

UNIVERSITY OF CALIFORNIA,
IRVINE

Parallel screening for rapid identification of orthogonal bioluminescent tools

DISSERTATION

submitted in partial satisfaction of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

in Organic Chemistry

by

Colin Michael Rathbun

Dissertation Committee:
Associate Professor Jennifer Prescher, Chair
Professor Scott Rychnovsky
Professor Chris Vanderwal

2012

DEDICATION

(Optional dedication page) xxx To ...

TABLE OF CONTENTS

	Page
LIST OF FIGURES	iv
LIST OF TABLES	v
ACKNOWLEDGMENTS	vi
CURRICULUM VITAE	vii
ABSTRACT OF THE DISSERTATION	ix
1 Parallel Screening for Rapid Identification of Orthogonal Bioluminescent Tools	1
1.1 Introduction	1
1.2 Results and discussion	3
1.3 Conclusions	11
A Appendix Title	12
A.1 Lorem Ipsum	12

LIST OF FIGURES

Page

LIST OF TABLES

Page

ACKNOWLEDGMENTS

I would like to thank the National Science Foundation for funding through the Graduate Research Fellowship Program (grant No. DGE-1321846, and Allergan for funding through the Allergan Graduate Fellowship. I would also like to thank the members of the Prescher lab, Bernard Choi, and Bruce Tromberg, for very helpful discussions. xxx

You also need to acknowledge any publishers of your previous work who have given you permission to incorporate that work into your dissertation. See Section 3.2 of the UCI Thesis and Dissertation Manual.)

CURRICULUM VITAE

Colin Michael Rathbun

EDUCATION

Doctor of Philosophy in Computer Science	2012
University name	<i>City, State</i>
Bachelor of Science in Computational Sciences	2007
Another university name	<i>City, State</i>

RESEARCH EXPERIENCE

Graduate Research Assistant	2007–2012
University of California, Irvine	<i>Irvine, California</i>

TEACHING EXPERIENCE

Teaching Assistant	2009–2010
University name	<i>City, State</i>

REFEREED JOURNAL PUBLICATIONS

Ground-breaking article

2012

Journal name

REFEREED CONFERENCE PUBLICATIONS

Awesome paper

Jun 2011

Conference name

Another awesome paper

Aug 2012

Conference name

SOFTWARE

Magical tool

<http://your.url.here/>

C++ algorithm that solves TSP in polynomial time.

ABSTRACT OF THE DISSERTATION

Parallel screening for rapid identification of orthogonal bioluminescent tools

By

Colin Michael Rathbun

Doctor of Philosophy in Organic Chemistry

University of California, Irvine, 2012

Associate Professor Jennifer Prescher, Chair

The abstract of your contribution goes here.

Chapter 1

Parallel Screening for Rapid Identification of Orthogonal Bioluminescent Tools

1.1 Introduction

Our understanding of living systems is profoundly shaped by our ability to see biology in action. Central to these efforts are robust and translatable imaging tools.(1, 2) Decades of work to engineer and optimize fluorescent proteins have provided a palette of designer probes for cellular studies. Using combinations of these tools, it is now possible to trace the orchestrated behaviors of immune cells,(3) nerve cell connections,(4) and other interactions.(5) Widespread application of fluorescent probes will continue to reveal unanticipated facets of biology. Concurrently, gaps in our knowledge will spur the development of innovative tools. Fluorescent probes endowed with novel functions (altered colors, photoswitches, etc.) are already enabling new pursuits.(6, 7)

Biological discoveries will be further bolstered by advances in bioluminescent probe development. Bioluminescence relies on light generation via luciferase enzymes and luciferin small molecules.(8, 9) Since no excitation light is required, this modality is attractive for studies involving large length or time scales.(10-12) Indeed, firefly luciferase (Fluc) and its cognate substrate (d-luciferin) are ubiquitously used in rodent models to interrogate molecular and cellular events.(13, 14) Fluc and its homologues have been further engineered to provide different colors of light.(15, 16) Such tools have been applied for multicomponent imaging in vitro. The wavelengths achieved, though, are unsatisfactory for routine use in vivo. Multicomponent imaging with even the most spectrally resolved probes remains challenging due to interference from surrounding tissue.(10) As a consequence, bioluminescence has lagged behind fluorescence for multicellular studies in tissues and whole organisms.

To realize multicomponent imaging in vivo, we turned to a more tractable parameter: substrate selectivity. Enzymes exhibiting mutually exclusive (i.e., orthogonal) substrate preferences should be readily distinguished in a variety of biological models. Fluc is remarkably tolerant of a variety of luciferin modifications, including both electronic(17-21) and steric(22-26) derivatives. While dozens of luciferin analogues have been crafted, most result in reduced photon outputs relative to d-luciferin, the native substrate, at saturating doses.(27) In some cases, boosts in light emission have been achieved using modified versions of the enzyme.(23, 25) These results set the stage for developing designer luciferase?luciferin pairs, but few methods to systematically generate orthogonal sets have been pursued.(25, 28) Substrate-selective luciferases are found in nature, and a handful have been coopted for dual imaging (e.g., combinations of d-luciferin- and coelenterazine-utilizing enzymes).(29-32) However, most remain suboptimal for use in vivo. The characterization of other naturally occurring luciferases and luciferins has also not kept pace with the demand for new pairs. Consequently, bioluminescence imaging has been limited by a lack of mutually orthogonal enzymes and substrates.

We aimed to expedite the search for luciferases that exhibit unique preferences for distinct luciferin analogues. Accessing enzymes with alternative substrate use is well preceded in directed evolution.(33-35) However, traditional applications of this technique have focused on optimizing one enzyme at a time. Selectivity for one molecule over another is often realized as a consequence, but is not typically the parameter being screened.(36-41) Here, we present a general and rapid approach to achieve substrate selectivity and engineer orthogonal luciferase?luciferin pairs. This strategy relies on parallel screening of functional luciferases with collections of chemically diverse luciferins (Figure 1a). The large data sets are then mined for orthogonal combinations using a custom computer script. Enzyme?substrate pairs are deemed orthogonal if robust reactivity is observed when complementary partners interact, but minimal to no reactivity is observed in all other cases (Figure 1a). Collectively, we screened 159 mutants and 12 analogues, generating a candidate list of greater than 800,000 possible pairs. We evaluated the orthogonality of 175 pairs in vitro. A subset was successfully applied in cultured cell and animal models, highlighting the feasibility and translatability of the approach. We also analyzed principles governing selective substrate use and identified methods to search for expanded collections of orthogonal imaging agents. Overall, this work greatly expands the number of viable bioluminescence probes for multicomponent imaging and presents a strategy to accelerate the identification of new ones. The parallel screening method is also applicable to other areas where selective substrate use is required.

1.2 Results and discussion

Expanding the Pool of Candidate Luciferins and Luciferases

As a starting point for substrate modification, we focused on d-luciferin derivatives with steric appendages at C4? and C7?. These positions lie in close proximity to the Fluc backbone,(42)

and preliminary work revealed that modifications here do not quench or otherwise impede photon emission.⁽²⁵⁾ We also previously identified a pair of luciferases that could discriminate between luciferins with modifications at these positions,⁽²⁵⁾ suggesting that they were good starting points for new probe development. However, attempts to optimize this pair via traditional directed evolution (focusing on one enzyme at a time) did not result in improved substrate selectivity (Figure S1).

We reasoned that screening for selectivity at the outset would provide a more rapid route to new bioluminescent pairs. Engineering luciferases to discriminate among structurally similar compounds can be difficult.^(23, 43) Thus, we initially focused on diversifying the enzyme and substrate inputs. Collections of both new and known^(24, 25) luciferins were assembled (Figure 1b). These molecules covered a broad range of chemical space and comprised both hydrophilic and hydrophobic functional groups. The luciferins were benchmarked for light emission with Fluc (Figure S2). All compounds were functional light emitters, though they varied in terms of photon output. Some level of enzyme activity is necessary for successful evolution, but weak performers can be advantageous starting points for evolving new functions.⁽³³⁾

In parallel with luciferin diversification, we targeted broad sectors of Fluc sequence space for mutagenesis. Twenty-three residues near the active site were selected, and the mutations were covered in 8 libraries (labeled in Figure 1c). The majority of the mutants would likely be nonfunctional, and thus not ideal starting points for probe development. We aimed to eliminate these luciferases early on and perform parallel screens with an enriched pool of viable mutants. Such an approach would save time and reagents as luciferases are not amenable to high-volume separations (e.g., FACS) or selections; rather, each mutant must be physically interrogated with a given substrate. We adapted a high-throughput method to traverse the luciferase libraries and cull nonfunctional members (Figure S3a).⁽²⁵⁾ The libraries were transformed into bacteria, and the transformants were grown on agar con-

taining one of four minimally perturbed luciferins: 4⁺/7⁻-BrLuc or 4⁺/7⁻-MeLuc (Figure 1b, Figure S3a). These analogues were selected for on-plate screens since they are among the “brightest” emitters and easy to access in bulk. Light-emitting colonies were picked and further assayed in lysate and by sequencing (Figure S3a). A variety of mutants were identified (Figure S4), including enzymes that were unique to each luciferin. Some hits were further diversified (1-3 generations) via random mutagenesis to enlarge the pool of luciferase mutants (Table S1 and Figure S3).

Screening for Orthogonal Luciferase-Luciferin Pairs in Silico

With enriched sets of functional luciferases, we aimed to screen the collection for orthogonal pairs. Testing each combination of two mutants and two substrates would have required 829,026 separate experiments (Figure 2a), an impractical number. Instead, we screened each analogue across the same panel of 159 luciferases, generating 1908 (12 substrates x 159 enzymes) individual data points (Figure S3b). An ideal orthogonal enzyme would be “positively” matched with a single substrate and “negatively” matched with all other luciferins. To identify such enzymes, we established a metric to quantify orthogonality and mine the data. We reasoned that perfect selectivity could be represented by an identity matrix (Figure S5, Supplementary Note). Orthogonality would be maximal if each enzyme was completely selective for its cognate substrate (represented by a “1” in the identity matrix) and nonfunctional with other luciferins (“0” in the identity matrix). An orthogonality score was determined by representing each set of two luciferases and two luciferins as a square matrix, with enzymes in rows and substrates in columns. These data were compared to the ideal case (identity matrix) via root-mean-square distance (RMSD). The RMSD values were then converted to numeric values (i.e., orthogonality scores) representing the fold resolution between the positive and negative pairings (see Supplementary Note for more details). We wrote a computer script to assemble each possible matrix from the screening data and cal-

culate the orthogonality of each pairing. The pairs were sorted by increasing RMSD, with the smallest value (highest orthogonality score) representing the most orthogonal pair.

The algorithm provided a ranked list of the 829,026 possible orthogonal sets (Figure 2a). The top pair comprised analogues 2 and 11 (4?-MorphoLuc and 7?-MorPipLuc) with mutants 81 and 104 (Figure 2a). Selective light emission with these enzymes and substrates was verified in vitro (Figure 2b). We further validated the top ten unique pairings on the ranked list, along with a handful of others in the data set (every tenth rank among the top 100, every 100th rank among the top 1000, and every 1000th rank down to position 5000). In all cases, orthogonality scores were measured in bacterial lysate (Figure 2c). Among the top 1000 pairs, 10-fold photon outputs were observed with the positively paired luciferase?luciferin set compared to the negatively paired set (Figure 2c). Diminishments in selectivity were observed farther down the list. These results suggest that the in silico rank order is a good predictor of orthogonal substrate use. The method also culled 99.9% (828,000 of the total 829,026) of irrelevant enzyme?substrate pairings (Figure S6), enabling fast convergence on important hits. As more luciferases and luciferins are screened, the data set can be expanded and continually mined for new orthogonal pairs.

Imaging with Orthogonal Pairs in Cultured Cell and Animal Models

We aimed to transition lead pairs from the screening analyses to mammalian cell imaging. In these more complex environments, issues of enzyme stability, substrate biocompatibility, and compound transport are of paramount concern. Fortunately, our approach to enriching functional luciferases preselects for luciferases and luciferins that are well behaved. Three of the top pairs from the script were analyzed in cultured cell (Figure S7) and animal models (Figure 3): (1) 4?-MorphoLuc/enzyme 81 (R218A, F250M, S314T, G316T)

with 7?-DMAMeLuc/enzyme 37 (R218K), (2) 7?-MeLuc/enzyme 87 (R218K, F250Y, S314T, G316T) with 4?-BrLuc/enzyme 53 (V240I, V241M, F243M, F247Y, S347G), and (3) 4?-BrLuc/enzyme 51 (F243M, S347G) with d-luciferin/enzyme 93 (R218K, M249L, S314T, G316S). These pairs were selected due to the ease of accessing the substrates, along with their relative brightness. The mutants were stably expressed in DB7 mouse mammary carcinoma cells. The cells were treated with relevant luciferins and imaged (Figure S7). Substrate specificity was maintained in all cases, highlighting the success of the parallel screening method.

Selectivity was also maintained in vivo. DB7 cells expressing the relevant mutants were implanted in FVB/NJ mice. Subsequent administration of the complementary luciferin analogues resulted in light emission for positively paired compounds with minimal cross reactivity (Figure 3). These images mark an initial demonstration of dual imaging with purely engineered luciferase?luciferin pairs. It is also important to note that perfect resolution is not required for multicomponent imaging applications. Rather, patterns of substrate use can serve as diagnostic fingerprints.⁽⁴⁴⁾ Photon outputs from the orthogonal pairs are in a useful range for monitoring bulk cell populations. The dimmest set (enzyme 37/7?-DMAMeLuc and enzyme 81/4?-MorphoLuc) emits enough photons to visualize $\sim 6 \times 10^6$ cells in subcutaneous models. The other orthogonal sets are substantially brighter and can enable more sensitive imaging. Collectively, these data show that parallel screens and in silico analyses can be used to identify and transition orthogonal sets to a variety of biological models.

Analyzing Trends in Orthogonal Substrate Use

To gain insight into principles governing orthogonality, we undertook a detailed analysis of the screening results. The highest-ranked pairs comprised a variety of enzymes and substrates. Seven unique luciferins (from both the 4? and 7? series) were found among the

top 10 pairs, along with luciferases comprising mutations at 18 unique sites (Supplementary Data; Figure S8). The diversity in hits implies that there are a variety of paths to achieve substrate resolution. Among the pairs, orthogonality was primarily realized not by markedly enhanced turnover of a preferred substrate. Rather, selectivity arose from reduced photon production with other compounds. As shown in Figure 4a, matched enzymes and substrates (positive pairs) were on par with native Fluc in terms of photon output. The unmatched enzymes and substrates (negative pairs), by contrast, demonstrated reduced activities ($\sim 10^3$ -fold lower). Thus, in a given orthogonal pair, selectivity is mostly achieved by reducing light emission with the negatively paired compound versus selectively increasing light emission with the positively paired compound. For example, mutant 81 provides ~ 4 -fold enhanced light output with 4'-MorphoLuc compared to Fluc. With every other luciferin screened, including the negatively paired compound 7'-MorPipLuc, mutant 81 emits ~ 10 -fold fewer photons than the native enzyme. So while light output with 4'-MorphoLuc is slightly improved with mutant 81, the decrease in light emission observed with 7'-MorPipLuc (~ 100 -fold) contributes more to orthogonality.

Since compound selectivity appears to be achieved by destroying enzyme-substrate interactions, structurally related compounds would be expected to exhibit similar trends in orthogonality. Indeed, bulky 7'-modified compounds tend to positively pair with the same types of enzymes (Figure S9). Many of these mutants (e.g., mutant 104) likely harbor space in the active site to accommodate 7' substituents (Figure S9a). Conversely, 4'-modified luciferins tend to produce less light with these same mutants and are thus negatively paired. Structurally divergent compounds were also more likely to comprise an orthogonal pair (Figure 4b). For example, substrates with a modification at the 4' position were rarely orthogonal to other 4' compounds. It is probably difficult to destroy activity with one 4'-modified compound without impacting others in the same series. When 4'- and 7'-modified substrates are paired, though, each substrate likely interacts differently with the enzyme, making it

easier to achieve orthogonality. These results suggest a strategy for continued orthogonal bioluminescent probe development: incorporate more diverse analogues in parallel screens.

Some compounds appear uniquely suited for orthogonal probe development. For example, 4?-BrLuc shows up nearly twice as often in the top 1,000 hits compared to other compounds (Figure S10). The mechanistic basis for this preference is unclear. The bromine substituent is roughly the same size as the methyl group in 4?-MeLuc, negating a pure steric argument. 4?-MeOHLuc is predicted to form orthogonal pairs (albeit less frequently) with similar compounds as 4?-BrLuc (Figure 4), suggesting that polarizable substituents might be preferred. Heavy halogen atoms (e.g., Br) are also known to quench the fluorescence of some molecules via intersystem crossing.⁽⁴⁵⁾ Thus, certain 4?-BrLuc conformations could result in internal quenching (and poor light emission) and thus pair negatively with several mutants. Additional compound screens and analyses will be necessary to discriminate among these possibilities and gain more insight.

We further analyzed the frequency of positive and negative pairings between luciferins and individual residues (Figure 5). Luciferases with mutations at residues 240, 247, or 347 seemed to prefer 4?-modified compounds. These residues are known to modulate the binding and light emission of the native substrate, d-luciferin.^(43, 46-48) Docking studies corroborated these findings, suggesting that the mutations (e.g., S347G in mutants 51 and 53) likely create space for bulky substituents (e.g., 4?-BrLuc, Figure S11). These residues are also negatively paired with most of the 7?-modified compounds, suggesting that they are good candidates for future orthogonal probe design. Surprisingly few hot spot residues correlated with selective 7? analogue use (Figure 5). Fluc residues near C7? primarily comprise backbone amides.^(42, 49) Thus, it is unclear how specificity for these analogues might arise.

Added Diversity Improves Orthogonality

Multicomponent imaging requires not just pairs of orthogonal enzymes and substrates, but also triplets, quadruplets, and higher order sets. Identifying such expanded collections requires structurally diverse enzyme and substrate architectures. If only a few privileged luciferases or luciferins from our data set could provide the desired selectivities, it would be difficult to achieve larger collections of orthogonal probes. To assess whether we were approaching an upper limit on orthogonality, we performed simulations within the existing data set. Random subsets of various sizes were selected from the full pool of substrates and enzymes. The sets were analyzed using the algorithm from above, and orthogonality scores were generated (Figure 6a). Regardless of the identities of enzymes or substrates used, scores increased with greater numbers of both enzymes (from 2 to 159) and substrates (from 2 to 12). This result implies that we have not reached a plateau in identifying orthogonal pairs. Exploring more sequence space with mutant luciferases and chemical space with modified luciferins should also improve the orthogonality of the top pairings.

As a next step, we modified the algorithm to search for not just two pairs of orthogonal probes but also triplets and multiple sets in general. A set of three adds significant complexity, as not only three positive pairings, but also six negative pairings, must be identified. From our current data set, this required sifting through 144 million combinations. We combed the original data set in search of three mutually orthogonal enzyme?substrate pairs. A total of 6171 potential sets were identified. The orthogonalities of the top ten were verified in bacterial lysate (Figure S11). The top triplet set is shown in Figure 6b and comprises two enzyme?substrate pairs previously validated *in vivo*. The overall orthogonality score for this set was lower than that of the individual pairs from above. However, this result represents key proof-of-concept and a starting point for the development of larger collections of mutually orthogonal luciferase?luciferin sets. Perfect selectivity is also not required for using the probes in biological environments. Rather, patterns in substrate use are most

important and can be discerned using standard imaging equipment and rates of change in photon output.

1.3 Conclusions

We developed a general and rapid strategy to engineer orthogonal luciferase?luciferin pairs. The method relies on developing an initial pool of functional enzymes and screening the collection with chemically diverse luciferins. Using this approach, we generated $\sim 800,000$ possible pairings and mined the data for orthogonal pairs with a custom computer algorithm. Dozens of candidates were identified and validated in vitro. A handful of hits were further translated into cultured cell and animal models, greatly expanding the number of bioluminescent probes for multicomponent imaging.

We further analyzed the principles governing orthogonal substrate use. Chemical and sequence diversity was key to eliciting high levels of selectivity. Thus, the addition of more luciferins and libraries to our ?living? data set should improve orthogonality and lower the barrier to identifying higher-order sets. A fleet of sensitive, selective pairs will bolster imaging capabilities and push the boundaries of what we can ?see? and learn about biological systems. The methods reported here are also applicable beyond the field of bioluminescence. Parallel screens and in silico analyses can expedite the search for other orthogonal enzyme?substrate or protein?ligand pairs relevant to optogenetics, cell signaling, and other disciplines.