

# Week 6

## Import and Export

### 6.1 Organelles and Transport

- 11/1:
- Warm-up activity: Quiz questions.
  - This class and next class: Protein localization mechanisms and inhibition.
  - Think about...
    - How a cell transports and localizes proteins;
    - Mechanisms of preventing things from going where they should.
  - Indeed, some protein inhibitors work not by inactivating proteins but by making sure they don't get to the right place.
  - **Chaperone:** A small molecule that allows a misfolded protein in the wrong place to fold and reach its site of action.
    - These are promising new drugs.
    - Example: There is a known risk gene for Parkinson's disease. A protein gets stuck in the ER. If there are small molecules you can use to get the protein to fold in the ER and be released, you win.
    - Example: Cardiovascular disease. Most drugs fail clinical trials right at the last stage of testing because they cause something called **long QT syndrome**.
      - Said drugs cause this syndrome by preventing ion channels from reaching the plasma membrane.
      - Chaperones could potentially help overcome this common barrier.
  - **Arrhythmia:** A fast, chaotic heartbeat.
  - **Long QT syndrome:** A heart signaling disorder that can cause arrhythmias.
    - Symptoms can be severe, up to death.
  - Today: Mechanisms of protein localization.
    - How do proteins reach the nucleus, mitochondria, and a mystery organelle? These are open questions in basic biology.
  - Organelles and membranes by the numbers.
    - Cytosol: 2% of the membrane in a cell, but 54% of total cell volume.
    - Thus, the plasma membrane spends a lot of energy keeping the cytosol happy.

- The mitochondria and ER have a huge amount of membrane but very little volume and contribute to keeping the cytosol happy as well.
  - The membrane content helps maintain homeostasis.
- What was the net point of all this??
- Evolution of compartments.
  - Helps us understand **topological equivalence**.
  - Yamuna believes that the endosymbiotic theory is just a hypothesis and that it's all up in the air and likely to change.
  - Current hypothesis: Archaea lost its cell wall making it easier for it to acquire DNA. Once it acquired enough valuable genes, the cell membrane underwent an invagination to form the nucleus and extra folds of the ER. This prevents the cell from losing the DNA it's acquired.
    - This is why the ER and extracellular matrix are topologically equivalent, i.e., because the former evolved from the latter.
  - Mitochondria are the cell intaking another bacteria that could produce energy.
- **Topologically equivalent** (compartments): Two compartments inside (or outside) a cell such that materials do not have to cross a membrane to get from one to the other.

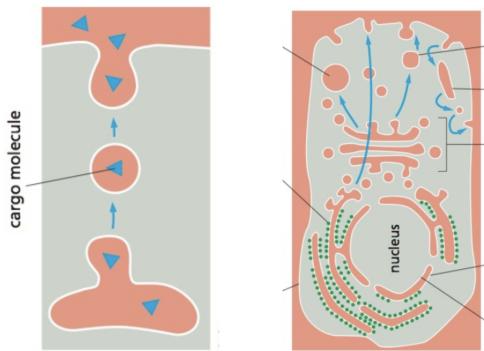


Figure 6.1: Topological equivalence.

- Example: Golgi and ER are topologically equivalent (and topologically equivalent to the extracellular matrix) but not to the cytoplasm. This is because materials in the ER move to the Golgi and then to the extracellular matrix within a **vesicle**, i.e., they never have to cross a plasma membrane so much as they get surrounded and moved by different membranes.
- Example: The extracellular matrix and the cytoplasm are not topologically equivalent. Notice that any material coming into the cytoplasm from the outside must cross through the plasma membrane using one of the mechanisms from last class (we're not talking endocytosis yet).
- Proteins can be transported between organelles either by being stuck in the membrane of a vesicle (at which point they will end up in the membrane of the target organelle) or within said vesicle's lumen (at which point they will end up in the lumen of the target organelle).
- Isolating organelles.
  - We discover transport mechanisms by carrying out a lot of mutations and then isolating specific target organelles and testing for a protein's presence (look for a ratio between the quantity of this protein present and a standard protein that you know will be there).

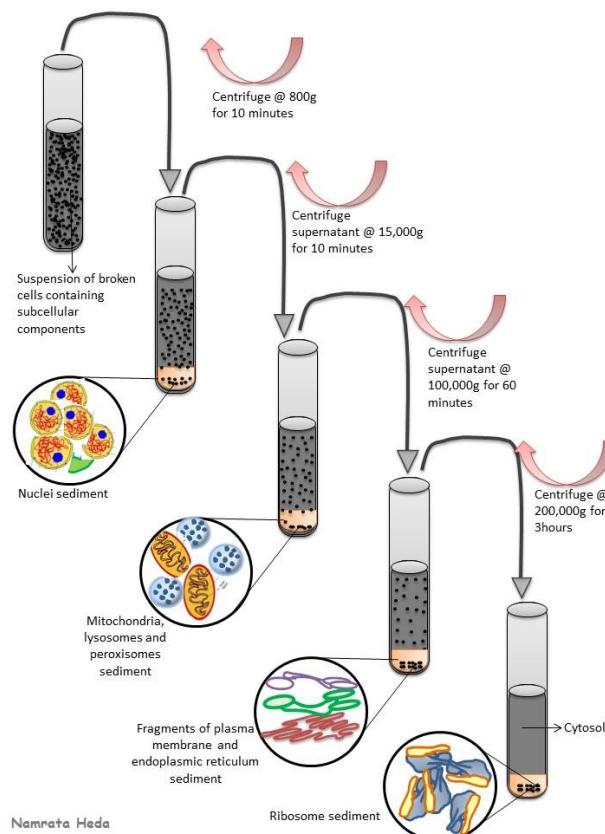


Figure 6.2: Isolating organelles.

- Isolation mechanism: We take a bunch of cells, dissolve the extracellular membrane to get a soup of organelles, centrifuge it (to let heavy organelles like the nucleus fall out), centrifuge again (to get the small organelles like the mitochondria, lysosomes, and peroxisomes), centrifuge it again (to get fragments of the plasma membrane and ER), and centrifuge it one last time (to get ribosomes).
- Alternative isolation mechanism: Use a matrix with a density gradient and just centrifuge once to get multiple layers.
- Three main ways to move proteins in a cell.
  - Recall passive and active exchange.
  - You can let physical equilibrium take hold, of course, but often that won't lead to great enough concentrations.
  - Thus, cells evolved the following three methods...
- **Gated transport:** A type of transport involving a gate and a condition (e.g., a binding or membrane potential) that must be satisfied for the gate to "lift open."
- **Translocation:** The movement between topologically nonequivalent compartments.
  - Example: Suppose you have a protein that's been made in the ER, has been deposited into the cytoplasm, and now needs to get into the mitochondria. We will consider this example in much greater detail shortly.
- **Vesicular transport:** The movement of biomolecules in vesicles between topologically equivalent compartments.

- **Localization sequence:** A molecular GPS. *Also known as nuclear localization sequence, NLS.*
  - Usually located on the N-terminus because it comes out first and needs to know where to go.
  - Localization sequences have different strengths. Some send proteins in a high fraction somewhere; some send proteins in a low fraction somewhere.
    - Strength is determined by the sequence's affinity for the transport protein. We will discuss this in more depth later.
  - Length: Tetrapeptides up to 20-30 AAs.
  - Very occasionally occur in the middle of a protein.
- **Translocation sequence:** A molecular GPS on the C-terminus that moves a protein after folding.
- Gated transport example: Movement from the cytosol into the nucleus.
  - This is also the most common type of gated transport.
  - The gate is the nuclear pore, and the condition is **karyopherin** binding
- **Karyopherin:** A protein involved in transporting molecules between the cytoplasm and the nucleus.
- Consider first the structure of the **nuclear pores**.
- **Nuclear pore:** A gateway from the cytosol to the nucleus.

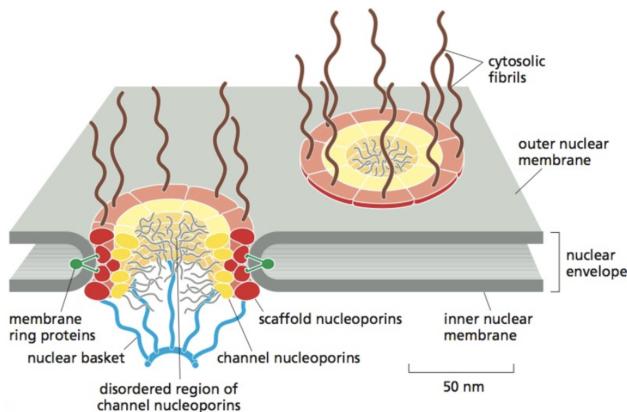


Figure 6.3: Nuclear pore structure.

- Nuclei have double plasma membranes and nuclear pores. All transport in and out of the nucleus occurs via nuclear pores.
- **Membrane ring proteins** make the membrane bend backward around nuclear pores.
- There are about 3000 nuclear pores per nucleus.
- About 1000 molecules transport both ways per nuclear pore per second.
- Active v. passive transport: Anything smaller than 40 nm will freely diffuse to a significant extent (smaller implies higher passive transport). Larger, you need something to capture it and drag it through (this is active transport).
- Nuclear pores are 8-fold symmetric bodies.
  - We still don't know the complete structure.
  - Composed of **nucleoporins**.
  - Hair-like **cytosolic fibrils** on the outside and a **nuclear basket** on the inside.
  - A porous plug in the center; still don't know what it is, but it's made of lots of FG repeats.

- **Nucleoporin:** A protein that is a constituent building block of the nuclear pore complex. *Also known as nap.*
  - There are permanent naps, but there are also naps which come off and on.
  - Approximately 30 exist.
  - Some are transmembrane.
- Probing NLS-enabled nuclear import.

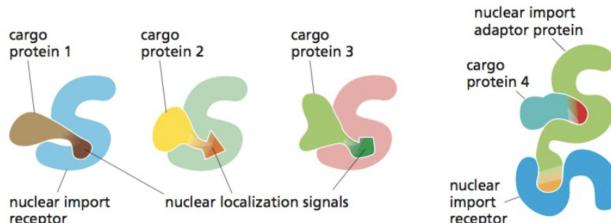
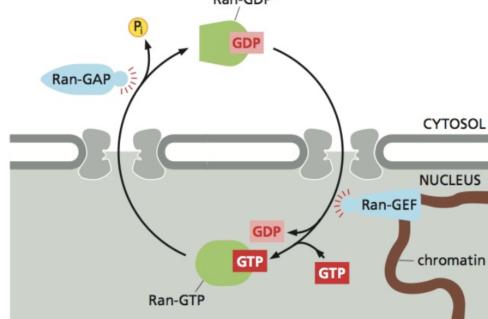


Figure 6.4: Nuclear import receptor binding.

- Suppose we fuse an NLS with GFP (for which a Nobel Prize has been awarded).
- If, in the NLS sequence, we change a K → T, then localization of the sequence is significantly decreased.
  - Review: Replacing positively charged lysine with polar threonine would certainly affect the interaction between the NLS and the **nuclear import receptor**!
- Conclusion: We can affect NLS efficiency by altering one's affinity for its karyopherin (or vice versa), or by altering the ability of the karyopherin to enter the nucleus.
- “It depends upon which bus you get on and upon the strength of your ticket.”
- Benefit of differential binding affinities: It is possible to have different concentrations of different proteins. You don't want all proteins in the nucleus to have the same concentration, after all.
- There also exist **nuclear import adaptor proteins** which link cargo proteins to their nuclear import receptors with higher binding affinities.
- Nuclear export is the reverse of nuclear import.

- Before we can discuss nuclear export directly, we should discuss the Ran proteins.



(a) The Ran proteins.

Figure 6.5: Nuclear import and export mechanism.

- Ran complexes have a domain called a **GTPase domain**.
- Ran's GTPase domain has GTP- and GDP-bound forms.

- There is a Ran-GDP / Ran-GTP gradient across the nuclear membrane: Ran-GTP is present in much higher concentrations within the nucleus, and Ran-GDP is present in much higher concentrations outside the nucleus.
- Ran-GAP is a **GAP** for Ran-GTP and Ran-GEF is a **GEF** for Ran-GDP.
- Ran-GAP is localized in the cytosol, and Ran-GEF is localized in the nucleus (it sits on chromatin inside the nucleus).
- We are now ready to discuss nuclear import and export.

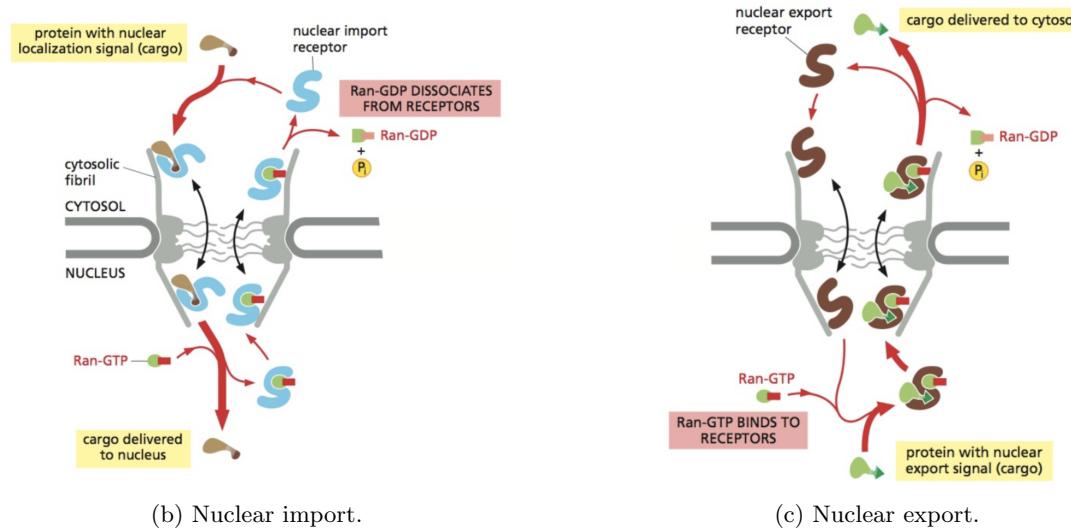


Figure 6.5: Nuclear import and export mechanism.

- Nuclear importers first bind their proteins. Their hydrophobic regions are then caught by the cytosolic fibrils. Moving downward into the FG repeats, the importer's movement once inside is a random walk.
- Once an importer arrives in the nucleus, Ran-GTP attacks. Ran-GTP has a higher affinity for it than its substrate, so it will bind and cause the substrate to fall off, completing delivery to the nucleus.
- When the Ran-GTP-bound importer diffuses back out of the nucleus, Ran-GAP promotes GDP hydrolysis, and Ran-GDP dissociates.
- Nuclear export receptors random walk into the nucleus, bind a Ran-GTP, engage the cargo, random walk out of the nucleus, Ran-GAP hydrolyzes Ran-GTP to RanGDP which leaves, and this kicks out the cargo.
- Note that as we would expect for an example of gated transport, a condition is met and only then does transport occur.
- **GTPase domain:** A region of a protein that hydrolyzes a GTP to release energy, accelerating and powering the function of the protein.
  - Carried by many kinds of proteins and is very powerful.
  - Can help a ribosome work, help proteins move from the nucleus to the cytosol, promote vesicle fusing, etc.
  - Essentially functions as a backpack with a battery.
- **GAP:** A protein that promotes GTP hydrolysis in a GTPase domain that's already bound to GTP. *Also known as GTPase activating protein.*
- **GEF:** A protein that exchanges GDP for GTP at the GTPase domain. *Also known as Guanine nucleotide exchange factor.*

- Translocation example: Movement from the cytosol into a mitochondrion.
  - Mitochondria have proteins that sit specifically on the outer membrane, inner membrane, in the interluminal space, or in the center of the matrix. This indicates very high accuracy and targeting.
    - Many mitochondrial diseases occur due to poor localization.
  - The lessons here are broadly applicable.
  - Before next class, brush up on translation.
  - This is an example of **post-translational** protein transport.
  - There is also **co-translational** protein transport, but we'll talk about that another day.
- **Post-translational** (protein transport): Having a protein cross a membrane after its ribosome has finished synthesizing it.
- **Co-translational** (protein transport): Having a protein cross a membrane as it is still being synthesized by a ribosome.
  - Usually happens in the ER.
- There are four main mitochondrial proteins/complexes to consider: **TIM**, **TOM**, **SAM**, and **OXA**.

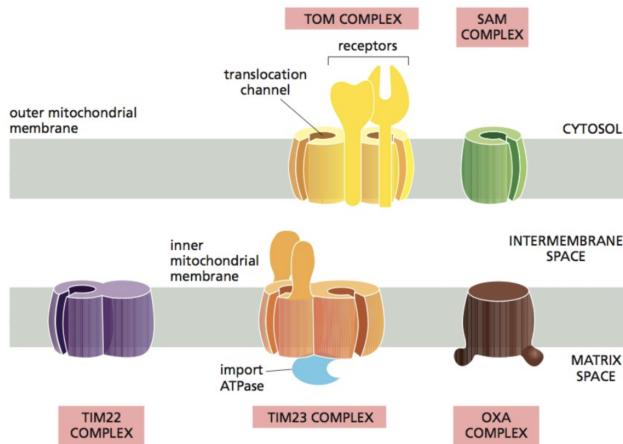


Figure 6.6: Mitochondrial translocators.

- These are all big, multiprotein complexes assembled on the various membranes.
- **TOM complex:** The mitochondrial complex of proteins — localized in the outer membrane — responsible for the movement of proteins through this barrier and into the interluminal space. *Also known as translocase of the outer membrane.*
- **TIM complex:** The mitochondrial complex of proteins — localized in the inner membrane — responsible for the movement of proteins through this barrier and into the matrix. *Also known as translocase of the inner membrane, TIM23.*
- **SAM complex:** The mitochondrial complex of proteins — localized in the outer membrane — responsible for the folding/embedding of proteins into the outer membrane. *Also known as sorting and assembly machinery complex.*
- **OXA complex:** The mitochondrial complex of proteins — localized in the inner membrane — responsible for the movement of proteins through this barrier and into the matrix. *Also known as oxidase assembly complex.*

- There is a way to get TIM and TOM to lock together so translocation happens all at once from the cytosol to the matrix (instead of having to pass through the interluminal space).

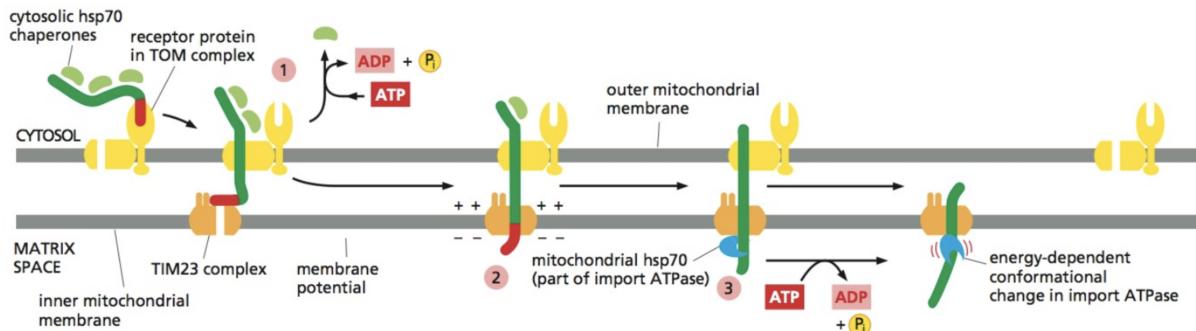


Figure 6.7: Translocation from the cytosol to the mitochondrial matrix.

- Role of energy in protein import into the mitochondrial matrix.
  - Every ATP hydrolyzed at TOM causes you to pull the protein through by a couple of peptides.
  - Membrane potential drives TIM.
- Once the whole protein has been pulled through TOM, TOM and TIM separate.
- Once the translocation sequence has completely entered the matrix, a signal peptidase cleaves it, trapping the protein in the matrix.
- We now talk about how proteins are sent to each membrane.
- Sending proteins to the outer membrane.

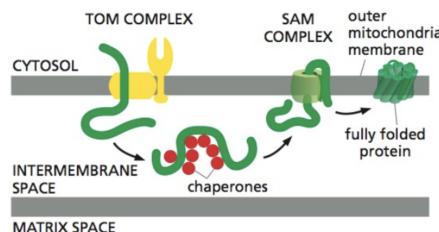


Figure 6.8: Translocation from the cytosol to the mitochondrial outer membrane.

- Many porins are present on the outer membrane.
- A protein first gets pulled into the interluminal space.
  - Aside: In bacteria, this space is known as the **periplasm**.
  - The periplasm is very nice for protein generation because there are very few proteins there and once you clone proteins there, all you have to do to release them is crack open the cell wall.
- When proteins get pulled into the intermembrane space, they are all hydrophobic (because they will reside in a phospholipid bilayer eventually). Thus, they are prone to aggregation in their water-based media, but chaperones latch on to separate them. Once stabilized in the intermembrane space, the protein then gets sent to SAM which folds it into the membrane. SAM has a slit, so as it pulls peptides in, it ejects them out laterally into the membrane.
  - Aside: In a bacteria, there is an analogous BAM complex.
  - Implication: This process is conserved between bacteria and mitochondria, further supporting the endosymbiotic theory.

- Sending proteins to the inner membrane.

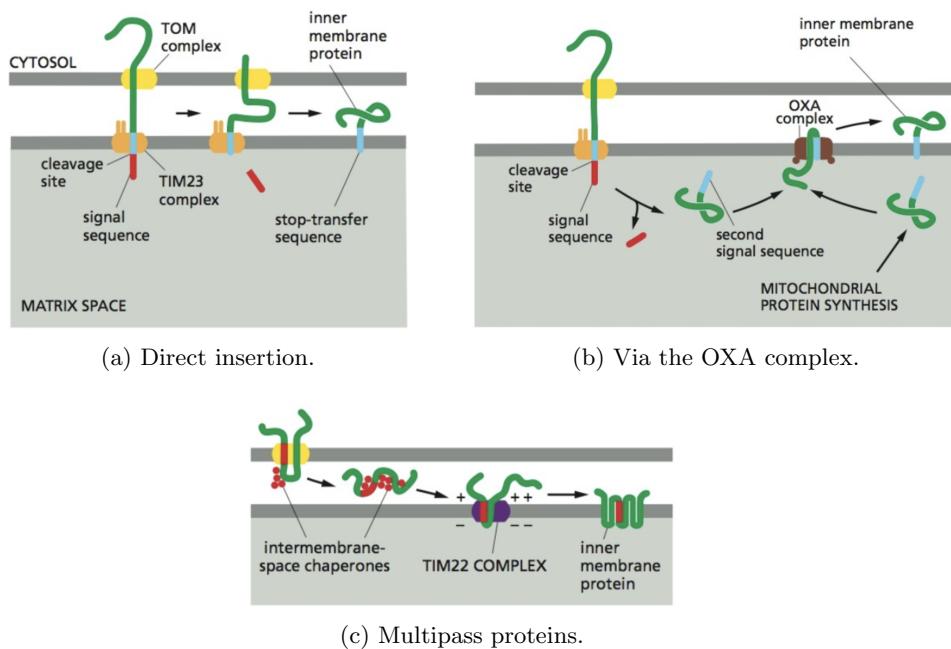
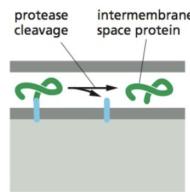


Figure 6.9: Translocation from the cytosol to the mitochondrial inner membrane.

- There are three methods by which this can occur.
- Method 1 (Figure 6.9a).
  - This method directly inserts single-pass transmembrane proteins into the inner membrane.
  - Translocation begins as if the protein is to be pulled into the matrix, but immediately following the localization sequence, there is a stop-transfer sequence. When TIM interacts with this, it stops pulling the protein through, signal peptidase cleaves off the localization sequence, and TIM ejects the hydrophobic stop-transfer sequence into the inner membrane.
  - Once TOM finishes pulling the bulk into the interluminal space and the protein refolds, we are done.
- Method 2 (Figure 6.9b).
  - Use the OXA complex.
  - The TIM/TOM complex moves a protein into the matrix and a single peptidase cleaves off the localization sequence.
  - A secondary tag following the localization sequence then engages the OXA complex. The OXA complex flips the protein so that the tag is in the inner membrane and the bulk of the protein is in the interluminal space.
- Method 3 (Figure 6.9c).
  - Multipass membrane proteins are introduced via the **TIM22 complex**.

- Import from the cytosol into the intermembrane space.



(a) Inner membrane cleavage.

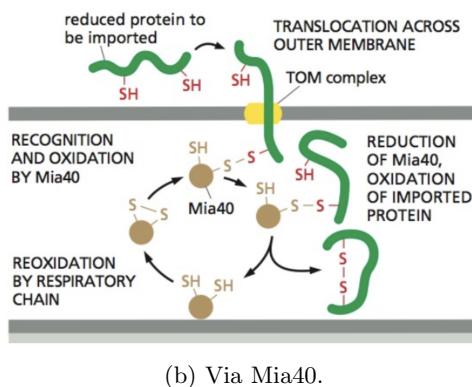


Figure 6.10: Translocation from the cytosol to the mitochondrial interluminal space.

- If after insertion into the inner membrane, the bulk is cleaved from the transmembrane region, it will float away in the interluminal space. This is a secondary mechanism by which proteins enter the interluminal space, in addition to direct import by TOM.
- A third (and very popular) mechanism leverages disulfide bonds. These help the protein fold, but if the protein is to be pulled through TOM, these will have been reduced to split them and unfold the protein. When the reduced disulfide bonds interact with **Mia40** in the interluminal space, they get reassembled and Mia40 gets regenerated (it is a catalyst). This refolding sticks the protein in place.
- Next time: Molecular mechanism of translocases; how they release transmembrane domains in the right origin.
- Peroxisomes.
  - We understand very little about their function, but if anything is wrong with them, it's deadly.
  - Take long lipid chains and cut them into shorter lipid chains by peroxidizing them (they have many reactive oxygen species).
  - Smallest organelle in the cell (50-100 nm) and has a very small number of proteins.
  - Peroxisomes are thought to be born from the ER via budding. Then the peroxisome must mature and acquire proteins, both in its membrane and in the lumen.
  - There are peroxisome targeting sequences, but who takes proteins to the peroxisomes and how they are transferred into the peroxisome is not clear.
  - Peroxisomes are thought to undergo fission for replication, but we have no way to distinguish early peroxisomes from mature, functional ones.
  - Peroxisomes carry their own catalase (which reduces oxygen into water and ROSs).
- Summary of today.
  - How compartments evolved; the evolution determines what is topologically equivalent to what. Nucleus and cytosol are equivalent (transport is facilitated by import and export factors), but most organelles are topologically equivalent to the extracellular matrix. You can take proteins to the nucleus or cytosol once they're born or to the mitochondria, or to specific places in the mitochondria.
  - Chemists are the best inventors, but they don't have a very good understanding of cell biology.
  - GTPs are used for big conformational changes.
    - The energy from hydrolyzing GTP and ATP is the same; it's just a question of how you use it molecularly.
  - Yamuna is very knowledgeable about a variety of topics in her field.

## 6.2 Quiz Prep

From Wu et al. (2020).

### Notes

11/2:

- **Single-stranded DNA:** DNA that is not currently bound and hydrogen bonded into a double helix.  
*Also known as ssDNA.*
- **Transcription:** A highly dynamic process that generates ssDNA as transcription bubbles.
- **Transcription bubble:** A portion of the double helix that has been unwound and separated for the purpose of transcribing one strand of it.
- **KAS-seq:** Kethoxal-assisted single stranded DNA sequencing, the subject of this paper, provides rapid (within 5 mins), sensitive, and genome-wide capture and mapping of ssDNA produced by transcriptionally active RNA polymerases or other processes *in situ* using as few as 1000 cells.
  - Kethoxal is a small molecule that rapidly and selectively binds with unpaired guanine.
  - Attaching an azide group to kethoxal allows it to be tagged with bio-orthogonal click chemistry.
  - Applications of KAS-seq.
    - Definition of a group of single-stranded enhancers that enrich unique sequence motifs.
      - Specifically, these enhancers are associated with the binding of specific transcription factors and exhibit elevated enhancer-promoter interactions.
    - Discovery: When **protein condensation** is inhibited, RNA polymerase II (Pol II) rapidly releases from a group of promoters.
    - Fast and accurate analysis of transcription dynamics and enhancer activities simultaneously in both low-input and high-throughput modalities.
- **Protein condensation:** Proteins sticking together.
- **Chromatin:** The material of which the chromosomes of organisms are composed, consisting of protein, RNA, and DNA.
- Transcription and its regulation (importance).
  - Determine physiological function and the cell's fate.
  - Regulation issues often lead to disease.
- **Global transcription regulation:** Regulation of transcription across the entire genome.
- How do we understand global transcription regulation?
  - Employ techniques like **ChIP-seq**.
  - Search for the presence and level of **nascent RNA**.
    - Based on **run-on assays**, **metabolic labeling**, and Pol II-associated or chromatin-associated RNA enrichment.
- **ChIP-seq:** A genome-wide sequencing approach that analyzes the occupancy of RNA polymerases.
- **Nascent RNA:** Newly made RNA, often still tethered to the DNA axis by elongating Pol II and being continuously altered by splicing and other processing events during its synthesis.
- **Assay:** An experimental method for assessing the presence, localization, or biological activity of a substance in living cells and biological matrices.

- **Run-on assay:** A method for measuring the frequency of transcription initiation. *Also known as nuclear run-on assay. Procedure*
  1. Take cells. At the time you want to measure the frequency of transcription initiation, freeze them.
  2. Reheat them and incubate at 37 °C in the presence of NTPs and radiolabeled UTP.
  3. Measure the amount of radiation given off by the products.
  4. The above measurement will be roughly proportional to the number of nascent transcripts on a gene at a certain time, which in turn is thought to be proportional to the frequency of transcription initiation.
- **Metabolic labeling:** The process of using the synthesis and modification machinery of living cells to incorporate detection or affinity tags into biomolecules.
- Limitations of the current methods for understanding global transcription regulation.
  - Run-on assays and enrichment require millions of cells as starting material.
  - Pol II ChIP-seq cannot distinguish whether RNA polymerases are simply bound or are actively engaged in transcription.
  - Metabolic labeling cannot measure low-abundance RNA species. Post-transcriptional processing can also alter results.
- If we want to understand global transcription regulation, then certainly it will be important to determine where transcription occurs.
- Goal: Locate where RNA polymerases engage in transcription.
  - Observation: RNA polymerases transform dsDNA to ssDNA bubbles as they move.
  - Method: Label/tag/identify/characterize ssDNA.
- Previous attempts: MnO<sub>4</sub><sup>-</sup> preferentially oxidizes single-stranded thymidine residues.
  - Has been used to reveal Pol II-induced promoter melting locally and on a genome-wide basis.
  - Works together with S1 nuclease digestion.
  - Doesn't work on B DNA.
  - Limitations:
    - Requires tens of millions of cells.
    - Shows low sensitivity for weak/broad signals at Pol II elongation sites.
- Outline of the paper.
  - Describe KAS-seq.
  - Prove that it simultaneously measures the dynamics of transcriptionally engaged Pol II, transcribing enhancers, Pol I and Pol III activities, and non-canonical DNA structures in which ssDNA plays a major role.
  - Prove that it works with as few as 1000 cells.
  - Prove that KAS-seq detects changes in transcription during quick environmental changes, e.g., inhibition of protein condensation.
- Note that most conclusions listed here have supporting data and correlation numbers given in the paper.
- Genome-wide profiling of ssDNA using N<sub>3</sub>-kethoxal-based labeling.
  - Prior literature: Kethoxal reacts with the N1 and N2 positions of guanines (the ones that form Watson-Crick interactions) in ssDNA and RNAs under physiological conditions.

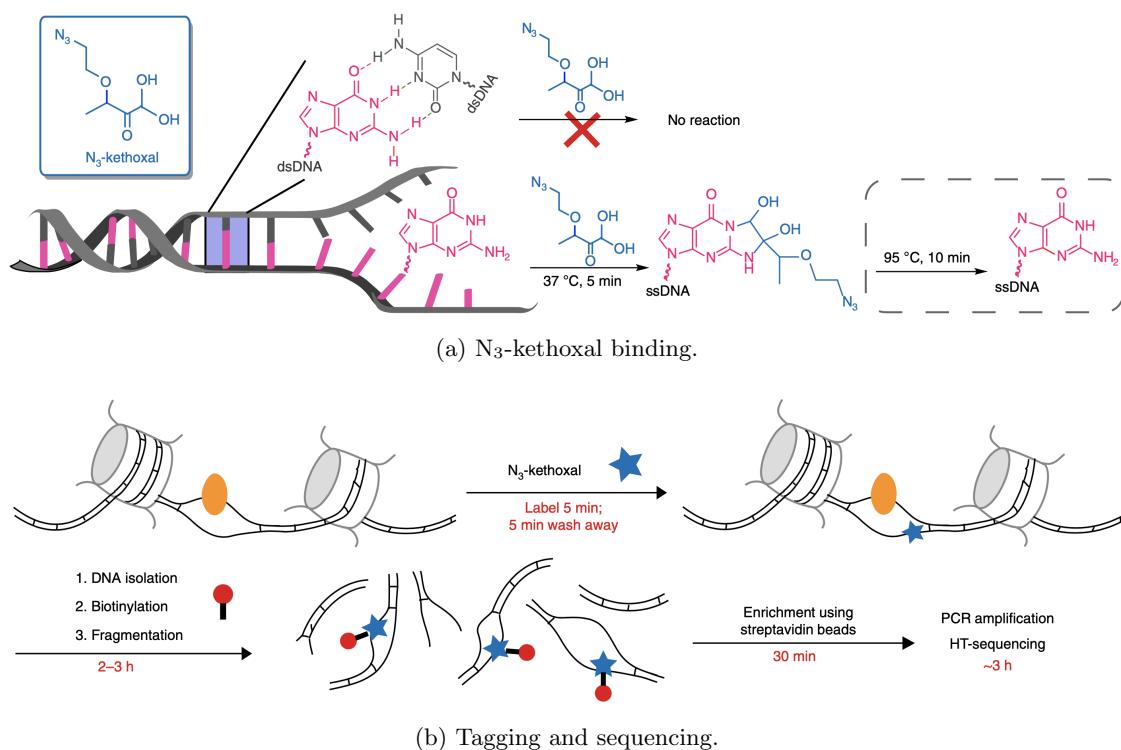


Figure 6.11: Locating ssDNA within the genome.

- This work: Attaching an azide “handle” to kethoxal to make  $\text{N}_3\text{-kethoxal}$ .
- Properties of  $\text{N}_3\text{-kethoxal}$ .
  - Retains high activity and selectivity for guanine.
  - Offers a bio-orthogonal handle that can readily be modified with a biotin or other FG.
- $\text{N}_3\text{-kethoxal}$  effectively maps the secondary structure of RNAs by selectively labeling guanines in ssRNAs under mild conditions in live cells.
- Hypothesis based on this result: The scope can be expanded to selectively labeling ssDNA (Figure 6.11a).
- Verification of  $\text{N}_3\text{-kethoxal}$ ’s high labeling reactivity.
  - Run an *in vitro* labeling assay using a synthetic DNA oligonucleotide with exactly four deoxyguanosine bases.
  - $37^\circ\text{C}$  and 5 mins incubation labels all deoxyguanosine bases.
- Optimization of the KAS-seq  $\text{N}_3\text{-kethoxal}$  introduction conditions.
  - Reaction occurs with deoxyguanosine within 2 mins.
  - Reaction occurs with L-arginine within 10 mins.
  - Thus, 5 mins is a good time to both tag deoxyguanosine and minimize protein labeling.
- After labeling, **genomic DNA** is isolated and biotinylated through click chemistry.
- Enrich the fragments with **streptavidin** beads.
- Subject them to **library construction**.
- Remove  $\text{N}_3\text{-kethoxal}$  labels with a short heating at  $95^\circ\text{C}$ .
- Perform a PCR amplification.
- The whole process takes about 1 day.

- **Genomic DNA:** Regular DNA inside the nucleus. *Also known as gDNA.*
- **Streptavidin:** A protein with an extraordinarily strong binding affinity for biotin.
- **DNA library:** A collection of DNA fragments that have been cloned into vectors.
- **Library construction:** The act of storing and/or propagating a DNA library in a population of micro-organisms through the process of molecular cloning.
- Control experiments (run on one million live HEK293T cells<sup>[1]</sup> and mouse embryonic stem cells [mESCs]).
  - Does N<sub>3</sub>-kethoxal labeling affect gDNA isolation yield and purity? No.
  - Do we still observe biotin signals in the absence of either N<sub>3</sub>-kethoxal or the biotinylation reagent (biotin-DBCO)? No.
- KAS-seq results are highly reproducible in replicate experiments.
- KAS-seq signals mark active transcription.
  - KAS-seq signals exhibit a similar distribution pattern to Pol II ChIP-seq signals along regions with different G/C contents. Indicates G-specific labeling isn't a major factor.
  - KAS-seq reads are very common at gene-coding regions, especially at gene promoters and transcription termination areas.
    - On the other hand, they are far less common at intergenic regions.
    - KAS-seq signals positively correlate with known histone modifications denoting active transcription, and negatively correlate with inactive chromatin markers.
    - KAS-seq also shows improvement over the permanganate method, particularly in the area of weak and broad ssDNA signals.
- **Transition start site.** *Also known as TSS.*
- **Transition end site.** *Also known as TES.*
- KAS-seq works with very small numbers of cells.
  - “Because of the high guanine labeling reactivity of N<sub>3</sub>-kethoxal and the high affinity between biotin and streptavidin, KAS-seq is expected to maintain its sensitivity when using low-input starting materials or primary tissue samples” (Wu et al., 2020, p. 516).
  - KAS-seq signals remain unchanged when using 10,000, 5,000, or even 1,000 HEK293T cells.
  - KAS-seq retains its strong TSS signals but loses some of its gene body and TES signals when mouse liver tissue is used.
  - Low-input of cells still yields similar enrichment efficiency.
- KAS-seq reveals the dynamics of transcriptionally engaged Pol II.
- Proof that what we're seeing is related to transcriptionally engaged Pol II.
  - KAS-seq results correlate well with results from **GRO-seq** and Pol II ChIP-seq.
  - Experiments with inhibitors (DRB and triptolide) confirm that “the strong and sharp KAS-seq peaks on gene promoters reflect transcription initiation and pausing of Pol II near the TSS, and that KAS-seq signals at gene bodies are derived from transcription elongation” (Wu et al., 2020, p. 517).
  - Treatment of the cells with DRB before performing KAS-seq decreased peak numbers by 57% overall, primarily in the gene body and termination regions (signals went up at the TSS).

<sup>[1]</sup>A derivative of a common strain of immortalized human kidney cells.

- This is the expected result, since DRB is known to inhibit Pol II release and keep it stuck at the TSS.
- Treatment of the cells with triptolide before performing KAS-seq decreased peak numbers by 93%.
  - This is the expected result, since triptolide is known to inhibit Pol II being recruited to and loaded onto promoter regions.
- **GRO-seq:** Global run-on sequencing, which is the most widely used method to measure nascent RNA.
- What the dynamics of Pol II are.
  - KAS-seq data from the promoter-proximal and gene body regions revealed that there are four classes of genes: Those for which Pol II pauses in the promoter or doesn't pause and those for which the gene is actively transcribed or isn't.
    - This is consistent with previously reported GRO-seq studies.
  - KAS-seq data shows considerably enriched signals at the TES.
    - DRB removes these, so they are from Pol II elongation and pausing at the end, not some attaching-in-a-different-place artifact.
    - KAS-seq reads density on the terminal regions are all about the same, so KAS-seq doesn't exhibit length-dependent bias.
    - KAS-seq gives a higher **termination index** than Pol II ChIP-seq and GRO-seq, suggesting Pol II accumulation at the TES is greater than previously expected.
- **Termination index:** The ratio of the reads density at the TES downstream regions relative to the density in the promoter-proximal regions.
- **RNA polymerase I:** The RNA polymerase that transcribes the 5.8S, 18S, and 28S rRNAs. *Also known as Pol I.*
- **RNA polymerase III:** The RNA polymerase that transcribes the 5S rRNAs, tRNAs, and some small RNAs. *Also known as Pol III.*
- KAS-seq detects Pol I- and Pol III-mediated transcription events and non-B form ssDNA structures in the same assay.
  - Pol I- and Pol III-mediated transcription events are detected with Pol II ones as expected.
  - These two do not respond to DRB or triptolide.
  - Only about 2/3 tRNAs are actively transcribed, hinting at a transcription-level regulation of codon usage.
  - Several KAS-seq peaks could not be paired to Pol I- or Pol III-mediated transcription events under DRB and triptolide conditions.
    - Hypothesis: These could be from other DNA forms and telomeric DNA.
    - Test: Used a previously reported method to predict where non-B form DNA species might exist in the genome and looked for overlaps with their mystery regions; found many.
    - Takeaway: Using KAS-seq to study other ssDNA-involved biological processes could be a cool avenue to pursue in future research.
- Many enhancer regions are single-stranded, which correlates with higher enhancer activity.
  - Since Pol II is known to bind at certain enhancers, we can use KAS-seq to identify enhancers that are being transcribed by Pol II.
  - Identified **ssDNA-containing enhancers** under DRB conditions to focus on the TSS region.
  - In mESCs, 25% of enhancers are SSEs.
  - Two SSE subtypes: KAS-seq signals span the whole enhancer, and the signals don't.

- SSEs include 94% of super-enhancers.
  - Genes associated with SSEs show higher expression levels.
  - SSEs possess much more long-range interactions, indicating that these transcribing enhancers may possess a stronger capability to activate their target genes.
  - SSEs enrich unique sequence motifs (??). Thus, they have distinct sequence features and transcription factor (TF) binding potentials.
  - Comparison of SSEs and enhancers with high TF binding.
    - ATAC-seq-positive enhancers are readily accessible.
    - 50% of these show no or very weak KAS-seq signals in mESCs.
    - Genes associated with the KAS-seq-positive group show a higher expression level.
    - There is a distinction between SSEs and motifs that are ATAC-seq-positive but KAS-seq-negative.
  - Pol II, histone modifications, and other transcription regulatory proteins are enriched on the SSEs.
  - In HEK293T cells, the ratio of SSEs to general enhancers is lower, but all characteristics (overlap with super-enhancers, DRB response, and correlation with transcription regulatory proteins) are preserved.
  - SSEs possess distinct genomic features and unique TF-binding footprints, as per our KAS-seq analysis.
- **ssDNA-containing enhancer. Also known as SSE.**
  - **Protein condensate:** A highly dynamic structure formed through interactions between mediators, TFs, and other transcription coactivators that have been shown to incorporate Pol II to activate transcription.
  - ssDNA dynamics upon the inhibition of protein condensates.
    - 1,6-hexanediol dissociates protein condensates.
    - The longer we let HEK293T cells sit in it, the more the KAS-seq signals diminish, supporting a role of protein condensate formation on transcription activation.
    - Novel observation: After 5 mins, there is an increase in ssDNA clustered around the TSS.
      - Leads to a slightly increased signal on the gene body and a coinciding decrease in signal at the TSS.
      - The clusters form in both directions for bidirectionally transcribed genes and downstream, only, for unidirectionally transcribed genes.
      - As time goes by, the clusters moved toward the TESs and gradually diminished.
    - Findings validated by Pol II ChIP-seq. KAS-seq even outdoes it in some places (e.g., detection of the above **fast-responsive genes**).
  - **Fast-responsive gene:** A gene with significant ssDNA cluster formation in the TSS region at 5 minutes.

## Q & A

1. The reason that KAS-Seq works on just 1000 cells as opposed to competing methods (e.g., ChIP-seq) that need millions of cells is:
  - Kethoxal is highly reactive and specific to guanines.
  - “Because of the high guanine labeling reactivity of N<sub>3</sub>-kethoxal and the high affinity between biotin and streptavidin, KAS-seq is expected to maintain its sensitivity when using low-input starting materials or primary tissue samples” (Wu et al., 2020, p. 516).

2. How were the authors able to assign opened DNA structures to transcription and not replication?
  - Experiments with inhibitors (DRB and triptolide) confirm that “the strong and sharp KAS-seq peaks on gene promoters reflect transcription initiation and pausing of Pol II near the TSS, and that KAS-seq signals at gene bodies are derived from transcription elongation” (Wu et al., 2020, p. 517).
3. Why do the authors incubate cells with kethoxal-N<sub>3</sub> for such a short time (5 minutes) when incubation for a longer time will capture more ssDNA while the polymerase is transcribing?
  - Kethoxal-N<sub>3</sub>’s high binding affinity for guanines means that it will almost immediately attach to the target species, i.e., not more than 2 minutes is really needed. Additionally, given enough time (circa 10 minutes), it will begin to attach to other species, such as L-arginine (which also has two adjacent nitrogens). This leads to undesired tagging of proteins.
4. Which other purposes can kethoxal-N<sub>3</sub> be used for?
  - “Provides an effective way to map RNA secondary structures by labeling guanines in single-stranded RNAs under mild conditions in live cells” (Wu et al., 2020, p. 515).
5. How did the authors show that the background contribution upon subjecting cells to KAS-seq was negligible?
  - “KAS-seq performed in the absence of N<sub>3</sub>-kethoxal or the biotinylation reagent (biotin-DBCO) resulted in negligible biotin signals shown by dot blot, nor sufficiently enriched DNA for library construction, suggesting minimum background of KAS-seq” (Wu et al., 2020, p. 516).
6. Why do the authors use excess kethoxal with a short reaction time rather than a small amount of kethoxal incubated for a long time?
  - See 3.
7. Glyoxal and methyl glyoxal are known cellular metabolites. Their accumulation is known to be disease causing. Can you explain how a disease might be caused?
  - Inhibiting protein condensation leads Pol II to rapidly release from a group of promoters (as if the cell fears using up all of its energy when there might not be as much around).
  - Likewise, perhaps it is possible that when there is too much protein, Pol II becomes hyperactive, consuming too much cellular energy, leading to oxidative stress and perhaps cell death.
  - Lead to more reactive oxygen species, hence more oxidative stress.
8. What is the major advantage of a 5-minute kethoxal exposure, i.e., giving the cells a “pulse” of excess kethoxal that is then washed away?
  - See 3.
9. KAS-Seq scores over ChIP-Seq in terms of its ability to work with frozen tissue samples. Why? Bear in mind that in frozen samples, the transcriptional bubbles remain.
  - “Because of the high guanine labeling reactivity of N<sub>3</sub>-kethoxal and the high affinity between biotin and streptavidin, KAS-seq is expected to maintain its sensitivity when using low-input starting materials or *primary tissue samples*” (Wu et al., 2020, p. 516).
  - Freezing denatures proteins, but does not alter transcriptional bubbles.
10. Which one(s) of the following descriptions is/are correct when we compare KAS-seq and ATAC-seq?
  - “Notably, KAS-seq signals correlate better with H3K36me3 than ATAC-seq results do, indicating that while ATAC-seq serves as a powerful tool to probe chromatin accessibility, KAS-seq directly measures transcription activities” (Wu et al., 2020, p. 516).

- ATAC-seq is a tool to probe chromatin accessibility more broadly.
- Out of all ATAC-seq-positive enhancers, 50% showed no (or very weak) KAS-seq signals. Thus, since KAS-seq is highly specific for ssDNA, this must mean that 50% of ATAC-seq-positive enhancers are composed of dsDNA. Indeed, KAS-seq is more selective for ssDNA than ATAC-seq.

## Proposed Answers

- 1-2, 2-1, 3-1, 4-(1), 4, 5-4, 6-3, 7-3, (4), 8-5, 9-3, 10-4.

### 6.3 Co-Translational Protein Transport

11/3:

- Today: Co-translational protein transport.
- The lumen of the ER is 10% of the volume of the cell.
- Any protein that needs to be secreted from the cell or is a membrane protein (which is about 30%) is born in the ER.
  - We have various ways to move proteins to where they need to go.
- The ER is the most structurally and functionally diverse organelle in the cell.
  - It has big flat sheets and tubular regions.
  - Very dynamic — sheetlike regions can become tubular and vice versa.
  - You can look at the ER via either electron microscopy or, now, GFP.
  - There are the rough and smooth ER. The rough one is so named because it has ribosomes on its surface.
  - Rough ER is where proteins are actively translated. Smooth is where vesicles full of proteins bud off and go to the golgi.
  - The cytosol is a huge bank of calcium. It contains 10,000 times higher concentrations than the extracellular matrix.
    - This signals to bacteria and viri that they have entered a cell.
    - However, the ER's level of calcium is comparable to the extracellular matrix.
    - A cell cannot tolerate so much calcium in the cytosol for a long time because **excitotoxicity** will take hold.
    - Thus, any release of calcium into the ER must be very coordinated.
- **Excitotoxicity:** Having a cell get too excited and expend so much energy that it dies.
- Isolating ER membranes.
  - The ER is very fragile — as soon as you remove the plasma membrane (e.g., by sonication), the ER breaks up into **microsomes**.
  - When you do organelle isolation, one of the smallest/last things you centrifuge out are the microsomes (see Figure 6.2).
- **Microsome:** A small fragment of the ER membrane, possibly containing ribosomes.
  - There are two types of microsomes: smooth and rough, depending on which ER they came from.
  - Very important to understanding how proteins are transported.
- There are two kinds of proteins present in the ER: soluble proteins (present in the lumen) and transmembrane proteins (not fully water soluble).

- Targeting proteins into the ER lumen.
  - Primary experimental verification of such transport.
    - Take an mRNA that will be secreted by the cell.
    - Do an *in vitro* translation in the absence of microsomes.
    - This protein is slightly larger than the one created in the presence of microsomes.
    - Thus, there must have been a small bit of the protein that got chopped off in the microsome.
    - This led to many more experiments, culminating in the following plausible model.
  - As an mRNA is being transcribed, its signal sequence gets associated with a translocator. The signal peptidase either chops off the signal sequence and then the protein is pushed through, or the protein is pushed through and then it loses its signal sequence.
  - Even as a translocator opens to receive a growing protein, it is water- and calcium-tight.
- What brings ribosomes to the ER when translation begins?

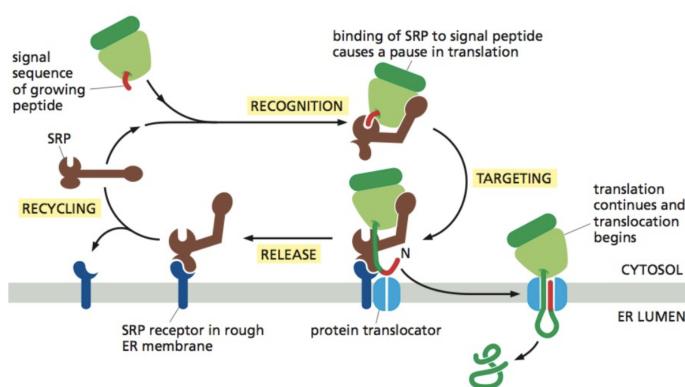


Figure 6.12: Signal recognition particle mechanism.

- A signal recognition particle (SRP) is an RNA protein complex that binds the emergent signal sequence.
- This induces a conformational change in the SRP (it hinges in the middle) that stops the ribosome from translating until the signal recognition particle takes the ribosome to the ER membrane.
- How one signal recognition particle recognizes all signal sequences: It has a hydrophobic pocket that the AA signal sequence fits in.
  - Think of the pocket like a ball of putty — no matter which pencil you stick in, you will lift up the putty.
  - However, the hydrophobic pocket is not too tight; thus, it can let go. In fact, it is very flexible and molds around the signal sequence.
- Recall from the translation lecture that at some point, the ribosome has to hydrolyze GTP to allow us to get to the next codon. This is what the SRP inhibits.
- Upon binding to a signal sequence, the SRP unmasks another binding site which can bind to SRP receptors on the ER membrane.
- This binding induces yet another conformational change that attracts a nearby protein translocator.
- When the translocator binds, a conformational change allows the signal peptide to get jammed into the translocator, causing the ribosome to fall off of the SRP and resuming translation.
- Now free of the ribosome, the SRP disengages from the SRP receptor and goes and looks for another free translating ribosome in the cytosol.

- Translocation on microsomes.
  - The translocator is made of two parts and opens only on one side.
  - You have a water-filled channel through which the polypeptide passes.
  - The translocator is called the Sec61 complex (Sec for secretion).
  - The channel is usually blocked by hydrophobicity (to be water- and calcium-tight).
  - When the peptide passes through and the signal peptidase cuts the molecule, the seam opens and pushes out the protein.
- Anchoring lumenally translated proteins to the membrane.
  - Three ways in which proteins are inserted into the ER membrane, depending on where in the protein the stop-transfer sequence is.
  - We can control the orientation of proteins in the membrane because we understand how it is put there so well.
  - Consider a protein that has a single transmembrane domain.
    - Such a protein must have (1) a signal sequence (2) at the N-terminus.
    - The start-transfer signal begins taking the protein into the ER, and then when the translocator hits a stop transfer, it stops and ejects the protein into the membrane.
    - There are well-known start and stop sequences. A stop sequence is typically also a hydrophobic transmembrane domain.
    - The start domain breaks off, and then the rest of the protein is synthesized in the cytosol.
    - The region between the start-transfer sequence and the stop-transfer sequence exists within the ER, and the region past the stop-transfer sequence exists outside the ER.
  - Suppose the start sequence is located in the middle of the protein AA sequence.
    - Two possible orientations: N-terminus in the ER and C-terminus in the cytoplasm, and vice versa.
    - Start means go to the ER and start translating; that's it.
    - How it gets oriented depends on which side is more positive and which side is more negative. More positive residues face the predominantly negative cytosol.
      - This allows us to control which side of our protein gets localized where.
      - Possible test question: How do I position a GFP in the cytosol but embedded on the ER membrane?
  - Multi-pass transmembrane proteins.
    - Start transfer is pulled in, goes along until you reach the stop codon, ejected.
    - Then the ribosome translates into the cytosol until the next start gets pulled into a translocator.
    - Signal peptidase does not chop at the end of every stop sign here; why is not known, but there must be some kind of “final” stop sign.
- Post-translational protein translocation: Signal sequence is on the C-terminus.
  - No SRP needed here.
  - Instead, we have a Get pathway.
  - **Snare proteins** (Nobel Prize 2013), which are very important in vesicle fusion and fission, decide how organelles come together.
  - Before the Get pathway was discovered, people hypothesized that proteins ram themselves into the ER membrane, but there were many holes in this theory (e.g., why not ram into the plasma membrane?).
  - The Get pathway is formed by 3 proteins: Get1, Get2, and Get3.

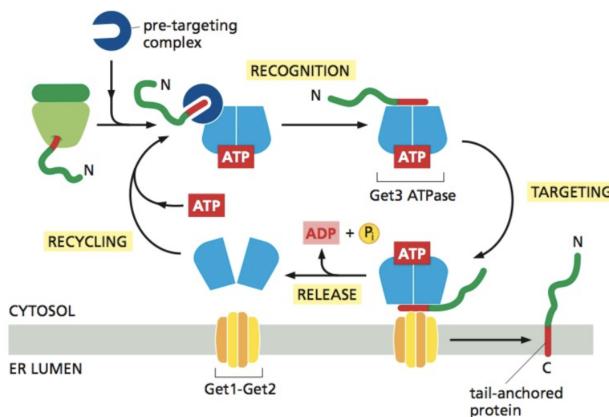


Figure 6.13: Post-translational protein translocation: Get pathway.

- Get1, Get2 only recognize Get3.
- Get3 is the key: It gets the C-terminus signal peptide, but only after it has been bound to a pre-targeting complex, inducing a conformational change that allows the pre-targeting complex to bind to specifically the ATP-bound form of Get3.
- Get3 is an ATPase, meaning that it hydrolyzes ATP.
- The pre-targeting complex binds the signal peptide sequence to Get3 ATPase; the mechanism is not well understood.
- Once this has happened, Get3 is competent to deliver the sequence to Get1-Get2, which is located in the ER membrane and functions as a translocator.
- In order to shove the sequence into Get1-Get2, we need energy; this energy comes from the hydrolysis of ATP to ADP + Pi.
- At this point, Get3 is ready to bind another ATP and restart the cycle.
- Contrast this with the GTPase cycle. Think about how an ATPase differs from a GTPase.
- Glycosylation (overview).
  - Many proteins need to be glycosylated (this will be the subject of next class).
  - Additionally, many proteins are anchored to the ER membrane not by a transmembrane region but by a lipid. How is this related??
  - Start with a complex sugar (we don't need to know the details), which is stuck onto a lipid (in particular, a steroid) called **dolichol** (which is a cholesterol) with phosphates.
  - This sugar gets transferred to any protein which gets glycosylated.
  - The transfer is carried out by oligosaccharide transferase.
  - The sugar ends up on an Asn side chain.
  - A specific signal leads to glycosylation at a specific Asn; in particular, you need a serine or threonine, then an arbitrary amino acid, then your asparagine.
  - The sugar is transferred *en bloc* (from French: all at once).
- GPI anchors.
  - These are lipoproteins.
  - Glycosylphosphatidylinositol: A lipid, phosphate, and many inositol groups (sugars).
  - A transamidation occurs, breaking the peptide bond connecting the transmembrane region of a protein to the rest of the protein and connecting the rest of the protein to the GPI anchor.

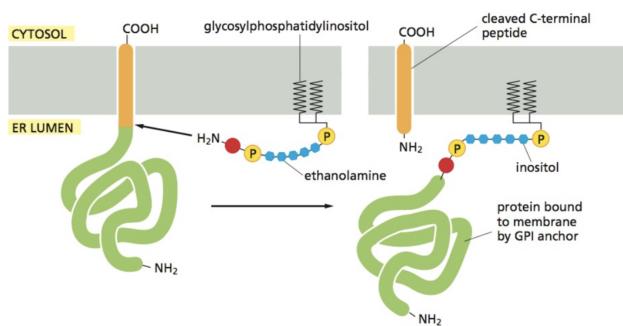


Figure 6.14: GPI anchoring.

- Thus, it is the sequence at the border of the transmembrane region (a GPI-anchoring signal) and the rest of the protein that decides whether or not a protein will have a GPI anchor.
- Once a protein becomes GPI-anchored, it gets moved to the extracellular surface of the cytosol.
- Differences between ATPases and GTPases.
  - GTPases: You need GAPs and GEFs.
  - ATPases: The rate limiting step (this is very important) is the dissociation of the NDP (usually ADP).