CASP - CBM56 - AG Seidel

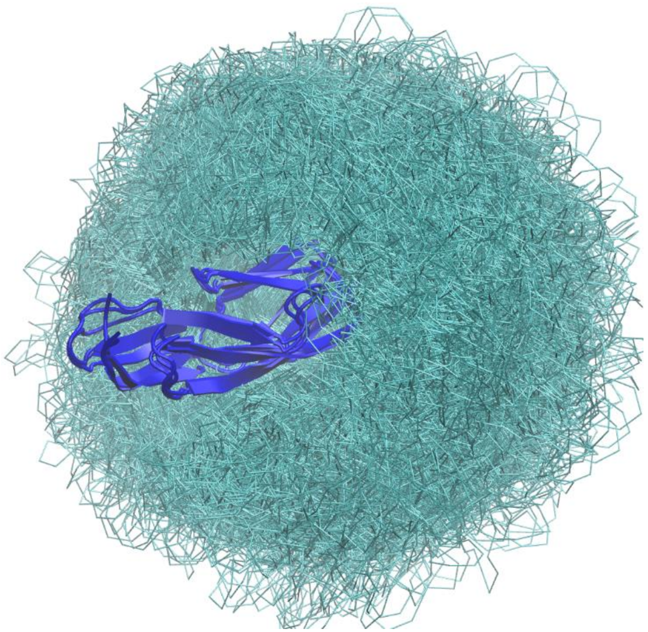
Table of contents:

FRET pair selection for FRET measurements

1. We started from the provided sequence (cf. WT in Table 1).

2. Using de-novo structure prediction we created initial models for two parts of the proteins (residues 1 to 85 and 86 to 184). Structural models for each individual part were generated separately using several templates.

3. From the initial structural model, coarse-grained simulations using NMSim were used to generate structural ensembles of possible conformations, focusing on the mutual orientation of the two parts. From these ensembles, the optimal FRET donor-acceptor pairs were determined following the method presented in Dimura et al., Current Opinion in Structural Biology, 2016. 40: p. 163-185 (DOI: 10.1016/j.sbi.2016.11.012).



Prior ensemble generated using coarse-grained NMSim simulations highlighting the mutual orientation of the two domains. The structures were overlaid on residues 1 to 85 (blue cartoon), while the backbone of residues 86 to 184 is shown in cyan. This ensemble was used to determine the optimal FRET donor-acceptor pairs.

Variants

Table 1: Sequences of the variants (mutations in brackets were introduced to allow for labeling; these residues are highlighted in red in the sequence)

Variant [a]	Sequence (mutated residues are colored in red)
WT [b]	NVPVTGVTVNPTTAQVEVGQSVQLNASVAPSNATNKQVTWSVSGSSIASVSPNGLVTGLAQGTTTVTATTADGNKAASATITVAPAPSTVIVIGDEVKGLK KIGDDLLFYVNGATFADLHYKVNNGGQLNVAMAPTGNGNYTYPVHNLKHGDTVEYFFTYNPGQGALDTPWQTYVHGVTQGTPE
1 (A29C, D168C)	NVPVTGVTVNPTTAQVEVGQSVQLNASV**C**PSNATNKQVTWSVSGSSIASVSPNGLVTGLAQGTTTVTATTADGNKAASATITVAPAPSTVIVIGDEVKGLK KIGDDLLFYVNGATFADLHYKVNNGGQLNVAMAPTGNGNYTYPVHNLKHGDTVEYFFTYNPGQGAL**C**TPWQTYVHGVTQGTPE
2 (N35C, T115C)	NVPVTGVTVNPTTAQVEVGQSVQLNASVAPSNAT**C**KQVTWSVSGSSIASVSPNGLVTGLAQGTTTVTATTADGNKAASATITVAPAPSTVIVIGDEVKGLK KIGDDLLFYVNGA**C**FADLHYKVNNGGQLNVAMAPTGNGNYTYPVHNLKHGDTVEYFFTYNPGQGALDTPWQTYVHGVTQGTPE
3 (A29C, N147C)	NVPVTGVTVNPTTAQVEVGQSVQLNASV**C**PSNATNKQVTWSVSGSSIASVSPNGLVTGLAQGTTTVTATTADGNKAASATITVAPAPSTVIVIGDEVKGLK KIGDDLLFYVNGATFADLHYKVNNGGQLNVAMAPTGNGNYTYPVH**C**LKHGDTVEYFFTYNPGQGALDTPWQTYVHGVTQGTPE
4 (S46C, T115C)	NVPVTGVTVNPTTAQVEVGQSVQLNASVAPSNATNKQVTWSVSGS**C**IASVSPNGLVTGLAQGTTTVTATTADGNKAASATITVAPAPSTVIVIGDEVKGLK KIGDDLLFYVNGA**C**FADLHYKVNNGGQLNVAMAPTGNGNYTYPVHNLKHGDTVEYFFTYNPGQGALDTPWQTYVHGVTQGTPE
5 (S46C, D106C)	NVPVTGVTVNPTTAQVEVGQSVQLNASVAPSNATNKQVTWSVSGS**C**IASVSPNGLVTGLAQGTTTVTATTADGNKAASATITVAPAPSTVIVIGDEVKGLK KIGD**C**LLFYVNGATFADLHYKVNNGGQLNVAMAPTGNGNYTYPVHNLKHGDTVEYFFTYNPGQGALDTPWQTYVHGVTQGTPE
6 (N35C, N124C)	NVPVTGVTVNPTTAQVEVGQSVQLNASVAPSNAT**C**KQVTWSVSGSSIASVSPNGLVTGLAQGTTTVTATTADGNKAASATITVAPAPSTVIVIGDEVKGLK KIGDDLLFYVNGATFADLHYKV**C**NGGQLNVAMAPTGNGNYTYPVHNLKHGDTVEYFFTYNPGQGALDTPWQTYVHGVTQGTPE
7 (N74C, D168C)	NVPVTGVTVNPTTAQVEVGQSVQLNASVAPSNATNKQVTWSVSGSSIASVSPNGLVTGLAQGTTTVTATTADG**C**KAASATITVAPAPSTVIVIGDEVKGLK KIGDDLLFYVNGATFADLHYKVNNGGQLNVAMAPTGNGNYTYPVHNLKHGDTVEYFFTYNPGQGAL**C**TPWQTYVHGVTQGTPE
8 (N35C, N138C)	NVPVTGVTVNPTTAQVEVGQSVQLNASVAPSNAT**C**KQVTWSVSGSSIASVSPNGLVTGLAQGTTTVTATTADGNKAASATITVAPAPSTVIVIGDEVKGLK KIGDDLLFYVNGATFADLHYKVNNGGQLNVAMAPTG**C**GNYTYPVHNLKHGDTVEYFFTYNPGQGALDTPWQTYVHGVTQGTPE
9 (A29C, N124C)	NVPVTGVTVNPTTAQVEVGQSVQLNASV**C**PSNATNKQVTWSVSGSSIASVSPNGLVTGLAQGTTTVTATTADGNKAASATITVAPAPSTVIVIGDEVKGLK KIGDDLLFYVNGATFADLHYKV**C**NGGQLNVAMAPTGNGNYTYPVHNLKHGDTVEYFFTYNPGQGALDTPWQTYVHGVTQGTPE
10 (S46C, Q164C)	NVPVTGVTVNPTTAQVEVGQSVQLNASVAPSNATNKQVTWSVSGS**C**IASVSPNGLVTGLAQGTTTVTATTADGNKAASATITVAPAPSTVIVIGDEVKGLK KIGDDLLFYVNGATFADLHYKVNNGGQLNVAMAPTGNGNYTYPVHNLKHGDTVEYFFTYNPG**C**GALDTPWQTYVHGVTQGTPE

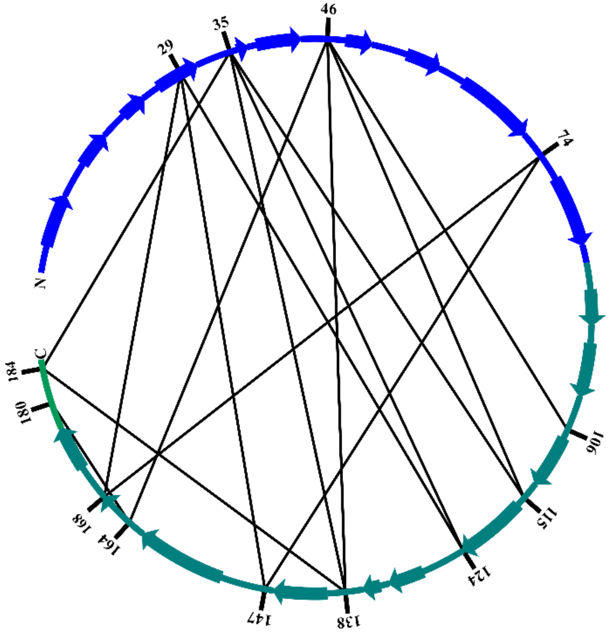
11 (S46C, N138C)	NVPVTGVTVNPTTAQVEVGQSVQLNASVAPSNATNKQVTWSVSGS**C**IASVSPNGLVTGLAQGTTTVTATTADGNKAASATITVAPAPSTVIVIGDEVKGLK KIGDDLLFYVNGATFADLHYKVNNGGQLNVAMAPTG**C**GNYTYPVHNLKHGDTVEYFFTYNPGQGALDTPWQTYVHGVTQGTPE
12 (N74C, N147C)	NVPVTGVTVNPTTAQVEVGQSVQLNASVAPSNATNKQVTWSVSGSSIASVSPNGLVTGLAQGTTTVTATTADG**C**KAASATITVAPAPSTVIVIGDEVKGLK KIGDDLLFYVNGATFADLHYKVNNGGQLNVAMAPTGNGNYTYPVH**C**LKHGDTVEYFFTYNPGQGALDTPWQTYVHGVTQGTPE
13 (N138C, E184C)	NVPVTGVTVNPTTAQVEVGQSVQLNASVAPSNATNKQVTWSVSGSSIASVSPNGLVTGLAQGTTTVTATTADGNKAASATITVAPAPSTVIVIGDEVKGLK KIGDDLLFYVNGATFADLHYKVNNGGQLNVAMAPTG**C**GNYTYPVHNLKHGDTVEYFFTYNPGQGALDTPWQTYVHGVTQGTP**C**
14 (Q164C,Q180C)	NVPVTGVTVNPTTAQVEVGQSVQLNASVAPSNATNKQVTWSVSGSSIASVSPNGLVTGLAQGTTTVTATTADGNKAASATITVAPAPSTVIVIGDEVKGLK KIGDDLLFYVNGATFADLHYKVNNGGQLNVAMAPTGNGNYTYPVHNLKHGDTVEYFFTYNPG**C**GALDTPWQTYVHGVT**C**GTPE

[a] Mutations in brackets were introduced to allow for labeling, counted from 1 (corresponds to residue 694 in the complete sequence).

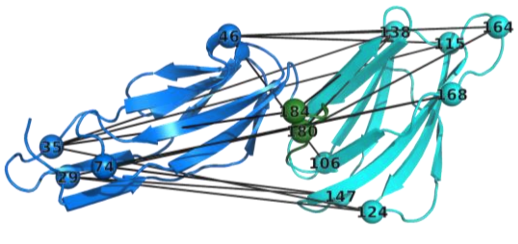
[b] Sequence as initially provided; Residues 694 to 877 of the original sequence (184 amino acids)

Network of donor-acceptor pairs

The following picture shows the sequence of the protein (in a circular representation) together with the selected FRET network. The two domains are colored blue and dark cyan, respectively. Labeling positions are indicated by the numbers and black lines between labeling positions indicate a donor-acceptor pair. The C-terminal section is colored in green.



Variant	
1 (A29C, D168C)	probes orientation of the two domains
2 (N35C, T115C)	probes orientation of the two domains
3 (A29C, N147C)	probes orientation of the two domains
4 (S46C, T115C)	probes orientation of the two domains
5 (S46C, D106C)	probes orientation of the two domains
6 (N35C, N124C)	probes orientation of the two domains
7 (N74C, D168C)	probes orientation of the two domains
8 (N35C, N138C)	probes orientation of the two domains
9 (A29C, N124C)	probes orientation of the two domains
10 (S46C, Q164C)	probes orientation of the two domains
11 (S46C, N138C)	probes orientation of the two domains
12 (N74C, N147C)	probes orientation of the two domains
13 (N138C, E184C)	probes structure of C-terminal loop/tail
14 (Q164C, Q180C)	probes structure of C-terminal loop/tail



Note: The structure shown on the left is for visualization of the labeling positions only. It is a member of the initial ensemble, but not necessarily a solution.

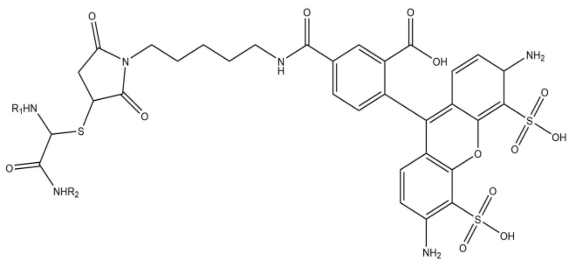
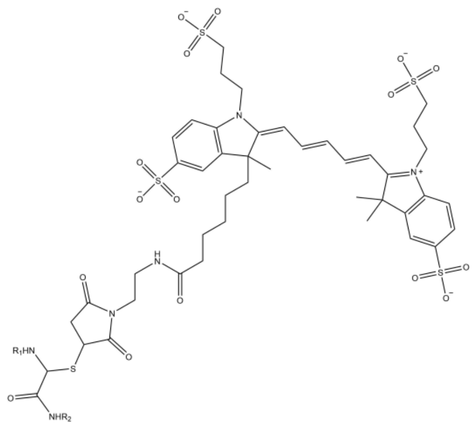
Labeling positions are shown as spheres and labeled according to their residue numbers. Black lines between labeling positions represents a donor-acceptor pair that is measured.

Dyes used in the FRET measurements

For the FRET measurements, a donor and an acceptor dye molecule are attached to the protein via flexible linkers.

We used as FRET-**donor**: Alexa488 with C5-maleimide linker
and as FRET-**acceptor**: Alexa647 with C2-maleimide linker

Both dyes are attached via their linkers to Cysteine residues in the protein.
The Cysteine mutations were introduced at the indicated positions (cf. Table 1)

<u>**Donor**</u> **Alexa488 with C5-maleimide linker attached to Cysteine**	<u>**Acceptor**</u> **Alexa647 with C2-maleimide linker attached to Cysteine**
	
SMILES: O=C(C(N[R1])SC1CC(N(CCCCCNC(C2=CC=C(C(C(C=CC(N)=C3S(O)(=O)=O)=C3O4)=C5C4=C(S(O)(=O)=O)C(N)C=C5)C(C(O)=O)=C2)=O)C1=O)=O)N[R2]	SMILES: O=C(C(N[R1])SC1CC(N(CCNC(CCCCCC2(C)/C(N(CCCS(=O)([O-])=O) C3=C2C=C(S(=O)([O-])=O)C=C3)=C\C=C\C=C\C4=[N+](CCCS(=O)([O-]) =O)C5=C(C=C(S(=O)([O-])=O)C=C5)C4(C)C)=O)C1=O)=O)N[R2]

Labeling of variants

The following list describes the status of the work on the selected variants. For nine of the variants, labeling was successful, while the remaining five variants still require some work.

Variant (mutations introduced for labeling)	Labeling successful	Dynamic shift visible?
1 (A29C, D168C)	OK	yes
2 (N35C, T115C)	OK	no
3 (A29C, N147C)	OK	no
4 (S46C, T115C)	OK	no
5 (S46C, D106C)	work in progress	work in progress
6 (N35C, N124C)	OK	no
7 (N74C, D168C)	work in progress	work in progress
8 (N35C, N138C)	work in progress	work in progress
9 (A29C, N124C)	work in progress	work in progress
10 (S46C, Q164C)	work in progress	work in progress
11 (S46C, N138C)	OK	yes
12 (N74C, N147C)	OK	yes
13 (N138C, E184C)	OK	yes
14 (Q164C, Q180C)	OK	yes

What we can provide

* NEW: The observed dynamics in the measurements suggests that we have a multi-state system (by PDB-Dev definition).
* Therefore, we would provide multi-state datasets (and time scales) (at least two datasets) consisting of:
    * Inter-dye distances and uncertainties
    * protein sequences (cf. Table 1; page 3-4)

* For information on the usage of inter-dye distance in structural modeling, please see the next page.

Using FRET data for structural modeling

FRET measurements are a label-based technique, where the distance between a donor dye and an acceptor dye are measured. The donor and acceptor dyes are flexibly coupled to specific protein residues.

Suggestions on how to process the FRET data:

Method	Description	Remark
Converting the distances to C_α-C_α distances	Note: As FRET measurements report on distances between the flexibly coupled dyes, the use of C_α-C_α results in high uncertainty and thus low resolution (cf. Kalinin 2012, Nat. Methods, **9**(12): p. 1218-1227. DOI: 10.1038/nmeth.2222). In this case uncertainty in FRET-based C_α-C_α distances is likely to be in the range of 20-30 Å if provided naively, this completely defeats the purpose of FRET-based restraints, which usually have inter-dye distance uncertainties of 2-5 Å.	Low resolution (loss of information)
Atomic description of dyes, e.g. in MD simulations	Inter-dye distance distributions can be obtained from Molecular Dynamics simulations with explicit dye representation (DOI: 10.1021/ct500869p, 10.1016/j.bpj.2015.04.038). However, convergence time of such simulations can be restrictive (>1 µs) and the accuracy of the fluorophore force field parameters is not yet well established. Examples of such estimations: DOI: 10.1039/c4cp01222d	Sampling problem unsolved; fluorophore force field parameters are challenging
Usage of rotamer library	DOI: 10.1039/c4cp01222d (however, implementation of AV in this paper is slower than it could be (1h/AV), our implementation is ~30ms/AV). This approach is used extensively for EPR, DOI: 10.1039/C0CP01865A. Application to FRET is less attractive, since our linkers are longer and linker dynamics is well studied and usually good-behaved.	computationally slower than accessible volume calculations; application to FRET not streamlined
CNS; Axel Brunger group	“Crystallography & NMR System” (CNS) includes support for FRET-based restraints (DOI: 10.1038/nsmb.1763). They also provide some documentation on FRET functionality in CNS: https://www.mrc-lmb.cam.ac.uk/public/xtal/doc/cns/cns_1.3/tutorial/fret/	
Coarse-grained dye representation: NPS (Nano positioning system); Jens Michaelis group	DOI: 10.1016/j.cpc.2017.05.027	Generates credible positional space of dye molecules
Coarse-grained dye representation: FPS (FRET Positioning and Screening); Claus Seidel group	Using accessible volume (AV) or accessible contact volume (ACV) calculations, which is possible in the FPS software (DOI: 10.1038/nmeth.2222) (https://github.com/Fluorescence-Tools/FPSv1; http://www.mpc.hhu.de/software/fps.html) or using LabelLib (https://github.com/Fluorescence-Tools/LabelLib). LabelLib is a library for the simulation of small probes flexibly coupled to biomolecules (AVs); It is the engine behind the FPS software. LabelLib is mainly intented to be used by programmers and provides APIs for C/C++ and Python.	Allows for rigorous quantitative comparison of structural models; fast positional sampling of dye linkers/fluorophore positional distributions