

Tuning the flexibility of glycine-serine linkers to allow rational design of multidomain proteins – Supporting information

Martijn van Rosmalen, Mike Krom and Maarten Merkx*

* corresponding author: m.merkx@tue.nl

Contents

Supplementary Figure S1 Sequences of proteins used in this study	S2
Supplementary Figure S2 Relationship between direct EYFP excitation at 514 and 420 nm	S4
Supplementary Figure S3 Comparison of experimental FRET efficiencies with those predicted with Gaussian Chain models	S4
Supplementary References	S4

CLY- (GS₅)₉

atgggcagcagccatcatcatcatcacagcagcggcctgggtgccgcgcgagccatgatggtgagcaagggc
M G S S H H H H H S S G L V P R G S H M V S K G
gaggagctgttcaccgggggtggtgcccatcctggtcgagctggacggcgacgtaaaccggccacaagttcagcgtg
E E L F T G V V P I L V E L D G D V N G H K F S V
ggcgagggcgagggcgatgccacctcctacggcaagctgacctgaagttcatctgcaccaccggcaagctgccc
G E G E G D A T S Y G K L T L K F I C T T G K L P
gtgccctggcccaccctcgtgaccacctgacctggggcgctgcagtgttcagccgctaccccgaccatgaag
V P W P T L V T T L T W G V Q C F S R Y P D H M K
cagcagcacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggc
Q H D F F K S A M P E G Y V Q E R T I F F K D D G
aactacaagaccgcgcccaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgac
N Y K T R A E V K F E G D T L V N R I E L K G I D
ttcaaggaggacggcaacatcctggggcacaagctggagtacaactacatcagccacaacgtctatatcaccggc
F K E D G N I L G H K L E Y N Y I S H N V Y I T A
gacaagcagaagaacggcatcaaggccaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgcc
D K Q K N G I K A N F K I R H N I E D G S V Q L A
gaccactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagcaccag
D H Y Q Q N T P I G D G P V L L P D N H Y L S T Q
tccgccctgagcaaagacccaacgagaagcgcgatcacatggctcctgctggagttcgtgaccgcccggggtac
S A L S K D P N E K R D H M V L L E F V T A A G I

EcoRI *BamHI* *BamHI*
actctcggcatggacgagctgtacaagtcgggaattcgtagtgatcctcgtcaagctcaggatcctcttcgagc
T L G M D E L Y K S G I R S G S S S S S S S S S S

BamHI *BamHI* *BamHI* *BamHI*
tcaggatcctcgtcgagctctggatccagcagttcaagcggatcctcttcgagctctggatccagttcaagttcg
S G S

BamHI *BamHI* *BamHI* *SacII*

ggatccagctcttctagtgatcctctagtagctcgggatccagttcaagcacaaccgggatgggtgagcaagggc
G S S S S S G S S S S S S G S S S S T P R M V S K G
gaggagctgttcaccgggggtggtgcccatcctggtcgagctggacggcgacgtaaaccggccacaagttcagcgtg
E E L F T G V V P I L V E L D G D V N G H K F S V
tccggcgagggcgagggcgatgccacctacggcaagctgacctgaagttcatctgcaccaccggcaagctgccc
S G E G E G D A T Y G K L T L K F I C T T G K L P
gtgccctggcccaccctcgtgaccaccttcggctacggcctgcagtgttcgcccgtaccccgaccatgaag
V P W P T L V T T F G Y G L Q C F A R Y P D H M K
cagcagcacttcttcaagtccgccatgcccgaaggctacgtccaggagcgcaccatcttcttcaaggacgacggc
Q H D F F K S A M P E G Y V Q E R T I F F K D D G
aactacaagaccgcgcccaggtgaagttcgagggcgacaccctggtgaaccgcatcgagctgaagggcatcgac
N Y K T R A E V K F E G D T L V N R I E L K G I D
ttcaaggaggacggcaacatcctggggcacaagctggagtacaactacaacagccacaacgtctatatcatggcc
F K E D G N I L G H K L E Y N Y N S H N V Y I M A
gacaagcagaagaacggcatcaaggtgaacttcaagatccgccacaacatcgaggacggcagcgtgcagctcgcc
D K Q K N G I K V N F K I R H N I E D G S V Q L A
gaccactaccagcagaacacccccatcggcgacggccccgtgctgctgcccgacaaccactacctgagctaccag
D H Y Q Q N T P I G D G P V L L P D N H Y L S Y Q
tccgccctgagcaaagacccaacgagaagcgcgatcacatggctcctgctggagttcgtgaccgcccggggtac
S A L S K D P N E K R D H M V L L E F V T A A G I
actctcggcatggacgagctgtacaagtgagccaccgcagttcgaaaaataa
T L G M D E L Y K W S H P Q F E K -

His-tag EYFP
ECFP Strep-tag
Linker

Supplementary Figure S1 Sequences of proteins used in this study

CLY-(G₂S₄)₉ linker

EcoRI *BamHI* *BamHI* *BamHI*
ggaattcgttagtggatcctcgggatcatcaggatcctctgggagctcaggatcctcggg
G I R S G S S G S S G S S G S S G S S G

 BamHI *BamHI* *BamHI*μ
agctctggatccagcgggttcaagcggatcctcagggagctctggatccagtggtagttcg
S S G S S G S S G S S G S S G S S G S S

BamHI *BamHI* *BamHI* *SacII*
ggatccagcgggatctagtggatcctctggtagttcgggatccagtggtagc ccgcgg
G S S G S S G S S G S S G S S G S P R

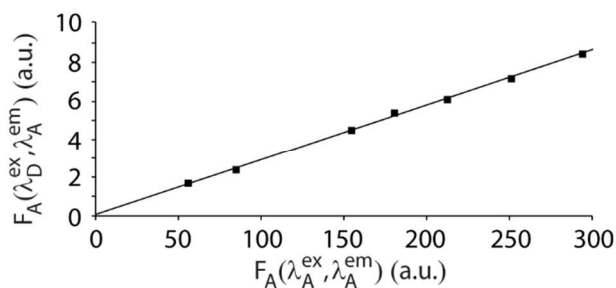
CLY-(S₆)₉ linker

EcoRI *SacI* *SacI* *SacI*
ggaattcgttccgagctcatccagttcttcgagctctagttcctcatcgcagctcctccagc
G I R S S S S S S S S S S S S S S S S S

 SacI *SacI* *SacI*
tcatcgcagctcaagttcatcctcgcagctcttcgctcatcttcgcagctctccagtagttcg
S S S S S S S S S S S S S S S S S S S

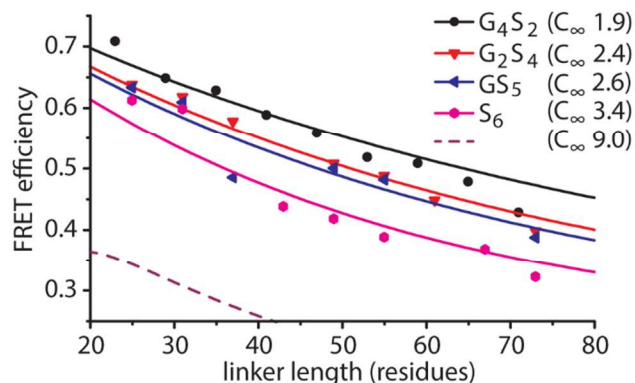
SacI *SacI* *SacI* *SacII*
agctctagttcgtcctcgcagctcatcttcaagttcgcagctccagttcatccaca ccgcgg
S S S S S S S S S S S S S S S S S S T P R

Supplementary Figure S1 (continued) Sequences of proteins used in this study.



Supplementary Figure S2 Relationship between direct EYFP excitation at 514 and 420 nm.

Fluorescence intensities for a concentration series of 100, 140, 230, 260, 310, 350 and 400 nM EYFP in 20 mM TrisHCl pH 8.0 containing 100 mM NaCl, 20 mM EDTA and 10% (v/v) glycerol, exciting at 420 nm (vertical axis) or at 514 nm (horizontal axis). Emission was recorded from 523 nm to 533 nm in both cases. A straight line was fit through the data (Equation 2 in the main text).



Supplementary Figure S3 Comparison of experimental FRET efficiencies with those predicted with Gaussian Chain models. Symbols show E_{obs} for all CLY proteins. Solid lines represent best fit of modelled $\langle E \rangle_{ensemble}$ vs. linker length curves with indicated values of C_{∞} . The purple dashed line ($C_{\infty} = 9$) is the modelled FRET efficiency vs. linker length curve for polyserine according to the Flory model^{1,2}.

Supplementary References

- (1) Brant, D. A., and Flory, P. J. (1965) The configuration of random polypeptide chains. II. Theory. *J. Am. Chem. Soc.* 87, 2791–2800.
- (2) Miller, W. G., Brant, D. A., and Flory, P. J. (1967) Random coil configurations of polypeptide copolymers. *J. Mol. Biol.* 23, 67–80.