Supplementary Material

for

Structure-guided design of a peptide lock for modular peptide binders

Patrick Ernst[1], Franziska Zosel[2], Christian Reichen[3], Daniel Nettels, Benjamin Schuler and Andreas Plückthun[*].

Department of Biochemistry, University Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland

Corresponding Author

*****Andreas Plückthun

Tel. +41-44-635 5570, Fax. +41-44-635 5712

E-mail: plueckthun@bioc.uzh.ch

Present Addresses

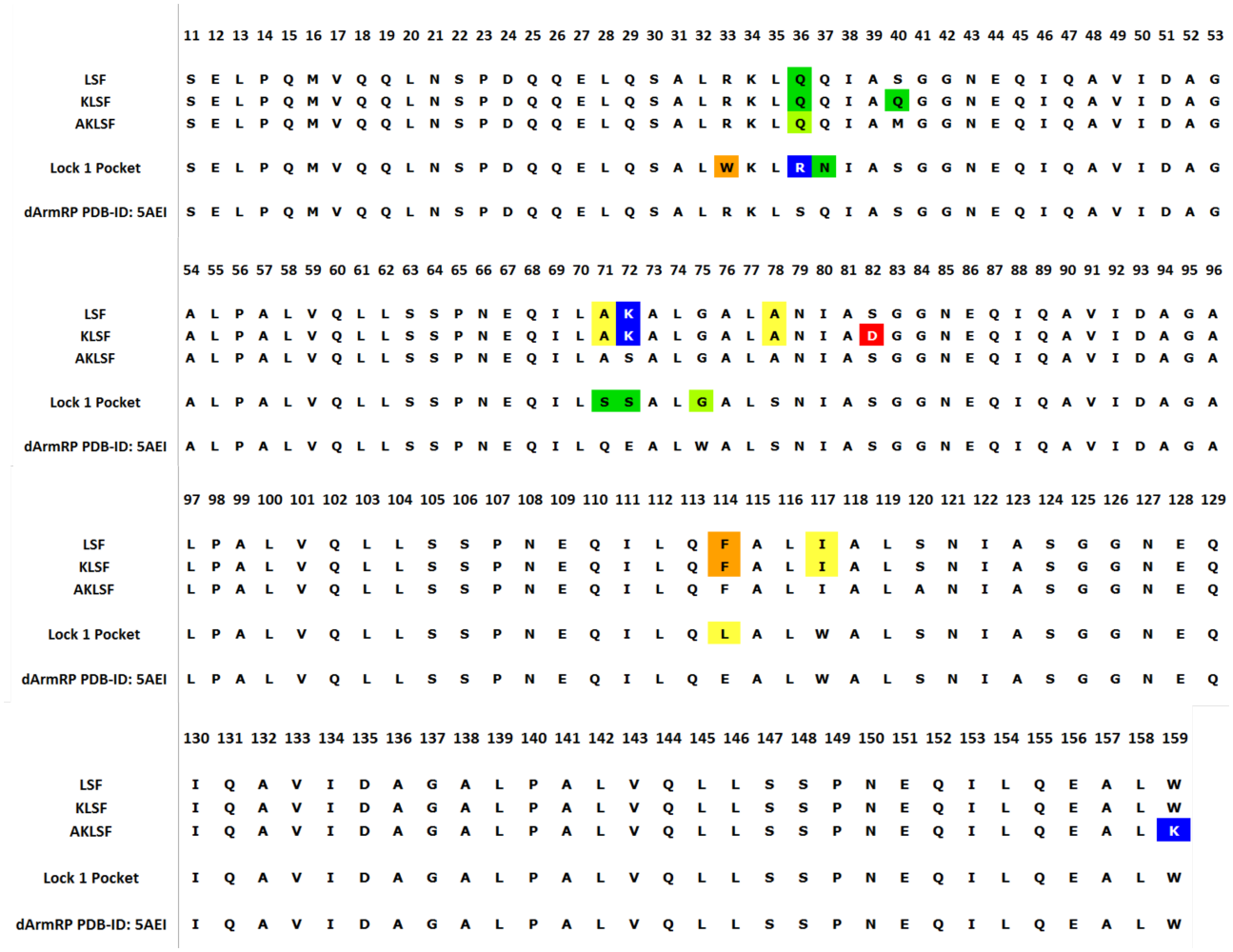
[1] Present address: University of Basel, Graduate Center, Petersgraben 35, 4051 Basel, Switzerland

[2] Present address: Novo Nordisk A/S, Novo Nordisk Park 1, 2760 Måløv, Denmark

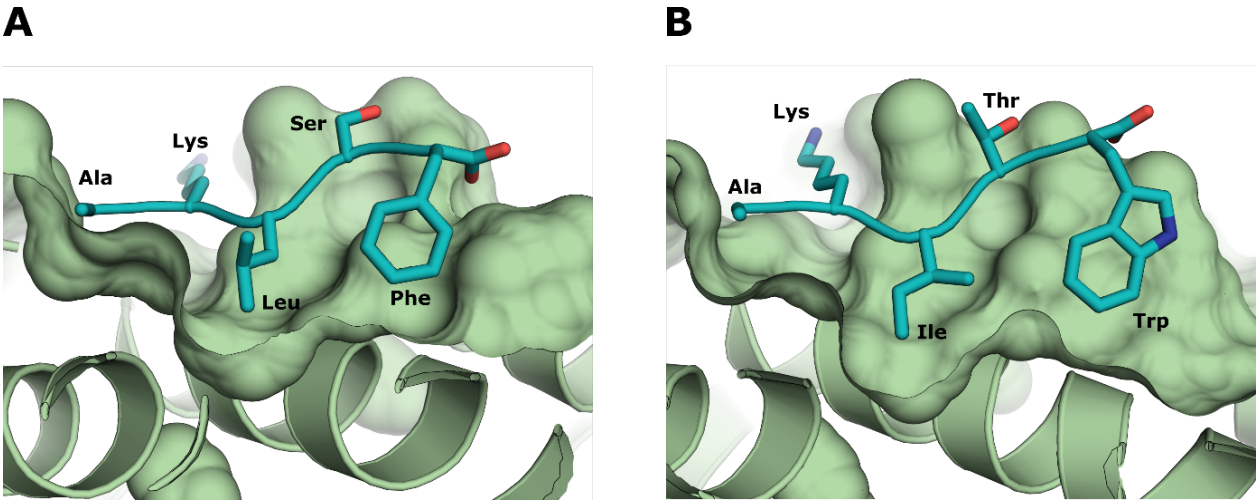
[3] Present address: Molecular Partners AG, Wagistrasse 14, 8952 Zürich-Schlieren, Switzerland

PDB-ID	Protein	Ligand	P3'	P4'	P5'	P6'	P7'	P8'	P9'
1G3J	β-catenin	XTCF3-CBD	D	E	L	I	R	F	K
1I7W	β-catenin	E-cadherin	D	S	L	L	V	F	D
1JDH	β-catenin	HTCF-4	D	E	L	I	S	F	K
1M1E	β-catenin	ICAT	D	V	V	M	A	F	S
1T08	β-catenin	APC	D	T	L	L	H	F	A
3OUX	β-catenin	pLef-1	D	E	M	I	P	F	K

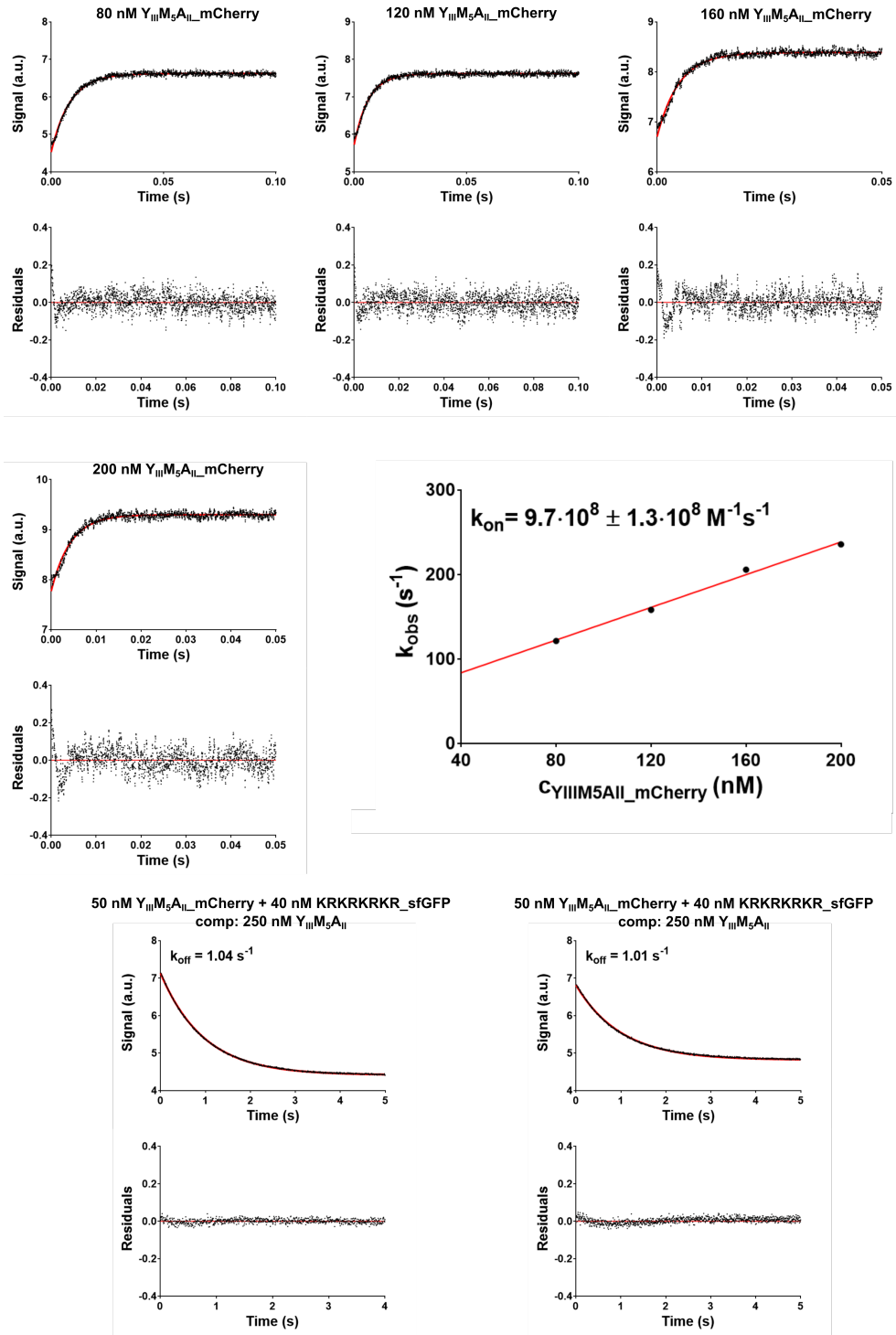
SI Figure 1: Sequence alignment of different natural ligands binding to β-catenin and plakoglobin structures. Amino acids are colored according to their type (red=negatively charged, green=polar uncharged, yellow=hydrophobic, cyan=positively charged and aromatic, orange=aromatic, blue=positively charged). Binding pocket numbers are shown above the respective amino acids.



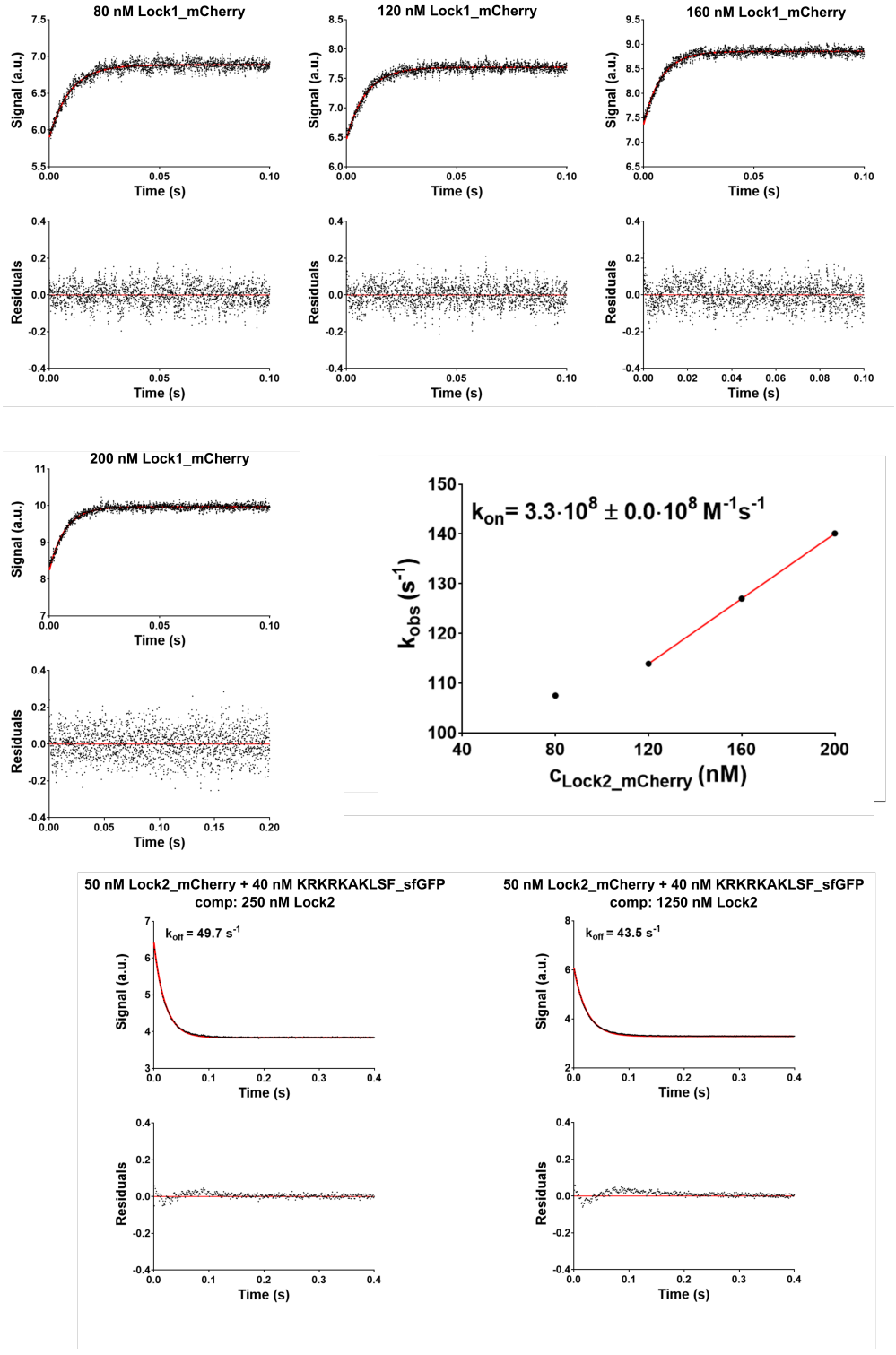
SI Figure 2: Sequence alignment of the different design models. Mutated amino acids are colored according to their type (red=negatively charged, green=polar uncharged, yellow=hydrophobic, orange=aromatic, blue=positively charged). Position W159K was not mutated in the AKLSF design as this lysine bound to the C-terminal carboxy group of the AKLSF peptide in the design model and was not involved in sidechain interactions.



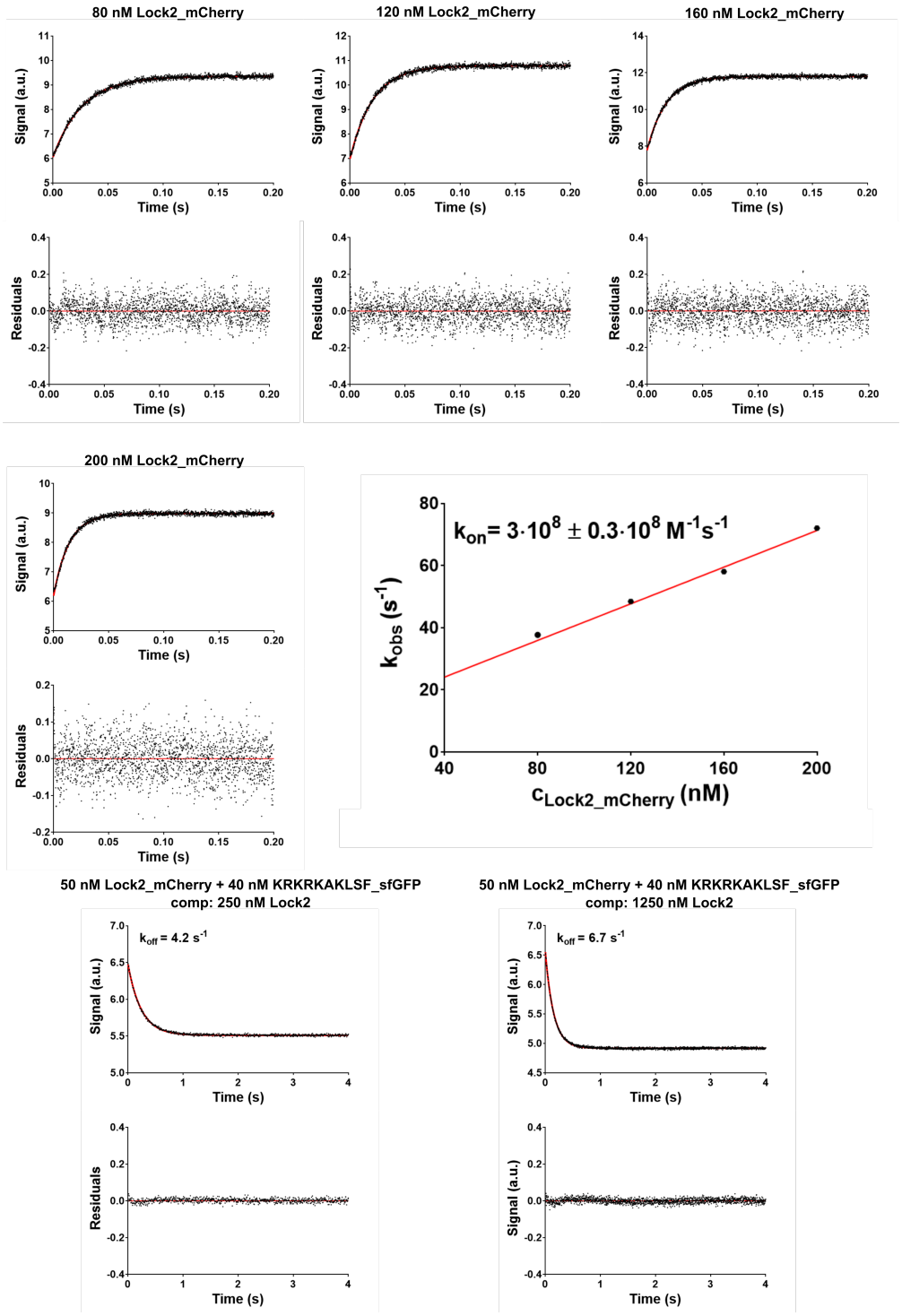
SI Figure 3: Surface representation of the second-generation lock bound to the LSF (A) or ITW (B) peptide moieties. dArmRP in green and bound peptide in cyan.



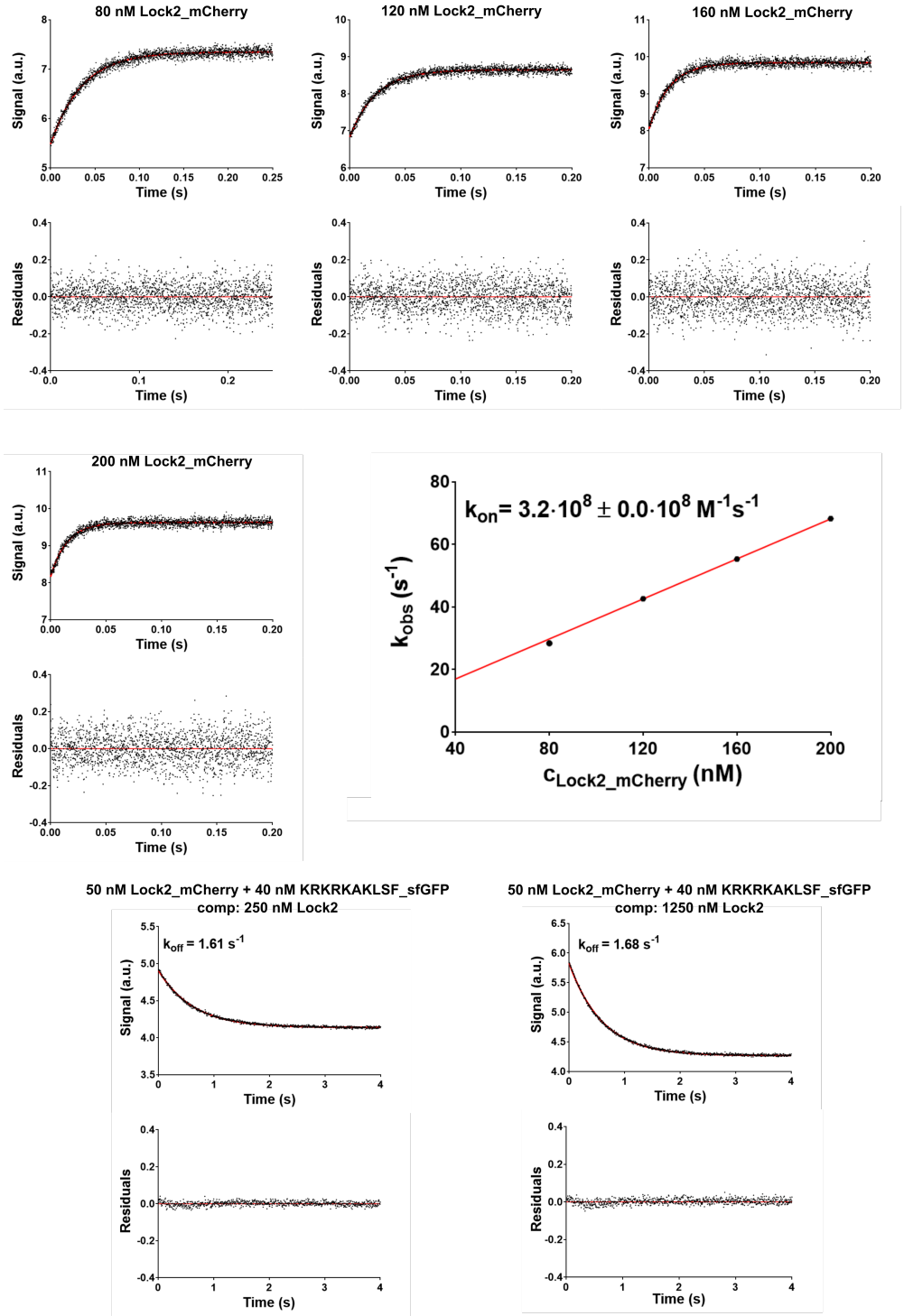
SI Figure 4: Determination of kinetic binding constants for the complex of $Y_{III}M_5A_{II}$_mCherry:KRKRKRKRKR-sfGFP. Upper panels: association kinetics at different concentrations of dArmRP with 40 nM KRKRKRKRKR-sfGFP. Lower panels: determination of dissociation rate by mixing complex with excess unlabeled competitor dArmRP. The residual plots of the respective fits (red) are shown below the measured kinetics.



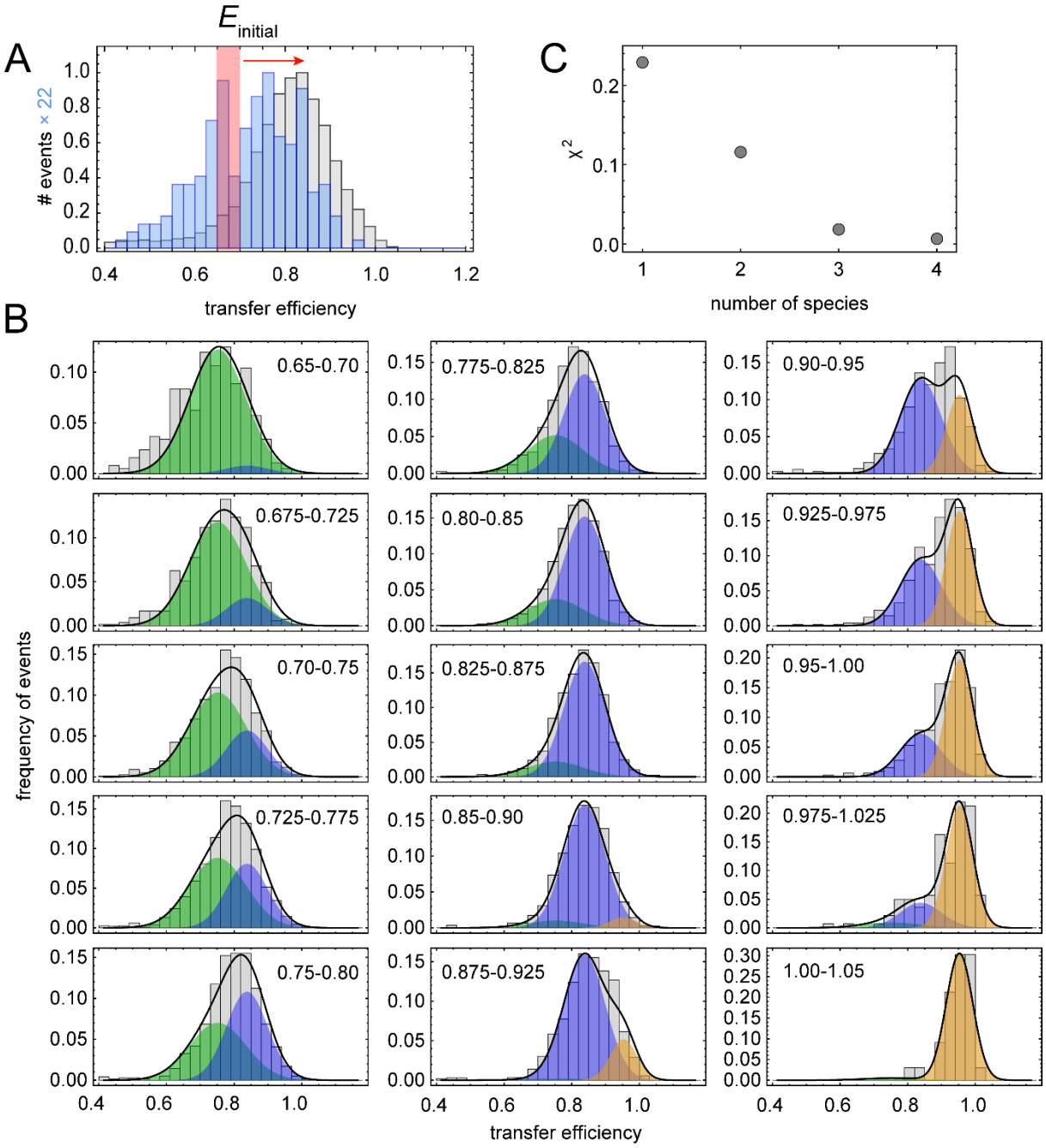
SI Figure 5: Fit of kinetic binding constants for the complex of mCherry-Lock 1:KRKAKRKLSF-sfGFP. Upper panels: association kinetics at different concentrations with 40 nM KRKAKRKLSF-sfGFP as present in the measurement. Lower panels: competition assays for the determination of the dissociation rate. The residual plots of the respective fits (red) are shown below the measured kinetics.



SI Figure 6: Fit of kinetic binding constants for the complex of mCherry-Lock 2:KRKAKRKLSF-sfGFP. Upper panels: association kinetics at different concentrations with 40 nM KRKAKRKLSF-sfGFP as present in the measurement. Lower panels: competition assays for the determination of the dissociation rate. The residual plots of the respective fits (red) are shown below the measured kinetics.

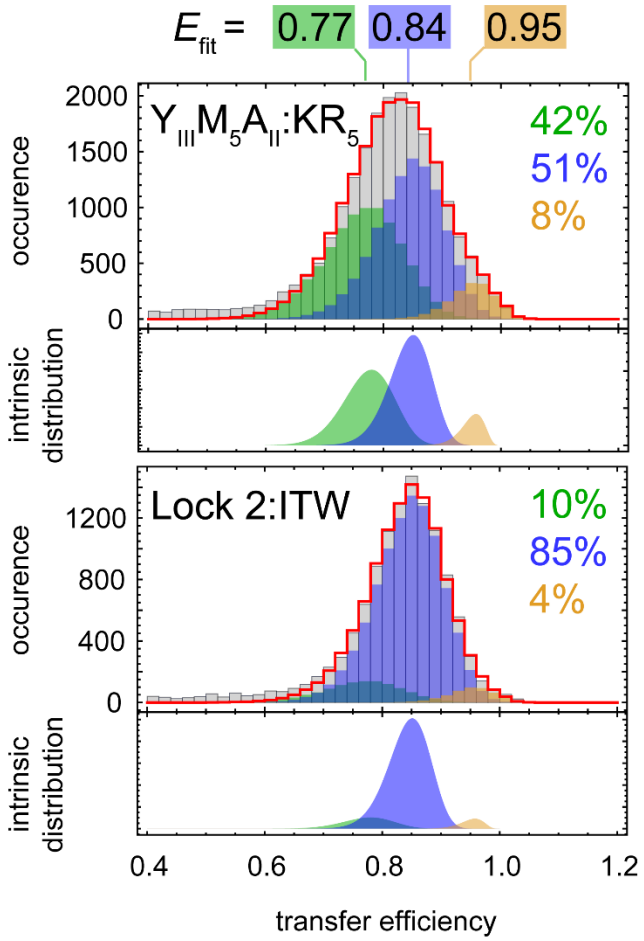


SI Figure 7: Fit of kinetic binding constants for the complex of mCherry-Lock 2:KRKAKRKITW-sfGFP. Upper panels: association kinetics at different concentrations with 40 nM KRKAKRKLSF-sfGFP as present in the measurement. Lower panels: competition assays for the determination of the dissociation rate. The residual plots of the respective fits (red) are shown below the measured kinetics.



SI Figure 8: Recurrence analysis of single particles (RASP) indicates that at least three species are required to fit the transfer efficiency histogram of the $Y_{III}M_5A_{II}$:KR_5 complex. (**A)** Principle of RASP. In a first step, all fluorescence bursts that fall in an initial transfer efficiency window ($E_{initial}$, shaded red) are located in the measurement. In a second step, all bursts occurring within 5 ms after the initial bursts are selected and used to construct the light blue transfer efficiency histogram (the "recurrence histogram"). Under these conditions, the probability that the first and the second burst come from the same molecule is >85%. If different molecular species with different transfer efficiency distributions contribute to a histogram, they can be isolated in this way (as long as they do not interchange in those 5 ms). The transfer efficiency histogram of the whole measurement is shown in the background in gray. **(B)** Burst pairs with $E_{initial}$ in the indicated range (on the right) are selected and the normalized transfer efficiency histogram of the bursts recurring

within 5 ms is plotted in gray. The transfer efficiency window for burst selection was moved from $E_{initial}$ = 0.65 to 1.0 in steps of 0.025. All 15 recurrence histograms are fitted globally with three Gaussian peak functions (green, blue and orange). (**C**) χ^2 analysis (sum of the squared residuals between histogram and fit) indicates that at least three Gaussian peak functions are required for a good fit of all recurrence histograms.



SI Figure 9: Transfer efficiency histograms of the $Y_{III}M_5A_{II}$:KR_5 complex (top) and the Lock 2:ITW complex (bottom), same as in Figure 5A. This time, measured transfer efficiency histograms (gray) are compared to a fit (red line) with three populations calculated from PDA assuming intrinsic peak broadening (in addition to shot-noise broadening). The intrinsic distributions of transfer efficiencies are given below the histograms. They share the same peak width and positions, and are shaded in green, blue and orange, respectively. Their fitted mean transfer efficiency E_{fit} is given on the top, their relative contribution to both histograms on the right.

Table 1 - Data collection and refinement statistics for structures. Values in parentheses show the data for the highest resolution shell.

	Lock_1:$(KR)_4$KLSF	Lock_2_$(GS)_6$_KRKRKAKLSF	Lock 2_$(GS)_6$_KRKRKAKITW	internalLock_2_$(GS)_6$_KRKRKLKFKR	internalLock_2_$(GS)_6$_KRKAKITWKR
PDB-ID	**6S9L**	**6S9M**	**6S9N**	**6S9O**	**6S9P**
condition	66 % (v/v) MPD	22.5 % (w/v) PEG smear medium	22.5 % (w/v) PEG smear medium	25 % (w/v) PEG smear low	10 % (v/v) Ethylene glycol, 18 % (w/v) PEG smear medium
		0.1 M $CaCl_2$, 0.1 M $MgCl_2$,	0.1 M $CaCl_2$, 0.1 M $MgCl_2$,	0.04 M CaCl2, 0.04 M Na-formate	0.1 M MgCl2, 0.1 M KCl
	0.1M HEPES NaOH pH 8	0.1 M Pipes pH 7	0.1 M Pipes pH 7	0.1 M Tris pH 8	
Data collection					
Resolution range	47.72 - 2.1 (2.175 - 2.1)	45.61 - 2.1 (2.175 - 2.1)	48.2 - 2.0 (2.071 - 2.0)	48.45 – 3.17 (3.283 – 3.17)	42.16 - 2.8 (2.9 - 2.8)
Space group	$P2_12_12_1$	$P2_1$	$P2_1$	C2	C2
Cell dimensions					
a, b, c (Å)	59.35, 80.26, 122.72	55.26, 85.5, 194.03	54.853, 85.474, 193.97	168.7, 82.4, 191.78	140.68 37.41 127.32
α, β, γ (°)	90, 90, 90	90, 96.334, 90	90, 96.323, 90	90, 90.364, 90	90 96.588 90
Total Reflections	472453 (48437)	557467 (46095)	793261 (77346)	307233 (31223)	110212 (9800)
Unique reflections	34965 (3443)	103532 (9974)	118079 (11689)	44419 (4437)	16544 (1601)
Multiplicity	13.5 (14.1)	5.4 (4.6)	6.7 (6.6)	6.9 (7.0)	6.7 (6.1)
Completeness (%)	99.95 (99.94)	98.57 (95.80)	97.98 (97.33)	98.16 (98.28)	98.59 (96.16)
I/σ(I)	6.79 (1.29)	6.84 (1.39)	14.23 (2.49)	5.82 (0.94)	7.12 (1.40)
Wilson B-factor	25.81	25.08	28.08	79.99	50.44
R_{merge}	0.3687 (2.324)	0.2449 (1.127)	0.1263 (0.9605)	0.3067 (1.962)	0.2997 (1.62)
R_{meas}	0.3831 (2.411)	0.2737 (1.274)	0.1373 (1.045)	0.3318 (2.119)	0.325 (1.768)
R_{pim}	0.1034 (0.6378)	0.1191 (0.5763)	0.0597 (0.4045)	0.1251 (0.793)	0.124 (0.6967)
$CC_{1/2}$	0.991 (0.506)	0.979 (0.618)	0.999 (0.92)	0.993 (0.778)	0.991 (0.637)
Refinement					
R-work	0.2071	0.2318 (0.3378)	0.1868 (0.2849)	0.2373 (0.3499)	0.2183 (0.3486)
R-free	0.2363	0.2677 (0.3509)	0.2189 (0.3007)	0.2921 (0.3847)	0.2689 (0.4082)
RMS(bonds)	0.004	0.018	0.021	0.003	0.004
RMS(angles)	0.94	1.80	1.93	0.53	0.94
Ramachandran plot (%)					
Favoured	99.66	99.78	99.78	98.66	99.25
Allowed	0.17	0.17	0.06	1.24	0.6
Outliers	0.17	0.06	0.17	0.1	0.15
Rotamer outliers (%)	0.21	2.37	2.28	0.75	0.77
Average B-factor ($Å^2$)	33.68	35.26	34.01	91.45	54.98
Non-hydrogen atoms	4900	14498	14653	15158	4962
Protein	4547	13302	13442	15081	4931
Ligand	56	22	139	56	16
Water	297	1174	1072	21	15

Rosetta Scripts Code for the redesign of Lock 1

XML-File:

```
<ROSETTASCRIPTS>

 <TASKOPERATIONS>
    ProteinInterfaceDesign name=pido repack_chain1=1 repack_chain2=1 design_chain1=0 design_chain2=1 interface_distance_cutoff=10/> task operation that designates
which residues are designable and repackable at the interface
 </TASKOPERATIONS>

 <FILTERS>
    <Ddg name=ddG scorefxn=talaris2014 threshold=-18.5 repeats=2/> binding energy calculation; an average of two repeats is computed for better numerical accuracy
    <Sasa name=sasa threshold=800  /> Buried surface area upon complex formation
    <Rmsd name=rmsd confidence=0/> confidence=0 means that the filter will be evaluated but not used as an acceptance criterion
    <CompoundStatement name=ddg_sasa> combine filters into a single logical statement
      <AND filter_name=ddG/>
      <AND filter_name=sasa/>
    </CompoundStatement>
 </FILTERS>

 <MOVERS>
    <Docking name=docking score_high=soft_rep fullatom=1 local_refine=1/> Invokes RosettaDock local-refinement (in full-atom) with a soft potential
    <BackrubDD name=backrub partner1=1 partner2=0 interface_distance_cutoff=8.0 moves=1000 sc_move_probability=0.25 scorefxn=talaris2014
small_move_probability=0.15 bbg_move_probability=0.25/> perturb the backbone of chain2
    <RepackMinimize name=des1 scorefxn_repack=soft_rep scorefxn_minimize=soft_rep minimize_bb=0 minimize_rb=1/>
    <RepackMinimize name=des2 scorefxn_repack=talaris2014 scorefxn_minimize=talaris2014 minimize_bb=0 minimize_rb=1/> Design & minimization at the interface
    <RepackMinimize name=des3 minimize_bb=1/>

     <ParsedProtocol name=design>
      <Add mover_name=des1/>
      <Add mover_name=des2/>
      <Add mover_name=des3/>
      <Add mover_name=backrub/>
      <Add mover_name=des3 filter_name=ddg_sasa/>
    </ParsedProtocol>
    <GenericMonteCarlo name=iterate scorefxn_name=talaris2014 mover_name=design trials=5/>
 </MOVERS>

 <PROTOCOLS>
    <Add mover=docking/>
    <Add mover=iterate/>
    <Add filter=ddG/>
    <Add filter=sasa/>
    <Add filter=rmsd/>
  </PROTOCOLS>
</ROSETTASCRIPTS>
```

Flag-file:

```
-s CR.pdb
-out::pdb_gz true
-correct
-ex1
-ex2
-extrachi_cutoff 5
-ignore_unrecognized_res
-run:constant_seed
-nstruct 200
-jd2:ntrials 10
-parser:protocol flexbb-interfacedesign.xml
-docking:no_filters
```

Rosetta Scripts Code for the redesign of the peptide on Lock 2

XML-File:

```
<ROSETTASCRIPTS>

 <TASKOPERATIONS>
    ProteinInterfaceDesign name=pido repack_chain1=1 repack_chain2=1 design_chain1=0 design_chain2=1 interface_distance_cutoff=10/> task operation that designates
which residues are designable and repackable at the interface
 </TASKOPERATIONS>

 <FILTERS>
    <Ddg name=ddG scorefxn=talaris2014 threshold=-24 repeats=2/> binding energy calculation; an average of two repeats is computed for better numerical accuracy
    <Sasa name=sasa threshold=950  /> Buried surface area upon complex formation
    <Rmsd name=rmsd confidence=0/> confidence=0 means that the filter will be evaluated but not used as an acceptance criterion
    <CompoundStatement name=ddg_sasa> combine filters into a single logical statement
      <AND filter_name=ddG/>
      <AND filter_name=sasa/>
    </CompoundStatement>
 </FILTERS>

 <MOVERS>
    <Docking name=docking score_high=soft_rep fullatom=1 local_refine=1/> Invokes RosettaDock local-refinement (in full-atom) with a soft potential
    <BackrubDD name=backrub partner1=1 partner2=0 interface_distance_cutoff=8.0 moves=1000 sc_move_probability=0.25 scorefxn=talaris2014
small_move_probability=0.15 bbg_move_probability=0.25/> perturb the backbone of chain2
    <RepackMinimize name=des1 scorefxn_repack=soft_rep scorefxn_minimize=soft_rep minimize_bb=0 minimize_rb=1/>
    <RepackMinimize name=des2 scorefxn_repack=talaris2014 scorefxn_minimize=talaris2014 minimize_bb=0 minimize_rb=1/> Design & minimization at the interface
    <RepackMinimize name=des3 minimize_bb=1/>

     <ParsedProtocol name=design>
      <Add mover_name=des1/>
      <Add mover_name=des2/>
      <Add mover_name=des3/>
      <Add mover_name=backrub/>
      <Add mover_name=des3 filter_name=ddg_sasa/>
    </ParsedProtocol>
    <GenericMonteCarlo name=iterate scorefxn_name=talaris2014 mover_name=design trials=5/>
 </MOVERS>

 <PROTOCOLS>
    <Add mover=docking/>
    <Add mover=iterate/>
    <Add filter=ddG/>
    <Add filter=sasa/>
    <Add filter=rmsd/>
  </PROTOCOLS>
</ROSETTASCRIPTS>
```

Flag-file:

```
-s CR.pdb
-out::pdb_gz true
-correct
-ex1
-ex2
-extrachi_cutoff 5
-ignore_unrecognized_res
-run:constant_seed
-nstruct 200
-jd2:ntrials 10
-parser:protocol flexbb-interfacedesign.xml
-docking:no_filters
```