Title
Statistical Coupling Analysis-Guided Library Design for the Discovery of Mutant Luciferases

Journal
Biochemistry, 57(5)

ISSN
0006-2960

Authors
Liu, Mira D
Warner, Elliot A
Morrissey, Charlotte E
et al.

Publication Date
2018-02-06

DOI
10.1021/acs.biochem.7b01014

Peer reviewed

HHS Public Access

Author manuscript

Biochemistry. Author manuscript; available in PMC 2019 February 06.

Published in final edited form as:

Biochemistry. 2018 February 06; 57(5): 663–671. doi:10.1021/acs.biochem.7b01014.

Statistical Coupling Analysis-guided library design for discovery of mutant luciferases

Mira D. Liu[#1], **Elliot A. Warner**[#1], **Charlotte E. Morrissey**[1], **Caitlyn W. Fick**[1], **Taia S. Wu**[1], **Marya Y. Ornelas**[1], **Gabriela V. Ochoa**[1], **Brendan Zhang**[2], **Colin M. Rathbun**[2], **William B. Porterfield**[2], **Jennifer A. Prescher**[2,3,4], and **Aaron M. Leconte**[1,*]

[1]W.M. Keck Science Department of Claremont McKenna, Pitzer, and Scripps Colleges, Claremont, California, 91711, United States of America

[2]Department of Chemistry, University of California – Irvine, Irvine, California, 92697, United States of America

[3]Department of Molecular Biology and Biochemistry, University of California – Irvine, Irvine, California, 92697, United States of America

[4]Department of Pharmaceutical Sciences, University of California – Irvine, Irvine, California, 92697, United States of America

[#] These authors contributed equally to this work.

Abstract

Directed evolution has proven to be an invaluable tool for protein engineering; however, there is still a need for developing new approaches to continue to improve the efficiency and efficacy of these methods. Here, we demonstrate a new method for library design that applies a previously developed bioinformatic method, Statistical Coupling Analysis (SCA). SCA uses homologous enzymes to identify amino acid positions that are mutable, functionally important, and engage in synergistic interactions between amino acids. We use SCA to guide a library of the protein luciferase and demonstrate that, in a single round of selection, we can identify luciferase mutants with several valuable properties. Specifically, we identify luciferase mutants that possess both red-shifted emission spectra and improved stability relative to the wild type enzyme. We also identify luciferase mutants that possess a >50-fold change in specificity for modified luciferins. To understand the mutational origin of these improved mutants, we demonstrate the role of mutations at N229, S239, and G246 in altered function. These studies show that SCA can be used to guide library design and rapidly identify synergistic amino acid mutations from a small library.

For Table of Contents Use Only

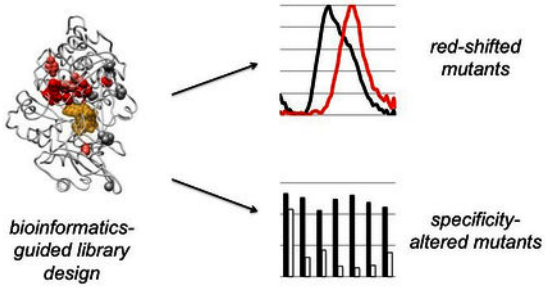
[*]**Corresponding Author**: Telephone: 909-607-9353. aleconte@kecksci.claremont.edu.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Supporting Information. Supporting Methods, Results, Discussion, Figures S1-S10, and Tables S1-S7. This material is available free of charge via the Internet at http://pubs.acs.org.

The authors declare no competing financial interests.



INTRODUCTION

Directed evolution has proven to be a powerful approach to engineering novel protein function.[1–3] Inspired by Nature, proteins can be engineered using iterative cycles of diversification and selection to identify mutant proteins with novel or improved function. While directed evolution is a highly effective approach, it is labor and resource intensive.

The encoding of genetic diversity (*i.e.* library design) is defined by both the amino acid positions and the identity of the amino acid changes. Library design is typically limited by the fact that i.) most mutations have neutral or negative impact on the property of interest, ii.) even mutations that do improve the property of interest often diminish one or more secondary enzyme parameters such as stability, dynamics, substrate recognition, or catalysis, and iii.) many mutations have a larger, nonadditive impact in the presence of a second mutation (i.e. the mutations are synergistic), which can be statistically unlikely to find via random mutagenesis. For enzymes that are not amenable to higher throughput selection based approaches, library sizes are often limited to 10^3-10^4 mutants, which is only a small fraction of the total sequence space.

The design of highly functional libraries can help maximize the probability of identifying a valuable mutant in a small library. DNA shuffling amongst homologous enzymes was the first attempt to gain a high quality library by only using natural diversity.[4] In the natural context, these mutations do not disrupt secondary properties, ensuring a larger proportion of functional mutants within the library. Many other methods, including, but not limited to, the use of neutral mutation,[5, 6] SCHEMA,[7, 8] and evolutionary consensus approaches,9 have similarly used purifying selection (either by Nature or by the experimenter) to remove mutations that disrupt secondary properties. While these approaches are effective at creating functional libraries by using minimally disruptive mutations, they do not, at least by design, focus on identifying positions that are most likely to be functionally important and/or behave synergistically. Currently, most researchers rely on empirical structural or biochemical data to identify mutable and functionally important amino acids; however, new methods that do not rely on prior knowledge could substantially improve library design.

The growing quantity of sequence data and the concomitant development of new and exciting bioinformatic methods to analyze that data increases opportunities to use computational bioinformatic approaches to identify key amino acids for library design. Here, we use a previously described bioinformatic approach, Statistical Coupling Analysis (SCA),

to guide library design. SCA is a computational technique that uses sequencing data from hundreds to thousands of homologs to identify networks (e.g. sectors) of co-evolving, functionally important amino acids.[10] By definition, these identified amino acids are mutable, are functionally important, and likely possess synergistic interactions with other amino acids in the sector; these properties are all extremely valuable for library design. SCA has previously been used to study protein folding dynamics, structure, and allostery.[10–12] It has further been used to identify minimal amino acid motifs for changing biochemical functions found in Nature.[13, 14] However, to date, there has been little effort to develop SCA as a tool to be used for engineering non-natural function. To date, we know of only a single example where SCA was used to guide directed evolution of a nonnatural activity.[15] In this case, SCA was used to identify co-varying amino acids within 5Å of an amino acid position that had been identified through other methods; these mutations were used to mitigate loss of secondary properties that occurred during selection. However, to our knowledge, SCA has not been used to create libraries to identify proteins with entirely new function.

To evaluate the efficacy of our approach to incorporating SCA into library design, we have used the methodology to identify mutants of Firefly luciferase (Fluc), the luciferase enzyme from the North American firefly (Photinus pyralis). Fluc is the most commonly used enzyme for bioluminescence imaging (BLI), a non-invasive method for monitoring diverse processes in living systems, which relies on coupling the presence and/or function of Fluc and/or the presence of its substrate luciferin (D-luc) to *in vitro* and *in vivo* biological events.[16] The potential detection limit, versatility, and temporal and spatial resolution of BLI assays often exceeds fluorescence-based methods for *in vivo* work;[17] however, much like green fluorescent protein required significant optimization through directed evolution to reach full utility,[18] Fluc will require additional engineering to reach its full potential.

Two of the most commonly pursued altered Fluc properties are red-shifted emission spectrum[19, 20] and altered recognition of modified luciferins[21, 22]. A more red-shifted emission spectrum would result in a lower signal-to-noise, as less of the light would be absorbed by endogenous biomolecules,[23] and enable more effective multi-color experiments.[20] Altered substrate recognition enables the use of modified luciferins and has the potential to improve BLI by enabling use of luciferins with improved pharmacokinetic and pharmacodynamic properties[24, 25] as well as to enable orthogonal luciferases,[21, 22] which could be used in multi-component imaging applications.

To date, engineering these properties has largely relied on mutations generated by either random mutagenesis or mutations at positions determined by past structural studies. To obtain altered properties, these studies have often required multiple rounds of diversification and screening and have resulted in diminished secondary properties, such as decreased stability, decreased substrate affinity, and decreased brightness.19 While some researchers have been able to later engineer improved secondary properties through separate experiments,9 this type of approach is labor and resource intensive with no guarantee of success.

Here, we test the hypothesis that using SCA as a guide for library design will lead to the rapid identification of high-value enzymes in a single round of screening. We performed

multiple screens for different properties and show, for each, a single round of selection can identify useful luciferase mutants, which can expand the capabilities of bioluminescent imaging.

MATERIALS AND METHODS

SCA.

The amino acid sequences of 38 luciferase and 245 acetyl CoA synthetase (Acs) homologs were collected from the NCBI non-redundant database using NCBI BLASTp (Table S1). Luciferase and Acs sequences were combined (283 sequences total) and aligned using CLC Sequence Viewer 6 (Table S2). The SCA version 5.0 MATLAB toolbox was downloaded (http://systems.swmed.edu/rr_lab/sca.html) and applied in MATLAB 2015b to the alignment to define four independent components with a cut-off stringency of 90%. Four final sectors were defined and evaluated individually by visual inspection of physical continuity and proximity to the enzyme active site when mapped onto the structure of *Luciola cruciata* luciferase (PDB 2D1S) as well as sector size and inclusion of literature-identified residues; sector 1 was selected for library generation. For full SCA5 parameters see Table S3.

Library design.

The 14 top-weighted amino acids of sector 1 were considered as potential positions to encode diversity in the library. From the 38 known luciferases in our MSA (Figure S1), we identified natural genetic diversity at each of these positions. Four amino acids (53, 243, 253, 435) showed no natural diversity amongst the panel of known luciferases and were removed. Three amino acids (120, 191, 477) were removed to aid in assembly of the library; because they are distant in primary sequence space, they would have considerably increased the cost and experimental difficulty of generating the library. Amino acid substituions found in the panel of luciferases occurring at the remaining nine amino acids (Figure S1) were encoded into primers bearing degenerate codons (Table S4).

Red-shift and Substrate Selectivity assays.

Cultures were resuspended in 150 μL lysis buffer. Each lysate was assayed with three substrates. For each, 15 μL of lysate was added to 15 μL of 3 mM ATP in lysis buffer and 15 μL of either 450 μM D-luciferin in lysis buffer, 300 μM 4’-bromo luciferin, or 900 μM 7’-pyridone luciferin, and 15 μL of cell lysate. Cells were shaken at room temperature for 30 seconds, followed by a 2 minute wait period. Luminescence emissions were recorded with a 528 nM (± 10 nM) filter, a 620 nM (± 20 nM) filter, and no filter. Reported color scores were the ratio of the luciferase’s emission using D-luciferin in the 620 nM filter divided by the emission in the 528 nM filter. Reported selectivity scores were the ratio of the luciferase’s emission without a filter with 4’-bromo luciferin divided by the emission without a filter with 7’-pyridone luciferin.

Stability assay.

Luciferases were expressed as described above. Colonies were lysed in 250 μL lysis buffer. Cell lysates were divided into 9 microcentrifuge tubes, each containing 20 μL of lysate. Lysate divisions were differentially temperature treated for 2 minutes at the follow

temperatures, 25 °C, 30 °C, 32 °C, 35.2 °C, 39.3 °C, 44.9 °C, 51.9 °C, and 54 °C. 15 μL of each heat-treated lysate was added to 15 μL of 3 mM ATP, 15 μL of 450 μM D-luciferin. Lysate divisions were shaken at room temperature for 30 seconds, followed by a 3 minute wait period. Luminescence emissions were recorded without a filter. Thermostability curves were non-linearly fitted according to the equation $y = a - \frac{(a-10) * x^b}{c^b + x^b}$, where x is temperature, y is luciferase emission at temperature x, and a, b, and c are constants. The constants were then used to calculate a T_{50} by calculating the temperature at which luminescence is 50% of the maximum luminescence.

RESULTS

SCAI library design and construction.

We performed SCA version 5.0 using a multiple sequence alignment consisting of 38 luciferase homologs and 245 homologs from the CoA synthetase family (for multiple sequence alignment, see Figure S1), a highly homologous family that catalyzes a similar reaction, possesses a similar fold, and has been shown to be possess latent luciferase activity.[26] From this SCA, we identified several potential sectors. Each protein sector, generated by SCA, is a group of residues that tend to mutate in a coordinated fashion among homologs, and, based on previous demonstrations of SCA, are highly likely to be functionally related. We chose one sector of 14 amino acids, primarily focused in and around the active site of the enzyme, to guide our library (Figure 1).

To incorporate diversity at these positions in a manner that is not disruptive to function, we only encoded amino acid residues found at the SCA-identified amino acid positions in 38 known luciferases. Of the 14 amino acids identified through SCA, 4 showed no variance among the native luciferases (amino acids 53, 243, 253, 435, *P. pyralis* numbering) and, thus, were removed from the library. Further, three amino acids (120, 191, 477) were removed due to their distance from the remainder of the mutations in primary sequence space and constraints of library size; due to this distance in primary sequence space, encoding the amino acid changes is difficult. To simplify library construction, they were removed.

The observed natural variation at the remaining 9 SCA-identified amino acids was synthetically shuffled using degenerate oligonucleotides to create a library of genes using previously described techniques.[27] We used synthetic shuffling to enable recombination independent of sequence space distance, which was necessary since the library is distributed across less than 200 bp of primary sequence. The library included 15 amino acid changes found in Nature, seven amino acid changes included to ease library synthesis, and nine wildtype amino acids (Figure S2). The library was cloned into F-luc using Gibson Assembly, and transformed into E. coli. Based on the number of transformants, we estimate the library contains 1×10^4 unique members. To validate that the encoded diversity was successfully shuffled, we sequenced nine randomly chosen mutants from the library and observed 28 of the 31 possible mutations in the library (Figure S3). Further, we observed that each mutant enzyme has 5.8± 1.5 mutations on average.

Screen for red-shifted mutants without loss of stability.

Previous efforts to engineer red-shifted luciferases have employed single-property screens, resulting in novel, red-shifted enzymes that, unfortunately, typically diminished one or more secondary biochemical properties, such as stability, luciferin affinity, or brightness. We sought to design a screen that could probe multiple parameters concurrently; specifically, we sought to develop a screen that could, with minimal labor, assess both emission color and stability. While selecting for multiple properties in a single screen step, especially from a small library, is challenging, we felt that our approach is well-suited for this challenge since interacting amino acids often are needed for improved stability. Thus, we developed a plate-based screen that could simultaneously measure emission color and stability (for a description of the development of the screen, see Supporting Results).

Prior to our plate-based screen, we pre-screened our library for functional mutants on agar plates containing D-luc (Figure S4C). Of 2,666 mutants assayed, 330 (12%) emitted light in the presence of D-luc; considering that our average mutant possesses 5.8 mutations, this is a high proportion of functional mutants. Each of the 330 active colonies was used to create a new culture that was screened for emission color and stability in a single assay using our plate-based screen.

We isolated the five mutants predicted to have the largest red-shifted emission (Figure 2A). After validating that the observed red-shift score was reproducible, we sequenced the five mutants (Figure 2B). Gratifyingly, by comparing our sequencing before selection (Figures S3) to our sequencing after selection, we observed substantial sequence convergence amongst the mutants, indicating a successful selection. All five mutants possessed N229T, and four of the five mutants also possess S239T. Further, we observed convergence back to the wild type amino acid residue at F250 and T251 in the recovered mutants.

To evaluate the stability of these five mutants, we measured the T_{50} after two-minute incubation at varying temperatures (Figure 2C). Three of the five mutants possessed improved thermal stability relative to Fluc, possessing an increase in the T_{50}. Most notably, COL4 (T_{50} = 47.1° C) possesses a 3.3° C increase in the T_{50} relative to Fluc (T_{50} = 43.8° C). We also observed that COL2 (T_{50} = 45.2° C) and COL3 (T_{50} = 44.5° C) possess increased temperatures while COL1 (T_{50} = 40.4° C) and COL5 (T_{50} = 40.3° C) possess decreased temperatures. Each mutant has four or more mutations; considering that mutations are typically destabilizing, the presence of highly mutated stable mutants with four or greater mutations indicates the efficacy of our library design approach.

To verify that our most stable mutants possess a genuine shift in their emission spectrum, we observed the emission spectrum of COL2 and COL4, the two most stable mutants. Protein was expressed in culture using autoinduction, and the lysate was incubated with 1mM D-luc and 1mM ATP and observed using a FluoroLog3 fluorometer at room temperature. While wildtype luciferase has an emission maximum of 555nm, which is in good agreement with past literature precedent,[28] both COL2 and COL4 possess emission maxima at 605nm, displaying a significant 50nm shift in emission at room temperature (Figure 2D, 2E). These data show that the wavelength emission score did accurately predict that these mutants are red-shifted.

Screen for substrate preference.

Having shown that the SCA1 library could identify interesting red-shifted mutants quickly from a small number of screened mutants, we wondered if the library could be screened for other high value luciferase properties. Another important area for Fluc engineering is the development of orthogonal luciferase/luciferin pairs, which would enable development of multicomponent imaging techniques.[22, 29] Thus, we examined whether our SCA1 library could be used to identify luciferases with altered modified luciferin specificity. 4'Br-Luc (Figure S6) has been previously shown to be uniquely suited for orthogonal probe development likely due to the size and polarizability of the Br atom. Like our identification of red-shifted mutants, we performed a multiple stage screen to first identify active mutants and then identify selective mutants (for a detailed description of the screen, see Supporting Results).

From 3375 transformants, we identified 101 active mutant enzymes that can luminesce using 4'Br-Luc as a substrate. While only approximately 3% of the mutants were functional on 4'Br-Luc, compared to approximately 12% for D-luc, that is to be expected considering that Fluc displays 3% of the luminescence when using 4'Br-Luc than with D-luc.[30]

The 101 mutant enzymes showing measurable luminescence using 4'Br-Luc were then screened to identify mutants with altered selectivity for luciferin analogs; specifically, we examined the relative luminescence using 4'Br-Luc relative to 7'Me-Luc. We observed that 13 of the mutants possessed >10-fold changes in specificity relative to the wild-type enzyme for these two luciferins (Figure S7). These mutants were then screened against a panel of 11 luciferin analogs displaying substitutions of variable steric and electronic properties at either the 4', 5', and 7' positions (Figure S8). Of all of the compounds, the most significant changes in recognition were for 4'Br-Luc relative to 7'Pyr-Luc; thus, we focused on six mutant enzymes and their recognition of these two luciferin analogs. While these compounds are chemically quite different, it is not clear, on first inspection, why 7'Pyr-Luc is discriminated against better than other 7'-substituted analogs, displaying the value of the luciferin screening based approach.

While the wild-type enzyme possesses a slight preference for 4'Br-Luc, generating 3-fold more light using 4'Br-Luc than 7'Pyr-Luc, the six most selective enzymes possess a 19- to 222-fold preference for 4'Br-Luc over 7'Pyr-Luc (Figure 3A). Interestingly, the selectivity gains were largely achieved through decrease in 7'Pyr activity; all six mutants showed greater than 10-fold decreases in 7'Pyr activity, but none of the mutants showed significant increases in luminescence when using 4'Br-Luc. For several mutants (4B3, 4B4, 4B5), the 7'Pyr signal possessed similar luminescence signal to a control in the absence of luciferin analog.

We sequenced the six most active mutants (Figure 3B). We observed significant sequence convergence relative to the pre-selection library; the four most selective mutants all had N229T, a mutation at F250 to either A or C, and mutation at G246A. We also observed convergence to the wildtype at T251 and M249. Further, the two most selective mutants, 4B3 and 4B4, differ only by a single mutation, S239T. The strong sequence convergence indicates a successful enrichment of active mutants.

Mutational origins of altered activity.

Both our red-shift screen and our screen for substrate specificity revealed that several amino acids (F250, N229, G246, S239) were frequently mutated (or not mutated) in the most promising recovered mutants from each screen, strongly suggesting that these positions are important for function. To understand the roles of the specific mutations on color, substrate specificity, and protein stability, we created single point mutants at these positions (Table S5) and further biochemically evaluated these enzymes for each property. To gain further sequence-function information, we also evaluated each mutant recovered in the original screens for emission color, stability, and specificity for 4'Br-Luc over 7'Pyr-Luc.

In our color and stability screen, we saw strong sequence convergence to N229T (5 of 5 mutants observed possessed this mutation) and S239T (4 of 5 mutants observed possessed this mutation). We also did not observe mutations in F250 (0 of 5 mutants observed possessed this mutation). To further understand the determinants of red-shifting, we measured the emission color using the 620:528 ratio for a panel of commonly observed mutants, N229T, S239T, G246A, F250A, F250C, and F250G (Figure 4A). Of these mutations, only N229T significantly red-shifts the emission color. Interestingly, several mutants that are frequently observed in the red-shift mutants, S239T and G246A, appear to slightly blue-shift the light, showing a decrease in the ratio of light at 620:528. Mutations at F250, which were observed in selectivity-altering mutants, but not in red-shifted mutants, do not appear to impact color. From these data, it appears that only N229T results in a significant shift in emission color. The importance of N229T was corroborated by examining the emission color of mutants identified in the selectivity screen (Figure S9); all mutants containing N229T were significantly red-shifted while the mutants that did not have the mutation did not show a red shift.

To further examine whether only N229 results in red-shifting, we also mutated each amino acid position in the library to alanine and evaluated these mutants for emission color (Figure S9). Of these mutants, N229A had the largest change in the emission wavelength; however, the red:green ratio observed (620:528 ratio = 68) was lower than both N229T (620:528 ratio = 140) and the red-shifted mutants (620:528 ratio ranged from 93-141). Other than N229A, only I237A (620:528 ratio = 7) had an appreciable change in the color of light emitted, although the effect was small compared to mutations at N229. Collectively, these data suggest that N229T alone is likely responsible for the red-shift that was observed.

In our color and stability screen, we saw strong convergence to N229T and S239T. We also observed that the most stable mutants recovered from the screen also possessed G246A. Specifically, the three stable mutants (COL2, COL3, COL4) all possess both S239T and G246A while the less stable mutants (COL1, COL5) have either S239T only or neither position mutated, implying that one or both of these mutations might be largely responsible for stability. To understand the determinants of stability, we examined the impact of the commonly observed mutations on enzyme stability (Figure 5). S239T and G246A both show increases of T_{50} relative to Fluc by 0.7 °C and 2.0 °C, respectively. Interestingly, although N229T has the beneficial property of shifting the emission wavelength, on its own, the mutation is fairly destabilizing, decreasing T_{50} by 2.7 °C. Similarly, despite their frequent selection in selectivity-altered mutants, F250 mutations were destabilizing as well,

decreasing T_{50} between 3.8 °C and 7.0 °C. Collectively, these data effectively explain the relative stability of the three most stable color-shifted mutants, which all possess the stabilizing mutations S239T and G246A. These data were corroborated by examining the stability of the mutant enzymes identified for their altered stability (Figure S10); the most stable of these enzymes all possess S239T, indicating the role of this mutation in stabilizing the enzyme.

In our selectivity screen, we observed recurring mutations at N229T (5 of 6 mutants possessed this mutation), G246A (6 of 6 mutants), and F250 (all mutants had a mutation at this position althought the identity of the mutation varied). To understand the determinants of specificity, we measured the luminescence using 4'Br-Luc and 7'Pyr-Luc from the previously described six point mutants (Figure 6). Perhaps unsurprisingly, G246A had the largest effect; this mutation was found in all of the isolated selective mutants and in the three most selective red-shifted mutants. Surprisingly, the selectivity gain of G246A was relatively modest; G246A has only a 6-fold increase in preference for 4'Br-Luc relative to Fluc, which is less selective than nearly all of the selectivity-altered mutants. Although all of the selectivity-altered mutants observed had a mutation at F250, in isolation, the three mutants observed (F250A, F250C, and F250G) showed only modest changes in selectivity. Further, N229T was found in the five most selective hit mutants, but, similarly, this mutation on its own does not show a significant change in selectivity. These data imply that, for these mutations to alter selectivity, they may require secondary synergistic mutations. Collectively, the relatively large differences between the specificity of the most active mutants and the point mutants suggest that synergistic interactions between mutliple mutations are likely necessary to see large shifts in specificity for these luciferin analogs. To further examine these effects, we also looked at the specificity of the red-shifted mutants (Figure S11); generally, the specificity of these mutants is significantly less, which we hypothesize is caused by the lack of mutations at F250.

DISCUSSION

SCA as a guide for library design.

SCA is a particularly attractive starting point for library design since it identifies amino acid positions that are i.) mutable ii.) functional and iii.) likely to engage in synergistic interactions. These are all important characteristics of mutant library design. Considering the results of our selection and biochemical experiments, we see evidence of all three of these features in our mutations, validating our use of SCA and natural diversity in library design.

It is important that mutations in a library are mutable without severely disrupting protein function. Our library possesses 5.8 mutations on average. If a library were created using either random mutagenesis or structure-guided mutations that possessed an average of 5.8 mutations, the overwhelming majority of the library (>99%) would be expected to be nonfunctional. Here, when tested for functional mutants, 12% of the library was found to be active, which is an impressively high amount considering the number of mutations per clone in the library. This strongly suggests that the library design approach identifies positions that are mutable.

Further, amongst our eleven sequenced post-selection mutants, we observe mutations at all positions except for T251, showing that nearly every individual amino acid position is mutable. Notably, T251 mutants have been identified in other selection experiments,[21] suggesting that this position is mutable; however, in our screens, it appears that there was strong selective pressure against mutating at this residue. It is unclear whether T251 is not mutable in general in our screens or whether the specific mutation in our library, T251I, is particularly disruptive. However, overall, it is clear that all nine SCA-identified positions are, indeed, mutable.

While there are many methods to identify mutable positions, typically employing natural diversity, there are relatively few methods that identify amino acids that are both mutable and functionally important. Here, even from a small library and a single round of selection, we have identified mutants with dramatically altered functions, suggesting that this library approach has selected functionally important mutations. Through our biochemical analysis, we were able to show that, for four of the eight mutable positions in our experiments, the mutations are functionally important. While some of these amino acids have been studied previously, in most cases, we would not have been likely to predict their effects based on past studies alone. G246A, which has been previously only observed in blue-shifted mutants,[31] here has been shown to both stabilize luciferase structure and to alter recognition of modified luciferin substrates. N229T has been previously identified to result in red-shifted emissions;[32] however, as we observe here, this mutation also destabilizes the enzyme. Unlike previous work, in our case, its coselection with other mutations helps to mitigate the unfavorable characteristics of this amino acid mutation. S239T has not been identified previously in past studies; here, this mutation was frequently identified in our screens and appears to improve stability. Lastly, F250 also seems to be under strong selective pressure; although prior studies have identified F250 mutants that alter color,[31] we saw strong convergence to the wild-type amino acid in our red-shift screen. This might be due to the observed destabilization that appears to occur with this mutation. F250M was recently observed to improve recognition of a 4’ morpholino substituted luciferin.[22] While we do not observe substantial changes for selectivity of the F250A/C/G in isolation, their recurrent mutation in our most promising mutants, as well as observation in past studies, points to a functionally important role in modified luciferin recognition. Collectively, our sequencing data and biochemical analysis show that mutations at nearly half of the amino acid positions have a key role in imparting altered function in Fluc. The relatively large number of mutations observed to be functionally important, even from a relatively small number of possible mutations (there are only 22 point mutations in the library), points to the efficacy of SCA as a tool in library design.

Finally, identification of interacting amino acids, especially those that are synergistic, can substantially improve the probability of identifying mutant enzymes with improved properties. Since SCA identifies co-evolving amino acid positons, it is possible that this can be used to identify synergistic interactions. We do observe interacting, synergistic amino acid changes here. In the case of the red-shifted mutants, secondary mutations such as S239T and G246A appear to directly offset the loss of stability of N229T in a nonadditive fashion, making the lead mutants useful for future study.

We see stronger synergistic interactions for the selectivity-altered mutants. The change in selectivity observed in the lead mutants (>50-fold change relative to Fluc wt for 4B4) is substantially greater than the increase in selectivity observed in the recurring mutations N229T (<2-fold), F250A/C (<3-fold), or G246A (6-fold). F250A/C/G and N229T are both found in all recovered specificity-altered mutants, but, in isolation, they do not substantially change specificity. Considering the strong convergence on these mutations, it seems likely that these mutations do functionally contribute to altered substrate recognition, but might require secondary mutations to impart function. We created double mutants containing just N229T and F250A/C/G mutations; however, these enzymes are inactive (unpublished results), likely due to the resultant instability of the protein, implying that, perhaps, these mutations require G246A to impart altered function. Although further studies are required, it is likely that synergistic interactions are responsible for the improved function seen here.

To further understand the interactions between these amino acids, we observed these amino acid positions in the context of the known crystal structure of the homologous enzyme, L. cruciata, bound to a high energy intermediate analog of D-luc (Figure 7). Observing the structure highlights both the challenges of protein design and the utility of this SCA-guided approach. While T251 is 3Å from the 4’ position of the native luciferin, and thus might be reasonably selected for mutagenesis for 4’Br-Luc recognition using traditional library design methods, we do not observe this amino acid mutated in our studies. However, the frequently mutated F250 and G246 pack on each other in the active site and, considering their proximity in primary sequence, they likely impact positioning of T251. Likewise, we observe N229 mutations in all selectivity-altered mutants; while this amino acid does not directly contact either the modified luciferin or T251, it does appear to contact another SCA-identified residue, I237, which makes contacts with T251. These interactions highlight the complexity of interactions in the active site of proteins and the inherent challenges of structure-based design. If choosing only based on structural knowledge, G246, F250, and N229T would be unlikely to have been chosen; however, they are identified in all of our most selective mutants.

Implications for library design.

Compared to traditional diversification techniques, such as error-prone PCR or site-directed mutagenesis, SCA-guided library design appears to be capable of identifying high-value mutants in fewer rounds of evolution and from smaller libraries. While prior studies have identified mutants with similar characteristics (see Supplementary Discussion), these prior studies have required multiple rounds of selection and even multiple studies to identify mutants with characteristics similar to those we describe here. Thus, considering our relatively rapid identification of these mutants, we can conclude that SCA-guided library design can improve the efficiency of directed evolution compared to more traditional approaches.

While we think SCA is a particularly useful bioinformatic method to guide library design, there are many bioinformatics approaches that also identify co-varying amino acids that could similarly be used.[33, 34] We hope our work here will inspire further efforts to apply bioinformatics, especially methods identifying co-varying amino acids, in library design

and, in time, comparative studies shou be performed to determine the most effective approach.

CONCLUSIONS.

Here, we have described a method that uses the bioinformatic method Statistical Coupling Analysis to identify mutant luciferase enzymes. In a single library, using a single round of screening, we have identified valuable lead mutants that possess characteristics similar to the most exciting mutants in the literature identified to date. Further, we have examined the mutational origins of key amino acid changes, both contributing to the known function of luciferase mutations and validating that our SCA-guided library contains mutable, functional, synergistically interacting amino acid changes. The methods here should be broadly applicable to other enzymes and other luciferase properties, and should find broad use in the protein engineering community.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements.

We thank the Prescher group for materials as well as Dr. Krysten Jones for helpful discussions.

Funding

This work was supported by the Research Corporation for Science Advancement (Cottrell Scholar Award #23717 to A.M.L.) and U.S. National Institutes of Health (R01 GM107630 to J.A.P.). C.M.R. was supported by the National Science Foundation Graduate Research Fellowship under Grant No. DGE-1321846. J.A.P. is also a Cottrell Scholar, Alfred P. Sloan Fellow, and Dreyfus Scholar.

ABBREVIATIONS

Fluc	Firefly Luciferase
D-luc	D-luciferin
wt	wild-type
SCA	Statistical Coupling Analysis

References

[1]. Wang L, Brock A, Herberich B, and Schultz PG (2001) Expanding the genetic code of Escherichia coli, Science 292, 498–500. [PubMed: 11313494]

[2]. Castle LA, Siehl DL, Gorton R, Patten PA, Chen YH, Bertain S, Cho HJ, Duck N, Wong J, Liu D, and Lassner MW (2004) Discovery and directed evolution of a glyphosate tolerance gene, Science 304, 1151–1154. [PubMed: 15155947]

[3]. Brustad EM, and Arnold FH (2011) Optimizing non-natural protein function with directed evolution, Curr Opin Chem Biol 15, 201–210. [PubMed: 21185770]

[4]. Ness JE, Welch M, Giver L, Bueno M, Cherry JR, Borchert TV, Stemmer WP, and Minshull J (1999) DNA shuffling of subgenomic sequences of subtilisin, Nat Biotechnol 17, 893–896. [PubMed: 10471932]

[5]. Bershtein S, Goldin K, and Tawfik DS (2008) Intense neutral drifts yield robust and evolvable consensus proteins, J Mol Biol 379, 1029–1044. [PubMed: 18495157]

[6]. Gupta RD, and Tawfik DS (2008) Directed enzyme evolution via small and effective neutral drift libraries, Nat Methods 5, 939–942. [PubMed: 18931667]

[7]. Heinzelman P, Snow CD, Smith MA, Yu XL, Kannan A, Boulware K, Villalobos A, Govindarajan S, Minshull J, and Arnold FH (2009) SCHEMA Recombination of a Fungal Cellulase Uncovers a Single Mutation That Contributes Markedly to Stability, J Biol Chem 284, 26229–26233. [PubMed: 19625252]

[8]. Bedbrook CN, Rice AJ, Yang KK, Ding XZ, Chen SY, LeProust EM, Gradinaru V, and Arnold FH (2017) Structure-guided SCHEMA recombination generates diverse chimeric channelrhodopsins, Proc Natl Acad Sci U. S. A. 114, E2624–E2633. [PubMed: 28283661]

[9]. Kim SB, and Izumi H (2014) Functional artificial luciferases as an optical readout for bioassays, Biochem Biophys Res Commun 448, 418–423. [PubMed: 24802399]

[10]. Socolich M, Lockless SW, Russ WP, Lee H, Gardner KH, and Ranganathan R (2005) Evolutionary information for specifying a protein fold, Nature 437, 512–518. [PubMed: 16177782]

[11]. Smock RG, Rivoire O, Russ WP, Swain JF, Leibler S, Ranganathan R, and Gierasch LM (2010) An interdomain sector mediating allostery in Hsp70 molecular chaperones, Mol Syst Biol 6, 414. [PubMed: 20865007]

[12]. Shulman AI, Larson C, Mangelsdorf DJ, and Ranganathan R (2004) Structural determinants of allosteric ligand activation in RXR heterodimers, Cell 116, 417–429. [PubMed: 15016376]

[13]. Russ WP, Lowery DM, Mishra P, Yaffe MB, and Ranganathan R (2005) Natural-like function in artificial WW domains, Nature 437, 579–583. [PubMed: 16177795]

[14]. McLaughlin RN, Jr., Poelwijk FJ, Raman A, Gosal WS, and Ranganathan R (2012) The spatial architecture of protein function and adaptation, Nature 491, 138–142. [PubMed: 23041932]

[15]. Pinheiro VB, Taylor AI, Cozens C, Abramov M, Renders M, Zhang S, Chaput JC, Wengel J, Peak-Chew SY, McLaughlin SH, Herdewijn P, and Holliger P (2012) Synthetic genetic polymers capable of heredity and evolution, Science 336, 341–344. [PubMed: 22517858]

[16]. Prescher JA, and Contag CH (2010) Guided by the light: visualizing biomolecular processes in living animals with bioluminescence, Curr Opin Chem Biol 14, 80–89. [PubMed: 19962933]

[17]. Paley MA, and Prescher JA (2014) Bioluminescence: a versatile technique for imaging cellular and molecular features, Medchemcomm 5, 255–267. [PubMed: 27594981]

[18]. Tsien RY (1998) The green fluorescent protein, Annu Rev Biochem 67, 509–544. [PubMed: 9759496]

[19]. Branchini BR, Southworth TL, Khattak NF, Michelini E, and Roda A (2005) Red- and green-emitting firefly luciferase mutants for bioluminescent reporter applications, Anal Biochem 345, 140–148. [PubMed: 16125663]

[20]. Branchini BR, Ablamsky DM, Murtiashaw MH, Uzasci L, Fraga H, and Southworth TL (2007) Thermostable red and green light-producing firefly luciferase mutants for bioluminescent reporter applications, Anal Biochem 361, 253–262. [PubMed: 17181991]

[21]. Harwood Katryn R., Mofford David M., Reddy Gadarla R., and Miller Stephen C. (2011) Identification of Mutant Firefly Luciferases that Efficiently Utilize Aminoluciferins, Chem Biol 18, 1649–1657. [PubMed: 22195567]

[22]. Jones KA, Porterfield WB, Rathbun CM, McCutcheon DC, Paley MA, and Prescher JA (2017) Orthogonal Luciferase-Luciferin Pairs for Bioluminescence Imaging, J Am Chem Soc 139, 2351–2358. [PubMed: 28106389]

[23]. Zhao H, Doyle TC, Coquoz O, Kalish F, Rice BW, and Contag CH (2005) Emission spectra of bioluminescent reporters and interaction with mammalian tissue determine the sensitivity of detection in vivo, J Biomed Opt 10.

[24]. Mofford DM, Adams ST, Reddy G, Reddy GR, and Miller SC (2015) Luciferin Amides Enable in Vivo Bioluminescence Detection of Endogenous Fatty Acid Amide Hydrolase Activity, J Am Chem Soc 137, 8684–8687. [PubMed: 26120870]

[25]. Adams ST, Mofford DM, Reddy G, and Miller SC (2016) Firefly Luciferase Mutants Allow Substrate-Selective Bioluminescence Imaging in the Mouse Brain, Angew. Chem.-Int. Edit 55, 4943–4946.

[26]. Mofford DM, Reddy GR, and Miller SC (2014) Latent luciferase activity in the fruit fly revealed by a synthetic luciferin, Proc Natl Acad Sci U S A 111, 4443–4448. [PubMed: 24616520]

[27]. Ness JE, Kim S, Gottman A, Pak R, Krebber A, Borchert TV, Govindarajan S, Mundorff EC, and Minshull J (2002) Synthetic shuffling expands functional protein diversity by allowing amino acids to recombine independently, Nat Biotechnol 20, 1251–1255. [PubMed: 12426575]

[28]. Branchini BR, Southworth TL, Murtiashaw MH, Boije H, and Fleet SE (2003) A mutagenesis study of the putative luciferin binding site residues of firefly luciferase, Biochemistry 42, 10429–10436. [PubMed: 12950169]

[29]. Rathbun CM, and Prescher JA (2017) Bioluminescent Probes for Imaging Biology beyond the Culture Dish, Biochemistry 56, 5178–5184. [PubMed: 28745860]

[30]. Steinhardt RC, Rathbun CM, Krull BT, Yu JM, Yang Y, Nguyen BD, Kwon J, McCutcheon DC, Jones KA, Furche F, and Prescher JA (2017) Brominated Luciferins Are Versatile Bioluminescent Probes, Chembiochem 18, 96–100. [PubMed: 27930848]

[31]. Branchini BR, Ablamsky DM, Rosenman JM, Uzasci L, Southworth TL, and Zimmer M (2007) Synergistic mutations produce blue-shifted bioluminescence in firefly luciferase, Biochemistry 46, 13847–13855. [PubMed: 17994766]

[32]. Viviani VR, Arnoldi FG, Neto AJ, Oehlmeyer TL, Bechara EJ, and Ohmiya Y (2008) The structural origin and biological function of pH-sensitivity in firefly luciferases, Photochem Photobiol Sci 7, 159–169. [PubMed: 18264583]

[33]. McMurrough TA, Dickson RJ, Thibert SM, Gloor GB, and Edgell DR (2014) Control of catalytic efficiency by a coevolving network of catalytic and noncatalytic residues, Proc Natl Acad Sci U S A 111, E2376–2383. [PubMed: 24912189]

[34]. Yip KY, Patel P, Kim PM, Engelman DM, McDermott D, and Gerstein M (2008) An integrated system for studying residue coevolution in proteins, Bioinformatics 24, 290–292. [PubMed: 18056067]

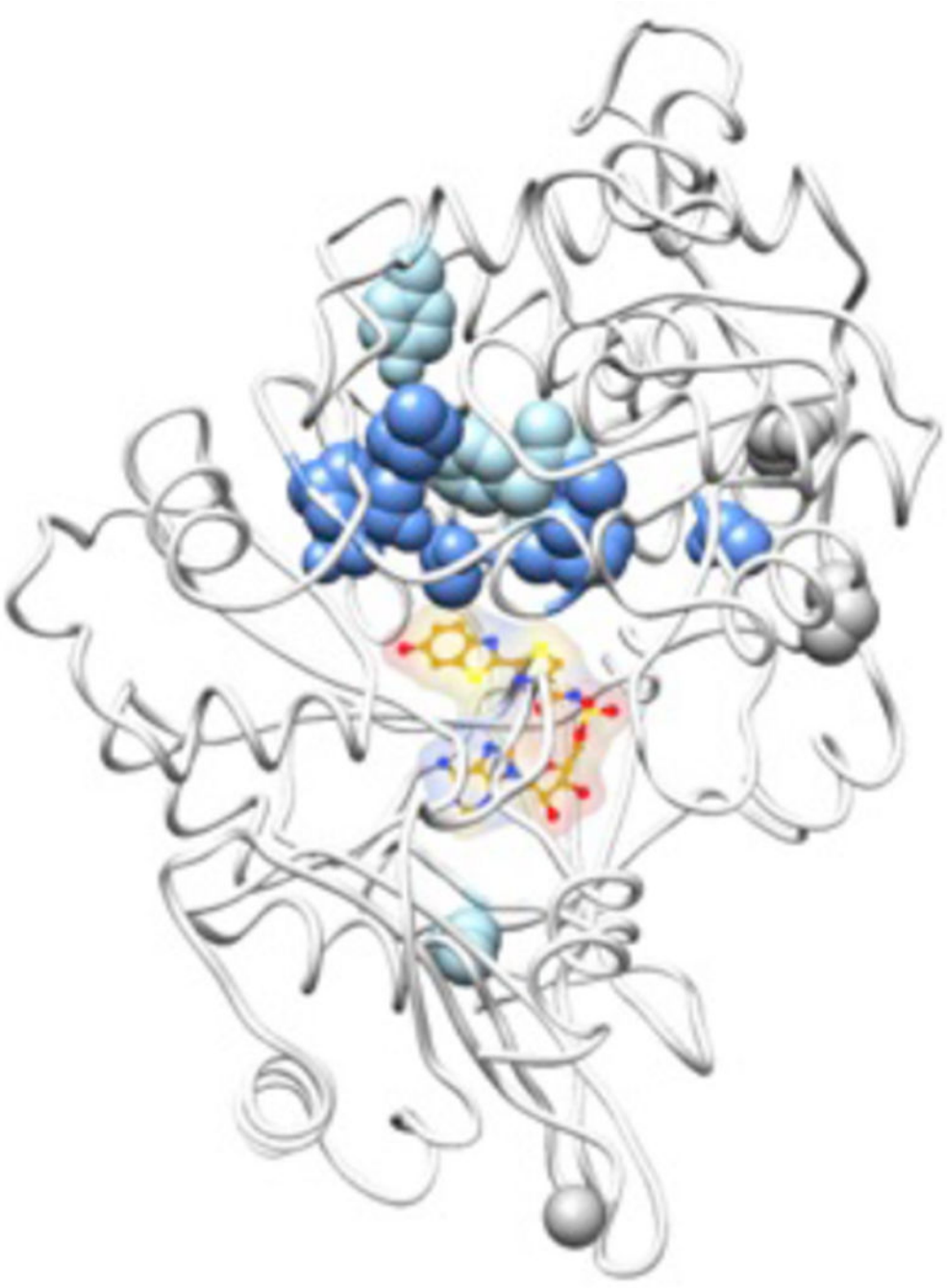


Figure 1.

SCA-identified amino acids used to encode library. The crystal structure of the homologous *Luciola cruciata* luciferase (PDB 2D1S) is shown with D-luciferin shown in goldenrod with surface. Amino acid positions included in the library (dark blue), amino acid positions removed due to lack of natural diversity in luciferases (light blue) or removed due to difficulty in coding amino acids (gray) are highlighted in the structure.

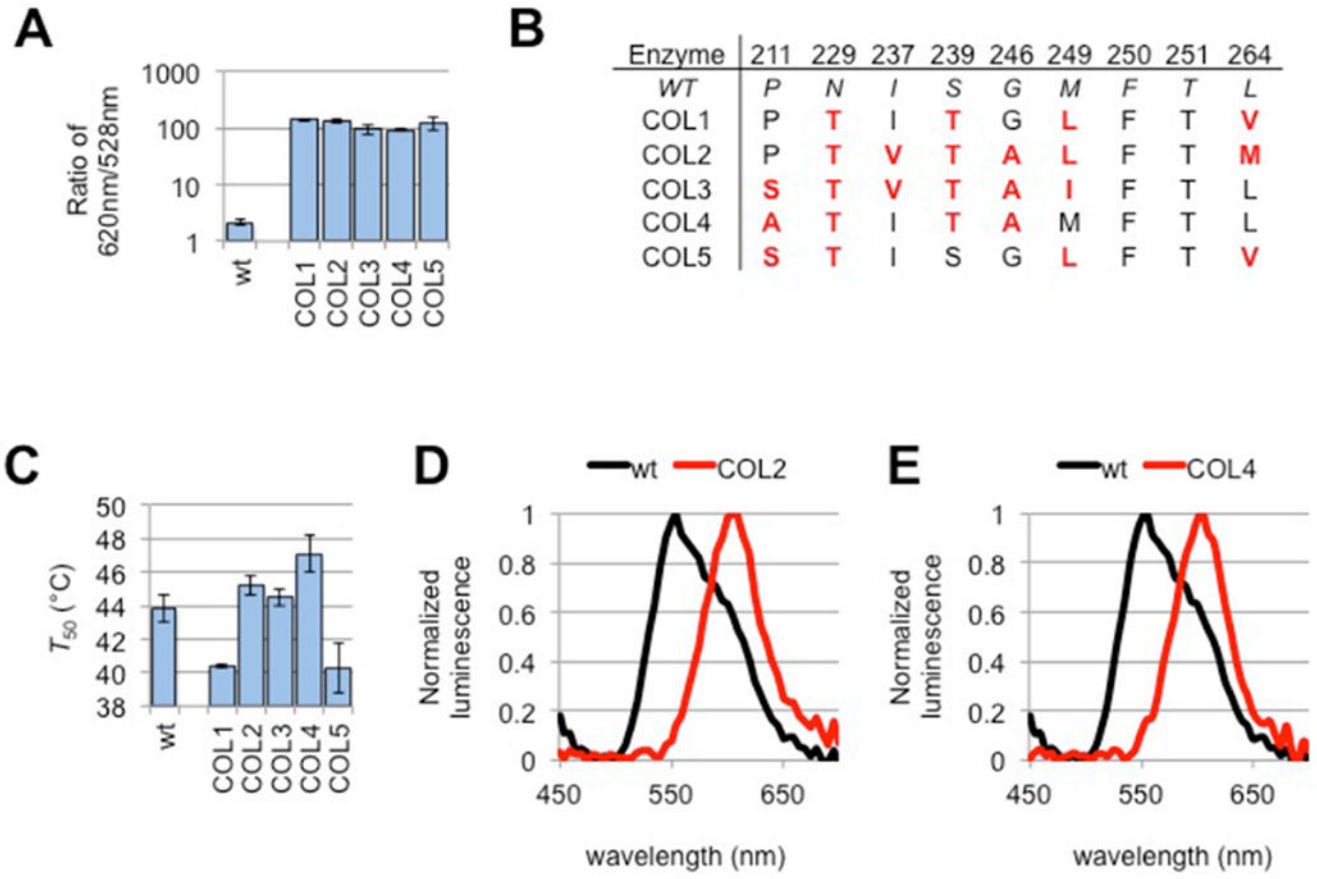
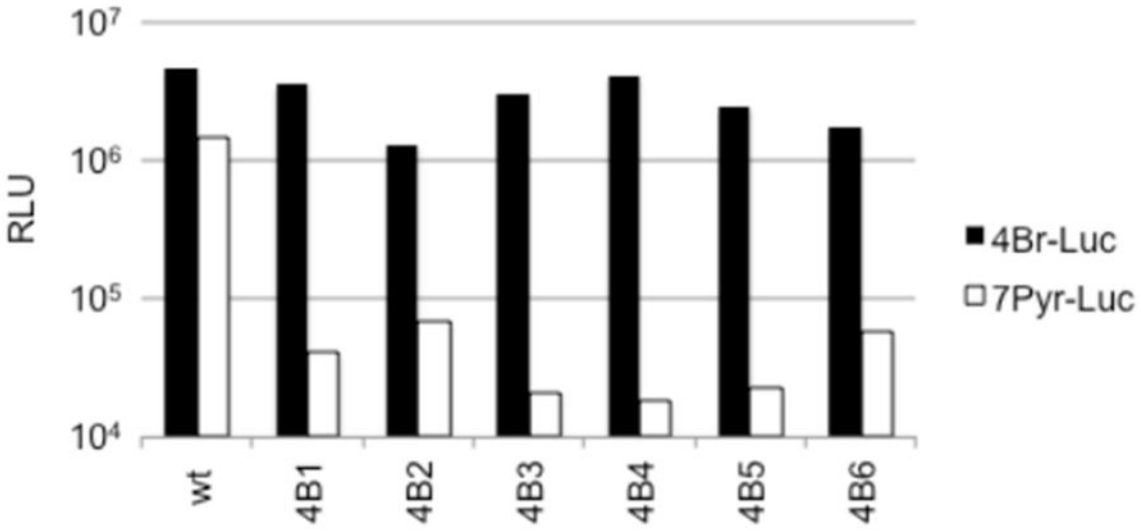


Figure 2.

Identified mutant luciferases with red-shifted emission spectra. (A) 620:528 emission ratio of wildtype and isolated mutants (*n*=3). For each mutant, individual lysates were incubated with D-luc for two minutes and luminescence recorded using either a 528nm filter or a 620nm filter and 1s integrations (B) Genotype of identified active mutants. Only amino acid positions mutated in the library are shown (mutations highlighted in red); no mutations outside of those encoded by the library were observed. (C) Thermal stability of Fluc wildtype and isolated mutants (*n*=3). For each replicate, a single lysate was split and portions were incubated at varying temperatures for 2 minutes and immediately assayed for luminescence using D-luc. (D) Emission spectrum of COL2. (E) Emission spectrum of COL4.

A



B

Sequence	RLU(4Br)/ RLU(7Pyr)	211	229	237	239	246	249	250	251	264
WT	*3*	*P*	*N*	*I*	*S*	*G*	*M*	*F*	*T*	*L*
4B1	87	S	T	I	S	A	M	C	T	V
4B2	19	A	T	I	S	A	M	G	T	L
4B3	143	A	T	V	S	A	M	A	T	L
4B4	222	A	T	V	T	A	M	A	T	L
4B5	109	T	T	I	T	A	M	C	T	V
4B6	30	T	N	V	T	A	M	C	T	L

Figure 3.

Selective luciferases identified in this study. (A) Relative luminescence units for emission in the presence of 4'Br-Luc and 7'Pyr-Luc. A single lysate was split and incubated with either modified luciferin and unfiltered emission was recorded for 10s. (B) Genotypes of isolated mutants at library-encoded positions. Only amino acid positions mutated in the library are shown (mutations highlighted in red); no mutations outside of those encoded by the library were observed.

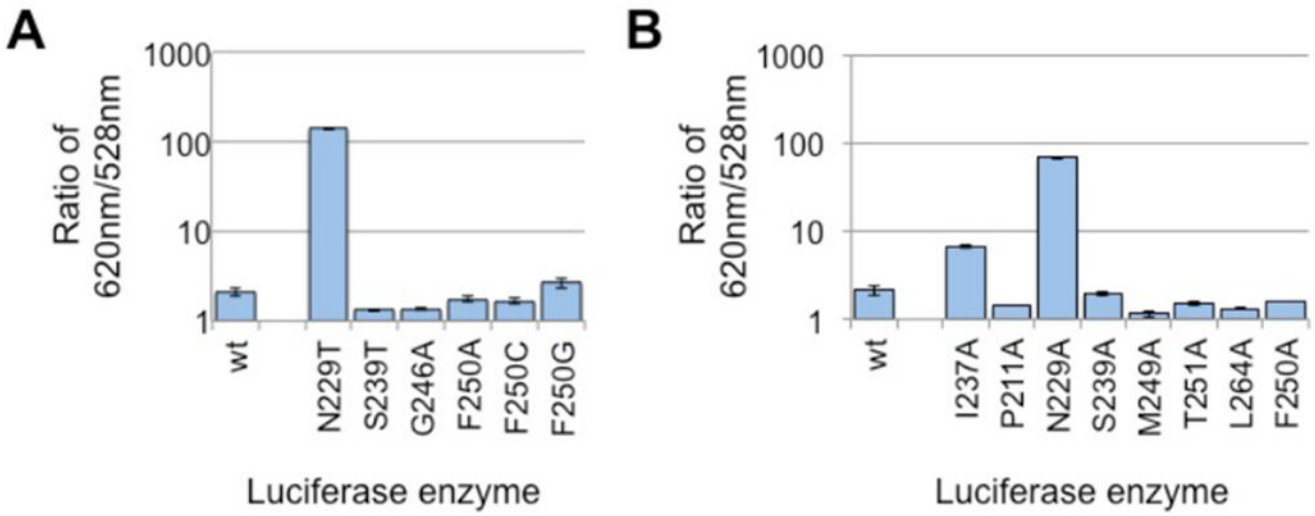


Figure 4.

620:528 ratio for mutants isolated in (A) frequently occurring mutants and (B) alanine mutants corresponding to SCA library positions. For each mutant, individual lysates were incubated with D-luc for two minutes and luminescence recorded using either a 528nm filter or a 620nm filter and 1s integrations. Each mutant was assayed in triplicate.

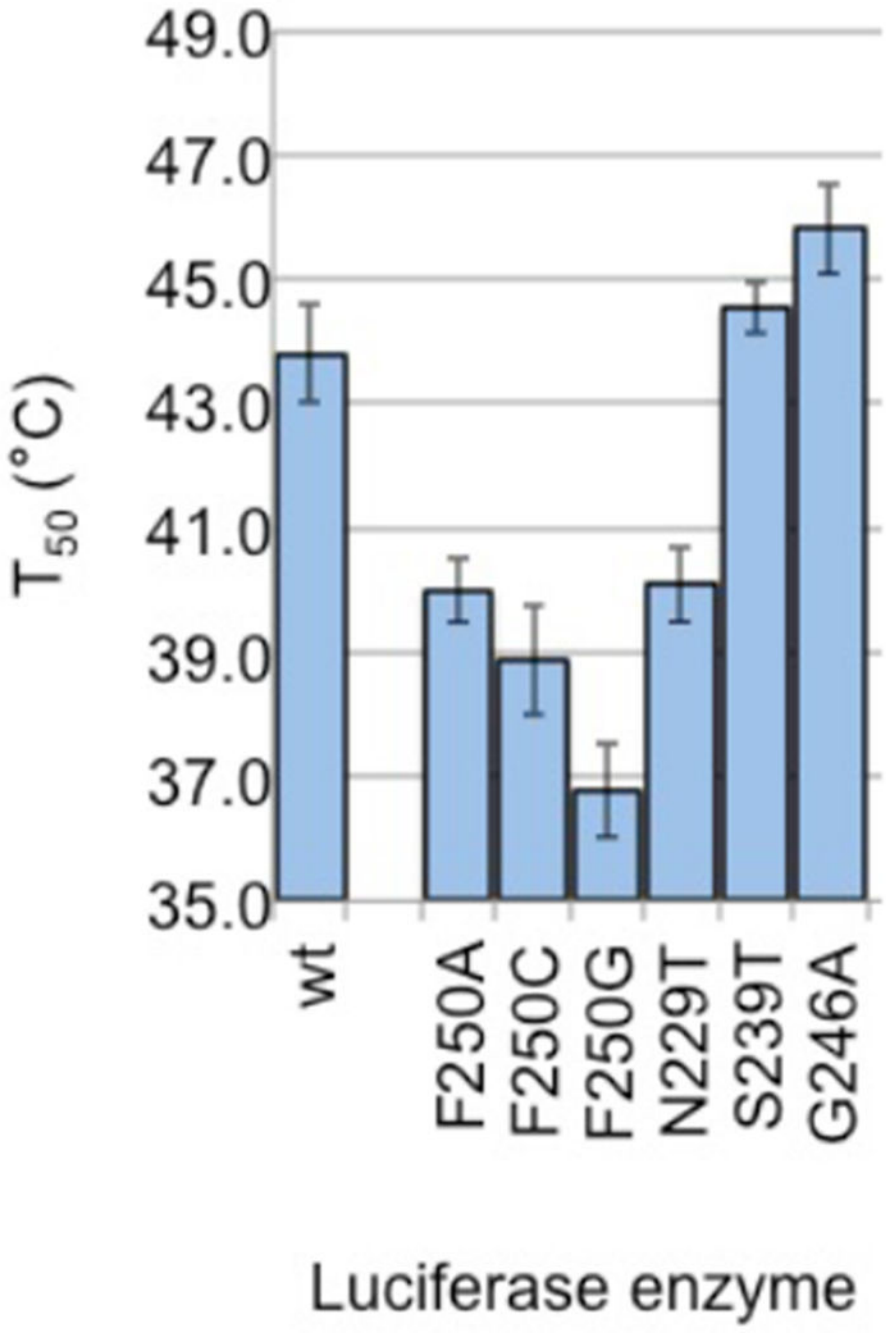


Figure 5.

T_{50} values for frequently observed mutations. For each replicate, a single lysate was split and portions were incubated at varying temperatures for 2 minutes and immediately assayed for luminescence using D-luc. Each luciferase was assayed in triplicate.

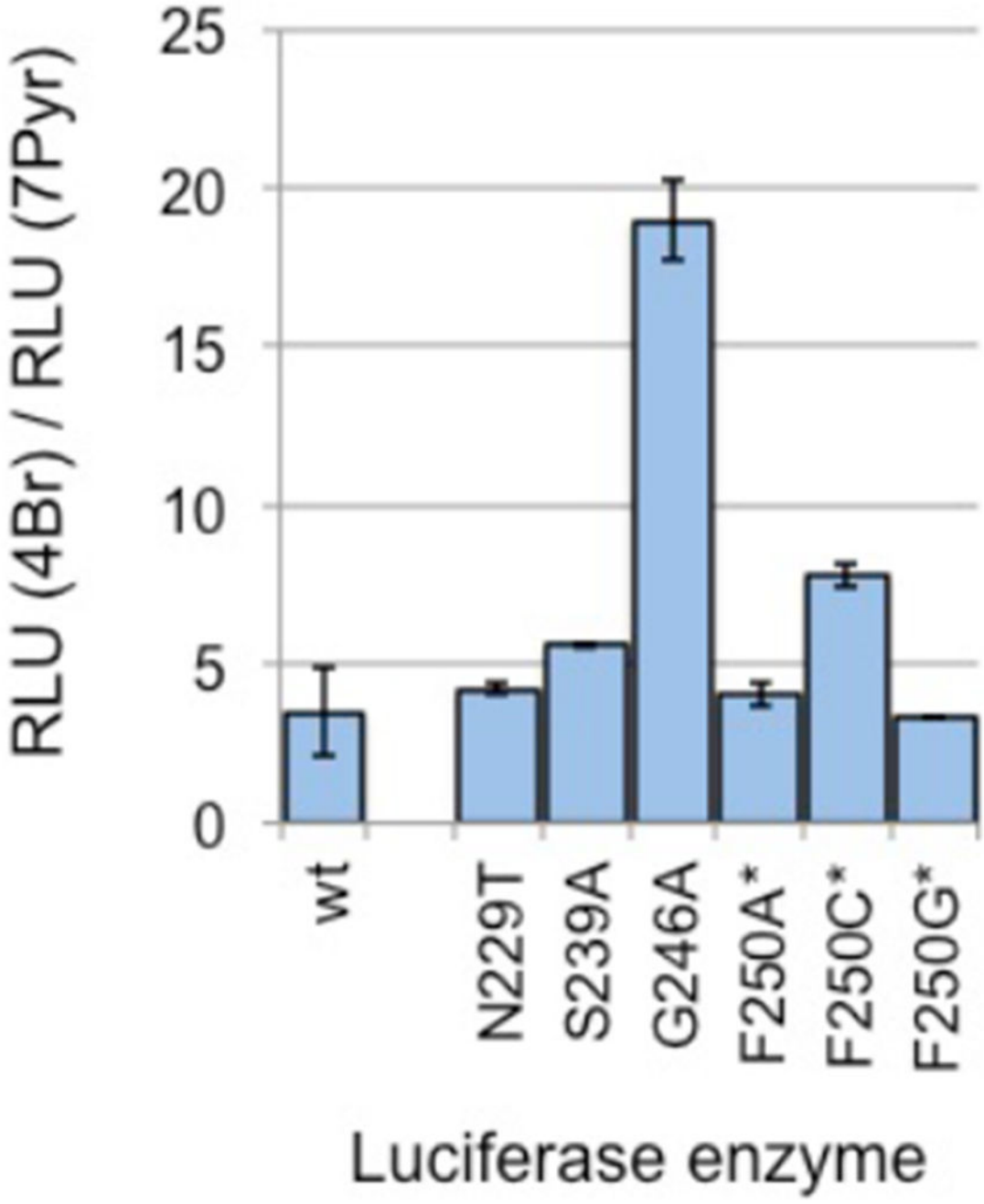


Figure 6.

Selectivity of frequently occurring mutants. For each replicate, a single lysate was split and incubated with either modified luciferin and unfiltered emission was recorded for 10s. Enzymes indicated with an asterisk indicate no 7'Pyr-Luc activity was observed; thus, the observed ratio is a lower limit on selectivity. All mutants were assayed in triplicate.

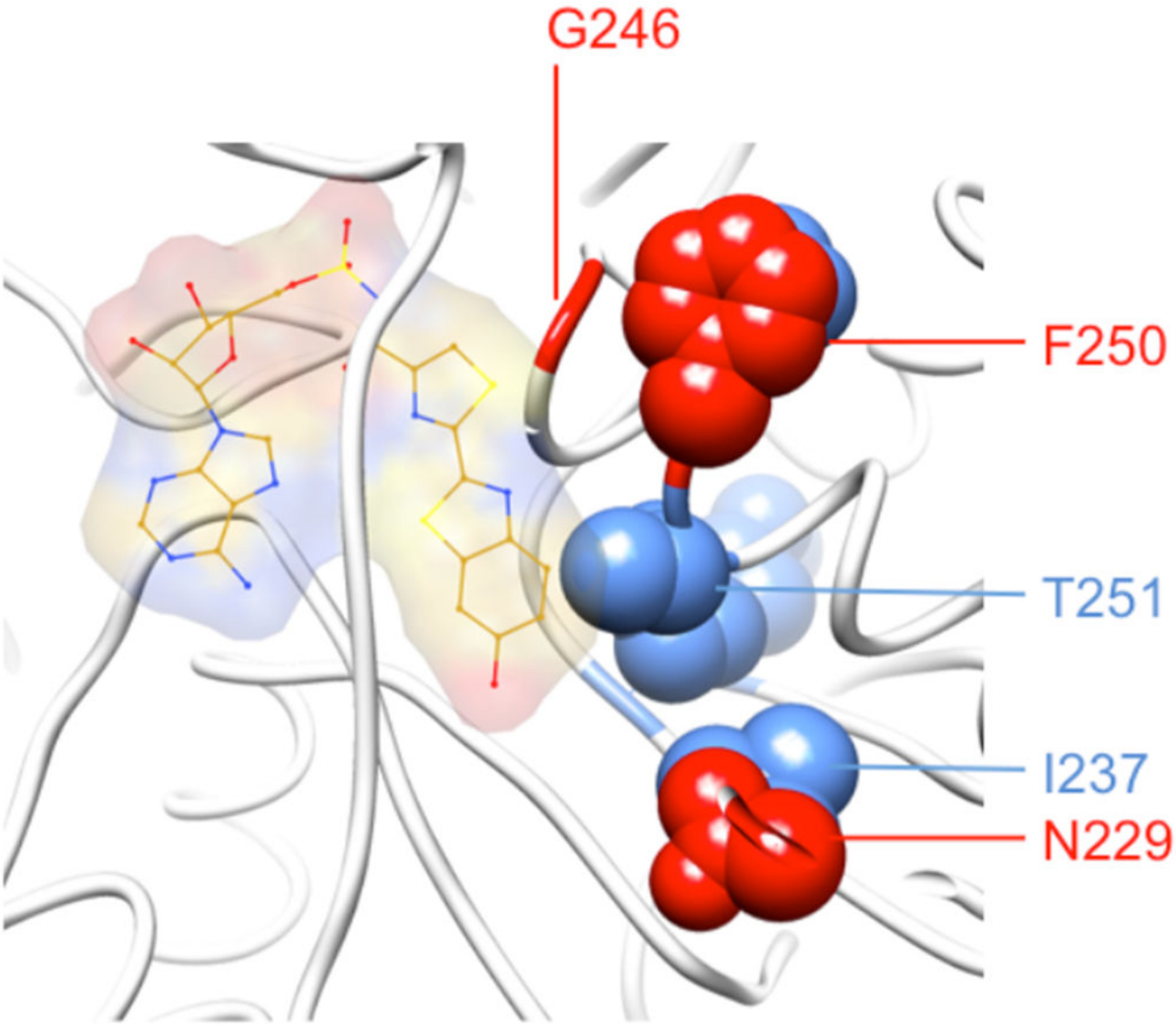


Figure 7.

Positions of frequently observed mutations. Active site of homologous enzyme L. cruciata (PDB ID: 2D1S) with luciferin analog shown in goldenrod with surface, key mutated amino acids shown in red, and other SCA-identified positions shown in blue. Although T251 is most proximal to the 4' position of luciferin, it is not observed to be mutated in our studies while other amino acid positions (G246, F250, N229T) are observed to be mutated in all recovered mutants.