

Supporting Information

Supramolecular Control over Split-Luciferase Complementation

*Ralph P. G. Bosmans, Jeroen M. Briels, Lech-Gustav Milroy, Tom F. A. de Greef,
Maarten Merkx, and Luc Brunsveld**

anie_201602807_sm_miscellaneous_information.pdf

Table of Contents

Supplementary Tables and Figures

Table S1.	Overview of the investigated split-luciferase constructs	2
Figure S1.	Activity assay split-luciferase fragments linked with a flexible linker	2
Figure S2.	SDS PAGE analysis split-luciferase constructs linked with a flexible linker	2
Figure S3.	SDS PAGE analysis individual split-luciferase fragments	3
Figure S4.	Activity screen of FGG-tagged split-luciferase fragments	3
Figure S5.	Enzymatic activity of individual split-luciferase fragments NFluc437 and CFluc398	4
Figure S6.	SDS PAGE and MS analysis NFluc437 and CFluc398 luciferase fragments	4
Figure S7.	SDS PAGE and MS analysis full length luciferase	5
Figure S8.	Emission spectrum full length luciferase and NFluc437 + CFluc398 split-luciferase combination	5
Figure S9.	Activity assay; Q8 titration to NFluc437 + CFluc398 split-luciferase fragments	6
Figure S10.	SDS PAGE and MS analysis of GGG-CFluc398 purification and corresponding activity assay	6

Thermodynamic binding model

	Derivation and Numerical solution	7
Figure S11	Binding equilibria	8
	Data analysis and parameter estimation	10
Figure S12	Plot linear regression of the calculated NFluc437-CFluc398 complex versus the intensity	10
Table S2	Optimized K_d and σ_2 values and calculated ternary equilibrium constants $K_{ter,hom}^d$ and $K_{ter,het}^d$.	11

Experimental methods

	Plasmid construction of linker constructs	12
	Expression and purification of split-luciferase fragments with linker	12
	Purification of GGG-CFluc398	12
	Expression and purification of individual FGG split-luciferase fragments	13
	Expression and purification full length luciferase	13
	QToF-MS analysis	13
	Luminescence activity assay	13
	Emission spectrum	13
	Binding constant NFluc437 and CFluc398	14
	Amino acid sequence protein constructs	14
Supplementary references		15

Table S1. Overview of the split-luciferase constructs investigated in this study. Three different split-luciferase constructs were screened for bacterial expression and cucurbit[8]uril (Q8) mediated complementation. The numbers represent the amino acid numbers of the full length firefly luciferase enzyme.

#	N-terminal fragment	C-terminal fragment
Fluc(1-416)/(398-550) ^[1]	1-416 (<i>NFluc416</i>)	398-550 (<i>CFluc398</i>)
Fluc(1-437)/(438-550) ^[2]	1-437 (<i>NFluc437</i>)	438-550 (<i>CFluc438</i>)
Fluc(1-475)/(265-550) ^[3]	1-475	265-550

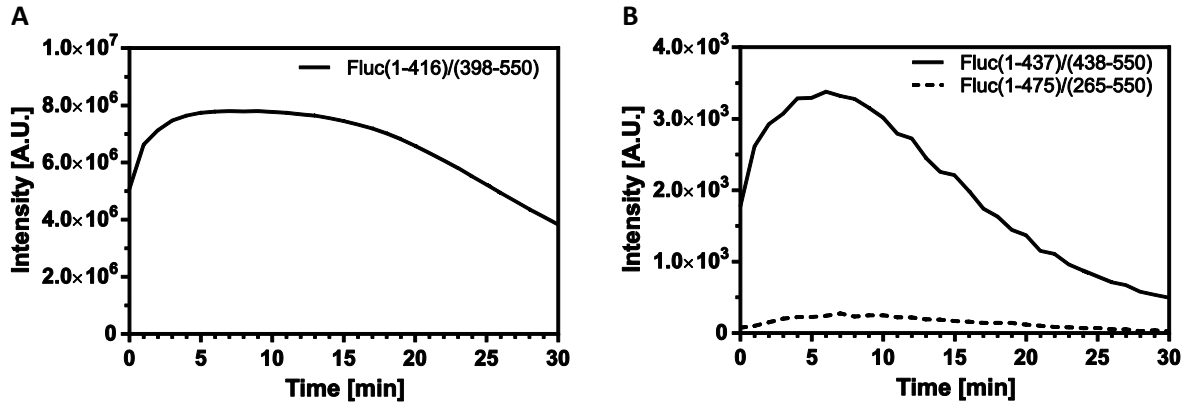


Figure S1. Activity assay of the split-luciferase tethered with a flexible linker. After addition of commercial Luciferase Assay Reagent (LAR) (Promega) to the split constructs in cell lysate, light emission was recorded for 30 minutes. (A) The Fluc(1-416)/(398-550) construct showed enzymatic activity. The observed flat line at $8 \cdot 10^6$ was caused by saturation of the plate reader's sensor. (B) Fluc(1-437)/(438-550) construct remained active after bacterial expression, but the activity was lower than Fluc(1-416)/(398-550). The Fluc(1-475)/(265-550) showed no significant activity.

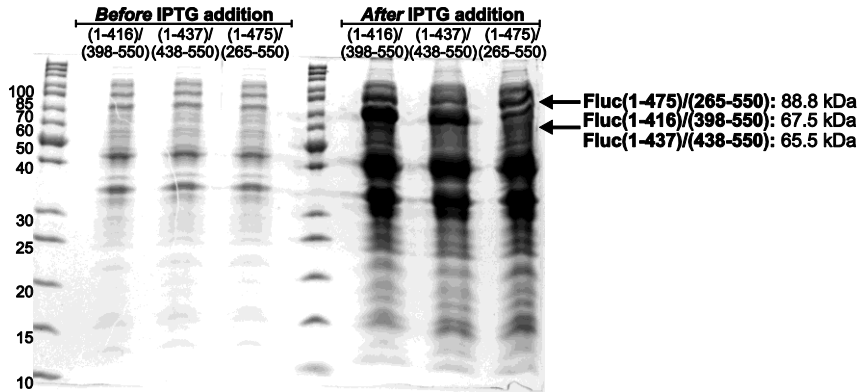


Figure S2. SDS PAGE gel shows the expression yields of the split-luciferase fragments linked with in flexible linker in the cell lysate before and after isopropyl- β -D-thiogalactopyranoside (IPTG) addition. A clear band of Fluc(1-416)/(398-550) and Fluc(1-437)/(438-550) was observed indicating the successful overexpression in *E. coli*. The Fluc(1-475)/(265-550) construct was not expressed by *E. coli*, because no band around 88.8 kDa was observed.

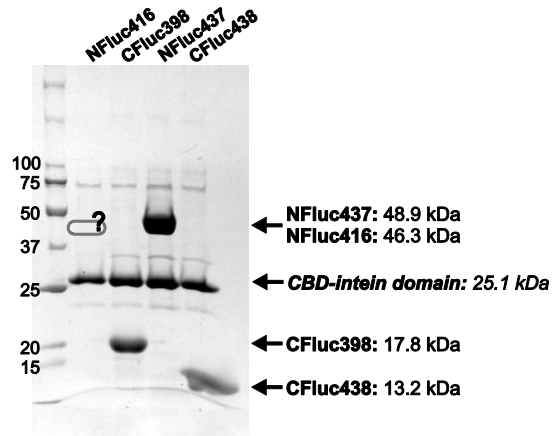


Figure S3. Individual split-luciferase fragments with N-terminal FGG motif after chitin column purification. Each lane represents a different split fragment. The bands at the expected height for the different constructs were observed, except for NFluc416. All the constructs showed also a band around 25 kDa representing the CBD-intein domain that co-elutes from the chitin column. The bands on the gel reveal the successful expression of the fragile split fragments by making use of intein splicing. It is plausible that the CBD-intein domain helps the expression by increasing the solubility of the construct. The absence of the protein band of NFluc416 was unexpected, because a CBD-intein domain band was present, showing that protein expression of this construct in itself had occurred. In order to improve the NFluc416 fragment expression, a stronger purification tag was used. In between the FGG sequence and the NFluc416 sequence a six histidine tag was introduced by site-directed mutagenesis. However, for this construct the desired product could not be observed using SDS PAGE and Q-ToF LC-MS (data not shown). A possible explanation is that the single NFluc416 fragment is not stable without fusion protein and therefore precipitates after intein splicing.

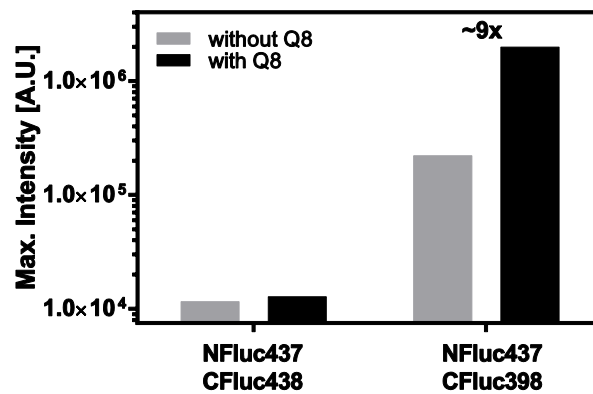


Figure S4. Activity screen of FGG-tagged split-luciferase fragments. Results show the peak luminescence intensity and are the average of two independent measurements. 2 μ M NFluc and 2 μ M CFluc were mixed with 5 μ M Q8. Light emission was recorded immediately after addition of LAR substrate. 20 mM sodium phosphate, 150 sodium chloride, 1 mM EDTA, pH 7, 30 $^{\circ}$ C.

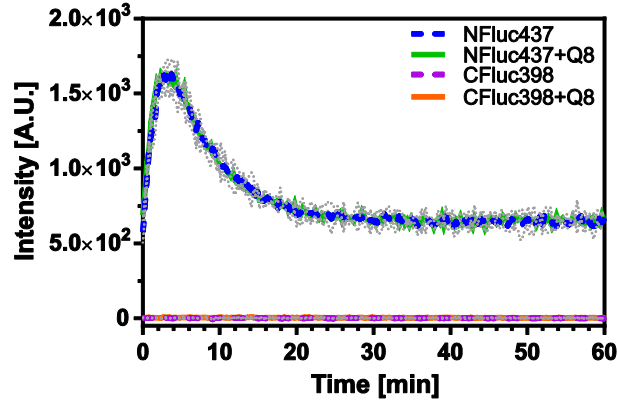


Figure S5. Activity assay of the individual split-luciferase fragments NFluc437 and CFluc398 in the presence and absence of cucurbit[8]uril (Q8). Whereas the individual CFluc398 did not show any activity (purple and orange line), slight activity was observed for the NFluc437 (green and blue line). No effect of Q8 was observed for the activity of the individual split-luciferase fragments. The split fragments and Q8 were present at 0.5 μ M each. The measurement was immediately started after addition of 65 μ L LAR substrate. The average of two measurements was depicted, the dotted gray line represents the standard deviation. 20 mM sodium phosphate, 150 mM sodium chloride, 1 mM EDTA, pH 7, 30 $^{\circ}$ C.

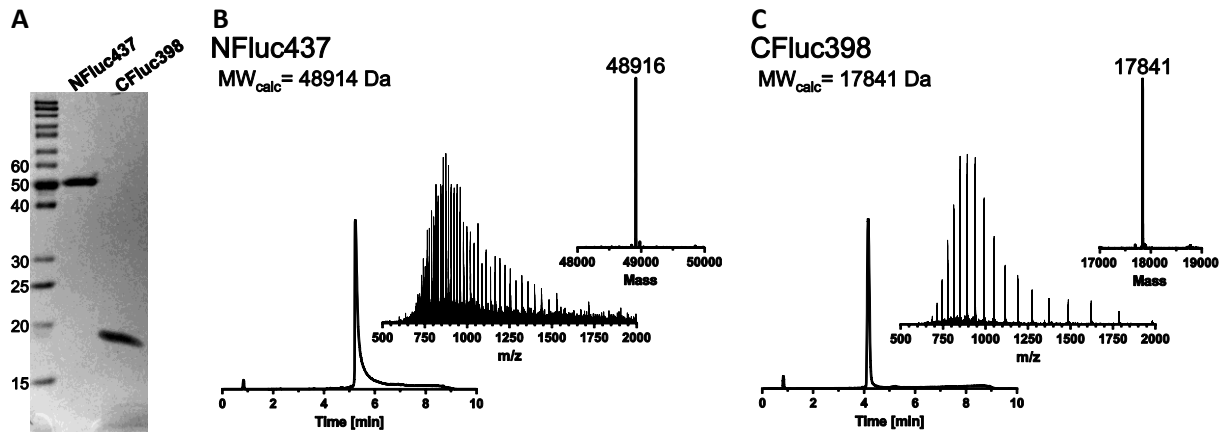


Figure S6. SDS PAGE and MS analysis of the individual NFluc437 and CFluc398 split-luciferase fragments. (A) SDS PAGE gel showing the high purity. Mass spectrometry confirms the high purity and exact mass of (B) NFluc437 and (C) CFluc398. The graphs represent from left to right: total ion count (TIC) chromatogram, m/z spectrum and deconvoluted mass spectrum.

Full-length luciferase

The full-length luciferase featured the exact same sequence as the split variants, but the overlapping region as well as the flexible linker were left out. At the N-terminus a His₆-tag was introduced for purification. The plasmid containing the sequence for the full length luciferase was obtained in a pET15b vector. After bacterial expression, the full length luciferase was purified using nickel affinity chromatography. The purity and correct mass were verified using SDS PAGE and Q-ToF LC-MS. The correct mass without the N-terminal methionine was observed, as well as an additional mass of +179 Da. The presence of an additional mass of +179 Da after recombinant protein expression has been reported before and could be explained by glycosylation of the His-tag.^[4] It was not expected that the glycosylation at the N-terminus of the full length luciferase would interfere with the measurements.

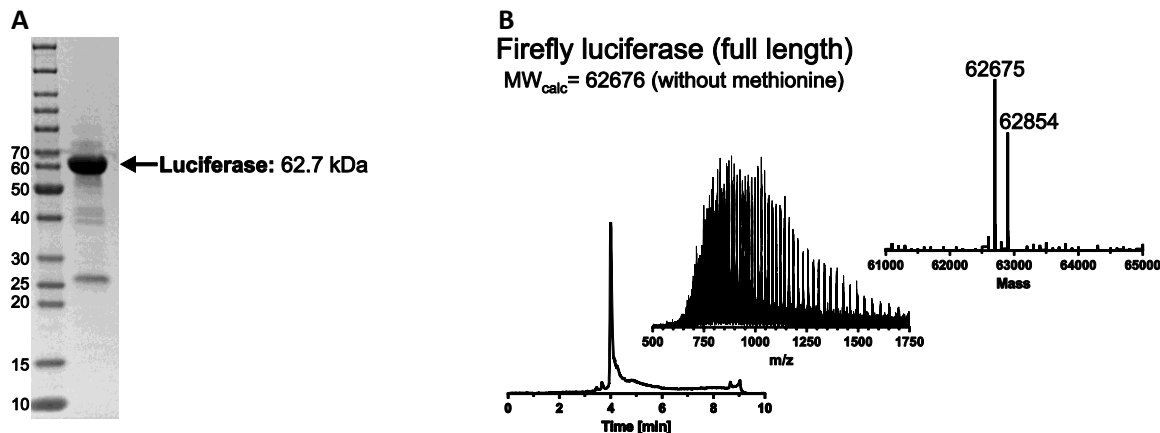


Figure S7. SDS PAGE and MS analysis of the full length luciferase. (A) SDS PAGE showed the purity of the full length firefly luciferase. (B) The observed mass was without the methionine. An additional mass of +179 Da was observed that might be caused by glycosylation of the His-tag.^[4] The graphs represent from left to right: total ion count (TIC) chromatogram, m/z spectrum and deconvoluted mass spectrum.

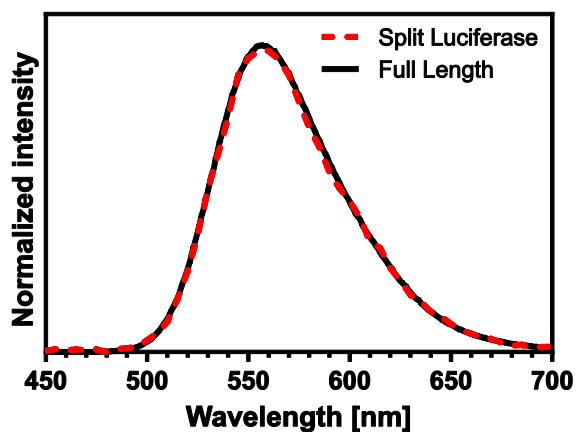


Figure S8. Emission spectrum of the split-luciferase fragments NFluc437 and CFluc398 combination and full length luciferase. The emission spectra of the split-luciferase and full length luciferase were recorded at pH 7. A maximum emission was observed at 555 nm. Data shows the normalized spectra at the maximum intensity at 555 nm. Average of five measurements was depicted. 20 mM sodium phosphate, 150 mM sodium chloride, 1 mM EDTA, pH 7.

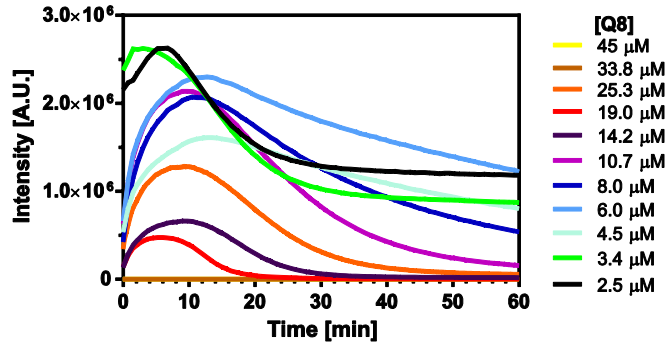


Figure S9. Titration of Q8 to the NFluc437 and CFluc398 split-luciferase combination. Up to 5 equivalents of Q8 compared to the split-luciferase fragments resulted in an gradual increase in enzymatic activity (Figure 2A). However, more than 5 equivalents of Q8 resulted in a loss of enzymatic activity with concomitant changes in the kinetic profile. This loss in activity might be explained by the binding of Q8 to other aromatic amino acids of the split-luciferase fragments and thereby altering the split-luciferase conformation and activity. The split fragments and Q8 were present at 0.5 μM each. The measurement was immediately started after addition of 65 μL LAR substrate. 20 mM sodium phosphate, 150 mM sodium chloride, 1 mM EDTA, pH 7, 30 $^{\circ}\text{C}$.

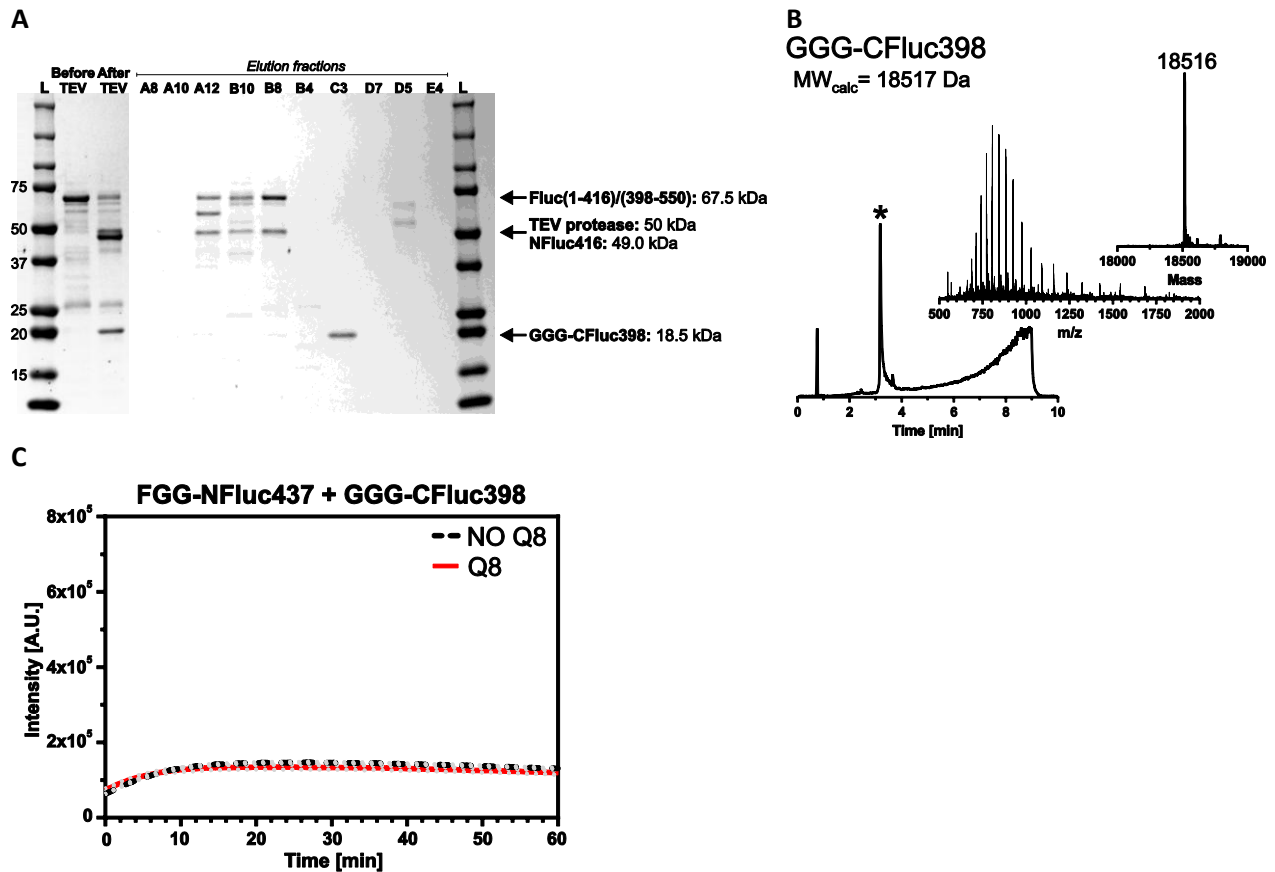


Figure S10. SDS PAGE of GGG-CFluc398 purification and activity assay (A) SDS PAGE showed the purification steps of the GGG-CFluc398 construct. TEV cleavage results in the cleavage of the linker construct Fluc(1-416)/(398-550) yielding GGG-CFluc398. The cleavage mixture was purified using Size Exclusion Chromatography. Elution fractions were analyzed on SDS PAGE. The GGG-CFluc398 was present in fraction C3. (B) Mass spectrometry confirms the high purity and exact mass of GGG-CFluc398. The graphs represent from left to right: total ion count (TIC) chromatogram, m/z spectrum and deconvoluted mass spectrum. The impurity eluting around 1 minute is an impurity originated from the buffer. (C) Activity assay of FGG-NFluc437 and GGG-CFluc398 in the presence and absence of cucurbit[8]Juri (Q8). Addition of Q8 did not result in an increase in enzymatic activity, showing that an N-terminal phenylalanine is required for ternary complex formation. The average of three measurements was depicted; the dotted grey line represents the standard deviation. The split fragments and Q8 were present at 0.5 μM each. The measurement was immediately started after addition of 65 μL LAR substrate. 20 mM sodium phosphate, 150 mM sodium chloride, 1 mM EDTA, pH 7, 30 $^{\circ}\text{C}$.

Thermodynamic model describing the binding of NFLuc and CFluc fragments to Q8 in the presence of FGG competitor

Derivation and Numerical Solution

Consider a system consisting of two protein fragments (*NFluc* and *CFluc*) each functionalized at the C or N-terminus with an FGG sequence, and a cucurbit[8]uril (Q8) receptor. The Q8 receptor contains two binding sites for FGG motifs. In addition, free FGG peptides are present that can compete with the proteins fragments for binding to Q8. When equilibrium is attained, the system consists of various ternary complexes containing identical and non-identical protein fragments, Q8 complexes containing either an *NFluc* or *CFluc* fragment and various complexes containing FGG inhibitor.

Q_{tot}	= initial concentration of Q8 in mol L ⁻¹
C_{tot}	= initial concentration of <i>CFluc</i> fragment in mol L ⁻¹
N_{tot}	= initial concentration of <i>NFluc</i> fragment in mol L ⁻¹
I_{tot}	= initial concentration of FGG inhibitor in mol L ⁻¹
K^d	= reference dissociation equilibrium constant for binding of a single FGG fragment to Q8 in mol
σ_1	= dimensionless cooperativity parameter for binding of two identical fragments to Q8 (i.e. binding of two <i>NFluc</i> or <i>CFluc</i> fragments to the same Q8 or binding of two FGG inhibitors to Q8).
σ_2	= dimensionless cooperativity parameter for binding of two non-identical fragments to Q8 (i.e. binding of one <i>NFluc</i> and one <i>CFluc</i> fragment to the same Q8).
$K_{ter_hom}^d$	= ternary dissociation constant (in mol ² L ⁻²) for binding of two identical fragments to Q8 (i.e. binding of two <i>NFluc</i> or <i>CFluc</i> fragments to the same Q8 or binding of two FGG inhibitors to Q8).
$K_{ter_het}^d$	= ternary dissociation constant (in mol ² L ⁻²) for binding of two non-identical fragments to Q8 (i.e. binding of one <i>NFluc</i> and one <i>CFluc</i> fragment to the same Q8).
[Q]	= free concentration of Q8 in mol L ⁻¹
[C]	= free concentration of <i>CFluc</i> fragment in mol L ⁻¹
[N]	= free concentration of <i>NFluc</i> fragment in mol L ⁻¹
[I]	= free concentration of FGG inhibitor in mol L ⁻¹
[QN]	= concentration of Q8· <i>NFluc</i> species in mol L ⁻¹ (see Figure S11)
[QC]	= free concentration of Q8· <i>CFluc</i> fragment in mol L ⁻¹ (see Figure S11)
[QI]	= free concentration of Q8· <i>Inh</i> species in mol L ⁻¹ (see Figure S11)
[QNN]	= concentration of Q8· <i>NFluc</i> · <i>NFluc</i> species in mol L ⁻¹ (see Figure S11)
[QCC]	= free concentration of Q8· <i>CFluc</i> · <i>CFluc</i> fragment in mol L ⁻¹ (see Figure S11)
[QII]	= free concentration of Q8· <i>Inh</i> · <i>Inh</i> species in mol L ⁻¹ (see Figure S11)
[QCN]	= free concentration of Q8· <i>CFluc</i> · <i>NFluc</i> species in mol L ⁻¹ (see Figure S11)
[QCI]	= free concentration of Q8· <i>CFluc</i> · <i>Inh</i> species in mol L ⁻¹ (see Figure S11)
[QNI]	= free concentration of Q8· <i>NFluc</i> · <i>Inh</i> species in mol L ⁻¹ (see Figure S11)

When the equilibrium is attained, the initial Q8 concentration, Q_{tot} , is partitioned into the various complexes as depicted in Figure S11:

$$Q_{tot} = [Q] + [QN] + [QC] + [QI] + [QNN] + [QCC] + [QCN] + [QCI] + [QNI] \quad \text{Eq. 1}$$

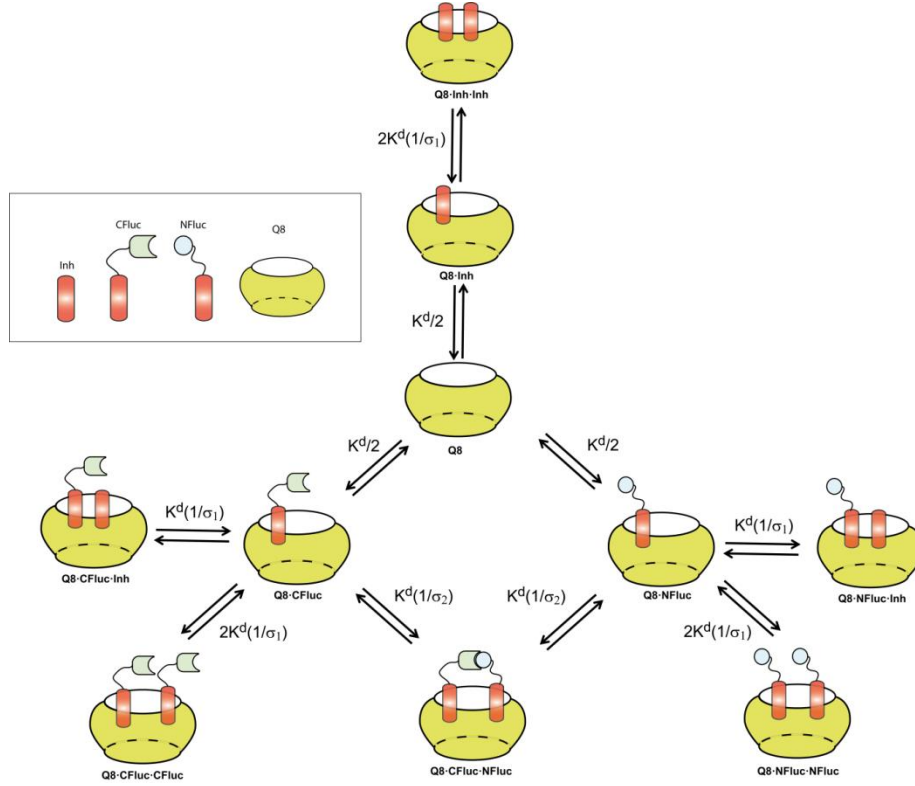


Figure S11. Considered binding equilibria of FG functionalized CFluc and NFluc fragments in the presence of monovalent FG inhibitor.

Next, we write down the mass-balances of the total concentration of *CFluc* and *NFluc* fragments partitioned between their respective free forms and the various complexes:

$$C_{tot} = [C] + [QC] + 2[QCC] + [QCN] \quad \text{Eq. 2}$$

$$N_{tot} = [N] + [QN] + 2[QNN] + [QCN] \quad \text{Eq. 3}$$

Finally, the mass-balance of the inhibitor that relates the total concentration of inhibitor to the free concentration of inhibitor and Q8-inhibitor complexes

$$I_{tot} = [I] + [QI] + [QCI] + [QNI] + 2[QII] \quad \text{Eq. 4}$$

Now, we derive expressions relating the concentrations of *C*, *N*, *Q*, *QC*, *QCC* and *QCN* to the equilibrium constants K^d and the cooperativity parameters σ_1 and σ_2 by taking into account the appropriate statistical factors:

$$\frac{1}{2}K^d = \frac{[Q][C]}{[QC]} \quad \text{Eq. 5}$$

$$\frac{2K^d}{\sigma_1} = \frac{[C][QC]}{[QCC]} \quad \text{Eq. 6}$$

$$\frac{K^d}{\sigma_2} = \frac{[N][QC]}{[QCN]} \quad \text{Eq. 7}$$

Next, the concentration of QC, QCC and QCN complexes are expressed as a function of the monovalent dissociation constant K^d , the dimensionless cooperativity parameters, the free concentrations of Q8 (Q) and the C and N-terminal Fluc fragments (C and N).

$$[QC] = \frac{2[Q][C]}{K^d} \quad \text{Eq. 8}$$

$$[QCC] = \frac{[Q][C]^2\sigma_1}{(K^d)^2} \quad \text{Eq. 9}$$

$$[QCN] = \frac{2[Q][C][N]\sigma_2}{(K^d)^2} \quad \text{Eq. 10}$$

In a similar way, we derive expressions relating the concentrations of C, N, Q, QN, QNN to the equilibrium constants K^d and the cooperativity parameters σ_1 and σ_2 by taking into account the appropriate statistical factors:

$$\frac{1}{2}K^d = \frac{[Q][C]}{[QN]} \quad \text{Eq. 11}$$

$$\frac{2K^d}{\sigma_1} = \frac{[N][QN]}{[QNN]} \quad \text{Eq. 12}$$

$$\frac{K^d}{\sigma_2} = \frac{[C][QN]}{[QCN]} \quad \text{Eq. 13}$$

Again, we express the concentrations of QN, QNN complexes as a function of the dissociation constant K^d , the dimensionless cooperativity parameters and the free concentrations of Q8 (Q) and the C and N-terminal Fluc fragments (C and N).

$$[QN] = \frac{2[Q][N]}{K^d} \quad \text{Eq. 14}$$

$$[QNN] = \frac{[Q][N]^2\sigma_1}{(K^d)^2} \quad \text{Eq. 15}$$

The ternary equilibrium constants $K^d_{ter_hom}$ and $K^d_{ter_het}$ are defined as:

$$K^d_{ter_hom} = \frac{K^d}{2} \frac{2K^d}{\sigma_1} = \frac{(K^d)^2}{\sigma_1} \quad \text{Eq. 16}$$

$$K^d_{ter_het} = \frac{K^d}{2} \frac{K^d}{\sigma_2} = \frac{(K^d)^2}{2\sigma_2} \quad \text{Eq. 17}$$

Finally, we derive expressions relating the concentrations of Q, Inh, QI, QII, QC, QN, QCI and QNI to the equilibrium constants K^d and the cooperativity parameter σ_1 by taking into account the appropriate statistical factors:

$$[QI] = \frac{2[Q][Inh]}{K^d} \quad \text{Eq. 18}$$

$$[QII] = \frac{[Q][Inh]^2\sigma_1}{(K^d)^2} \quad \text{Eq. 19}$$

$$[QCI] = \frac{2[Q][C][Inh]\sigma_1}{(K^d)^2} \quad \text{Eq. 20}$$

$$[QNI] = \frac{2[Q][N][Inh]\sigma_1}{(K^d)^2} \quad \text{Eq. 21}$$

Next, we substitute eq. 8-10, eq. 14-15, and eq. 18-21 in the mass-balance equations (eq. 1-4) to arrive at the following expressions:

$$Q_{tot} = [Q] + 2\frac{[Q][N]}{K^d} + 2\frac{[Q][C]}{K^d} + \frac{[Q][N]^2\sigma_1}{(K^d)^2} + \frac{[Q][C]^2\sigma_1}{(K^d)^2} + \frac{[Q][C][N]\sigma_2}{(K^d)^2} + 2\frac{[Q][Inh]}{K^d} + \frac{[Q][Inh]^2\sigma_1}{(K^d)^2} + \frac{2[Q][C][Inh]\sigma_1}{(K^d)^2} + \frac{2[Q][N][Inh]\sigma_1}{(K^d)^2} \quad \text{Eq. 22}$$

$$C_{tot} = [C] + 2\frac{[Q][C]}{K^d} + 2\frac{[Q][C]^2\sigma_1}{(K^d)^2} + \frac{[Q][C][N]\sigma_2}{(K^d)^2} + \frac{2[Q][C][Inh]\sigma_1}{(K^d)^2} \quad \text{Eq. 23}$$

$$N_{tot} = [N] + 2\frac{[Q][N]}{K^d} + 2\frac{[Q][N]^2\sigma_1}{(K^d)^2} + \frac{[Q][C][N]\sigma_2}{(K^d)^2} + \frac{2[Q][N][Inh]\sigma_1}{(K^d)^2} \quad \text{Eq. 24}$$

$$Inh_{tot} = [Inh] + 2\frac{[Q][Inh]}{K^d} + 2\frac{[Q][C][Inh]\sigma_1}{(K^d)^2} + \frac{2[Q][N][Inh]\sigma_1}{(K^d)^2} + \frac{2[Q][Inh]^2\sigma_1}{(K^d)^2} \quad \text{Eq. 25}$$

Using custom-written scripts in Matlab, the coupled non-linear equations 22-25 were solved for the free concentrations Q , C , N and Inh by employing an iterative root-finding procedure based on successive substitution. The script uses the known initial concentrations of the NFluc and CFluc fragments (C_{tot} and N_{tot}), the known initial concentration of Q8 (Q_{tot}), the known total concentration of FGG competitor (Inh_{tot}), the dissociation constant of monovalent FGG binding (K^d) and the cooperativity parameters (σ_1 and σ_2) as an input.

Data analysis and parameter estimation

We performed non-linear least square analysis on the FGG titration data as reported in Figure 3C by comparing the fluorescence data to the computed fluorescence. The computed fluorescence was calculated by multiplying the concentration of ternary complex QCN in the model with a conversion factor, i.e. $F_{cal} = F^*[QCN]$. The conversion factor F was determined to be $1.325e+13$ A.U. using the following procedure; Based on the determined binding constant ($K_d^{NC} = 112 \pm 12 \mu\text{M}$) between the NFluc437 and CFluc398 in the absence of Q8 (Figure 1), a new graph is constructed where the calculated concentration of dimeric NFluc437-CFluc398 complex is plotted against the luminescence intensity. The intensity values were corrected by subtracting the background activity of NFluc437. These data points were fitted using linear regression (Figure S12) resulting in a conversion factor of $1.325 \cdot 10^{13}$ i.e. the light intensity in A.U. produced by 1 M of NFluc437-CFluc398 complex. The determination of the conversion factor rests on the assumption that Q8 itself does not contribute to the enzymatic complemented split-luciferase reaction.

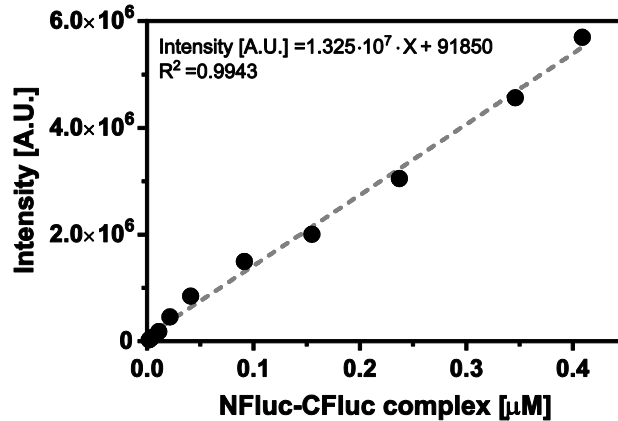


Figure S12. Linear regression of the calculated NFluc437-CFluc398 complex versus the intensity. The dots represent the data points and the gray line represents the linear fit. In the equation, X is the concentration of NFluc437-CFluc398 complex in μM .

During initial non-linear least square analysis of the data, we found that the parameter σ_1 describing the cooperativity in the binding of two identical FGG fragments to Q8 had negligible effects on the shape of the titration curve. As a result, we decided to fix this parameter to a value of 1, reflecting the fact that binding of the second *identical* FGG peptide or fragment is non-cooperative. Because our model does not take into account background fluorescence, we subtracted the residual fluorescence at high FGG concentrations from the raw fluorescence data. Non-linear least square minimization of the data

was performed with the function *lsqnonlin*, a subspace trust region method based on the interior-reflective Newton method. In order to prevent entrapment in a local minimum, 20 different starting values of K_d and σ_2 were defined, and the best fit (defined as the fit with the lowest squared 2-norm of the residuals) is taken as the final solution for the optimized values. The different initial parameter sets are defined using a *latin hypercube sampling* method (Matlab function *lhsdesign*). The standard deviation of the parameters were calculated using the Fischer information matrix. The optimized values of K_d and σ_2 and their standard deviations are reported in Table S2. Using these values and equation 16 and 17 we calculated the ternary equilibrium constants $K_{ter_hom}^d$ and $K_{ter_het}^d$.

Table S2. Optimized values and their standard deviation

K_d	$7.4 \pm 0.6 \times 10^{-5} \text{ M}$
σ_2	301 ± 48
$K_{ter_hom}^d$	$5.5 \times 10^{-9} \text{ M}^2$
$K_{ter_het}^d$	$9.1 \times 10^{-12} \text{ M}^2$

Experimental

Plasmid construction

Based on the amino acid sequence of the firefly luciferase plasmid (pGL4.10-luc2) of Promega, DNA constructs were obtained in a pUC57 vector (Genscript). Codon usage was optimized for *E. coli* using GeneDesign 2.0 software. Treatment of this construct with the restriction enzymes *NcoI* and *XhoI* (New England Biolabs) yielded insert fragments that were cloned in a pET28a vector that was treated with the same restriction enzymes. Prior to ligation, the cleaved fragments were purified on agarose gel. Ligation was performed using T4 DNA ligase (TaKaRa Mighty Mix, TaKaRa Bio Inc, Japan), by incubating different vector:insert ratios for 2 hours at 16 °C. Successful cloning was confirmed by DNA sequencing (Baseclear, Leiden, the Netherlands).

The individual split-luciferase fragments NFluc416, CFluc398, NFluc437 and CFluc438 were cloned in a pTWIN-1 plasmid (New England Biolabs). Cloning was performed by using the In-Fusion® HD EcoDry™ Cloning Kit (Clontech). The pET28a vectors containing the fragments with linker served as a template. The NFluc437 was amplified by PCR using the forward primer 5'-atcattgtacacaacttttggtggcgaggacgcgaaaaatatcaagaaag-3' and reverse primer 5'-gccggatccccttccttagtcgacccctatcgacaatgaagaaa-3'. The NFluc416 was amplified by PCR using the forward primer 5'-atcattgtacacaacttttggtggcgaggacgcgaaaaatatcaagaaag-3' and reverse primer 5'-gccggatccccttccttagtcgacaccatctttgtctatcaac-3'. The CFluc398 and CFluc438 were amplified by PCR using the forward primer 5'-atcattgtacacaactttgggggtgggagcggcggttcagaat-3' and reverse primer 5'-gccggatccccttccttactcgagcagcgcaatcttcccgcct-3'. The forward and reverse primers contained the nucleotide sequence which translated for the FGG motif as well as the complement sequence of the pTWIN-1 vector. The pTWIN-1 vector was linearized by treating with the *SapI* and *PstI* restriction enzymes (New England Biolabs). The PCR product was ligated in the linearized vector using the In-Fusion enzyme. The successful cloning was verified by sequencing (BaseClear, Leiden, the Netherlands).

Based on the amino acid sequence of the firefly luciferase of Promega (pGL4.10-luc2), DNA constructs translating for the full length luciferase were designed and cloned in a pET-15b vector using the *NdeI* and *BamHI* restriction sites (GenScript). The same DNA sequence was used as for the split fragments NFluc437 and CFluc398.

Expression and purification of split-luciferase fragments with linker

The split-firefly luciferase constructs connected via a flexible GGS₁₂ linker were expressed in *E. coli* BL21(DE3) competent cells (Novagen). The bacteria were cultured in 500 mL LB medium containing 30 µg mL⁻¹ kanamycin at 37 °C and 250 rpm until an OD₆₀₀ of ~0.7. Subsequently, protein expression was induced by adding isopropyl-β-D-thiogalactopyranoside (IPTG) to a final concentration of 0.2 mM. Cells were incubated overnight at 15 °C and 250 rpm before being harvested by centrifugation (8000 g for 10 minutes). Bacterial cells were lysed by resuspending the pellet in Bugbuster Protein Extraction Reagent supplemented with benzonase (Novagen) and incubated for 30 minutes at room temperature. The insoluble fraction was removed by centrifugation (40.000 g for 30 minutes). The split-luciferase fragments with flexible linker were purified by gravity flow on a nickel affinity column (His-bind resin, Merck Millipore) according to the manufacturer's protocol.

Cleavage and purification of the GGG-CFluc398 fragment

To obtain GGG-CFluc398, the linker construct Fluc(1-416)/(398-550) was cleaved using TEV protease. This results in cleavage of the flexible linker yielding the CFluc398 fragment with a glycine at the N-terminus. The nickel affinity purified linker constructs were incubated overnight in the presence of 40 U TEV (proTEV plus, Promega). The overnight cleavage product was purified using size exclusion chromatography on a superdex 75 column (GE Healthcare) connected to an AKTA FPLC. The column was pre-equilibrated with buffer A (20 mM sodium phosphate, 150 mM sodium chloride, 1 mM ethylenediaminetetraacetic acid (EDTA), pH 7). Flow rate was 1 ml min⁻¹. The eluted protein fractions were analyzed for protein content on SDS-PAGE. Fractions containing the GGG-CFluc398 fragments were pooled and concentrated using a centrifugal filter device (EMD Millipore, MWCO 3000 Da). Concentration was determined using the Nanodrop ND-1000 spectrophotometer and the theoretical extinction coefficient at 280 nm of 18450 M⁻¹ cm⁻¹. Molecular weight of GGG-CFluc398 is 18517 Da.

Expression and purification of FGG split-luciferase fragments

The pTWIN-1 plasmids containing the individual split-luciferase fragments were transformed in *E. coli* BL21(DE3) host strain (Novagen). The bacteria were cultured in 2 L LB medium containing 100 $\mu\text{g mL}^{-1}$ ampicillin at 37 °C and 180 rpm until an OD₆₀₀ of ~0.7. Subsequently, protein expression was induced by adding isopropyl- β -D-thiogalactopyranoside (IPTG) to a final concentration of 0.5 mM. Cells were incubated for 7 hours at 25 °C and 180 rpm before being harvested by centrifugation (8000 g for 10 minutes). Bacterial cells were lysed by resuspending the pellet in Bugbuster Protein Extraction Reagent supplemented with benzonase (Novagen) and incubated for 1 hour at room temperature. The insoluble fraction was removed by centrifugation (40.000 g for 30 minutes). The soluble fraction was purified by applying it to a column filled with chitin beads (New England Biolabs) by gravity flow. The column was washed with 10 column volumes buffer A (20 mM sodium phosphate, 150 mM sodium chloride, 1 mM ethylenediaminetetraacetic acid (EDTA), pH 7). The on-column intein splicing was incubated overnight at room temperature by closing the column. The cleaved proteins were eluted from the column by washing with the same buffer.

The successful split-luciferase pairs NFluc437 and CFluc398 were further purified using an additional anion exchange chromatography step. Prior to loading the protein on the columns (Strong Ion Exchange Spin Columns (Pierce Thermo Scientific)), the buffer was changed to buffer B (20 mM sodium phosphate, 50 mM sodium chloride, pH 8) by filtration (Millipore centrifugal filter device MWCO 10 kDa). Columns were washed with increasing amounts of sodium chloride in buffer B. Both split-luciferase fragments eluted at 145 mM sodium chloride. The purity and correct mass were confirmed by SDS-PAGE and Q-ToF LC-MS. Yields of purified split-luciferase fragments were low: ~0,5 mg per liter culture medium for NFluc437 and ~1 mg per liter culture medium for CFluc398. Concentrations were determined using the Nanodrop ND-1000 spectrophotometer and theoretical extinction coefficients at 280 nm of 35090 $\text{M}^{-1} \text{cm}^{-1}$ for NFluc437, 18450 $\text{M}^{-1} \text{cm}^{-1}$ for CFluc398. Proteins were stored at -80 °C.

Expression and purification of full-length luciferase fragments

The full-length firefly luciferase was expressed in *E. coli*, using the same protocol as has been described for the split-luciferase fragments. The full length luciferase was purified by gravity flow using a NiNTA column (Clontech). After loading the soluble fraction, the column was washed with 5 column volumes buffer B (50 mM sodium phosphate, 300 mM sodium chloride, 2 mM 2-mercaptoethanol, 25 mM imidazole, pH 7.5). Thereafter the column was washed with 5 column volumes buffer B supplemented with 0.1% Triton X100. In order to remove the Triton X100, an additional wash step of 5 column volumes buffer B was performed. The protein was eluted using buffer B with 250 mM imidazole. The buffer was changed to buffer A (20 mM sodium phosphate, 150 mM sodium chloride, 1 mM ethylenediaminetetraacetic acid (EDTA), pH7) using a centrifugal filter device (Millipore, MWCO. 50.000). The purity and correct mass were confirmed by SDS-PAGE and Q-ToF LC-MS. Final yield of full-length luciferase was 10 mg per liter culture. Concentrations were determined using the Nanodrop ND-1000 spectrophotometer using a theoretical extinction coefficient ϵ_{280} of 39310 $\text{M}^{-1} \text{cm}^{-1}$.

Q-ToF LC-MS analysis

Purity and exact mass of the proteins were determined using a High Resolution LC-MS system consisting of a Waters ACQUITY UPLC I-Class system coupled to a Xevo G2 Quadrupole Time of Flight (Q-ToF). The system was comprised of a Binary Solvent Manager and a Sample Manager with Fixed-Loop (SM-FL). Proteins were separated (0.3 mL min^{-1}) by the column (Polaris C18A reverse phase column 2.0 x 100 mm, Agilent) using a 15% to 75% acetonitrile gradient in water supplemented with 0.1% v/v formic acid before analysis in positive mode in the mass spectrometer. Deconvolution of the m/z spectra was done using the MaxENT1 algorithm in the Masslynx v4.1 (SCN862) software.

Luminescence activity assay

All the activity assays were performed on a Tecan Safire II plate reader at 30 °C in a white 96-wells plate (LUMITRAC™ 200, half area, GreinerBio). The split-luciferase fragments were pre-incubated at 0.5 μM each in the presence or absence of cucurbit[8]uril (Q8) (Sigma) in buffer A (20 mM sodium phosphate, 150 mM sodium chloride, 1 mM EDTA, pH 7). After addition of 65 or 80 μL Luciferase Assay Reagent (LAR) (Promega), the measurement was immediately started (1000 ms integration time). Final volume was 100 μL . Competitors were added after 65 minutes of incubation when the light emission was stable. The FGG peptide was ordered from Caslo ApS, Denmark and the memantine was obtained from Acros Organics. The exact concentrations of Q8 were determined by titration as described by Kaifer et. al.^[5]

Emission spectrum

Emission spectra were recorded in a Varian Cary Eclipse spectrometer. The concentration of the split-luciferase was 1 μM each complemented with 8 μM Q8. Full-length luciferase was measured at concentration of 1 nM. Emission spectra were recorded in buffer A (20 mM sodium phosphate, 150 mM NaCl, 1 mM EDTA, pH 7.0). The reaction was initiated with 65 μL LAR in a total volume of 125 μL .

Binding constant NFluc437 and CFluc398

Addition of increasing concentrations of CFluc398 to NFluc437 in the absence of Q8 resulted in an increase in activity (Figure 1). Light intensity at $t = 50$ minutes was plotted against the CFluc398 concentration. The data was corrected for background activity of NFluc437 by subtracting the activity of NFluc437 at $t = 50$ minutes from the other data points. Data was fitted using a one-site specific binding model (Figure 1). The non-linear least-square analysis resulted in a dissociation constant $K_d^{NC} = 112 \pm 12 \mu\text{M}$. The model is based on the assumption that the luminescence signal is linearly correlated with the amount of active complexes and the CFluc398 itself did not show any activity.

The data was fitted using a one-site specific binding model using Graphpad Prism v.6.01. The model was described by:

$$\text{Intensity} = B_{\max} \cdot \frac{x}{K_d^{NC} + x}$$

With B_{\max} the maximum specific binding in the same units as the intensity [A.U], x the CFluc398 concentration [M] and K_d^{NC} the dissociation constant [M].

Amino acid sequence of the used constructs

Fluc(1-416)/(398-550)

MW = 67457 Da

The flexible linker is underlined and the TEV cleavage site is depicted in bold

MEDAKNIKKG PAPPFYPLEDG TAGEQLHKAM KRYALVPGTI AFTDAHIEVD ITYAEYFEMS VRLAEAMKRY GLNTNHRIVV
CSENSLQFFM PVLGALFIGV AVAPANDIYN ERELLNSMGI SQPTVVFVSK KGLQKILNVQ KKLPIIQKII IMDSKTDYQG
FQSMYTFVTS HLPFGFNEYD FVPESFDRDK TIALIMNSSG STGLPKGVAL PHRTACVRF S HARDPIFGNQ IIPDTAILS
VPFHGFGMF TTLGYLICGF RVVLMYRFEE ELFLRSLQDY KIQSALLVPT LFSFFAKSTL IDKYDLSNLH EIASGGAPLS
KEVGEAVAKR FHLPGIRQGY GLTETTSAIL ITPEGDDKPG AVGVVPPFE AKVVDLDTGK TLGVNQRGEL CVRGP MIMSG
YVNNPEATNA LIDKDGVDGG SGGSGGSGGS GGSGGSGGS GGSGGSGGS **ENLYFQGGGSG** SEFM SGYVNN PEATNALIDK
DGWLHSGDIA YWDEDEHFFI VDRLKS LIKY KGYQVAPAE ESILLQHPNI FDAGVAGLPD DDAGELPAAV VVLEHGKTMT
EKEIVDYVAS QVTTAKKLRG GVVVFVDEVPK GLTGKLDARK IREILIKAKK GKIAVLEHH HHHH

Fluc(1-437)/(438-550)

MW = 65466 Da

The flexible linker is underlined and the TEV cleavage site is depicted in bold

MEDAKNIKKG PAPPFYPLEDG TAGEQLHKAM KRYALVPGTI AFTDAHIEVD ITYAEYFEMS VRLAEAMKRY GLNTNHRIVV
CSENSLQFFM PVLGALFIGV AVAPANDIYN ERELLNSMGI SQPTVVFVSK KGLQKILNVQ KKLPIIQKII IMDSKTDYQG
FQSMYTFVTS HLPFGFNEYD FVPESFDRDK TIALIMNSSG STGLPKGVAL PHRTACVRF S HARDPIFGNQ IIPDTAILS
VPFHGFGMF TTLGYLICGF RVVLMYRFEE ELFLRSLQDY KIQSALLVPT LFSFFAKSTL IDKYDLSNLH EIASGGAPLS
KEVGEAVAKR FHLPGIRQGY GLTETTSAIL ITPEGDDKPG AVGVVPPFE AKVVDLDTGK TLGVNQRGEL CVRGP MIMSG
YVNNPEATNA LIDKDGWLHS GDIAYWDEDE HFFIVDRVDG GGSGGSGGS GGSGGSGGS GGSGGSGGS **ENLYFQGGGSG**
GSEFLKSLIK YKGYQVAPAE LESILLQHPN IFDAGVAGLP DDDAGELPAA VVLEHGKTM TEKEIVDYVA SQVTTAKKLR
GGVVVFVDEVP KGLTGKLDAR KIREILIKAK KGGKIAVLEH HHHHH

Fluc(1-475)/(265-550)

MW = 88755 Da

The flexible linker is underlined and the TEV cleavage site is depicted in bold

MEDAKNIKKG PAPPFYPLEDG TAGEQLHKAM KRYALVPGTI AFTDAHIEVD ITYAEYFEMS VRLAEAMKRY GLNTNHRIVV
CSENSLQFFM PVLGALFIGV AVAPANDIYN ERELLNSMGI SQPTVVFVSK KGLQKILNVQ KKLPIIQKII IMDSKTDYQG
FQSMYTFVTS HLPFGFNEYD FVPESFDRDK TIALIMNSSG STGLPKGVAL PHRTACVRF S HARDPIFGNQ IIPDTAILS
VPFHGFGMF TTLGYLICGF RVVLMYRFEE ELFLRSLQDY KIQSALLVPT LFSFFAKSTL IDKYDLSNLH EIASGGAPLS
KEVGEAVAKR FHLPGIRQGY GLTETTSAIL ITPEGDDKPG AVGVVPPFE AKVVDLDTGK TLGVNQRGEL CVRGP MIMSG
YVNNPEATNA LIDKDGWLHS GDIAYWDEDE HFFIVDRLKS LIKYKGYQVA PAELESILLQ HPNIFDAGVA GLPDDVDGGS
GGSGGSGGS GGSGGSGGS GGSGGSGGS **ENLYFQGGGSG** EFM YRFEEEL FLRSLQDYKI QSALLVPTLF SFFAKSTLID
KYDLSNLHEI ASGGAPLSKE VGEAVAKRFH LPGIRQGYGL TETTSAILIT PEGDDKPGAV GKVVPPFEAK VVDLDTGKTL
GVNQRGELCV RGP MIMSGYV NNPEATNALI DKDGWLHSGD IAYWDEDEHF FIVDRLKS LI KYKGYQVAPA ELESILLQHP
NIFDAGVAGL PDDAGELPA AVVVLEHGKT MTEKEIVDYV ASQVTTAKKL RGGVVVFVDEV PKGLTGK LDA RKIREILIKA
KGGKIAVLE HHHHHH

NFluc416

MW = 46281 Da

FGGEDAKNIK KGPAPFYPLE DGTAGEQLHK AMKRYALVPG TIAFTDAHIE VDITYAEYFE MSVRLAEAMK RYGLNTNHRI
VVCSENSLQF FMPVLGALFI GVAVAPANDI YNERELLSNM GISQPTVVV FV SKKGLQKILN VQKKLP I I QK I I I M D S K T D Y
QGFQSMYTFV TSHLPPGFNE YDFVPESFDR DKTIALIMNS SGSTGLPKGV ALPHRTACVR FSHARDPIFG NQIIPDTAIL
SVVFFHHGFG MFTTLGYLIC GFRVVMYRF EEELFLRSLQ DYKIQSALLV PTLFSFFAKS TLIDKYDLSN LHEIASGGAP
LSKEVGEAVA KRFHLPGIRQ GYGLTETTSA ILITPEGDDK PGAVGKVVVF FEAKVVLDLT GKTLGVNQRG ELCVRGPMIM
SGYVNNPEAT NALIDKDGVD

CFluc398

MW = 17841 Da

FGGGSGGSEF MSGYVNNPEA TNALIDKDGW LHSGDIAYWD EDEHFFIVDR LKSLIKYKGY QVAPAELESI LLQHPNIFDA
GVAGLPDDDA GELPAAVVVL EHGKTMTEKE IVDYVASQVT TAKKLRGGVV FVDEVPKGLT GKLDARKIRE ILIKAKKGGK
IAVLE

NFluc437

MW = 48914 Da

FGGEDAKNIK KGPAPFYPLE DGTAGEQLHK AMKRYALVPG TIAFTDAHIE VDITYAEYFE MSVRLAEAMK RYGLNTNHRI
VVCSENSLQF FMPVLGALFI GVAVAPANDI YNERELLSNM GISQPTVVV FV SKKGLQKILN VQKKLP I I QK I I I M D S K T D Y
QGFQSMYTFV TSHLPPGFNE YDFVPESFDR DKTIALIMNS SGSTGLPKGV ALPHRTACVR FSHARDPIFG NQIIPDTAIL
SVVFFHHGFG MFTTLGYLIC GFRVVMYRF EEELFLRSLQ DYKIQSALLV PTLFSFFAKS TLIDKYDLSN LHEIASGGAP
LSKEVGEAVA KRFHLPGIRQ GYGLTETTSA ILITPEGDDK PGAVGKVVVF FEAKVVLDLT GKTLGVNQRG ELCVRGPMIM
SGYVNNPEAT NALIDKDGWL HSGDIAYWDE DEHFFIVDRV D

CFluc438

MW = 13217 Da

FGGGSGGSEF LKSLIKYKGY QVAPAELESI LLQHPNIFDA GVAGLPDDDA GELPAAVVVL EHGKTMTEKE IVDYVASQVT
TAKKLRGGVV FVDEVPKGLT GKLDARKIRE ILIKAKKGGK IAVLE

Full-length luciferase

MW = 62676 Da (MW without methionine)

MGSSHHHHH SSGLVPRGSH MEDAKNIKKG PAPPFYPLEDG TAGEQLHKAM KRYALVPGTI AFTDAHIEVD IITYAEYFEMS
VRLAEAMKRY GLNTNHRIV CSENSLQFFM PVLGALFIGV AVAPANDIYN ERELLNSMGI SQPTVVVFSK KGLQKILNVQ
KKLP I I QK I I I M D S K T D Y Q G F Q S M Y T F V T S H L P P G F N E Y D F V P E S F D R D K T I A L I M N S S G S T G L P K G V A L P H R T A C V R F S
H A R D P I F G N Q I I P D T A I L S V P F H H G F G M F T T L G Y L I C G F R V V L M Y R F E E L F L R S L Q D Y K I Q S A L L V P T L F S F F A K S T L
I D K Y D L S N L H E I A S G G A P L S K E V G E A V A K R F H L P G I R Q G Y G L T E T T S A I L I T P E G D D K P G A V G K V V P F F E A K V V L D L T G K
T L G V N Q R G E L C V R G P M I M S G Y V N N P E A T N A L I D K D G W L H S G D I A Y W D E D E H F F I V D R L K S L I K Y K Y Q V A P A E L E S I L L Q
H P N I F D A G V A G L P D D D A G E L P A A V V V L E H G K T M T E K E I V D Y V A S Q V T T A K K L R G G V V F V D E V P K G L T G K L D A R K I R E I L I
K A K K G G K I A V

References

- [1] K. E. Luker, M. C. P. Smith, G. D. Luker, S. T. Gammon, H. Piwnica-Worms, D. Piwnica-Worms, *Proc. Natl. Acad. Sci. U. S. A.* **2004**, *101*, 12288–12293.
- [2] R. Paulmurugan, Y. Umezawa, S. S. Gambhir, *Proc. Natl. Acad. Sci. U. S. A.* **2002**, *99*, 15608–15613.
- [3] R. Paulmurugan, S. S. Gambhir, *Anal. Chem.* **2005**, *77*, 1295–1302.
- [4] K. F. Geoghegan, H. B. F. Dixon, P. J. Rosner, L. R. Hoth, A. J. Lanzetti, K. A. Borzilleri, E. S. Marr, L. H. Pezzullo, L. B. Martin, P. K. LeMotte, et al., *Anal. Biochem.* **1999**, *267*, 169–184.
- [5] S. Yi, A. E. Kaifer, *J. Org. Chem.* **2011**, *76*, 10275–10278.