Problem Set 2 – Mechanistic Biology – Fall 2026

Do not use AI in preparing your answers (e.g., ChatGPT or any similar LLM).

1. **Submitted by Dr. Christina Gladkova:**

Using AlphaFold 3 for Discovery

Post-lecture exercise: triaging hits from a KEAP1 pull-down

Scenario

NRF2 is a transcription factor that upregulates antioxidant and stress response pathways, and is upregulated by cancer cells to safeguard their growth. KEAP1 promotes NRF2 degradation through recruitment of the CUL3 E3 ubiquitin ligase and other factors, and is therefore mutated in many cancers.

You want to understand which proteins KEAP1 interacts with, and whether those interactions are compatible with, or mutually exclusive with NRF2 binding. You performed an immunoprecipitation with KEAP1 followed by mass spectrometry and found the candidates below.

Your task: Use AlphaFold3 to decide which candidates are likely to interact directly with the Kelch domain of KEAP1, which interact with KEAP1 somewhere else, and which are likely contaminants. Then design an experiment to test one prediction.

Sequences

Protein	UniProt	Notes
KEAP1 (bait)	Q14145	Domains: BTB (N-terminal), IVR (middle), Kelch (C-terminal, roughly residues 321–609). Look these up on the UniProt page.
NRF2 (gene NFE2L2)	Q16236	Candidate
CUL3	Q13618	Candidate
HSPA8	P11142	Candidate
p62 (gene SQSTM1)	Q13501	Candidate

Problem Set 2 – Mechanistic Biology – Fall 2026

Part 1: Fold each candidate with KEAP1

1. For each candidate, run one job containing full-length KEAP1 (copies = 1) and the candidate (copies = 1).
2. For each result, record ipTM and look at the PAE plot. Find the off-diagonal blocks between KEAP1 and the candidate.
3. Work out where on KEAP1 the candidate binds. Use the residue numbers along the PAE plot axes, or open the model in Mol*, PyMOL, or ChimeraX and see which KEAP1 domain the candidate touches.
4. If the candidate binds KEAP1, work out which residue range of the candidate mediates the interaction. This may be short (10 – 50 residues) if it binds as a peptide (without secondary structure), or relatively long (50 – 400 residues) if it binds as a folded domain
5. Fill in the table below.

Candidate	ipTM	If it binds, which candidate residue range? (start – end)	Where on KEAP1? (BTB / IVR / Kelch / unclear)	How does the inter-chain PAE look?	Direct Kelch binder? (yes / no / unsure)
NRF2					
CUL3					
HSPA8					
p62					

Part 2: Fold just the binding region of each candidate with just its interacting KEAP1 domain

Long disordered regions can lower ipTM even when a real interface is predicted, because ipTM considers the whole chain.

1. Using the residue ranges and KEAP1 domain you identified in the table above, fold just the KEAP1-interacting peptide of p62 with its interacting KEAP1 domain. *Optional: do this for the other candidates too*
2. Record ipTM and look at the PAE plot. Find the off-diagonal blocks between KEAP1 and the candidate.
3. Fill in the table below.

Problem Set 2 – Mechanistic Biology – Fall 2026

Candidate	Folded?	ipTM	ipTM change from before	How does the inter-chain PAE compare with before?
P62 truncated				

Part 3: p62 and phosphorylation

In cells, p62 binds the Kelch domain much more strongly when serine 349 is phosphorylated.

1. Repeat Part 2 with Ser349 on p62 phosphorylated. (On AlphaFold Server, after entering the protein sequence, click the 3 dots on the right, click 'PTMs' then click the serine in the peptide sequence and add the phosphoserine modification. Check that you picked the residue corresponding to Ser349 of full-length p62.)
2. Compare the two predictions: ipTM, PAE, and the pLDDT of the peptide residues at the interface.

Questions

1. Which candidate(s) do you think bind the Kelch domain directly? How confident are you, and which pieces of evidence (ipTM, PAE, pLDDT, location) were most convincing?
2. CUL3 is a genuine KEAP1 partner. What did AlphaFold3 tell you about how it binds, and why does that matter for interpreting a pull-down?
3. HSPA8 is a very common contaminant in pull-downs. What did its prediction look like, and how would you tell this kind of result apart from a real interaction?
4. How did ipTM scores change when disordered regions were removed from p62?
5. Did phosphorylation change the prediction/scores? What does that tell you about whether AlphaFold3 can say under which conditions binding happens?
6. Design an experiment. Using your NRF2–KEAP1 model, pick one or two residues in NRF2 at the interface that you would mutate to disrupt binding, and explain your choice. What experiment would you do, and what result would support the prediction? Bonus: look up NFE2L2 mutations in lung squamous cell carcinoma on cBioPortal (cbioportal.org). Do any patient mutations fall in the region you chose?

Problem Set 2 – Mechanistic Biology – Fall 2026

2. From Dr. Christopher Lima:

1) The 5' cap and polyA tail protect messenger RNA from degradation via decapping or deadenylation in the cytoplasm. What chemical features of the 5' cap prevents decay? What additional protein factors are involved in preventing decapping enzymes from accessing the cap in the nucleus and cytoplasm?

2) Name two of the most common post-transcriptional modifications made to nuclear RNAs to promote decay?

As you know from the class discussion paper (Chang HM, Triboulet R, Thornton JE, Gregory RI. A role for the Perlman syndrome exonuclease Dis3l2 in the Lin28-let-7 pathway. Nature 2013 497: 244-248.), Dis3L2 can degrade RNAs.

3) What post-transcriptional modification to microRNA promotes interaction with Dis3L2 and what family of enzymes catalyze this post-transcriptional modification?

4) Why is this modification needed for Dis3L2 to degrade substrates?

Take a look at Ustianenko et al., EMBO J. 2016 Oct 17;35(20):2179-2191 (https://www.ncbi.nlm.nih.gov/pubmed/27647875). The authors claim that Dis3L2 is involved in degradation of many other structured RNAs.

5) What do you consider as the best experimental evidence supporting this claim? Justify your answer (don't just tell me the experiment they performed, tell me why the experiment supports the claim). Based on their data, estimate the fraction of cytoplasmic structured RNAs that are degraded by this pathway?

Play with Claude to see how things vary over time if you consider the following:

Rate of synthesis in X per minute in the nucleus, mRNA export takes 35-65 minutes from transcription through export to translation, rate of decay in X per minute in the cytoplasm

Graph the steady state levels of a protein from 0 to 6 hours with a synthesis rate of 5 per minute and a degradation rate of 1 per minute, then at 6 hours change the degradation rate to 20 per minute and graph to 12 hours.

Graph the steady state levels of a protein from 0 to 12 hours with a synthesis rate of 5 per minute and a degradation rate of 1 per minute, then at 12 hours change the degradation rate to 5 per minute and graph to 24 hours.

Problem Set 2 – Mechanistic Biology – Fall 2026

Continue this graph after 24 hours with degradation rate to 5 per minute for 6 hours, then change the degradation rate to 1 per minute and continue for 6 hours.

Consider two scenarios, one requires upregulation of transcription which requires an average of 50 minutes before the mRNA is exported to the cytoplasm for translation, in the other scenario the mRNA is already in the cytoplasm ready to be translated. Transcription rate of 100 per minute and degradation rate is 50 per minute, plot accumulation of translated products over 2 hours.

After 2 hours upregulate translation for scenario 1 by increasing transcription to 200 per minute, in the second scenario upregulate translation by decreasing decay rates to 25 per minute.

Now plot a scenario that starts with mRNA being transcribed and exported with a synthesis rate of 10 per minute and a decay rate of 20 per minute over 2 hours.

Now include two scenarios, one that increases transcription by 2 fold (including the delay required for mRNA export), the other that decreases the decay rate by 2 fold.

Problem Set 2 – Mechanistic Biology – Fall 2026

3. From Dr. Naama Aviram:

Imagine you challenge two bacterial strains with the same temperate phage.
Strain A carries a type II CRISPR-Cas9 system and strain B carries a type III CRISPRCas10 system. After several days, strain A survivors have acquired spacers and are resistant to the phage. Strain B survivors have also acquired spacers, but many of them are stable lysogens carrying the phage in their chromosome.

Explain this result. What property of type III targeting makes lysogeny compatible with immunity, and why is the same outcome much harder to achieve in a type II system? What does this suggest about what CRISPR immunity is actually optimizing for?

Problem Set 2 – Mechanistic Biology – Fall 2026

4. From Drs. An and Ordureau:

1. Imagine that you are working on a project that requires a Ribo-seq experiment after knocking out your gene of interest. Before collecting the cells, you need to stop the protein translation by adding inhibitors. What would be the proper inhibitor: Cycloheximide or puromycin? Why do you choose this inhibitor? What happens to 80S ribosomes after the Cycloheximide (>100 uM) or Puromycin (1ug/ml) treatment, respectively, in relation to mRNA?

2. Describe the mechanism of action for Cycloheximide in mammalian cells in detail (which site does it bind, and what happens after the binding).

3. What will happen to the actively translating ribosomes if you add a low concentration of Cycloheximide, Emetine, or Anisomycin (~1-2uM)? What will happen to the actively translating ribosomes if you add high concentration of Cycloheximide, Emetine, or Anisomycin (>100uM)?

Problem Set 2 – Mechanistic Biology – Fall 2026

5. Based on Dr. Overholtzer's lecture (submitted by Zaza Gelashvili):

In class we discussed the possibility that cancer cells can store protein within lysosomes to fuel growth under nutrient-limiting conditions. Design an approach to test *in vivo* the hypothesis that cancer cells can "store" and utilize "stored" protein in lysosomes. Discuss your approach in terms of design, potential outcomes, and key limitations. Consider: How will you distinguish and define what is "stored" protein? How will you determine where it is stored, and what state it is in? Will you be able to tell how long protein is stored, and if so, for how long? How will you examine experimentally the effects of its usage? Finally, can you design an orthogonal approach?