Structural Characterization of the Baculoviral Minor Capsid Protein Ac132

Kyra Klijmeij, BSc. – Coulibaly Laboratory, Monash University

Supervised by: Assoc. Prof. Fasseli Coulibaly,

Bishwa Subedi, PhD, Jungmin Ha, MSc.

06/11/2022

MONASH University

Utrecht University

Abstract

Baculoviruses are insect viruses and have been widely applied for agricultural and medicinal purposes[1,2]. During the viral replication cycle, the baculoviral nucleocapsid is believed to transport to the nucleus via actin polymerization[3]. However, the precise mechanism is not yet fully determined. Understanding more about the replication cycle and in particular baculoviral motility via actin polymerization can help lead to important biotechnological advancements. One of the proteins involved in this mechanism is the nucleocapsid base protein Ac132[4]. This study aimed to determine the structure of Ac132 via X-ray crystallography to elucidate its role in baculoviral replication. Before the protein could be crystallized, Ac132 needed to be expressed, soluble, and purified using affinity (HIS-Tag) and size exclusion chromatography. The protein was expressed successfully in BL21(DE3) and C41(DE3) bacteria after IPTG induction, but was insoluble in various buffers with pH 6.5-9.2, detergent, and/or a reducing agent and therefore unsuitable for purification and crystallization. Co-expression with the nucleocapsid protein p78/83 and protein refolding did not increase the protein solubility. The protein remains – to date - insoluble, but several methods remain that may increase its solubility. Moreover, protein refolding may be optimized by refolding on a larger scale or under different refolding conditions. Therefore, the structural characterization of Ac132 is not deemed impossible. It would be valuable to see the biochemical reasoning behind the faced challenges and understand more about its role in baculovirus-induced actin polymerization.

Layman's Summary

Baculoviruses are viruses that are often used in agriculture, as biopesticides, or in gene therapy. However, how these viruses replicate in the host cell is not yet fully understood. They use a specific mechanism to transport themselves to the cell's nucleus, where they will create more viral particles. The virus itself consists of various proteins of which several are known to play an important role in this method of transportation. One of these proteins is called Ac132. It is still not known what this protein looks like, and determining its structure may help understand its role in actin polymerization and baculoviral replication. This study aimed to establish this structure by creating protein crystals. Before these crystals could be formed, the protein needed to be soluble and purified. However, after the protein was produced in bacteria, it was not soluble in the chosen solution. Changing the solvent conditions did not improve the solubility. Other methods, such as producing the protein together with a second protein, or unfolding and refolding the protein, did not result in increased protein solubility. The protein thus remains insoluble, but several techniques may still be tried to solubilize the protein. Therefore, determining the structure of Ac132 is still deemed highly achievable. It would be of great value to discover why this protein is so insoluble and what role it plays in the baculovirus-induced actin polymerization.

Table of Contents

List of Common Abbreviations

*Ac*MNPV	*Autographa Californica* multiple nucleopolyhedrovirus
Amp	Ampicillin
BME	Beta-mercaptoethanol
BV	Budded virion
Cam	Chloramphenicol
CV	Column volume
IPTG	Isopropyl-beta-D-thiogalactopyranoside
Kan	Kanamycin
ODV	Occlusion-derived virion
pI	Isoelectric point
SEC	Size-Exclusion Chromatography
TEV	Tobacco Etch Virus

1. Introduction

Baculoviruses are used extensively in agriculture and biomedicine as they can be easily manipulated, produced, and purified in large titers, and can carry large and multiple DNA inserts[5]. One of the most common applications of these viruses is biopesticides. They are insect-specific, safe for the environment and non-target organisms, and can be used in sustainable farming practices, making them suitable for insect control[1]. However, the high production cost and the long-term storage problems prevent they are more commonly used[1]. Baculoviruses can also be employed as delivery systems for gene therapy as they enable site-specific delivery and are easily modified. They have been successfully applied in several studies for cancer treatment, vaccines, and regenerative medicine[2,5].

Even though baculoviruses have been applied in various fields, much is still unknown about the biology of these rod-shaped viruses, in particular their viral assembly. Several aspects contribute to this problem. For example, high-resolution molecular structures are not yet available and the compositions of sub-viral structures (e.g. nucleocapsid cap and base) have yet to be fully established. A structure-function study on the major and minor capsid proteins of baculoviruses can provide more information on the role of these proteins in the multiple functions of the nucleocapsid. It can shine a

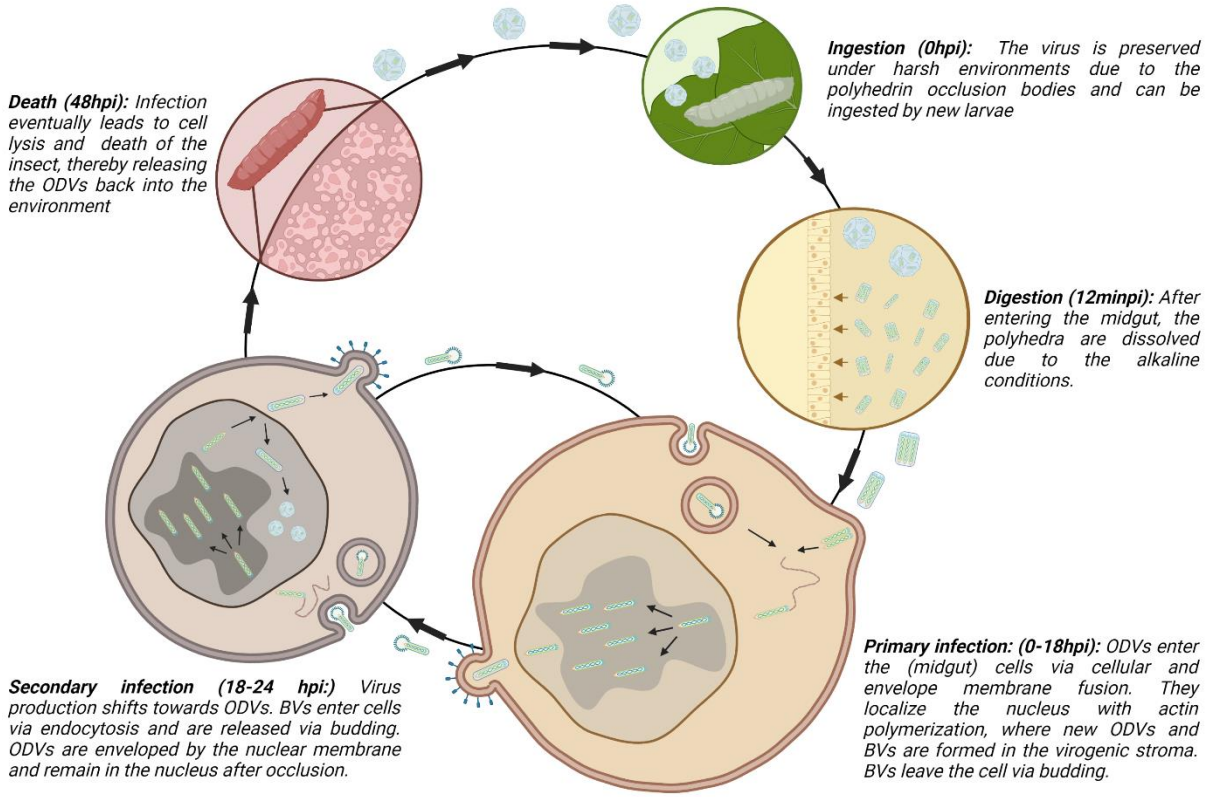


Figure 1. 1 – Baculoviral life cycle. A baculoviral particle is ingested by larvae (0 hours post-infection (hpi)) after which it is digested in the midgut. Here, the polyhedral occlusion bodies of the particle are dissolved and the enveloped particle can enter the midgut cells via membrane fusion (12 minutes pi (minpi)), starting the primary infection phase. After membrane fusion, the occlusion-derived virions (ODVs) move towards the nucleus via actin polymerization. The nucleocapsid then enters the nucleus and initiates the DNA replication in the virogenic stroma. Progeny nucleocapsids are assembled, leave the nucleus as enveloped viruses, and exit the cell via budding. These budded virions (BVs) can further transport through the organism, infect other tissues, thereby starting the secondary infection. The BVs enter these cells via endocytosis and the production of BVs and ODVs shifts from predominantly BVs to more ODVs. These ODVs are enveloped and occluded in the nucleus. Infection eventually leads to cell lysis and the death of the organism. The ODVs are then released into the environment to be taken up by new larvae. Created with BioRender.com (2022).

light on the mechanisms of baculoviral assembly and therefore provide new insights for biotechnological developments. One of the most well-described baculoviruses is the *alphabaculovirus Autographa Californica* (*Ac*MNPV), from which the first complete *alphabaculovirus* genome sequence was determined[6,7]. This virus has often been the baculoviral 'prototype' and will also be the subject of this study.

Even though there are significant knowledge gaps, a general replication cycle of the virus has been established (Fig.1.1). The virus itself is occluded (i.e. enclosed in a 10–15-µm protein crystal formed by polyhedrins[8]), enveloped, and has supercoiled, circular dsDNA with a genome size of ~134 kbp[9]. The morphology of the virus is often referred to as 'rod-shaped' even though the viral particles (AcMNPV) have also been observed to be ovoidal[10]. After ingestion of the virus by the host insect, the occlusion bodies (a.k.a. polyhedra) that protect the virus from environmental strains (e.g. temperature, UV) are dissolved due to the alkaline conditions of the insect midgut[11,12,13]. These occlusion-derived virions (ODVs) are then free to infect the midgut cells by membrane fusion using the nine *per os* infectivity factors (pifs) on the viral envelope[12,14]. The nucleocapsids are released into the cytoplasm and transported to the nucleus via actin polymerization due to the association of major and minor capsid proteins with actin[15,3]. It is believed that they can enter the nucleus intact using the nuclear pore complex (NPC)[15,16]. In the nucleus, they replicate in the virogenic stroma – a virus-induced subnuclear structure that can act as a DNA and protein factory – using the virus-encoded DNA polymerase (dnapol)[15,17]. The progeny viruses are then thought to exit the nucleus and the cell via budding using and are referred to as budded virions (BVs)[12,14,15]. These BVs can induce the secondary infection by spreading further through the insect entering cells via endocytosis using the GP64 fusion protein [11,15]. Eventually, the equilibrium of BVs and ODVs will slowly move towards the production of new ODVs. These ODVs are enveloped and occluded in the nucleus and are released back into the environment after virus-induced cell lysis, resulting in host death[13,12,18].

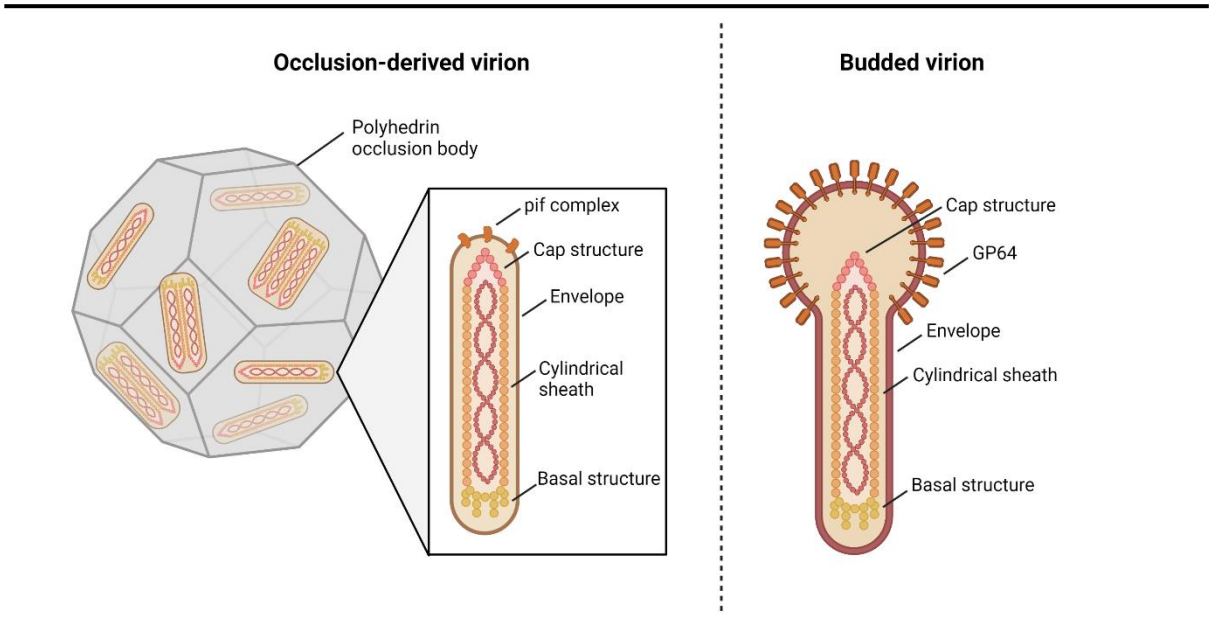


Figure 1. 2 – Schematic representation of the baculovirus structure (AcMNPV). The baculovirus can be subdivided into two structures: The occlusion derived virions (ODVs; primary infection) and the budded virions (BVs; secondary infection). The ODVs are enclosed by a polyhedral occlusion body and can be enveloped with one or multiple nucleocapsids. The envelopes of the ODVs contain the *per os* infectivity factors (pif), whilst the BVs are not occluded and contain a GP64 fusion protein for cell entry. Created with BioRender.com (2022).

Each baculovirus encodes approximately 90-181 genes, of which 30 core genes are highly conserved in the baculovirus family[19]. Many of these core genes encode proteins that are part of the nucleocapsid[19]. The nucleocapsid itself is identical between ODVs and BVs and can be subdivided into a cylindrical sheath, an apical cap, and a basal structure (Fig. 1.2)[20]. The most predominant is the major capsid protein VP39, which can be detected in the capsid sheath[21]. The proteins in the cap are more diverse and are known to be involved in, for example, nucleocapsid assembly and genome packaging[19]. The basal structure in its turn consists of various minor capsid proteins of which many have been associated with the actin-driven transportation of the virus. As mentioned before, baculoviruses induce actin polymerization to transport the nucleocapsid to the nucleus to initiate replication and to move towards the cell surface to exit the host cell. This phenomenon is believed to be unique for baculoviruses and is to date not fully understood[22]. One of the more characterized minor capsid proteins that are involved in actin polymerization is the p78/83[23]. This protein is known to play an important role in the transportation of the nucleocapsid to the nucleus as it mimics the characteristics of the well-characterized Wiskott-Aldrich syndrome (WASP)-proteins. WASP proteins can bind G-actin as well as the Arp2/3 complex – responsible for the nucleation of branched actin filaments – and can therefore accelerate actin polymerization[23,24]. It is known to bind and interact with several other baculoviral proteins and forms a complex with the minor capsid proteins C42, EC27, and AC102, all associated with actin polymerization and baculoviral motility[25,26]. Knock-out of C42 and Ac102 also lead to abnormal actin levels[27,28]. One of the less extensively studied minor capsid proteins presumed to be located in the base is Ac132. This protein is believed to have actin-binding properties, but its precise role in the baculovirus-induced actin polymerization is yet to be determined.

Several studies have been performed to determine the function of this protein, which is first located in the cytoplasm from 12 hours post-infection (hpi) onwards and ends up in the nucleus with a peak expression level at 24 hpi, which is in line with the expression times (up to 24hpi) of the late genes[29–31]. Although the results are inconsistent, Ac132 has also been detected in ODV envelope fractions[30]. Knock-out of the Ac132 gene led to the delayed formation of the virogenic stroma and a reduction in the number of enveloped nucleocapsids[29]. Additionally, it resulted in a reduction in the number of occluded virions and the number of envelopes containing multiple nucleocapsids (instead of a single nucleocapsid)[29]. It must be noted that there were cases of occluded virions in the cells at 96 hpi, implying that the protein does not impact the progression of the infection in the very late phase. Besides affecting the number of enveloped and occluded virions, Ac132 deletion does seem to negatively impact BV production and infectivity[29,30]. However, is not believed to affect DNA replication or the integrity of the viral genomic DNA[29,30]. Also, in contrast to the deletion of other nucleocapsid proteins, the deletion of Ac132 does not seem to impact viral morphogenesis[30]. The protein is also believed to associate with both viral envelope protein E18 and DNA-binding protein p6.9, and may therefore impact BV development and nuclear assembly due to the interaction with each respective protein[29].

Further experiments showed that BVs are not able to enter nuclei after Ac132-knockout, indicating that this capsid protein plays a role in the transport of the protein through the nuclear pores[30]. Moreover, Ac132 is believed to play a role in F-actin polymerization for virion transportation to the nucleus by accelerating the accumulation of actin in the nucleus, thereby delaying nuclear exportation[32]. It contains a sequence homologous to a NEBU-like domain, which is known to bind actin. This domain is present in the thin-filament-associated protein nebulin that is comprised of ~35 amino acid repeats (named: nebulin-like repeats)[4]. Mutation of the conserved SDXXYK motif of this NEBU-like domain in Ac132 led to decreased viral titer and thus most likely impacts viral infectivity[30]. As the NEBU-like domain is involved in actin binding, it is hypothesized that Ac132 may use this domain

to stabilize F-actin, thereby attaching the F-actin to the nucleocapsids and inducing the entry of the nucleocapsids into the nucleus[30].

There are many speculations regarding the functioning of Ac132, but a molecular structure has yet to be determined. Establishing the structure of Ac132 can provide more information on the residues involved in Ac132's mode of action and confirm or deny its anticipated roles in nuclear entry and virion transportation. It is expected that especially the NEBU-like domain (p103-134) plays an important role in Ac132 functioning due to its actin-binding ability[30]. However, it is not excluded that other domains may be essential as well. In this study, we aimed to determine the structure of Ac132 via X-ray crystallography and identify the crucial residues of this protein to further elucidate the molecular mechanisms behind the baculoviral replication cycle.

2. Results

To obtain pure protein suitable for crystallization, the Ac132 gene was inserted in a pET23a(+) vector together with a C-terminus extension containing a TEV cleavage site (ENLFQ↓S) and a 6xHIS-tag (see Supplementary Material 1.1 for details). The TEV cleavage site was inserted to enable the removal of the His-Tag and further purify the protein. The gene was codon-optimized for production in *Escherichia Coli* and checked for unwanted XhoI and NbeI restriction sites.

Before the protein was to be expressed and purified, the structure of Ac132 was predicted using the AlphaFold2 algorithm to provide insight into potential important residues, secondary structures, and the oligomerization potential of the protein.

The predicted structure of Ac132 is highly accurate for the C-terminus domains and demonstrates the potential for dimerization

With the use of programs such as Alphafold2, the Ac132 structure can be predicted with high accuracy[33]. AlphaFold2 predicted a clear C-terminus structure with high accuracy (>50 on the plDDT scale) containing a beta-sandwich with two antiparallel beta-sheets (β1, β4, β5, β8 and β2, β3, β6, β7 (as stated in Fig.2.1A, B)) with several α-helices on the N-terminus side (Fig.2.1A, B), here visualized in a hook-like structure. However, the neighbouring residues scored low on the plDDT scale, indicating low confidence behind the estimates of these residues (Fig.2.1A). It must also be noted that even though the separate residues may be predicted with high accuracy, the long-distance spatial arrangement of the domains is uncertain. The Predicted Aligned Error provided by AlphaFold2 illustrates the error behind the spatial arrangement between each residue compared to a second residue. This predicted position error showed little certainty (error = >15) on the 3D position of especially the N-terminus residues, whilst the positions of the C-terminus residues were predicted with more confidence (error = <15) (Fig.2.1A).

To establish whether there is potential for oligomerization of this protein, several oligomers were predicted using the AlphaFold2 algorithm (Fig.2.1C; Fig.S2.1). The trimer and tetramer were predicted with a high error in the spatial arrangement between the two monomers and are therefore deemed unlikely to occur. Contrastingly, the dimer form of the protein had a lower predicted aligned error rate, particularly in between the C-terminus residues of both monomers. The structure – as visualized in PyMol – showed that the main interaction between the monomers is predicted to occur at the tail-end of the C-terminus structures and between two α-helices (α1 and α4 (as named in Fig.2.1B)). As the single cysteine (p119) present in the Ac132 sequence does not seem to be at the interaction sites of this prediction model (Fig.2.1C), disulfide bonds are not expected to contribute to the dimerization of Ac132.

Despite AlphaFold2's ability to predict structures with relatively high accuracy, several parts of the structure, particularly at the N-terminus, are predicted with low accuracy and low certainty regarding the spatial arrangement of the residues. Therefore, it is important to determine the structure of Ac132 to confirm the predicted model and provide more information on Ac132's mode of action.

The C-terminus is well-conserved and is therefore likely to be essential for Ac132 functioning

There are several homologs of Ac132 in other baculoviruses (e.g. Bm109; plxy128; TaOrf-119) of which Bm109 from the *Bombyx mori nucleopolyhedrovirus* (BmNPV) is the most studied. Even though they share +/- 97% sequence similarity, there are clear differences between the two proteins as Bm109 s localized in the cytoplasm and can only be found in ODVs, whereas Ac132 is present in the nucleus and both ODVs and BVs[29,3429,30,34,35]. A multiple sequence alignment was made to define the conserved and non-conserved regions of Ac132 (Fig. S3.1) across 50 different alphabaculovirus sequences to

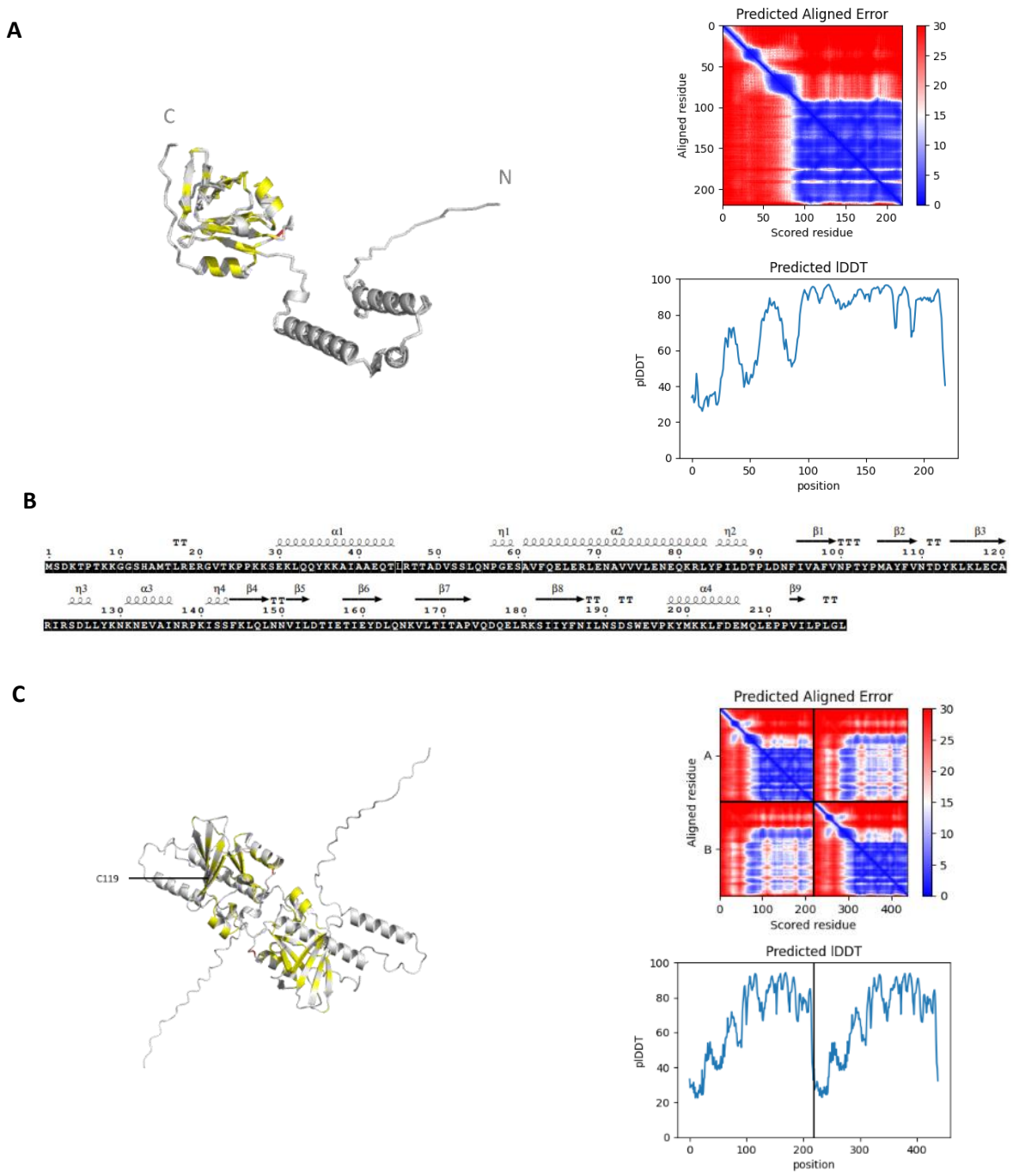


Figure 2. 1 – AlphaFold2 prediction of Ac132 structure. A) The structure of Ac132 is predicted and constructed by AlphaFold2_advanced.ipynb. and visualized with PyMol using the amino acid sequence obtained from NCBI ref#: NP_054162. The plDDT scale portrays the confidence behind the prediction from low (<50) to very high (>90) in numerical values. The Predicted Aligned Error illustrates the error behind the spatial arrangement of each residue compared to a second residue ranging from a high (red) to low error rate (blue). The structure has been colour-coded based on residue conservation (yellow: conserved; red: strictly conserved; grey: not conserved) obtained from a multiple sequence alignment (Fig. S3.1). B) Amino acid sequence of Ac132 accompanied by the secondary structures as predicted by AlphaFold2. Sequence and structures are visualized by ESPript (Robert, X. & Gouet, P, 2014). C) Ac132 dimerization as predicted by AlphaFold2 (visualized with PyMol). The plDDT scale illustrates the confidence in the prediction per monomer. The Predicted Aligned Error visualizes the error rate in the spatial arrangement per individual monomer (monomer A: residues 0-219; monomer B: residues 220-438) and between the two monomers. The single cysteine present C(119) in the sequence has been indicated on the structure of monomer A. The structure has been colour-coded based on residue conservation (yellow: conserved; red: strictly conserved; grey: not conserved) obtained from a multiple sequence alignment (Fig. S3.1).

elucidate which fractions of the protein may play an important role in Ac132 functioning[36,37]. Most residues were not conserved across most virus species, especially on the N-terminus, the 'tail-end',

with several sequences not containing the N-terminal α-helices. These sequences (Malacosoma_neustria_NPV to Spodoptera_litura_NPV_II; Fig. S3.1) are classified as Group II alphabaculoviruses, in contrast to AcMNPV which belongs to Group I. The C-terminus – base structure – of the protein appears to be more conserved. Only one residue (W194) was strictly conserved across all sequences. (Fig. 2.1). The NEBU-like domain (p103-134) falls within this base structure and several residues of this domain are well-conserved. The non-conserved nature of the N-terminus residues implies that this region is less essential for the primary function of the protein and that the C-terminus contains the critical functional domains. No direct relationship between conservation and potential dimerization could be identified as the conserved residues do not appear to be consistently present in the predicted interaction surfaces(Fig. 2.1C).

Ac132 can be well-expressed, but further optimization is needed to obtain soluble protein

pET23a+_Ac132_TEV_HIS was transformed in competent BL21(DE3) and Origami™ 2(DE3) pLysS *E. Coli* bacteria to assess the expression and solubility of the protein. As the BL21(DE3) and Origami™ 2(DE3) pLysS bacterial strains both contain the lambda DE3 prophage, T7 RNA polymerase expression will be under the control of the lacUV5 promoter, which is naturally inhibited by the lac5 repressor. Isopropyl β-D-1-thiogalactopyranoside (IPTG) can bind to this repressor and enable T7 polymerase expression, allowing for the controlled expression of our target protein. The Origami 2(DE3) pLysS bacteria also contain the T7 lysozyme that can inactivate transcription by the T7 polymerase and prevent any 'leaky' expression of Ac132 (Fig. 2.2A).

After transformation, only three colonies were observed for the Origami 2(DE3) pLysS bacteria. The BL21(DE3) bacteria produced full plates. Repetition of the transformation produced similar results, hinting toward potential toxicity of the protein. Analysis of the samples (non-induced, induced, precipitant, supernatant) via SDS PAGE showed expression of a protein at approximately 25 kDa, which is similar to the calculated size of Ac132 (25.1 kDa; ProtParam)[37], for both bacterial strains (Fig. 2.2B). However, little difference can be seen between the non-induced and IPTG-induced bands of the BL21(DE3) samples, indicating that there might have been some leaky expression. There is a clear difference between the non-induced and induced bands for the Origami 2(DE3) pLysS bacteria could be seen. For both strains, it appears that Ac132 is present in the insoluble fraction. No protein was visible at 25 kDa in the samples containing the HIS-tagged proteins.

Due to suspicions of potential protein-induced toxicity, the experiment was repeated in C41(DE3) bacteria. These bacteria are mutated to prevent cell death, making them suitable for the expression of toxic proteins. After making these cells chemically competent for plasmid transformation, there was abundant colony growth, comparable to the BL21(DE3) bacteria. As these bacteria are similar to the BL21(DE3), these results were anticipated. Liquid cultures of the C41(DE3) bacteria also appeared to grow faster than the other bacterial strains. The results of the following expression test show clear induction and expression of a protein at approximately 25 kDa, again present only in the insoluble fraction (Fig.2.2B). This improvement in expression may be due to the change in bacterial strains but is more likely to be the result of a more skilful performance of the experiment.

Ac132 expression was confirmed with western blotting and sequenced via mass spectrometry (MS). Two identical expression tests were performed in C41(DE3) bacteria of which one gel was stained with Coomassie Blue and the other was further used for western blotting. The western blot result shows an intense chemiluminescent signal for the induced and insoluble fractions, and a relatively faint, but clear signal for the non-induced and supernatant (Fig. 2.2C). Mass spectrometry further confirmed the expression of Ac132 (Fig. 2.2C). A band in the precipitant fraction was excised from the SDS PAGE gel

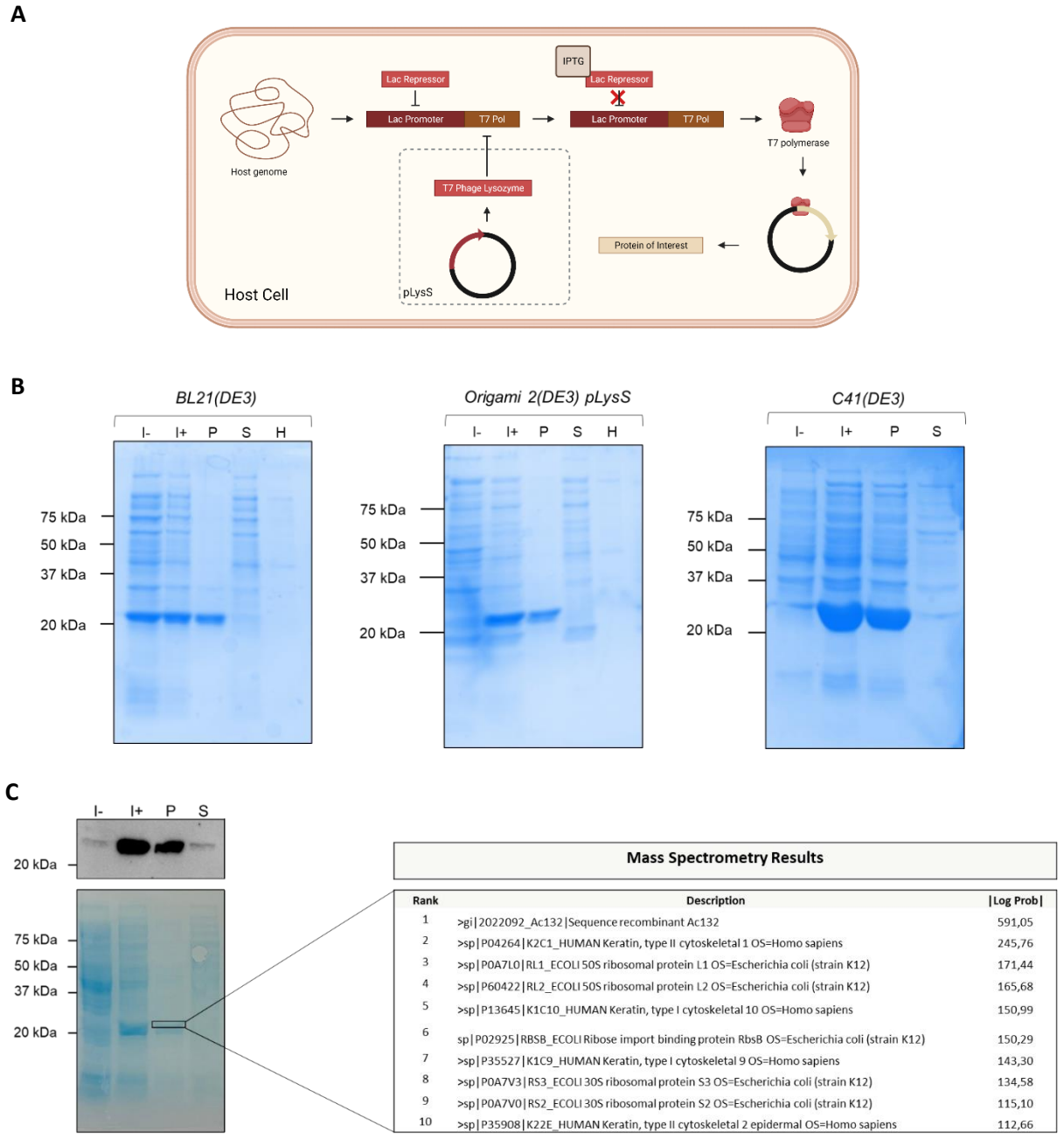


Figure 2. 2 – SDS PAGE analyses of the Ac132 expression tests in various BL21(DE3), Origami 2(DE3) pLysS, and C41(DE3) E. Coli bacteria. A) Schematic representation of IPTG-induced protein expression in DE3 and DE3 pLysS bacteria. IPTG can inhibit the lac5 repressor, thereby activating the lacUV5 promotor. This results in the expression of the T7 polymerase. The pLysS plasmid (present in the Origami 2(DE3) pLysS bacteria) encodes a T7 lysozyme that can inactivate T7 polymerase transcription and thereby prevent any leaky expression. Created with BioRender.com (2022) B) Ac132 has an expected size of 25.1 kDa. The pET23a(+)_Ac132_TEV_HIS plasmid was transformed in the above-mentioned bacterial strains. The samples were divided in IPTG-induced (I+) and non-induced (I-) samples. The induced samples were separated in a supernatant (soluble fraction, S) and a precipitant (insoluble fraction, P). The HIS-tagged proteins in the supernatant were pulled down using Ni^{2+} beads (H) to assess whether the proteins originated from the host or the pET23a(+)_Ac132_TEV_HIS plasmid. A) comparative analysis of the protein expression in three different bacterial strains (BL21(DE3); Origami 2(DE3) pLysS; C41(DE3)). The samples were separated in a reducing SDS PAGE and stained with Coomassie Blue. No pull-down was performed for the samples obtained from the C41(DE3) bacteria. C) Ac132 identification via western blotting and mass spectrometry sequencing. A band was excised (black box) from the gel and send for sequencing to the Monash Proteomics and Metabolomics Facility. The box shows the top 10 hits from the mass spectrometry analysis.

and sent for sequencing to the Monash Proteomics and Metabolomics Facility using LC-MS. 100 proteins were identified with the Byonic (ProteinMetrics) search engine using the Uniprot Human and Ecoli databases as well as the Ac132_TEV_HIS sequence. Half of the identified sequences were common host proteins or contaminants. The Ac132 sequence was deemed most probable with a 95.3% coverage match.

Most (+/- 95%) of the Ac132 seems to be present in the insoluble fraction using the current lysis buffer (50 mM HEPES, pH 6.5; 150 mM NaCl). A buffer screen was performed to assess the behaviour of the protein in various buffer combinations as buffer conditions – such as pH, detergents, and reducing agents – can impact protein solubility.

Variations in lysis buffer conditions do not greatly impact Ac132 solubility

The buffer screen was performed with a high salt concentration (500 mM NaCl), either 50 mM HEPES, pH 7.2 or 50 mM CHES, pH 9.0, and/or a reducing agent (beta-mercaptoethanol (BME)) and detergent (Triton X-100). Reducing agents can disrupt potential disulfide bonds and detergents can incorporate proteins in detergent micelles, and are therefore commonly used to increase protein solubility.

As expected, Ac132 did not appear to be present in the soluble fractions of the samples without BME or Triton-X100 (Fig.2.3). However, the samples with Triton-X100 (both with and without BME) showed small signs of Ac132 in the supernatant, particularly for pH 7.2. This was less clear for the samples containing only BME. Therefore, a scale-up was performed using a 50 mM CHES, pH 9.0, 500 mM NaCl, 1% Triton-X100, and 5mM BME lysis buffer to see if these bands would be more clearly defined at a

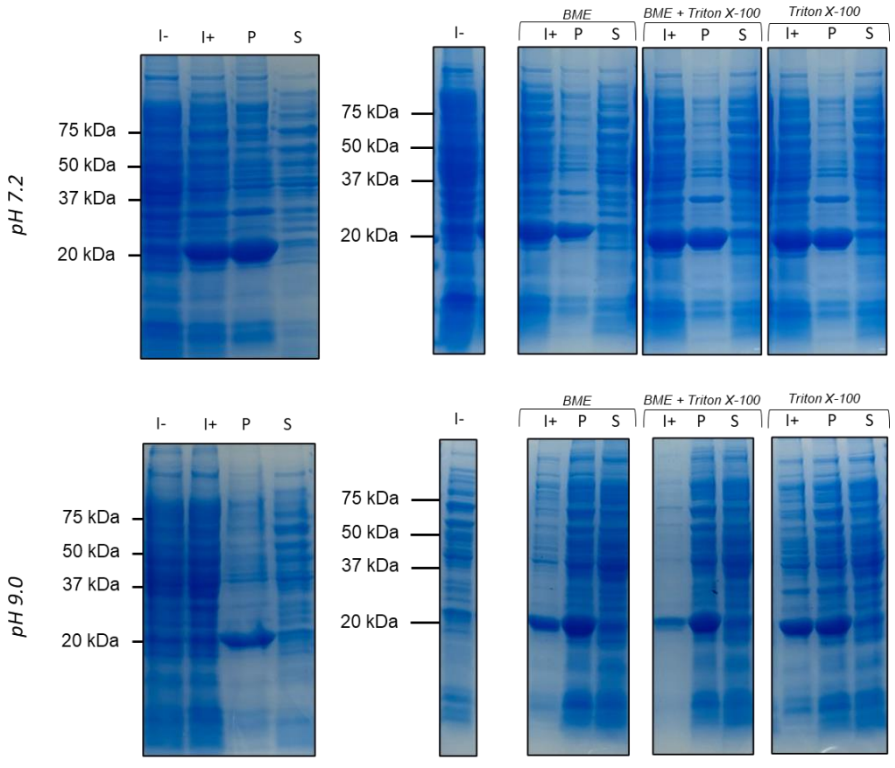


Figure 2. 3 – SDS PAGE analysis of a buffer screen with 500 mM NaCl, 50 mM HEPES, pH 7.2 or 50 mM CHES, pH 9.0, and beta-mercaptoethanol (BME) and/or Triton X-100. Ac132 has an expected size of 25.1 kDa. The pET23a(+)_Ac132_TEV_HIS plasmid was transformed in C41(DE3) bacteria. The samples have been divided in IPTG-induced (I+) and non-induced (I-) samples. The induced samples have been separated into a supernatant (soluble fraction, S) and a precipitant (insoluble fraction, P). A) Results from the buffer screen. Each gel contained samples from one pH condition: no BME and Triton X-100 (gel 1); BME and Triton X-100 conditions (gel 2).

larger scale (500 mL culture). The collected protein sample was purified with a HIS Trap FF column and the proteins in all samples and fractions were separated via SDS PAGE (Fig. S4.1 (left)).

Ac132 cannot be clearly detected in the sample. Further purification was therefore not performed. Re-lysing the precipitant via sonication did not appear to release more Ac132 into the soluble fraction (Fig. S4.1 (right)).

Co-expression with p78/83 hints toward protein-protein interaction, but does not increase Ac132 solubility

To improve the solubility of the protein, Ac132 was co-expressed with the baculoviral protein p78/83. As both proteins are thought to share a similar function and localization (nucleus, >24hpi)[38], co-expression of p78/83 and Ac132 may increase the solubility of Ac132 via protein-protein interactions. (see Fig.2.5A for the predicted complex)

P78/83 was previously cloned into a pET28 vector with a kanamycin resistance gene to adequately select for the bacteria containing both plasmids by Jungmin Ha (Coulibaly Laboratory). The plasmids were transformed into BL21(DE3), C41(DE3), and Origami 2(DE3) pLysS bacteria, but colonies were only formed for the BL21(DE3) bacteria. As the transformation was performed with relatively old C41(DE3) cells, the lack of colonies may be due to a loss of competence. The transformation results of the Origami 2(DE3) pLysS bacteria are in line with previous results. pET23a(+)_Ac132_TEV_HIS and pET28_p78/83_TEV_HIS were also individually transformed to act as positive controls. As both proteins have varying predicted pI values (Ac132: 8.45; p78/83: 5.78), three different pH levels (50 mM HEPES, pH 6.5; 50 mM HEPES, pH 7.2; 50 mM CHES, pH 9.0) were used with a low salt concentration (150 mM NaCl).

The SDS PAGE results show solubility for p78/83 for the positive control. No soluble Ac132 can be detected in the singly-expressed samples (Fig. 2.5). Co-expression of both proteins does not appear to change the solubility of Ac132. In previous experiments performed in the lab (data not shown) the solubility of p78/83 decreased after co-expression with several other base proteins. This may indicate that there are interactions between p78/83 and Ac132 that negatively influence the solubility of p78/83. Nevertheless, Ac132 remains insoluble, making it unsuitable for purification and crystallization. Ac132 was therefore refolded to improve solubility.

Protein refolding appears to increase protein solubility, but Ac132 cannot be successfully purified

Protein refolding is often used to obtain soluble protein from the insoluble bacterial inclusion bodies. By adding compatible additives to the refolding buffer, denatured (unfolded) proteins may refold and remain in a soluble state[39]. The experiment was performed with single expression of Ac132 and co-expression with p78/83. As Ac132 has only a single cysteine and p78/83 has none, no intramolecular disulfide bonds needed to be restored during the refolding process. Unfolding was first performed with urea on a small scale to assess whether Ac132 could be extracted from the soluble fraction after denaturation by a high concentration of urea. A clear band was visible after the pull-down with agarose Ni^{2+} beads, indicating that the protein was soluble in its unfolded state (Fig.S5.1). The protein was then expressed on a larger scale. Large portions of the protein were not purified properly after Ni-NTA affinity chromatography. Nevertheless, an adequate amount of protein could be extracted from the 250 mM imidazole fractions W2.1-W2.3 (totalling approximately 17 mg and 6 mg protein for the Ac132 and Ac132+p78/83 conditions respectively).

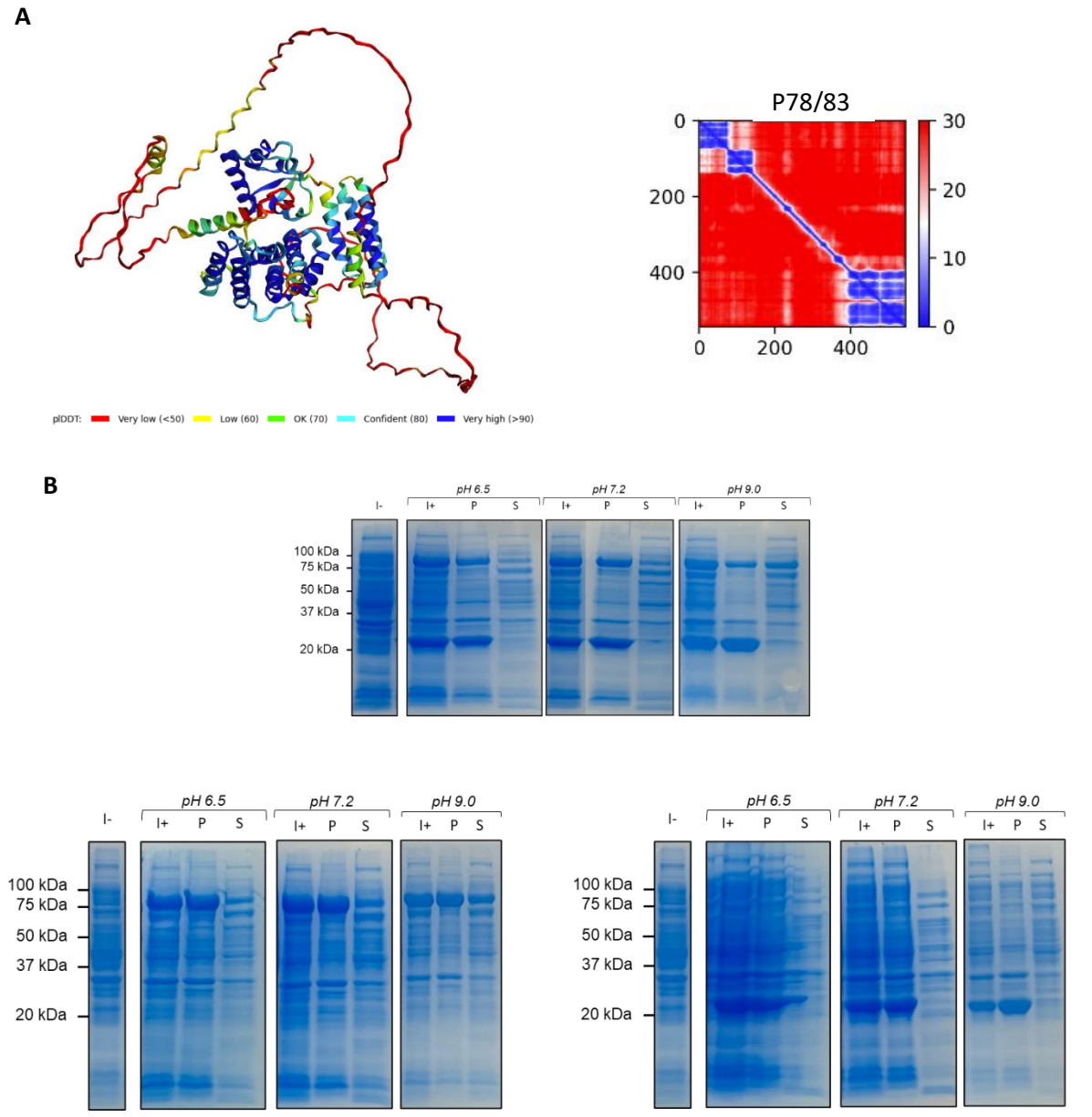
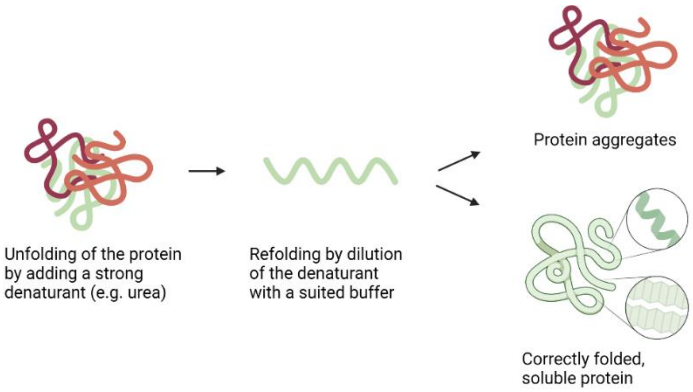


Figure 2. 4 – Co-expression of Ac132 with p78/83. A) The structure of p78/83 is predicted and constructed by AlphaFold2_advanced.ipynb using the amino acid sequence obtained from NCBI: NP_054038.1 (p78/83) coloured by the lDDT scale (certainty of prediction). The Predicted Aligned Error illustrates the error behind the spatial arrangement of each residue compared to a second residue ranging from a high (red) to low error rate (blue). B) SDS PAGE analyses of Ac132 expression, p78/83 expression, and p78/83-Ac132 co-expression at pH 6.5, 7.2, and 9.0. Ac132 and p78/83 have expected sizes of 25.1 kDa and 60.7 kDa respectively. The samples have been divided in IPTG-induced (I+) and non-induced (I-) samples. The induced samples have been separated in a supernatant (soluble fraction, S) and a precipitant (insoluble fraction, P).

The protein was added to each well of the 96-well plate containing the refolding buffers. To analyse the solubility of the protein, the OD (340nm) of the solution was measured at 0h and 24h. A blank measurement was taken before the addition of the protein and subtracted from the final measurements. To distinguish between soluble and insoluble conditions, a cut-off value was set at 0.05. Precipitated (insoluble) proteins were presumed to increase the OD due to the scattering of the UV light. At 0h, the results showed the most solubility for the conditions containing a higher pH level and 200 mM NaCl concentration for ac132. The proteins appeared more soluble for the ac132-p78/83 co-expression with most 'soluble' conditions including the 200 mM NaCl (Fig.2.5). The solubility

A



B

Single-expression

0h

		1	2	3	4	5	6	7	8	9	10	11	12
		pH 5.0		pH 6.0		pH 7.0		pH 8.0		pH 9.0		pH 10.0	
	NaCl	0 mM	200 mM	0 mM	200 mM	0 mM	200 mM	0 mM	200 mM	0 mM	200 mM	0 mM	200 mM
A	L-arginine	0,3335	0,41	0,6365	0,066	0,1445	0,0675	0,229	0,004	0,2485	0,003	-0,902	0,0055
B	Glucose	0,102	0,158	0,322	0,137	0,201	0,1525	0,1965	0,0695	0,291	0,0185	0,179	0
C	Glycerol	0,144	0,0635	0,2735	0,1045	0,239	0,1145	0,2015	0,042	0,1505	0,039	0,1305	0,0095
D	EDTA/ Ions	0,1095	0,1325	0,561	0,341	0,272	0,137	0,375	0,028	0,173	0,0275	0,0925	0,01
E	PEG4000	0,111	0,085	0,306	0,1905	0,1785	0,1735	0,211	0,0665	0,197	0,041	0,0705	0,0055
F	PEG400	0,164	0,0965	0,325	0,178	0,289	0,1095	0,186	0,0855	0,2565	0,0705	0,1175	0,0225
G	Triton X100	0,0325	0,059	0,1695	0,138	0,241	0,066	0,1465	0,0315	0,11	-0,0535	0,048	0,016
H	Urea	0,091	0,0685	0,3285	0,162	0,356	0,1415	0,4165	0,0155	0,1965	0,055	0,0625	0,045

24h

		1	2	3	4	5	6	7	8	9	10	11	12
		pH 5.0		pH 6.0		pH 7.0		pH 8.0		pH 9.0		pH 10.0	
	NaCl	0 mM	200 mM	0 mM	200 mM	0 mM	200 mM	0 mM	200 mM	0 mM	200 mM	0 mM	200 mM
A	L-arginine	0,791	0,973	0,775	0,159	0,172	0,117	0,24	0,034	0,296	0,024	-0,886	0,017
B	Glucose	0,705	1,098	0,383	0,394	0,172	0,236	0,217	0,127	0,301	0,057	0,184	0,006
C	Glycerol	1,287	1,301	0,444	0,159	0,242	0,193	0,224	0,113	0,244	0,095	0,11	0,029
D	EDTA/ Ions	1,948	1,364	0,598	0,363	0,277	0,312	1,035	0,058	0,186	0,09	0,062	0,029
E	PEG4000	2,143	1,693	0,391	0,374	0,219	0,238	0,235	0,117	0,189	0,087	0,03	0,028
F	PEG400	2,301	1,906	0,335	0,267	0,273	0,218	0,281	0,119	0,313	0,062	0,044	0,04
G	Triton X100	1,52	1,239	0,632	0,173	0,399	0,133	0,353	0,107	0,228	-0,037	0,192	0,026
H	Urea	0,651	0,636	0,415	0,191	0,335	0,245	0,421	0,052	0,217	0,047	0,02	0,058

Co-expression

0h

		1	2	3	4	5	6	7	8	9	10	11	12
		pH 5.0		pH 6.0		pH 7.0		pH 8.0		pH 9.0		pH 10.0	
	NaCl	0 mM	200 mM	0 mM	200 mM	0 mM	200 mM	0 mM	200 mM	0 mM	200 mM	0 mM	200 mM
A	L-arginine	0,021	0,0105	0,0885	0,012	0,079	0,015	0,0645	0,0035	0,1035	-0,006	-0,4105	0,0055
B	Glucose	0,057	0,1025	0,522	0,0345	0,1325	0,0355	0,154	0,1375	0,1065	0,008	0,045	0,007
C	Glycerol	0,065	0,035	0,1815	0,034	0,118	0,059	0,1625	0,01	0,175	0,019	0,08	0,0165
D	EDTA/ Ions	0,017	0,036	0,0765	0,031	0,209	0,041	-0,022	0,011	0,1495	0,0265	0,07	0,016
E	PEG4000	0,034	0,0165	0,233	0,062	0,1565	0,0315	0,154	0,0185	0,137	0,048	0,1035	0,0185
F	PEG400	0,047	0,0185	0,2805	0,051	0,1275	0,042	0,155	0,022	0,148	0,026	0,106	0,0075
G	Triton X100	-0,042	0,007	0,105	0,0505	0,0725	0,0265	0,113	0,0055	0,0975	-0,037	0,0245	0,01
H	Urea	0,0205	0,0195	0,2505	0,055	0,134	0,0355	0,1595	0,014	0,2055	0,027	0,0695	0,014

24h

		1	2	3	4	5	6	7	8	9	10	11	12
		pH 5.0		pH 6.0		pH 7.0		pH 8.0		pH 9.0		pH 10.0	
	NaCl	0 mM	200 mM	0 mM	200 mM	0 mM	200 mM	0 mM	200 mM	0 mM	200 mM	0 mM	200 mM
A	L-arginine	0,195	0,091	0,168	0,889	0,133	0,436	0,457	0,051	0,457	0,028	-0,126	0,034
B	Glucose	0,069	0,075	0,757	0,298	0,083	0,081	0,247	0,341	0,289	0,038	0,06	0,028
C	Glycerol	0,322	0,239	0,153	0,198	0,1	0,201	0,159	0,07	0,438	0,07	0,378	0,039
D	EDTA/ Ions	0,242	0,078	-0,084	0,343	0,151	0,135	0,122	0,022	0,432	0,093	0,312	0,044
E	PEG4000	0,391	0,071	0,258	0,297	0,124	0,255	0,259	0,077	0,371	0,56	0,158	0,039
F	PEG400	0,328	0,082	0,187	0,929	0,12	0,293	0,263	0,824	0,369	0,286	0,17	0,016
G	Triton X100	0,004	0,052	0,076	0,834	0,093	0,694	0,444	0,04	0,378	0,014	0,042	0,023
H	Urea	0,243	0,055	0,168	0,551	0,146	0,216	0,363	0,029	0,323	0,041	0,104	0,031

Figure 2. 5 – Protein refolding. A) Schematic representation of the refolding process. Proteins are denatured by a strong denaturant and refolded using various refolding buffers, leading to the formation of soluble protein or protein aggregates. Created with BioRender.com (2022) B) Optical density (OD) measurements of the 96 refolding buffer mixes after addition of ac132 or ac132-p78/83. Measurements were taken 0h and 24h after addition of 5 µL protein (+/- 5 mg/mL) to the buffers. The OD measurements were subtracted from a blank measurement of each buffer without protein. Each refolding buffer contained a unique additive (50 mm arginine, 1mM EDTA/Ions, 500 mM glucose, 20% glycerol, 0.05% PEG400, 0.05% PEG4000, 0.01% Triton X-100, or 100 mM urea), and 0 mM or 200 mM NaCl. The pH was adjusted by several buffers (MES, pH 5.0; HEPES, pH 6.0; HEPES, pH 7.0; Tris-HCl, pH 8.0; CHES, pH 9.0; CHES, pH 10.0). The absorbance was measured with the PheraStar at 340 nm. The solubility cut-off was set at 0.05. Wells meeting this cut-off are marked green. Wells exceeding this cut-off value are marked red. Wells bordering the cut-off (0.05-0.1) are marked orange.

decreased for most conditions after 24h with 11 and 18 wells meeting the cut-off point for the single and co-expression respectively. Most (read: >50%) of these conditions contained the CHES, pH 10 buffer and/or 200 mM NaCl. However, a lower pH level is preferred in future experiments. Two conditions (L-arginine and EDTA) at pH 8.0, 200 mM NaCl were chosen for larger-scale refolding. Due to the low concentration of the Ac132+p78/83 condition, the W1.2-W1.6 fractions were pooled and concentrated to obtain more protein for further refolding (~27 mg protein).

Approximately 2 mg of protein was diluted in the chosen refolding buffers (dilution factor = 8x). The samples were concentrated and filtered for analysis by gel filtration. Little to no precipitation was observed after the concentration of the protein samples. However, the protein concentration did appear to decrease (~50%) after filtration for all chosen conditions. As no large differences could be observed between conditions, the pH 8.0, 200 mM NaCl, 50 mM L-arginine condition was chosen for further analysis by gel filtration.

The sample was injected onto the size exclusion chromatography (SEC) column to identify and isolate the refolded protein. The SEC results showed a large peak at ~1.08 Column Volume (CV) and several minor peaks between 0.88 and 1.04 CV (Fig. 2.6). This was unexpected as the column calibration curve projected the peaks to appear at ~0.90 (Ac132, monomer), ~0.81 (Ac132, dimer), ~0.79 (p78/83, monomer), and ~0.75 (p78/82-Ac132 complex) (Fig. S5.2). Moreover, only one large peak (above the 10 mAU threshold) was observed for the Ac132-p78/83 samples, where at least 2-3 large peaks – representing Ac132, p78/83, and Ac132-p78/83 – were expected. Repetition of the gel filtration with the pH 8.0, 200 mM NaCl, 1 mM EDTA condition only produced one larger peak at ~1.8 CV (Fig. S5.4). No calibration was performed for this column, leaving the possibility open that this substance eluted this late due to problems with the column. To validate the presence of the protein in the large peaks, the collected fractions were concentrated (4-5x) and analysed via SDS PAGE and western blotting. The protein did not appear to be concentrated as the concentrations in the 'concentrated' sample and the

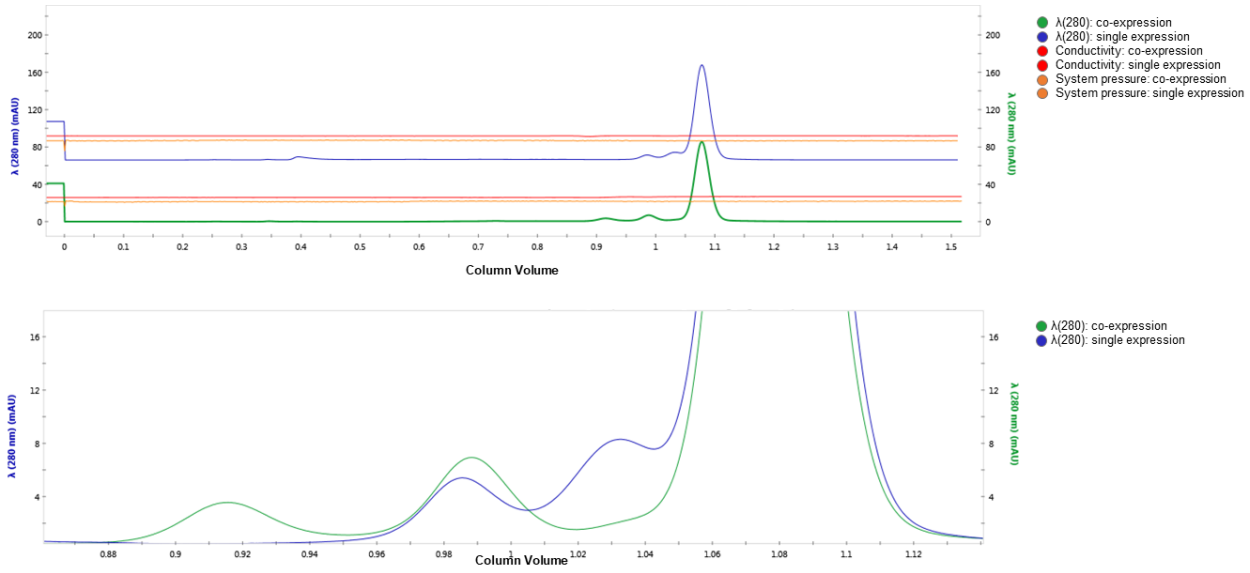


Figure 2. 6 – Size exclusion chromatogram from the single- and co-expression of Ac132 and Ac132-p78/83 in a pH 8.0, 200 mM NaCl, 50 mM L-arginine refolding buffer. The protein sample was injected in the BioRad NGC chromatography system. Gel filtration was performed on a Superdex 16/600 200 pg column with a 100 mM Tris-HCl, pH 8.0, 200 mM NaCl, 50 mM L-arginine running buffer. The chromatogram was visualized with ChromLab software.

flow-through were comparable (Fig. S5.4). To analyse whether these peaks were caused by the buffer components, the UV spectra of several components were compared to the UV spectra of the protein samples, but no direct match in absorbance at 280 nm could be observed (Fig. S5.5). No protein could be observed by Coomassie-stained SDS PAGE or western blotting with an anti-HIS6x antibody (Fig. S5.3), leading to the conclusion that neither Ac132 nor p78/83 is present in the collected fractions.

3. Discussion and Conclusion

Ac132 has been predicted to be a small (~25.1 kDa, ProtParam[37]), relatively soluble (~0.616, Protein-Sol[40]) protein. Nevertheless, it appears to be difficult to dissolve. After adjusting the buffer conditions, co-expressing with p78/83, and even refolding the protein, Ac132 did not appear to reach a solubility level suitable for purification and crystallization, potentially due to oligomerization or aggregation.

Even though the refolding did not – to date – provide a sufficient amount of soluble protein, new refolding conditions may still be tested. The higher pH levels appeared to be preferable in the refolding buffer screen and need to be included in future experiments. Moreover, even though the larger peaks of the size exclusion chromatograms did not contain Ac132 or p78/83, the same cannot be said for the smaller peaks. These peaks contained – according to their UV spectra – a low amount of protein, which may not have been sufficient for SDS PAGE analysis. As these peaks appear slightly earlier, they may – particularly the peak at ~23.5 kDa for co-expression – contain the refolded Ac132 monomer. This, therefore, needs to be repeated with a larger amount of Ac132, to account for the large loss of protein during the refolding experiment.

Several other methods may be tried to increase protein solubility and protein truncation may be a suitable next step. As the N-terminus does not appear to be conserved amongst other *alphabaculoviruses*, it may be removed without losing any essential residues, either in its entirety (creating Ac132$^{p94\text{-}219}$) or by truncating only the most N-terminal part (creating Ac132$^{p29\text{-}21}$), without removing any secondary structures (as predicted by AlphaFold2). Removing several parts of the protein may increase the hydrophilicity – and therefore solubility – of the protein. Other approaches may include the addition of a solubility tag – such as a maltose binding protein (MBP)-tag which has been shown to increase protein solubility[41]–, changing the culture media by adding e.g. additional growth factors[42] or expressing the protein in its natural host (insect cells). The latter may provide the necessary post-translational modifications that are needed to increase the solubility of the protein. No PTMs have been described in literature for Ac132, but phosphorylation and glycosylation have been frequently observed in AcMNPV proteins (including p78/83) and may also occur for Ac132[43]. However, the potential toxicity of the protein may prevent sufficient protein expression in an insect expression system.

The complex formation between p78/83 and Ac132 is also yet to be confirmed. Determining an interaction between both proteins will further cement the role of Ac132 in baculovirus-induced actin polymerization. Besides refolding, this complex can be established by adding a second affinity tag on one of the proteins. If, after performing affinity chromatography, both proteins are present in the eluted fractions, the interaction between Ac132 and p78/83 can be confirmed. Due to the expected size of the complex, the structure can then be determined via cryo-electron microscopy (Cryo-EM). The involvement of the complex, or both separate proteins, in actin polymerization can be further studied in an actin polymerization assay by using pyrene-bound actin to identify and quantify the actin polymerization – by measuring the fluorescent intensity – after the addition of the purified protein. Other base proteins involved in actin polymerization, such as EC27, C42, and Ac102, can be included to assess the interaction between all five base proteins and the necessity of the proteins in baculoviral-induced actin polymerization.

Algorithms such as AlphaFold2 and RoseTTAFold have already made major improvements in the past years and are currently able to predict a protein structure with relatively high accuracy. However, they are not without fault[44]. Researchers still depend on the class wet-lab techniques (X-ray Crystallography, Cryo-EM, NMR) for the determination of the protein structure, which – as shown in this project – can have its own set of challenges.

Even though we were not able to produce soluble Ac132 suitable for crystallization, this project was able to bring to light the difficulties surrounding this protein and it will be valuable to discover the biochemical reasoning behind these challenges. There are still many options that can be tried to increase the protein's solubility and crystallizing Ac132 is not deemed impossible. The structure of Ac132 will eventually provide more insight into baculovirus-induced actin polymerization, which can lead to important technological advancements, by, e.g., using this machinery as a protein transportation method in biomedicine. The structural characterization of Ac132 is therefore not only important but also believed to be achievable in the near future.

4. Methods

Alphafold2 Predictions

Structure predictions were performed in AlphaFold2_advanced with (variations of) the Ac132 amino acid sequence (NCBI ref#: NP_054162):

MSDKTPTKKGGSHAMTLRERGVTKPPKKSEKLQQYKKAIAAEQTLRTTADVSSLQNPGESAVFQELERLENAVVVLENEQKRLYPILDTPLDNF
IVAFVNPTYPMAYFVNTDYKLKLECARIRSDLLYKNKNEVAINRPKISSFKLQLNNVILDTIETIEYDLQNKVLTITAPVQDQELRKSIIYFNILNSDS
WEVPKYMKKLFDEMQLEPPVILPLGL

And the p78/83 sequence (NCBI ref#: NP_054038.1):

MTNRRYESVQSYLFNNRNNKIDAHQFFERVDTAEAQIIKNNIYDNTVVLNRDVLLNILKLANDVFDNKAYMYVDDSEVSRHYNAVVKMKRLV
IGVRDPSLRQSLYNTIAYIERLLNIGTVNDSEITMLIADFYDLYSNYNIELPPPQALPRSRRPSVVQPAAPAPVPTIVREQTKPEQIIPAAPPPPPSP
VPNIPAPPPPPPPSMSELPPAPPMPTEPQPAAPLDDRQQLLEAIRNEKNRTRLRPVKPKTAPETSTIVEVPTVLPKETFEPKPPSASPPPPPPPPP
PPAPPAPPPMVDLSSAPPPPPLVDLPSEMLPPPAPSLSNVLSELKSGTVRLKPAQKRPQSEIIPKSSTTNLIADVLADTINRRRVAMAKSSSEATS
NDEGWDDDDNRPNKANTPDVKYVQALFNVFTSSQLYTNDSDERNTKAHNILNDVEPLLQNKTQTNIDKARLLLQDLASFVALSENPLDSPAI
GSEKQPLFETNRNLFYKSIEDLIFKFRYKDAENHLIFALTYHPKDYKFNELLKYVQQLSVNQQRTESSA

The predicted structures were visualized in PyMol2 (Version 2.5.2).

Multiple Sequence Alignment

The AcMNP Ac132 sequence (NP_054162) was compared and aligned with NCBI PSI-BLAST (non-redundant proteins only; filtered for viruses (taxid: 10293)). Iterations were performed until no new sequences were found. Only 1 sequence per virus species was selected (lowest E-score). The sequences were further processed in Clustal Omega after which the alignment file was visualized and processed with ESPript (no defined groups; %Equivalent similarity colouring scheme). The secondary structures visualized in the alignment were obtained from the AlphaFold2 prediction.

Plasmids, Cells, and Chemicals

pET23a+_Ac132_TEV_HIS was constructed and obtained from GenScript® (Fig.S1.1). The AC132 (NCBI ref#: NC_001623.1) and TEV cleavage site (ENLYFQ↓S) sequences were inserted between the NdeI and XhoI restriction sites (see Supplementary Material 1.1 for details). The pET28a_p78/83_TEV_HIS plasmid was previously constructed and cloned in this lab.

BL21(DE3) and Origami™2(DE3) pLysS (ref#: 71346; Sigma-Aldrich) cells were commercially or previously made competent. C41(DE3) cells were chemically made competent by resuspension in $CaCl_2$ and $MgCl_2$.

The antibiotics were dissolved in MilliQ to obtain the noted concentrations, filtered before use, and stored at -20°C.

All buffer components were filtered before use with a 22µm Steritop® Filter (Millipore Express® Plus) and stored at room temperature. Buffers used for gel filtration were degassed by stirring at a vacuum for >15 minutes.

The Coomassie Blue stain was made with 20% acetic acid, 40% ethanol, 0.25% Brilliant Blue G, and MilliQ (q.s). Destaining was performed with 100% MilliQ or 50% MilliQ, 40% ethanol, and 10% acetic acid.

Protein Production

The designated E. Coli bacteria were transformed with the pET23a+_Ac132_TEV_HIS plasmid and/or pET28_p78/83_TEV_HIS (co-expression test). 1 µL plasmid (100 ng/µL) was added to 50 µL bacterial stock and kept on ice for 20 minutes. The samples were heat-shocked at 42°C for 60-90 seconds and put on ice. 1 mL of LB medium containing 1:1000 dilutions of the appropriate antibiotics (Ampicillin (Amp) (100 ng/mL); Chloramphenicol (Cam) (34 ng/mL); Kanamycin (Kan) (.. ng/mL) was added to the bacteria. They were incubated for 1h at 37°C and 220 rpm. The cells were concentrated by removing 900 µL of the medium after centrifugation. The concentrated sample was streaked on lab-made agar plates (+antibiotics) and incubated at 37°C overnight. One colony was added to 2 mL of LB (+ antibiotic) and incubated overnight at 37°C (220 rpm). The main culture was prepared with the seed culture and fresh LB medium (+antibiotic) in a 1:10 ratio and incubated until it reached an optical density (OD) of approximately 0.6 (measured with the Biowave Cell Density Meter CO8000). The I+ samples were then induced with 0.5M IPTG (2 mL per condition) to initiate protein expression. Both the non-induced and induced conditions were incubated overnight at 20°C. The media was removed from the samples via centrifugation (2 minutes, 13000 rpm) and 500 µL of the designated lysis buffer was added to the pelleted cells. The cells were lysed via sonication (10 sec on, 10 sec off; 10 Amp). The I+ samples were centrifuged to separate the cell debris from the soluble proteins (20 min, 13000 rpm). A sample of the I+ sample was taken before centrifugation. The pellet was resuspended in 500 µL lysis buffer.

Protein production on a larger scale was performed by culturing 750 mL of LB medium (+ antibiotic) with 5 or 10 mL of seed culture until an OD of 0.6 was reached. Cells were centrifuged after induction for 30 minutes at 14000 rpm with a high-speed centrifuge and lysed via sonication. Besides the described components, the lysis buffer contained 5 mg/mL lysozyme (from chicken egg white) and 5 µL Nuclease® Benzonase. The supernatant was separated from the pellet for 30 minutes at 13000 rpm with the high-speed centrifuge and prepared for further processing.

Protein Expression Test

The proteins in the samples for the protein expression test were separated via SDS PAGE. Before the gel electrophoresis step, the His-tagged proteins (including Ac132) in I+ supernatant were isolated with immobilized metal ion affinity chromatography (IMAC) using the Ni Sepharose™ 6 Fast Flow beads (ref#: 17-5318-02; GE Healthcare) or an agarose equivalent. Equilibration and washing steps were performed with the designated buffer. Next, 3x loading dye was added to the samples (I-; I+; precipitant; supernatant; IMAC sample) and the samples were denatured at +/- 95°C for 10 minutes. The samples were added to the SDS PAGE gel (12%; 0.75 mm; freshly prepared using the TGX FastCast Acrylamide Kit, 12% (ref#: 1610175, BioRad)) together with the Precision Plus Protein Dual Xtra Standards (ref#: 1610377; BioRad) as a protein marker. The proteins were separated by gel electrophoresis (15 min, 100V; 45 min, 150V) and the gel was stained with Coomassie Blue for 1h. It was then destained overnight with either 100% MilliQ or 50% MilliQ, 40% ethanol, and 10% acetic acid to remove the remaining dye.

Western Blot

The SDS PAGE gel was transferred on an activated (100% EtOH, 3 minutes; Transfer Buffer (25mM Tris, 200 mM glycine, 20% ethanol), until the membrane sank) PVDF membrane for 1h at 100V. The membrane was blocked for 1h at room temperature in 5% skim milk (in TBS-T) (whilst rocking). The

membrane was incubated with 5 mL 1:4000 HIS-Tag HRP-conjugated antibody (ref#: MAB050H, Bio-techne) in 1% skim milk (TBS-T) overnight at 4 ° whilst rocking. The membrane was then washed three times with TBS-T for 10 minutes. ECL™ Prime Western Blotting Detection Reagent (ref#: RPN2232; GE Healthcare) was added to the membrane. The chemiluminescent signal was measured with the ChemiDoc™ Touch Imaging System (Bio-Rad).

Mass spectrometry

The sample for mass spectrometry analysis was obtained from the insoluble fraction on a Coomassie-stained SDS PAGE gel. The excised band was sent for protein identification to the Monash Proteomics and Metabolomics Facility in a clean Eppendorf tube. The protein was analyzed with a Nano LC System (Dionex Ultimate 3000 RSLCnano), a Fusion (Thermo Scientific) mass spectrometer, an Acclaim PepMap RSLC (75 µm x 50 cm, nanoViper, C18, 2 µm, 100Å; Thermo Scientific) analytical column, and an Acclaim PepMap 100 (100 µm x 2 cm, nanoViper, C18, 5 µm, 100Å; Thermo Scientific) trap column. The sequences were found with the Byonic (ProteinMetrics) search engine using the Uniprot Human and Ecoli databases and the Ac132_TEV_HIS sequence.

Ni^{2+}-NTA Affinity Chromatography

The protein was purified via nickel affinity chromatography using the Ni-NTA 5 mL HisTrap™ Fast Flow column (ref#: 17-5255-01, Cytiva). The supernatant was collected and filtered using a 22 µm syringe filter. The column was equilibrated with 10 mL MilliQ (filtered), 10 mL $NiCl_2$ (0.1M, filtered), 10 mL MilliQ (filtered), and 10 mL equilibration buffer. The filtered supernatant was loaded on the column after and the column was washed with 50 mM, 150 mM (buffer screen only), 250 mM, and 500 mM imidazole concentration in the elution buffer. All steps were collected individually on ice. The fractions were sampled for SDS PAGE and stored at 4 °C.

Refolding

The proteins were produced in 750 mL using the previously described scale-up protocol. The cells were induced for 4 hours at 37 °C with 0.5M IPTG. After induction, the cells were centrifuged at 4000 rpm for 30 minutes, resuspended in 30 mL 50 mM Tris-HCl, pH 8.0; 150 mM NaCl; 5 mg/mL lysozyme, and stored O/N at -80°C. After thawing the cell suspensions, 5µL Nuclease® Benzonase was added to the solution. Protein unfolding and refolding were performed as described by *Vincentelli et al., 2004*[45]. After lysis by sonication, the pellets were centrifuged for 30 minutes at 14000 rpm, resuspended in 20 mL 50 mM Tris-HCl, pH 8.0; 150 mM NaCl; 10 mM Imidazole; 8M urea, and centrifuged at 14000 rpm for 30 minutes. The supernatants were diluted to a concentration of 6M urea with a 50 mM Tris-HCl, pH 8.0; 150 mM NaCl; 10 mM imidazole buffer. The supernatants were loaded onto a preequilibrated (50 mM Tris-HCl, pH 8.0; 150 mM NaCl; 10 mM imidazole; 6M urea) His-TRAP FF column (ref#: 17-5255-01, Cytiva) and the proteins (Ac132; p78/83) were obtained via washing/elution with 24 mL 50mM, 250 mM, and 500 mM imidazole buffer (50 mM Tris-HCl, pH 8.0; 150 mM NaCl; 6M urea). Fractions of 4 mL were obtained throughout each step. Protein purity and integrity were assessed in all fractions via SDS PAGE. The fractions containing the protein(s) of interest were pooled and concentrated to a concentration of >5mg/mL with Amicon® Ultra-15 centrifugal filters (ref#: UFC900324; Merck Millipore). The imidazole concentration was diluted with 50 mM Tris-HCl, pH 8.0; 150 mM NaCl; 6M urea via centrifugal filtration. Protein concentrations were determined with the Implen-80 nanospectrophotometer and diluted to a concentration of 5 mg/mL using the predicted

molar extinction coefficient of 18910 (ProtParam). 5 µL of protein was added to the 96 refolding mixes (see: table S.5.) in a round-bottom 96-well plate. The buffers were premixed and the plates were stored at -20°C until use. DTT was added after thawing. The absorbance was measured directly after the addition of the protein with the PheraStar plate reader at 340 nm. A blank was measured before adding the protein and was subtracted from the final value. A second measurement was taken after 24 hours. Results were further processed in RStudio (v4.1.2).

2 mg of the protein was diluted 8x or 20x in refolding buffer for the scale-up. The sample was centrifuged for 30 minutes at 13000 rpm after 2h and after 72h to check for unwanted precipitation. The sample was then concentrated to approximately 4 mg/mL and filtered before injection on the size exclusion column. The concentrations were measured with the Implen-80 nanospectrophotometer after each step.

Size Exclusion Chromatography

The column (HiLoad® 16/600 Superdex® 200 pg; ENrich™ SEC 650 10x300) was equilibrated with 1.5 CV of 100% Buffer A (refolding buffer; filtered and degassed). 500 µL of the filtered sample was injected into the BioRad NGC Chromatography System and eluted at 0.5 mL/min with 100% Buffer A. Fractions were collected manually or when the threshold of 40 mAU was reached. The fractions were stored at 4°C and analysed with SDS PAGE and western blotting.

A calibration curve was constructed using the high molecular weight standards of the Gel Filtration HMW Calibration Kit (ref#: 28403842; Cytiva). Chicken egg white lysozyme (3 mg/mL, filtered) was added as a low molecular weight standard. The standards were diluted according to the manual in a 50 mM Tris-HCl, pH 8.0, 200 mM NaCl buffer (buffer A) and injected in the HiLoad® 16/600 Superdex® 200 pg column.

References

1. Popham, H. J. R., Nusawardani, T. & Bonning, B. C. Introduction to the Use of Baculoviruses as Biological Insecticides. *Methods Mol. Biol.* **1350**, 383–392 (2016).

2. Schaly, S., Ghebretatios, M. & Prakash, S. Baculoviruses in Gene Therapy and Personalized Medicine . *Biol. Targets Ther.* **15**, 115–132 (2021).

3. Overview of the Role of Actin in the Intracellular Transport of Baculovirus. *Adv. Biochem. Biotechnol.* (2021) doi:10.29011/2574-7258.010105.

4. Ogut, O., Hossain, M. M. & Jin, J.-P. Interactions between Nebulin-like Motifs and Thin Filament Regulatory Proteins*. *J. Biol. Chem.* **278**, 3089–3097 (2002).

5. Airenne, K. J. *et al.* Baculovirus: an Insect-derived Vector for Diverse Gene Transfer Applications. *Mol. Ther.* **21**, 739–749 (2013).

6. Ayres, M. D., Howard, S. C., Lopez-Ferber, M., Lopez-Ferber, M. & Possee, R. D. The complete DNA sequence of Autographa californica nuclear polyhedrosis virus. *Virology* **202**, 586–605 (1994).

7. Carstens, E. B. AcMNPV As A Model for Baculovirus DNA Replication *. *Virol. Sin.* **24**, 243–267 (2009).

8. Coulibaly, F. *et al.* The atomic structure of baculovirus polyhedra reveals the independent emergence of infectious crystals in DNA and RNA viruses. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 22205–22210 (2009).

9. Introduction to the Use of Baculoviruses as Biological Insecticides. doi:10.1007/978-1-4939-3043-2_19.

10. Wang, Q. *et al.* Budded baculovirus particle structure revisited. *J. Invertebr. Pathol.* **134**, 15 (2016).

11. Blissard, G. W. & Theilmann, D. A. Baculovirus Entry and Egress from Insect Cells. *https://doi-org.proxy.library.uu.nl/10.1146/annurev-virology-092917-043356* **5**, 113–139 (2018).

12. Monteiro, F., Carinhas, N., Carrondo, M. J. T., Bernal, V. & Alves, P. M. Toward system-level understanding of baculovirus-host cell interactions: From molecular fundamental studies to large-scale proteomics approaches. *Front. Microbiol.* **3**, 391 (2012).

13. Slack, J. & Arif, B. M. The Baculoviruses Occlusion-Derived Virus: Virion Structure and Function. *Adv. Virus Res.* **69**, 99 (2006).

14. Rohrman, G. Structural proteins of baculovirus occlusion bodies and virions - Baculovirus Molecular Biology - NCBI Bookshelf. *NCBI* 21–56 https://www.ncbi.nlm.nih.gov/books/NBK543455/ (2019).

15. Wang, M. & Hu, Z. Advances in the Molecular Biology of Baculoviruses. doi:10.21775/cimb.034.183.

16. Fay, N. & Panté, N. Nuclear entry of DNA viruses. *Front. Microbiol.* **6**, 467 (2015).

17. Okano, K., Vanarsdall, A. L., Mikhailov, V. S. & Rohrmann, G. F. Conserved molecular systems of the Baculoviridae. *Virology* **344**, 77–87 (2006).

18. Schultz, K. L. W. & Friesen, P. D. Baculovirus DNA Replication-Specific Expression Factors Trigger Apoptosis and Shutoff of Host Protein Synthesis during Infection. *J. Virol.* **83**, 11123–11132 (2009).

19. Miele, S. A. B., Garavaglia, M. J., Belaich, M. N. & Ghiringhelli, P. D. Baculovirus: Molecular Insights on Their Diversity and Conservation. *Int. J. Evol. Biol.* **2011**, 1–15 (2011).

20. Zhao, S., He, G., Yang, Y. & Liang, C. Nucleocapsid Assembly of Baculoviruses. *Viruses* **11**, (2019).

21. Vanarsdall, A. L., Pearson, M. N. & Rohrmann, G. F. Characterization of baculovirus constructs lacking either the Ac 101, Ac 142, or the Ac 144 open reading frame. *Virology* **367**, 187 (2007).

22. Ohkawa, T., Volkman, L. E. & Welch, M. D. Actin-based motility drives baculovirus transit to the nucleus and cell surface. *J. Cell Biol.* **190**, 187 (2010).

23. Goley, E. D. *et al.* Dynamic nuclear actin assembly by Arp2/3 complex and a baculovirus WASP-like protein. *Science (80-.).* **314**, 464–467 (2006).

24. Machesky, L. M. & Insall, R. H. WASP homology sequences in baculoviruses. *Trends Cell Biol.* **11**, 286–287 (2001).

25. Hepp, S. E., Borgo, G. M., Ticau, S., Ohkawa, T. & Welch, M. D. Baculovirus AC102 Is a Nucleocapsid Protein That Is Crucial for Nuclear Actin Polymerization and Nucleocapsid Morphogenesis. *J. Virol.* **92**, 111–129 (2018).

26. Wang, Y. *et al.* Autographa californica Multiple Nucleopolyhedrovirus Nucleocapsid Protein BV/ODV-C42 Mediates the Nuclear Entry of P78/83. *J. Virol.* **82**, 4554–4561 (2008).

27. Zhang, Y. *et al.* Ac102 Participates in Nuclear Actin Polymerization by Modulating BV/ODV-C42 Ubiquitination during Autographa californica Multiple Nucleopolyhedrovirus Infection. *J. Virol.* **92**, (2018).

28. Li, K. *et al.* The Putative Pocket Protein Binding Site of Autographa californica Nucleopolyhedrovirus BV/ODV-C42 Is Required for Virus-Induced Nuclear Actin Polymerization. *J. Virol.* **84**, 7857 (2010).

29. Yang, M., Wang, S., Yue, X.-L. & Li, L.-L. Autographa californica Multiple Nucleopolyhedrovirus orf132 Encodes a Nucleocapsid-Associated Protein Required for Budded-Virus and Multiply Enveloped Occlusion-Derived Virus Production. *J. Virol.* **88**, 12586 (2014).

30. Fang, Z., Li, C., Wu, W., Yuan, M. & Yang, K. Y. The Autographa californica multiple nucleopolyhedrovirus Ac132 plays a role in nuclear entry. *J. Gen. Virol.* **97**, 3030–3038 (2016).

31. Berretta, M. F. *et al.* Baculovirus Gene Expression. *Curr. Issues Mol. Virol. - Viral Genet. Biotechnol. Appl.* (2013) doi:10.5772/56955.

32. Peng, Y., Liang, A. & Fu, Y. Function and Application Analysis of Ac132 Protein in Autographa californica Multiple Nucleopolyhedrovirus. *Biotechnol. Bioprocess Eng.* **23**, 655–661 (2018).

33. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nat. 2021 5967873* **596**, 583–589 (2021).

34. Zhou, Y. *et al.* Characterization of bombyx mori nucleopolyhedrovirus ORF109 that encodes a 25-kDa structural protein of the occlusion-derived virion. *Curr. Microbiol.* **61**, 451–457 (2010).

35. Ono, C. *et al.* Phenotypic grouping of 141 BmNPVs lacking viral gene sequences. *Virus Res.* **165**, 197–206 (2012).

36. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539 (2011).

37. Walker, J. M. *et al.* Protein Identification and Analysis Tools on the ExPASy Server. *Proteomics*

Protoc. Handb. 571–607 (2005) doi:10.1385/1-59259-890-0:571.

38. Functional Analysis of AcMNPV P78/83 Protein. https://www.virosin.org/en/article/id/VS20041905.0498.

39. Hill, A. F., Barnham, K. J., Bottomley, S. P. & Cappai, R. *Protein Folding, Misfolding, and Disease*. vol. 752 (Humana Press, 2011).

40. Hebditch, M., Carballo-Amador, M. A., Charonis, S., Curtis, R. & Warwicker, J. Protein–Sol: a web tool for predicting protein solubility from sequence. *Bioinformatics* **33**, 3098 (2017).

41. Esposito, D. & Chatterjee, D. K. Enhancement of soluble protein expression through the use of fusion tags. doi:10.1016/j.copbio.2006.06.003.

42. Burgess-Brown Editor, N. A. Heterologous Gene Expression in E. coli Methods and Protocols Methods in Molecular Biology 1586.

43. Chen, N., Kong, X., Zhao, S. & Xiaofeng, W. Post-translational modification of baculovirus-encoded proteins. *Virus Res.* **279**, 197865 (2020).

44. Yin, R., Feng, B. Y., Varshney, A. & Pierce, B. G. Benchmarking AlphaFold for protein complex modeling reveals accuracy determinants. *Protein Sci.* **31**, (2022).

45. Vincentelli, R. *et al.* High-throughput automated refolding screening of inclusion bodies. *Protein Sci.* **13**, 2782 (2004).

Supplementary Material

S1. Plasmid Construct (pET23a+_Ac132_TEV_HIS)

S1.1. Plasmid Map

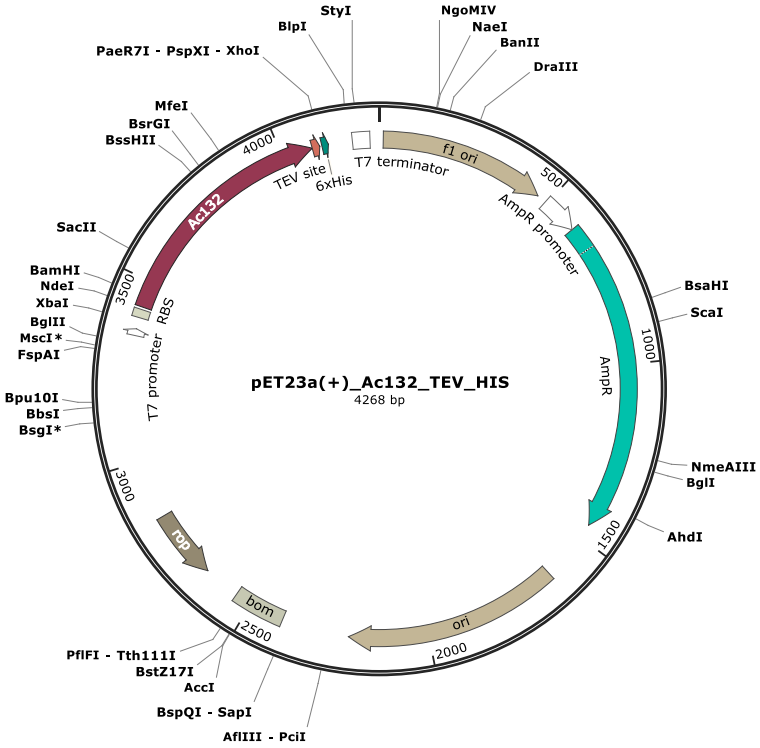


Figure S1. 1 – Plasmid map of pET23a(+)_Ac132_TEV_HIS. Constructed in SnapGene. The plasmid (4268 bp) is composed of the pET23a(+) vector with the AC132 gene and TEV cleavage site as inserts. The plasmid further contains an ampicillin resistance gene (AmpR), a ribosome binding site (RBS), and a His-Tag (6xHis). The plasmid was synthesized by GenScript®.

S2. Oligomerization

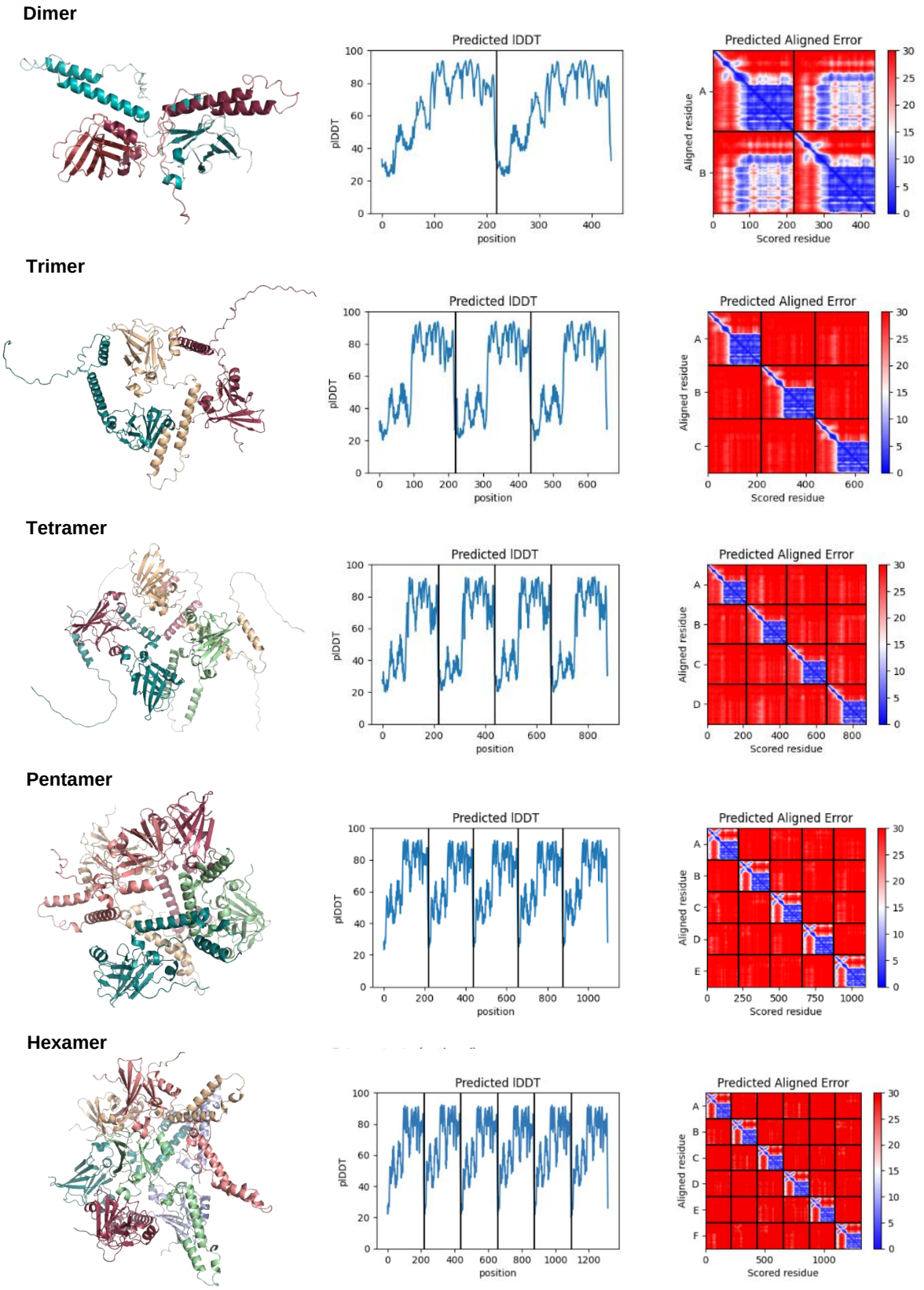


Figure S2. 1 – Predicted oligomerization structures using the AlphaFold2 algorithm. Ac132 oligomerization as predicted by AlphaFold2 (visualized with PyMol). The pIDDT scale illustrates the confidence in the prediction per monomer. The Predicted Aligned Error visualizes the error rate in the spatial arrangement per individual monomer (monomer A: residues 0-219; monomer B: residues 220-438; etc.) and between monomers. The monomers are coloured individually.

S3. Multiple Sequence Alignment

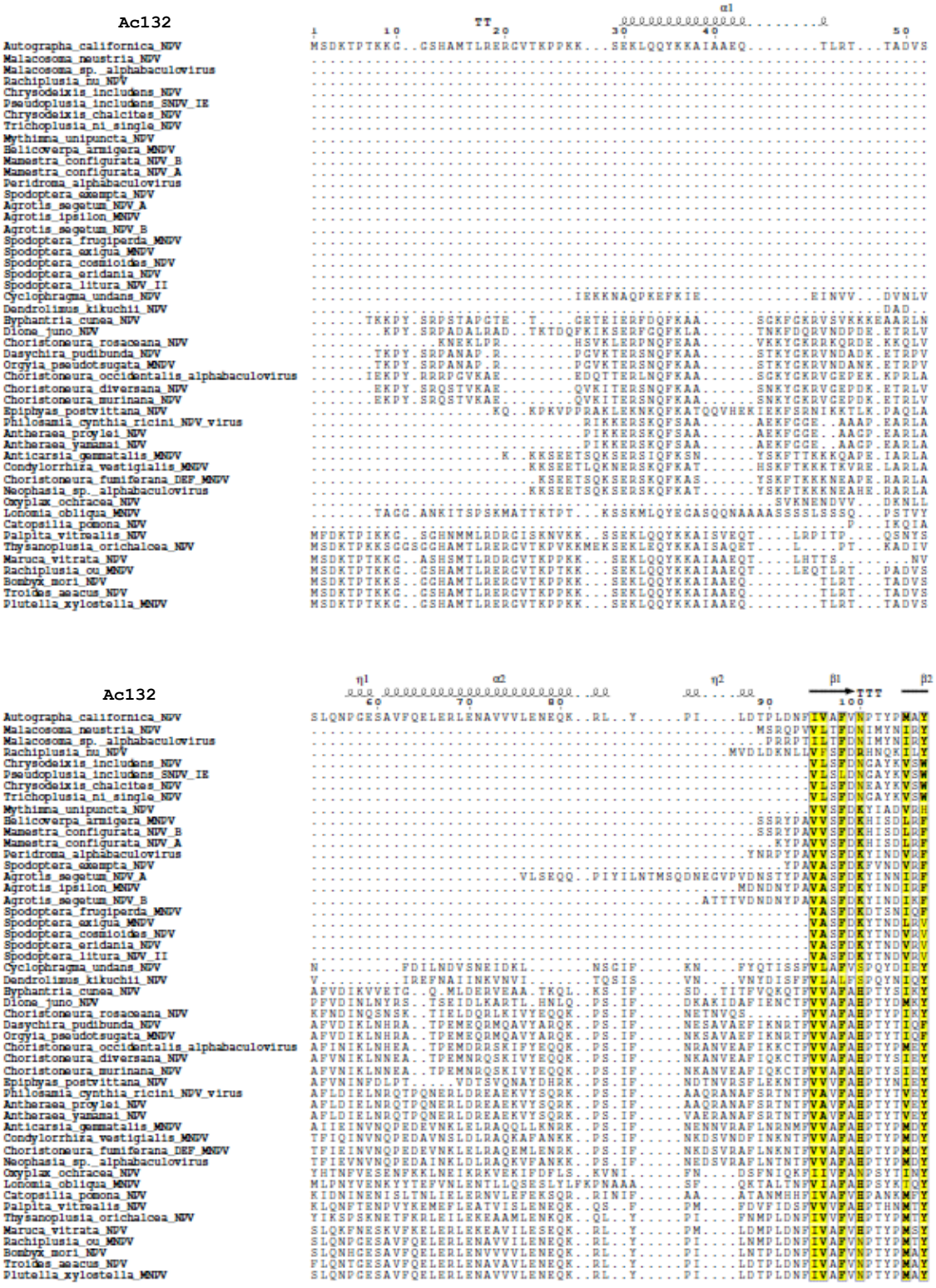


Figure S3.1 – Sequence alignment of 50 nonredundant Ac132 sequences. Sequences were obtained with NCBI PSI-BLAST and Clustal Omega and processed in ESPript. Secondary structures were predicted via AlphaFold2 and represented above the alignments. The numbering and secondary structures correspond with the Autographa Californica nucleopolyhedrovirus (NPV) sequence 1. Residues in yellow boxes with blue frames are similar accros the various sequences. Strictly conserved sequence are illustrated with white characters in a red box.

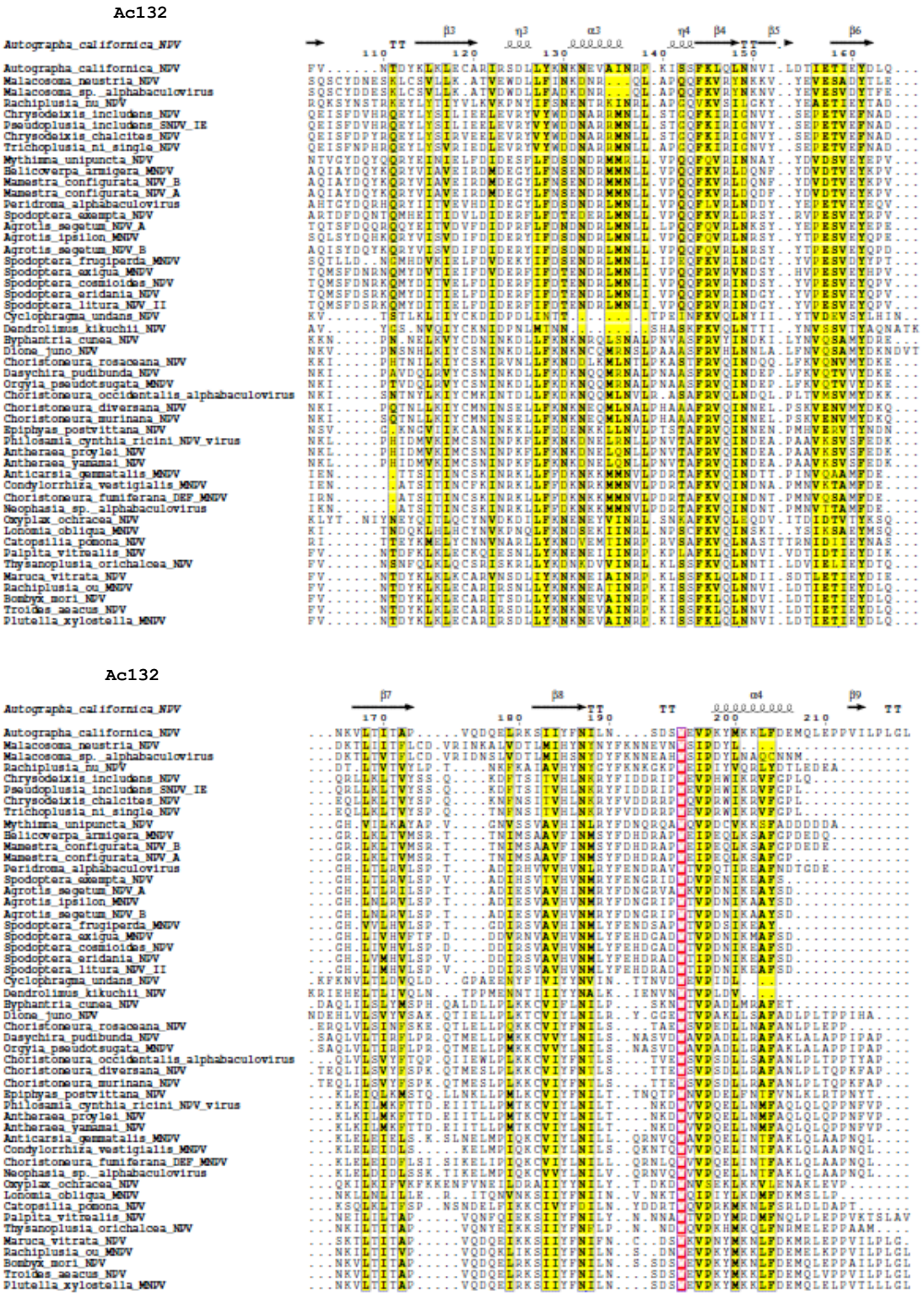


Fig. S3.1 continued.

S4. Buffer Screen

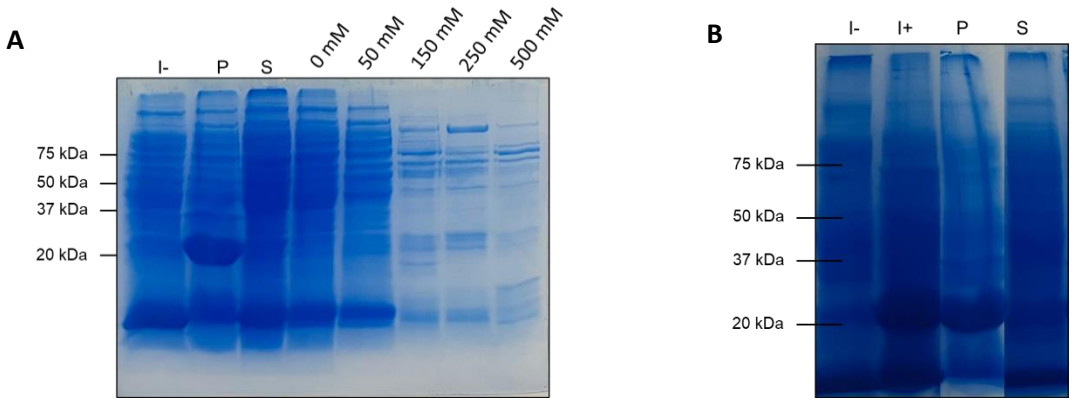


Figure S4.1 – SDS PAGE analysis of large scale Ac132 expression with 500 mM NaCl, 50 mM CHES, pH 9.0, beta-mercaptoethanol (BME) and Triton X-100. Ac132 has an expected size of 25.1 kDa. The pET23a(+)_Ac132_TEV_HIS plasmid was transformed in C41(DE3) bacteria and Ac132 was expressed in 500 mL culture. The bacteria were lysed via sonication using a 50 mM CHES, pH 9.0, 500 mM NaCl, 1% Triton X-100, 5mM BME lysis buffer with 5mg/mL lysozyme and 5 µL Benzonase® Nuclease. The samples have been divided in IPTG-induced (I+) and non-induced (I-) samples. The induced samples have been separated in a supernatant (soluble fraction, S) and a precipitant (insoluble fraction, P). The samples were separated in a reducing SDS PAGE and stained with Coomassie Blue Scale-up of the protein expression. The supernatant was filtered and purified with a 5 mL HIS-Trap FF column using various imidazole concentrations (0-500 mM) (A). The pellet was re-lysed via sonication using the same lysis buffer (B).

S5. Protein Refolding

S5.1 Ni-NTA Purification Results

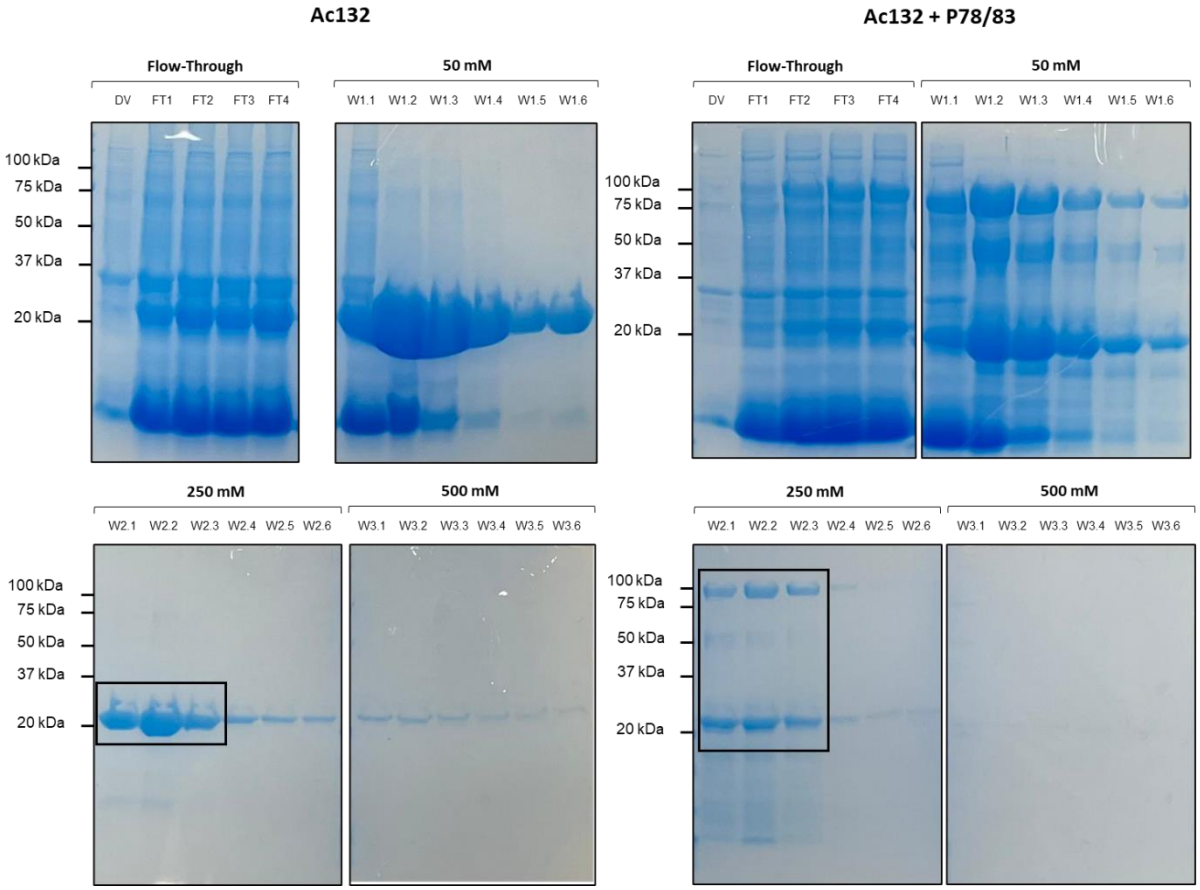


Figure S5. 1 – SDS PAGE analysis of the fractions obtained by Ni-NTA affinity chromatography purification. Purification was performed with a 5mL His-TRAP FF column and a peristaltic pump using 50 mM, 250 mM, and 500 mM imidazole (+ 50 mM Tris-HCl, pH 8.0, 150mM NaCl, 6M urea) as wash/elution buffers. The supernatant was unfolded in 8M urea (50 mM Tris-HCl, pH 8.0, 150mM NaCl, 10 mM imidazole) and diluted to 6M urea with a 50 mM Tris-HCl, pH 8.0, 150mM NaCl, 10 mM imidazole buffer before administration to the system. Fractions were obtained each 4mL and protein purity was assessed by SDS PAGE. The fractions chosen for the refolding experiment are indicated by a black box.

S5.2 Refolding Buffer Mixes

Table S5.1 – Plate layout of the 96 refolding buffers. Each buffer contains a buffer to obtain the correct pH value, 0 or 200 mM NaCl, 1 µL DTT, 5 µL protein, and the designated chemical additive. The wells are filled to 100 µL with MilliQ.

		1	2	3	4	5	6	7	8	9	10	11	12
		pH 5	pH 5	pH 6	pH 6	pH 7	pH 7	pH 8	pH 8	pH 9	pH 9	pH 10	pH 10
	NaCl	0 mM	200 mM	0 mM	200 mM	0 mM	200 mM	0 mM	200 mM	0 mM	200 mM	0 mM	200 mM
A	50 mM Arginine		NaCl		NaCl		NaCl		NaCl		NaCl		NaCl
		L-arginine	L-arginine	L-arginine	L-arginine	L-arginine	L-arginine	L-arginine	L-arginine	L-arginine	L-arginine	L-arginine	L-arginine
		MES	MES	HEPES	HEPES	HEPES	HEPES	Tris-HCl	Tris-HCl	CHES	CHES	CHES	CHES
B	500 mM glucose		NaCl		NaCl		NaCl		NaCl		NaCl		NaCl
		Glucose	Glucose	Glucose	Glucose	Glucose	Glucose	Glucose	Glucose	Glucose	Glucose	Glucose	Glucose
		MES	MES	HEPES	HEPES	HEPES	HEPES	Tris-HCl	Tris-HCl	CHES	CHES	CHES	CHES
C	20 % glycerol		NaCl		NaCl		NaCl		NaCl		NaCl		NaCl
		Glycerol	Glycerol	Glycerol	Glycerol	Glycerol	Glycerol	Glycerol	Glycerol	Glycerol	Glycerol	Glycerol	Glycerol
		MES	MES	HEPES	HEPES	HEPES	HEPES	Tris-HCl	Tris-HCl	CHES	CHES	CHES	CHES
D	1mM EDTA /1mM ions		NaCl		NaCl		NaCl		NaCl		NaCl		NaCl
		MgCl2	EDTA	NiCl2	CaCl2	EDTA	MgCl2	NiCl2	EDTA	CaCl2	MgCl2	CaCl2	MgCl2
		MES	MES	HEPES	HEPES	HEPES	HEPES	Tris-HCl	Tris-HCl	CHES	CHES	CHES	CHES
E	0.05% PEG4000		NaCl		NaCl		NaCl		NaCl		NaCl		NaCl
		PEG 4000	PEG 4000	PEG 4000	PEG 4000	PEG 4000	PEG 4000	PEG 4000	PEG 4000	PEG 4000	PEG 4000	PEG 4000	PEG 4000
		MES	MES	HEPES	HEPES	HEPES	HEPES	Tris-HCl	Tris-HCl	CHES	CHES	CHES	CHES
F	0.05% PEG 400		NaCl		NaCl		NaCl		NaCl		NaCl		NaCl
		PEG 400	PEG 400	PEG 400	PEG 400	PEG 400	PEG 400	PEG 400	PEG 400	PEG 400	PEG 400	PEG 400	PEG 400
		MES	MES	HEPES	HEPES	HEPES	HEPES	Tris-HCl	Tris-HCl	CHES	CHES	CHES	CHES
G	0.1% Triton X100		NaCl		NaCl		NaCl		NaCl		NaCl		NaCl
		Triton X100	Triton X100	Triton X100	Triton X100	Triton X100	Triton X100	Triton X100	Triton X100	Triton X100	Triton X100	Triton X100	Triton X100
		MES	MES	HEPES	HEPES	HEPES	HEPES	Tris-HCl	Tris-HCl	CHES	CHES	CHES	CHES
H	100 mM urea		NaCl		NaCl		NaCl		NaCl		NaCl		NaCl
		Urea	Urea	Urea	Urea	Urea	Urea	Urea	Urea	Urea	Urea	Urea	Urea
		MES	MES	HEPES	HEPES	HEPES	HEPES	Tris-HCl	Tris-HCl	CHES	CHES	CHES	CHES

S5.3 Calibration Curve

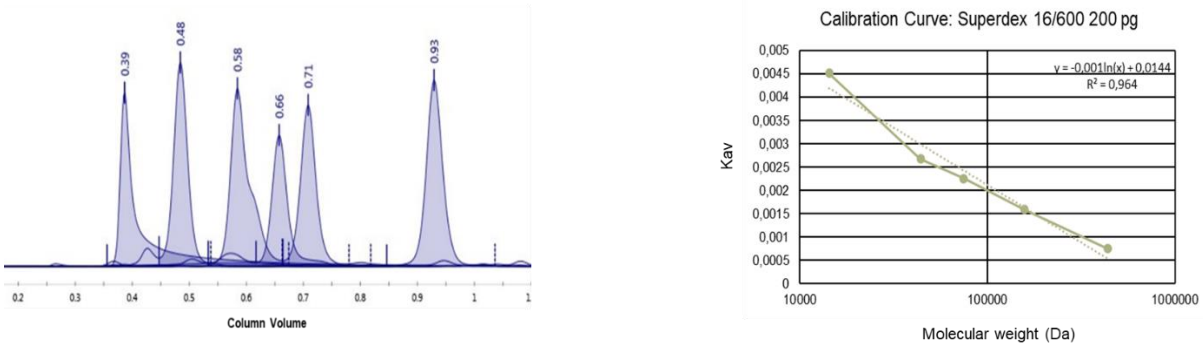
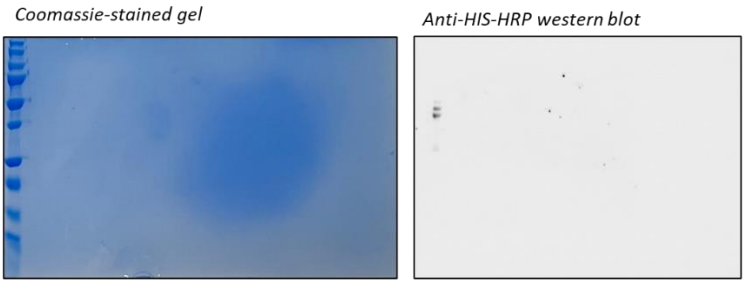


Figure S5. 2 – Calibration curve with protein standards on the Superdex 16/600 200 pg column. Standards were obtained from the Cytiva high molecular weight protein standards (0.39, void volume, Blue Dextran; 0.48, Ferritin; 0.58, Aldolase; 0.66, Conalbumin; 0.71, Ovalbumin). Chicken egg white lysozyme (14.4 kDa; 0.93) was separately included as a low molecular weight standard. Samples were diluted in 50 mM Tris-HCl, pH 8.0, 200 mM NaCl. 500µL of protein sample was injected in the BioRad NGC chromatography system. Gel filtration was performed on a Superdex 16/600 200 pg column with a 50 mM Tris-HCl, pH 8.0, 200 mM NaCl running buffer. The chromatogram was visualized with ChromLab software and the data was processed in Excel. The calibration curve was constructed with the molecular weights of the standards and their corresponding elution volumes (kav = Ve-V0/Vc-Vo; with Ve = elution volume, V0 = void volume, Vc = column volume). As it the curve is based on molecular weight, not the hydrodynamic surface, the molecular weights of nonglobular proteins may defer from the predicted values.

S5.5 Protein Validation



Column	Selected samples	CV	Additive
1	Concentrated Ac132 (single expression)	1.08	L-arginine
2	Flow-through Ac132 (single expression);	1.08	L-arginine
3	Fraction 2 (co-expression)	0.99	L-arginine
4	Fraction 3 (co-expression);	0.99	L-arginine
5	Fraction 4 (co-expression);	0.99	L-arginine
6	Concentrated Ac132 (Co-expression)	1.08	L-arginine
7	Flow-through Ac132 (Co-expression)	1.08	L-arginine
8	Fraction 16 (co-expression)	0.75	EDTA
9	Concentrated Ac132 (co-expression)	1.8	EDTA
10	Flow-through Ac132 (Co-expression)	1.8	EDTA

Figure S5. 3 – Protein validation of Size Exclusion Chromatography (SEC) samples. The samples were analysed via SDS PAGE and western blotting. The samples were separated in a reducing SDS PAGE and stained with Coomassie Blue or incubated with an anti-HIS-HRP antibody on a PVDF membrane. The table indicates the samples that were loaded with the column volume (CV) indicating the location of the fractions on the corresponding chromatogram.

S5.6 UV spectra

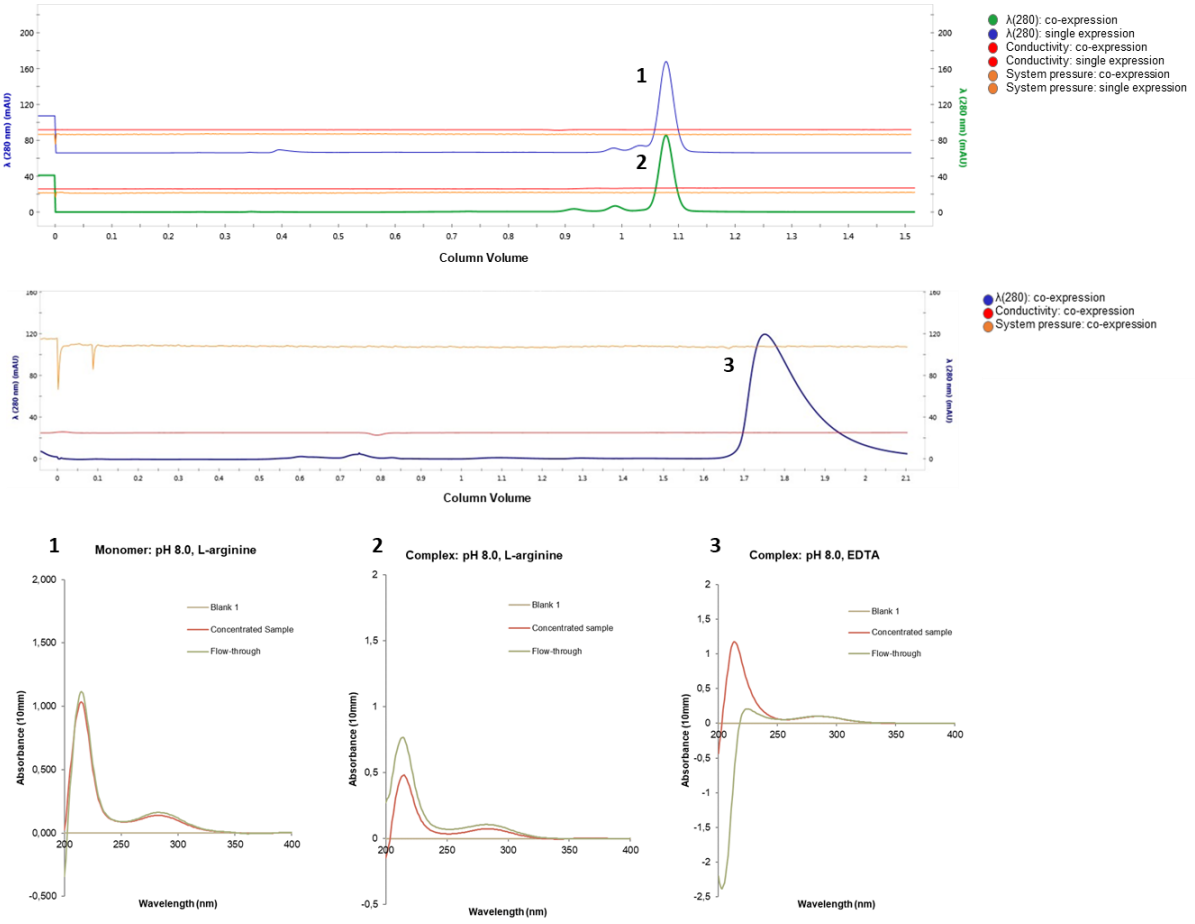


Figure S5. 4 – Nanospectrophotometer UV spectra of selected Size Exclusion Chromatography (SEC) fractions. The fractions obtained with SEC as numbered in the chromatograms (top: pH 8.0, 200 mM NaCl, 50 mM L-arginine (Ac132: blue; Ac132-p78/83: green); bottom: pH 8.0, 200 mM NaCl, 1 mM EDTA (Ac132-p78/83: blue)) were collected. The UV spectra were measured by the Implen-80 nanospectrophotometer and analysed in Excel. The blank is the corresponding refolding buffer.

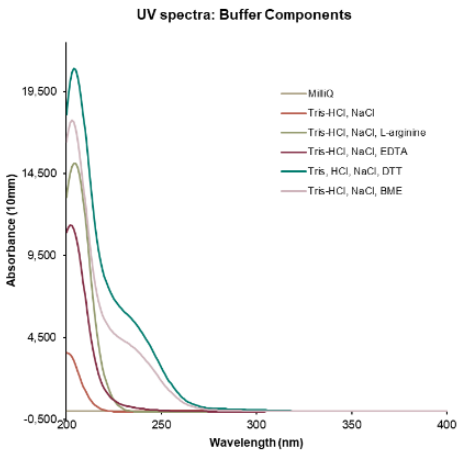


Figure S5. 5 – Nanospectrophotometer UV spectra of buffer components. The UV spectra of the chemicals in the refolding buffers were separately measured with the Implen-80 nanospectrophotometer with MilliQ as the blank. The spectra were analysed in Excel.

SPECIAL THANKS TO:

COULIBALY LABORATORY, MONASH BIOMEDICINE DISCOVERY INSTITUTE:

A/PROF. FASSÉLI COULIBALY

BISHWA SUBEDI, PHD

JUNGMIN HA, MSC

BENJAMIN DARGAN, BSC

VIROLOGY DIVISION, UTRECHT UNIVERSITY:

PROF. BEREND-JAN BOSCH