SUPPORTING INFORMATION FOR

A delicate balance between functionally required flexibility and aggregation risk in a β-rich protein

Mylene C. Ferrolino, [§‡⊥] Anastasia Zhuravleva,[§⊥] Ivan Budyak,[§] Beena Krishnan, and Lila M. Gierasch[§‡#*]

Table S1. Aggregation propensity (a fraction of insoluble protein) of different CRABP1 WT* mutants and protein destabilization ($\Delta\Delta G^o$) upon single point mutations

CRABP 1 variant	Aggregation propensity	$\Delta\Delta G^o$ (kcal/mol)
WT	0	0 [1]
R29A	0	0.8 ± 0.3[1]
E69A	10	1.3 ± 0.2[1]
T75A	20	2.4 ± 0.3[1]
R79A	30	3.1 ± 0.2[1]
F50M	50	2.0 ± 0.2[2]
L118V	50	2.3 ± 0.4[1]
G78A	50	2.5 ± 0.4[1]
G68A	50	2.6 ± 0.2[1]
F71M	50	2.7 ± 0.2[2]
F65M	50	3.3 ± 0.2[2]
R135G	90	3.9 ± 0.3[1]
Y133S	90	4.8 ± 0.2[1]

Figure legends

Figure S1. NMR revealed a conformational transition at low urea concentrations. Chemical shifts (CSs) for amide ^{1}H (left) and ^{15}N (middle) atoms and normalized CSs (right, calculated as for Figure 3E) as a function of urea concentration. Only CRABP1 WT* residues with large CS perturbations (those highlighted in red in Figure 5) are shown.

Figure S2. Cysteine accessibility in different near-native CRABP1 conformations.
A. Three cysteine residues mapped on the CRABP1 near-native conformations [overlay of the X-ray structures of holo-CRABP1 (PDB ID 1CBR, chain A, white, represent the closed conformation) and apo-CRABP1 (PDB ID 1CBI, chain A, red, represent the open conformation and B, pink)]: (Right) the CRABP1 near-native conformations with the RA molecule shown in green (same as Figure 1 A). (Left) The cysteine residues are shown

as grey spheres with SG atoms highlighted in yellow. B. Accessibility of cysteine residues. (Left) SG atoms (shown in yellow, not visible) of three cysteine residues are completely protected from the solvent; the rest atoms are shown as white spheres (all-atom space-filling representation of CRABP1, PDB ID 1CBR, chain A). (Middle and right) Atoms within 5 Å cut-off from cysteine sulfur atoms are shown as a white spheres; sulfur atoms are highlighted in yellow, mapped on the structure of the closed conformation (PDB ID 1CBR, chain A) and open conformation (PDB ID 1CBI, chain A). C. Aggregation-prone mutations L18A, F71M and L118V do not significantly perturb solvent accessibility of CRABP1 cysteine residues: (Left) Atoms within 5 Å cut-off from cysteine SG atoms are shown as spheres; residues L18, F71 and L118 are highlighted in green and cysteine sulfur atoms are shown in yellow. (Right) Same as (Left) but the residues L18, F71 and L11 were mutated as L18A, F71M and L118V. Atoms within 5 Å cut-off from cysteine sulfur atoms are shown as white spheres, cysteine sulfur atoms are highlighted in yellow. The single-point mutations were made in Pymol.

Figure S3. Solvent accessibility of cysteine residues is increased in the open, aggregation-prone state of CRABP1.

A. Significant reduction in fluorescein fluorescence is observed for the fluorescein maleimide (Flu-Mal) treated CRABP1 WT* in the presence of RA. This suggests that a fraction of protein under native condition exists in a near-native state with solvent exposed sulfhydryl of the otherwise buried intrinsic cysteines of the protein. The solvent exposure of the cysteines in the near-native state involves the ligand (RA)-binding cavity and therefore the presence of RA precluded the fluorescein labeling of the holo-protein, or in other words stabilizes the native state and reduces the population of the near-native state. B-C. An increased fluorescein labeling of cysteine residues in native protein for the aggregation prone CRABP1 WT* mutants is observed indicative of an increased population of the near-native conformation for the variants. The reduced labeling of the cysteine mutants, C129A and Cys95 (a C81A/C129A mutant containing only a single cysteine) provide evidence for C129 and C95 are the residues with increased solvent accessibility in the near-native state of CRABP1. The fluorescein fluorescence is normalized with respect to amount of protein is represented in panel C,

and the error bars represent standard deviation from at least two independent experiments. The lanes with Flu-Mal quenched labeling in B are control experiment to show that labeling is thiol-mediated covalent linkage of the dye. In panels A and C, the image on the right represents UV exposed fluorescent image of Coomassie-stained gel image shown in the left, and the arrow indicates band corresponding to the unreacted Flu-Mal reagent.

Figure S4. Conformational transitions between closed and open conformations for CRABP1 and its aggregation-prone mutant F65M.
Distance variations between strands 4 and 5 (reported on barrel opening[2]) extracted from 24-ns MD trajectory of CRABP1 WT and its aggregation-prone F65M variant. The interstrand distance was defined as the distance between C^{α} atoms of T61 (strand 4) and E74 (strand 5).[2]

Figure S5. Aggregation-prone regions for different members of the iLBP family.
A. Experimental aggregation cores of CRABP 1 (strand 3-turn II-strand 4 and strand 9 and 10, shown in orange) mapped onto holo-CRABP1 (PDB ID 1CBR, chain A) (as reported by Budyak et al.[1] based on H/D exchange data). B-D. Aggregation-prone regions (in orange) predicted by the Zyggregator[3] based sequence features of CRABP1 (B), IFABP (C) and ILBP (D) mapped on the X-ray structures of the corresponding proteins. PDB code: 1CBR, chain A for CRABP1 (B), 2JU8 for IFABP (C) and PDB code 1EIO, chain A for ILBP (D) The ligand molecules (RA for CRABP1, oleate for IFABP and glycocholate for ILBP) are shown in green.

SI REFERENCES:

1. Budyak, I. L., Krishnan, B., Marcelino-Cruz, A. M., Ferrolino, M. C., Zhuravleva, A., and Gierasch, L. M. (2013) Early folding events protect aggregation-prone regions of a β-rich protein, *Structure 21*, 476-485.
2. Budyak, I. L., Zhuravleva, A., and Gierasch, L. M. (2013) The Role of Aromatic-Aromatic Interactions in Strand-Strand Stabilization of β-Sheets, *J Mol Biol 425*, 3522-3535.
3. Tartaglia, G. G., and Vendruscolo, M. (2008) The Zyggregator method for predicting protein aggregation propensities, *Chem Soc Rev 37*, 1395-1401.

Fig. S1

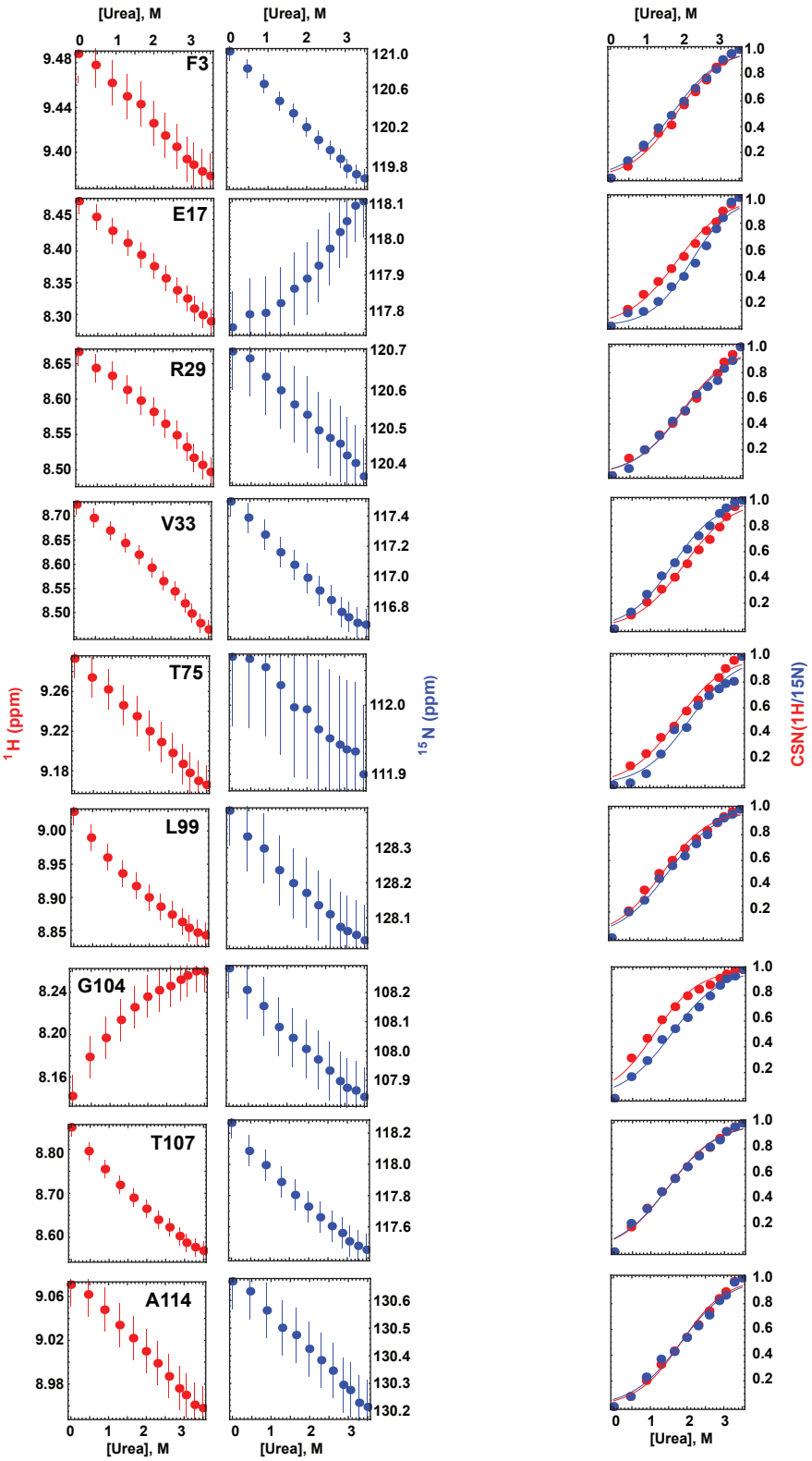

[Urea], M
[Urea], M
[Urea], M
F3
E17
R29
V33
T75
L99
G104
T107
A114
^{1}H (ppm)
^{15}N (ppm)
CSN(1H/15N)

Fig. S2

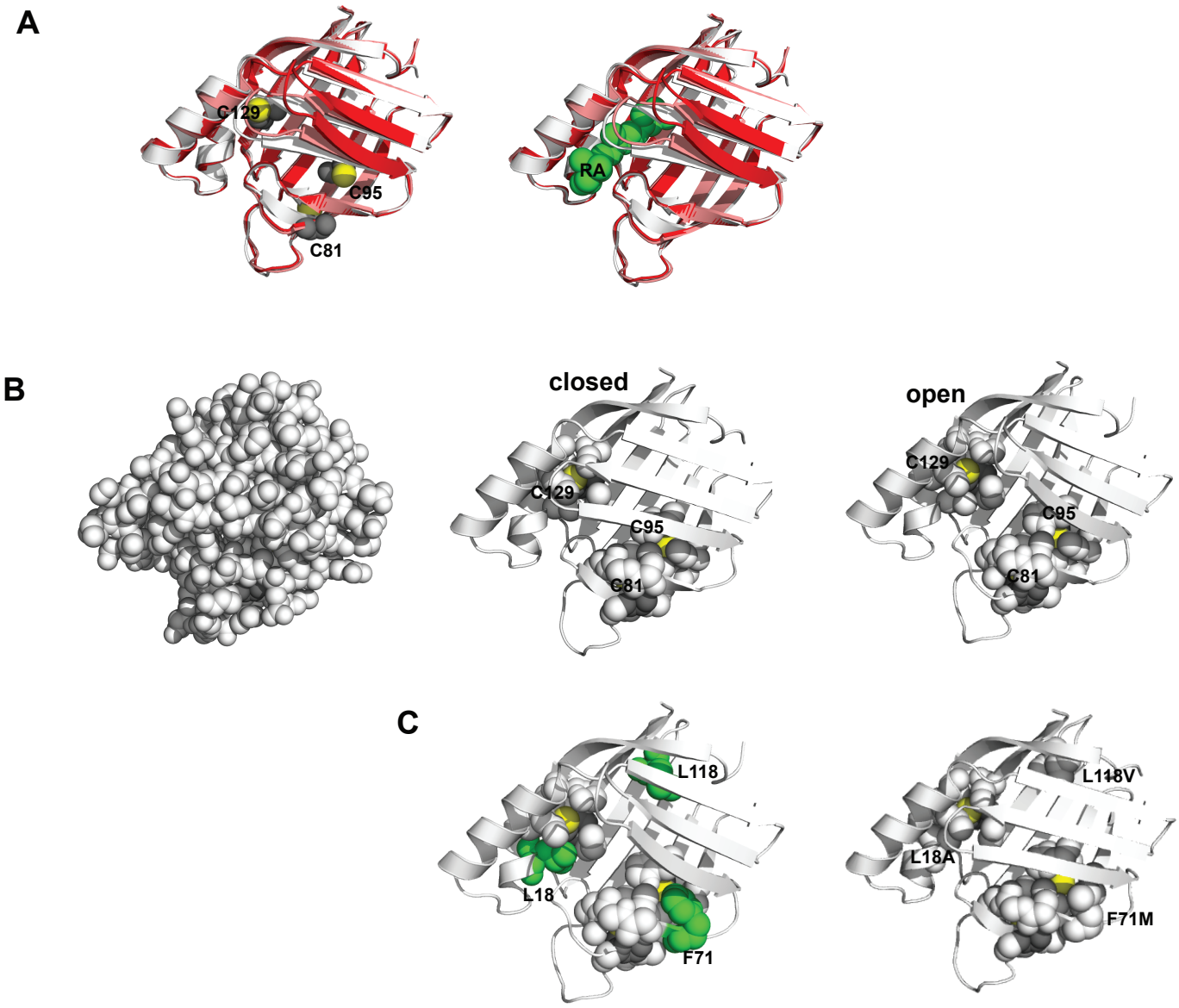
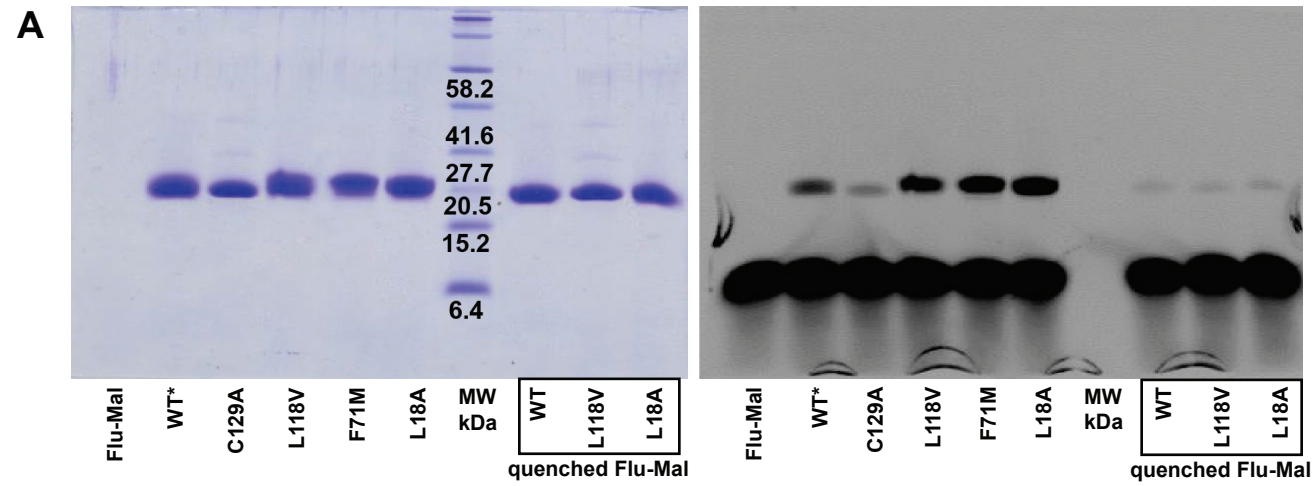
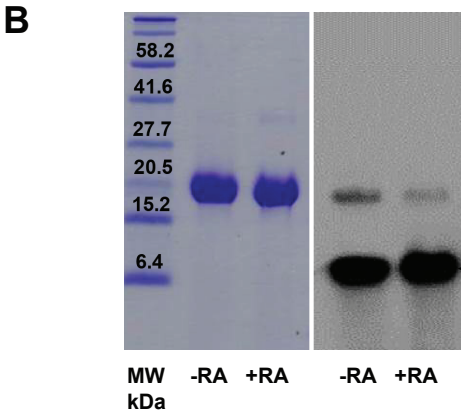

A
C129
C95
C81
RA
B
closed
C129
C95
C81
open
C129
C95
C81
C
L118
L18
F71
L118V
L18A
F71M

Fig. S3


A
58.2
41.6
27.7
20.5
15.2
6.4
Flu-Mal
WT*
C129A
L118V
F71M
L18A
MW kDa
WT
L118V
L18A
quenched Flu-Mal
Flu-Mal
WT*
C129A
L118V
F71M
L18A
MW kDa
WT
L118V
L18A
quenched Flu-Mal


B
58.2
41.6
27.7
20.5
15.2
6.4
MW kDa
-RA
+RA
-RA
+RA

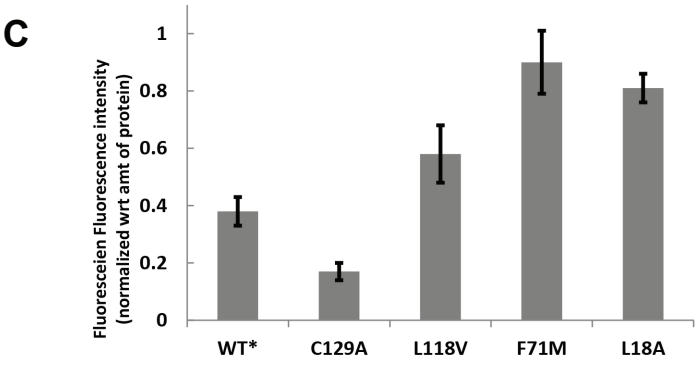

C
Fluoresceien Fluorescence intensity (normalized wrt amt of protein)
1
0.8
0.6
0.4
0.2
0
WT*
C129A
L118V
F71M
L18A

Fig. S4

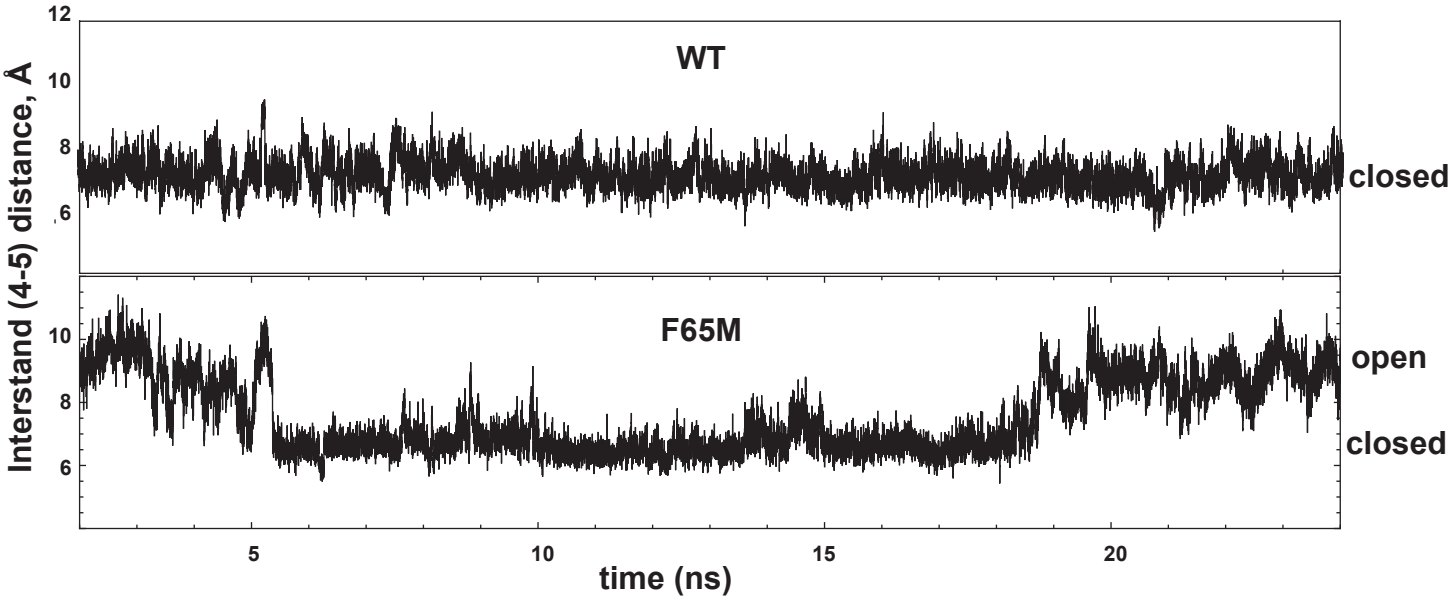

WT
F65M
closed
open
closed
Interstand (4-5) distance, Å
time (ns)
12
10
8
6
10
8
6
5
10
15
20

Fig. S5