

Supplementary Information

Programmable Protein Stabilization with Language Model-Derived Peptide Guides

Lauren Hong,^{1,*} Tianzheng Ye,^{2,*} Tian Wang,¹ Divya Srijay,¹ Lin Zhao,¹ Rio Watson,¹ Sophia Vincoff,¹ Tianlai Chen,¹ Kseniia Kholina,¹ Shrey Goel,¹ Matthew P. DeLisa,²⁻⁴ Pranam Chatterjee^{1,5,6,†}

1. Department of Biomedical Engineering, Duke University
2. Robert F. Smith School of Chemical and Biomolecular Engineering, Cornell University, Ithaca, NY, USA
3. Meinig School of Biomedical Engineering, Cornell University, Ithaca, NY, USA
4. Cornell Institute of Biotechnology, Cornell University, Ithaca, NY, USA
5. Department of Computer Science, Duke University
6. Department of Biostatistics and Bioinformatics, Duke University

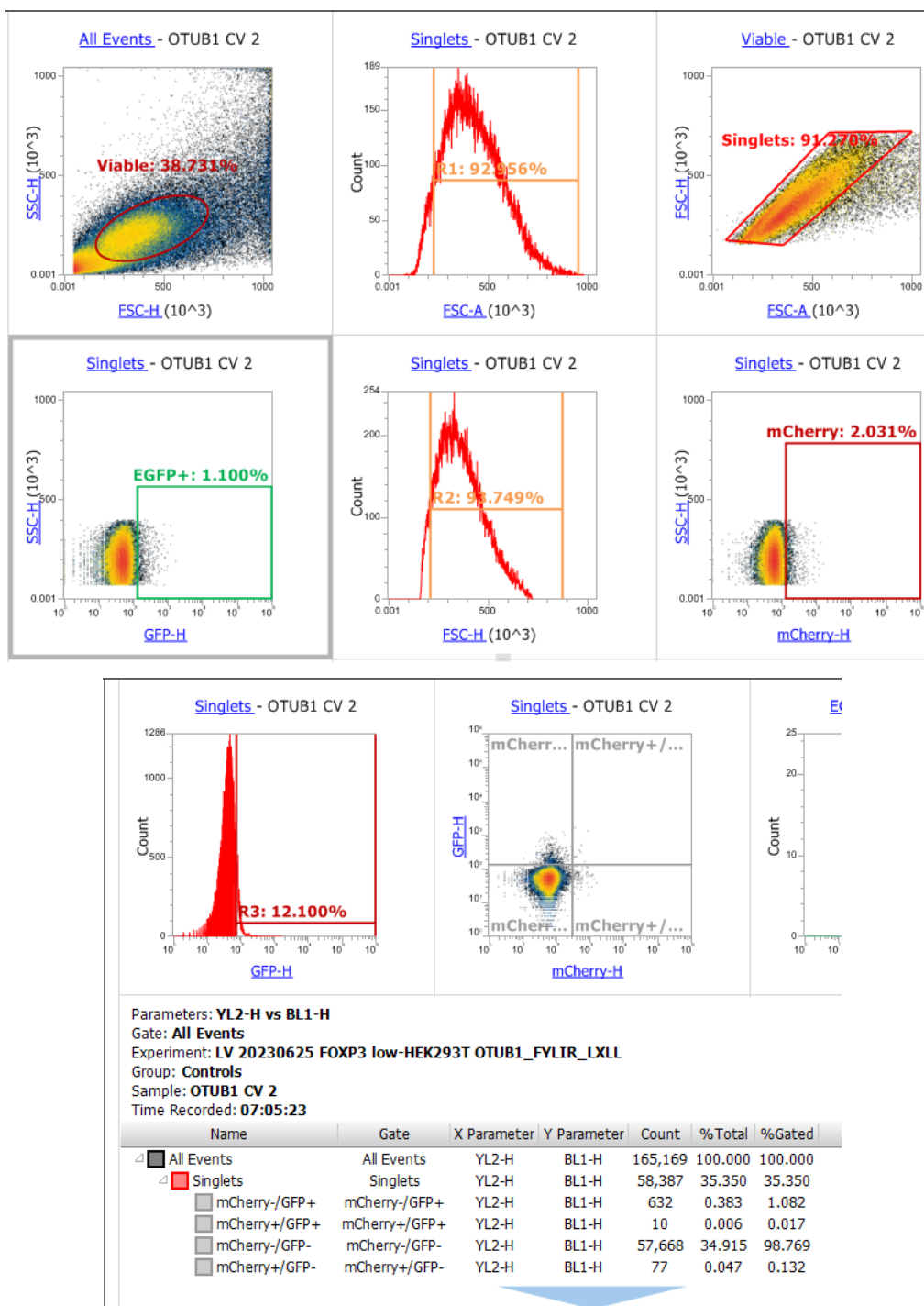
†Corresponding author: pranam.chatterjee@duke.edu

Supplementary Figures

1. Example gating strategy for flow cytometry analysis.

Supplementary Tables

1. Linker and DUB catalytic domain sequences used in study.
2. Binder sequences used in study.



Supplementary Figure 1. Example gating strategy for flow cytometry. ~10,000 gated events for data analysis based on the default FSC/SSC parameters for HEK293T cells. The GFP+ and mCherry+ gates were established both by a GFP- negative control and an mCherry- control. All analysis was conducted in FlowJo.

Supplementary Table 1. Linker and DUB catalytic domain sequences used in study. Sequences are provided with their type and UniProt ID, if applicable. duAb = Binder + Linker + OTUB1 catalytic domain.

Label/Name	Type	Amino Acid Sequence	UniProt ID
L1	Linker	GAPGSG	N/A
L2	Linker	GSGSG	N/A
OTUB1	DUB Catalytic Domain	SYIRKTRPDGNCFYRAFGFSHLEALLDDSKELQRFKAVSAKSKEDL VSQGFTEFTIEDFHNTFMDLIEQVEKQTSVADLLASFNDQSTSDYLV VYLRLLTSGYLQRESKFFEHFIEGGRTVKEFCQQQEVEPMCKESDHI HIIALAQALSVSISQVEYMDRGEGGTTNPHIFPEGSEPKVYLLYRPGH YDILYK	Q96FW1
OTUD1	DUB Catalytic Domain	SDSEGVNCLAYDEAIMAQQDRIQQEIAVQNPLVSRLELSVLYKEYA EDDNIYQQKIKDLHKKYSYIRKTRPDGNCFYRAFGFSHLEALLDDSK ELQRFKAVSAKSKEDLVSQGFTEFTIEDFHNTFMDLIEQVEKQTSV ADLLASFNDQSTSDYLVVYLRLLTSGYLQRESKFFEHFIEGGRTVKE FCQQQEVEPMCKESDHIHIIALAQALSVSISQVEYMDRGEGGTTNPHIF PEGSEPKVYLLYRPGHYDILYK	Q5VV17
UCHL1	DUB Catalytic Domain	MQLKPMEINPEMLNKVLSRLGVAGQWRFDVLGLEEESLGSVPAP ACALLLLFPLTAQHENFRKKQIEELKGQEVSPKVYFMKQTIGNSCG TIGLIHAVANNQDKLGFEDGSVLKQFLSETEKMSPEDRAKCFEKNE AIQAAHDAVAQEGQCRVDDKVNHFILFNNVDGHLIELDGRMPFP VNHGASSEDTLKDAAKVCREFTEREQGEVRFSVALCKAA	P09936
UBC9	SUMOlase Catalytic Domain	MSGIALSRLAQERKAWRKDHPFGFVAVPTKNPDGTMNLMNWECAL PGKKGTPWEGGLFKLRMLFKDDYPSSPPKCKFEPPLFHPNVYPSG TVCLSILEEDKDWRPATIKQILLGIQELLNEPNIQDPAQAEAYTIYCQ NRVEYEKRVRAQAKKFAPS	P63279

Supplementary Table 2. Binder sequences used in this study. UniProt IDs are provided to enable location of target amino acid sequence, as are the derivation method and sequence information. duAb = Binder + Linker + OTUB1 catalytic domain.

Binder Name	Target UniProt ID	Target	Binder Sequence	Derivation
YFP Nb	N/A	YFP	MQVQLVESGGALVQPGGSLRLSCAA SGFPVNRYSMRWYRQAPGKEREW VAGMSSAGDRSSYEDSVKGRFTISR DDARNTVYLQMNSLPEDTAVYYCN VNVGFEYWGGQTQVTVSS	Kanner, et al. ¹¹
β cat_SnP_7	P35222	β -catenin	DPTAPPYDSLLVFDYEGS	Brixi, et al. ¹⁰
β cat_SnP_8	P35222	β -catenin	PTAPPYDSLLVFDYEG	Brixi, et al. ¹⁰
P60D2A	Q9BZS1	FOXP3	RAFQSFRKMWPFFAM	Lozano, et al. ²⁰
FOXP3_SnP_1	Q9BZS1	FOXP3	YVINGKRK	SaLT&PepPr
FOXP3_SnP_2	Q9BZS1	FOXP3	NFYVINGKRK	SaLT&PepPr
FOXP3_SnP_3	Q9BZS1	FOXP3	KVNFYVINGKRK	SaLT&PepPr
FOXP3_SnP_4	Q9BZS1	FOXP3	VKVNFYVING	SaLT&PepPr
FOXP3_SnP_5	Q9BZS1	FOXP3	TPVKVNFY	SaLT&PepPr
FOXP3_SnP_6	Q9BZS1	FOXP3	FYVINGKRKRSQ	SaLT&PepPr
FOXP3_SnP_7	Q9BZS1	FOXP3	NGKRKRSQ	SaLT&PepPr
FOXP3_SnP_8	Q9BZS1	FOXP3	RSQPQHFTYH	SaLT&PepPr
WEE1_PpC_1	P30291	WEE1	VMVVYPPGSYLTASGSN	PepPrCLIP
WEE1_PpC_2	P30291	WEE1	SGGAGGYGFCTFSGSAL	PepPrCLIP
WEE1_PpC_3	P30291	WEE1	APMYRGYMRDSPGLPVK	PepPrCLIP
WEE1_PpC_4	P30291	WEE1	EMPGCKTHKGELVWETFNG	PepPrCLIP
WEE1_PpC_5	P30291	WEE1	YPGGGSWGQPHGGSWGQK	PepPrCLIP
WEE1_PpC_6	P30291	WEE1	RMPDDLHYHTGELVQCTPHA	PepPrCLIP
p53_pMLM_1	P04637	p53	GRTRRLGGGKLCRGS	PepMLM
p53_pMLM_2	P04637	p53	SDTYLTGAGCCCCGR	PepMLM
p53_pMLM_3	P04637	p53	KPTRRYGGGKLRCRR	PepMLM
p53_pMLM_4	P04637	p53	GPTRRTGAGRKTCRK	PepMLM
p53_pMLM_5	P04637	p53	SPTYPLSACKKCRGR	PepMLM
p53_pMLM_6	P04637	p53	SDTYRLLGRKKRCGS	PepMLM
p53_pMLM_7	P04637	p53	EPTYLIDCCY	PepMLM
p53_pMLM_8	P04637	p53	DDTYRQGCPCT	PepMLM
P3F1_pMLM_1	N/A	PAX3::FOXO1	MVTQLSSDDYTSLVCSN	PepMLM

P3F1_pMLM_2	N/A	PAX3::FOXO1	ALTSSTSVLYTTYLLISN	PepMLM
P3F1_pMLM_3	N/A	PAX3::FOXO1	MVTQLSDDLSTYYDDLPK	PepMLM
P3F1_pMLM_4	N/A	PAX3::FOXO1	NVTSDTSDSGSTYDDCPS	PepMLM
P3F1_pMLM_5	N/A	PAX3::FOXO1	ALVSSSDVLYTYSLLISN	PepMLM
P3F1_pMLM_6	N/A	PAX3::FOXO1	NLVDLSSVLYSSYDLCPK	PepMLM
P3F1_pMLM_7	N/A	PAX3::FOXO1	NLVSLSSDLSTYYLLCKS	PepMLM
P3F1_pMLM_8	N/A	PAX3::FOXO1	ALVQDTSVLGSYYLLCSN	PepMLM
P3F1_pMLM_9	N/A	PAX3::FOXO1	NLTSLDDDLGSYKDLLKK	PepMLM
P3F1_pMLM_10	N/A	PAX3::FOXO1	NLTSLSSDSYTYKLDLPN	PepMLM