Week 8

Bioengineering Small Molecules

8.1 RNA Interference and CRISPR

11/15:

- For exam prep on the last four lectures, look at past exams; there will not be a pset on them.
- Tang used to spend 10 lectures on biotechnology. She's gonna try to cover 6-7 and cut out some of the details in the last four lectures.
- Goal: She wants to show us what's out there so we know to look for more details.
- Controlling genetic information flow.
  - In addition to DNA, RNA, and protein sequence diversity, how to *regulate* transcription and translation is even more important for higher organisms.
  - Approaches to regulate/control DNA transcription, RNA translation, and protein activities are promising research tools and drug candidates.
  - How we can perturb this process:
    - Suppose we want to either up- or downregulate the level of a protein.
    - If we want to do this, we do have to do something specifically (we don't want to affect everything).
    - One strategy is a specific inhibitor (this is the most successful thus far).
    - Another approach is to degrade proteins (not regulate it, but change its abundance). Protac is an example; one part binds, the other part attracts a degrader.
    - Another: Antibodies. Different ways to use them PD-1 and PD-L1 example from cancer therapy (what we do in the Lin lab!). Outcome can be elimination of an entire cell.
    - Upregulation is more difficult at every stage, and thus more limited. There are some examples, but nothing is generally applicable.
    - Upregulating protein levels is hard to do by injection; your proteases will usually degrade anything injected. Even proteins wrapped in lipid bylayers are liable to face difficulty entering cells. Notable example: Insulin; only works tho because this protein exists in the bloodstream in general, not in the cells.
    - At the RNA level, we can degrade/cleave RNA, inhibit translation, etc. to downregulate.
    - We can stabilize the RNA and upregulate translation at the ribosomes, in theory. V hard to do tho.
    - Dickinson developed a technology to upregulate translation in a cool way.
    - mRNA delivery is another way to upregulate. Tried for 10 years but not successful until recently via mRNA vaccines.
    - At the DNA level, we can downregulate by changing the epigenetics; methylation corresponds to downregulation; demethylation corresponds to upregulation.

    - DNA editing/knockout to downregulate.
    - Gene therapy allows you to add genes. Why does this even work? The genes that need to be edited aren't that lethal because they didn't kill us in the womb??
  - What level is more programmable?
    - The first programmable interference that went to clinical trials was at the RNA level. At the RNA level, we can use W-C interactions to our benefit.
    - At the protein level, this is harder to get because every protein has a slightly different shape. We can't yet design protein inhibitors; we still find them by screening.
    - Histone binders at the DNA level help us a lot.
    - Today's lecture: Programmable degradation and inhibition at the RNA level; DNA editing. Also RNA inhibit translation and gene therapy to some extent??

- What is RNA interference (**RNAi**)?
  - 1995 (initial observation): Guo and Kemphues introduce antisense RNA into *C. elegans* (worms) to suppress gene expression.
  - 1998 (correct mechanism): Fire and Mello show that dsRNA, not sense or antisense ssRNA, led to degradation of homologous mRNA. The effect is so longlasting that you observe it even in the progeny.
  - Intriguing properties of RNAi: 100% gene inhibition, highly specific, only a few dsRNA molecules are required (catalytic?)
  - Mechanism: Ainylam and Ionis are companies researching this technology. Their stock price has increased much more quickly than the Dow Jones average.
  - Timeline:
    - RNAi discovery, then first RNAi clinical trial a couple of years later.
    - Nobel prize 2006 to Fire and Mello (just 8 years later!).
    - This ignites the field on fire and causes big pharma to enter. They buy up many small companies.
    - Several advanced trials fail, causing big pharma to leave.
    - Innovation of delivery tech revitalizes the field around 2012-2013.
    - There were still some more hiccups, but in 2019, the first RNAi therapeutics were approved (Onpattro).
    - Alnylam's stock price is still doing very well. Ionis (formerly ISIS lol) has switched focus to targeting SMA (spinal muscular atrophy). This disease is caused by one particular DNA mutation. They have a DNA stabilization gene though that causes transcription to occur correctly.

- **RNAi**: Degradation of mRNA triggered by homologous dsRNA.

- Gene silencing in *C. elegans*.
  - Fire and Mello targeted specifically the unc-22 gene. If the gene is intact, the organism won't twitch; if it is disrupted, the organism will twitch. Because *C. elegans* is so simple, there are only a few phenotypes you can follow.
  - Sense and anti-sense RNA show no pheonotype when gel-purified prior to injection.
  - When injected simultaneously or in rapid succession, sense + antisense RNA gives altered phenotypes (worms twitch).
  - dsRNAs that target introns do not affect phenotype, suggesting that the silencing agent acts on spliced, mature mRNAs.
  - ...

  - The gene silencing occurs in *C. elegans* when individuals are fed bacteria that express dsRNA targeting GFP.
  - Embryos are affected by dsRNA, as seen by comparison with RNAi-defective individuals.

- If you inject DNA into yourself, your innate immune response will be activated. If you injest it, you will degrade it and use it as starting materials for building your own genome.
- Why do we need to start with dsRNA?
- Dicer (a protein) chops dsRNA into siRNAs.
  - RNAi can be triggered by...
- Dicer structure.
  - Dicer acts as a "molecular ruler."
  - The RNA-binding PAZ domain lies 65 Å from the catalytic domain.
    - ...
- siRNAs target mRNA for degradation by RISC.
  - The RNA-Induced Silencing Complex (RISC) is a protein-RNA complex that directs siRNAs to their homologous target mRNAs.
  - How does RISC decide to incorporate the target strand or the complementary strand? Open question, but it looks rn like both are used so some is wasted.
  - Argonaute is the RISC...
- The RNA trigger must be double stranded and highly homologous. *picture; what's injected vs. the interference you get*
  - Experimental design: Correct hypothesis is you need double-stranded RNA to get interference.
  - We want to target two genes: GFP and un-22. We can introduce interference RNA for either in all of its forms (dsDNA, ssDNA, with a loop, etc.).
  - See the figure for verification that dsDNA is importance.
  - dsGFP still activates even with an ss sense/antisense loop appended.
  - If you inject dsGFP plus sense/antisense loops, your unc-22 interference varies: We need both to be injected simultaneously and in high enough concentrations so that after dsGFP gets edited out, the sense and antisense strands are present in high enough concentrations to hybridize. If tested and we say yes, that's perfectly fine; it's an experimental condition thing.
  - The farther your gene is from the wildtype, the less interference you get because less complementarity means less binding.
- RdRP may enable transitive RNAi.
  - Why the $3'$ OH group is important: The RNA-dependent RNA polymerase (RdRP) is essential for RNAi in *C. elegans*. RdRP use RNA to synthesize more RNA; *C. elegans* uses RNA to synthesize DNA?? Example: COVID-19 uses RdRP to replicate its genome. Remdesivir (the initial COVID drug) inhibits RdRP.
  - 80%-90% inhibition in humans, but also not as long-lasting.
- Translation inhibition: miRNA.
  - MicroRNAs: Small regulatrory RNAs that block translation of mRNA.
  - Precursors are $\sim$ 70 bp hairpin stem loops processed by Dicer into $\sim$ 22 bp miRNAs.
  - RISC uses the antisense strand of the miRNA to target mRNA and prevent translation.

  - Some RISC components for miRNAs are different than those required for RNAi.

- How does RISC discriminate between RNAi and translation inhibition? *table*
  - Predicting the correct result here will be easy because we know the correct hypothesis, but in research, dreaming up the correct hypothesis and designing an experiment to exclusively prove it is very difficult. Example with ringing a bell, crab runs away, break the crabs legs, crab doesn't run after bell rung, therefore: crab hears through its legs. Things like this happen quite often in research.
  - Hypothesis: Perfect siRNA complementation leads to RNA degredation/cleavage; some siRNA complementation leads to translation inhibition.
  - Engineered siRNA and miRNA lead to RNA degradation; natural (imperfect ones) just downregulate protein levels. All downregulate protein levels (degredation $\Rightarrow$ protien downreuglation)
  - This is a very well designed experiment; Tang encourages us to go through it in detail.

- RNAi summary.
  - ...

- Programmed DNA editing: A search problem.
  - ...

- General approaches for gene editing.
  - ZFNs, TALENs, and CRISPR/Cas 9 can do targeted double strand breaks (DSBs).
  - The repair process is typically done by non-homologous end joining (NHEJ) without caring too much which ends are joined; you may have insertions, deletions, or other changes. This is Gene Disruption by NHEJ.
  - Gene correction by homology-directed repair (HDR) if you have WT donor DNA. Less efficient.

- Genome engineering has a long history.
  - Begins with the discovery of zinc fingers. If you have a long strand of zinc fingers, it is possible to locate one specific DNA sequence.
  - Timeframe of knockout mouse came down from 10 years to six months using ZFs.
  - Knockout mouse is knocking out a specific mouse gene. Can sustain a whole lab, because then you're the only lab that can really investigate the effect of that gene.
  - One TALEN molecule recognizes one base pair (2009).
  - CRISPR arises in 2012.

- Programmable DNA binding domains: Zn fingers.
  - Each ZF: ~ 30 amino acid residues, binds a zinc atom.
  - Highly prevalent in eukaryotes.
  - Each ZF usually recognizes 3-4 bp sequences: It inserts its α-helix into the major groove of the double helix.
  - Can be engineered for different sequences, but design and selection are laborious and time consuming.
  - Normally, 3-6 such ZFs are linked together in tandem to generate a ZFP.

- Zn finger nucleases.
  - FokI nuclease domain: Dimer to function, no sequence specificity.
  - You get cleavage somewhere in the middle, but not single-nucleotide resolution.

- Zn finger nucleases in clinical trials.
  - Primarily driven by Sangamo Therapeutics.
  - Good data for Hemophilia.
  - Partners: Pfizer, Sanofi, Kite-Gilead, Takeda, etc.
  - Zn fingers developed so far before CRISPR that they still have a significant market interest.
- TALENs: Transcription activator-like effector nucleases.
  - Arrived too close to the advent of CRISPR to have a significant market share today.
  - The general structural organization is similar to that of ZFs...
- One of the most beautiful structures of all time.
  - TALE structure: 11 monomers that wrap around a double-helical structure. Different amino acids interact with different kinds of bases.
- Problems with ZFNs and TALENs:
  - Difficult to clone (highly repetitive sequences).
  - Some context dependency.
  - High "activation" barrier to widespread use.
- CRISPR was initially found by microbiologists researching yoghurt in the 1980s.
  - They want fermentation plus phage protection for their bacteria.
  - The exponential trend of CRISPR hasn't stopped, and basically no one is working on Zn fingers and TALENs any more.
- More on CRISPR next lecture.

8.2 Unnatural Amino Acid Incorporation

11/17:

- The final will be cumulative.
- Summary of DNA editing technologies from last time. *table*
  - ...
- **Innate immunity**: We don't have to get trained to respond it.
  - If a liposaccharide invades your blood, you will respond.
- **Adaptive immunity**: Our immune system has to be trained to respond to it.
  - If a virus invades your blood, you have to figure out how to respond to it.
- How CRISPR works in bacteria.
- There is a diversity of types...
- The discovery of the mechanism of CRISPR/Cas9 was quickly recognized as huge by the scientific community.
  - Jennifer Doudna and Emmanuel Charpentier quickly receive the Breakthrough Prize and several years later get the Nobel.
  - Tang heard that a lot of Berkeley students sought to drop their lab work and join Doudna's lab immediately because it was so exciting.

- A two-component system.
  - ...
  - The original report by Doudna: "We propose an alternative methodology based on RNA-programmed Cas9 that could offer considerable potential for gene-targeting and genome-editing applications."
  - ...and the race begins.
  - The original paper got accepted by Science/Nature within three weeks.
  - When Tang was a post-doc, though, she did see a faster acceptance: A way to use CRISPR to do A2G replacement was submitted on a Thursday and after a long weekend submitted on Tuesday. All three reviewers said, you should do this.
- History is never simple; Another guy got a manuscript of the same idea submitted first, but it was stuck in review for a while and published second.
- The key experiment: Cas9 works in mammalian cells.
  - Just 6 months after the Doudna-Charpentier, Feng Zhang (cleavage) and George Church (DNA activation) showed that it did work beyond bacteria in human cells.
- The mechanism.
  - ...
- Comparison of CRISPR/Cas9 with its predecessors.
  - Not as important (science has decreed CRISPR is the winner), but Tang asks us to take a look.
- Why is this such a huge deal?
  - Integrating DNA through homologous integration is a very slow and error-prone progress.
  - Now you can do this making use of double-stranded breaks.
  - Allows you to get model organisms much more quickly.
- The FDA doesn't recognize a single point mutation as GMO.
- Delivering this complex into people could help correct a lot of diseases which are caused by single point mutations.
  - For example, sickle cell disease. The gene editing therapy looks to reactivate fetal hemoglobin (which is more potent than normal hemoglobin so as to steal oxygen from the mother). Early results have been very successful.
- Cas9 and its derivatives for multiple functions.
  - Gene editing/regulation, epigenome editing, chromatin imaging (via fusing fluorescent proteins), target RNAs, chromatin topology, etc.
- We now end our discussion of CRISPR and switch to today's content.
- Overview.
  - Strategies to increase protein functional diversity.
    - Chemical approaches.
    - Site-directed mutagenesis (won the NObel along with PCR).
    - Unnatural amino-acid incorporation.
  - ...
- Increasing protein functional diversity.

  - Why? Consider the serine protease. If we want it to be responsive to light, ...

- Chemical approaches.
  - What kind of functionality can we look to invoke?
  - Suppose we have a cysteine. Use its thiol to interact with another compound as a Michael acceptor. Lysine can form amide bonds.
  - ...
  - If you do this *in vivo* post-translationally, you will have no selectivity (every relevant amino acid will be activated toward the reaction). Thus, yuo want to do this in a test-tube whenever possible.
  - Synthetic or semi-synthetic approaches. We don't need to know how solid phase peptide synthesis works.
  - Native Chemical Ligation: Conjugating short peptides modified at the C-terminus with a (highly reactive) thioester. If your second protein starts with a thiol, you will form a ...
  - We will not be directly tested on NCL (if she gives us an example, she will show us the mechanism).

- Site-directed mutagenesis.
  - Engineering the DNA sequences to obtain desired mRNA sequences.
  - Limited to 20 natural amino acids.

- Protein translation.
  - Machinery for protein translation: mRNA (genetic information), tRNA, tRNA synthetase, and the ribosome.
  - Unnatural amino acid incorporation: The codon, the stop codon, accept tRNA charged with uAAs, ribosome, tRNA, tRNA synthetase (1. not recognized endogenous tRNA; only cognate tRNA; 2. not recognize natural amino acids, only uAA, so we don't get a mixture but get pure replacement).
  - We can't use any codons that are already coding (we will get massive competition).

- Genetic code.
  - We can hijack the UGA stop codon.

- How do you know that you won't get some actual stop codons?
  - You do get some truncated protein, but these typically have short half-lives and do not interfere.
  - You do want high efficiency, though.

- Two rare amino acids in nature beyond the 20.
  - In nature, selenocysteine and pyrrolysine are already incorporated from time to time.

- *In vitro* incorporation of unnatural amino acids.
  - ...

- An example of uAA on protein function: Acetylcholine receptor.
  - ...

- Fluorination...

- Incorporation of uAAs to acetylcholine receptors of oocyte.
  - ...

- Binding of acetylcholine to Trp149 receptor analogs.
  - ...
- Challenge of this *in vivo* method.
  - Frog eggs are 10x larger than human eggs and can be seen with the naked eye.
  - This allows you to do your microinjection by hand.
- Requirements for genetically encoded uAAs.
  - A unique tRNA with a unique codon (amber nonsense codon UAG).
  - A corresponding aminoacyl-tRNA synthetase (aaRS).
  - Sufficient...
- An orthogonal tRNA-codon pair.
  - Extended loop in *E. coli* tRNA.
  - Shorter loop in an archea.
  - Gives you orthogonality: Negative selectino against aminoacylation by endogenous *E. coli* synthetase using a cytotoxic nonsense barnase gene.
  - Positive selection: for tRNA that can be charged by the archaea synthetase...
  - Something here that Tang really likes to test??
- tRNA synthetase.
  - Barnase is a toxin (if the gene is expressed, the bacteria will die).
  - Suppression and no suppression of TAG stop codons; you need to read through this stop codon sometimes.
  - Types of tRNA: non-functional tRNA, tRNA only charged by endogenous synthetase, tRNA only charged by orthogonal synthetase, tRNA charged by both.
- An orthogonal tRNA.
  - Validating that your tRNA is orthogonal to the entire *E. coli* system.
- More on directed evolution next time.