# Team 4 Workspace

Teams can collaborate and add ideas, data, code, etc. to capture their work on this Dropbox Paper Workspace during the Biotech + Design workshop on Thursday, January 5th.

Code of Conduct: <https://hackcodeofconduct.org/379-biotech_design_2017>

# Workshop Schedule

|  |  |  |
| --- | --- | --- |
| 10:00 AM | Workshop Introduction & Logistics, including DropBox Paper intro | Susan Baxter |
| 10:15 AM | Problem Pitches | Participants |
| 10:30 AM | Design Thinking Process & Outline for the Day | Susan Baxter |
| 11:00 AM | Teams Get To Work! | Teams |
| Noon | Box Lunches Available |  |
| 4:30 PM | Start Wrapping Up  - Craft Lightning Talks! | Teams |
| 5:15 PM | Lightning Talks | Team Reps |
| 5:45 PM | Crowd Voting: Most Creative Idea, the Broadest Potential Impact & Biggest Potential Collaboration | All |
| 6:00 PM | Workshop Close |  |

Design Thinking Process Prompts: <http://www.csuperb.org/csuicorps/wp-content/uploads/2016/12/Take-Home-Poster-November-2016-CSUPERB.pdf>

<https://www.dropbox.com/s/ava47c76djrcurb/2017%20CSUPERB%20Biotech%2BDesign%20LIghtning%20Talks.pdf?dl=0>

# Team 4:

Jamil Momand (CSU Los Angeles)

“Predicting sites of protein oxidation can help to identify sites of damage, enzyme active sites and allosteric sites. We created and published a software program that is 80% accurate at predicting sites of thiol oxidation.  In 2008, we used structure data from the Protein Data Bank (PDB) to create a database of protein cysteine thiols that were either oxidation susceptible or non-oxidation susceptible. We collected physicochemical properties from the structure data and used a decision tree-generating algorithm to optimize prediction. A software program was created and is available on the internet. The drawbacks to the software program is that it requires a structure for prediction, and it is not accurate for certain types of oxidation (for example nitrosothiols, mediated by nitric oxide). Since 2008, we have collected more data from the PDB and hope to create a software program that is more accurate than 80% and can predict oxidation from primary sequence data.  I want the Biotech + Design Team to help **create a more accurate algorithm for predicting a variety of thiol oxidations**. I want the Team to help create an algorithm that can operate on primary protein sequence. I can supply an excel spreadsheet with data from >150 thiol sites (half are not oxidation susceptible and half are oxidation susceptible).  My current algorithm is the most accurate available at this time.”

*Initial Team 4 Members*: Jamil Momand, Jason Grunberger, Nate Jue, Benigno Mojica, Michael Morikone & Aishani Chittoor Prem, Salvador Mireles

# Team Notes

1. EMPATHIZE: Meet your team members to discover their skills, knowledge and talents!

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Name | Campus | Department | Email | Strengths | Weakness | Interests |
| Aishani Chittoor Prem | SDSU | Bioinformatics | aishaniprem@gmail.com | Protein modeling, programming | Genomics | Programming and  Improving problem solving skills |
| Jason Grunberger | SSU | Bio | jwgrunberger@gmail.com | Cys sulfur redox chemistry  structural protein biochem | algorithms, computational work | Improving my weaknesses |
| Salvador Mireles | SJSU | ChE | salvador.mireles@sjsu.edu | Some programming, chemistry | genetics | Solving interesting problems using software |
| Nate Jue | CSUMB | Biology | njue@csumb.edu | Statistics, Programming, Genomics | Structural Protein Biology | Interesting biological questions |
| Jamil | CSULA | Chem and Biochem | jmomand@calstatela.edu | Protein structure | Programming | Therapies to combat cancer; fundamental knowledge of molecular switches |
| Michael Morikone | CSUSB | Biology/Bioinformatics | morikonm@coyote.csusb.edu | A bit of programming/algorithms, genetics/genomics | Protein biology | Application of computer science and mathematics to solve biological questions. Interdisciplinary research. |

What is out there already?

1. Existing 3-D models (PDBs)
2. Support Vector Machine Learning vs. Associative Neural Networks
3. Random Forest Approach?
4. Other predictors from PDB models (based on 3-D Homology)

What are we really trying to do?

1. User-friendly algorithm/interface? Easily implemented?
2. Visual interpretator
3. Best Oxidation Predictor Possible
4. General Protein Residue Modification Predictor
   * Oxidation
   * Phosphorylation
   * Acetylation
   * Methylation
   * SUMOlyation
   * Ubiquitination
   * Proline-hydroxylation
5. Apply to proteins outside of PDB database - ab initio prediction

What are our hopes and challenges?

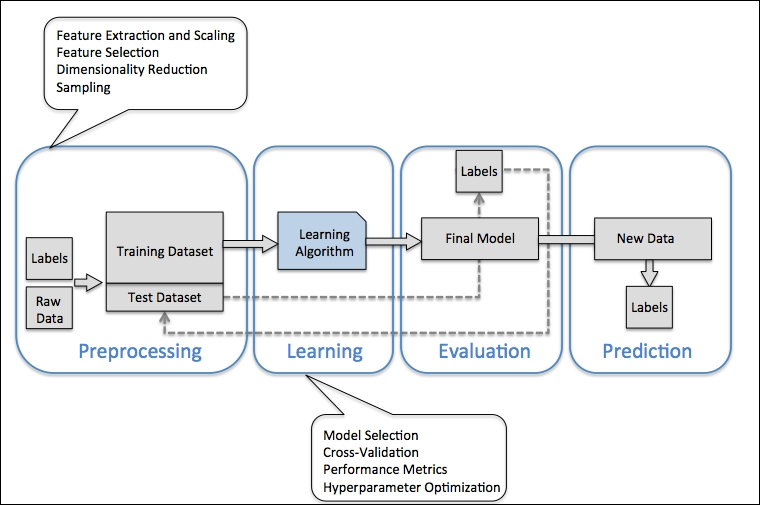
* Big picture
  + Something anyone working on protein modifications will find useful
  + Something new
  + Something fundable or basis for broader project
  + Improve accuracy and involvement of all/many protein modifications
* Small Picture
  + Today - let’s find a good idea today
  + Learn more about protein modification
  + Learn more about tools available for protein analysis
  + Learn more Python and Machine Learning

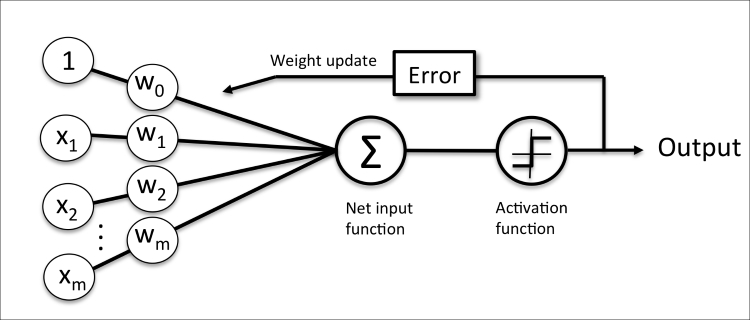
2. DEFINE: Talk about the problem space for a while

* **Identify and map programming/algorithm approach and identify areas of improvement**
  + Develop project outline for predicting cysteine oxidation
  + Expand data base sources past PDB to sites like NCBI
  + Confirmation of significant parameters with other models
  + Brainstorm on whether other parameters or approaches may be important for oxidation
* **Identify a more user-friendly/useful and visual interface approach**
  + Come up with ways to include a visual intepretator.
  + Learn how to  create a visual of the protein structure with modification sites
* **How do we achieve wide range applicability**
  + Brainstorm on whether other parameters may be important for types of protein modification other than oxidation
  + Figure out if oxidation methods would be applicable to other protein modification prediction and how to test
  + Find out if others have used PDB to predict protein modification (other than oxidation).
  + Sketch out how to approach other types of modification for prediction

1. IDEATE: Brainstorm solutions for a while. Seek QUANTITY. Suspend judgment. No idea is to far-fetched and no one’s ideas can be rejected. Ideating is all about creativity and fun. Become silly, savvy, risk takers, wishful thinkers and dreamers of the impossible...and the possible.

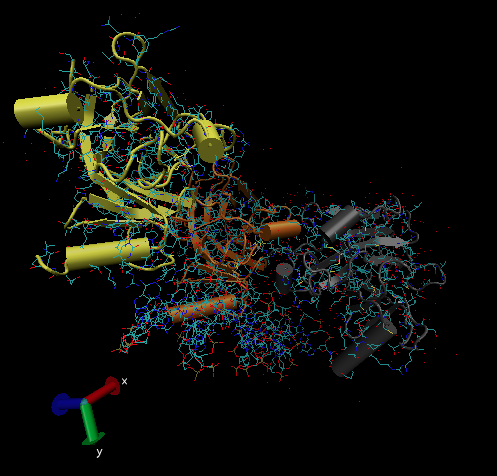
* Algorithm approach and improvement
  + Add more parameters to model prediction
    - Data other than PDB - do these things affect predictions?
      * Site conservation
      * Functional Domains
        + Receptors?
      * Polarity
      * Intra/extracellular or transmembrane region
  + Explore alternative machine learning algorithms





1. * + J48 alternative → non-binomial structure: how to parameterize probability function?
     + Get a good map of current modeling process
     + Identify algorithms we want to test
       - Research this, for example Generative Adversarial Network or
     + Describe algorithm biases and optimal usage strategy
     + Model robustness to training data errors and how to identify them
   * Model comparison and/or sensitivity
     + Approximate Bayesian Computations
     + Bayes Factors or different information criterium
     + Sensitivity or Uncertainty analysis
   * Confidence of outcome
     + Probability associated with oxidation prediction

* Interface design (Web-based tool)
  + Input
    - Text files
    - PDB file
      * automate protein characteristics
    - Amino acid sequence (unmatched)
    - Amino acid sequence (returns all PDB best matched models)
      * Overlapping vs non-overlapping
    - Accession #
    - List of “something”
  + Execution
    - User-friendly usage
    - Progress indicator
    - Built-in parameter optimization/recommendation
    - Easy to understand running condition and run evaluation recommendations
    - Options for “advanced” options (controlled variations of running conditions)
    - no pop-ups
    - Run Finished Notification
      * Link
      * Email/Twitter/Text Message
    - Provide UNIX instructions for remote/local usage
    - Usage controls
  + Output
    - Interactive, 3-D representation



* + - * Degree of confidence of prediction → e.g., info for each cysteine
      * Explore other data visualization tools
        + D3.js implementation?
    - Text-version of output (exportable)
      * Standardized format?
        + One file or multiple files? Optional?
        + Table of results
        + FASTA with stuff
        + Annotated PDB file
        + What are the end-points for users? Other programs, analyses, or visualization tools.
    - Database cross-references
    - Specific bound ligand (from PDB reference file?)
* Broader Applicability:
  + Effects of mutation on:
    - protein characteristics? which ones?
    - functional capabilities
    - changing residue to cysteine for tagging purposes
    - Speed issues?
    - How to estimate structural changes? Or just melting temp changes, for example?
  + Applying cysteine oxidation algorithm to other protein modification predictions
    - Parameterize for other modifications
      * Get examples
      * Iterate training
      * Iterate evaluation and rule establishment

1. SELECT/PROTOTYPE: Rough and rapid portion of the workshop. “[Sacrificial designs](http://www.sciencedirect.com/science/article/pii/S0092867411012086).” A prototype can be a sketch, model, or a cardboard box. It is a way to convey an idea quickly. It is better to fail early and often as you create prototypes.

<https://www.dropbox.com/s/adhbafvjl2zu39q/Combineddata.xls?dl=0>

<http://copa.calstatela.edu/>

1. TEST: An iterative process to get feedback. Learn what works and what doesn’t, and then iterate. Iterate means modifying based on feedback. The final test iteration will be your presentation during the workshop wrap-up!

# Lightning Talks (5:15 PM)

At 5:15 PM EACH TEAM WILL GIVE A 1-2 MINUTE LIGHTNING TALK TO TELL THEIR STORY OF THE DAY!

Write a Lightning Talk to last ~90 seconds.  Create a “poster” to complement your Lightning Talk.

Try to cover the 9 thoughts (listed below) in about 200 words.

• That’s about 22 words per subject for the 9 points.

• The order is not absolute, but each point should be covered.

• The talk should definitely end with points 8 (What’s Next – what should the

listener do?) and 9 (a “wow” idea or tag line that will be the listener’s takeaway

about the idea and/or the team).

1. **Who?** Who are you or who is on the team? From what campus/department? Contact Information?

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Name | Campus | Department | Email | Strengths | Weakness | Interests |
| Aishani Chittoor Prem | SDSU | Bioinformatics | aishaniprem@gmail.com | Protein modeling, programming | Genomics | Programming and  Improving problem solving skills |
| Jason Grunberger | SSU | Bio | jwgrunberger@gmail.com | Cys sulfur redox chemistry  structural protein biochem | algorithms, computational work | Improving my weaknesses |
| Salvador Mireles | SJSU | ChE | salvador.mireles@sjsu.edu | Some programming, chemistry | genetics | Solving interesting problems using software |
| Nate Jue | CSUMB | Biology | njue@csumb.edu | Statistics, Programming, Genomics | Structural Protein Biology | Interesting biological questions |
| Jamil | CSULA | Chem and Biochem | jmomand@calstatela.edu | Protein structure | Programming | Therapies to combat cancer; fundamental knowledge of molecular switches |
| Michael Morikone | CSUSB | Biology/Bioinformatics | morikonm@coyote.csusb.edu | Programming/algorithms, genetics/genomics | Protein biology | Application of computer science and mathematics to solve biological questions. Interdisciplinary research. |

2. **What?** What is the problem or need addressed by your team?

Prediction of reversible oxidation sites on proteins is critical to understanding how proteins are regulated. Current methods are 80% accurate and should be improved to provide better information to protein chemists.

3. **How?** How are you going to address the problem or need? What are your materials

and methods? Do you have institutional data, biological datasets, control data or test bed data?

We will use data collected from the Protein Data Bank that has already tagged cysteine thiols as being susceptible to oxidation. The data consists of distance of thiols to other atoms, exposure to solvent, calculated pKa. The data is divided into two categories: cysteines that are oxidation susceptible and those that are not oxidation susceptible. We will use a machine learning program known as a neural network to classify the cysteine thiols in this data set. We can develop a software program accessible through a user-friendly web site to provide a prediction tool to the science community.

4. **Better?** What are the alternatives out there? Don’t reinvent the wheel! Are they better than what you’re doing now? Do you need an Easier solution? Cheaper? More Sensitive? Faster?

There are no better software tools available (as far as we know).

5. **So What?** Will people accept your solution or results? Are they better enough or

more informative than information already available? Does it really matter?

If the software program is more accurate it can be used by researchers to predict cysteine oxidation sites in proteins. This approach may be generalized to predict other post-translational modifications such as phosphorylation, SUMOylation, ubiquitylation, acetylation etc.

6. **What Now?** What is the status of your solution or project at this time and what are

your needs?

Physiochemical data has been compiled and decision-tree based algorithm implemented online, but there is plenty of room for improvement.

7. **Wishing?** What would you like the workshop participants to think about your team or your idea?

That our team wants the audience to appreciate the importance of cysteine oxidation in protein regulation and that our website will be used by researchers that want to predict sites of protein regulation. If successful, this approach can be used to predict other sites of protein modification.

8. **What’s Next?** What do you want the workshop participants or your team members to do next (the “Ask”)? Contact you? View your Dropbox Paper workspace? Collaborate with you? Write a proposal with you? Review your plans or test your demo/prototype?

View our Dropbox Paper workspace and offer suggestions.

9. **Wow.** Use a memorable “tag” line or “wow” line or mantra that the listener will

remember and take away with them (so they remember to follow up or vote for your team!).

This tool could be the gold-standard for predicting molecular switches that control life.

# Follow-On Reading about Creativity and Design Thinking in Research and Education

CONTEXT and PROJECTS TO WATCH

AAC&U, General Education and Assessment: Design Thinking for Student Learning: <https://www.aacu.org/meetings/generaleducation/gened2017/cfp>

CIRTL Network, Integrating Creativity, Innovation, and Design Thinking in STEM Courses: <http://www.cirtl.net/Creativity-Innovation-Design-Thinking-STEM-Courses> (new website being constructed in November 2016)

Michigan State University, Applying Design Thinking to Academic Plans: <http://hub.msu.edu/after-the-post-it-notes-are-gone-thoughts-on-design-thinking-in-academia/> (and <http://amanyadav.org/CSEER/nsf-supports-project-to-support-first-generation-engineering-students-design-thinking/>)

Smith College, The Design Thinking Initiative: <http://smith.edu/design-thinking/>

University of California San Francisco (2012): <https://www.ucsf.edu/news/2012/12/13250/design-science-ucsf-project-applies-innovative-thinking-research>

University of Cincinnati, Center for Clinical & Translational Science & Training: <https://cctst.uc.edu/funding/designthinking>

BLOGS & ARTICLES

Steve Wilson (2014) “My Stanford Design Thinking Bootcamp Takeaway: Innovation Can Be Repeatable,” Wired. <http://www.wired.com/insights/2014/09/innovation-can-be-repeatable/>

Steve Fyffe (2015) “Design Thinking is About Doing,” featuring Stefanos Zenios. Stanford Business Magazine. <https://www.gsb.stanford.edu/insights/design-thinking-about-doing>

Emi Kolawole (2015) A place for design thinking in academic research. Stanford D.School Blog

<http://dschool.stanford.edu/fellowships/2015/03/24/a-place-for-design-thinking-in-academic-research/>

Rieko Yajima (2015) “Catalyzing Scientiﬁc Innovation with Design Thinking.” DMI Review. <http://onlinelibrary.wiley.com/doi/10.1111/drev.10310/epdf>

Robert I. Sutton and David Hoyt (2016) “Better Service, Faster: A Design Thinking Case Study.” Harvard Business Review. <https://hbr.org/2016/01/better-service-faster-a-design-thinking-case-study>

Daniel Shneiderman (2016) “Can design thinking challenge the scientific method?”  UOP Blog. <http://blog.oup.com/2016/03/design-thinking-scientific-method/>

Nicholas Tenhue (2016) “Why we all need design thinking.” CMS Wire. <http://www.cmswire.com/customer-experience/why-we-all-need-design-thinking/>

BOOKS

Tim Brown (2009) Change by Design: How Design Thinking Transforms Organizations and Inspires Innovation, HarperBusiness ISBN-10: 0061766089 Available at: <https://www.amazon.com/Change-Design-Transforms-Organizations-Innovation/dp/0061766089>

Scott G. Isakson and K. Brian Dorval (2010) Creative Approaches to Problem Solving, SAGE Publications, Inc.; 3rd ed. Edition. ISBN-10: 1412977738 Available at: <https://www.amazon.com/Creative-Approaches-Problem-Solving-Innovation/dp/1412977738>

Want More? See John Spencer’s recommended reading list here: <http://www.spencerauthor.com/2015/06/my-favorite-fifty-books-on-creativity.html/>