

De novo designed luciferase

5

Reference to Related Applications

This application claims priority to U.S. Provisional Patent Application Serial Nos. 63/300,171 filed January 17, 2022 and 63/381,922 filed November 1, 2022, each incorporated by reference herein in their entirety.

10

Federal Funds Statement

This invention was made with government support under Grant No. K99EB031913, awarded by the National Institutes of Health. The government has certain rights in the invention.

15

Sequence Listing Statement

A computer readable form of the Sequence Listing is filed with this application by electronic submission and is incorporated into this application by reference in its entirety. The Sequence Listing is contained in the file created on January 8, 2023 having the file name "21-1622-WO.xml" and is 638 kb in size.

20

Background of the Invention

Bioluminescent light produced by the enzymatic oxidation of a luciferin substrate is widely used for bioassays and imaging in biomedical research. Because no excitation light source is needed, luminescent photons are produced in the dark which results in higher sensitivity than fluorescence imaging in live animal models and in biological samples where autofluorescence or phototoxicity is a concern. However, the development of luciferases as molecular probes has lagged behind that of well-developed fluorescent protein toolkits for a number of reasons: (i) very few native luciferases have been identified; (ii) many of those that have been identified require multiple disulfide bonds to stabilize the structure and are therefore prone to misfolding in mammalian cells; (iii) most native luciferases do not recognize synthetic luciferins with more desirable photophysical properties; and (iv) multiplexed imaging to follow multiple processes in parallel using mutually orthogonal luciferase-luciferin pairs has been limited by the low substrate specificity of native luciferases.

35

Summary of the Invention

In one aspect, the disclosure provides proteins having luciferase activity, comprising the secondary structure arrangement H1-L1-H2-L2-E1-L3-E2-L4-H3-L5-E3-L6-E4-L7-E5-L8-E6, wherein "H" is a helical domain, "L" is a loop domain, and "E" is a beta strand domain; wherein:

- (a) the H1 domain is at least 18 or 19 amino acids in length; residue 14 of the H1 domain is Y, D, or E, and residue 18 of the H1 domain is D or E;
- (b) the E3 domain is at least 6, 7, 8, 9, or 10 amino acids in length and residue 2 of the E3 domain is R; and
- (c) the E5 domain is at least 10, 11, 12, 13, or 14 amino acids in length and residue 9 of the E5 domain is H or N.

In various embodiments, 7 of the E5 domain is M; the E6 domain is at least 9, 10, 11, 12, or 13 amino acids in length and wherein residue 5 of the E6 domain is V; residue 1 of the L5 domain is S; residue 7 of the E5 domain is M and residue 5 of the E6 domain is V; and/or residue 7 of the E5 domain is M, residue 5 of the E6 domain is V, and residue 1 of the L5 domain is S.

In other embodiments, the H2 domain is at least 5, 6, or 7 amino acids in length, the H3 domain is at least 9, 10, 11, 12, 13, or 14 amino acids in length, the E1 domain is at least 3 or 4 amino acids in length, the E2 domain is at least 3 or 4 amino acids in length, and/or the E4 domain is at least 8, 9, 10, 11, or 12 amino acids in length.

In one embodiment, 1, 2, 3, 4, or all 5 of the following is true:

- (a) residue 13 of domain H1 is F;
- (b) residue 1 of domain L3 is W;
- (c) residue 5 of domain E5 is V or another hydrophobic residue;
- (d) residue 8 of domain E5 is A or L or another hydrophobic residue; and/or
- (e) residue 11 of domain E5 is W.

In another embodiment, 1, 2, 3, 4, 5, or all 6 of the following are true:

- (a) residue 2 of domain E1 is I or another hydrophobic residue;
- (b) residue 4 of domain H3 is F;
- (c) residue 6 of domain E4 is V or another hydrophobic residue;
- (d) residue 8 of domain E4 is L or another hydrophobic residue;
- (e) residue 5 of domain E6 is M or V or another hydrophobic residue; and/or
- (f) residue 7 of domain E6 is V or another hydrophobic residue.

In a further embodiment, the protein comprises an amino acid sequence at least 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence selected from the group consisting of SEQ ID NO:1-181, or SEQ ID NO:1-3. In another embodiment, the
 5 protein comprises the amino acid sequence of SEQ ID NO:4.

In one aspect, the disclosure provides proteins having luciferase activity, and comprising an amino acid sequence at least 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of SEQ ID NO:1, wherein:
 10 Residue 14 is Y, D, or E and residue 98 is H or N; and
 Residue 18 is D or E and residue 65 is R.

In various embodiments, the protein comprises one or both of A96M and M110V substitutions relative to SEQ ID NO:1; both of A96M and M110V substitutions relative to SEQ ID NO:1; and/or an R60S substitution relative to SEQ ID NO:1; R60S, A96M, and
 15 M110V substitutions relative to SEQ ID NO:1. In other embodiments, the protein comprises an amino acid sequence at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence selected from the group consisting of SEQ ID NO:1-3, or 1-181.

In another aspect, the disclosure provides a protein comprising the formula X1-Z1-
 20 X2-Z2-X3-Z3-X4-Z4-X5-Z5-X6-Z6-X7-Z7-X8-Z8, wherein:

X1 has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of MSEEQIRQFL RRFYEALD (SEQ ID NO: 182), wherein residue 14 is Y, D, or E and residue 18 is D or E;

25 X2 has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of ADTAASLF (SEQ ID NO: 183);

X3 has an amino acid sequence at least 50%, 75%, or 100% identical to the amino acid sequence of TIHL (SEQ ID NO: 184);

30 X4 has an amino acid sequence at least 33%, 66%, or 100% identical to the amino acid sequence of VTF;

X5 has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of EEFR EWFERLFST (SEQ ID NO: 185) ;

5 X6 has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of QREIKSL EVR (SEQ ID NO: 186) , wherein residue 2 is R;

X7 has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of VEVH VQLHATH (SEQ ID NO: 187) ;

10 X8 has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of KHTVDATHHW HFR (SEQ ID NO: 188) , wherein residue 8 is H or N;

X9 has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid
15 sequence of VTEM RVHINPTG (SEQ ID NO: 189) ; and

wherein Z1, Z2, Z3, Z4, Z5, Z6, Z7, and Z8 are independently present or absent, and when present may comprise any amino acid sequence.

In another embodiment, the disclosure provides self-complementing multipartite protein shaving luciferase activity, comprising at least a first polypeptide component and a
20 second polypeptide component, wherein the at least first polypeptide component and the second polypeptide component are not covalently linked, wherein in total the at least first polypeptide component and the second polypeptide component comprise domains X1-Z1-X2-Z2-X3-Z3-X4-Z4-X5-Z5-X6-Z6-X7-Z7-X8-Z8-X9, wherein each domain is as defined herein;

25 wherein (a) each X domain is fully present within one polypeptide component of the at least first polypeptide component and the second polypeptide component, and (b) none of the at least first polypeptide component and the second polypeptide component include each of X1, X2, X3, X4, X5, X6, X7, X8, and X9.

In a further embodiment, the disclosure provides self-complementing multipartite
30 protein having luciferase activity, comprising at least a first polypeptide component and a second polypeptide component, wherein the at least first polypeptide component and the second polypeptide component are not covalently linked, wherein in total the at least first polypeptide component and the second polypeptide component comprise the secondary

structure arrangement H1-L1-H2-L2-E1-L3-E2-L4-H3-L5-E3-L6-E4-L7-E5-L8-E6, wherein each domain is as defined herein

In a further embodiment, the disclosure provides proteins having luciferase activity, comprising the secondary structure arrangement H1-L1-H2-L2-E1-L3-E2-L4-H3-L5-E3-L6-E4-L7-E5-L8-E6, wherein “H” is a helical domain, “L” is a loop domain, and “E” is a beta strand domain; wherein:

- (a) the H1 domain is at least 18 or 19 amino acids in length; residue 14 of the H1 domain is Y, D, or E, and residue 18 of the H1 domain is D or E;
- (b) the E3 domain is at least 6, 7, 8, 9, or 10 amino acids in length and residue 2 of the E3 domain is R; and
- (c) the E5 domain is at least 10, 11, 12, 13, or 14 amino acids in length and residue 9 of the E5 domain is H or N.

The disclosure also provides fusion proteins comprising:

- (a) the protein or polypeptide component of any embodiment; and
- (b) one or more additional functional domains.

The disclosure further provides nucleic acids encoding the protein, polypeptide component, or fusion proteins of the disclosure, expression vector comprising the nucleic acids operatively linked to a suitable control element, host cells comprising a protein, polypeptide component, fusion protein, nucleic acid, and/or expression vector of the disclosure; and kits comprising a protein, polypeptide component, fusion protein, nucleic acid, expression vector, and/or host cell of the disclosure; and instructions for their use. The disclosure also provides methods for use of a protein, polypeptide component, fusion protein, nucleic acid, expression vector, host cell, and/or kit of the disclosure.

Brief Description of the Figures

Figure 1: Generation of idealized scaffolds and computational design of de novo luciferases. (a) Family-wide hallucination. Sequences encoding proteins with the desired topology are optimized by Monte Carlo sampling with a multicomponent loss function. Structurally conserved regions are evaluated based on consistency with input residue-residue distance and orientation distributions obtained from 85 experimental structures of NTF2-like proteins, while variable non-ideal regions are evaluated based on the confidence of predicted inter-residue geometries calculated as the KL-divergence between network predictions and the background distribution. The sequence-space MCMC-sampling incorporates both sequence changes and insertions/deletions (see Methods) to guide the hallucinated sequence

towards encoding structures with the desired folds. Hydrogen-bonding networks are incorporated into the designed structures to increase structural specificity. (b-d) The design of luciferase active sites. (b) Generation of DTZ conformers using AIMNet. (c) Generation of a rotamer interaction field (RIF) to stabilize the anionic DTZ and form hydrophobic packing interactions around the DTZ conformers. (d) Docking of the RIF into the hallucinated scaffolds, and optimization of substrate-scaffold interactions using position-specific score matrices (PSSM)-biased sequence design. (e) Selection of the NTF2 topology. The RIF was docked into 4000 native small molecule binding proteins, excluding proteins that bind the luciferin substrate using more than 5 loop residues. Most of the top hits were from the NTF2-like protein superfamily. Using the family-wide hallucination scaffold generation protocol, we generated 1615 scaffolds and found that these yielded better predicted RIF binding energies than the native proteins. (f) Scaffolds generated with family-wide hallucination sample more within the space of the native structures than previous blueprint generated scaffolds, and (g) have stronger sequence to structure relationships than native or blueprint de novo NTF2 scaffolds.

Figure 2. Biophysical characterization of LuxSit. (a) Coomassie-stained SDS-PAGE of purified recombinant LuxSit from *E. coli*. (b) Size-exclusion chromatography of purified LuxSit suggested monodispersed and monomeric properties. (c) Far-ultraviolet CD spectra at 25 °C, 95 °C, and cooled back to 25 °C. Insert: CD melting curve of LuxSit at 220 nm. (d) Luminescence emission spectra of DTZ in the presence and absence of LuxSit. (e) Structural alignment of the design model and AlphaFold2 predicted model, which are in close agreement at both backbone (left) and sidechain level (right). (f-i) Site saturation mutagenesis of substrate interacting residues. Zoomed-in views (left) of design and AlphaFold2 models at sidechain level illustrated the designed enzyme-substrate interactions of (f), Tyr14-His98 core HBNets, (g), Asp18-Arg65 dyad, (h), π -stacking, and i, hydrophobic packing residues. Sequence profiles (right) are scaled by the activities of different sequence variants: (activity for the indicated amino acid) / (the sum of activities over all tested amino acids at the position). Substitutions with increased activity (Ala96 and Met110) are highlighted.

Figure 3. Characterization of luciferase activity in vitro and in human cells. (a) Substrate concentration dependence of LuxSit, LuxSit-f, and LuxSit-i activity. Numbers indicate the signal-to-background ratio at V_{max} . (b) Fluorescence and luminescence imaging of live HEK293T cells transiently expressing LuxSit-i-mTagBFP2; LuxSit-i activity can be detected at single-cell resolution. Left: fluorescence channel representing mTagBFP2 signal. Right: total luminescence photons were collected during a course of 10 s exposure. Inserts:

negative control, untransfected cells with DTZ. The luminescence images were acquired immediately after adding 25 μ M DTZ without excitation light. Scale bar: 20 μ m. 40X.

Figure 4. High substrate specificity of designed luciferases allows multiplexed bioassay. (a) Chemical structures of Coelenterazine substrate analogs. (b) Activity of LuxSit-i on selected luciferin substrates. Luminescence image (top) and signal quantification (bottom) of the indicated substrate in the presence of 100 nM LuxSit-i. LuxSit-i has high specificity for the design target substrate, DTZ. (c) Heatmap visualization of the substrate specificity of LuxSit-i, *Renilla* luciferase (RLuc), *Gussia* luciferase (GLuc), engineered NLuc from *Oplophorus* luciferase. The heatmap shows luminescence for each enzyme on each substrate; values are normalized on a per-enzyme basis to the highest signal for that enzyme over all substrates. (d) Luminescence emission spectrum of LuxSit-i/DTZ and RLuc/PP-CTZ can be spectrally resolved by 528/20 and 390/35 filters (shown in dashed bars) and only recognize the cognate substrate. (e) Schematic of the multiplex luciferase assay. HEK293T cells transiently transfected with CRE-RLuc, NFkB-LuxSit-i, and CMV-CyOFP plasmids were treated with either Forskolin or human tumor necrosis factor alpha (TNF α) to induce the expression of labeled luciferases. (f-g) Luminescence signals from cells can be measured under either substrate-resolved or spectrally resolved methods by a plate reader. (f) For the substrate-resolved method, luminescence intensity was recorded without a filter after adding either PP-CTZ or DTZ. (g) For the spectrally resolved method, both PP-CTZ and DTZ were added, and the signals were acquired using 528/20 and 390/35 filters simultaneously. In (f) and (g), the lower panel indicates the addition of Forskolin or TNF α . Luminescence signals were acquired from the lysate of 15,000 cells in CellLytic™ M reagent while CyOFP fluorescence signal was used to normalize cell numbers and transfection efficiencies. All data were normalized to the corresponding non-stimulated control. Data are presented as mean $\pm$ SD (n = 3).

Figure 5. Proposed catalytic mechanism of coelenterazine-utilizing luciferases. Density-functional theory (DFT) calculation suggested that the formation of an anionic state is the essential electron source for the activation of triplet oxygen ($^3\text{O}_2$). Supported by both theoretical^{26,27} and experimental evidence^{28,29}, the next oxygenation process is likely through a single electron transfer (SET) mechanism in which the surrounding reaction field could highly influence the change of Gibbs free energy (ΔG_{SET}). Finally, the thermolysis of a dioxetane light emitter intermediate can produce photons via the mechanism of gradually reversible charge-transfer-induced luminescence (GRCTIL), which is generally exergonic. Since all the historical pieces of evidence are based on calculations in the virtual solvents or

chemiluminescence in ideal organic solvents. The detailed mechanism of a luciferase-catalyzed luminescence reaction has remained unclear. We proposed that the key step of the enzyme is to promote the formation of an anionic state and create a suitable environment to facilitate efficient SET. Hence, the goal of this study is to design an enzyme reaction field surrounding the substrate to stabilize the anionic substrate state and alter the local proton activity, solvent polarity, and hydrophobicity for the efficient activation of $^3\text{O}_2$.

Figure 6. Schematic representative of colony-based luciferase screening.

Computationally designed DNA sequences were provided in an oligo array, where the fragments were amplified by PCR, assembled, and ligated into a pBAD bacterial expression vector. The plasmid library was used to transform DH10B cells. Each colony grown on the LB agar plate represented one luciferase design. The plates were sprayed with DTZ solution and imaged to identify active colonies using a ChemiDoc™ imager. All active colonies were inoculated in 96-well plates, expressed, and purified to confirm individual luciferase activity. Selected plasmids can then be sequenced to point out active design models that provide insights into the design principle and enzyme functions or can be subjected to random mutagenesis for further evolution. Insert: three luciferases were identified from this screening. We refer to the most active and DTZ-specific luciferase as “LuxSit”.

Figure 7. Expression, purification, and structural characterization of LuxSit variants.

(a-c) The recombinant expression of (a) LuxSit, (b) LuxSit-i, and (c) LuxSit-f in *E. coli*. Annotations for each lane are the following – 1: Pre-IPTG; 2: Post-IPTG; 3: Soluble lysate; 4: Flow-through; 5: Wash; 6: Elution; 7: Post-TEV cleavage; 8: Post-SEC. (d-f) Size-exclusion chromatography of the purified (d) LuxSit; (e) LuxSit-i; and (f) LuxSit-f monomer. (g-i) Deconvoluted mass spectrum of (g) LuxSit, (h) LuxSit-i, and (i) LuxSit-f. (j-k) Far-ultraviolet circular dichroism (CD) spectra (Left panel) of (j) LuxSit-i; and (k) LuxSit-f at 25 °C, 95 °C and cooled back to 25 °C. CD melting curve at 220 nm (Right panel). (l) Dimeric SEC peak was observed when LuxSit-i was concentrated to high concentration (~50 μM) in Tris pH 8.0 buffer. Both dimeric and monomeric SEC fractions showed the expected size on SDS PAGE and both peaks were catalytically active to emit luminescence in the presence of 25 μM DTZ.

Figure 8. Screening of a randomized NNK library at 60, 96, and 110 positions and sequence alignment between LuxSit and its variants. We generated a fully randomized library at 60, 96, and 110 positions to exhaustively screen all possible combinations. After the colony-based screening, we identified many colonies with strong luciferase activities with DTZ. Each colony was expressed individually in each well of 96-

well plates (1 mL culture) and purified accordingly (see Methods). (a) Individual luminescence activity of each selected mutant was plotted and compared to the parent LuxSit. Luminescence activities were measured in the presence of 25 μ M DTZ. Luminescence activity (RLU) was shown as the integrated signal over the first 15 min. Statistical analysis of the amino acid frequency versus the luciferase activity at residue (b) 60, (c) 96, and (d) 110. Among all selected mutants, Arg60 is confirmed to be mutable as Arg60 may be structurally less well defined as it emanates from a loop and has no hydrogen-bonding partner. Ala96 prefers larger sidechain (Leu, Ile, Met, and Cys), and Met110 favors hydrophobic residues (Val, Ile, and Ala). A newly discovered variant (R60S/A96L/M110V) with more than 100-fold higher photon flux over LuxSit was assigned LuxSit-i for its high brightness.

Figure 9. Sequence alignment of Lux-Sit (SEQ ID NO: 1), Lux-Sit-i (SEQ ID NO: 2), and Lux-Sit-f (SEQ ID NO: 3). In the sequence alignment, mutations are highlighted. The conserved catalytic dyads of Asp18-Arg65 and Tyr14-His98 are shown.

Figure 10. Additional characterization of LuxSit variants. (a) Normalized emission kinetics of 15,000 intact HeLa cells expressing LuxSit-i, 100 nM purified LuxSit-i, or 100 nM purified LuxSit-f in the presence of 50 μ M DTZ. The more extended emission kinetics in HeLa cells is likely due to the diffusion rate of DTZ across cell membranes. (b) Normalized luminescence decay curves of LuxSit-i in various pH buffers revealed a pH-dependent catalytic mechanism. (c) Luminescent quantum yield was estimated from the integrated luminescence signal until completely converting 125 pmol substrates to photons in the presence of 50 nM corresponding luciferase (see Methods). All data points were plotted as the average of triplicate measurements.

Figure 11. Expression, localization, and luminescence activity of LuxSit-i in live HEK293T and HeLa cells. (a-b) Fluorescence imaging of live (a) HEK293T and (b) HeLa cells expressing LuxSit-i-mTagBFP2, which is untargeted or localized to the nucleus (Histone2B), plasma membrane (KRasCAAX), or mitochondria (DAKAP) cellular compartments. Scale bar: 10 μ m. (c-d) Luminescence signals were measured with 15,000 intact (c) HEK293T or (d) HeLa cells in the presence of 25 μ M DTZ in DPBS. Transfection efficiencies range from 60-70% for HEK293T cells and 5-10% for HeLa cells. (e) Luminescence emission spectra acquired from LuxSit-i expressing HEK293T cells is consistent with the emission spectra of recombinant LuxSit-i purified from *E. coli*. (f-g) Luminescence signals were measured with 15,000 (f) intact LuxSit-i expressing HEK293T cells or (g) cell lysate in the presence of 25 μ M indicated substrate in DPBS. Luminescence intensities were normalized to DTZ signal, showing high DTZ specificity over other

substrates in cell-based assays. Data were shown as total luminescence signal over the first 20 min and were done in technical triplicates. (h) Normalized luminescence intensity profile of lines traversing across different cells (n=10) of main Fig. 3b luminescence image; lines represent untransfected cells. Error bars represent $\pm$ SEM.

5 **Figure 12. Substrate specificity of LuxSit-i and spectrally resolved luciferase-luciferin pairs allow multiplexed bioassay.** (a) The orthogonality relationship between LuxSit-i-DTZ and RLuc-PP-CTZ (Prolume Purple, methoxy e-Coelenterazine) luminescent pairs. Indicated amounts of each luciferase were mixed at different ratios totaling 100%. (b) After the addition of both 25 μ M DTZ and PP-CTZ substrates, filtered light from 528/20 and 10 390/35 were measured simultaneously. Data are presented as mean $\pm$ SD (n = 3). Heatmap shows the luminescence signal for individual luciferase (100 nM) or 1:1 mixture in the presence of the cognate or non-cognate (DTZ or PP-CTZ or both) substrates. Response signals were acquired by a Neo2™ plate reader with 528/20 and 390/35 filters simultaneously. (c) Multiplex luciferase assay in live HEK293T after co-transfection of CRE- 15 RLuc, NF κ B-LuxSit-i, and CMV-CyOFP plasmids and stimulation by Forskolin (FSK) or human tumor necrosis factor alpha (TNF α). (d,e) 15,000 intact cells were assayed (see Methods) by either (d) substrate-resolved or (e) spectrally resolved modes after adding DTZ, PP-CTZ, or both DTZ and PP-CTZ in DPBS without cell lysis. Area-scanning of CyOFP fluorescence signal was used to estimate cell numbers and transfection efficiency. The 20 reported unit was RLU/a.u.; relative light units/fluorescence intensity measurements at Ex./Em.=480/580 nm. All data were normalized to the corresponding non-stimulated control. Data are presented as mean $\pm$ SD (n = 3).

25 **Figure 13.** Secondary structure is shown mapped onto an exemplary protein of the disclosure (SEQ ID NO: 1).

Detailed Description

All references cited are herein incorporated by reference in their entirety.

As used herein, the singular forms "a", "an" and "the" include plural referents unless the context clearly dictates otherwise.

30 As used herein, the amino acid residues are abbreviated as follows: alanine (Ala; A), asparagine (Asn; N), aspartic acid (Asp; D), arginine (Arg; R), cysteine (Cys; C), glutamic acid (Glu; E), glutamine (Gln; Q), glycine (Gly; G), histidine (His; H), isoleucine (Ile; I), leucine (Leu; L), lysine (Lys; K), methionine (Met; M), phenylalanine (Phe; F), proline (Pro;

P), serine (Ser; S), threonine (Thr; T), tryptophan (Trp; W), tyrosine (Tyr; Y), and valine (Val; V).

In all embodiments of polypeptides disclosed herein, any N-terminal methionine residues are optional (i.e.: the N-terminal methionine residue may be present or may be
5 deleted).

All embodiments of any aspect of the disclosure can be used in combination, unless the context clearly dictates otherwise.

Unless the context clearly requires otherwise, throughout the description and the claims, the words 'comprise', 'comprising', and the like are to be construed in an inclusive
10 sense as opposed to an exclusive or exhaustive sense; that is to say, in the sense of "including, but not limited to". Words using the singular or plural number also include the plural and singular number, respectively. Additionally, the words "herein," "above," and "below" and words of similar import, when used in this application, shall refer to this application as a whole and not to any particular portions of the application

15 In a first aspect, the disclosure provides proteins having luciferase activity, comprising the secondary structure arrangement H1-L1-H2-L2-E1-L3-E2-L4-H3-L5-E3-L6-E4-L7-E5-L8-E6, wherein "H" is a helical domain, "L" is a loop domain, and "E" is a beta strand domain; wherein:

(a) the H1 domain is at least 18 or 19 amino acids in length; residue 14 of the H1
20 domain is Y, D, or E, and residue 18 of the H1 domain is D or E;

(b) the E3 domain is at least 6, 7, 8, 9, or 10 amino acids in length and residue 2 of the E3 domain is R; and

(c) the E5 domain is at least 10, 11, 12, 13, or 14 amino acids in length and residue 9 of the E5 domain is H or N.

25 As disclosed in the examples that follow, the proteins of the disclosure are non-naturally occurring, have luciferase activity and share this recited secondary structure arrangement. The arrangement is shown with respect to the amino acid sequence of SEQ ID NO:1 in Figure 13. The inventors have conducted extensive studies to assess key residues in the polypeptides for retaining luciferase activity and made a large number of modified
30 versions of the polypeptides as detailed in SEQ ID NO:4 and the examples that follow. The required amino acids noted above are those involved in the catalytic dyads, as described below and in the examples.

Except as noted, the different domains may be any suitable length.

In one embodiment, residue 7 of the E5 domain is M. In another embodiment, the E6 domain is at least 9, 10, 11, 12, or 13 amino acids in length and residue 5 of the E6 domain is V. In a further embodiment, residue 1 of the L5 domain is S. In one embodiment, residue 7 of the E5 domain is M and residue 5 of the E6 domain is V. In a further domain, residue 7 of the E5 domain is M, residue 5 of the E6 domain is V, and residue 1 of the L5 domain is S. In another embodiment, the H2 domain is at least 5, 6, or 7 amino acids in length, the H3 domain is at least 9, 10, 11, 12, 13, or 14 amino acids in length, the E1 domain is at least 3 or 4 amino acids in length, the E2 domain is at least 3 or 4 amino acids in length, and the E4 domain is at least 8, 9, 10, 11, or 12 amino acids in length. In all these embodiments, one or more of the recited domains may independently include amino acid residues. In one embodiment, one or more of the domains may independently include an additional 1, 2, 3, 4, or 5 residues.

In one embodiment:

the H1 domain is 19 amino acids in length;
the H2 domain is 7 amino acids in length;
the E1 domain is 4 amino acids in length;
the E2 domain is 4 amino acids in length;
the H3 domain is 14 amino acids in length;
the E3 domain is 10 amino acids in length;
the E4 domain is 12 amino acids in length;
the E5 domain is 14 amino acids in length; and
the E6 domain is 12 or 13 amino acids in length.

The loop domains may be of any length and may include insertions, relative to the sequences exemplified herein, of any residues or functional domains as deemed appropriate, including but not limited to metal binding domains, drug binding domains, GPCR receptors, protein switches, and small molecule binding domains.

In another embodiment, the proteins comprise an amino acid sequence at least 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence selected from the group consisting of SEQ ID NO:1-3, wherein residues in parentheses are optional and may be present or may be deleted.

SEQ ID NO:1 is the Lux-Sit construct disclosed herein. **Figure 13** shows the domain structure mapped onto the SEQ ID NO:1 amino acid sequence.

(M) SEEQIRQFL RRFYEALDSG DADTAASLFH PGVTIHLWDG VTFTSREEFR
EWFERLEFSTR KDAQREIKSL EVRGDTVEVH VQLHATHNGQ KHTVDATHHW
HFRGNRVTEM RVHINPT(G) (SEQ ID NO:1)

SEQ ID NO:2 is the Lux-Sit-i construct disclosed herein.

5 (M) SEEQIRQFL RRFYEALDSG DADTAASLFH PGVTIHLWDG VTFTSREEFR
EWFERLEFSTR KDAQREIKSL EVRGDTVEVH VQLHATHNGQ KHTVDLTHHW
HFRGNRVTEV RVHINPT(G) (SEQ ID NO:2)

SEQ ID NO:3 is the Lux-Sit-f construct disclosed herein.

(M) SEEQIRQFL RRFYEALDSG DADTAASLFH PGVTIHLWDG VTFTSREEFR
10 EWFERLEFSTR KDAQREIKSL EVRGDTVEVH VQLHATHNGQ KHTVDMTHHW
HFRGNRVTEV RVHINPT(G) (SEQ ID NO:3) .

In each of the annotated sequences shown for SEQ ID NO:1-3:

- (a) Bold and underlined and increased font size positions are Dyad 1 (catalytic
15 residues) Y14 (H1 domain residue 14) + H98 (E5 domain residue 9);
- (b) Bold and increased font size positions are Dyad 2 (catalytic residues) D18 (H1
domain residue 9) + R65 (E3 domain residue 2);
- (c) Increased font and not bolded positions are core packing (recognition residues)
F13 (residue 13 of domain H1), I35 (residue 2 of domain E1), W38 (residue 1 of domain L3),
20 F49 (residue 4 of domain H3), V81 (residue 6 of domain E4), L83 (residue 8 of domain E4),
V94 (residue 5 of domain E5), A/L 97 (residue 8 of domain E5), W100 (residue 11 of domain
E5), M/V110 (residue 5 of domain E6), V112 (residue 7 of domain E6); and
- (d) Underlined and not bolded positions are regions (loop domains or immediately
adjacent) for splitting the enzyme or inserting other functional domains.

25 In some embodiments of the proteins, 1, 2, 3, 4, or all 5 of the following is true:

- (a) residue 13 of domain H1 is F;
(b) residue 1 of domain L3 is W;
(c) residue 5 of domain E5 is V or another hydrophobic residue;
(d) residue 8 of domain E5 is A or L or another hydrophobic residue; and/or
30 (e) residue 11 of domain E5 is W.

In other embodiments of the proteins, the H2 domain is 7 amino acids in length, the
H3 domain is 14 amino acids in length, the E1 domain is 4 amino acids in length, the E2
domain is 4 amino acids in length, and/or the E4 domain is 12 amino acids in length. In
further embodiments, 1, 2, 3, 4, 5, or all 6 of the following are true:

- (a) residue 2 of domain E1 is I or another hydrophobic residue;
- (b) residue 4 of domain H3 is F;
- (c) residue 6 of domain E4 is V or another hydrophobic residue;
- (d) residue 8 of domain E4 is L or another hydrophobic residue;
- 5 (e) residue 5 of domain E6 is M or V or another hydrophobic residue; and/or
- (f) residue 7 of domain E6 is V or another hydrophobic residue.

In another embodiment, the protein comprises the amino acid sequence of SEQ ID NO:4.

10 **SEQ ID NO:4**

position 1: M, or absent
position 2: S, T
position 3: A, E, K, S
position 4: A, D, E, K, Q, R, S, T
position 5: A, D, E, Q
position 6: I, Q
position 7: E, K, R
position 8: A, D, E, K, N, Q, R
position 9: F
position 10: C, N, V, W, Y, L
position 11: A, D, E, K, Q, R
position 12: K, Q, R, F
position 13: F
position 14: Y
position 15: A, D, E, K, L, N, Q, R, S
position 16: A
position 17: L
position 18: D
position 19: A, D, S
position 20: G
position 21: D
position 22: A
position 23: D, E, N, V

position 24: T
position 25: A
position 26: A, S
position 27: A, S
position 28: L
position 29: F
position 30: K, P, R, H
position 31: D, P
position 32: G
position 33: T, V
position 34: E, I, K, L, N, Q, R, V, T
position 35: I
position 36: D, E, H, K, Y
position 37: L
position 38: W
position 39: D
position 40: G
position 41: I, K, R, T, V
position 42: E, I, T, V
position 43: F
position 44: E, F, H, K, N, R, S, T, Y
position 45: K, T, S
position 46: K, Q, R
position 47: A, E
position 48: E, Q
position 49: F
position 50: K, Q, R
position 51: A, D, E, K, Q, S
position 52: W
position 53: F
position 54: E, K, R, V
position 55: A, D, E, K, Q, R, T

position 56: L
position 57: F, H, K, R, Y
position 58: A, S
position 59: E, K, L, Q, R, T
position 60: S, R
position 61: A, D, E, K, N, Q, S, T
position 62: A, D, E, G, N, Q
position 63: A
position 64: A, K, R, S, Q
position 65: R
position 66: D, E, H, K, R, S
position 67: I, V
position 68: E, I, T, V, K
position 69: A, D, E, K, N, Q, R, S
position 70: F, L, M
position 71: D, E, K, Q, R, S, T, V
position 72: V
position 73: A, D, E, H, I, N, R
position 74: G
position 75: D, N
position 76: D, E, F, I, K, R, T, V
position 77: A, S, V
position 78: D, E, F, H, K, L, N, R, T, Y
position 79: I, V
position 80: E, I, K, N, R, S, T, V, H
position 81: V
position 82: I, V, Q
position 83: L
position 84: D, E, H, K, R, V
position 85: A
position 86: D, E, F, I, K, N, R, S, T, V, Y
position 87: E, F, H, I, K, V, Y

position 88: A, D, E, K, N, Q, R
position 89: G
position 90: E, K, Q, R, T
position 91: A, D, E, H, K, P, Q
position 92: E, H, K, L, R, V
position 93: E, I, K, R, T, V
position 94: V
position 95: A, E, F, G, H, K, L, N, R, S, D
position 96: L, A, M
position 97: E, H, K, N, R, T, V, A, L
position 98: H
position 99: E, F, I, K, L, N, Q, R, T, V, W, Y, H
position 100: A, F, T, Y, W
position 101: E, F, H, K, L, Q, R, V, Y
position 102: F, W
position 103: D, E, K, R
position 104: G
position 105: D, S, N
position 106: E, H, K, Q, R
position 107: L, V
position 108: K, R, T, V
position 109: E, K, R
position 110: V, M
position 111: D, E, F, H, K, N, R, S, T, W, Y
position 112: V
position 113: A, D, E, H, K, R, S, T, V
position 114: I
position 115: D, E, F, H, I, K, N, Q, R, S, T, V, Y
position 116: P

position 117: D, L, M, V, T
Position 118: G or is absent

- In another embodiment, the proteins comprise an amino acid sequence at least 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid selected from the group
- 5 consisting of SEQ ID NO:1-181, as shown in Table 1. SEQ ID NO:5-181 in Table 1 are re-designed amino acid sequences based on LuxSit-i (SEQ ID NO:2), with their luciferase activities shown.

Table 1

Name	SEQ ID NO: (AA)	Amino acid Sequence	Activity	SEQ ID NO: (NT)	Nucleotide sequence
LuxSit	1	(M) SEEQIRQFLRRFY EALDSGDADTAASLFH PGVTIHLWDGVTFTSR EEFREWFERLFSTRKD AQREIKSLEVRGDTVE VHVQLHATHNGQKHTV DATHHWHFRGNRVTEM RVHINPT (G)	75.3	200	>LuxSit (codon optimized for E. coli expression) (ATG)AGCGAAGAACAGATTCGTCAGTTTCTGC GTCGTTTTTATGAAGCGCTGGATAGCGGCGATG CGGATACCGCTGCGAGCCTGTTTCATCCGGGCG TGACAATTCATCTGTGGGATGGCGTTACCTTTA CCAGCCGTGAAGAATTCGTGAATGGTTGAAC GTCTGTTTAGCACCCTGAAAGATGCGCAGCGTG AAATTAAGAGCCTGGAAGTACGTGGCGATACCG TGGAAGTGCATGTGCAGTTGCACGCGACCCATA ATGGCCAGAAACATACCGTAGATGCAACCCATC ATTGGCATTTCGTGGCAATCGTGTGACCGAAA TGCGTGTGCATATCAATCCGACC (GGCTAA)
LuxSit-i	2	(M) SEEQIRQFLRRFY EALDSGDADTAASLFH PGVTIHLWDGVTFTSR EEFREWFERLFSTSKD AQREIKSLEVRGDTVE VHVQLHATHNGQKHTV DLTHHWHFRGNRVTEV RVHINPT (G)	763.6	201	>LuxSit-i (codon optimized for E. coli expression) (ATG)AGCGAAGAACAGATTCGTCAGTTTCTGC GTCGTTTTTATGAAGCGCTGGATAGCGGCGATG CGGATACCGCTGCGAGCCTGTTTCATCCGGGCG TGACAATTCATCTGTGGGATGGCGTTACCTTTA CCAGCCGTGAAGAATTCGTGAATGGTTGAAC GTCTGTTTAGCACCAGTAAAGATGCGCAGCGTG AAATTAAGAGCCTGGAAGTACGTGGCGATACCG TGGAAGTGCATGTGCAGTTGCACGCGACCCATA ATGGCCAGAAACATACCGTAGATTTGACCCATC ATTGGCATTTCGTGGCAATCGTGTGACCGAAG TTCGTGTGCATATCAATCCGACC (GGCTAA)
				202	>LuxSit-i (codon optimized for human cell expression) (ATG)AGCGAGGAGCAGATCAGACAGTTCCTGA GGAGATTCTACGAGGCCCTGGATAGCGGAGACG CCGACACAGCTGCCAGCCTTTCCATCCTGGCG TGACCATCCACCTGTGGGACGGCGTCACCTTCA CTAGCAGGGAGGAGTTCAAGGAGTGGTTCGAGA GACTGTTCAAGCACCAGCAAGGACGCCAGAGAG

					AGATCAAGAGCCTGGAAGTTAGAGGCACACCG TGGAAGTGCACGTGCAGCTGCACGCCACACACA ACGGACAGAAGCACACCGTCGACCTGACCCACC ACTGGCACTTCAGAGGCAACAGAGTGACCGAGG TGAGAGTGCACATCAATCCCACC (GGCTAA)
LuxS it-f	3	(M) SEEQIRQFLRRFY EALDSGDADTAASLFH PGVTIHLWDGVTFTSR EEFREWFERLFSTRKD AQREIKSLEVRGDTVE VHVQLHATHNGQKHTV DMTHHWHFRGNPVTEV RVHINPT (G)	534. 5	203	> LuxSit-f (codon optimized for E. coli expression) (ATG) AGCGAAGAACAGATTTCGTGATTTCTGC GTCGTTTTTATGAAGCGCTGGATAGCGGCGATG CGGATACCGCTGCGAGCCTGTTTCATCCGGGCG TGACAATTCATCTGTGGGATGGCGTTACCTTTA CCAGCCGTGAAGAATTTCGTGAATGGTTTGAAC GTCTGTTTAGCACCCGCAAGATGCGCAGCGTG AAATTAAGAGCCTGGAAGTACGTGGCGATACCG TGGAAGTGCATGTGCAGTTGCACGCGACCCATA ATGGCCAGAAACATACCGTAGATATGACCCATC ATTGGCATTTCGTGGCAATCGTGTGACCGAAG TGCGTGTGCATATCAATCCGACC (GGCTAA)
luxs iti_ 0.3_ 386	181	MSEQAQRDFVKKFYEA LDAGDAETASALFPDG TQIYLWDGQTFTTQEQ FRAWFVELRSTSKDAK REVVSFKVDGNNAEVE VVLHIVIDGEEPTVKL KHVFQWEGDQLVKVTV SIEPL	152. 4478 231	204	(ATG) AGCGAACAGGCGCAGCGCATTTTGTGA AGAAATTTTATGAAGCCCTGGATGCGGGCGATG CCGAAACTGCAAGCGCACTGTTTCCGGATGGCA CCCAGATTTATCTGTGGGATGGCCAGACCTTTA CCACCCAGGAACAGTTTCGCGCGTGGTTTGTGG AACTGCGCAGCACCAGCAAGATGCGAAACGCG AAGTGGTGAGCTTTAAAGTGGATGGTAACAACG CGGAAGTGGAAAGTTGTGCTGCATGCGGTGATTG ATGGCGAAGAACGACCGTGAAACTGAAACATT ATTTTCAGTGGGAAGCGCATCAGCTGGTGAAAG TGACCGTGAGCATTGAACCGCTGGG (TAA)
luxs iti_ 0.3_ 389	5	MSEEQIREFCKRFYEA LDAGDAVTASALFPNG TRIHLWDGITTFTTQEE FREWFERLYSQSEDAK REIVSLEVEGNVAYVE VILHASFRGEEKTVRL RHVFQFEGDKLVEVTV EIEPL	421. 3061 507	205	(ATG) TCTGAAGAACAATTCGCGAATTTTGCA AACGCTTTTATGAAGCGCTGGATGCGGGCGATG CGGTGACTGCTAGCGCGTTGTTTCCGAATGGCA CCCGCATTTCATCTGTGGGATGGCATTACCTTTA CCACCCAGGAAGAATTTCTGGAATGGTTTGAAC GCCTGTATAGCCAGAGCGGAAGATGCGAAACGCG AAATTGTGAGCCTGGAAGTGGAAAGGCAACGTGG CGTATGTGGAAGTTATCTGCATGCGAGCTTTC GCGGTGAAGAGAAAACCGTGCGCCTGCGCCATG TTTTTCAGTTTGAAGCGATAAACTGGTTCGAAG TGACCGTGGAATTTGAACCGCTGGG (TAA)
luxs iti_ 0.3_ 392	6	MSEEAQREFWKRIFYEA LDAGDAETAAALFPDG TEIHLWDGTTFTTQAE FRAWFEELYSTSQNAK REIVEFRVDGNKSTVT VVLHAIYEGEEKVLL EHFTEWEGDRLVRVEV SIVPL	2.03 0407 74	206	(ATG) AGCGAAGAAGCGCAGCGCAATTTTGGA AACGCTTTTATGAAGCGCTGGATGCGGGCGATG CCGAAACTGCAGCGCGCTTGTTCCTGATGGCA CCGAAATCCATCTGTGGGATGGTACCACCTTTC GCACCCAGGCCGAATTTGCGCGTGGTTTGAAG AACTGTATTTCAGCAGCCAGAACGCGAAACGTG AAATTGTGGAATTCGCGTGGATGGCAACAAA GCACCGTGACCGTGGTGTGCATGCGATTACG AAGGTGAAGAAAAGAAAGTGTGCTGGAACATT TTACCGAATGGGAAGCGCATCGCCTGGTGCGCG TTGAAGTGAGCATCTGCGCGCTGGG (TAA)
luxs iti_ 0.3_ 395	7	MSEEEQREFVRRIFYEA LDAGDADTASALFPDG TVIELWDGTTFTTRAE FKAWFEALHSKSENAV RHVVEFEVEGNKAKVR VVLHADYKGGKRVVEL EHFTEFEGDRLVKVSV DIHPI	5.15 4803 041	207	(ATG) TCCGAAGAAGAAGCGTGAATTTGTGC GCCGCTTTTATGAAGCGCTGGATGCGGGTGATG CGGATACCGCGAGCGCACTGTTTCCAGATGGCA CCGTGATTGAACTGTGGGATGGTACCACCTTTC GCACCCGCGCGGAATTTAAAGCGTGGTTTGAAG CCCTGCATAGCAAAAGCGAAAACGCGTGCGCC ATGTGGTGGAAATTTGAAGTGGAAAGGCAACAAAG CCAAAGTGCGCGTGGTGTGCATGCTGATTATA AAGGCAAGAAACGCGTTGTGGAAGTGGAAACATG

					AATTCGAGTTCGAAGCGATCGCCTGGTGAAAG TGTCTGTGGATATTACCCGATTGG (TAA)
luxs iti_... 0.3_... 402	8	MSEEEIRKFCERFYAA LDAGDAETASALFPDG TEIHLWDGKTFRTQAE FREWFVRLRSTSDAR RHITKLKVDGNVAEVE VVLRSYDGEKVVVAL KHVFKFEGDKLVEVKV EITPL	95.7 8507 256	208	(ATG)AGCGAAGAAGAAATTCGCAAATTTTGTGC AACGCTTTTATGCGGCGCTGGATGCGGGCGATG CGGAAACTGCAAGCAGCCTGTTTCCTGATGGCA CCGAAATTCATCTGTGGGATGGCAAAACCTTTC GCACCCAGGCGGAATTCGCGAATGGTTTGTTC GTCTGCGCAGCACCAGCGATGATGCGCGCCGTC ATATTACCAAAGTGAAGTGGATGGTAACGTGG CGGAAGTGAAGTGTGTCTGCGCGCGAGCTATG ATGGTGAAGAAAAAGTGGTGGCGCTGAAACATG TGTTTTAAATTTGAAGGCGATAAACTGGTTGAAG TGAAAGTTGAAATTACCCCGCTGGG (TAA)
luxs iti_... 0.3_... 405	9	MSEEAQREFVRRFYEA LDAGDAETASALFPDG TQIYLWDGKTFRTREE FRSWFVKLHSTSDNAK RRVVEFRVEGNKAKVK VVLKASINGKERTVLL EHFFEFEGDKLVKVEV RIRPL	119. 6136 835	209	(ATG)AGCGAAGAAGCCAGCGTGAATTTGTGC GCCGCTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACCGCAAGCGCACTGTTTCCAGATGGCA CCCAGATTTATCTGTGGGATGGCAAAACCTTTA CCACCCGTGAAGAATTCGCGAGCTGGTTTGTGA AATTGCATAGCACCAGCGATAACGCGAAACGCC GCGTGGTTGAATTTAGAGTGAAGGCAACAAAG CGAAAGTAAAAGTGGTGTGAAAGCGAGCATT ACGGTAAAGAAGCGACCGTGTCTGCTGGAACATT TCTTTGAATTCGAAGGCGATAAACTGGTTAAAG TAGAAGTGGCGATTTCGCCGCTGGG (TAA)
luxs iti_... 0.3_... 443	10	MSEEEQREFVRRFYEA LDAGDAETASALFPDG TEIYLWDGRVFTQAE FRAWFVELRSKSDAR REVTSEFVVDGNKADVR VVLHANYKGEKVVVEL RHVYEFEGDRLVRVEV TINPL	45.8 8113 338	210	(ATG)TCTGAAGAAGAACAGCGCAATTTTGGC GTGCTTTTATGAAGCGCTGGATGCGGGCGATG CCGAAACCGCAAGCGCGTTGTTTCCTGATGGCA CCGAAATTTATCTGTGGGATGGCCGCGTGTTC GCACCCAGGCGGAATTCGCGCATGGTTTGTGG AACTGCGCAGCAAAAGCGATGATGCGCGCCGCG AAGTGACCACTTTTCGCGTGGATGGCAACAAAG CGGATGTGCGCGTGGTGTGATGCGCAACTATA AAGGCGAAAAGAAAGTGGTCGAACTGAGACATG TGTATGAATTTGAAGGCGATCGCCTGGTGGCTG TGGAAGTTACCATTAAACCGCTGGG (TAA)
luxs iti_... 0.3_... 454	11	MSEEEIREFVKRFYEA LDAGDAETASALFPDG TKIHLWDGTVFTTREE FKSWFVELRSTSDAK REVVDLKVGNKAYVE VVLRAIVNGEDKVVKL RHVFKFEGDKLVEVHV EIDPL	520. 3628 196	211	(ATG)AGCGAAGAAGAAATTCGCGAATTTGTGA AACGCTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACCGCGAGCGCATTTGTTCCGGATGGCA CCAAAATTCATCTGTGGGATGGTACCGTGTTTA CCACCCGTGAAGAATTTAAAGCTGGTTTGTGG AACTGCGCAGCACCAGCGATGATGCGAAACGCG AAGTGGTGGATCTGAAAGTGAAGGCAACAAAG CGTATGTGGAAGTGTGTCTGCGCGCGATCGTGA ACGGCGAAGATAAAGTGGTTAACTGCGTCATG TGTTTTAAATTTGAAGGCGATAAACTGGTCGAAG TGCAFGTTGAAATTTGATCCGCTGGG (TAA)
luxs iti_... 0.3_... 462	12	MSKEEQRKFFVERFYKA LDAGDAETASALFPDG TKIHLWDGTTFHTRAE FRAWFVLDLRSKSENAK REVVAFEVDGNKAKVV VVLKANYKGEERTVKL EHEFYFEGDHLVEVNV KIEPV	65.7 8991 016	212	(ATG)AGCAAAGAAGAACAGCGCAAATTTGTGG AACGCTTTTATAAAGCGCTGGATGCGGGCGATG CGGAAACTGCGAGCGCATTTGTTCCGGATGGCA CCAAAATTCATCTGTGGGATGGTACCACCTTTC ATACCCGCGCGGAATTCGCGCGTGGTTCTGTGG ATCTGCGCAGCAAAAGCGAAAACGCGAAACGTG AAGTGGTGGCGTTTGAAGTTGATGGCAATAAAG CCAAAGTGGTGTGGTGTCTGAAAGCAAACCTATA AAGGCGAAGAAGCGACCGTGAACCTGGAACATG AATTTTATTTTGAAGGCGATCATCTGGTGGAG TGAACGTTAAATTTGAACAGTGGG (TAA)

luxs iti_... 0.3... 468	13	MSEEEQREFVWRFYEA LDAGDAETASALFPDG TEIHLWDGKTEHTREE FRAWFVDLYSKSKDAS REITKFEVEGNPAFVE VVLRSYDGEDRTVKL RHIFFEGDHLVEVYV EIEPL	165. 0207 326	213	(ATG)AGCGAAGAAGAACAGCGCGAATTTTGGA AACGCTTTTATGAAGCGCTGGATGCGGGCGATG CCGAAACTGCCAGCGCACTGTTTCCAGATGGCA CCGAAATTCATCTGTGGGATGGCAAAACCTTTC ATACCCGTGAAGAATTCGCGCGTGGTTGTGG ATCTGTATTGAAAAGCAAAGATGCGAGCCGCG AAATTACCAAATTTGAAGTGAAGGCCAACCGCG CGTTTGTGAGTTGTGCTGCGCGCGAGCTATG ATGGCGAAGATCGCACCGTGAACTGAGACATA TTTTTGAATTCGAAGGCGATCATCTGGTGAAG TGTATGTGGAATTTGAACCGCTGGG (TAA)
luxs iti_... 0.3... 483	14	MSAEQIREFVARFYAA LDAGDADTASALFPDG TEIHLWDGRTFRTQAE FRAWFETLRAQSADAR RHVVALEVTGNTADVE VVLHASVDGEERTVAL RHRFQFEGDRLVRVTV EITPL	79.9 1015 895	214	(ATG)AGCGCGGAACAAATTCGCGAATTTGTGG CGCGCTTTTATGCGGCCTTGGATGCCGGTGATG CGGATACGGCAAGCGCGTTGTTTCCGGATGGCA CCGAAATTCATCTGTGGGATGGCCGACCTTTC GCACCCAGGCGGAATTCGCGCATGGTTTGA CCCTGCGCGCGCAGAGCGCATGCGCGTAGAC ATGTGGTGGCATTTGAAGTGACCGGCAACACCG CGGATGTGGAAGTTGTGCTGCATGCGAGCGTGG ATGGCGAAGAAGCGCACCGTGCGCGTGAGACATC GCTTTCAGTTTGAAGGCGATCGCCTGGTGGCGG TGACCGTTGAAATTACCCGCTGGG (TAA)
luxs iti_... 0.3... 485	15	MSEEEQKEFVKRFYEA LDAGDAETASALFKDG TEIYLWDGTTFRTRA FRAWFRLLKSTSEEAR RSVVSFKVDGNVADVE VVLRARWKGEERVVVL RHRFVFEGDEVVRVEV EIRPL	75.7 9958 535	215	(ATG)TCGGAAGAAGAAGCAGAAAGAAATTTGTGA AACGCTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACCGCGAGCGCATTTGTTTAAAGATGGCA CCGAAATTTATCTGTGGGATGGTACCACCTTTC GTACCCGCGCGGAATTTGCGCGTGGTTTCCGC GTCTGAAAAGCACCAGCGAAGAAGCCCGCGTA GCGTGGTGAGCTTTAAAGTGGATGGCAATGTGG CGGATGTGGAAGTGGTGGTGGTGGCGGTGGGA AAGGTGAAGAAGCGGTAGTGAAACTGCGCCATC GCTTTGTGTTTGAAGGCGATGAAGTTGTACGCG TTGAAGTGGAAATTCGCCCGCTGGG (TAA)
luxs iti_... 0.3... 494	16	MSAEQREFVVKRFYEA LDAGDAETASALFPDG TKIHLWDGTVFNTQEE FKAWFVKLRSTSENAK REVTHFEVDGDKAVVE VVLHANIKGKEKTVRL RHEFLFEGDKIVEVNV EIEPL	6.14 5818 936	216	(ATG)AGCGCGGAAGAACAGCGCGAATTTGTGA AACGTTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACTGCAAGCGCACTGTTTCCGGATGGCA CCAAATTCATCTGTGGGATGGTACCGTGTGTTA ACACCCAGGAAGAATTTAAAGCGTGGTTTGTGA AACTGCGCAGCACCAGCGAAGAACGGAACGCG AAGTGACCCATTTTGAAGTGGATGGCGATAAAG CGGTGGTGGAGTGGTGGTGGATGCGAACATTA AAGGCAAAGAAAGACCGTGCGCCTGCGCCATG AATTTCTGTTTGAAGGCGACAAAATTTTGAAG TTAATGTGGAATTTGAACCGCTGGG (TAA)
luxs iti_... 0.3... 500	17	MSAEQREFVVKFYDA LDAGDAETASALFPDG TEIYLWDGKVFKTQAE FRAWFVELYSKSDNAQ RSVVRFEVDGNVAYVE VVLRFNDGEEKTVRL RHIFFYFEGDGLVKVEV TIEPL	8.46 0262 612	217	(ATG)AGCGCGGAAGCGCAGCGCGAATTTGTGA AGAAATTTTATGATGCGCTGGATGCGGGCGATG CGGAAACTGCAAGCGCGTTGTTTCCGGATGGAA CCGAAATTTATCTGTGGGATGGCAAGTGTGTTA AAACCCAGGCGGAATTTGCGCGTGGTTTGTGG AACTGTATAGCAAAGCGATAACCGCGAGAGAA GCGTGGTGGCTTTTGAAGTGGATGGTAACGTGG CGTATGTGGAAGTGGTGGTGGCGCGGAACCTTG ATGGCGAAGAAAAGACCGTGCGCCTGCGCCATA TTTTCTACTTTGAAGTGATAGCCTGGTGAAAG TTGAAGTTACCATTTGAACCGCTGGG (TAA)
luxs iti_... 0.3... 506	18	MSEEEIKEFWRRFYQA LDAGDAETASALFPDG TEIYLWDGTTFRTRA FRAWFEALRATSDAAS RHIVKLEVKGNPAEVE VVLRARVDGEERVVRL	49.7 9820 318	218	(ATG)AGCGAAGAAGAAATTAAGAATTTTGGC GCCGCTTTTATCAGGCGCTGGATGCGGGTGATG CCGAAACCGCAAGCGCATTTGTTTCCGGATGGCA CCGAAATTTATCTGTGGGATGGTACCACCTTTC GCACCCGCGCGGAATTTCTGTCATGGTTTGAAG CGCTGCGTGCAACAGCGATGCGCGCTAGAC

		RHIFEFEGDRLKRVEV EIEPL			ATATTGTGAAACTGGAAGTGAAAGGCAACCGCG CCGAAGTGGAGTTGTGCTGCGCGGAGAGTTG ATGGTGAAGAACGCGTTGTGCGCCTGCGCCATA TTTTTGAATTTGAAGGCGATCGTCTGAAACGCG TCGAAGTTGAAATTGAACCGCTGGG (TAA)
luxs iti_0.3_515	19	MSEEEIREFWRRFYEA LDAGDAETASALFPDG TKIYLWDGIVFSTKEE FRSWFVELKSKSDNAK RHVVS LRVDGNVANVK VVLEAEINGEKKVVEL THTTKFEGDKLVEVNV DINPL	2.34 8306 842	219	(ATG)AGCGAAGAAGAAATTCGTGAATTTTGGC GCCGCTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACCGCAAGCGCACTGTTTCCAGATGGTA CCAAAATTTATCTGTGGGATGGCATTGTGTTTA GCACCAAAGAAGAGTTTCGCGAGCTGGTTTGTGG AACTGAAAAGCAAAAGCGATAATGCGAAACGCC ATGTGGTGAGCCTGCGCGTGGATGGTAACGTGG CCAACGTGAAAGTGGTGTGGAAGCGGAAATCA ACGGCGAAAAGAAAGTTGTGGAGCTGACCCATA CCACCAAATTTGAAGGCGATAAACTGGTGGAG TGAACGTGGATATTAACCGCTGGG (TAA)
luxs iti_0.3_517	20	MSEEEQRKFWEKFYEA LDAGDAETASALFKDG TKIYLWDGTVFETQEE FPAWFEKLYSTSDNAQ RHIVSFKVDGNISYVE VVLHANVNGEKT VKL THLKFEGDHHVVEVKV KIEPL	26.0 5459 572	220	(ATG)AGCGAAGAAGAACAGCGCAAAATTTGGG AAAAATTTTATGAAGCGCTGGATGCGGGCGATG CCGAAACCGCAAGCGCATTGTTTAAAGATGGCA CCAAAATTTATCTGTGGGATGGTACCGTTTTTG AAACCCAGGAAGAATTCGCGCGTGGTTTGA AACTGTATAGCACCAGCGATAACGCGCAGCGCC ATATTGTGAGCTTTAAAGTGGATGGCAACATTA GCTATGTGGAAGTGGTGTGATGCGAACGTGA ACGGTGAAGAAAAGACCGTGAACACTGACCCACC TGTTCAAATTTGAAGGCGATCATGTGTTGAGG TGAAAGTGAAAATTTGAACCGCTGGG (TAA)
luxs iti_0.3_519	21	MSEEEQKEFWKIFYDA LDAGDAETASALFPDG TKIHLWDGKTFTTQAE FKAWFVLDKSKSDDAK RYITKFKVDGNKAYIE VVLKASVKGKKTKVKL KHTAQFEGDKLVEVDV TINPL	2.20 4561 161	221	(ATG)AGCGAAGAAGAACAGAAAGAAATTTGGA AAAAGTTTTACGATGCGCTGGATGCGGGCGATG CAGAACTGCGAGCGCACTGTTTCCGATGGCA CCAAAATTCATCTGTGGGATGGTAAACCTTTA CCACCCAGGCGAATTTAAAGCGTGGTTTGTGG ATCTGAAAAGCAAAAGCGATGATGCGAAACGCT ATATTACCAAATTCAAAGTGGATGGAACAAAG CGTATATTGAAGTGGTGTGTAAGCGAGCGTGA AAGGCAAGAAAAGACCGTGAACACTGAAACATA CCGCGCAGTTTGAAGGCGATAAACTGGTGGAG TGGACGTGACCATTAAACCGCTGGG (TAA)
luxs iti_0.3_521	22	MSEEAIREFWRRFYEA LDAGDAETASALFPDG TRIHLDGTTFRTRAE FRAWFVELHKSNDNAQ REVVRLEVEGNPARVE VVLKANYKGKTKVRL RHEFLFEGDALVEVRV EIEPL	1647 .027 643	222	(ATG)AGCGAAGAAGCGATTCGCGAATTTGTGC GCCGCTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACTGCGAGCGCACTGTTTCCGGATGGCA CCCGCATTATCTGTGGGATGGTACCACCTTTC GTACCCGTGCGGAATTCGCGCGTGGTTTGTGG AACTGCATAGCAAAAGCGATAACGCGCAACGCG AAGTGGTGCGCTGGAAGTGAAGGCAACCGTG CAAGAGTTGAAGTTGTGCTGAAAGCGAATA AAGGCAAGAAAAGACCGTGCCTCTGCGCCATG AATTTCTGTTTGAAGGCGACGCGTGGTGAAG TGCGCGTTGAAATTGAACCGCTGGG (TAA)
luxs iti_0.3_541	23	MSAEAIREFWRRFYEA LDAGDAETASALFPDG TEIHLWDGTTFRTREE FRAWFVELYSTSDNAK RKIVSLRVDGDRALVE VVLRAVVGEEKTVKL KHEALFEGDRLVEVRV QIEPL	2.97 6503 11	223	(ATG)AGCGCGGAAGCGATCCGCGAATTTTGGC GCCGCTTTTATGAAGCGCTGGATGCGAGCGATG CGGAAACCGCAAGCGCTCTGTTTCCGGATGGCA CCGAAATTCATCTGTGGGATGGTACCACCTTTC GTACGCGCAAGAAATTCGCGCGTGGTTTGTGG AACTGTATAGCACCAGCGATAACGCGAAACGCA AAATTGTGAGCCTGCGCGTTGATGGCGATCGCG CGCTGGTTGAAGTGGTGTGAGAGCAAGCGTGG GTGGTGAAGAAAAGACCGTGAACACTGAAACATG AAGCCCTGTTTGAAGGAGATCGCCTGGTGAAG TGCGCGTGCAGATTGAACCGCTGGG (TAA)

luxs iti_... 0.3_... 543	24	MSEEAQRKFVERFYKA LDAGDADTASALFPDG TKIYLWDGKVFETTQEE FRAWFVELYSKSENAK REVVSFNVGDGVAEVE VVLHANVRGELKTVKL KHTFKFEGDKLVEVNV TIKPL	90.7 7332 412	224	(ATG)AGCGAAGAAGCGCAGCGTAAATTTGTGG AACGCTTTTACAAAGCGCTGGATGCGGGTGATG CGGATACCGCAAGCGCGCTGTTTCCTGATGGCA CCAAAATTTATCTGTGGGATGGCAAAGTGTTTA CCACCCAGGAAGAATTCGCGCGTGGTTTGTG AACTGTATAGCAAAAGCGAAAACGCGAAACGTG AAGTGGTGTCTTTAACGTGGATGGCGATGTGG CGGAAGTGGAGTTGTGCTGCATGCGAACGTGC GCGGCGAACTGAAAACCGTGAAATTGAAACATA CCTTTAAATTCGAAGGCGATAAACTGGTTGAAG TGAACGTGACCATTAAACCGCTGGG (TAA)
luxs iti_... 0.3_... 564	25	MSAEQREFWRPFYAA LDAGDAETASALFPDG TKIHLWDGKTFTTAE PREWFEKLYSTSDNAK REITELKVDGDKSYVK VVLKANINGEEKTVNL THEFQFEGDKLVEVNV TINPL	398. 2017 968	225	(ATG)AGCGCGGAAGCGCAGCGCAATTTTGGC GCCGCTTTTATGCGGCGTGGATGCGGGCGATG CGGAAACTGCAAGCGCGCTGTTTCCAGATGGCA CCAAAATTCATCTGTGGGATGGCAAACCTTTA CCACCCGCGCGCAATTCGCGAATGGTTTGAAA AACTGTATAGCACCTCTGATAACGCGAAACGTG AAATTACCGAACTGAAAGTGGATGGCGATAAAA GCTATGTGAAAGTTGTGCTGAAAGCGAACATTA ACGGCGAAGAAAAGACCGTGAAACCTGACCCATG AATTTTCAGTTTGAAGGCGACAAACTGGTGAAG TGAACGTGACCATTAAACCGCTGGG (TAA)
luxs iti_... 0.3_... 568	26	MSEQAIREFWKRFYNA LDAGDADTASALFPDG TVIHLWDGKTFHTQAE ERKWFVDLYSRSDNAK REIVSLEVNDGNHAKIV VVLNALINGEKKTVLL THETLFEGDHLVEVNV EIKPL	1.36 0055 287	226	(ATG)AGCGAACAGGCGATTGCGCAATTTTGGG AACGCTTTTATAACGCGCTGGATGCTGGTGATG CGGATACCGCGAGCGCGCTGTTTCCTGATGGTA CCGTGATTCACTGTGGGATGGCAAACCTTTC ATACCCAGGCGGAATTCGCAAAATGGTTTGTGG ATCTGTATTCCCGCAGCGATACGCCAAACGCG AAATTGTGAGCCTGGAAGTGGATGGTAACCATG CGAAAATTGTGTGGTGGTGAATGCCATTATTA ACGGCGAAAAGAAAACCGTGCTGCTGACCCATG AAACCGTGTTTGAAGGCGATCATCTGGTGAAG TTTTTGTGAAATTAACCGCTGGG (TAA)
luxs iti_... 0.3_... 582	27	MSEEQREFFWKRFYEA LDAGDAETASALFPDG TKIHLWDGTTFTTQAE FRAWFVRLHSTSENAS RSIVSFKVEGNKALVE VVLRAVKGEEKTVKL THEFQFEGDKLVEVNV DIKPL	146. 1630 961	227	(ATG)AGCGAAGAAGAACAGCGCAATTTTGGG AACGCTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACTGCAAGCGCACTGTTTCCGGATGGCA CCAAAATTCATCTGTGGGATGGTACCACCTTTA CCACCCAGGCGGAATTCGCGCGTGGTTTGTGC GCCTGCATAGCACCAGCGAAACGCGACCGGTA GTATTGTGAGCTTTAAAGTGAAGGCAACAAAG CGTTGGTGAAGTGGTGTGCTGCGCGGAGCGTGA AAGGTGAAGAAAAGACCGTGAAACTGACCCATG AATTTTCAGTTTGAAGGCGATAAACTGGTTGAAG TGAACGTGGATATTAAACCGCTGGG (TAA)
luxs iti_... 0.3_... 584	28	MSEEDIRRFVERFYAA LDAGDAETASALFPDG TRIHLWDGTTFTTREE FRAWFEKLHATSENAK RHVVKLKVEGNKAYVE VILHASFKGKEKTVKL THVFLFEDDRLVEVNV KIEPL	2.87 4222 529	228	(ATG)AGCGAAGAAGATATTGCGCGCTTTGTGG AACGCTTTTATGCAAGCGCTGGATGCGGGCGATG CGGAAACTGCGAGCGCATGTTTCCGGATGGCA CCCGCATTCACTGTGGGATGGAACACCTTTA CCACCCGTGAAGAATTCGCGCGTGGTTTGAAA AACTGCACGCGACAGCGAAAACGCGCGCGCC ATGTGGTGAAGTGAAGTGAAGGTAAACAAAG CGTATGTGGAAGTGATTCTGCATGCCAGCTTTA AAGGCAAAGAAAAGACCGTTAAACTGACCCATG TGTTTCTGTTTGAAGATGATCGCCTGGTTGAAG TGATGTGAAAATTGAACCGCTGGG (TAA)
luxs iti_... 0.3_... 596	29	MSEEQREFWEKIFYNA LDAGDAETASSLFPDG TEIYLWDGTVFKTREE FRAWFEKLHSTSENAK RHIVSFEVDGNKSKVK VVLHAKINGEETVEL	1756 .351 071	229	(ATG)AGCGAAGAAGAAGCGCGCAATTTTGGG AAAAATTTTATAACGCGCTGGATGCGGGTGATG CCGAAACCGCAAGTAGCCTGTTTCCGGATGGCA CCGAAATTTATCTGTGGGATGGTACCGTGTTTA AAACCGTGGAAGAATTCGCGCGTGGTTTGAAA AACTGCATAGCACCAGCGAAAACGCGAAACGCC

		EHVYEFEGDKLKKVEV EIKPL			ATATTGTGAGCTTTGAAGTGGACGGCAACAAAA GCAAAGTGAAAGTGGTGCATGCGAAAATTA ACGGAGAAGAAGTGACCGTGGAACTGGAAACATG TGTATGAATTTGAAGGCGATAAACTGAAAAAGG TGGAAAGTTGAAATTAAACCGCTGGG (TAA)
luxs iti_ 0.4_ 600	30	MSEIAQREFWRRFYAA LDAGDAETASALFPDG TEIYLWDGRTFHTREE FEAWFRELYSKSENAR REITSFTVTGDIAEIR VVLRAIDGKRTVSL NHVTHFEGDQLVRVEV SIFPL	2.52 3151 348	230	(ATG)AGCGAAATTGCGCAGCGCGAATTTTGGC GCCGCTTTTATGCGGCGCTGGATGCGGGTGATG CGGAAACTGCAAGCGCGCTGTTTCCAGATGGCA CGGAAATTTATCTGTGGGATGGCCGCACCTTTC ATACCCGCGAAGAATTTGAAGCGTGGTTTCGCG AACTGTATAGCAAAAGCGAAAACGCGCTCGCG AAATCACCTCTTTTACCGTGACCGGCGATATTG CCGAAATTCGCGTGGTGTGCGCGCGACCATG ATGGCGAAAAACGACCGTGAGCCTGAACCATG TGACCCATTTTGAAGGCGATCAGCTGGTGCAGC TGGAAAGTGAGCATTTTCCGCTGGG (TAA)
luxs iti_ 0.4_ 601	31	MSEEEQKEFWKRFYEA LDSGDADTASALFPDG TKIYLWDGTVFETKAQ FKAWFVELYSRSDNAQ RSITKFEVDGNTSRVV VVLKATVRGKEKVVEL EHVFEFEGDELKEVKV EIHPL	1.47 2011 057	231	(ATG)AGCGAAGAAGAACAGAAAGAAATTTTGGG AACGCTTTTATGAAGCGCTGGATAGCGCGATG CGGATACAGCGTCCGCTCTGTTTCCGGATGGCA CCAAATTTATCTGTGGGACCGCACCGTGTGTTG AAACCAAAGCGCAGTTTAAAGCGTGGTTTGTGG AACTGTATAGCGCGACGGATACCGCGCAGCGCA GCATTACCAAATTTGAAGTGGATGGTAACACCA GCCGCGTGGTGGTGTGCTGAAAGCGACCGTGC GCGGCAAGAAAAAGTGGTTGAACTGGAACATG TGTTTCAATTCGAAGGCGATGAAGTGAAGAAG TGAAAGTTGAAATTCATCCGCTGGG (TAA)
luxs iti_ 0.4_ 605	32	MSEEAIREFVKRFYDA LDAGDADTASALFPDG TLIHLWDGKTFRTQAE FRAWFEELYSKSENAR RHVVKLQVDGDRAQVE VVLHANIDGQEVVRL RHEFLFEGDRLVEVYV QIQPL	409. 7712 509	232	(ATG)AGCGAAGAAGCGATTGCGGAATTTGTGA AACGCTTTTATGATGCCCTGGATGCGGGCGATG CGGATACCGCCTCTGCTTGTGTTCCAGATGGCA CCCTGATTCACTCTGTGGGATGGCAAAACCTTTC GTACCCAGGCGGAATTTCCGCGCTGGTTTGAAG AACTGTATAGCAAAAGCGAAAACGCGAAACGCC ATGTGGTGAACTGCAGGTGGATGGCGATCGCG CGCAGGTGGAAGTGGTGTGCTGCAAGCAACATG ATGGCCAGGAAGTTGTGGTGCCTGCGCCATG AATTTCTGTTTGAAGGCGATAGACTGGTGAAG TGTATGTGCAGATTGACCCGCTGGG (TAA)
luxs iti_ 0.4_ 606	33	MTAEAQRAFWDRFYRA LDAGDAETASSLFPDG TEIHLWDGKTFHTREE FRQWFENLRKSSENAR RHIVDFRVDGDPADV VVL RANYQGEERVVQL HHEFLFEGDQLVRVEV RIIPE	293. 3476 158	233	(ATG)ACCGCGAAGCGCAGCGCGCTTTTGGG ATCGCTTTTATGCGCGCTTGATGCGGGCGATG CGGAAACCGCAAGCAGCCTGTTTCCAGATGGCA CCGAAATTCATCTGTGGGATGGTAAACCTTTC ATACCCGCGAAGAATTTCCGCGAGTGGTTGAAA ACCTGCGCAGCAAAAGCGAAAACGCGCGCCGCC ATATTGTGGATTTTCCGCTGGATGGCGATCGCG CAGATGTGCGCGTGGTGTGCGTGCAAACTATC AGGGTGAAGAACGCGTTGTGCGAGCTGCATCATG AATTTCTGTTTGAAGGCGATCAGCTGGTGCCTG TGGAAAGTGCGCATTATTCGGAAGG (TAA)
luxs iti_ 0.4_ 610	34	MSEKEQREFCARFYAA LDAGDAETASALFPDG TRIHLWDGRTFTTQAE FGAWFRRLRSTSDNAR RHIVDFRVDGDVAKVR VVLHASVGGREPTVQL EHVFI FEGDRLVEVHV SIQPE	2.78 7836 904	234	(ATG)AGCGAAAAAGAAGCGCGAATTTTGGC CGCGCTTTTATGCGGCCCTGGATGCGGGTGATG CGGAAACTGCAAGCGCGTGTGTTTCCGGATGGCA CCCGCATTCATCTGTGGGATGGCCGCACCTTTA CCACCCAGGCGGAATTTGGCGCGTGGTTTCGTC GTCTGCGTAGCACAAAGCGATAACGCGCGCGTC ATATTGTGGATTTTCCGCTGGATGGCGATGTGG CGAAAGTGCGCGTAGTGTGATGCGAGCGTGG GTGGTAGAGAACGACCGTGACGTGGAACATG TGTTTATTTTGAAGGCGATCGCCTGGTGAAG TGCATGTGAGCATTGACCGGAAGG (TAA)

luxs iti_... 0.4_... 611	35	MSEEEQRKFWKIFYNA LDAGDAETASALFPDG TEIYLWDGKVFTRAE FREWFERLYSKSDNAQ RSITSFEVNGNVAEIK VILKANVNGEEKEVSL NKHTEFEGDKLVKVEV DISPL	5.10 6427 091	235	(ATG)AGCGAAGAAGAACAGCGCAAATTTTGGG AAAAATTTTATAACGCGCTTGATGCGGGCGATG CGGAAACCGCAAGCGCACTGTTTCCGGATGGTA CCGAAATTTATCTGTGGGATGGCAAAGTGTTTC GCACCCGCGCGGAATTCGCGAATGGTTTGAAC GCCTGTATTGAAAAGCGATAACGCCAGCGCA GCATTACCAGCTTTGAAGTGAACGGCAACGTGG CGGAAATTAAGTGATTCTGAAAGCGAACGTTA ACGGTGAAGAAAAAGAAGTGAGCCTGAACCATA AAACCGAATTTGAAGGCGATAAACTGGTGAAAG TGGAAGTGGATATTAGCCCGCTGGG (TAA)
luxs iti_... 0.4_... 613	36	MSEKEQREFWKKFYEA LDAGDAETAAALFPDG TVIHLWDGKTFHTQEE FKWFIKLRSTSENAK RKIISFKVDGNISYVE VELKAKIKGESKVVKL RHIKFEFGDKLVEVDV TIKPI	6.80 3040 774	236	(ATG)AGCGAGAAAGAACAGCGCGAATTTTGGG AAAAGTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACCGCAGCAGCCTTGTTTCCAGATGGTA CCGTGATTCACTGTGGGATGGCAAACCTTTC ATACCCAGGAAGAATTTAAAGAATGGTTTATTA AACTGCGCAGCACCTCTGAAAACGCGAAACGCA AAATTATTAGTTTTAAAGTGGATGGTAACATTA GCTATGTTGAAGTGGAACTGAAAGCGAAAATTA AAGGTGAAAGCAAAGTGGTGAAACTGAGACATA TTTTTAAATTTGAAGGTGATAAACTGGTGGAAG TCGATGTGACCATTAAACCGATTGG (TAA)
luxs iti_... 0.4_... 614	37	MSAEEIKEFCRRFYEA LDAGDAETASALFPDG TIIHLWDGRTFHTQAE FRAWFRELSTSEDAK REIERLEVNDGNQAKVE VILHANRDGEKKVVRL THEFLFEGDRLVEVTV AIRPL	68.1 7069 8	237	(ATG)AGCGCGGAAGAATTAAGAATTTTGGC GCCGCTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACCGCAAGCGCGTTGTTTCCCTGATGGCA CCATTATTCACTGTGGGATGGCCGTACCTTTC ATACCCAGGCGGAATTTGCGCGGTGGTTTCGCG AACTGCGCAGCACCAGCGAAGATGCGAAACGCG AAATTGAACGCTGGAAGTGGATGGCAACACAGG CCAAAGTGGAAGTTATCTGCGATGCGAACCGCG ATGGCGAAAAGAAAGTGGTGCGCCTGACCCATG AATTTCTGTTTGAAGGCGATCGCCTGGTTGAAG TGACCGTGGCGATCCGCCGCTGGG (TAA)
luxs iti_... 0.4_... 639	38	MSKQEIKDFCEKIFYNA LDAGDADTASALFKDG TKIYLWDGKTFETQAE FKAWFTKLHSTSDNAK RYITSLEVNDGDKAFVK VVLKANVNGEEKVVEL THIFEFEGDELVKVQV KIKPL	1.21 6309 606	238	(ATG)AGCAAACAGGAAATTAAGATTTTGGC AAAAATTTTATAACGCGCTTGATGCGGGCGATG CGGATACCGCGAGCGCACTGTTCAAAGATGGCA CCAAATTTATCTGTGGGATGGCAAACCTTTG AAACCCAGGCGGAATTTAAAGCGTGGTTTACCA AACTGCATAGCACCAGCGATTAATGCGAAACGCT ATATTACCAGCCTGGAAGTGGATGGCGATAAAG CCTTTGTGAAAGTGGTGCTGAAAGCCAACGTGA ATGGTGAAGAAAAGACCGTGGAACTGACTCATA TTTTTGAATTTGAAGGCGATGAACTGGTTAAAG TGCAGGTTAAATTAAGCCGCTGGG (TAA)
luxs iti_... 0.4_... 670	39	MSEEEIRQFIKRFYDA LDAGDAETASALFPDG TEIDLWDGTTFHTQEE FRSWFVRLKSTSDNAK RHVVALEVQGNRAKVE VVLEASIDGKKITVQL THEFYFEGDKLVKVV TIKPL	673. 3462 336	239	(ATG)AGCGAAGAAGAATTCGCCAGTTTATTA AACGCTTTTATGATGCGCTGGATGCGGGCGATG CCGAAACTGCGAGCGCATTTGTTCCGGATGGCA CCGAAATTGATCTGTGGGATGGTACCACCTTTC ATACCCAGGAAGAATTTGCGCTGCTGTTTGTGC GTCTGAAAAGCACCAGCGATAACCGGAAACGCC ATGTGGTGGCGTTAGAAGTGCAGGGCAACCGTG CGAAAGTGAAGTGGTGCTGGAAGCCAGCATTG ATGGCAAGAAAATTACCGTGCACTGACCCATG AATTTTATTTTGAAGGCGATAAACTGGTGAAAG TGAACGTGACCATTAAACCGCTGGG (TAA)
luxs iti_... 0.4_... 679	40	MSEEEIREFVRRFYAA LDAGDAETASALFPDG TVIHLWDGRTFHTQAE FRAWFEELYSRENAK REVTQLEVNDGDHAFVK VVLRAHNGEEKVVGL	625. 7152 73	240	(ATG)AGCGAAGAAGAATTCGCCGAATTTGTGC GCCGCTTTTATGCGGGCGCTGGATGCGGGTGAATG CGGAAACTGCAAGCGCGCTGTTTCCAGATGGCA CCGTGATTCACTGTGGGATGGCCGCACTTTC ATACCCAGGCGGAATTTCTGCGCTGGTTTGAAG AACTGTATAGCCGAGCGAAAACGCGAAACGTG

		EHYFHFEGDQLKEVEV VIRPE			AAGTGACCCAGCTGGAAGTGGATGGCGATCATG CGTTTTGTGAAAGTGGTGTGCGCGCGAACATTC ATGGTGAAGAAAAAGTTGTGGGCGCTGGAACATT ATTTTCATTTTGAAGGTGATCAGCTGAAAGAAG TTGAAGTGGTTATTCTCGCGGAAGG (TAA)
luxs iti_0.4_687	41	MSKEEIEEFWKIFYQA LDAGDAETASALFPDG TVIHLWDGKVFHTKAE FREWFVKLKSESEDAK REIVSLRVDGNKAEVE VVLKAKYKGEEKEVRL KHESLFEGDRLVEVDV DIFPL	2.56 8762 958	241	(ATG)AGCAAAGAAGAAATTGAAGAATTTTGGG AAAAGTTTTATCAGGCGCTGGATGCGGGCGATG CGGAAACCGCAAGCGCCTTGTTCCTGATGGCA CCGTGATTCATCTGTGGGATGGCAAAGTGTTC ATACCAAAGCGGAATTTCCGGAATGGTTTGTGA AACTGAAAAGCGAGAGCGAAGATGCGAAACGCG AAATTGTGAGCTGCGCGTGGATGGTAACAAAG CCGAGGTGGAAGTGGTGTGAAAGCGAAATATA AAGGCGAAGAAAAAGAGTGGCGCTGAAACATG AATCCCTGTTTGAAGGCGATCGCCTGGTCGAAG TGGATGTGGATATTTTCCGCTGGG (TAA)
luxs iti_0.4_695	42	MSEEEIREFWKIFYEA LDAGDAETASALFKDG TKIYLWDGITFFTRDE FRAWFEKLYSTSEDAK REIVKFNVDGNKAKVE VVLHAIVDGGQKKTVKL THTSYFEGDELKKVEV SIEPM	1.76 8486 524	242	(ATG)AGCGAAGAAGAAATCCGCGAATTTTGGG AACGCTTTTATGAAGCGCTGGATGCGGGCGATG CCGAAACCGCGAGCGCACTGTTTAAAGATGGCA CCAAAATTTATCTGTGGGATGGCATTACCTTCT TTACCCGCGATGAATTTCCGCGGTGGTTTGA AACTGTATAGCACCTCAGAAGATGCGAAACGCG AAATTGTGAAATTTAACGTGGATGGTAACAAAG CGAAAGTGGAAAGTGGTGTGCGATGCGATTGTTG ATGGCCAAAAGAAAACCGTGAACCTGACCCATA CCAGCTATTTTGAAGGTGATGAACGAAAAGG TCGAAGTGAGCATTGAACCGATGGG (TAA)
luxs iti_0.4_707	43	MSEEEIREFIKIFYEA LDAGDAETASALFPDG TKIYLWDGTVFSTQAE FRDWFVRLRSTSQDAR RKVVBLKVDGVDVADVE VVLRAVVEGEERVVRL RHVFHFEGDKVVEVEV EIEPE	89.9 0393 918	243	(ATG)AGCGAAGAAGAAATCCGCGAATTTATTA AACGCTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACCGCAAGCGCATTTGTTCCGGATGGCA CCAAAATTTATCTGTGGGATGGTACCGTGTTC CGACCCAGGCGGAATTTCCGCGATTGTTTGTGC GTCTGCGTAGCACCAGCCAGGATGCCCGCCGTA AAGTGGTTGATCTGAAAGTGGATGGTGTGTTGG CGGATGTGGAAGTGGTGTGCGCGCGAACGTGG AAGGTGAAGAACGCGTGGTGCAGCTGAGACATG TGTTTCATTTTGAAGGCGATAAAGTTGTTGAAG TCGAAGTGGAAATTGAACCGGAAGG (TAA)
luxs iti_0.4_712	44	MSEEEQREFWRRFYAA LDAGDAETASALFPDG TKIYLWDGKTFTTQAE FRAWFRKLRSQSEGAS RSVEYFEVDGNKSHVE VVLRASVNGEQHVVKL RHEFLFEGDKLVEVEV KIEPL	5.81 7553 559	244	(ATG)AGCGAAGAAGAACAGCGCGAATTTTGGC GCCGCTTTTATGCGGCGCTGGATGCGGGTGATG CGGAAACCGCATCAGCGCTGTTTCTCTGATGGTA CCAAAATTTATCTGTGGGATGGCAAAACCTTTA CCACCCAGGCGGAATTTCCGCGGTGGTTTCGCA AACTGCGCAGCCAAAGCGAAGGCGCTCTAGAA GCCGTGGAATATTTTGAAGTGGATGGTAACAAAA GCCATGTGGAAGTGGTGTGCGCGCCAGCGTGA ACGGCGAACAGCATGTGGTGAAACTGAGACATG AATTTCTGTTTGAAGGCGATAAACTGGTTGAAG TCGAAGTGAAATTTGAACCGCTGGG (TAA)
luxs iti_0.4_718	45	MSEKAQREFWKIFYEA LDAGDADTASALFKDG TEIYLWDGKVFKKQEE FKEWFEKLHSTSENAK RKIVKFEVKGNSVYVE VVLEASLAGKKTVKL KHMFFEGDKLVKVTV DIYPL	7.18 8666 206	245	(ATG)AGCGAAAAAGCGCAGCGCAATTTTGGG AAAAGTTTTATGAAGCGCTGGATGCGGGCGATG CCGATACCGCGAGCGCGCTGTTTAAAGATGGCA CCGAAATTTATCTGTGGGATGGCAAAGTGTTC AAAAGCAGGAAGAAATTCAAAGAATGGTTTGA AACTGCATAGCACCGAGAGAACGCGAAACGTA AAATTGTGAAATTTGAAGTGAAGGCAACGTGA GCTATGTGGAAGTGGTGTGGAAGCGAGCCTGG CGGGTGAAAAAGAAACCGTGAACCTGAAACACA TGTTTCAGTTTGAAGGCGATAAACTGGTGAAAG TGACCGTGGATATTTATCCGCTGGG (TAA)

luxs iti_... 0.4_... 721	46	MSEKAQREFVKKFYDA LDAGDADTASALFPDG TIIYLWDGKTFRTQTE FRWFVLDLKSTSENAK RKVTDFRVDGNVANVT VVLEANINGKKETVKL NHI FHWEGDRLVEVEV TIEPL	3.71 8728 404	246	(ATG)AGCGAGAAAGCGCAGCGGAATTTGTGA AACGCTTTTATGATGCGTGGATGCAGGCGATG CGGATACCGCGAGCGCACTGTTCCGGATGGCA CCATTATTTATCTGTGGGATGGTAAACCTTTC GCACCCAGACCGAATTTCGCCGCTGGTTGTGG ATCTGAAAAGCACCAGCGAAAACGCGAAACGCA AAGTTACCGATTTTCGCCGTGGATGGAACGTTG CGAACGTGACCGTGGTTCTGGAAGCGAACATTA ACGGCAAGAAAGAAACCGTGAACTGAACCATA TTTTTCATTGGGAAGGCGATCGCCTGGTGAAG TTGAAGTGACCATTGAACCGCTGGG (TAA)
luxs iti_... 0.4_... 724	47	MSAEAREFVRDFYEA LDAGDADTASALFPDG TEIHLWDGTVFRTAE PKAWFTKLYSTSEDAR RKVVRELEVGDVARVE VVLHASHAGKESTVKL QHEFSFEGDRLVEVRV VIEPL	201. 8825 155	247	(ATG)TCTGCGGAAGCGATTGCGGAATTTGTGC GCGATTTTATGAAGCGCTGGATGCGGGTGATG CGGATACCGCGAGCGCGTTATTTCGGATGGCA CGGAAATTCACCTGTGGGATGGTACCGTGTTC GTACCCGCGCGGAATTAAGCGTGGTTACCA AACTGTATAGCACCAGCGAAGATGCGCGTCGCA AAGTGGTGCGTCTGGAAGTGGAAGCGGATGTGG CGCGTGTGAAGTGGTTCTGCATGCGAGCCATG CGGGCAAAGAAAGCACCGTGAAACTGCAGCATG AATTTAGCTTTGAAGGTGATCGCCTGGTGAAG TTCGTGTGGTGATTGAACCGCTGGG (TAA)
luxs iti_... 0.4_... 747	48	MSEEEIKKFCEKFYEA LDAGDAETASALFPDG TEIYLWDGRVFKTREE FRAWFVELYSKSDNAQ RKIVDLKVDGNKAKVK VVLHANHNGEQHKVAL THEFLFEGDKLVEVNV EIKPL	21.9 3365 584	248	(ATG)TCTGAAGAAGAAATCAAGAAATTTGTG AAAAATTTTATGAAGCGCTGGATGCGGGCGATG CCGAAACTGCGAGCGCGTTGTTTCTCTGATGGCA CCGAAATTTATCTGTGGGATGGCGCGGTGTTA AAACCCGCGAAGAATTTGCGCGGTGGTTGTGG AACTGTATAGCAAAAGCGATATGCGCAGCGCA AAATTGTGGATCTGAAAGTGGATGGTAACAAAG CGAAAGTGAAAGTTGTGCTGCATGCGAACCATA ACGGCGAACAGCATAAAGTGGCGCTGACCCATG AATTTCTGTTTGAAGCGGATAAACTGGTGAAG TGAACGTGGAAATTAACCGCTGGG (TAA)
luxs iti_... 0.4_... 748	49	MSEEAAREFWKKFYEA LDAGDAETASALFPDG TIIHLWDGTTFTTQDE FKKWFVELFSKSENAS RKIVKLEVKNVAYVE VVLKAKIDGKEKTVRL KHVTKFEGDKLVEVEV DIDPL	1.24 9481 686	249	(ATG)AGCGAAGAGCGATTGCGGAATTTTGA AAAAGTTCTATGAAGCACTGGATGCGGGCGATG CGGAAACCGCAAGCGCACTGTTTCCGGATGGCA CCATTATTCATCTGTGGGATGGTACCACCTTTA CCACCCAGGATGAATTTAAAAAGTGGTTGTGG AACTGTTTAGTAAAGCGAAGCGAGCCGCA AAATTGTGAACTGGAAGTGAAGGCAACGTGG CGTATGTGGAAGTTGTGCTGAAAGCCAAAATCG ATGGCAAAGAAAGACCGTGCGCCTGAAACATG TGACCAAATTTGAAGCGGATAAACTGGTTGAAG TCGAAGTGGATATTGATCCGCTGGG (TAA)
luxs iti_... 0.4_... 750	50	MSADDQKEFWKRFYEA LDAGNADEASALFPDG TEIYLWDGTTFTKTAQ FKAWFCQLRSTSENAK REIVKFEVDGNFSYVE VVLEATVNGKKFIVSL DHYFEFEQQLVEVYV KIRPL	1.43 6074 637	250	(ATG)AGCGCGGATGATCAGAAAGAAATTTGGA AACGTTTTTACGAAGCGCTGGATGCGGGCAATG CAGATGAAGCGAGCGCGCTGTTTCCGGATGGCA CCGAAATCTATCTGTGGGATGGTACCACCTTTA AAACCAAAGCGCAGTTTAAAGCGTGGTTCTGCC AGCTGCGCAGCACCAGCGAAAACGCGAAACGTG AAATTGTTAAATTTGAAGTGGATGGGAACTTTA GCTATGTGGAAGTGGTGTGGAAGCGACCGTGA ACGGCAAGAAATTTATGTGAGCCTGGATCAT ATTTTGAATTTGAGGGCGAACAGCTGGTTGAAG TCTATGTGAAAATTCGCCGCTGGG (TAA)
luxs iti_... 0.4_... 766	51	MSAKAQREFWKKFYEA LDAGDADTAAALFPDG TRIHLWDGRTFTTQAE FRAWFTLHRSRDNK RSIVAFNVSGDVSDV VVLRSYEGEERTVRL	2.56 6689 703	251	(ATG)TCTGCGAAAGCGCAGCGTGAATTTGGA AAAAGTTTTATGAAGCGCTGGATGCGGGTGATG CGGATACCGCAGCAGCACTGTTTCCGGATGGCA CCCGCATTCATCTGTGGGATGGCGCTACCTTTA CCACCCAGGCGGAATTTGCGCGCTGGTTTCAGA CCCTGCATAGCCGAGCGATAACGCGAAACGCT

		RHTFRFEGDHLVEVDV AIDPL			CTATCGTGGCGTTTAAACGTGAGCGGTGATGTGA GCGATGTGGTGGTTGTGCTGCCGCGGAGCTATG AAGGCGAAGAACGCACCGTGGCTCTGCGTCATA CCTTTTCGCTTTGAAGGTGATCATCTGGTGGAAAG TTGATGTGGCGATTGATCCGCTGGG (TAA)
luxs iti_ 0.4_ 770	52	MSEAEQREFWRRFYEA LDAGDAETASALFPDG TEIHLWDGKTFHTQAE FRAWFVELRSTSEAR RHVVAFNVGDGRARVE VVLHAVRDGEARTVQL THETVFAGDQVVRVRV QIKPL	1.00 7601 935	252	(ATG)AGCGAAGCGGAACAGCGCAATTTTGGG GCCGCTTTTATGAAGCGCTGGATGCCGGCGATG CGGAAACTGCAAGCGCGTTGTTTCCAGATGGCA CCGAAATTCATCTGTGGGATGGCAAAACCTTTC ATACCCAGGCGGAATTTCCGCGCTGGTTTGTGG AATTACGCAGCACCTCTGAAGCGCGCGCCGTC ATGTGGTTGCGTTTAAACGTGGATGGCGATCGCG CTAGAGTGGAASTCGTGCTGCATGCCGTGCGTG ATGGTGAAGCGAGAACCCTGCAGCTGACCCATG AAACCGTGTTCGCGGTGATCAGGTGGTGGCGG TGCGTGTGCAGATTAAACCGCTGGG (TAA)
luxs iti_ 0.4_ 812	53	MSEEAVEREFVARFYEA LDAGDADTASALFPDG TVIHLWSGQTFHTRAE FRAWFERLYAESAAAR RRVEMRVDGDVAFVE VVLHASHQGGQPVVRL RHI FRFEGDRLREVEV SIDPL	1.84 3814 789	253	(ATG)AGCGAAGAACCGGTGCCGCAATTTGTTC CGCGCTTTTATGAAGCGCTGGATGCCGGTGATG CGGATACCGCGAGCGCATTGTTTCCGGATGGCA CCGTGATTTCATCTGTGGAGCGGCCAGACCTTTC ATACCCGCGCGGAATTTCCGCGCTGGTTTGAAC GCCTGTATGCGGAAAGCGCGCGCGGAGAGAAGAA GAGTGGTGGAAATGCGCGTTGATGGCGATGTGG CGTTTGTGGAAGTGGTTCTGCATGCGAGCCATC AGGGCCAGCCGAGGTGGTTCTGCTGAGACATA TTTTTCGCTTTGAAGCGCATGCTGTGGTGAAG TCGAAGTGAGCATCGATCCGCTGGG (TAA)
luxs iti_ 0.4_ 837	54	MSAEQREFWDRFYAA LDAGDADTASALFNDG TEIYLWDGKVFHTQAE FKAWFVLDLRSRSTGAS REIVAFKVTGNRAEIE VVLHADVDGQPKTVRL THYTEFEGDRLVRVEV KINPL	1.08 0165 86	254	(ATG)AGCGCGGAACAGCAGCGCAATTTTGGG ATCGCTTTTATGCGCGCGCTGGATGCCGGCGATG CGGATACCGCAAGCGCATTTGTTAACGATGGCA CCGAAATTTATCTGTGGGATGGCAAAAGTGTTC ATACCCAGGCGGAATTTAAAGCGTGGTTTGTGG ATCTGCGCAGCGCAGCACCGCGCGGAGCAGAG AAATTGTGGCTTTTAAAGTGACCGCAACCGCG CCGAAATCGAAGTGGTGTGCTGCATGCAGATGTTG ATGGCCAGCCGAAAACCGTGCGCCTGACCCATT ATACCGAATTTGAAGCGCATCGCCTGGTGGCGG TAGAAGTGAAAATTAATCCGCTGGG (TAA)
luxs iti_ 0.4_ 838	55	MSEQAIREFVRRFYDA LDAGDAETAAALFPDG TVIHLWDGRTFHTRAE FRAWFVRLRSTSDAR RRVVELRVVGDPARVR VVLQANVDGEPVVVEL EHEFEFEGDRLVRVRV DIRPL	1.07 6019 35	255	(ATG)AGCGAACAGCGGATTCTGTAATTTGTGC GCCGCTTTTATGATGCGCTGGATGCCGGTGATG CCGAAACCGCAGCTGCACTGTTTCCGGATGGCA CCGTGATTTCATCTGTGGGATGGCCGACCTTTC ATACCCGCGCGGAATTTCCGCGCTGGTTTGTGA GGCTGCGTAGCACTAGCGATGATGCCCGCCGTA GGGTGGTGGAACTGCGTGTGGTTGGTGATCGTG CTAGAGTGCGCGTTGTGCTGCAGGCCAACGTGG ATGGCGAACCAGAGTGGTTGAACTGGAACACG AATTTGAATTCGAAGGCGATCGTCTGCGCGAAG TGCGTGTGATATTCCGCCGCTGGG (TAA)
luxs iti_ 0.4_ 841	56	MSAEEQRKFVEKFYTA LDSGDAETASSLFPDG TLIYLWDGRVFTQAE FRAWFEKLYSQSANAK RSVVRFVEVEGNEAYVE VVLHAVVDGKETVVRL KHNYLFEGDRLVRVEV EIKPM	235. 7553 559	256	(ATG)AGCGCGGAAGAACAGCGTAAATTTGTGG AAAAATTTTATACCGCGCTGGATAGCGCGGATG CCGAAACCGCAGCAGCCTGTTTCCCTGATGGCA CCCTGATTTATCTGTGGGATGGCCGCGTGTTC ATACCCAGGCGGAATTTCCGCGCTGGTTTCAAAA AACTGTATAGCCAGAGCGCGAAGCGGAAACGCA GCGTGGTGGCTTTGAAGTGGAAAGGCAATGAAG CGTACGTGGAAGTGGTGTGCTGCATGCCGTGGTGG ATGGCAAAGAAACCGTGGTGAGACTGAAACATA ACTATCTATTTGAAGGCGATCGCCTGGTTCGCG TCGAAGTTGAAATTAACCGATGGG (TAA)

luxs iti_... 0.4_... 842	57	MSEEAQREFAKRFYAA LDAGDADTAAALFPDG TEIYLWDGKTFETRAE FRAWFEELRSTSDRAK RRVDRFEVNGDTAHVE VTLDAYVNGENRTVRL RHFVHFEGDRVKRVEV EIKPL	13.1 3061 507	257	(ATG)AGCGAAGAAGCGCAGCGGAATTTGCGA AACGCTTTTATGCGGCGCTGGATGCGGGTGATG CGGATACCGCAGCAGCACTGTTCCAGATGGCA CCGAAATTTATCTGTGGGATGGCAAAACCTTTA CCACCCGCGCGGAATTCGCGCGTGGTTGAAG AACTGCGCAGCACCAGCGATCGCGCGAAAAGGC GCGTTGATCGTTTGAAGTGAACGGCGATACGG CGCATGTGGAAGTGACCCCTGGACGCGTATGTTA ACGGTGAAAACCGTACCGTGCGCCTGCGCCATG TGTTTCATTTTGAAGGCGATCGTGTGAAACGCG TCGAAGTGGAAATTAACCGCTGGG (TAA)
luxs iti_... 0.4_... 844	58	MSEEEIREFVRPFYEA LDAGDAETASALFPDG TEIHLWDGTTTFRQAQ FRAWFERLRAQSANAR REIVDLKVEGDRAKVE VILRASFDGEEKVVNL THEFLFEGDRLVRVSV TITPL	1953 .540 428	258	(ATG)TCCGAAGAAGAAATTCGCGAATTTGTGC GCCGCTTTTATGAAGCGCTGGATGCGGGTGATG CGGCAACCGCAAGCGCATTTGTTCCGGATGGCA CCGAAATTCATCTGTGGGATGGTACCACCTTTC GCACCCAGGCGCAGTTTCGCGCGTGGTTTGAAC GCCTGCGTGCAAAAGCGCAACGCGCGCAGAG AAATTGTCGATCTGAAAGTGGAAAGCGATCGCG CGAAAGTTGAAGTGATTCTGCGCGCAGCTTTG ATGGTGAAGAAAAAGTGGTGAACCTGACCCATG AATTTCTGTTTGAAGTGATCGCCTGCTGCGCG TGAGCGTGACCATTAACCGCTGGG (TAA)
luxs iti_... 0.4_... 862	59	MSEEEIKEFWKRFYEA LDAGDAETASALFPDG TEIHLWSGKTFRTREE FRAWFEELYQSSENAK RHIIVSLEVDGDEAYVR VVLHAFVKGESRTVEL EHFFRFEGDHLVEVKV DIIPL	7.78 9219 074	259	(ATG)AGCGAAGAAGAAATTAAGAATTTTGGGA AACGCTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACCGCAAGCGCATTTGTTCCGGATGGCA CCGAAATTCATCTGTGGGATGGTACCACCTTTC GCACCCGCTGAAGAATTCGCGCGTGGTTTGAAG AACTGTATAGCCAGAGCGAAACGCGAAACGCC ATATTGTGAGCCTGGAAGTGGATGGCGATGAAG CCTATGTGCGCGTTGTGCTGATGCGTTTGTGA AAGGCGAAAGCCGACCGTGGAACTGGAACATT TCTTTCGCTTTGAAGGTGATCATCTGTTGGAAG TTAAAGTGGACATTATTCGCTGGG (TAA)
luxs iti_... 0.4_... 868	60	MSEAAIREFVDKFYAA LDAGDAETASALFPDG TKIYLWDGRTFTTQAE FKAWFEELHATSDNAR REVVSMEVDGDVADVE VVLHASFRGEQREVRL RHFVHFEGDELREVEV EILPL	2361 .967 519	260	(ATG)AGCGAAGCGGCGATTTCGCGAATTTGTTG ATAAATTTTATGCGGCGCTGGATGCGGGCGATG CGGAAACAGCAGCAAGCCTGTTTCCAGATGGCA CCAAATTTATCTGTGGGATGGCCGACCTTTA CCACCCAGGCGGAATTTAAAGCGTGGTTTGAAG AACTGCATGCGACCGCATACGCGCGCGCGCG AAGTGGTGAGCATGGAAGTTGATGGCGATGTGG CAGATGTGGAAGTCTGCTGCACGCGAGCTTTC GCGGCGAACAGCGTGAAGTTCTGCTGCGTCATG TGTTTCATTTTGAAGGTGATGAACTGCGGGAAG TGGAAGTAGAAATTCGCGCTGGG (TAA)
luxs iti_... 0.4_... 871	61	MSEEQQREFWARFYAA LDAGDADTASALFPDG TKIHLWDGKTFETRAE FRDWFVRLHSRSENAK RRITSFKVEGDISYVE VVLHASRDGQEHVVKL KHVAKFEGDELREVEV EITPL	2.84 9343 469	261	(ATG)AGCGAAGAACAGCAGCGGAATTTTGGG CGCGCTTTTATGCGGCGTTAGATGCGGGTGATG CGGATACCGCAGCGCGCTGTTTCCAGATGGCA CCAAATCCATCTGTGGGATGGCAAAACCTTTA CCAGCCGCGCGGAATTCGCGCATTTGTTGTGC GCCTGCATAGCCGCTCTGAAAACGCCAAACGCC GCATTACGAGCTTTAAAGTGGAAAGCGATATTA GCTATGTGGAAGTGGTGTGATGCGAGCCGCG ATGGCCAGGAACATGTAGTGAACCTGAAACATG TGGCGAAATTTGAAGGTGATGAACTGGTTGAAG TGAAAGTTGAAATTAACCGCTGGG (TAA)
luxs iti_... 0.4_... 872	62	MSKEEQKNFCKKFYEA LDAGDAETASALFPDG TKIYLWDGKTFSTQEE FRAWFEKLRSANAS RRIVSFNVDGNVAKVK VVLKANYRGKSTVKL	1.56 7380 788	262	(ATG)AGCAAGAAGAACAAGAACTTTTGCA AAAAGTTTATGAAGCGCTGGATGCGGGTGATG CGGAAACCGCTAGCAGCCTGTTTCCGGATGGCA CCAAATTTATCTGTGGGATGGTAAACCTTTA GCACCCAGGAAGAATTCGCGCGTGGTTTGAAG AACTGCGCAGCGAATCGGCGAATGCGAGCCGCC

		EHKFEFEGDKLVKVEV TIEPL			GTATTGTGAGCTTTAAAGTGGATGGGAACGTGG CGAAAGTGAAAGTGGTGTGAAAGCGAAGTATC GCGGCAAGAAAAGCACCGTGAAACTGGAACATA AATTTGAATTCGAAGGCGATAAACTGGTTAAAG TAGAAGTGACCATTGAACCGCTGGG (TAA)
luxs iti_0.4_880	63	MSEEAIREFVKRFYEA LDAGDADTASALFPDG TRIYLWDGTTTFRTOAE FRAWFRELYSKSEGAR REVINLKVDGVDVAYVE VVLHAVYKGEEKTVKL VHVFKFEGDKVVEVHV EIVPL	1070 .718 728	263	(ATG)AGCGAAGAAGCGATTTCGCGAATTTGTGA AACGCTTTTATGAAGCGCTGGATGCGGGTGATG CGGATACCGCAAGCGCGTTGTTTCCGGATGGCA CCCGCATTTATCTGTGGGATGGTACCACCTTTC GCACCCAGGCGGAATTTCCGCGCTGGTTTCGTG AACTGTATAGCAAAAGCGAAGCGCGCGCCGCG AAGTTATTAACTGAAAGTGGATGGTGTATGTGG CGTATGTGGAAGTGGTGTGCATGCGGTGTATA AAGGTGAAGAAAAGACCGTGAAACTGGTGCATG TGTTTAAATTTGAAGGCGATAAACTGGTTGAAG TGCACGTGGAAATTTGCGCGCTGGG (TAA)
luxs iti_0.4_887	64	MSEEEIREFVKRFYTA LDAGDAETASALFPDG TKIYLWDGTTTETREG FAAWFRKLKSQSEEA RHVVKLVKVEGNVAYIE VVLHAKHKGEKTVRL THIYQFEGDKLVEVRV EIKPL	42.4 9136 144	264	(ATG)AGCGAAGAAGAAATTCGCGAATTTGTGA AACGCTTTTATACCGCGCTGGATGCGGCGGATG CGGAAACTGCGAGCGCACTGTTTCCAGATGGCA CGAAATTTATCTGTGGGATGGTACCACCTTTG AAACCCGTGAAGGCTTTGCGCGCTGGTTTCGCA AACTGAAAAGCCAGTCGGAAGAACGAAACGCC ATGTGGTGAAATTGAAAGTGAAGGCAACGTGG CCTATATTGAAGTGGTGTGCATGCGAAACATA AAGGTGAAGAAAAGACCGTGCGCCTGACCCATA TCTATCAGTTTGAAGGCGATAAACTGGTTGAAG TTCCGCTGGAAATTAACCGCTGGG (TAA)
luxs iti_0.4_889	65	MSEAEIKNFVKRFYEA LDAGDAETASALFPDG TEIHLWDGTTTFRTRA FRAWFEELYSRSEDAK REVTKLEVKGNVAFVE VVLKANLNGKERTVKL KHIFHFEGDSLVRVEV SIVPL	779. 2273 67	265	(ATG)AGCGAAGCGGAAATTAACCTTTGTGA AACGCTTTTATGAAGCGCTGGATGCGAGCGATG CGGAAACCGCTAGCGCACTGTTTCCGATGGCA CCGAAATTCATCTGTGGGATGGTACCACCTTTC GCACCCGCGCGAATTTCCGCGCTGGTTTGAAG AACTGTATAGCGCAGCGAAGATGCGAAACGCG AAGTGACCAACTGGAAGTGAAGGCAACGTGG CGTTTGTGGAAGTTGTGCTGAAAGCGAACCCTGA ACGGCAAGAACGACCGTGAAACTGAAACATA TTTTTCATTTTGAAGGCGATAGCCTGGTGCAGG TTGAAGTGTGATTGTGCGCTGGG (TAA)
luxs iti_0.4_893	66	MSEQEQKEFWKIFYQA LDAGDAETASALFKDG TEIYLWDGKVPKTQEE FKAWFVELYSKSLNAK RKIVEHEVKGNESYVK VVLRASVDGEERTVLL EHKFLFEGDKLVKVS EITPL	60.6 5169 316	266	(ATG)AGCGAACAGGAACAGAGAAATTTTGGG AAAAATTTTATCAGGCGCTGGATGCGGGCGATG CGGAAACCGCGAGCGCTGTTTAAAGATGGCA CCGAAATTTATCTGTGGGATGGCAAACTGTTCA AAACCCAGGAAGAATTCAAAGCGTGGTTCTGTG AACTGTATAGCAAAAGCCTGAACGCGAAACGCA AAATTGTGGAACATGAAGTGAAGGCAACGAAT CTTATGTGAAAGTGGTGTGCGCGCCAGCGTGG ATGGCGAAGAACGCACCGTGTGTTGGAACACA AATTTCTGTTTGAAGGCGATAAACTGGTTAAAG TGAGCGTTGAAATTACCCCGCTGGG (TAA)
luxs iti_0.4_894	67	MTAEEQREFVKRFYEA LDAGDAETASALFPDG TLIHLWDGKTFHTREE FRAWFVELRSTSDDAK REVTAFEDVDGVARVE VVLHANIQGNKKTVAL THEFLFEGDKLVEVRV DIKPL	3.09 3296 475	267	(ATG)ACCGCGGAAGAAGCAGCGCAATTTGTGA AACGCTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACTGCGAGCGCTCTGTTTCCAGATGGCA CCCTGATTCTCTGTGGGATGGCAAAACCTTTC ATACCCGCGAAGAATTTCCGCGCTGGTTTGTGG AACTGCGCAGCACGCGATGATGCGAAACGCG AAGTGACCGCCTTTGAAGTGGATGGCGATGTGG CGCGCGTTGAAGTTGTGTGCATGCGAACATTC AGGGCAACAAAAGACCGTGCGCTGACCCACG AATTTCTGTTTGAAGGCGATAAACTGGTGGAAG TGCGGTGGATATTAAACCGCTGGG (TAA)

luxs iti_... 0.4_... 914	68	MSEEEQKNFVRKFYDA LDAGDSETASALFKDG TKIYLWDGQVFETQEE FRAWFEKLYSTSTDAK REIVSFKVDGNKAVVE VVLKAIKDGEKVVNL KHVFLFEGDEVVEVEV DIVPS	652. 4934 347	268	(ATG)AGCGAAGAAGAACAGAAAACTTTGTGC GCAAATTTTATGATGCGCTGGATGCGGGCGATA GCGAAACCGCGAGCGCGCTGTTTAAAGATGGCA CCAAAATTTATCTGTGGGATGGCCAGGTTTTTG AAACCCAGGAAGAATTTCGCGCGTGGTTTGAAA AATTGTATAGCACCAGCACCAGATGCCAAACGCG AAATTGTGAGCTTTAAAGTGGATGGCAACAAAG CGGTGGTTGAGGTGGTGTGAAAGCGATTAAAG ACGGTGAAGAAAAAGTGGTGAACCTGAAACATG TGTTTCTGTTTGAAGGTGATGAAGTGGTAGAAG TGGAAGTTGATATTGTGCCGAGCGG (TAA)
luxs iti_... 0.4_... 915	69	MTAEQQRNFWARFYAA LDAGDAETAAALFPDG TVIHLWDGRTEHTQAE FRAWFEALRSTSENAR REIVFFQVDGNTADVE VVLHASVDGDARTVRL RHI FQFEGDHLVEVHV DIRPL	4.01 5203 87	269	(ATG)ACCGCGGAACAGCAGCGCAACTTTTGGG CGCGCTTTTATGCGCGCTGGATGCGGGTGTG CGGAAACTGCAGCAGCTTATTTCCTGATGGCA CCGTGATTCATCTGTGGGATGGCCGACCTTTC ATACCCAGGCGGAATTTCGCGCGTGGTTTGAAG CGCTGCGTAGCACCAGCGAAAACGCGCTCGTG AAATTGTGTTTTCCAGGTGGATGGCAACACCG CCGATGTGGAAGTGGTGTGTCATGCGAGCGTGG ACGGTGACGCAAGAACCGTGGCTCTGAGACATA TTTTTCAGTTCGAAGGCGATCATCTGTTGAAG TGCATGTGGATATTCCGCCGCTGGG (TAA)
luxs iti_... 0.4_... 921	70	MSEEEQKNFVAIFYKA LDAGDAETASALFPDG TEIHLWDGKTFKTQAE FKAWFEKLSTSDNAK RKVVSPKVNIGNIADVQ VVLEANINGEPVKVKL NHTFKFKGDKLVEVNV QIEPL	907. 3780 235	270	(ATG)AGCGAAGAAGAACAAAGAACTTTGTGG CGGCGTTTTATAAAGCGCTGGATGCGGGCGATG CCGAAACTGCGAGCGCGTGTGTTCCAGATGGCA CCGAAATTCATCTGTGGGATGGCAAAACCTTTA AAACCCAGGCGGAATTAAAGCGTGGTTTGAAA AACTGTTTAGCACCAGCGATACGCGAAACGCA AAGTGGTGAGCTTTAAAGTGAACGGCAACATTG CGGATGTGCAAGTGGTGTGGAAGCGAATATTA ACGGCGAACCAGTGAAAGTTAAACTGAACCATA CCTTCAAATTCAAAGGCGATAAACTGGTGAAG TTAACGTGCAGATTGAACCGCTGGG (TAA)
luxs iti_... 0.4_... 923	71	MSEEEIREFVRKFYAA LDAGDADTASALFPDG TEIHLWDGVTFTKTQAE FKAWFTELKSKSDNAK REVVKLKVDGNKADVE VVLHATIDGKEVTVHL RHIFEFEGDKLVKVEV VIEPL	366. 0525 225	271	(ATG)AGCGAAGAAGAAATTCGTGAATTTGTGC GCAAATTTTATGCGCGCTGGATGCGGGTGTG CGGATACCGCAAGCGCATTGTTTCCGATGGCA CCGAAATTCATCTGTGGGATGGCGTGACCTTTA AAACCCAGGCGGAATTAAAGCGTGGTTTACCG AACTGAAAAGCAAAAGCGATAACGCGAAACGCG AGGTGGTGAAATTGAAAGTGGATGGTAACAAAG CGGATGTGGAAGTGGTGTGTCATGCGACCATTG ATGGCAAAGAAGTGACCGTGCATCTGCGCCATA TTTTTGAATTGGAAGGCGATAAACTGGTTAAAG TTGAAGTTGTGATTGAACCGCTGGG (TAA)
luxs iti_... 0.4_... 945	72	MSAEAQREFVRKFYDA LDAGDADTASALFADG TRIYLWDGREFRTRAE FRAWFERLRSRSDAAK RHVTSFKVEGNVAQVV VVLKANFRGKEHVVEL LHEFVFDGRLVEVHV EIRPR	4.07 5328 265	272	(ATG)AGCGCGGAAGCGCAGCGTGAATTTGTGA AACGCTTTTATGATGCCCTGGATGCGGGTGTG CGGATACCGCGAGTGCCTTGTGTTGCCGATGGCA CCCGCATTTATCTGTGGGATGGCCGCGAATTC GCACCCGTGCGGAATTTAGAGCGTGGTTTGAAC GTCTGCGCAGCCGAGCGATGCGGCGAAACGTC ATGTGACCAGCTTTAAAGTGAAGGCAACGTGG CGCAGGTGGTGGTGTGCTGAAAGCGAACTTTC GCGGCAAAGAACATGTGGTGAACGTGCTGCATG AATTCGTGTTTGAAGGCGATCGCCTGGTTGAAG TGCATGTGGAATTCGCCCGCGCGG (TAA)
luxs iti_... 0.4_... 955	73	MSEEEIREFVRKFYEA LDAGDAETASALFPDG TKIHLWDGITFNTQEE FQKWFVELRSQSDNAK REVVS LKVDGNHAKVE VVLKANIGGEDRVVVKL	32.4 8928 818	273	(ATG)AGCGAAGAAGAAATCCGCGAATTTGTGA AACGCTTTTATGAAGCGCTGGATGCGGGTGTG CGGAAACCGCAAGCGCACTGTTTCCGGATGGGA CCAAAATTCATCTGTGGGATGGCATTACCTTTA ACACCCAGGAAGAATTTAGAAATGGTTTGTGG AACTGCGCAGCCAGAGCGATAACGCAAAACGTG

		THHFLFEGDKLIEVRV EIEPL			AAGTGGTGAGCCTGAAAGTGGATGGTAACCATG CGAAAGTTGAAGTTGTGCTGAAAGCGAACATTG GCGGCGAAGATCGCGTGGTGAAACTGACCCATC ATTTTCTGTTTGAAGCGGATAAACTGATTGAAG TGCGCGTGGAAATTGAACCGCTGGG (TAA)
luxs iti_0.4_957	74	MSAEAIREFVERFYAA LDAGDAETASALFPDG TIIHLWDGRVFHTREE FRRWFVDLFTSDNAS RSVVDLHVDGNVAHV VRLKASWRGKKRTVLL THVFEFEGDHLVKVTV SIQPV	11.6 8071 873	274	(ATG)AGCGCGGAAGCGATTGCGTAATTTGTG AACGTTTTTATGCGCGCTGGATGCGGGCGATG CGGAAACTGCAAGCGCACTGTTTCCTGATGGCA CCATTATTCACTCTGTGGGATGGCCGCGTGTTC ATACCGCGAAGAATTTCGCCGCTGGTTTGTGG ATTTGTTTAGCACCAGCGATAACCGAGCCGCA GCGTGGTGGATCTGCATGTGGATGGCAACGTGG CTCATGTGGTGGTGGCGCTGAAAGCGAGCTGGC GTGGCAAGAAACGACCGTGTGTGACCCATG TGTTTGAATTGCAAGGTGATCATCTGGTGAAG TGACCGTGAGCATTCAGCCGCTGGG (TAA)
luxs iti_0.4_967	75	MSAEIREFVARFYAA LDAGDADTASALFPNG TKIYLWDGTVFTTQAE FRAWFVKLRSTSENAK REVVFLEVEGNEAFVE VVLHAEIKGQKKTVRL RHVFKFEGDHLVEVSV DIEPL	11.3 2826 538	275	(ATG)AGCGCGGAAGAAATTCGCGAATTTGTGG CGCGCTTTTATGCGCGCTGGATGCGGGTGATG CGGATACCGCAAGCGCGCTGTTTCCGAACGGTA CCAAATTTTATCTGTGGGATGGCACCCTGTTTA CCACCCAGGCGGAATTCGCGCGTGGTTTGTGA AACTGCGCAGCACCAGCGAAACGCGAAACGTG AAGTGGTGTCTGGAAGTGAAGGCAACGAAG CGTTTGTGGAAGTTGTGCTGCATGCGGAAATTA AAGGCCAGAGAAACCGTGCGCCTGCGCCATG TGTTTAAATTTGAAGCGCATCATCTGGTTGAAG TGCTGTGGATATTGAACCGCTGGG (TAA)
luxs iti_0.4_969	76	MSKEDIEEFCKFYDA LDAGDAETASALFPDG TEINLWDGKVTFTKSE FRAWFRELYARSDHAK REITHLEVVDGNFATIE VVLNASVDGEQKTVSL KHFFQFEGDRLVRVDV KINPL	97.6 1368 348	276	(ATG)AGCAAAGAAGATATTGAAGAAATTTGCA AAAAGTTTTATGATGCGCTGGATGCGGGCGATG CCGAAACAGCAAGCGCCTGTTTCCAGATGGCA CCGAAATTAACCTGTGGGATGGCAAAGTGTTA CCACCAAAGCGAATTTGCGCGTGGTTTCGCG AACTGTATGCCCGCAGCGATCATGCCAAACGTG AAATTACCCATCTGGAAGTGGATGGTAACCTTG CGACCATTTGAAGTGGTGTGAACGCGAGCGTGG ACGGCGAACAGAAAACCGTGAGCCTGAAACATT TCTTTCAGTTTGAAGCGCATCGCCTGGTGGCGG TTGATGTGAAATCAACCCGCTGGG (TAA)
luxs iti_0.4_977	77	MSEEEIREFWQRFYEA LDAGDAETASALFPDG TEIYLWDGKTFRTAE FRAWFEELHSTSENAK RHVTALEVVDGNVAVQ VVLHASINGEEKTVKL DHETHFEGDRLVRVTV DIKPL	2.06 7035 245	277	(ATG)AGCGAAGAAGAAATTCGCGAATTTTGGC AGCGCTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACCGCGAGCGCATTTTCTCTGATGGCA CCGAAATTTATCTGTGGGATGGCAAACCTTTC GCACCCGCGCGGAATTCGCGCGTGGTTTGAAG AACTGCATAGCACCAGCGAAACGCCAAACGTC ATGTGACCGCTCTGGAAGTGGATGGTAACGTGG CGCGCGTTTCAAGTGGTGTGCATGCGAGCATTA ATGGTGAAGAAAAGACCGTGAAACTGGATCATG AAACCCATTTTGAAGGTGATCGCCTGGTGGCGG TGACCGTGGATATTAAACCGCTGGG (TAA)
luxs iti_0.4_998	78	MSEEEQRNFVKRFYEA LDAGDAETASALFPDG TVIHLWDGTTTFTQAE FRAWFTELIRSTSENAK REVTKFAVDGNVAHVE VVLHANHNGEPRVVRL THEFHFEGDHLVEVTV RIDPL	4.93 7111 265	278	(ATG)AGTGAAGAAGAACAGCGCAACTTTGTGA AACGCTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACCGCGAGCGCGTTATTTCTCTGATGGCA CCGTGATTCACTGTGGGATGGTACCACCTTTC ATACCCAGGCGGAATTCGCGCGTGGTTTACCG AACTGCGCAGCACCAGCGAAACGCGAAACGTG AAGTGACCAAATTTGCGGTGGATGGCAACGTGG CGCATGTGGAAGTTGTGTGCATGCGAACCATA ACGGTGAACCGCGCGTTGTGCGCCTGACCCATG AATTTCAATTTTGAAGCGCATCATCTGGTTGAAG TTACCGTGGCATCGATCCGCTGGG (TAA)

luxs iti_... 0.2_... 7	79	MSEEEQREFVRRFYEA LDAGDAETASALFPDG TVIDLWDGKTFHTRAE FREWFVKLKSTSDNAK REVTNFKVDGDVAHVE VVLKASINGEEKVVKL THTFQFEGDRLVRVSV KIEPL	78.3 3724 948	279	(ATG)AGCGAAGAAGAACAGCGCGAATTTGTGC GCCGCTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACTGCAAGCGCGCTGTTTCCAGATGGCA CCGTGATCGATCTGTGGGATGGCAAAACCTTTC ATACCCGCGCCGAATTCGCGAATGGTTTGTGA AACTGAAAAGCACCAGCGATAACGCGAAACGTG AAGTGACCAACTTTAAAGTGGATGGCGATGTGG CGCATGTGGAAGTGGTGCTGAAAGCGAGCATTA ACGGTGAAGAAAAAGTGGTTAACTGACCCATA CCTTCCAGTTTGAAGGCGATCGCCTGGTGCGCG TGAGCGTGAAATTTGAACCGCTGGG (TAA)
luxs iti_... 0.2_... 35	80	MSEEEQKEFCKKFYEA LDAGDADTASALFPDG TKIHLWDGKTFTTQAAQ PKAWFVELYSKSENAK REITKFEVDGNVADVE VVLHAIVNGEKKTVKL KHVFKFEGDKLVEVNV EIKPL	170. 6966 137	280	(ATG)TCGGAAGAAGAACAGAAAGATTTTGCA AAAAGTTTTATGAAGCATTTGGATGCGGGTGATG CCGATACGGCGAGCGCCTGTTTCCAGATGGCA CCAAAATCCATCTGTGGGATGGCAAAACCTTTA CCACCCAGGCGCAGTTTAAAGCCTGGTTTGTGG AACTGTATAGCAAAAGCGAAAACGCCAAACGTG AAATTACGAAATTTGAAGTGGATGGTAACGTGG CGGATGTGGAAGTGGTGCTGCATGCGATCGTGA ACGGCGAAAAGAAAACCGTGAAACTGAAACATG TTTTTAAATTCGAAGGTGATAAACTGGTTGAAG TCAACGTGCAAAATCAAACCGCTGGG (TAA)
luxs iti_... 0.2_... 41	81	MSAEEIKNFCDKFYKA LDAGDADTASALFPDG TEIYLWDGKTFTKQAE FKAWFEKLYSTSKDAK RKITSLEVNGNVAKVK VVLEALIDGEKKT VNL EHIFYFEGDKLVKVEV SIEPL	1348 .814 098	281	(ATG)AGCGCGGAAGAAATTAATAAACTTCTGCG ACAAATTTTATAAAGCGCTGGATGCGGGCGATG CGGATACCGCAAGCGCCTTGTTCAGATGGCA CCGAAATTTATCTGTGGGATGGCAAAACCTTTA AAACCCAGGCGGAATTTAAAGCGTGGTTTGAAA AACTGTATAGCACCAGCAAAAGATGCGAAACGCA AAATTACCAGCCTGGAAGTGAATGGCAATGTTG CGAAAGTGAAAGTGGTGCTGGAAGCGATTATTG ATGGCGAAAAGAAAACCGTGAACTGGAACATA TTTTCTATTTTGAAGGCGATAAACTGGTCAAAG TTGAAGTGAGCATTTGAACCGCTGGG (TAA)
luxs iti_... 0.2_... 44	82	MSEEEQREFVARFYAA LDAGDAETASALFPDG TEIHLWDGKTFTTRA FPAWFEKLHSLSDNAS RHVTSFKVDGNVAEVE VVLHADFKGKKLT VKL RHRYQFEGDRLVRVDV EIFPL	5.90 5321 355	282	(ATG)AGCGAAGAAGAACAGCGCGAATTTGTGG CGCGCTTTTATGCGCGCTGGATGCGGGTGATG CGGAAACTGCAAGCGCGTTGTTTCCGATGGCA CCGAAATTCATCTGTGGGACGGCAAAACCTTTA CCACCCGCGCGGAATTTGCGCGTGGTTTGAAA AATTGCATAGCTTGAGCGATTAATGCGAGCCGCC ATGTGACCAGCTTTAAAGTGGATGGTAACGTGG CGGAAGTGGAAGTTGTGCTGCATGCGGATTTTA AAGGCAAGAACTGACCGTGAAATTCGCCATC GCTATCAGTTTGAAGGCGATCGCCTGGTGCGCG TCGATGTGGAATTTTTCGCTGGG (TAA)
luxs iti_... 0.3_... 478	83	MSADAQREFVRRFYEA LDAGDAETAAALFPDG TEIHLWDGKTFTTQAE FPAWFEKLRAQSENAK RHVVNFQVDGNDADVE VVLHATYNGEQKTVRL KHKHFHFEQQLVRVDV TIEPL	1.48 9979 267	283	(ATG)AGCGCGGATGCGCAGCGCGAATTTGTGC GCCGCTTTTATGAAGCGCTGGATGCGGGTGATG CGGAAACCGCAGCCCGCTTGTTCCTGATGGCA CCGAAATTCATCTGTGGGATGGCAAAACCTTTC GCACCCAGGCGGAATTTCTGCGTGGTTTGAAA AACTGCGCGCGCAGAGCGAAAACGCGAAACGCC ATGTGGTGAACCTTTCAGGTGGATGGTAACGATG CCGATGTGGAAGTGGTGCTGCATGCGACCTATA ACGGCGAACAGAAAACCGTGCGCCTGAAACATA AATTTTCATTTTGAAGGCGATCAGCTGGTGCGCG TTGATGTGACCATTGAACCGCTGGG (TAA)
luxs it_R 65A	84	MSEEQIRQFLRRFYEA LDSGDADTAAALFHPG VTIHLWDGVTFSTRREE FREWFERLFSRDKDAQ AEIKSLEVVRGDTVEVH VOLHATHNGQKHTVDA	1	284	(ATG)AGCGAAGAACAATTCGCCAGTTTCTGC GCCGCTTTTATGAAGCGTTAGATTCTGGCGATG CGGATACCGCGGCCAGCCTCTTTCATCCGGCGG TGACCATTCATCTGTGGGATGGCGTTACCTTTA CCAGCCGTGAAGAATTTCCGAATGGTTTGAAC GCCTGTTTAGCACCCGCAAGATGCGCAGGCGG

		THHWHFRGNRVTEM RV HINPT			AAATTAAAAGCCTGGAAGTGC GCGGCGATACCG TGGAAGTTCACGTGCAGCTGCATGCGACCCATA ACGGCCAGAAACATACCGTTGATGCGACGCATC ATTGGCATTTCGCGGCAACCGCGTGACGGAAG TGCGCGTG CATATTAACCCGACCGG (TAA)
luxs iti	85	MSEEQIRQFLRRFYEA LDSGDADTAASLFHPG VTIHLWDGVTFTSREE FREWFERLFTSKDAQ REIKSLEVRGDTVEVH VQLHATHNGQKHTVDL THHWHFRGNRVTEVRV HINPT	633. 2411 887	285	(ATG)AGCGAAGAACAATCCGCCAGTTTCTGC GCCGCTTTTATGAAGCGCTGGATAGCGGCGACG CGGATACCGCAGCGAGCTTATTTTCATCCGGGCG TGACCATTTCATCTGTGGGATGGCGTTACCTTTA CCAGCCGTGAAGAATTCGCGAATGGTTTGAAC GCCTGTTTAGCACCAGCAAAGATGCGCAGCGCG AAATTAAAAGCCTGGAAGTGC GCGGCGATACCG TGGAAGTTCATGTGCAGCTGCATGCGACCCATA ACGGCCAGAAACATACCGTTGATCTGACCCATC ATTGGCATTTCGCGGCAACCGCGTGACGGAAG TGAGAGTG CATATTAACCCGACCGG (TAA)
luxs iti 0.2 35	86	MSEEEQKEFCCKFYEA LDAGDADTASALFPDG TKIHLWDGKTFTTQAE FKAWFVELYSKSENAK REITKFEVDGNVADVE VVLHAI VNGEKKTVKL KHVFKFEGDKLVEVNV EIKPL	318. 4996 545	286	(ATG)AGCGAAGAAGAACAAGAAATTTTGCA AGAAATTTTATGAAGCGCTGGATGCGGCTGATG CGGATACCGCAAGCGCACTGTTTCCCGATGGTA CCAAATCCATCTGTGGGATGGCAAAACCTTTA CCACCCAGGCGCAGTTTAAAGCGTGGTTTGTGG AACTGTACAGCAAAAGCGAAATGCGAAACGTG AAATTACCAAATTTGAAGTGGATGGTAACGTGG CGGATGTGGAAGTGGTGCTGCATGCGATTGTGA ACGGCGAAAAGAAAACCGTGAAACTGAAACATG TGTTTAAATTCGAAGCGATAAACTGTTTGAAG TTAATGTTGAAATCAAACCGCTGGG (TAA)
luxs iti 0.2 41	87	MSAEEIKNFCDKFYKA LDAGDADTASALFPDG TEIYLWDGKTFTQAE FKAWFEKLYSTSKDAK RKITSLEVNGNVAKVK VVLEAIDGEEKTVNL EHI FYFEGDKLVKVEV SIEPL	1719 .627 505	287	(ATG)AGCGCGGAAGAAATTAACCTTTTGCG ATAAATTTTATAAAGCGCTGGATGCGGGTGATG CGGATACCGCGAGTGCACTGTTTCTGATGGCA CCGAAATTTATCTGTGGGATGGCAAAACCTTTA AAACCCAGGCGGAATTTAAAGCGTGGTTTGAA AACTGTATAGCACCAGCAAAGATGCGAAACGTA AAATTACCAAGCTGGAAGTGAACGCAACGTTG CGAAAGTGAAAGTGGTGCTGGAAGCGATTATTG ATGGCGAAAAGAAAACCGTTAACCTGGAACATA TTTTCTATTTTGAAGGTGATAAACTGGTTAAAG TGGAAGTATCTATTGAACCGCTGGG (TAA)
luxs iti 0.2 44	88	MSEEEQREFVARFYAA LDAGDAETASALFPDG TEIHLWDGKTFTTRA FRAWFEKLHSLSDNAS RHVTSFKVDGNVAEVE VVLHADFKGKKLTVKL RHRYQFEGDRLVRVDV EIFPL	1317 .533 518	288	(ATG)AGCGAAGAAGAACACGCGTGAATTTGTGG CGCGCTTTTATGCGGCGCTGGATGCGGGTGATG CGGAAACTGCAAGCGCGCTGTTTCTGATGGCA CCGAAATTCATCTGTGGGATGGCAAAACCTTTA CCACCCGCGCGGAATTCGCGCGTGGTTTGAAA AATTGCATAGCCTGAGCGATAACGCGAGCCGCC ATGTGACCAGCTTTAAAGTGGATGGTAACGTGG CGGAAGTGGAGTTGTGCTGCATGCGGATTTTA AAGGCAAAAAGCTGACCGTGAAACTGAGACATC GCTATCAGTTTGAAGGCGATCGCCTGGTGCGCG TTGATGTGGAATTTTCCGCTGGG (TAA)
luxs iti 0.2 170	89	MSEEEIRQFVKKFYEA LDAGDAETASALFPDG TKIYLWDGTVFNTQAE FKAWFVELYSTSENAK REVVKFEVDGNKAKVE VVLHASIDGEEKTVKL KHEFYFEGDKVKEVKV EIEPL	228. 1472 011	289	(ATG)AGCGAAGAAGAGATTCCGCCAGTTTGTGA AGAAATTTTATGAAGCGCTGGATGCAGGTGATG CGGAAACCGCAAGCGCCCTGTTTCCGGATGGCA CCAAATTTATCTGTGGGATGGTACCGTGTTTA ACACCCAGGCGGAATTTAAAGCGTGGTTTGTGG AACTGTATTCGACGAGCGAAACGCGAAACGTG AAGTGGTGAAATTTGAAGTCGATGGTAACAAAG CGAAAGTGGAAAGTCGTGCTGCATGCGAGCATTG ATGGCGAGGAAAAGACCGTGAAACTGAAACATG AATTTTACTTTGAAGGCGATAAAGTGAAAGAAG TTAAAGTTGAAATTTGAACCGCTGGG (TAA)

luxs iti_... 0.2_... 196	90	MSEEEQREFWRRFYEA LDAGDAETASALFPDG TEIYLWDGRTEFTQAE FRAWFERLRATSEDA REIVSFRVEGNRSEVE VVLHANVNGEKKTVRL RHYFEFEGDRLVRVEV EIVPL	684. 0103 663	290	(ATG)TCCGAAGAAGAACAGCGCGAATTTTGGC GCCGCTTTTATGAAGCCCTGGATGCGGCGATG CTGAAACCGCGAGCGCGTTATTTCCGGATGGCA CCGAAATTTATCTGTGGGATGGCCGTACCTTTC GCACCCAGGCGGAATTCGCGCATGGTTGAAC GCCTGCGCGCCACTAGCGAAGATGCGCGCAGAG AAATTTGTGAGCTTTCGCGTGAAGGCCAATCGCA GCGAAGTGAAGTTGTGCTGCATGCGAACGTGA ACGGCGAAAAAGAAACCGTGCCTGAGACATT ATTTTGAATTTGAAGGCGATCGCCTGGTGC TTGAAGTTGAAATCGTGCCGCTGGG (TAA)
luxs iti_... 0.3_... 453	91	MSEEEIKEFCCKFYEA LDAGDADTASALFKDG TEIYLWDGRTEFKTRAE FKAWFTRLRSTSDNAK REIVSLSVNGNVAKVK VVLRASINGEERTVLL THEFLFEGDELVEVRV SIKPL	71.8 1271 596	291	(ATG)AGCGAAGAAGAAATTAAGAATTTTGCA AAAAGTTTATGAAGCGCTGGATGCGGGCGATG CGGATACCGCGAGCGCGCTGTTTAAAGATGGCA CCGAAATTTATCTGTGGGATGGTACCTTTTA AAACCGCGCGGAATTTAAAGCGTGGTTTACCC GTCTGCGCAGCACCTCTGATAACGCGAAACGTG AAATTTGTGAGCTGTGCGTGAACGCGAACGTGG CGAAAGTGAAGTGGTCTGCGCGCGAGCATTA ACGGTGAAGAAGCGACCGTGTGCTGACCCATG AATTTCTGTTTGAAGGCGATGAACGTGGGAAG TGCGCGTGAGCATCAAACCGCTGGG (TAA)
luxs iti_... 0.2_... 63	92	MSKEEQEEFVKQFYEA LDAGDAETASALFPDG TVIHLWDGKTFHTQAE FRAWFEELKSTSENAK REVTKFEVDGDVADVE VVLKANINGEEKVVNL KHKFKFEGDKVVEVWV EIEPL	281. 5210 781	292	(ATG)AGCAAAGAAGAACAGAGGAATTTGTGA AACAGTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACCGCAAGCGCACTGTTTCCAGATGGCA CCGTGATCCATCTGTGGGATGGCAAAACCTTTC ATACCCAGGCGGAATTTGCGCGCTGGTTTGAAG AACTGAAAAGCACCAGCGAAACGCGAAACGTG AAGTGACCAAAATTTGAAGTGGATGGCGATGTGG CGGATGTGGAAGTGGTGTGAAAGCGAACATTA ACGGCGAAGAAAAAGTGGTTAACCTGAAACATA AATTTAAATTCGAAGGCGATAAAGTTGTCGAAG TTTGGGTGGAATTTGAACCGCTGGG (TAA)
luxs iti_... 0.2_... 138	93	MSEDAIREFVKRFYEA LDAGDADTASALFPDG TEIHLWDGRVFRTRAE FPAWFVELRSQSENAK REVTDLVEGNVARVE VVLHANYKGEERTVRL RHEFQFEGDKVVEVRV EIEPL	73.0 1105 736	293	(ATG)AGCGAAGATGCCATTCGCGAATTTGTGA AACGTTTTTATGAAGCGCTGGATGCGGGCGATG CGGATACCGCGAGCGCATTATTTCCGGATGGCA CCGAAATTCATCTGTGGGATGGCCGCGTGTTC GCACCCGCGCGGAATTTCTGCGCTGGTTTGTGG AACTGCGCAGCCAGAGCGGAAACGCGAAACGTG AAGTGACCGATCTGGAAGTGAAGGCCAACGTGG CGCGCGTGGAAGTCTGCTGCATGCGAACTATA AAGGCGAAGAACCGACCGTGCCTGCGCCATG AATTTCAGTTTGAAGGCGATAAAGTGGTTGAAG TGCGCGTTGAAATTTGAACCGCTGGG (TAA)
luxs iti_... 0.4_... 1046	94	MSEEEQKEFCRRFYLA LDAGDADTASALFKDG TKIYLWDGKVENTQEE FRKWFVELRSTSDQAK RSIEDFKVEGNTAKVK VVLEASKNGEKRTVGL EHI FEFEGDEVKEVYV KIKPL	10.5 2246 026	294	(ATG)TCAGAAGAAGAACAGAAAGATTTCTGCC GCCGCTTTTATCTGGCGCTGGATGCGGGTGATG CGGATACCGCAAGCGCACTGTTTAAAGATGGCA CCAAATTTATCTTTGGGATGGCAAAGTGTTTA ACACCCAGGAAGAATTTGCGAAATGGTTTGTGG AACTGCGCAGCACCAGCGATCAGGCGAAACGCA GCATTGAAGATTTTAAAGTGAAGGCCAACACCG CGAAAGTGAAGTGGTGTGGAAGCGAGCAAAA ACGGCGAAAAACGCACGGTGGGCTGGAAACATA TTTTTGAATTTGAAGGCGATGAGGTGAAAGAAG TGATGTGAAAAATTAACCGCTGGG (TAA)
luxs iti_... 0.4_... 1061	95	MSATEQREFNKRIFYEA LDAGDADTASALFPDG TEIYLWDGTTFTTQEE FRAWFVELRSTSENAR RHIVSFTVDGNKADVE VVLEATVKGEKRVRL	17.7 1596 406	295	(ATG)AGCGCGACCGAACAGCGCGAATTTAACA AACGCTTTTATGAAGCGCTGGATGCGGGTGATG CGGATACCGCAAGCGCACTGTTTCCGGATGGCA CCGAAATTTATCTGTGGGATGGTACCACTTTC ATACCCAGGAAGAATTTGCGCGTGGTTTGTGG AACTGCGCAGCACCTCTGAAAACGCGCGCAGAC

		THNYQWEGDKLVKVTV KIVPL			ATATTGTGAGCTTTACCGTGGACGGCAACAAAG CGGATGTGGAAGTGGTGCTGGAAGCGACCGTGA AAGGCGAACCAGAAACGCGTGCCTGACTCATA ACTATCAGTGGGAAGCGGATAAACTGGTGAAAG TGACCGTTAAAATTGTGCCGCTGGG (TAA)
luxs iti_ 0.4_ 1066	96	MSEEEIREFVKRFYQA LDAGDADTASALFPDG TKIYLWDGVI FTKREE FRKWFVELKSKSDNAK REVVDLRVEGDTADVS VVLKAKYKGEPKVVKL NHKFFYFEGDKLVEVEV EIVPM	4.92 8818 245	296	(ATG)AGCGAAGAAGAAATTCGCGAATTTGTGA AACGCTTTTATCAGGCGCTGGATGCGGGTGATG CGGATACCGCAAGTGCGCTGTTTCCGGATGGTA CCAAAATTTATCTGTGGGATGGCGTTATTTTFA CCAAACGCGAGGAATTCGCAAAATGGTTCGTGG AACTGAAAAGCAAAAGCGATAACGCGAAACGAG AAGTGGTGGATCTGCGCGTGGAAGCGGATACGG CGGATGTGAGCGTGGTGCTGAAAGCGAAATATA AAGGCGAACCAGAAAGTGGTTAACTGAACCATA AATTTTATTTTGAAGGTGATAAACTGGTTGAAG TGGAAAGTTGAAATTGTGCCGATGGG (TAA)
luxs iti_ 0.4_ 1081	97	MTEEEQREFWERFYKA LDAGDAETASALFPDG TEIYLWDGTVFKTQEE FRSWFVDLYSKSNNNAK RHIVEFRVEGDVSYVE VVLHASENGTEKTVRL THITEFEGDKVVKVEV SITPL	1.31 9281 272	297	(ATG)ACCGAAGAAGAACAGCGCAATTTTGGG AACGTTTTTATAAGCGCTGGATGCGGGTGATG CGGAAACCGCGAGCGCATTGTTTCCGGATGGCA CCGAAATCTATCTGTGGGATGGTACCGTGTTTA AAACCCAGGAAGAATTCGCGAGCTGGTTTGTGG ATCTGTATAGCAAAAGCAACACGCGCAACGCC ATATTGTGGAATTTAGGGTGGAAAGCGGATGTGA GCTATGTGGAAGTTGTGCTGCATGCGAGCGAAA ACGGCACGGAAGACCGTGCGCCTGACCCATA TTACCGAATTTGAAGGTGATAAAGTGGTTAAAG TTGAAGTGAGCATTACCCCGCTGGG (TAA)
luxs iti_ 0.4_ 1100	98	MSEDEIREFWKIFYEA LDAGDADTASALFPDG TKIYLWDGKVFHTQAE FREWFVKLYSTSENAK REITRFEVDGNVANIE VVLEANEENGEQKKVKL RHIAEFEGDKLVEVNV EIEPL	1.21 4927 436	298	(ATG)AGCGAAGATGAAATTCGCGAATTTTGGG AAAAGTTTTATGAAGCGCTGGATGCGGGTGATG CGGATACCGCGAGCGGCTGTGTTTCTGATGGCA CCAAAATTTATCTGTGGGATGGCAAGTGTTTC ATACCCAGGCGGAATTCGCGAATGGTTTGTGA AACTGTATAGTACCGAGCGAAATGCGAAACGTG AAATCACCCGCTTTGAAGTGGATGGTAACGTGG CGAACATTGAAGTTGTGCTGGAAGCGAATGAAA ACGGTGAACAGAAGAAAGTTAACTGCGTCATA TTGCCGAATTTGAAGCGGATAAACTGGTGGAAG TGAACGTGGAAATTGAACCGCTGGG (TAA)
luxs iti_ 0.4_ 1111	99	MSEKEIREFVERFYKA LDDGDAETASSLFPDG TRIYLWDGTEFSTQAE FRAWFEELHSRSEAAAR RRVVRLEVDGDEADVE VVLDADEFEGEHHRVRL RHRFFFEGDRLVEVEV TIEPL	7.23 2895 646	299	(ATG)AGCGAAAAAGAAATTCGCGAATTTGTTG AACGTTTTTATAAGCGCTGGATGATGGCGATG CGGAAACCGCGAGCAGCCTGTTTCCGGATGGCA CCCGCATTTATCTGTGGGATGGTACCGAATTTA GCACCCAGGCGGAATTCGCGCGTGGTTTGAAG AACTGCATAGCCGAGCGAAGCCGCGCGCAGAA GAGTGGTGCGTTTGAAGTTGATGGTGATGAAG CGGATGTGGAAGTGGTTCTGGATGCTGATTTTG AAGGCGAACATCATCGCGTGCGCCTGCGCCATC GCTTTTTCTTCGAAGGTGATCGCCTGGTTGAAG TCGAAGTGACCATTGAACCGCTGGG (TAA)
luxs iti_ 0.4_ 1122	100	MSEKEIREFVDKIFYKA LDSGDAETASALFPDG TNIYLWDGTVFKTKEE FRKWFVELKSTSDNAK RKVTKLEVHGNTAHVE VVLKASVNGEEKVVKL KHIFLFEGDKLREVYV EIEPL	296. 7740 152	300	(ATG)AGCGAAAAAGAAATTCGCGAATTTGTGG ATAAATTTTATAAGCGCTGGATAGCGGCGATG CCGAAACCGCTAGCGCGTGTGTTTCCGGATGGCA CCAACATTTATCTGTGGGATGGTACCGTGTTTA AAACCAAAGAAGAAATTCGCAAAATGGTTTGTG AACTGAAAAGCACCGCGATAACGCGAAACGCA AAGTGACCAAACTGGAAGTGATGGTAATACCG CGCATGTGGAAGTTGTGCTGAAAGCGAGCGTGA ACGGTGAAGAAAAAGTGGTGAAATTTGAAACATA TTTTTCTGTTTGAAGCGGATAAACTGCGTGAAG TTTATGTTGAAATCGAACCGCTGGG (TAA)

luxs iti_... 0.4_... 1125	101	MSEEEQREFWRKFYEA LDAGDAETASALFPDG TEIYLWDGTVFRTAE FRAWFRQLYSQSDNAK RRITSFVVEGNPAYVE VVLVANIKGKKEVSL KHLFEFEGSRLVRVYV EIKPL	385. 2584 658	301	(ATG)AGCGAAGAAGAACAGCGCGAATTTTGGC GCAAATTTTATGAAGCGCTGGATGCGGGCGATG CCGAAACTGCTAGCGCACTGTTTCCGGATGGCA CCGAAATTTATCTGTGGGATGGTACCGTGTTTC GCACCCGCGCGGAATTCGCGCGTGGTTTCGCC AGCTGTATAGCCAGAGCGATAACCGGAAACGCC GCATTACCAGCTTTGTGGTGAAGGCAACCGCG CGTATGTGGAAGTGGTGGTGGTTCGGAACATTA AAGGCAAGAAAGTTGAAGTGAGCCTGAAACATT TGTTTGAATTTGAAGGCGAGCCGCTGGTGGCGG TGTATGTTGAAATTAACCGCTGGG (TAA)
luxs iti_... 0.4_... 1126	102	MSEQEQRDFVARFYAA LDAGDAETASALFRDG TKIYLWDGKTFRTQAE FRSWFVELHSQSKDAK RPVVSFRVDGNVADVN VILHASVEGEERTVRL NHVFKFEGDHLVEVRV DIEPL	2.89 1499 654	302	(ATG)AGCGAACAGGAACAGCGCGATTTTGTGG CGCGCTTTTATGCGGCGCTGGATGCGGGTGATG CGGAAACTGCAAGCGCGCTGTTTCGTGATGGCA CCAAAATTTACCTGTGGGATGGCAAAACCTTTC GCACCCAGGCGGAATTCGCGAGCTGGTTTGTGG AACTGCATAGCCAGAGCAAGATGCGAAACGCC GCGTGGTGAGCTTTCGCGTGGATGGTAACGTGG CGGATGTGAACGTGATTCTGCATGCGAGCGTGG AAGGCGAAGAACGACCGTTGCGCTGAATCATG TGTTTAAATTTGAAGGTGATCATCTGGTGGAG TGCGCGTTGATATCGAACCGCTGGG (TAA)
luxs iti_... 0.4_... 1134	103	MSEEEIRAFWERFYSA LDAGDAETASSLFPDG TEIYLWDGKTFRTQEE FRAWFEELRSTADAS RHITSLKVEGNRADIE VVLRAFNGEKEKTVRL RHEAFEGDRLVRVEV RIDPL	2.27 0905 321	303	(ATG)AGCGAAGAAGAAATTCGCGCGTTTGGG AACGCTTTTATAGCGCGCTGGATGCGGCGATG CCGAAACGCGGAGCAGCCTGTTTCCAGATGGCA CCGAAATTTATCTGTGGGATGGCAAAACCTTTC GCACCCAGGAAGAATTCGCGCCTGGTTTGAAG AACTGCGCAGCACCAGCGCGATGCAAGCAGAC ATATTACCAGCCTGAAAGTGAAGGCAACCGCG CGGACATTGAAGTGGTGGTGGCGCGAACTTTA AAGGTGAAGAAAAGACCGTGGCGCTGCGCCATG AAGCGGAATTTGAAGGCGATCGCTTGGTGGCGG TGGAAGTGCGCATTGATCCGCTGGG (TAA)
luxs iti_... 0.4_... 1135	104	MSAAEQREFVDRFYKA LDAGDAETASALFPDG TVIDLWDGRTFRTAE FRAWFERLHATSTDAK REVTQMKVDGNTVDVE VVLRAVNGEEKVVKL RHQFHFEGDQLVRVRV DITPL	24.6 1161 023	304	(ATG)AGCGCGGCGGAACAGCGCGAATTTGTTG ATCGCTTTTATAAAGCGCTGGATGCGGGCGATG CTGAAACCGCAAGCGCGTTGTTCCCTGATGGCA CCGTGATTGATCTGTGGGATGGCCGACCTTTC GCACCCGCGCAGAATTCGCGCGTGGTTTGAAC GCCTGCATGCGACGACGACCGATGCGAAACCGG AAGTGACCCAGATGAAAGTGGATGGCAACACCG TGGATGTTGAAGTGGTGGTGGCGCGACCGTGA ACGGTGAAGAAAAGTGGTTAAACTGCGCCATC AGTTTCATTTTGAAGGCGATCAGCTGGTGGCGG TGCGTGTGGATATTACCCGCTGGG (TAA)
luxs iti_... 0.4_... 1139	105	MSESEIKEFVRKFYEA LDAGDADTAAALFPDG TVIKLWDGTTFTQAE FRAWFKLYSTSDAS REVTSCLKVEGDKA EVE VVLKAKINGEEKTVKL KHIFEFKGDKLVRVEV SISPL	530. 8161 714	305	(ATG)AGCGAAAGCGAAATTAAGAATTTGTGC GCAAATTTTATGAAGCGCTGGATGCGGGTGATG CCGATACCGCAGCGCACTGTTTCCGGATGGCA CCGTGATTAAACTGTGGGATGGTACCACCTTTT ATACCCAGGCGGAATTCGCGCGTGGTTTGTGA AACTGTATAGCACCAGCGATGAAGCCAGCCGCG AAGTGACCGACTGAAAGTGAAGGCGATAAAG CGGAAGTTGAAGTGGTGGTGAAGCCAAAATTA ACGGTGAAGAAAAGACCGTGAATTTGAACATA TTTTTGAATTTAAAGGCGACAACTGGTGGCGG TCGAAGTTAGCATTAGCCCGCTGGG (TAA)
luxs iti_... 0.4_... 1141	106	MSEEEQREFWKRFYEA LDAGDAETAAALFPDG TEIHLWDGKTFRTQAE FRAWFQLRSTSEAAK RKIKFEVIGNKSFEV VVLKASINGEEKVFL	2.40 4284 727	306	(ATG)AGCGAAGAAGAACAGCGCGAATTTTGGG AACGCTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACCGCAGCAGCGTTGTTTCCAGATGGCA CCGAAATTCATCTGTGGGATGGCAAAACCTTTC GCACCCAGGCTGAATTCGCGCGTGGTTTGTGC AGCTGCGCAGCACCAGTGAAGCGGCGAAACGCA

		RHYFEFEGDKLVRVYV TIKPL			AAATTATTAAATTTGAAGTGATTGGCAACAAAA GCTTTTGTGGAAGTGGTGCTGAAAGCGAGCATTA ACGGTGAAGAAAAAGAAGTGTTCCTGCGCCAT ATTTTGAATTCGAAGCGGATAAACTGGTGCGCG TGTATGTGACCATCAAACCGCTGGG (TAA)
luxs iti_ 0.4_ 1160	107	MSEREIRQFVDRFYAA LDAGDAETAAALFPDG TEIHLWDGRTFTTQAE FRAWFEKLRRASDNAR REVVDLQVDGDEADVRL VVLDAIFRGEARRVRL THRFLFEGDQVTRVEV EIRPD	1.34 8306 842	307	(ATG)AGCGAACGCGAAATTCGCCAGTTTGTGG ATCGCTTTTATGCGGCGCTGGATGCGGGCGATG CGGAAACTGCAGCAGCGTTGTTTCCAGATGGCA CCGAAATCCATCTGTGGGATGGCCGCACCTTTA CCACCCAGGCGGAATTCGCGCGTGGTTTGAAA AACTGCGTGCGCGCAGCGATAACGCGCTCGCG AAGTGGTTGATCTGCAGGTTGATGGCGATGAAG CGGATGTGCGCGTTGTGCTGGACGCGACCTTTC GCGGTGAAGCGAGAAGAGTGGCTGTGACCCATC GCTTCCTGTTTGAAGGCGATCAGGTGACCCGCG TGGAAAGTGGAAATTAGGCCGGATGG (TAA)
luxs iti_ 0.4_ 1162	108	MSEAAQREFWDRFYRA LDAGDAETASALFPDG TEIHLWDGTTFTTQAE FRAWFRDLHARSDAAR RSIVSFEVDGDDARVE VVLHANVDGQERTVRL RHEAHFEGDRLVRVHV VIQPL	3.52 9371 113	308	(ATG)AGCGAAGCGGCGCAGCGCAATTTTGGG ATCGCTTTTATCGTGCGCTGGATGCGGGCGATG CGGAAACCGCAAGCGCACTGTTTCCGGATGGCA CCGAAATCCATCTGTGGGATGGTACCACCTTTC GCACCCGTGCGGAATTCGCGCGTGGTTTCGCG ATCTGCATGCGAGAAGCGATGCGGCCCGTAGAA GCATTGTGAGCTTTGAAGTGGATGGCGATGATG CCCGTGTGGAAGTGGTGCTGCACGCCAACGTGG ACGGTCAGGAACGTACCGTGCGTCTGAGACATG AAGCGCATTTTGAAGGCGACCGCTGGTGCGCG TGCATGTGGTGATTACGCCGCTGGG (TAA)
luxs iti_ 0.4_ 1169	109	MSETAIRQFVERFYEA LDAGDAETAAALFPDG TRIYLDGTTFTTQAE FRAWFEALHATSSGAK RHVVALQVDGVDADVE VVLHADVDGKRTVHL KHTFKFRGDRIVEVEV HIQPL	786. 2695 232	309	(ATG)AGCGAAACCGCGATCCGCCAGTTTGTGG AACGCTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACTGCAGCAGCGCTGTTTCCAGATGGCA CCCGCATTTATCTGTGGGATGGTACCACCTTTC ATACCCAGGCGGAATTCGCGCGTGGTTTGAAG CCCTGCATGCGACAGCAGCGGTGCGAAACGTC ATGTGGTGGCGTTGCAGGTTGATGGCGATGTGG CGGATGTGGAAGTGGTGCTGCACGCCGATGTAG ATGGTGAAAAACGACCGTTCATCTGAAACATA CCTTTAAATTTGCTGGCGATCGCCTGGTTGAAG TCGAAGTGCAATTTACGCCGCTGGG (TAA)
luxs iti_ 0.4_ 1172	110	MSSDAQRAFVDRFYRA LDAGDAETASALFPDG TRIHLWDGTTFTTQAE FRAWFVLDLRSSENAA REVVSFDVDGDVAHVE VVLKAVIEGEEVVVRL RHVFEWEGDRLVEVYV EIDPL	571. 7076 71	310	(ATG)AGCAGCGATGCGCAGCGCGCTTTGTGG ATCGCTTTTATCGTGCGCTGGATGCGGGCGATG CCGAAACTGCGAGCGCGTTGTTTCCAGATGGCA CCCGCATTCATCTGTGGGATGGTACCACCTTTA CCACCCGCGAAGAATTCGCGCGTGGTTTGTG ATCTGCGCTCTCGCAGCGAAACGCGCGCGTG AAGTGGTGAGTTTGTGATGTGGATGGCGATGTGG CGCATGTGGAAGTTGTGCTGAAAGCGGTGATTG AAGGTGAAGAAGTCTGGTTTCGCTGCGCCATG TGTTTGAATGGGAAGGCGATCGCCTGGTTGAAG TGTATGTAGAAATTGATCCGCTGGG (TAA)
luxs iti_ 0.4_ 1177	111	MSEKDQKEFVKFYEA LDAGDAETASSLFPDG TEIHLWDGKVFHTQAE FKAWFEELYSRSDDAK RSIISFKVDGNIKVI VVLKAFVDGKLEVL EHEFKFEGDKLVEVKV KIYPL	423. 7816 171	311	(ATG)AGCGAAAAAGATCAGAAAGAATTTGTGA AGAAATTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACCGCAAGCAGCTTGTTCGGATGGCA CCGAAATCCATCTGTGGGATGGTAAAGTGTTC ATACCCAGGCGGAATTTAAAGCGTGGTTTGAAG AACTGTATAGCGCAGCGATGATGCGAAACGCA GCATTATTAGCTTCAAAGTGGACGGCAACATTG CGAAAGTGATTGTGGTGCTGAAAGCGTTTGTG ATGGTGAAAAACTGGAAGTGCTGCTGGAACATG AATTCAAATTTGAAGGCGATAAACTGGTGGAAG TAAAGTGAAAAATTTATCCGCTGGG (TAA)

luxs iti_... 0.4_... 1178	112	MSEEEQREFVRRFYDA LDAGDAETASALFPDG TQIELWDGKTFHTREE FRAWFVKLHSTSDAR REVVSFSVAGDVANVE VVLNANIRGEKKVVRL RHI FEFEGDRLVRVRV DIEPL	417. 7097 443	312	(ATG)AGCGAAGAAGAACAGCGTGAATTTGTGC GCCGCTTTTATGATGCGCTGGATGCGGGCGATG CGGAAACTGCAAGCGCGCTGTTTCCAGATGGCA CCCAGATTGAACTGTGGGATGGCAAAACCTTTC ATACCCGTGAAGAATTCGCGCGTGGTTTGTGA AATTGCATAGCACCTCGGATGATGCCCGCCGCG AAGTGGTGAGCTTTAGCGTGGCAGGTGATGTGG CGAACGTGGAAGTTGTGCTGCGTGCGAACATTC GCGGCGAAAAAGAAAGTGGTTGCGCTGCGCCATA TTTTTGAATTCGAAGGCGATCGCCTGGTGCGCG TGCGTGTGGATATTGAACCGCTGGG (TAA)
luxs iti_... 0.4_... 1186	113	MSEDEIREFVARFYAA LDAGDANTASALFPDG TVIHLWDGITFHTQAE FRAWFEKLHSTSENAS REVVS LKVDGNVAHVE VVLRASVDGEERTVRL RHVFRFEGDKLVEVSV SITPL	732. 0027 643	313	(ATG)AGCGAAGATGAAATTCGCGAATTTGTGG CGCGCTTTTATGCGGCGCTGGATGCGGGTGATG CCAACTGCAAGCGCGCTGTTTCCCTGATGGCA CCGTGATTCATCTGTGGGATGGTATTACCTTTC ATACCCAGGCGGAATTCGCGCGTGGTTTGAAA AACTGCATAGCACCGAGCGAAAAACGCGAGCCGCG AAGTGGTGAGCTTGAAGTGGATGGCAACGTGG CGCATGTGGAAGTTGTTCTGCGCGGAGCGTGG ACGGTGAAGAACGACGGTTCGTCTGCGTCATG TGTTTCGCTTTGAAGGCGATAAACTGTTGAAG TGAGCGTGAGCATTACCCGCTGGG (TAA)
luxs iti_... 0.2_... 4	114	MSEEEIREFWKRFYEA LDAGDAETASALFPDG TRIYLWDGKEFTTQAE FRAWFEELYSTSENAS REIVKLEVEGNVAYVE VVLKANINGEEKVVKL KHVFHFEGDRLVEVEV EIEPL	6.06 9108 5	314	(ATG)TCAGAAGAAGAATTCGCGAATTTGGGA AACGCTTTTATGAAGCGTTAGATGCGGGTGATG CGGAAACCGCTAGTGCCCTGTTTCCGGATGGCA CCCGCATTTATCTGTGGGATGGTAAAGAATTTA CCACCCAGGCGGAATTCGCGCGTGGTTTGAA AACTGTATAGCACCGAGCGAAGATGCTAGCCGCG AAATGTGAACTGGAAGTGAAGGCAACGTGG CGTATGTGGAAGTTGTGCTGAAAGCGAACATTA ACGGCGAAGAAAAAGTGGTTAACTGAAACATG TGTTTCATTTTGAAGGCGATCGCCTGGTTGAAG TCGAAGTTGAAATTAACCGCTGGG (TAA)
luxs iti_... 0.2_... 16	115	MSAEQRRFVDRFYAA LDAGDADTASALFPDG TEIHLWDGRFTFTRAE FRAWFRELPARSDNAR REVVAFEVDGDTAHVE VVLRASIDGEERVVRL RHTFYFEGDRLVRVEV EIEPL	2569 .767 795	315	(ATG)AGCGCGGAAGCGCAGCGCGCTTTGTGG ATCGCTTTTATGCGGCGCTGGATGCGGGCGATG CGGATACCGCAAGTGCGCTGTTTCCAGATGGCA CCGAAATTCATCTGTGGGATGGCCGACCTTTC GCACCCGCGCGAATTCGCGCATGGTTTCGTG AACTGCGTGCGCGTAGCGATACGCGCGCGCTG AAGTGGTGCGCTTTGAAGTTGATGGCGATACGG CGCATGTGGAAGTTGTGCTGCGCGGAGCATTG ATGGTGAAGAACGCGTGGTGCGCCTGCGTCATA CCTTTTATTTTGAAGGTGATCGCCTGGTGCGTG TTGAAGTCGAAATTAACCGCTGGG (TAA)
luxs iti_... 0.2_... 37	116	MSEEEQREFVDRFYAA LDAGDAETASALFPDG TKIYLWDGKVFTTREE FRAWFEKLYSTSENAK RHVVSFKVDGNKADVE VVLHANINGEKKTVRL RHVFFYFEGDKLVEVKV EIKPL	1702 .834 14	316	(ATG)AGCGAAGAAGAACAGCGCAATTTGTTG ATCGCTTTTATGCGGCGCTGGATGCGGGCGATG CAGAAACCGCGAGCGCATTTGTTTCCCTGATGGCA CCAAAATTTATCTGTGGGATGGCAAAGTGTTTA CCACCCGTGAAGAATTCGCGCGTGGTTTGAAA AACTGTATAGCACCGAGCGAAAAACGCGAAACGCC ATGTTGTGAGCTTTAAAGTGGATGGTAACAAAG CGGATGTGGAAGTGGTGCTGCATGCGAACATTA ACGGCGAAAAAGAAACCGTGCGCCTGCGCCACG TGTTTTATTTTGAAGGCGATAAACTGGTTGAAG TGAAAGTTGAAATTAACCGCTGGG (TAA)
luxs iti_... 0.2_... 38	117	MSEEEQREFVKRFYEA LDAGDAETASALFPDG TEIHLWDGKTFYTREE FRAWFEKLYSTSDNAS RSVTEFKVDGNKAKVK VVLKANINGEKKTVKL	448. 5729 095	317	(ATG)AGCGAAGAAGAACAGCGTGAATTTGTGA AACGCTTTTATGAAGCGTTAGATGCGGGCGATG CCGAAACCGCGAGCGCATTTGTTTCCGGATGGCA CCGAAATTCATCTGTGGGATGGTAAACCTTTT ATACCCGCGAAGAGTTTCGCGCGTGGTTTGAAA AACTGTATAGCACCGAGCGATAACGCGAGCCGTA

		EHYFEFEGDKLVRVNV TIKPL			GCGTGACCGAATTTAAAGTGGATGGGAACAAAG CGAAAGTGAAAGTGGTGCTGAAAGCCAAACATTA ACGGCGGAAAAGAAAACCGTGAAACTGGAACATTT ATTTTGAATTCGAAGGCGGATAAACTGGTGCGCG TGAACGTGACCATTAAACCGCTGGG (TAA)
luxs iti_ 0.2_ 39	118	MSEEEQRRFVERFYAA LDAGDADTASALFPDG TEIHLWDGTVFRTRAE FRAWFERLYSTSENAK RHVVRFVEGNNVARVE VVLHANIDGEERTVRL THVFEFEGDRLVRVNV TINPL	2.42 2944 022	318	(ATG)AGCGAAGAAGAAGACAGCGCCGCTTTGTGG AACGCTTTTATGCGGCGCTGGATGCGGGTGATG CGGATACCGCAAGCGCATTGTTTCCGGATGGCA CCGAAATTCATCTGTGGGATGGTACCGTGTTTC GCACCCGCGCGGAATTTGCGCGCTGGTTTGAAC GCCTGTATAGTACCAGCGAAAACGCGAAACGCC ATGTGGTGCGCTTTGAAGTGAAGGCCAACGTGG CGCGCGTTGAAGTTGTGCTGCATGCGAACATTG ATGGTGAAGAAGCAGCGCTGCGCCTGACCCATG TGTTTGAATTTGAAGGCGACCGCTGGTGCGTG TGAACGTGACCATTAAACCGCTGGG (TAA)
luxs iti_ 0.2_ 65	119	MSEEEIKEFVKRFYEA LDAGDAETASALFPDG TRIYLWDGTVFRTRAE FRAWFVELHSTSEDAK REVIELKVEGNVAKVK VVLHANINGEKKTVLL EHYFEFEGDRLVEVRV EIKPL	1532 .327 574	319	(ATG)AGTGAAGAAGAGATTAAAGAAATTTGTGA AACGCTTTTATGAAGCCCTGGATGCGGCGGATG CGGAAACCGCGAGCGCTTTGTTTCCAGATGGCA CCCGCATTTATCTGTGGGATGGTGCCTGTTTC GCACCCGTGCGGAATTTGCGCGATGGTTTGTGG AACTGCATAGCACCAGCGAAGATGCGAAACGCG AAGTGATTGAACTGAAAGTGAAGGCCAACGTGG CGAAAGTGAAAGTTGTGCTGCATGCCAACATCA ACGGCGAAAAGAAAACCGCTTCTGCTGGAACATT ATTTTGAATTTGAAGGCGATCGCCTGGTGGAAG TGCGCGTGGAATTTAAACCGCTGGG (TAA)
luxs iti_ 0.2_ 66	120	MSEEAQKEFVKRFYEA LDAGDADTASALFPDG TEIYLWDGKTFHTKAE FKAWFVKLKSTSDNAK RSVVKFEVDGNVAYVE VVLHANINGEEKVVKL THIFEFEFGDKLVKVV TIKPL	11.0 4077 402	320	(ATG)AGCGAAGAAGCGCAGAAAGAAATTTGTGA AGAAATTTTATGAAGCGCTGGATGCGGGCGATG CGGATACCGCATCGGCTTTGTTTCCGGATGGCA CCGAAATTTATCTGTGGGATGGTAAACCTTTC ATACCAAAGCGGAATTTAAAGCGTGTTTGTGA AACTGAAAAGCACCAGCGATAACGCCAAACGCA GCGTTGTGAAATTTGAAGTGGATGGGAACGTGG CGTATGTGGAAGTGGTGCTGCATGCGAACATTA ACGGTGAAGAGAAAGTGGTCAAACCTGACCCATA TTTTTGAATTCGAAGGCGATAAATTAGTGAAAG TGAACGTGACCATTAAACCGCTGGG (TAA)
luxs iti_ 0.2_ 73	121	MSEEEIREFVRRFYEA LDAGDAETASALFPDG TVIYLWDGKTFHTREE FRAWFVELRSKSENAK RHVVSRLRVDGNVADVE VVLDAADINGEKKTVKL RHEFRFEGDRLVEVRV EIEPL	246. 8369 039	321	(ATG)AGCGAAGAAGAAATTCGCGAATTTGTGC GCCGCTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACTGCGAGCGCATTGTTTCCCTGATGGCA CCGTGATTTATCTGTGGGATGGCAAAACCTTTC ATACCCGTGAAGAATTTCTGCGTGTTTGTGG AACTGCGCAGCAAAAGCGAAAACGCGAAACGCC ATGTGGTGAGCCTGCGCGTGATGGTAATGTGG CGGATGTGGAAGTGGTGCTGGACGCGGATATTA ACGGCGAAAAGAAAACCGTGAAACTGAGACATG AATTTAGATTTGAAGGCGATCGCCTGGTTGAAG TGCGCGTCGAAATTGAACCGCTGGG (TAA)
luxs iti_ 0.2_ 76	122	MSEEEIREFVRRFYEA LDAGDADTASALFPDG TKIYLWDGKTFSTQAE FKAWFVKLKSTSENAK RKIVKLVKVDGVAEVE VVLHANVNGEEKVVKL KHKFKFEGDKLVEVEV EIKPL	211. 3918 452	322	(ATG)AGCGAAGAAGAAATTCGCGAATTTGTGA AACGTTTTTATGAAGCGCTGGATGCGGGTGATG CGGATACCGCATCTGCGCTGTTTCCGGATGGCA CCAAAATTTACCTGTGGGATGGTAAACCTTTA GCACCCAGGCGGAATTTAAAGCCTGGTTTGTGA AACTGAAAAGCACCAGCGAAAACGCCAAACGCA AAATTGTGAAGCTGAAAGTGGATGGCGATGTGG CGGAAGTGGAAGTTGTGCTGCATGCGAACGTGA ACGGTGAAGAAAAAGTGGTGAAATTTGAAACATA AATTCAAATTTGAAGGCGATAAACTGGTTGAGG TTGAAGTTGAAATTTAAACCGCTGGG (TAA)

luxs iti_... 0.2_... 83	123	MSEEEQREFWKR FYEA LDAGDADTASALFPDG TKIHLWDGKTFTTQAE FRAWFVELHSTSDDAK REIVKFEVDGNKSYVE VVL RANINGEEKVVRL THETQFEGDKLVEVKV TIEPL	1.21 1472 011	323	(ATG)AGCGAAGAAGAAGCAGCGGAATTTTGGGAACGCTTTTATGAAGCGCTGGATGCGGGTGATGCGGATACCGCGAGCGCTTGTTCAGATGGCA CCAAATTCATCTGTGGGATGGCAAAACCTTTA CCACCCAGGCGGAATTCGCGCGTGGTTGTG AACTGCATAGCACCAGCGATGATGCGAAACGCG AAATTTGTGAAATTTGAAGTGGATGGTAACAAA GCTATGTGGAAGTGGTGCTGCGCGCGAACATTA ACGGTGAAGAAAAAGTGGTTCGCTGACCCATG AAACCCAGTTTGAAGGCGATAAACTGGTTGAAG TTAAAGTGACCATTGAACCGCTGGG (TAA)
luxs iti_... 0.2_... 98	124	MSEEAQREFVRRFYAA LDAGDAETASALFPDG TVIHLWDGRTRFTQAE FRAWFVELYSKSEDAK REVVRFEVDGDRARVE VVLKANFKGKKEVVKL VHYFLFEGDRLVEVRV EIKPL	184. 6724 257	324	(ATG)AGTGAAGAAGCGCAGCGGAATTTGTGC GCCGCTTTTATGCGCGCTGGATGCGGGTGATG CGGAAACAGCAGCAGCGTTGTTCGGATGGCA CCGTGATTCTCTGTGGGATGGCCGACCTTTC GCACCCAGGCGGAATTCGCGCTGGTTGTGG AACTGTATAGCAAAAGCGAAGATGCGAAACGTG AAGTGGTGCGTTTGAAGTTGATGGCGATCGCG CGCGTGTGGAAGTTGTGCTGAAAGCCAATTTTA AAGGCAAGAAAAAGTGGTGAACCTGGTGCAAT ATTTTCTGTTTGAAGGCGATAGACTGGTTGAAG TGCGCGTGGAAATTAACCGCTGGG (TAA)
luxs iti_... 0.2_... 100	125	MSAEAQREFVRRFYAA LDAGDAETASALFPDG TRIHLWDGRVFTREE FRAWFVKLYSTSEDAK REVTAFRVDGDRADVE VVL RASINGEEKTVRL RHVFQFEGDRLPREVRV AIEPL	746. 0829 302	325	(ATG)AGCGCGGAAGCGCAGCGGAATTTGTGC GCCGCTTTTATGCGCGCTGGATGCGGGTGATG CGGAAACCGCAAGCGCGCTGTTTCAGATGGCA CCCGATTCTCTGTGGGATGGCCGCGTGTTC GCACCCGTGAAGAATTCGCGCGTGGTTGTGA AACTGTATAGCACCAGTGAAGATGCGCGCCGCG AAGTGACCGCGCTTTCGTGTTGATGGCGATCGTG CCGATGTGGAAGTGGTGCTGCGCGCGAGCATTA ACGGCGAAGAAAAAGCCGTGCGCCTGCGTCATG TGTTCAGTTTGAAGGGGATCGTCTGCGTGAAG TGCGCGTCCGCTTGAACCGCTGGG (TAA)
luxs iti_... 0.2_... 108	126	MSEEEIREFCKR FYEA LDAGDADTASALFPDG TEIYLWDGRTRFTREE FRAWFVELHSTSDDAK RSIVKLEVEGNVAYVE VVL RANVDGEEKVVRL VHIFYFEGDKLVKVYV SIRPL	64.7 8921 907	326	(ATG)AGCGAAGAAGAAATTCGCGAATTTTGCA AACGCTTTTATGAAGCGCTGGATGCGGGCGATG CGGATACCGCAAGTGCCTGTTTCAGATGGCA CCGAAATTTATCTGTGGGATGGCCGACCTTTC GCACCCGTGAAGAATTCGCGCGTGGTTGTGG AACTGCATAGCACCAGCATGATGCCGCGCGCA GCATTGTGAACTGGAAGTGAAGGCAACGTGG CGTATGTGGAAGTTGTGCTGCGCGCAACGTGG ATGGGGAAGAAAAAGTGGTGCCTGGTGATA TTTTCTATTTTGAAGGCGATAAACTGGTGAAG TGTATGTGAGCATTGCCCGCTGGG (TAA)
luxs iti_... 0.2_... 118	127	MSEEEIREFWKK FYEA LDAGDAETASALFPDG TKIHLWDGKTENTQAE FKAWFVKLYSESDDAK REIVELEV DGNVAYVK VVLHANYKGEQKTVEL KHIFKFEVDKLVVNV EIKPL	106. 1672 426	327	(ATG)AGCGAAGAAGAAATTCGCGAATTTTGGGA AAAAGTTTTATGAAGCGCTGGATGCGGGTGATG CGGAAACCGCGAGCGCACTGTTTCTGATGGTA CCAAATTCATCTGTGGGATGGCAAAACCTTTA ACACCCAGGCGGAATTTAAAGCGTGGTTGTGA AACTGTATAGCGAAAGCGATGATGCGAAACGCG AAATTTGTGGAAGTGAAGTGGATGGTAACGTGG CGTATGTGAAAGTGGTGCTGCATGCGAACTATA AAGGCGAACAGAAAACCGTTGAACGTGAACATA TTTTAAATTTGAAGGCGATAAACTGGTGGAG TTAACGTTGAAATTAACCGCTGGG (TAA)
luxs iti_... 0.2_... 119	128	MSEEEQREFWKK FYEA LDAGDAETASALFPDG TKIYLWDGKVFTQEE FRAWFVELRSQSDDAK RKIVDFKVEGDVAYVE VVLEANINGEKKTVKL	1059 .310 297	328	(ATG)AGCGAAGAAGAAGCAGCGGAATTTTGGGA AAAAGTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACCGCAAGCGCTTGTTCAGATGGCA CCAAATTTATCTGTGGGATGGCAAGTGTCT ATACCCAGGAAGAAATTCGCGCGTGGTTGTGG AACTGCGCAGCCAGCGATGATGCGAAACGCA

		KHIYKFEGDKLVEVKV KIEPL			AAATTGTGGATTTTAAAGTGAAGGCGATGTGG CGTATGTGGAAGTGGTGTGGAAGCGAACATTA ATGGCGAAAAAGAAACCGTGAAACTGAAACATA TTTATAAATTCGAAGGTGATAAACTGGTTGAAG TGAAAGTTAAATCGAACCGCTGGG (TAA)
luxs iti_ 0.2_ 123	129	MSEEEQREFWRRFYEA LDAGDAETASALFPDG TKIHLWDGRTFTTQEE FRAWFVDLHSRSDDAK REIVSFKVEGNKARVE VVLRAEDGGEERVVAL RHETEFEGDRLVEVNV DIRPL	2.00 8293 02	329	(ATG)AGCGAAGAAGAACAGCGCGAATTTTGGC GCCGCTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACCGCCTCTGCACTGTTTCCAGATGGCA CCAAAATTCATCTGTGGGATGGCCGCACCTTTA CCACCCAGGAAGAATTTGCGCGCTGGTTTGTGG ATCTGCATAGCCGAGCGATGATGCGAAACGCG AAATTGTGAGCTTTAAAGTGAAGGCAACAAAG CGCGCGTTGAAGTGGTGTGCGCGCACGTGAAG ATGGTGAAGAAGCGCTGTGGCGCTGCGTCATG AAACCGAATTTGAAGGCGATCGCCTGGTGAAG TGAAAGTGAATATTCGCCCGCTGGG (TAA)
luxs iti_ 0.2_ 125	130	MSEEAQREFVRRFYAA LDAGDADTASALFPDG TRIYLWDGRTFTTRA FRAWFERLHATSADAK REVVDVEVEGNKAKVK VVLHANVNGEEKTVLL EHWFEFEGDRLVRVEV TIKPL	12.3 0131 306	330	(ATG)AGCGAAGAAGCGCAGCGTGAATTTGTGC GCCGCTTTTATGCGCGCTTGGATGCGGGTGATG CGGATACCGCTAGCGCCTTATTTCCGGATGGCA CCCGCATTTATCTGTGGGATGGCCGCACCTTTA CCACGCGCGCGGAATTTGCGCGATGGTTTGAAC GCCTGCATGCGACGAGCGAGATGCGAAACGGG AAGTGGTGGATTTTGAAGTGAAGGCAACAAAG CGAAAGTGAAGTGGTCTGCAACGCGAACGTGA ACGGTGAAGAAAAGACCGTGTGCTGGAACATT GGTTCAATTTGAAGGCGATCGCCTGGTGGCGG TGGAAGTGACCATTAACCGCTGGG (TAA)
luxs iti_ 0.2_ 126	131	MSEKEQREFWKKFYEA LDAGDAETASALFPDG TKIYLWDGKTFTTQAE FRAWFVELRSTSENAK RKIVSFKVDGNKSEVE VVLEANINGEKVVVKL KHIYKFEGDKLVEVKV EIIPL	159. 0704 907	331	(ATG)AGCGAAAAAGAACAGCGCGAATTTTGGG AAAAGTTTTATGAAGCGTGGATGCGGGTGATG CCGAAACCGCGAGCGCGCTGTTTCCAGATGGCA CCAAAATTTATCTGTGGGACGGCAAAACCTTTA CCACCCAGGCGAATTTGCGCGCTGGTTTGTGG AACTGCGCAGCACCGAGCGAGAACGCCAAACGCA AAATTGTGAGCTTTAAAGTGGATGGCAACAAAA GCGAAGTGAAGTGGTCTGGAAGCGAACATCA ACGGCGAAAAAGTGGTTAAACTGAAACACA TTTTTAAATTTGAAGGCGATAAACTGGTTGAAG TGAAAGTTGAAATTATCCGCTGGG (TAA)
luxs iti_ 0.2_ 127	132	MSEEEQREFWKKFYEA LDAGDADTASALFPDG TEIHLWDGKTFRTREE FRAWFEELYSTSEDAK REIVKFEVEGNKSFEV VVLKANVNGKKVVVKL RHVFEFEGDKVVKVEV EIEPL	1.08 7076 71	332	(ATG)TCCGAAGAAGAACAGCGCGAATTTTGGC GCAAATTTTATGAAGCGCTGGATGCGGGTGATG CGGATACCGCTCTGCGCTGTTTCCAGATGGCA CCGAAATTCATCTGTGGGATGGCAAAACCTTTC GCACCCGCGAAGAATTTGCGCGCTGGTTTGAAG AACTGTATAGCACCAGCGAAGATGCGAAACGCG AAATTGTGAAATTTGAAGTGAAGGCAACAAAA GCTTTGTGGAAGTGGTGTGAAAGCGAACGTCA ACGGCAAGAAAGTGGTTGTCAAACCTGCGCCATG TGTTTGAATTCGAAGGCGATAAAGTTGTGAAGG TTGAAGTTGAAATTGAACCGCTGGG (TAA)
luxs iti_ 0.2_ 132	133	MSEEEIKKFWEKFYNA LDAGDADTASALFPDG TEIHLWDGKTFKTQAE FREWFVKLYSTSDDAK REIVKLEVDGNVAYVE VVLKANVDGEEKTVKL RHVAVFEGDKLKEVKV TIKPL	1.70 2833 449	333	(ATG)AGCGAAGAAGAAATTAAGTTTGGG AAAAATTTTATAACGCCCTGGATGCGGGTGATG CGGATACTGCAAGCGCTCTGTTTCCGACCGGTA CCGAAATTCATCTGTGGGATGGCAAAACCTTTA AAACCCAGGCGGAATTTGCGGAATGGTTTGTGA AACTGTATAGCACCAGCGATGATGCGAAACGCG AAATTGTTAAACTGGAAGTGGATGGTAACGTGG CGTATGTGGAAGTTGTGTGCGCGCAACGTGG ACGGTGAAGAAAAGACCGTAAACTGCGCCATG TGGCGGTGTTTGAAGGCGATAAACTGAAAGAAG TGAAAGTGACCATTAACCGCTGGG (TAA)

luxs iti_... 0.2_... 142	134	MSEEEQREFVRRFYEA LDAGDAETASALFPDG TKIHLWDGKTFSTRAE FRAWFVELRSTSENAK RRVTSFRVEGNEADVE VVLEATVNGEKKRVRL RHRFLFEGDRLVEVYV EIKPL	100. 1679 337	334	(ATG)AGCGAAGAAGAACAGCGCGAATTTGTGC GCCGCTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACTGCGAGCGCGTTGTTCCAGATGGCA CCAAAATTCATCTGTGGGATGGCAAAACCTTTA GCACCCGCGCGGAATTCGTGCGTGGTTGTGG AACTGCGTAGCACCAGCGAAAACGCCAAACGCC GCGTGACCAGCTTTCGTGTGAAGGCAACGAAG CGGATGTTGAAGTGGTGTGGAAGCGACCGTGA ACGGCGAAAAGAAACGCGTGCCTGCGTCACC GCTTTCTGTTTGAAGGCGATCGCCTGGTGAAG TGTATGTGGAAATTAACCGCTGGG (TAA)
luxs iti_... 0.2_... 144	135	MSEEEQREFVRRFYEA LDAGDAETASALFPDG TLIYLWDGKTFTTQAE FRAWFEKLRSTSEDAK REVVEFKVDGNVADV VVLEANINGEKKVVKL RHRFHFEKDKLVKEV EIEPL	257. 4740 843	335	(ATG)AGCGAAGAAGAACAGCGCGAATTTGTGA AACGTTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACTGCAAGCGCACTGTTCCGGATGGCA CCCTGATTTATCTGTGGGATGGGAAAACCTTTA CCACCCAGGCGGAATTCGCGCATGTTTGA AACTGCGCAGCACCAGCGAGGATGCGAAACGCG AAGTGGTGGAAATTTAAAGTTGATGTTAACGTGG CGGATGTGGTGGTTGTGCTGGAAGCGAACATTA ACGGCGAAAAGAAAGTGGTTAAGCTGCGTCATC GCTTTTCATTTTGAAGGCGATAAACTGGTCAAAG TGGAAAGTTGAAATTAACCGCTGGG (TAA)
luxs iti_... 0.2_... 147	136	MSEEEQREFVRRFYEA LDAGDADTASALFPDG TKIHLWDGTTFTTQAE FKAWFEKLYSKSDNAK RHVVSPKVEGNVAYVE VVLHANFKGEEKTVSL KHIFKFEKDKLVEVEV KIKPL	376. 0767 104	336	(ATG)AGCGAAGAAGAACAGCGCGAATTTGTAA AACGCTTTTATGAAGCGTTGGATGCCGCTGATG CGGATACCGCGAGCGCGTTGTTCCGGATGGCA CCAAAATTCATCTGTGGGATGGCAAAACCTTTA CCACCCAGGCGGAATTTAAAGCGTGGTTTGA AACTGTATAGCAAAAGCGATACGCCAAACGCC ATGTGGTGAGCTTTAAAGTGGGAAGGCAACGTGG CGTATGTGGAAAGTGGTGTGCTGCATGCCAATTTA AAGGTGAAGAAAAGACCGTGAGCCTGAAACATA TCTTTAAATTTGAAGGCGATAAACTGGTTGAAG TCGAAGTGAATTAACCGCTGGG (TAA)
luxs iti_... 0.2_... 149	137	MSEEEQREFVRRFYEA LDAGDAETASALFPDG TVIHLWDGKTFTTQAE FRAWFVELKSTSEDAK RRVVSFRVDGDEAEV VVLHANIDGEERVRL THRFLFEGDRVVEVWV EIEPL	3.70 6288 874	337	(ATG)AGCGAAGAAGAACAGCGCGAATTTGTGC GCCGTTTTTATGAAGCGCTGGATGCGGGTGATG CGGAAACTGCGAGCGCACTGTTTCTGATGGCA CCGTGATTCATCTGTGGGATGGCAAAACCTTTC GCACCCAGGCGGAATTCGCGCGTGGTTGTGG AACTGAAAAGCACCAGCGAGGATGCGAAACGTC GCGTTGTGAGCTTTCGTGTGGATGGCGATGAAG CCGAAGTGGTGGTTGTGCTGCATGCGAACATTG ATGGTGAAGAACGCGTGTGCGCCTGACCCATC GCTTTCTGTTTGAAGGTGATCGCGTAGTGAAG TGTGGGTGGAAATTAACCGCTGGG (TAA)
luxs iti_... 0.2_... 158	138	MSEEQREFVDRFYRA LDAGDADTASALFPDG TEIYLWDGKTFTTQAE FREWFKLHSTSEDAK RRVVSFSVDGNVADVE VVLEANVNGEKKTVKL KHRFHFEKDKVVRVEV SIEPL	1.76 7104 354	338	(ATG)AGCGAACAGGAACAGCGCGAATTTGTGG ATCGCTTTTATCGTGCGCTGGATGCGGGTGATG CGGATACCGCGAGCGCACTGTTTCTGATGGTA CCGAAATTTATCTGTGGGATGGCAAAACCTTTC GCACCCAGGCGGAATTCGCGAATGTTTGTGA AACTGCATAGCACCAGCGAAGATGCCAAACGCC GCGTGGTGAGCTTTAGCGTGGATGGTAACGTGG CGGATGTGGAAGTGGTGTGGAAGCGAATGTGA ACGGCGAAAAGAAAACCGTCAAACGTAAACATC GCTTCCATTTTGAAGGCGATAAAGTGGTTCGCG TTGAAGTGAGCATTGAACCGCTGGG (TAA)
luxs iti_... 0.2_... 161	139	MSEEEIREFVRRFYAA LDAGDAETASALFPDG TRIHLWDGRTFTTQAE FREWFKLYSESDDAR REVVSLEVDGDPALVR VVLRSYKGEERVVEL	397. 2411 887	339	(ATG)AGCGAAGAAGAAATTCGCGAATTTGTGC GCCGCTTTTATGCGGGCGCTGGATGCGGGTGATG CGGAAACTGCAAGCGCGTTGTTTCCGGATGGCA CCCGCATTCATCTGTGGGATGGCGCACCTTTA CCACCCAGGCGGAATTCGTGAATGGTTGTGA AACTGTATAGCGAAAGCGATGATGCGCGCCGCG

		RHEFLFEGDELVEVRV EIRPL			AAGTGGTGAGCCTGGAAGTGGATGGCGATCGTG CATTGGTGCGCGTGGTCTGCGTGCAAGCTATA AAGGTGAAGAACGCGTCTGGAACTGCGCCATG AATTTCTGTTTGAAGGCGATGAACTGGTGGAAAG TTCGCGTGGAAATTAGACCGCTGGG (TAA)
luxs iti_ 0.2_ 164	140	MSEEEQREFVKRFYEA LDAGDAETASALFPDG TEIHLWDGTVFRTRAE FRAWFEQLYAESDDAK REIVSFKVEGNVSDVE VVLHASYGKEKKTVKL RHRFFFEGRDLVEVEV EIEPL	174. 2522 46	340	(ATG)AGCGAAGAAGAACAGCGCAATTTGTGA AACGCTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACTGCAAGCGCGCTGTTTCCAGATGGCA CCGAAATTCATCTGTGGGATGGTACCGTGTTTC GCACCGCGCGGAAATTTCGCGCGTGGTTTGAAC AGCTGTATGCCGAAAGCGATGATGCCAAACGCG AAATTGTGAGCTTTAAAGTGAAGGCCAACGTTA GCGATGTGGAAGTGGTGTGCATGCGAGCTATA AAGGCGAAGAAACCGTGAACCTGAGACATC GCTTTTCTTTTGAAGGCGATCGTCTGGTTGAAG TCGAAGTTGAAATTGAACCGCTGGG (TAA)
luxs iti_ 0.2_ 165	141	MSEEEQREFVKRFYEA LDAGDAETASALFPDG TEIHLWDGITFHTQAE FREWFVKLRSTSENAK REVVKFEVDGNKAKVE VVLKANINGEKEVKL THTFEFEGDKLVKVVN DIKPL	1.13 4070 491	341	(ATG)AGCGAAGAAGAACAGCGTGAATTTGTGA AACGCTTTTATGAAGCGCTGGATGCGGGTGATG CGGAAACCGCAAGCGCGCTGTTTCCAGATGGCA CCGAAATTCATCTGTGGGATGGCATTACCTTTC ATACCCAGGCGGAATTTCCGCAATGGTTTGTGA AACTGCGCAGCACCAGCGAAGAACGCGAAACGCG AAGTGGTGAAATTTGAAGTTGATGGTAACAAAG CGAAAGTGAAGTTGTGCTGAAAGCAAACATTA ACGGTGAAGAAAAAGTGAACCTGACCCATA CCTTTGAATTCGAAGGCGATAAATTAGTGAAG TGAACGTGGATATTAAACCGCTGGG (TAA)
luxs iti_ 0.2_ 173	142	MSAEEIREFVRIFYEA LDAGDADTASALFPDG TEIHLWDGRTFRTQAE FRAWFVRLRATSDAS REVVSLEVDGNVADVE VVLHANVNGEKKVVRL RHRFYFEGDRLVRVEV TIVPL	25.3 1720 802	342	(ATG)AGCGCGGAAGAAATTCGCGAATTTGTGC GTGCGTTTTATGAAGCGCTGGATGCGGGTGATG CCGATACCGCGAGTGCGCTGTTTCCATGATGGCA CCGAAATTCATCTGTGGGATGGCCGACCTTTC GCACCCAGGCGGAATTTCCGCGATGGTTTGTGA GGCTGCGCGCGACAAAGCGATGATGCGAGCCGTG AAGTGGTGAGCTTGAAGTGGATGGCAACGTTG CGGATGTGGAAGTTGTGCTGCATGCGAACGTGA ACGGCGAAGAAAGTGGTTCGCTGCGCCATC GCTTCTATTTTGAAGGCGATGCGCTGGTGCGCG TTGAAGTGACCATTGTGCGCTGGG (TAA)
luxs iti_ 0.2_ 176	143	MSKEEIKEFVKRFYQA LDAGDAETASSLFPDG TKIYLWDGKVPKQEE FRAWFEKLHSTSENAK REVTKLEVEGNVAYIE VVLKANVNGEEKIVNL THKFYFEGDKLVEVEV KIVPL	127. 8949 551	343	(ATG)AGCAAAGAAGAAATTAAGAAATTTGTGA AACGCTTTTATCAGGCGCTGGATGCGGGCGATG CGGAAACAGCGAGCAGCCTGTTTCCAGATGGCA CCAAATTTATCTGTGGGATGGCAAGTGTTTA AAACCCAGGAAGAGTTTCGCGCGTGGTTTGAAA AACTGCATAGCACCAGCGAAGAACGCGAAACGCG AAGTGACCAAACCTGGAAGTTGAAGGCAACGTGG CGTATATTGAAGTGGTGTGAAAGCGAACGTTA ACGGCGAAGAAAAGATTGTGAACCTGACCCATA AATTTTATTTTGAAGGCGATAAAGTGGTGAAG TGGAAGTGAAATTTGTCCGCTGGG (TAA)
luxs iti_ 0.2_ 184	144	MSEEEIREFVKRFYEA LDAGDAETASALFPDG TKIHLWDGTVFTTQAE FRAWFEKLYSTSDNAS RSVVSFKVDGNVAEVE VVLKANIKGREKVVKL KHIFHEGDKLVEVEV SIEPL	246. 1610 228	344	(ATG)AGCGAAGAAGAAATCCGCAATTTGTGA AACGCTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACGCGGAGCGCATGTTTCCAGATGGCA CCAAATTCATCTGTGGGATGGTACCGTGTTTA CCACCCAGGCGGAATTTCCGCGCTGGTTTGAAA AACTGTATAGCACCAGCGATAACGCGAGCCGCA GCGTGGTGAGCTTTAAAGTGGATGGCAACGTGG CGGAAGTGAAGTTGTGCTGAAAGCGAACATTA AAGGCGCGGAAAAAGTGGTGAACCTGAAACATA TTTTTCATTTTGAAGGCGATAAAGTGGTGAAG TCGAAGTGAGCATTGAACCGCTGGG (TAA)

luxs iti_... 0.2_... 186	145	MSEEEIREFVKRFYEA LDAGDAETASALFPDG TVIHLWDGITEHTQAE FRAWFEALYSTSENAR REVVALRVENRPARVE VVLHANVDGEERTVRL RHEFLFEGDRLVEVRV EIEPL	800. 6247 408	345	(ATG)AGCGAAGAAGAAATTCGCGAATTTGTGA AACGCTTTTATGAAGCGCTGGATGCGGGTGATG CCGAAACAGCGTCCGCGCTGTTCCAGATGGCA CCGTGATTCATCTGTGGGATGGCATTACCTTTC ATACCCAGGCGGAATTCGCGCGTGGTTGAAG CCCTGTATAGCACCAGCGAAAATGCGCGCCGCG AAGTGGTGGCGTTGAGAGTGAAGGTAACCGTG CTCGTGTGGAAGTTGTGCTGCATGCGAACGTGG ATGGTGAAGAACGCACCGTTCCGCTGCGCCATG AATTTCTGTTGGAAGGCGATCGCCTGGTGAAG TGCGCGTTGAAATTGAACCGCTGGG (TAA)
luxs iti_... 0.2_... 191	146	MSEEAQREFVKRFYEA LDAGDAETASALFPDG TRIHLWDGRTFTTREE FRAWFEELRSKSEDAK REVTSFRVDGNVAEVE VVLHANINGEEKTVLL KHFVFEFGDRLVEVHV EIKPL	221. 3282 654	346	(ATG)TCCGAAGAAGCGCAGCGCGAATTTGTGA AACGCTTTTATGAAGCGCTGGATGCGGGTGATG CGGAAACCGCGCAGCGCTGTTTCCAGATGGCA CCCGCATTCATCTGTGGGATGGCCGACCTTTA CCACCCGTGAAGAATTCGCGCGTGGTTGAAG AACTGCGCAGCAAAAGCGAAGATGCGAAACGCG AAGTGACCAGCTTTCGCGTGGATGGCAACGTGG CGGAAGTGGAGTTGTGCTGCATGCCAACATCA ACGGCGAAGAAAAGACCGTGTGCTGAAACATC GCTTTGTGTTGGAAGGCGATCGCCTGGTTGAAG TGCAATGTGGAATTAACCGCTGGG (TAA)
luxs iti_... 0.2_... 193	147	MSEEAQRQFVRRFYEA LDAGDADTASALFPDG TEIHLWDGRTFRTRA FRAWFERLRATSADAR RSVTSFEVDGDFARVE VVLRA SFAGEERVVRL RHYFQFEGDRLVRVEV EIRPL	33.6 5376 641	347	(ATG)AGCGAAGAAGCCAGCGTCAGTTTGTGC GCCGCTTTTATGAAGCCCTGGATGCGGGCGATG CGGATACCGCTAGCGCACTGTTTCCAGATGGCA CCGAAATTCATCTGTGGGATGGCCGACCTTTC GCACCCGCGCGGAATTCGTGCTGCTGGTTGAAC GCCTGCGCGCGACAAGCGCAGATGCCCGCAGAA GCGTGACTTCTTTTGAAGTGGATGGCGATTTTG CGCGTGTGGAAGTGGTGTGCTGCGTGCGAGCTTG CCGGAGAAGAAGCGTGGTGGTGTGCTGCGTCATT ATTTTCAGTTGGAAGGCGATCGCCTGGTGGCGG TTGAAGTTGAAATTCGCCCGCTGGG (TAA)
luxs iti_... 0.2_... 194	148	MSKKAQEEFWKRFYEA LDAGDADTASALFPDG TKIYLWDGKTFTTQAE FRAWFVELHKSNDNAK RRITKFEVDGNKSYVE VVLEANVNGKKEVVKL VHETLFEQDKLVEVRV DIKPL	1.76 5031 099	348	(ATG)AGCAAAAAGCGCAGGAAGAATTTTGA AACGCTTTTACGAAGCGCTGGATGCGGGTGATG CGGATACCGCAAGCGCGCTGTTTCCAGATGGCA CCAAAATTTATCTGTGGGATGGCAAAACCTTTA CCACCCAGGCGGAATTCGCGCGTGGTTGTGG AACTGCATAGCAAAAGCGATACCGCAAAACGCC GCATTACCAAAATTTGAAGTGGATGGTAACAAAA GCTATGTGGAAGTGGTGTGGAAGCGAACGTGA ACGGCAAGAAAGAGTTGTGAAACTGGTGCATG AAACCCTGTTTGAAGGCGATAAATTTGTTGAAG TTCGCGTGGATATTAAACCGCTGGG (TAA)
luxs iti_... 0.2_... 198	149	MSEEEQREFVRRFYEA LDAGDAETASALFPDG TEIHLWDGKVFRTQAE FRAWFERLRATSADAK REIVSFRVDGDRADVE VVLRAVVDGEEKEVKL NHRFFFEGDRLVRVEV EIKPL	959. 2121 631	349	(ATG)AGCGAAGAAGAAGCGCGAATTTTGGC GTGCTTTTATGAAGCGCTGGATGCGGGCGATG CCGAAACTGCTAGCGCACTGTTTCCAGATGGCA CCGAAATTCATCTGTGGGATGGCAAAAGTGTTC GCACCCAGGCGGAATTCGCGCGTGGTTGAAC GTCTGCGCGCGACAAGCGCAGATGCGAAACGCG AAATTGTGAGCTTTCGCGTGGATGGCGATCGCG CGGATGTGGAAGTGGTGTGCTGCGTGCAGTTGTTG ATGGTGAAGAAAAAGAGTGAACCTGAACCATC GCTTCTTTTTCGAAGGCGATAGACTGGTGGCGG TTGAAGTGGAAATTAACCGCTGGG (TAA)
luxs iti_... 0.2_... 201	150	MSEEAIREFWKRFYEA LDAGDAETASALFPDG TEIHLWDGKTFRTRA FRAWFEELYSTSENAR REITSLKVEGNPAEVE VVLHANIEGKEKTVNL	37.8 9011 748	350	(ATG)TCCGAAGAAGCGATTGCGGAATTTTGA AACGCTTTTATGAAGCCCTGGATGCGGGCGATG CCGAAACCGCGAGCGCATTTTCCAGATGGCA CCGAAATTCATCTGTGGGATGGCAAAACCTTTC GCACCCGCGCGGAATTTAAAGCGTGGTTGAAG AACTGTATAGCACCAGCGAAAATGCCAACGTG

		KHWFEFEGDHLVRVRV DIKPL			AAATTACCAGCCTGAAAGTGGAAAGGCAACCGCG CCGAAGTTGAAGTGGTGTGCATGCGAACATTG AAGGTAAAGAAAAGACCGTGAACTGAAACATT GGTTCGAATTTGAAGGCGATCATCTGGTGCGCG TGCGGTGGATATTAAACCGCTGGG (TAA)
luxs iti_0.2_202	151	MSEEEQREFVRRFYEA LDAGDAETASALFPDG TVIHLWDGRTFRTQAE FRAWFEELYSTSEDAS RHVVSFKVDGNKARVE VVLHANINGEHTVNL THVFFHEGDKLVEVEV DIRPL	82.4 9067 035	351	(ATG)TCGGAAGAAGAACAGCGCGAATTTGTGC GCCGTTTTATGAAGCGCTGGATGCGGGCGATG CCGAAACTGCGAGCGCTCTGTTTCCAGATGGCA CCGTGATTCACTGTGGGATGGCCGCACCTTTC GCACCCAGGCGGAATTTCCGCCCTGGTTTGAAG AACTGTATAGCACCAGCGAAGATGCGTCTCGCC ATGTGGTGAGCTTTAAAGTGGATGGCAACAAAG CGCGCGTGGAGTGGTGTGCATGCCAACATTA ACGGCGAAGAGCATACCGTTAACTGACCCATG TGTTTCATTTTGAAGGCGATAAACTGGTTGAAG TCGAAGTGGACATTGCTCCGCTGGG (TAA)
luxs iti_0.2_204	152	MSEEEQREFVRRFYEA LDAGDAETASALFPDG TVIHLWDGKTFHTREE FRAWFVELKSKSEDAK REIVDFEVDGNVSRVK VVLHANYKGEEKVVEL EHVFFHEGDKLVEVEV KIKPL	133. 8949 551	352	(ATG)AGCGAGGAAGAACAGCGCGAATTTTGGGA AACGCTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACTGCGAGCGCGTGTGTTTCTGATGGCA CCGTGATTCACTGTGGGATGGCAAAACCTTCC ATACCGGTGAAGAATTTCCGCGGTGGTTTGTGG AACTGAAAAGCAAAAGCGAAGATGCGAAACGCG AAATTGTGGATTTTGAAGTTGACGGCAACGTGA GCCGCGTGAAGTGGTGTGCATGCCAACTATA AAGGCGAAGAAAAAGTTGTTGAAGTGGAAACATG TGTTTCATTTTGAAGGCGATCGCTGGTGGAAAG TCGAAGTTAAATTTAAACCGCTGGG (TAA)
luxs iti_0.2_206	153	MSEEEIREFVKRFYAA LDAGDAETASALFPDG TEIYLWDGTTTFTQAE FRAWFVELYSRSDAS REVVDLKVNDGNKAEVE VVLKASVDGEERVVKL KHFFEFEGDKLVKVEV TIEPL	126. 0138 217	353	(ATG)AGCGAAGAAGAAATTCGTGAATTTGTGA AACGCTTTTATGCGGCGCTGGATGCGGGCGATG CGGAAACCGCAAGCGCACTTTTCTGATGGCA CCGAAATTTATCTGTGGGATGGTACCACCTTTC GCACCCAGGCGGAATTTCCGCGGTGGTTTGTGG AACTGTATAGCGCAGCGATGATGCGAGCGCGG AAGTGGTGGATCTGAAGTGGATGGCAATAAAG CGGAAGTGGAAAGTTGTGCTGAAAGCGAGCGTAG ATGGTGAAGAACGCGTAGTGAAGTGAACATT TCTTTGAATTCGAAGGCGATAAACTGGTGAAG TTGAAGTGACCATTGAACCGCTGGG (TAA)
luxs iti_0.2_229	154	MSEEEQREFVKKFYEA LDAGDAETASALFPDG TEIHLWDGKTFRTQAE FRAWFEKLKSTSENAR RKIVSFKVDGNKADVE VVLKANINGEETVKL KHTFEFEGDKLKVNV EIVPL	1279 .815 48	354	(ATG)AGCGAAGAAGAACAGCGCGAATTTGTGA AGAAATTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACCGCAAGCGCGTGTGTTTCCAGATGGCA CCGAAATTCATCTGTGGGATGGTAAACCTTTC GCACCCAGGCGGAATTTCTGCGGTGGTTTGAAG AACTGAAAAGCACCAGCGAAACGCGAAACGCA AAATTGTGAGCTTTAAAGTGGATGGCAACAAAG CGGATGTGGAGTTGTGCTGAAAGCGAATATTA ACGGGGAAGAAAAGACCGTGAAATTTGAACATA CCTTTGAATTTGAAGGCGATAAGCTGAAAAGG TGAACGTGGAAATTTGTCCGCTGGG (TAA)
luxs iti_0.2_237	155	MSEEAQREFVRRFYEA LDAGDAETASALFPDG TKIYLWDGRVFETQEE FRAWFVELRSTSENAR RRVVDFFEVDGNVARVE VVLHANIDGEEKTVRL RHIKFEFGDKLVEVEV TIEPL	1.87 8369 039	355	(ATG)AGCGAAGAAGCGCAGCGCGAATTTGTGC GCCGTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACCGCAAGCGCACTGTTTCTGATGGCA CCAAAATTTATCTGTGGGATGGCCGCGTGTG AAACCCAGGAAGAATTTCCGCGGTGGTTTGTGG AACTGCGCAGCACCAGCGAAAACGCCGAGAA GGGTGGTGGATTTTGAAGTGGATGGCAATGTTG CGCGCGTTGAAGTTGTGCTGCATGCGAACATTG ATGGTGAAGAAAAGACCGTGCGCCTGCGCCATA TTTTTAAATTTGAAGGCGATAAACTGGTGAAG TCGAAGTGACCATTGAACCGCTGGG (TAA)

luxs iti_... 0.2_... 243	156	MSEEEIREFVRRFYEA LDAGDAETASALFPDG TKIYLWDGITEFTQAE FRAWFVELRSTSDAAR REVVRLEVDGNVAFVE VVLHASHRGEERTVLL RHFVQFEGDKLVKVEV SIDPL	1.30 0621 977	356	(ATG)AGCGAAGAAGAAATTCGCGAATTTGTGC GCCGCTTTTATGAAGCGCTGGATGCGGGTGTATG CGGAAACAGCAAGCGCGCTGTTTCCCGATGGCA CCAAAATTTATCTGTGGGATGGCATTACCTTTA CCACCCAGGCGGAATTCGCGCGTGGTTGTGG AACTGCGCAGCACCAGCGATGCGGCGAGAAGAG AAGTGGTGGCGTTGGAAGTGGATGGTAACGTGG CGTTTGTGTAAGTTGTGCTGCATGCGAGCCATC GCGGTGAAGAACGCACCGTGTGCTGCGTCATG TGTTTCAGTTTGAAGGCGATAAACTGGTGAAAG TGGAAGTTAGCATCGATCCGCTGGG (TAA)
luxs iti_... 0.2_... 248	157	MSEKEQKEFWKKFYDA LDAGDAETASALFPDG TIIYLWDGIVFRTREE PKWVFKLKSTSENAK RKIVSFKVDGNKSDVE VVLEAVNNGEKKKVKL KHVFHFEEDKLVVEVE KIDPL	30.5 0172 771	357	(ATG)AGCGAAAAAGAACAGAAAGAAATTTTGGGA AAAAGTTTTATGATGCGCTGGATGCGGGCGATG CGGAAACTGCGAGCGCACTGTTTCCAGATGGCA CCATTATCTATCTGTGGGATGGCATTGTGTTTC GCACGCGCGAAGAATTCAAAGAATGGTTTGTGA AACTGAAAAGCACCTCGGAAAACGCCAAACGCA AAATTGTGAGCTTTAAAGTGGATGGTAACAAAT CTGATGTGGAAGTGGTGTGGAAGCGAACGTGA ACGGCGAAAAAGAAAGGTTAAATTGAAACATG TGTTCCATTTTGAAGGCGATAAACTGGTTGAAG TCGAAGTGAAAATTGATCCGCTGGG (TAA)
luxs iti_... 0.2_... 258	158	MSEEEIRNFVKRFYEA LDAGDAETASALFPDG TEIYLWDGKTFRTQAE FRAWFEELRSTSDDAK REVVKFEVEGNVAYVE VVLKANHKKGKKEVVKL KHEFEFEGDKLVKVRV EIKPL	29.5 1209 399	358	(ATG)AGCGAAGAAGAAATTCGTAACCTTTGTGA AACGCTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACCGCCAGTGGTGTGTTTCCAGATGGCA CCGAAATTTATCTGTGGGATGGCAAAACCTTTC GCACCCAGGCGGAATTCGCGCGTGGTTTGAAG AACTGCGCAGCACCAGTGATGATGCGAAACGTG AAGTGGTTAAATTTGAAGTTGAAGGCAACGTGG CGTATGTGGAAGTTGTGCTGAAAGCGAACCATTA AAGGCAAGAAAGAGTCGTGAAACTGAAACATG AATTTGAGTTCGAAGGCGATAAAGTGGTGAAAG TGCGCGTTGAAATCAAACCGCTGGG (TAA)
luxs iti_... 0.2_... 262	159	MSEEEQREFVVKRFYEA LDAGDAETASALFPDG TEIHLWDGRTFRTRAE FRAWFEELYSQSENAR REITRFEVDGNVSHVE VVLEAEYEGERVVRRL RHFVEFEGDRLVRVEV EIEPL	5.55 9778 853	359	(ATG)AGCGAAGAAGAACAGCGCAATTTTGGGA AACGCTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACCGCAAGCGCGTGTGTTTCTGATGGCA CCGAAATTCATCTGTGGGATGGCCGACCTTTC GCACCCGCGCCGAATTCGTGCGTGGTTTGAAG AACTGTATAGCCAGAGCGGAAACGCGCGCCGCG AAATTACCCGCTTTGAAGTGGATGGCAACGTGT CACATGTGGAAGTGGTGTGGAAGCGGAATATG AAGGCGAAAAACGTGTGGTGGCCTGCGCCATG TGTTTGAATTTGAAGGTGATCGCCTGGTGGTG TTGAAGTCGAAATTGAACCGCTGGG (TAA)
luxs iti_... 0.2_... 263	160	MSEEEIREFVVKRFYEA LDAGDAETASALFPDG TEIHLWDGTVFRTREE FRAWFVELRSKSDDAK REVVKFEVDGNKAYVE VVLHANVDGEEKTVRL VHEFLFEGDKLVKVEV DIEPL	7.66 7588 113	360	(ATG)AGCGAAGAAGAAATTCGCGAATTTGTGA AACGCTTTTATGAAGCGCTGGATGCGGGCGATG CCGAAACTGCGAGCGCACTATTTCCGATGGCA CCGAAATTCATCTGTGGGATGGTACCGTGTTC GCACCCGTGAAGAATTCGCGCGTGGTTTGTGG AACTGCGCTCAAAAAGCGATGATGCGAAACGTG AAGTGGTGAAATTTGAAGTTGATGGTAACAAAG CGTATGTGGAAGTTGTGCTGCATGCGAACGTGG ATGGGGAAGAAAAGACCGTGGCGCTGGTGCATG AATTTCTGTTTGAAGGCGATAAACTGGTCAAG TGAAAGTGGATATTGAACCGCTGGG (TAA)
luxs iti_... 0.2_... 273	161	MSEEDIREFVRRFYEA LDAGDADTASALFKDG TKIHLWDGKTFRTREE FREWVRLHSTSDDAK REVVRLEVEGNPAHVE VVLRSYQGEDRVVRL	574. 8009 675	361	(ATG)AGCGAAGAAGATATTCGCGAATTTGTGC GCCGCTTTTATGAAGCGCTGGATGCGGGCGATG CGGATACCGCAAGCGCGCTGTTTAAAGATGGCA CCAAAATTCATCTGTGGGATGGCAAAACCTTTA CCACCGTGAAGAATTTCTGTAATGGTTTGTGA GACTGCATAGCACCAGCGATGATGCGCGCCGGG

		THEFLFEGDEVVEVHV RIEPL			AAGTGGTACGCCTGGAAGTTGAAGGCAACCGTG CACATGTGGAAGTCGTGCTGCGCGGAGCTATC AGGGCGAAGATCGCGTGGTTAGACTGACCCATG AATTTCTGTTTGAAGGTGATGAAGTTGTTGAAG TGCATGTGCGCATTGAACCGCTGGG (TAA)
luxs iti_ 0.2_ 276	162	MSEEAQREFWRRFYEA LDAGDAETASALFPDG TEIHLWDGRTFRTQAE FRDWFVKLHSTADAR REITRFRVEGDVAHVE VVLHASIDGEEKTVKL RHVAHFEGDKLVEVHV DIEPL	4.37 6641 327	362	(ATG)AGCGAAGAAGCGCAGCGCGAATTTCTGGC GCCGCTTTTATGAAGCGCTGGATGCGGGTGATG CGGAAACCGCCTCTGCACTGTTTCCGGATGGTA CCGAAATTCATCTGTGGGATGGCCGCACCTTTC GCACCCAGGCAGAAATTCGCGATTGTTTGTGA AATTGCATAGCACCAGCGCGGATGCGCGCCGCG AAATTACCCGCTTTAGAGTGAAGGTGATGTGG CGCATGTTGAAGTGGTCTGTCATGCGAGCATTG ATGGCGAAAGAAACCGTGAAACTGCGCCATG TGGCCCATTTTGAAGGCGATAAACTGGTGAAG TGCATGTGGATATTGAACCGCTGGG (TAA)
luxs iti_ 0.2_ 277	163	MSEEEIREFVRRFYEA LDAGDAETASALFKDG TKIYLWDGTVFETREE FRAWFVELYSKSENAR RRVVSFKVDGNVAEVE VVLHASFQGEDKVVRL KHRFKFEGDEVVEVEV DIEPL	1922 .585 349	363	(ATG)AGCGAAGAAGAAATTCGCGAATTTGTGC GCCGCTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACCGCAAGCGCGCTGTTTAAAGATGGCA CCAAATTTATCTGTGGGATGGTACCGTGTTTG AAACCCGTGAAGAATTTCCGCGCTGGTTTGTGG AACTGTATAGCAAAAGCGAAGAACGCGCGCCGCG GAGTGGTGAGCTTTAAAGTGGATGGCAACGTGG CTGAAGTGAAGTGGTGTGTCATGCGAGCTTTC AGGGCGAAGATAAAGTTGTGCGTCTGAAACATC GTTTTAAATTTGAAGCGCATGAAGTTGTTGAAG TAGAAGTTGATATTGAACCGCTGGG (TAA)
luxs iti_ 0.2_ 280	164	MSEEAQREFWKRFYEA LDAGDAETASALFPDG TRIYLWDGRVFTTQAE FRAWFVELRSTSDAK REIVKFEVDGDVSYVE VVLKANVDGEEKIVRL RHVFHFEGDRLVEVEV DIEPL	3.35 6599 862	364	(ATG)AGCGAAGAAGCGCAGCGCGAATTTTGGGA AACGCTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACCGCGAGCGCATTTGTTCCAGATGGCA CCCGCATTTATCTGTGGGATGGCCGCGTTTTTA CCACCCAGGCGGAATTTCCGCGCTGGTTTGTGG AACTGCGTTCGACCAGCGCCGATGCGAAACGTG AAATTGTTAAATTTGAAGTGGATGGCGATGTGT CCTATGTGGAAGTGGTGTGTAAGCGAAGCTAG ATGGTGAAGAAAAGATTGTGCGCCTGCGCCATG TGTTTCATTTTGAAGCGCATGCGCTGGTTGAAG TCGAAGTTGAAATCGAACCGCTGGG (TAA)
luxs iti_ 0.2_ 283	165	MSEEEQREFWKRFYEA LDAGDAETASALFPDG TRIYLWDGRTFTTRA FRAWFEELHSTSENAR REIVSFRVDGDVADVE VVLRAVNGEERTVRL RHVFHFEGDRLVEVHV DIEPL	143. 1679 337	365	(ATG)TCTGAAGAAGAACAGCGCGAATTTTGGGA AACGCTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACCGCGAGCGCACTTTTCCAGATGGCA CCCGCATTTATCTGTGGGATGGCCGCACCTTTA CCACCCGCGCGAATTTCCGCGTGGTTTGAAG AACTGCATAGCACCAGTGAACCGCGCGCCGCG AAATTGTGAGCTTTCGTGTGGATGGCGATGTGG CGGATGTGGAAGTGGTGTGCGTGCGAATGTGA ATGGCGAAGAAGTACCGTGCGCCTGCGTCATG TGTTTCATTTTGAAGCGCATGCGCTGGTTGAAG TGCATGTTGAAATTGAACCGCTGGG (TAA)
luxs iti_ 0.2_ 286	166	MSEEAQREFWRRFYDA LDAGDAETASALFPDG TRIHLWDGRTFHTRAE FRAWFVELRSTSDDAK REIISFEVDGNRSRVE VVLHANIDGEEKVLL RHEFLFEGDRLVEVHV EIKPL	2.29 9930 891	366	(ATG)AGCGAAGAAGCGCAGCGCGAATTTTGGC GCCGCTTTTATGATGCGCTGGATGCGGGCGATG CGGAAACCGCAAGCGCGTGTGTTTCTGATGGCA CCCGCATTCATCTGTGGGATGGCCGCACCTTTC ATACCCGCGCGGAATTTCCGCGCTGGTTTGTGG AACTGCGTAGCAGCAGCGATGATGCCAAACGCG AAATTATTAGCTTTGAAGTGGATGGCAACCGCA GCCGCGTGAAGTGGTGTGTCATGCGAACATCG ATGGTGAAGAAAAGAAAGTGTGCTGCGCCATG AATTTCTGTTTGAAGCGCATGCGCTGGTTGAAG TTCATGTGGAATTAACCGCTGGG (TAA)

luxs iti_... 0.2_... 290	167	MSEEEIREFVDRFYKA LDAGDAETASALFPDG TEIHLWDGTTFTQAE FRAWFVKLKSTSDNAK REIVKLEVDGNKAKVE VVLKANVNGEEKVVKL THEFEFEGDRLVKVTV KIEPL	26.8 7145 819	367	(ATG)AGCGAAGAAGAAATCCGCGAATTTGTTG ATCGCTTTTATAAAGCGCTGGATGCGGGTGATG CGGAAACCGCAAGCGCTTATTTCCCGATGGCA CCGAAATTCATCTGTGGGATGGTACCACCTTTC ATACCCAGGCGGAATTCGCGCGTGGTTTGTGA AACTGAAATCGACCAGCGATAACGCGAAACGTG AAATTTGTTAAACTGGAAGTGGATGGCAACAAAG CGAAAGTGGAGTTGTGCTGAAAGCCAACGTTA ACGGTGAAGAAAAAGTGGTCAAAGTACCCATG AATTTGAATTCGAAGGCGATCGCCTGGTGAAAG TGACCGTGAAATTTGAACCGCTGGG (TAA)
luxs iti_... 0.3_... 305	168	MSAEAIRQFVERFYAA LDAGDADTASALFPDG TEIHLWDGRTFRTQAE FRAWFEELYSTSDGAS REVVALEVDGNRARVE VVLHASVGGEEKTVRL RHEFEFEGDQLVRVTV EIEPL	54.5 0034 554	368	(ATG)AGCGCGGAAGCAATTCGCCAGTTTGTGG AACGCTTTTATGCGGCGCTGGATGCGGGTGATG CGGATACTGCATCAGCGCTGTTTCCGGATGGTA CCGAAATTCATCTGTGGGATGGCCGTACCTTTC GCACCCAGGCGGAATTCGCGCGTGGTTTGAAG AACTGTATAGCACCAGCGATGGCGGAGCCGCG AAGTGGTGGCACTTGAAGTTGATGGTAACCGCG CGAGAGTGGAGTTGTGCTGCATGCGAGCGTTG GCGGCGAAGAAAAGACCGTGCCTGCGTCATG AATTTGAATTCGAAGGCGATCAGCTGGTGGCGG TGACCGTGGAATTTGAACCGCTGGG (TAA)
luxs iti_... 0.3_... 306	169	MSEKEQKEFWKKFYDA LDAGDADTASALFPDG TKIYLWDGKVFTQEE FREWFELKYSKSDNAK REIVKFKVDGNIAYVE VILKANVDGEEKVVKL KHVFKFEGDKVKEVKV KIVPL	439. 9108 5	369	(ATG)AGCGAAAAAGAACAGAAAGAAATTTGGGA AAAAGTTTATGATGCGCTGGATGCGGGTGATG CGGATACCGCGAGCGCACTGTTTCCAGATGGCA CCAAAATTTATCTGTGGGATGGCAAGTGTTC GCACCCAGGAAGAAATTCGCTGAATGGTTTGAAA AACTGTATAGCAAAAGCGATACGCCAAACGCG AGATTGTGAAATTTAAAGTTGATGGTAACATCG CCTATGTGGAAGTGATTCTGAAAGCGAACGTGG ATGGCGAAGAGAAAGTGGTGAAACTGAAACATG TGTTTAAATTTGAAGGCGATAAAGTGAAAGAAG TTAAAGTCAAAATTTGTCCCGCTGGG (TAA)
luxs iti_... 0.3_... 308	170	MSEESQREFWKRIFYAA LDAGDAETASALFPDG TEIHLWDGTVFRTAE FRAWFVLDLHKSNDAS REITSFKVEGNKALVE VVLHASFKGEERTVKL THVFEFEGDRLVRVEV EIKPL	1116 .827 92	370	(ATG)AGCGAAGAAAGCCAGCGCAATTTTGGA AACGCTTTTATGCGGCGCTGGATGCGGGCGATG CGGAAACTGCAAGCGCATTGTTTCCGGATGGCA CCGAAATTCATCTGTGGGATGGTACCCTGTTTC GCACCCGCGCGGAATTCGCGCGTGGTTTGTGG ATCTGCATAGCAAAAGCGATAACGCGAGCCGCG AAATTACCTCTTTTAAAGTGAAGGCAACAAAG CGCTGGTTGAAGTGGTGTGCTGCATGCGAGCTTA AAGGTGAAGAACGCACCGTGAACCTGACCCACG TGTTTGAATTTGAAGGCGATCGCCTGGTGGCGG TCGAAGTTGAAATTAACCGCTGGG (TAA)
luxs iti_... 0.3_... 309	171	MSEEAVEREFVRIFYEA LDAGDAETASALFPDG TEIHLWDGTTFRTAE FRAWFRRLRATSEDAR RRVVRLEVDGNVADVE VVLHASVYRGEERVVRL RHRYEFEGDRLVRVDV EIRPL	53.3 1582 585	371	(ATG)TCGGAAGAAGCGGTGCGTGAATTTGTGC GCCGCTTTTATGAAGCGCTGGATGCGGGCGATG CGGAAACTGCGAGCGCTTGTTCCTGATGGCA CCGAAATTCATCTATGGGATGGTACCACCTTTC GCACCCGCGCGGAATTCGTCGCTGGTTTCGTC GTCTGAGAGCAACTAGCGAAGATGCGGCGCGTC GTGTGGTGGCTCTGGAAGTGATGGCAACGTTG CGGATGTCGAAGTTGTGCTGCATGCGAGCTATC GCGGTGAAGAACGCGTGTGCGGCTGCGCCATC GCTACGAATTTGAAGGCGATCGCCTGGTGGCGG TTGATGTTGAAATTCGCCCGCTGGG (TAA)
luxs iti_... 0.3_... 317	172	MSEEEQRQFVKRIFYEA LDAGDSETASALFPDG TVIHLWDGTTFTQEE FRAWFEKLRSTSDNAR REVVFHEVEGNVAYVE VVLRASVGGEEVVRL	10.1 7346 234	372	(ATG)AGCGAAGAAGAAGCGCCAGTTTGTGA AACGCTTTTACGAAGCGCTGGATGCGGGTGATA GCGAAACCGCAAGCGCGCTGTTTCCGGATGGCA CCGTGATTCATCTGTGGGATGGTACCACCTTTC ATACCCAGGAAGAATTCGCGCGTGGTTTGAAA AACTGCGCAGCACCAGCGATAACGCGCGCGTG

		RHVHFHFECDHLVEVEV EIEPL			AAGTGGTGCATTTTGAAGTTGAAGGCAACGTGG CGTATGTGGAAGTTGTTCTGCGTGGCAGCGTGG GCGGTGAAGAACGCGTTGTGCGTCTGCGTCATG TGTTTCATTTTGAAGGCGATCATCTGGTTGAAG TCGAAGTGGAAATTGAACCGCTGGG (TAA)
luxs iti_ 0.3_ 323	173	MSEEEQRNFWKFFYEA LDAGDAETASALFPDG TEIYLWDGTVFKTREE FRAWFVELYSTENAK RHIVDFKVDGNEKVK VILKANIDGEEKVVEL EHYFLFEGDHLVEVKV EIHPL	5.97 4429 855	373	(ATG)AGCGAAGAAGAAGCAGCGCAACTTTTGGG AAAAGTTTATGAAGCGCTGGATGCGGGTGATG CGGAAACCGCAAGCGCGCTGTTTCTCTGATGGCA CCGAAATTTATCTGTGGGATGGTACCGTGTTF AAACCGCGCAAGAGTTCCGTCGCTGGTTTGTGG AACTGTATTCGACCTCGGAAAACGCGAAACGCC ATATTGTGGATTTTAAAGTGGATGGCAACGAAA GCAAAGTGAAGTGATTCTGAAAGCGCAACATTG ATGGTGAAGAAAAGTGGTTGAACTGGAAACATT ATTTTCTGTTTGAAGGCGATCATCTGGTTGAAG TGAAGTGGAAATTCATCCGCTGGG (TAA)
luxs iti_ 0.3_ 324	174	MSEEEIREFVERFYKA LDAGDAETASSLFPDG TEIHLWDGKTFHTQEE FRSWFVELKSKSDNAK RKVVSFEVEGNKAKVK VVLKASYKGEEKVKL EHEFEFEGDKLVEVNV KIEPL	179. 7187 284	374	(ATG)AGCGAGGAAGAAATTCGAGAAATTTGTGG AACGCTTTTATAAGCGCTGGATGCGGCGGATG CGGAAACTGCGAGCAGCCTGTTTCCAGATGGCA CCGAAATTCATCTGTGGGATGGCAAAACATTTT ATACCCAAGAAGAATTTGCGAGCTGGTTTGTGG AACTGAAAAGCAAAAGCGATACGCGCAACGCA AAGTGGTGAGCTTTGAAGTGAAGGCAACAAAG CGAAAGTGAAGTTGTGCTGAAAGCGAGCTATA AAGGGGAGGAAAAAGGTTAACTGGAACATG AATTTGAATTCGAAGGCGATAAATGGTTGAAG TCAACGTGAAAATTCGAACCGCTGGG (TAA)
luxs iti_ 0.3_ 330	175	MSEKEQREFWKFFYEA LDAGDAETASALFPDG TEIHLWDGKTFHTREE FKEWFKLYSRSENAK RHIVKFEVDGDKAYVE VVLKAYIGGEEKVVKL KHEFLFEGDKLVEVNV KIEPL	560. 5245 335	375	(ATG)AGCGAAAAAGAACAGCGCGAAATTTTGGG AGAAATTTTATGAAGCGCTGGATGCGGGTGATG CGGAAACCGCGCAGCGGTTATTTCCCGATGGCA CCGAAATTCATCTGTGGGATGGCAAAACCTTTC ATACCCGCGAAGAATTTAAAGAATGGTTTGTGA AACTGTATAGCGGTTTCAAGAAACGCGAAACGTC ATATTGTGAAGTTTGAAGTGGATGGCGATAAAG CGTATGTGGAAGTGGTGTGATGCGAAAATTTG GCGGTGAAGAAAAAGTGGTTAACTGAAACATG AATTTCTGTTTGAAGGCGACAACTGGTTGAAG TTAACGTGAAAATTCGAACCGCTGGG (TAA)
luxs iti_ 0.3_ 338	176	MSEEEIKEFVERFYKA LDAGDADTASSLFPDG TEIFLWDGKTFKTKKE FREWFVKLKSESDDAK REVVSFKVEGNKAYVE VVLHANYKGEKKEVKL THIFEFEGDKVVKVEV TIVPL	53.7 5120 94	376	(ATG)AGCGAAGAAGAAATTTAAAGAATTTGTGG AACGCTTTTATAAGCGCTGGATGCGGGTGATG CGGATACCGCAAGCAGCCTGTTTCCAGATGGCA CCGAAATCTTCTGTGGGATGGCAAAACCTTTA AAACCAAAGAGGAATTTCCGGAATGGTTTGTGA AACTGAAAAGCGAAAGCGATGATGCGAAACGCG AAGTGGTGTGCTGAAAGTGAAGGCAACAAAG CGTATGTGGAAGTAGTGTGATGCGAACTATA AAGGCGAAAAGAAAGAAAGTTAACTGACCCATA TTTTTGAATTTGAAGGTGATAAAGTCGTGAAAG TTGAAGTGACCATTGTGCCGCTGGG (TAA)
luxs iti_ 0.3_ 366	177	MSEEEQRQFVKFFYEA LDAGDAETASALFPDG TEIHLWDGKTFHTQAE FRAWFERLYSTENAR RHVVDFRVEGNVADVE VVLHATKDGEEHTVRL RHKFEFEGDELVRVEV VIEPL	243. 3538 355	377	(ATG)AGCGAAGAAGAGCAGCGCCAGTTTGTGA AAGAATTTTATGAAGCGCTGGATGCGGGTGATG CGGAAACCGCGAGCGCGTGTGTTTCCAGATGGCA CCGAAATTCATCTGTGGGATGGTAAACCTTTA CCACCCAGGCGGAATTTCCGCGCTGGTTTGAAC GCCTGTATAGCACCAGCGAAATGCGCGCGGCC ATGTCGTGGATTTTCTGTGGAAGGCAACGTGG CCGATGTTGAAGTGGTGTGATGCGACCAAAG ATGGTGAAGACATACCGTGCCTGCGTCATA AATTTGAATTCGAAGGCGATGAACTGGTGCAG TGAAGTTGTGATTGAACCGCTGGG (TAA)

luxs iti_... 0.3... 381	178	MSAEAQREFWARFYAA LDAGDADTASALFPDG TEIYLWDGKVFRTREE FRAWFVELRSTSA REIVSFEVDGNPASVD VVLHASVNGEEHTVHL HHEFEFEGDRLVRVRV TIQPL	1.97 5120 94	378	(ATG)AGCGCGGAAGCGCAGCGCGAATTTTGGG CGCGCTTTTATGCGGCGTGGATGCGGGTGATG CGGATACCGCTAGCGCGCTGTTTCTTGATGGCA CCGAAATTTATCTGTGGGATGGCAAAGTGTTTC GCACCCGCGAAGAATTTCGCGCGTGGTTGTGG AACTGCGCAGCACCAGCGCCGATGCGAAACGTG AAATTTGTGAGCTTTGAAGTGGATGGTAACCGTG CGAGCGTGGATGTGGTGCTGCATGCCAGCGTGA ACGGTGAAGAACATACCGTGCATCTGCATCAG AATTTGAATTCGAAGGCGATCGCCTGGTGCGCG TGCGTGTGACCATTTCAGCCGCTGGG (TAA)
luxs iti_... 0.3... 383	179	MSAEQQREFVRFYEA LDAGDADTASALFPDG TEIHLWDGTTFRTRAE FRAWFEELYSTSENAS REVTSFSVDGDVADVE VVLRLNLGGEDRTVSL RHVFHFAGDRLVRVEV SIRPL	1119 .172 771	379	(ATG)AGCGCAGAACAGCAGCGCGAATTTGTGA AACGCTTTTATGAAGCGCTGGATGCGGGTGATG CGGATACCGCAAGCGCACTGTTTCCGGATGGCA CCGAAATTCATCTGTGGGATGGTACCACCTTTC GCACCCGCGCGGAATTTCGCGCGTGGTTTGAAG AACTGTATAGCACCTCGGAAAACGCGAGCCGCG AAGTGACCAGCTTTTCGGTGGATGGCGATGTGG CAGATGTGGAAGTGGTGCTGCGCGCGAACCTGG GTGGTGAAGATCGCACCGTTAGCCTGAGACATG TGTTTCATTTTGCGGGCGATCGCCTGGTGCGCG TTGAAGTGAGCATTTCGCCCGCTGGG (TAA)
luxs iti_... 0.3... 385	180	MSEEEIKEFVRRFYEA LDAGDADTASALFPDG TRIHLWDGTTFTTRAE FRAWFVDLYSKSDAAK REVTSLKVEGNVAKVK VVLKANYKGEEKTVEL EHKFEFEGDRLVRVDV TIKPM	347. 1257 775	380	(ATG)AGCGAAGAAGAAATTAAAGAAATTTGTGC GCCGCTTTTATGAAGCGCTGGATGCGGGTGATG CGGATACCGCAAGCGCACTGTTTCCGGATGGCA CCCGCATTCATCTGTGGGATGGTACCACCTTFA CCACCCGCGCGGAATTTTCGCGCGTGGTTTGTGG ATCTGTATAGCAAAAGCGATGCGGCGAAACGTG AAGTGACCAGCCTGAAAGTGGAAGGTAATGTGG CGAAAGTGAAAGTAGTGCTGAAAGCGAACTATA AAGGTGAAGAAAAGCCGTGGAAGTGAACATA AATTTGAATTCGAAGGCGATCGCCTGGTGCGCG TTGATGTTACCATTAAACCGATGGG (TAA)

In another aspect, the disclosure provides protein having luciferase activity, comprising an amino acid sequence at least 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100%

5 identical to SEQ ID NO:1, wherein:

Residue 14 is Y, D, or E and residue 98 is H or N;

Residue 18 is D or E and residue 65 is R.

The proteins of this aspect are non-naturally occurring. In one embodiment, the percent identity relative to the reference sequence is carried out by sequence alignment with the Needleman-Wunsch algorithm, a common sequence alignment tool for those of skill in the art, which allows for insertions and deletions.

In one embodiment, the protein comprises one or both of A96M and M110V substitutions relative to SEQ ID NO:1. In another embodiment, the protein comprises comprising an R60S substitution relative to SEQ ID NO:1. In a further embodiment, the protein comprises R60S, A96M, and M110V substitutions relative to SEQ ID NO:1. In

another embodiment, any substitutions relative to SEQ ID NO:1 at residues F12, I35, W38, F49, V81, L83, V94, A 97, W100, M110, V112 are conservative amino acid substitutions. In one embodiment, the protein comprises an amino acid sequence at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence selected from SEQ ID NO:1-3. In another embodiment, the protein comprises an amino acid sequence at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence selected from SEQ ID NO: 1-181.

In another embodiment, the proteins comprise the formula X1-Z1-X2-Z2-X3-Z3-X4-Z4-X5-Z5-X6-Z6-X7-Z7-X8-Z8, wherein:

X1 has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of MSEEQIRQFLRRFYEALD (SEQ ID NO: 182), wherein residue 14 is Y, D, or E and residue 18 is D or E;

X2 has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of ADTAASLF (SEQ ID NO: 183);

X3 has an amino acid sequence at least 50%, 75%, or 100% identical to the amino acid sequence of TIHL (SEQ ID NO: 184);

X4 has an amino acid sequence at least 33%, 66%, or 100% identical to the amino acid sequence of VTF;

X5 has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of EEFREWFERLFST (SEQ ID NO: 185);

X6 has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of QREIKSLEVR (SEQ ID NO: 186), wherein residue 2 is R;

X7 has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of VEVHVQLHATH (SEQ ID NO: 187);

X8 has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of KHTVDATHHWHFR (SEQ ID NO: 188), wherein residue 8 is H or N;

X9 has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of VTEMRVHINPTG (SEQ ID NO: 189); and

wherein Z1, Z2, Z3, Z4, Z5, Z6, Z7, and Z8 are independently present or absent, and
 5 when present may comprise any amino acid sequence.

In one embodiment, 1, 2, 3, 4, 5, 6, 7, or all 8 of the following are true

Z1 comprises SGD;

Z2 comprises HPGV (SEQ ID NO: 190);

Z3 comprises WDG;

10 Z4 comprises TSR;

Z5 comprises RKDA (SEQ ID NO: 191);

Z6 comprises GDT;

Z7 comprises NGQ;

Z8 comprises GNR; and

15 wherein 0, 1, 2, 3, 4, 5, 6, 7, or all 8 of Z1, Z2, Z3, Z4, Z5, Z6, Z7, and Z8 further comprising an additional polypeptide domain.

In one embodiment of all of the proteins of the disclosure, amino acid substitutions relative to the reference protein are conservative amino acid substitutions. As used herein, “conservative amino acid substitution” means a given amino acid can be replaced by a
 20 residue having similar physiochemical characteristics, e.g., substituting one aliphatic residue for another (such as Ile, Val, Leu, or Ala for one another), or substitution of one polar residue for another (such as between Lys and Arg; Glu and Asp; or Gln and Asn). Other such conservative substitutions, e.g., substitutions of entire regions having similar hydrophobicity characteristics, are known. Proteins comprising conservative amino acid substitutions can be
 25 tested in any one of the assays described herein to confirm that a desired activity, is retained. Amino acids can be grouped according to similarities in the properties of their side chains (in A. L. Lehninger, in Biochemistry, second ed., pp. 73-75, Worth Publishers, New York (1975)): (1) non-polar: Ala (A), Val (V), Leu (L), Ile (I), Pro (P), Phe (F), Trp (W), Met (M); (2) uncharged polar: Gly (G), Ser (S), Thr (T), Cys (C), Tyr (Y), Asn (N), Gln (Q); (3) acidic:
 30 Asp (D), Glu (E); (4) basic: Lys (K), Arg (R), His (H). Alternatively, naturally occurring residues can be divided into groups based on common side-chain properties: (1) hydrophobic: Norleucine, Met, Ala, Val, Leu, Ile; (2) neutral hydrophilic: Cys, Ser, Thr, Asn, Gln; (3) acidic: Asp, Glu; (4) basic: His, Lys, Arg; (5) residues that influence chain orientation: Gly,

Pro; (6) aromatic: Trp, Tyr, Phe. Non-conservative substitutions will entail exchanging a member of one of these classes for another class. Particular conservative substitutions include, for example; Ala into Gly or into Ser; Arg into Lys; Asn into Gln or into His; Asp into Glu; Cys into Ser; Gln into Asn; Glu into Asp; Gly into Ala or into Pro; His into Asn or into Gln; Ile into Leu or into Val; Leu into Ile or into Val; Lys into Arg, into Gln or into Glu; Met into Leu, into Tyr or into Ile; Phe into Met, into Leu or into Tyr; Ser into Thr; Thr into Ser; Trp into Tyr; Tyr into Trp; and/or Phe into Val, into Ile or into Leu.

In another embodiment, the disclosure provides self-complementing multipartite protein having luciferase activity, comprising at least a first polypeptide component and a second polypeptide component, wherein the at least first polypeptide component and the second polypeptide component are not covalently linked, wherein in total the at least first polypeptide component and the second polypeptide component comprise domains X1-Z1-X2-Z2-X3-Z3-X4-Z4-X5-Z5-X6-Z6-X7-Z7-X8-Z8-X9, wherein each domain is as defined above; and

wherein (a) each X domain is fully present within one polypeptide component of the at least first polypeptide component and the second polypeptide component, and (b) none of the at least first polypeptide component and the second polypeptide component include each of X1, X2, X3, X4, X5, X6, X7, X8, and X9.

The split proteins are non-naturally occurring. The split proteins comprise at least a first polypeptide component and a second polypeptide component in which X domains are preserved while split points are taken only in the Z domains. In other words, each X strand or (X1, X2, X3, X4, X5, X6, X7, X8, and X9) is fully present within one polypeptide component of the at least first polypeptide component and the second polypeptide component, while the protein is split into separate components at a Z domain (of Z1, Z2, Z3, Z4, Z5, Z6, Z7, and Z8), wherein the Z domain that the split occurs at may be absent, or may be partially present in one or both of the first and second polypeptide components. By way of non-limiting example, in various embodiments of a split luciferase protein, the first polypeptide component and the second polypeptide component may comprise components as exemplified in Table 2.

Table 2

Example	First polypeptide component comprises	Second polypeptide component comprises
1: Split at Z2	X1-Z1-X2-(Z2)	(Z2)-X3-Z3-X4-Z4-X5-Z5-X6-Z6-X7-Z7-X8-Z8-X9

2: Split at Z4	X1-Z2-X2-Z2-X3-Z3-X4-(Z4)	(Z4)-X5-Z5-X6-Z6-X7-Z7-X8-Z8-X9
3: Split at Z6	X1-Z2-X2-Z2-X3-Z3-X4-Z4-X5-Z5-X6-(Z6)	(Z6)-X7-Z7-X8-Z8-X9
4: Split at Z8	X1-Z2-X2-Z2-X3-Z3-X4-Z4-X5-Z5-X6-Z6-X7-Z7-X8-(Z8)	(Z8)-X9

In various embodiments, the split may occur at Z4, Z5, Z6, or Z7.

In another embodiment, the disclosure provides self-complementing multipartite protein having luciferase activity, comprising at least a first polypeptide component and a second polypeptide component, wherein the at least first polypeptide component and the second polypeptide component are not covalently linked, wherein in total the at least first polypeptide component and the second polypeptide component comprise the secondary structure arrangement H1-L1-H2-L2-E1-L3-E2-L4-H3-L5-E3-L6-E4-L7-E5-L8-E6, wherein each domain is as defined above;

wherein (a) each H and E domain is fully present within one polypeptide component of the at least first polypeptide component and the second polypeptide component, and (b) none of the at least first polypeptide component and the second polypeptide component include all of the H and E domains.

In this embodiment, the split proteins comprise at least a first polypeptide component and a second polypeptide component in which H and E domains are preserved while split points are taken only in the L domains. In various embodiments, the split occurs at L4, L5, L6, L7, or L8.

The split proteins of these embodiments are only active when they are brought together, and thus are conditionally active.

In another embodiment the disclosure provides fusion proteins comprising:

(a) the protein or polypeptide component of any embodiment or combination of embodiments herein; and

(b) one or more additional functional domains.

As used herein, a “functional domain” is any polypeptide that can be usefully fused to the luciferase protein or split protein component of the disclosure. By way of non-limiting examples, the one or more additional functional domains may comprise a diagnostic polypeptide, any protein that one might want to localize within a cell, tissue, or organism; etc.

In another aspect the disclosure provides nucleic acids encoding the protein, protein component, or fusion protein of any embodiment or combination of embodiments of the disclosure. The nucleic acid sequence may comprise single stranded or double stranded RNA

(such as an mRNA) or DNA in genomic or cDNA form, or DNA-RNA hybrids, each of which may include chemically or biochemically modified, non-natural, or derivatized nucleotide bases. Such nucleic acid sequences may comprise additional sequences useful for promoting expression and/or purification of the encoded polypeptide, including but not limited to polyA sequences, modified Kozak sequences, and sequences encoding epitope tags, export signals, and secretory signals, nuclear localization signals, and plasma membrane localization signals. It will be apparent to those of skill in the art, based on the teachings herein, what nucleic acid sequences will encode the polypeptides of the disclosure. In various non-limiting embodiments, the nucleic acid may comprise the nucleotide sequence of any one of SEQ ID NO:200-380, wherein residues in parentheses are optional and may be present or absent.

In a further aspect, the disclosure provides expression vectors comprising the nucleic acid of any aspect of the disclosure operatively linked to a suitable control sequence. "Expression vector" includes vectors that operatively link a nucleic acid coding region or gene to any control sequences capable of effecting expression of the gene product. "Control sequences" operably linked to the nucleic acid sequences of the disclosure are nucleic acid sequences capable of effecting the expression of the nucleic acid molecules. The control sequences need not be contiguous with the nucleic acid sequences, so long as they function to direct the expression thereof. Thus, for example, intervening untranslated yet transcribed sequences can be present between a promoter sequence and the nucleic acid sequences and the promoter sequence can still be considered "operably linked" to the coding sequence. Other such control sequences include, but are not limited to, polyadenylation signals, termination signals, and ribosome binding sites. Such expression vectors can be of any type, including but not limited plasmid and viral-based expression vectors. The control sequence used to drive expression of the disclosed nucleic acid sequences in a mammalian system may be constitutive (driven by any of a variety of promoters, including but not limited to, CMV, SV40, RSV, actin, EF) or inducible (driven by any of a number of inducible promoters including, but not limited to, tetracycline, ecdysone, steroid-responsive). The expression vector must be replicable in the host organisms either as an episome or by integration into host chromosomal DNA. In various embodiments, the expression vector may comprise a plasmid, viral-based vector, or any other suitable expression vector.

In another aspect, the disclosure provides host cells that comprise the nucleic acids, expression vectors (i.e.: episomal or chromosomally integrated), non-naturally occurring polypeptides, fusion protein, or compositions disclosed herein, wherein the host cells can be

either prokaryotic or eukaryotic. The cells can be transiently or stably engineered to incorporate the nucleic acids or expression vector of the disclosure, using techniques including but not limited to bacterial transformations, calcium phosphate co-precipitation, electroporation, or liposome mediated-, DEAE dextran mediated-, polycationic mediated-, or
5 viral mediated transfection.

The disclosure also provide kits, comprising:

- (a) the protein, polypeptide component, fusion protein, nucleic acid, expression vector, and/or host cell of any embodiment or combination of embodiments herein; and
 - (b) instructions for their use.
- 10 In one embodiment, the kits further comprise diphenylterazine (DTZ).

In another aspect, the disclosure provides methods for use of the protein, polypeptide component, fusion protein, nucleic acid, expression vector, host cell, and/or kit of any preceding claim for any suitable purpose, including but not limited to use luminescent reporting assays, diagnostic assays, cellular localization of targets of interest, cellular
15 imaging, gene editing, live animal imaging, cancer labeling, CART-cells reporting, secreted assay, gene delivery, tissue engineering, etc. Additional details can be found in the examples.

In another aspect, the disclosure provides methods for making a luciferase, comprising de novo design using the methods of any embodiment disclosed herein, starting with the protein comprising the amino acid sequence of SEQ ID NO:381. The examples
20 provide detailed methods for de novo design of luciferases for DTZ. As described in the examples, the methods involve designing a shape complementary catalytic site that stabilizes the anionic state of DTZ and lowers the SET energy barrier, assuming that the downstream dioxetane light emitter thermolysis steps are spontaneous. To stabilize the anionic species of DTZ, we focused on the placement of the positively charged guanidinium group of an
25 arginine residue to interact with the anionic imidazopyrazinone core. To computationally design such active sites, the inventors first used AIMNet to generate an ensemble of anionic DTZ conformers (Fig. 1b). Next, around each conformer, the inventors used the RIFgen method to enumerate Rotamer Interaction Fields (RIFs) on 3D grids consisting of millions of placements of amino acid sidechains making hydrogen bonding and nonpolar interactions
30 with DTZ (Fig. 1c). Additionally, the inventors included an arginine guanidinium group near the deprotonation site at the nitrogen of imidazopyrazinone (N1 atom) in the RIF. RIFdock was then used to dock each DTZ conformer and associated RIF in the central cavity of each scaffold to maximize protein-DTZ interactions. An average of eight sidechain rotamers including an arginine to stabilize the anionic imidazopyrazinone core were positioned in each

pocket (Fig. 1d). For the top 50,000 docks with the most favorable sidechain-DTZ interactions, the inventors optimized the remainder of the sequence (Fig. 1d) for high-affinity binding to DTZ with a bias towards the naturally observed sequence variation to ensure foldability. During the design process, pre-defined hydrogen bond networks (HBNet) in the scaffolds were kept intact for structural specificity and stability, and interactions of these HBNet side chains with DTZ were explicitly required in the RIFdock step to ensure preorganization of residues essential for catalysis. In the first sequence design phase, the identities of all RIF and HBNet residues were kept fixed, and the surrounding residues were optimized to hold the sidechain-DTZ interactions in place and maintain structural specificity. In the second sequence design step, the RIF residue identities (except the arginine) were also allowed to vary, to identify apolar and aromatic packing interactions missed in the RIF due to binning effects. During sequence design, the scaffold backbone, sidechains, and DTZ substrate were allowed to relax in Cartesian space.

15 Examples

Summary

De novo enzyme design has sought to introduce active sites and substrate binding pockets predicted to catalyze a reaction of interest into geometrically compatible native scaffolds^{1,2}, but has been limited by a lack of suitable protein structures and the complexity of native protein sequence-structure relationships. Here we describe a deep-learning based “family-wide hallucination” approach that generates large numbers of idealized protein structures containing diverse pocket shapes and designed sequences that encode them. We use these scaffolds to design artificial luciferases that selectively catalyze the oxidative chemiluminescence of the synthetic luciferin substrates, diphenylterazine (DTZ)³ through the placement of an arginine guanidinium group adjacent to an anion species that develops during the reaction in a high shape complementarity binding pocket. For both luciferin substrates, we obtain designed luciferases with high selectivity; the most active of these is a small (13.9 kDa) and thermostable ($T_M > 95^\circ\text{C}$) enzyme with a catalytic efficiency on DTZ ($k_{cat}/K_M = 10^6 \text{ M}^{-1}\text{s}^{-1}$) comparable to native luciferases but with much higher substrate specificity. The design of highly active and specific biocatalysts from scratch with broad applications in biomedicine is an important milestone for computational enzyme design, and our approach should enable the design of a wide range of new and useful luciferases and other enzymes.

35

Study

Bioluminescent light produced by the enzymatic oxidation of a luciferin substrate is widely used for bioassays and imaging in biomedical research. Because no excitation light source is needed, luminescent photons are produced in the dark which results in higher sensitivity than fluorescence imaging in live animal models and in biological samples where autofluorescence or phototoxicity is a concern^{4,5}. However, the development of luciferases as molecular probes has lagged behind that of well-developed fluorescent protein toolkits for a number of reasons: (i) very few native luciferases have been identified; (ii) many of those that have been identified require multiple disulfide bonds to stabilize the structure and are therefore prone to misfolding in mammalian cells; (iii) most native luciferases do not recognize synthetic luciferins with more desirable photophysical properties; and (iv) multiplexed imaging to follow multiple processes in parallel using mutually orthogonal luciferase-luciferin pairs has been limited by the low substrate specificity of native luciferases.

We sought to use de novo protein design to create new luciferases that are small, highly stable, well-expressed in cells, specific for one substrate, and need no cofactors to function. As we are not constrained to natural luciferase substrates, we chose a synthetic luciferin, Diphenylterazine (DTZ) as the target substrate due to its good quantum yield, red-shifted emission³, favorable *in vivo* pharmacokinetics^{14,15}, and lack of required cofactors for emission. Previous computational enzyme design studies have primarily repurposed native protein scaffolds in the PDB, but there are few native structures with binding pockets appropriate for DTZ, and the effects of sequence changes on native proteins can be unpredictable. To circumvent these limitations, we set out to generate large numbers of ideal protein scaffolds with pockets of the appropriate size and shape for DTZ, and with clear sequence-structure relationships to facilitate subsequent active site incorporation. To identify protein folds capable of hosting such pockets, we first docked DTZ into 4000 native small molecule binding proteins. We found that many NTF2 (nuclear transport factor 2)-like folds have binding pockets with appropriate shape-complementary and size for DTZ placement (Fig. 1e), and hence selected the NTF2-like superfamily as the target topology.

Family-wide hallucination

Native NTF2 structures have a range of pocket sizes and shapes but also contain non-ideal features such as long loops which compromise stability. To create large numbers of ideal NTF2-like structures, we developed a deep-learning based “family-wide hallucination”

approach that integrates unconstrained de novo design^{19,20} and fixed backbone sequence design approaches²¹ to enable the generation of an essentially unlimited number of proteins having a desired fold (**Fig. 1a**). The family-wide hallucination approach utilizes the de novo sequence and structure discovery capability of unconstrained protein hallucination^{19,20} for loop and variable regions, and structure-guided sequence optimization for core regions. We employed the trRosettaTM structure prediction neural network²², which is effective in identifying experimentally successful de novo designed proteins and hallucinating new globular proteins of diverse topologies. Starting from sequences and predicted structures of 2,000 naturally occurring NTF2s, we used trRosettaTM to optimize the amino sequence of conserved core and variable loop regions. Protein core idealization was carried out with a topology-specific loss function over core residue pair geometries (see Methods) and variable loop optimization, by optimizing sequence length and identity to maximize the confidence of the neural network in the predicted structure. To further encode structural specificity, we incorporated buried, long-range hydrogen-bonding networks. The resulting 1615 family-wide hallucinated NTF2 scaffolds provided more shape complementary binding pockets for DTZ than native small-molecule protein binding proteins (**Fig. 1e**). This approach samples protein backbones closer to native NTF2-like proteins (**Fig. 1f**) and with better scaffold quality metrics than a previous non deep-learning approach²³ (**Fig. 1g**).

We chose the NTF2 scaffold of SEQ ID NO:381 from which to design luciferases for DTZ, as described in detail below.

```
>2692_0_0.35_5_19_1806_2_0.45_5_Y14_H98_W100d1o7nb____clean__0001
__D18_R65.rdl.pdb
MSEEEIRQFLRRFYEAFDKGDVDTFASLFHPGVTIHVWQGITFTSREELREWVERFLRNFKD
MQREILSLEVRGDTVEVHVQVHTTHNGQKYTFDVTHHWHFRGHRVTEIRVHVNP (SEQ
ID NO:381)
```

De novo design of luciferases for DTZ

Standard computational enzyme design generally starts from an ideal active site or theozyme consisting of protein functional groups surrounding the reaction transition state that is then extrapolated into a set of existing scaffolds^{1,2}. However, the detailed mechanism of native marine luciferases is not well defined as only a handful of apo-structures and no holo-structures have been solved^{24,25} (excluding calcium-regulated photoproteins). Both quantum chemistry calculations^{26,27} and experimental data^{28,29} suggest that the chemiluminescent

reaction proceeds through an anionic species and that the polarity of the surroundings can substantially alter the free energy of the subsequent single electron transfer (SET) process with triplet molecular oxygen ($^3\text{O}_2$). Guided by these data (Fig. 5), we sought to design a shape complementary catalytic site that stabilizes the anionic state of DTZ and lowers the SET energy barrier, assuming that the downstream dioxetane light emitter thermolysis steps are spontaneous. To stabilize the anionic species of DTZ, we focused on the placement of the positively charged guanidinium group of an arginine residue to interact with the anionic imidazopyrazinone core.

To computationally design such active sites into large numbers of hallucinated NTF2 scaffolds, we first used AIMNet³⁰ to generate an ensemble of anionic DTZ conformers (Fig. 1b). Next, around each conformer, we used the RIFgen method^{31,32} to enumerate Rotamer Interaction Fields (RIFs) on 3D grids consisting of millions of placements of amino acid sidechains making hydrogen bonding and nonpolar interactions with DTZ (Fig. 1c). Additionally, we included an arginine guanidinium group near the deprotonation site at the nitrogen of imidazopyrazinone (N1 atom) in the RIF. RIFdock was then used to dock each DTZ conformer and associated RIF in the central cavity of each scaffold to maximize protein-DTZ interactions. An average of eight sidechain rotamers including an arginine to stabilize the anionic imidazopyrazinone core were positioned in each pocket (Fig. 1d). For the top 50,000 docks with the most favorable sidechain-DTZ interactions, we optimized the remainder of the sequence using RosettaDesignTM (Fig. 1d) for high-affinity binding to DTZ with a bias towards the naturally observed sequence variation to ensure foldability. During the design process, pre-defined hydrogen bond networks (HBNet) in the scaffolds were kept intact for structural specificity and stability, and interactions of these HBNet side chains with DTZ were explicitly required in the RIFdock step to ensure preorganization of residues essential for catalysis. In the first sequence design phase, the identities of all RIF and HBNet residues were kept fixed, and the surrounding residues were optimized to hold the sidechain-DTZ interactions in place and maintain structural specificity. In the second sequence design step, the RIF residue identities (except the arginine) were also allowed to vary, to identify apolar and aromatic packing interactions missed in the RIF due to binning effects. During sequence design, the scaffold backbone, sidechains, and DTZ substrate were allowed to relax in Cartesian space. Following sequence optimization, the designs were filtered based on ligand-binding energy, protein-ligand hydrogen bonds, shape complementarity, and contact molecular surface, and 7982 designs were selected and ordered as pooled oligos for experimental screening.

Screening and characterization of DTZ specific luciferases

Oligonucleotides encoding the two halves of each design were assembled into full-length genes and cloned into an *E. coli* expression vector (see Methods). A colony-based screening method was used to directly image active luciferase colonies from the library and the activities of selected clones were confirmed using a 96-well plate expression (Fig. 6). Three active designs were identified; we refer to the most active of these as LuxSit (*Latin: let light exist*); LuxSit is the smallest known luciferase with 117 residues (13.9 kDa). Biochemical analysis, including SDS-PAGE and size exclusion chromatography (Fig. 2ab and Fig. 7), indicated that LuxSit is highly expressed, soluble, and monomeric from *E. coli* expression. Circular dichroism (CD) spectroscopy showed a strong far UV CD signature, suggesting an organized α - β structure. CD melting experiments showed that the protein is not fully unfolded at 95 °C, and the full structure is regained when the temperature is dropped (Fig. 2c). Incubation of LuxSit with DTZ resulted in luminescence with an emission peak at ~480 nm (Fig. 2d), consistent with the DTZ chemiluminescence spectrum. While we were not able to determine the crystal structure of LuxSit, the AlphaFold2³³ predicted structure is very close to the design model at the backbone level (RMSD = 1.3 Å) and over the side chains interacting with the substrate (Fig. 2e). The designed LuxSit active site contains Tyr14-His98 and Asp18-Arg65 dyads; with the imidazole nitrogen atoms of His98 making hydrogen bond interactions with Tyr14 and the O1 atom of DTZ (Fig. 2f). The center of the Arg65 guanidinium cation is 4.2 Å from the N1 atom of DTZ and Asp18 forms a bidentate hydrogen bond to the guanidinium group and backbone N-H of Arg65 (Fig. 2g).

Activity optimization

To better understand the contributions to catalysis of LuxSit, the most active of our luciferase designs, we constructed a site saturation mutagenesis (SSM) library in which every mutation was made at every pocket residue one at a time (see Methods). Fig. 2f-i illustrate the amino-acid preferences at key positions. Arg65 is highly conserved, and its dyad partner Asp18 can only be mutated to Glu (which reduces activity), suggesting the carboxylate-Arg65 hydrogen bond is important for luciferase activity. In the Tyr14-His98 dyad, Tyr14 can be substituted with Asp and Glu, while His98 can be replaced with Asn. As all active variants had hydrogen bond donors and acceptors at these positions, the dyad may help mediate the electron and proton transfer required for luminescence. Hydrophobic (Fig. 2h) and π -stacking (Fig. 2i) residues at the binding interface tolerate other aromatic or aliphatic

substitutions and generally prefer the amino acid in the original design consistent with model-based affinity predictions of mutational effects. The A96M and M110V mutants increase activity by 16-fold and 19-fold over LuxSit respectively (Table 4). Optimization guided by these results yielded LuxSit-f (A96M/M110V) with strong initial flash emission and LuxSit-i (R60S/A96L/M110V) with more than 100-fold higher photon flux over LuxSit (Fig. 9). Overall, the active site saturation mutagenesis results support the design model, with the Tyr14-His98 and Asp18-Arg65 dyads playing key roles in catalysis and the substrate-binding pocket largely conserved.

The most active catalysts, LuxSit-f and LuxSit-i were both expressed solubly in *E. coli* at high levels and are monomeric (some dimerization was observed at the high protein concentration, Fig. 7l) and thermostable (Fig. 7j-k). Similar to native CTZ-utilizing luciferases, the apparent Michaelis constants K_M of both LuxSit-f and LuxSit-i are in the low μM range (Fig. 3a) and the luminescent signal decays over time due to fast catalytic turnover (Fig. 10a). LuxSit-i is a very efficient enzyme with a k_{cat}/K_M of $10^6 \text{ M}^{-1}\text{s}^{-1}$. The luminescence signal is readily visible to the naked eye, and the photon flux (photon s^{-1}) is 38% greater than the native *Renilla reniformis* luciferase (RLuc) (Table 3). The DTZ luminescent reaction catalyzed by LuxSit-i is pH-dependent (Fig. 10b), consistent with the proposed mechanism.

Cellular imaging and multiplexed bioassay

As luciferases are commonly used genetic tags and reporters for the study of cellular functions, we evaluated the expression and function of LuxSit-i in live mammalian cells. LuxSit-i-mTagBFP2-expressing HEK293T cells had DTZ specific luminescence (Fig. 3b), which was maintained following targeting of LuxSit-i to the nucleus, membrane, and mitochondria (Fig. 11). Native and previously engineered luciferases are quite promiscuous with activity on many luciferin substrates (Fig. 4ac), possibly due to their large and open pockets (a luciferase with high specificity to one luciferin substrate has been difficult to control even with extensive directed evolution³⁵). In contrast, LuxSit-i exhibited exquisite specificity to its target luciferin with 50-fold selectivity for DTZ over bis-CTZ (which differ only in a benzylic carbon). Overall, the specificity of our designed luciferases is much greater than native luciferases^{36,37} or previously engineered luciferases³⁸.

The high substrate specificity of LuxSit-i might allow multiplexing of luminescent reporters through substrate-specific or spectrally resolved luminescent signals (Fig. 4d and Fig. 12ab). To explore this possibility, we tracked two independent signaling pathways (cAMP/PKA and NF- κ B) by placing the expression of either RLuc or LuxSit-i downstream

of the NF- κ B or cAMP response element promoters, respectively (Fig. 4e). Imaging in the presence of the substrates for the two luciferases (PP-CTZ for RLuc and DTZ for LuxSit-i) one at a time (Fig. 4f) can clearly distinguish known activators of the two pathways. Because the luminescence of the two reactions occurs at different wavelengths, we were also able to (Fig. 4g) simultaneously assess the activation of the two signaling pathways in the same sample in either intact HEK293T cells or cell lysates (Fig. 12c-e) by providing both substrates together and monitoring luminescence at different wavelengths.

Conclusion

Computational enzyme design to date has been constrained by the number of available scaffolds, which limits the extent to which catalytic configurations and enzyme-substrate shape complementarity can be achieved¹⁶⁻¹⁸. The use of deep-learning to generate large numbers of de novo designed scaffolds here eliminates this restriction; moving forward, the more accurate RoseTTAfold^{TM39} and AlphaFold2^{TM33} should enable still more effective protein scaffold generation by leveraging family-wide hallucination abilities. The diversity of scaffold pocket shapes and sizes enabled the exploration of a range of catalytic geometries and the maximization of substrate-enzyme shape complementarity; to our knowledge, no native luciferases have folds similar to LuxSit, and the two enzymes have high specificity for fully synthetic luciferin substrates that do not exist in nature. With the incorporation of 2-3 substitutions that provide a more complementary pocket to stabilize the transition state, LuxSit-i has higher activity than any previously de novo designed enzyme; the k_{cat}/K_M of 10^6 $M^{-1}s^{-1}$ is in the range of native luciferases. This is a notable advance for computational enzyme design, as tens of rounds of directed evolution were required to obtain catalytic proficiencies in this range for a designed retroaldolase, and the structure was remodeled considerably⁴⁰; in contrast, the predicted differences in ligand-sidechain interactions between LuxSit and LuxSit-i are very subtle. Achieving such high activities directly from the computer remains an outstanding goal for computational enzyme design. The small size of LuxSit makes it well suited as a genetic tag for capacity-limited viral vectors, biosensor development, and fusions to proteins of interest. On the basic science side, the small size, simplicity, and high activity make LuxSit-i an excellent model system for computational and experimental studies aimed at improving understanding of the luciferase catalytic mechanism. Extension of the approach used here to create similarly specific new luciferases for synthetic luciferin substrates beyond DTZ and h-CTZ would considerably extend the multiplexing opportunities illustrated in Fig. 4 or with the microscopy phasor⁴¹, leading to

widely useful multiplexed luminescent toolkits. More generally, our family-wide hallucination method opens up an almost unlimited number of new scaffold possibilities for substrate binding and catalytic residue placement, which is particularly important in cases where the reaction mechanism, and how to promote it, are not completely understood: many structural and catalytic hypotheses can be readily enumerated with different catalytic residue placements in shape and chemically complementary binding pockets. While luciferases are unique in catalyzing the emission of light, the chemical transformation of substrates into products is common to all enzymes, and the approach developed here should be readily extendable to a wide variety of chemical reactions.

Table 3. Photoluminescence properties of selected LuxSit variants

	LQY ^a (%)	k_{cat} (1/s)	K_M (μ M)	V_{max} (photon s^{-1} molecule $^{-1}$)	k_{cat} / K_M ($\times 10^6 M^{-1} s^{-1}$)
LuxSit-i	14.5 $\pm$ 1.8	2.5	2.5	0.36	1.0
LuxSit-f	10.9 $\pm$ 1.5	1.8	8.9	0.20	0.2
RLuc	5.3 ^b	4.9	1.5	0.26	3.3

a. mean $\pm$ s.d., n=3 (technical triplicates). LQY (luminescent quantum yield) measurements were performed by consuming 125 pmol of DTZ (w/ LuxSit-i, and LuxSit-f) or CTZ (w/ RLuc) in 50 μ L PBS with 50 nM corresponding recombinant luciferases. **b.** All LQY values were estimated relative to the reported quantum yield of RLuc⁴². All values were calculated by the assumptions of the simplest Michaelis-Menten kinetics model, excluding potential substrate/product inhibition or enzyme modification during the reaction^{43,44}.

Methods

1. Materials and general methods

Synthetic genes and oligonucleotides were purchased from Integrated DNA Technologies or GenScript. The synthetic gene was inserted between NdeI and XhoI sites of a pET29b+ vector, containing an N-terminal hexahistidine tag followed by a TEV protease cleavage site and a C-terminal stop codon. Restriction endonucleases, Q5 PCR polymerase, and T4 ligase were purchased from NEB. Plasmid DNA, PCR products, or digested fragments were purified by Qiagen DNA purification kits. DNA sequences were analyzed by Genewiz. Coelenterazine (CTZ) was purchased from Gold Biotechnology. Diphenylterazine (DTZ), pyridyl diphenylterazine (8pyDTZ), and Furimazine (FRZ) were purchased from MedChemExpress. All other coelenterazine analogs (bis-CTZ: bisdeoxycoelenterazine; f-CTZ: f-Coelenterazine; e-CTZ: e-Coelenterazine-F; PP-CTZ: methoxy e-Coelenterazine; v-CTZ: v-Coelenterazine. All other chemicals were purchased from Sigma-Aldrich or Fisher

Scientific and used without further purification. To identify the molecular mass of each protein, intact mass spectra were obtained via reverse-phase LC/MS on an Agilent 6230B TOF on an AdvanceBio RP-Desalting column and subsequently deconvoluted by Bioconfirm software using a total entropy algorithm. ÄKTA pure M with UNICORN 6.3.2 Workstation control (GE Healthcare) coupled with a Superdex™ 75 Increase 10/300 GL column was used for size exclusion chromatography. DNA and protein concentrations were determined by a NanoDrop™ small-volume 8 channel UV/vis spectrometer. CD spectra and CD melting experiments were performed by the default setting on a J-1500 Circular Dichroism Spectropolarimeter (Jasco). All luminescence measurements were acquired by a Biotek Synergy Neo2™ Multi-Mode Plate Reader. To convert relative arbitrary unit (RLU) to the number of photons, Neo2 plate reader was calibrated by determining the chemiluminescence of luminol with known quantum yield in the presence of horseradish peroxidase and hydrogen peroxide in K₂CO₃ aqueous solution as previously described⁴⁵. SDS PAGE and luminescence images were captured by a Bio-Rad ChemiDoc™ XRS+. Images were analyzed using the Fiji image analysis software.

2. General procedures for protein production and purification

Lemo21(DE3) strain was used for transformation with the pET29b+ plasmid encoding the gene of interest. Transformed cells were grown for 12 h in TB medium supplemented with kanamycin. Cells were inoculated at 1:50 ratio in 100 mL fresh TB medium, grown at 37 °C for 4 h, and then induced by IPTG for an additional 18 h at 16 °C. Cells were harvested by centrifugation at 4,000g for 10 min and resuspended in 30 mL lysis buffer (20 mM Tris-HCl pH 8.0, 300 mM NaCl, 30 mM imidazole, and Pierce™ Protease Inhibitor Tablets). Cell resuspensions were lysed by sonication for 5 min (10 s per cycle). Lysates were clarified by centrifugation at 24,000g at 12 °C for 40 min and pre-equilibrated with 1 mL of Ni-NTA nickel agarose at 4 °C for 1 h. The resin was washed twice with 10 mL wash buffer and then eluted in 1 mL elution buffer (20 mM Tris-HCl pH 8.0, 300 mM NaCl, 300 mM imidazole). The eluted proteins were purified by size exclusion chromatography in PBS. Fractions were collected based on A280 trace, snap-frozen in liquid nitrogen, and stored at -80 °C,

3. Computational design of idealized scaffolds

Our generation of idealized NTF2-scaffolds can be divided into four parts: (3.1) Generation of seed-structures, (3.2) optimization of backbone geometries using trRosetta™.

based hallucination, (3.3) generation of structure-conditioned sequence models to bias design, (3.4) design and filtering.

3.1 Generation of seed structures

5 We thought to increase the set of NTF2 structures by complementing experimentally resolved structures from the PDB with highly accurate models generated by trRosetta^{TM22}. To achieve this, we first collected 85 NTF2-like protein structures from the PDB based on SCOPe annotation (d. 17.4 SCOPe v2.05). Corresponding sequences were then used as queries to collect sequence homologs from UniProtTM by performing 8 iterations of hhblits at
10 1e-20 e-value cutoff against uniclust30_2018_08 database; default filtering cutoffs were relieved (-maxfilt 100000000 -neffmax 20 -nodiff -realign_max 10000000) to maximize the number of the output hits. All the hits were redundancy reduced using cd-hit⁴⁶ with a sequence identity cutoff of 60% yielding a set of 7,573 candidates for modeling.

To generate inputs for structure modeling with trRosettaTM, we built multiple
15 sequence alignments (MSAs) for each of the 7,573 selected sequences with *hhblits* using a more conservative e-value cutoff of 1e-50; the resulting MSAs were also complemented by hits from *hmmsearch* against uniref100 (release-2019_11) with the bit-score threshold of 115 (i.e. ~1bit per position). After joining the above two sets of alignments and filtering them at 90% sequence identity and 75% coverage cutoffs, only sequences with more than 50
20 homologs in the corresponding MSAs were retained for modeling (2,005 sequences). The filtered MSAs along with information on the top 25 putative structural homologs as identified by *hhsearch* against the PDB100 database of templates were used as inputs to the template-aware version of trRosetta^{TM47} to predict residue pair distances and orientations. Network predictions were then used to reconstruct full atom 3D structure models using a RosettaTM-
25 based folding protocol described previously²².

3.2 Hallucination of idealized NTF2s

Seeking to idealize the native structure seeds, we reasoned that trRosettaTM, a
convolutional residual neural network, which predicts residue-residue orientations and
30 distances from sequence, could serve as a key component in a protein idealizer. Previously, this network has been used to generate diverse proteins that resemble the “ideal” structures of de novo designed proteins by changing the protein sequence to optimize the contrast (KL-divergence) between the predicted geometry and that of randomly generated sequences¹⁹.

For our purpose, the desired fold-space is not diverse but instead focused on the NTF2-like topology. To guarantee generation of ideal structures within this fold-space, we implemented a new fold-specific loss-function, which biased hallucinations based on observed geometries in native crystal structures. As many experimentally characterized NTF2s contain non-ideal regions, we began by creating a set (χ) of trimmed but ideal NTF2s by manually removing non-ideal structural elements such as kinked helices, and long or rarely observed loops. For each seed structure, we then used a structure-based sequence alignment method (see 3.3) to find equivalent positions between the seed structure and χ . Residue pairs were considered to be in a conserved tertiary motif (TERM) if there were 5 or more equivalent positions in χ . The smooth probability distributions based on observed geometries in χ were then computed. For distances we used a Gaussian distribution with mean equal to the true distance denoted by D and standard deviation denoted by σ equal to 0.5 Å. The probability density function for distances d is given by:

$$f(d; D, \sigma) = \frac{1}{\sqrt{2\pi\sigma^2} \exp\left(-\frac{(d-D)^2}{2\sigma^2}\right)}$$

Using this density function one can construct a categorical distribution for binned distances by evaluating this function at the centers of the bins and then normalizing by a sum of all values in different bins. Similarly, a von Mises distribution was used for omega angle smoothing with probability density function given by $f(\omega; \Omega, \kappa) = N(\kappa) \exp[\kappa \cos(\omega - \Omega)]$ where $N(\kappa)$ is a normalizing constant, Ω is the crystal value, κ is the inverse variance chosen to be 100, and ω is the smoothed angle. For phi and theta angles a von Mises-Fisher blur is given by $f(x; \mu, \kappa) = N(\kappa) \exp[\kappa \mu^T x]$ where $N(\kappa)$ is a normalizing constant, μ is a unit vector on a 3D sphere corresponding to the phi and theta angles from the crystal structure, x is a smoothed unit vector, and κ is the inverse variance chosen to be 100.

Next, we converted those probability distributions to energy landscapes (ie - negative log likelihoods) and sought to minimize the expected energy. This soft restraint encouraged the network to seek out the consensus structure, while still allowing deviations where needed. Specifically, we formulated the fold-specific loss as:

$$L_{fold} = \sum_{x \in \{d, \omega, \theta, \phi\}} \left[\sum_{i,j=1}^L \sum_{k=1}^{N_x} -m_{ij} p_{x,ijk} \ln(s_{x,ijk}) \right] / \sum_{i,j=1}^L m_{ij}$$

$$m_{ij} = \{ 1 \text{ if } i \text{ and } j \text{ are in a TERM; else } 0$$

where p is the network prediction and s is the smoothed probability distribution of the conserved residue pairs. For the second part of the loss function and similar to previous work¹⁹, we sought to maximize the Kullback–Leibler (KL) divergence between the predicted probability distribution and a background distribution for all ij residue pairs not in a TERM.

$$L_{hall} = - \sum_{x \in \{d, \omega, \theta, \phi\}} \left[\sum_{i,j=1}^L \sum_{k=1}^{N_x} (1 - m_{ij}) p_{x,ijk} \ln(p_{x,ijk}/b_{x,ijk}) \right] / \sum_{i,j=1}^L (1 - m_{ij})$$

where b is the background distribution and N_x is the number of bins in each probability distribution ($N_d = 37$, $N_{\omega, \theta} = 25$, $N_{\phi} = 13$). Briefly, b is calculated by a network of similar architecture to trRosettaTM trained on the same training data, except it is never given sequence information as an input. The final loss is given by:

$$L = L_{fold} + L_{hall}$$

We used a Markov Chain Monte Carlo (MCMC) procedure to search for sequences that trRosettaTM predicted to fold into structures that minimize this loss function. We allowed four types of moves with different sampling probabilities: mutations ($p=0.55$), insertions ($p=0.15$), deletions ($p=0.15$), and moving segments ($p=0.15$). Mutations randomly changed one amino acid to another, with an equal transition probability for all 20 amino acids. Insertions inserted a new amino acid (all equally likely) into a random location subject to the KL-divergence loss. Deletions deleted a random residue from the same locations. Finally, we also allowed “segments” to move, cutting and pasting themselves from one part of the sequence to another, while maintaining the same overall segment order. Here, a “segment” is a continuous stretch of amino acids all subject to fold specific loss, often composed of a single strand or helix. Starting from a random sequence of an initial length (typically 120 amino acids), we used the standard Metropolis criteria to accept or reject moves:

$$A_i = \min[1, \exp(-(L_i - L_{i-1}) / T)]$$

where A_i is the chance of accepting the move at step i , L_i is the loss at the current step, L_{i-1} is the loss at the previous step and T is the temperature. The temperature started at 0.2 and was reduced by half every 5k steps. Generally, it took 30k steps to converge.

3.3 Structure-conditioned multiple sequence alignment

Given the complexity of the NTF2-like protein fold, we hypothesized that it was necessary to impose sequence design rules to disfavor alternative states (negative design). Towards this end, we computed a structure-conditioned multiple sequence alignment based on native NTF2-like proteins. Specifically, we used TMalign⁴⁸ to superimpose each of the 2005 predicted native structures (from 3.1) onto each hallucinated backbone (from 3.2). Next, to find structurally corresponding positions, we implemented a structure-based dynamic programming algorithm, similar to the Needleman-Wunsch algorithm⁴⁹. However, instead of using the amino acid similarity as the scoring metric, we used a tunable structure-based score function. After aligning the two structures, we scored the structural similarity of any two residues by empirically weighting several metrics: (1) Distance between Ca atoms, (2) differences between backbone torsion angles (phi and psi) backbone torsion angles and (3) the angle (degrees) between the vectors pointing from Ca to Cβ in each residue. To calculate the unweighted score for each component, we normalized each by a maximum possible value (180 degrees for angles and 10Å for distances) and included a “set point” that approximately delineated when we judged a metric to indicate two residues to be more similar than not. Values above this setpoint are positive, indicating two residues are similar and values below the set point indicated two residues are dissimilar.

$$Score_{unweighted} = (set_point - value) / max_value$$

Each value was scaled by its normalized weight and summed to give an overall similarity score between any two amino acids.

These similarity scores were used as the similarity metric in our dynamic programming algorithm, in place of the typical BLOSUM62 similarity metric. We used a gap penalty of 0.1 and an extension penalty of 0.0. Finally, after concatenating all the structure-conditioned aligned sequences, we used PSI-BLAST-cxB^{50,51} to compute sequence redundancy weighted log-odds scores for each amino acid at each position (position-specific scoring matrices, PSSMs).

3.4 Design

To design the resulting backbones, we sought, in addition to the sequence patterns captured in the PSSM (3.3), to further specify the backbone conformation and functionalize the pocket, by installing entire hydrogen bonding networks from native NTF2-like proteins. We compiled two sets of hydrogen bonding networks: a set for the cavity containing 85 networks and another set of networks connecting the C-terminal region of the first helix with

the third beta-strand containing 25 networks. In 20 independent attempts for each backbone, we randomly grafted a network from each set, fixed the identities of hydrogen bonding residues, and designed the sequences for all other positions under PSSM constraints. The resulting models were filtered for various backbone quality metrics and for maintenance of hydrogen bonding networks in the absence of constraints, resulting in a total of 1615 idealized scaffolds.

4 RIFdock tuning files

The hierarchical search framework of RifDock is a powerful way to search through 6-dimensional rigid body orientations. While originally designed to work with physics-based forcefields, the scoring machinery can easily be modified to do other things. A system was added called “Tuning Files” that allows one to tune the energetics of rifdock by “requiring” specific interactions. Specified interactions can range from specific hydrogen bonds, to specific bidentates, and even to specific hydrophobic interactions. The specifics are that during the RifGen stage, each stored rotamer is compared against a list of definitions in the Tuning File. If the rotamer satisfies a definition, it is stored into the RIF with a “Requirement Number”. Later during RifDock, these Requirement Numbers are available during scoring and the presence or absence of certain rotameric interactions may be used to penalize or even completely discard dock solutions. In this work, the Tuning Files were used to require the specific hydrogen bond interactions between the arginine and the secondary amine in the pyrazine ring of the colenterazine-like substrate.

5 Designing theozyme architectures into *de novo* NTF2 scaffolds

De novo design of luciferases can be divided into three main steps – scaffold construction, substrate placement with required interactions, and sequence design. With the idealized NTF2-like scaffolds in hand, we selected 5 diverse rotamers from AIMNet and used the Rotamer Interaction Field (RIF) docking method³¹ to exhaustively search a large space of interacting side chains to the anionic form of DTZ. Chemically, deprotonation of N1 hydrogen is the first step to forming an anionic species (Fig. 5). We first generated RIF using RifGen³¹ to guide placement in the protein scaffolds. We required the placement of a positively charged Arginine sidechain by a tuning file (see below) to stabilize the formation of negative charge N1 atom where the deprotonation occurs initially and enumerated large numbers of possible sidechain interactions with the rest of DTZ.

```

HBOND_DEFINITION
N1 1 ARG
END_HBOND_DEFINITION

5  REQUIREMENT_DEFINITION
1 HBOND N1 1
END_REQUIREMENT_DEFINITION

```

Rifdock was then used to hierarchically search for the best combination of RIF to place on the input backbone. Although the negative charge can move to another electronegative atom O1 via resonance of the imidazopyrazinone core, it is unclear which anionic species is more critical for the luciferase-catalyzed luminescence emission. Thus, we let Rifdock place the polar rotamers on the basis of hydrogen-bond geometry to O1 and apolar rotamers to DTZ without specific requirements. In the next docking step, we parsed the -scaffold_res argument with a list of residue numbers as scaffold backbone positions that were annotated as pocket residues to allow a hierarchical search of RIF placement. We only allowed the RIF placements in the pocket residues and left pre-defined hydrogen bond networks (HB Nets) intact. After Rifdock, we continued for RosettaTM sequence design where the score function was reweighted for higher buried_unsat_penalty⁵² and the amino acid selection was biased by giving a pre-generated PSSM file via SeqprofConsensus task operation. This would minimize buried unsatisfied residues and increase pre-organized architectures in the core that are known to be beneficial for a catalytic pocket⁵³. Two rounds of FastDesign calculation were included: we restricted the RIF rotamers and core HB Nets to repacking in the first round while we allowed other residues for re-design based on PSSM during the Monte Carlo simulated annealing procedure. After the surrounding residues were optimized to retain the RIF interactions, we allowed the re-design of RIF rotamers, to find efficient aromatic and hydrophobic packing around DTZ while catalytic residues (the N1 requirement) were still limited to only repacking. The final set of designs was obtained after filtering by ligand-binding interface energy, shape complementarity, contact molecular surface, number of HbondsToResidue, and the presence of N1_hbond.

6. Structure prediction of LuxSit with AlphaFold2 and comparison to design model

To computationally assess the accuracy of the LuxSit design model, we performed single sequence structure prediction using AlphaFold2. All models were run with 12 recycles and generated models were relaxed using AMBER⁵⁴. The model with the highest pLDDT was used for comparison to the RosettaTM design model and structural superpositions were

performed using the Theseus alignment tool to determine backbone RMSD between the design model and AlphaFold2 model³³.

8. Computational SSM experiment to estimate mutation binding free energy

5 RosettaTM cartesian_ddg application^{55,56} was used to computationally estimate enzyme and substrate binding free energy. The LuxSit design model was relaxed beforehand in cartesian space with the substrate-bound. For the 21 positions that were experimentally screened for single mutation effects on luciferase activity, each residue was computationally mutated into other amino acid types and packing and cartesian relaxation was performed to
10 evaluate the final score in REU. This procedure was applied three times in parallel for both substrate-bound and apo-states. The average of the three calculation results was used to calculate the relative binding free energy (ddG_{bind}) by subtracting the total score of the apo-state from the complex state.

15 9. Construction and screening of designed luciferase libraries

The construction of assembled gene libraries was described previously in detail⁵⁷. In brief, the amino acid sequences of all designed luciferases were first reverse-translated into E. coli codon-optimized DNA sequences. All DNA sequences were categorized into multiple sub-pools by the gene length (~500 designs per sub-pool). Each gene was subsequently split
20 into two fragments (fragment A and fragment B) and added outer and inner primer sequences to the 5' and 3' end (e.g., Outer_oligoA_5primer + design_half_A + Inner_oligoA_3primer and Inner_oligoB_5primer + design_half_B + Outer_oligoB_3primer). All oligos were ordered in one Twist 250nt Oligo Pool. To construct the library of each sub-pool, polymerase chain reaction (PCR) with oligoA_5primer/oligoA_3primer or
25 oligoB_5primer/oligoB_3primer oligonucleotide pairs was used to amplify the individual fragment A or fragment B from each sub-pool. The pool-specific sequences were removed with Uracil Specific Excision Reagent (USER) followed by NEB End Repair kit. Outer primers (oligoA_5primer and oligoB_3primer) were then used for fragment A and fragment B assembly and amplification. The assembled full-length fragment was digested with
30 XhoI/HindIII and ligated into a predigested pBAD/His B vector. All ligation products were used to transform ElectromAXTM DH10B Cells, which were next plated on 150 mm × 15 mm LB agar plates supplemented with carbenicillin and L-arabinose. We sequenced 30 random colonies and 11 of the sequences were in our designed library. The plates (~2000 colonies per plate) were incubated at 37 °C overnight to form bacterial colonies and left at 4 °C for

another 24 h. To directly image luminescence activity from bacterial colonies, we sprayed the PBS solution containing 30 μ M DTZ to each agar plate, waited for 2 min, and the luminescence images were acquired and processed with Bio-Rad ChemiDoc XRS+. After screening 15 plates, active colonies were collected for sequencing, protein expression, and other downstream characterization where LuxSit was selected from three active designs shown catalytic signal above background.

10. Construction and evaluation of LuxSit site saturation mutagenesis libraries

To create libraries of each single amino acid substitution at residues 13, 14, 17, 18, 35, 37, 38, 49, 52, 53, 56, 60, 65, 81, 83, 94, 96, 98, 100, 110, and 112, forward oligos mixture with degenerate codons (NDT, VHG, and TGG = 1:1:0.1 ratio) and an overlapped reverse oligo were used to amplify the plasmid of LuxSit. The resulting PCR products were circularized by Gibson Assembly protocol and were subsequently used to transform ElectroMAX™ DH10B Cells. The cells were plated on 150 mm $\times$ 15 mm LB agar plates supplemented with carbenicillin and L-arabinose, incubated at 37 °C overnight, and left at 4 °C for another 24 h. As described in the screening of luciferase libraries, colony-based screening by spraying DTZ solution was used to identify active colonies. Inactive colonies were also randomly picked. As a result, a total of 32 colonies were picked for each residue library. 32 $\times$ 21 individual colonies were grown in 1 mL of TB supplemented with carbenicillin and L-arabinose in 96-well deep-well culture plates. The plates were shaken at 37 C overnight (~16-18 h) on 96-well plate shakers at 1,100 rpm. Cells were pelleted by centrifugation at 4,000g for 15 min in a tabletop centrifuge. Media was discarded and the cell pellets were resuspended in 0.2 mL BugBuster HT Protein Extraction buffer. The plates were transferred back to 96-well plate shakers and incubated at 1,100 rpm for an additional 30 min. Cellular debris was pelleted again by centrifugation at 4,000g for 15 min, soluble lysates were transferred to a new semi-deep 96-well plate, and incubated with 10 μ L of magnetic Ni-NTA beads for 30 min to allow binding. The magnetic extractor was used to first transfer the beads from the binding plates to wash plates with 200 μ L IMAC wash buffer in each well, and then transfer the beads to elution plates containing 30 μ L IMAC elution buffer in each well. The concentrations of all proteins in each well were determined by the Bradford assay directly. The elution solution in each well was used to make a 25 μ L protein solution at indicated concentration and mixed with 25 μ L of 50 μ M DTZ PBS solution. The luminescence signals were acquired over a course of 15 min while the actual point mutation was identified by sequencing. Thus, the mutation-to-activity relationship can be mapped. To

evaluate whether these beneficial mutations are synergistic, we ordered individual mutants with combinatorial mutations at residue 14, 60, 96, 98, and 110 (see **Table 4**), expressed, and purified these LuxSit variants for kinetic, emission spectra, and luminescence intensity. We identified four mutants that can produce 47 to 77-fold more photons than the parent LuxSit.

- 5 We assigned one of which, LuxSit-f (A96M/M110V), for its strong initial flash emission. Since the mutations at residue 96 and 110 are robust and mutations at residue 60 are versatile, we generated a fully randomized library at 60, 96, and 110 positions to exhaustively screen all possible combinations. After the colony-based screening, we identified many colonies with strong luciferase activities with DTZ (**Fig. 9**). Among all selected mutants, Arg60 is
10 confirmed to be mutable, Ala96 prefers larger hydrophobic sidechains (Leu, Ile, Met, and Cys), and Met110 favors hydrophobic residues (Val, Ile, and Ala). A newly discovered mutant R60S/A96L/M110V with more than 100-fold higher photon flux over LuxSit was assigned LuxSit-i for its high brightness.

15 11. *In vitro* characterization of photoluminescence properties

- For Michaelis–Menten kinetics measurements, 25 μ L of serial diluted DTZ substrate in Tris pH 8.0 buffer was added into the wells of a white 96-well half-area microplate containing 25 μ L of purified luciferases (final enzyme concentration: 100 nM; substrate concentration: 0.78 to 50 μ M). Measurements were taken every 1 min (0.1 s integration and
20 10 s shaking between each interval) for a total of 20 min. Initial velocities were estimated as the average of the light intensities from the first three data points to fit the Michaelis-Menten equation. All relative arbitrary unit (RLU) per second values were converted to photon/s by the luminol-H₂O₂-HRP calibration method⁴⁵. Following the equation: $I_{\max} = \text{LQY} \times k_{\text{cat}} \times [\text{E}]$, $I_{\max}$ is the maximal photon flux (photon s⁻¹), [E] is the total enzyme concentration, and $V_{\max}$
25 is the maximum photon flux per molecule (photon s⁻¹ molecule⁻¹) from the fitting of the Michaelis-Menten equation. To determine the luminescent quantum yields, 25 μ L of 5 μ M individual substrate in PBS was injected into 25 μ L PBS containing 100 nM corresponding luciferase. DTZ was used for all LuxSit variants while CTZ was used as the substrate of native RLuc. The luminescence signals were monitored until the reactions were completed
30 (0.1 s integration and measurements were taken every 5 s for a total of 40 min). The sum of luminescence photon counts was normalized to the total photon counts of RLuc/CTZ pair (LQY = $5.3 \pm 0.1\%$)⁵⁸ to derive relative luminescent quantum yields of LuxSit variants (**Fig. 10c**). k_{cat} values for each individual enzyme were calculated using the equation: $k_{\text{cat}} = V_{\max} / \text{LQY}$. To record emission spectra, 25 μ L of 50 μ M DTZ in PBS were injected into 25 μ L of

200 nM pure luciferases and the emission spectra were collected with 0.1 s integration and 2 nm increments from 300 to 700 nm. *In vitro* luminescence activity measurements of LuxSit-i expressing HEK293T or HeLa cells were done similarly as 15,000 intact cells or lysates were used in the assay instead of purified luciferases. To evaluate the substrate specificity, 25 μ L of 50 μ M substrate analogs in PBS were added to 25 μ L of 200 nM indicated luciferases, and the signals were recorded over 20 min. Data were shown as the total luminescence signal over the first 10 min. We normalized the data by setting the highest emission substrate at 100%.

12. Circular dichroism (CD)

Purified protein samples were prepared at 15 μ M in pH 7.4 10 mM phosphate buffer. Spectra from 190 nm to 260 nm were recorded at 25 °C, 50 °C, 75 °C, 95 °C, and after cooling back to 25 °C. Thermal denaturation was monitored at 220 nm from 25 °C to 95 °C (1°C per min increments). T_m values were not reported because no obvious inflection points of the melting curves.

13. Mammalian cell culture and transfection

HEK293T and HeLa cell lines were maintained at 37 °C with humidified 5% CO₂ atmosphere and cultured in Dulbecco's Modified Eagle's Medium (DMEM, GIBDO) supplemented with 10% fetal bovine serum (FBS, Sigma). Cells were transfected with Turbofectin™ 8.0 (Origene) with 500 μ g of plasmid DNA. After 24 h at 37 °C in a CO₂ incubator, the medium was removed, and cells were collected and resuspended in Dulbecco's phosphate-buffered saline (DPBS).

14. Fluorescence Microscopy and image analysis

Cells were washed twice with HBSS and subsequently imaged in HBSS in the dark at 37°C. Right before imaging, cells were incubated with 25 μ M DTZ. Epifluorescence imaging was conducted on a Yokogawa CSU-X1 microscope equipped with a Hamamatsu ORCA-Fusion scientific CMOS camera and Lumencor Celesta light engine. Objectives used were: 10 $\times$, NA 0.45, WD 4.0 mm, 20 $\times$, NA 1.4, WD 0.13 mm, and 40 $\times$, NA 0.95, WD 0.17–0.25 mm with correction collar for cover glass thickness (0.11 mm to 0.23 mm) (Plan Apochromat Lambda). Imaging for BFP utilized a 408 nm laser, 432/36 nm dichroic, and a 440/40 nm emission filter (Semrock). Exposure times were 200 ms for BFP and 10 s for luminescence. All epifluorescence experiments were subsequently analyzed using NIS Elements software.

15. Multiplex dual-luciferase reporter assay for the cAMP/PKA and NF- κ B pathways

HEK293T cells were grown in a tissue culture-grade white 96-well plate and transfected with indicated CRE-RLuc, NF κ B-LuxSit-i, and CMV-CyOFP plasmids. 24 h after transfection, the medium was replaced by 2 μ M of Forskolin (FSK) or 300 ng/mL human tumor necrosis factor alpha (TNF α) in regular cell media. 23 h after stimulation, the cells were resuspended in DPBS by pipette mixing. 25 μ L of DPBS containing 30,000 intact cells was mixed with 25 μ L of CellLytic M for 15 min to make cell lysates. For intact cell assay, 25 μ L of DPBS containing 15,000 intact cells was mixed with 25 μ L of PP-CTZ (2 μ M) or/and DTZ (10 μ M) in DPBS. For cell lysate assay, 25 μ L of cell lysate was added to 25 μ L of PP-CTZ (2 μ M) or/and DTZ (10 μ M) to initiate luminescence reactions. The signals were recorded every 1 min for a total of 10 min. The light signals were collected in the substrate-resolved mode without filters and with 528/20 and 390/35 filters under the spectrally resolved mode. Area scanning the fluorescence intensity of CyOFP at 480 nm (excitation wavelength) and 580 nm (emission wavelength) was used to estimate the total cell numbers and transfection efficiency. The reported unit was the average of the first 10 min luminescence (RLU) over the relative fluorescence units (a.u.). To derive fold-of-activation, all values were normalized to the corresponding non-stimulated control.

Statistical analysis

No statistical methods were used to pre-determine the sample size. No sample was excluded from data analysis. Results were reproduced using different batches of pure proteins on different days. Unless otherwise indicated, data are shown as mean $\pm$ s.d., and error bars in figures represent s.d. of technical triplicate. Data were analyzed and plotted using GraphPad Prism 8, seaborn, and matplotlib.

Supplementary Information

Table 4. Enzymatic and photoluminescence properties of LuxSit and its mutants

	$\lambda_{\max}$ (nm)	K_M (μ M) ^a	$V_{\max}$ ($\times 10^{-2}$ photon s ⁻¹ molecule ⁻¹) ^a	Relative emission intensity ^b
LuxSit (60R/96A/98H/110M)	478	18.3(14.1-21)	0.32(0.29-0.37)	1
Y14E	478	12.6(9.9-16.1)	0.19(0.17-0.21)	0.95
R60C	476	45.2(28.7-78.9)	0.57(0.44-0.8)	1.55
A96I	476	17.5(12.1-25.9)	1.3(1.1-1.5)	4.32
A96M	476	4.8(3.1-7.3)	6.0(5.3-6.9)	16.4

H98N	476	29.3(22.1-39.8)	0.17(0.15-0.2)	0.48
M110C	478	67.5(43-121)	0.4(0.3-0.6)	0.64
M110V	480	30.3(22.7-41.5)	7.1(6.2-8.4)	19.3
60C/96I/98H/110C	476	11.3(6.5-20.2)	0.12(0.1-0.16)	0.39
60R/96I/98N/110C	476	8.7(3-7.4)	0.02(0.017-0.022)	0.12
60R/96I/98H/110C	480	4.3(2.7-6.8)	0.05(0.045-0.059)	0.27
60C/96I/98H/110V	478	6.3(4.8-8.5)	14.7(13.4-16.2)	77.1
60C/96M/98N/110V	474	9.7(5.6-17.4)	0.13(0.1-0.16)	0.5
60C/96M/98H/110C	478	30.9(18.3-57.2)	2.44(1.9-3.4)	4.26
60R/96I/98N/110V	476	7.2(5.4-9.5)	0.32(0.29-0.35)	1.39
60R/96I/98H/110V	482	10.8(8-14.6)	12.9(11.7-14.5)	47.5
60C/96M/98H/110V	476	7.0(4.8-10.2)	15.9(14-18.1)	66.3
14E/60C/96M/98H/110V	478	8.0(4.0-15.8)	0.16(0.12-0.2)	1.0
60R/96M/98N/110V	478	8.0(5.3-12.2)	0.044(0.04-0.05)	0.21
60C/96M/98N/110C	476	17.4(13.4-22.8)	0.016(0.015-0.018)	0.08
60C/96I/98N/110C	476	24.7(18.7-33.4)	0.03(0.026-0.035)	0.11
60R/96M/98H/110C	474	23.4(15.9-35.8)	0.28(0.24-0.35)	0.58
60C/96I/98N/110V	480	80.9(43.3-222)	0.83(0.55-1.8)	1.81
60R/96M/98H/110V	480	8.9(6.7-11.9)	19.7(17.9-21.9)	60.9
(LuxSit-f)				
60R/96M/98N/110C	476	9.3(7-12.2)	0.019(0.017-0.021)	0.96
60S/96L/98H/110V	482	2.5(1.9-3.3)	36.1(33.7-38.6)	153
(LuxSit-i)				

a. mean (95% CI), n=3 (technical triplicates); b. Integrated luminescence intensity over the first 20 min. All values were normalized to the signal of LuxSit with 25 μ M DTZ in Tris pH 8.0 buffer.

- 5 **Table 5.** Amino acid and DNA sequences of *de novo* luciferases in this work
The sequences (underlined) below contain a PolyHis-TEV or PolyHis tag for protein purification (which are optional and may be present or deleted)

>LuxSit
MGSHHHHHHSGSENLYFQGMSEEQIRQFLRRFYALDSDGADTAASLFHFGVTIHLWDGVTFTS
REEFWERFLFSTRKDAQREIKSLEVRGDTVEVHVQLHATHNGQKHTVDATHHWHFRGNRVTEMR
VHINPTG (SEQ ID NO: 192)
>LuxSit (codon optimized for E. coli expression)
ATGGGTAGCCATCACCACCATCACCATGGTAGCGGTAGCGAGAACTTGTACTTCCAAGGAAGCATG
AGCGAAGAAGAGATTTCGTGAGTTTCTGCGTCGTTTTATGAAGCGCTGGATAGCGGCGATGCGGAT
ACCGCTGCGAGCCTGTTTCATCCGGGCGTGACAATTCATCTGTGGGATGGCGTTACCTTTACCAGC
CGTGAAGAATTTTCGTGAATGGTTTGAACGTCTGTTAGCACCCGTAAAGATGCGCAGCGTGAAATT
AAGAGCCTGGAAGTACGTGGCGATACCGTGGAAGTGCATGTGCAGTTGCACGCGACCCATAATGGC
CAGAAACATACCGTAGATGCAACCCATCATTGGCATTTCGTGGCAATCGTGTGACCGAAATGCGT
GTGCATATCAATCCGACCGCTAA (SEQ ID NO: 193)
>LuxSit-f
MGSHHHHHHSGSENLYFQGMSEEQIRQFLRRFYALDSDGADTAASLFHFGVTIHLWDGVTFTSR
EEFWERFLFSTRKDAQREIKSLEVRGDTVEVHVQLHATHNGQKHTVDMTHHWHFRGNRVTEVRV
HINPTG (SEQ ID NO: 194)
> LuxSit-f (codon optimized for E. coli expression)
ATGGGTAGCCATCACCACCATCACCATGGTAGCGGTAGCGAGAACTTGTACTTCCAAGGAATGAGC
GAAGAAGAGATTTCGTGAGTTTCTGCGTCGTTTTATGAAGCGCTGGATAGCGGCGATGCGGATACC
GCTGCGAGCCTGTTTCATCCGGGCGTGACAATTCATCTGTGGGATGGCGTTACCTTTACCAGCCGT

GAAGAATTTTCGTGAATGGTTTGAACGTCTGTTTAGCACCCGCAAAGATGCGCAGCGTGAAATTAAG AGCCTGGAAGTACGTGGCGATACCGTGGGAAGTGCATGTGCAGTTGCACGCGACCCATAATGGCCAG AAACATACCGTAGATATGACCCATCATTGGCATTTCGTGGCAATCGTGTGACCGAAGTGCCTGTG CATATCAATCCGACCGGCTAA (SEQ ID NO: 195)
>LuxSit-i
MGSHHHHHHGSGSENLYFQGMSEEQIRQFLRRFYEALDSGDADTAASLFHPGVTIHLWDGVTFTSR EEFREWFERLESTSKDAQREIKSLEVRGDTVEVHVQLHATHNGQKHTVDLTHHHWHFRGNRVTEVRV HINPTG (SEQ ID NO: 196)
>LuxSit-i (codon optimized for E. coli expression)
ATGGGTAGCCATCACCACCATCACCATGGTAGCGGTAGCGAGAACTTGTACTTCCAAGGAATGAGC GAAGAACAGATTTCGTCACTTTCTGCGTCGTTTTTATGAAGCGCTGGATAGCGGCGATGCGGATACC GCTGCGAGCCTGTTTCATCCGGGCGTGACAATTCATCTGTGGGATGGCGTTACCTTTACCAGCCGT GAAGAATTTTCGTGAATGGTTTGAACGTCTGTTTAGCACCAAGTAAAGATGCGCAGCGTGAAATTAAG AGCCTGGAAGTACGTGGCGATACCGTGGGAAGTGCATGTGCAGTTGCACGCGACCCATAATGGCCAG AAACATACCGTAGATTGACCCATCATTGGCATTTCGTGGCAATCGTGTGACCGAAGTTCGTGTG CATATCAATCCGACCGGCTAA (SEQ ID NO: 197)
>LuxSit-i (codon optimized for human cell expression)
ATGAGCGAGGAGCAGATCAGACAGTTCTGAGGAGATTCTACGAGGCCCTGGATAGCGGAGACGCC GACACAGCTGCCAGCCTTTTCCATCCTGGCGTGACCATCCACCTGTGGGACGGCGTCACCTTCACT AGCAGGGAGGAGTTTCAGGGAGTGGTTCGAGAGACTGTTTCAGCACCAGCAAGGACGCCCCAGAGAGAG ATCAAGAGCCTGGAAGTTAGAGGCGACACCGTGGAAAGTGCACGTGCAGCTGCACGCCACACACAAC GGACAGAAGCACACCGTCGACCTGACCCACCACTGGCACTTCAGAGGCAACAGAGTGACCGAGGTG AGAGTGACATCAATCCACCGGCTAA (SEQ ID NO: 198)

References

1. Jiang, L. *et al.* De novo computational design of retro-aldol enzymes. *Science* **319**, 1387–1391 (2008).
2. Rothlisberger, D. *et al.* Kemp elimination catalysts by computational enzyme design. *Nature* **453**, 190–195 (2008).
3. Yeh, H. W. *et al.* Red-shifted luciferase-luciferin pairs for enhanced bioluminescence imaging. *Nat. Methods* **14**, 971–974 (2017).
4. Love, A. C. & Prescher, J. A. Seeing (and Using) the Light: Recent Developments in Bioluminescence Technology. *Cell Chemical Biology* **27**, 904–920 (2020).
5. Syed, A. J. & Anderson, J. C. Applications of bioluminescence in biotechnology and beyond. *Chem. Soc. Rev.* **50**, 5668–5705 (2021).
6. Yeh, H.-W. & Ai, H.-W. Development and Applications of Bioluminescent and Chemiluminescent Reporters and Biosensors. *Annu. Rev. Anal. Chem.* **12**, 129–150 (2019).
7. Zambito, G., Chawda, C. & Mezzanotte, L. Emerging tools for bioluminescence imaging. *Curr. Opin. Chem. Biol.* **63**, 86–94 (2021).

8. Markova, S. V., Larionova, M. D. & Vysotski, E. S. Shining Light on the Secreted Luciferases of Marine Copepods: Current Knowledge and Applications. *Photochem. Photobiol.* **95**, 705–721 (2019).
9. Wu, N. *et al.* Solution structure of Gaussia Luciferase with five disulfide bonds and
5 identification of a putative coelenterazine binding cavity by heteronuclear NMR. *Sci. Rep.* **10**, (2020).
10. Jiang, T. Y., Du, L. P. & Li, M. Y. Lighting up bioluminescence with coelenterazine: strategies and applications. *Photochem. Photobiol. Sci.* **15**, 466–480 (2016).
11. Shakhmin, A. *et al.* Coelenterazine analogues emit red-shifted bioluminescence with
10 NanoLuc. *Org. Biomol. Chem.* **15**, 8559–8567 (2017).
12. Michelini, E. *et al.* Spectral-resolved gene technology for multiplexed bioluminescence and high-content screening. *Anal. Chem.* **80**, 260–267 (2008).
13. Rathbun, C. M. *et al.* Parallel screening for rapid identification of orthogonal bioluminescent tools. *ACS Cent. Sci.* **3**, 1254–1261 (2017).
- 15 14. Yeh, H.-W., Wu, T., Chen, M. & Ai, H.-W. Identification of Factors Complicating Bioluminescence Imaging. *Biochemistry* **58**, 1689–1697 (2019).
15. Su, Y. C. *et al.* Novel NanoLuc substrates enable bright two-population bioluminescence imaging in animals. *Nat. Methods* **17**, 852–860 (2020).
16. Lombardi, A., Pirro, F., Maglio, O., Chino, M. & DeGrado, W. F. De Novo design of
20 four-helix bundle metalloproteins: One scaffold, diverse reactivities. *Acc. Chem. Res.* **52**, 1148–1159 (2019).
17. Chino, M. *et al.* Artificial diiron enzymes with a DE Novo designed four-helix bundle structure. *Eur. J. Inorg. Chem.* **2015**, 3352–3352 (2015).
18. Basler, S. *et al.* Efficient Lewis acid catalysis of an abiological reaction in a de novo
25 protein scaffold. *Nat. Chem.* **13**, 231–235 (2021).
19. Anishchenko, I. *et al.* De novo protein design by deep network hallucination. *Nature* (2021) doi:10.1038/s41586-021-04184-w.
20. Wang, J. *et al.* Scaffolding protein functional sites using deep learning. *Science* **377**, 387–394 (2022).
- 30 21. Norm, C. *et al.* Protein sequence design by conformational landscape optimization. *Proc. Natl. Acad. Sci. U. S. A.* **118**, (2021).
22. Yang, J. Y. *et al.* Improved protein structure prediction using predicted interresidue orientations. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 1496–1503 (2020).

23. Basanta, B. *et al.* An enumerative algorithm for de novo design of proteins with diverse pocket structures. *Proc. Natl. Acad. Sci. U. S. A.* **117**, 22135–22145 (2020).
24. Loening, A. M., Fenn, T. D. & Gambhir, S. S. Crystal structures of the luciferase and green fluorescent protein from *Renilla reniformis*. *J. Mol. Biol.* **374**, 1017–1028 (2007).
- 5 25. Tomabechi, Y. *et al.* Crystal structure of nanoKAZ: The mutated 19 kDa component of *Oplophorus* luciferase catalyzing the bioluminescent reaction with coelenterazine. *Biochem. Biophys. Res. Commun.* **470**, 88–93 (2016).
26. Ding, B. W. & Liu, Y. J. Bioluminescence of Firefly Squid via Mechanism of Single Electron-Transfer Oxygenation and Charge-Transfer-Induced Luminescence. *J. Am. Chem. Soc.* **139**, 1106–1119 (2017).
- 10 27. Isobe, H., Yamanaka, S., Kuramitsu, S. & Yamaguchi, K. Regulation mechanism of spin-orbit coupling in charge-transfer-induced luminescence of imidazopyrazinone derivatives. *J. Am. Chem. Soc.* **130**, 132–149 (2008).
28. Kondo, H. *et al.* Substituent effects on the kinetics for the chemiluminescence reaction of 6-arylimidazo[1,2-a]pyrazin-3(7H)-ones (Cypridina luciferin analogues): support for the single electron transfer (SET)-oxygenation mechanism with triplet molecular oxygen. *Tetrahedron Lett.* **46**, 7701–7704 (2005).
- 15 29. Branchini, B. R. *et al.* Experimental Support for a Single Electron-Transfer Oxidation Mechanism in Firefly Bioluminescence. *J. Am. Chem. Soc.* **137**, 7592–7595 (2015).
- 20 30. Zubatyuk, R., Smith, J. S., Leszczynski, J. & Isayev, O. Accurate and transferable multitask prediction of chemical properties with an atoms-in-molecules neural network. *Science Advances* **5**, (2019).
31. Dou, J. Y. *et al.* De novo design of a fluorescence-activating beta-barrel. *Nature* **561**, 485–491 (2018).
- 25 32. Cao, L. *et al.* Design of protein-binding proteins from the target structure alone. *Nature* **605**, 551–560 (2022).
33. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596**, 583–+ (2021).
34. Dauparas, J. *et al.* Robust deep learning based protein sequence design using ProteinMPNN. *bioRxiv* (2022) doi:10.1101/2022.06.03.494563.
- 30 35. Yeh, H.-W. *et al.* ATP-Independent Bioluminescent Reporter Variants To Improve in Vivo Imaging. *ACS Chem. Biol.* **14**, 959–965 (2019).
36. Bhaumik, S. & Gambhir, S. S. Optical imaging of *Renilla* luciferase reporter gene expression in living mice. *Proc. Natl. Acad. Sci. U. S. A.* **99**, 377–382 (2002).

37. Szent-Gyorgyi, C., Ballou, B. T., Dagnal, E. & Bryan, B. Cloning and characterization of new bioluminescent proteins. in *Biomedical Imaging: Reporters, Dyes, and Instrumentation* (eds. Bornhop, D. J., Contag, C. H. & Sevick-Muraca, E. M.) (SPIE, 1999). doi:10.1117/12.351015.
- 5 38. Hall, M. P. *et al.* Engineered luciferase reporter from a deep sea shrimp utilizing a novel imidazopyrazinone substrate. *ACS Chem. Biol.* **7**, 1848–1857 (2012).
39. Back, M. *et al.* Accurate prediction of protein structures and interactions using a three-track neural network. *Science* **373**, 871–+ (2021).
40. Giger, L. *et al.* Evolution of a designed retro-aldolase leads to complete active site remodeling. *Nat. Chem. Biol.* **9**, 494–498 (2013).
- 10 41. Yao, Z. *et al.* Multiplexed bioluminescence microscopy via phasor analysis. *Nat. Methods* **19**, 893–898 (2022).
42. Loening, A. M., Dragulescu-Andrasi, A. & Gambhir, S. S. A red-shifted Renilla luciferase for transient reporter-gene expression. *Nat. Methods* **7**, 5–6 (2010).
- 15 43. Dijkema, F. M. *et al.* Flash properties of Gaussia luciferase are the result of covalent inhibition after a limited number of cycles. *Protein Sci.* **30**, 638–649 (2021).
44. Schenkmyerova, A. *et al.* Engineering the protein dynamics of an ancestral luciferase. *Nat. Commun.* **12**, (2021).
45. Ando, Y. *et al.* Development of a quantitative bio/chemiluminescence spectrometer determining quantum yields: Re-examination of the aqueous luminol chemiluminescence standard. *Photochem. Photobiol.* **83**, 1205–1210 (2007).
- 20 46. Li, W. & Godzik, A. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics* **22**, 1658–1659 (2006).
47. Farrell, D. P. *et al.* Deep learning enables the atomic structure determination of the Fanconi Anemia core complex from cryoEM. *IUCrJ* **7**, 881–892 (2020).
- 25 48. Zhang, Y. & Skolnick, J. TM-align: a protein structure alignment algorithm based on the TM-score. *Nucleic Acids Res.* **33**, 2302–2309 (2005).
49. Needleman, S. B. & Wunsch, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**, 443–453 (1970).
- 30 50. Oda, T., Lim, K. & Tomii, K. Simple adjustment of the sequence weight algorithm remarkably enhances PSI-BLAST performance. *BMC Bioinformatics* **18**, 288 (2017).
51. Altschul, S. F. *et al.* Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res.* **25**, 3389–3402 (1997).

52. Coventry, B. & Baker, D. Protein sequence optimization with a pairwise decomposable penalty for buried unsatisfied hydrogen bonds. *PLoS Comput. Biol.* **17**, (2021).
53. Smith, A. J. T. *et al.* Structural Reorganization and Preorganization in Enzyme Active Sites: Comparisons of Experimental and Theoretically Ideal Active Site Geometries in the Multistep Serine Esterase Reaction Cycle. *J. Am. Chem. Soc.* **130**, 15361–15373 (2008).
54. Salomon-Ferrer, R., Case, D. A. & Walker, R. C. An overview of the Amber biomolecular simulation package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* **3**, 198–210 (2013).
55. Kellogg, E. H., Leaver-Fay, A. & Baker, D. Role of conformational sampling in computing mutation-induced changes in protein structure and stability. *Proteins* **79**, 830–838 (2011).
56. Park, H. *et al.* Simultaneous optimization of biomolecular energy functions on features from small molecules and macromolecules. *J. Chem. Theory Comput.* **12**, 6201–6212 (2016).
57. Klein, J. C. *et al.* Multiplex pairwise assembly of array-derived DNA oligonucleotides. *Nucleic Acids Res.* **44**, (2016).
58. Loening, A. M., Wu, A. M. & Gambhir, S. S. Red-shifted Renilla reniformis luciferase variants for imaging in living subjects. *Nat. Methods* **4**, 641–643 (2007).
59. Liang, J., Feng, X., Hait, D. & Head-Gordon, M. Revisiting the performance of time-dependent density functional theory for electronic excitations: Assessment of 43 popular and recently developed functionals from rungs one to four. *J. Chem. Theory Comput.* **18**, 3460–3473 (2022).
60. Chai, J.-D. & Head-Gordon, M. Long-range corrected hybrid density functionals with damped atom-atom dispersion corrections. *Phys. Chem. Chem. Phys.* **10**, 6615–6620 (2008).
61. Ditchfield, R., Hehre, W. J. & Pople, J. A. Self-consistent molecular-orbital methods. IX. An extended Gaussian-type basis for molecular-orbital studies of organic molecules. *J. Chem. Phys.* **54**, 724–728 (1971).
62. Grimme, S. Exploration of chemical compound, conformer, and reaction space with meta-dynamics simulations based on tight-binding quantum chemical calculations. *J. Chem. Theory Comput.* **15**, 2847–2862 (2019).

63. Pracht, P., Bohle, F. & Grimme, S. Automated exploration of the low-energy chemical space with fast quantum chemical methods. *Phys. Chem. Chem. Phys.* **22**, 7169–7192 (2020).
64. Luchini, G., Alegre-Requena, J. V., Funes-Ardoiz, I. & Paton, R. S. GoodVibes:
5 automated thermochemistry for heterogeneous computational chemistry data. *F1000Res.* **9**, 291 (2020).
65. Li, Y.-P., Gomes, J., Mallikarjun Sharada, S., Bell, A. T. & Head-Gordon, M. Improved force-field parameters for QM/MM simulations of the energies of adsorption for molecules in zeolites and a free rotor correction to the rigid rotor harmonic oscillator
10 model for adsorption enthalpies. *J. Phys. Chem. C Nanomater. Interfaces* **119**, 1840–1850 (2015).
66. Götz, A. W. *et al.* Routine microsecond molecular dynamics simulations with AMBER on GPUs. I. Generalized Born. *J. Chem. Theory Comput.* **8**, 1542–1555 (2012).
67. Becke, A. D. Density-functional thermochemistry. III. The role of exact exchange. *J.*
15 *Chem. Phys.* **98**, 5648–5652 (1993).
68. Grimme, S., Antony, J., Ehrlich, S. & Krieg, H. A consistent and accurate ab initio parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu. *J. Chem. Phys.* **132**, 154104 (2010).
69. Grimme, S., Ehrlich, S. & Goerigk, L. Effect of the damping function in dispersion
20 corrected density functional theory. *J. Comput. Chem.* **32**, 1456–1465 (2011).
70. Meiler, J. & Baker, D. ROSETTALIGAND: Protein-small molecule docking with full side-chain flexibility. *Proteins* **65**, 538–548 (2006).
71. Davis, I. W. & Baker, D. RosettaLigand docking with full ligand and receptor flexibility. *J. Mol. Biol.* **385**, 381–392 (2009).
- 25 72. Davis, I. W., Raha, K., Head, M. S. & Baker, D. Blind docking of pharmaceutically relevant compounds using RosettaLigand. *Protein Sci.* **18**, 1998–2002 (2009).
73. Wang, J., Wolf, R. M., Caldwell, J. W., Kollman, P. A. & Case, D. A. Development and testing of a general amber force field. *J. Comput. Chem.* **25**, 1157–1174 (2004).
74. Bayly, C. I., Cieplak, P., Cornell, W. & Kollman, P. A. A well-behaved electrostatic
30 potential based method using charge restraints for deriving atomic charges: the RESP model. *J. Phys. Chem.* **97**, 10269–10280 (1993).
75. Besler, B. H., Merz, K. M. & Kollman, P. A. Atomic charges derived from semiempirical methods. *J. Comput. Chem.* **11**, 431–439 (1990).

76. Singh, U. C. & Kollman, P. A. An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* **5**, 129–145 (1984).
77. Jorgensen, W. L., Chandrasekhar, J., Madura, J. D., Impey, R. W. & Klein, M. L. Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **79**, 926–935 (1983).
- 5 78. Maier, J. A. *et al.* Ff14SB: Improving the accuracy of protein side chain and backbone parameters from ff99SB. *J. Chem. Theory Comput.* **11**, 3696–3713 (2015).
79. Darden, T., York, D. & Pedersen, L. Particle mesh Ewald: An $N \cdot \log(N)$ method for Ewald sums in large systems. *J. Chem. Phys.* **98**, 10089–10092 (1993).
- 10 80. Roe, D. R. & Cheatham, T. E., III. PTRAJ and CPPTRAJ: Software for processing and analysis of molecular dynamics trajectory data. *J. Chem. Theory Comput.* **9**, 3084–3095 (2013).

We claim

1. A protein having luciferase activity, comprising the secondary structure arrangement H1-L1-H2-L2-E1-L3-E2-L4-H3-L5-E3-L6-E4-L7-E5-L8-E6, wherein "H" is a helical domain, "L" is a loop domain, and "E" is a beta strand domain; wherein:
 - (a) the H1 domain is at least 18 or 19 amino acids in length; residue 14 of the H1 domain is Y, D, or E, and residue 18 of the H1 domain is D or E;
 - (b) the E3 domain is at least 6, 7, 8, 9, or 10 amino acids in length and residue 2 of the E3 domain is R; and
 - (c) the E5 domain is at least 10, 11, 12, 13, or 14 amino acids in length and residue 9 of the E5 domain is H or N.
2. The protein of claim 1, wherein residue 7 of the E5 domain is M.
3. The protein of claim 1 or 2 wherein the E6 domain is at least 9, 10, 11, 12, or 13 amino acids in length and wherein residue 5 of the E6 domain is V.
4. The protein of any one of claims 1-3, wherein residue 1 of the L5 domain is S.
5. The protein of any one of claims 3-4, wherein residue 7 of the E5 domain is M and residue 5 of the E6 domain is V.
6. The protein of any one of claims 3-4, wherein residue 7 of the E5 domain is M, residue 5 of the E6 domain is V, and residue 1 of the L5 domain is S.
7. The protein of any one of claims 1-6, wherein the H2 domain is at least 5, 6, or 7 amino acids in length, the H3 domain is at least 9, 10, 11, 12, 13, or 14 amino acids in length, the E1 domain is at least 3 or 4 amino acids in length, the E2 domain is at least 3 or 4 amino acids in length, and/or the E4 domain is at least 8, 9, 10, 11, or 12 amino acids in length.
8. The protein of any one of claims 1-7, wherein:
 - the H1 domain is 19 amino acids in length;
 - the H2 domain is 7 amino acids in length;
 - the E1 domain is 4 amino acids in length;

- the E2 domain is 4 amino acids in length;
 the H3 domain is 14 amino acids in length;
 the E3 domain is 10 amino acids in length;
 the E4 domain is 12 amino acids in length;
 5 the E5 domain is 14 amino acids in length; and
 the E6 domain is 12 or 13 amino acids in length.

9. The protein of any one of claims 1-8, wherein 1, 2, 3, 4, or all 5 of the following is true:
- 10 (a) residue 13 of domain H1 is F;
 (b) residue 1 of domain L3 is W;
 (c) residue 5 of domain E5 is V or another hydrophobic residue;
 (d) residue 8 of domain E5 is A or L or another hydrophobic residue; and/or
 (e) residue 11 of domain E5 is W.
- 15 10. The protein of any one of claims 8 or 9, wherein 1, 2, 3, 4, 5, or all 6 of the following are true:
- (a) residue 2 of domain E1 is I or another hydrophobic residue;
 (b) residue 4 of domain H3 is F;
 20 (c) residue 6 of domain E4 is V or another hydrophobic residue;
 (d) residue 8 of domain E4 is L or another hydrophobic residue;
 (e) residue 5 of domain E6 is M or V or another hydrophobic residue; and/or
 (f) residue 7 of domain E6 is V or another hydrophobic residue.
- 25 11. The protein of any one of claims 1-10, comprising an amino acid sequence at least 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence selected from the group consisting of SEQ ID NO:1-181, or SEQ ID NO:1-3.
- 30 12. The protein of claim 11, comprising the amino acid sequence of SEQ ID NO:4.
13. A protein having luciferase activity, comprising an amino acid sequence at least 30%, 35%, 40%, 45%, 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%,

95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of SEQ ID NO:1, wherein:

Residue 14 is Y, D, or E and residue 98 is H or N; and

Residue 18 is D or E and residue 65 is R.

5

14. The protein of claim 13, comprising one or both of A96M and M110V substitutions relative to SEQ ID NO:1.

15. The protein of claim 13, comprising both of A96M and M110V substitutions relative to SEQ ID NO:1.

10

16. The protein of any one of claims 13-15, comprising an R60S substitution relative to SEQ ID NO:1.

17. The protein of claim 16, comprising R60S, A96M, and M110V substitutions relative to SEQ ID NO:1.

15

18. The protein of any one of claims 10-14 and 16, wherein any substitutions relative to SEQ ID NO:1 at residues F12, I35, W38, F49, V81, L83, V94, A97, W100, M110, and/or V112 are conservative amino acid substitutions.

20

19. The protein of any one of claims 13-18, comprising an amino acid sequence at least 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence selected from the group consisting of SEQ ID NO:1-3, or 1-181.

25

20. The protein of any one of claims 11 and 13-19, wherein any substitutions relative to the reference sequence are conservative amino acid substitutions.

30

21. A protein comprising the formula X1-Z1-X2-Z2-X3-Z3-X4-Z4-X5-Z5-X6-Z6-X7-Z7-X8-Z8, wherein:

X1 has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid

sequence of MSEEQIRQFL RRFYEALD (SEQ ID NO: 182), wherein residue 14 is Y, D, or E and residue 18 is D or E;

X2 has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid
5 sequence of ADTAASLF (SEQ ID NO: 183);

X3 has an amino acid sequence at least 50%, 75%, or 100% identical to the amino acid sequence of TIHL (SEQ ID NO: 184);

X4 has an amino acid sequence at least 33%, 66%, or 100% identical to the amino acid sequence of VTF;

10 X5 has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of EEFREWFERLFST (SEQ ID NO: 185);

X6 has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid
15 sequence of QREIKSLEVR (SEQ ID NO: 186), wherein residue 2 is R;

X7 has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of VEVHVQLHATH (SEQ ID NO: 187);

X8 has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%,
20 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of KHTVDATHHWHR (SEQ ID NO: 188), wherein residue 8 is H or N;

X9 has an amino acid sequence at least 50%, 55%, 60%, 65%, 70%, 75%, 80%, 85%, 90%, 91%, 92%, 93%, 94%, 95%, 96%, 97%, 98%, 99%, or 100% identical to the amino acid sequence of VTEMRVHINPTG (SEQ ID NO: 189); and

25 wherein Z1, Z2, Z3, Z4, Z5, Z6, Z7, and Z8 are independently present or absent, and when present may comprise any amino acid sequence.

22. The protein of claim 21, wherein 1, 2, 3, 4, 5, 6, 7, or all 8 of the following are true

Z1 comprises SGD;

30 Z2 comprises HPGV (SEQ ID NO: 190);

Z3 comprises WDG;

Z4 comprises TSR;

Z5 comprises RKDA (SEQ ID NO: 191);

Z6 comprises GDT;

Z7 comprises NGQ; and/or

Z8 comprises GNR; and

wherein 0, 1, 2, 3, 4, 5, 6, 7, or all 8 of Z1, Z2, Z3, Z4, Z5, Z6, Z7, and Z8 further
5 comprise an additional polypeptide domain.

23. A self-complementing multipartite protein having luciferase activity, comprising at least a first polypeptide component and a second polypeptide component, wherein the at least first polypeptide component and the second polypeptide component are not covalently linked,
10 wherein in total the at least first polypeptide component and the second polypeptide component comprise domains X1-Z1-X2-Z2-X3-Z3-X4-Z4-X5-Z5-X6-Z6-X7-Z7-X8-Z8-X9, wherein each domain is as defined in claims 21-22;

wherein (a) each X domain is fully present within one polypeptide component of the at least first polypeptide component and the second polypeptide component, and (b) none of
15 the at least first polypeptide component and the second polypeptide component include each of X1, X2, X3, X4, X5, X6, X7, X8, and X9.

24. The self-complementing multipartite protein of claim 23, wherein the split occurs at
20 Z4, Z5, Z6, or Z7.

25. A self-complementing multipartite protein having luciferase activity, comprising at least a first polypeptide component and a second polypeptide component, wherein the at least first polypeptide component and the second polypeptide component are not covalently linked,
25 wherein in total the at least first polypeptide component and the second polypeptide component comprise the secondary structure arrangement H1-L1-H2-L2-E1-L3-E2-L4-H3-L5-E3-L6-E4-L7-E5-L8-E6, wherein each domain is as defined in any one of claims 1-9;

wherein (a) each H and E domain is fully present within one polypeptide component of the at least first polypeptide component and the second polypeptide component, and (b)
30 none of the at least first polypeptide component and the second polypeptide component include all of the H and E domains.

26. The self-complementing multipartite protein of claim 25, wherein the split occurs at L4, L5, L6, L7, or L8.

27. A fusion protein comprising:
- (a) the protein or polypeptide component of any preceding claims; and
 - (b) one or more additional functional domains.
- 5 28. A nucleic acid encoding the protein, polypeptide component, or fusion protein of any preceding claim.
29. The nucleic acid of claim 28, comprising the nucleotide sequence of any one of SEQ
10 ID NO:200-380.
30. An expression vector comprising the nucleic acid of claim 28 or 29 operatively linked to a suitable control element.
- 15 31. A recombinant host cell comprising the protein, polypeptide component, fusion protein, nucleic acid, and/or expression vector of any preceding claim.
32. A kit comprising:
- (a) the protein, polypeptide component, fusion protein, nucleic acid, expression
20 vector, and/or host cell of any preceding claim; and
 - (b) instructions for their use.
33. The kit of claim 32, further comprising diphenylterazine (DTZ).
- 25 34. A method for using the protein, polypeptide component, fusion protein, nucleic acid, expression vector, host cell, and/or kit of any preceding claim for any suitable purpose, including but not limited to luminescent reporting assays, diagnostic assays, cellular localization of targets of interest, cellular imaging, gene editing, live animal imaging, cancer labeling, CART-cells reporting, secreted assay, gene delivery, and tissue engineering.
- 30 35. A method for making a luciferase, comprising de novo design using the methods of any embodiment disclosed herein, starting with the protein comprising the amino acid sequence of SEQ ID NO:381.