ENGINEERING SYNTHETIC PROTEINS FOR BIOREMEDIATION OF RARE EARTHS AND APPLICATIONS IN MOLECULAR IMAGING

By

Harvey Dohyun Lee

A DISSERTATION

Submitted to
Michigan State University
in partial fulfillment of the requirements
for the degree of

Biomedical Engineering – Doctor of Philosophy

2023

ABSTRACT

The use of synthetic biology to carry out functions achieved through conventional means are often met with higher performance and cheaper production costs, such as the modern development and production of recombinant human insulin in E. coli. This dissertation focuses on utilizing synthetic biology to access the versatility of proteins and unique features of Rare Earth Elements (REEs) through binding motifs found in nature.

REEs are an essential resource for modern technology – anything with a screen, lens, glass, lights, magnets, steel alloys, or batteries require the use of REEs. In addition to their properties in magnetism, chemical reactivity, and temperature durability, REEs are also heavily utilized for their unique spectroscopic properties, making them crucial for almost every sector of industry, as well as molecular imaging assisted diagnostics. Current methods for mining these resources involve the extensive use of harsh chemicals and intense labor, not to mention low yields and excessive byproducts. Moreover, not only do REEs exist on Earth in a finite amount, but their mining and distribution is alarmingly reliant on very few sources, leaving the availability of such resources vulnerable to unforeseen circumstances.

It would therefore be beneficial for nations to develop REE recycling technology with higher yields, lower costs, and environmentally friendlier methods. Herein, REE binding motifs were integrated into newly designed synthetic proteins for the bioremediation of rare earths and applications in molecular imaging.

This dissertation is dedicated to friends and family who have always believed in me.

ACKNOWLEDGEMENTS

First and foremost, I would like to express my humble gratitude to my PhD advisor Dr. Assaf Gilad for his invaluable supervision, support, and tutelage throughout the years. It is his personalized mentorship and kind caring, as well as high standards that have made me the researcher I am today.

Special thanks are in order to committee members Dr. Adam Alessio, Dr. Christina Chan, Dr. Rupali Das, and Dr. Taeho Kim for investing their expertise, time, and energy into helping me along this path – I am flattered to have received their collective guidance, which is something truly unique.

I would also like to thank Dr. Jon Debling for providing great ideas and opportunities, as well as being a crucial part in the exploration of the commercial potential of our research.

Last but not least, I would like to thank Connor Grady, Katie Krell, and Cooper Strebeck for their invaluable contributions as coworkers and lab members.

This research would not have been possible without the financial support from the following sources:

- NIH/NINDS: R01-NS098231; R01-NS104306
- NIH/NIBIB: R01-EB031008; R01-EB030565; R01-EB031936; P41-EB024495
- NSF 2027113
- This work was also supported under the MTRAC Program by the State of Michigan 21st Century Jobs Fund, received through the Michigan Strategic Fund, and administered by the Michigan Economic Development Corporation

TABLE OF CONTENTS

LIST OF SYMBOLS AND ABBREVIATIONS

Abbreviation	Meaning
3D	3-dimensional
AU	Arbitrary Unit
CAGR	Compound Annual Growth Rate
CEST	Chemical Exchange Saturation Transfer
cfDNA	Circulating free DNA
cpEGFP	Circularly permuted Enhanced Green Fluorescent Protein
ctDNA	Circulating tumor DNA
DDIW	Double De-ionized Water
DNA9	Deoxyribonucleic Acid
FC	Fragment Crystallizable
FDA	Food and Drug Administration
GBCA	Gadolinium-based Contrast Agent
GLamouR	Green Lanmodulin-based Reporter
GLamouR-rs	Green Lanmodulin-based Reporter, red-shifted
HPRM1	Human Protamine-1
ICP-MS	Inductively Coupled Plasma, Mass Spectrometry
ICP-OES	Inductively Coupled Plasma, Optical Emission Spectrometry
IEDB	Immune Epitope Database
IgG	Immunoglobulin-G
LED	Light Emitting Diode
LRP	Lysine-rich Protein
mAb	Monoclonal Antibody

MHC	Major Histocompatibility Complex
MRI	Magnetic Resonance Imaging
NIR	Near-Infrared
PBMC	Peripheral Blood Mononuclear Cells
PCR	Polymerase Chain Reaction
PET	Positron Emission Tomography
POET	Protein Optimization Evolutionary Tool
PPPR	Pattern-Recognition Receptor
REE	Rare Earth Elements
RFU	Relative Fluorescence Unit
SPECT	Single-photon emission computed tomography
SPR	Surface Plasmon Resonance
TE	Echo Time
TR	Repetition Time
TRIS	Tris(hydroxymethyl)aminomethane Buffer
USD	United States Dollar
UTA	Universal Theranostic Adaptor
UV	Ultraviolet
WWTP	Wastewater Treatment Plant

* The International Union of Pure and Applied Chemistry's Periodic Table of Elements[1] contains all abbreviations for elements found in this dissertation

1. PREFACE

This dissertation encloses four independent projects with different applications that can be tied together with only one overarching theme: synthetic proteins. Each project has thus been divided into separate chapters as they deserve their own introduction and conclusion.

The first chapter, titled “A Novel Protein for the Bioremediation of Gadolinium Waste” contains details on a project that resulted in a high impact publication, patents pending internationally, funding from various government agencies, a prototype device built for widespread accessibility, invitations for collaboration from around the world, and even a potential startup business that is currently in the works. This technology is thought to better public health through the prevention of environmental gadolinium pollution and subsequent bioaccumulation in humans, as well as the reversal of pollution (or “bioremediation”) of wastewaters that have been already polluted. In addition to gadolinium recovery, industrial applications in REE recycling and biomining are being considered as well, which is anticipated to generate a considerable amount of these resources and add wealth to human society.

The second chapter, titled “Development of Non-immunogenic CEST MRI Proteins” describes a project that was carried out during the first year of graduate studies. The immunogenicity of newly designed proteins is an important factor to consider for medical applications and *in-vivo* studies because it can also provide properties such as prolonged half-life of the protein for applications such as long-term cell tracking. Although new sequences were designed, tested, and ranked *in-silico*, the project still remains to be further pursued through *in-vitro* and *in-vivo* validation.

The third chapter, titled “Accurate Quantification for Synthetic Proteins” was a small pet project but an important one, especially for researchers working with synthetic proteins. To gauge the performance of new proteins, assays have to be conducted at equimolar concentrations so that each molecule can be fairly assessed on a level playing field. It is anticipated that the completion of this project will revolutionize the field of *de novo* synthetic protein engineering through accurate data acquisitions that in turn may facilitate the creation of other newly designed proteins that benefit human health.

The fourth chapter, titled “A Molecular Imaging Adaptor Protein” was the most coveted and fascinating study to personally pursue that yielded the most exciting and outstanding results but unfortunately did not receive the care it deserved due to the hot attention that was already generated towards the study highlighted in the first chapter. Through the development of such adaptor, however, it is anticipated that the development of molecular theranostics will be accelerated years ahead of time through the increased accessibility of attachable molecular imaging agents.

Rather than requiring the exploration of innovative methods in protein design or getting lucky with a particular sequence, the projects described in this dissertation required one’s ability to optimize assays and troubleshoot errors to develop new and practical technologies, with a multidisciplinary approach that can be expected from a true biomedical engineer.

2. CHAPTER 1) A NOVEL PROTEIN FOR THE BIOREMEDIATION OF GADOLINIUM WASTE

2.1 INTRODUCTION

Gadolinium in MRI

Gadolinium is widely used in medicine due to its unparalleled paramagnetic properties, which generate contrast due to its seven unpaired electrons in the S ground state with a large magnetic moment and long electronic relaxation time. In successful GBCAs the Gd^{3+} is bound to eight coordinating atoms of the chelate, and the ninth position interacts with rapidly exchanging water molecules. The Gd^{3+} shortens the water molecule's T1 relaxation time, and due to the fast exchange of the water molecule with the bulk water molecules, it shortens the average T1 of bulk water. When optimal TR and TE are used, GBCA labeling appears as hyperintensity on the MR image[2,3]. The bonding of Gd ions with chelating compounds decreases the toxicity of the element, as the Gd^{3+} would not be available for any other biochemical binding events other than those with higher affinity. Chelating compounds used to form complexes for contrast agents have certain properties such as high affinity and specificity of Gd^{3+} binding, good water solubility, and increased stability. Such properties provide quick and almost complete excretion of Gd complexes from the patient's body with normal renal excretion.

GBCAs are classified according to their chemical structure and pharmacokinetic properties. Linear GBCAs have the lowest stability and the highest risk of NSF, a rare but serious condition that affects the skin and internal organs. On the other hand, macrocyclic GBCAs have a higher stability and a lower risk of NSF. Ionic GBCAs have not been associated with NSF but have a higher risk of anaphylactic reactions. The most common route of administration for GBCAs is intravenous injection, but oral, rectal, and intra-

articular administration have also been reported. The amount of GBCAs used in an MRI scan is typically small, in the range of 0.1-0.3 mmol/kg body weight. The elimination of GBCAs is mainly through the kidneys, and the elimination half-life varies depending on the type of the agent.

There is an ongoing demand to improve the selectivity and sensitivity (T1 relaxivity) of these contrast agents for the detection of targets on the molecular level[4]. T1 relaxivity is also known as spin-lattice relaxation, or longitudinal relaxation, which is the time constant for the regrowth of proton spins to recover to their original aligned state inside the magnetic field in relation to the "lattice", or surrounding environment. T2 is known as spin-spin relaxation or transverse relaxation, which is the time constant for the decay of transverse magnetization. This may occur without T1 relaxation in the case of a dipolar interaction known as a spin-spin "flip-flop", where a pair simultaneously exchanges their longitudinal angular momentum, resulting in no net T1 but a loss of T2 coherence. Correspondingly, protein-based MRI contrast agents have been reported with attributes such as enhanced relaxivity, stability, and targeting capabilities[5]. In a pioneering work in the field, Louie et al. (1999) was able to monitor gene expression with beta-galactosidase as the reporter, which cleaves off a portion of an activatable MRI contrast agent to introduce a sudden increase in water availability, turning it into an "on" state[6]. This study set the stage for a variety of genetically encoded reporters based on T1 relaxation. Bartelle, Turnbull, and colleagues took a creative approach to express a multi-component gadolinium based reporter in endothelial cells[7]. Another strategy relied on overexpression of an organic anion transporting protein which facilitates the uptake of Gd^{3+}-EOB-DTPA [8,9] . An alternative approach focused on the development of a protein-based contrast

agent that exhibits high stability and greater selectivity for Gd^{3+} over Zn^{2+}, which was used as an injectable agent[10].

Anthropogenic gadolinium pollution

Many critical decisions in the clinic today rely on MRI. Due to its large magnetic moment and a long electronic relaxation time, gadolinium (^{64}Gd) is commonly used to enhance the contrast of the MR image[2,3]. It is estimated that more than 30 million doses of GBCAs are administered to patients annually[11], which are then excreted into the environment. Any unused portions of each bottle/syringe are also discarded, totaling up to several hundreds of tons of GBCAs being emptied into our wastewaters every year. The accumulation of gadolinium in our environment is being addressed with increasing urgency since the second decade of the 21st century; recent studies show anthropogenic gadolinium to be lurking in our wastewaters, crops, tap water, and even marine wildlife[12,13].

The current state-of-art for disposing radioactive waste is to collect and store for 10 half-lives, or when activity has decayed to less than 0.1%[14]. Pharmaceutical waste is treated and incinerated at legally registered, licensed facilities[15]. GBCAs, on the other hand, do not have any regulations in terms of proper disposal – the rational being that gadolinium is safe to use when properly chelated. Many GBCAs have been FDA approved to use in the clinic for their safety during bodily circulation and the unparalleled contrast they provide for better healthcare. Although gadolinium retention inside the body has recently come to light, the benefits of GBCAs still undoubtedly outweigh the possible side effects for repeated administration in critical patients[16]. However, the long-term effects of GBCAs being disposed into the environment (and, ultimately, the human body, which consumes products of the environment) have not been fully considered.

Accumulation of gadolinium in the environment has now been going on for more than 30 years since the introduction of gadopentetate in 1988, with increasing GBCA administration every year, but there are currently no plans for the prevention of further pollution or efforts to reverse the contamination in our natural water sources.

Gadolinium retention *in-vivo*

Administering gadolinium containing products without any regulation for disposal may someday lead to the bioaccumulation of gadolinium inside humans, as seen in crops and marine wildlife. It has been reported that repeated administration of GBCAs leads to gadolinium retention inside the skin, bones, brain, and muscle[17]. The species of which the gadolinium exists (chelated, non-chelated, bound to other compounds such as phosphates, etc.) has yet to be proven with hard evidence. However, should gadolinium break free from their chelates, it would be possible for non-chelated, free gadolinium to bind to receptors which should otherwise be binding calcium[18]. This would hinder any biological functions requiring calcium such as muscle contraction or neuronal activity, as typically observed in symptoms associated with gadolinium deposition and Nephrogenic Systemic Fibrosis[19-21]. Consequently, individuals with these diseases may experience swelling, thickening, and tightening of the skin, burning sensations and sharp pains, brain fog, restriction in mobility from affected muscles and joints, reduced internal organ functions that include the heart, liver, lungs, and gastrointestinal tract, and yellow plaque formation on the sclera of the eyes.

Evo-HLn: A genetically engineered microbe for the bioremediation of gadolinium waste

Methylotrophic bacteria have previously been reported to possess the ability to mine rare-earth elements (REEs)[22]. However, low levels of attention have been directed towards cleaning the environment of gadolinium.

To this end, we developed a genetically modified strain of *Methylorubrum extorquens* AM1 that had been evolved to be able to utilize gadolinium as its sole REE source for methanol growth, termed *evo*-HLn[23]. In-out allelic exchange with 5% sucrose for counterselection was used to generate the mutation, which was confirmed by PCR and genomic Sanger sequencing of the META1_1800 locus. Growth with gadolinium was only apparent for the mutated strain albeit 3.7 times slower than with lanthanum. The evo-HLn that had consumed gadolinium was apparently more vivid in magenta color when compared to the wildtype strain and ancestral strain that was fed lanthanum, which are closer to pink.

Figure 1a shows the first evidence of an intracellular lanthanide containing compartment we termed the “lanthasome. The size of the surrounding lanthasome for gadolinium (bottom right) is much bigger than lanthanum (top right), suggesting an increased capacity for storing gadolinium, which was confirmed with MRI and ICP. Figure 1b shows the T1 relaxation times of the cells that were acquired with MRI and their corresponding T1 maps, and it is evident that the *evo*-HLn (denoted as evo) that was fed gadolinium (in blue) possessed striking contrast when compared to other strains that were fed lanthanum (pink). Interestingly, the *evo*-HLn showed a preference for the lighter REE lanthanum when fed both gadolinium and lanthanum at equimolar concentrations (grey),

as T1 relaxation time did not decrease as much as it did for when gadolinium was the sole REE source. Figure 1c shows the intracellular REEs quantified with ICP-MS per cell dry weight, and the evolved strain could store up to 36 times more gadolinium than the wildtype, 3-fold less lanthanum, and 82-fold more gadolinium than lanthanum, but severely impaired storage for gadolinium when provided both gadolinium and lanthanum at equimolar concentrations. Finally, we put the *evo*-HLn to the test by providing gadolinium chloride (denoted as free Gd) as one experimental group, and gadopentetic acid (Gd-DTPA, denoted here as GBCA) as another. The evolved strain showed remarkable growth with the linear GBCA at a stunning rate and was on par with even unchelated gadolinium.

Unfortunately, growth with gadoteric acid was not observed, suggesting that the macrocyclic GBCA was much more stable and unwilling to release gadolinium ions from their chelates than their linear counterparts. Overall, strains such as the *evo*-HLn show promise for the bioremediation of gadolinium from GBCAs, although slow growth rates, the requirement of heavy instruments, deterred storage capacity for REE mixtures, and extensive labor to purify REEs from cells would be hurdles to overcome when considering the development of such pipelines.

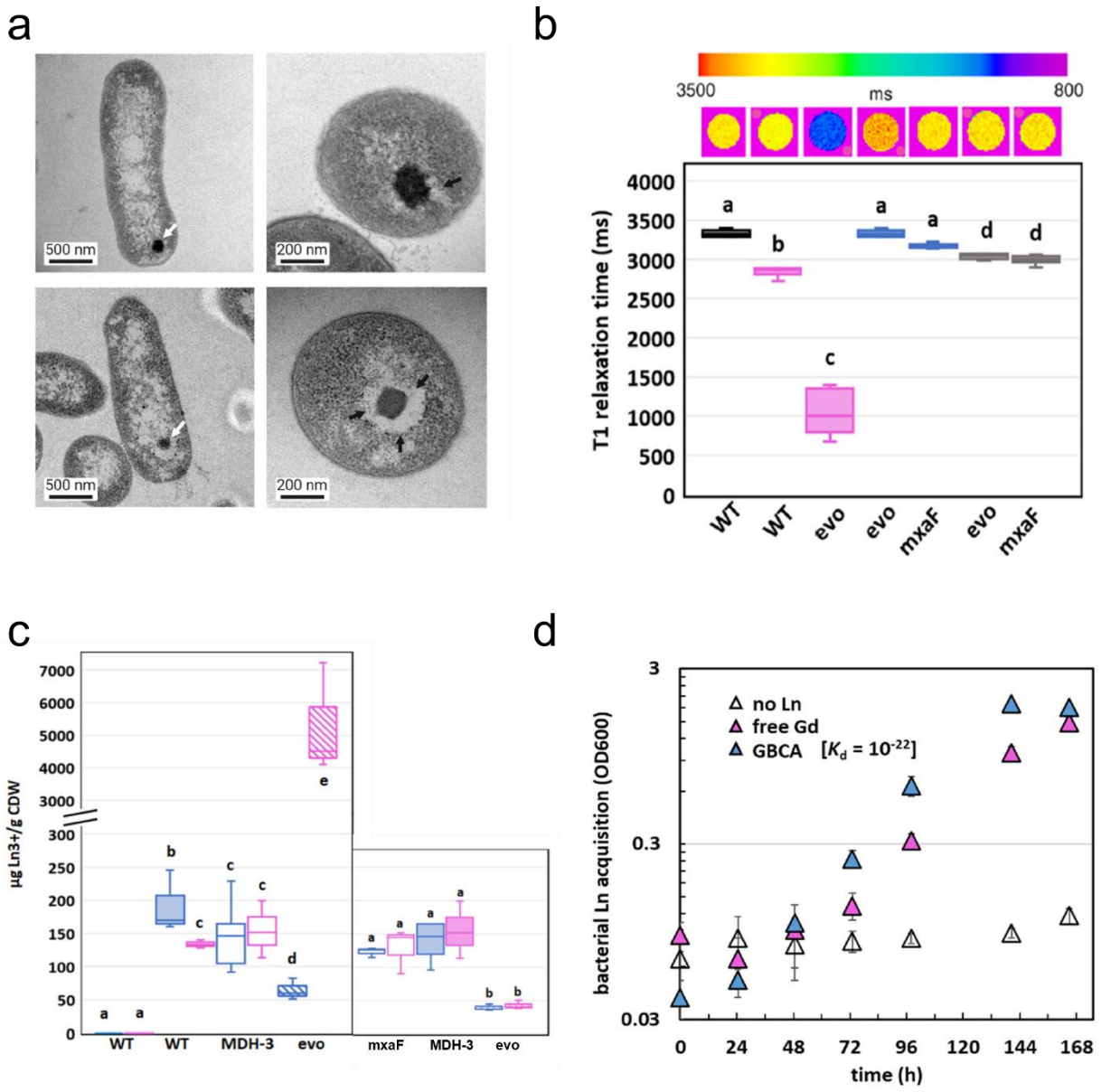

FIGURE 1. Evo-HLn, a mutant strain of genetically modified bacteria for the hyperaccumulation of gadolinium. (a) TEM of *evo*-HLn. White arrows show electron-dense lanthanide deposits while black arrows indicate the surrounding lanthasome. (b) Whole-cell MRI of strains fed with gadolinium, lanthanum, and both. (c) ICP-MS was used to quantify lanthanide content of the cells imaged with MRI. (d) Growth curve of *evo*-HLn when given gadolinium in free and linearly-chelated forms.

Other Rare Earth Elements

REEs are an essential resource for modern technology – anything with a screen, lens, glass, lights, magnets, steel alloys, or batteries require the use of REEs, as well as conventional “non-electric” vehicles that run on a form of fossil fuel (which also require a catalytic converter)[24]. Current methods for mining REEs require extensive use of harsh chemicals and intense labor, not to mention low yields and excessive byproducts[25]. Moreover, not only do REEs exist on Earth in a finite amount, but their distribution is alarmingly reliant on REE-rich countries with affordable labor costs and looser regulations on environmental pollution arising from industrial practices[26]. On the bright side, efforts are being made to develop new technologies for mining REEs at a lower cost and for recycling them[27]. The market for recycling REEs was evaluated at 248 million USD in 2021, and it is projected to grow at a rapid pace with a CAGR of 11.2%, reaching 422 million USD by 2026[28].

It would thus be beneficial for nations to develop technology for recycling REEs, preferably at higher yields, cheaper costs, and hopefully with environmentally friendlier methods. In addition to utilizing microbes, cutting-edge technology involving the use of peptides and proteins for biomining are being developed by researchers as well [29-35]. Aspects to consider for industrializing the application of such technologies would be the identification of reliable REE sources, performance and reproducibility across various sample types, and cost effectiveness.

Creating a synthetic protein for the bioremediation of gadolinium waste

As a potential solution to mitigate Rare Earth Element (REE) pollution, methylotrophic bacteria have been shown to uptake and store light lanthanides and are being proposed as a potential tool for bioremediation and biomining[36,37]. We have recently developed a mutant of *Methylobacterium*[38]/*Methylorubrum extorquens* AM1 that has acquired the ability of hyperaccumulating heavy lanthanides such as gadolinium[23]. However, as mentioned earlier, slow growth rates, the requirement of heavy instruments, deterred storage capacity for REE mixtures, and extensive labor to purify REEs from cells are major drawbacks to utilizing these strains of bacteria.

As an alternative, we shifted our focus to the proteins expressed in these bacteria to carry out the specific function of REE bioremediation. The methylotrophic bacteria express a unique protein with the ability to bind lanthanides with picomolar affinity, termed lanmodulin[39], which is thought to play an important role in its metabolic processes. Lanmodulin binds lanthanides via EF-hand motifs that are 12 amino acids long[40]. Due to the similarity of these lanthanide-binding motifs to calcium-binding motifs, we used the backbone of a well-established genetically encoded calcium indicator in the GCaMP family[41] to construct a new synthetic protein termed Green Lanmodulin-based Reporter (GLamouR), which reports gadolinium binding through green fluorescence.

2.2 MATERIALS AND METHODS

MATERIALS

Reagents

All materials were purchased from Thermo Fisher or its subsidiaries, with the exception of REEs and calcium chloride purchased from Sigma-Aldrich, centrifugal filtration units and artificial urine diluent from Merck, and FPLC columns from Cytiva.

Cell lines

- For DNA replication and other cloning purposes: E, Coli, TOP10 (Thermo Fisher), 5-alpha (New England Biolabs), 10-beta (New England Biolabs)
- For protein expression: E. Coli, BL21* (Thermo Fisher)

Equipment

- Fluorescence measurements: Cytation 5 (Biotek), Spark (Tecan)
- MRI: 7T 70/30 Biospec (Bruker)
- ICP: Varian 710-ES (ICP-OES), Thermo Fisher iCAP-Q (ICP-MS)
- Gel visualization: ChemiDoc Imaging System (BioRad)

Software

Prism (GraphPad) was used for generating graphs and performing statistical analysis. SnapGene was used for designing vectors containing new proteins and generating their corresponding maps, and BioRender was used for creating illustrations. Paravision 360 was used for obtaining T1 and T2 relaxation times and their corresponding maps.

METHODS

Gene design and cloning

Proteins were cloned via Gibson Assembly or TOPO cloning of DNA fragments obtained from PCR of existing constructs (GCaMP6m, rGECO1, Lanmodulin) and custom ordered fragments from IDT. Cloning results were confirmed after transforming the final vector into TOP10 to obtain sufficient DNA for Sanger sequencing (provided by Genewiz) and other downstream applications.

Protein expression

Proteins were expressed by E. coli (BL21*) that had been transformed with the cloned pET101 vectors containing the GLamouR constructs. Cells were incubated with ampicillin-spiked (100 µg/mL) Magic Media for 24hrs at 30^{o}C, shaking at 300-360 RPM. Expression and purification were verified via Western Blot against the V5 tag.

Protein purification

Purification was performed via HIS-tag purification with cobalt resin. For small (<50 mL) cultures, 200 µL columns were used, whereas larger volumes (>400 mL) were purified via FPLC (AKTA by Cytiva). Buffer exchange was done with either centrifugal filtration units (3-10kD, 4-15 mL), desalting columns (7kD), or dialysis cassettes (10kD) at least three consecutive times with 25 mM TRIS buffer at pH 7.0. Further purification via size exclusion was performed as necessary, with HiLoad 16/600 Superdex 200pg columns connected to the FPLC system.

Fluorescence Spectrometry

Fluorescence was measured with the Cytation5 (Biotek) with excitation at 488 nm and emission at 510 nm, with monochromators and/or filters. Wells were prepared in triplicates (unless indicated otherwise) with a 10-200 nM concentration of GLamouR (quantified via sequence-specific a205) in TRIS buffer (25 mM, pH7). After the second read, REEs/negative controls were introduced to reach desired concentrations (with ten averages per read).

Inductively Coupled Plasma

ICP data were acquired with either ICP-OES (Varian 710-ES) or ICP-MS (Thermo iCAP-Q), depending on the required limit of detection. Samples were brought to a volume of 1.0 mL using 2% nitric acid (Optima grade), then further diluted 10-fold into the 2% nitric matrix. Samples were spiked with 2 ppb of an internal standard mixture of Li, Sc, In, Y, and Bi. In was selected as the internal standard for analysis of Gd. For each analyte, the dwell time was 0.3 S and 6 main runs of 15 sweeps were performed. Each analyte was measured in KEDS mode by introduction to a collision cell, using helium as the collision gas. Isotopes for each analyte were selected by Qtegra software (ThermoFisher Scientific) to maximize signal abundance while minimizing potential interferences. Error bars were generated with data acquired at different wavelengths (OES) or isotopes (MS). Quantification was performed against calibration curves generated for each analyte of interest. Prior to analysis, the mass spectrometer was tuned, and instrument calibration was verified using automated routines. In-run analytical precision and accuracy were assessed by reintroduction of aliquots of a selected calibration standard interspersed

throughout the run. Data were corrected to a matrix blank of 2% nitric acid (Optima grade) prepared in ultrapure water and analyzed using Qtegra.

Magnetic Resonance Imaging

MRI data were acquired using a 7 Tesla 70/30 Biospec (Bruker) equipped with 12-inch gradients and an 86 mm transmit/receive volume coil. T1 maps were generated via Paravision 360, with ten TR experiments ranging from 100 ms-17500 ms with an echo time of 6.89 ms and three averages each. Multiple 1 mm-thick slices were acquired across samples contained in PCR tubes, 0.5 mL tubes, and 1.5 mL microcentrifuge tubes. Samples were prepared by allowing GLamouR (volumes from 400 uL-1 mL, concentrations from 0.5 mg/mL-1 mg/mL) to bind with gadolinium (0.5 mM) for 30 mins at RT, followed by several washing steps, identical to the methods used for buffer exchange following protein purification.

Relaxivity (r_i) values were calculated with the following equation[42] :

$$R_i = R_i^0 + r_i \times [CA]; \qquad i = 1,2$$

where concentration is denoted by [CA], which is the Gadolinium concentration measured by ICP. R_i is the relaxation rate of the solution in presence of the contrast agent (=1/T_i). R_i^0 is the relaxation rate of the solution without the contrast agent.

2.3 RESULTS

DESIGN, EXPRESSION, AND PURIFICATION

Many genetically encoded calcium indicators incorporate two pieces of a circularly permuted fluorescent protein that come together to constitute a fully functioning state[43-50]. In such proteins, calmodulin is known to wrap around and interact with the M13/RS20 upon binding calcium, which effectively restricts water access to the chromophore, resulting in deprotonation to an anionic, fluorescent state[51]. Due to the similarity in primary structure for calcium binding motifs to lanthanide-binding motifs, it was hypothesized that a hybrid of lanmodulin and calmodulin would be compatible with the M13 while maintaining its selectivity for gadolinium over calcium. Figure 2a depicts the protein map on the left, with a 2-dimensional representation of conformational change on the right. After constructing the protein *in-silico*, DNA was purchased and cloned into an expression vector for E. coli, which were transformed with the purified plasmid. Expression and purification of the protein were subsequently carried out before assessing function and characteristics.

2.3.2 RAPID DETECTION AND QUANTIFICATION OF GADOLINIUM

Gadolinium detection capabilities of the GLamouR were assessed with fluorescence spectrometry. Upon binding gadolinium, the protein increases its green fluorescence (Figure 2b). Ten readings were measured and averaged out for each datapoint every ten seconds, and the reaction was found to be within this window of time, suggesting that binding occurs almost instantly. As seen in Figure 2c, the protein reaction time occurs almost instantaneously. The fluorescence increase is denoted on the y axis as Δf/f0, which is the change in fluorescence. This is calculated by subtracting the baseline

fluorescence value established in the first read(f_0) from the current read(f) resulting in Δf, which is then divided by f_0 and converted into a percentage. It can be seen that immediately after injection the fluorescence increases, with slight photobleaching resulting from repeated reads. TRIS buffer and calcium were also tested as negative controls, showing a decrease due to the lowered concentration of fluorescent protein upon administration of the negative controls followed by slight photobleaching over consecutive reads.

With increasing concentrations of gadolinium in separate wells, we observed the protein's saturation kinetics. Figure 2d, shows that the GLamouR has a linear response up to 50 µM, suggesting its capability of quantifying gadolinium in the micromolar range. Furthermore, GLamouR is reliably detectable at 10 nM (Figure 3a) and is sensitive enough for 200 nM concentrations of gadolinium as shown in Figure 2e. Assuming a limit of detection at 3 standard deviations, it is suggested that a sample containing gadolinium can be reliably differentiated from one without gadolinium when the concentration is at least 98 nM. A limit of quantification assumed at 10 standard deviations suggests quantification can be performed at concentrations starting from 425 nM. GLamouR was also tested for its thermal/temporal stability by leaving a 20 µM solution of GLamouR at room temperature for a period of one week. Although there was a slight decline in performance increase over time, GLamouR was able to maintain functionality for the entirety of the week (Figure 3b).

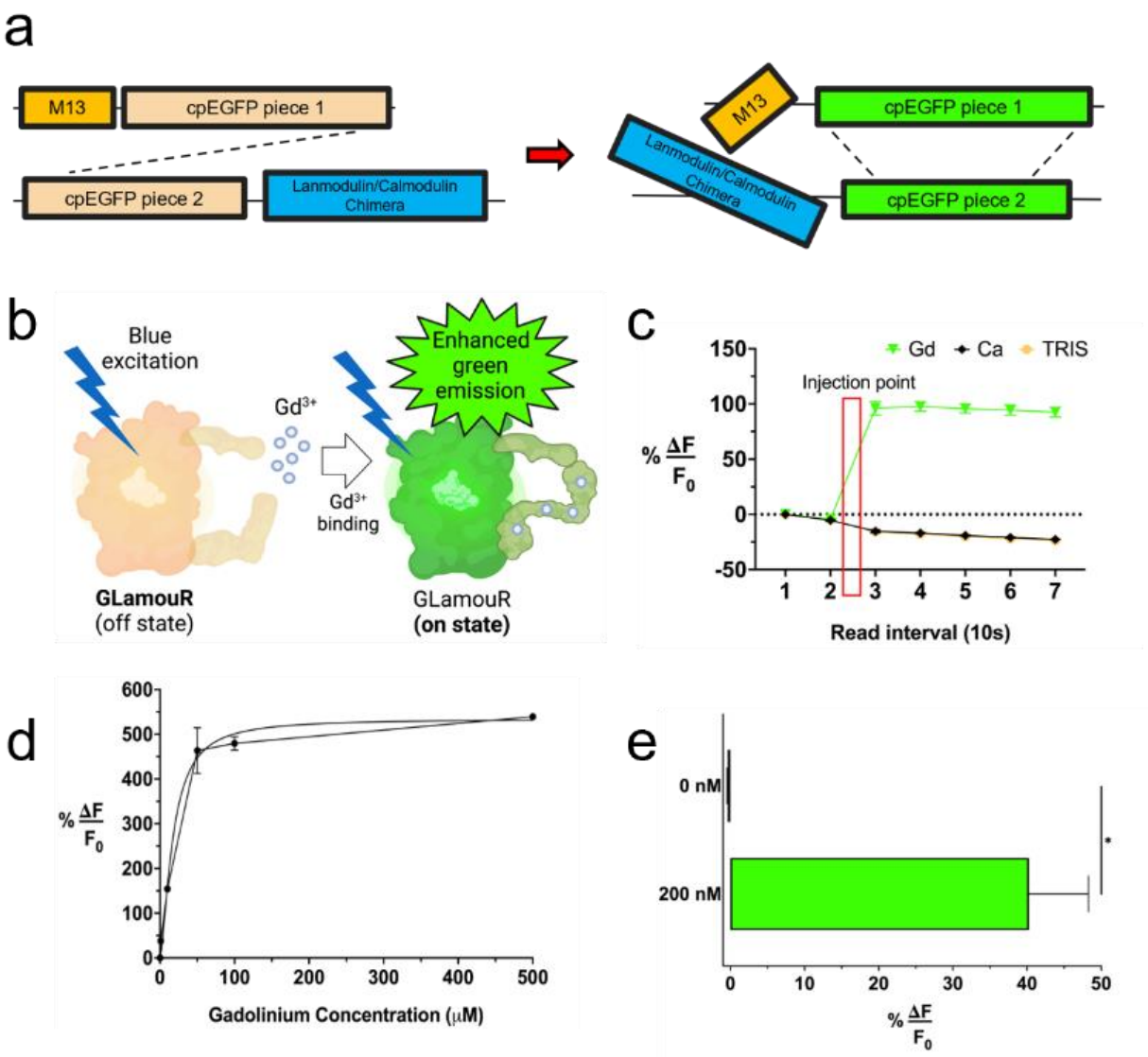

FIGURE 2. Fluorescence properties of GLamouR. (a) A map of the protein in its native state on the left, with the hypothesized compatibility of the lanmodulin/calmodulin chimera with M13 on the right. (b) Schematic representation of fluorescence and GLamouR-Gd^{3+} binding. GLamouR in its "on" state will possess enhanced green emission upon excitation with blue light. (c) Fluorescence kinetics of GLamouR with gadolinium (Gd), calcium (Ca), and TRIS buffer injection, which were administered after the second datapoint (ten reads

FIGURE 2. (cont'd)

per datapoint). Percent change in fluorescence, or $\Delta F/F_{0,}$ is calculated by subtracting the first read (F_0) from the observed fluorescence at any given time point (F), divided by F_0 and then given as a percentage by multiplying a hundred as follows: $100(F-F_0)/F_0$ (d) Fluorescence saturation curve for GLamouR with gadolinium shows a linear response up to 50 µM. (e) 200 nM concentrations of gadolinium were detectable with GLamouR with a 40% increase in fluorescence upon injection. Statistical significance was calculated by an unpaired t-test with Welch's correction, p value=0.0128 (two-tailed).

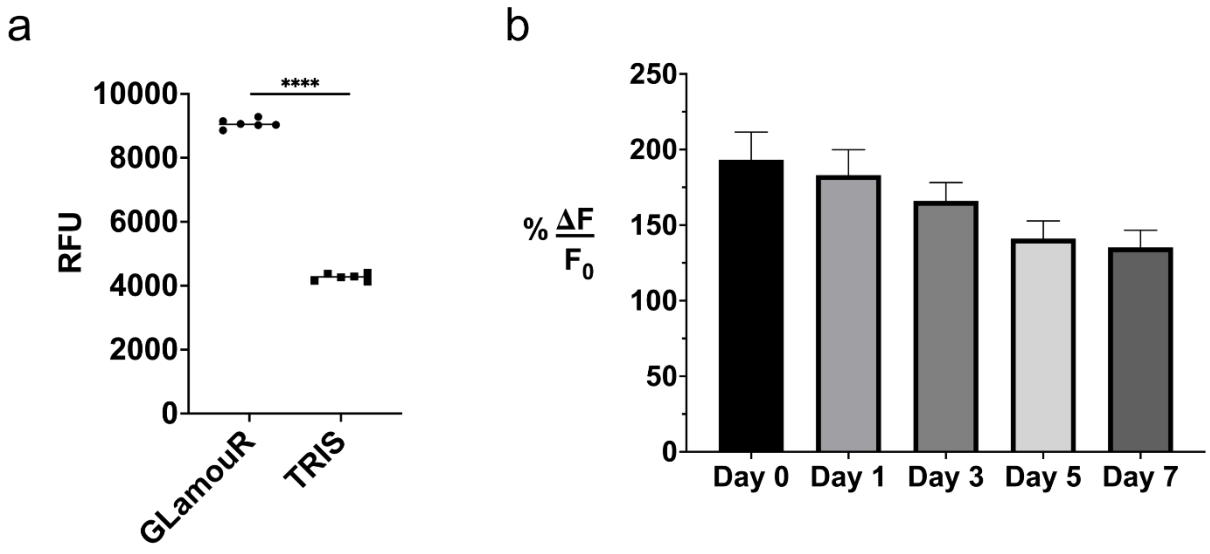

FIGURE 3. Baseline readings and thermal/temporal stability of GLamouR. (a) GLamouR can be reliably detected at 10 nM (n=6 wells, $p < 0.0001$ with Welch's correction) in TRIS buffer and used as baseline readings for assays. (b) GLamouR solutions at 1mg/mL were left at room temperature for up to 7 days and frozen accordingly before thawing for the assay. 2 µL of GLamouR mixed with 243 µL of TRIS buffer were used per well, along with 5µL injections of 5mM gadolinium (n=5 wells).

VERIFICATION OF GADOLINIUM BINDING

Membrane filtration and MRI were used to verify that GLamouR was in fact capturing gadolinium and not just reporting its presence in an indirect manner. After binding GLamouR with gadolinium, the solution was dialyzed to remove excess gadolinium and subsequently filtered with washed membranes of centrifugal filtration units to isolate the protein (Figure 4). The retentate would therefore hold GLamouR-Gd conjugates while the filtrate would be the surrounding buffer. T1 and T2 relaxation times are shown in Figures 5a and 5b, with their corresponding T1 and T2 maps (Figures 5c, 5d). Figures 5a-d show the filtrate containing insignificant levels of gadolinium after dialysis, whereas the protein was able to hold on to gadolinium at high levels of centrifugal force and possess MRI contrast. By plotting relaxation rates as a function of gadolinium concentration, we obtained the r1 and r2 relaxivities of GLamouR-Gd conjugates, which were 6.0 $mM^{-1}s^{-1}$ and 41.85 $mM^{-1}s^{-1}$ respectively. These numbers represent the enhanced relaxivity of gadolinium upon binding GLamouR (Table1) [42,52-55], which is likely to be resultant from the tumbling time shifting closer to the Larmor frequency as a conjugate, when compared to free gadolinium. To obtain higher concentrations of GLamouR-Gd conjugates, GLamouR was lyophilized, reconstituted at a lower volume with DDIW and dialyzed with fresh TRIS buffer at 25 mM to match the salt concentrations before initiating gadolinium binding. Compatibility with centrifugal force, lyophilization, and long dialysis times demonstrate the durability and stability of GLamouR. Thermal properties were also assessed with a thermal shift assay suggesting tighter folding at temperatures closer to 45^{o}C (n=5 wells), but minimal variance in performance was found due to temperature (Figure 6).

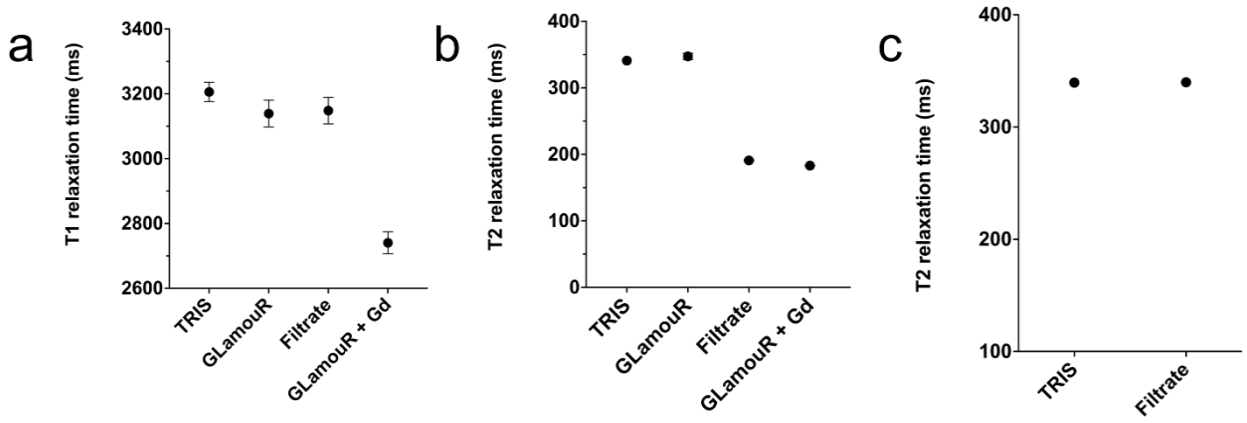

FIGURE 4. Washing membranes for filtration and MR Imaging. The effect of washing membranes before filtration can be seen with MRI. (a) T1 relaxation time does not seem to be lower for the filtrate sample. (b) On the other hand, T2 relaxation time is significantly decreased. (c) Upon washing the membrane before filtration, the filtrate T2 is returned to its normal values. It was found that glycerin on the membranes (for moist storage) were contributing to T2, which could be easily washed off by running DDIW through the membrane.

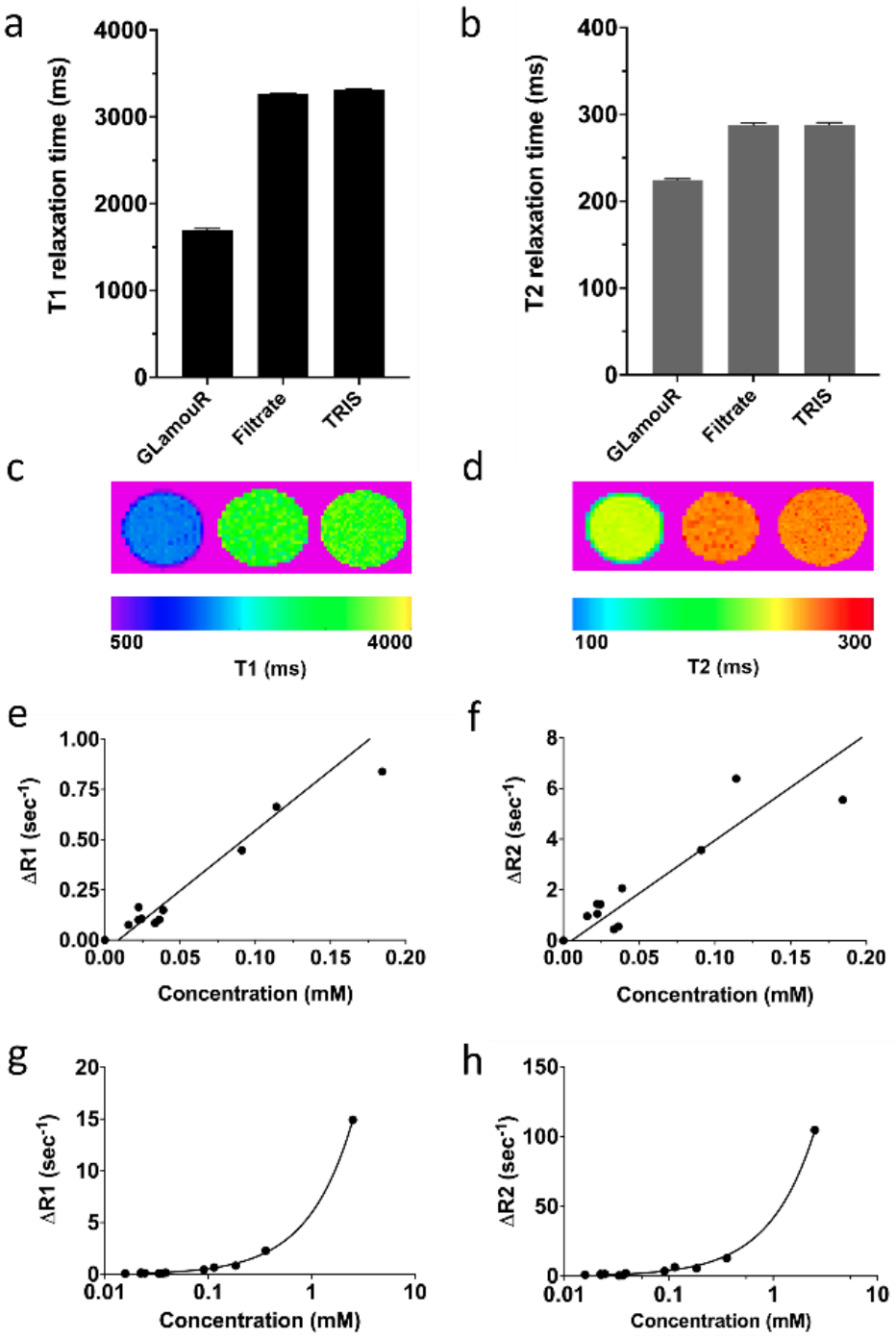

FIGURE 5. MRI data confirm protein-Gd binding. GLamouR was bound with gadolinium and dialyzed thoroughly prior to separation of filtrate (surrounding buffer) and retentate (GLamouR-Gd conjugates). T1 (a) and T2 (b) relaxation times were compared

FIGURE 5. (cont'd)

against TRIS buffer by averaging the ROIs of their corresponding T1 (c) and T2 (d) maps. r1 and r2 relaxivity (per-gadolinium) were derived from thirteen dilutions across six different batches of GLamouR. Linear regression for R1 and R2 datapoints as a function of gadolinium concentration are shown in linear (e, f) and logarithmic scale (g, h), with $p < 0.00001$ and $r^2 = 0.999$.

TABLE 1. Relaxivity values for GLamouR-Gd conjugates compared to gadolinium chloride hexahydrate and other chelators. Asterisks (*) denote values calculated with data collected from this study.

Compound	Buffer	Temperature	Field Strength	r1 (L/mmol-s)	Reference
GLamouR	TRIS 25mM, pH7	RT	7 T	6.0	*
GdCl3*6H20	TRIS 25mM, pH7	RT	7 T	4.6	*
Gadoxetate Disodium	Water	37 ºC	4.7 T	4.9	[55]
	Water	37 ºC	3 T	4.3	[55]
	TRIS 50mM, pH8	RT	7 T	4.9	*
Gadobutrol	Water	37 ºC	4.7 T	3.2	[55]
	Water	37 ºC	3 T	3.2	[55]
	Human plasma	37 ºC	7 T	3.8	[54]
	Human plasma	37 ºC	7 T	4.7	[53]
	Human blood	37 ºC	7 T	4.2	[52]
	TRIS 50mM, pH8	RT	7 T	3.7	*
Gadobenate Dimeglumine	Water	37 ºC	4.7 T	4.0	[55]
	Water	37 ºC	3 T	4.0	[55]
	TRIS 50mM, pH8	37 ºC	7 T	3.6	*
Gadoterate Meglumine	Water	37 ºC	4.7 T	2.8	[55]
	Water	37 ºC	3 T	2.8	[55]
	Human plasma	37 ºC	7 T	2.8	[54]
	Human plasma	37 ºC	7 T	3.2	[53]
	Human blood	37 ºC	7 T	2.8	[52]
	TRIS 50mM, pH 8	RT	7 T	3.5	*

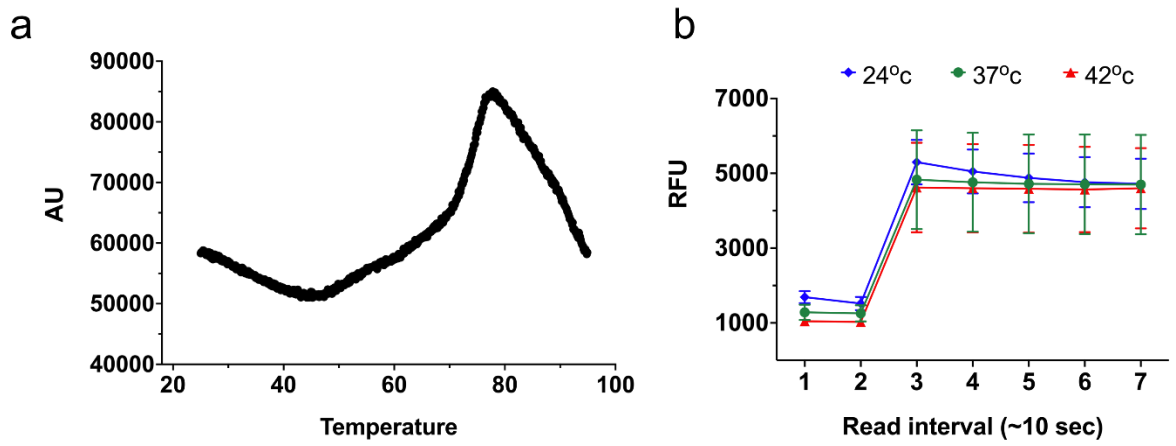

FIGURE 6. Thermal properties of GLamouR. (a) Thermal shift assay was conducted for GLamouR. SYPRO orange dye binds to hydrophobic regions of protein which are increasingly exposed upon denaturation, until aggregation takes place. (b) Performance of GLamouR in different temperatures show minimal variance across tested conditions (n=5 wells).

Macrocyclic chelates have been proven to be more stable than linear chelates, and thus considered safer to use[56]. Although GBCAs are used immediately upon opening in clinical practice, it is not uncommon for researchers to keep bottles of GBCAs for prolonged periods of time for *in-vitro* studies.

In Figure 7a, various contrast agents that had been open for an unknown period of time were tested for their stability. GLamouR was used to detect unchelated free gadolinium from unstable chelates that had released gadolinium. As expected, macrocyclic agents had less free gadolinium than linear configurations, with Gadopentetate Dimeglumine triggering the highest fluorescence increase (which is currently discontinued for use in the clinic). However, upon 24 hours of UV irradiation, the GBCAs had similar amounts of free gadolinium in solution regardless of configuration (with the exception of Gadopentetate Dimeglumine, which was much more susceptible to UV breakdown than other chelates).

To emulate urine samples from GBCA administered MRI patients, artificial urine was spiked with the most stable GBCA of the group (Gadoterate Meglumine) to a concentration of 2.25 mM, equivalent to a dose for a 50 kg patient with 90% of GBCA eliminated in 2 L (Figure 7b) for conservative calculation. Samples were then pre-treated by removing phosphates which compete with GLamouR for gadolinium binding, prior to UV irradiation. Once gadolinium had been released from their chelates, GLamouR was introduced to the solution for binding the released ions, which was then subsequently dialyzed to exclusively obtain gadolinium-protein conjugates. After the protein is denatured to release gadolinium, the solution can then be filtered so the final product is gadolinium in TRIS buffer. Figure 7c shows that 37% of total gadolinium was bound to

GLamouR with partial removal of phosphates, whereas 69% was bound when they were removed more efficiently, showing that pre-treatment of samples is an important factor for effective extraction.

Based on data represented in Figure 7a, we have found that gadolinium is most likely released from their chelates once GBCAs are excreted into the environment, especially in circumstances involving prolonged UV irradiation from the sun. However, we have also demonstrated that gadolinium can potentially be collected from GBCA-administered MRI patient urine samples before they get excreted into the environment, which can be a preventative measure for such pollution.

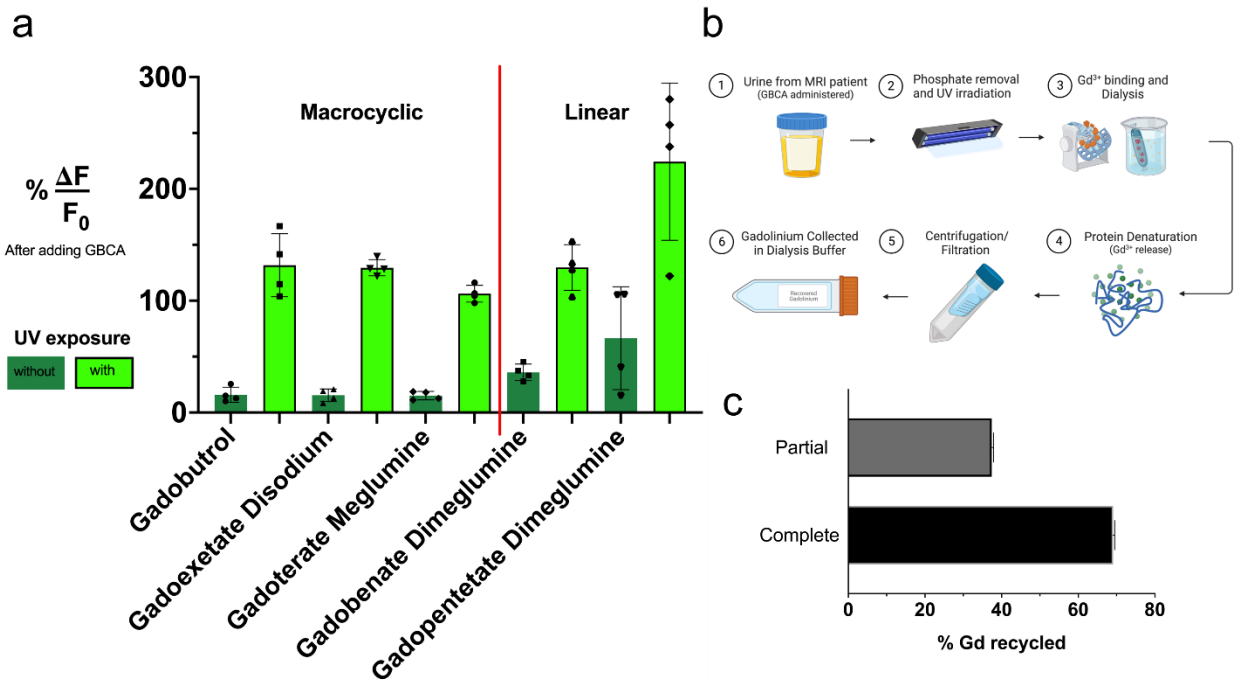

FIGURE 7. **Collection of gadolinium from clinical GBCAs.** (a) Clinical GBCAs were tested for unchelated gadolinium ions with GLamouR, before and after UV irradiation. (n=4 individual wells) (b) Schematic showing the proposed procedure for collecting gadolinium from MRI patient urine samples, which was substituted for Gadoterate Meglumine-spiked artificial urine in this study. (c) Recycled gadolinium after partial removal and complete removal of phosphates from Gadoterate Meglumine-spiked artificial urine samples. For complete removal, Gadoterate Meglumine was added after passing artificial urine through phosphate removal resin, whereas partial removal was emulated by adding Gadoterate Meglumine before passing through the column with a lower resin ratio. Mean values were calculated by averaging out data that was acquired at 5 different wavelengths associated with gadolinium.

RECYCLING LOW CONCENTRATIONS OF GADOLINIUM WASTE WITH AFFORDABLE TECHNOLOGY

Gadolinium has been reported to exist at the picomolar scale in various polluted bodies of water, even including tap water[12]. To demonstrate the possibility of gadolinium collection from environmentally relevant concentrations, gadolinium was introduced into 1 L of TRIS buffer to a concentration of 200 pM. GLamouR was then used to collect the gadolinium, which was subsequently concentrated down to 100 µL. Final gadolinium concentration was measured via ICP-MS and total number of moles was calculated to be 155.17 picomoles (SD=4.43, n=4 measured isotopes), equivalent to 77% of the initial number of moles of gadolinium. Proper pre-treatment is indeed important to obtain a high percentage of collection, as seen in Figure 7c. However, it is demonstrated in Figure 8a that low concentrations of gadolinium do not pose a hurdle in collecting or filtering anthropogenic gadolinium from water.

This process involves multiple filtration steps, many plastic consumables, and a centrifuge. We therefore designed a device for collecting and detecting gadolinium which further reduces such labor and need for additional resources. Figure 8b depicts a device setup in which the sample would flow through a pre-treatment chamber for removing competing compounds such as phosphates, which may then also include UV-irradiation to break any pre-existing bonds that could interfere with gadolinium collection. GLamouR column would be placed inside the detection unit, where increase of fluorescence would indicate remaining gadolinium, and saturation of fluorescence would indicate the need for a fresh column. The flow of sample would be continuously repeating until fresh columns do not increase in fluorescence, and remaining waste could then be redirected for

disposal. The existing columns could then be further processed to release gadolinium as depicted in Figure 7b, steps 4-6. The advantage of this system would be that extremely low concentrations of gadolinium could be concentrated to higher concentrations and subsequently eluted for isolating gadolinium.

The prototype device was constructed with a budget for less than 500 USD (Figure 7c). The outer shell was 3D-printed, with the bulk of the cost coming from the Arduino microcontroller and Hamamatsu spectrometer. Samples can be pumped in and out of the device by creating a pressure differential inside a bottle containing the sample by utilizing an air pump inflator typically found on sphygmomanometers. Figure 7d shows data collected from the device, where gadolinium concentration was brought up to 100 μM by injecting a 1 mM solution of gadolinium equaling $1/10^{th}$ of the final volume.

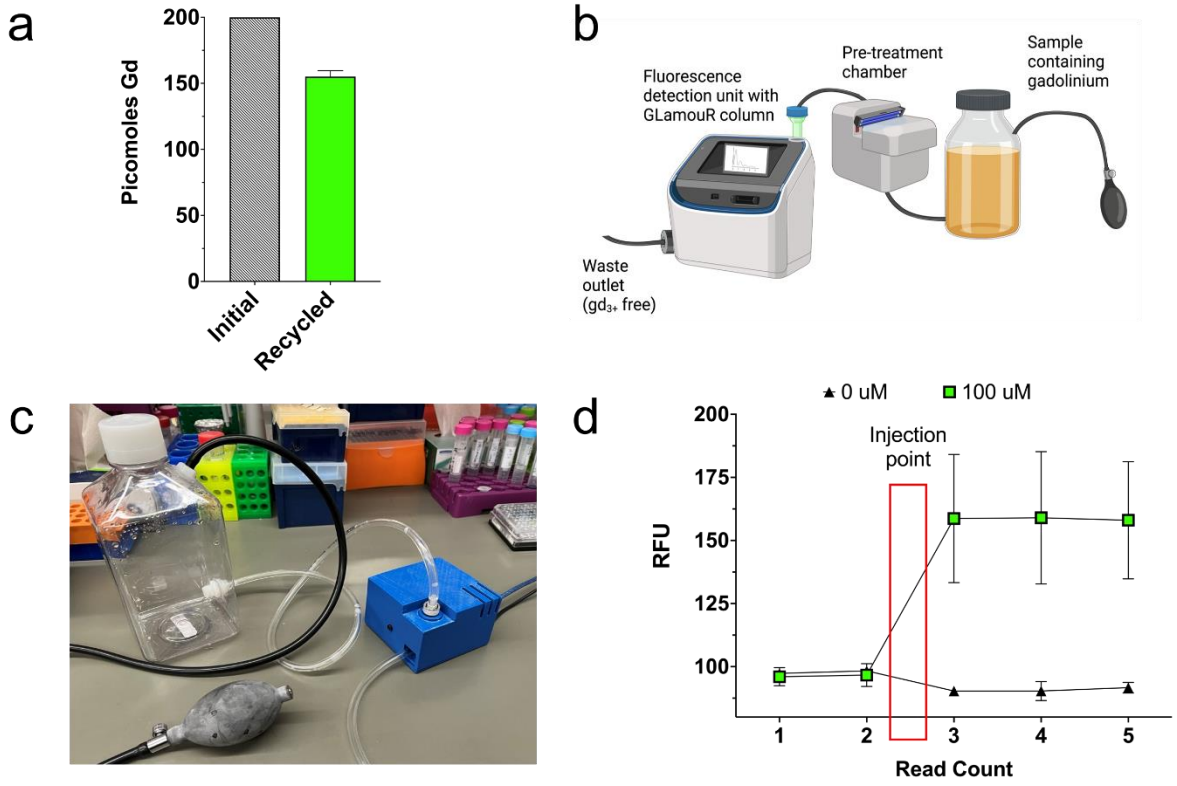

FIGURE 8. Collection and concentration of picomolar concentrations suggest GLamouR can be used to recover gadolinium from natural water sources.

(a) Gadolinium scavenging capabilities at 200 pM (measured with a multimodal plate reader). Gadolinium was gathered and subsequently concentrated down to nanomolar concentration while retaining 77% of total gadolinium mass. (b) Schematic diagram showing the proposed system for mining gadolinium with affordable technology. (c) Photo of the actual prototype built in the lab, to demonstrate the simplicity of the technology and affordable price point. (d) Data produced with the prototype device for gadolinium detection.

IDENTIFICATION AND COLLECTION OF REES BEYOND GADOLINIUM

Based on Figures 2-8, GLamouR is a novel candidate protein for the detection, quantification, and recycling of gadolinium waste. However, gadolinium is not the sole REE pollution resulting from industrial activities[57], and thus further experiments were carried out to investigate the protein's responses to different REEs. As seen in Figure 9b, GLamouR was effective in detecting 11 REEs and distinguishing them from calcium. Change in fluorescence was not particularly unique for different REEs, as most of them lingered around the 100% region.

A red-shifted GLamouR was then created (Figure 10) to see if REEs could be identified due to unique REE responses when used in conjunction with green. While the GLamouR-rs has a much lower brightness in both its on and off states than the original green version, it produces a much higher delta across all REEs with the exception of Lanthanum (Figure 9), where response was recorded to be very minimal (+8.26%, SD=0.68). In this case, green fluorescence could be measured to quantify the REE, while the lack of fluorescence increase in red could specify the REE to be Lanthanum.

The creation of GLamouR proteins with unique responses to different REEs suggest the possibility of future GLamouR variants that bind with a stronger affinity or react more efficiently to specific REEs. This would make it feasible to utilize the system depicted in Figure 8c for collecting and isolating specific REEs from environmental samples that contain a mixture of several REEs, as it is often the case in nature. Even with moderate specificities, low concentration mixtures could be accumulated to high enough concentrations for downstream separation to achieve higher yields for cost effectiveness with minimum impurities of undesirable REEs.

In addition to mining and recycling specific REEs, it is hoped that GLamouR variants will allow users to determine REE composition and identify REE-rich sources in a rapid, affordable, and simplistic manner. Although urine samples from GBCA-administered patients have a pre-defined matrix with a guaranteed range of gadolinium concentration, this would not be the case for samples with diverse and unique compositions for every batch, such as electronic waste or samples obtained from the environment.

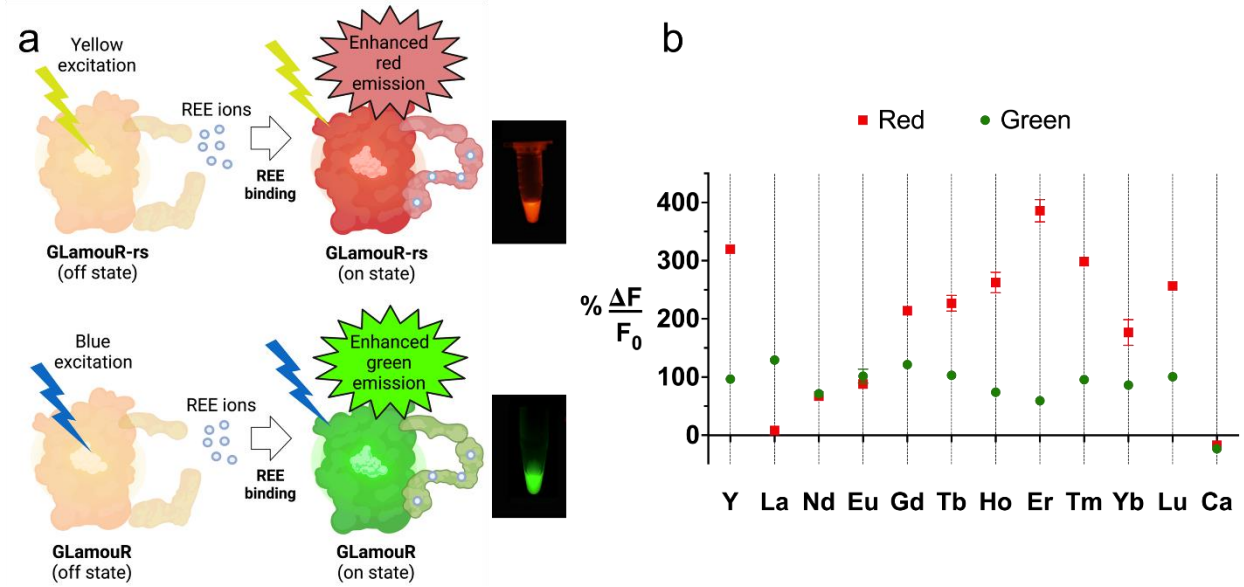

FIGURE 9. REE binding with GLamouR-rs and GLamouR. (a) Schematic of the red-shifted GLamouR (GLamouR-rs) and original GLamouR binding REEs, with their corresponding fluorescence images to the right. (b) Fluorescence increase of green and red shifted GLamouR upon addition of eleven different REEs show a difference in their preference for certain REEs such as lanthanum (calcium added as negative control).

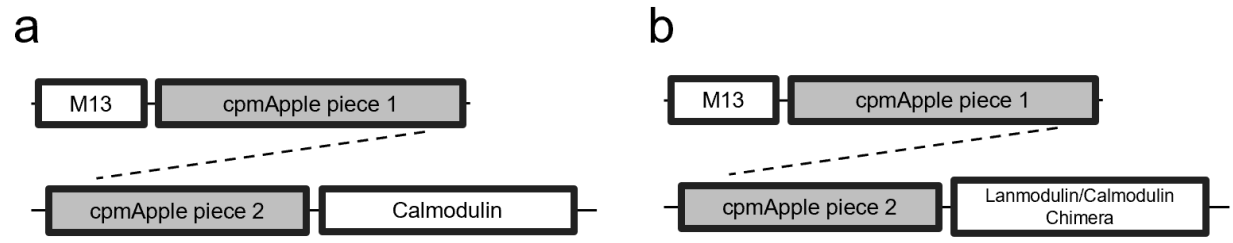

FIGURE 10. Protein maps for rGECO1 and GLamouR-rs. Similar to the original GLamouR, the protein design for GLamouR-rs is based on a well-established calcium-sensing protein, but with an emission peak in red. The resulting protein has lost its strong affinity for calcium and has developed a specificity for REEs. (a) Protein map for the rGECO1. (b) Protein map for the GLamouR-rs, where the calmodulin domain of rGECO1 has been replaced by the lanmodulin/calmodulin hybrid.

2.4 FUTURE DIRECTIONS

In this study, we have demonstrated gadolinium recycling from artificial urine at a physiologically accurate GBCA concentration, and collection and concentration of gadolinium from picomolar to nanomolar while retaining roughly 70% of total gadolinium mass for both instances. We have also shown that gadolinium can be detected with a simple setup including a spectrometer, LED light source, and a microcontroller inside a 3d-printed housing for less than 500 USD, making the technology widely accessible and scalable. GLamouR has been evaluated to be a highly stable protein by retaining function for the entirety of a week-long assay at room temperature (Figure 3b), while being compatible with lyophilization and high centrifugal force.

Once established on Earth, it is anticipated that GLamouR-like proteins could be used in outer space for extraterrestrial biomining. The near side of the Earth's moon is reported to possess massive reserves composed of a mixture of Potassium, REEs, and Phosphorus (KREEP), estimated to be 220 million cubic kilometers[58]. A recent study by Cockell et.al., demonstrated extraterrestrial biomining of REEs where microorganisms were tested for their mining capabilities from basaltic rock in various micro-g environments on the international space station[59]. As it is for any type of travel but especially air and space travel, reducing cargo weight is essential for increasing distance per unit fuel. Compatibility with lyophilization to reduce weight during takeoff, simple mining procedures with light and affordable equipment, and instant feedback during REE collection could be desirable characteristics of GLamouR for REE mining in outer space.

The use of biological tools to carry out functions otherwise achieved through chemical, electrical, or mechanical means are often met with higher performance and

cheaper production costs, such as the modern development and production of recombinant human insulin in E. coli[60]. Furthermore, biological therapies are the best-selling drugs on the market today[61], and new products that mimic (biosimilars) and even improve upon the precision and efficacy of traditional medications (biobetters) are being developed as well[62]. Such technologies involving biological tools are expected to be tailored for their respective applications in each field, whether it be industrial practices, agriculture, precision medicine, or even unearthing new scientific discoveries.

3. CHAPTER 2) DEVELOPMENT OF NON-IMMUNOGENIC CEST MRI PROTEINS

3.1 INTRODUCTION

In 2007, a landmark paper in CEST MRI was published in Nature Biotechnology[63] by Gilad et al., highlighting the application of an entirely genetically encoded contrast agent for non-invasive cell tracking via synthetic biology and molecular imaging. CEST MRI begins with selectively saturating exchangeable protons (hydrogen nuclei of a contrast agent) with a radiofrequency pulse, and relies on the dynamic exchange of saturation with its surrounding water protons[64]. Contrast can be enhanced by increasing the number of exchangeable protons and/or facilitating quicker exchange. Unfortunately, most CEST MRI proteins to date have been relying on possessing as many exchangeable protons as possible, leading to a high content of basic amino acids in the protein.

An advantage of genetically encoded, protein based CEST MRI contrast agents would be that the signal does not dilute over cellular replication[65]. For example, iron-oxide nanoparticle labeling can be diluted down to 50% of its original contrast after one round of cell division as they does not pass the contrast agent down to their daughter cells, whereas genetically encoded systems possess the capability to do so. This makes genetically encoded contrast agents useful for tracking cells even after several division cycles.

However, one hurdle for any protein-based therapy or molecular imaging-based diagnostic aid is immunogenicity[66]. Proteins recognized as foreign get targeted and rejected through the immune system, which poses a threat to applications that require longer periods of time. Any contrast of therapeutic effect the protein would otherwise be

providing would disappear as proteolytic activity takes place, breaking down the protein into harmless units of cellular waste.

To address the issue of possible immunogenicity or cell toxicity for long-term studies, a natural protein known as HPRM1 was assessed for its contrast capabilities under the assumption that a protein of human origin would elicit less of an immune response when compared to a synthetically designed protein with no sequence homology to anything found in nature. Although the CEST MRI signal was proven to be significant[67], the immunogenic properties of this natural protein were never assessed. This raised the question of whether genetically encoded CEST contrast proteins could benefit from an adjusted immunogenicity.

A stealth option where the protein can evade the immune system for a prolonged period of time can be imagined for long-term contrast, whereas a super immunogenic protein can be used for providing short-term contrast while ultimately eliciting immune activity near the targeted environment. In general, innate immune response to peptides has been ignored in the process of this project as historically it has been somewhat difficult to synthesize immunogenic peptides towards the innate immune response. In order to make these peptides, they must have a very specific sequence that binds to receptors called Pattern-Recognition Receptors (PRRs)[68], which are coded in genes that have been passed down through many generations. It was therefore evaluated to be difficult to create peptides that have an immediate innate immune response, and the focus was shifted into creating non-immunogenic peptide sequences with CEST contrast. To this end, millions of sequences were tested with IEDB T-cell epitope predictor tools to develop sequences with low binding affinity to MHC-I and MHC-II, under the premise that

if peptides cannot bind to either target, they would be unable to activate T-cells through T-cell receptor/MHC interactions to make an adaptive immune response.

3.2 MATERIALS AND METHODS

MATERIALS

Reagents

All materials were purchased from Thermo Fisher or its subsidiaries, with the exception of centrifugal filtration units from Merck. Human blood was purchased through the Gulf Coast Regional Blood Center.

Equipment

- Flow Cytometer: Attune NxT
- Custom-built server (specs unknown) for computational calculations.

Software

Prism (GraphPad) was used for generating graphs and performing statistical analysis.

METHODS

PBMCs were obtained from human blood and incubated for 12-18 hrs with peptides, proteins, and viruses. The PBMCs were then subsequently screened with Flow Cytometry to identify surface markers. Other methods and strategies for deriving non-immunogenic protein sequences for CEST MRI are described within the results section of this chapter.

3.3 RESULTS

The objective was to create a peptide string of 12 amino acids long (which is divisible by 2,3,4, and 6 while being long enough for diverse compositions, but not excessively burdensome in terms of computing power) with at least 33% lysine content to ensure CEST contrast, while avoiding any possible MHC binding. Based on investigation with strings comprised of single amino acids, a general trend towards certain amino acids were discovered, but when put into a different context with other neighboring amino acids, the binding scores were slightly shifted for that amino acid. It was also evident that having a diverse spread of many amino acids not only provided additional weak points (high binding windows) throughout the chain, but also increased computing time when determining the best permutation, or ordering, of amino acids. Acidic amino acids tended to improve solubility and also decrease likelihood of MHC binding, so another 33% was dedicated towards exploring the use of aspartic acid and glutamic acid, and finally, amino acids of neutral charge and lower binding tendencies were also evaluated. The resulting peptides were originally designed to be repeats of a 6 consecutive amino acid chain, so that they could be evaluated in windows of 9 or 10 amino acids, as it is known to be how antigens are presented. Ideally, the peptides should be designed to be repeats of 10 consecutive amino acids so that peptide sequences could be linked together in succession to produce proteins of unlimited length with predetermined binding affinities for every possible location. However, due to limited computing capacity at the time, proof-of-concept was established for 6 amino acid repeats first, and later on a novel 10 amino acid repeat with a satisfactory score was discovered. The best sequence derived from human trial and error (conducted throughout the summer by an intern student) for a 10 amino acid repeat was KKKEDKKKED, which had a

minimum MHC I percentile rank (score) of 4.95. Herein, we found not only a better composition of amino acids, but also hundreds of better permutations for the same composition, and thousands of permutations for different compositions. In addition to lysine-rich proteins, the same process was repeated for arginine-rich proteins and histidine-rich proteins, although only 12-mers were evaluated, and variance was kept at a minimum of three unique residues. Further investigation is needed in order to explore every possible protein permutation composed of amino acids of desired properties, but this study has shown that it is possible to derive the highest binding scores for each permutation of a given composition. After designing the peptides and protein sequences in-silico, the next step would be to test *in-vitro*. The presence or absence of specific surface markers can be used to gauge potential immunogenicity. For example, CD25 and CD69 can be probed for T-cell activation[69], CD107a and FasL for cytotoxic activity[70], INF-gamma for cytokine assessment, and 45RO for assessing memory T-cell population. In addition, technology such as Luminex xMAP can provide a multiplexed assessment of PBMC cytokine release profiles such as IFN-gamma, TNF-alpha, IL-2, IL-10, or TGF-beta.

A sample of the results is shown in Table 2. For a 12-mer peptide with five unique residues and 33% lysine content, a score of 30.5 was achieved with the sequence KQDNKEKQDNKE. For an unlimited length protein with 50% lysine content, the maximum score was 15. A SCORE greater than 1 (percentile rank of 1%) is considered to pass immunogenicity testing[71], and several thousands of permutations that could be non-immunogenic have been generated. Table 2 cuts the list short to a percentile rank of 10% or above.

TABLE 2. Sample of derived sequences

12-mer peptide			Unlimited length protein		
ID	**SEQ**	**SCORE**	**ID**	**SEQ**	**SCORE**
>aa248	KQDNKE	30.5	>aa356	DQKKKDDKQK	15
>aa12	KKNDQE	29	>aa475	KDDKQKDQKK	15
>aa4	KKNQDE	25	>aa535	DDKQKDQKKK	15
>aa85	QEKKND	25	>aa817	QKDQKKKDDK	15
>aa117	NKKQDE	24	>aa1432	KKKDDKQKDQ	15
>aa358	KKQNED	23	>aa1631	KQKDQKKKDD	15
>aa48	NKEQDK	22	>aa1997	QKKKDDKQKD	15
>aa26	QDNEKK	21	>aa2006	KDQKKKDDKQ	15
>aa87	EKKDNQ	21	>aa2127	DKQKDQKKKD	15
>aa143	DKKNQE	21	>aa2514	KKDDKQKDQK	15
>aa201	QKEKDN	20.5	>aa106	QDKDQKKKKD	12
>aa90	QEDKKN	19.5	>aa1167	DKDQKKKKDQ	12
>aa212	EDNQKK	19.5	>aa1603	QKKKKDQDKD	12
>aa231	QEKDNK	19.5	>aa1630	KKDQDKDQKK	12
>aa347	QKKNDE	19	>aa1897	KKKKDQDKDQ	12
>aa167	EQKKDN	18.5	>aa1956	KDQDKDQKKK	12
>aa221	KEQDKN	18.5	>aa2055	KDQKKKKDQD	12
>aa258	KNQKED	18.5	>aa2344	DQKKKKDQDK	12
>aa353	QNDKKE	18.5	>aa2425	DQDKDQKKKK	12
>aa54	DQNKEK	18	>aa2441	KKKDQDKDQK	12
>aa199	KKEDNQ	18	>aa92	KKQKDQKKDD	11.45
>aa153	EQNDKK	17.5	>aa342	KKDDKKQKDQ	11.45
>aa30	NDQKEK	17	>aa352	KDQKKDDKKQ	11.45
>aa239	NKEKQD	17	>aa824	QKKDDKKQKD	11.45
>aa131	EKNQDK	16.5	>aa969	KQKDQKKDDK	11.45
>aa310	QKNKED	16.5	>aa1135	DQKKDDKKQK	11.45
>aa31	QNKEDK	16	>aa1406	DDKKQKDQKK	11.45
>aa63	QNDEKK	16	>aa1914	QKDQKKDDKK	11.45
>aa103	KNEDQK	16	>aa2109	KDDKKQKDQK	11.45
>aa150	KKNDEQ	16	>aa2191	DKKQKDQKKD	11.45
>aa276	KQNKED	16	>aa432	KKKDDKKDQQ	11
>aa277	KQNDEK	16	>aa1292	KKDQQKKKDD	11
>aa45	ENQKKD	15.5	>aa1360	DDKKDQQKKK	11

TABLE 2. (cont'd)

12-mer peptide			Unlimited length protein		
ID	**SEQ**	**SCORE**	**ID**	**SEQ**	**SCORE**
>aa270	KEKDQN	15.5	>aa1718	DKKDQQKKKD	11
>aa357	KDEKQN	15.5	>aa1727	DQQKKKDDKK	11
>aa28	KDNEQK	15	>aa1819	KDQQKKKDDK	11
>aa33	QDNKKE	15	>aa2118	KKDDKKDQQK	11
>aa59	KDNKEQ	15	>aa2403	QKKKDDKKDQ	11
>aa92	DQKKNE	15	>aa45	KKQDKQKKDD	10.6
>aa261	NDKKQE	15	>aa77	KQDKQKKDDK	10.6
>aa269	KNEDKQ	15	>aa414	DKKQDKQKKD	10.6
>aa283	NQKDEK	15	>aa507	KQKKDDKKQD	10.6
>aa319	DQNKKE	15	>aa914	QKKDDKKQDK	10.6
>aa345	QNKDEK	15	>aa977	DDKKQDKQKK	10.6
>aa246	DNEQKK	14.6	>aa1188	KDDKKQDKQK	10.6
>aa287	DKQKNE	14.6	>aa1507	DKQKKDDKKQ	10.6
>aa66	KDKNQE	14	>aa1643	KKDDKKQDKQ	10.6
>aa68	KKDENQ	14	>aa2077	QDKQKKDDKK	10.6
>aa172	DQKENK	14	>aa65	KDQDKKKQDK	10.45
>aa177	QENDKK	14	>aa323	KKQDKKDQDK	10.45
>aa216	KKNEDQ	14	>aa498	KQDKKDQDKK	10.45
>aa360	KKQDNE	14	>aa732	KKKQDKKDQD	10.45
>aa94	DNKEKQ	13.5	>aa753	DKKDQDKKKQ	10.45
>aa278	QEDKNK	13.5	>aa758	KKDQDKKKQD	10.45
>aa49	DKKQNE	13	>aa1382	DKKKQDKKDQ	10.45
>aa130	NDQKKE	13	>aa1975	QDKKKQDKKD	10.45
>aa179	EKKNQD	13	>aa2049	DQDKKKQDKK	10.45
>aa43	DNKEQK	12.9	>aa2348	QDKKDQDKKK	10.45
>aa164	KNKEQD	12.9	>aa174	KDDKKQDQKK	10.35
>aa195	KDQENK	12.9	>aa274	DDKKQDQKKK	10.35
>aa266	QKNEKD	12.9	>aa527	QKKKDDKKQD	10.35
>aa268	KDQNEK	12.9	>aa618	DQKKKDDKKQ	10.35
>aa348	NKEDQK	12.4	>aa937	KKDDKKQDQK	10.35
>aa356	KEDQKN	12.4	>aa970	KKKDDKKQDQ	10.35
>aa55	KDQNKE	12.25	>aa1124	QDQKKKDDKK	10.35
>aa108	KEDQNK	12.25	>aa1172	KKQDQKKKDD	10.35

TABLE 2. (cont'd)

12-mer peptide

ID	SEQ	SCORE
>aa314	EKNDQK	12.25
>aa359	QNKKED	12.25
>aa105	KNDEQK	12
>aa291	KQEDNK	11.75
>aa62	KKQEND	11.65
>aa128	KDNEKQ	11.65
>aa303	DKKNEQ	11.65
>aa320	DEKKQN	11.65
>aa209	DKQNEK	11.6
>aa271	NKKQED	11.6
>aa32	EKDQKN	11.45
>aa51	DNKKEQ	11.45
>aa218	DNEKKQ	11.2
>aa302	QNDKEK	11.2
>aa321	NEQDKK	11.2
>aa354	KKNEQD	11.2
>aa67	QNKKDE	11.15
>aa24	DKENQK	11.05
>aa91	KQNDKE	11.05
>aa171	QEKKDN	11.05
>aa223	NKEKDQ	11.05
>aa217	KENQDK	10.9
>aa292	KQDNEK	10.9
>aa99	EKQDNK	10.85
>aa244	EKNKDQ	10.85
>aa140	NDKKEQ	10.8
>aa6	EDKKQN	10.45
>aa46	KKDNEQ	10.45
>aa134	KKENQD	10.45
>aa147	NEDKKQ	10.45
>aa245	NKDKQE	10.2
>aa254	DENKQK	10.2
>aa355	NKDEQK	10.2
>aa295	KQNEKD	10

Unlimited length protein

ID	SEQ	SCORE
>aa1533	KQDQKKKDDK	10.35
>aa1807	DKKQDQKKKD	10.35
>aa165	KKQDDKQKKD	10.3
>aa776	QKKDKKQDDK	10.3
>aa783	DKKQDDKQKK	10.3
>aa1068	KQDDKQKKDK	10.3
>aa1313	KQKKDKKQDD	10.3
>aa1608	KDKKQDDKQK	10.3
>aa1723	KKDKKQDDKQ	10.3
>aa1810	DDKQKKDKKQ	10.3
>aa17	QDQKKKKDDK	10
>aa36	DDKQDQKKKK	10
>aa169	KQDQKKKKDD	10
>aa762	KKKDDKQDQK	10
>aa870	KKDDKQDQKK	10
>aa1151	DQKKKKDDKQ	10
>aa1436	QKKKKDDKQD	10
>aa2114	DKQDQKKKKD	10
>aa2168	KDDKQDQKKK	10
>aa2375	KKKKDDKQDQ	10

3.4 FUTURE DIRECTIONS

The next steps for this study would be testing the *in-silico* designed peptides *in-vitro* as seen in Figure 11 or assessing cytokine release profiles with xMAP. Peptides in the order of high to low SCORE should be tested for immunogenicity and a CEST MRI scan can be performed to verify contrast. Furthermore, tools such as the POET[72] can be utilized to pick the highest contrast peptide/proteins beforehand, to prioritize candidates for *in-vitro* evaluation of CEST contrast and subsequent immunogenicity testing. Once a satisfactory candidate has been identified, the motif can be cloned into a plasmid to translate a long protein of high molecular weight, which should also be evaluated. As seen in the original LRP paper by Gilad et al., transducing a cancer cell line and implanting the xenograft inside rat brain would be ideal for demonstrating feasibility for *in-vivo* imaging in terms of contrast, but not immunogenicity. After conducting *in-vitro* assessments of immunogenicity, the final stage would be creating a stable cell line that produces the protein at a high enough concentration, and then monitoring its distribution inside a mouse model that is compatible with our human-origin xenograft initially but possesses an intact immune system after successful transplant of the xenograft. This has been achieved by injecting CD34+ HSCs[73,74] or PBMCs[73] into immunocompromised mice as shown in the humanized NSG mice of Jackson Laboratories, or through blockade of T-cell activation during inoculation of tumors in immunocompetent mice[75].

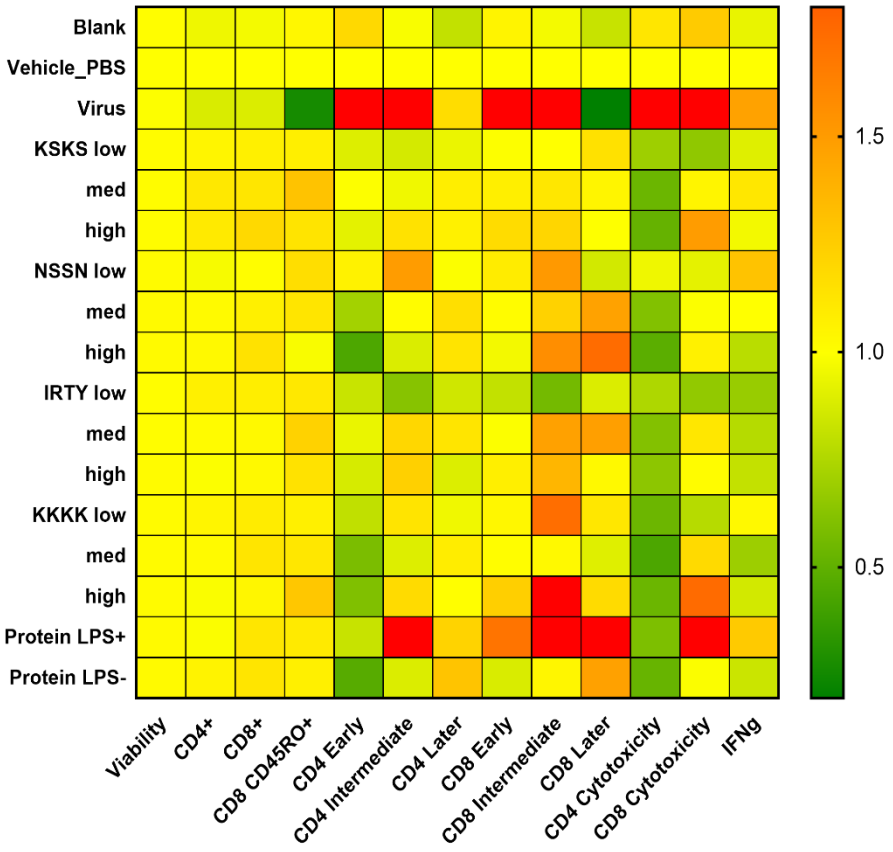

FIGURE 11. Heatmap representation of *in-vitro* immunogenicity screening. Different concentrations of the peptides and HPRM1 were assessed at three different levels: low, mid, and high, which were set at a final well concentrations of 8.125 µM, 16.25 µM, and 32.5 µM respectively, and plotted as a heat map on a relative scale from 0 to 2. Results did not show any immunogenic peptides at 8.125 µM concentration and should be increased for future studies (at minimal concentrations required for CEST MRI detection). Positive control (denoted “Virus”) and HPRM1 without endotoxin removal were the only groups considered to be immunogenic in this dataset, although the concentration of LPS negative HPRM1 is unknown and should not be considered non-stimulatory without further investigation.

4. CHAPTER 3) ACCURATE QUANTIFICATION FOR SYNTHETIC PROTEINS

4.1 INTRODUCTION

Protein quantification is arguably one of the most important, yet trickiest things to obtain at high accuracy without a sample of the exact same protein at a known mass/concentration for comparison. Even then, the accuracy of the value for the known sample is questionable as there are several different methods for measurement, none of which are foolproof [76].

For natural proteins, this problem is less pressing than it is for synthetic proteins, as the abundance of amino acids in naturally occurring proteins is somewhat similar, therefore a calibration curve produced with BSA standards might suffice. For synthetically designed proteins, however, the abundance may not necessarily fall within the estimated range; for example, the rdLRP above is comprised of 57 lysines that constitute more than 55% of the entire sequence[77], which is unprecedented in nature. The mechanism of which current protein quantification assays are performed mostly rely on the existence of specific amino acids within the protein sequence. For example, the Bradford Assay[78] is largely dependent on the presence of certain basic amino acids such as arginine, lysine, and histidine and is known to be approximately proportional to the number of positive charges found on the protein of interest. Therefore, a sequence-specific approach is highly recommended for proteins of synthetic origin, which may not contain the amino acids required or do not necessarily possess a natural distribution of amino acids in their sequence.

An example of such approach would be the study conducted by Anthis et al. in 2013 [79]. Absorbance with UV at 280 nm relies on the presence of tryptophan or tyrosine

residues, but a shift down to 205 nm provides the advantage of being able to measure absorptivity values primarily from the peptide bond. Because of the fact that absorbance from some the aromatic side chains (such as tryptophan, tyrosine, phenylalanine, and histidine) remain significant as well as minor contributions by arginine, cysteine, and cystines, a sequence-specific approach was adopted that calculates the molar absorptivity values of each component. This can then be calculated into a single molar extinction coefficient with the given formula, where ε is molar absorptivity value, i is for each individual amino acid, n is the number of times they appear in the sequence, bb is backbone, and r is the length of the protein.

$$\varepsilon_{205} = \sum(\varepsilon_i n_i)\varepsilon_{bb}(r-1)$$

In the following study, a sequence-specific approach similar to adjusting the molar extinction coefficient when measuring proteins with ultraviolet wavelengths was adopted, but with the Bradford assay. Arginine, lysine, and histidine residues in the protein primary structure were counted for the protein of interest and a ratio to BSA was derived and applied to the final readout of the Bradford assay, which is performed with BSA standards.

4.2 MATERIALS AND METHODS

MATERIALS

Reagents

All materials were purchased from Thermo Fisher or its subsidiaries, with the exception of centrifugal filtration units from Merck.

Equipment

- Absorbance measurements: Cytation 5 (Biotek), Spark (Tecan), Nanodrop Spectrometer

Software

Prism (GraphPad) was used for generating graphs and performing statistical analysis.

METHODS

Absorbance measurements

Wells were prepared in triplicates with 5µL of protein and 250µL of Coomassie Reagent. Wells were incubated with shaking for 10 minutes at room temperature. Absorbance was then measured with the Cytation5 (Biotek) and/or Spark (Tecan) at a wavelength of 595 nm. Blank well averages (wells without protein) were subtracted from each measurement, and a standard curve was plotted to calculate corresponding protein concentration.

Absorbance at 205 nm was performed with the Nanodrop Spectrometer by placing a single drop on the pedestal and obtaining a read with the protein's corresponding molar extinction coefficient. Each read was performed three times and averaged to obtain a mean value.

4.3 RESULTS

Myoglobin (from equine heart) and ovalbumin were adopted for the assay, as myoglobin is compatible with Bradford, whereas ovalbumin is not due to its small molecular weight of less than 3kDa. 1mg/mL solutions were prepared by starting with 1mg of dry protein which were then dissolved into 1mL of DDIW. With a UV spectrometer, sequence-specific absorbance was measured and then compared to conventional Bradford, as well as the new “corrected” method. As seen in Figure 12, the corrected method seems to work best for this particular protein. As expected, the Bradford assay was not compatible with ovalbumin and no change in absorbance was observed. The UV spectrometer readout for ovalbumin was also 0.66 mg/mL, which demonstrates that proteins require unique approaches for accurate measurements. Additionally, the protein of interest should be dialyzed against the same buffer as the standards after reconstitution. This process will eliminate any unwanted salts, glycerol, other stabilizing agents, or impurities to replicate the conditions for each protein and minimize variance.

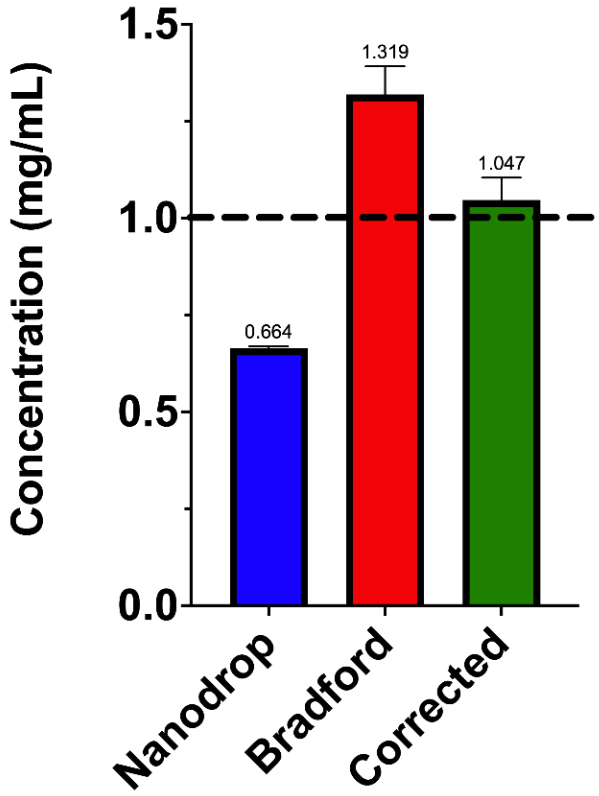

FIGURE 12. Myoglobin measured at 1mg/mL with various methods. The new "corrected" method with a sequence-specific approach utilizing the Bradford assay was the best performing.

4.4 FUTURE DIRECTIONS

Measuring the weight of the lyophilized form of an ultrapure batch of protein would be a straightforward way to quantify proteins; however, the sample would have to be completely dry, and any salts resulting from the buffer would have to be taken into account. In addition, the process would take a couple of days, or even weeks, and require a relatively large sample volume. Further investigation is required to fully develop a method for a wider range of proteins, and it is likely that the primary/tertiary structures of the protein would be evaluated prior to executing quantification. A one-size-fits-all approach would be desirable but unlikely, and it is estimated that specific assays would be performed based on protein makeup, with different correction methods for different characteristics, with some being compatible with more than one method.

Another hurdle for accurate protein quantification is the purity of the protein sample. In many cases, proteins are not purified to the extent of the high standards that can be seen for industrial applications that require medical grade purity. In those instances, yield can sometimes be greatly sacrificed for purity because of the importance of batch-to-batch consistency as well as the fact that medicine should be subject to the highest scrutiny in terms of safety. Any misfolded proteins of the same size, sequence, and charge are difficult to separate with conventional practices and these impurities will contribute towards a lower efficacy of the whole batch of protein, as they are counted in the quantification process but do not provide equivalent function.

5. CHAPTER 4) A MOLECULAR IMAGING ADAPTOR PROTEIN

5.1 INTRODUCTION

As a standardized solution for inconsistent pre-treatment diagnostics, there are currently four biologics paired with two different automated immunohistochemistry platforms that have been FDA approved to predict patient response to PD-L1 blockade in cancer immunotherapy[80]. Although these systems eliminate laboratory environmental factors and antibody usage variance, pathologist interpretation and scoring methodologies still present an issue. Response prediction has thus started relying on neo-antigen analysis results by measuring the Tumor Mutational Burden (TMB) through whole exome sequencing and targeted panel sequencing[81,82].

These sequencing techniques are considered to be more accurate in predicting patient response[83] to immune checkpoint blockade, however, the process may take up to 12 weeks[84] depending on the healthcare provider, a timeframe in which the nature of the tumor can be different. Moreover, these methods still require tissue samples derived through invasive biopsies which fail to take into account sampling errors due to the inter- and intra-tumoral heterogeneity of tumors, not to mention temporal variations[85,86]. Liquid biopsies are being used as minimally invasive procedures in which circulating cell free DNA (cfDNA) are derived from blood samples to assess TMB, but these approaches are also limiting due to the fact that the amount of circulating tumor DNA (ctDNA) is a small fraction of cfDNA[87]. It would thus be ideal to evaluate immunotherapy potency with a rather direct method that not only is impervious to sampling error, but also takes into account the pharmaco -dynamics and -kinetics of the drug within the native environment of the biological system, and which can be monitored in a rapid manner.

Molecular imaging can provide real-time imaging in a non-invasive manner while eliminating sampling error. However, current methods of tagging monoclonal antibodies require experts to perform harsh chemical modifications that may alter targeting specificity[14]. Moreover, the tagging itself may be non-specific, where commonly used heterobifunctional crosslinkers will latch onto any amine group of the antibody including the functional region, leading to decreased functionality and stability[88-90]. Even assuming a satisfactory tagging yield, the resulting theranostic is usually compatible for one specific imaging modality such as PET, which requires additional expertise in handling radioactive materials.

The synergistic effect of specific conjugation methods such as protein-protein interactions and cross confirming target localization through multiple imaging modalities[91] is anticipated to solve such issues. With the proposed Universal Theranostic Adaptor (UTA) technology, the ability to elucidate key components of the tumor microenvironment will be easily achievable with simple, gentle, and rapid tagging of a single theranostic adaptor, compatible with almost any type of imaging modality with the versatility of proteins and lanthanides. Proteins are the essence of biological function as we can observe in ligand binding and oxidation-reduction reactions throughout biological systems, and lanthanides possess unique physical properties which in turn allow us to utilize them in a wide variety of precedent applications that include biomedical imaging. With a mixture of unique protein sequences and lanthanides, multimodal imaging can be easily achieved and utilized by researchers in various fields.

For example, gadolinium is the most commonly used lanthanide in imaging, used in MRI contrast agents with more than 30 million injections a year[92]. Europium can be

found as a security measure in Euro banknotes emitting a high and narrow peak of long-lived luminescence in the NIR range with resistance to photobleaching, suggesting their use for deep tissue imaging[93]. Lutetium can be found in the market as ^{177}Lu dotatate, which has been FDA approved for gastroenteropancreatic neuroendocrine tumors. As an attachable theranostic, ^{177}Lu has been conjugated to anti-PSMA as a form of successful radioimmunotherapy for metastatic castration-resistant prostate cancer in a phase II clinical trial[94] and has also been FDA approved as of March 2022[95]. Proteins have also been steadily used as imaging agents. Bioluminescent reporters such as the firefly luciferase and NanoLuc are used in optical imaging to produce light[96]. The connection between such imaging agents and biological therapies can be of two different ways: Direct fusion (amino acid sequence incorporation), or the use of specifically binding proteins. Derived from the bacterial species *S. Aureus*, Protein A binds onto the FC regions of antibodies through simple protein-protein interaction[97], allowing us to tag a wide variety of mAbs used as biological therapeutics without hindering targeting capabilities. There are numerous aspects of innovation involved with this study. First, quick, and easy tagging of any mAb based biological therapy will be achievable with the theranostic adaptor protein. This opens room for re-evaluation of “inaccurate” biomarkers such as PD-L1 for assessing patient response to blockade through cancer immunotherapies. Furthermore, europium is known to be inefficient at being excited directly; it luminesces based on Dexter Electron Transfer and Forster Resonance Energy Transfer[93] with variable excitation peaks depending on the surrounding matrices[98]. In addition, an anticipated effect of the theranostic adaptor includes prolonged half-life of the biological therapy[99], as the FC regions of mAbs have been shown to play an important role in

macrophage accrual / elimination by the immune system[100]. Finally, gadolinium relaxivity is anticipated to dramatically increase upon binding with biologics, as demonstrated with protein-bound gadolinium complexes[10,101-103].

Signal intensities will depend on the number of lanthanide ions tightly bound inside the protein. This naturally urges one to include as many binding motifs (12-amino acids long) back-to-back as possible. However, highly repeated motifs may lead to a loss in protein function due to different folding kinetics, as well as problems in the translation process due heavy usage of specific tRNAs. Different combinations of binding tags can be explored, as well as varying lengths in different regions throughout the tag.

Spacer or linker sequences are also another factor to consider when constructing a protein with multiple functional regions[104,105]. Empirical linkers provide mechanical characteristics to the protein in terms of flexibility and cleavability. Flexible linkers can allow interaction between functional domains, whereas rigid linkers are good for maintaining distance and structure. Cleavable linkers allow for cleavage at the linker site by proteases produced *in-vivo* or administered *in-vitro*. Adequate use of linkers has been shown to increase stability by altering folding kinetics, increase expression and solubility, alter PK, and ultimately, improve biological activity of the functional domains by increasing accessibility[104,105].

5.2 MATERIALS AND METHODS

MATERIALS

Reagents

All materials were purchased from Thermo Fisher or its subsidiaries, with the exception of REEs and calcium chloride purchased from Sigma-Aldrich, centrifugal filtration units from Merck, and FPLC columns from Cytiva.

Cell lines

- For DNA replication and other cloning purposes: E, Coli, TOP10 (Thermo Fisher), 5-alpha (New England Biolabs), 10-beta (New England Biolabs)
- For protein expression: E. Coli, BL21* (Thermo Fisher)

Equipment

- Fluorescence measurements: Cytation 5 (Biotek), Spark (Tecan)
- MRI: 7T 70/30 Biospec (Bruker)
- ICP: Varian 710-ES (ICP-OES), Thermo Fisher iCAP-Q (ICP-MS)
- Gel visualization: ChemiDoc Imaging System (BioRad)

Software

Prism (GraphPad) was used for generating graphs and performing statistical analysis. SnapGene was used for designing vectors containing new proteins and generating their corresponding maps, and BioRender was used for creating illustrations. Paravision 360 was used for obtaining T1 and T2 relaxation times and their corresponding maps.

METHODS

Gene design and cloning

Proteins were cloned via TOPO cloning of a custom ordered double stranded DNA from IDT (sequence in appendix). Cloning results were confirmed after transforming the final vector into TOP10 to obtain sufficient DNA for Sanger sequencing (provided by Genewiz) and other downstream applications.

Protein expression

Proteins were expressed by E. coli (BL21*) that had been transformed with the cloned pET101 vectors containing the UTA constructs. Cells were incubated with ampicillin-spiked (100 µg/mL) Magic Media for 24hrs at 30^{o}C, shaking at 300-360 RPM. Expression and purification were verified via Western Blot against the V5 tag.

Protein purification

Purification was performed via HIS-tag purification with cobalt resin. For small (<50 mL) cultures, 200 µL columns were used, whereas larger volumes (>400 mL) were purified via FPLC (AKTA by Cytiva). Buffer exchange was done with either centrifugal filtration units (3-10kD, 4-15 mL), desalting columns (7kD), or dialysis cassettes (10kD) at least three consecutive times with 25 mM TRIS buffer at pH 7.0. Further purification via size exclusion was performed as necessary, with HiLoad 16/600 Superdex 200pg columns connected to the FPLC system.

Time Resolved Luminescence

Time Resolved Luminescence was measured with the Cytation5 (Biotek) with excitation at 280 nm and emission at 535-620 nm with monochromators and/or filters. Wells were prepared in triplicates (unless indicated otherwise) with 1-10 µL of UTA and 5-25 µL of REEs inside TRIS buffer (25 mM, pH7), resulting in a total well volume of 250

µL. Both delay times and exposure times were set at the maximum of 2000 µs to boost signal to noise ratio.

Inductively Coupled Plasma

ICP data were acquired with either ICP-OES (Varian 710-ES) or ICP-MS (Thermo iCAP-Q), depending on the required limit of detection. Samples were brought to a volume of 1.0 mL using 2% nitric acid (Optima grade), then further diluted 10-fold into the 2% nitric matrix. Samples were spiked with 2 ppb of an internal standard mixture of Li, Sc, In, Y, and Bi. In was selected as the internal standard for analysis of Gd. For each analyte, the dwell time was 0.3 S and 6 main runs of 15 sweeps were performed. Each analyte was measured in KEDS mode by introduction to a collision cell, using helium as the collision gas. Isotopes for each analyte were selected by Qtegra software (ThermoFisher Scientific) to maximize signal abundance while minimizing potential interferences. Error bars were generated with data acquired at different wavelengths (OES) or isotopes (MS). Quantification was performed against calibration curves generated for each analyte of interest. Prior to analysis, the mass spectrometer was tuned, and instrument calibration was verified using automated routines. In-run analytical precision and accuracy were assessed by reintroduction of aliquots of a selected calibration standard interspersed throughout the run. Data were corrected to a matrix blank of 2% nitric acid (Optima grade) prepared in ultrapure water and analyzed using Qtegra.

Magnetic Resonance Imaging

MRI data were acquired using a 7 Tesla 70/30 Biospec (Bruker) equipped with 12-inch gradients and an 86 mm transmit/receive volume coil. T1 maps were generated via Paravision 360, with ten TR experiments ranging from 100 ms-17500 ms with an echo time of 6.89 ms and three averages each. Multiple 1 mm-thick slices were acquired across samples contained in PCR tubes, 0.5 mL tubes, and 1.5 mL microcentrifuge tubes. Samples were prepared by allowing GLamouR (volumes from 400 uL-1 mL, concentrations from 0.5 mg/mL-1 mg/mL) to bind with gadolinium (0.5 mM) for 30 mins at RT, followed by several washing steps, identical to the methods used for buffer exchange following protein purification.

Relaxivity (r_i) values were calculated with the following equation[42] :

$$R_i = R_i^0 + r_i \times [CA]; \qquad i = 1,2$$

where concentration is denoted by [CA], which is the Gadolinium concentration measured by ICP. R_i is the relaxation rate of the solution in presence of the contrast agent (=1/T_i). R_i^0 is the relaxation rate of the solution without the contrast agent.

5.3 RESULTS

Protein purification and binding

Figure 13a is a 3-dimensional representation of the estimated protein structure, rendered by Robetta and visualized by PyMOL. REE binding sites are shown in blue, with inserted Tryptophan residues labeled in red. The rest of the protein is shown in green, which consists of an antibody binding domain and several linker sequences between the REE binding sites and other domains as spacers. As seen in figure 13b, UTA was purified on the leftmost lane of the native SDS page, with an anti-mouse IgG in the center, and UTA-IgG conjugate on the rightmost lane showing one protein slightly above IgG, suggesting the formation of a conjugate. Excess UTA remaining from the 1:2 molar ratio binding is shown further down the lane at a lower molecular weight. Figure 13c shows an overlay of all three proteins, where it is evident that the IgG elutes before UTA. This follows the laws of physics and basic principles of size-exclusion chromatography in that larger molecules elute faster than smaller molecules when passed through SEC resin. The UTA-IgG conjugate has a single elution peak at the IgG point indicating a similar size, but with no additional peak at the UTA region.

Further investigation is needed to confirm UTA-IgG binding with irrefutable proof such as ITC or SPR experiments. Collecting the UTA-IgG peak separately and running molecular imaging assays may also confirm the presence of UTA on a molecule as large as the IgG. Furthermore, running the peak with a non-native SDS page could help in determining the presence of UTA, but since IgG is comprised of 25kDa and 50kDa subunits, the 40kDa-sized UTA would be very close to the 50kDa mark of the all-blue ladder. A different ladder (in smaller range) could be used in combination with a longer

gel to increase the resolution, or a native gel could be used to try telling the difference between the 150kDa-sized IgG from a 200kDa-sized conjugate, with a larger-range ladder to aid in analysis such as the “PAGE-ruler” ladder. Separation of UTA from the IgG could also occur, revealing the UTA band underneath the IgG band, which would confirm the formation of a conjugate, but with weaker binding than probably desired. The advantage of ITC or SPR experiments is that not only will they confirm binding events between two proteins, but also provide the binding affinity values (and in the case of SPR, real-time kinetics).

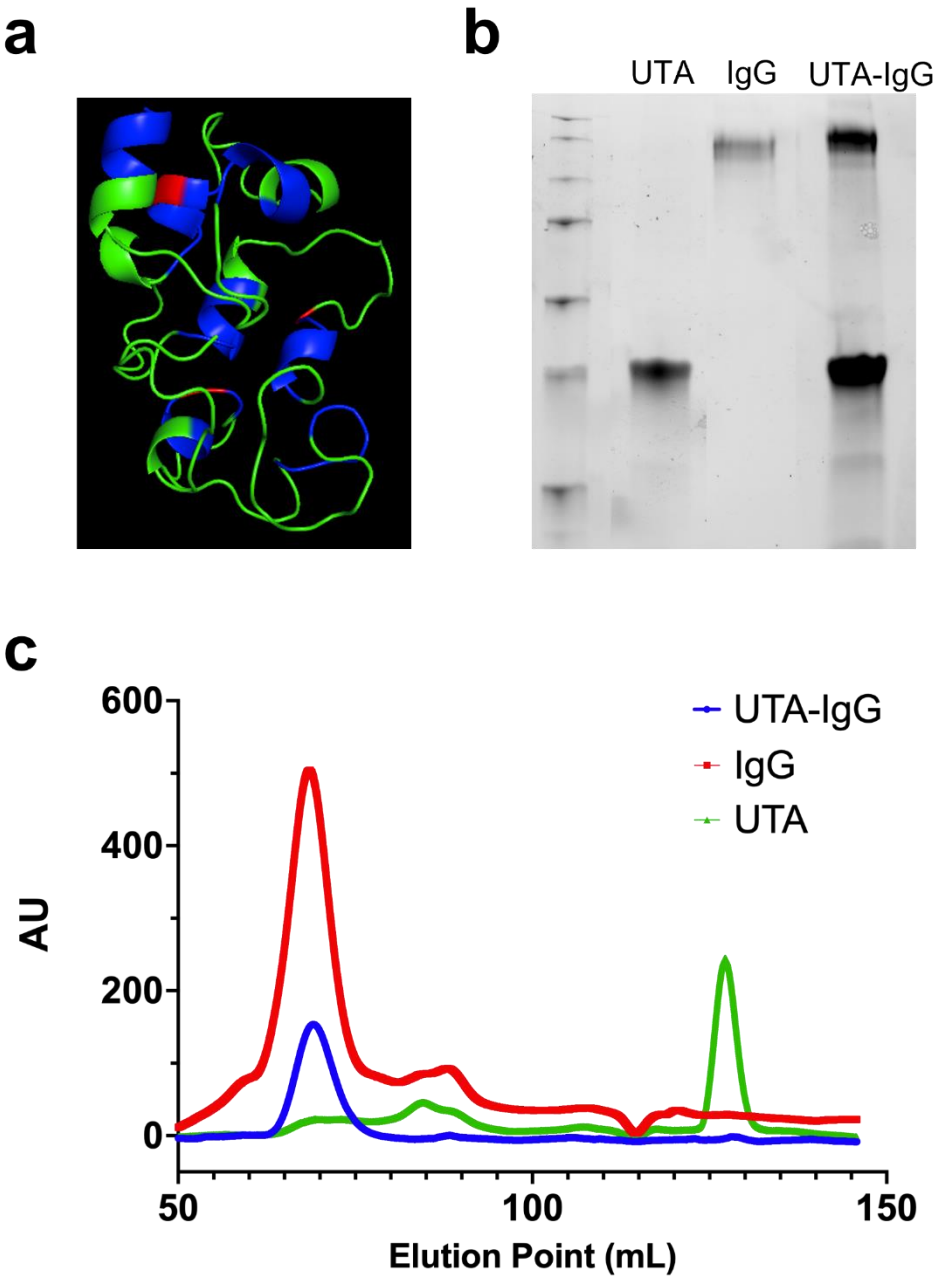

FIGURE 13. Purification and binding of the UTA protein. (a) A 3-dimensional rendition of the protein with PyMOL. (b) An SDS-page of UTA, IgG, and UTA-IgG conjugate. (c) an SEC chromatogram overlayed for UTA, IgG, and IgG conjugate.

Molecular Imaging

Time-resolved Luminescence

Optical imaging capabilities of UTA were tested with a multi-mode plate reader (n=3 individual wells, 10 averages per read). In Figure 14a, it is shown that the UTA possesses minimal luminescence properties, with a high concentration (2X) of terbium being detectable when set at a high enough gain. When UTA and terbium are mixed, however, RFU is significantly increased to approximately 15,000. Even when terbium is increased to 2X, the luminescence remains at the same level, which suggests that the limiting factor in achieving higher RFU is the amount of UTA, and not the terbium. Figure 14b demonstrates a setting where parameters have been optimized so that UTA is virtually non-detectable, whereas terbium is reduced to approximately $1/9^{th}$ of the original RFU. Once more, the UTA+Tb group shows that luminescence is significantly increased and suggests that with a much higher concentration of protein, the RFU can be gained even higher. Such experiment was carried out in Figure 14c, where terbium was reduced even further down (to less than $1/25^{th}$ of the original value) while enough UTA was added to further boost the RFU to approximately 18,000, displaying an impressive 8,380% increase in signal. In addition to terbium, europium was also tested as seen in Figure 14d. Although the signal increase is 219%, which is lower than terbium, it is worth noting that the stokes shift is much larger at more than 300nm (with red emission), which separates europium further apart from autofluorescence or green light. Further exploration into the optimization for europium would be highly desirable where signal is increased even more, up to the level of terbium.

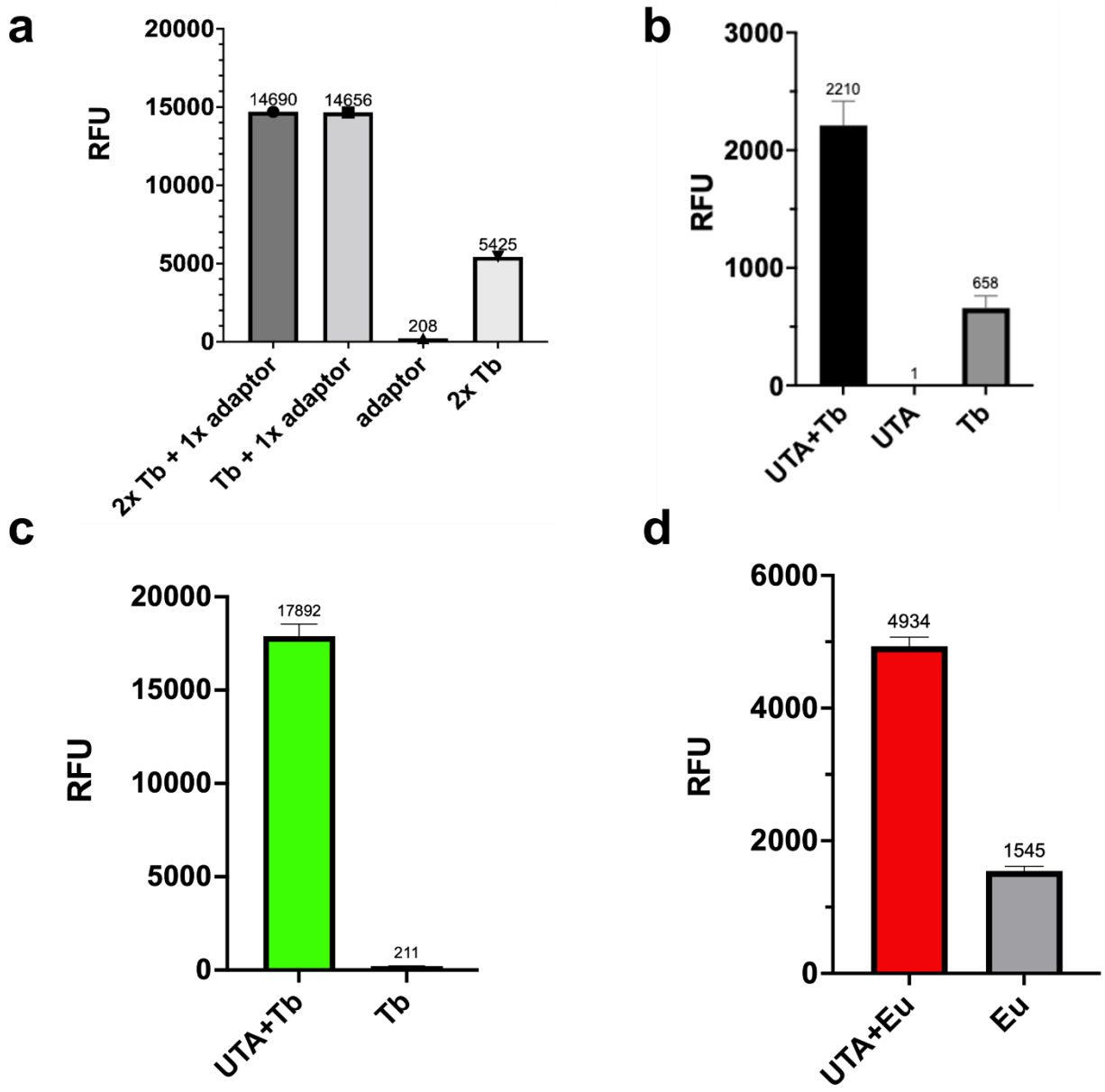

FIGURE 14. Time-resolved luminescence of UTA. (a) Concentration of UTA was determined to be the limiting factor in increasing signal to noise ratio. (b) Optimized parameters where UTA signal is virtually non-existent. (c) 8,380% increase in TRL with almost no background signal with terbium. (d) 219% increase in TRL with europium at 650 nm emission wavelength.

Enhanced relaxivity of gadolinium

As anticipated, gadolinium was seen to have a higher relaxivity when bound to UTA with an r1 of 7.5 (Figure 15) which is higher than that of GLamouR. Similarly, it is anticipated that a higher molecular weight associated with the conjugated form brings the molecular correlation time closer to the Larmor frequency so that tumbling is more in tune, therefore significantly increasing relaxivity. It is worth noting that UTA possess a smaller size than that of GLamouR, so additional factors such as inner or outer sphere effects, or water availability and hydrophilicity of the protein may have influenced the per-gadolinium relaxivity to be higher.

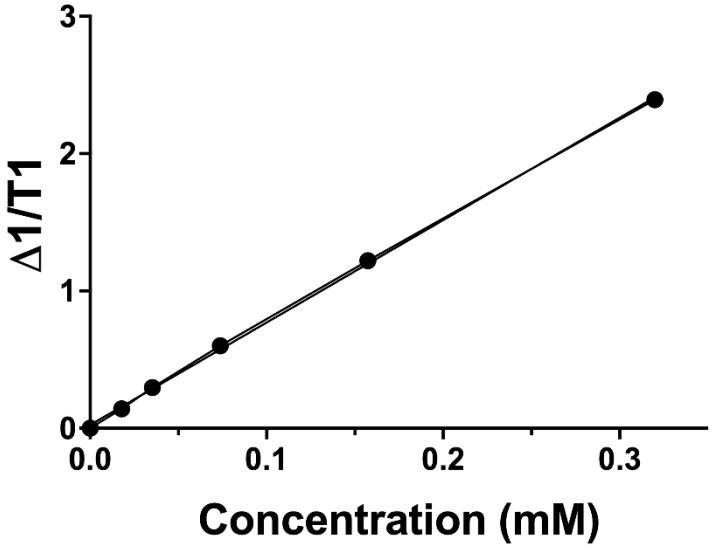

FIGURE 15. Relaxivity of UTA. R1 relaxivity of UTA was calculated to be 7.5.

5.4 FUTURE DIRECTIONS

The successful completion of this project lays the groundwork for any protein-based biological therapy to be easily tagged for observation via non-invasive imaging modalities without the need for extensive labor or cost. However, assays for protein stability and immunogenicity that were described in previous chapters should be performed before considering *in-vivo* experimentation. Humanized mouse models hold the potential for many more targets and biologics such as Rheumatoid Arthritis disease models with Adalimumab, or immune-cell targeting biologics for cancer immunotherapies such as anti PD-1 or anti CTLA-4. Future studies shall include SPECT/CT to drive this point further, which can be demonstrated with ^{177}lutetium as the imaging lanthanide and radioactive therapy, coupled with anti-PSMA for the targeting system which has been FDA approved. Similarly, it is anticipated that isotopes for Yttrium may be used for use with PET/MRI. Molecular imaging is a powerful non-invasive tool that holds the key to personalized and lesion-specific medicine by elucidating individual patient uptake and response, and this study aims to facilitate widespread research to tag any mAb-based biological therapy for multiple modalities.

6. CONCLUDING REMARKS

Synthetic proteins hold promise in a wide range of applications with endless possibilities. With modern technology, it is possible to explore this infinite search space with calculated approaches and has to this day resulted in many novel proteins that carry out specifically desired functions in applications such as molecular imaging, therapy, or biological detection, which have been demonstrated herein.

With recent advancements in artificial intelligence, quantum computing, and robotics, it is anticipated that we may someday crack the code for protein design. However, as required for the projects described herein, technologies must be developed and optimized to certain thresholds for the proper utilization of new synthetic proteins designed for specific purposes. Further efforts are also required for technologies to be cost effective, practical, and attractive to industries and consumers alike. Only when all aspects have been thoroughly considered will new discoveries actually make an impact on the world.

REFERENCES

1 Vol. Critical Metals: The Chemistry of Light, Box FC-5, Folder Periodic Tables (Assorted) (International Union of Pure and Applied Chemistry).

2 Caravan, P., Ellison, J. J., McMurry, T. J. & Lauffer, R. B. Gadolinium(III) Chelates as MRI Contrast Agents: Structure, Dynamics, and Applications. *Chem Rev* **99**, 2293-2352, doi:10.1021/cr980440x (1999).

3 Le Fur, M. & Caravan, P. The biological fate of gadolinium-based MRI contrast agents: a call to action for bioinorganic chemists. *Metallomics* **11**, 240-254, doi:10.1039/c8mt00302e (2019).

4 Caravan, P. Strategies for increasing the sensitivity of gadolinium based MRI contrast agents. *Chem Soc Rev* **35**, 512-523, doi:10.1039/b510982p (2006).

5 Salarian, M. *et al.* Early detection and staging of chronic liver diseases with a protein MRI contrast agent. *Nat Commun* **10**, 4777, doi:10.1038/s41467-019-11984-2 (2019).

6 Louie, A. Y. *et al.* In vivo visualization of gene expression using magnetic resonance imaging. *Nat Biotechnol* **18**, 321-325 (2000).

7 Bartelle, B. B. *et al.* Novel genetic approach for in vivo vascular imaging in mice.

8 Patrick, P. S. *et al.* Dual-modality gene reporter for in vivo imaging.

9 Nyström, N. N. *et al.* Longitudinal Visualization of Viable Cancer Cell Intratumoral Distribution in Mouse Models Using Oatp1a1-Enhanced Magnetic Resonance Imaging.

10 Xue, S. *et al.* Protein MRI contrast agent with unprecedented metal selectivity and sensitivity for liver cancer imaging. *Proc Natl Acad Sci U S A* **112**, 6607-6612, doi:10.1073/pnas.1423021112 (2015).

11 Runge, V. M. Commentary on T1-Weighted Hypersignal in the Deep Cerebellar Nuclei After Repeated Administrations of Gadolinium-based Contrast Agents in Healthy Rats. *Invest Radiol* **50**, doi:10.7892/boris.77250 (2015).

12 Ebrahimi, P. & Barbieri, M. Gadolinium as an Emerging Microcontaminant in Water Resources: Threats and Opportunities. *Geosciences* **9**, doi:10.3390/geosciences9020093 (2019).

13 Brunjes, R. & Hofmann, T. Anthropogenic gadolinium in freshwater and drinking water systems. *Water Res* **182**, 115966, doi:10.1016/j.watres.2020.115966 (2020).

14 Khan, S. Radioactive Waste Management in A Hospital. *Int J Health Sci* **4** (2010).

15 Windfeld, E. S. & Brooks, M. S. Medical waste management - A review. *J Environ Manage* **163**, 98-108, doi:10.1016/j.jenvman.2015.08.013 (2015).

16 Fraum, T. J., Ludwig, D. R., Bashir, M. R. & Fowler, K. J. Gadolinium-based contrast agents: A comprehensive risk assessment. *J Magn Reson Imaging* **46**, 338-353, doi:10.1002/jmri.25625 (2017).

17 Gulani, V. *et al.* Gadolinium deposition in the brain: summary of evidence and recommendations. *Lancet Neurol* **16**, 564-570, doi:10.1016/S1474-4422(17)30158-8 (2017).

18 Sherry, A. D., Caravan, P. & Lenkinski, R. E. Primer on gadolinium chemistry. *J Magn Reson Imaging* **30**, 1240-1248, doi:10.1002/jmri.21966 (2009).

19 Semelka, R. C. *et al.* Gadolinium deposition disease: Initial description of a disease that has been around for a while. *Magn Reson Imaging* **34**, 1383-1390, doi:10.1016/j.mri.2016.07.016 (2016).

20 Ramalho, M., Ramalho, J., Burke, L. M. & Semelka, R. C. Gadolinium Retention and Toxicity-An Update. *Adv Chronic Kidney Dis* **24**, 138-146, doi:10.1053/j.ackd.2017.03.004 (2017).

21 Ersoy, H. & Rybicki, F. J. Biochemical safety profiles of gadolinium-based extracellular contrast agents and nephrogenic systemic fibrosis. *J Magn Reson Imaging* **26**, 1190-1197, doi:10.1002/jmri.21135 (2007).

22 Bonificio, W. D. & Clarke, D. R. Rare-Earth Separation Using Bacteria. *Environmental Science & Technology Letters* **3**, 180-184, doi:10.1021/acs.estlett.6b00064 (2016).

23 Good, N. M. *et al.* Hyperaccumulation of Gadolinium by Methylorubrum extorquens AM1 Reveals Impacts of Lanthanides on Cellular Processes Beyond Methylotrophy. *Front Microbiol* **13**, 820327, doi:10.3389/fmicb.2022.820327 (2022).

24 Balaram, V. Rare earth elements: A review of applications, occurrence, exploration, analysis, recycling, and environmental impact. *Geoscience Frontiers* **10**, 1285-1303, doi:10.1016/j.gsf.2018.12.005 (2019).

25 Liu, T. & Chen, J. Extraction and separation of heavy rare earth elements: A review. *Separation and Purification Technology* **276**, doi:10.1016/j.seppur.2021.119263 (2021).

26 Abraham, D. S. *The Elements of Power*. (Yale University Press, 2015).

27 Jowitt, S. M., Werner, T. T., Weng, Z. & Mudd, G. M. Recycling of the rare earth elements. *Current Opinion in Green and Sustainable Chemistry* **13**, 1-7, doi:10.1016/j.cogsc.2018.02.008 (2018).

28 Rare Earth Metals Recycling Market by Application, Technology, and Region. (Markets and Markets, 2022).

29 Lederer, F. L., Curtis, S. B., Bachmann, S., Dunbar, W. S. & MacGillivray, R. T. Identification of lanthanum-specific peptides for future recycling of rare earth elements from compact fluorescent lamps. *Biotechnol Bioeng* **114**, 1016-1024, doi:10.1002/bit.26240 (2017).

30 Park, D. M. *et al.* Bioadsorption of Rare Earth Elements through Cell Surface Display of Lanthanide Binding Tags. *Environ Sci Technol* **50**, 2735-2742, doi:10.1021/acs.est.5b06129 (2016).

31 Gao, Y. *et al.* Peptide–carbon hybrid membranes for highly efficient and selective extraction of actinides from rare earth elements. *Journal of Materials Chemistry A* **9**, 14422-14431, doi:10.1039/d1ta03549e (2021).

32 Mattocks, J. A. & Cotruvo, J. A. Biological, biomolecular, and bio-inspired strategies for detection, extraction, and separations of lanthanides and actinides. *Chem Soc Rev* **49**, 8315-8334, doi:10.1039/d0cs00653j (2020).

33 Mattocks, J. A., Ho, J. V. & Cotruvo, J. A., Jr. A Selective, Protein-Based Fluorescent Sensor with Picomolar Affinity for Rare Earth Elements. *J Am Chem Soc* **141**, 2857-2861, doi:10.1021/jacs.8b12155 (2019).

34 Deblonde, G. J. *et al.* Selective and Efficient Biomacromolecular Extraction of Rare-Earth Elements using Lanmodulin. *Inorg Chem* **59**, 11855-11867, doi:10.1021/acs.inorgchem.0c01303 (2020).

35 Featherston, E. R., Issertell, E. J. & Cotruvo, J. A., Jr. Probing Lanmodulin's Lanthanide Recognition via Sensitized Luminescence Yields a Platform for Quantification of Terbium in Acid Mine Drainage. *J Am Chem Soc* **143**, 14287-14299, doi:10.1021/jacs.1c06360 (2021).

36 Skovran, E. & Martinez-Gomez, N. C. Microbiology. Just add lanthanides. *Science* **348**, 862-863, doi:10.1126/science.aaa9091 (2015).

37 Peplow, M. Unlocking the Lanthanome. *ACS Central Science*, doi:10.1021/acscentsci.1c01357 (2021).

38 Leducq, J. B. *et al.* Comprehensive Phylogenomics of Methylobacterium Reveals Four Evolutionary Distinct Groups and Underappreciated Phyllosphere Diversity. *Genome Biol Evol* **14**, doi:10.1093/gbe/evac123 (2022).

39 Cotruvo, J. A., Jr., Featherston, E. R., Mattocks, J. A., Ho, J. V. & Laremore, T. N. Lanmodulin: A Highly Selective Lanthanide-Binding Protein from a Lanthanide-Utilizing Bacterium. *J Am Chem Soc* **140**, 15056-15061, doi:10.1021/jacs.8b09842 (2018).

40 Gutenthaler, S. M. *et al.* Lanmodulin peptides - unravelling the binding of the EF-Hand loop sequences stripped from the structural corset. *Inorg Chem Front* **9**, 4009-4021, doi:10.1039/d2qi00933a (2022).

41 Chen, T. W. *et al.* Ultrasensitive fluorescent proteins for imaging neuronal activity. *Nature* **499**, 295-300, doi:10.1038/nature12354 (2013).

42 Wahsner, J., Gale, E. M., Rodriguez-Rodriguez, A. & Caravan, P. Chemistry of MRI Contrast Agents: Current Challenges and New Frontiers. *Chem Rev* **119**, 957-1057, doi:10.1021/acs.chemrev.8b00363 (2019).

43 Kostyuk, A. I., Demidovich, A. D., Kotova, D. A., Belousov, V. V. & Bilan, D. S. Circularly Permuted Fluorescent Protein-Based Indicators: History, Principles, and Classification. *Int J Mol Sci* **20**, doi:10.3390/ijms20174200 (2019).

44 Shen, Y. *et al.* A genetically encoded Ca(2+) indicator based on circularly permutated sea anemone red fluorescent protein eqFP578. *BMC Biol* **16**, 9, doi:10.1186/s12915-018-0480-0 (2018).

45 Akerboom, J. *et al.* Genetically encoded calcium indicators for multi-color neural activity imaging and combination with optogenetics. *Front Mol Neurosci* **6**, 2, doi:10.3389/fnmol.2013.00002 (2013).

46 Junichi Nakai, M. O., Keiji Imoto. A high signal-to-noise Ca2+ probe composed of a single green fluorescent protein. *Nat. Biotechnol.* **19** (2001).

47 Tian, L. *et al.* Imaging neural activity in worms, flies and mice with improved GCaMP calcium indicators. *Nat Methods* **6**, 875-881, doi:10.1038/nmeth.1398 (2009).

48 Akerboom, J. *et al.* Optimization of a GCaMP calcium indicator for neural activity imaging. *J Neurosci* **32**, 13819-13840, doi:10.1523/JNEUROSCI.2601-12.2012 (2012).

49 Masatoshi Inoue, A. T., Shin-ichiro Horigane, Masamichi Ohkura, Keiko Gengyo-Ando,, Hajime Fujii, S. K., Sayaka Takemoto-Kimura, Masanobu Kano, Junichi Nakai, & Kazuo Kitamura, H. B. Rational design of a high-affinity, fast, red calcium indicator R-CaMP2. *Nat. Methods* **12** (2015).

50 Yongxin Zhao, S. A., Jiahui Wu, Takayuki Teramoto, Yu-Fen Chang,, Masahiro Nakano, A. S. A., Manabi Fujiwara, Takeshi Ishihara, & Takeharu Nagai, R. E. C. An Expanded Palette of Genetically Encoded Ca2+ Indicators. *Science* **333**, 1888-1891 (2011).

51 Akerboom, J. *et al.* Crystal structures of the GCaMP calcium sensor reveal the mechanism of fluorescence signal change and aid rational design. *J Biol Chem* **284**, 6455-6464, doi:10.1074/jbc.M807657200 (2009).

52 Shen, Y. *et al.* T1 relaxivities of gadolinium-based magnetic resonance contrast agents in human whole blood at 1.5, 3, and 7 T. *Invest Radiol* **50**, 330-338, doi:10.1097/rli.0000000000000132 (2015).

53 Noebauer-Huhmann, I. M. *et al.* Gadolinium-based magnetic resonance contrast agents at 7 Tesla: in vitro T1 relaxivities in human blood plasma. *Invest Radiol* **45**, 554-558, doi:10.1097/RLI.0b013e3181ebd4e3 (2010).

54 Szomolanyi, P. *et al.* Comparison of the Relaxivities of Macrocyclic Gadolinium-Based Contrast Agents in Human Plasma at 1.5, 3, and 7 T, and Blood at 3 T. *Invest Radiol* **54**, 559-564, doi:10.1097/rli.0000000000000577 (2019).

55 Rohrer, M., Bauer, H., Mintorovitch, J., Requardt, M. & Weinmann, H. J. Comparison of magnetic properties of MRI contrast media solutions at different magnetic field strengths. *Invest Radiol* **40**, 715-724, doi:10.1097/01.rli.0000184756.66360.d3 (2005).

56 Morcos, S. K. Extracellular gadolinium contrast agents: differences in stability. *Eur J Radiol* **66**, 175-179, doi:10.1016/j.ejrad.2008.01.025 (2008).

57 Gwenzi, W. *et al.* Sources, behaviour, and environmental and human health risks of high-technology rare earth elements as emerging contaminants. *Sci Total Environ* **636**, 299-313, doi:10.1016/j.scitotenv.2018.04.235 (2018).

58 McLeod, C. & Krekeler, M. Sources of Extraterrestrial Rare Earth Elements: To the Moon and Beyond. *Resources* **6**, doi:10.3390/resources6030040 (2017).

59 Cockell, C. S. *et al.* Space station biomining experiment demonstrates rare earth element extraction in microgravity and Mars gravity. *Nat Commun* **11**, 5523, doi:10.1038/s41467-020-19276-w (2020).

60 Johnson, I. S. Human Insulin from Recombinant DNA Technology. *Science* **219**, 632-637, doi:doi:10.1126/science.6337396 (1983).

61 Urquhart, L. Top companies and drugs by sales in 2019. *Nat Rev Drug Discov* **19**, 228, doi:10.1038/d41573-020-00047-7 (2020).

62 Beck, A. Biosimilar, biobetter and next generation therapeutic antibodies. *MAbs* **3**, 107-110, doi:10.4161/mabs.3.2.14785 (2011).

63 Gilad, A. A. *et al.* Artificial reporter gene providing MRI contrast based on proton exchange. *Nature Biotechnology* **25**, 217-219, doi:10.1038/nbt1277 (2007).

64 van Zijl, P. C. & Yadav, N. N. Chemical exchange saturation transfer (CEST): what is in a name and what isn't?

65 Bulte, J. W. M., Wang, C. & Shakeri-Zadeh, A. In Vivo Cellular Magnetic Imaging: Labeled versus Unlabeled Cells. *Advanced Functional Materials* **32**, 2207626, doi:https://doi.org/10.1002/adfm.202207626 (2022).

66 De Groot, A. S. & Scott, D. W. Immunogenicity of protein therapeutics. *Trends in Immunology* **28**, 482-490, doi:https://doi.org/10.1016/j.it.2007.07.011 (2007).

67 Bar-Shir, A. *et al.* Human Protamine-1 as an MRI Reporter Gene Based on Chemical Exchange. *ACS Chemical Biology* **9**, 134-138, doi:10.1021/cb400617q (2014).

68 Li, D. & Wu, M. Pattern recognition receptors in health and diseases. *Signal Transduction and Targeted Therapy* **6**, 291, doi:10.1038/s41392-021-00687-0 (2021).

69 Cibrián, D. & Sánchez-Madrid, F. A.-O. CD69: from activation marker to metabolic gatekeeper.

70 Aktas, E., Kucuksezer, U. C., Bilgic, S., Erten, G. & Deniz, G. Relationship between CD107a expression and cytotoxic activity. *Cellular Immunology* **254**, 149-154, doi:https://doi.org/10.1016/j.cellimm.2008.08.007 (2009).

71 Vita, R. *et al.* The Immune Epitope Database (IEDB): 2018 update.

72 Bricco, A. R. *et al.* A Genetic Programming Approach to Engineering MRI Reporter Genes. *ACS Synth Biol*, doi:10.1021/acssynbio.2c00648 (2023).

73 Pearson, T., Greiner, D. L. & Shultz, L. D. Creation of "humanized" mice to study human immunity. *Curr Protoc Immunol* **Chapter 15**, Unit 15 21, doi:10.1002/0471142735.im1521s81 (2008).

74 Rosato, R. R. *et al.* Evaluation of anti-PD-1-based therapy against triple-negative breast cancer patient-derived xenograft tumors engrafted in humanized mouse models. *Breast Cancer Res* **20**, 108, doi:10.1186/s13058-018-1037-4 (2018).

75 Samantha Semenkow, S. L., Ulf D. Kahlert, Eric H. Raabe, Jiadi Xu, Antje, Arnold, M. J., Byoung Chol Oh, Gerald Brandacher, Jeff W.M. & Bulte, C. G. E., Piotr Walczak. An immunocompetent mouse model of human glioblastoma. *Oncotarget* **8** (2017).

76 Vaudel, M., Sickmann, A. & Martens, L. Peptide and protein quantification: a map of the minefield. *Proteomics* **10**, 650-670, doi:10.1002/pmic.200900481 (2010).

77 Perlman, O. *et al.* Redesigned reporter gene for improved proton exchange-based molecular MRI contrast. *Scientific Reports* **10**, 20664, doi:10.1038/s41598-020-77576-z (2020).

78 Bradford, M. M. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. *Analytical Biochemistry* **72**, 248-254, doi:https://doi.org/10.1016/0003-2697(76)90527-3 (1976).

79 Anthis, N. J. & Clore, G. M. Sequence-specific determination of protein and peptide concentrations by absorbance at 205 nm. *Protein Sci* **22**, 851-858, doi:10.1002/pro.2253 (2013).

80 Kashima, J. & Okuma, Y. Harmonization study of antibodies and platforms for programmed death ligand 1 immunostaining in non-small cell lung cancer: does shuffling couples settle the troubles? *J Thorac Dis* **10**, S2124-S2126, doi:10.21037/jtd.2018.06.116 (2018).

81 Chalmers, Z. R. *et al.* Analysis of 100,000 human cancer genomes reveals the landscape of tumor mutational burden. *Genome Med* **9**, 34, doi:10.1186/s13073-017-0424-2 (2017).

82 Xu, Z. *et al.* Assessment of tumor mutation burden calculation from gene panel sequencing data. *Onco Targets Ther* **12**, 3401-3409, doi:10.2147/OTT.S196638 (2019).

83 Zhu, J. *et al.* Association Between Tumor Mutation Burden (TMB) and Outcomes of Cancer Patients Treated With PD-1/PD-L1 Inhibitions: A Meta-Analysis. *Front Pharmacol* **10**, 673, doi:10.3389/fphar.2019.00673 (2019).

84 UCSF Health Center for Clinical Genetics and Genomics. Exome Sequencing. (2020) at <https://genomics.ucsf.edu/content/exome-sequencing>

85 Melendez, B. *et al.* Methods of measurement for tumor mutational burden in tumor tissue. *Transl Lung Cancer Res* **7**, 661-667, doi:10.21037/tlcr.2018.08.02 (2018).

86 Truesdell, J., Miller, V. A. & Fabrizio, D. Approach to evaluating tumor mutational burden in routine clinical practice. *Transl Lung Cancer Res* **7**, 678-681, doi:10.21037/tlcr.2018.10.10 (2018).

87 Fenizia, F. *et al.* Measuring tumor mutation burden in non-small cell lung cancer: tissue versus liquid biopsy. *Transl Lung Cancer Res* **7**, 668-677, doi:10.21037/tlcr.2018.09.23 (2018).

88 Dammes, N. & Peer, D. Monoclonal antibody-based molecular imaging strategies and theranostic opportunities. *Theranostics* **10**, 938-955, doi:10.7150/thno.37443 (2020).

89 D. Shahbazi-Gahrouei, S. M. R., M.A. Williams, B.J. Allen. In vitro studies of gadolinium-DTPA conjugated with monoclonal antibodies as cancer-specific magnetic resonance imaging contrast agents. *AUSTRALAS PHYS ENG S* **25** (2002).

90 Bensch, F. *et al.* (89)Zr-atezolizumab imaging as a non-invasive approach to assess clinical response to PD-L1 blockade in cancer. *Nat Med* **24**, 1852-1858, doi:10.1038/s41591-018-0255-8 (2018).

91 Kircher, M. F. *et al.* A brain tumor molecular imaging strategy using a new triple-modality MRI-photoacoustic-Raman nanoparticle. *Nat Med* **18**, 829-834, doi:10.1038/nm.2721 (2012).

92 Lohrke, J. *et al.* 25 Years of Contrast-Enhanced MRI: Developments, Current Challenges and Future Perspectives. *Adv Ther* **33**, 1-28, doi:10.1007/s12325-015-0275-4 (2016).

93 Mara, M. W. *et al.* Energy Transfer from Antenna Ligand to Europium(III) Followed Using Ultrafast Optical and X-ray Spectroscopy. *J Am Chem Soc* **141**, 11071-11081, doi:10.1021/jacs.9b02792 (2019).

94 Hofman, M. S. *et al.* TheraP: a randomized phase 2 trial of (177) Lu-PSMA-617 theranostic treatment vs cabazitaxel in progressive metastatic castration-resistant prostate cancer (Clinical Trial Protocol ANZUP 1603). *BJU Int* **124 Suppl 1**, 5-13, doi:10.1111/bju.14876 (2019).

95 Zang, J. *et al.* 177Lu-EB-PSMA Radioligand Therapy with Escalating Doses in Patients with Metastatic Castration-Resistant Prostate Cancer. *Journal of Nuclear Medicine* **61**, 1772, doi:10.2967/jnumed.120.242263 (2020).

96 England, C. G., Ehlerding, E. B. & Cai, W. NanoLuc: A Small Luciferase Is Brightening Up the Field of Bioluminescence. *Bioconjug Chem* **27**, 1175-1187, doi:10.1021/acs.bioconjchem.6b00112 (2016).

97 Atkins, K. L. *et al.* S. aureus IgG-binding proteins SpA and Sbi: host specificity and mechanisms of immune complex formation. *Mol Immunol* **45**, 1600-1611, doi:10.1016/j.molimm.2007.10.021 (2008).

98 Eduardo J. Nassar, K. J. C., Paulo S. Calefi, Lilian L. Avila, Lucimara C. Bandeira, Alexandre Cestari, Emerson H. de Faria, Alex L. Marçal, Marcela G. Matos. *Europium III: Different Emission Spectra in Different Matrices, the Same Element*. (Nova Science Publishers, Inc., 2010).

99 Hutt, M., Farber-Schwarz, A., Unverdorben, F., Richter, F. & Kontermann, R. E. Plasma half-life extension of small recombinant antibodies by fusion to immunoglobulin-binding domains. *J Biol Chem* **287**, 4462-4469, doi:10.1074/jbc.M111.311522 (2012).

100 Sean P. Arlauckas, C. S. G., Rainer H. Kohler,Maya Kitaoka,, Michael F. Cuccarese, K. S. Y., Miles A. Miller, Jonathan C. Carlson, & Gordon J. Freeman, R. M. A., Ralph Weissleder, Mikael J. Pittet. In vivo imaging reveals tumor-associated macrophage-mediated resistance pathway in anti-PD-1 therapy. *Science Translational Medicine* (2017).

101 Caravan, P. Protein-Targeted Gadolinium-Based MAgnetic Resonance Imaging (MRI) Contrast Agents: Design and Mechanism of Action. *Acc Chem Res* **42** (2008).

102 Jenny J. Yang, J. Y., Lixia Wei, Omar Zurkiya, Wei Yang,, Shunyi Li, J. Z., Yubin Zhou, Anna L Wilkins Maniccia, Hui Mao,, Fuqiang Zhao, R. M., Shumin Zhao, Julian Johnson, Xiaoping Hu, & Eirik Krogstad, Z.-R. L. Rational Design of Protein-Based MRI Contrast Agents. *J Am Chem Soc* (2008).

103 Li, H. & Meade, T. J. Molecular Magnetic Resonance Imaging with Gd(III)-Based Contrast Agents: Challenges and Key Advances. *J Am Chem Soc* **141**, 17025-17041, doi:10.1021/jacs.9b09149 (2019).

104 Chen, X., Zaro, J. L. & Shen, W. C. Fusion protein linkers: property, design and functionality. *Adv Drug Deliv Rev* **65**, 1357-1369, doi:10.1016/j.addr.2012.09.039 (2013).

105 van Rosmalen, M., Krom, M. & Merkx, M. Tuning the Flexibility of Glycine-Serine Linkers To Allow Rational Design of Multidomain Proteins. *Biochemistry* **56**, 6565-6574, doi:10.1021/acs.biochem.7b00902 (2017).

APPENDIX

DNA SEQUENCES

> GLamouR

atggtcgactcatcacgtcgtaagtggaataagacaggtcacgcagtcagagctataggtcggctgagctcactcgaga

acgtctatatcaaggccgacaagcagaagaacggcatcaaggcgaacttcaagatccgccacaacatcgaggacgg

cggcgtgcagctcgcctaccactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccacta

cctgagcgtgcagtccaaactttcgaaagaccccaacgagaagcgcgatcacatggtcctgctggagttcgtgaccgcc

gccgggatcactctcggcatggacgagctgtacaagggcggtaccggagggagcatggtgagcaagggcgaggagc

tgttcaccggggtggtgcccatcctggtcgagctggacggcgacgtaaacggccacaagttcagcgtgtccggcgaggg

tgagggcgatgccacctacggcaagctgaccctgaagttcatctgcaccaccggcaagctgcccgtgccctggcccacc

ctcgtgaccaccctgacctacggcgtgcagtgcttcagccgctaccccgaccacatgaagcagcacgacttcttcaagtc

cgccatgcccgaaggctacatccaggagcgcaccatcttcttcaaggacgacggcaactacaagacccgcgccgagg

tgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgacttcaaggaggacggcaacatcctgg

ggcacaagctggagtacaacctgccggaccaactgactgaagagcagatcgcagaatttaaagaggctttctccctattt

gataaggacggcaccattgacctgaaagagctggggacggtgatgcggtctctggggcagaaccccacagaagcag

agctgcaggacatgatcaatgaagtagacccggacaaagatggcaccctggacgcgaaggagttcctgacaatgatg

gcaagaaaagggagctacagggacacggaagaagaaattagagaagcgttcggtgtgtttgacccggataacgacg

gcaccctggacaagaaagagcttcgccacgtgatgacaaaccttggagagaagttaacagatgaagaggttgatgaaa

tgatcagggaagcaaacccggataacgacggcaccattgatgcgcgtgaatttgtacaaatgatgacagcgaagaagg

gcgagctcaattcgaagcttgaaggtaagcctatccctaaccctctcctcggtctcgattctacgcgtaccggtcatcatcac

catcaccat

> UTA

atggaccctgacaaggatggaacgattgacctgaaagaggatcctgataaagacggtaccctggacgcaaaggagga

tccggataatgatggaacgttagacaaaaaagaagatccggataaggacggtactatcgatctgaaagaggacccag

acaacgatggtactctggacaagaaagaagatccagataaagacgggaccttagatgctaaggaggacccggataac

gatgggactttagataagaaggaagtcttcactcttgaggatttcgttggggactggcgacagacagccggctacaacctg

gaccaagtccttgaacagggaggtgtgtccagtttgtttcagaatctcggggtgtccgtaactccgatccaaaggattgtcct

gagcggtgaaaatgggctgaagatcgacatccatgtcatcatcccgtatgaaggtctgagcggcgaccaaatgggcca

gatcgaaaaaatttttaaggtggtgtaccctgtggatgatcatcactttaaggtgatcctgcactatggcacactggtaatcga

cggggttacgccgaacatgatcgactatttcggacggccgtatgaaggcatcgccgtgttcgacggcaaaaagatcactg

taacagggaccctgtggaacggcaacaaaattatcgacgagcgcctgatcaaccccgacggctccctgctgttccgagt

aaccatcaacggagtgaccggctggcggctgtgcgaacgcattctggcggcagcgcagcacgacgaggcgcagcaa

aacgcattctatcaagttttgaatatgcccaatttaaacgccgaccagcgtaatggtttcatccagtcgttgaaagatgaccc

gtcccaatccgctaatgtactgggtgaggcgaagaaactgaatgaaagtcaagcccctaaggccgataataacttcaat

aaagagcagcagaacgccttctacgagattcttaacatgcctaacttaaatgaggaacagcgcaatgggttcattcaatcc

cttaaggacgatccctcccagagtgcgaacttactttcagaggcaaaaaaattgaatgagagccaagctccgaaggctg

acaacaaattcaataaagaacaacaaaatgctttctatgaaattcttcatttaccaaatttgaatgaggagcaacgtaatgg

gttcattcagtctttaaaggatgacccttcgcagtcagcaaatcttctggcggaagcaaaaaagctgaatgatgcccaagc

cccgaaggcagataataaatttaataaggaacagcagaacgctttttatgagattttacatttgccaaatttgaccgaagaa

cagcgtaacggatttatccagagcctgaaggacgatcccagcgtgtcgaaggaaattttagcagaggcaaagaaactg

aatgacgcgaagggcgagctcaattcgaagcttgaaggtaagcctatccctaaccctctcctcggtctcgattctacgcgta

ccggtcatcatcaccatcaccat

> GLamouR-rs

atggtcgactcatcacgtcgtaagtggaataaggcaggtcacgcagtcagagctataggtcggctgagctcacccgtggtt
tccgagcggatgtaccccgaggacggcgccctgaagagcgagatcaagaaggggctgaggctgaaggacggcggc
cactacgccgccgaggtcaagaccacctacaaggccaagaagcccgtgcagctgcccggcgcctacatcgtggacat
caagttggacatcgtgtcccacaacgaggactacaccatcgtggaacagtgcgaacgcgccgagggccgccactcca
ccggcggcatggacgagctgtacaagggaggtacaggcgggagtctggtgagcaagggcgaggaggataacatggc
catcatcaaggagttcatgcgcttcaaggtgcacatggagggctccgtgaacggccacgagttcgagatcgagggcgag
ggcgagggccgcccctacgaggcctttcagaccgctaagctgaaggtgaccaagggtggcccccctgcccttcgcctgg
gacatcctgtcccctcagttcatgtacggctccaaggcctacattaagcacccagccgacatccccgactacttcaagctgt
ccttccccgagggcttcaggtgggagcgcgtgatgaacttcgaggacggcggcattattcacgtgaaccaggactcctcc
ctgcaggacggcgtattcatctacaaggtgaagctgcgcggcaccaacttcccccccgacggccccgtaatgcagaag
aagaccatgggctgggaggctacgcgtgacgacctgactgaagagcagatcgcagaatttaaagaggctttctccctatt
tgacccggataaggacggcaccattgacctgaaagagcttgggacggtgttccgatcgctggggcagaaccccacaga
agcagagctgcaggacatgatcaatgaagtagacccggacaaagatggcaccctggacgcgaaggagttcctgacg
atgatggcaagaaaaatgaatgacacagacagtgaagaggaaattcgcgaagcgttccgcgtgtttgacccggataac
gacggcaccctggacaagaaagagcttcgccacgtgatgacagaccttggagagaagttaacagatgaggaggttgat
gaaatgatcagggtagcaaacccggataacgacggcaccattgatgcgcgtgaatttgtacaaatgatgacagcgaag
aagggcgagctcaattcgaagcttgaaggtaagcctatccctaaccctctcctcggtctcgattctacgcgtaccggtcatc
atcaccatcaccat

PLASMID MAPS

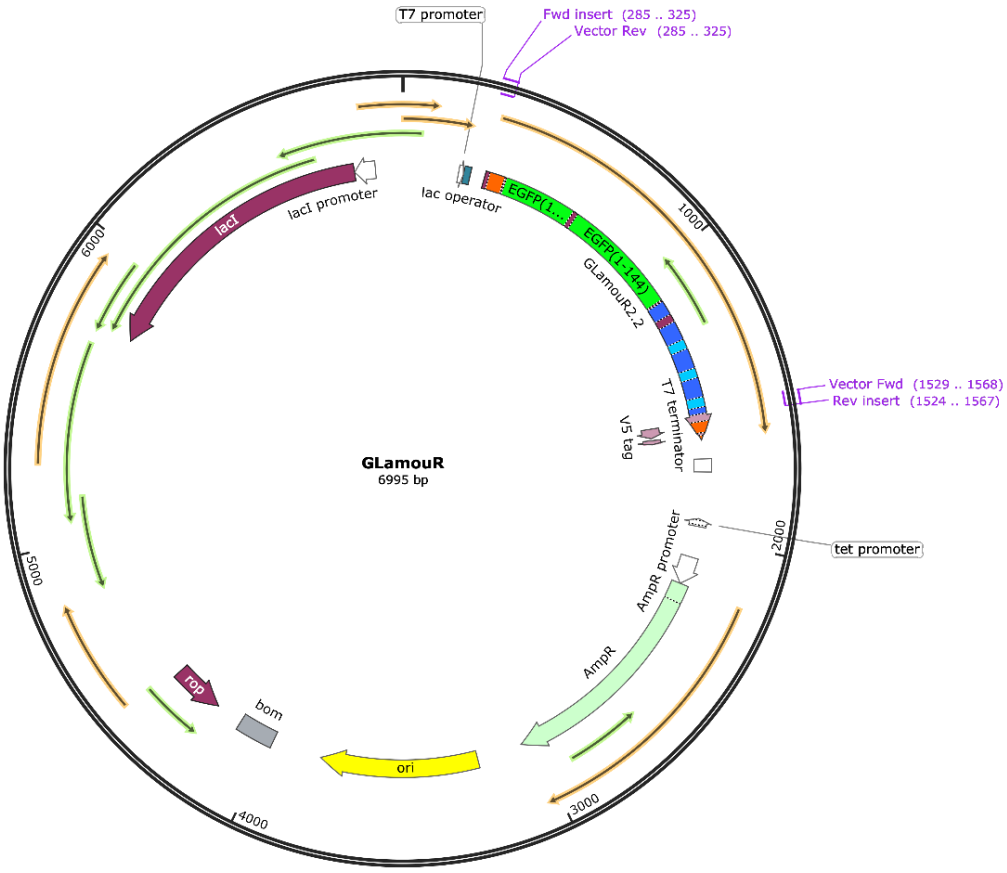

FIGURE 16. Plasmid map of GLamouR.

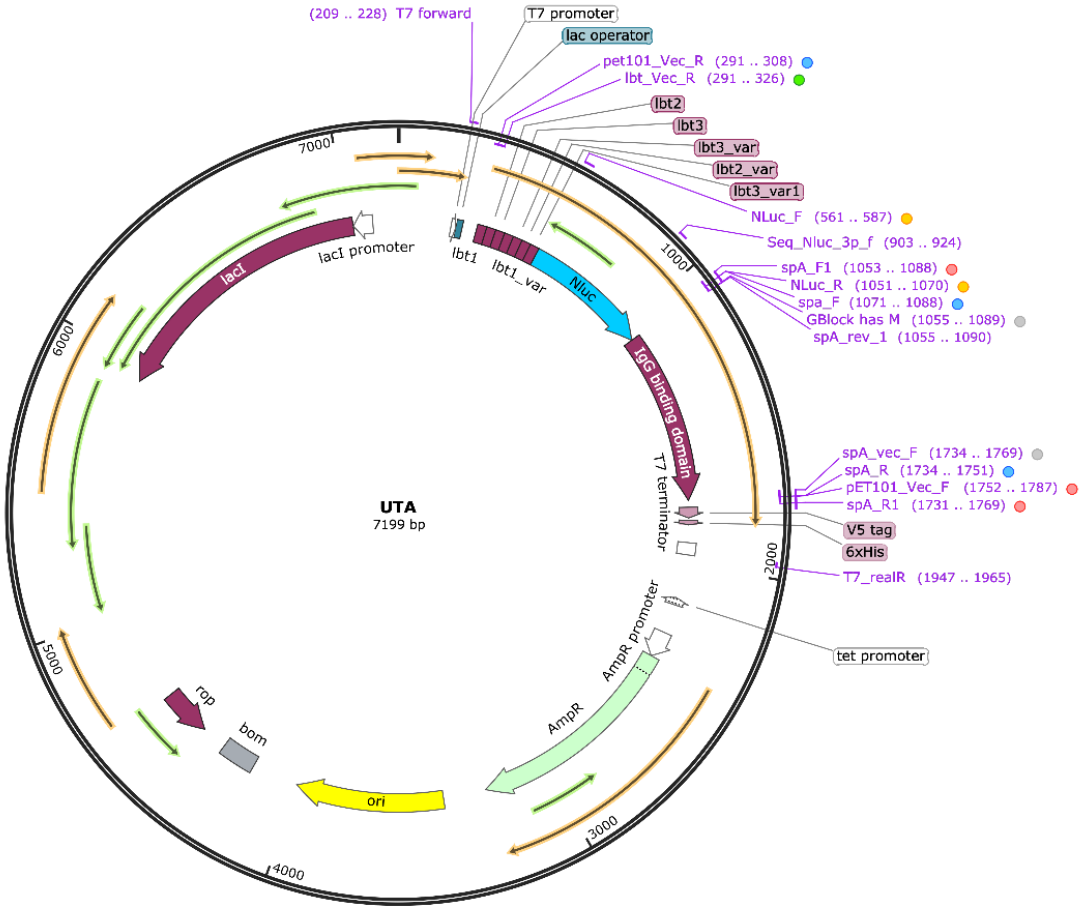

FIGURE 17. Plasmid map of UTA.

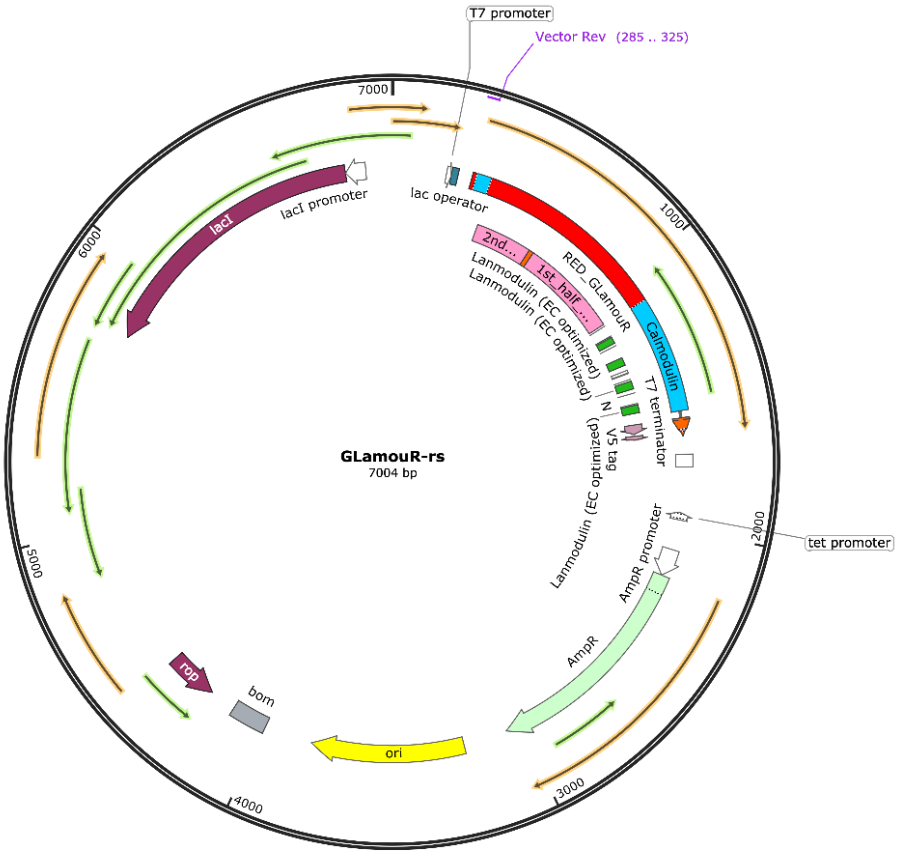

FIGURE 18. Plasmid map of GLamouR-rs.

PATENTS, PRESENTATIONS, AND PUBLICATIONS

Patents

(International) Patent Pending: PCT/US22/*****, “Lanmodulin-Based Protein”. <Lee, Grady, Gilad>

Notable Presentations

- The International Society for Magnetic Resonance in Medicine (ISMRM) - Poster Presentation, Molecular & Cellular Imaging SG Award
- World Molecular Imaging Congress (WMIC) – Poster Award Finalist
- Korean Society for Biomaterials - Poster Presentation, Best Paper Award
- MSU Imaging Research Symposium - Oral Presentation, featured on the Advanced Imaging Facility Website
- National Science Foundation BEACON Congress – Oral Presentation

Publications

1. **Lee HD**, Grady CJ, Krell K, Strebeck C, Good NM, Martinez-Gomez NC, Gilad AA*. A Novel Protein for the Bioremediation of Gadolinium Waste. PMID: 36711778; PMCID: PMC9881998. *in revision, Nat. Commun.*

2. Fillion AJ, Bricco AR, **Lee HD**, Korenchan D, Farrar CT, Gilad AA. Development of a Synthetic Biosensor for Chemical Exchange MRI Utilizing In-Silico Optimized Peptides. bioRxiv; 2023 Mar 8. doi: https://doi.org/10.1101/2023.03.08.531737

3. Good NM, **Lee HD**, Hawker ER, Su MZ, Gilad AA. Martinez-Gomez NC*. Hyperaccumulation of Gadolinium by Methylorubrum extorquens AM1 Reveals Impacts of Lanthanides on Cellular Processes Beyond Methylotrophy. Front. Microbiol., 2022, 13.

4. Hwang J, Son J, Seo Y, Jo Y, Lee K, **Lee D**, Khan MS, Chavan S, Park C, Sharma A, Gilad AA*, Choi J*. Functional silica nanoparticles conjugated with β-glucan to deliver anti-tuberculosis drug molecules, J Ind Eng Chem, 2018, 53, 376-385.

5. **Lee D**†, Seo Y†, Hwang J, Jo Y, Son J, Lee K, Khan MS, Chavan S, Park C, Gilad AA*, Choi J*. Use of nanoscale materials for the effective prevention and extermination of bacterial biofilms, Biotechnol. Bioprocess Eng., 2018, 23, 1-10.

6. Son J, Hwang J, **Lee D,** Khan MS, Jo Y, Lee K, Park C, Chavan S, Seo Y, Choi Y, Kim S, Kim D, Na D*, Choi J.* Strategiesfor the optimization of bead-immunoassays for the effective detection of target biomolecules, Korean J Chem Eng, 2018, 35(3), 805-811

7. **Lee D**†, Hwang J†, Seo, Y, Gilad AA*, Choi J*. Optical immunosensors for the efficient detection of target biomolecules, Biotechnol. Bioprocess Eng., 2018, 23(2), 123-133.

8. Seo Y†, Park C†, Son J, Lee K, Hwang J, Jo Y, **Lee D**, Khan MS, Chavan S, Choi Y, Kim S, Kim D, Gilad AA*, Choi J.* Synthesis of multi-walled carbon nanotubes with silver nanoparticles and evaluation of their antibacterial activities and cytotoxic properties, J. Vis. Exp., 2018, (135), e57384

9. **Lee D†**, Hwang J†, Son J, Seo Y, Jo Y, Lee K, Park C, Choi J*. Engineered nanomaterials for their applications in theranostics, J Ind Eng Chem, 2018, 66, 20-28