university of groningen / faculty of science and engineering

N-Terminally tagging a substrate for ClpA-mediated unfolding

Bachelor Research Project (Molecular Life Sciences)

Malachy Bayston (s5259991)

Examiners:

Sumayra Sultanji

Dr. Giovanni Maglia

23/06/2025

Abstract

ClpA is a bacterial AAA+ unfoldase that recognises and unfolds substrates tagged with degradation signals. It has been proposed as a candidate for its use the upcoming nanopore-based protein sequencing technology due to potentially having more consistent unfolding steps compared to ClpX. Previous work has highlighted ClpA's ability to degrade substrates bearing an N-terminal RepA tag. In this project we expressed and purified a RepA-mNG fusion protein and tested its degradation by various ClpA constructs. Our results showed that although full-length RepA-mNG could be purified using a SUMO-tag strategy, degradation remained low in fluorescence assays. While N-terminal RepA tagging shows potential, mNG may be a limiting factor in the assay. Future work should explore GFP as an alternative reporter to improve fluorescence degradation rates and evaluate degradation under high-salt concentrations to better mimic the conditions required in electrophysiology experiments. These experiments could help determine whether ClpA is a suitable unfoldase for nanopore protein sequencing applications.

Acknowledgements

I would like to thank Sumayra for her excellent supervision throughout this project. Thank you to Mihai, you were an excellent lab partner. I would also like to thank Giovanni Maglia for the opportunity to work in your lab. Lastly, I would like to thank the rest of the Chemical Biology I group, you were very helpful in the lab and a pleasure to be around.

Table Of Contents

Introduction

ClpA is a member of the Hsp100 protein family, which is a group of AAA+ chaperones that are capable of unfolding proteins (Saibil, 2013). They are found in bacterial cells such as *Escherichia Coli.* ClpA monomers assemble to form hexameric rings, with a central channel containing conserved pore loops (see Fig. 1b) (Zuromski et al., 2021). Each subunit contains two ATPase sites (see Fig. 1a). As the protein substrate enters the channel, the pore loops of ClpA interact with the backbone of the substrate (Shih et al., 2024). Unfolding is mediated by a sequential hand over hand movement caused by the conformational changes of the ClpA subunits due to the hydrolysis of ATP, which mechanically pulls the substrate through the channel in a stepwise fashion (Shih et al., 2024). ClpA can form a complex with the serine protease ClpP (Shih et al., 2024). If they are complexed together then the protein substrate will get degraded after its unfolding as it passes from ClpA to ClpP. In nature, the unfolding capabilities of ClpA are very important in regulating various cellular processes. For instance, the ClpAP complex can degrade misfolded or damaged proteins in the cell (Hoskins, Singh, et al., 2000). ClpA is also capable of dissociating oligomeric proteins through its unfolding action. One such protein is RepA (see Fig. 1c), which is inactive as a dimer, but once monomerised by ClpA is capable of binding to plasmid DNA to initiate DNA replication (Hoskins, Singh, et al., 2000).

Previous studies have shown that ClpA is capable of unfolding the fusion proteins RepA(1-70)-GFP and RepA(1-15)-GFP (based on fluorescence degradation experiments) (Hoskins, Kim, et al., 2000), which suggests that ClpA recognizes a region in the first 15 amino acids of the RepA protein. This is supported by their additional finding that shows that RepA(16-286) is degraded significantly slower than wildtype RepA (determined by SDS-PAGE analysis). Furthermore, it was shown to be capable of unfolding GFP when RepA is used as a C-terminal tag, the construct being GFP-RepA(1-15) (Hoskins et al., 2002a). Interestingly, ClpA was also

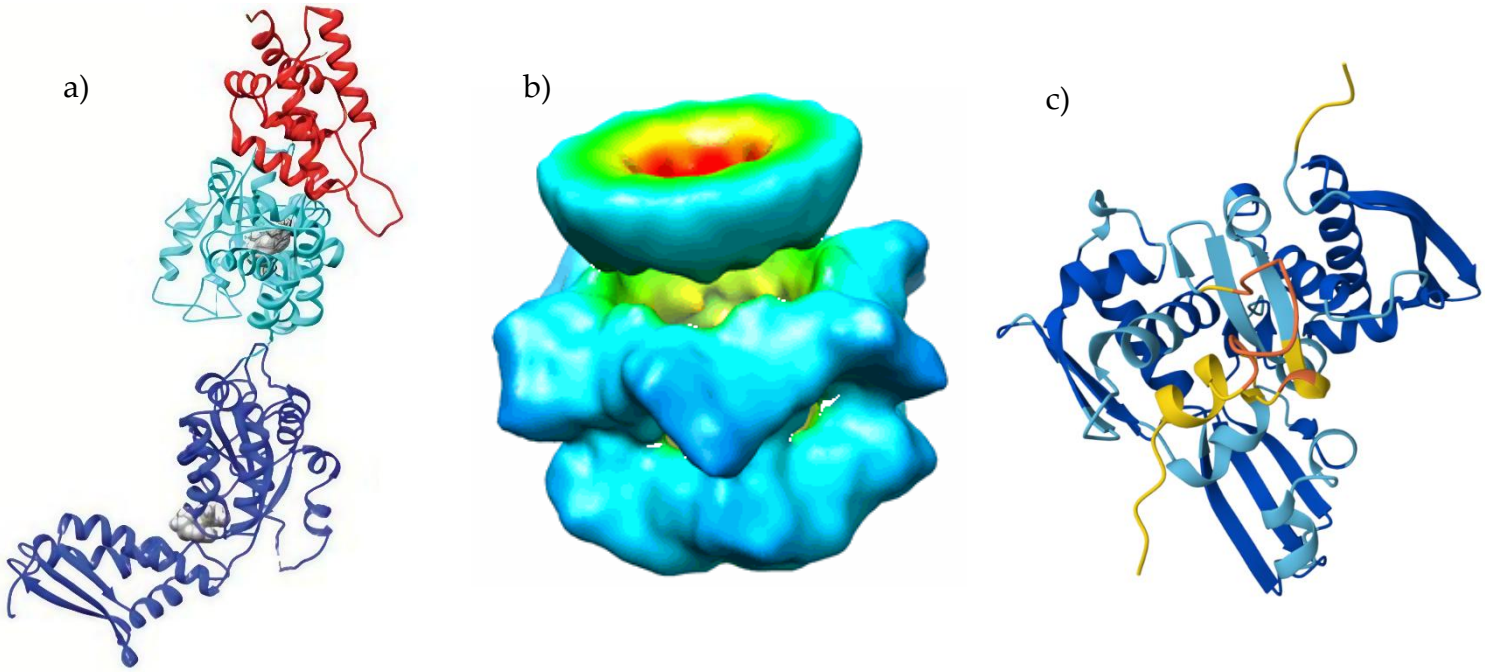

Fig. 1: (a) Ribbon diagram of ClpA monomer with ATP bound. Red = N-Domain, Cyan = D1 ATPase domain, Blue = D2 ATPase domain (Image adapted from Clare & Saibil, 2013). (b) Cryo-EM structure of ClpA hexamer (EMDB entry EMD-1673). (c) Monomeric structure of RepA, predicted by AlphaFold (https://alphafoldserver.com/).

shown to be able to unfold GFP-RepA(1-286), which indicates that ClpA can unfold substrates with recognition sequences located in the interior of the primary sequence (Hoskins et al., 2002a).

In addition to RepA-derived tags, ClpA also recognizes other well-characterized degrons. One such tag is ssrA, an 11-amino acid sequence that is added to the C-terminus of proteins that stall at the ribosome which tags the protein for degradation (Hoskins, Kim, et al., 2000). This occurs when the messenger RNA is damaged and does not have a stop signal. It has been shown that ClpA recognises the fusion protein GFP-ssrA (Flynn et al., 2001).

Unfoldases are being studied for their use in single molecule protein sequencing (Motone et al., 2024). This upcoming technology uses nanopores to determine the amino acid sequence of a protein by measuring how the ionic current changes based on which amino acids are in the channel of the nanopore (Motone et al., 2024). To

achieve this, unfoldases can be used to unfold the protein to be sequenced so that it can be passed through the nanopore at a consistent rate. For the unfoldase to be able to recognise the substrate, the C-terminal tag ssrA is predominantly used (Motone et al., 2024). ClpA is a promising unfoldase to use in this context, since it has been proposed that it takes shorter more consistent steps compared to the more commonly used ClpX (Olivares et al., 2014), which would provide more consistent information potentially resulting in a higher resolution.

In this project, we aim to test the ability of ClpA to unfold a protein substrate containing the first 70 amino acids of RepA as a recognition tag. Having an N-terminus tag such as RepA would provide two benefits for nanopore protein sequencing. Firstly, the ability to unfold and measure the change in current of a known fusion protein from the N->C direction (as opposed to C->N as is the case when using ssrA) would provide more information on how the current changes for different amino acids. Secondly, in practical nanopore protein sequencing applications, tags need to be added to the unknown proteins for their recognition and unfolding by unfoldases. This could be achieved enzymatically through the use of an omniligase for example. Importantly, if the C-terminus of the protein is structurally buried or post translationally modified, then having a functional N-terminal tag becomes essential.

Materials and Methods

Plasmids and cloning:

Primers were ordered from IDT (Table 1). ThermoFisher PhusionU Hot Start DNA polymerase protocol was followed for USER cloning. A pET24 plasmid expressing RepA-TwinStrep-mNG fusion protein was created with USER cloning. Starting with a RepA-mNG-Strep construct, a twin strep tag was inserted between the RepA and mNG sequences of the fusion protein, then the C-terminal strep tag was removed.A pBAD plasmid expressing $ClpA_{\Delta C}$ was mutated to M169T with USER cloning. Additionally, a 26AA linker between T142 and T169 was removed and replaced with GSGS. This part formed the linker between the N-domain and the rest of the protein. This construct will be written as $ClpA_{\Delta L \Delta C}$. A $ClpA_{\Delta N \Delta C}$ plasmid was already available for us. DNA purification of PCR product was performed using Thermo Fisher's GeneJET Genomic DNA Purification Kit. Agarose gel electrophoresis was done to check if the PCR and USER cloning was successful.

The *E. coli* strain E. cloni 10G chemically competent cells used for transformation for all cloning and plasmid purification purposes. Cells were incubated at 37°C overnight on plates with the appropriate antibiotic (Carbenicillin 100 μg/ml for our pBAD plasmids, Kanamycin 50 μg/ml for our pET24 plasmids). Individual colonies were selected to inoculate 10ml of LB medium with the appropriate antibiotic and grown overnight at 37°C. Their plasmids were purified using ThermoFisher's GeneJET Plasmid Miniprep Kit. To confirm that the USER cloning was successful, DNA sequencing was done on these plasmids.

Table 1: Cloning Plan						
Antibiotic Resistance	Final construct / Purpose		>	Primer Name	Primer Sequence	Template
Ampicillin	pET24_RepA_TwinStrep_mNG_strep	1. Backbone prep	Fw	RepA_mNG_NoLZ_Fw	AGCGGUGGCGGTGGGGTATC	pET24_RepA_mNG_strep
			Rv	mNG_bb_forTwinstrep_Rv	ACGAACUAGAACCTTCCGCTTTCGCGATTTC	
		2. TwinStrep	Fw	TwinStrep-ins-Fw	agttcgUCTGCATGGAGTCATCCCC	pT7 SC1_His6-MBP-ssrA
			Rv	TwinStrep_ins_Rv	ACCGCUTTTCTCGAACTGTGGATGT	
	pET24_RepA_TwinStrep_mNG	Remove C-term strep	Fw	RepA_removeCstrep_Fw	ACAAATAAUAAAAGCTTGCGGCCGC	pET24_RepA_TwinStrep_mNG_strep
			Rv	RepA_removeCstrep_Rv	ATTATTTGUACAGTTCGTCCATGCCCATC	
Kanamycin	pBAD_6His_ClpAdC_M169T_Strep	M169T Mutation	Fw	ClpA-M169T_Fw	AGGAACGUAcGGAGAATTTCACGA	pBAD_6His_ClpAdC_Strep
			Rv	ClpA-M169T_Rv	ACGTTCCUCCCCACCAGCT	
	pBAD_ClpAdC_M169T_Strep	Remove N terminus his tag	Fw	ClpA_removeHistag_Fw	aaccaugCTCAATCAAGAACTGGAAC	pBAD_6His_ClpAdC_M169T_Strep
			Rv	ClpA_removeHistag_Rv	ATGGTUAATTCCTCCTGTTAG	
	pBAD_ClpAdLdC_M169T_Strep	Remove N domain linker	Fw	ClpA_removeLinker_Fw	aggatcUACGGAGAATTTCACGACG	pBAD_ClpAdC_M169T_Strep
			Rv	ClpA_removeLinker_Rv	agatccugaaccCGTGCCATGAGAGATAAAG	

All ClpA constructs used had the M169T mutation and the 9AA C-terminal deletion to aid in overexpression, here-on it can be assumed that when any ClpA construct is mentioned it has these changes.

Proteins:

For all ClpA constructs, the *E. coli* BL21-AI chemically competent cells strain was used. For RepA-mNG, the *E. coli* strains: SoluBL21 electrocompetent cells and SG1146a chemically competent cells were used.

A single colony from transformed cells was taken to inoculate 10ml of LB-Antibiotic media and grown overnight at 37°C with shaking. 7.5ml of this starter culture was

added to 300ml of LB-Antibiotic media, grown at 37°C with shaking for 2-3 hours until the OD600 measurement was between 0.4-0.8. Then these cultures were induced for protein production. 0.1% Arabinose was used for pBAD plasmids, 1mM IPTG for pET24 plasmids, grown overnight at 22°C with shaking. These cultures were centrifuged in 50ml falcon tubes at 8000g for 10 minutes, supernatant discarded. Then the pellet was resuspended in 20ml of lysis buffer (Table 2). One protease inhibitor tablet was added. Next, the cells were sonicated in ice water with settings : 5s on, 5s off, 10 mins total ON time, 22% power. Then they were balanced and centrifuged at 4°C for 40 minutes at 10,000g. The supernatant was incubated with 500µl of Strep-Tactin 4Flow high capacity beads for 20 minutes in a rotating tube mixer at 4°C. Supernatant was loaded onto a gravity flow column, flow-through collected and reloaded onto the column. 2 CVs of wash buffer were added. 200µl of elution buffer was added five times, collecting each elution separately.

The 6His-SUMO-RepA-mNG was expressed in *E. Coli* SoluBL21 and purified similarly. For SUMO cleavage, Nickel beads were added to the supernatant, this was incubated for 1 hour at 4°C. Supernatant was loaded onto a gravity flow column. 1CV of wash buffer was added, the flow through was discarded. A cap was placed on the bottom of the column. 2.5ml of wash buffer was added. Excess Ulp1 protease was added. A cap was placed on the top of the column. This was rotated for 1 hour at room temperature. The elution was collected and concentrated using a 10kDa amicon filter.

A Bradford assay was performed to determine the concentration of the proteins in the elutions. Furthermore, samples from the pellet, flow-through, wash and elution fractions of the purifications were analysed by SDS-PAGE.

ClpP and mNG-ssrA proteins were obtained from colleagues in the Chemical Biology Lab.

Table 2: Buffer Compositions				
Protein:	ClpA/ClpA$_{\Delta N}$		RepA-TwinStrep-mNG (SoluBL21)	
Reagent	Lysis/Wash Buffer	Elution Buffer	Lysis/Wash Buffer	Elution Buffer
HEPES (pH 7.5)	50mM	50mM	50mM	50mM
NaCl	300mM	300mM	300mM	300mM
Desthiobiotin	-	5mM	-	5mM
TCEP	2mM	2mM	2mM	2mM
EDTA	2mM	2mM	1mM	1mM
Protein:	$_{\Delta His}$ClpA		RepA-TwinStrep-mNG (SG1146a)	
Reagent	Lysis/Wash Buffer	Elution Buffer	Lysis/Wash Buffer	Elution Buffer
HEPES (pH 7.5)	50mM	50mM	50mM	50mM
NaCl	300mM	300mM	300mM	300mM
Desthiobiotin	-	5mM	-	5mM
DTT	1mM	2mM	1mM	2mM
EDTA	1mM	1mM	1mM	1mM
Protein:	-		6His-SUMO-RepA-mNG	
Reagent	1x ATP Regen Buffer (pH 7.5)		Lysis Buffer	Wash/Elution buffer
HEPES	50mM		50mM	50mM
NaCl	-		0.3M	0.3M
DTT	1mM		1mM	-
EDTA	0.5mM		1mM	-
TCEP	-		-	2mM
Imidiazole	-		-	20mM
$MgCl_2$	10mM		-	-
Creatine Kinase	4µM		-	-
Creatine Phosphate	16mM		-	-

Table 3: Protein Sequences	
Protein	Amino acid sequence (N->C)
RepA-TwinStrep-mNG 37.72 kDa	MNQSFISDILYADIESKAKELTVNSNNTVQPVALMRLGVFVPKPSKSKGESKEIDAT KAFSQLEIAKAEGSSSSAWSHPQFEKGGGSGGGSGGSAWSHPQFEKSGGGGVSKGE EDNMASLPATHELHIFGSINGVDFDMVGQGTGNPNDGYEELNLKSTKGDLQFSP WILVPHIGYGFHQYLPYPDGMSPFQAAMVDGSGYQVHRTMQFEDGASLTVNYRY TYEGSHIKGEAQVKGTGFPADGPVMTNSLTAADWCRSKKTYPNDKTIISTFKWSYT TGNGKRYRSTARTTYTFAKPMAANYLKNQPMYVFRKTELKHSKTELNFKEWQKA FTDVMGMDELYK
6His-ClpA$_{\Delta C}$-M169T-Strep 85.61 kDa (Monomer) =513.66 kDa (Hexamer)	MKHHHHHHGSGLNQELELSLNMAFARAREHRHEFMTVEHLLLALLSNPSAREA LEACSVDLVALRQELEAFIEQTTPVLPASEEERDTQPTLSFQRVLQRAVFHVQSSGR NEVTGANVLVAIFSEQESQAAYLLRKHEVSRLDVVNFISHGTRKDEPTQSSDPGSQP NSEEQAGGEERTENFTTNLNQLARVGGIDPLIGREKELERAIQVLCRRRKNNPLLV GESGVGKTAIAEGLAWRIVQGDVPEVMADCTIYSLDIGSLLAGTKYRGDFEKRFKA LLKQLEQDTNSILFIDEIHTIIGAGAASGGQVDAANLIKPLLSSGKIRVIGSTTYQEFS NIFEKDRALARRFQKIDITEPSIEETVQIINGLKPKYEAHHDVRYTAKAVRAAVELAV KYINDRHLPDKAIDVIDEAGARARLMPVSKRKKTVNVADIESVVARIARIPEKSVSQ SDRDTLKNLGDRLKMLVFGQDKAIEALTEAIKMARAGLGHEHKPVGSFLFAGPTG VGKTEVTVQLSKALGIELLRFDMSEYMERHTVSRLIGAPPGYVGFDQGGLLTDAVI KHPHAVLLLDEIEKAHPDVFNILLQVMDNGTLTDNNGRKADFRNVVLVMTTNAG VRETERKSIGLIHQDNSTDAMEEIKKIFTPEFRNRLDNIIWFDHLSTDVIHQVVDKFI VELQVQLDQKGVSLEVSQEARNWLAEKGYDRAMGARPMARVIQDNLKKPLANE LLFGSLVDGGQVTVALDKEKNELTYGFQSAGSSWSHPQFEK
Strep-ClpA$_{\Delta N\Delta C}$-M169T 66.31 kDa (Monomer) = 397.86 kDa (Hexamer)	MWSHPQFEKGSGGGEERTENFTTNLNQLARVGGIDPLIGREKELERAIQVLCRRRK NNPLLVGESGVGKTAIAEGLAWRIVQGDVPEVMADCTIYSLDIGSLLAGTKYRGDF EKRFKALLKQLEQDTNSILFIDEIHTIIGAGAASGGQVDAANLIKPLLSSGKIRVIGST TYQEFSNIFEKDRALARRFQKIDITEPSIEETVQIINGLKPKYEAHHDVRYTAKAVRA AVELAVKYINDRHLPDKAIDVIDEAGARARLMPVSKRKKTVNVADIESVVARIARIP EKSVSQSDRDTLKNLGDRLKMLVFGQDKAIEALTEAIKMARAGLGHEHKPVGSFL FAGPTGVGKTEVTVQLSKALGIELLRFDMSEYMERHTVSRLIGAPPGYVGFDQGGL LTDAVIKHPHAVLLLDEIEKAHPDVFNILLQVMDNGTLTDNNGRKADFRNVVLV MTTNAGVRETERKSIGLIHQDNSTDAMEEIKKIFTPEFRNRLDNIIWFDHLSTDVIH QVVDKFIVELQVQLDQKGVSLEVSQEARNWLAEKGYDRAMGARPMARVIQDNL KKPLANELLFGSLVDGGQVTVALDKEKNELTYGFQSA
ClpA$_{\Delta L\Delta C}$- M169T- Strep 81.82 kDa (Monomer) = 490.92 kDa (Hexamer)	MLNQELELSLNMAFARAREHRHEFMTVEHLLLALLSNPSAREALEACSVDLVALR QELEAFIEQTTPVLPASEEERDTQPTLSFQRVLQRAVFHVQSSGRNEVTGANVLVAIF MSEQESQAAYLLRKHEVSRLDVVNFISHGTGSGSTENFTTNLNQLARVGGIDPLIG REKELERAIQVLCRRRKNNPLLVGESGVGKTAIAEGLAWRIVQGDVPEVMADCTIY SLDIGSLLAGTKYRGDFEKRFKALLKQLEQDTNSILFIDEIHTIIGAGAASGGQVDA ANLIKPLLSSGKIRVIGSTTYQEFSNIFEKDRALARRFQKIDITEPSIEETVQIINGLKPK YEAHHDVRYTAKAVRAAVELAVKYINDRHLPDKAIDVIDEAGARARLMPVSKRK KTVNVADIESVVARIARIPEKSVSQSDRDTLKNLGDRLKMLVFGQDKAIEALTEAIK MARAGLGHEHKPVGSFLFAGPTGVGKTEVTVQLSKALGIELLRFDMSEYMERHTV SRLIGAPPGYVGFDQGGLLTDAVIKHPHAVLLLDEIEKAHPDVFNILLQVMDNGTL TDNNGRKADFRNVVLVMTTNAGVRETERKSIGLIHQDNSTDAMEEIKKIFTPEFRN

	RLDNIIWFDHLSTDVIHQVVDKFIVELQVQLDQKGVSLEVSQEARNWLAEKGYDR AMGARPMARVIQDNLKKPLANELLFGSLVDGGQVTVALDKEKNELTYGFQSAGSS WSHPQFEK
ClpP-Strep 24.43 kDa (Monomer) = 342.02 kDa (Tetradecamer)	MSYSGERDNFAPHMALVPMVIEQTSRGERSFDIYSRLLKERVIFLTGQVEDHMANLI VAQMLFLEAENPEKDIYLYINSPGGVITAGMSIYDTMQFIKPDVSTISMGQAASMG AFLLTAGAKGKRFSLPNSRVMIHQPLGGYQGQATDIEIHAREILKVKGRMNELMA LHTGQSLEQIERDTERDRFLSAPEAVEYGLVDSILTHRNGSSWSHPQFEK
6His-GFP-ssrA 29.15 kDa	MGHHHHHHSSASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLK FICTTGKLPVPWPTLVTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKD DGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQ KNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPN EKRDHMVLLEFVTAAGITHGMDELYKAANDENYALAA
GFP-Strep 28.17 kDa	MASKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKLPV PWPTLVTTFSYGVQCFSRYPDHMKRHDFFKSAMPEGYVQERTIFFKDDGNYKTRA EVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNF KIRHNIEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVL LEFVTAAGITHGMDELYKAAWSHPQFEK
Inv.ssrA-10K-GFP-Strep 31.34 kDa	MAALAYNEDNAASGSGKKKKKKKKKKGSGSGSASKGEELFTGVVPILVELDGDV NGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFSYGVQCFSRYPDHM KRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDG NILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPI GDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYKGSGW SHPQFEK
D-inv.ssrA-10K-GFP-Strep 31.38 kDa	MDALAYNEDNAASGSGKKKKKKKKKKGSGSGSASKGEELFTGVVPILVELDGDV NGHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFSYGVQCFSRYPDHM KRHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDG NILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPI GDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYKGSGW SHPQFEK
E-inv.ssrA-10K-GFP-Strep 31.4 kDa	MEALAYNEDNAASGSGKKKKKKKKKKGSGSGSASKGEELFTGVVPILVELDGDVN GHKFSVSGEGEGDATYGKLTLKFICTTGKLPVPWPTLVTTFSYGVQCFSRYPDHMK RHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVKFEGDTLVNRIELKGIDFKEDGN ILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIEDGSVQLADHYQQNTPIGD GPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITHGMDELYKGSGWSHP QFEK
6His-SUMO-RepA-mNG-Strep 50.86 kDa (Full size) = 37.16 kDa (After cleavage) ↑ = Ulp1 cleavage site	MGSSHHHHHHSSGLVPRGSHMASMSDSEVNQEAKPEVKPEVKPETHINLKVSDG SSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDII EAHREQIGG↑NQSFISDILYADIESKAKELTVNSNNTVQPVALMRLGVFVPKPSKSK GESKEIDATKAFSQLEIAKAEGSSSSGGGGVSKGEEDNMASLPATHELHIFGSINGV DFDMVGQGTGNPNDGYEELNLKSTKGDLQFSPWILVPHIGYGFHQYLPYPDGMSP FQAAMVDGSGYQVHRTMQFEDGASLTVNYRYTYEGSHIKGEAQVKGTGFPADGP VMTNSLTAADWCRSKKTYPNDKTIISTFKWSYTTGNGKRYRSTARTTYTFAKPMAA NYLKNQPMYVFRKTELKHSKTELNFKEWQKAFTDVMGMDELYKGGGGSSGSGSS GSSGGGSAGSAWSHPQFEK

Fluorescence degradation assay:

Fluorescence degradation was measured using a CLARIOstar Plus plate reader. Reaction mixtures contained 250nM ClpA, 250nM ClpP, 66.6nM substrate, 1mM ATP, 1x ATP Regen Buffer, 0.2 or 0.5M KCl (as indicated where necessary) and had a total volume of 166µl. Reactions were carried out at 37°C. mNG was measured with excitation at 470nm wavelength and emission at 515nm wavelength. Fluorescence was measured for 10 minutes in total. ATP was added at t=33s. Normalization was performed so that the value of the fluorescence signal at t=33s is set to 1.

Results

RepA-mNG construct not recognised by ClpA nor ClpA$_{\Delta N}$

A degradation assay was done to see if ClpA and ClpA$_{\Delta N}$ would have activity for the substrate RepA-mNG (Fig. 2). The RepA-TwinStrep-mNG construct used was purified from the *E. coli* SoluBL21 strain. No fluorescence degradation was observed at all. This may be due to an issue with the substrate, because when it was run on the SDS PAGE gel the RepA-TwinStrep-mNG protein appeared below the 35kDa band of the ladder, yet its weight is expected to be 37.7 kDa (Fig. 3).

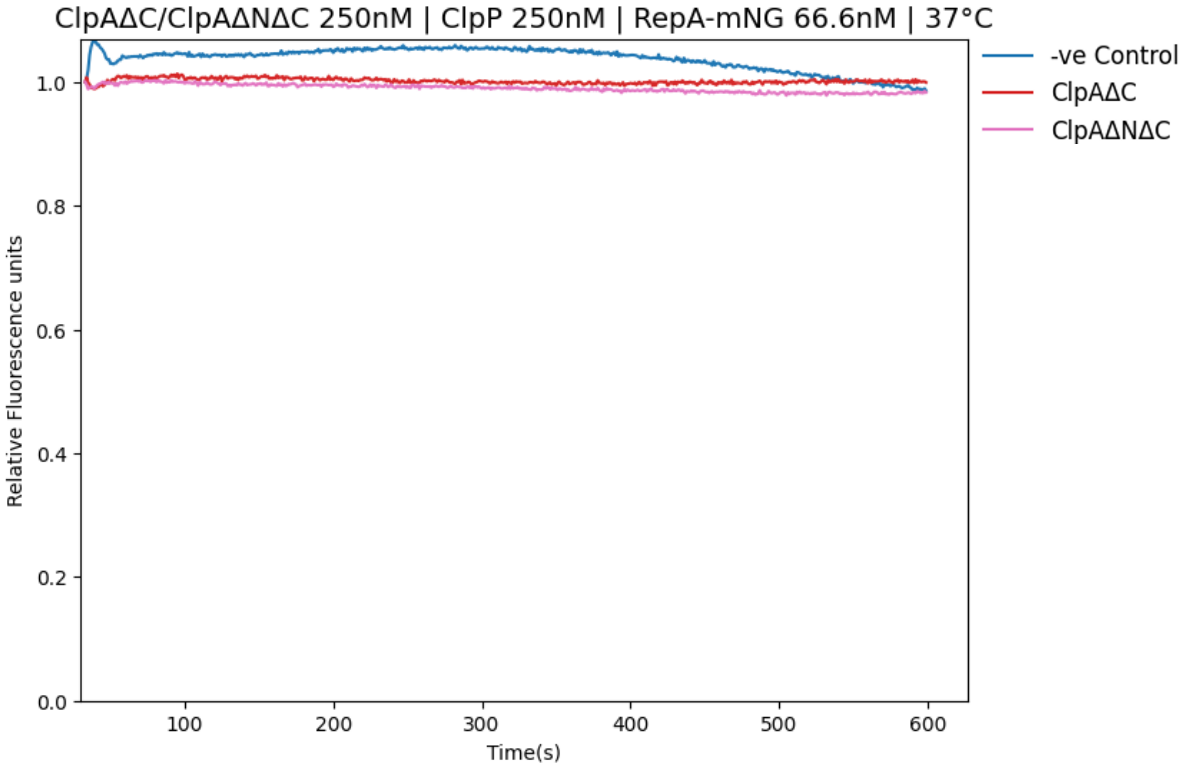

Fig. 2: Fluorescence degradation assay with 250nM ClpA/ClpA$_{\Delta N}$, 250nM ClpP, 66.6nM RepA-mNG, 0.2M KCl, 1mM ATP, 1x ATP regen at 37°C. The negative control contained all components except the unfoldase.

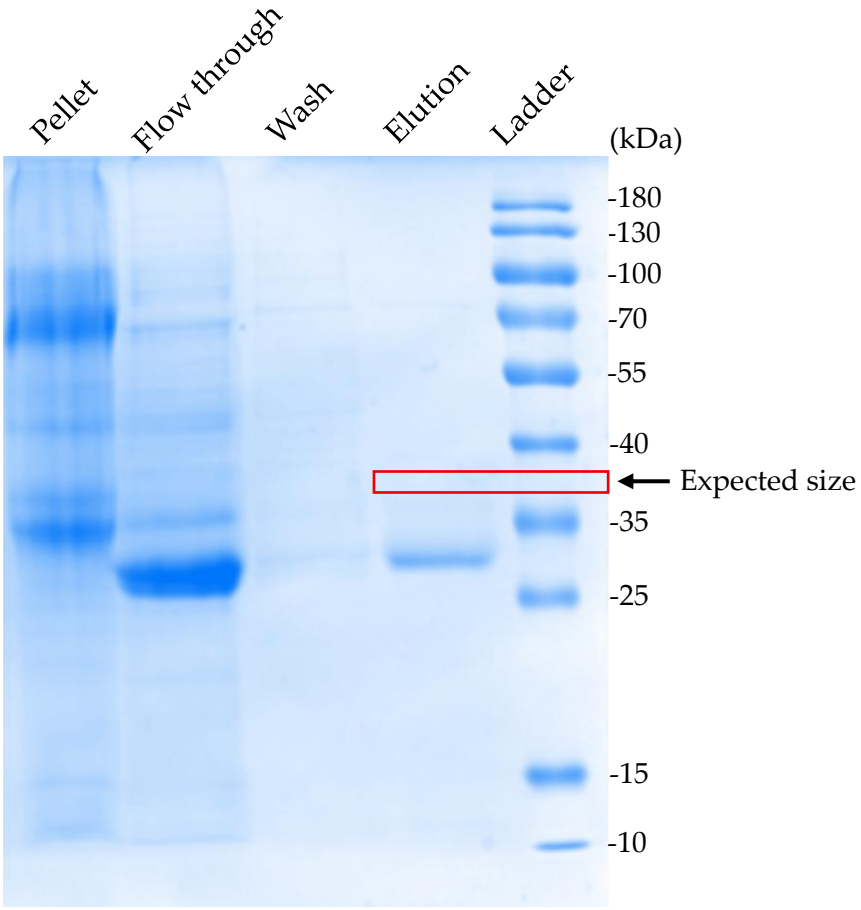

Fig. 3: SDS-PAGE gel of RepA-TwinStrep-mNG purification. The ladder is PageRuler™ Prestained Protein Ladder. The expected size of RepA-TwinStrep-mNG (~38kDa) is shown.

ClpA$_{\Delta N}$ shows higher activity than ClpA for an ssrA-tagged substrate

We also tested the substrate mNG-ssrA. Unfolding activity was observed for both ClpA and ClpA$_{\Delta N}$ (Fig. 4). ClpA$_{\Delta N}$ degraded >90% of the substrate by 10 minutes. In comparison, ClpA only degraded ~30% of the substrate by 10 minutes. This shows that the unfoldase was working, and that ClpA$_{\Delta N}$ had higher activity towards mNG-ssrA than ClpA.

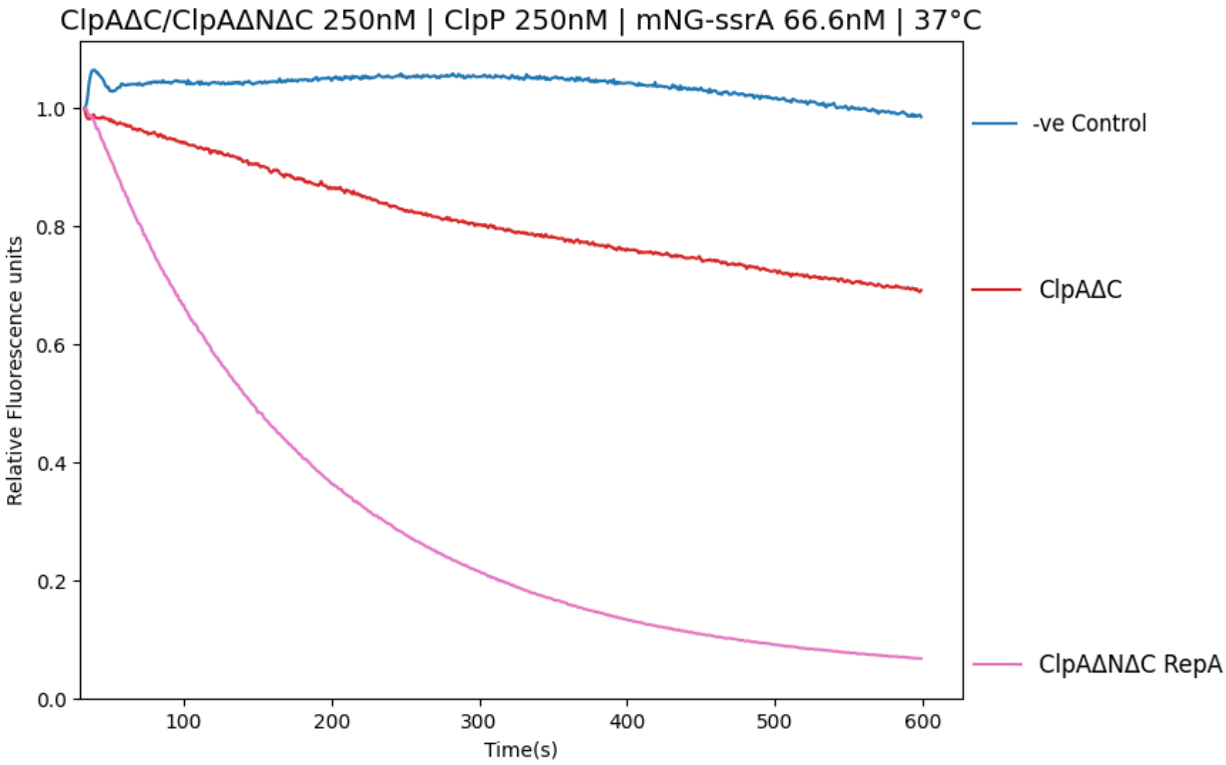

Fig. 4: Fluorescence degradation assay with 250nM ClpA/ClpA$_{\Delta N}$, 250nM ClpP, 66.6nM mNG-ssrA, 0.2M KCl, 1mM ATP, 1x ATP regen at 37°C. The negative control contained all components except the unfoldase.

ClpA and ClpA$_{\Delta N}$ show no activity towards varying concentrations of RepA-mNG

To test if the concentration of the RepA-mNG substrate was the reason no activity was observed, a fluorescence degradation assay with varying concentrations of substrate was performed for ClpA and ClpA$_{\Delta N}$ (Fig. 5). No activity was observed for any concentration of RepA-mNG for either unfoldase. The positive control ClpA$_{\Delta N}$ with mNG-ssrA showed a ~50% decrease in fluorescence. This differs to the >90% decrease in fluorescence that can be observed in Fig. 4. This is likely because the unfoldase loses activity when stored for longer periods; it was not made fresh for this experiment.

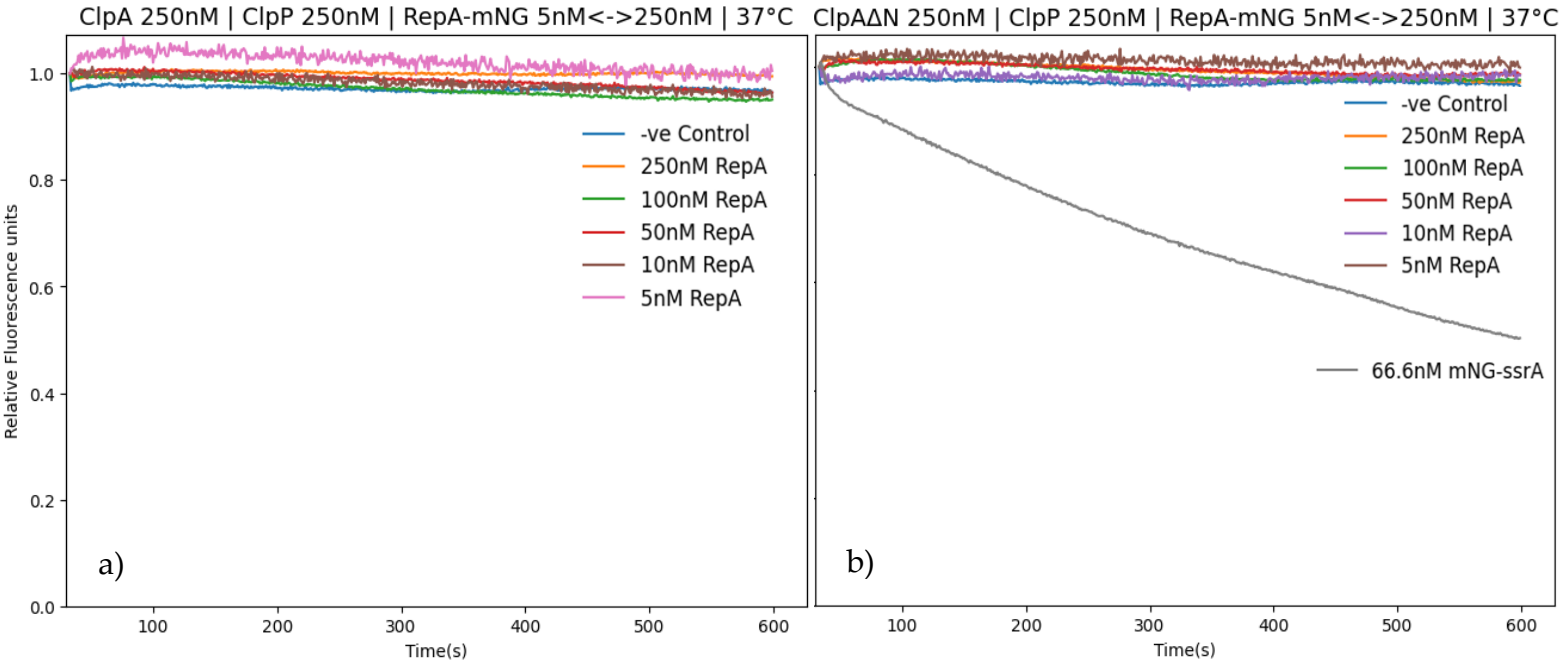

Fig. 5: Fluorescence degradation assays with (a) 250nM ClpA or (b) 250nM ClpA$_{\Delta N}$, 250nM ClpP, 5nM<->250nM RepA-mNG, 0.2M KCl, 1mM ATP, 1xATP regen at 37°C. The negative controls contained all components except the unfoldase. For (b), 66.6nM of mNG-ssrA was used as a positive control.

$_{\Delta His}$ClpA shows slight activity towards RepA-mNG obtained from two different strains

The N-domain of ClpA was proposed to have an accessory function in RepA recognition (Cranz-Mileva et al., 2008). The ClpA construct used previously had a 6His tag on its N-terminal which may have interfered with substrate recognition. To determine if the 6His tag on ClpA was the reason no activity was observed, it was removed with USER cloning. Additionally, RepA-TwinStrep-mNG was expressed in the *E. coli* strain SG1146a, this strain that lacks ClpP (Sauciuc et al., 2024). $_{\Delta His}$ClpA was tested in a fluorescence degradation assay with RepA-mNG from the SoluBL21 and SG1146a strains. A slight decrease in fluorescence was observed (~5%) for both RepA purifications in both the 0.2M and 0.5M KCl conditions compared to the negative control (Fig. 6). For the positive control mNG-ssrA, there was a ~35%

decrease in fluorescence in 0.2M and 0.5M KCl conditions, which appears to be a slightly bigger decrease than for ClpA in Fig. 4. To determine if this is a significant change, more replicates would need to be done to see if the difference lies within the margin of error or not.

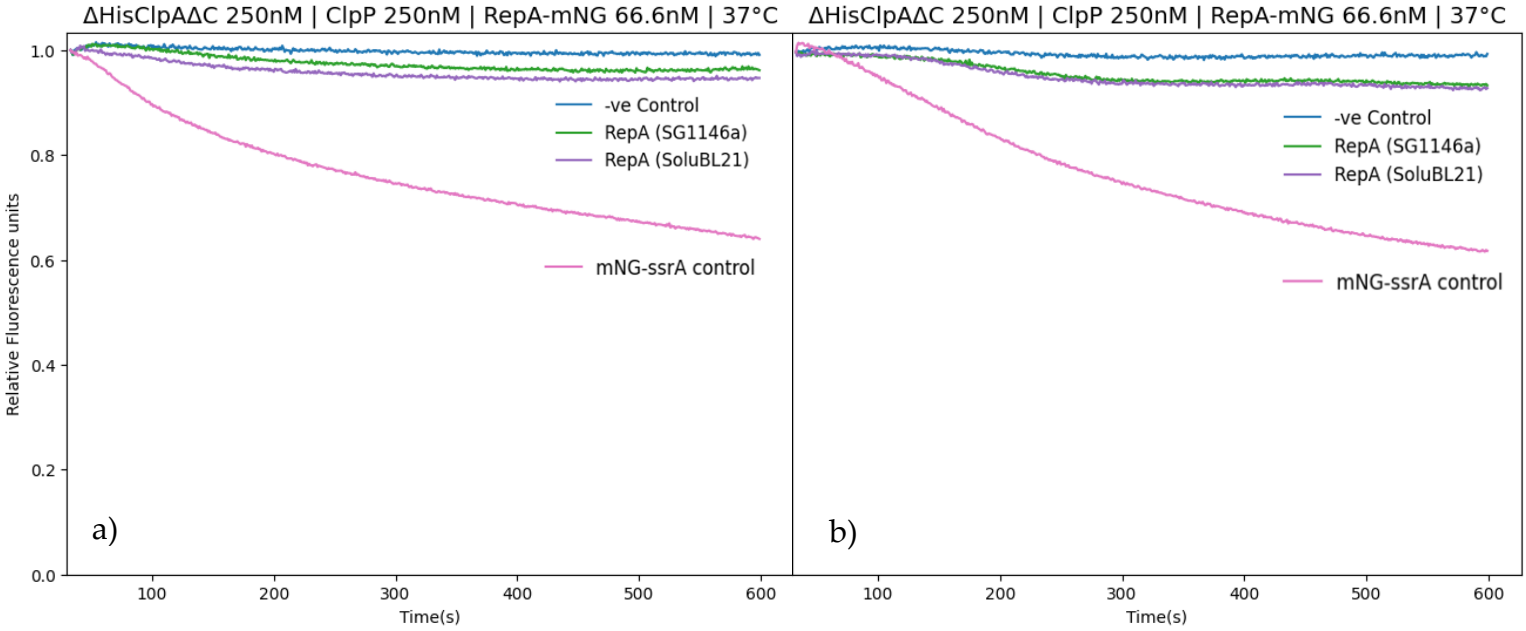

Fig. 6: Fluorescence degradation assay with 250nM $_{\Delta His}$ClpA, 250nM ClpP, 66.6nM RepA-mNG (from SG1146a/SoluBL21 strains), (a) 0.2M KCl or (b) 0.5M KCl, 1mM ATP, 1x ATP regen at 37°C. The negative controls contained all components except the unfoldase.

RepA-mNG from SG1146a runs lower than expected on SDS-PAGE gel

An SDS PAGE gel was done, comparing RepA-mNG from each strain with mNG-ssrA (Fig. 7). The RepA-mNG from SG1146a appeared slightly higher up on the gel compared to RepA-mNG from SoluBL21 (likely an insignificant difference), however it still ran below the 35kDa band of the ladder. A faint band can be seen above the 35kDa band which may be the full-size construct. Both of the RepA constructs appeared at a similar level to the mNG-ssrA band even though there is supposed to be a 10kDa difference between them.

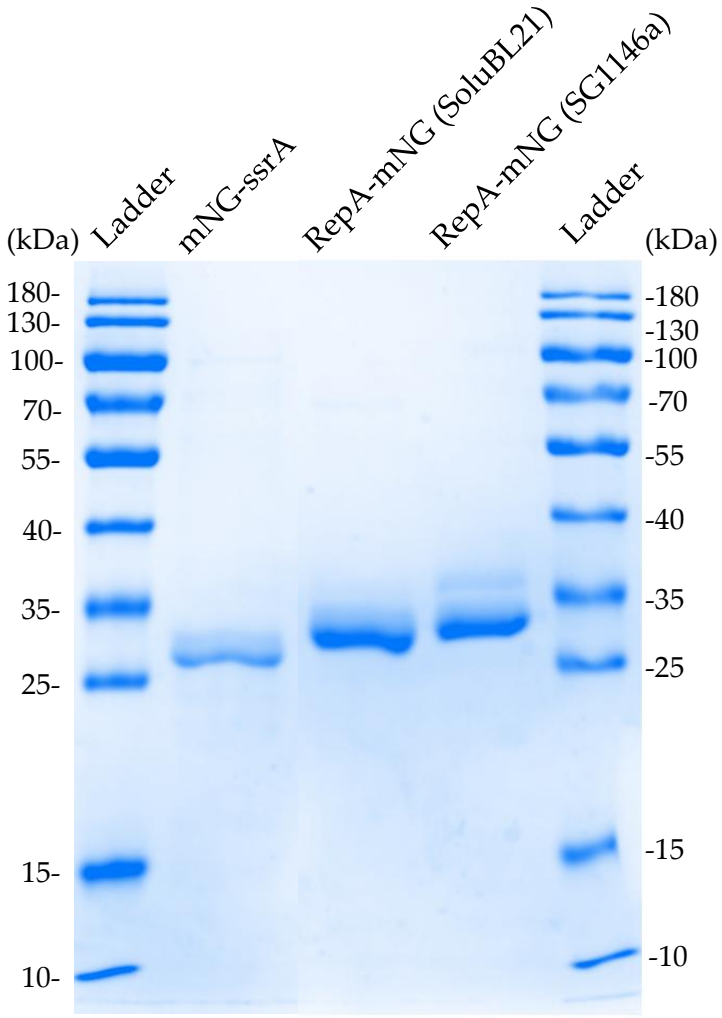

Fig. 7: SDS-PAGE gel comparing mNG-ssrA, RepA-mNG-Strep and RepA-TwinStrep-mNG from SoluBL21 and SG1146a *E. coli* strains. The ladder is PageRuler™ Prestained Protein Ladder.

Purification of full-length RepA-mNG via SUMO-tag cleavage

A His-SUMO tag was added to the N-terminus of RepA-mNG with USER cloning. During purification, while the fusion protein was bound to the Ni^{2+} beads, Ulp1 protease was added to cleave the SUMO tag. Ulp1 cleaves precisely at the junction between the SUMO tag and the target protein, so the eluted product should be the full-length protein. Samples from the purification were ran on an SDS-PAGE gel (Fig. 8). The bands from the elutions appear to be at the correct size of ~37kDa. They are clearly higher than the band for RepA-TwinStrep-mNG.

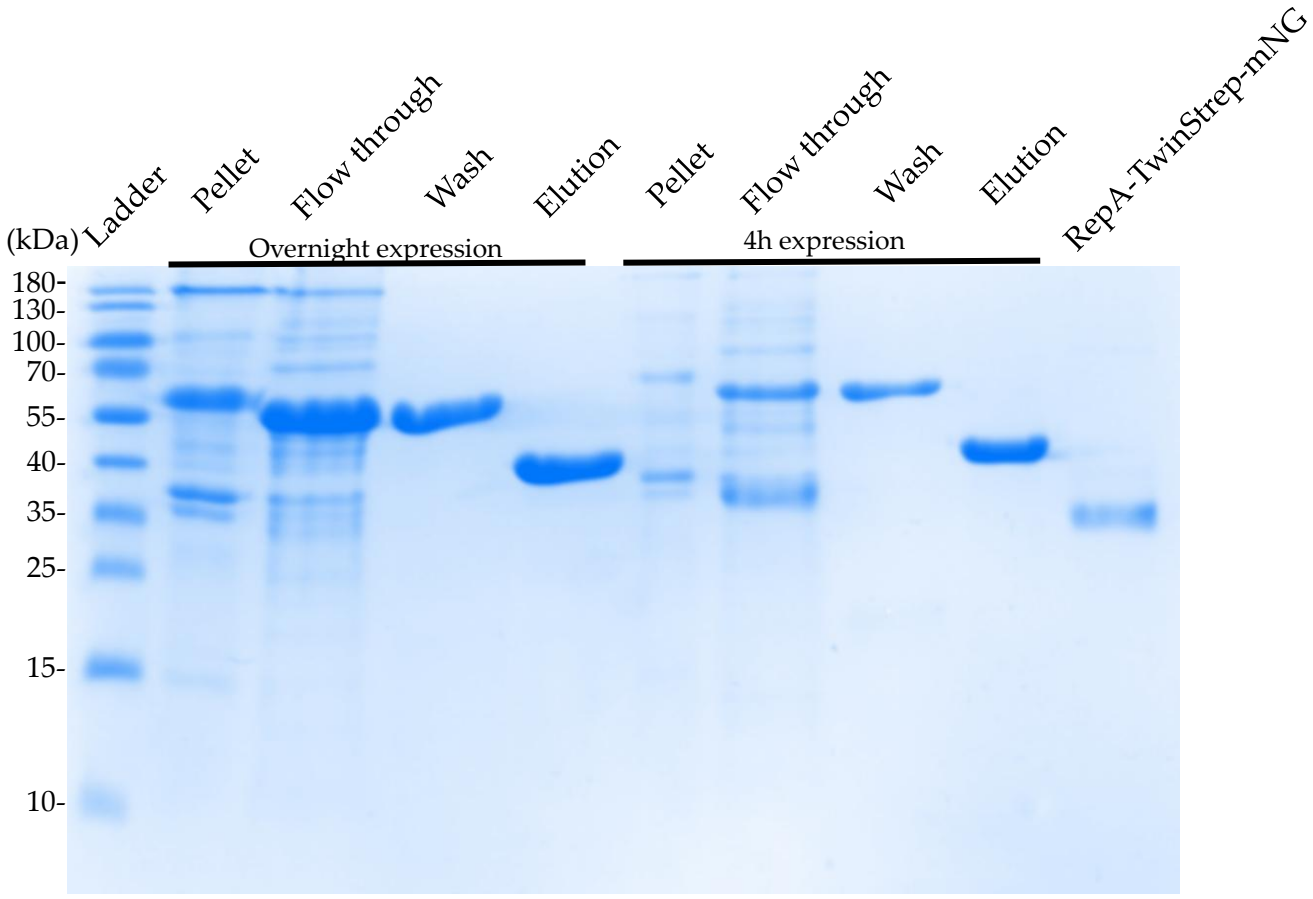

Fig. 8: SDS-PAGE gel of two SUMO-RepA-mNG purifications. One was expressed overnight, the other for 4 hours. The last lane has RepA-TwinStrep-mNG. The ladder is PageRuler™ Prestained Protein Ladder.

SUMO-derived RepA-mNG shows minor fluorescence degradation

The RepA-mNG obtained from the SUMO construct was compared with the RepA-mNG obtained from before in a fluorescence degradation assay with $_{\Delta His}$ClpA and ClpA$_{\Delta N}$ (Fig. 9). ClpA with RepA-mNG (SUMO) showed a ~10% decrease in fluorescence. Fluorescence for ClpA$_{\Delta N}$ with either RepA substrate didn't differ significantly from the negative control. Surprisingly, fluorescence for $_{\Delta His}$ClpA with RepA-mNG from SG1146a increased slightly compared to the negative control, which differs to what was observed in Fig. 6.

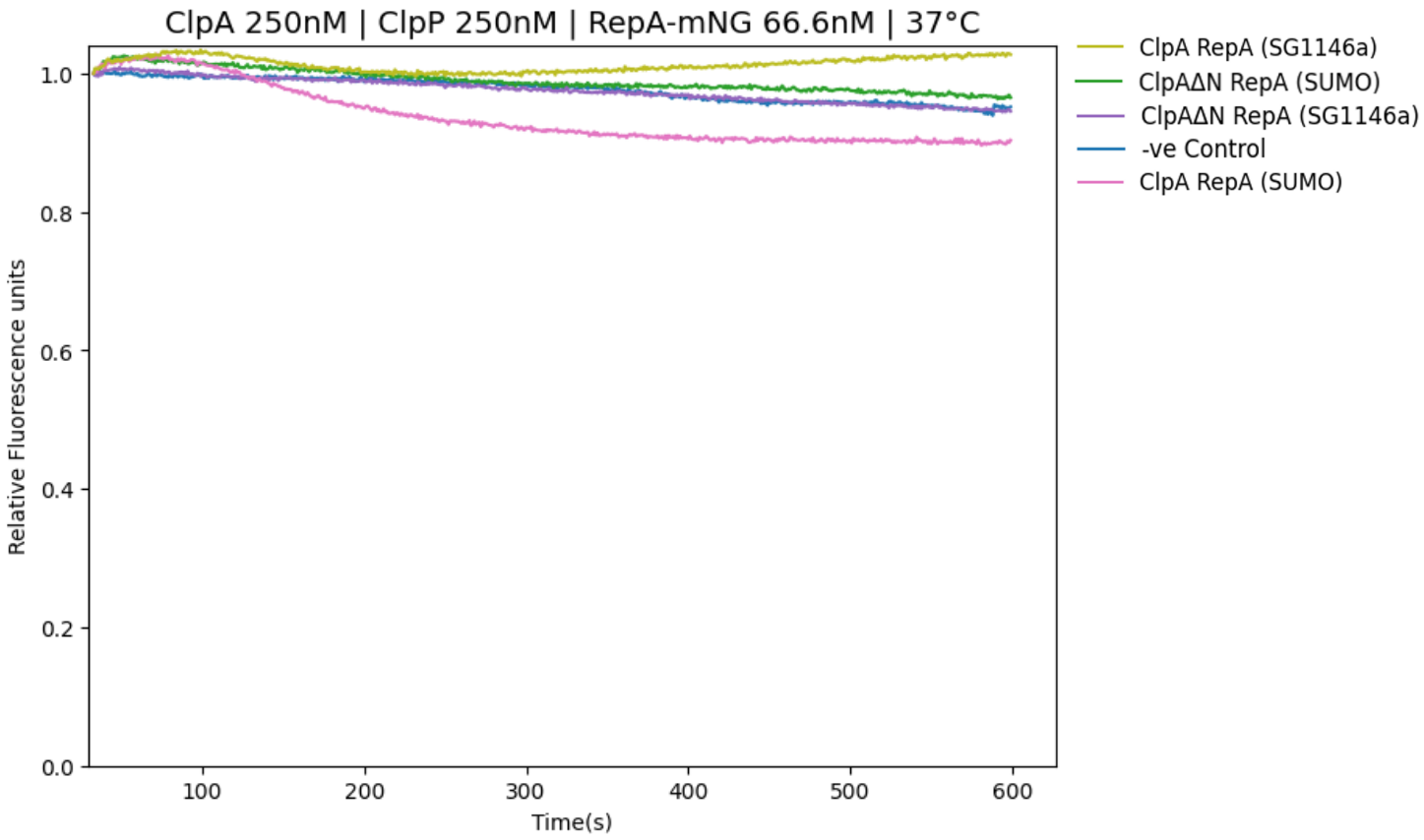

Fig. 9: Fluorescence degradation assay with 250nM $_{\Delta His}$ClpA/ClpA$_{\Delta N}$, 250nM ClpP, 66.6nM 66.6nM RepA-mNG (SUMO)/RepA-mNG (SG1146a), 0.2 KCl, 1mM ATP, 1x ATP regen at 37°C. The negative control contained all components except the unfoldase.

ClpA$_{\Delta L}$ shows similar activity to ClpA$_{\Delta N}$ for an ssrA-tagged substrate

We removed the linker of the N-domain of our $_{\Delta His}$ClpA construct with USER cloning, because previous research suggested that doing so lowered its K_m for the two substrates measured (Cranz-Mileva et al., 2008). We compared $_{\Delta His}$ClpA$_{\Delta L}$ with $_{\Delta His}$ClpA and ClpA$_{\Delta N}$ in a fluorescence degradation assay with the substrates mNG-ssrA and RepA-mNG (SUMO) (Fig. 10a and b). Fig. 10a shows that ClpA$_{\Delta L}$ has almost as much activity as ClpA$_{\Delta N}$ towards mNG-ssrA. Fig. 10b shows that ClpA

and ClpA$_{\Delta L}$ have similar activity (~10% fluorescence degradation) for RepA-mNG (SUMO).

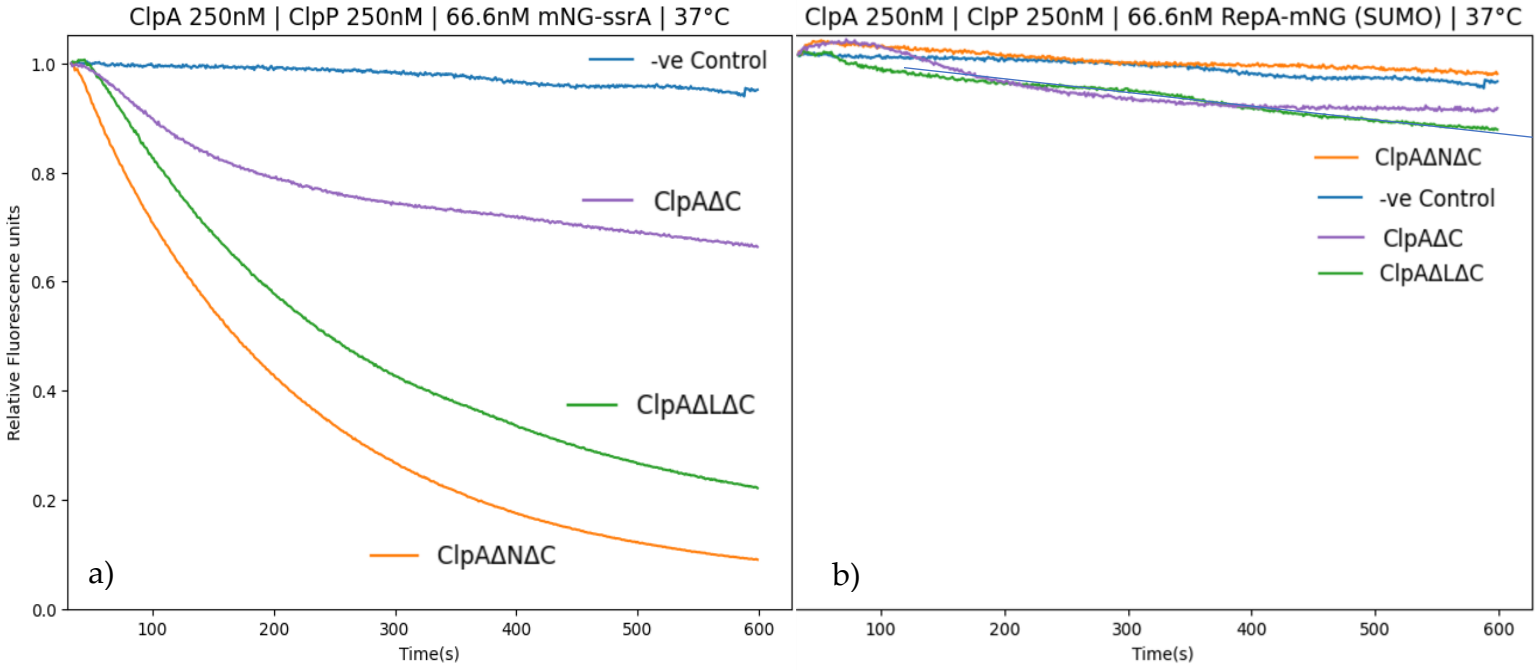

Fig. 10: Fluorescence degradation assay with 250nM $_{\Delta His}$ClpA/$_{\Delta His}$ClpA$_{\Delta L}$/ClpA$_{\Delta N}$, 250nM ClpP, (a) 66.6nM mNG-ssrA or (b) 66.6nM RepA-mNG (SUMO), 0.2M KCl, 1mM ATP, 1x ATP regen at 37°C. The negative controls contained all components except the unfoldase.

ClpA$_{\Delta L}$ without ClpP shows no activity for ssrA substrate

We decided to test ClpA$_{\Delta L}$ without ClpP to see if a decrease in fluorescence could still be observed (Fig. 11). We expected to see at least a partial decrease in fluorescence since the unfoldase should still be able to unfold the substrate, resulting in a loss of fluorescence, until it refolds. These results showed that the unfoldase did not work at all without ClpP.

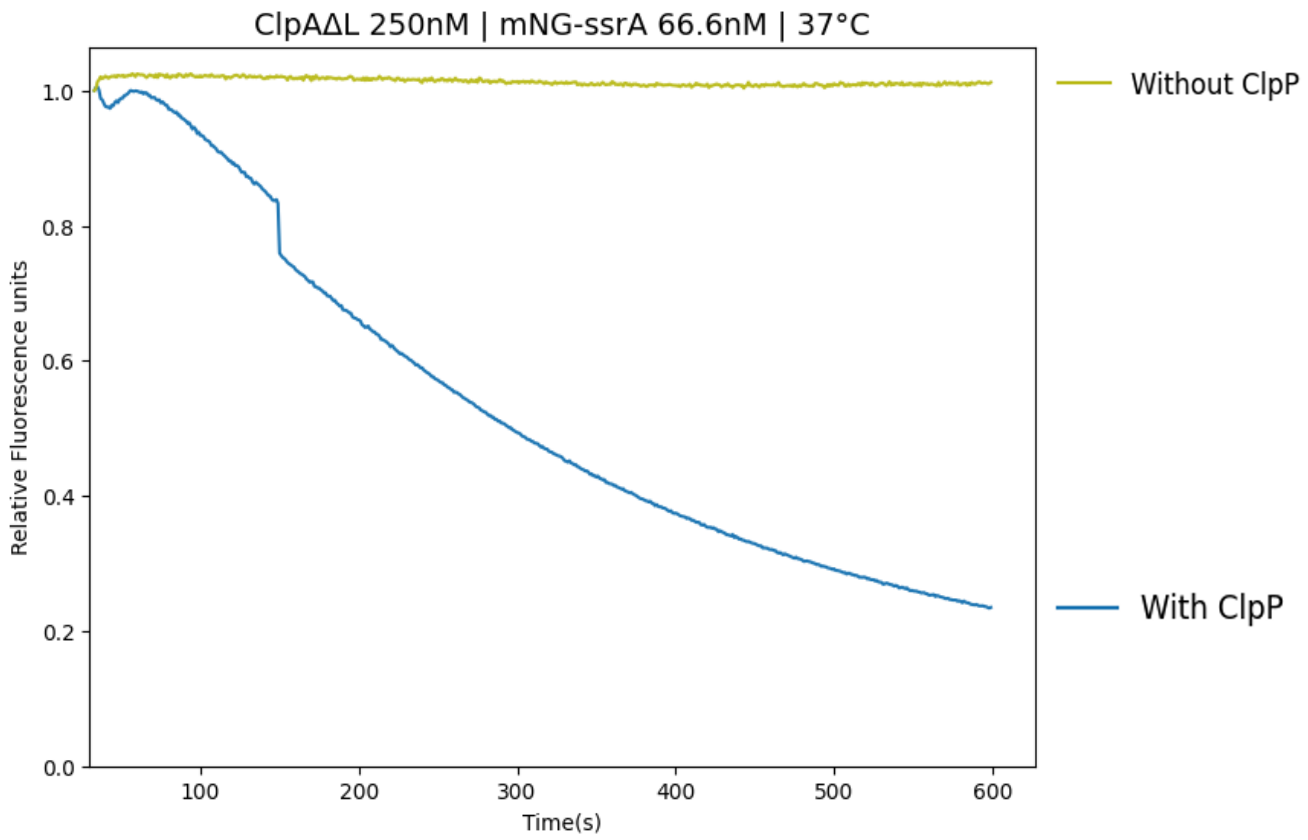

Fig. 11: Fluorescence degradation assay with 250nM $_{\Delta His}$ClpA$_{\Delta L}$, 250nM ClpP, 66.6nM mNG-ssrA, 0.2M KCl, 1mM ATP, 1x ATP regen at 37°C.

ClpA$_{\Delta N}$ shows little activity towards various ssrA-GFP constructs

SsrA is normally added to the C-terminus of proteins so they can be recognised by ClpA/ClpX (Hoskins, Kim, et al., 2000). To determine if ssrA could work as a tag on the N-terminus, we did a fluorescence degradation assay for ClpA$_{\Delta N}$ with three different N-terminally tagged ssrA substrates that were provided to us by a colleague (see Table 3). There was no decrease in fluorescence for the negative control (GFP-ssrA with no unfoldase) (Fig. 12). There was a >90% decrease in fluorescence for GFP-ssrA, the positive control, mirroring what was observed in Fig. 4. The three ssrA-GFP constructs and untagged GFP all showed slight decreases in fluorescence (~10%). This may be a result of some non-specific activity by ClpA for these constructs.

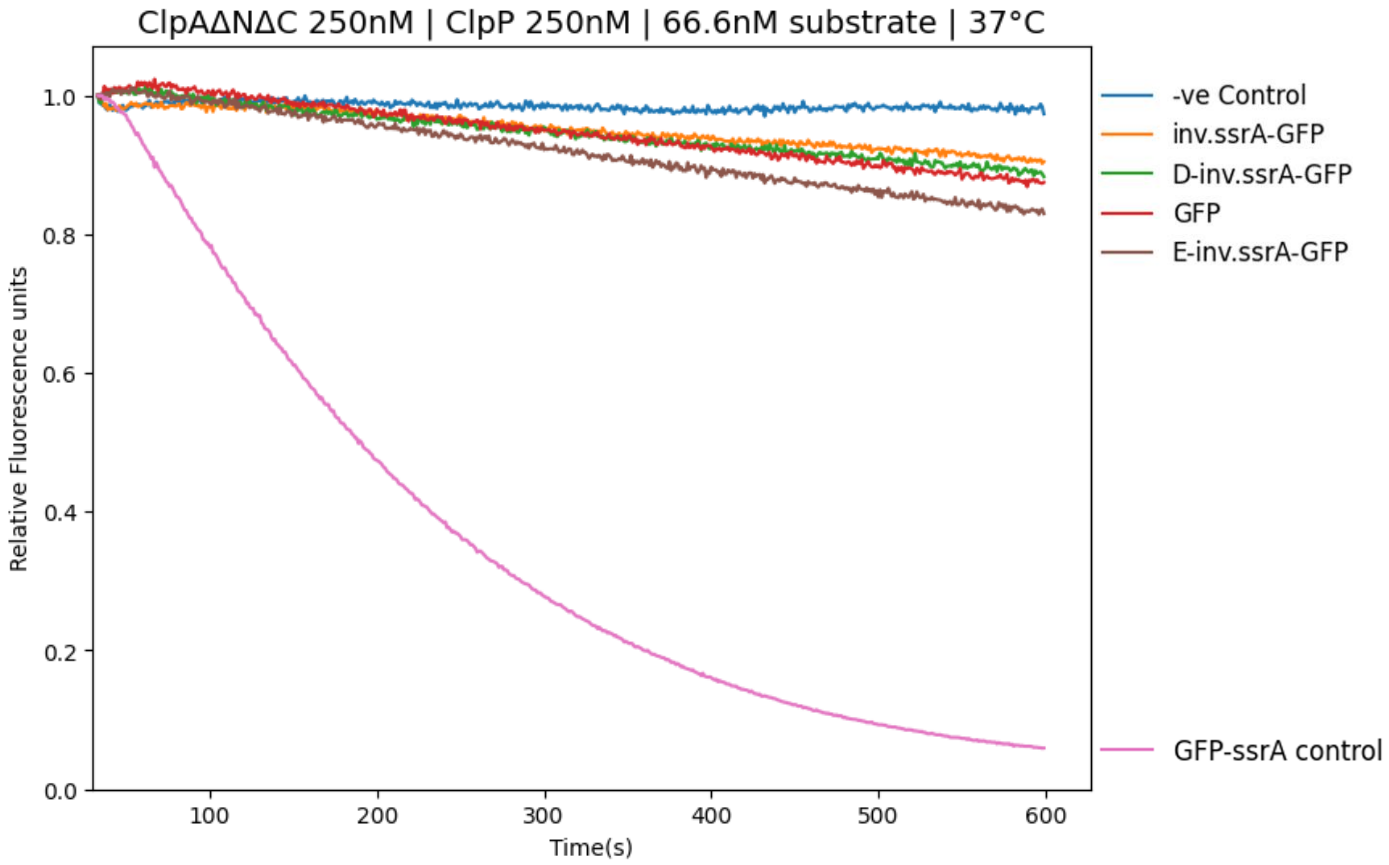

Fig. 12: Fluorescence degradation assay with 250nM ClpA$_{\Delta N}$, 250nM ClpP, 66.6nM substrate, 0.5M KCl, 1mM ATP, 1xATP regen, 37°C. The negative control contained all components except the unfoldase.

Discussion

In this project, our goal was to purify a RepA-mNG construct and then measure its unfolding with the unfoldase ClpA in a fluorescence degradation assay. Ultimately, we were successful in purifying the full-length RepA-mNG fusion protein, however it could not be degraded by more than 10% (Fig. 10b).

Optimising RepA-mNG purification

The initial assays showed no ClpA activity for the RepA-mNG substrate (Fig. 2 and 5). This was unexpected since previous studies demonstrated ClpA's ability to degrade RepA-tagged substrates Hoskins et al., 2000.

One possible explanation for this lies in our RepA-mNG substrate. In Fig. 3 it ran ~7kDa lower than expected. We suspected that the RepA portion of the fusion protein was degraded by the cells' endogenous ClpAP, which would make sense as the RepA tag used has a molecular weight of 7.62kDa. Without this degradation tag on the substrate ClpA would not be able to recognise it. Thus, we decided to express RepA-mNG in *E. Coli* SG1146a (a strain lacking ClpP (Sauciuc et al., 2024)), however as seen in Fig. 7 this did not affect the size of the fusion protein. This indicates that the size discrepancy is not due to degradation by the endogenous ClpAP complex. Instead, it suggests that a peptidase cleavage site may exist near the C-terminus of the RepA segment, leading to truncation of the fusion protein.

Subsequently, we tried to purify RepA-mNG by using a construct with a 6His-SUMO tag on the N-terminus, since it would allow us to selectively purify the un-cleaved fusion proteins (as the cleaved mNG constructs wouldn't have the 6His tag to facilitate binding to the Ni^{2+} beads). As shown in Fig. 8, this approach successfully yielded the full-length RepA-mNG, indicating that not all of the substrates get cleaved during expression.

6His tag of ClpA interfering with substrate recognition

At the same time, we also considered whether the unfoldase was contributing to the lack of fluorescence degradation. Cranz-Mileva et al., 2008 highlighted that the N-domain of ClpA played an accessory role in RepA degradation; our ClpA construct had a 6His tag on its N-terminal so we speculated that it could be interfering with substrate recognition. Fig. 6 shows that $_{\Delta His}$ClpA had some activity, confirming that the 6His tag was probably interfering previously. The sigmoidal shape of the curves suggests heterogeneity of the RepA-mNG from SoluBL21/ SG1146a strains: an initial decline reflects degradation of the full-length RepA-mNG, followed by a plateau as only the unrecognizable truncated substrates remain. The faint band at ~37kDa for RepA-mNG (SG1146a) in Fig. 7 may be the full-length substrate, which would support this hypothesis. This faint band may have been more apparent for RepA-mNG (SoluBL21) if a higher concentration was loaded.

Limitations of mNG as a ClpAP substrate

When testing $_{\Delta His}$ClpA on SUMO-derived RepA-mNG, we expected to see much more fluorescence degradation than what was observed in Fig. 6, however only ~10% fluorescence degradation occurred (Fig. 9). It could be that the concentration of RepA-mNG we used was too high. In the future it's worth doing an assay with varying concentrations of the SUMO-derived RepA-mNG construct (similar to what was done in Fig. 4). However, and unexpectedly, this curve still appears to plateau, similarly to how it did in Fig. 6. If the concentration of substrate was too high, the fluorescence should still decrease steadily, just at a low rate. It doesn't make sense for the curve to plateau at 90% fluorescence. To explain this, we hypothesised that ClpA might not be able unfold mNG from the N->C direction consistently. This would mean that the ClpAP complex degrades the RepA part of the substrate until it reaches mNG, where occasionally it manages to unfold and degrade it, but more often than not it stalls and then dissociates from the substrate. This would explain the curve plateauing; eventually all the RepA tags of the substrate will get degraded,

leaving behind mNG that ClpAP cannot degrade. To test this, the reaction could be performed whilst taking samples for SDS-PAGE at regular intervals. If the ~37kDa band of the RepA-mNG substrate shifts to ~28kDa - the size of mNG itself - then this hypothesis would be supported. It may be hard to detect the presence of these bands due to the presence of ClpP and Creatine Kinase, whose monomers have a molecular weight of ~24 and ~40kDa respectively. To overcome this, you could do a Western blot with anti-mNG antibodies and compare it to a lane with the RepA-mNG substrate on its own.

Modulating ClpA activity via linker deletion

In Fig. 10a ClpA$_{\Delta L}$ had activity almost as high as ClpA$_{\Delta N}$, showing that the presence of the linker of the N-domain makes it harder for ClpA to recognise ssrA-tagged substrates. This can be attributed to the findings by Zuromski et al., 2021 which highlighted that ssrA binds directly to the D1 pore loop of ClpA, so the presence of the linker would make this part less accessible. This also explains why ClpA$_{\Delta N}$ has greater activity than ClpA$_{\Delta L}$, not having the N-domain at all would make the D1 region even more accessible. In Fig. 10b the ClpA$_{\Delta L}$ curve plateaus much less than $_{\Delta His}$ClpA, and it has ~5% more fluorescence degradation. These differences are small but it may indicate that that ClpA$_{\Delta L}$ has greater unfolding capabilities for RepA-mNG. This assay should be repeated for longer time durations to see if the difference becomes clearer.

SsrA on the N-terminus

Results from Fig. 12 indicated that non-specific degradation occurring because the untagged GFP had a similar amount of fluorescence degradation to the three N-terminus ssrA constructs. All four of these proteins had a strep tag on their C-terminus, this may have been what ClpA$_{\Delta N}$ was recognising. The N-terminal ssrA tags may have even been cleaved off the GFP during expression; there was a 10K linker in between which may have been recognised by peptidases as a cleavage site.

ClpA without ClpP

Fig. 11 shows that ClpA$_{\Delta L}$ did not work at all for mNG-ssrA without ClpP, yet Hoskins et al., 2000 showed that ClpA can work without ClpP. This result may be indicating that ClpA is not forming its hexamer without ClpP, perhaps ClpP is a 'docking' site that facilitates ClpA oligomerisation. A sample of ClpA that has been incubated with ATP-γ-S and its substrate (so the substrate can bind without being unfolded) should be analysed with Cryo-EM to determine if ClpA is forming a hexamer without ClpP, which may provide insight on the next steps to be taken.

Outlook and Future directions

Although RepA-mNG was successfully purified, degradation in the fluorescence assay remained low throughout, even after testing different ClpA variants, the most promising being ClpA$_{\Delta L}$. Future work should explore other fluorescent proteins such as GFP. If degradation rates of RepA-tagged substrates improve then the tag can be shortened and the impact of higher salt concentrations (relevant for electrophysiology experiments) can be assessed. Lastly, it is worth testing the efficacy of RepA as a C-terminal degradation tag, as shown by Hoskins et al., 2002. If RepA proves effective on both the N- and C-termini, it would simplify the tagging of unknown proteins in nanopore protein sequencing applications by eliminating the need for separate. Overall, these findings outline clear next steps that, if successful, position ClpA as a viable tool in nanopore-based protein sequencing.

References

Cranz-Mileva, S., Imkamp, F., Kolygo, K., Maglica, Ž., Kress, W., & Weber-Ban, E. (2008). The Flexible Attachment of the N-Domains to the ClpA Ring Body Allows their Use On Demand. *Journal of Molecular Biology*, *378*(2), 412–424. https://doi.org/10.1016/j.jmb.2008.02.047

Flynn, J. M., Levchenko, I., Seidel, M., Wickner, S. H., Sauer, R. T., & Baker, T. A. (2001). Overlapping recognition determinants within the ssrA degradation tag allow modulation of proteolysis. *Proceedings of the National Academy of Sciences of the United States of America*, *98*(19), 10584–10589. https://doi.org/10.1073/pnas.191375298

Hoskins, J. R., Kim, S.-Y., & Wickner, S. (2000). Substrate Recognition by the ClpA Chaperone Component of ClpAP Protease. *Journal of Biological Chemistry*, *275*(45), 35361–35367. https://doi.org/10.1074/jbc.M006288200

Hoskins, J. R., Singh, S. K., Maurizi, M. R., & Wickner, S. (2000). Protein binding and unfolding by the chaperone ClpA and degradation by the protease ClpAP. *Proceedings of the National Academy of Sciences*, *97*(16), 8892–8897. https://doi.org/10.1073/pnas.97.16.8892

Hoskins, J. R., Yanagihara, K., Mizuuchi, K., & Wickner, S. (2002). ClpAP and ClpXP degrade proteins with tags located in the interior of the primary sequence. *Proceedings of the National Academy of Sciences of the United States of America*, *99*(17), 11037–11042. https://doi.org/10.1073/pnas.172378899

Motone, K., Kontogiorgos-Heintz, D., Wee, J., Kurihara, K., Yang, S., Roote, G., Fox, O. E., Fang, Y., Queen, M., Tolhurst, M., Cardozo, N., Jain, M., & Nivala, J. (2024). Multi-

pass, single-molecule nanopore reading of long protein strands. *Nature*, *633*(8030), 662–669. https://doi.org/10.1038/s41586-024-07935-7

Olivares, A. O., Nager, A. R., Iosefson, O., Sauer, R. T., & Baker, T. A. (2014). Mechanochemical basis of protein degradation by a double-ring AAA+ machine. *Nature Structural & Molecular Biology*, *21*(10), 871–875. https://doi.org/10.1038/nsmb.2885

Saibil, H. (2013). Chaperone machines for protein folding, unfolding and disaggregation. *Nature Reviews Molecular Cell Biology*, *14*(10), 630–642. https://doi.org/10.1038/nrm3658

Sauciuc, A., Morozzo della Rocca, B., Tadema, M. J., Chinappi, M., & Maglia, G. (2024). Translocation of linearized full-length proteins through an engineered nanopore under opposing electrophoretic force. *Nature Biotechnology*, *42*(8), 1275–1281. https://doi.org/10.1038/s41587-023-01954-x

Shih, T.-T., Sauer, R. T., & Baker, T. A. (2024). How the double-ring ClpAP protease motor grips the substrate to unfold and degrade stable proteins. *Journal of Biological Chemistry*, *300*(11), 107861. https://doi.org/10.1016/j.jbc.2024.107861

Zuromski, K. L., Kim, S., Sauer, R. T., & Baker, T. A. (2021). Division of labor between the pore-1 loops of the D1 and D2 AAA+ rings coordinates substrate selectivity of the ClpAP protease. *Journal of Biological Chemistry*, *297*(6), 101407. https://doi.org/10.1016/j.jbc.2021.101407