and metabotropic neurotransmitter receptors to better understand neurotransmitter receptor mechanisms and interactions with ligands. We began this research due to a common interest in the neurotransmitter receptor proteins involved in neurodegenerative disorders, and we certainly hope that our manuscript, once published in a scientific journal, will serve as an outlet for further research and potentially treatment of disorders like Alzheimer's, Parkinson's, and ALS."

"I choice a hearing loss protein structure because my husband lost a lot of his hearing during his time in the military and will now need hearing aids. During this research class, I feel that I have learned and accomplished so much! I learned how to successfully obtain and run a job on LONI supercomputer and if it weren't for this class teaching me, I feel as if I would never be able to do it on my own. I do believe this class will help me in my future career because I want to do lab work running samples and collecting data."

"Initially I wanted to compare aging proteins, but this was much more complicated to find. I am very interested in why we age and if age is reversible and preventable. However, I am also interested in mood. As a college student, we are always under pressure and occasionally we put our mental health to the side to focus on other things. This can lead to depression, in this generation, it seems as if my fellow students/peers are not fulfilled or satisfied in life. I also feel that when we do seek professional help or guidance, they are quick to prescribe us multiple medications. I wanted to find the actual pathway of how depression medications worked. I'm grateful to have been able to participate in a small class with wonderful and bright students, this class had many students that I already had connections with which was a plus. Overall, I've gained knowledge that I did not have before, and am confident in how to obtain and compare proteins. When learning to run a job, you may run into some problems – but once you've figured out the issue, it's a great feeling. Once I was finally able to overcome small hurdles and completed my first run, was a feeling of success and gratification. Curiosity then took over for me, and I wanted to know and learn more behind the scientific uses of each protein. In my future pursuits, I plan on going to PA-school, however, I wish to obtain a master's degree. Following graduation this May, I plan to take a gap year to shadow, gain more hands-on experience, and find my spark in medicine again. I feel a little unmotivated and unsure of my future, so this time off, should help me. If I were to enter research in my gap year, I feel like my new found skill of being able to run a job, is useful. I'm also aware that research is also attractive on a resume. I also want to mention that we, as a class, love and respect our professor. We see and understand your love in our growth and potential! So we thank you for all that you have done for us and providing us a space to grow."

"I had originally chosen heat shock proteins within different parasites because I had just done a presentation for my Senior Seminar class on Plasmodium falciparum and thought it might be interesting to see. However, I kept running into issues while trying to submit my job and initially thought that it was my dataset causing the problems. So, I deleted the original data set and chose GABA receptors just so I could show that I could successfully run a job. If I were to run another job, I would choose the heat shock proteins or something that I am more interested in. I really enjoyed taking this course this semester. I liked learning how to work the LONI supercomputer even if I had trouble and I enjoyed working alongside others in the class. I also got to learn about a bunch of different proteins as almost everyone focused on a different kind. This course

did allow me to graduate with a minor in chemistry and I am glad I took it as it helped me branch out more as I never thought I would take a research course."

"I chose the protein structures, Dopamine, because I recently learned about them in one of my classes and thought it would be interesting to dive deeper into the proteins and their structures. This is my second year, so I already knew how to run a job. This semester, I felt more comfortable running jobs and was able to have more fun with it. The first-year students were also creative with their proteins, which inspired me to find more creative protein structures. I think that this class will help me in the future not only with biochemistry but also learning a completely new system and building on it on my own time. In May, I will be going to PA school and this class has taught me many leadership skills as well as keeping myself on track."

"I have an interest in forensic science and pathology which are both fields that usually involve death or examination of organs and humans usually after death. I picked caspases as they relate to cell death, a process that is always occurring in us. I learned more about bioinformatics and how it is used to help us compare proteins based on structure and sequence. I knew of it, but I never knew much about the process until now. I think this has given me a new outlook on research, in the sense that not everything has to necessarily be an experiment in a lab. This form of research allows us to use technology to put together information we already know which is pretty cool. I think this experience has helped me with my future as I am still figuring out what I want to do and biomedical research has been something that I have considered. It also helped me get a small background in coding, which is something that I may need to learn in my future career if I choose research."

"I chose bacterial proteins because I was always interested in microbiology, and I hope to learn more about bacterial proteins through my research. This semester, I have learned important research skills and I have also gained an interest in doing research. I will be applying to dental school this summer and I hope that having taken this class will relay my interest in sciences and chemistry to the dental school. This class has opened the possibility of conducting my own research while I am in dental school to further my knowledge of dentistry and science in general. I will hopefully be able to continue researching in Chem 462 in the fall 2023 semester."

"There was no large meaning behind the protein I selected, I mainly went with keratin because it is a well-known protein, and I hear about it a lot in the cosmetic world. Next semester I would like to choose one I can understand/ relate to and use in my future career better. I learned a lot this semester about proteins, technology, research, and learning to work independently. Before this class I had never been involved with research or anything related to it, so I had never seen this side of chemistry or biology. I was very confused about what everything was at the beginning, but once I started reading the PowerPoints and practicing the jobs, I understood it better. It also helped being in Chem 317 at the same time, so I could use the class knowledge for research and vice versa. I am not the best at technology, so I struggled when trying to run the job, but after going to your office a couple of times I got the hang of it and was able to complete it on my own. I do not plan to do research in my future career, but I learned a lot of skills that will be useful in my future. There are a lot of proteins in the mouth, and as a dentist it will be beneficial to have the background knowledge. It is important to learn problem solving, and independent study skills before becoming a dentist. I enjoyed this course and 317 a lot, and I learned a lot from you along with other students this semester! Thank you for being such a great professor, and all the help you have given me."

"I chose mostly thermoreceptors receptors in the TRPV subfamily, because I have a special interest in them after taking Neurobiology. I also chose a few other examples of proteins that I was interested in, such as miraculin in human gustation. I learned quite a bit about computational methods, a little coding, and several different methods for visualizing protein structures, and how structure determines function. Learning the TSR method was difficult, but with practice and repetition, it became easier and actually enjoyable. I also learned how to use MEGA and PDB, which are fantastic resources for understanding proteins, their 3D structures, ligands, agonists, antagonists, similarities, and so much more. This research program will surely help me in future research efforts, and with other course studies, such as R coding for Biologists, and Python for chemistry. I quite enjoyed this research course."

"I have always been interested in neurobiology and the mechanisms of neurotransmitters. I watched other classmate's presentations about dopamine and serotonin, so I wanted to choose one that nobody had researched yet. That is how I landed on melatonin receptors and I would love to continue my research on these receptors in the next semester of chem 462. I was overwhelmed when we were first shown what the research in this class looked like and the process of using the supercomputer. It only took me about two classes to actually understand what was going on and I was definitely worried for no reason. I was extremely grateful for the help provided by second and third year research students. It took me a long time of trial and error to successfully run a job by myself, but I appreciated it because I was able to learn from my mistakes. I feel like I actually know how to do it now and I didn't just get lucky on my first try. I will look forward to helping the new research students next semester! I really enjoyed listening to the research presentations of the graduate students and it made me think about if I wanted to go that deep into my research. Either way, I would love to continue my research on melatonin receptors for another semester. This class made me realize how much I really enjoy learning about neurotransmitters and that I might want to go into a field that involves neurobiology."

"I chose epidermal growth factor receptor proteins because I have always been interested how cell division can be controlled. I also enjoy learning about the skin since it is the largest organ and our first line of defense for infection. EGFR is expressed throughout the epidermis, so I knew it would be something that would further my knowledge of the skin. This semester, I learned that even if you have been researching a topic for a while, you can still encounter obstacles that hinder the process. For example, I had trouble running one of my jobs even though I have been running jobs for over a year now. However, I was able to fix my issues and continue. Trial and error is a big part of research, and so is the amount of time and effort that you put in. This experience will definitely help me in my future career, which I anticipate to be filled with research."

"I chose to research the protein structure norepinephrine due to my interest in the fight or flight response. I work in the ER and many of the patient's that I see arrive at the peak of their body's fight or flight response. I thought it would be interesting to look further into the structure of this vasoconstrictor. This semester was more interesting than the last semester of research that I did more so because I had a better idea of what to do with LONI and why it is important to study these protein structures. I learned how to run a job more effectively and efficiently this semester mostly because my laptop cooperated with the programs this semester unlike last semester. Also, I learned more about norepinephrine after running a job and doing background research. Although the use of supercomputers may not be the focus of my future career, a deeper understanding of proteins, their structures, and how they work was a valuable experience to

have as part of my background looking toward my career aspirations in medicine. Thank you Dr. Wu for all of your knowledge and help!"

"I am interested in learning about the genetics/mechanisms involved with cancer so I chose to analyze oncoproteins common in humans. I thought it was interesting to learn that even if proteins have similar sequences they can be quite different in their structures, meaning they probably have different functions. As someone interested in genetics, I used to focus on how related the gene sequences are rather than the structure of proteins but now I feel that structure of proteins can tell us more about proteins and how they are behaving. I think this course has taught me valuable information that I can use in my future educational journey."

"I chose the protein structure that I did because I wanted to explore something that isn't really talked about in class and understand something new. Throughout the semester, I learned a new appreciation for this system and a new understanding for TSR. For the future of my career, it could help with future research as I plan to continue to a graduate program. It is helpful to see how programming can be incorporated into research to assist and discover new similarities."

"I chose glutamate receptors structures because it was a topic I was fairly familiar with prior to the class so I just had to do a bit more research to be able to make a cluster map from it. My experience this semester was great, at first glance, I did think that preparing a job and running it would be complicated, but it turned out to be simple. I do hope to go more advanced with running these jobs though, as they did interest me a lot. I also hope to one day go on a job interview and tell them all about the research I did in this class, therefore I do believe it will be beneficial in my future career."

"I was interested in different neurotransmitters, since that is what I was learning about in class. I definitely learned a lot this semester. I have always been interested in coding and have had very minimal experience, but I enjoyed learning the technique to compare the different kinds of proteins to see how similar and different they were. Yes, it will help me because I am not completely sure what I want to do in the future, and this has shown me a new option I hadn't considered before. Thank you for all your help this semester!"

"I wanted to pick a protein that would be a beneficial cancer therapeutic target, and I wanted my topic to be novel and evolving. I decided to look up possible therapeutic targets for glioblastoma, since I have been interested in glioblastoma for years now due to the very short average life expectancy of patients following a glioblastoma diagnosis, and thus the patients needing better treatment options. I previously analyzed a biomath paper that used differential equations to model glioblastoma tumor growth rates under different treatment conditions, so I thought this research could build on what I learned previously. Upon researching receptor targets of glioblastoma, I came across a 2022 published article by Yale Cancer Biology Institute on extracellular EGFR mutations in glioblastoma. This is a newly evolving topic and has significant promise for future therapeutics, as EGFR inhibitors have not been a very successful glioblastoma therapeutic like they are for other cancers, such as lung cancer. I thought comparing the 3D structures of different EGFRs ("normal" and cancer-mutated receptors bound to different ligands) would produce new information that could be useful to other researchers working on glioblastoma treatments. Regarding running a job, I learned how to look for and delete spaces in my file on the supercomputer instead of on excel to correct an error code. I also learned not to use capital letters in the data set headers. As for my results, I learned a lot about different epitopes, therapeutics and how they inhibit the EGFR, as well as different

conformations of the EGFR. I know the two research classes I took with you and the research I am currently doing with you will be beneficial for my career in medicine. I am very thankful to have had this opportunity to learn many new computer programs and the TSR method, as well as the opportunity to learn from you. I knew I wanted to do research under you specifically after I took your gen chem II and biochem I classes, and you told us you previously did your postdoc research at St. Jude. I know I will do research later in my career and hope to use the research methods/programs we learned here. Also, I did not know much about EGFRs prior to my research in your class, so I have learned a ton about a very valuable target for cancer therapy. I also believe that our research will be beneficial for glioblastoma therapies in the future and that this research can be expanded as more mutations are discovered and more therapeutics are tested."

"At first, I didn't like that I didn't know how to do it, which was the reason why I wanted to give up. After knowing how to do it I actually enjoyed it because it was tedious. It was nice that everything was orderly and laid out. I learned how to make my own dataset and how to run a job. I didn't even know what the terminal was before this summer. It was hard to get used to at first since it was all computer based and I'm not good with computers or have ever even used excel. This experience helped me learn a lot and I think doing this was beneficial for later. Doing this gave me the push out of my comfort zone that I needed and everyone being so supportive made this the best research experience I could ever imagine. Even though I didn't do much, I still gained a lot of knowledge and even confidence in myself."

"I have learned that I chose the right major (Biology) as I do not think I would succeed as a computer science major. I did learn the process of how to generate a protein cluster map. I have learned that a more straightforward and user-friendly software needs to be developed to generate cluster maps easier. I now know what website I can use to access information about specific proteins of interest."

"I really appreciated how the computer lab taught me how to navigate the protein database and search for different proteins, different chains of proteins, etc. I enjoyed the 3D configuration provided by the database that illustrated the way hundreds of alpha helices or beta pleated chains can assemble together to function as a unit. I also enjoyed learning how to run a dataset and play with the coding page, as I had only seen language like that in media. Lastly, I enjoyed looking on google for examples and had randomly chosen those that were familiar to me."

"I learned that supercomputer skills are very important. Coding is not easy and very sensitive to small errors or lack of full attention. This was the most difficult lab for me, but I do understand the importance of this lab. Polygenetic sequence is important, it provides us with more accurate descriptions of patterns of relatedness than was available before the advent of molecular sequencing. These programs break down the makeups of every protein sequence known to mankind. This is important in chemistry and biology to compare the functions of different molecules that are aiding biological features in humans and animals. These programs give heavy insight that allows for many doctors and scientists to figure out ways to stop, start or regulate different necessary functions that needed when dealing with diseases, cloning, and other important factors. This lab was unfortunately difficult for me but overall, I have learned the importance of it all."

"Through this lab, the experimenter better understood practical aspects of computational and structural biology research. A portion of the work was done in a remote Linux environment,

using SSH, Bash, and other software. Experience was gained using websites such as the RCSB, UniProt, and Ensembl. Additionally, the experimenter learned about how ERs and ERRs operate in cells."

"The proteins I have chosen is proteoglycan. I chose the proteins because I was currently learning about their functions in the extracellular matrix in BIOL 457. I thought that the structures of aggrecan bound to hyaluronan looked interesting and its function useful and surprising. I thought that using proteoglycan in this lab would be a great opportunity to learn more about it. I learned how to set up the supercomputer on my personal computer and how to process data through. I also learned how to use MEGA to upload sequences and to create a phylogenetic tree from them. I liked using MEGA since it was easy to learn and my results were easy to understand and looked more accurate."

"During this computer lab, I have learned how scavenger through the protein database and how to even find proteins that are similar to each other in the database. I also learned how to read the cluster map, because prior to this experiment I was not very familiar with cluster maps and how to read them, but this experiment helped me a bunch. Lastly, I learned the importance of using modern day technology to your own advantage because how to do this is very important and extremely fascinating to see how far technology has advanced over the years."

"This lab showed that even proteins of completely different tissue types, functions and sizes can have a relatively high amount of similarity. These proteins were not large, but still ranged from 112 to 850 residues. Despite this range of sizes, they maintained certain structural motifs or folding properties that led to structural similarities, It was interesting to see that proteins with completely different functions could have even relatively similar structures."

"For my dataset, I chose keratin because it was a protein that I remembered learning about in terms of function, but never really in structure. In addition, I did not know specifically what kinds of structures or complexes existed with it, so I thought it would be interesting to learn more about them and use this software to compare them to one another visually. Because there was not an overwhelming variety of keratin structures to choose from, that also made it easier and less intimidating to choose structures to run my job on. From this computer lab, I have gotten the chance to experience and learn more about the computational side of biochemistry. Since my roommate has used this software, so getting the opportunity to try doing it myself was completely different than I thought it was going to be like. I really enjoyed learning about how clustermap can structurally compare protein structures to one another and display the relationships in terms of shading. It helped me better understand how proteins can be either similar or different to one another yet still serve similar functions. In addition to learning about the processes involved in running jobs and how the computer program work, I also learned how different keratin structures compare to one another based off my clustermap. Ultimately, I really enjoyed how this lab exposed me to a different perspective on chemistry that is not strictly formulas and calculations, instead showing me how structural comparisons are also a way to study chemistry."

"I chose to do Titin protein structures because I felt like titin is an overlooked protein that is studied and focused on less than actin and myosin are when talking about sarcomeres. During this semester, I learned how to actually run a job using the TSR method and the LONI supercomputer. I also learned how to get into research and how to really enjoy learning through research. This was my first time doing research, so it was a great experience learning new

techniques and information about different proteins. I believe this experience will help me in my future when I go to medical school because I now know techniques that can help me with research about protein structure and function."

"I learned how to run jobs on WinSCP and Putty and MEGA. It was interesting to learn how to produce results using the supercomputer. I am my own point of view, I think it is essential and highly useful for undergraduate students to participate in computational studies because it seems that a lot of future research will be heading towards computational means. This is my first time doing research and I am very glad I was able to work with someone else on the project. I would have been lost if Ryan didn't help me. I believe this will help me in my career, especially if I decide to apply for grad school as I do not have any research experience besides this. I am very grateful for the opportunity and appreciate everyone's help in this."

"I was inspired by one of the other students who used caspases. So I decided to find relations among subtypes. I learned a lot about the programming of how to code and run a job using a the TSR method. I can now run jobs through a supercomputer and run jobs to find similarities or differences between different proteins found in the world. I had a very positive experience seeing others work on interesting types of proteins. Everyone was super nice and was very patient with helping me setup my job."

"I choose subunits of laminin proteins for my data set because I am currently doing a research project on regenerative bandages for diabetic wounds that utilize laminin proteins. It is a current development of a team at Northwestern University being led by Dr. Ameer. I think I ended up learning a lot about different types of proteins from other presentations, but I also learned a lot about the process of conducting protein research. At the beginning of the semester, I felt very confused about downloading the right programs and did not understand the concept of "running a job", but by the second month of meetings I felt much more familiar with how to gather protein data and use it to run a job. If I did continue to use this method of research in the future, I would be very capable of creating a larger protein data set and running a job. I believe that I have the tools I would need to study a particular protein I learn about it medical school if I wanted to."

"I learned how to prepare a very large data set and interpret my results. This was my first time doing a research project and presenting it at a conference. It was a wonderful experience and helped me gain confidence in my research and presentation abilities. This has been very helpful for my future career as I would like to be a scientist in the biomedical field which would primarily involve research. I hope to one day get a PhD in a biological science and this research is only the beginning of that long journey."

"This protein was an important protein to study and it is necessary to do more research on it because it provides an insights into fundamental biological processes and has significant implications for medicine, drug development, and understanding various diseases. This semester has been incredibly rewarding. I have had the chance to assist fellow students by sharing what I've learned. Presenting at a conference was an incredible honor and a fantastic opportunity to impart knowledge. Moreover, I have expanded my understanding by delving into new information, enriching my educational journey for the future."

"Attending the national conference at Harvard allowed me to meet other undergraduate researchers and learn about their research. I was also able to hear from accomplished researchers who discussed the type of work they do and their journey through their academic and professional careers. It gave me a new perspective on the importance of research and that

sometimes, what you deem a mistake may be a discovery. As someone hoping to pursue graduate school, this conference introduced me to the many research paths that are to be explored. From this caspase research, I have not only learned more about the proteins themselves but also how to analyze their structural similarities through computational methods. I had no previous experience in computational biochemistry and this research has given me a broader understanding of it. I feel confident now in using the TSR-based method to compare proteins of interest and how to interpret my results. Presenting at conferences has also taught me how to explain results to others. These methods can allow us to compare protein substructures and how they interact with certain drugs so we can better predict treatment outcomes."

"I have learned several things from attending the National Collegiate Research Conference at Harvard University. The first thing that I learned is that it is very important to receive feedback on your most recent work. I got to make several new connections and exchange ideas with so many people, but I also heard some valuable criticism regarding the poster I presented. With the feedback I received, I can refine my research poster and make sure that I am putting my best work out for all to see. I also learned how important it is to stay up to date with all of the latest research. Hearing about the research that others are doing have been very inspiring for my own research ideas. Doing research on the caspase enzyme has taught me just how important of a role different caspases play in inflammation and cell death. Cell death can be normal, such as when the tissue between our fingers and toes is removed during fetal development so that they are no longer webbed, but it can also be abnormal. An example of abnormal cell death can be observed when cells in the heart and brain undergo apoptosis after their blood supply is lost. In the research that we do, different types of caspases are compared using structural analyses. Knowing what difference in structure between the different types of caspases allows us to better understand their function and it could lead to new treatment options for several diseases that are associated with apoptosis. Knowing the structural differences between the different types of caspases can be helpful when it comes to targeting a specific type of caspase. With this knowledge, pharmaceutical drugs could be developed to promote or prevent cell death. For example, medicines that trigger apoptosis in cancer cells while sparing healthy ones can be developed. This can help save lives, and, in turn, would affect the lives of people at the University, in our community, and all over the world. Without participating in this research, I would have never even thought about the importance of the caspase enzyme and the large effect it can have on cancers, autoimmune disorders, and neurodegenerative disorders."

*"This semester I finalized the SH2 dataset and showed other students how to check the SH2 data using Interpro, Zhang group website, download files from RCSB and display files from RCSB. I have also made the powerpoint of the information about the SH2 domain. I met you to go over the results you produced. I found the formula you used to determine the number of different triangles (3aa) you can form from 4 aa is 4, because you used the formula for combinations: Number of combinations (order does NOT matter and gives number of possible subsets/samples possible w/ NO replacements)= ((n!)/(r!(n-r)!)). So, # combinations of triangles (r=3aa) can get from (n=4aa) = ((4*3*2*1)/((3*2*1)(4-3)!)=24/6=4. FOR SIZE of total number of objects (n=100, r=3), total number of combinations/ number of triangles possible= 161700. I have also written the introduction and methods sections for the research paper and will be sending to you ASAP. Yes, I believe this information was very helpful as I learned how to use Interpro and the zhanggroup website, along with the display and download files from RCSB. I also learned a lot about cell signaling pathways regarding the SH2 domains, receptor tyrosine kinases and cancer*

formation. I am also learning how to write a research paper that is to be submitted to be published."

"I feel like I learned very many things this semester. This is my second time doing this research, and I learned about the new functions that the supercomputer is able to do when comparing proteins. I also learned about SH2 domains, their function in signaling cascades, and the role they play in forming cancer. I also learned how to double check proteins and their amino acid sequence to find where certain domains started and ended. I believe that this experience will definitely help me with my future career in medicine. With the abilities that I learned from this class, I can do more research in medical school, be familiar with more proteins, and examine proteins that interact with drugs or proteins that can lead to diseases."

"I was going through PDB, and had just stumbled on oncoproteins, and as I did more research, I thought I could focus on one type of gene of oncoproteins. My experience during this course was great. It was my first-time doing research, so I was confused at the beginning, but as presentations started, I began to understand the purpose. Learning and researching about proteins/drugs on the Protein Data Base, I learned specific details and visuals that I have never seen before. I really enjoyed learning how to run the job and figuring out how to run it via FileZilla and terminal by myself, learning from the mistakes and figuring out what I did wrong. I also enjoyed learning about the proteins that my classmates have been working on. I would like to do research again and create a more complex cluster map(s)."

"My experience was amazing. I truly enjoyed the research. During my time working with Elijah, I learned many things. Although I was never able to run a job, Elijah demonstrated how the programs were utilized and taught me that at the molecular level, biochemistry research relied on principles from biology, chemistry, physics, and computational science. He also showed me how many lab techniques including how to mix and create solutions from scratch. If I had to conduct the research again, I would spend more time gathering information and running several jobs to get a wide range of accurate data. In conclusion, this experience taught me the value and impact of research as well as strengthened my abilities to work with others."

"This semester I have learned how to coordinate research with multiple peers. This will help me in my future endeavors when managing tasks as a team. I have also learned that research can help me get into Dental School where I can possibly begin new projects."

"I chose the enzyme pepsin because I am currently taking Biol 318- Adv. Anatomy and physiology and we are starting the digestive system. So I wanted to choose a protein that would benefit me in learning about and help me with my presentation in this course. I thoroughly enjoyed this class this semester. Although we only met one a week, I feel like I've learned a good bit on how research actually takes place on a basic level. The main thing I learned is how to complete a sample_detail using PDB and how to run a job on that specific sample_detail. This allows me to compare and contrast the structures of different proteins. I don't believe this information will help me with my future career as a chiropractor but who's to say it won't help with future courses because I don't know what I will be taking."

"I chose these proteins because they are quite common, and they are super important in the human body. The main thing I learned this semester was how to read the cluster maps and understand what each part means. I also feel very confident in running jobs this semester, and I was even able to help some of the new students. Last semester, I barely knew how to do this on my own, so I was happy to feel confident in it enough to help others. I start Dental school in

August, and I do not plan to continue doing protein research but being part of this class has taught me a lot about problem solving. The research did not come easy to me but after working on it for a while I was able to get good at running jobs and problem solve through any type of error I may have had. Thank you so much for this semester, and I learned so much from being in your class and part of your research!"

"I chose myoglobin because I like to work out and understand the mechanism involved in muscle contraction. Muscles need an abundance of ATP, which means they also need a large concentration of oxygen to complete oxidative phosphorylation. I wanted to understand how they gain or retain a high amount of oxygen, hence leading me to investigate myoglobin. My experience in the Wu research lab was interesting. I learned how to run a job to find similarity between proteins and how to read cluster maps. This will help understand future research as chemistry studies advance. This lab will help my future career with understanding how to compare proteins. This lab also helped build connections with other students, who are doing research. I will be conducting future research with Walter, which will be great on my resume."

"I chose Acetylcholine Binding Protein because I am particularly interested in endocrinology and neurotransmitters such as Acetylcholine. This semester I learned about the TSR method which is used to analyze proteins based on sequences from the PDB data bank. This was my first time doing this research and I learned a lot of useful skills involving chemistry, biology and computational skills. These skills will help me in my future and career as I apply to dental school this cycle."

"I really was open to any opportunity to work with a grad student, but I was lucky to actually enjoy studnet's research. With my desire to work in the medical field, the SH2 domains application in cancer treatments was interesting to learn about. Learning about the application of these protein domains in cancer treatments was so interesting to learn about. I knew little about any research going into cancer treatments, so this gave so much insight to the actual mechanisms that go on behind the treatments. I would love to see how much this research develops by the time. I am working in the professional field. I don't have much interest in oncology, but I know it can be applied to many different areas."

"I chose Insulin because it is a hormone that I do research on already in the Biology Lab. I work with urinary tract infections (UTI) and I mock diabetic urine to see the pathophysiology of Klebsiella pneumoniae and susceptibility of UTIs in diabetic patients. I learned how to submit a job using the LONI supercomputer and how to analyze cluster map data. I also learned how certain proteins are derived from animals and how they can be used for human therapeutic care. I really enjoyed doing research with you Dr. Wu and meeting other people who are performing research in the biochemistry area. It allowed me to attain another research opportunity with Walter and to study another protein known as calmodulin. This experience will help me reach other research opportunities and increase my chances of getting accepted into more competitive medical schools."

"My grandpa just recently passed away from Dementia, and the specific type he struggled with was affected by the Lewy Body Protein, which is the protein I chose to study. I really enjoyed this different experience of research. I have worked with the psychology department doing research on humans, but never on such a small scale thing like proteins. I will definitely take this with me in the future because I am going to medical school. I feel like this research can help to have a deeper and better understanding of specific types of proteins, which can be extremely beneficial

in the medical field, when I specialize. This semester just helped me to have a more rounded approach to research, and see a different side of it. I also feel like it widened my tools and ability to do things like coding and computer programming because that was such a big part of the class."

"I have done research in the past however, the research that I have done involves zebrafish and is very different that what we are doing here. Learning how to run a job exposed me to a new type of computer program that I did not even know existed. Learning how to work WinSCP, putty, VMD etc. has been very insightful. I feel like this could be helpful with my future career. I want to attend medical school for psychiatry and there are many proteins that are used in psychology research that could be helpful for my future studies. Also, from listening to everyone's presentations I have learned so many facts about different proteins and diseases. Overall, I really enjoyed this class semester, and I felt like it was a breath of fresh air. Thank you so much for the experience, Dr. Wu!"

"For my presentation and my own dataset, I chose collagen because of the dental field. I am fascinated with dentistry, so wanted to choose a protein that aligned with the workings related to the dental field. I am grateful for this opportunity to join Dr. Xu's research lab. This semester was my first time doing research, so it was all very new for me. At first, I was very nervous, but with my amazing colleagues and professor, I was able to feel reassured quickly. This semester I learned how to run jobs on the LONI supercomputer which was something I was not familiar with, but after much practice, it began to get easier. In addition, I am now familiar with creating my own dataset and finding proteins through PDB. This experience was not only great in providing me with knowledge about running and monitoring jobs but presenting. I have not done many presentations, so presenting in research has allowed me to gain more experience and feel more comfortable when it is time to present for future courses. This experience will most definitely help me in my future career or course studies."

"I chose the protein structure (enamelysin) for my dataset because I am interested in pursuing dental field. I decided to learn more about my future aspirations through my research and fuel my curiosity. As I was involved in this research experience, I was also taking biochemistry 1 at the same time, therefore I got the opportunity to exercise the skills I learned. When analyzing the cluster maps that were generated, I was able to understand how the structures correlated with the amino acid sequence and what made these features similar. The concepts of secondary structure that I was taught in biochemistry provided me the ability to decipher their interactions and recognize how their functions are related. I gained primary knowledge of biochemistry practices concerning proteins that will have a direct impact on my future path. Besides conceptual information that I gained, there was also the practical computational expertise that I was never exposed to before. Previously, I knew how to handle computer programs, but was never fluent in the work. This course gave me the capacity to master a system of computational biology, which I may be able to use later on for better understanding of complex facets."

"I chose the tumor suppressor proteins p53 and p73 for my dataset because I am currently in a cancer biology class and have learned in depth about the proteins and their pathways. I thought it would be interesting to see how their structures compared and how significantly different mutations affect their structure. I really enjoyed learning about your research this semester! This semester, I learned how to run jobs on a supercomputer, and I learned how the supercomputer can assist in studying different protein structures and their similarities. I found it very interesting that two seemingly similar proteins can actually have significantly different structures. While I'm

not incredibly sure whether this experience will significantly help me as I continue into the medical field, I still found it to be interesting and valuable. If I were to ever go into medical research, this experience will for sure be useful! Thank you so much for your time and dedication! I really appreciated the opportunity.

"I chose Titin because I knew a little bit about it beforehand. We had gone over it very briefly in my Anatomy class. I learned a good bit in this research class. I learned how to run a job and check that the job is running properly. I learned how to view the structure of a protein in VMD. This course will help me in my course study and future because it properly demonstrates how to work with others."

"I chose the proteins that I used because I was curious to see how many different proteins would be related or not related to each other. I learned how to successfully submit a job and use the LONI super computer. Although it was very stressful to learn how to submit a job in the end, I was able to complete it. This experience taught me how specific typing codes are and if there is any smallest detail missing the entire job will not work."

"I chose beta amyloid because of my interest in Alzheimer's disease. I think research focusing on a cure for Alzheimer's is very important, and it starts by preventing beta-amyloid plaque in the brain. I learned a lot about research this semester. I've conducted research in the biology department for the past 3 years, but I think I prefer chemistry research more. I think I was able to learn more about research and its resources in this lab than I have in other labs. I think learning about the supercomputer and about these different resources will help me in the future in any research I conduct. I am grateful for the things I have learned working under Dr. Wu and I hope I will carry these teachings with me throughout my career and future schooling."

"I chose to do my research on the CFTR protein for cystic fibrosis because I had came across this topic in my microbiology course. I found it quite unfortunate how it is inherited and how it is chronic. Further investigating this issue is necessary to assist those who are affected. I chose those specific protein structures because most of them belonged to the same group. I want to see how close proteins of the same group were in structure. This semester, I was introduced to several knew proteins and their functions. As a biology major, studying medical and allied heaths, this was very informational. Something knew that I learned is that through programming, we ourselves can compute structures of the proteins. There were several students in my class who did this but I was unaware that we were able to do so. This course study will definitely help me in my course study. This will expand my knowledge in research skills, problem solving, data analysis, etc. Learning research and analytical skills is essential in the medical field and advancements in the medical field."

Supplementary File 5: Outcomes of the project-oriented student learning

1. Background and course objective

Background: The shortage of science, technology, engineering, and mathematics (STEM) workers to meet the demands of the market is a critical concern. A lack of career potential and low salary in STEM fields, especially K-12 teachers, are some of the factors responsible for this shortage, probably across demographics. Fundamental changes may be needed for resolving the STEM workforce shortage situation. Although beyond the scope of this manuscript, this situation motivates Dr. Xu to develop specific strategies, focused on the contributions Dr. Xu can make in encouraging more students to choose STEM disciplines. During the review of the first paper on the TSR-based method [1], Dr. Xu designed a 1-credit hour course, CHEM 362, to offer an undergraduate project-based learning research experience. Sophomores, juniors or seniors majoring in biology or chemistry can take this course. Project-based learning introduces real-world problems and captures the students' curiosity, motivating them to recognize and investigate abstract concepts and principles. This experience is exactly what they need to develop critical thinking skills during the course of their training in order to prepare them for their future careers. Federal funding agencies, e.g. NSF and NIH, invest significant money to support undergraduate students to do research during the summer (Summer Undergraduate Research Experience). The Department of Chemistry at the University of Louisiana at Lafayette has not received funds from either NSF or NIH to support Summer Undergraduate Research Experience. Instead, the Department of Chemistry offers course (CHEM 362) (often 1 credit hour and could be 2 credit hours) to gain research experience during spring, summer or fall semesters. The University of Louisiana at Lafayette initiates Advance Student Research Experience (ASRE) Pathways. These pathways include Basic, Distinction and Excellence. Basic, Distinct and Excellence pathways target ~80%, <40% and 10% of chemistry and biology students, respectively. The pathway will be shown in their undergraduate transcript. The Department of Chemistry chooses CHEM 362 as part of ASRE pathways.

Course Objectives: The CHEM 362 designed by Dr. Xu has *three objectives* with a focus on interdisciplinary training. (i) The first objective is training. Students need to have the skills for doing undergraduate student research. The students will receive hands-on training on protein structure database, e.g., PDB, protein sequence (SnapGene, DNASTAR and MEGA) and structure analysis or visualization software (VMD), and supercomputer skills (WinSCP, putty); (ii) The second objective is for students to learn biochemistry from doing biochemistry research through individual and group study. Students learning skills from students is an important part of the training. By doing so, students will gain informal teaching experiences. To promote students' interest and critical thinking, the students will choose their proteins and build their datasets for protein sequence and structures analyses. Students are required to INDEPENDENTLY search for their proteins, build their dataset, run jobs, analyze the results and interpret the results based on the data; (iii) The third objective is to gain experiences in literature search, writing reports and presenting their results. Each student will write a report and do a student presentation at the end. The main objective of this course is not focused on a solution of a specific scientific problem but achieving a broad impact on the future of science and research. It is achieved through providing a first-hand research experience especially to undergraduate students who cannot do research in Research Experience for Undergraduates (REU) sites due to student summer plans and/or limit number of students being accepted by REU sites. However, Camille Reaux, the first author of this manuscript, wants to achieve scientific discoveries using the hands-on experiences she has

learned from the first semester. She initiated the neurotransmitter receptor project and has led other undergraduate students to achieve scientific discoveries through a two-year journey of education and research. Their study shows that different types of neurotransmitter receptors share a little sequence similarity, but have significant portions of similar local structures. It is our first time to achieve a synergistic effect by simultaneously using two feature engineering methods: amino acid grouping and size filtering, in perfectly matching protein structure-based clusters with their functional classification. The clusters from the sequence-based clustering study failed in perfectly matching with their functional classification.

2. Courses and students

Courses: Dr. Xu teaches the following courses from 2021 to 2023

Biochemistry I, CHEM 317 (Three credit hours), Spring 2021, 2022, 2023 and Fall 2021, 2022 and 2023 (Scheduled)

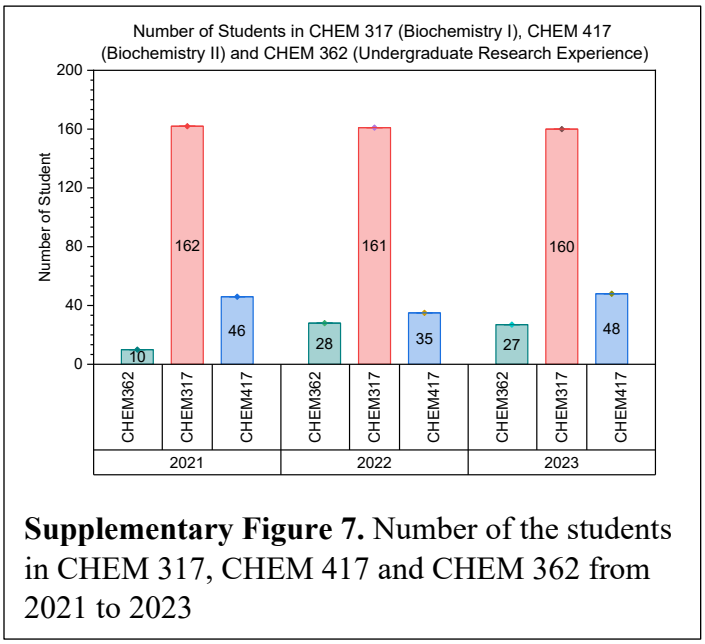
Biochemistry II, CHEM 417 (Three credit hours), Spring 2021, 2022, 2023

Dr. Xu also teaches Undergraduate Research Experience, CHEM 362 in the spring, summer and fall of 2021 to 2023.

It should be pointed out that the Department of Chemistry at the University of Louisiana at Lafayette was focused on undergraduate education and research until a new Master's program was approved in 2020, and a new shared Ph.D. program with physics and geoscience was approved in 2019. There are no teaching assistants (TAs) for CHEM 317 and CHEM 417. CHEM 362 is not a required course for biology or chemistry majors.

Students: The numbers of the students in CHEM 317, CHEM 417 and CHEM 362 are shown in Supplementary Figure 12. There are a total of 483 students who have taken or will take CHEM 317, Biochemistry I, from 2021 to 2023. The number of CHEM 317 students for the fall of 2023 is based on currently registered students. Biology and chemistry students are required to take CHEM 317. There are roughly ~160 students who take CHEM 317 each year (approximately 70 students in each spring semester and roughly 90 students in each fall semester). The Department of Chemistry offers CHEM 417 only in spring semesters. 129 students, 26.7% (129/483) of students who took CHEM 317, Biochemistry I, took CHEM 417, Biochemistry II, from 2021 to 2023.

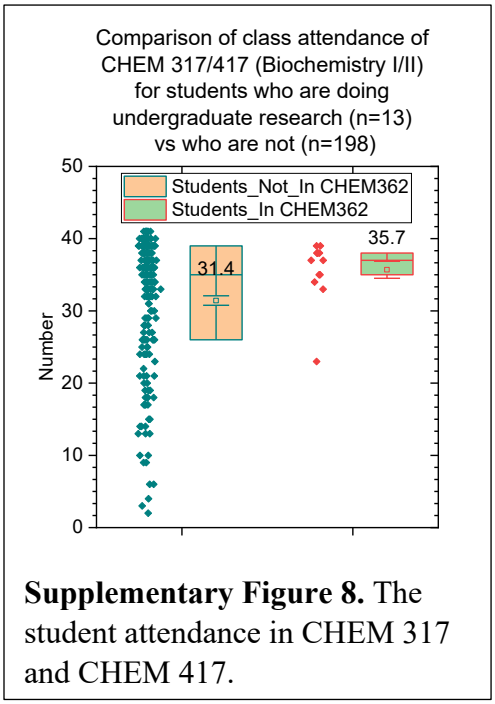
Students can take CHEM 362 before they take CHEM 317, or at the same time or after they take CHEM 317. Most students take CHEM 362 at the same time when they take CHEM 317 or after they took CHEM 317 in the previous semester. Students can register CHEM 362 one or more semesters. Dr. Xu has not put efforts to recruit students to CHEM 362. The students who are interested in biochemistry will normally talk with Dr. Xu. Dr. Xu will introduce them the course content. They will register the class if they are interested. The numbers of students who took CHEM 362 and will take CHEM 362 from 2021 to 2023 are shown in Supplementary Figure 7. The number of students who will take CHEM 362 during the summer or fall of 2023 is based on the currently registered students. A total of 65 students took or will take CHEM 362 from 2021 to 2023. However, some students may take CHEM 362 twice or more. If we count only one for the students who took or will take CHEM 362 twice or more, there will be 43 students. 8.9% (43/483) or 13.5% (65/483 based on the numbers of students in each CHEM 362 class) of the students who take CHEM 317 also take CHEM 362 (Supplementary Figure 7).



Supplementary Figure 7. Number of the students in CHEM 317, CHEM 417 and CHEM 362 from 2021 to 2023

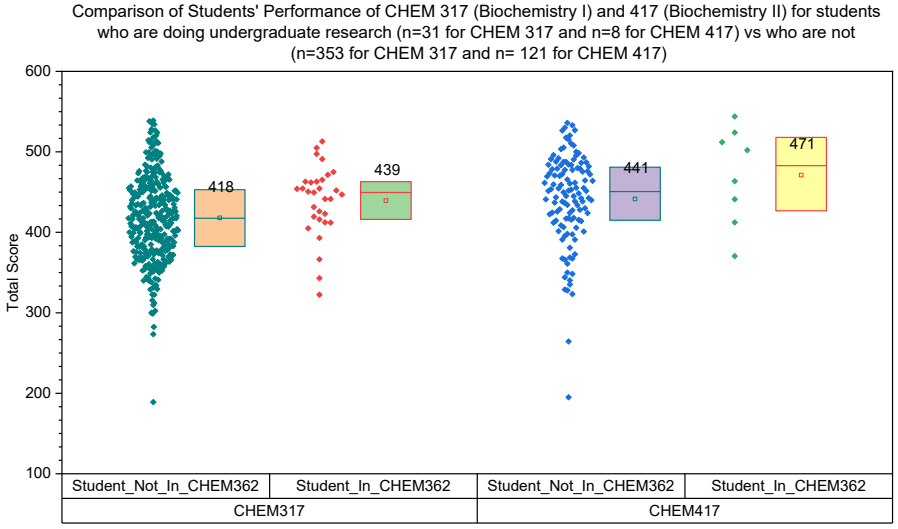
3. Comparison of the student performance in CHEM 317 (Biochemistry I) and CHEM 417 (Biochemistry II) between the students who take CHEM 362 vs who do not take CHEM 362 with Dr. Xu.

Supplementary Figure 8 shows the student attendance in CHEM 317 and CHEM 417 by comparing the students who took CHEM 362 with Dr. Xu and students who did not take CHEM 362 with Dr. Xu. It should mention that some of the students may take CHEM 362 with other professors in the Department of Chemistry at the University of Louisiana at Lafayette. The comparison in the manuscript is specifically for the students who take CHEM 362 with Dr. Xu. The class attendance was done only for the fall of 2022 and the spring of 2023. Dr. Xu did not do class roles from the spring of 2021 to the spring of 2022 due to the pandemic. The class attendance in CHEM 317 and CHEM 417 for the students who took CHEM 362 is 89.3% (35.7/40) is slightly higher on average than 78.5% (31.4/40) of class attendance for the students who did not take CHEM 362. There are a total of 40 classes in CHEM 317 and CHEM 417.



Supplementary Figure 8. The student attendance in CHEM 317 and CHEM 417.

Dr. Xu compared the student performance in CHEM 317 and CHEM 417 for the students who took CHEM 362 vs the students who did not have CHEM 362 with Dr. Xu. There are four quizzes and four exams in both CHEM 317 and CHEM 417. There are 25 points for each quiz and each mid-term exam is worth of 100 points. Final exam is 150 points. Therefore, the total point is 550 for either CHEM 317 or CHEM 417. The student performance of CHEM 417 (441/550) is better on average than that of CHEM 317 (418/550). The performance of CHEM 317 (418/550 vs 439/550) and CHEM 417 (441/550 vs 471/550) for the students who took CHEM 362 is better than the students who did not take CHEM 362. The data presented in Supplementary Figures 8 and 9 show class attendance and student performance in CHEM 317 and CHEM 417 by comparing the students who took CHEM 362 with the students who did not take CHEM 362. It should be pointed out that the observation (higher class attendance and higher performance for the students who took CHEM 362 in CHEM 317 and CHEM 417 than the students who did not take Chem 362) does not lead to a conclusion that conducting undergraduate research in biochemistry will enhance teaching in biochemistry. Instead, we can conclude that the students who perform better in CHEM 317 or CHEM 417 want to gain undergraduate research experiences in biochemistry by taking CHEM 362.



Supplementary Figure 9. The student performance in CHEM 317 and CHEM 417.

4. The detailed procedure of how to incorporate the TSR-based method in CHEM 362.

As stated earlier, one of the objectives of CHEM 362 is for students to learn biochemistry from doing biochemistry. Learning is designed as through a project-oriented approach. The purpose is for students to gain hands-on experience, skills and critical thinking. It is also expected that students retain knowledge or information acquired from CHEM 362 for a long term because of the hands-on approach and an opportunity of choosing proteins based on their own interests. Knowledge learned from attending regular lectures or reading textbooks is often short term. Integration of learning from lectures with project-oriented learning will be more effective in transferring from short-term to long-term knowledge.

Camille Reaux wrote the detailed procedure of how to connect students' computers to the supercomputer clusters, download the software, prepare datasets, submit jobs, monitor jobs and analyze results. Students also prepared a YouTube tutorial to show how to use the TSR-based method for studying protein structures step by step.

Dr. Xu will be happy to help you if you have any questions about how to use the TSR-based method in your teaching or research.

5. The outcome of introducing the TSR-based method in CHEM 362.

Dr. Xu conducted a survey of the factors influencing students' decision on their future careers.

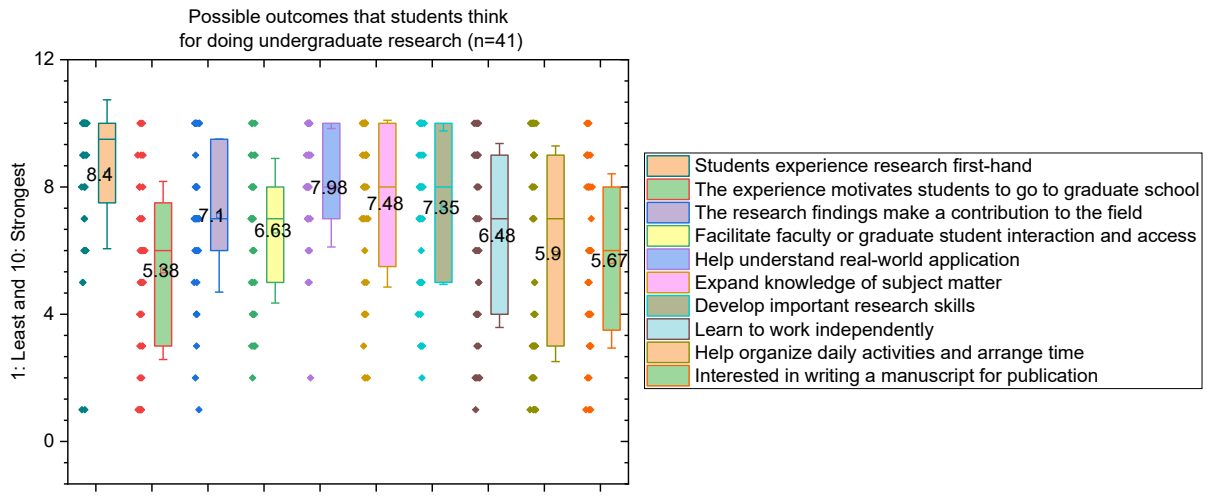
Supplementary Figure 10. A survey showing the factors influence students' decision on their future careers.

The result shows large variations among the forty students who completed the survey. On average, students own interest ranks the highest in influencing their future careers and the influence from their friends and colleagues received the lowest point (Supplementary Figure 10).

Twelve undergraduates, Julia Daigle (a book chapter) [2], Camille Reaux [3, 4], Laura Moore [3], Caleb Collette [5], Sophia Zhou [6], Ali Faust [7], Ashlin Naquin [7], Avery Walton [7], Peter Kishbaugh [7], Ryan L. Fontenot [8], Antoinette Myers [9] and Sara Furman [10] are co-authored in nine TSR related publications since 2020.

Dr. Xu emphasizes the requirement of independently and successfully running jobs and performing analysis after the students received hands-on training. Majority of the students need multiple hands-on training sessions. In the first semester, the spring of 2021, when Dr. Xu started to introduce the TSR-based method in CHEM 362, it took nearly three months for everyone to independently and successfully run jobs and perform the analyses. Now, it took about one month (four one-hour hands-on sessions) for everyone to be able to independently run a job in the spring of 2023. The improvement is largely due to the tutorials prepared by the co-authors and student-student training. Dr. Xu has not introduced the TSR-based method to every student in CHEM 317 and CHEM 417 except the students who want to earn honor credits for CHEM 317 or CHEM 417. To do so, he needs to develop homework assignment related to protein structure analyses. Currently, he is still focused on project-oriented learning in CHEM 362 for stimulating

students' interest in doing research and building their confidence in accomplishing things although could be small rather than lecture-based learning in CHEM 317 or CHEM 417.



Supplementary Figure 11. Possible outcome from a student survey.

The learning outcomes presented is based on the evaluation of student report and presentation. Dr. Xu also conducted a survey for determining what the students have learned from the students' point of view. Again, the variation is high (Supplementary Figure 11), similar to the survey presented in Supplementary Figure 10. On average, the first-hand research experience ranks the highest, and this experience motivating them to go to a graduate school ranks the lowest.

6. References for supporting the educational components

[1] S. Kondra, T. Sarkar, V. Raghavan, W. Xu, Development of a TSR-Based Method for Protein 3-D Structural Comparison With Its Applications to Protein Classification and Motif Discovery, Frontiers in Chemistry 8 (2021).
[2] L. Luo, J. Daigle, Y. Chen, X. Duan, Y. Wang, A. Chistoserdov, Y. Wang, W. Xu, Structural and functional studies of a eukaryotic type Ser/Thr kinase, Slr0599, of Synechocystis sp. PCC 6803 using a combination of experimental and computational approaches, in: H.J.M. Hou, S.I. Allakhverdiev (Eds.), Photosynthesis From Plants to Nanomaterials, Academic Press2023.
[3] T. Sarkar, Y. Chen, Y. Wang, Y. Chen, F. Chen, C.R. Reaux, L.E. Moore, V. Raghavan, W. Xu, Introducing mirror-image discrimination capability to the TSR-based method for capturing stereo geometry and understanding hierarchical structure relationships of protein receptor family, Computational Biology and Chemistry 103 (2023) 107824.
[4] T. Sarkar, C.R. Reaux, J. Li, V.V. Raghavan, W. Xu, The specific applications of the TSR-based method in identifying Zn2+ binding sites of proteases and ACE/ACE2, Data in Brief 45 (2022) 108629.
[5] S. Kondra, F. Chen, Y. Chen, Y. Chen, C.J. Collette, W. Xu, A study of a hierarchical structure of proteins and ligand binding sites of receptors using the triangular spatial relationship-based structure comparison method and development of a size-filtering feature designed for comparing different sizes of protein structures, Proteins 90 (2022) 239-257.
[6] T. Sarkar, V.V. Raghavan, F. Chen, A. Riley, S. Zhou, W. Xu, Exploring the effectiveness of the TSR-based protein 3-D structural comparison method for protein clustering, and structural

motif identification and discovery of protein kinases, hydrolase, and SARS-CoV-2's protein via the application of amino acid grouping, Computational Biology and Chemistry (2021) 107479.
[7] W. Xu, X.-J. Xie, A. Faust, M. Liu, X. Li, F. Chen, A. Naquin, A. Walton, P. Kishbaugh, J.-Y. Ji, All-Atomic Molecular Dynamic Studies of Human and Drosophila CDK8: Insights into Their Kinase Domains, the LXXLL Motifs, and Drug Binding Site, International Journal of Molecular Sciences 21 (2020).
[8] T.I. Milon, Y. Wang, R.L. Fontenot, P. Khajouie, F. Villinger, V. Raghavan, W. Xu, Development of a novel representation of drug 3D structures and enhancement of the TSR-based method for probing drug and target interactions, Computational Biology and Chemistry 112 (2024) 108117.
[9] F. Chen, I.T. Milon, P. Khajouie, A. Myers, W. Xu, A Parallel Implementation for Large-Scale TSR-based 3D Structural Comparisons of Protein and Amino Acid, Current Bioinformatics 19 (2024) 1-16.
[10] T.I. Milon, K. Rauniyar, S. Furman, K.H. Orthi, Y. Wang, V. Raghavan, W. Xu, Representing and Quantifying Conformational Changes of Kinases and Phosphatases Using the TSR-Based Algorithm, Kinases and Phosphatases 2 (2024) 315-339.