

1 **Supplementary file:**

2 **UTag, a cysteine-free thermostable tagging system for tracking single mRNA translation**  
3 **live**

4 Luis U Aguilera<sup>1,†</sup>, Suzhsuan Chen<sup>2,†</sup>, Rhiannon M. Sears<sup>1</sup>, Jake Yarbro<sup>1</sup>, Jacob DeRoo<sup>2</sup>,  
5 Hunter A. Ogg<sup>4</sup>, Brian J. Geiss<sup>2,3</sup>, Timothy J. Stasevich<sup>4</sup>, Christopher D. Snow<sup>2,\*</sup>, and Ning  
6 Zhao<sup>1,\*</sup>

7

8

9 Supplementary Figures S1-S9

10 Supplementary Tables S1-S9

11 Supplementary Movies S1-S4

12 **Supplementary Movie S1: Translation tracking using the UTag tagging system.** A  
13 representative cell expressing anti-UTag-IB-mEGFP and 24xUTag-KDM5B-24xMS2 to visualize  
14 translation. The movie is related to Figs. 7C-7E. Field of view is 66.51 x 66.51  $\mu\text{m}$ . 5 s interval.

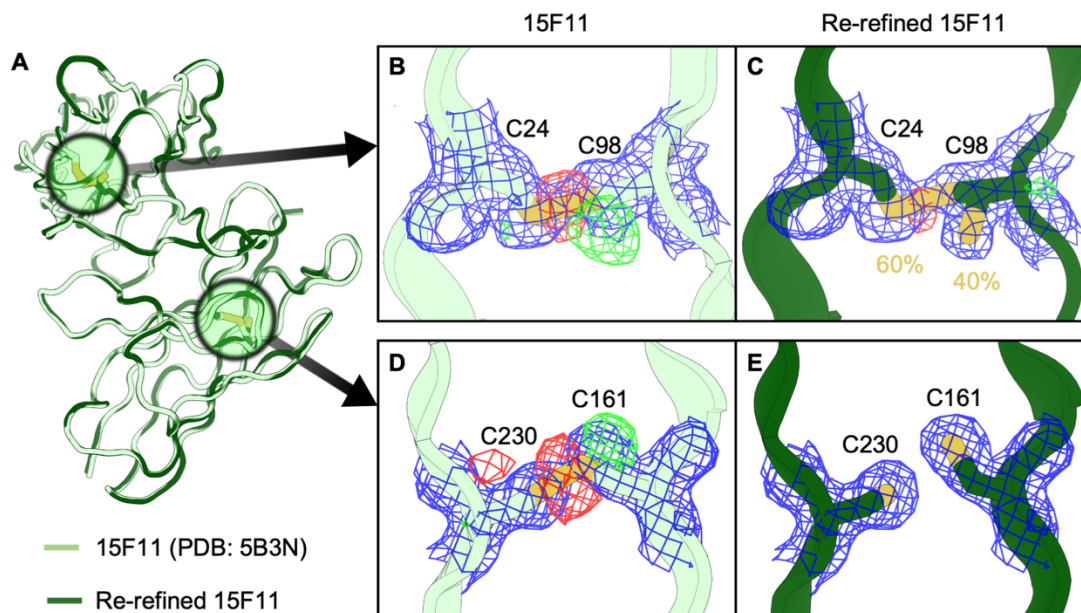
15 **Supplementary Movie S2: Translation tracking using the cysteine-free UTag tagging**  
16 **system.** A representative cell expressing anti-UTag-IB( $\Delta\text{Cys}$ )-mEGFP and 24xUTag-KDM5B-  
17 24xMS2 to visualize translation. The movie is related to Figs. 7C-7E. Field of view is 66.51 x 66.51  
18  $\mu\text{m}$ . 5 s interval.

19 **Supplementary Movie S3: Translation tracking using the SunTag system.** A representative  
20 cell expressing anti-SunTag-IB-mEGFP and 24xSunTag-KDM5B-24xMS2 to visualize translation.  
21 The movie is related to Figs. 7C-7E. Field of view is 57.67 x 50.53  $\mu\text{m}$ . 5 s interval.

22 **Supplementary Movie S4: Translation tracking using the ALFA-tag tagging system.** A  
23 representative cell expressing anti-ALFA-IB-mEGFP and 24xALFA-KDM5B-24xMS2 to visualize  
24 translation. The movie is related to Figs. 7C-7E. Field of view is 66.51 x 66.51  $\mu\text{m}$ . 5 s interval.

25

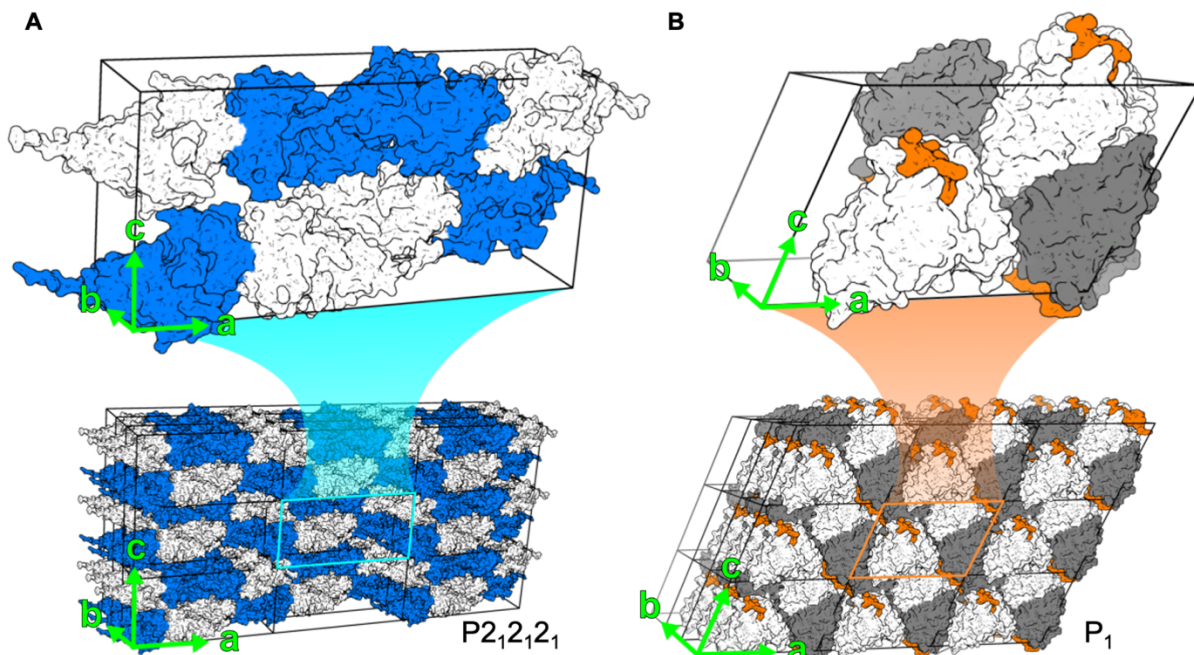
26



27

28 **Figure S1. Structural comparison of 15F11 scFv (PDB: 5B3N) before and after re-refinement. (A)**  
 29 **Structural superposition of the original RCSB PDB: 5B3N entry (light green) and the re-refined model**  
 30 **(dark green). (B–C) Electron density and difference Fourier maps centered on the first disulfide bond**  
 31 **(C24-C98) for the original (B) and re-refined (C) structures. (D–E) Close-up views of the second potential**  
 32 **disulfide bond for the original (D) and re-refined (E) models. Electron density is shown as a  $2F_o-F_c$  map**  
 33 **(blue mesh, contoured at  $1.0 \sigma$ ), while model discrepancies are highlighted by the  $F_o-F_c$  difference map**  
 34 **(contoured at  $\pm 3.0 \sigma$ ). Green density indicates positive  $F_o-F_c$  peaks where additional atomic model may**  
 35 **be warranted. Red density indicates negative  $F_o-F_c$  peaks where the model over-represents the**  
 36 **experimental density.**

37



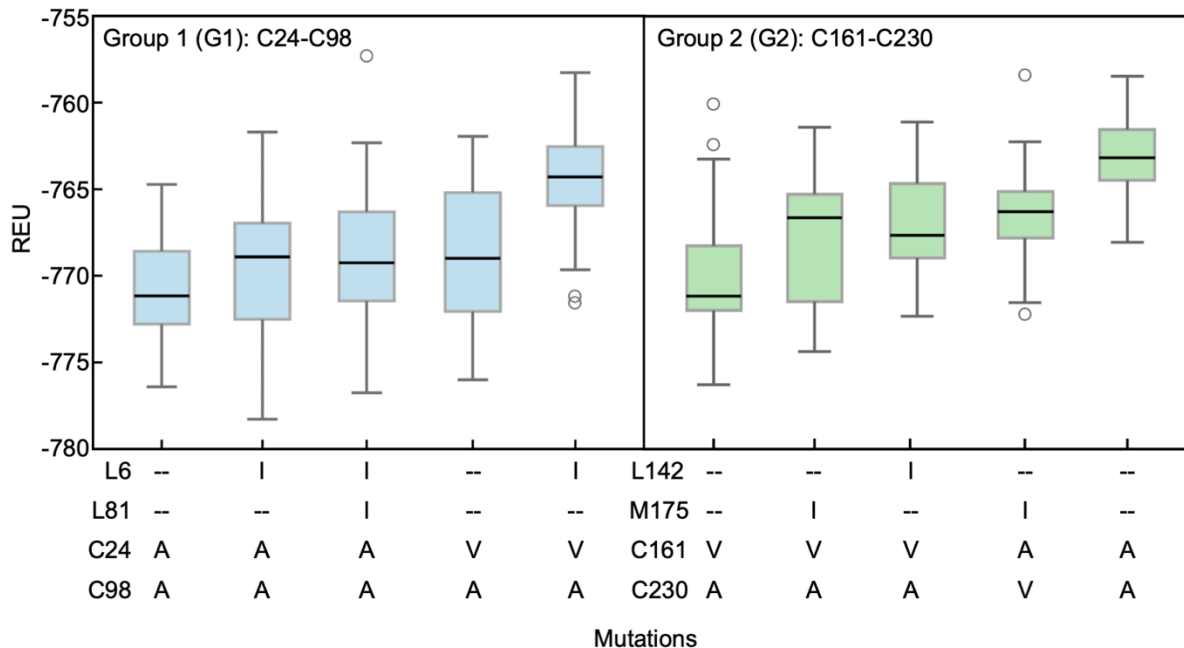
38

39 **Figure S2. Supercell structures of anti-UTag intrabody with and without UTag. (A)** Anti-UTag-IB  
 40 (PDB: 7URL) arranged in a 3 x 3 x 3 unit-cell supercell containing 216 protein chains. Each unit cell  
 41 contains four copies of two asymmetric molecules with space group P<sub>2</sub><sub>1</sub>2<sub>1</sub>2<sub>1</sub>. The two asymmetric copies  
 42 of protein chains are shown in blue and white. **(B)** Anti-UTag-IB/UTag complex (PDB: 8V8F) arranged in  
 43 a 3 x 3 x 3 unit-cell supercell containing 108 protein chains and 108 peptide chains. Each P<sub>1</sub> unit cell  
 44 contains four copies of intrabody-peptide complexes. Protein chains are shown in gray and white. UTag  
 45 peptide is in orange. Both panels show the crystal packing with crystallographic axes labeled (a, b, c in  
 46 green).

47

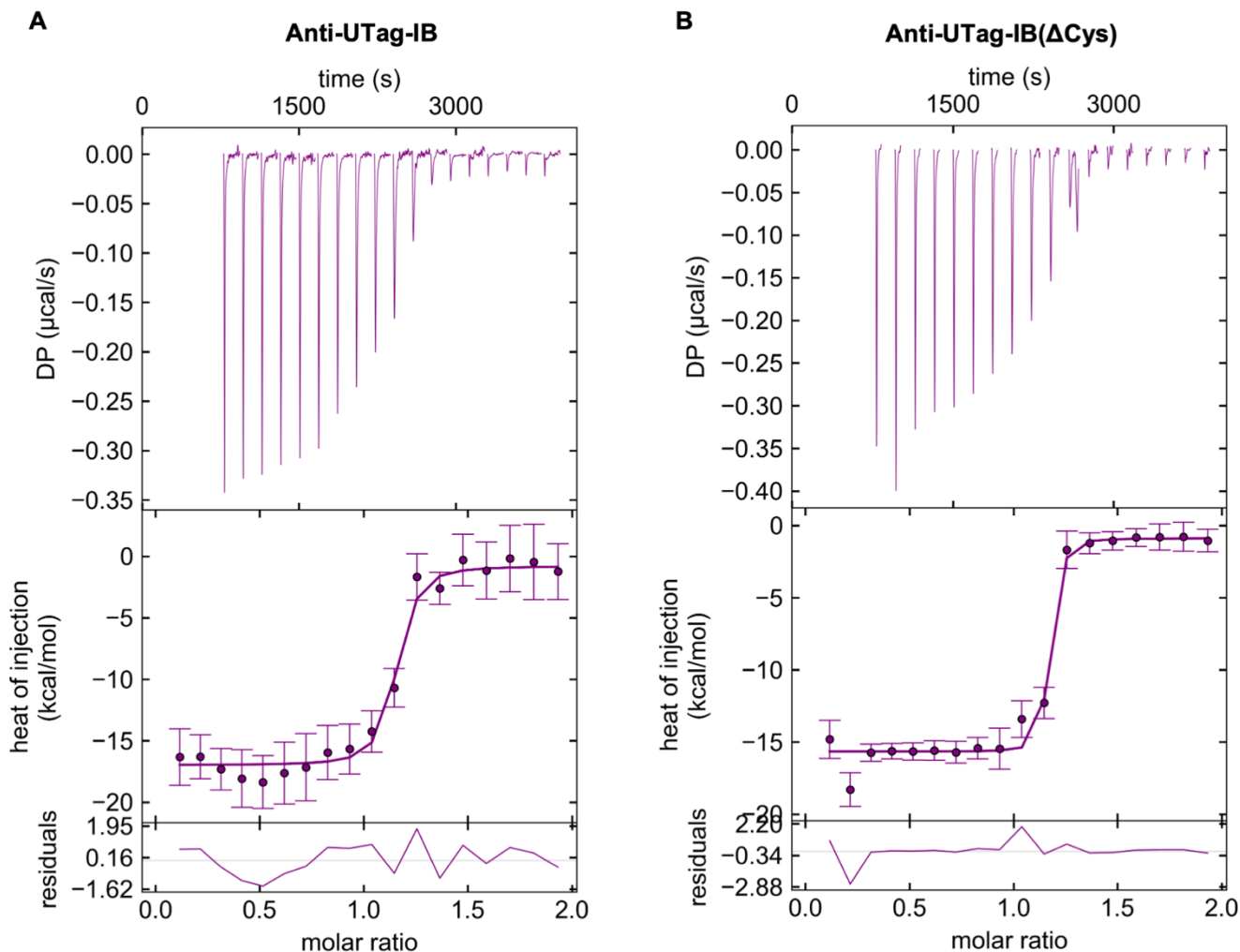
48

49



50

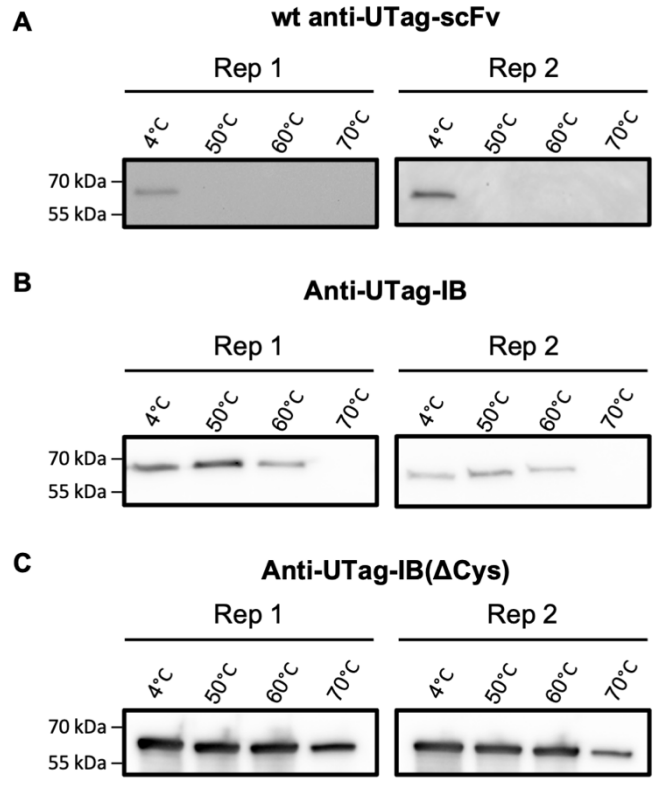
51 **Figure S3. Computational stability analysis of cysteine-free intrabody variants.** Rosetta Energy  
 52 Units (REU) for the top five candidates in **(A)** Group 1 (C24–C98) and **(B)** Group 2 (C161–C230). The  
 53 cysteine residues are numbered according to the sequence of 15F11 (PDB: 5B3N). Data represent 60  
 54 independent FastRelax replicates per design. Boxes indicate the interquartile range (IQR), and the  
 55 central line represents the median. The maximum and minimum values observed within 1.5 x IQR are  
 56 shown as whiskers (error bars), with circles indicating statistical outliers defined as points falling outside  
 57 this range.



58

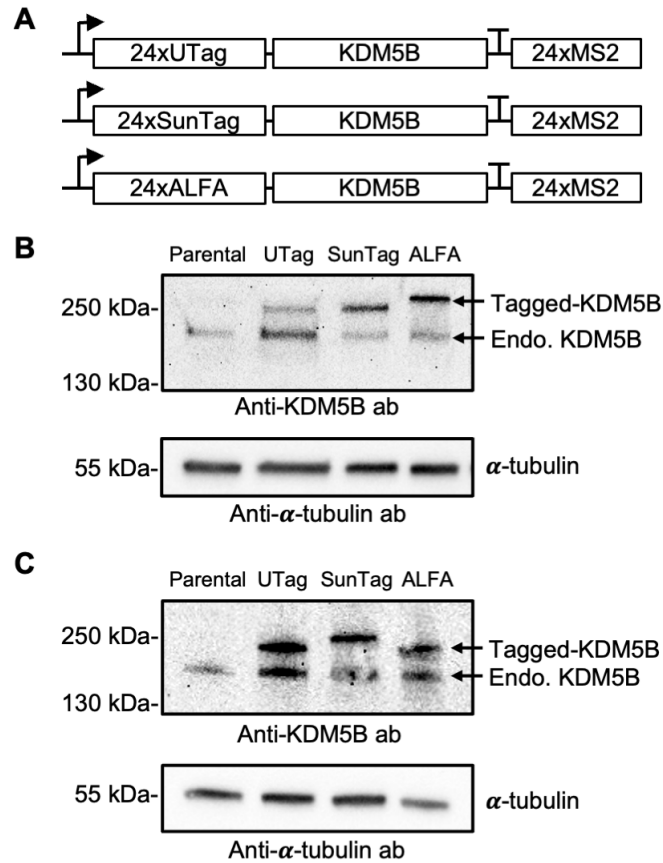
59 **Figure S4. Anti-UTag-IB and anti-UTag-IB( $\Delta\text{Cys}$ ) bind to UTag with low nM binding affinity. (A)**  
 60 **Representative isothermal titration calorimetry (ITC) experiment of anti-UTag-IB. (B) Representative ITC**  
 61 **experiment of anti-UTag-IB( $\Delta\text{Cys}$ ).** For each intrabody, the top panel shows the raw thermogram, the  
 62 middle panel shows the integrated injection heats (kcal/mol) versus molar ratio fitted to a single-site  
 63 sigmoid binding model (solid line), and the bottom panel shows the residuals of the fit. For each intrabody,  
 64 the ITC experiment was repeated three times.

65



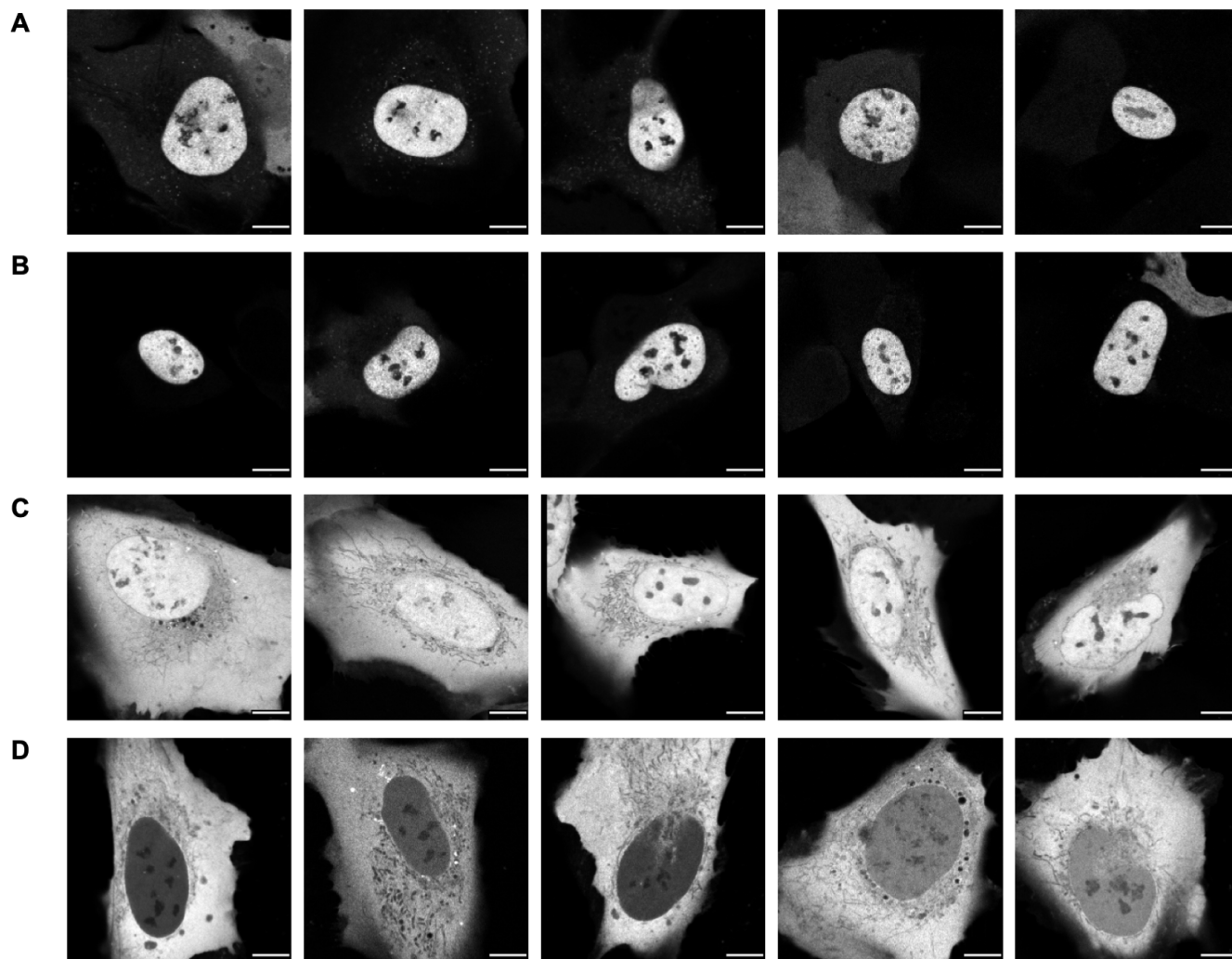
66

67 **Figure S5. Western blots of characterizing the thermostability of anti-UTag intrabodies.** Two  
 68 independent western blots of soluble fractions after incubation for 15 min at 4 °C, 50 °C, 60 °C, and 70  
 69 °C for **(A)** wild-type anti-UTag-scFv, **(B)** anti-UTag-IB, and **(C)** anti-UTag-IB( $\Delta$ Cys).



70

71 **Figure S6. Validation of full-length KDM5B expression with different tagging systems. (A)**  
 72 Schematic illustration of reporter constructs containing 24xUTag-, 24xSunTag-, or 24xALFA-tagged  
 73 KDM5B. **(B)** Western blot analysis of the reporter constructs shown in (A). **(C)** Independent replicate of  
 74 the western blot analysis shown in (B).  
 75



76

77

78 **Figure S7. Cellular localization of tagged-KDM5B visualized by cognate fluorescent intrabodies.**

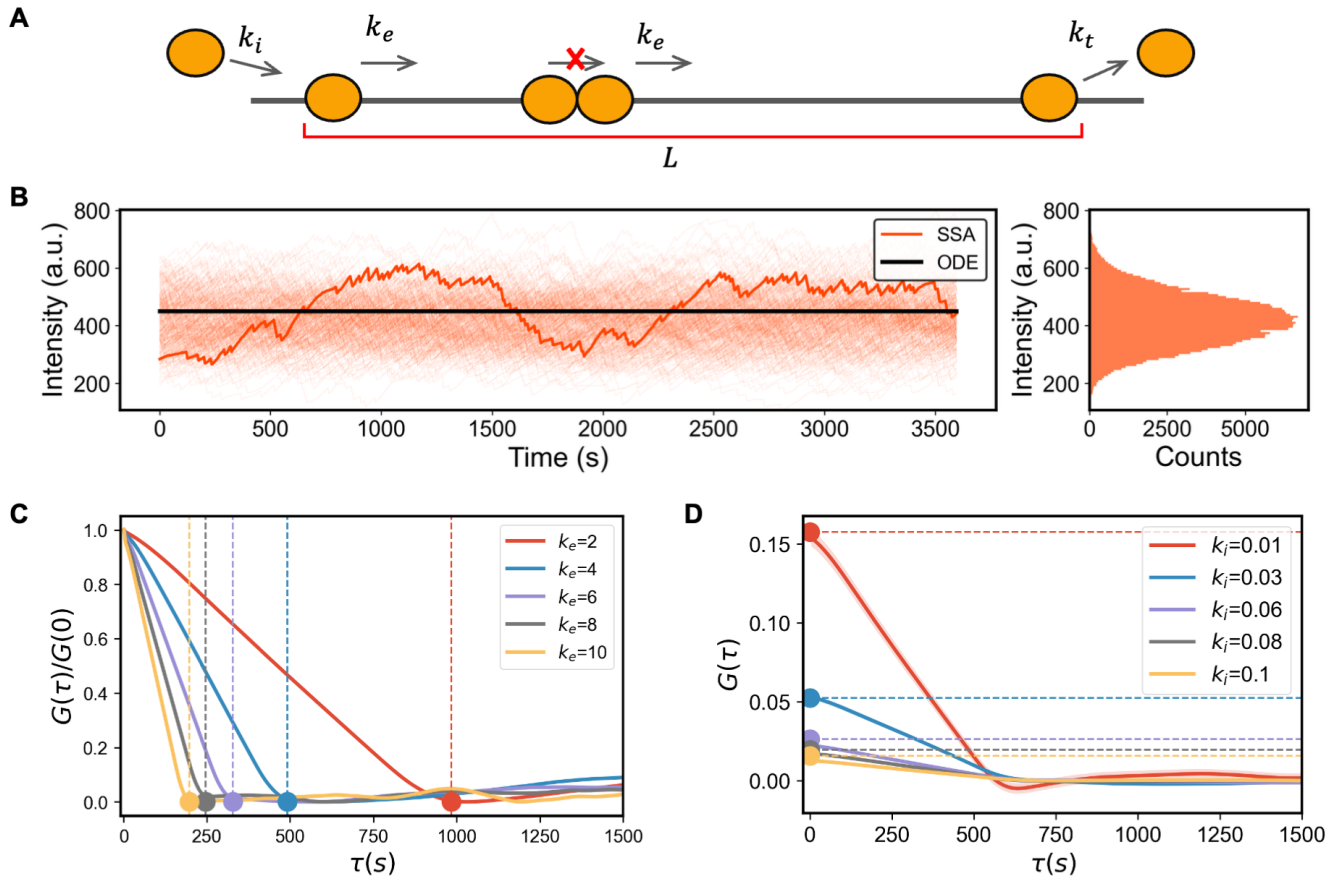
79 Representative cell images showing different tagged KDM5B cellular localization 24 h post-transfection.

80 **(A)** Anti-UTag-IB. **(B)** Anti-UTag-IB( $\Delta$ Cys). **(C)** Anti-SunTag-IB. **(D)** Anti-ALFA-IB. Scale bars: 10  $\mu$ m.

81

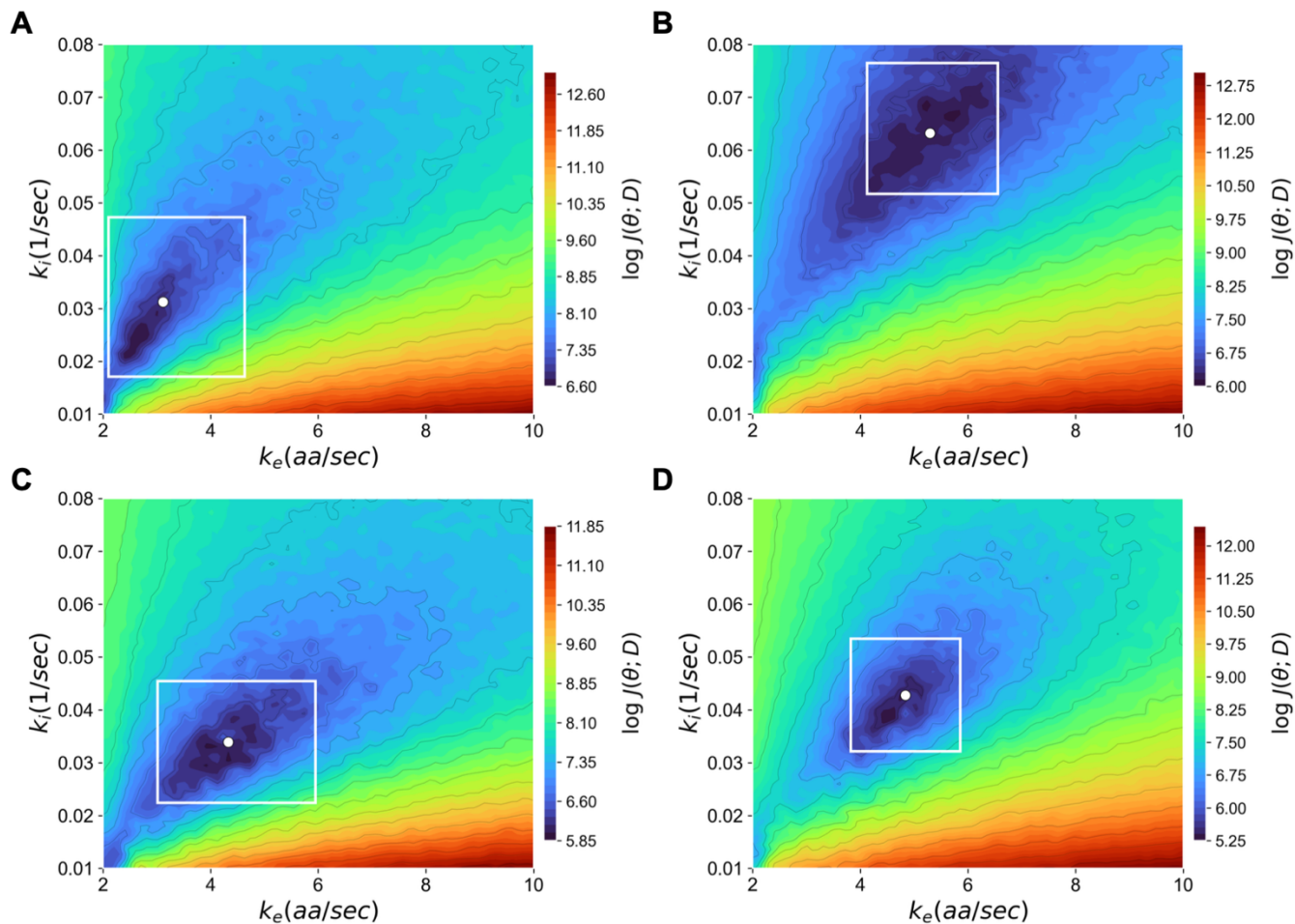
82

83



84

85 **Figure S8. TASEP modeling simulates single-mRNA translation.** (A) Schematic illustrating key steps  
 86 and main factors affecting translation included in the TASEP model, including initiation ( $k_i$ ), elongation  
 87 ( $k_e$ ), gene length ( $L$ ), ribosome exclusion ( $n_f$ ), and termination ( $k_t$ ). (B) The TASEP model was solved  
 88 using 300 repetitions in the Gillespie Algorithm, using representative parameters obtained from the  
 89 parameter optimization process ( $k_e = 3.1$  aa/s and  $k_i = 0.031$  s $^{-1}$ ). (C) Normalized autocorrelation function  
 90 (ACF) calculation from stochastic simulations using a parameter sweep for  $k_e \in [2, 10]$  aa/s (solid lines).  
 91 (D) ACF calculation from stochastic simulations using a parameter sweep for  $k_i \in [0.01, 0.1]$  s $^{-1}$  (solid  
 92 lines). (C, D) Dashed lines show the corresponding approximated analytical solutions derived from the  
 93 closed-form expressions for  $k_i = 1/(G(0)\tau)$  and  $k_e = L/\tau$ , where  $\tau$  represents the dwell time.  
 94  
 95



96

97 **Figure S9. Parameter optimization strategy using an 80 x 80 grid.** For a better visualization, the color  
 98 map is plotted using the log of the objective function. The parameter estimates (white dot) represent the  
 99 median  $k_i$  and  $k_e$  values from the top 5% of grid search solutions ranked by cost. The rectangle delineates  
 100 the 90% confidence interval, defined by the 5th and 95th percentiles of the parameter distributions within  
 101 the top 5% of this population. **(A)** Anti-UTag-IB-mEGFP. **(B)** Anti-UTag-IB( $\Delta$ Cys)-mEGFP. **(C)** Anti-  
 102 SunTag-IB-mEGFP. **(D)** Anti-ALFA-IB-mEGFP.

103

104

105

106

107

108

109

110

111

112

113

**Table S1. Structure wwPDB validation metrics comparison of PDB 5B3N vs re-refined 5B3N.**

Metric	PDB 5B3N	Re-Refined
$R_{\text{free}}$	0.253	0.234
Clash score	3	1
Ramachandran outliers	0	0
Sidechain outliers	2.1%	0
RSRZ outliers	3.0%	1.3%

114

115

116

117

**Table S2.  $\text{RMSD}_{\text{C}\alpha}$  comparing the local structural environments of mutated residues between parental and cysteine-free intrabody structures.**

Mutation	Before	After	$\text{RMSD}_{\text{C}\alpha}$ (Å)
L4I	VK <b>L</b> VE	VK <b>I</b> VE	0.05
C22A	LS <b>C</b> AA	LS <b>A</b> AA	0.08
C96A	YY <b>C</b> AR	YY <b>A</b> AR	0.13
C163V	IS <b>C</b> RA	IS <b>V</b> RA	0.13
C232A	YY <b>C</b> QQ	YY <b>A</b> QQ	0.14

118

119

120  
121

**Table S3. Phenix.Table\_one for anti-UTag intrabody (PDB: 7URL).**  
Statistics for the highest-resolution shell are shown in parentheses.

<b>Wavelength</b>	1.1
<b>Resolution range</b>	37.03 - 1.411 (1.462 - 1.411)
<b>Space group</b>	P 21 21 21
<b>Unit cell</b>	41.725 76.12 160.628 90 90 90
<b>Total reflections</b>	176609 (2042)
<b>Unique reflections</b>	88732 (1114)
<b>Multiplicity</b>	2.0 (1.8)
<b>Completeness (%)</b>	89.33 (11.37)
<b>Mean I/sigma(I)</b>	8.56 (0.91)
<b>Wilson B-factor</b>	16.78
<b>R-merge</b>	0.05586 (0.5636)
<b>R-meas</b>	0.079 (0.7971)
<b>R-pim</b>	0.05586 (0.5636)
<b>CC1/2</b>	0.998 (0.369)
<b>CC*</b>	0.999 (0.734)
<b>Reflections used in refinement</b>	88713 (1114)
<b>Reflections used for R-free</b>	2000 (25)
<b>R-work</b>	0.1836 (0.3413)
<b>R-free</b>	0.2225 (0.2710)
<b>CC(work)</b>	0.974 (0.577)
<b>CC(free)</b>	0.967 (0.638)
<b>Number of non-hydrogen atoms</b>	6077
<b>  macromolecules</b>	5661
<b>  ligands / solvent</b>	24 / 392
<b>Protein residues</b>	488
<b>RMS(bonds) / RMS(angles)</b>	0.018 / 1.75
<b>Ramachandran favored (%)</b>	96.65
<b>Ramachandran allowed (%)</b>	3.09
<b>Ramachandran outliers (%)</b>	0.26
<b>Rotamer outliers (%)</b>	5.05
<b>Clashscore</b>	5.38
<b>Average B-factor</b>	27.92
<b>  macromolecules</b>	27.16
<b>  ligands</b>	70.79
<b>  solvent</b>	36.20
<b>Number of TLS groups</b>	14

122  
123

124  
125

**Table S4. Phenix.Table\_one for cysteine-free anti-UTag intrabody (PDB: 9N97).**

Statistics for the highest-resolution shell are shown in parentheses.

<b>Wavelength</b>	0.99999
<b>Resolution range</b>	42.9 - 1.45 (1.49 - 1.45)
<b>Space group</b>	P 21 21 21
<b>Unit cell</b>	40.679 74.521 157.394 90 90 90
<b>Total reflections</b>	167334 (8893)
<b>Unique reflections</b>	84215 (4729)
<b>Multiplicity</b>	2.0 (1.9)
<b>Completeness (%)</b>	98.05 (78.06)
<b>Mean I/sigma(I)</b>	11.24 (0.78)
<b>Wilson B-factor</b>	16.57
<b>R-merge</b>	0.03664 (0.8496)
<b>R-meas</b>	0.05182 (1.201)
<b>R-pim</b>	0.03664 (0.8496)
<b>CC1/2</b>	0.999 (0.394)
<b>CC*</b>	1 (0.752)
<b>Reflections used in refinement</b>	84182 (4718)
<b>Reflections used for R-free</b>	2000 (113)
<b>R-work</b>	0.1712 (0.3428)
<b>R-free</b>	0.2120 (0.3542)
<b>CC(work)</b>	0.959 (0.673)
<b>CC(free)</b>	0.943 (0.536)
<b>Number of non-hydrogen atoms</b>	4539
<b>macromolecules</b>	3927
<b>ligands / solvent</b>	25 / 587
<b>Protein residues</b>	512
<b>RMS(bonds) / RMS(angles)</b>	0.006 / 0.91
<b>Ramachandran favored (%)</b>	98.21
<b>Ramachandran allowed (%)</b>	1.79
<b>Ramachandran outliers (%)</b>	0.00
<b>Rotamer outliers (%)</b>	0.00
<b>Clashscore</b>	2.84
<b>Average B-factor</b>	25.90
<b>macromolecules</b>	23.89
<b>ligands</b>	102.40
<b>solvent</b>	36.09
<b>Number of TLS groups</b>	30

126

127

128  
129

**Table S5. Phenix.Table\_one for co-crystal of anti-UTag intrabody and UTag (PDB: 8V8F).**  
Statistics for the highest-resolution shell are shown in parentheses.

<b>Wavelength</b>	1.0
<b>Resolution range</b>	45.71 - 2.27 (2.351 - 2.27)
<b>Space group</b>	P 1
<b>Unit cell</b>	54.467 61.824 85.5767 68.8341 77.8985 71.375
<b>Total reflections</b>	78817 (7454)
<b>Unique reflections</b>	42832 (3973)
<b>Multiplicity</b>	1.8 (1.9)
<b>Completeness (%)</b>	94.23 (87.85)
<b>Mean I/sigma(I)</b>	4.77 (0.78)
<b>Wilson B-factor</b>	33.85
<b>R-merge</b>	0.2431 (1.059)
<b>R-meas</b>	0.3438 (1.498)
<b>R-pim</b>	0.2431 (1.059)
<b>CC1/2</b>	0.0186 (0.304)
<b>CC*</b>	0.191 (0.682)
<b>Reflections used in refinement</b>	42688 (3968)
<b>Reflections used for R-free</b>	1998 (186)
<b>R-work</b>	0.2284 (0.3273)
<b>R-free</b>	0.2755 (0.3759)
<b>CC(work)</b>	0.944 (0.627)
<b>CC(free)</b>	0.906 (0.641)
<b>Number of non-hydrogen atoms</b>	7868
<b>macromolecules</b>	7647
<b>ligands / solvent</b>	45 / 176
<b>Protein residues</b>	998
<b>RMS(bonds) / RMS(angles)</b>	0.002 / 0.58
<b>Ramachandran favored (%)</b>	96.71
<b>Ramachandran allowed (%)</b>	3.29
<b>Ramachandran outliers (%)</b>	0.00
<b>Rotamer outliers (%)</b>	0.62
<b>Clashscore</b>	3.66
<b>Average B-factor</b>	40.56
<b>macromolecules</b>	40.56
<b>ligands</b>	57.60
<b>solvent</b>	36.19
<b>Number of TLS groups</b>	27

130  
131

132

**Table S6. Binding affinity characterization of each tagging system using FRAP.**

Tagging system	Recovered intensity (Mean $\pm$ SEM)	Half-recovery time ( $t_{1/2}$ , sec)
Anti-UTag-IB	84% $\pm$ 1%	177 $\pm$ 7
Anti-UTag-IB( $\Delta$ Cys)	78% $\pm$ 1%	194 $\pm$ 6
Anti-HA-IB	90% $\pm$ 1%	39 $\pm$ 4
Anti-SunTag-IB	55% $\pm$ 2%	N.A.
Anti-ALFA-IB	39% $\pm$ 2%	N.A.

133

134

135

**Table S7. Translation kinetics determined with different tagging systems.**

Tagging system	$k_i$ ( $s^{-1}$ )	$k_e$ (aa/s)	Protein length (aa)	Rib. density (%)	No. ribosomes	Rib. distance (aa)
Anti-UTag-IB	0.031 [0.017, 0.047]	3.11 [2.10, 4.63]	1968	9.0	19.8	99.6
Anti-UTag-IB( $\Delta$ Cys)	0.063 [0.052, 0.076]	5.29 [4.13, 6.56]	1968	10.7	23.5	83.8
Anti-SunTag-IB	0.034 [0.022, 0.045]	4.33 [3.01, 5.94]	2133	7.1	16.7	127.6
Anti-ALFA-IB	0.043 [0.032, 0.053]	4.84 [3.82, 5.85]	2016	8.0	17.8	113.0

136

$k_i$  and  $k_e$ : Parameters calculated with top 5% ranked solutions (shown in brackets).

137

Protein length: Total length of the reporter protein plus the 24xtag array.

138

Ribosome density: Percentage of codons occupied by ribosomes at steady state.

139

Number of ribosomes: Average number of ribosomes simultaneously translating each transcript.

140

Ribosome distance: Average distance between consecutive ribosomes along the mRNA.

141

142

143

**Table S9. Statistical summary and bootstrap analysis of Rosetta Energy Units (REU) for cysteine-free intrabody variants.**

Name	Mutations	REU
G1_mutant_000	L6L-L81L-C24A-C98A	-770.9 $\pm$ 0.7
G1_mutant_002	L6L-L81L-C24V-C98A	-768.4 $\pm$ 1.0
<b>G1_mutant_009</b>	<b>L6I-L81L-C24A-C98A</b>	<b>-769.3 <math>\pm</math> 0.8</b>
G1_mutant_011	L6I-L81L-C24V-C98A	-764.5 $\pm$ 0.7
G1_mutant_012	L6I-L81L-C24A-C98A	-768.8 $\pm$ 0.9
G2_mutant_004	L142L-M175I-C161A-C230V	-766.4 $\pm$ 0.6
G2_mutant_005	L142L-M175I-C161V-C230A	-767.8 $\pm$ 0.9
G2_mutant_006	L142L-M175M-C161A-C230A	-763.0 $\pm$ 0.5
<b>G2_mutant_008</b>	<b>L142L-M175M-C161V-C230A</b>	<b>-770.1 <math>\pm</math> 0.7</b>
G2_mutant_017	L142I-M175M-C161V-C230A	-766.9 $\pm$ 0.7

144

Energetic stability was evaluated for the top five variants within each mutation group. The mean and standard deviation ( $\pm$  SD) were calculated from a bootstrap analysis (1,000 iterations), where 20 samples out of 60 total were drawn with replacement in each iteration. Final designs selected for experimental testing based on REU scores and structural intuition are indicated in bold.

145

146

147

148