SYNTHETIC BIOLOGY- BIOENG-320
MOCK EXAM

Note: This mock exam question paper is set to give you an idea of the type of questions that are most likely to appear in the final exam scheduled on June 27th. A solution set will also be provided to refer to the correct/expected answer. Please reach out to the TAs for any doubts in the concepts or questionnaire.

PROTEIN DESIGN

Question 1. You are running your first Monte Carlo based protein design calculations. You notice that your simulations have a hard time to converge and do not give you low energy design solutions. Which parameter could you tune to improve the convergence and increase the chance that you will identify low energy structures and why?

SOLUTION: To improve convergence and increase the likelihood of finding low-energy structures, you can decrease the annealing temperature. In Monte Carlo simulations, the temperature parameter controls the probability of accepting higher-energy conformations. A lower temperature reduces the acceptance of such unfavorable moves, thereby biasing the search toward lower-energy conformations. This can help the simulation focus more effectively on refining promising solutions and improve convergence toward energetically favorable designs.

Question 2. Protein-protein binding: you would like to design a protein X that can bind to protein A but not to proteins B and C. What kind of computational design strategy could you envision and why? You can draw a schematic figure to illustrate the design strategy if necessary.

SOLUTION: To achieve binding specificity, you can use a multi-state design approach. This strategy involves simultaneously optimizing the sequence of protein X across multiple binding states:

- **Positive design state: X bound to the target protein A — the sequence is optimized to stabilize this interaction.**
- **Negative design states: X bound to off-target proteins B and C — the sequence is optimized to destabilize these interactions.**

By considering both positive and negative binding states during sequence optimization, multi-state design enables the identification of sequences that are selectively compatible with the target while disfavoring non-specific interactions.

Question 3. The chemokine receptor CXCR4 and its ligand CXCL12 play crucial roles in cancer progression, with their dysregulation being linked to tumor growth, metastasis, and therapy resistance. The interaction between CXCR4 and CXCL12 promotes cancer cell migration, invasion, and metastasis, making it a key target in cancer treatment and diagnosis. Notably, cancer cells often exhibit increased sensitivity to shallower gradients of the chemokine, allowing them to respond to low concentrations of CXCL12 present in distant tissues. Enhancing the binding of CXCR4 to CXCL12 can serve as a potent biosensor, enabling the development of molecular systems that specifically recognize and report the presence of cancer cells, even in early stages or when they are sparsely distributed.

In Figure 1, you have a partial vision of the interface between receptor and the chemokine. Here, you are requested to design one mutation (specifically on the CXCR4 side) that aims at stabilizing the binding between CXCR4 and CXCL12. Please explain the logic behind it.

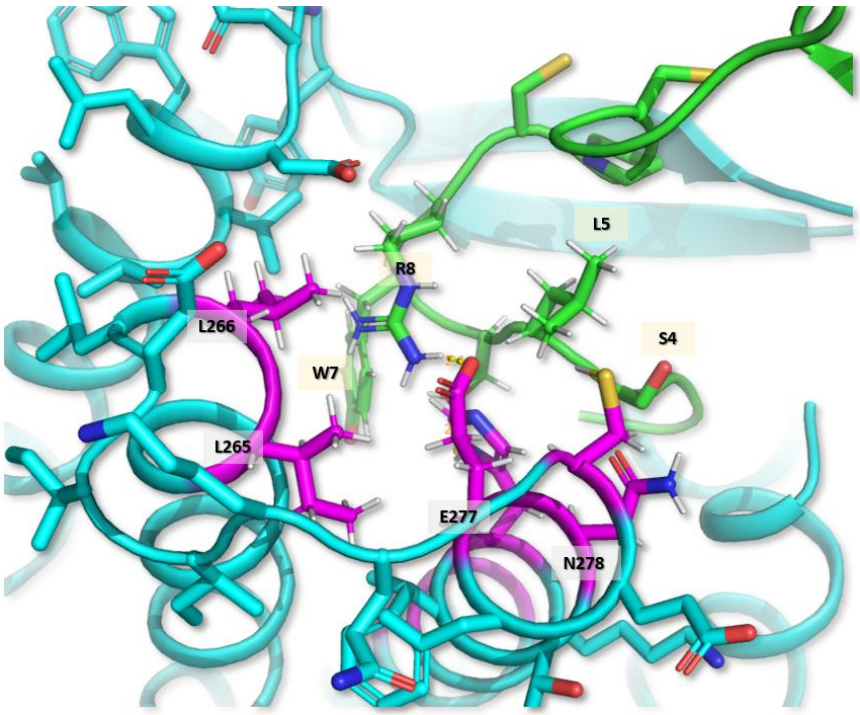


Figure 1. Extracellular loop 3 region of CXCR4 (light blue) and N-terminal region of CXCL12 (green). Residues belonging to CXCR4 that are within 5 A from the chemokine are highlighted in Magenta. Hydrogens are shown in white.

SOLUTION: One strategy to stabilize the CXCR4–CXCL12 interaction is to introduce a polar or charged residue at the receptor interface to form a new favorable interaction with the chemokine. Specifically, mutating leucine 265 (L265) or leucine 266 (L266) on CXCR4 to a polar residue such as glutamine (Q) or asparagine (N), or even better aspartic acid (D) or glutamic acid (E), could enhance binding. These mutations would introduce the potential for hydrogen bonding or electrostatic interactions with arginine 8 (R8) of CXCL12, a positively charged residue near the interface.

Question 4. A different scientific approach identified three hotspots in the interface region between the receptor and the ligand. These hotspots interact with the first 6 residues of the chemokine.

Which mutations achieved the goal of creating a more sensitive receptor to CXCL12 gradients? Do you observe any non-additive effect in sensitivity and potency when all the mutations are combined in the Cdyn design?

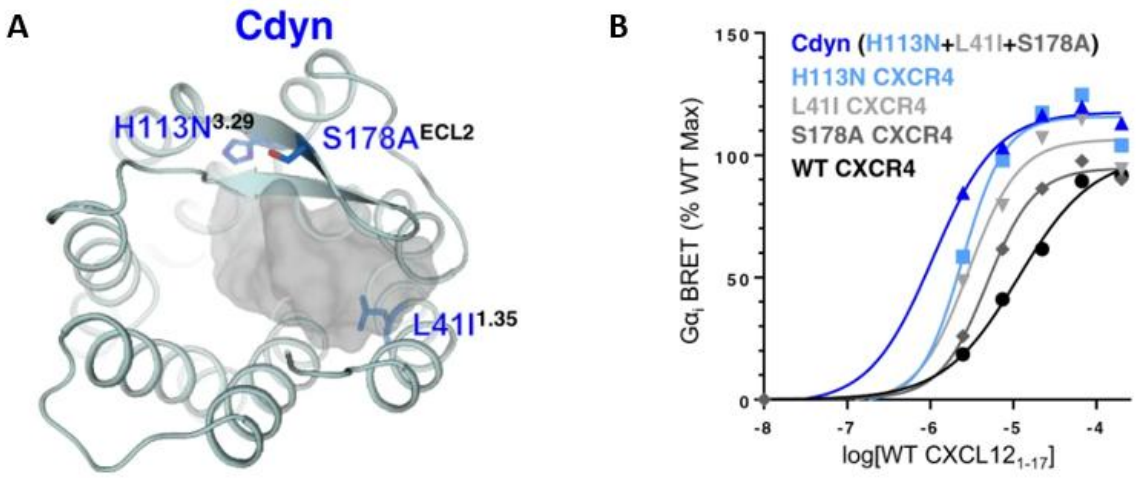


Figure 2. A) Overhead view of the Cdyn design and localization of hotspots. B) Dose response curve of point mutants and Cdyn design with CXCL12

SOLUTION:

All mutations introduced at the three identified interface hotspots led to increased sensitivity of the receptor to CXCL12 when tested individually.

- **Sensitivity refers to the ability of the receptor to respond to lower concentrations of the chemokine. On a dose–response curve, this is typically reflected by a leftward shift, indicating activation at shallower gradients.**
- **Potency is defined as the concentration of ligand required to elicit a half-maximal response (commonly represented by EC_{50}). A lower EC_{50} value indicates higher potency.**

The Cdyn design shows a cumulative effect in both sensitivity and potency. It outperforms each of the individual mutants, indicating that combining the mutations leads to an overall stronger enhancement of receptor function.

Question 5. During the COVID-19 pandemic, protein engineering offered rapid and powerful methods for building therapeutics for SARS-COV-2. One of the possible therapeutics is a binder that traps the receptor binding domain (RBD) of the spike protein on the surface of the virus and thus neutralizing it. To build this binder, you decide to redesign a naturally occurring binding protein, namely angiotensin-converting enzyme II (ACE2), a membrane-bound receptor that's the target of the spike protein.

a) While searching for ideas to start your design, you read in a paper that the researchers performed computational alanine scanning (where every interface residue was mutated in-silico to alanine and then have the interface energy measured), what can such a scan tell you? What do you aim to learn from it?

SOLUTION: Computational alanine scanning identifies hotspot residues — positions at the binding interface that contribute significantly to binding energy. By mutating each residue to alanine and measuring the change in binding energy, you can pinpoint which residues are critical for binding. Alanine is chosen because it removes side-chain interactions (e.g. hydrogen bonds, salt bridges, hydrophobic contacts) while preserving backbone geometry. This helps guide redesign by focusing on key positions to either enhance or preserve during affinity maturation.

b) Shown below is the structure of ACE2 where residues that contribute strongly to binding the spike RBD are highlighted in sphere representation (Fig. 1). Out of the highlighted residues, H34, Q42, and K353 were chosen for computational saturation mutagenesis (Fig. 2). Which of the positions (H34, Q42, or K353) would you choose to mutate if you aim to optimize the binding interface between ACE2 and the RBD? And which of the mutations shown in Fig. 2 would you use as a starting point for design that would optimize the binding interface? Justify your answer.

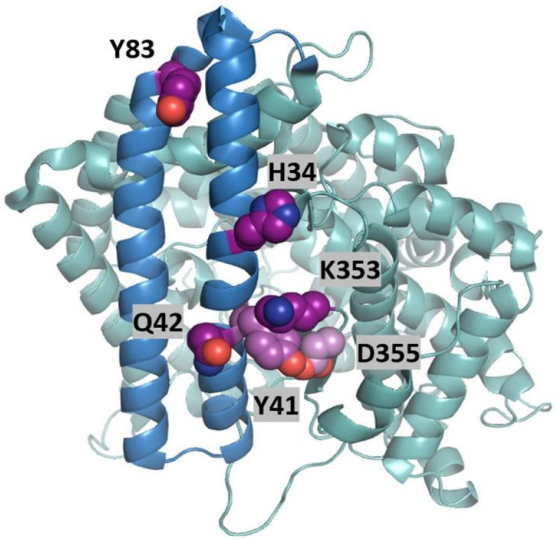
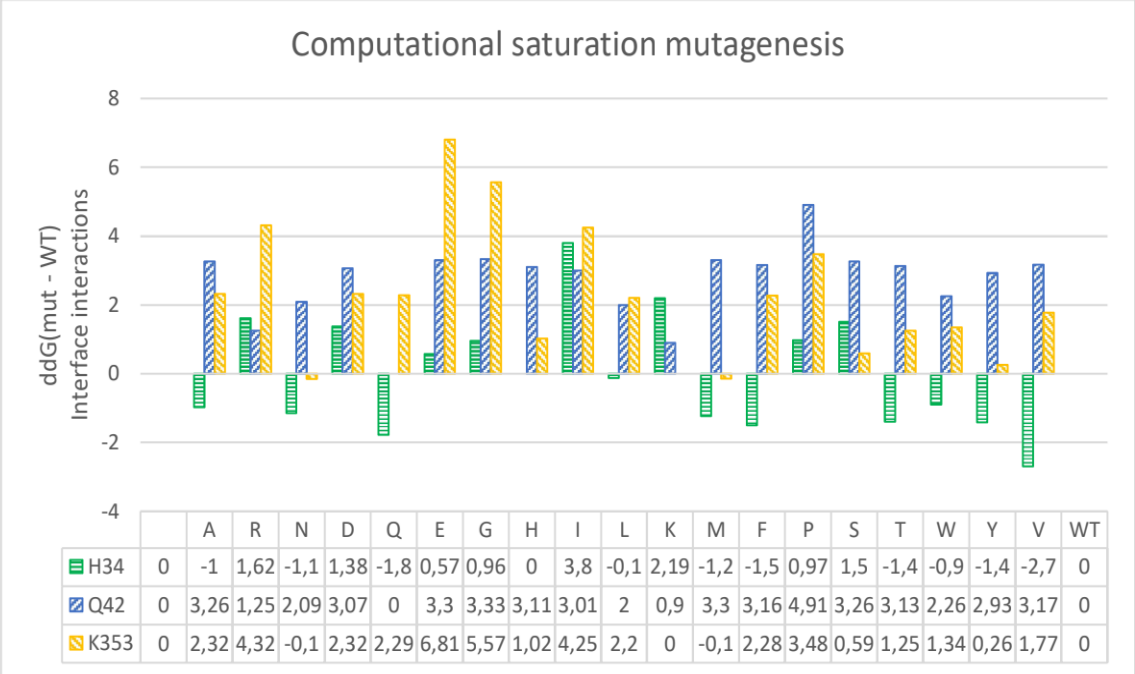


Figure 1: ACE2 interface where interesting residues from the computational alanine scanning are shown as spheres



		A	R	N	D	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	WT
H34	0	-1	1,62	-1,1	1,38	-1,8	0,57	0,96	0	3,8	-0,1	2,19	-1,2	-1,5	0,97	1,5	-1,4	-0,9	-1,4	-2,7	0
Q42	0	3,26	1,25	2,09	3,07	0	3,3	3,33	3,11	3,01	2	0,9	3,3	3,16	4,91	3,26	3,13	2,26	2,93	3,17	0
K353	0	2,32	4,32	-0,1	2,32	2,29	6,81	5,57	1,02	4,25	2,2	0	-0,1	2,28	3,48	0,59	1,25	1,34	0,26	1,77	0

Figure 2: computational saturation mutagenesis of the three residues that were selected from the alanine scan

SOLUTION: H34 is a good candidate for mutation because it is solvent-exposed and its side chain directly contacts the RBD, making it accessible for interaction tuning. Among all substitutions, H34V dramatically reduces ΔΔG, suggesting a significantly improved binding affinity. Therefore, H34V would be a strong starting point for interface optimization. Q42 should be kept as wild type, as none of the tested mutations improve the binding energy compared to the native glutamine. K353 is also solvent-exposed and forms part of the binding interface. While the improvements are more modest, K353N or K353M slightly lower the ΔΔG and could be considered as starting points for further optimization.

c) The interface of ACE2 and the RBD is shown in the figure below (Fig. 3). By looking at the amino acids at the interface:

- Pick two disruptive mutations from the table above (Fig. 2) and explain their destabilizing effect based on the binding interface shown in Fig. 3.

- Suggest one mutation that would enhance binding except mutations on position H34 and explain your answer.

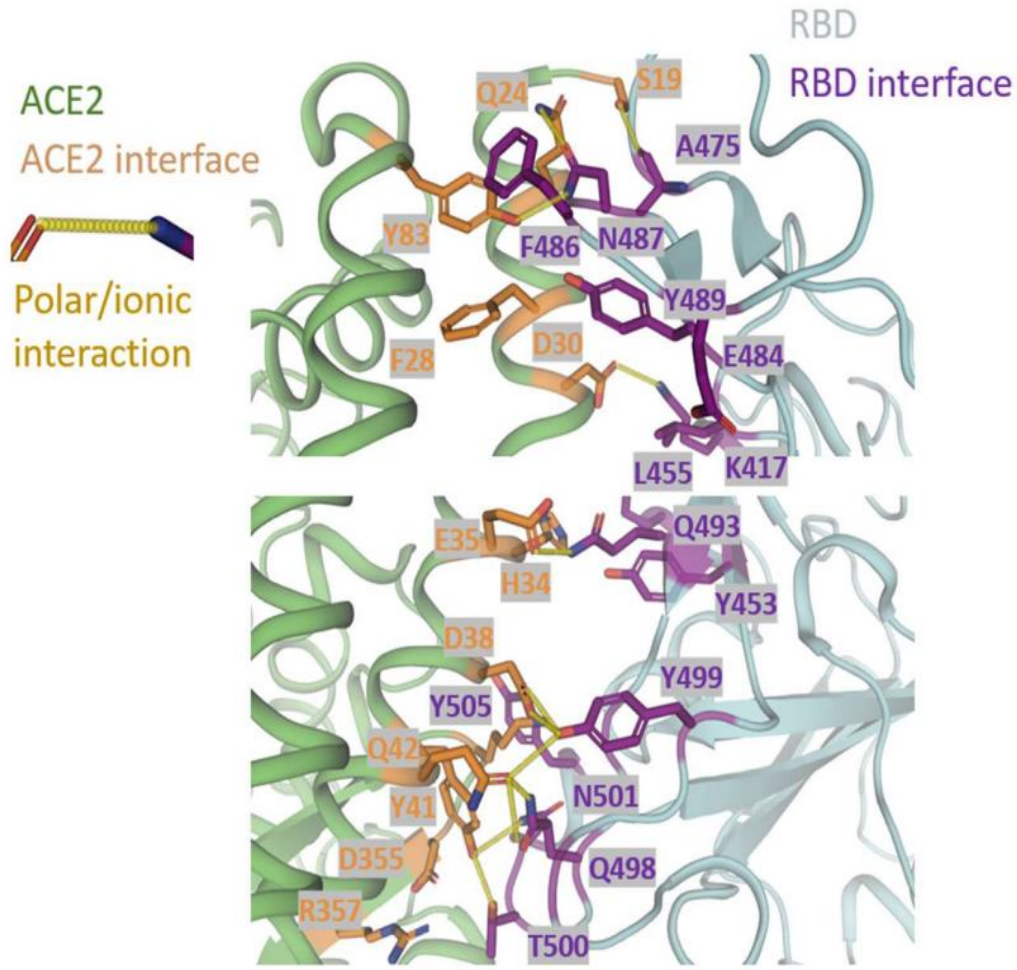


Figure 3

SOLUTION:

Disruptive mutations:

- **H34I: Histidine at position 34 forms polar interactions with Q493. Replacing it with a bulky hydrophobic residue like isoleucine disrupts these polar contacts, weakening the interface.**
- **Q42A: Glutamine at position 42 is involved in hydrogen bonding with Q498. Mutating it to alanine removes the side-chain hydrogen bond donor, destabilizing the interaction.**

Enhancing mutation:

- **F28S: Phenylalanine at position 28 is near Y489 of the RBD. Mutating F28 to serine could introduce a new hydrogen bond with Y489 or nearby polar residues, potentially increasing binding affinity through favorable polar contacts.**

PROTEIN CIRCUITS AND CELL ENGINEERING

Question 1. You are trying to predict the side-chain conformations of a protein starting from the structure of the protein backbone and using a side-chain rotamer sampling method. You are testing the method on a protein for which you already know the exact side-chain positions from an experimental structure (as a benchmark for the method). It turns out that you cannot accurately predict all side-chain conformations using this rotamer library approach. Can you explain possible reasons and propose 2 solutions to improve prediction accuracy?

SOLUTION: 1. Finer grained rotamer library , 2. Minimization over all conformational degrees of freedom after discrete rotamer repacking

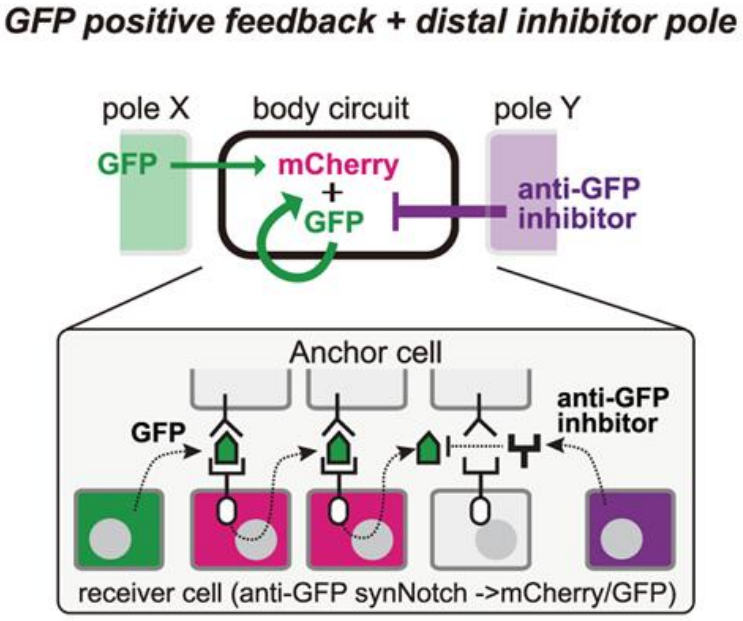
Question 2. You are designing an enzyme by stabilizing a model of the highest energy state of the reaction, i.e. the transition state (TS), using a simple single state design approach. Despite achieving great stabilization of TS, your designed enzyme displays poor catalytic function. What could go wrong (identify 3 potential issues) and what design strategies could you envision to address these challenges?

SOLUTION:

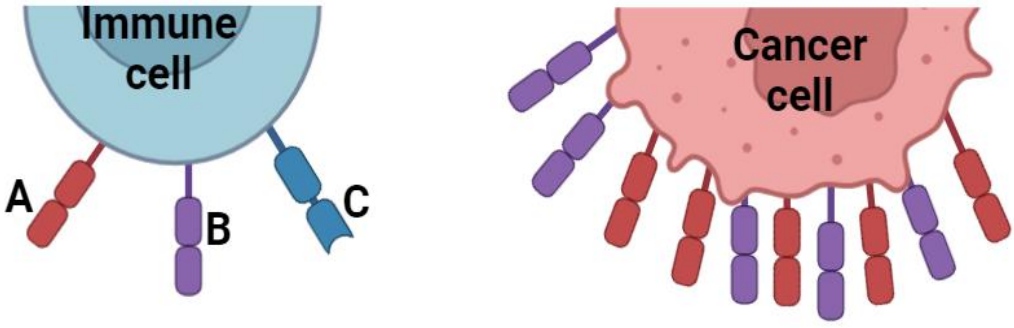
1. **Inaccurate TS model → Fix: Run quantum mechanical (QM) calculations**
2. **Poor substrate binding (high Km) → Fix: Include the enzyme–substrate (E–S) complex as a positive design state**
3. **Product inhibition (E-P is more stable than E-S) → Fix: Include the enzyme–product (E–P) complex as a negative design state**

Question 3. In our last lecture, we have seen how to build synthetic morphogen systems that can program multicellular patterning. For example, in a positive feedback circuit, GFP morphogen activates receiver cells to induce the secretion of more GFP. In a negative feedback circuit, GFP morphogen induces the expression of antimorphogen inhibitor by receiver cells. Can you propose a circuit combining these features to program two-domain patterns such as this? (please draw a circuit):

SOLUTION:

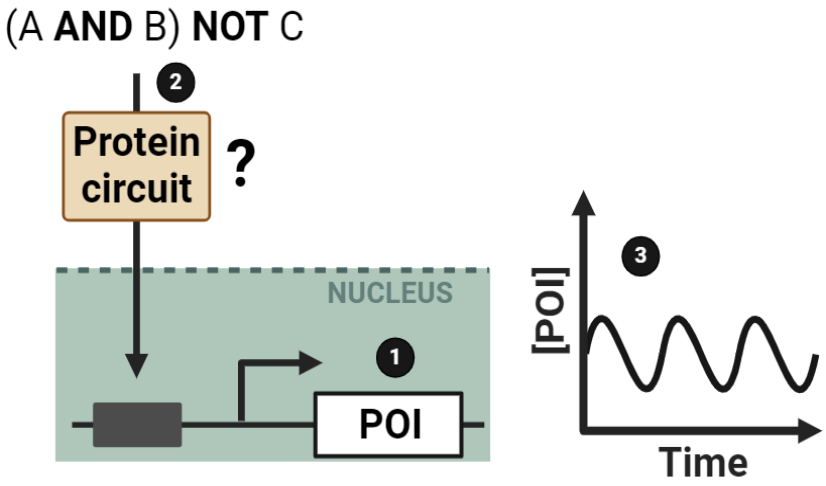


Question 4. The company you are working for is trying to develop a CAR T-cell therapy against a certain type of cancer. They have identified two antigens (A and B) that are highly overexpressed on the surface of the tumor cells. However, A and B are also present (in lower levels) on some immune cells, which could cause undesired off-target effects of CAR T-cell therapy. Luckily, the cancer cell seems to lack protein C, which is expressed on the surface of the immune cells (see image below).



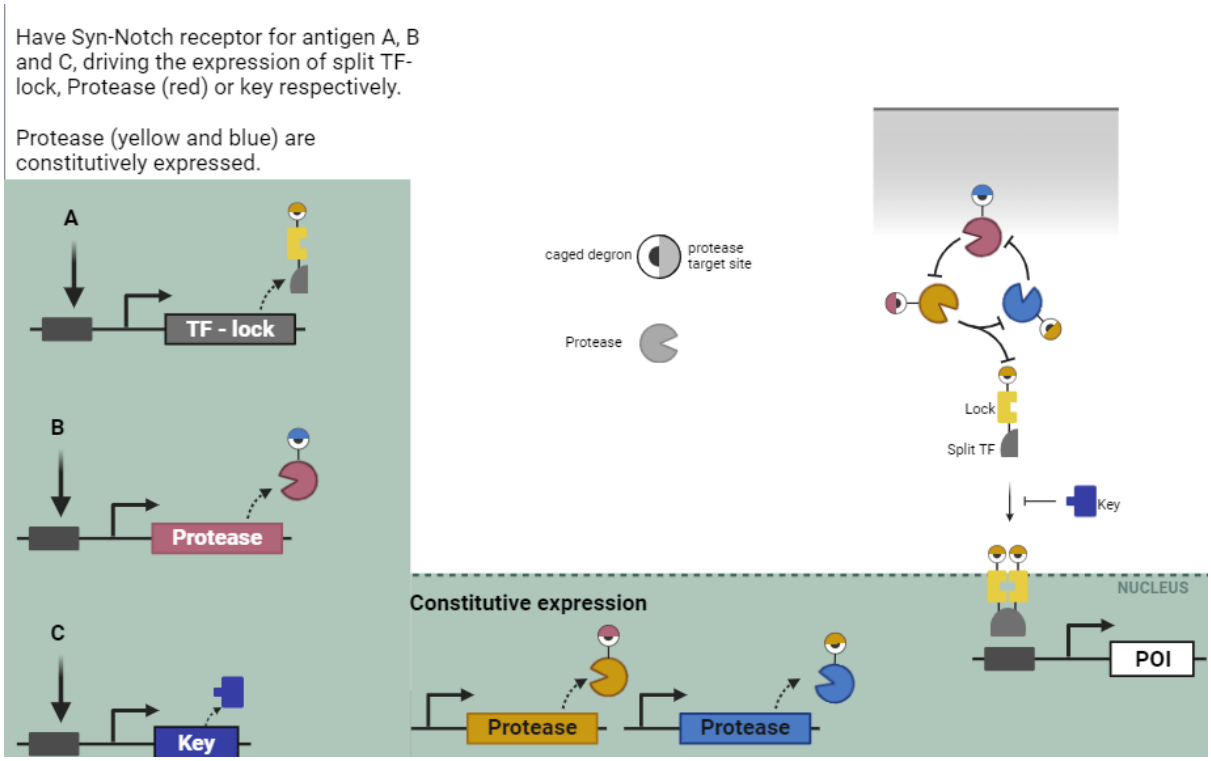
Your company has given you the task to engineer a CAR T-cell with the following properties (see schematic figure below)

1. The CAR T-cell should express a therapeutic protein of interest (POI)
2. The CAR T-cell contains a protein circuit where expression of the POI is controlled by an *(A **AND** B) **NOT** C* logic gate
3. If (and only if) *(A **AND** B) **NOT** C* is fulfilled, expression of the POI should oscillate

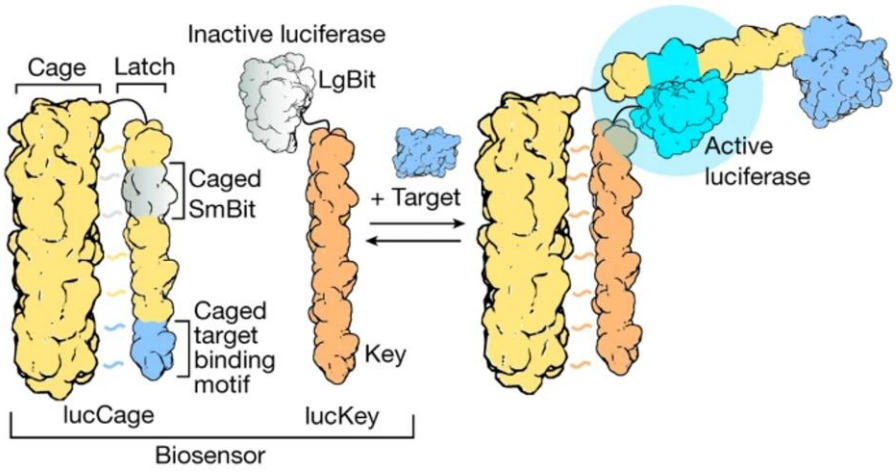


Please provide a schematic figure of your designed circuit where all the components are clearly labeled. Explain how your circuit works and how the three required properties mentioned above are achieved.

POTENTIAL SOLUTION



Question 5. The LOCKR (Latching Orthogonal Cage–Key) system is a modular protein switch that controls activity via conformational changes. How does the system work? The system consists of two protein components: first, a 'lucCage' that comprises a cage domain and a latch domain that contains a target-binding motif and a split luciferase fragment (small BiT); and second, a 'lucKey' that contains a key peptide that binds to the open state of lucCage and the complementary split luciferase fragment (large BiT). Consequently, lucCage has two states: a closed state, in which the cage domain binds to the latch and sterically occludes the binding motif from binding the target and SmBiT from combining with LgBit to reconstitute luciferase activity, and an open state, in which these binding interactions are not blocked and lucKey can bind to the cage domain. The association of lucKey with lucCage results in the reconstitution of luciferase activity. The thermodynamics of the system are tuned such that the binding free energy of lucKey to lucCage (ΔG_{CK}) is insufficient to overcome the free energy cost of lucCage opening (ΔG_{open}) in the absence of target ($\Delta G_{open} - \Delta G_{CK} >> 0$), but in the presence of the target, the additional binding free energy of the latch to the target (ΔG_{LT}) drives latch opening and luciferase reconstitution ($\Delta G_{open} - \Delta G_{CK} - \Delta G_{LT} << 0$). Central to this protein biosensor are these nearly isoenergetic states, the equilibrium between which is modulated by the analyte being sensed.



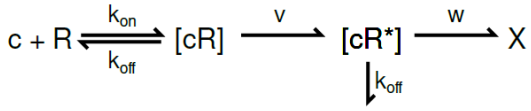
Compare the lucCage-lucKey technology with the negative-feedback system discussed in class, where a transactivator fused to the cage domain activates the expression of a YFP reporter and the key leads to the degradation of the cage-transactivator fusion, reducing YFP expression. Specifically, for each system, identify and discuss the key elements that influence the final output.

SOLUTION:

- **lucCage - lucKey: binding affinity Cage-Latch (most importantly caged target binding motif), binding affinity Cage-Key, binding affinity Target - caged target binding motif, binding cage - latch > binding target - caged target binding motif but at the same time not >>> or it doesn't open up, signal luciferase, binding affinity SmBit-LgBit, etc…**
- **Negative-feedback system: binding Cage - latch with degron, binding transcriptional activator - DNA, transcription and translation efficiency of YFP and Key, binding Cage - Key, temporal aspect (how fast are the proteins expressed?), efficiency of degradation (leaky degradation?), etc…**

GENE CIRCUITS

Background



Recall the kinetic proofreading scheme shown above, which we studied in class. We calculated the rate at which X is produced as summarized below.

Assuming v is much smaller than k_{off}, we have,

$$[cR] = \frac{k_{\text{on}}}{k_{\text{off}}} c \times R \quad , \tag{1}$$

where c and R stand for the the concentrations of the species c and R, and we omitted the brackets $[\ldots]$ to simplify the notation.

Next, assuming w is much smaller than k_{off}, we have,

$$[cR^*] = \frac{v}{k_{\text{off}}}[cR] = \frac{v k_{\text{on}}}{k_{\text{off}}^2} c \times R \quad , \tag{2}$$

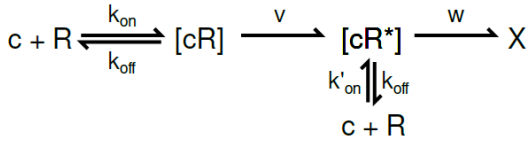
and finally,

$$\text{rate of production of } X = w[cR^*] = \frac{vw k_{\text{on}}}{k_{\text{off}}^2} c \times R \quad . \tag{3}$$

This gave us a k_{off}^2 term in the denominator, which we needed to explain the roughly squared reduction in the error rate of kinetic proofreading compared to the equilibium error rates.

We emphasized that the reactions parameterized by v, w, and the k_{off} dissociation of $[cR^*]$ be uni-directional reactions, with no reverse reaction.

Practice problem



But how important was it for the reactions to be unidirectional? Now consider the dissociation of $[cR^*]$ with rate k_{off} to be reversible with rate k'_{on}. Adapt the calculation above for the rate of production of X, assuming that $k'_{\text{on}} c \times R$ is a term that represents the direct production of $[cR^*]$, while everything else stays the same. How small must k'_{on} be compared to the other rates in this problem to be negligible?

Solution:

The only thing that changes is the steady-state equation for $[cR^*]$,

$$0 = v[cR] + k'_{\text{on}} c \times R - k_{\text{off}}[cR^*] \tag{4}$$

$$[cR^*] = \frac{v}{k_{\text{off}}}[cR] + \frac{k'_{\text{on}}}{k_{\text{off}}} c \times R \tag{5}$$

$$= \frac{v k_{\text{on}}}{k_{\text{off}}^2} c \times R + \frac{k'_{\text{on}}}{k_{\text{off}}} c \times R \tag{6}$$

$$= \frac{1}{k_{\text{off}}} \left(k'_{\text{on}} + \frac{v k_{\text{on}}}{k_{\text{off}}} \right) c \times R \tag{7}$$

The direct production of $[cR^*]$ is negligible if the following is satisfied:

$$k'_{\text{on}} \ll \frac{v k_{\text{on}}}{k_{\text{off}}} \quad . \tag{8}$$