
Supplementary information

De novo design of cysteine proteases

In the format provided by the
authors and unedited

Supplementary table 1 | Sequences for designed protease and substrate

Supplementary table 2 | LC-MS data

Supplementary table 3 | Crystallography data collection and refinement statistics

Supplementary table 4 | Foldseek analysis for crystal structures

Supplementary table 5 | Kinetics summary of sequence specific cysteine proteases

Supplementary Fig. 1 | Full image for SDS-page gel.

Supplementary Fig. 2 | Gel-based kinetic analysis of Ulp1 cleavage activity.

Supplementary Fig. 3 | Raw image for cell-based reporter assay

Supplementary information 1. pCool6 plasmid DNA sequence

Supplementary information 2. ZJ101 plasmid DNA sequence

Supplementary table 1. Sequences for designed protease and substrate

design_id	design_aa_seq	substrate sequence
HC04 (Gawi)	TVKSFPATVTEYTTADGKTYEIKLKDFSLTDEELEKIYQKVAETDPENKEKITAYGKALAEAG VASKADPEKLVALAKKYGVTDRCVCLLVALLYIQADPSWSLFYTPNNHFVALKVDADGKVLVR DLNSNNTALTESSYPSREAFSLASLGIVSSKTFETTIVSLTASNGLPAIKIE	GSKIISGG/S
HC05	TVKSFPATVTEYTTADGKTYEIKLKDFSLTDEELEKIYQKVAETDPENKEKITAYGKALAEAG VASKADPEKLVALAKKYGVTDRCVCLLVALLYIQADPSWSLFYTPNNHFVALKVDADGKVLVR DLNSNQALTTESSYPSREAFSLASLGIVSSKTFETTIVSLTASNGLPAIKIE	GSKIISGG/S
HC06	TVKSFPATVTKYTTADGKEYEYIDLKDFKLTDEELAAIYDKVAETDPSRKEAITAYGKALAEA GVASKADPEKLVALAKQYGVTDRCVCLLVALLYIQADPSWKLFTYTPNNHFVAIKEDENGKVLVR DLNANNTALTERSYPSREAFSLASLGIVSSETFETTIVSLTASNGLPAIKIE	GSKIISGG/S
HC07	TVKSFPATVTKYTTADGKEYEYIDLKDFKLTDEELAAIYDKVAETDPSRKEAITAYGKALAEA GVASKADPEKLVALAKQYGVTDRCVCLLVALLYIQADPSWKLFTYTPNNHFVAIKEDENGKVLVR DLNANQALTALTERSYPSREAFSLASLGIVSSETFETTIVSLTASNGLPAIKIE	GSKIISGG/S
HC10	VVKSFPATVTKYTTADGKEYEYIDLKDFRLTDEELKAIYDKVAETDPSRADAIRAYGDALAEA KVAEKADPELLVALAKEYGVTDIVCLLVALLYIQADPSWKLFTYTPNNHFVALKVDENGKVT RDLNSNNTALTERSYPSREAFSLASLGIVSSETFETTIVSLTASNGLPAIKIE	GSKIISGG/S
HC12 (Kal)	ATQTLTVKVPAGVVEVVSFEGTEEQREKAQAIIVDEGNKTLDSVDLGKGLYKDKNGNIIVEISR EQLGKAAELAKTRLPIYIDGVAAGAPGGSCLPLSIIRAYEEYKAEELGKELGVPVSDKVLVV NFNNQHFVVQVPVGDKLYVFDVDENNNPVYYTEPLPELTYVADVNVTFYASGISGSIEAIP	SIVLTGSG/S
HC13	ATQTLTVEVAPGVSVVEVKFEGTEEQREKAQAIIDEAQNLSSVNLGKGLYKDKEGNTLVEIT VEQLKAAELAKTIKPAIDKVKALGAAGGSCLPLGVIRAVERYEKLKEEAKKKGLEVSETVFIK NINDSHFVILPVGDKWYVVDVFNEKNEPVEYFIDKPELTYVGDVDVTFWVGGISGAIEITP	SIVLTGSG/S
HC14	ATQTLTVRVAPGVSVTVEFEGTEAQRDAAQAIIAEGEKILSSVDLGKGRLYRDKENIVVEITR EQLEKAAELAKKLLPYIDAVAAAGAPGGSCLPLSLIEAVKKIEEEEEAKKKGLEVSEKVIYVNI DNSHFVVLIPVGDKAFAVVDVDANNPNVLSEVPLPELEFVGEVNITFYAGGISGAIKADL	SIVLTGSG/S
HC32	SLEERYKKRAEGVLEIYEVYKEVGKELAKKYPDAEVQLVNYNNSHFYLRNLNTSKGVVTVDDG ENNNPVVTDAGKTTYDEKAKEKLDKIAEEYDIDLSTTCLIRSLAIEAFYEVNDGKSPVSLVTEG DGVPLQEVSDANLSGKVKYKNDKSEYLELKGSWNSDGKVTLEIRNPKGEVLAKIDIKV	VPTLTGG/G
HC33	SEKAIEYKRAEGVKEIYEVYETGKKLAEYYPDATVNLVNYDNSHFYLELETPEGTIVVDGDE NNYPVVTKDGKKTVDDEEAKELKEIAEKYDIDLSTTCLIRSLAIVQAYWETRNGKAPVRLYTE GTGVPRQEVAEANLSGIYKKDDKSKYLKVTGAWRPDGVLTNLDDPEGNVLASYTIRY	VPTLTGG/G
HC34 (Dokki)	SEKAIEYKRAEGVKEIYEVYETGKKLAEYYPDATVNLVNYDNSHFYLELETPEGTIVVDGDE NQYPVVTKDGKKTVDDEEAKELKEIAEKYDIDLSTTCLIRSLAIVQAYWETRNGKAPVRLYTE GTGVPRQEVAEANLSGIYKKDDKSKYLKVTGAWRPDGVLTNLDDPEGNVLASYTIRY	VPTLTGG/G
HC37	SAKEEQKKRVEGAVEIYKVYYEAGKKLAKKYPDAKVKLVDNSHFYLELETTPKGTVKVDG DSNNNIVVTEGGKKTVDDEEAKKKLDEIAEKYNIDLSTTCLIRSLAIIKAYQETNNGKTPVELYTE GTGVPKKKIDEANLSGKIKEKNDSSKYINLKGSWNKDGVITLLELDENGEVLKKIEIKV	VPTLTGG/G
HC38	SAKEEQKKRVEGAVEIYKVYYEAGKKLAKKYPDAKVKLVDNSHFYLELETTPKGTVKVDG DSNQNVVTEGGKKTVDDEEAKKKLDEIAEKYNIDLSTTCLIRSLAIIKAYQETNNGKTPVELYTE GTGVPKKKIDEANLSGKIKEKNDSSKYINLKGSWNKDGVITLLELDENGEVLKKIEIKV	VPTLTGG/G
Gawi-09	TVKSFPATVTKYTTADGKEYEYVDLKDFKLTDEELAAIYDKVAETDPSRKEAITAYGKALAEA GVASKADPEKLVALAKQYGVTDRCVCLLVALLYIQADPSWKVFFTPNNHVAIKEDENGKVLV KDLNANNTALTERSYPSREAFSLASLGIVSSETFESTASLTASDGLPAIEVD	VPTLTGG/G
Dokki-15	GLDAIEYKLAEGVKEIYEVYETGKKLAEYYPDATVNLVNYDNSHFYLELETTPKGTIVVDGDE NQYPIVTKNGKKTEDKEAKAKLKEIAEKYDIDLSTTCLIRSLAIVQAYWETRNGKAPVKLYSEG TGVPRQELAAANLVSKIYKKGDKSKYLKVTGAWRPDGVLTNLDDPNGNVLASYSITY	VPTLTGG/G
Dokki-v2	GLDAIEYKLAEGVKEIYEVYETGKKLAEYYPDATVNLVNTDNVHFYLELETTPKGTIVVDGDE NQYVIVTKNGKKTEDKEAKAKLKEIAEKYDIDLSTFCLIRSLAIVQAYWETRNGKAPVKLYSEG TGVPRQELAAANLVSKIYKKGDKSKYLKVTGAWRPDGVLTNLDDPNGNVLASYSITY	VPTLTGG/G

Supplementary table 2. LC-MS data

Designs	Expected cleaved MW (Da)	Observed MW (Da)	Difference (Da)
hc04 (Gawi)	26913.33	26913.06	0.27
hc05	26913.33	26913.07	0.26
hc06	26913.33	26913.05	0.28
hc07	26913.33	26913.06	0.27
hc10	26913.33	26913.06	0.27
hc12 (Kal)	26970.38	26969.99	0.39
hc13	26970.38	26970	0.38
hc14	26970.38	26970.03	0.35
hc32	26970.38	26970.02	0.36
hc33	26970.38	26970.4	-0.02
hc34 (Dokki)	26970.38	26970.04	0.34
hc37	26970.38	26970.05	0.33
hc38	26970.38	26970.04	0.34
TEV	27272.67	27272.17	0.5
TEV C141A	53830.1849	53830.22	-0.0351

Supplementary table 3. Crystallography data collection and refinement statistics

	Kal WT complex PDB Code: 9YOZ	Kal C103A complex PDB Code: 9YOY	Dokki WT apo PDB Code: 9YP0
Data collection			
Space group	P 4 ₃ 2 ₁ 2	P 4 ₃ 2 ₁ 2 ₁	P 2 ₁ 2 ₁ 2 ₁
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	77.23, 77.23, 104.97	77.29, 77.29, 104.66	37.44, 57.04, 75.75
α , β , γ (°)	90, 90, 90	90, 90, 90	90, 90, 90
Resolution (Å)	104.97 - 2.19 (2.26 - 2.19)	104.66 - 1.96 (2.01 - 1.96)	37.88 - 1.77 (1.86 - 1.77)
<i>R</i> _{sym} or <i>R</i> _{merge}	0.437 (1.518)	0.105 (1.956)	0.284 (1.875)
<i>I</i> / σ <i>I</i>	11.3 (2.7)	21.1 (2.0)	9.5 (2.0)
Completeness (%)	100 (100)	100 (100)	99.9 (99.6)
Redundancy	25.9 (27.6)	25.9 (26.6)	9.9 (10.2)
Refinement			
Resolution (Å)	62.21 - 2.19 (2.33 - 2.19)	62.17 - 1.96 (2.05 - 1.96)	37.88 - 1.77 (1.82 - 1.77)
No. reflections	16675 (2734)	22604 (2699)	0.2452 (0.3386)
<i>R</i> _{work} / <i>R</i> _{free}	0.1976 (0.3072) / 0.2546 (0.3361)	0.1857 (0.3229) / 0.2193 (0.3577)	0.1973 (0.3160) / 0.2452 (0.3386)
No. atoms			
Protein	1517	1528	1507
Ligand/ion	n/a	n/a	n/a
Water	69	64	54
<i>B</i> -factors			
Protein	47	55	21
Ligand/ion	n/a	n/a	n/a
Water	46	52	29
R.m.s. deviations			
Bond lengths (Å)	0.005	0.013	0.012
Bond angles (°)	0.71	1.22	1.15

*Single xtal used for each data/structure.

*Values in parentheses are for highest-resolution shell.

Cont...

	Gawi-09 WT apo PDB Code: 9YOX	Dokki-15 WT apo PDB Code: 9YNM	Dokki-15 C100A complex PDB Code: 9YNL
Data collection			
Space group	P 1 2 ₁ 1	P 2 ₁ 2 ₁ 2 ₁	P 4 ₁ 2 ₁ 2
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	42.28, 61.19, 63.36	38.23, 61.86, 74.76	61.63, 61.63, 374.44
α , β , γ (°)	90, 105.24, 90	90, 90, 90	90, 90, 90
Resolution (Å)	32.88 - 2.17 (2.28 - 2.17)	34.04 - 1.34 (1.42 - 1.34)	34.48 - 3.38 (3.56 - 3.38)
<i>R</i> _{sym} or <i>R</i> _{merge}	0.242 (2.212)	0.094 (1.888)	0.755 (4.883)
<i>I</i> / σ <i>I</i>	5.7 (1.6)	11.0 (0.4)	5.0 (1.4)
Completeness (%)	99.5 (98.8)	85.6 (46.1)	99.9 (100)
Redundancy	7.0 (7.0)	11.4 (4.7)	25.2 (27.0)
Refinement			
Resolution (Å)	32.88 - 2.17 (2.23 - 2.17)	34.04 - 1.56 (1.61 - 1.56)	34.48 - 3.38 (3.53 - 3.38)
No. reflections	16577 (1386)	25703 (2197)	10991 (1315)
<i>R</i> _{work} / <i>R</i> _{free}	0.2079 (0.2886) / 0.2655 (0.3424)	0.1715 (0.2117) / 0.2017 (0.2625)	0.2226 (0.2974) / 0.2794 (0.3333)
No. atoms			
Protein	2868	1512	4543
Ligand/ion	10	n/a	n/a
Water	55	207	6
<i>B</i> -factors			
Protein	38	22	94
Ligand/ion	47	n/a	n/a
Water	37	34	56
R.m.s. deviations			
Bond lengths (Å)	0.002	0.010	0.002
Bond angles (°)	0.48	1.15	0.44

*Single xtal used for each data/structure.

*Values in parentheses are for highest-resolution shell.

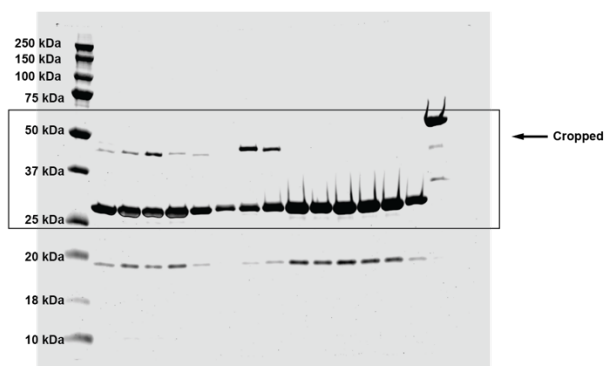
Supplementary table 4. Foldseek analysis for crystal structures

Designs	TM score	Database	Target
Kal WT complex	0.48	AFDB50	AF-A0A923ZJ62-F1-model_v6
Kal C103A complex	0.49	AFDB50	AF-A0A923ZJ62-F1-model_v6
Gawi-09 WT apo	0.46	AFDB50	AF-A0A0X8G3I4-F1-model_v6
Dokki WT apo	0.50	All databases	A0A346A683
Dokki-15 WT apo	0.50	MGNIFY_ESM30	MGYP001077622663
Dokki-15 C100A complex	0.48	CATH50	2.170.210.20 Q9N4L1 1-108

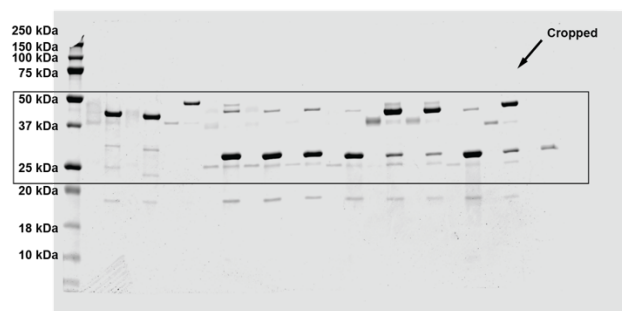
Supplementary table 5. Kinetics summary of sequence specific cysteine proteases

Enzymes	substrate	k_{cat} (s ⁻¹)	K_M (μM)	k_{cat}/K_M (M ⁻¹ s ⁻¹)
Dokki	DABCYL-GSVPTLTGG↓GS – EDANS	0.0003 ± 0.0001	1900 ± 700	0.15 ± 0.08
Dokki-15	DABCYL-GSVPTLTGG↓GS – EDANS	0.0011 ± 0.0001	89 ± 10	12 ± 2
Dokki-v2	DABCYL-GSVPTLTGG↓GS – EDANS	0.064 ± 0.005	154 ± 14	413 ± 50
Ulp1	eGFP-SUMO-REQIGG↓G-mScarlet	6.1 ± 0.5	17 ± 2	3.5 ± 0.4 × 10 ⁵
	eGFP-REQIGG↓G-mScarlet	Not measurable	Not measurable	Not measurable
TEV(Kapust et al. 2001)	ENLYFQ↓S	0.16 ± 0.01	61 ± 10	2620 ± 460
mPro(Kenneson et al. 2025)	DABCYL-KTSAVLQ↓SGFRKME-(EDANS)-NH2	1.23	190	6,600
HRV3C(Kenneson et al. 2025)	MBP-LEVLFQ↓X-6His-GFP/GST	0.021-0.32	34 - 49	390 - 4730

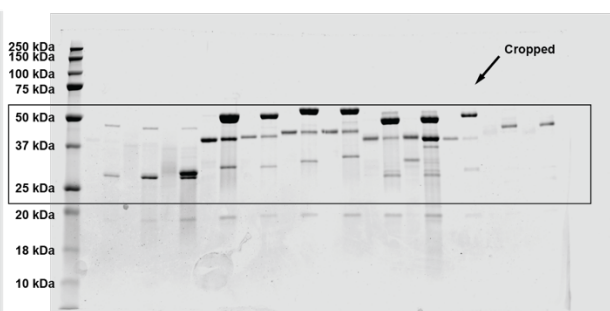
a. Figure 1d: hc04,05,06,07,10,12,13,14,32,33,34,37,38,TEV WT, TEV C151A



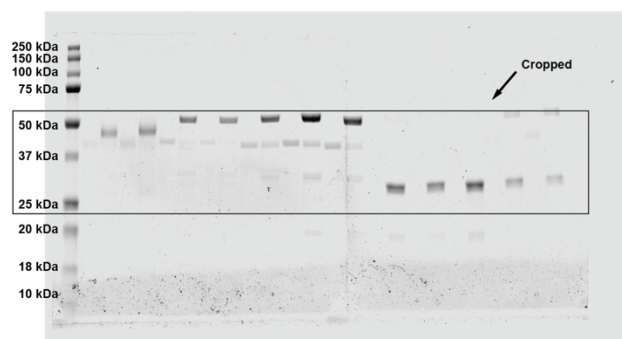
b. Extended data figure 2: hc01-12



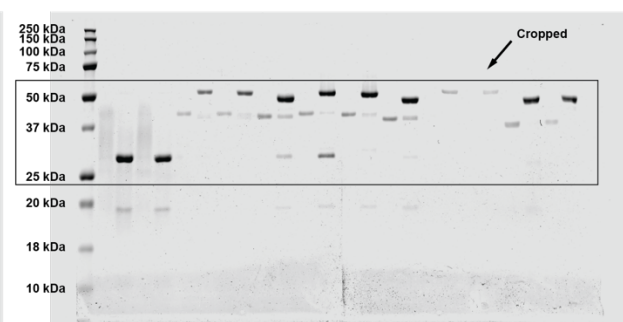
c. Extended data figure 2: hc13-24



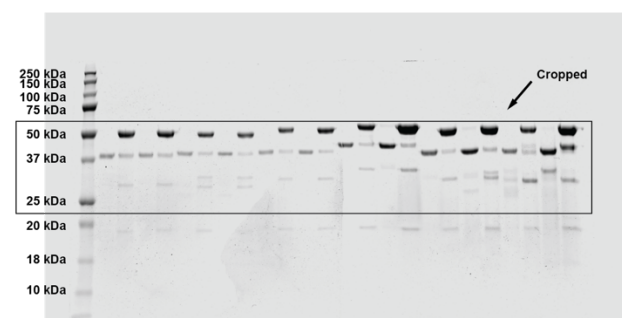
d. Extended data figure 2: hc25-36



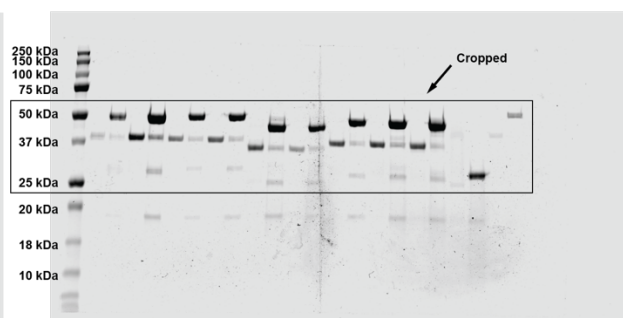
e. Extended data figure 2: hc37-48



f. Extended data figure 2: hc49-60

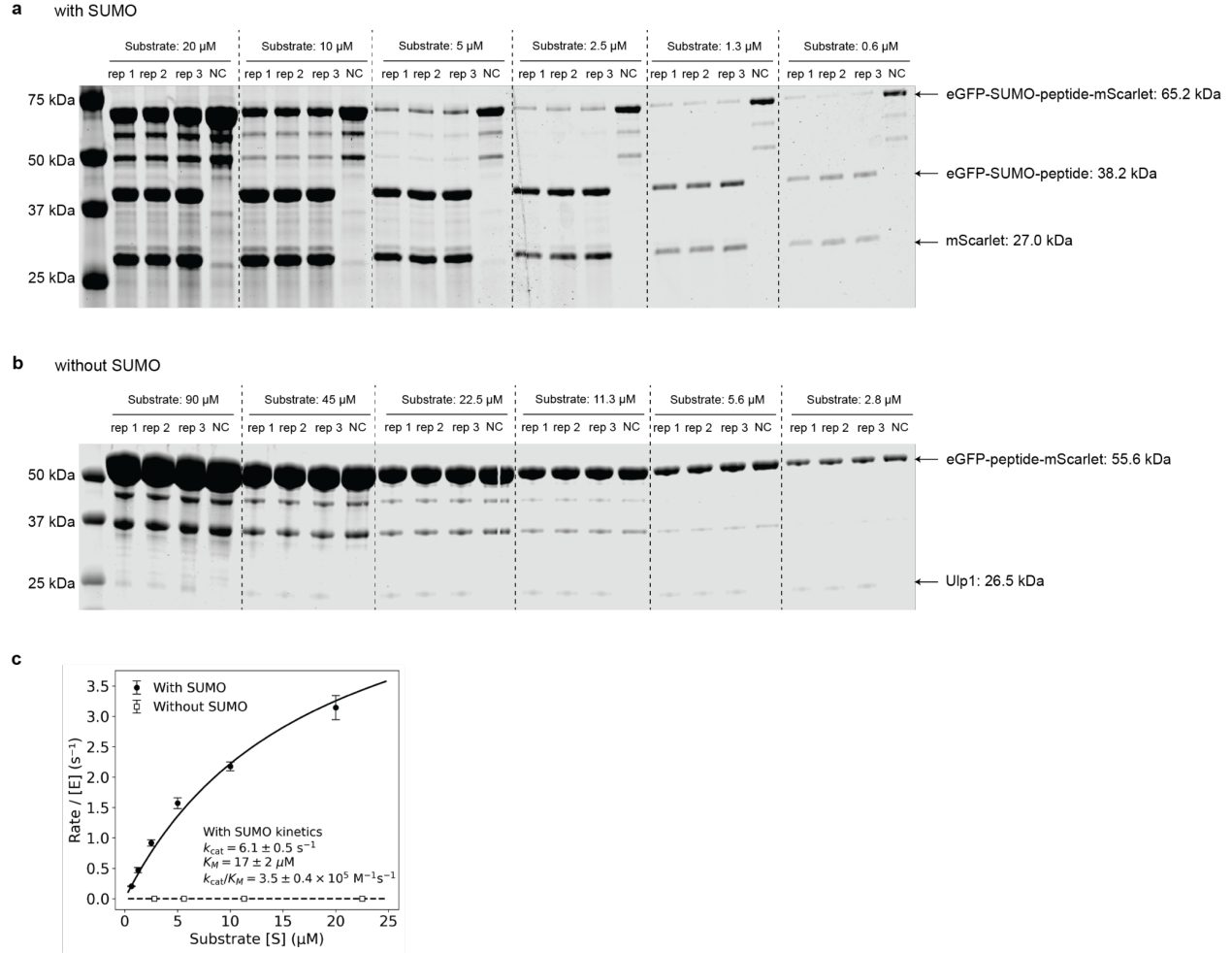


g. Extended data figure 2: hc61-69 & TEV WT, TEV C151A



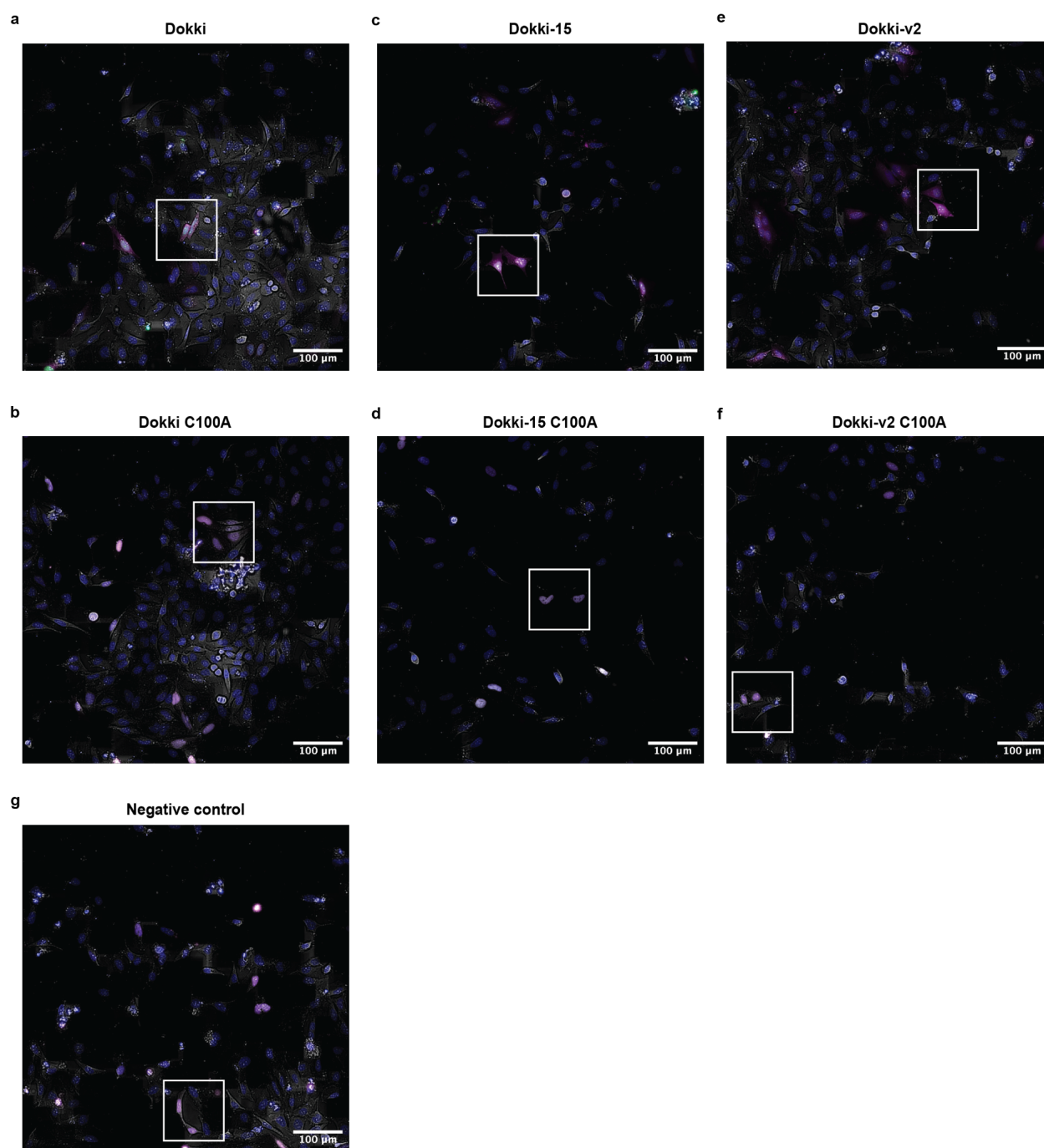
Supplementary Fig. 1 | Full image for SDS-page gel.

Full SDS-page gel image for **a**, Figure 1d, **b–g**, Extended data Figure 2



Supplementary Fig. 2 | Gel-based kinetic analysis of Ulp1 cleavage activity.

Ulp1 activity against eGFP-peptide-mScarlet substrates was measured using a gel-based cleavage assay. **a**, For SUMO-containing substrates, reactions contained 30 nM Ulp1 and two-fold serial dilutions of substrate from 20 to 0.6 μM . Reactions were incubated for 1 min at room temperature. Each concentration performed three technical replicates. **b**, For substrates lacking SUMO, reactions contained 1 μM Ulp1 and two-fold serial dilutions of substrate from 90 to 2.8 μM . Reactions were incubated for 1 h at room temperature. Each concentration performed three technical replicates. Reaction products were resolved by SDS-PAGE, stained with Coomassie blue and imaged on a LI-COR imaging system using Coomassie detection. **c**, Steady-state kinetic analysis of the Ulp1 to with/without SUMO substrate, fitted to the Michaelis-Menten substrate inhibition model to derive apparent k_{cat} , K_M , K_I , and catalytic efficiency (k_{cat}/K_M). Error bars represent the standard deviation of three technical replicates.



Supplementary Fig. 3 | Raw image for cell-based reporter assay

a–g, Representative fluorescence microscopy images of HeLa cells before crop. Cropped region is shown as white box **a**, Dokki; **b**, Dokki-15; **c**, Dokki-v2; **d**, Dokki C100A; **e**, Dokki-15 C100A; **f**, Dokki-v2 C100A; **g**, Negative control.

Supplementary information 1. pCool6 plasmid DNA sequence

actcctgcattaggaagcagcccagtagtaggtgaggccgttgagcaccgccgccgaaggaatggtgcatgcaaggagatggcgcccaacagtccccggcca
cggggcctgccaccataccacgccgaaacaagcgtcatgagcccgaagtggcgagcccgatctcccatcggtgatgtcggcgatataggcgccagcaaccg
cacctgtggcgccggtgatccggccacgatcgctccggcgtagaggatcgagatcgatctcgatcccggaattaatacactactataggggaattgtgagcg
gataacaattcccctctagaataattttgttaactttaagaaggagatatacatatgTCAGGAGGAGACCGAaagcgtattacagtgcaggtgacagcg
acagctatcagttgctcaaggcatatatgatgtcaatatctccggtctggtatTgcacaaccaagaatgaagcccATGCAGTTTAAGGTTTACACCT
ATAAAAGAGAGAGCCGTTATCGTCTGTTTGTGGATGTACAGAGTGATATTATTGACACGCCCCGGGCG
ACGGATGGTGATCCCCCTGGCCAGTGCACGTCTGCTGTCAGATAAAAGTCTCCCGTGAACTTTACCCG
GTGGTGCATATCGGGGATGAAAGCTGGCGCATGATGACCACCGATATGGCCAGTGTGCCGGTtTCCG
TTATCGGGGAAGAAGTGGCTGATCTCAGCCACCGCGAAAATGACATCAAAAACGCCATTAACCTGA
TGTTCTGGGGAATatgaACGGTCTCGTTCCGGCAGCGGCAGCGGCAGCATGGATAGCACCGAAGCGGT
GATTAAAGAATTTATGCGCTTTAAAGTGCATATGGAAGGCAGCATGAACGGCCATGAATTTGAAATT
GAAGGTGAAGGCGAAGGCCGCCCCGATGAAGGCACCCAGACCGCGAAACTGAAAGTGACCAAAGG
CGCCCCGCTGCCGTTTAGCTGGGATATTCTGAGCCCCGAGTTTATGTATGGCAGCCGCGCGTTTATTA
AACATCCGGCGGATATCCCGGATTATTGGAAACAGAGCTTTCCGGAAGGTTTTAAATGGGAACGCGT
GATGATCTTTGAAGATGGTGGCACCGTGAGCGTGACCCAGGATACCAGCCTGGAAGATGGCACCCCT
GATTTATAAAGTGAAACTGCGCGGCGGCAACTTTCCGCCGGATGGCCCCGGTGATGCAGAAACGCAC
CATGGGCTGGGAAGCGAGCACCGAACGCCTGTATCCGGAAGATGTGGTGCTGAAAGGCGATATTAA
AATGGCGCTGCGCCTGAAAGATGGCGGCCGCTATCTGGCGGATTTTAAAACACCTATAAAGCGAA
GAAACCGGTGCAGATGCCGGGCGCCTTTAACATTGATCGCAAACCTGGATATTACCAGCCATAACGA
AGATTATACCGTGGTGGAACAGTATGAACGCAGCGTGGCGCGCCATAGCACCGGCTCTcatcatcatcaca
TCATaatagatccggctgctaacaagccccgaaggaagctgagttggctgctgccaccgctgagcaataactagcataacccttggggcctctaaacgggt
cttgaggggtttttgctgaaaggaggaactatatacggattggcgaaatgggacgcgcctgtagcggcgcaataagcgcggcggtgtgtgttgacgcgagcgt
gaccgctacacttgcagcgccttagcgcgcctctcttcccttcttctcgcacggttcgcccgttccccgtaagctctaatacggggctcccttta
gggttccgatttagtgctttacggcacctcgacccccaaaaacttgattaggggtgatggttcacgtagtgggccaatgcgcctgatagacggttttcccttgacgttgg
agtccacgttcttaatagttgactctgttccaaactggaacaacactcaacctatctcggtctattcttttgattataagggattttgccgatttcgcctattggtaaaa
aatgagctgatttaacaaaaatttaacgcgaatttaacaaaatfaaacgtttacaatttcagggtgcacttttcggggaaatgtgcgcggaacccctatttgttttttcta
aatacatcaaatatgtatccgctcatgaattaattctagaaaaactcatcgagcatcaaatgaactgcaatttattcatatcaggattatcaatacatattttgaaaaag
ccgtttctgtaataaggagaaaaactcaccgagcgagttccataggatggcaagatcctggtatcggttcgattccgactcgtccaacatcaatacaacctattaattt
ccccctgtaaaaaaagggttatcaagtgagaatcaccatgagtgacgactgaatccggtgagaaatggcaaaagtttatgcatttcttccagactgttcaacaggcca
gccattacgctcgtcatcaaaatcactcgcataaccaaacctgtattcattcgtgattgcgcctgagcgagacgaatacgcgacgctgttaaaaggacaattacaaa
caggaatcgaatgcaaccggcgaggaacactgccagcgcatcaacaatattttacctgaatcaggatattcttataacctggaatgctgttttccggggatcgca
gtggtgagtaacatgcatcatcaggagtagcgtgataaaatgcttgatggcggaagagcgataaattccgtcagccagtttagtctgacctctcatctgtaacatcattg
gcaacgctacctttgccatgttcagaaaactctggcgcatcgggcttccatacaatcgatagattgtcgacactgattgcccgacattatcgagccatttatac
ccataataatcagcatccatgttggaatttaacgcggcctagagcaagacgtttcccggtgaatatggctcataacacccctgtattactgtttatgtaagcagacagtttt
attgttcagacaaaaatcccttaacgtgagtttctgtccactgagcgtcagaccccgtagaaaagatcaaaaggatcttcttgagatcctttttctgcgcgtaattctgctg
cttgcaaaaaaaaaccaccgctaccagcgggtggttgtttgccgatcaagagctaccaactcttttccgaaggtaactggcttcagcagagcgagatacaaat
actgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgtaactcctgttaccagtggctgctgccagtgccgat
aagtcgtgtcttaccgggttgactcaagacgatagttaccggataaggcgagcggtcgggctgaacggggggtcgtgcacacagcccagcttggagcgaaacg
acctacaccgaactgagatacctacagcgtgagctatgagaaagcgccacgcttccgaaggagaaaggcgacaggtatccggttaagcggcaggggtcggaac
aggagagcgcacgaggagcttcagggggaacgcctggtatctttatagctctgctgggttccgacacctctgacttgagcgtcgattttgtgatgctcgcaggg
ggcgagcctatgaaaaacgccagcaacgcggcctttacggcttctggccttttctgctgacatgttcttctgcggtatccctgattctgttgataa
ccgtattaccgctttgagtgagctgatacgtcgcgcagccgaacgacgagcgcagcgagtgagtgaggaagcgggaagagcgctgatcggtatttt
ctccttacgcatctgtgctgtatttcaaccgcatatatggtgcactctcagtaacaatctgctctgatccgcatagttaagccagtatacactccgctatcgctacgtgact
gggtcatggctgcgccccgaccccccaacacccgctgacgcgcctgacgggcttctgctcggcatccgcttacagacaagctgtgaccgtctccgggag
ctgcatgtgtcagaggtttaccgctatcaccgaaacgcgcgagcgagctgcggttaagctcatcagcgtggtcgtgaagcgattcacagatgtctcgttcatcc

gcgtccagctcgttgagtttccagaagcgtaattgtctggcttctgataaagcgggcatgtaaggcggttttctgtttggtcactgatgcctccgtgaagggg
gatttctgttcatgggggtaataaccgatgaaacgagagaggtgctcacgatacgggttactgatgataacatccccgttactggaacgttgtagggtaaca
actggcggtatggatgcggcgaggaccagagaaaaactcactagggtcaatgccagcgcttcgtaatacagatgtaggtgtccacagggtagccagcagcatcctg
cgatgcagatccggaacataatggtgcagggcgctgacttccgctttccagactttacgaaacacggaaccggaagaccattcatgtttgtctcaggtcgcagacg
ttttgcagcagcagtcgcttcacgttcgctcgcgtatcgggtattcattctgctaaccagtaaggcaacccccgccagcctagccgggtcctcaacgacaggagcagcat
catgcgcacccgtggggcgccatgccggcgataatggcctgcttctgccgaaacgtttgggtggcgggaccagtgcgaaggcttgagcgaggggcggtgcaagat
tccgaataaccgaagcgacaggccgatcgtcgcgctccagcgaaagcggtcctgccgaaatgaccagagcgctccggcacctgtcctacgagttgcat
gataaagaagacagtcataagtgcggcgacgatagtcacccccgcgccaccggaaggagctgactgggtgaaggctcgaaggcatcggctgagatccggg
tgcctaagtgtgagtaacttacattaattgcttgcgctcactgccgcttccagtcgggaaacctgtcgtgccagctgcattaatgaatggccaacgcgcggggga
gaggcggtttgcgtattggggcgccagggtggttttctttaccagtgagacgggcaacagctgattgcccttaccgcctggccctgagagagttgcagcaagcggt
ccacgctggtttccccagcaggcgaaaaactcgtttgatggtggttaacggcgggataaatacagctgtcttgcgtatcgtctatccactaccgagatgtccga
ccaacgcgcagccccggactcggtaattggcgcgcaattgcgccagcgccatctgatcgttggcaaccagcatcgcagtggaacgatgccctcattcagcatttgc
ggtttgttgaacacggacatggcactccagtcgccttccggttccgctatcggtgaatttgatgcgagtgagatatttatgccagccagccagacgcagacgcgc
gagacagaacttaaggcccgtaacagcgcggttgcgtggtgacctaatgcgaccagatgctccacgccagtcgctaccgttctatgggagaaaataact
gttgatgggtgtctgtgcagagacatcaagaataacccgggaacattagtgcaggcagcttccacagcaatggcatcctggtcatccagcggtatgtaattgatcag
ccactgacgcgttgcgcgagaagattgtcaccgccgctttacaggcttcgacgccgcttcttaccatcgacaccaccagctggcaccagttgatcggcg
gagatthaatgccgcgacaatttgcgacggcgctgcagggccagactggaggtgcaacgccaatcagcaacgactgtttcccgcagttgttgcacgcg
gttgggaatgtaattcagctccgcatcgcgcttccacttttcccgcttttcgcagaacgtggctggcctgttaccacggcgaaacggtctgataagagacac
cggcactactctgcgacatctataacgttactggtttacattcaccacctgaattgactcttccggcgctatcatgccataccgcgaagggtttgcgccattcgt
ggtgtccgggatctcgacgtctcccttatgcg

Supplementary information 2. ZJ101 plasmid DNA sequence

gaagatcctttgatcttttctacggggtctgacgctcagtggaacgaaaactcacgttaagggttttggatgagattatcaaaaaggatcttcacctagatccttttaaa
taaaaatgaagtttaaatcaatctaaagtatatatgagtaaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatctcagcgatctgtctatttctgtcatcca
tagttgcctgactccccgtcgtgtagataactacgatacgggaggggttaccatctggccccagtgctgcaatgataccgcgGgaccacgctcaccggctccagatt
tatcagcaataaaccagccagccggaaggccgagcgcagaagtggctcgaactttatccgctccatccagcttattaattgttgcgggaagctagagtaagta
gttcgccagttaatgtttgcgaacgttgttgcattgtctacaggcatcgtgtgtcacgctcgtctgttggatggcttcattcagctccggttccaacgatcaaggcg
agttacatgatccccatgttgtgcaaaaagcgggttagctccttcggtcctccgatcgttgcagaagtaagttggccgcagtggtatcactcatggtatggcagcactg
cataattctctactgtcatgccatccgtaagatgctttctgtgactggtgagtagtactaaccaagtcattctgagaatagtgtatgcggcgaccgaggtgctcttgcggcg
gtcaatacgggataataccgcgccatagcagaactttaaaagtctcatcattggaacgttcttcggggcgaaaactctcaaggatcttaccgctgttgatcca
gttcgatgtaaccacactcgtgcaccaactgatcttcagcatctttactttaccagcggttctgggtgagcaaaaacaggaaggcaaaaatgccgcaaaaagggaata
agggcgacacggaaatgttgaactcactcttcttttaataattattgaagcatttatcagggttattgtctcatgagcggatacatatttgaatgtatttagaaaaataa
acaaataggggttccgcgcacatttccccgaaaagtgccacctgacgtggatccccgagggggccccatgggctagaggatccggcctcggcctctgcataaat
aaaaaaaattagtcagccatgagcttggccattgcatacgttgatccatatcataatgtacattatattggctcatgtccaacattaccgcatgttgacattgattatt
gactagttataatagtaataacacggggtcattagttcatagccatatatggagttccgcttacataacttacggtaaatggcccgctggctgaccgccaacga
ccccgccattgacgtcaataatgacgtatgtcccatagtaacgccaatagggactttccattgacgtcaatgggtggagatttaccgtaaaactgccacttggcag
acatcaagtgtatcatatgccaaagtacccccctattgacgtcaatgacggtaaatggcccgctggcattatgccagtcacatgaccttatgggactttctacttggca
gtacatctacgtattatgcatcgtattaccatggtgatcgggttttggcagtcacatcaatggcgctggatagcgggttgactcacggggattccaagtctccacccatt
gacgtcaatgggagtttgtttggcaccaaaataacgggactttccaaaatgtcgtacaactccgccattgacgcaaatggcggttagggcgtgtacggtgggag
gtctatataagcagagctcgtttagtgaaccgtcagatcgctggagacgccatccacgtgtttgacctccatagaagacaccgggaccgatccagcctcccccg
agcttcatgttggtaccgagctcggatcctgagaactcagggtgagctctatgggacccttgatgttttcttcccccttctttctatggttaagtcatgtcataggaagggg
agaagtaacagggtacacataattgacaaatcagggttaatttgcatttgaattttaaaaaatgcttcttcttttaataatactttttgttatcttatttctaatactttccctaact
ctttcttccagggaataatgataaatgtatcatgccttctgacacattctaaagaataacagtgataatttctgggttaaggcaatagcaatttctgcataaaatatttct
gcataaaattgtaactgatgtaagaggttcatattgctaatagcagctacaatccagctaccattctgctttttttatggttgggataaggctggaatttctgagtccaa
gctaggccccctttgtaatactgttcatactcttcttctccacagctcctgggcaacgtgctggtctgtgtgctggcccatcatttggcaaacgacgtgagatctga
attctgacactAAGGTTGAGACCgccaccaagcgtattacagtgacagttgacagcgacagctatcagttgctcaaggcatatatgatgtaataatctccggt

ctggtaTgcacaaccaagaatgaagcccATGCAGTTTAAAGGTTTACACCTATAAAAGAGAGAGCCGTTATCGTCTGT
TTGTGGATGTACAGAGTGATATTATTGACACGCCCGGGCGACGGATGGTGATCCCCCTGGCCAGTGC
ACGTCTGCTGTCAGATAAAGTCTCCCGTGAACTTTACCCGGTGGTGCATATCGGGGATGAAAGCTGG
CGCATGATGACCACCGATATGGCCAGTGTGCCGGTtTCCGTTATCGGGGAAGAAGTGGCTGATCTCA
GCCACCGCGAAAATGACATCAAAAACGCCATTAACCTGATGTTCTGGGGAATAAtgaGGTCTCcCCTActc
aaatcctgcacaacagattcttcatgtttggaccaaatcaacttgtgataccatgctcaagaggcctcaattatattgagtttttaattttatgaaaaaaaaaaaaaaaa
cgggaattcaccaccagtgaggctgcctatcagaaagtgggtggtggtgctaatgccctggcccacaagtatcactaagctcgctttctgctgtccaatttctat
taaagggtcctttgttccctaagtccaactactaaactgggggatattatgaagggccttgagcatctggattctgcctaataaaaaacatttttcttgcattgcaatgatgtatt
aaattatttctgaatattttactaaaaaggggaatgtgggaggtcagtgcatttaaacataaagaaatgaagagctagtccaaccttgggaaaatacactatatcttaact
ccatgaaagaaggtgaggctgcaaacagctaatagcacattggcaacagcccctgatgcctatgccttattcatccctcagaaaaggattcaagtagaggcttgatttgg
aggttaaagttttgcctatgtctgtattttacattactattgttttagctgtcctcatgaatgtcttttactaccatttgcctatcctgcacctcagccttgactccactcagttctc
ttgcttagagataccacctttccctgaagtgttcccttccatgttttacggcgagatggtttctcctgcctggccactcagccttagttgtctctgttgccttatagaggcttac
ttgaagaaggaaaaacagggggcatggtttgactgtcctgtgagcccttcttccctgcctccccactcacagtgaccggaatccctcgacatggcagcttagcacta
gtgcggccgcagatctgttctcctgcctcactgactcgtcgtcgtcgttcggctgcggcgagcgggtatcagctcactcaaaggcggtaatacgggtatccacaga
atcaggggataacgcaggaaagaacatgtgagcaaaaaggccagcaaaaaggccaggaaccgtaaaaaggccgcttgcgtgcgttttccataggctccgcccccc
tgacgagcatcacaaaaatcgacgctcaagtcagaggtggcgaaacccgacaggactataaagataccaggcggttccccctggaaagctccctcgtcgtcctcctg
ttccgaccctgccgttaccggatacctgtccgccttttcccttcgggaagcgtggcgcttttctcatagctcacgctgtaggtatctcagttcgggtgtaggtcgttcgctc
caagctgggctgtgtcacgaaccccccggtcagcccaccgctgcgccttatccggtaactatcgtcttgagtccaacccggtaagacacgacttatcgcactggc
agcagccactggtaacaggatttagcagagcgaggtatgtaggcgggtgctacagagttcttgaagtgggtggcctaactacggctacactagaagaacagtatttggat
ctgcgctctgctgaagccagttaccttcggaaaaagagttggtagctcttgatccggcaacaaccaccgctggtagcgggtggttttttgttgaagcagcagattac
gcgcagaaaaaaggatctcaa