Week 9

Protein Engineering

9.1 Protein Post-Translational Modification

11/29:

- Review from last time.
  - What you need for unnatural amino acid incorporation: TGA (amber stop codon), special tRNA to both selectively bind the unnatural amino acid and TGA.
- We now study work by Wang and Schultz (2001).
- tRNA selection.
  - If we do negative selection (correct read through generates toxic barnase), nonfunctional tRNA will cause the bacteria to survive.
  - If the tRNA is only charged by E. coli synthesis, the bacteria will die.
  - If tRNA is only charged by mj syn, your bacteria will die.
  - If the tRNA can be charged by both, the bacteria will die.
  - Now for positive selection: Nonfunctional tRNA? No survival gene, so bacteria dies.
  - But this time, tRNA only charged by mj syn survives! Thus, if you do negative selection and then positive selection, you will select specifically for what you want.
  - We can also do positive selection first and then negative selection. In this case, three bacteria survive first, and then two get eliminated.
  - Positive selection seldom is the opposite of negative selection; negative does specificity, positive does ??
  - This is a very testable topic.
- An orthogonal tRNA.
  - You now have a tRNA that can only be charged by exogenous amino acid residues.
  - Orthogonal tRNA and the right gene gives you great ampicillin resistance
- An orthogonal aminoacyl-tRNA synthetase.
  - Change the specificity of *Mj* TyrRS so that it charges the selected tRNA with O-methyl-l-tyrosine.
  - The crystal structure had been determined for the homologous TyrRS from *Bacillus stearothermophilus* bound to a free tyrosine residue.
  - Five active site residues were found to be within 6.5 Å of the *para*-hydroxyl (yellow).
- Aminoacyl tRNA synthetase selection.

  - Similarly to before, it doesn't matter here if you do positive or negative selection first.
  - If you get read through for the Cm gene (chloramphenicol resistance), the bacteria survives.
  - Survivors contain aaRSs capable of charging any natural or unnatural aa onto the orthogonal tRNA.
  - We don't include the unnatural amino acid in solution for negative selection. Thus, when we have an orthogonal tRNA that only accepts O-MeTyr, these bacteria will survive because they cannot synthesize barnase. Any tRNA that incorporates an*other* amino acid will synthesize barnase and die.

- A mutant tRNA synthetase.
  - A mutant synthetase was selected byat selectively charged the *Mj* suppressor tRNA with O-methyl-L-tyrosine.
  - The substitution removes a hydrogen bond, creating a hydrophobic pocket.

- O-MeTyr as UAA incorporation into DHFR.
  - Dihydrofolate reductase (DHFR) was generated with TAG in place of the third codon and purified by metal-affinity chromatography.
  - Tandem MS of DHFR tryptic digest unambiguously shows complete incorporation of O-methyl-L-tyrosine in the third amino acid position.

- UAAs incorporated *in vivo*.
  - Reactive amino acids for further post-translational modification.
  - Caged (photo-activable) amino acids to study kinetics and mechanisms.
  - Heavy atom-labeled amino acids for structural elucidaiton.
  - Incorporated in *E. coli*, yeast, ...
  - Incorporating a bioorthogonal handle (alkyne) can be done.
  - Heavy-carbon side chain for cross-linking is possible.

- Recoding *E. coli* genome.
  - ...

- Summary.
  - Strategies to increase protein functional diversity.
    - Chemical approaches.
    - Site-drirected mutagenesis: Changing one amino acid to another.
    - Unnatural amino acid incorporation.
  - Non-sense suppression *in vitro* and *in vivo*.
  - Orthogonal tRNA and aminoacyl tRNA synthetase.
    - Directed evolution (negative and positive).
  - Applications of unnatural amino acid incorporation.

- We now start on protein post-translational modification.

- Humans are the most complex organisms on earth. But where does that complexity come from?

- Gene number and genome size.
  - We have 3 billion base pairs in our genome, containing 20k-25k protein coding genes. These constitute about 1% of our genome.

  - Do we have the highest number of genes? No! The most common species of water flea (tiny and translucent) has more than 31k genes (8000 more than us). It has so many genes due to extensive gene duplication, even though we're more complex. We also have this, but it has it more.
  - *C. elegans* is a model organism that is often used, especially in aging studies since it contains all of the human aging genes. It's life cycle is just much shorter and therefore easier to study. It has a similar amount of protein-coding genes to humans!
  - Do we have the largest genome by base pairs? The Japanese flower *Paris japonica* has a 149 billion base pair genome, 50 times the size of our haploid genome.
  - So why are we so complex?
  - We are complex due to meticulous regulation of transcription and translation (the other 99% that's not non-coding is largely regulatory; we don't fully understand it, but we know it's not just junk).
  - Proteome complexity also plays a role.

- **Proteome**: The entire set of proteins that is, or can be, expressed by a genome, cell, tissue, or organism at a certain time.
- **Proteomics**: The study of the proteome, such as proteome profiling.
- There's genome, proteome, **transcriptome**, **kinome** (all kinases), **methylome** (all methylated DNA), etc.
- **Kinome**: The complete set of protein kinases encoded in its genome.
- Post-translational modifications augment proteome complexity.
  - Chemical changes that happen after translation.
  - Lactylation, other example??
  - Common types: Phosphorylation, hydroxylation, glycosylation (N-glycosylation from Dr. Krishnan; super important for cell recognition; Dr. Tang use to have a lecture on it), Lysine acetylation, lysine ubiquitinylation (adding a 76-77 aa peptide instead of a small functional group; usually precedes degradation).
  - You don't have to memorize these; if tested, we will be given a chemical structure.
- Protein phosphorylation.
  - Catalyzed by kinases.
  - Within the kinase active site, you have a base that deprotonates the hydroxyl group, leading the O^- nucleophile to attack the γ phosphate in an ATP via nucleophilic acyl substitution.
  - Protein phosphatase removes phosphates from phosphorylated protein residues, resulting in...
- Functions of protein phosphorylation.
  - 1/3-2/3 of the proteome in eukaryotes can be phosphorylated.
  - Not all of this leads to activity (as far as we know at this point).
  - Phosphorylation alters...
- The human kinome comprises 518 kinases.
  - Tyrosine kinases have their own corner of the evolutionary phylogenetic tree.
  - There are 48 FDA approved kinase inhibitors; more than half are for tyrosine kinases.
  - What's the difficulty? Do we want to inhibit one tyrosine kinase or many? Probably just one, belonging to one protien. Thus, *specificity* is the largest challenge of developing novel kinase inhibitors.

  - Gleevic (or Imatinib) is a drug for Leukemia. Once the...
- MAP kinase pathways: An example of kinase cascade complexity.
  - MAPKKKK lol.
  - One protein can be phosphorylated by many kinases.
- Mass spectrometry-based **phosphoproteomics**.
  - ...
  - There are about 200 phosphorylation sites in the human genome.
  - We don't know the functions of most phosphorylation sites.
- **Phosphoproteomics**: A branch of proteomics that characterizes phosphorylated proteins.
- How can cellular substrates of specific kinases be identified?
  - *In vitro* peptide microarray screens.
  - You take a chip, divide it into 1000-2000 wells.
  - Each well encodes a specific peptide sequence.
  - Then you flow on the kinase you want to study and supply radioactive ^{32}P. Then you can take a radiograph of the chip.
  - No longer widely used because it's not super biologically relevant.
- Tyrosine kinases.
  - Shokat's bump and hole strategy for kinase substrate identification.
  - His lab developed one of the first covalent drugs for ?? inhibition (cancer related).
  - Make ATP larger so that...
- Interaction between mutated kinase and modified ATP.
  - ...
- Kinase/ATP analog engineering allows substrate tagging in cell lysates but not in live cells.
  - ...
- Use chemical tagging...