Dickinson College
Chemistry Department
Data Analytics Department

Finding Orthogonal Bioluminescent Probes via Statistical Model with the ultimate goal of Multicomponent Imaging

by Suong B. A. Tran

Professor Colin Rathbun, Supervisor - Chemistry Department
Professor Eren Bilen, Supervisor - Data Analytics Department

Carlisle, Pennsylvania
May 2023

Contents

Introduction **3**

Methods **5**

Statistical model to predict selective-binding mutations of LgBiT 5

Testing "hits" from the developed model 5

Results and Discussion **5**

Predicted single-mutation of LgBiT via GLM 6

Single-mutation of LgBiT results in SmBiT binding difference 7

Conclusion **8**

List of Figures

1 Luciferase Reaction and Multicomponent Bioluminescence Imaging 3
2 Bioluminescene Resonance Energy Transfer (BRET) 4
3 Predict selective-binding mutation of LgBiT by GLM 6
4 Testing predicted mutants of LgBiT 7

Introduction

Bioluminescent imaging has been developed over the last decade as a tool for studying and analyzing biological functions, which demands probes that can report on processes over time within living organisms. It offers sensitive imaging in cells and tissues without the need for surgery.[1] With bioluminescent tools, researchers can "see" biological processes in real-time, providing critical insights into the functions of living organisms. Bioluminescence occurs when a luciferase enzyme catalyzes the oxidation of its luciferin substrate, resulting in the visible production of light (Figure 1A).[2] Unlike other imaging techniques, such as fluorescence imaging, bioluminescence offers various advantages, including lower background signal and no need for excitation light.[3] Bioluminescence imaging is a highly versatile technology that has gained popularity in cancer research and studying complex interactions (e.g. host-pathogen) *in vivo*.[4] It can also be employed in numerous applications, such as investigating protein-protein and protein-ligand interactions, exploring gene regulation, and studying cell signaling.[5, 6]

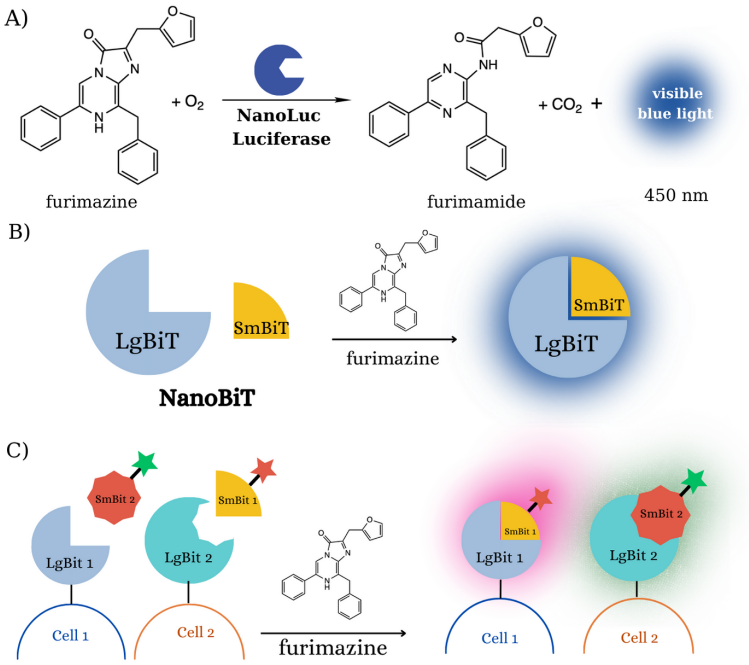

Figure 1: (A) Luciferase reaction of NanoLuc in the presence of furimazine. (B) The association of SmBiT and LgBiT in the presence of furimazine generates blue light. (C) Multi-component imaging of two different cell types with engineered NanoBiT enzymes will require the SmBiT-LgBiT binding interaction and tuning emission wavelength with fluorophores.

The major limitation of bioluminescence imaging *in vivo* is that it can only image one population of cells at a time. The development of a multicomponent bioluminescence imaging tool would enable researchers to study complex interactions in tandem. With at least two new colors, the bioluminescent tools could be developed to enable visualizing multiple cell types in mouse models (Figure 1C). Because of their utility, enhanced bioluminescence imaging techniques are needed to allow researchers to "see" new biological processes.

To develop enhanced bioluminescent tools, it is important to start with a bright and robust luciferase enzyme. We chose NanoLuc, the newest commercially available luciferase enzyme, derived from the deep-sea shrimp *Oplophorus gracilirostris*. It emits glow-type luminescence 150-fold greater than firefly-based enzymes.[7] Particularly, we are interested in NanoBiT, a split reporter system based on the luciferase enzyme NanoLuc. NanoBiT is composed of two subunits: a small peptide consisting of 11-13 amino acids (SmBiT) and a larger 18 kDa enzyme (LgBiT). Similiar to NanoLuc, the association of SmBiT and LgBiT gives bioluminescence with blue light in the presence of furimazine – the luciferase substrate (Figure 1B).[8] NanoBiT was chosen for our research because of the ease in modulating the bioluminescence properties only by modifying its subunit - SmBiT,

instead of dealing with the entire luciferase enzyme.

In addition to the development of multicomponent bioluminescence imaging techniques, it is significant to note another drawback of NanoBiT that its emitted light is blue (450 nm). Within the visible spectrum, hemoglobin is the primary chromophore in blood cell that absorbs light. Hemoglobin absorbs primarily in the blue and green part of the visible spectrum.[9] This phenomenon can be attributed to short-wavelength light, such as blue light, is less permeable through tissues than longer-wavelength light.[9, 10] Therefore, to enable the visualization of cell populations in live mice, it is necessary to red-shift the color of light emission of NanoBiT probes. To modulate the light emission color of NanoBiT, we will take advantage of bioluminescence resonance energy transfer (BRET). BRET is the process of dipole-dipole non-radiative energy transfer from a luciferase energy donor to an acceptor fluorophore.[11] Previous work concatenated NanoLuc with fluorescent proteins to red-shift emission via BRET.[12] Another study showed that the color of NanoLuc is also tunable via BRET with small-molecule fluorophores via HaloTag and SNAP-tag probes.[13] However, these techniques require bulky proteins, and do not enable true multi-component imaging. Based on the principle of BRET, we have developed an easier technique to tune the color of bioluminescence by taking advantage of BRET and NanoBiT. We modulated the emission wavelength of NanoBiT by chemically appending a fluorophore to the N-terminus of SmBiT peptide (Figure 2). In the presence of furimazine, the association of SmBiT and LgBiT enabled BRET, shifting the emitted blue light to near-red light. With this technique, I successfully modulated the color of NanoBiT from blue to three new colors (cyan, green, and yellow).

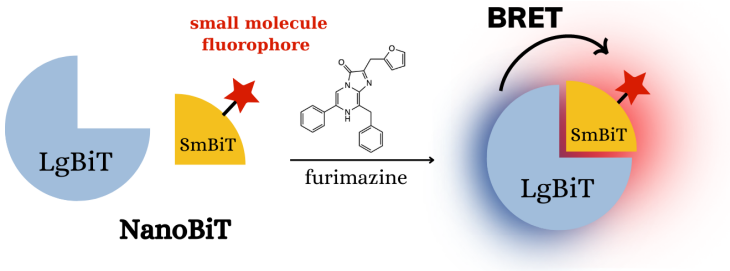

Figure 2: SmBiT and LgBiT bind together in the presence of furimazine, and blue light is emitted. Resonance nergy is transferred to the fluorophore which then emits a longer wavelength of light.

To enable the use of these probes in multicomponent imaging, our laboratory has introduced mutations to the amino acids at the binding pocket of LgBiT to selectively bind each SmBiT. We utilized deep mutational scanning to analyze the frequency of each mutation appeared before and after binding to two different SmBiTs.[14] We developed generalized linear model (GLM), a statistical model, to predict the contribution of binding preference between two SmBiTs. Predicted mutations preferring specific SmBiTs are also known as "hits". Investigating the light emission of significant "hits" predicted by the model with the equal combination of two SmBiTs showed that even single mutations of LgBiT can result in distinction of binding preference between SmBiTs.***This study has demonstrated that we are able to red-shift the emitted light of NanoBiT and develop an easy-to-use technique for multicomponent bioluminescence imaging.***

Methods

Statistical model to predict selective-binding mutations of LgBiT

To utilize the NanoBiT-fluorophore conjugates for multi-component imaging, a statistical model was developed to finding LgBiTs that selectively bind specific SmBiTs. SmBiT 1 (VSGWRLFKKIS) and SmBiT 2 (VEGYRLFEKIS) were screened separately against the library of LgBiT mutants and their sequences were determined by high throughput sequencing. Deep mutational scanning was used to analyzed the observed sequences, and how frequent each mutation appeared before (Count library) and after (Count SmBiT) binding to SmBiT.[14, 15] Generalized Linear Model (GLM) was adapted to better estimate the contribution of each mutation to the binding difference between SmBiT 1 and SmBiT 2 based on their sequence count, normalized to the count of wild type (WT).[16, 17] The difference in observed count after binding each SmBiT was calculated by:

$$\text{Difference} = \frac{\frac{\text{Count}_{\text{Mutation\_SmBiT2}}}{\text{Count}_{\text{WT\_SmBiT2}}} - \frac{\text{Count}_{\text{Mutation\_SmBiT1}}}{\text{Count}_{\text{WT\_SmBiT1}}}}{\frac{\text{Count}_{\text{Mutation}}}{\text{Count}_{\text{WT}}}}$$

$\text{Count}_{\text{Mutation\_SmBiT1}}$ and $\text{Count}_{\text{WT\_SmBiT1}}$ are, respectively, the observed counts of LgBiT mutant and WT when they were screened with SmBiT 1; similar to SmBiT 2. $\text{Count}_{\text{Mutation}}$ and Count_{WT} are, respectively, the observed counts of LgBiT mutant and WT when they were screened alone. GLM was applied on provided data using the following call in R.

```
> glm( changeB2B1~pos*mut, data, family ='gaussian')
```

All experiment data and coding scripts for GLM can be found on GitHub via the link (*https://github.com/tranJen/RaBi-Collaboration*). The README file will be helpful to navigate the data.

Testing "hits" from the developed model

Flourophore Cou343 (λ_{em} = 490 nm, cyan light) was appended on SmBiT 1 and fluorophore TAMRA (λ_{em} = 580 nm, yellow light) was appended on SmBiT 2. The similar experiment of bioluminescence imaging described above was executed. Instead of adding only one SmBiT to each well on 96-well plate, two SmBiT-flourophores were combined equally to each well with the total concentration of 100 nM, and a total volume of 50 µL.

Results and Discussion

I successfully tuned the bioluminescence of NanoBiT from blue to three new colors (cyan, green, and yellow) and proved that near-red light emissions are more permeable through tissue. *Further discussion in chemistry synthesis of NanoBiT-fluorophore conjugates and their permeability through tissue can be found in my Chemistry Honors Thesis.*

Here in this report, I just summarize related data of the statistical model with the ultimate goal of multicomponent imaging.

Predicted contribution of single-mutation of LgBiT via GLM

In order for these NanoBiT-fluorophore conjugates to be used in multi-component imaging, it is important to have LgBiTs that only bind certain SmBiTs. Our lab has been mutating amino acids at the binding pocket of LgBiT to have a selective binding for each SmBiT.

We have been investigating SmBiT 1 (VSGWRLFKKIS) and SmBiT 2 (VEGYRLFEKIS). SmBiT 1 (K_D = 700 pM) has a lower dissociation constant, indicating a higher binding affinity to native LgBiT compared to SmBiT 2 (K_D = 1.3 μM). This also implies that the association of SmBiT 1 and LgBiT generates a stronger light emission than that of SmBiT 2. Each of two other lab members, Abigail McGahan and Roberta Akrong, evaluated a library containing 736 LgBiT single-mutants at the binding pocket on the C-terminus. SmBiT 1 and SmBiT 2 were screened separately against the library of LgBiT mutants and their sequences were determined by high throughput sequencing. Deep mutational scanning was used to analyzed the observed sequences, and how frequent each mutation appeared before and after binding to each SmBiT.[14] We hypothesize that the difference in the number of mutants count for each SmBiT would significantly affect the selectivity for each SmBiT.

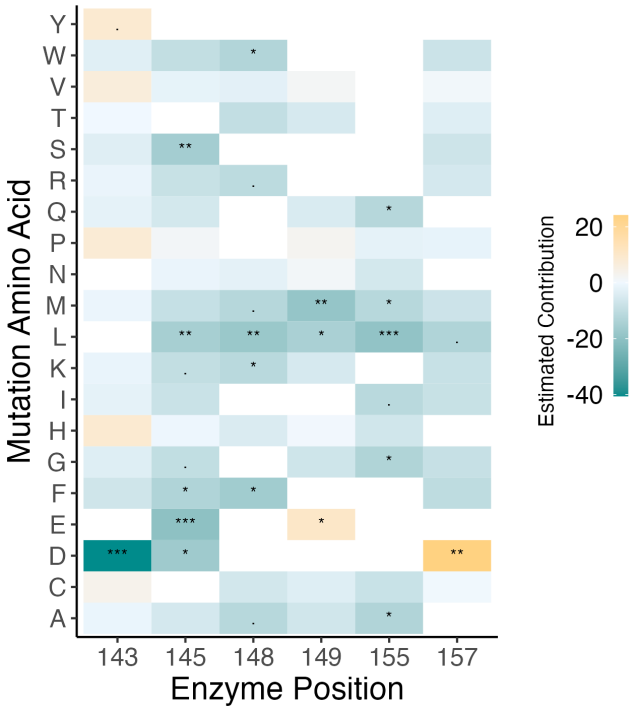

Figure 3: Heat map of the contribution of each mutation at each position screened to the binding difference between SmBiT 1 (cyan) and and SmBiT 2 (yellow) predicted by generalized linear model (GLM). *Significant level code: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1.*

After high throughput sequencing the library, the data set included 429 observations of LgBiT mutants. Generalized Linear Model (GLM) was developed to handle this large data set and predict the contribution of each mutation to the binding preference between SmBiT 1 and SmBiT 2.[16] GLM is a statistical framework for modeling relationships between explanatory variables and response variables. The modelling was developed in collaboration with Professor Eren Bilen from the Data Analytics Department.

After running the analysis, the model predicted several mutations contributing to the binding difference between SmBiT 1 and and SmBiT 2 (Figure 3). At the 0.01 level of significance, the mutants hypothesized to prefer binding SmBiT 1 (cyan block) were L143D, T145E, T145L, T145S, G148L, S149M, T155L. The mutant hypothesized to prefer binding SmBiT 2 (yellow block) was N157D.

Single-mutation of LgBiT results in SmBiT binding difference

Abigail McGahan and Professor Rathbun cloned several of these mutants of interest for further testing. Due to the unavailability of all mutations, only three mutants were used to test the prediction of the GLM, which were T145E, T145S, N157D. Even though mutant T145S was not predicted to as significantly contribute to binding difference between SmBiT 1 and SmBiT 2 as mutant T145E, this mutant was examined to see if the model correctly predicted the relative contribution to the difference. Mutant T145E was predicted to have larger contribution than mutant T145S.

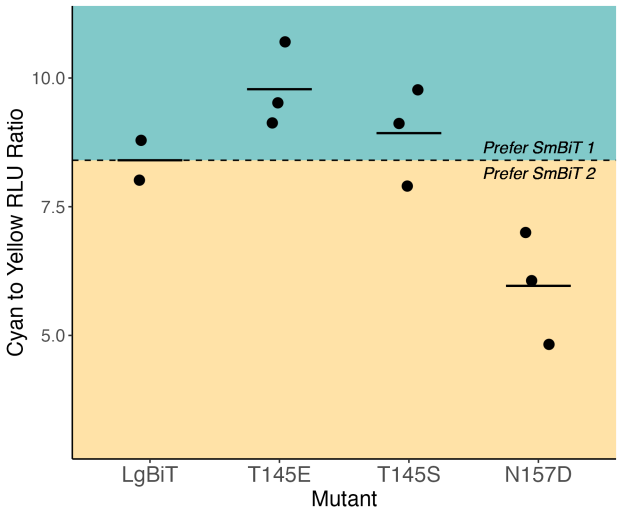

Figure 4: Mutation N157D prefers binding SmBiT 2. Mutation T145E and T145S prefer binding SmBiT 1, and T145E causes a larger difference. The dashed line corresponds to the binding preference of wild type.

To test if the GLM predicted correctly the contribution of each LgBiT mutation in binding difference between SmBiT 1 and SmBiT 2, an experiment of bioluminescence imaging was executed. Flourophore Cou343 (λ_{em} = 490 nm, cyan light) was appended on SmBiT 1 and fluorophore TAMRA (λ_{em} = 580 nm, yellow light) was appended on SmBiT 2. SmBiT 1 and SmBiT 2 were then equally added to the LgBiT and recorded spectra. Cyan to yellow luminescence ratio (the ratio of luminescence at 490 nm and at 580 nm), was used to analyze how the binding difference changed (Figure 4).

Even though the result indicated that all three mutations (T145E, T145S, N157D) selectively bound to SmBiT 1 more than SmBiT 2 because the ratio is larger than 1, they did show difference relative to wild type (Figure 6B, dashed line) in the binding preference toward SmBiT. Mutants T145S and T145E were predicted to prefer binding SmBiT 1, and indeed, we were able to make the mutations to prefer the SmBiT 1 more than the wild type because their cyan to yellow ratios were higher than that of wild type. Furthermore, the model accurately predicted the relative impact of mutations on the binding difference between SmBiT 1 and SmBiT 2. As expected, mutant T145E resulted in a greater increase in the cyan to yellow RLU ratio than mutant T145S. Mutants N157D was expected to have a preference for SmBiT 2, but it was still preferred SmBiT 1. Even though this mutation did not completely switch preference to SmBiT 2, the experiment showed that the cyan to yellow RLU ratio was decreased below that of the wild type, indicating that it caused the LgBiT to prefer SmBiT 1 less. Noticeably, a single-mutation of LgBiT has already resulted in binding preference toward SmBiT 2. Introducing multiple mutation would would result in a greater influence.

After observing that single mutations in the C-terminus of LgBiT can cause difference in the binding preferences between SmBiT 1 and SmBiT 2, we anticipate that combining multiple mutations could result in even larger differences. To validate the accuracy of the GLM predictions, we plan to test other significant mutants (Figure 3). Additionally, future studies will also focus on exploring mutations at the N-terminus of LgBiT.

Conclusion

Tested LgBiT mutants T145E, T145S, N157D behaved as predicted by GLM. To ensure the accuracy of the model's predictions for all mutants, further significant mutants will be tested in future studies. With the ultimate goal of developing multi-component imaging technique, a combination of predicted mutants will be studied in order to enhance the binding preference between SmBiT 1 and SmBiT 2. Furthermore, the binding selectivity between two SmBiTs can be improved by exploring mutations on the N-terminus of LgBiT.

We believe our findings hold great significance for advancing bioluminescence imaging. The development of small and easy-to-use red NanoLuc-based probes enables multicomponent imaging, which can facilitate the study of complex interactions in living organisms. Ultimately, this research may shed light on the study of interaction between cancer cells and the immune system *in vivo*.

References

(1) Paley, M. A.; Prescher, J. A. Bioluminescence: a versatile technique for imaging cellular and molecular features. *MedChemComm* **2014**, *5*, 255–267.

(2) Roda, A., *Chemiluminescence and bioluminescence: past, present and future*; Royal Society of Chemistry: 2011.

(3) Choy, G.; O'Connor, S.; Diehn, F. E.; Costouros, N.; Alexander, H. R.; Choyke, P.; Libutti, S. K. Comparison of noninvasive fluorescent and bioluminescent small animal optical imaging. *Biotechniques* **2003**, *35*, 1022–1030.

(4) Lewis, M. D.; Fortes Francisco, A.; Taylor, M. C.; Burrell-Saward, H.; McLatchie, A. P.; Miles, M. A.; Kelly, J. M. Bioluminescence imaging of chronic T rypanosoma cruzi infections reveals tissue-specific parasite dynamics and heart disease in the absence of locally persistent infection. *Cellular Microbiology* **2014**, *16*, 1285–1300.

(5) Song, G.; Wu, Q.-P.; Xu, T.; Liu, Y.-L.; Xu, Z.-G.; Zhang, S.-F.; Guo, Z.-Y. Quick preparation of nanoluciferase-based tracers for novel bioluminescent receptor-binding assays of protein hormones: using erythropoietin as a model. *Journal of Photochemistry and Photobiology B: Biology* **2015**, *153*, 311–316.

(6) Yasuzaki, Y.; Yamada, Y.; Ishikawa, T.; Harashima, H. Validation of mitochondrial gene delivery in liver and skeletal muscle via hydrodynamic injection using an artificial mitochondrial reporter DNA vector. *Molecular Pharmaceutics* **2015**, *12*, 4311–4320.

(7) Hall, M. P.; Unch, J.; Binkowski, B. F.; Valley, M. P.; Butler, B. L.; Wood, M. G.; Otto, P.; Zimmerman, K.; Vidugiris, G.; Machleidt, T., et al. Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. *ACS Chemical Biology* **2012**, *7*, 1848–1857.

(8) Dixon, A. S.; Schwinn, M. K.; Hall, M. P.; Zimmerman, K.; Otto, P.; Lubben, T. H.; Butler, B. L.; Binkowski, B. F.; Machleidt, T.; Kirkland, T. A., et al. NanoLuc complementation reporter optimized for accurate measurement of protein interactions in cells. *ACS Chemical Biology* **2016**, *11*, 400–408.

(9) Zhao, H.; Doyle, T. C.; Coquoz, O.; Kalish, F.; Rice, B. W.; Contag, C. H. Emission spectra of bioluminescent reporters and interaction with mammalian tissue determine the sensitivity of detection in vivo. *Journal of Biomedical Optics* **2005**, *10*, 41210–41210.

(10) Wang, J.; Liu, G.; Cham-Fai Leung, K.; Loffroy, R.; Lu, P.-X.; Wang, X. J., et al. Opportunities and challenges of fluorescent carbon dots in translational optical imaging. *Current Pharmaceutical Design* **2015**, *21*, 5401–5416.

(11) Dale, N. C.; Johnstone, E. K.; White, C. W.; Pfleger, K. D. NanoBRET: the bright future of proximity-based assays. *Frontiers in Bioengineering and Biotechnology* **2019**, *7*, 56.

(12) Suzuki, K.; Kimura, T.; Shinoda, H.; Bai, G.; Daniels, M. J.; Arai, Y.; Nakano, M.; Nagai, T. Five colour variants of bright luminescent protein for real-time multicolour bioimaging. *Nature Communications* **2016**, *7*, 13718.

(13) Hiblot, J.; Yu, Q.; Sabbadini, M. D.; Reymond, L.; Xue, L.; Schena, A.; Sallin, O.; Hill, N.; Griss, R.; Johnsson, K. Luciferases with tunable emission wavelengths. *Angewandte Chemie* **2017**, *129*, 14748–14752.

(14) Fowler, D. M.; Fields, S. Deep mutational scanning: a new style of protein science. *Nature Methods* **2014**, *11*, 801–807.

(15) McGahan, A. F.; Rathbun, C. Identifying mutations of NanoBiT to increase selectivity for multicomponent cellular tracking. *Dickinson BCMB Thesis* **2022**.

(16) Unger, E. K.; Keller, J. P.; Altermatt, M.; Liang, R.; Matsui, A.; Dong, C.; Hon, O. J.; Yao, Z.; Sun, J.; Banala, S., et al. Directed evolution of a selective and sensitive serotonin sensor via machine learning. *Cell* **2020**, *183*, 1986–2002.

(17) Melnikov, A.; Rogov, P.; Wang, L.; Gnirke, A.; Mikkelsen, T. S. Comprehensive mutational scanning of a kinase in vivo reveals substrate-dependent fitness landscapes. *Nucleic Acids Research* **2014**, *42*, e112–e112.