De novo design of GCG, NPY, and PYY biosensors

Introduction –

Biosensors are becoming increasingly important in cell and neurobiology and remain highly important in the medical field [1][2]. Protein biosensors have the potential to revolutionize the way industry and clinical analytics are performed, while still offering a crucial role in monitoring a patients' status and detecting disease quickly. Current design methods for protein biosensors have been limited to reengineering natural proteins, which results in limited sensitivity and specificity for the target [3]. Finding compatible binding domains to engineer a sensor is challenging, and therefore a *de novo* design of a stable protein biosensor was explored. The Latching Orthogonal Cage-Key pRotein (LOCKR) system utilizes the principle that proteins fold to their lowest free energy state. Instead of focusing on conformational changes upon binding, the focal point was maximizing the free energy gap between the desired and resting structure [4]. Glucagon (GCG), Neuropeptide Y (NPY), and Peptide YY (PYY) were chosen as targets to test the LOCKR system since they are small in size, their binding domain sequences and structures are known, and all biomolecules show strong capabilities of being biomarkers for disease [5][6][7].

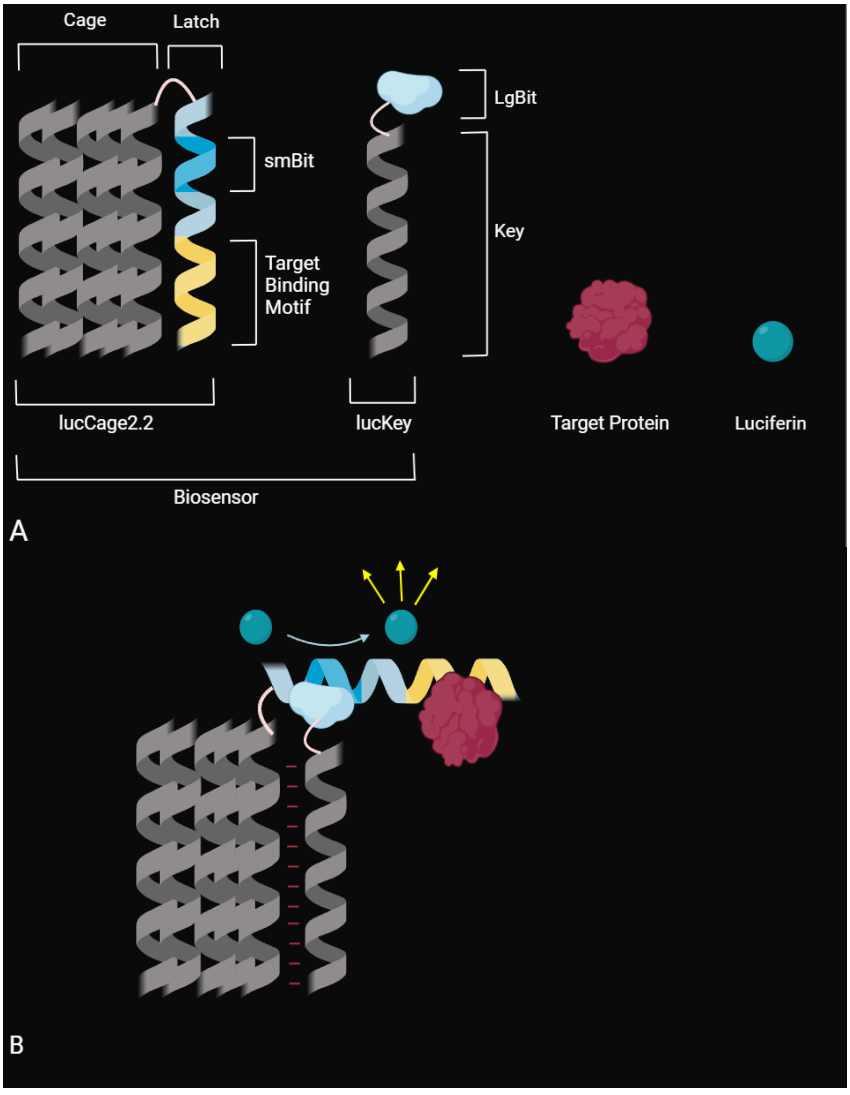

Fig 1. | *De novo* design of multi-state biosensor.
Fig 1.A: Schematic for inactive state of LOCKR biosensor. The closed form of lucCage2.2 cannot bind to lucKey, preventing the split luciferases (smBit and LgBit) from interacting with one another. *Fig 1.B:* Schematic for active state of LOCKR biosensor. The open state of lucCage2.2 binds to lucKey, recoupling the split luciferases to generate light in the presence of all other components.

Results –

Designs of the LOCKR system with the corresponding target protein's binding motif were successfully generated via a 'GraftSwitchMover' Rosetta-based method. Target binding peptides were placed along the latch domain with a total of twenty-four constructs manually chosen based on optimal interactions between the cage and latch regions. Synthetic gene fragments were ordered based on these designs and subsequently joined together into a linearized pET plasmid with Gibson cloning. Twenty out of the twenty-four constructs were successfully transformed into BL21 (DE3) competent cells, with the correct design sequence confirmed by Sanger sequencing.

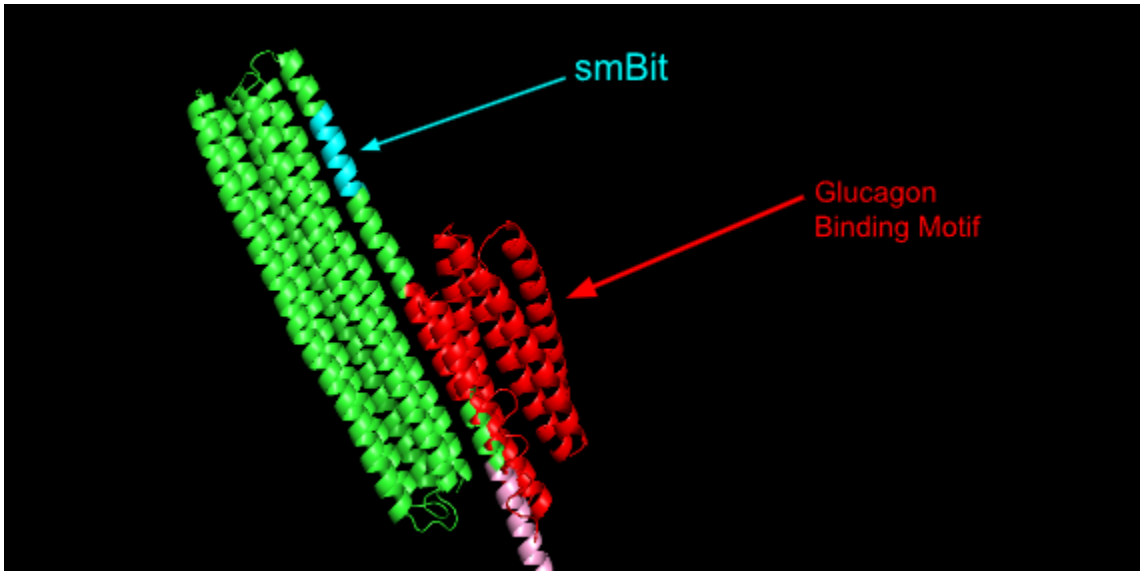

Fig 2. | Generating LOCKR system with corresponding peptide motifs.
Fig 2: Visualization of biomolecules for lucCage2.2. The completely green helical bundle represents the cage domain, and is connected to the latch domain with a linker (slightly above the smBit). The cyan region is the split luciferase that recouples with LgBit to generate light. The red region is an example of where a peptide binding motif can be placed along the latch. Ignore the pink region (Poly-A tail for PyRosetta design purposes).

Protein expression and purification were performed on the successful constructs and were separated into two batches. The first batch was left untouched after purification while the second was run through size-exclusion chromatography (SEC). SEC was used to separate monomeric proteins from aggregates and other impurities, ensuring that only properly folded and functional proteins were used in downstream assays. Two to three fragments of each construct were collected from the SEC and immediately flash frozen in liquid nitrogen. The concentrations were calculated for each crude construct by using a standard curve obtained by a Bradford assay.

Bioluminescence binding assays on the crude proteins with a serial dilution of the target protein were performed using a 96-well plate reader. The binding assays showed signaling from expected constructs and stronger signaling with higher target protein concentration. The constructs PYY334 (C1) and NPY334 (D1) both show the strongest signal under the 200nM concentration, and PYY330 (A1) and NPY330 (B1) show the second highest signal under the same concentration.

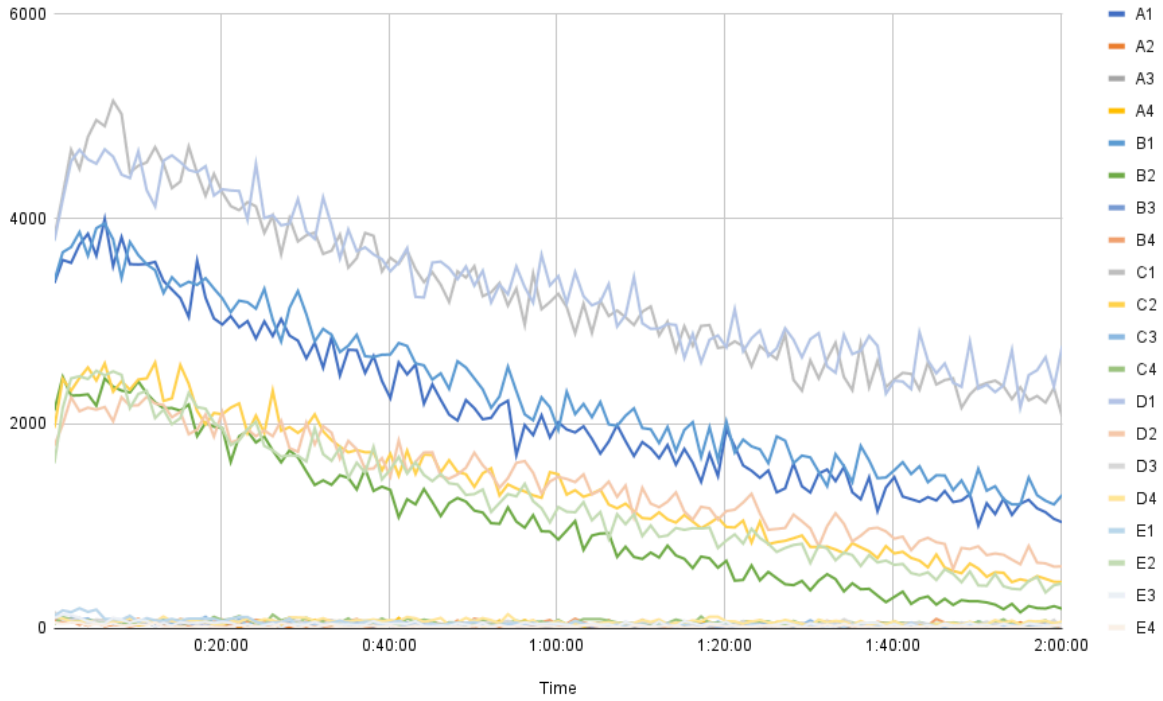

Fig.3 | Bioluminescence binding assay for NPY and PYY crude proteins.
Numbers correspond to specific constructs; 1: PYY330, 2: NPY330, 3: PYY334, 4:NPY334. Letters correspond to dilution of the peptide; A: 200nM, B: 100nM, C: 50nM, D: 25nM, E: 12.5 nM.

Discussion –

There has been considerable evidence that *de novo* protein biosensors work based on the basic principle that proteins fold to their lowest free energy state. The binding assay performed verifies bioluminescence activity, which can only appear if binding of lucCage, lucKey, and the protein of interest has occurred. Successful design of a protein biosensor without concern for the conformational changes of the binding pocket confirms our theory above and gives strong confirmation that *de novo* protein design can be an extremely effective tool for creating widely tunable and modular biosensors.

Early data shows that luminescence activity is present only in construct designs that are expected to work, while all the negative control designs have no luminescence activity. Additionally, the luminescence signal is typically stronger for higher concentrations of the target protein. Further assays with different batches of proteins would need to be performed to assess the characteristics of our GCG, NPY, and PYY biosensors. These biosensors also have a potential to be used in humans rather than an external diagnostic tool. Current research shows that the LOCKR system does not interfere with bodily or cellular processes, but extensive *in vivo* testing must be performed to ensure human viability, such as immunogenicity.

A wide range of sensing domains may be integrated into the LOCKR system to create new *de novo* biosensors for biomolecules of interest. This platform allows the creation of versatile protein biosensors that may deliver near instantaneous, no wash, high accuracy readouts. The successful design and validation of the LOCKR system as demonstrated in this

study highlights its potential for application in multiple fields including academia, clinical applications, and commercial use. There are many small molecules which may serve as important biomarkers for disease to use as targets.

Methods –

Design and Computational Grafting of Sensing Domain into lucCage

GCG, NPY, and PYY were chosen as target proteins to evaluate the efficacy of the LOCKR biosensor system because of their diverse biological functions, relevance to metabolic regulation and appetite control, small size, and known sequences and binding domains. The target peptide's sensing domain was grafted onto the latch of lucCage, using GraftSwitchMover, a RosettaScripts-based protein design algorithm, between the residues 323 and 359. The resulting designs were visually inspected which is crucial to verify the structural integrity and potential binding efficacy of each construct, reducing the likelihood of non-functional designs progressing to the synthesis stage. The constructs are then selected for gene synthesis, protein expression, and biochemical analysis.

Synthetic Gene Construction

The designed protein sequences were codon optimized for *E. coli* expression and ordered as synthetic genes for the pET29b+ expression vector. Codon optimization was necessary to maximize the efficiency of protein expression in E. coli, a widely used host for recombinant protein production due to its rapid growth and well-characterized genetics. The synthetic gene was then inserted at the NdeI and XhoI sites of each vector, including an N-terminal His6-tag.

Bacterial Cloning Procedure

In order to successfully clone the synthetic gene fragments, the pET29b+ plasmid, serving as the backbone for constructing the LOCKR biosensor, was amplified and linearized using polymerase chain reaction (PCR). PCR products were analyzed through gel electrophoresis to verify successful amplification and to assess the purity of the DNA fragments. Band sizes corresponding to the amplified backbone length were extracted and mini prepped for downstream use. Gibson assembly was subsequently employed to join the linearized lucCage backbone with the ordered synthetic gene fragments. The Gibson assembly products were then transformed into electrically competent BL21 *E. coli* cells and plated onto LB agar plates with kanamycin resistance and incubated overnight at 37℃.

Following the appearance of colonies on the LB/Kan plates, individual colonies were selected for further analysis. A single colony for each construct was picked and mixed with LB/Kan to start an overnight culture at 250 rpm and 37℃. A glycerol stock was also made for each colony picked for easier protein expression and purification later on. The cells were then harvested by centrifugation and mini prepped to isolate the desired plasmid. The extracted DNA plasmid's concentration was read using a microplate reader and sent off for Sanger sequencing to verify the correct sequences were transformed.

Protein Expression & Purification

E. coli from glycerol stocks were used to make overnight cultures in LB and kanamycin at 250 rpm and 37℃. 4 mL of overnight culture was then transferred to 100 mL of TB and kanamycin to grow overnight at 250 rpm and 37℃. IPTG was administered for protein expression once our cultures entered their log growth phase (0.6 - 0.8 OD). After administering

IPTG, the culture continued to incubate at 250 rpm and 37℃ for an hour and then left overnight at 250 rpm and 18℃. Cells were harvested the next day by centrifugation and resuspended with lysis buffer (20 mM Tris-HCl pH 8.0, 300 mM NaCl, 30 mM imidazole, 1 mM PMSF, 0.02 mg ml−1 DNase). Cell resuspensions were lysed by sonication for 2.5 min (5s cycles). Lysates were then centrifuged at 24,000g at 4°C for 20 minutes and passed through 2 mL of Ni-NTA nickel resin pre-equilibrated with wash buffer (20 mM Tris-HCl pH 8.0, 300 mM NaCl, 30 mM imidazole). The resin was washed twice with 10 column volumes (CV) of wash buffer, and then eluted with 3 CV of elution buffer (20 mM Tris-HCl pH 8.0, 300 mM NaCl, 300 mM imidazole). The eluted proteins were further purified using size exclusion chromatography. The fractions containing monomeric proteins were collected, snap-frozen with liquid nitrogen, and stored at -80°C.

In Vitro Bioluminescence Characterization

A BioTek Gen 5 microplate reader was used for all *in vitro* bioluminescence measurements. A final concentration mixture of 10 nM lucCage, 10 nM lucKey, and 25 nM DTZ was used to perform this assay. LucCage, lucKey, DPBS, and the target proteins were incubated at room temperature for an hour before DTZ was added and the assay was performed. Serial dilutions of the target protein in 200nM, 100nM, 50nM, 25nM, and 12.5nM were used for each construct. Bioluminescence measurements were taken every minute and continued for two hours.

References –

[1] C. Jackson, A. Anderson, and K. Alexandrov, “The present and the future of protein biosensor engineering,” *Current Opinion in Structural Biology*, vol. 75, p. 102424, Aug. 2022, doi: 10.1016/j.sbi.2022.102424.

[2] P. Mehrotra, “Biosensors and their applications – A review,” *J Oral Biol Craniofac Res*, vol. 6, no. 2, pp. 153–159, 2016, doi: 10.1016/j.jobcr.2015.12.002.

[3] A. Quijano-Rubio *et al.*, “De novo design of modular and tunable protein biosensors,” *Nature*, vol. 591, no. 7850, pp. 482–487, Mar. 2021, doi: 10.1038/s41586-021-03258-z.

[4] R. A. Langan *et al.*, “De novo design of bioactive protein switches,” *Nature*, vol. 572, no. 7768, pp. 205–210, Aug. 2019, doi: 10.1038/s41586-019-1432-8.

[5] A. Kolb *et al.*, “Glucagon/insulin ratio as a potential biomarker for pancreatic cancer in patients with new-onset diabetes mellitus,” *Cancer Biol Ther*, vol. 8, no. 16, pp. 1527–1533, Aug. 2009, doi: 10.4161/cbt.8.16.9006.

[6] S. Galli *et al.*, “Neuropeptide Y as a Biomarker and Therapeutic Target for Neuroblastoma,” *Am J Pathol*, vol. 186, no. 11, pp. 3040–3053, Nov. 2016, doi: 10.1016/j.ajpath.2016.07.019.

[7] F. Jing *et al.*, “PYY modulates the tumorigenesis and progression of colorectal cancer unveiled by proteomics,” *Am J Cancer Res*, vol. 12, no. 12, pp. 5500–5515, Dec. 2022.