Structural studies of myxoma virus antagonists of innate immunity

Peter Eric O'Byrne

Student ID: 11300046

Submitted in part for a Ph.D. in Biochemistry

School of Biochemistry and Immunology

Trinity College Dublin

Trinity College Dublin
Coláiste na Tríonóide, Baile Átha Cliath
The University of Dublin

July 2019

Funded by Science Foundation Ireland

Declaration

I declare that this thesis has not been submitted as an exercise for a degree at this or any other university and it is entirely my own work. I agree to deposit this thesis in the University's open access institutional repository or allow the library to do so on my behalf, subject to Irish Copyright Legislation and Trinity College Library conditions of use and acknowledgement.

_____

Peter O'Byrne

Abstract

Viral evasion of the host immune system is achieved via antagonism of antiviral signalling pathways by the expression of immunomodulating proteins. Vaccinia virus protein C7 and its homologues function as determinants of host range by antagonising the immune response via interaction with its poorly understood target SAMD9. Myxoma virus (*MYXV*) (the causative agent of myxomatosis), encodes 3 C7 homologues: M062, M063 and M064. M062 alone can complement C7 by binding to SAMD9. Compared to the lab strain Lausanne, *MYXV* vaccine strain Mav, contains a single residue mutant in M062: I79T, which has been shown to dramatically reduce the binding of M062 to SAMD9 (unpublished data, Dr. Jia Liu). While the crystal structures of C7 and *MYXV* homologue M064 are known, the structural basis of C7 antagonism remains poorly understood.

In this study, the crystal structures of M062-Lau (2.45 Å) and M062-Mav (3.2 Å) are presented. The presence of a novel putative dimerization interface is described, supported by *in vitro* evidence determined by size-exclusion chromatography multi-angle light scattering. The elongated C terminal tail preserved in clade II C7 homologues is identified in the M062 crystal structures, and shown to mediate interaction between individual M062 proteins. The newly described intermolecular interface comprises a hydrophobic pocket formed by β strands 11 (β_{11}) and 12 (β_{12}), into which Y152 of the C terminal tail is buried, supported by C-H$\cdots\pi$ bond stabilisation of cis-Pro (P136) by W151. Phylogenetic analysis of the residues proposed to mediate this intermolecular interaction indicates a compensatory adaptation in the OPV C7 homologues (which lack the C terminal tai), introducing an aromatic residue in β_{12} in order to stabilise the conserved cis-Pro of β_{12}, highlighting the $\beta_{11,12}$ interface as a potential region of structural importance in the C7 family.

Structural analyses of the pyrin domain of M013 by NMR and circular dichroism are also described, with the presentation of the first viral pyrin domain structural models. The NMR structural models of the M013 pyrin domain identify surface charge residues predicted to influence M013's antagonism by homotypic Type I interactions with its target: ASC. This study provides a more accurate model with which the molecular basis of M013's immunomodulation can be predicted.

Table of Contents

Abbreviations

Table I: Amino acids and their abbreviations

Amino acid	3-letter code	Single letter code
Alanine	Ala	A
Arginine	Arg	R
Asparagine	Asn	N
Aspartate	Asp	D
Cysteine	Cys	C
Glutamate	Glu	E
Glutamine	Gln	Q
Glycine	Gly	G
Histidine	His	H
Isoleucine	Ile	I
Leucine	Leu	L
Lysine	Lys	K
Methionine	Met	M
Phenylalanine	Phe	F
Proline	Pro	P
Serine	Ser	S
Threonine	Thr	T
Tryptophan	Trp	W
Tyrosine	Tyr	Y
Valine	Val	V

Table II: Abbreviations, Scientific

Abbreviation	Full name
^{13}C	Isotope 13 of carbon
^{15}N	Isotope 15 of nitrogen
2xYT	2x Yeast-Tryptone microbial growth medium
3C-Protease	Recombinant human rhinovirus 3C protease
3CS	Cysteine to serine mutations at position 90, 107 and 143 of M062
4x LB	4x strength SDS-PAGE loading buffer
A_{280}	Absorbance at 280 nm
AIM2	absent in melenoma 2
ASC	Apoptosis-associated speck-like protein containing a CARD
ASU	Asymmetric unit
CARD	Caspase activation and recruitment domain
CCP4	Collaborative Computational Project no. 4
CMM	Check My Metal
CSD	Cambridge Structural Database
CSH	Cold Spring Harbour
CSI	Chemical shift index
CV	Column volumes
DD	Death domain
ddH2O	Double-distilled water
DTT	Dithiothreitol
E. coli	*Escherichia coli*

Abbreviation	Full name
EDTA	Ethylene-diamene-tetraacetic acid
El	Eluate
EtOH	Ethanol
FL	Full-length
FPLC	Fast protein liquid chromatography
FT	Flow-through
HDACs	Histone deacetylases
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
His-	N-terminal hexahistidine sequence
IEX	Ion exchange chromatography
IFN	Type I interferon
IKK	Inhibitor of NFκB kinase
IMAC	Immobilised metal affinity chromatography
iPrOH	Isopropanol
IPTG	Isopropyl ß-D-1-thiogalactopyranose
IRF1	IFN regulatory factor 1
ISGs	Interferon-stimulated genes
ITR	Inverse tandem repeat
KD	Knockdown
KO	Knockout
LB	Lysogeny broth
LIC	Ligation-independent cloning
LN2	Liquid nitrogen
LS	Light scatter

Abbreviation	Full name
MALS	Multi-angle light scattering
MBP	Maltose-binding protein
Milli-Q	Filtered by Millipak Express 40 system (0.22 µm)
MIRAGE	Myelodysplasia, infection, restriction of growth, adrenal hypoplasia, genital phenotypes, and enteropathy
MR	Molecular replacement
Mw	Molecular weight
MWCO	Molecular weight cut-off
NEB	New England Biolabs
NECAT	Northeastern collaborative access team
NFTC	Normophosphataemic tumoral calcinosis
NFκB	Nuclear factor kappa-light-chain-enhancer of activated B cells
NMR	Nuclear magnetic resonance
NTA	Nitrilotriacetic acid
OD_{600}	Optical density at 600 nm
ORF	Open reading frame
PAMP	Pathogen-associated molecular patter
PDB	Protein data bank
PEG	Polyethylene glycol
POP	Pyrin-only protein
PP5	Protein phosphatase 5

Abbreviation	Full name
PRR	Pathogen recognition receptor
PYD	Pyrin domain
RAPD	Rapid automated processing of data
RT	Root temperature, 293 K
SAMD9	Sterile alpha motif domain containing 9
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate-polyacrylamide gel electrophoresis
SEC	Size-exclusion chromatography
SEC-MALS	Size-exclusion chromatography multi-angle light-scattering
STAND	Signal transduction ATPases with numerous domains
TLR	Toll-like receptor
tNCS	Translational non-crystallographic symmetry
TPRs	Tetratricopeptide repeats
TRIS	Tris(hydroxymethyl)aminomethane
vPOP	Viral pyrin-only protein
WT	Wild-type
XRF	X-ray fluorescence scan
β-ME	β-mercaptoethanol
ε_{280}	Extinction co-efficient at 280 nm

Table III: Abbreviations, Viruses

Abbreviation	Full Name	NCBI ref. seq.
APV	Avipoxvirus	Haller *et al.* 2014[1]
CaPV	Capripoxvirus	Haller *et al.* 2014[1]
CePV	Cervidpoxvirus	Haller *et al.* 2014[1]
COPV	Cotia virus	*YP_005296271.1*
CrP	Crocodylidpoxvirus	Haller *et al.* 2014[1]
CWPV	Cowpox virus	*SNB53759.1*
DPV	Deerpox virus	*ABI99060.1*
EPV	Entomopoxvirus B	Haller *et al.* 2014[1]
EPTV	Eptesipox virus	*YP_009408022.1*
FGPV	Flamingopox virus	*YP_009448142.1*
GTPV	Goatpox virus	*YP_001293258.1*
HSPV	Horsepox virus	*ABH08123.1*
LPV	Leporipoxvirus	Haller *et al.* 2014[1]
LSDV	Lumpy skin disease virus	*AAN02635.1*
MOPV	Molluscipoxvirus	Haller *et al.* 2014[1]
MPXV	Monkeypox virus	*AGR34755.1*
MSPV	Moosepox virus	*AYC44651.1*
MYXV	Myxoma virus	Haller *et al.* 2014[1]
Lau	Lausanne	*NP_051776.1*
Mav	Mav	*ADK63702.1*
PEPV	Penguinpox virus	*YP_009046213.1*
PGPV	Pigeonpox virus	*YP_009046445.1*
PPV	Parapoxvirus	Haller *et al.* 2014[1]

Abbreviation	Full Name	NCBI ref. seq.
RCNV	Racoonpox virus	*YP_009143330.1*
SFV	Rabbit fibroma virus	*NP_051951.1*
SFV	Rabbit fibroma virus	*NP_051953.1*
SPPV	Sheeppox virus	*AAM95246.1*
SuPV	Suipoxvirus	Haller *et al.* 2014[1]
SWPV	Swinepox virus	*NP_570224.1*
TPV	Tanapoxvirus	*ABQ43542.1*
VACV	Vaccinia virus	*5CYW*
YLDV	Yaba-like disease virus	*NP_073452.1*
YMTV	Yaba monkey tumour virus	*NP_938322.1*
YPV	Yatapoxvirus	Haller *et al.* 2014[1]
YPV	Yokapox virus	*YP_004821524.1*

Chapter 1: Introduction

1.1. Poxviral immune evasion: antagonism of the immune system by host range factors

1.1.1. Poxviruses and subversion of the immune response

Poxviruses are a group of large, dsDNA viruses that replicate in the cytoplasm of diverse species which they infect, from insects (*Entomopxvirinae*) to vertebrates (*Chordopoxvirinae*)[2]. Productive replication of viruses relies on evasion of the host's defences, which poxviruses achieve through their encoding of antagonists of the immune system. The interplay between the host immune system and the viral antagonists determines the tropism and host range of a virus[2,3]. Poxviruses encode a range of host restriction factors: proteins whose effective moderation of the immune system allows for permissive infection at the cellular, tissue/organ or species specific level[1]. Poxviruses represent a risk to human[4] and domesticated animal populations[5], and understanding the underlying mechanisms that determine host range may prove useful in the case of a zoonotic crossover event.

The presence and specific function of viral host range factors determines the variety of host species in which a virus can establish productive infection. Host range factors represent traits which have dramatic influence on viral replication and reproductive success[3], and so an evolutionary arms race between the host immune system and viruses is established[1,5,67]. The most successful host range genes would result from the evolutionary pressure exerted on viral probing of the host immune system for targets whose modulation enhances viral replication and transmission. Investigation of the method a host range factor has developed to alter a signalling pathway has the potential to reveal key regulatory steps in host immunity and inflammation which mediate antiviral responses.

1.1.1.1. Poxviral host range factors

Host range factors are immunomodulatory proteins encoded by viruses which determine in which species a virus can establish a productive infection. Pathogenesis depends on the interaction between the virus and host, and a specific interaction between the viral antagonist and host target protein is required for effective immunomodulation. Variance in host or viral protein sequences, or the regulation of expression of host target proteins in different cell or tissue types all influence the permissibility of viral infection.

Identification of host range factors is usually achieved by genetic screening: knocking-out viral open reading frames (ORFs), and observing the effect on viral growth in cell culture, or in host pathogenesis[8]. Bioinformatic screening of viral ORFs can also identify immunomodulatory proteins by their homology to conserved structural motifs from host or viral proteins[9,10]. By clustering viruses into groups based on the presence of suspected host range genes, correlation between groups which contain a member of the host range 'family' and the development of pathogenesis can indicate the function of the host range[1112].

1.1.1.2. Phylogenetic classification of poxviruses

Poxviruses as a family have incredibly broad host range, however the specific host range of individual members varies widely[3]. All sequenced poxviral genomes have been found to possess a single copy of a core 33 gene families, analysis of the poxviral family using these genes allowed for phylogenetic classification of the *poxvirinae*[13]. Analysis of the phylogenetic relationship between poxvirus host range factors further classifies poxviruses into different subgroups[1,12] (see Fig. 1).

Consistently, the poxvirus family is divided primarily into *Entomopoxvirinae* and *Chordopoxviriniae* groups. Phylogenetic analysis consistently sub-divides the *Chordopoxviriniae* into the orthopoxvirus and clade II sub-groups[1,12–14]. It is within the clade II family that *myxoma virus* (MYXV) is found (see Fig. 1).

1.2. ***Myxoma virus***

Myxoma virus (MYXV), the prototype of the *Leporipoxvirus* genus, is the causative agent of myxomatosis, a disease lethal in European rabbit (*Oryctolagus cuniculus*), but benign in its natural hosts[5]. *O. cuniculus* and *Sylvilagus* rabbits diverged approximately 10 million years ago[15]. There are two major, geographically distinct strains of *MYXV*: the Central and South American strain, with natural host *Sylvilagus brasiliensis*, and the Californian strain which infects *Sylvilagus bachmani*[5].

1.2.1.1. The myxoma virus genome

The fully sequenced genome of the *myxoma virus* standard lab strain (SLS) was published in 1999, and predicted 159 co-linearly expressed open-reading frames (ORFs), 12 of which are diploid, encoded twice, once in each inverse tandem repeat (ITR), giving a total of 171 genes. Analysis of the location of genes and their function shows that more highly conserved housekeeping and structural genes are located in the central region of the genome, while the immunomodulatory and host range genes mainly cluster in or nearer to the ITRs. The central core of the genome spanning M012L – M140R bears significant homology to the *vaccinia virus* (*VACV*) genome from F2L – A55R, however, a significant number of *VACV* genes are excluded in MYXV[16]. Of particular note is the presence of 3 copies of homologues of the *C7L* family, arising from two separate gene duplication events[11].

1.2.2. Myxomatosis

Myxomatosis is the name given to the disease caused by myxoma virus in the European rabbit. Its name derives from the mucoid appearance of the fibroma-like lesions the disease produces in the rabbits. In its natural hosts, in the Americas, myxoma virus causes a minor disease, with cutaneous fibroma, and self-resolves. Both strains are fatal in European rabbits however, causing swelling of the face and head, and mucoid cutaneous tumours, known as fibroma[5].

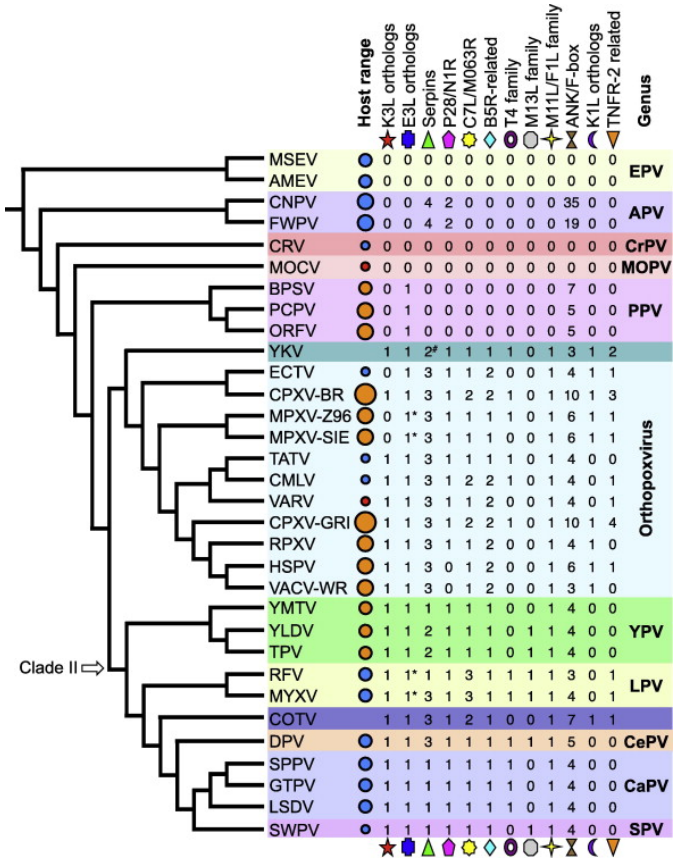


Figure 1: Presence of host range gene families in poxviruses. Depicted is a phylogenetic analysis of poxviruses, published by Haller *et al.* 2014[1]. Branch lengths are not drawn to scale. Circle size correlates with the relative host range of each virus: limited host range (small), broad host range (large), absence of a circle indicates insufficient data to project host range. Human-specific viruses are marked by red circles, viruses that can cause zoonotic infections in humans, by orange circles. The number of copies of likely functional genes of each host range gene family present in each virus is listed. Annotations: (*) mark E3L orthologs that do not encode for a functional Za domain, (#) an additional serpin gene in *YKV* that is probably non-functional. Abbreviations for poxvirus genera shown are: (*EPV*) *Entomopoxvirus B*; (*APV*) *Avipoxvirus*; (*CrP*) *Crocodylidpoxvirus*; (*MOPV*) *Molluscipoxvirus*; (*PPV*) *Parapoxvirus*; (*YPV*) *Yatapoxvirus*; (*LPV*) *Leporipoxvirus*; (*CePV*) *Cervidpoxvirus*; (*CaPV*) *Capripoxvirus*; (*SPV*) *Suipoxvirus*. For virus abbreviations see Haller *et al.*[1].

Myxomatosis pathogenesis progresses rapidly following infection via intradermal inoculation (transmission is by manual inoculation by blood-feeding arthropods, mainly mosquitos, which act as telmophagic vectors, feeding directly from the lesions). Within 2 days the virus can be detected in the adjacent draining lymph node, spread via leukocyte infection. 3-4 days post infection a high viral titre is seen in the draining lymph node, however the virus is not found in the serum. 3-5 days post infection, the virus spreads to the lungs, lymph nodes, spleen, testes and other organs. A raised red swelling occurs at the site of infection, followed by swollen eyelids. Secondary lesions follow at mucocutaneous junctions, such as the conjunctiva, the ears, and later all over the body as the disease progresses. The exact cause of death is unknown, as general organ failure occurs. Often secondary infection via pulmonary cutaneous lesions leads to $Gram^-$ infection[5].

The strain released in Australia in 1950 as a biological control agent had an estimated case fatality rate of 99.7% and caused the exposed rabbit populations to decrease by approximately 90%[17]. The near 100% fatality rate attenuated over time, as less pathogenic strains which allowed for longer host survival increased the chance of transmission to a new host[18].

Different strains of *MYXV* have been fully sequenced. Lausanne strain (Lau) was isolated in Lausanne, Switzerland, originating from the Brazilian strain released in Mallebois, France in 1952, which became endemic in Europe[16]. Investigations into the attenuation of *MYXV*, and genetic changes responsible for the reduction in viral pathogenicity in the European rabbit have found a surprisingly low level of genetic variability in isolates of attenuated field virus when comparing these, and attenuated viral strains 6918, and Mav, to Lau[19,20].

The live attenuated vaccine strain Mav was generated by serial passage of the MSD Californian strain in rabbit kidney cells[21]. Comparison of the fully sequenced Mav strain to the Lau strain as a refence, show it possesses, in contrast to the attenuated field strains, significant genetic differences to Lau. Phylogenetic analysis indicate that Mav arose as a result of genetic recombination of the Lau and MSD strains[22].

Mav features a reduced genome size of 147,574 bp vs 161,777 bp in the Lau strain includes the deletion of 11 ORFs, 1,455 a.a. deletions, and 1,307 a.a. substitutions[22]. Mav, in comparison to various field, challenge and Lau strain has shortened ITRs, 7.7 kbps in comparison to 11.6 kbps in Lau. The genomic

loss of the ITR regions includes the deletion of 11 ORFs, and truncation of 3 ORFs in Mav. ORFs *M007L – M008.1L* of the left ITR, and *M147R – M156R*, and *M008.1R* of the right ITR are deleted, with *M006L*, *M009L* and *M148R* truncated. All of these ORFs are known immunomodulatory proteins[22].

In addition to the ITR immunomodulatory gene differences above, mutations in the C7 proteins of Mav are present when compared to the Lau proteins. The identification of differences in the repertoire and sequences of immunomodulatory proteins of the attenuated Mav strain and the pathogenic Lau strain may give further insight into the host range factors essential for productive *MYXV* infection.

1.3. The *C7L* host range superfamily

1.3.1. Overview of C7 and its homologues

The *C7L* family is a group of poxviral genes found in most *Chordopoxviriniae*, with at least 1 copy present in all sequenced orthopox and clade II viruses[11]. The *C7L* gene and its orthologs encode host range factors which span from 18 - 31 kDa[8]. C7 and its homologues are essential for *in vivo* poxviral replication in a variety of mammalian cell cultures[23], and are determinants of pathogenesis in host infection[8]. C7 was initially identified by its ability to complement the function of host range factor K1 in restoring tropism of the WR strain of *VACV* in a variety of human cells[23]. Bearing no sequence similarity or homology to proteins outside the *Chordopoxvirus* genera[11], the origin of the C7 family is unknown[8].

Homologues of C7 have been shown to have distinct affects on host range, as the *C7L* gene in VACV *K1L*$^{-}$ does not permit for replication in rabbit kidney cells[23], whereas *MYXV* replication is permissive in RK13s, but restricted by the deletion of its C7 homologue M062[24]. Comparison of the C7 homologues detected in the *Chordopoxviriniae* reveals there can be high variability between homologues, sharing sequence identity as low as 19.5% to C7[11].

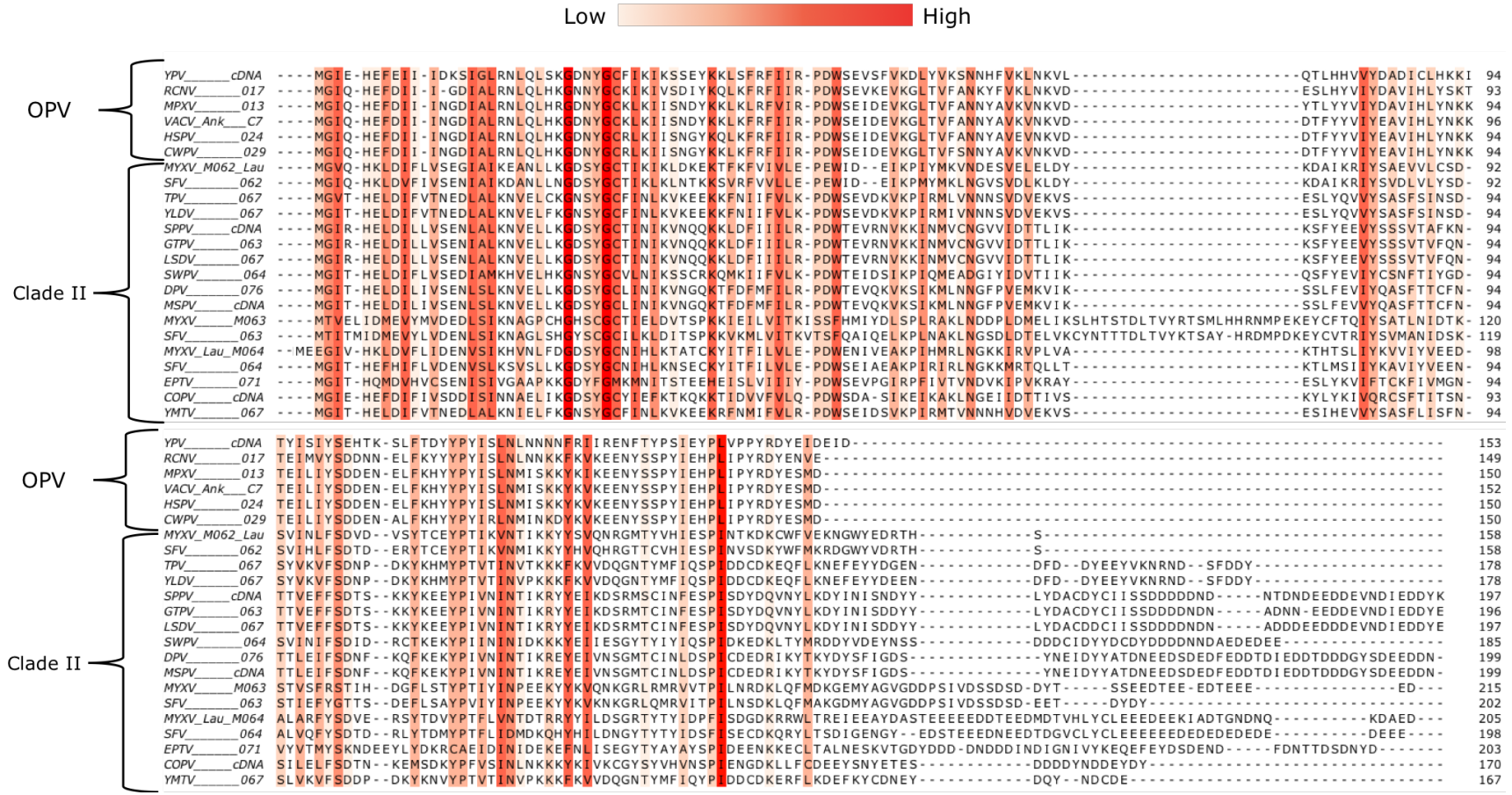


Figure 2: Multiple sequence alignment of members of the C7 superfamily. A multiple sequence alignment was carried out using Clustal Omega[25]. Amino acids are coloured by conservation in a gradient (above). Poxvirus abbreviations described in Table III. (OPV) orthopoxviruses. Sequence limits shown on right-hand side. Figure generated with Aline[26].

While the sequences vary significantly, the C7 homologues have a conserved motif of mainly hydrophobic residues (see Fig. 2). There exists a dichotomy in the C7 family between the OPV and clade II homologues, the latter possessing extended C terminal regions compared to those of the OPV. The C terminal region exhibit no distinct sequence similarity or homology, and contain a high proportion of acidic residues. It has been demonstrated that despite the low sequence identity, the majority of C7 homologues can complement the host restriction of C7, by inserting homologues into *VACV C7*$^{-}$ *K1*$^{-}$[8].

1.3.2. Antagonism of antiviral processes by the C7 family

1.3.2.1. Identification of the cellular target of the C7 family

A timeline summarising the major findings from studies on the effect of the C7 family on poxviral tropism is presented in Figure 3.

The first indication of a target of the C7 family was found when it was observed that pre-treatment of Huh7 or MCF-7 cells (which are normally permissive for both WT and C7^{-} K1^{-} *VACV*) with type I interferon (IFN) aborted replication of *VACV* C7^{-} K1^{-} but not WT *VACV*[27]. Interestingly treatment with IFN at the same time as infection with *VACV* C7^{-} K1^{-} did not prevent expression of interferon-stimulated genes (ISGs)[27], indicating the target of K1 and C7 is itself an ISG.

A genome-wide RNAi approach to identify host targets whose knockdown rescues *VACV* C7^{-} K1^{-} replication in non-permissive HeLa cells identified SAMD9 as a putative host restrictive gene.[28] CRISPR/Cas9 KO of SAMD9 in HeLa cells also recovered *VACV C7L*$^{-}$ K1L^{-} replication, while co-transfection of *VACV* C7^{-} K1^{-} with either IFN regulatory factor 1 (IRF1) or SAMD9 restored the restriction of viral tropism in *SAMD9*$^{-/-}$ HeLa cells[28]. This indicates that while SAMD9 was sufficient to restrict *VACV* tropism in a C7 (or K1)-dependent manner, an ISG was present whose induced expression can complement the activity of SAMD9.

The interaction of the C7 family with SAMD9 has been identified in a variety of studies. An *in vitro* interaction between C7 and K1 with SAMD9 was identified by the ability of C7 of K1 to facilitate coimmunoprecipitation of SAMD9 independently[28]. The *in vitro* interaction of the C7 family with SAMD9 was also shown by the identified of SAMD9 by mass spectrometry fragment analysis of proteins isolated by coimmunoprecipitation with M062 during infection of human cell lines with recombinant *MYXV* expressing V5-tagged M062[24]. Repeating this mass spectrometry analysis of proteins which coimmunoprecipitated with

SAMD9 by using an anti-SAMD9 antibody, M062 and M063 were identified. However, M063 was unable to interact with SAMD9 in the absence of M062[24], indicating the formation of a complex which includes both SAMD9 and M063 is M062-dependent.

The ability to interact with SAMD9 was later confirmed to be conserved in many members of the C7 family by co-IP of SAMD9 following infection of HeLa cells by C7 homologues inserted in the *VACV* C7$^-$ K1$^-$ background[29]. The key role of SAMD9 in the antiviral response targeted by the C7 family was further supported by the finding that constitutive knockdown of SAMD9 in U87 glioblastoma cells recovered the otherwise abortive replication of *MYXV* M062^{-}[24]. This study demonstrates the complementarity between M062 and SAMD9, as KD of SAMD9 alone was sufficient to rescue *MYXV* replication.

Evidence for the direct interaction between the C7 family and SAMD9 has been presented by biophysical analysis of recombinantly expressed M062 and SAMD9 using surface plasmon resonance[30]. The study found a low binding affinity between the N terminal region of SAMD9 (residues 1-385) with M062, with a K_D of approximately 10 μM. The study also carried out a series of co-IPs, screening different SAMD9 fragments, in order to localise the interaction domain of SAMD9, and found that the N terminal 1-385 region was sufficient to co-IP with V5-tagged M062. SAMD9 constructs 1-110, 78-285, 390-1170 and 1170-1589 failed to co-IP with V5-tagged M062.

Furthermore, Surface plasmon resonance of the SAM domain of SAMD9 comprising residues 10-78 showed no specific interaction with M062 compared to SAMD9 1-385. Failure of SAMD9 10-78, 1-110 and 78-285 indicates that the SAM domain of SAMD9 alone does not mediate interaction with the C7 family, and a domain encompassing the N terminal 385 residues is the potential target of C7 antagonism.

1.3.2.2. The molecular basis of the C7 family host range factors

Identification of SAMD9 as the cellular target for antagonism by the C7 family provided a marker where interaction of C7 homologues with SAMD9 could further elucidate the molecular basis of antagonism.

Sheeppox virus (*SPPV*) C7 homologue SPPV-063 was unable to rescue the restricted tropism of *VACV* C7$^-$ K1$^-$ in murine cells, while it did extend the host range in human cell lines[31]. By comparing and substituting regions of SPPV-063

with its most similar C7 homologue from swinepox (SWPV) SWPV-064 which confers host range in both human and murine cells to *VACV* C7$^-$ K1$^-$. 2 residues were found to be responsible for the host range restriction of SPPV-063: N134 F135, which when replaced with SWPV-064 equivalents enabled the resulting mutant SPPV-063 N134Y F135I to restore tropism in murine cells[31]. It was later found that SPPV-063 WT failed to bind the mouse paralog of SAMD9: SAMD9L, while the mutant SPPV-063 N134Y F135I interacted with SAMD9, as seen with the C7 homologues which confer host range in human and murine cells[32].

The crystal structure of C7 [PDB: 5CYW, 2.0 Å] and M064 [PDB: 5CZ3, 2.5 Å] were solved by Meng *et al*. in 2015[29], and revealed a novel structural fold, unique to the C7 family (see Fig. 4). Characterised as a β sandwich fold, the structure is composed of 2 sets of 6 antiparallel β sheets which form a curved interface, flanked by 2 short α helices. This fold is highly conserved between the 2 homologues with and RMSD of only 0.74 Å^2 over 136 aligned C_α's[29].

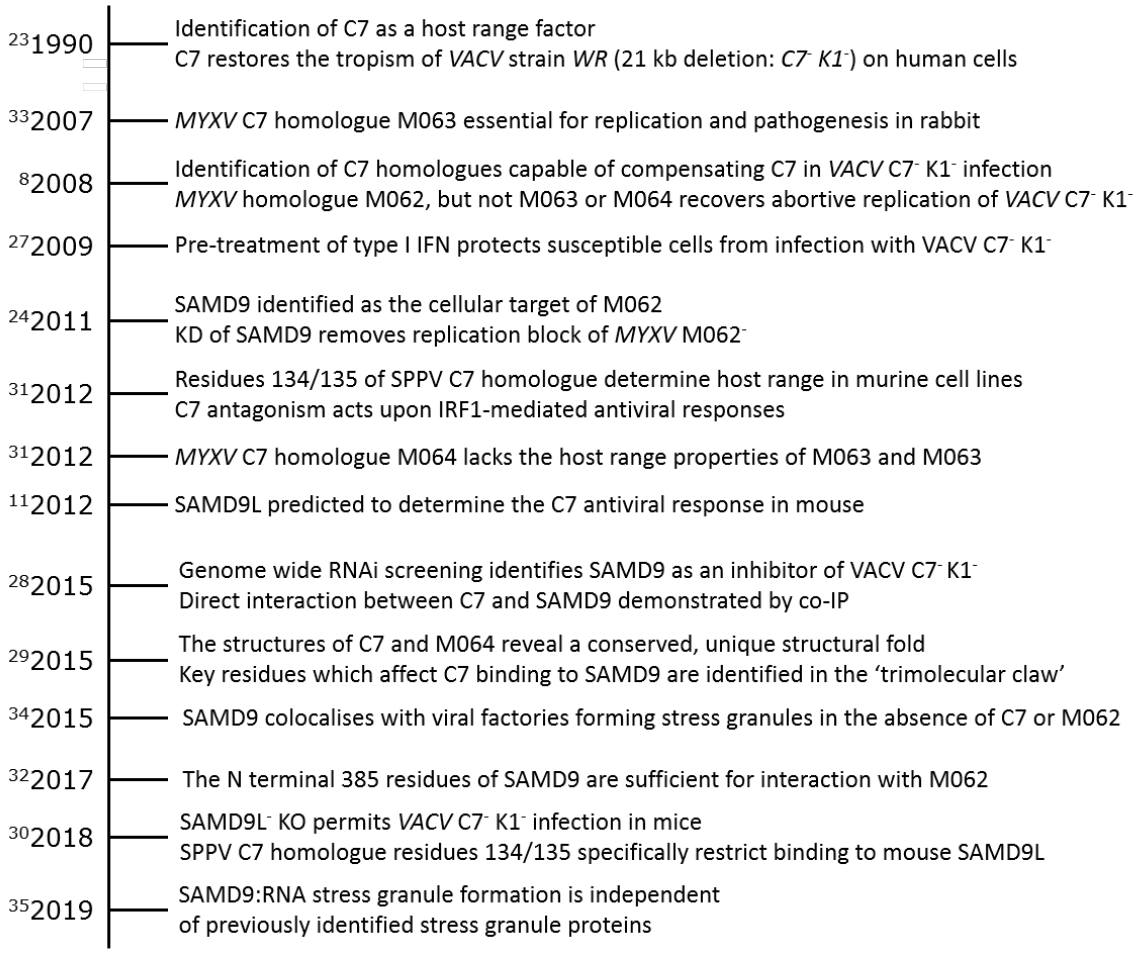


Figure 3: Timeline highlighting the major findings of studies of the C7 host range factor family.

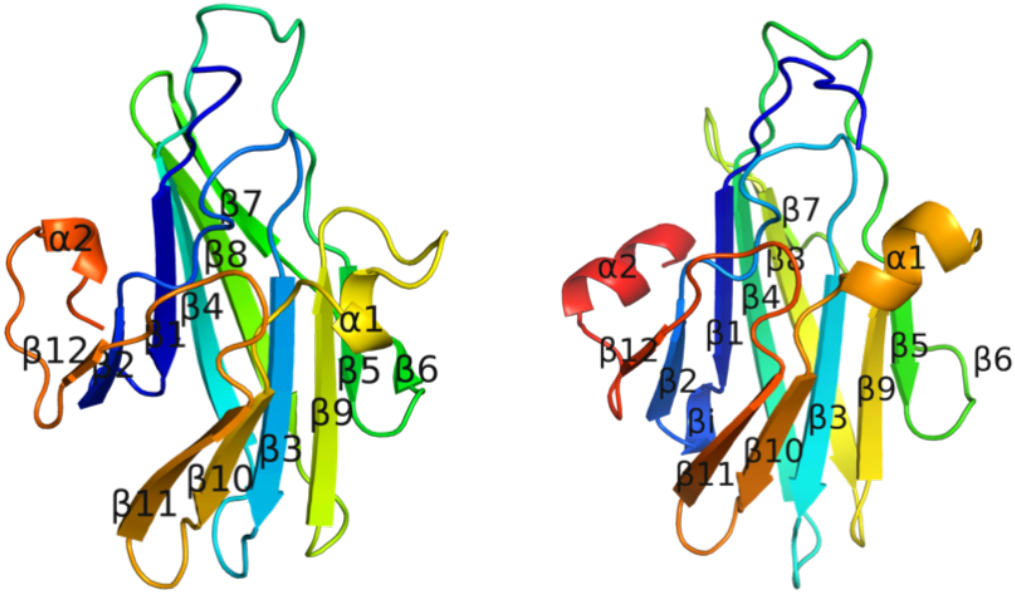


Figure 4: The structures of C7 proteins *VACV* C7 and *MYXV* M064. Ribbon models of the crystal structures are shown in rainbow colour (N terminus blue, C terminus red), with the secondary structures annotated.

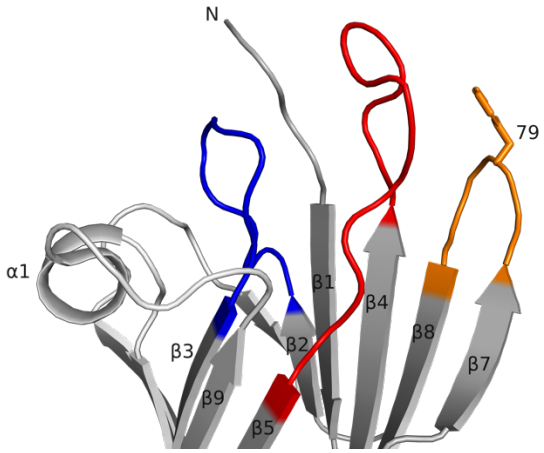


Figure 5: The loops of the Trimolecular claw of C7. Shown is a ribbon model of C7, highlighting the location of the 3 loops which form the trimolecular claw. The N terminus, position of residue 79, and secondary structural features which flank the loops are annotated. Loops coloured: basic loop $\beta_{2,3}$ (blue), acidic loop $\beta_{4,5}$ (red), and neutral loop $\beta_{7,8}$ (orange).

Structure-based systematic mutagenesis of surface residues of C7 identified 3 adjacent loops: $\beta_{2,3}$, $\beta_{4,5}$ and $\beta_{7,8}$, in which mutation of 9 key residues disrupted interaction with SAMD9. 5 of these residues: K24; D51, W52, E54; and F79 (C7 sequence) are thought to facilitate direct interaction with SAMD9[29]. This interaction motif was dubbed the 'trimolecular claw', and is located at the N terminal side of the β sandwich, as shown in Figure 5, (see Fig. 61 for sequence alignment position of the loops). M063 and M064, *MYXV* homologues of C7 that fail to complement C7 in restoring the host range of *VACV* $C7^-$ $K1^{-8}$ both feature mutations of these key residues, which may account for their loss of host restriction. It was also shown that deletion of the C terminal 10 residues of C7 appeared to not affect viral replication, however it did significantly reduce its interaction with SAMD9[29].

Interestingly, the SPPV-063 locus which has been shown to disrupt its binding to murine SAMD9L: residues 134, 135, are located quite distal to the trimolecular claw, at the beginning of β strand 12[32]. It was found that the interaction between SAMD9 and M062 was preserved with expression of the primary 385 residues of SAMD9, by co-IP[30]. This highlights a distinct mode of interaction between the C7 homologues and the other identified *VACV* $C7^-$ $K1^-$ host range rescuing poxviral proteins K1 and CP77[23]. These ankyrin repeat domain antagonists of SAMD9 have contrastingly been shown to bind a central region of SAMD9[33].

1.3.3. The effect of *MYXV* C7 homologues on viral tropism

MYXV is of particular interest among poxviruses, as it contains 3 orthologs of *C7L*. *MYXV C7L* orthologs *M062R, M063R and M064R* are arranged tandemly, resulting from 2 duplication events following the divergence of the *Leporipoxviruses*[12], similar copies are found in Shope fibroma virus (*SFV*)[34]. The encoded homologues are M062: 18.4 kDa, M063: 24.6 kDa and M064: 23.8 kDa. M063 contains a 24-residue insertion unique to the *Leporipoxviruses* (see Fig. 2). The homologues vary in their similarity to C7 (see Table 1), with M062 the closest in both sequence identity and similarity. The ability of *MYXV* homologues to complement *VACV* $C7^-$ $K1^-$ in mammalian cell culture correlates with their similarity to C7[8].

	C7	M062	M063	M064
VACV C7		27.4 (47.4)	14.4 (31.5)	20.4 (38.3)
MYXV M062			22.8 (42.8)	31.0 (44.8)
MYXV M063				19.3 (37.3)
MYXV M064				

Table 1: Comparison of *VACV* and *MYXV* C7 homologues. *VACV* strain Ankara vs *MYXV* strain Lau C7 homologue percentage sequence identity and similarity (brackets), as calculated by EMBOSS needle[35].

1.3.3.1. M062

M062 has the highest sequence similarity to C7 of the 3 *MYXV C7L* orthologs, and M062, but not M063 or M064 has been shown to restore *VACV* C7^{-} K1^{-} tropism in a range of human cell lines[8]. Knockout of M062 from *MYXV* completely abrogates its pathogenesis in the European rabbit, infection with the resulting virus is asymptomatic, demonstrating M062 is an essential host restriction factor for MYXV infection of *O. cuniculus*[24]. Interestingly, infection with the *MYXV* M062^{-} virus did not confer protection against subsequent challenge with WT *MYXV*. Replication *in vitro* in RK13s is also abortive with *MYXV* M062^{-}[24].

It has been proposed that M062 forms a complex with M063 in its antagonism of SAMD9, as the 3 proteins co-IP together with immunoprecipitation of either M062 or SAMD9 during *MYXV* WT infection[24]. It is unclear if M063 and M062 interact in the absence of SAMD9, as although they co-IP one another in murine cell lines[24], it is possible mouse SAMD9L facilitates their interaction. Whether the complex is a result of the preformation of a complex between M062 and M063, or if the interaction between SAMD9 and M062 facilitate the cooperative binding of M063 remains to be resolved.

The failure of C7 and Yaba-like disease virus (YLDV) C7 homologue M67 to compensate for *MYXV* M062^{-} host range in RK13s may imply a unique collaborative antagonism has evolved between *MYXV* C7 homologues M062 and M063. It has, however, been shown that M062, but not M063 or M064 can antagonise human SAMD9L in IFN- stimulated HeLa SAMD9$^{-/-}$ cells, and that M062 interacts with both human and mouse SAMD9L, while M063 and M064 cannot bind mouse SAMD9L[32].

The live attenuated *MYXV* strain variant Mav[19] contains a point mutation of I79T in M062 compared to the Lau strain. It has been shown that this single

residue mutant significantly reduces the interaction of M062 Mav with SAMD9 (Dr. Jia Liu, personal correspondence, see Fig. 16).

1.3.3.2. M063

M063 was shown to be an early expressed protein of *MYXV*, with expression continuing throughout infection of RK13s[36]. The absence of a signal sequence, glycosylation sites or a transmembrane region indicate M063 is a cellular protein, which interestingly shows up as a doublet when probed by Western blot following infection[36]. *MYXV* M063^{-} shows it acts as a classical host range factor, as it restricts replication in rabbit cells, but doesn't affect viral replication in certain human cancer or primate cell lines[36]. Furthermore, European rabbits infected with *MYXV* M063^{-} do develop myxomatosis, and are protected against subsequent infection with WT *VACV*, with the M063 KO virus acting instead as a protective, non-replicating vaccine[36].

Multiple sequence alignment of the C7 family (see Fig. 2) shows a 24 residue insertion in M063. These residues are located in the precise location of the neutral loop of the trimolecular claw, and the effect of the overall structure is unknown, but is likely to contribute to the unique function of M063 among C7 homologues. M063 is unable to interact with SAMD9 in the *VACV* C7^{-} K1^{-} background[29]. Comparison of the aligned residues of C7 and M063 also show semiconservative (†) and nonconservative (*) differences in 4 of the 5 key residues of the trimolecular claw[29], which may affect its host range function: K24H†, D51S*, W52F† and E54M*, F79 is preserved.

1.3.3.3. M064

M064, while more similar in sequence and homology to both C7 and M062 than M063 (see Table 1) was found by Liu *et al.*[37] to exhibit the least impact upon *MYXV* tropism, or pathogenesis of the *MYXV* C7 homologues. M064 is expressed as an early protein during *MYXV* infection. Infection of *MYXV* M064$^-$ in rabbits (*O. cuniculus*) showed a delay in the onset of symptoms, but pathogenesis progressed in a nearly identical fashion to infection with WT *VACV*. Little to no effect was seen in the *in vitro* infection of rabbit, human and mice cells, with a minor decrease in viral titres[37]. It was concluded that M064 does not function as a host range factor, and may influence the kinetics of *MYXV* infection, without being required for effective infection.

M064 has repeatedly been shown to fail to act as a host range factor in recovering *VACV* C7$^-$ K1$^-$ tropism[8], ability to interact with SAMD9[29], or human or mouse SAMD9L[32]. Nonconservative (*) differences are found in 3 of the 5 key residues of the trimolecular claw[29] in M064, which may affect its host range function: K24D*, E54N* and F79T*. The structure of M064 is surprisingly similar to C7, showing the C7 fold is highly conserved between the functional and non-functional homologue.

The M064 crystal structure is missing 53 residues from its C terminal tail, however, which are not represented in the crystal structure. The authors presumed these residues to be disordered, or to have degraded over the course of crystallisation[29]. The M064 structure, while showing the protein which crystallised has a structural fold consistent with the C7 crystal structure, may not capture differences between the structures adopted by C7 and M064 *in vivo*. The C terminal tail of M064 is composed of a significant number of acidic residues (see Fig. 2), and the authors did not comment on the potential impact of these residues on the structure of M064, or hypothesise about the role these residues may have in explaining the inability of M064 to complement C7 in *VACV* C7$^-$ K1$^-$ viral tropism.

1.3.4. Overview of the structure and function of sterile alpha motif domain containing 9

1.3.4.1. Expression and role in human health

The target of C7 antagonism was been identified by the interaction between M062 and SAMD9[24], and is supported by knockdown of SAMD9 recovering the

tropism of *VACV* C7^{-} K1^{-} in HeLa cells[28]. SAMD9 (sterile alpha motif domain containing 9) is a 1,589 residue protein located on the long arm of chromosome 7[38] expressed in the cytoplasm by most tissues, with the exception of the brain, and with lower expression in skeletal muscle[394041]. Its expression is upregulated by both IFN[42,43], TNF- α, and in response to hyperosmotic shock[44].

SAMD9 has been identified as a key regulator of a variety of cellular processes, including its role as a tumour suppressor, regulator of cell proliferation and apoptosis, endosomal fusion, congenital inflammatory disease, developmental disorders, and the antiviral response. While SAMD9 has been shown to have diverse functions, its mechanism of action remains largely unknown. The N terminal domain (1-124) has been shown to interact with the Ras effector protein RGL2 however, together inhibiting expression of the transcription factor early growth response gene-1[43], which plays a crucial role in myeloid differentiation[45].

A congenital disease called normophosphataemic tumoral calcinosis (NFTC) was identified in patients of Jewish-Yemenite ancestry resulting from a SNP in *SAMD9.* The disease caused by these mutations presents with the deposition of calcium in skin and mucosa, leading to the development of painful calcified tumours in the joints, and has an associated reduction in expression of SAMD9[4044]. These calcified tumours occur following injury, and together with NFTC patients frequently presenting with severe gingivitis indicate a role of SAMD9 in the inflammatory response[44].

Reduced expression of SAMD9 is seen in a variety of cancers[4139], with the finding that knockdown of SAMD9 increases the invasiveness, migration and proliferation of lung cancer *in vitro*, and conversely, its overexpression reduced these phenotypes[39]. Tumorigenesis was also enhanced with the injection of cancer cells with SAMD9 KD in nude mice[39]. Expression of miRNA-96 which has been shown to downregulate the expression of SAMD9 has been found to be increased in non-small cell lung cancer tissue in comparison to the surrounding tissue, and this decrease in SAMD9 expression significantly reduced the effective treatment of these cancers *in vitro*[46].

It has been found that loss of chromosome 7, or 7q is the most frequently represented chromosomal abnormality in a range of myeloid malignancies[47]. Loss of *SAMD9* and *SAMD9L* in monosomy 7, and following microdeletion of the long arm of chromosome 7 (7q21.3-) were found to be associated with myeloid

disorders such as acute myeloid leukaemia, myelodysplastic syndrome and juvenile myelomonocytic leukaemia[38]. Subsequently, it was seen that deletion, or even haploid deletion of the mouse paralog SAMD9L caused a series of myeloid malignancies and major mortality in SAMD9L$^{-/-}$ and SAMD9L$^{-/+}$ mice[47].

A highly fatal genetic disease caused by gain-of-function mutations in SAMD9, called MIRAGE syndrome is a multisystem disorder characterised by **m**yelodysplasia, **i**nfection, **r**estriction of growth, **a**drenal hypoplasia, **g**enital phenotypes, and **e**nteropathy[48,49,50]. The varied phenotypes which present in MIRAGE syndrome are thought to be a result of dysregulation of SAMD9's role in endosomal fusion[48]. The location of these mutants are not located near the N terminal region shown to be the target of C7 antagonism[30,48]. Investigation of the mutant phenotypes *in vitro* found SAMD9 plays a role in cell growth restriction, and the fusion of endosomes[48]. Similarly, gain-of-function mutants identified in SAMD9L cause disorders of myeloid proliferation, restrict cell growth, and immune and neurological disorders[51].

1.3.4.2. Evolution of SAMD9, and its paralog SAMD9L

SAMD9 is highly conserved among mammals, along with its paralog: SAMD9L, which is located tandemly in a head-to-tail configuration to SAMD9[52], including in opossum, indicating a gene duplication which occurred after the radiation of mammals, between 175 – 200 MYA[41]. However, while the presence of SAMD9 or its closely related paralog SAMD9L was initially thought to be restricted to mammals[52], the existence of SAMD9 orthologs have been found in a diverse range of species, including invertebrates and bacteria[53]. SAMD9 and SAMD9L appear to be highly conserved at both the protein and genetic level among diverse mammalian species, indicating they have been under positive selection over the course of their evolutionary divergence[52,53]. Interestingly, SAMD9 was lost in mouse, while SAMD9L was maintained, and it appears that mouse SAMD9L functions similarly to human SAMD9, indicating a level of redundancy between the paralogues[32].

1.3.4.3. Predicted domain architecture of SAMD9

SAMD9 was named following the prediction of the presence of an N terminal sterile alpha motif domain. This small, approximately 70 residue long domain is one of the most abundant motifs involved in protein-protein interactions in

Eukaryotes[54]. The structure of the SAM domain consist of 5 α helices, arranged in a bundle, with an unstructured N terminal region[55–57]. While highly conserved, SAM domains have diverse functions, including RNA binding[58], control of subcellular localisation through protein oligomerisation[59], and regulation of cell migration via protein-protein interactions[60]. The versatility of SAM domains in protein-protein polymer interactions is proposed to be enhanced as the N and C termini flanking the α helical bundle project outwards from the polymer interaction face, so SAM domains are not restricted to the N or C terminus of the protein[61]. The SAM domain of SAMD9 has been shown to polymerise *in vitro* by EM and native gel analysis[61].

Recently, a computational analysis of the domain architecture of the SAMD9 family was reported by Mekhedov *et al*.[53]. This report reveals that SAMD9 is a highly complex protein, comprising different motifs with a variety of functions (see Fig. 6). The modular organisation of the SAMD9 family is evidenced by the domain architecture of homologues varying from species to species. It appears that SAMD9, being present across different domains of life: bacteria, cnidaria, and vertebrates, has been preserved in a highly consistent state over the course of evolutionary divergence, indicating it has an important, and evolutionarily conserved role. The presence, and potential swapping of these 'modular domains' for others that provide similar function in distantly related homologues of SAMD9 may have allowed for adaptive variation in their specific function. There are indeed instances of death domains present in place, or in addition to the SAM domain in the SAMD9 family[53].

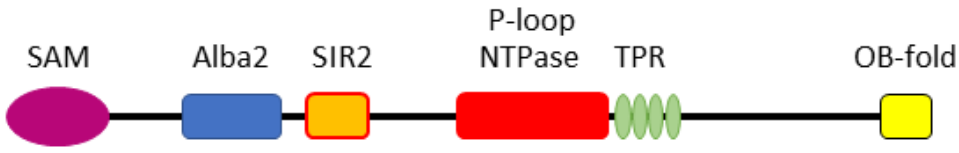


Figure 6: The predicted domain architecture of human SAMD9 adapted from Mekhedov et al.[53]. Multiple sequence searches with the SAMD9 family from a variety of species were conducted against the CDD, HHPred, Pfam and SMART databases. SAM: sterile alpha motif (magenta), Alba2 (blue), SIR2: sirtuin homologue (orange), P-loop NTPase (red), TPR: tetratricopeptide repeat (green), OB-fold: oligonucleotide/oligosaccharide-binding fold (yellow). Figure adapted from Mekhedov et al.[53].

SAMD9 possess multiple nucleic acid binding domains, with regions sharing homology with SAM, Alba2, SIR2, and an oligonucleotide/oligosaccharide-binding fold (see Fig. 6)[53]. While SAM domains have been demonstrated to bind RNA[58,62,63], the residues identified in Smaug[62] and Vts1p[58] which mediate the protein:nucleotide interaction are not conserved in the SAM domain of SAMD9, indicating it may function as a protein-protein interaction domain only.

Alba2 is a conserved archaeal chromatin (DNA-binding) protein, named for the regulation of the Alba family of proteins by acetylation (acetylation lowers binding affinity)[64]. Alba2 are basic, dimeric proteins, thought to bind DNA in a sequence independent fashion, but to also bind RNA, and can form ribonucleo-protein complexes[65].

Alba2 can in fact be regulated by Sir2[64], which itself is conserved from archaea to humans, and is a regulator of gene expression, chromatin remodelling[66], and is an NAD-dependent histone deacetylase (HDACs)[67]. The presence of these protein domains in tandem in SAMD9 indicates that SAMD9 is a nucleic acid binding protein, which could potentially regulate its activity through trans-, or auto- deacetylation via interaction of the Sir2 and Alba2 domains.

The SAMD9 P-loop NTPase motif is a member of the AAA+ superfamily of NTPases, that are characterised by the AAA+ motif, a nucleotide-binding, and catalytic module[68]. NTPases bind nucleoside triphosphates, catalysing the hydrolysis of the β-γ phosphate bond[69]. AAA+ NTPases (ATPases associated with diverse cellular activities) specifically commonly form oligomeric (frequently hexameric) ring structures, and while their catalytic function and structural motifs are conserved, the functions of AAA+ NTPases are diverse, and little can be predicted by their presence alone[68].

Tetratricopeptide repeats (TPRs) form protein-protein interaction motifs. Consisting of 34 amino acid degenerative repeats (although TPRs possess conserved spacing, size, and hydrophobicity), TPRs consist of a pair of antiparallel α helices, usually repeated in tandem, and with the subsequent TPR α helical pair interacting with the previous TRP, in an antiparallel manner. The packing of the tandem TPRs forms a right-handed helical extended structure, with an amphipathic channel, thought to potentially form a docking site, capable of accommodating an interacting protein's α helix[70,71]. TPRs in protein phosphatase 5 (PP5) are known to interact with arachidonic acid, inducing a conformational change which regulates PP5's catalytic activity[70].

The propensity of the individual SAMD9 motifs to form large oligomers via their individual protein-protein interactions indicates that SAMD9 itself may form a higher-order supramolecular complex upon activation, and that it may function as a large protein assemblage, similar to inflammasomes[72].

The complex domain architecture of SAMD9 correlates with its involvement in expansive cell pathways. The presence of multiple protein, DNA and RNA interaction modules, and the potential for its diverse regulation by post-translation modification, binding of nucleotides, nucleic acids or arachidonic acid (a key inflammation-dependent metabolite[73]) indicate that SAMD9's activity could integrate a host of cellular signals. The extent of the functional roles SAMD9 plays in cellular processes is still under investigation, but with such diverse and profound impacts seen in its genetic disease states, its regulation at an expression level, and its obligatory antagonism by the C7 family, SAMD9 already appears to be a highly influential protein of the innate immune system, with much still to be discovered.

1.4. M013: an immunomodulating, dual-effector pyrin domain-containing protein

1.4.1. Prediction of the pyrin-domain of M013

MYXV contains a wide repertoire of host immunomodulatory protein, amongst them is M013: a 126 amino acid protein which encodes an N terminal pyrin domain (PYD) with its first 80-90 residues[9,74,75]. The remaining ~30 residues form a positively charged 'tail', and are not conserved in PYDs, nor do they bear significant homology to known proteins, other than the closely related S013 pyrin-only protein (POP) from SFV[76]. M013 is a potent immunomodulatory protein, whose deletion attenuates *MYXV* infection in rabbits, leading to resolution of the disease[75]. M013 directly targets both ASC[75] (Apoptosis-associated Speck-like protein containing a CARD) and NFκB[77], acting as a dual-effector, by modulating these distinct, but overlapping pathways. The precise molecular mechanisms employed by M013 in its modulation of these distinct signalling pathways are not fully understood, and under investigation.

1.4.2. M013 interacts with ASC to inhibit antiviral immune responses

M013 has been shown to attenuate antiviral processes, modulating the immune response, leading to lower levels of inflammation at the site of infection, and decreasing the production of proinflammatory cytokines[75]. These activities appear to be dependent primarily on the M013 PYD, which has been shown to determine the subcellular localisation of M013, and furthermore facilitates the co-localisation of M013 with ASC-1[77]. This direct interaction was confirmed by co-IP[75], and to specifically occur between the PYDs of M013 and ASC[77]. M013 was shown to antagonise the activation and cleavage of procaspase-1 to caspase-1, and consequently prevents processing and release of pro-inflammatory cytokine IL-1β[77], both in a PYD-dependent mechanism[75].

1.4.2.1. ASC induces apoptosis via homotypic protein-protein interactions

ASC, also named PYCARD after its 2 domains: pyrin and CARD, is a 22 kDa protein identified by (and named after) its formation of speck-like aggregates in cells undergoing apoptosis[78]. The presence of an N terminal PYD and C terminal CARD positioned ASC as an adaptor protein capable of connecting signalling by PYD and CARD domain containing proteins. It was found that expression of either the PYD or CARD of ASC led to the formation of cellular aggregates, however the ASC-PYD aggregates were speck-like, while the ASC-CARD aggregates were more filamentous in morphology[79]. ASC is differentially expressed in a variety of cells with higher expression levels in epithelial cells and leukocytes, indicating a role in host-pathogen interactions[79].

ASC signalling was linked to apoptosis by finding its KD reduced chemotherapeutic induced apoptosis[78]. ASC apoptotic signalling was found to negatively regulated by the PYD of pyrin, presumably by interaction of the PYDs sequestering ASC, delaying the downstream activation of apoptosome machinery[80]. While caspase-1 interaction was isolated to ASC-CARD, expression of FL ASC, but neither ASC-PYD or ASC-CARD alone was found to activate caspase-1, leading to the proteolytic cleavage and activation of IL-1β[81]. AIM2, a cytosolic DNA sensor via its OB-fold, has been shown to initiate formation of a PYD-PYD dependent assembly with ASC following its activation, with DNA sensing leading to the formation of an AIM2:ASC large protein assembly, resulting in caspase-1 activation and induction of pyroptosis[82].

Activation of the p53-Bax mitochondrial apoptotic pathway was also shown to be dependent on ASC, with ASC found to translocate to the mitochondria in a PYD-dependent manner. Furthermore, Bax KO significantly reduced ASC-induced processing of caspases 2, 3 and 9[83].

Structural studies of the ASC indicate its function is dependent on the independent interactions of its two death domains (DD). Separation of the PYD and CARD by a flexible linker, the absence of any intramolecular interactions between the domains, and the distinct effects seen in expression of ASC-PYD or ASC-CARD implies ASC's functions are controlled by homotypic DD interactions alone[84].

1.4.3. Structure and function of the pyrin domain

1.4.3.1. Structure of the pyrin domain

The pyrin domain is a member of the death domain (DD) superfamily: an all α protein, it consists of a compact bundle of 6 α-helices, arranged in an anti-parallel fold with internal pseudosymmetry, centred around a hydrophobic core[9,74,85,86]. The domain was discovered nearly simultaneously following bioinformatic identification of proteins homologous to a conserved motif described initially in the eponymous pyrin protein. The NMR structure of the ASC-PYD reveals the PYD surface is composed of charged residues (see Fig. 7 (A)), which form localised polar faces, capable of mediating homo- ad heterotypic bonding via electrostatic interactions[87].

While members of the DD superfamily have low sequence homology, the overall structure of the members of the subfamilies are relatively conserved, in their formation of the 6 membered α-helical bundle, individual members of the subfamilies exhibit structural variation from one another, differing in helix length, interhelical angle, and the distribution of surface charged residues[88,89].

Functioning by enabling homotypic interaction with other PYDs, the activity of PYD-containing proteins is difficult to predict from homology modelling, or sequence comparison alone, without further structural characterisation. As the tertiary structures of individual DDs vary, and as the specific distribution of charged surface residues determines the interfaces of PYDs, the interactions of novel PYD-containing proteins must be investigated using structural and functional experiments in order to determine its molecular role.

1.4.3.2. Homotypic interaction interfaces of the DD superfamily

The compact α-helical bundle characterised by the DD produces 6 distinct faces thought to mediate homo- and heterotypic interactions which are grouped into 3 subtypes, Type I, II and III[88]. The interfaces of each subtype form complementary binding structures in opposed sides of the protein, which allows for the formation of large protein assemblies via repeated protein-protein interaction, as seen in the ASC-PYD (see Fig. 7 (C))[88]. Oligomerisation of the members of the DD superfamily is proposed as a method of signal amplification, and a mechanism of recruitment of downstream effector proteins through the assembly of large protein scaffolds, termed 'SCAF': signalling by cooperative assembly formation[88].

The 3 interaction interfaces are asymmetrical, with DD's forming described as bipolar proteins. The Type I interface results from the interaction of the complementarily charges interfaces formed by helices 1 and 4 (the Type Ia interface), and helices 2 and 3 (the Type Ib interface)[90]. The Type I interface is the dominant intermolecular interaction face of the DDs, mediating homotypic interactions within the helical assemblages identified as the speck-like oligomers formed ASC *in vitro*[91] (see Fig. 7 (C)).

The homotypic interaction seen in crystal structure complex between the CARDs of the Caspase-9 prodomain with the CARD of Apaf-1 shows the role of complementary surface electrostatic interfaces in DD binding (see Fig. 7 (B)). Furthermore, the negative regulation of apoptotic signalling of ASC by pyrin was predicted to be possible due to the complementary surface charged faces of their PYDs[80].

The importance of surface charged residues in mediating, and determining the specificity of Type I interactions has been shown by the different conservation of surface charge residues between the PYDs of ASC and AIM2, exhibiting distinct, but complementary sequence, and subsequent surface charge conservation which enables their interaction[92]. A role has also been implicated for a conserved hydrophobic patch within the Type Ib interface, thought to complement the complementary charge interactions[88,92,93]. Mutation of the key residues of the Type I interfaces can be seen to disrupt the antagonism of ASC by M013, however identifying the surface residues on the interaction faces currently relies on structural predictions and homology modelling[76], as there is of yet no published M013 structure.

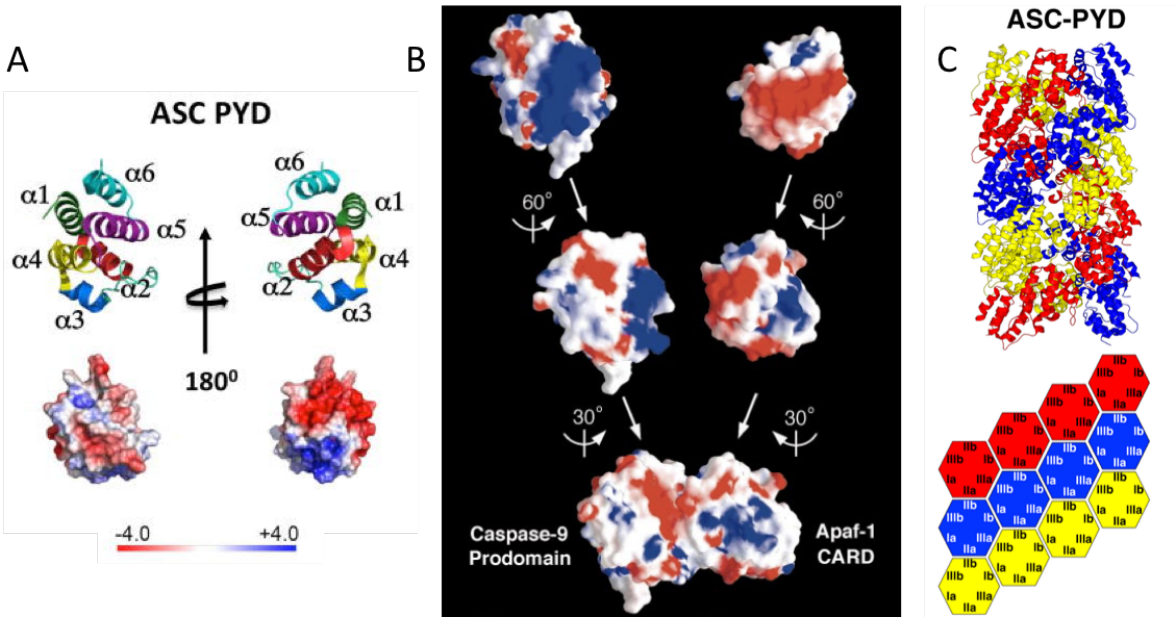


Figure 7: The DD superfamily display distinct surface charged faces, which mediate electrostatic-based interactions. (A) The structure of the ASC-PYD (1UCP)[87] is shown as a ribbon model, with helices coloured in green (α1), red (α2), blue (α3), yellow (α4), purple (α5) and turquoise (α6). The electrostatic surface is displayed on a scale from -4 kT/e (red) to 4kT/e (blue). Figure adapted from Dorfleutner *et al.* 2016[94]. (B) The electrostatic surface charge of the Caspase-9 prodomain:Apaf-1 CARD interaction faces which mediate complex formation, determined from the crystal structure (3YGS)[95] are shown, with positive-charge (blue), negative charge (red). Figure adapted from Qin *et al.* 1999[95]. (C) The ASC-PYD assemblage is shown as observed in the EM structure (3J63)[91]. Ribbon models of ASC-PYD are coloured by grouping into helical contact assemblies (red, blue or yellow). Below, a schematic diagram of the DD interface types. Figure adapted from Nanson *et al.* 2019[88].

1.4.4. M013 antagonises NFκB signalling independent of its inhibition of ASC activity

M013 has been shown to antagonise NFκB activated signalling by a mechanism independent of its antagonism of ASC and inflammasome formation[96]. M013 expression during *MYXV* infection prevents the phosphorylation and degradation of IKKs, preventing translocation of RelA/p65 to the nucleus, and the release of NFκB-controlled pro-inflammatory cytokines IL-1, IL-6 and TNF[97]. Activation of NFκB signalling by M013-deficient *MYXV* was shown to be independent of cytokine production via activation of the inflammasome. This indicates that M013 acts as a dual-effector, modulating cytokine production induced by *MYXV* infection by these distinct signalling pathways[97].

Consistent with the finding that truncation of the C terminal tail of M013 downstream of the PYD reduces the pathogenicity of the revertant *MYXV* M013^{-} virus in rabbits[75], a recent study found that deletion of residues 94-126 of M013 completely disrupted its interaction with NFκB[76]. While this finding indicates the C terminus determines M013's antagonism of NFκB, evidence of the PYD of ASC associating with IKKs, modulating their activation, and the production of NFκB-dependent cytokines[98], the mechanism of NFκB antagonism may include the PYD *and* the C terminal tail. Overall, the details of this interaction remain to be described at a structural level.

1.4.5. The structural basis of M013 dual effector functions

The current understanding of M013's mechanism of antagonism of ASC and NFκB is incomplete. It is proposed that while the N terminal PYD mediates interaction with ASC via Type I interface interactions, the N terminus facilitates the antagonism of NFκB[76]. It has also been demonstrated that revertant transfection of the PYD of M013 alone into *MYXV* M013^{-} results in an attenuated infection, reducing the mortality in rabbits by half compared to infection transfection of FL M013, furthermore, the C terminal portion of M013 without the PYD cannot establish a lethal infection[75]. This indicates that M013's antiviral function depends on both the PYD and the C terminal region, but the PYD is essential for its overall function.

1.5. Aims and objectives of this study

The objective of this study is to determine structural bases of *MYXV* antagonism of the human innate immune system by M062 and M013. This study aims to investigates the functional impact of a single residue change of the host range factor M062, a member of the C7 superfamily of poxviral host range factors. Determination of the crystal structures of a functional and non-functional strain variant of M062, and comparison of the structures attempts to describe how the structural differences between the M062 variants impacts the interaction of the C7 family with the target of C7 antagonism, SAMD9.

This study also aims to carry out structural characterisation of M013, a dual-effector *MYXV* pyrin-only protein that independently antagonises NFκB and ASC. The mechanism of M013's antagonism of NFκB and ASC is currently under investigation, and this study aims to provide structural information of M013, which may assist in understanding the interactions with it's the host target proteins. The objective of this study was to carry out primary, secondary, and tertiary structural studies of M013, using bioinformatic analysis, circular dichroism and nuclear magnetic resonance spectroscopy to generate a structural model of the pyrin domain of M013.

1.5.1. Structural studies of the Lausanne and Mav Myxoma virus strain variants of M062

A recent finding made by Liu *et al.* (unpublished data, see Fig. 16) shows that the single amino acid substitution of residue 79 of M062 from the reference, virulent Lausanne 'Lau' strain residue Ile to Thr, the residue present at position 79 in the vaccine 'Mav' strain M062 sequence is sufficient to significantly reduce the interaction between M062 and SAMD9. This finding supports mutagenesis studies by Meng *et al*.[29], which identified the C7 equivalent mutant F79S among a set of mutants which restricted the replication of revertant *VACV* $C7^-$ $K1^-$ virus, and prevented coimmunoprecipitation of SAMD9. Meng *et al.* propose the trimolecular claw, composed of three loops at the N terminal pole of the globular fold of C7 are responsible for mediating interaction with SAMD9.

The novel finding that the equivalent nonconservative substitution I79T in M062 facilitates the interaction between M062 and SAMD9 *in vitro*, and prevents viral replication, presents the possibility of elucidating the functional impact of this locus upon the C7 family's function by comparison of the Lau and Mav M062

structures. This study aims to express, purify, and crystallise recombinant Lau and Mav variants of M062 (see Chapter 3), to solve their crystal structures, and to compare the M062 structures to one another, and the published C7 [5CYW] and M064 [5CZ3] structures (see Chapter 4).

1.5.2. Structural studies of M013

This study aims to express and purify the pyrin domain of M013, in order to carry out structural analyses by circular dichroism and NMR spectroscopy. This study describes bioinformatic structural analysis and the expression, purification and preparation of native, and ^{15}N, ^{13}C double labelled M013 1-93 4Mut for CD and NMR spectroscopic analysis respectively. Generation of a structural model of the pyrin domain of M013 is described, to facilitate analysis of the surface charge residues and their potential role in mediating protein-protein interactions.

Chapter 2:
Materials and Methods

2.1. Materials

Table 2: Instruments and Equipment

Equipment	Manufacturer	Model (no.)
-20°C freezer	Hotpoint	RZA31
-80°C freezer	Thermo Fischer Scientific	Forma 900
16°C Incubator	Rumed	3000 Series
4°C cold cabinet	Angelantoni Industrie	EKOFRIGOLAB 1500
4°C cold cabinet	Thermo Scientific	NESLAB RTE 7
4°C cold cabinet	Angelantoni Industrie	EKOBASIC 700
Anion exchange column	GE Healthcare	sourceQ 15Q 4.6/100 PE
Anion exchange column	GE Healthcare	monoQ 5/50
Autoclave	Priorclave	Tactrol 2
Balance	Sartorius	BP110S
Balance	Ranger	RD6RM
Biological safety cabinet	ESCO	Labculture, Class II, Type A2
Camera	Nikon	Coolpix 4500
Cell culture flask	Fischer Scientific	Erlenmeyer, 3 L
Cell density meter	WPA Biowave	CO800
Centrifuge	Sorvall	RC6 Plus

Equipment	Manufacturer	Model (no.)
Centrifuge	Eppendorf	5415 D
Centrifuge	Eppendorf	5810 R
Centriugal rotor	Thermo Scientific	F21s-8x50Y
Centriugal rotor	Thermo Scientific	F10s-6x500Y
Dounce tissue grinder	Kontes Glass Co.	40 mL
Electrophoresis equipment	ATTO	AE-6450
Electrophoresis power supply	Consort	E132
Flow cell sonicator	Wyatt Technological Corporation	Comet
FPLC system	GE Healthcare	ÄKTA 100
Homogeniser	Avestin	Emulsiflex-C5
Incubation shaker	INFORS HT	Multitron
Incubator	Memmert	Incubator I
Liquid handling robot	TTP Labtech	mosquito crystal
Magnetic hotplate stirrer	DLab	MS7-H550-Pro
Milli-Q water purification system	Merck Millipore	Elix 10 Advantage, Millipak Express 40 filter
Multi-angel light scatter detector	Wyatt Technological Corporation	Tristar miniDAWN

Equipment	Manufacturer	Model (no.)
Nickel resin	Thermo Fischer Scientific	HisPur Ni-NTA Superflow Agarose, 25216
pH meter	Thermo Scientific	Orion 420A
Pipettes	Gilson	P1000, P200, P20, P2
Probe sonicator	Branson	Sonifier 250
Refractive index detector	Wyatt Technological Corporation	Optilab rEX
Rotator	Labinco	LD79
Scanner	Canon	Pixma MG2550S
Size exclusion column	GE Healthcare	Superdex 75 10/300
Size exclusion column	GE Healthcare	Superdex 75 16/60
Size exclusion column	GE Healthcare	Superdex 200 16/60
Spectrophotometer	Nanodrop	ND-1000
Spectropolarimeter	Jasco	J-810
Stereomicroscope	Nikon	SMZ800
Thermacycler	Eppendorf	Mastercycler Personal
Thermomixer	Eppendorf	Compact, 1.5 mL block
Uni-pucks	Molecular Dimensions	V1-Puck
Vortexer	VWR	Mini vortexer

Table 3: Consumables

Item	Manufacturer	Reference no.
2-Amino-2-(hydroxymethyl)-1,3-propanediol (Tris)	Sigma	T1503
2-mercaptoethanol (β-mercaptoethanol)	Sigma	M6250
2-Methyl-2,4-pentanediol (MPD)	Molecular Dimensions	MD2-250-24
24-well XRL Plate	Molecular Dimensions	MD3-11 XRL
2xYT	Formedium	YDB0102
4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES)	Sigma	H3375
96-well MRC 2-well crystallisation plate	Jena Biosciences	CPL-153L
Acetic acid	Fluka	27225
Acrylamide/N'N'-Methylenebisacrylamide 37.5:1, 40%	Acros	330225000
Amino acids	Sigma	-
Ammonium chloride	Sigma	A9434
Ammonium chloride (^{15}N, 99%)	Cambridge Isotope Laboratories, Inc.	NLM-467-PK
Ammonium persulfate	Sigma	A3678

Item	Manufacturer	Reference no.
Ammonium sulfate	Fluka	9978
Ampicillin sodium	Formedium	AMP25
Bis-tris propane	Sigma	B6755
BME vitamins 100x solution	Sigma	B6891
Bromophenol blue	Sigma	11A39-1
Calcium chloride	Sigma	21108
Cell culture dish 60 x 15 mm	Greiner	628160
Centifugal concentrator, Amicon Ultra-15, 10 kDa MWCO, 15 mL	Merck Millipore	UFC901024
Centifugal concentrator, Amicon Ultra-15, 10 kDa MWCO, 4 mL	Merck Millipore	UFC801024
Circular glass cover slips, siliconized, 22 mm	Jena Biosciences	CSL-106
cOmplete™, Mini, EDTA-free Protease Inhibitor Cocktail	Roche	4693159001
Coomaisse brilliant blue R-250	Fischer Scientific	BP101
Crystal clear sealing tape, 3"	Hampton Research	HR4-506
D-GLUCOSE (U-^{13}C6, 99%)	Cambridge Isotope Laboratories, Inc.	CLM-1396
Dialysis tubing, cellulose membrane (12 kDa MWCO)	Sigma	D9777

Item	Manufacturer	Reference no.
Dithiothreitol	FluroChem	M02712
DpnI	New England Biolabs	R0176
Ethylene glycol	Fluka	63750
Ethylenediamenetetraacetic acid	Fischer Scientific	BP120
Glutathione Sepharose 4B	GE Healthcare	GE17-0756-01
Glycerol	Acros	56-81-5
Glycine	Sigma	G8898
Hydrochloric acid	Fluka	7102
Imidazole	Sigma	56750
Isolate II plasmid mini kit	Bioline	BIO-52057
Isopropylthiogalactospyranose (IPTG)	VWR	437144N
Kanamycin monosulfate	Formedium	KAN0025
L-glutathione (reduced)	Sigma	G4251
L-selenomethionine	Acros	259960026
LB-Agar	Formedium	LMM0202
Lysogeny broth	Formedium	LMM0102
Membrane Filter, 0.22 µm	Merck Millipore	GSWP14250
Methanol	Sigma	67-56-1
Micro-prober 45°	Hampton Research	HR4-817
MRC 2-drop Crystallization Plate	Molecular Dimensions	MD11-00U-10/40/100

Item	Manufacturer	Reference no.
N-Cyclohexyl-2-aminoethanesulfonic acid (CHES)	Fluka	29311
N,N,N,N-Tetramethylethylenediamine (TEMED)	Sigma	T2.250-0
Nickel (II) chloride	Sigma	223387
Phenylmethanesulfonyl fluoride	Sigma	93482
Phusion DNA polymerase	New England Biolabs	M0530
Pierce™ Unstained Protein MW Marker	Thermo Fischer Scientific	26610
Polyethylene glycol 3,350	Fluka	88276
Polyethylene glycol 4,000	Fluka	95904
Polyethylene glycol 400	Fluka	91893
Polyethylene glycol monomethyl ether 5,000	Sigma	81323
Potassium chloride	Sigma	P3911
Potassium phosphate monobasic monohydrate	Sigma	P0662
Potassium thiocyanate	Fischer Scientific	333-20-0
Quick Start™ Bradford 1x Dye Reagent	BioRad	500-0205

Item	Manufacturer	Reference no.
Silicone grease, medium viscosity	Jena Biosciences	CGR-101
SnakeSkin™ Dialysis Tubing, 3.5K MWCO	Thermo Scientific	88242
Sodium dodecyl sulfate	Fischer Scientific	S/5200/53
Sodium hydroxide	Sigma	58045
Sodium phosphate dibasic	Sigma	S0876
Syringe filter, Filtropur S, 0.2 µm	Starstedt	83.1826.001
Thrombin protease	GE Healthcare	27-0846-01
Virkon	VWR	330003
X-ray capillary tube, 0.12 x 3 mm	MiTeGen	MicroRTTM Tubing Kit

2.2. **Methods**

2.2.1. Preparation and composition of buffers

All buffers were prepared at RT using ddH_2O (Elix 10 Advantage, Millipak Express 40 filter), their pH adjusted to within 0.01 pH unit (Orion 420A) by titration with conc. HCl (37%, Fluka) or NaOH (5 M, Sigma), and stored at 4°C for up to 3 days prior to use. Buffers for chromatography use were filtered (0.22 µm, Millipore) and degassed under vacuum prior to use.

Cell lysis and protein isolation buffers

Extraction Buffer A

500 mM NaCl, 20 mM tris, 20 mM imidazole, 5 mM β-ME[†], pH 8.0 (at 4°C)

[†] For the purification of M013, β-ME was not used.

Dialysis Buffer A

150 mM NaCl, 50 mM Tris, 1 mM DTT, 10 mM EDTA, 20% (v/v) glycerol, pH 8.0 (at 4°C)

Extraction Buffer B

300 mM NaCl, 20 mM tris, 20 mM imidazole, 5 mM β-ME[†], pH 8.0 (at 4°C)

Extraction Buffer C

1 M NaCl, 50 mM Tris, 1 mM DTT, 1 mM EDTA, pH 8.0 (at 4°C)

Elution Buffer A

300 mM NaCl, 20 mM tris[†], 200 mM imidazole, 5 mM β-ME[†], pH 8.0 (at 4°C)

[†] For the purification of M013, 10 mM Tris was used and β-ME was omitted

Elution Buffer B

150 mM NaCl, 50 mM tris, 10 mM GSH, pH 8.0 (at 4°C)

Phosphate-buffered saline [1 L]

8 g NaCl, 0.2 g KCl, 1.44 g Na_2HPO_4, 0.24 g KH_2PO_4, pH 7.5

FPLC buffers

IEX Buffer A1 [M062-Lau 3CS]

25 mM NaCl, 20 mM tris, 1 mM DTT, pH 7.5 (at RT)

IEX Buffer B1 [M062-Lau 3CS]

800 mM NaCl, 20 mM tris, 1 mM DTT, pH 7.5 (at RT)

IEX Buffer A2 [M062-Mav 3CS]

10 mM NaCl, 10 mM tris, 1 mM DTT, pH 7.8 (at RT)

IEX Buffer B2 [M062-Mav 3CS]

800 mM NaCl, 10 mM tris, 1 mM DTT, pH 7.8 (at RT)

IEX Buffer A3 [M013 1-93 4Mut]

120 mM NaCl, 10 mM tris, pH 7.9 (at 4°C)

IEX Buffer B3 [M013 1-93 4Mut]

800 mM NaCl, 10 mM tris, pH 7.9 (at 4°C)

Dialysis Buffer B [M013 1-93 4Mut]

90 mM NaCl, 10 mM Tris, pH 7.9 (at 4°C)

SEC Buffer

150 mM NaCl, 20 mM Tris, 1 mM DTT, pH 7.5 (at RT)

[†] For the purification of M013, 10 mM Tris was used and DTT was omitted

SDS-PAGE Solutions

Coomaisse Brilliant Blue Stain

0.4% (w/v) Coomaisse Brilliant Blue R-250, 40% (v/v) methanol, 10% (v/v) acetic acid

Destaining solution

40% (v/v) ethanol, 10% (v/v) acetic acid

4xSDS-PAGE loading buffer (4xLB)

200 mM Tris-Cl (pH 6.8), 400 mM DTT, 8% (w/v) SDS, 0.4% (w/v) bromophenol blue, 40% (v/v) glycerol

2.2.1.1. Preparation of media

Liquid media utilised for the growth of *E. coli*, was prepared by dissolving powdered stock of 2xYT (Formedium) or LB (Formedium) into ddH_2O to a working concentration as per the manufacturer's instructions. The media was then autoclaved (Priorclave, Tactrol 2) and stored at RT for up to 1 week before use.

LB-Agar solid phase medium (Formedium) was used for the selective growth of *E. coli* transformed with antibiotic resistant gene containing plasmids, which were used as the recombinant protein expression vector. LB-Agar was prepared in the same fashion as the liquid media above. Following autoclaving, in a biological safety cabinet (ESCO, Labculture, Class II, Type A2), the molten medium was allowed to cool to approximately ~50°C, after which was the antibiotic stock added, in order to minimise thermal degradation of the antibiotic. Following addition of the antibiotic, the solution was mixed, poured into cell culture dishes (Greiner), and allowed to cool. It was then used immediately or stored at 4°C for up to 2 weeks before use.

2.2.1.2. Preparation of antibiotic stocks

The vectors used in the transformation of *E. coli* confer resistance to antibiotics, via the inclusion of an antibiotic resistance gene (see Fig. 8, 9). These resistance factors allow for the selection of bacteria which contain the vector of interest, by culturing the cells with media containing the antibiotic. pET15b and pace3-MBP plasmids confer ampicillin and kanamycin resistance respectively. Ampicillin and kanamycin have working concentrations of 50 µg/mL and 300 µg/mL respectively. Antibiotic stocks were prepared at 1000x (50 mg/mL and 300 mg/mL for ampicillin and kanamycin respectively), by solubilising the powdered stocks in ddH_2O. Stocks were frozen at -20°C for up to 1 month.

2.2.2. Preparation of recombinant expression vectors

The genes encoding the proteins of interest in this project have been cloned into the well-established plasmids pET15b (Genscript) and pace3-MBP, see Figure 8 and 10 respectively. These plasmids have been engineered for use in bacterial recombinant protein expression and utilize a resistance gene to allow for antibiotic selection, a T7 polymerase promoter and terminator sequence, and a lac operator. The expression of protein is repressed by lacI, until the addition of IPTG, which induces expression by binding lacI, and preventing its binding to the lac operon, allowing initiation of transcription by the binding of the T7 polymerase[99].

2.2.2.1. M062 expression vector

Expression vector pET15b containing full-length wild-type, or 3CS constructs of M062-Mav inserted between the *NdeI* and *BamHI* sites were purchased from Genscript (see Fig. 8). The Lau variants of WT and 3CS M062 were generated by site-directed mutagenesis, mutagenizing Thr79 to Ile.

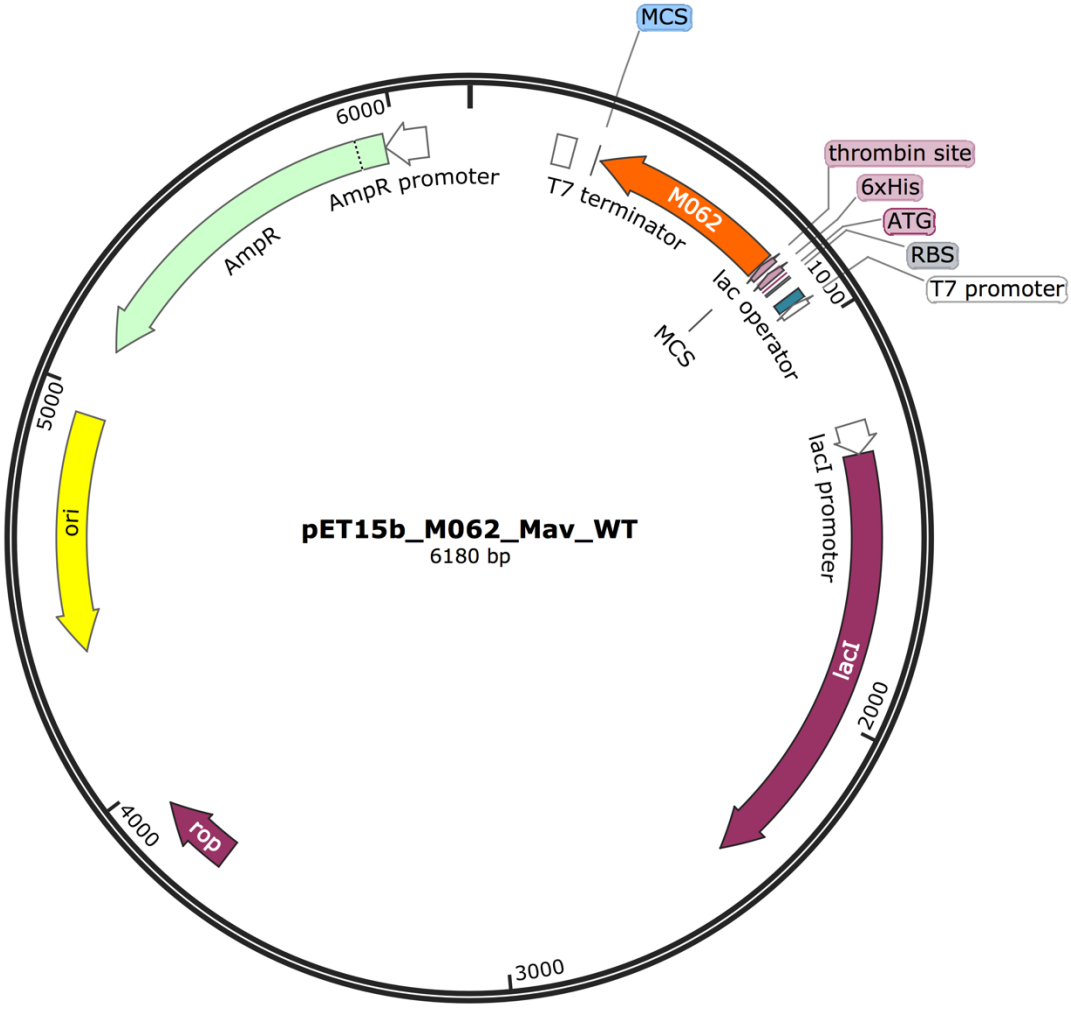


Figure 8: Plasmid map of pET15b encoding His-M062. Features are shown as coloured arrows. Yellow: ori, origin of replication; green: AmpR, ampicillin resistance gene (ß-lacamase); orange: M062; red: rop gene; lacI gene, white: promoters. Annotations highlight features of the multi-cloning site (MCS), RBS: ribosome-binding site. Figure generated with SnapGene(GSL Biotech).

2.2.2.2. Sequence and biochemical properties of His-M062-Lau 3CS

MGSSHHHHHHSSGLVPR/GSHMGVQHKLDIFLVSEGIAIKEANLLKGDSYGCTIKIKLD
KEKTFKFVIVLEPEWIDEIKPIYMKVNDESVELELDYKDAIKRIYSAEVVLSSDSVINLFS
DVDVSYTSEYPTIKVNTIKKYYSVQNRGMTYVHIESPINTKDKSWFVEKNGWYEDRTH
S-

Figure 9: The sequence of His-M062-Lau 3CS is shown, with the position of the introduced Cys → Ser mutations, and the I79 locus of the Lau-Mav mutant. Hexahistidine tag (grey), N terminus following proteolytic cleavage (underlined), thrombin cleavage site (red), I79, Mav-Lau variant locus (magenta), S90, S107 and S143, Cys → Ser mutations (gold).

Amino acid	Number of residues (cut)	Composition % (cut)
Ala	4 (4)	2.2 (2.5)
Arg	4 (3)	2.2 (1.9)
Asn	7 (7)	3.9 (4.3)
Asp	12 (12)	6.7 (7.5)
Cys	1 (1)	0.6 (0.6)
Gln	2 (2)	1.1 (1.2)
Glu	14 (14)	7.9 (8.7)
Gly	9 (7)	5.1 (4.3)
His	10 (4)	5.6 (2.5)
Ile	16 (16)	9.0 (9.9)
Leu	11 (10)	6.2 (6.2)
Lys	18 (18)	10.1 (11.2)
Met	4 (3)	2.2 (1.9)
Phe	5 (5)	2.8 (3.1)
Pro	5 (4)	2.8 (2.5)
Ser	19 (15)	10.7 (9.3)
Thr	8 (8)	4.5 (5.0)
Trp	3 (3)	1.7 (1.9)
Tyr	10 (10)	5.6 (6.2)
Val	16 (15)	9.0 (9.3)
Total number of residues	178 (161)	
Molecular weight (kDa)	20.507 (18.625)	
Extinction co-efficient of a 1% solution ($^{1\%}\varepsilon_{280}$)	15.31 (16.86)	
Theoretical pI	6.19 (5.42)	

Table 4: Biochemical properties of His-M062-Lau 3CS. Mw, 1% extinction co-efficient and theoretical pI calculated by protparam[101].

2.2.2.3. Site-directed mutagenesis of M062

Site-directed mutagenesis of both WT and 3CS M062-Mav constructs was performed to introduce a single residue change at position 79 (T79I, see Fig. 9), to generate M062-Lau variants. The protocol was adapted from that of the QuikChange II Site-Directed Mutagenesis kit[102] (see Table 5).

PCR primers for site-directed mutagenesis were ordered from Eurofins Genomics. Phusion polymerase and reaction reagents were purchased from NEB. A final reaction volume of 50 µL was used, Mastercycler personal (Eppendorf).

I

Reagent	**Final concentration**
Forward primer	~125 ng
Reverse primer	~125 ng
Plasmid template	5 – 50 ng
DMSO	4% v/v
dNTP	2 mM
Phusion polymerase	1 U

II

Step	**Process**	**Time (sec)**	**Temperature**
1	Initial denaturation	60	95°C
2	Denaturation	45	95°C
3	Annealing	90	55°C
4	Elongation	240	72°C
5	Repeat step 2-5 x 16	-	-
6	Hold	∞	4°C

Table 5: Mutagenesis PCR protocol

The resulting PCR products were treated with *DpnI* at 37°C for 1 hr, prior to transformation into DH5α cells for downstream processes. Mutations were confirmed by Sanger-sequencing, carried out by Eurofins Genomics.

2.2.2.4. M013 expression vector

The M013 construct used in this project had previously been generated and cloned using LIC into the pace3-MBP plasmid[103] by Dr. Patrick Lall. The construct, downstream of a T7 promoter, comprises an N-terminal hexahistidine tag, followed by a linker containing a 3C protease site, MBP, followed by residues 1-93 of M013, containing mutations at 4 loci: M21A, Y24G, Y31S, F68S, C121S (see Fig. 10).

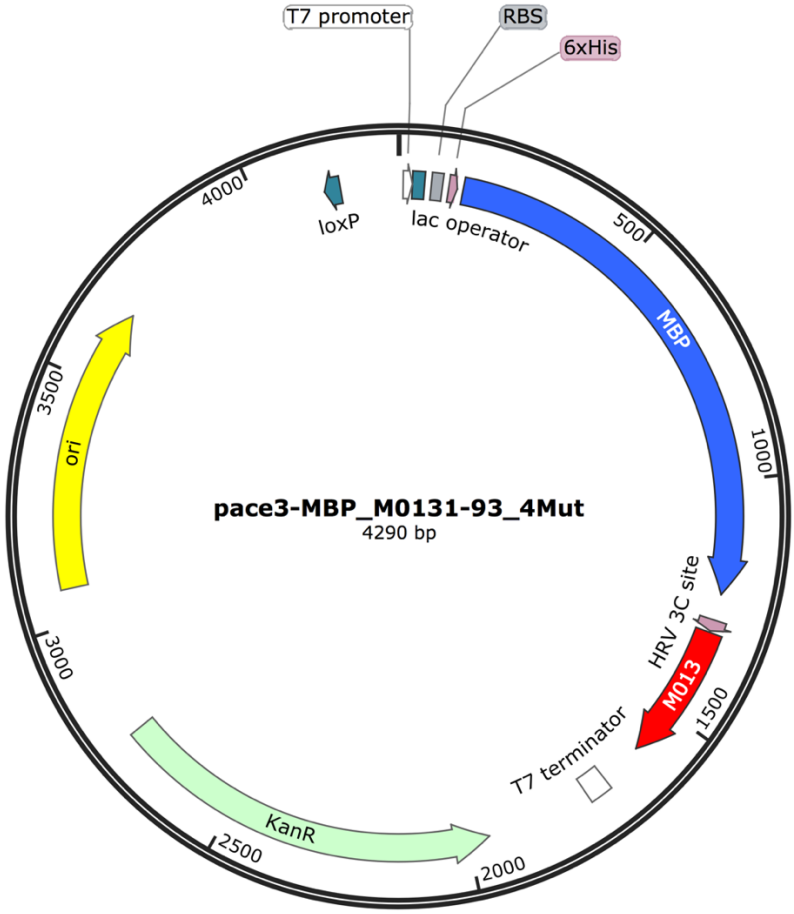


Figure 10: Plasmid map of pace3-MBP_M013_1-94_4Mut. Features are shown as coloured arrows. Yellow: ori, origin of replication; light green: KanR, kanamycin resistance gene; blue: maltose-binding protein (MBP) gene; pink: human rhinovirus (HRV) 3C protease cleavage site; red: M013 1-93 4Mut. RBS: ribosome-binding site. Figure generated with SnapGene(GSL Biotech).

2.2.2.5. Sequence and biochemical properties of His-MBP-M013 1-93 4Mut

MKHHHHHHSASPMKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEE
KFPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLI
AYPIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADG
GYAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETA
MTINGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLEN
YLLTDEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWY
AVRTAVINAASGRQTVDEALKDAQTNSSSNNNNNNNNNNPMSKLEVLFQ/GPAMMEH
RGVIITVLENLSDYQFKAFIGLAMEDLSIERAEKEKIDRIDLAHKISEQYLGTDYIEFMKR
VTDSIPNKVYVDSLLARAEADAEATMEAV-

Figure 11: The sequence of His-BP-M013 1-93 4Mut is shown, with the position of the 4 introduced mutations annotated. Hexahistidine tag (grey), MBP (blue), N terminus following proteolytic cleavage (underlined), 3C PreScission protease cleavage site (red), I79, Mav-Lau variant locus (magenta), M21A, Y24G, Y31S, F68S, C121S, Cys → Ser mutations (gold).

Amino acid	Number of residues (cut)	Composition % (cut)
Ala	56 (11)	11.2 (11.3)
Arg	10 (5)	2.0 (5.2)
Asn	34 (2)	6.8 (2.1)
Asp	32 (8)	6.4 (8.2)
Cys	0 (0)	0.0 (0.0)
Gln	12 (2)	2.4 (2.1)
Glu	39 (11)	7.8 (11.3)
Gly	33 (4)	6.6 (4.1)
His	11 (2)	2.2 (2.1)
Ile	31 (9)	6.2 (9.3)
Leu	40 (8)	8.0 (8.2)
Lys	44 (6)	8.8 (6.2)
Met	14 (5)	2.8 (5.2)
Phe	19 (3)	3.8 (3.1)
Pro	25 (2)	5.0 (2.1)
Ser	23 (5)	4.6 (5.2)
Thr	23 (4)	4.6 (4.1)
Trp	8 (0)	1.6 (0.0)
Tyr	19 (4)	3.8 (4.1)
Val	27 (6)	5.4 (6.2)
Total number of residues	500 (97)	
Molecular weight (kDa)	55.417 (10.981)	
Extinction co-efficient of a 1% solution ($^{1\%}\varepsilon_{280}$)	13.05 (5.43)	
Theoretical pI	5.22 (4.63) His-MBP: 5.58	

Table 6: Biochemical properties of His-MBP-M013 1-93 4Mut. Mw, 1% extinction co-efficient and theoretical pI calculated by protparam[101].

2.2.2.6. Preparation of chemically competent *E. coli* DH5α and BL21 DE3 cells

An overnight culture of DH5α or BL21 DE3 was prepared by the addition of a glycerol stock scraping to 2 mL LB (Formedium), incubated at 37°C, 180 rpm (Infors, Multitron HT) for 16 – 20 hrs. 50 mL LB was inoculated with 1 mL overnight culture and incubated at 37°C, 180 rpm for 3 hr. Rapid growth, indicating cells were in the log phase was confirmed by taking OD_{600} measurements at 20 min intervals (WPA Biowave, CO800). Once cell density was roughly doubling in 20 min, and at an optical density at 600 nm (OD_{600}) of 0.3 – 0.6, the culture was cooled on ice for 10 min, then cells pelleted by centrifugation at 3000 rcf, 4°C, 15 min (Eppendorf 5810 R).

The supernatant was decanted, and the pellet resuspended by gentle pipetting up and down using a cut-off tip, in 25 mL $CaCl_2$ (Sigma), 100 mM, 4°C. The cells were incubated on ice for 30 min, then cells were pelleted by centrifugation at 3000 rcf, 4°C, 15 min (Eppendorf 5810 R). The supernatant was decanted, and the pellet resuspended by gentle pipetting up and down using a cut-off tip, in 4 mL $CaCl_2$ 100 mM, 25% glycerol (Acros). The cells were incubated on ice for 15 min, then aliquoted in 100 µL into 1.5 mL reaction tubes, on ice. Cells were frozen at -80°C.

2.2.2.7. Transformation of chemically competent *E. coli*

LB-agar plates were pre-incubated at 37°C, to warm them prior to plating of transformed cells. In a 1.5 mL reaction tube, 35 µL of competent DH5α, or BL21 DE3 cells were thawed on ice from storage at -80°C. To this, 25-100 ng of plasmid was added, and the tube gently agitated to mix the contents. A negative control was set-up in which ddH_2O was added instead of the plasmid.

The cells were allowed to rest on ice for 15 min, followed by placing at 42°C for 1 min heat-shock (Thermomixer compact, Eppendorf). The cells were immediately returned to ice, for 3 min. To this, 1 mL of sterile LB media was added, and the cells recovered by incubation at 37°C for 45 minutes, shaking at 700 rpm (Infors, Multitron HT). The cells were briefly centrifuged (<30 sec) at 16,100 rcf, RT (Eppendorf, 5415 D), to pellet the cells. 800 µL of the supernatant LB was removed, and the pellet resuspended in the remaining volume by gentle pipetting up and down using a cut-off tip, to avoid subjecting the cells to excessive shear forces.

Using sterile technique, 100 µL of the cell suspension was pipetted onto the LB-agar, and spread upon the surface using a glass spreader. The plates were incubated at 37°C for 16 – 20 hr without shaking (Memmert) to allow the cells to recover, and grow on the antibiotic-containing medium. Successful transformations yielded single colonies in the cells transformed with plasmid, with the control cells yielding no colonies. A single colony was then chosen for plasmid amplification.

2.2.2.8. Plasmid amplification and isolation

Using sterile technique, 5 mL LB was transferred to a 20 mL sample tube, and supplemented with 5 µL 1000x stock ampicillin (Sigma). The LB was then inoculated, by the addition of a single colony of successfully-transformed plasmid-containing cells. The cells were incubated at 37°C, 180 rpm for 16 – 20 hrs. The cells were pelleted by centrifugation at 3,000 rcf for 15 min at 4°C (Eppendorf, 5810 R). The supernatant was decanted, and the tube left inverted for 5 min to allow remaining supernatant to drain. The plasmid was then isolated using the Isolate II plasmid mini kit (Bioline).

2.2.3. Solubility expression test of M062

The crystallisation of M062 required large-scale expression of recombinant protein in BL21 DE3 cells, followed by downstream purification in order to yield a sufficiently pure solution of homogenous protein which could be subjected to crystallisation screening. Protein expression levels and solubility are affected by a multitude of factors, not limited to, but including the temperature at which expression is carried out. To assess the effect of the temperature of expression on M062 solubility, small-scale expression tests were initially carried out.

2.2.3.1. Small-scale expression

The solubility of the recombinantly expressed M062 was assessed at small scale (15 mL cell culture), prior to scale up to large-scale expression (1 L) required to produce enough protein for crystallisation screening. This allowed for screening of the construct's solubility at different temperatures. Expression at 37°C and 18°C were compared.

2.2.3.2. Preparation of overnight inoculation culture

To achieve this, BL21 DE3 cells were transformed with pET15b containing M062-Mav WT. in a 5 mL tube, 2 mL LB with 2 µL ampicillin (1000x) was inoculated with a single-colony of BL21 DE3 cells containing a plasmid with the gene of interest. The cells were incubated in a rotary incubator (Infors, Multitron HT) at 37°C, 180 rpm for 16 – 20 hrs. This produced an 'overnight' culture with high cell density, which can be used to inoculate a larger volume.

2.2.3.3. Inoculation and cell growth

To a new 20 mL tube, 15 mL LB with 15 µL ampicillin (1000x) was added. The LB was then inoculated by addition of 500 µL of the 'overnight' culture. The culture was incubated at 37°C, 180 rpm (Infors, Multitron HT) for 1 ½ hr. The optical density at 600 nm (OD_{600}) of the cell culture was monitored from this timepoint (WPA Biowave, CO800). At an OD_{600} 0.6 – 0.8, the culture was divided into two tubes, at a final volume of 5 mL each. A glycerol stock was prepared at this stage, by mixing 400 µL cells with 200 µL 100% glycerol (Acros), snap frozen in LN2 and stored at -80°C for future use.

2.2.3.4. Induction of protein expression and cell lysis

One culture was moved to incubate at 18°C, 180 rpm (Infors, Multitron HT) for 15 min. Recombinant protein expression was induced in both cultures by the addition of 2.5 µL of a 1 M stock IPTG (VWR), 0.5 mM final concentration. Expression was carried out for 3 hr at 37°C, 180rpm, and for 16-20 hr at 18°C, 180 rpm.

Cells were pelleted by centrifugation at 3,000 rcf for 15 min at 4°C (Eppendorf, 5810 R). The supernatant was decanted, and the tube left inverted for 5 min to allow remaining supernatant to drain.

The pellet was fully resuspended in 1 mL Extraction Buffer B (see Section 2.2.1), 4°C by repeated pipetting, and transferred to a cooled 1.5 mL reaction tube. Cells were lysed by sonication using a probe sonicator (Branson, Sonifier 250), 3 cycles at output 3, duty cycle 30, for 30 sec, resting in an ice-bath for 1 min between each cycle. The cell lysate was then fractionated into soluble and insoluble components by centrifugation at 16,100 rcf for 15 min, 4°C (Eppendorf 5415 D).

2.2.3.5. Protein isolation by immobilised metal affinity chromatography

The supernatant was transferred to a new, cooled 1.5 mL reaction tube. The pellet was resuspended in 1 mL Extraction Buffer B, 4°C by repeated pipetting, and labelled as the 'insoluble' fraction. To the supernatant, 30 µL of a 1:1 slurry Ni-NTA agarose beads (Thermo Fischer Scientific, HisPur Superflow Agarose) pre-equilibrated with Extraction Buffer B was added. The beads were mixed with the supernatant by rotation at 20 rpm for 2 min at 4°C (Labinco, LD79 rotator). The mixture was briefly centrifuged (Eppendorf, 5415 D) (<30 sec) at 2 rcf, 4°C, to pellet the beads. The supernatant was removed, being careful not to agitate the pellet. The pellet was then washed 3 times using the following method: Addition of 1 mL Extraction Buffer B; resuspension and mixing by rotation at 20 rpm for 2 min at 4°C; brief centrifugation (<30 sec) at 2 rcf, 4°C to pellet the beads; and removal of the supernatant, being careful not to agitate the pellet. The pellet is labelled as the 'soluble' fraction.

2.2.3.6. Sample preparation for SDS-PAGE

Following the previous steps, a 'soluble' fraction comprising ~30 µL of a 1:1 slurry of Ni-NTA agarose, and an 'insoluble' fraction comprising a resuspension of the pelleted insoluble components of the whole cell lysate were obtained. For analysis of the solubility of M062, the 'soluble' and 'insoluble' fractions were prepared for SDS-PAGE. 15 µL of the 'insoluble' fraction was transferred to 1.5 mL reaction tube, to this 5 µL of 4xLB was added. To the soluble fraction, 10 µL of 4xLB was added directly to the slurry. The SDS-PAGE samples were then heated at 95°C for 2 min (Eppendorf, Thermomixer Compact), briefly centrifuged (Eppendorf, 5415 D) (<30 sec at 16,100 rcf) and allowed to cool to RT. Samples were analysed by SDS-PAGE immediately or stored at -20°C for analysis at a later date.

2.2.4. Large-scale expression and purification of M062

In order to obtain mg quantities of M062 3CS required for crystallisation screening, expression was scaled up to 1 L cultures. IMAC, IEX and SEC chromatography steps were utilised in the purification of M062 for crystallisation (see Fig. 12, 13). The protocol described here was arrived at following optimisation of the purification strategy, and represents the methodology used

to obtain the crystals of M062-Lau and M062-Mav 3CS used in determination of the structures (Chapter 4).

2.2.4.1. Inoculation and cell growth

An overnight inoculum was prepared as follows: to a 20 mL tube, 15 mL LB (Formedium) with 15 µL ampicillin (Formedium) (1000x) was added. The LB was then inoculated from a glycerol stock, and incubated at 37°C, 180 rpm for 16 – 20 hrs (Infors, Multitron HT). 1 L 2xYT (Formedium) was pre-warmed to 37°C for 15 minutes, followed by the addition of 1 mL ampicillin (1000x) and the overnight culture. The culture was incubated at 37°C, 180 rpm for 1 ½ hr. The optical density at 600 nm (OD_{600}) of the cell culture was monitored from this timepoint (WPA Biowave, CO800).

2.2.4.2. Induction of protein expression and cell lysis

Protein expression was induced at an OD_{600} of 0.6 – 0.8, at 37°C by the addition of 500 µL of a 1 M stock IPTG (VWR), 0.5 mM final concentration. Expression was carried out for 3 hr at 37°C, 180rpm. Cells were pelleted by centrifugation at 2,160 rcf for 15 min at 4°C (Sorvall, RC6 Plus). The supernatant was decanted and treated with Virkon (VWR) prior to disposal. Pellets were either used immediately for downstream protein isolation and purification, or washed with 1x PBS, re-pelletted, and stored at -20°C for use at a later date.

The pellet was fully resuspended in 15 mL Extraction Buffer A (see Section 2.2.1), 4°C by vortexing (VWR, Mini vortexer), and transferred to a cooled 40 mL Dounce homogenizer (Kontes). Proteolysis was minimised by the addition of PMSF (Sigma), 0.5 mM final concentration. Cells were homogenized, and the homogenate transferred to a cooled 50 mL glass beaker.

Cells were lysed utilising a high-pressure cell disruptor (Avestin Emulsiflex C3) coupled to a chiller, 3 cycles at output 10,000 – 15,000 psi, 4°C, resting in an ice-bath for 1 min between each cycle. The cell lysate was then clarified by centrifugation at 26,640 rcf for 45 min, 4°C (Sorvall, RC6 Plus), fractionating the lysate into soluble and insoluble components.

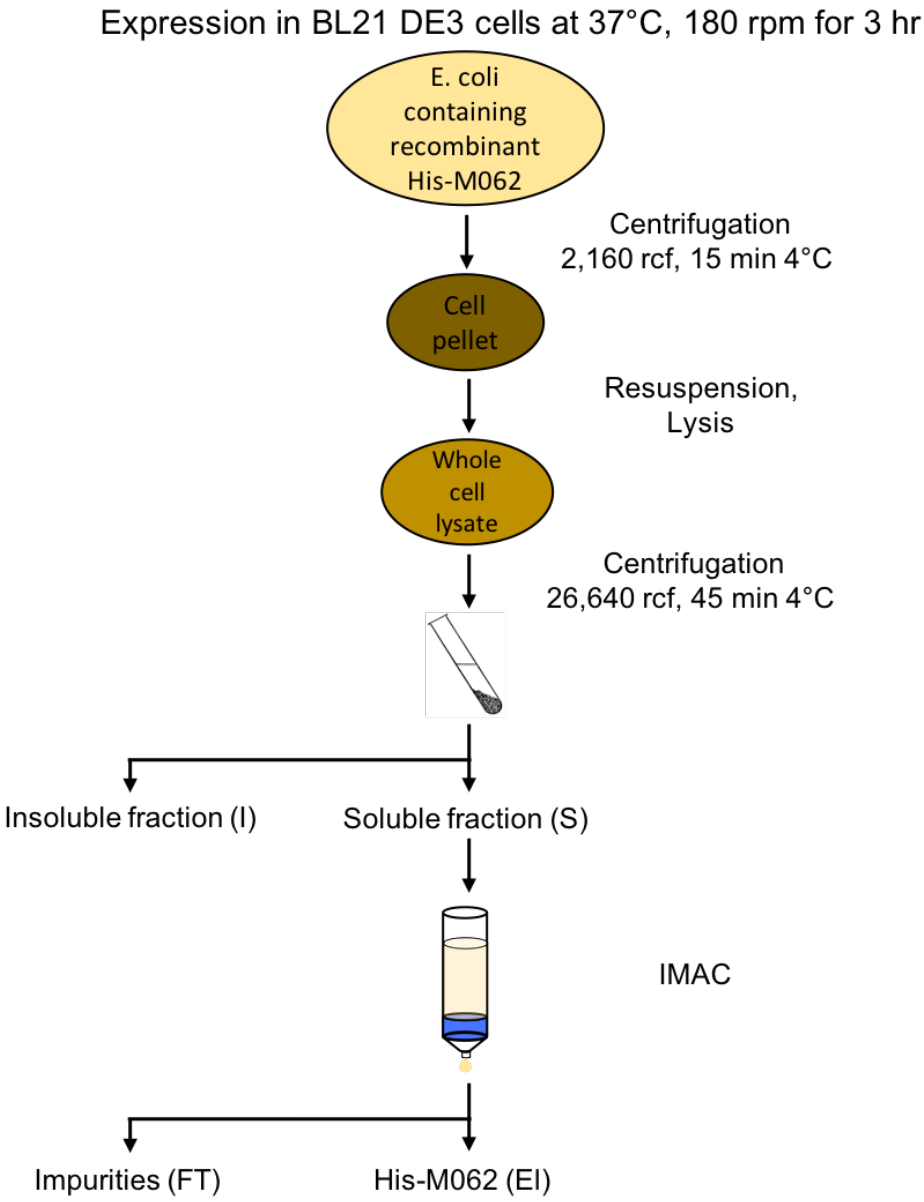


Figure 12: Flowchart of the purification of His-M062. The purification strategy used to isolate recombinantly expressed His-M062 from *E. coli* is depicted. Following recombinant expression of His-M062, cells were pelleted by centrifugation. The cell pellet was then resuspended, and lysed. Soluble (S) and insoluble (I) fractionation was achieved by centrifugation. Application of the soluble fraction to Ni-NTA agarose beads in a gravity column allows for binding of His-M062. Impurities flow-through (FT) the column, allowing for the elution (El) of purified His-M062.

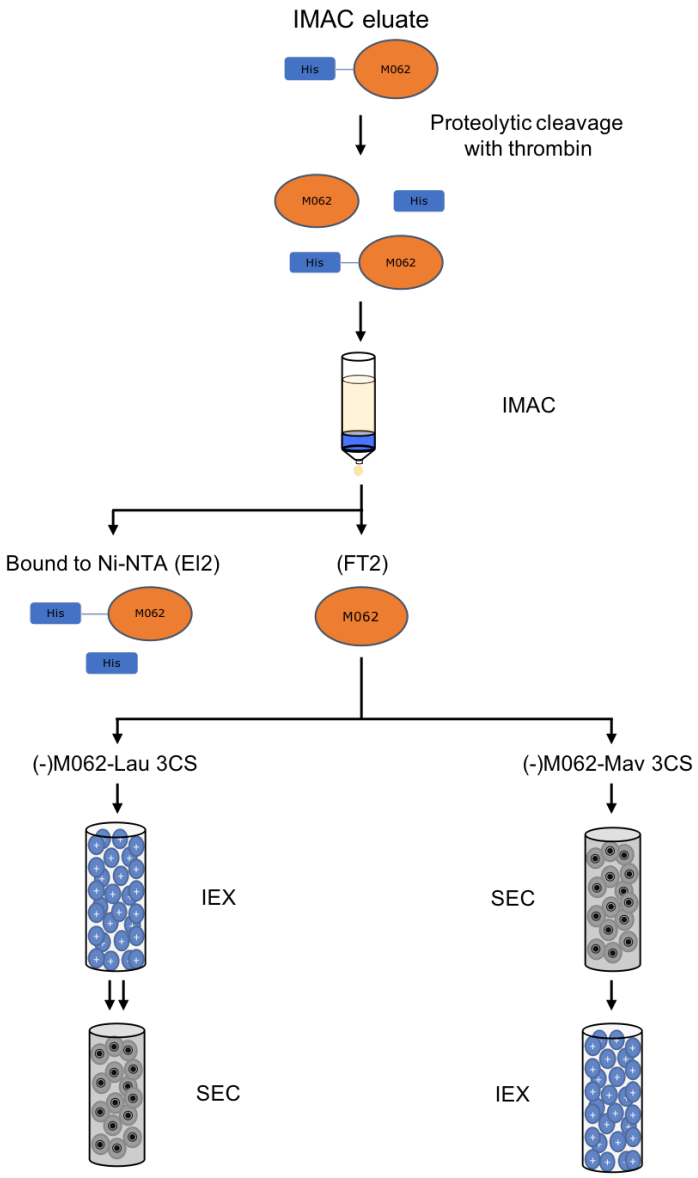


Figure 13: Flowchart of the purification of (-)M062-Lau 3CS and (-)M062-Mav 3CS. The purification strategies used downstream of IMAC eluted His-M062 3CS. Proteolytic cleavage of the His tag (blue) by thrombin yields a mixture of cleaved M062 (-)M062 (orange), His tags, and un-cleaved His-M062 (Cut). Application of the products of proteolysis to IMAC allows for isolation of (-)M062 in the flow-through (FT2). His-tags and un-cleaved His-M062 retained on the Ni-NTA (El2). (-)M062-3CS Lau was purified by IEX, followed by SEC of two fractions, A and B. (-)M062-3CS Lau was purified by IEX, followed by SEC of two fractions. (-)M062-3CS Mav was purified by SEC, followed by IEX.

2.2.4.3. Protein isolation by immobilised metal affinity chromatography

The isolation of His-tagged protein from large-scale expression was carried out with IMAC using gravity flow column chromatography. To a 25 mL gravity column, 4 mL of a 1:1 Ni-NTA agarose (Thermo Fischer Scientific, HisPur Superflow Agarose) bead slurry was added. The beads were pre-equilibrated with ddH_2O, followed by Extraction Buffer A, at RT. The soluble fraction was then added gently, so as not to disturb the agarose resin bed. The soluble fraction is a viscous, protein-rich solution which was allowed to flow through the Ni-NTA bed by gravity, allowing for selective binding the imidazole groups of the hexahistidine tag to the nickel atoms co-ordinated to the NTA-agarose beads, while the majority of non-tagged proteins passed through the column.

Affinity-based chromatography requires minimisation of non-specific targets which have affinity for the immobile phase. Following loading of the Ni-NTA resin with the soluble fraction, the resin was washed with Extraction Buffer A, in order to remove non-specifically bound proteins. The 'wash' flow-through was monitored for its protein content by visual assessment of Bradford's assay. Bradford's assay is a dye-based, colorimetric assay used to measure the total protein content of a sample[104]. In the unbound form, the dye, Coomaisse Blue G-250 appears red, whereas the protein-bound form appears blue. 2 µL of the flow-through were added and mixed by pipetting into 30 µL of Bradford's reagent (BioRad). Washing of the Ni-NTA continued until the flow-through elicited no perceptible colour change.

Elution of the bound His-M062 was achieved by the addition of Elution Buffer A (see Section 2.2.1). The high concentration of imidazole in the elution buffer out-competes for the Ni-NTA, displacing the His-M062. Elution of protein was monitored by Bradford's assay as above. The eluate was collected when a colour change was observed, until a return to the baseline colour change was achieved.

2.2.4.4. Isolation of proteolytically cleaved M062

Removal of this His-tag was carried out for crystallisation screening of M062. pET15b vectors encode a thrombin cleavage site after the N terminal His-tag (see Fig. 8). This allows for selective proteolysis to break the peptide bond leaving native M062, with a residual GSH tripeptide sequence prior to the N terminal Met.

Following proteolysis, a mixture of cleaved protein ((-)M062), His tags, and un-cleaved protein are in solution together (assuming <100% protease efficiency). Isolation of the cleaved protein is achieved by purification of the proteolytically treated mixture by IMAC. The cleaved protein, now lacking a His-tag should have lower affinity for the Ni-NTA and flow-through, whereas the His-M062 (and any His tags) should rebind. However, as the eluate contains a high concentration of imidazole, proteolysis is carried out concurrent to dialysis, in order to lower the concentration of imidazole to prevent its reduction of thrombin's enzymatic activity, and to allow for the His-M062 to bind to the Ni-NTA.

Proteolysis and dialysis of His-M062

To the eluate, thrombin (GE Healthcare) (10 U per 5 mL eluate) was added, transferred to dialysis tubing (Sigma) (12 kDa MWCO), and dialysed against 30x volume Extraction Buffer B, with magnetic stirring (DLab) of the buffer at 200 rpm for 16 – 20 hr at RT. Precipitated protein was removed by centrifugation at 3,000 rcf, 15 min, 4°C prior to IMAC.

IMAC purification of proteolyzed His-M062

To a 4 mL 1:1 Ni-NTA slurry (pre-equilibrated with ddH_2O, followed by Extraction Buffer B), the thrombin-treated eluate was added. The flow-through was collected, washing the resin with an excess of Extraction Buffer B to collect all non-binding protein (monitored by Bradford's assay).

2.2.4.5. Purification of M062 by ion-exchange chromatography

Ion-exchange chromatography allows for the separation of proteins by the selective binding, and dissociation of proteins from the stationary phase by charge-based affinity interactions. (-)M062 3CS Lau and Mav have a pI of ~5.4, (see Table 4) giving them a net negative charge at pH 7.5, at which the IEX was carried out. Anion-exchange chromatography refers to the interaction of negatively charged ions with the immobile phase. A sourceQ 15Q 4.6/100 PE (GE Healthcare), or monoQ 5/50 (GE Healthcare) anion-exchange column was used for the purification (-)M062-Lau and (-)M062-Mav 3CS for crystallisation trials respectively.

(-)M062 was bound to the column, and eluted by increase of the concentration of anions in the mobile phase, achieved by mixing IEX Buffer A and IEX Buffer B which contained a low and high concentration of NaCl respectively. Stable interaction between a (-)M062 and the positively charged anion-exchange resin requires a low enough ionic strength of the buffer to prevent the protein being competed off by the anions in the buffer, therefore a dialysis step was required prior to loading of (-)M062 onto the column, to reduce the concentration of NaCl.

Dialysis of IMAC purified (-)M062

(-)M062 was transferred to dialysis tubing (Sigma)(12 kDa MWCO) and dialysed against 30x volume of IEX Buffer A1 with stirring (magnetic stir box, DLab) at 200 rpm of the buffer for ≥3 hr at RT.

Step-wise gradient elution of (-)M062 by ion-exchange chromatography

IEX was carried out on ÄKTA Explorer 100 (GE Healthcare), at 4°C. FPLC pumps and lines were primed with buffers before connection of the anion exchange sourceQ/monoQ column. The column was equilibrated with 5 CV each: ddH_2O, IEX Buffer A, IEX Buffer B, IEX Buffer A. Following dialysis, (-)M062 was centrifuged at 3,000 rcf, 15 min, 4°C in order to pellet any insoluble aggregates. The protein sample was loaded manually, with injections of 3.5 mL into a 5 mL loop, and applied to the column with IEX Buffer A at a flow rate of 1 mL/min, repeating until all protein was applied.

Step-wise gradient elution of (-)M062-Lau 3CS was carried out at 1 mL/min, with the eluate collected in 1.5 mL fractions (see Table 7 (A)). (-)M062-Mav 3CS was eluted by the application of a linear gradient at 1 mL/min, with the eluate collected in 0.5 mL fractions. (see Table 7 (B)).

A: (-)M062-Lau 3CS

Buffer composition (%B)	16	20	22	25	100		100
CV	2.6	12	16	8	5		1

B: (-)M062-Mav 3CS

Buffer composition (%B)	0	-	60		100
CV	0	30	5		5

Table 7: IEX elution gradients for the purification of (-)M062 3CS Lau (I) and Mav (II) for crystallisation. The composition of the mobile phase, expressed as %B was altered by mixing of IEX Buffer A and IEX Buffer B. The duration of each gradient step, isocratic (#) or linear (-) is shown in CV (I: source, 1.7 mL, II: monoQ, 1 mL). The double-line denotes cessation of fractionation.

2.2.4.6. Size exclusion chromatography

SEC separates proteins based on their size. Porous beads allow proteins smaller than the diameter of their pore to pass through them, which increases their overall pathlength compared to larger proteins which by-pass the pores when flowed over a bead bed in a column. For crystallisation, a pure, homogenous population of protein is desirable, and SEC serves to remove soluble protein aggregates, and separate proteins that vary in size.

A superdex 75 16/60 (GE Healthcare) SEC column was equilibrated with SEC Buffer, 1.2 CV, at 0.8 mL/min. (-)M062 was concentration by centrifugal filtration (Merck Millipore, Amicon-Ultra-15) (10 kDa MWCO), 3000 rcf, 15 min, 4°C (Eppendorf 5810 R) to a final volume ≤5 mL. The concentrated protein was centrifuged at 16,100 rcf, 15 min, 4°C (Eppendorf 5415 D) in order to pellet any insoluble aggregates, and applied to the column, flow rate 1 mL/min, with the eluate collected in 1.5 mL fractions.

2.2.5. Multi-angle light scattering of His-M062 WT

Multi-angle light scattering instruments [Wyatt: refractive index detector (Optilab rEX) and MALS detector (Tristar miniDAWN), with a flow cell sonicator (Comet)] were coupled to an ÄKTA Explorer100 (GE Healthcare) with a Superdex75 10/300 (GE Healthcare) SEC column. The column and MALS instruments were equilibrated with SEC Buffer, 0.7 mL/min. The flow cell was sonicated, and purged and flow continued until a background LS signal of <0.3 was obtained.

His-M062-Lau WT and His-M062-Mav WT were isolated by IMAC as per Section 2.2.4. IMAC eluates were dialysed against 30x volume SEC Buffer for 3 hr at RT. insoluble particulates were pelleted by centrifuged at 3,000 rcf for 15 min at 4°C (Eppendorf, 5810 R). 0.8 mg was injected (His-M062-Lau WT: 0.4 mg/mL, His-M062-Mav WT: 0.46 mg/mL), with a flow rate of 0.7 mL/min, for 1.5 CV. Data were collected and processed using ASTRA for windows, version 4.90.08 QELLS 2.

2.2.6. Crystallisation of M062

Crystallisation of (-)M062 was achieved by the vapour-diffusion method. Sparse matrix screening[105], using commercially available screens to sample different crystallisation conditions of various buffering reagent, pH, salts, anions, cations, precipitants and additives was initially carried out, followed by refined screening around initial 'hits', assessed by scoring each crystallisation experiment over time. While initial crystals were obtained with relative ease, the size and morphology of the crystals required optimisation to allow for high resolution x-ray diffraction quality crystals.

The application of crystallization techniques such as reverse vapour-diffusion, seeding, salting-out and temperature-gradient crystallisation screening were employed in an attempt to produce a crystal with higher resolution diffraction, from which the structures of Lau and Mav (-)M062 3CS could be solved.

2.2.6.1. Concentration of (-)M062 3CS for crystallisation screening

Purified (-)M062-Lau and (-)M062-Mav were prepared as described in Section 2.2.4. Pure fractions were pooled and concentrated by centrifugal filtration (Merck Millipore, Amicon-Ultra-15) (10 kDa MWCO), at 3000 rcf, 4°C (Eppendorf 5810 R) in 15 min periods. The solution was mixed by gentle pipetting in between centrifugation step to maintain homogenous dispersion of protein in the solution, as centrifugal filtration can produce a concentration gradient, with proteins accumulating at the semi-permeable membrane.

It was found that (-)M062 3CS was highly susceptible to precipitation at concentrations > 1 mg/mL, with protein solubility appearing to correlate with temperature, and inversely with NaCl concentration, i.e. (-)M062 3CS at > 1mg/mL would precipitate to a higher degree at 4°C than at RT, and at 150 mM NaCl than at ~300 mM NaCl. Due to this temperature-dependent solubility behaviour, (-)M062 3CS was kept at RT in between centrifugation steps, and during crystallization set-up. The concentration of NaCl at which crystallisation was screened was varied, with protein concentration between 1-3 mg/mL at 150 mM NaCl, and between 3 – 12 mg/mL at ~300 mM NaCl.

The crystals from which structural diffraction data were obtained were screened with (-)M062-Lau 3CS concentrated in SEC Buffer (150 mM NaCl), to a final concentration of 2.7 mg/mL, and (-)M062-Mav 3CS concentrated in mixture of IEX Buffer A2 and B2 (~300 mM NaCl), to a final concentration of

10.5 mg/mL. Concentration was determined following centrifugation at 16,100 rcf, 15 min, 4°C (Eppendorf 5415 D) in order to pellet any insoluble aggregates by Abs_{280} (Nanodrop, ND-1000), using the estimated ε_{280} of a 1% (w/v) solution: (-)M062-Lau 3CS, 16.86 $1\%^{-1}$ cm^{-1}, (-)M062-Mav 3CS 16.87 $1\%^{-1}$ cm^{-1}.

2.2.6.2. Sparse matrix screening

Sparse matrix screening was set up using the sitting-drop vapour-diffusion method in 96-well plates (Jena Biosciences, MRC SD-2) by an automated liquid-handling robot (TTP Labtech, mosquito crystal). To each sub-well, 70 µL of a crystallisation condition screening solution was transferred by manual pipetting from a commercial 96-well screen (see Table 8). Using the automated liquid-handling robot, into each sitting well, 150 - 200 nL of protein was transferred, followed by the addition of an equal volume of the corresponding well's reservoir, containing the crystallisation screening solution. Plates were sealed (Hampton Research, Crystal clear), and transferred to a 16°C cooled incubator (Rumed, 3000) for the duration of crystallisation.

Screen	Manufacturer
Wizard	Jena Biosciences
Pact premiere	Molecular Dimensions
Structure	Molecular Dimensions
JCSG+	Molecular Dimensions
Midas	Molecular Dimensions
JBScreen 1-4	Jena Biosciences
JBScreen 5-8	Jena Biosciences

Table 8: Sparse matrix commercial screens.

2.2.6.3. Scoring crystallisation conditions

Identification of sparse matrix screening conditions with which targeted-screening for crystallisation could be carried out required evaluating each condition's impact on protein phase-behaviour. Conditions were screened by manual inspection of the protein : reservoir drops by stereomicroscope (Nikon, SMZ800). Visual signs of potential crystallisation-promoting conditions were monitored, noting changes in protein solubility such as levels of precipitation,

spherulite formation, or the presence of crystals. Screen conditions were evaluated immediately following set-up, daily for 3-5 days, weekly for ~ 2 months, and intermittently thereafter for up to 1 year. 'Hits' were identified as conditions which facilitated crystallisation, by the growth of crystals, or nuclei with crystal-like morphology.

2.2.6.4. Optimisation of crystallisation

Following the identification of conditions which promote crystallization from sparse matrix screening, targeted-screening was carried out by modifying the parameters of the screening condition. Incremental variation of the pH, precipitant and additive concentrations identified in the initial 'hit' was carried out using the hanging-drop vapour-diffusion method. In 24-well plates (Molecular Dimensions, XRL plate). Targeted-screening condition solutions were prepared at RT, the pH adjusted to ± 0.01 units, and syringe filtered (Starstedt, Filtropur, 0.2 µm).

To each sub-well, 500 µL of a screening condition was transferred by manual pipetting. Drops were set-up by pipetting a mixture of reservoir and protein with total volume 1 – 4 µL onto siliconized circular glass cover slips (Jena Biosciences, 22 mm). A ratio of 1:1 reservoir : protein was usually used, however alternate ratios were frequently used (on the same cover slip) to sample a range of protein : reservoir final concentrations. The cover slips were then placed, inverted, upon the silicon grease-coated (Jena Biosciences) rim of the corresponding well, forming an airtight seal between well and cover slip. Plates were transferred to a 16°C cooled incubator (Rumed, 3000) for the duration of crystallisation.

2.2.6.5. Seeding

A crystal seed stock was prepared by harvesting and crushing a polynucleated crystal grown in a condition used for the crystallisation of (-)M062-Mav 3CS (see Fig. 14), followed by serial dilution in the crystallisation solution. From the hanging-drop well, the cover slip was removed, inverted and the drop exposed. To the drop, 2 µL of reservoir solution was added to prevent rapid evaporation. 250 µL of reservoir solution was added to a 1.5 mL reaction tube, and the crystal transferred using a harvesting loop. The crystal was crushed by manual manipulation with a micro-prober tool (Hampton Research),

followed by brief vortexing. Three 10-fold dilutions of this seed stock were prepared by the addition, using a cut-off pipette tip, of 10 µL of stock to 90 µL of reservoir, mixing, and repeating the dilution twice using the newly diluted stock solution (see Fig. 14).

Seeding was carried out in 24-well targeted-screening conditions, either at initial setting-up of the screen, or by the addition of the seed stock to drops following an equilibration period. In order to screen the seed stock dilution series for appropriate nucleation, replicate reservoirs were set up. Drops 4:4:1 - protein : reservoir : seed stock with final volume 2.5 µL were set-up as per section 2.2.6.4.

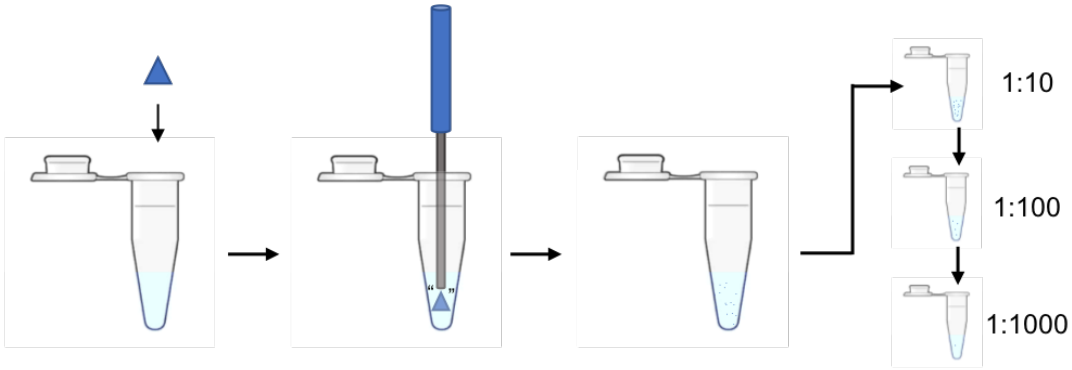


Figure 14: Preparation of seed stock of varying concentration. A crystal (blue triangle) is transferred to a 1.5 mL reaction tube containing reservoir buffer. The crystal is crushed by manipulation with a microprobe tool, followed by three 10-fold dilutions of the seed stock with reservoir buffer.

2.2.6.6. Salting-out

Reduction in the concentration of NaCl by dialysis of purified of M062-Lau 3CS was carried out to investigate whether crystallisation could be induced by 'salting-out'. 500 µL of a 1 mg/mL sample of (-)M062-Lau 3CS was transferred to dialysis tubing (Sigma) (12 kDa MWCO), and dialysed against 30x volume modified SEC Buffer, final concentration NaCl, 50 mM for 3 days at RT. The dialysis tubing was inspected daily for changes in solubility, or for crystal growth.

2.2.6.7. Reverse vapour diffusion

Reverse vapour diffusion was trialled as a means of crystallisation by setting up (-)M062-Mav 3CS, final concentration 3.4 mg/mL in SEC Buffer against a

reservoir of lower net osmotic strength. A 2 µL drop of a 1:1 ratio of protein : ddH_2O was set up against a reservoir of ddH_2O.

2.2.6.8. Temperature-dependent crystallisation trial

Controlled decrease of the temperature of a saturated solution of (-)M062-Mav 3CS was trialled as a means of crystallisation. 10 µL of (-)M062-Mav 3CS, final concentration 3.4 mg/mL in SEC Buffer was transferred to an x-ray capillary tube, and the end heat sealed. The tube was transferred to a 500 µL PCR reaction tube, and submerged in 200 µL ddH_2O. The reaction tube was transferred to a thermocycler (Eppendorf, Mastercycler personal) where the cycle in Table 9 was applied.

Step	Process	Time (min)	Temperature
1	Initial temperature	5	20°C
2	Cooling step	20	-0.1°C
3	Repeat step 2 x 99	-	-
4	Hold	∞	10°C

Table 9: Temperature-gradient crystallisation protocol

2.2.6.9. Expression of selenomethionine-derivatised (-)M062-Mav 3CS

Selenomethionine-derivatised M062-Mav 3CS was expressed as per the protocol used by Clardy et al.[106]. An overnight culture was prepared by inoculating 5 mL LB (Formedium) with 5 µL ampicillin (Formedium) (1000x), and incubating at 37°C, 180 rpm for 16 – 20 hrs (Infors, Multitron HT). To exclude methionine, expression was carried out by inoculation of minimal media, supplemented with amino acids to prevent methionine biosynthesis (500 mL, see Table 10).

Prior to induction, the media was supplemented with methionine feedback inhibiting amino acids Lys, Thr, and Phe (100 mg / L media each), Leu, Ile, and Val (50 mg / L media each), and L-(+)-selenomethionine (Acros) (50 mg/ L media). Expression and purification was then carried out as per the native protein, with supplementation of β-ME to 7 mM during lysis, to further prevent oxidation of selenomethionine.

Minimal media

10x M9 salts, 1 L (CSH receipe[107]): 37 g sodium phosphate dibasic, 30 g potassium phosphate monobasic, 5 g NaCl, 1 g NH_4Cl, autoclaved.

Volume	Reagent	Stock concentration	Final concentration
50 mL	M9 salts	10x	1x
1 mL	$MgSO_4$	1 M	2 mM
10 mL	Glucose	20% w/v	0.4%w/v
5 mL	BME vitamins	100x	1x
5 mL	†trace elements	100x	1x
50 µL	$CaCl_2$	1 M	0.1 mM
0.4 mL	$Fe(III)Cl_3$	5% w/v	0.004% w/v
25 mL	‡Amino acid mix	2 mg/mL	0.1 mg/mL
3.6 mL	ddH_2O	-	-

Table 10: Composition of 0.5 L minimal media for selenomethioinine protein expression. † Trace elements (100x): 5 g/L EDTA, ZnCl2 50 mg/L, CuCl2 10 mg/L, CoCl2 10 mg/L, H3BO3 10 mg/L, MnCl2 1.6 g/L, pH 7.0. ‡ Amino acid mix: 2 mg each, 20 common amino acids apart from methionine. All reagents sourced from Sigma.

2.2.6.10. Harvesting, cryoprotection and storage of crystals

Crystals were harvested by manually 'fishing' the crystals from drops using microloops/micromounts (MiTeGen). Loop size was chosen to allow for the crystal to be transferred with adequate mother liquor to envelope the crystal on the loop. Removal of crystal clear tape from 96-well trays, or inversion of the cover slip from 24-well trays exposed the drop of interest. 1-2 µL of reservoir was added to the drop to prevent rapid evaporation. Cryoprotectant was then supplemented to the drop, and/or the crystal transferred to cryoprotectant solution to 'soak' prior to snap freezing in LN2. Cryoprotective reagent, concentration, and soak times were varied in an attempt to optimise this step for downstream diffraction data quality.

Crystals were snap frozen by submersion in LN2 for 30 – 60 sec, and transferred to cryovials under LN2 for storage for up to 1 year. Crystals were transferred to Uni-Pucks (Molecular Dimensions) and shipped by dry LN2 Dewar to synchrotrons for x-ray diffraction data collection.

2.2.7. X-ray diffraction data collection

2.2.7.1. Synchrotron

Advanced Photon Source, Argonne National Laboratory

Diffraction data used for structure solution of M062-Lau 3CS and M062-Mav 3CS were collected using NE-CAT beamline 24-ID-C (GM124165), a Pilatus 6M-F detector (RR029205), an Eiger detector (OD021527) at the APS (DE-AC02-06CH11357).

Synchrotron Soleil

Selenomethionine energy scan and x-ray fluorescence scan were performed using beamline Proxima 2-A, an Eiger X 9 M detector, and Ketek SDD detector for X-ray fluorescence scans.

2.2.7.2. Diffraction data collection

M062-Lau 3CS

Diffraction data for M062-Lau 3CS was collected by Dr. Amir Khan at APS. 600 images were collected at a wavelength of 0.979100 Å, at 100 K, exposure time 0.2 sec, 0.3° oscillation with a detector distance of 379.89 mm.

M062-Mav 3CS

Diffraction data for M062-Lau 3CS was obtained with 600 images collected at a wavelength of 0.97910 Å, at 100 K, exposure time 0.2 sec, 0.2° oscillation with a detector distance of 600 mm.

Selenomethionine

Se incorporation into M062-Mav 3CS was assessed by carrying out an energy scan from 12,608 – 12,728 eV, at the synchrotron Soleil.

X-ray fluorescence scan

An x-ray fluorescence energy scan of M062-Lau 3CS was carried out from 8,238 – 8,403 eV, at the synchrotron Soleil.

2.2.8. Diffraction data processing and M062 structure refinement

2.2.8.1. Data processing

All M062-Lau 3CS and M062-Mav 3CS data processing was carried out by Dr. Amir Khan using CCP4 suite software[108].

2.2.8.2. Molecular replacement of M062-Lau 3CS using M064 as a model

M064 was used as a model for MR of M062-Lau 3CS. The CCP4 utility Chainsaw[109,110] was used to modify the coordinate PDB file (5CZ3). Residues 4 – 138 of chain A, truncated to C_β was used as a search model for MR of M062-Lau 3CS. MR was carried out using PHENIX: Phaser[111].

2.2.8.3. Molecular replacement of M062-Mav 3CS using M062-Lau 3CS as a model

M062-Mav 3CS was phased using PHENIX: Phaser[111], using M062-Lau 3CS as an search model, following refinement of the M062-Lau 3CS structure.

2.2.8.4. Structure refinement

Model building was performed using Coot[112] for both Lau and Mav M062 3CS structures. Structure refinement was performed with PHENIX.refine[113]. TLS groups were calculated with TLSMD embedded in PHENIX.refine[114115].

2.2.8.5. Validation

PHENIX validation tools, and Molprobity[116] were used to analyse structural parameters during, and after structure refinement.

The CheckMyMetal webserver[117118] was used to validate the Ni-binding by residues -2 – 0 of M062-Lau 3CS (insert PDB code X).

2.2.9. Expression and purification of ^{15}N, ^{13}C-labelled M013 1-93 4Mut for NMR structural analysis

The structure of M013 was investigated by NMR spectroscopy in this project. NMR is a spectroscopic technique, which acquires information about the chemical environment of an atom, based on the electromagnetic frequency at which collation between nuclei occurs. NMR spectroscopy relies on the nuclei within the sample possessing integral spin quantum numbers greater than ½[119].

2.2.9.1. Preparation of isotopically-enriched minimal media, for double-labelling of M013

In order to increase the proportion of NMR-interacting nuclei in protein, incorporation of isotopes ^{13}C and ^{15}N was carried out, to produce a double-labelled protein. Recombinant expression of M013 was carried out by *E. coli* cells cultured in minimal media supplemented with ^{13}C glucose and ^{15}C NH_4Cl (Cambridge Isotope Laboratories, Inc.). All other reagents from Table 11 from Sigma.

Minimal media

10x M9 salts, 1 L (ammonium chloride free) (CSH receipe[107]): 37 g sodium phosphate dibasic (Sigma), 30 g potassium phosphate monobasic, 5 g NaCl, autoclaved.

Volume	Reagent	Stock concentration	Final concentration
100 mL	M9 salts	10x	1x
2 mL	$MgSO_4$	1 M	2 mM
20 mL	^{13}C glucose	20% w/v	0.4%w/v
10 mL	BME vitamins	100x	1x
10 mL	†trace elements	100x	1x
100 µL	$CaCl_2$	1 M	0.1 mM
0.8 mL	$Fe(III)Cl_3$	5% w/v	0.004% w/v
4 mL	^{15}N NH_4Cl	25% w/v	0.1% w/v
853.1 mL	ddH_2O	-	-

Table 11: Composition of 1 L minimal media for ^{15}N, ^{13}C double-labelled protein expression. † Trace elements (100x): 5 g/L EDTA, ZnCl2 50 mg/L, CuCl2 10 mg/L, CoCl2 10 mg/L, H3BO3 10 mg/L, MnCl2 1.6 g/L, pH 7.0.

2.2.9.2. Inoculation and cell growth

In order to minimize the abundance of naturally occurring isotopes of carbon and nitrogen the cell culture, a restricted inoculum was prepared (10 mL overnight per 2 L minimal media). 10 mL LB, 15 µL kanamycin (Formedium) (1000x) was inoculated from a glycerol stock, and incubated at 37°C, 180 rpm (Infors, Multitron HT) for 16 – 20 hrs. 2 L minimal media was pre-warmed to 37°C for 15 minutes, followed by the addition of 2 mL kanamycin (1000x) and the overnight culture. The culture was incubated at 37°C, 180 rpm for 3 hr. The growth rate of *E. coli* in minimal media is greatly reduced, requiring longer incubation times before reaching the required cell density for protein expression. The optical density at 600 nm (OD_{600}) of the cell culture was monitored from this timepoint (WPA Biowave, CO800).

2.2.9.3. Induction of protein expression and cell lysis

Induction

At an OD_{600} 0.6 – 0.8, the temperature of incubation was reduced to 18°C. Recombinant protein expression was induced at 18°C by the addition of 1 mL of a 1 M stock IPTG (VWR), 0.5 mM final concentration. Expression was carried out for 18-22 hr at 18°C, 180rpm. Cells were pelleted by centrifugation at 2,160 rcf for 15 min at 4°C (Sorvall RC6 Plus). The supernatant was decanted and treated with Virkon (VWR) prior to disposal. Pellets were washed with 1x PBS, re-pelletted, and stored at -20°C for use at a later date.

Cell lysis was performed as described in Section 2.2.4.2.

2.2.9.4. Protein isolation by immobilised metal affinity chromatography

In order to maximise the yield of double-labelled M013, prior to gravity column IMAC, the batch method was used to ensure maximal binding of His-M013 1-93 4Mut to the Ni-NTA agarose.

In a 50 mL falcon tube (Greiner), the soluble fraction was added. To this, 10 mL 1:1 slurry of Ni-NTA resin (Thermo Fischer Scientific, HisPur Superflow Agarose), pre-equilibrated with Extraction Buffer A, was added. The beads were mixed with the supernatant by rotation at 5 rpm for 1 hr at 4°C. The Ni-NTA : soluble fraction slurry was then transferred to a gravity column, and His-M013 1-93 4Mut isolated as described in Sections 2.2.9.6 and 2.2.9.7, carried out at 4°C.

2.2.9.5. Expression and purification of GST-tagged HRV 3C protease

GST-tagged HRV 3C protease was transformed into BL21 DE3s as per Section 2.2.2.7. An overnight culture was prepared by inoculating 20 mL LB (Formedium) with 20 µL ampicillin (Formedium) (1000x), and incubating at 37°C, 180 rpm for 16 – 20 hrs (Infors, Multitron HT). 1 L 2xYT (Formedium) was pre-warmed to 37°C for 15 minutes, followed by the addition of 1 mL ampicillin (1000x) and the overnight culture. The culture was incubated at 37°C, 180 rpm for 1 ½ hr. The OD_{600} of the cell culture was monitored from this timepoint (WPA Biowave, CO800).

At OD_{600} ~0.45 the incubator temperature was dropped to 20°C and expression induced with IPTG (VWR), 0.2 mM after 10 minutes. Expression was

carried out for 16-20 hr at 20°C, 180rpm. Cells were pelleted by centrifugation at 2,160 rcf for 15 min at 4°C (Sorvall, RC6 Plus). The supernatant was decanted and treated with Virkon (VWR) prior to disposal.
Cell lysis was carried out as per Section 2.2.4.2.

The soluble fraction was then incubated with 20 mL 1:1 slurry glutathione sepharose resin (GE Healthcare), pre-equilibrated with Extraction Buffer C, by rotation at 7 rpm for 30 min at 4°C. The resin : soluble fraction slurry was then transferred to a gravity column, and washed with Extraction Buffer C, and eluted with Elution Buffer B. The eluate was dialysed against Dialysis Buffer A for 16-20 hr at 4°C. Alliquots of 100 µL, ~5 mg/mL were snap frozen in LN2 and stored at -80°C.

2.2.9.6. Proteolytical cleavage of His-MBP-M013 1-93 4Mut

Removal of the His and MBP solubility tags was required prior to carrying out the NMR structural analysis of M013. The pace3-MBP vector encodes a HRV 3C protease cleavage site located downstream of the His-MBP tag. Proteolytic cleavage yields a mixture of cleaved M013 ((-)M013), His-MBP, and un-cleaved protein.

To the Ni-NTA eluate, HRV 3C PreScission protease was added to a final concentration of ~1.4 mM. The eluate was transferred to dialysis tubing (Thermo Scientific, Snakeskin) (3.5 kDa MWCO), and dialysed against 30x volume Dialysis Buffer B, with magnetic stirring (DLab) of the buffer, 200 rpm for 36 hr at 4°C. An extended contact time with the protease was carried out to increase the yield of (-)M013.

2.2.9.7. Isolation of (-)M013 by ion-exchange chromatography

Isolation of (-)M013 required optimisation of the purification strategy downstream of His-MBP-M013 cleavage. The final strategy utilised is presented in Fig. 15.

IEX

The protease-treated eluate was centrifuged at 3,000 rcf for 15 min at 4°C (Eppendorf 5810 R) to remove any insoluble material, prior to being loaded manually onto a sourceQ 15Q 4.6/100 PE column (GE Healthcare). Large-scale preparations yielded significant quantities of protein, to prevent overloading of

the column, <25 mg/mL resin was injected. The FT, containing unbound His-MBP was collected for analysis by SDS-PAGE.

Step-wise gradient elution of (-)M013 was carried out at 1.5 mL/min, with the eluate collected in 1 mL fractions (see Table 12). Fractions were analysed by SDS-PAGE, and (-)M013 pooled for downstream SEC.

Buffer composition (%B)	0	-	9	-	11		100
mL	0	1.8	5	8.3	0		Waste

Table 12: IEX elution gradient for the purification of (-)M013. The composition of the mobile phase expressed as %B was altered by mixing of IEX Buffer A3 and IEX Buffer B3. The duration of each gradient step, isocratic (#) or linear (-) is shown in mL. The double-line denotes cessation of fractionation.

2.2.9.8. Size exclusion chromatography

A superdex 75 16/60 column (GE Healthcare) was equilibrated with SEC Buffer, 1.2 CV, at 0.8 mL/min.

IEX fractions containing (-)M013 were pooled and concentrated by centrifugal filtration (Merck Millipore, Amicon-Ultra-15) (3 kDa MWCO), 3000 rcf, 15 min, 4°C to a final volume ≤5 mL. The concentrated protein was centrifuged at 16,100 rcf, 15 min, 4°C (Eppendorf 5415 D) in order to pellet any insoluble aggregates, and applied to the column, flow rate 1 mL/min, with the eluate collected in 1.5 mL fractions.

IEX fractions of (-)M013 contained substantial levels of high Mw impurities, therefore two rounds of SEC were utilised as a final purification step. A single run did not have sufficient resolving capability to fully separate (-)M013 from the higher Mw impurities. Following an initial SEC step, fractions were analysed by SDS-PAGE. Fractions of pure (-)M013 which eluted earliest were kept aside. Fractions containing both (-)M013 and high Mw impurities were pooled, concentrated and a secondary SEC step applied, the pure (-)M013 fractions pooled together with those from the initial run, concentrated to a final concentration of 6-7 mg/mL, and snap frozen in LN2.

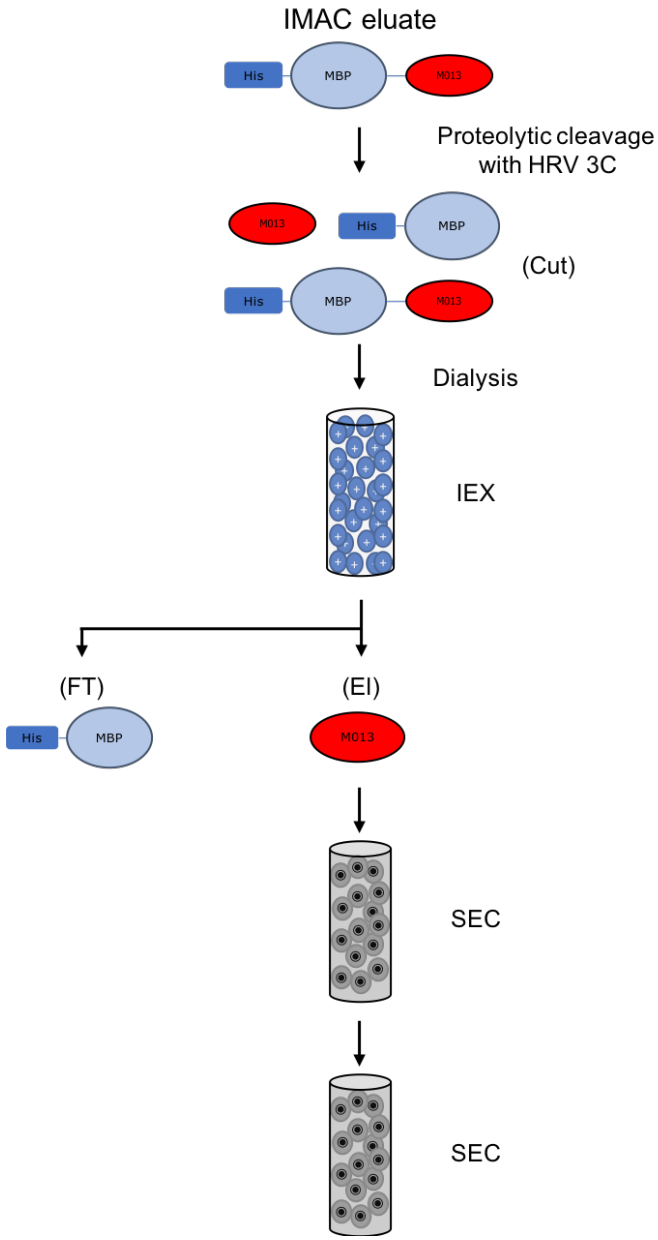


Figure 15: Flowchart of the purification of (-)M013 1-93 4Mut. The purification strategy used downstream of IMAC eluted His-MBP-M013 1-93 4Mut. Proteolytic cleavage of the His-MBP solubility tag (blue) by HRV 3C protease yields a mixture of (-)M013 (red), His-MBP, and un-cleaved His-MBP-M013 1-93 4Mut (Cut). Dialysis of the products of proteolysis allowed for the removal of His-MBP (FT), and isolation of relatively pure (-)M013 1-93 4Mut by IEX. Sequential SEC steps were required in order to remove un-cleaved, high Mw contaminants to yield pure M013 1-93 4Mut.

2.2.10. NMR structural analysis of double-labelled M013 1-93 4Mut

Purified ^{15}N,^{13}C-double labelled M013 1-93 4Mut recombinant protein was flash frozen in LN2 and shipped on dry ice to Dr. Arnout Kalverda of the Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, UK, for NMR structural analysis.

2.2.11. Circular dichroism analysis of M013

Single mutant (Y24G) or 4Mut (M21A, Y24G, Y31S, F68S) M013 1-93 was purified following the scheme presented in Figure 15. Proteins were concentrated to 1.5 mg/mL and dialysed overnight at 4°C against 1xPBS. CD experiments were conducted with a protein concentration of 0.5 mg/mL, 1:2 dilutions made into ddH_2O for the temperature gradient experiment, or buffered solutions: 20 mM phosphate pH 7.0, 20 mM Tris at pH 7.4, 7.6, 7.8 or 8.0. Ellipticity measurements were collected with a Jasco J-810 spectropolarimeter from 250 – 190 nm, Pitch: 1 nm, Scanning speed: 50 nm/min, Bandwidth: 1 nm, Accumulation: 4.

2.2.12. Analysis of protein samples by SDS-PAGE

SDS-PAGE of proteins in this project was carried out as per the CSH protocol[120]. Proteins were visualised with Coomaisse Brilliant Blue staining. Gels were stained by transferring to a plastic dish and covering in Coomaisse Brilliant Blue stain. Gels were incubated in the stain with rocking for 30 minutes. The stain was decanted, and the gel covered with, and incubated in Destaining Solution until the solution was saturated. The saturated solution was decanted, fresh Destaining Solution added, and the gel incubated until the background was clear. Gels were then transferred to water, and allowed to sit overnight before imaging (Canon, Pixma MG2550S).

Chapter 3: Expression, purification and crystallisation of Lausanne and Mav variants of M062

3.1. Aims

The structural basis of the poxviral host range factor C7 family's antagonism of host immunity remains poorly understood. MXYV protein M062, a functional homologue of C7[8] has a naturally occurring single residue mutation at position 79, with the Lausanne strain[16] encoding an Ile in this position, and the Mav strain encoding a Thr[19]. Co-immunoprecipitation studies by our collaborators, Dr. Bernince Nounano in the lab of Dr. Jia Liu, at the University of Arkansas for Medical Sciences have demonstrated that this single amino acid difference significantly alters M062's ability to pull-down its cellular target: SAMD9 (see Fig. 16).

While crystal structures of C7 (PDB: 5CYW) and M064 (5CZ3)[29] have been solved, the underlying structural basis for the C7 family's antagonism is not fully understood. In order to compare the Lau and Mav M062 variants at a structural level, high resolution crystal structures are required. This chapter aims to describe overexpression and purification of recombinant Lau and Mav variants of M062, and the carrying out crystallisation trials from which crystals capable of obtaining high resolution diffraction data were obtained.

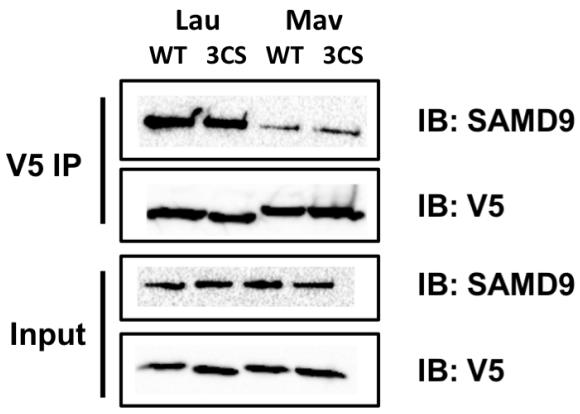


Figure 16: Co-immunoprecipitation of SAMD9 with V5-tagged M062 variants. Hela cells were infected with Lau or Mav variants of M062, either WT or 3CS mutant at an MOI of 3pfu/cell. 24hr post-infection, cells were harvested and either subjected to co-IP or loaded to a 10% SDS PAGE. Whole cell extracts and immunoprecipitates were analyzed by western blotting. Each blot (IP or Input) was first stained for anti-SAMD9 antibody, stripped, and then probed for anti-V5 tag antibody. Expression of V5-M062 appears equivalent for all variants. Co-IP of SAMD9 with V5-M062-Lau WT and 3CS variants appears greater than V5-M062-Mav WT and 3CS. WT and 3CS variants appear to immunoprecipitate SAMD9 to comparable levels. Data provided by Dr. Jia Liu, unpublished.

3.2. Solubility expression test of M062-Mav WT

M062-Mav WT was transformed into BL21DE3s (see Materials and Methods 2.2.2.7). The quantity and solubility of recombinant M062 was first assessed at small-scale (see Materials and Methods 2.2.3), prior to scaling up to produce the mg quantities of M062 required for crystallisation screening. His-M062-Mav WT expression and solubility is greater at 37°C than at 18°C (Fig. 17). Lanes containing high levels of His-M062-Mav WT contain a band at the interface between the stacking and separating gel, indicating the presence of high Mw aggregates which cannot enter the resolving gel.

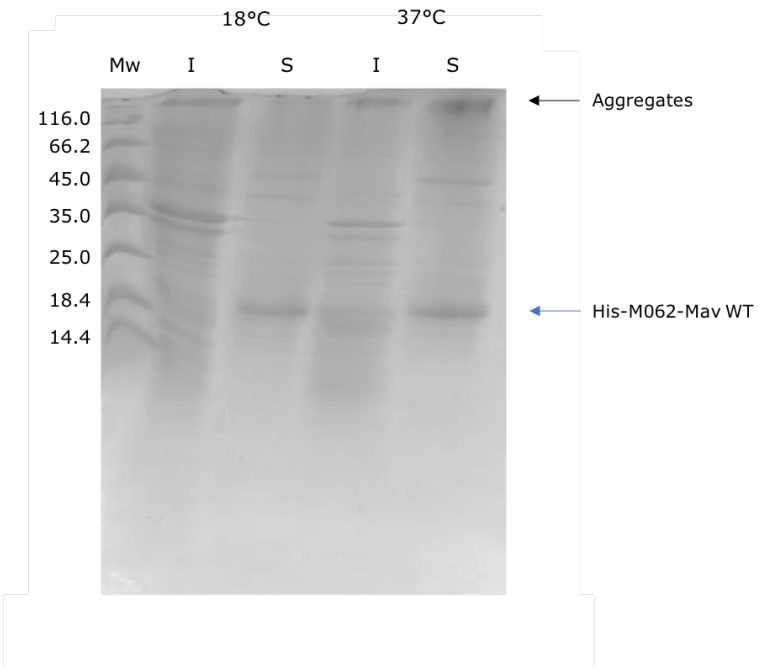


Figure 17: Small-scale expression test of His-M062-Mav WT. 15% SDS-PAGE analysis of insoluble (I) and soluble (S) fractions of BL21DE3s expressing recombinant His-M062-Mav WT. Samples prepared as per Materials and Methods 2.2.12. Molecular weight ladder (Mw) shows His-M062-Mav WT runs as a single band at ~20 kDa (blue arrow). Aggregates of high Mw (black arrow) are present at the interface between the stacking and separating gel.

3.3. Large-scale expression and purification of His-M062-Mav WT

Scaling up expression to 0.5 L, it was possible to obtain 92% pure His-M062-Mav WT by IMAC (see Materials and Methods 2.2.4.3), (see Fig. 18). Protein purity was calculated by densitometry, using imageJ[121], disregarding the high Mw aggregates present at the stacking and separating gel interface. Samples were taken during purification to monitor protein purity and degradation by SDS-PAGE.

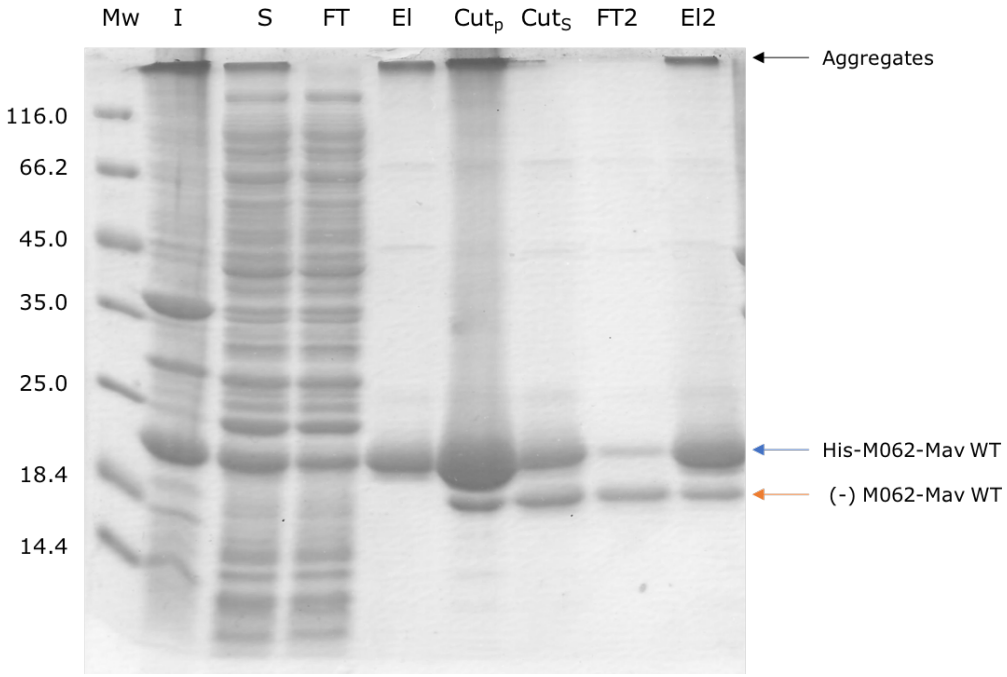


Figure 18: IMAC purification of His-M062-Mav WT. 12% SDS-PAGE anlaysis of His-M062-Mav WT through IMAC purification. His-M062-Mav WT (blue arrow) was isolated by IMAC (El) from the soluble fraction (S). High Mw aggregates are seen at the interface between the stacking and separating gel (black arrow). After thrombinisation of the eluate, precipitation was removed by centrifugation. A sample of the pellet (Cut$_p$) and supernatant (Cut$_S$) were run. Cut$_P$ shows a high degree of band smearing, and high Mw aggragates. Cleaved M062 (orange) runs as a singel band at ~18.4 kDa. Application of Cut$_S$ to IMAC yields mostly pure (:) (-)M062-Mav WT in the flow-through (FT2).

3.3.1. Solubility of His-M062 during IMAC purification

Following purification of mg quantities of His-M062-Mav WT by IMAC, it was observed that within 10 – 30 minutes post-elution from the Ni-NTA, significant levels of precipitation of the eluate occurred. This effect was exacerbated on ice, and at lower temperatures, with the finding that returning the solution to RT would re-solubilise some of the precipitated protein, and reduce the opacity of the solution. In Figure 18, the high Mw aggregate band in lane 'El' is ~55% the intensity of His-M062-Mav WT, a significant portion of the overall IMAC eluate.

Proteolysis by thrombin of the eluted protein resulted in further precipitation, as a result of the duration required for cleavage of the His-tag. The proteolyzed solution was clarified by centrifugation, resulting in a significant white pellet (Cut_P), with the supernatant (Cut_S) applied to IMAC for downstream purification.

3.3.2. SEC of (-)M062-Mav WT

SEC was utilised in order to investigate whether the aggregates observed by SDS-PAGE (see Fig. 18) were present as soluble aggregates in solution. SEC could also reduce the heterogeneity of the sample, as well as the potential induction of precipitation by separation of aggregates from natively folded M062. (-)M062-Mav WT was applied to a superdex 200 16/60 and run as described in Section 2.2.4.6.

Gel filtration of the IMAC eluate of the thrombin-treated M062-Mav WT (see Fig. 19) indicates that the aggregates seen by SDS-PAGE (see Fig. 18) were not present in solution, and may be a product of sample preparation for SDS-PAGE. Molecules larger than the pores of the SEC beads are excluded from the gel matrix, and will elute at the void volume (V_O), ~45 mL for superdex 200 16/60. No peak is present at V_O.

Significant separation of cleaved and un-cleaved protein was achieved, with His-M062-Mav WT eluting at ~80 mL, and (-)M062-Mav WT at ~93 mL. Baseline separation wasn't achieved, but subsequent IEX of (-)M062-Mav WT fractions yielded pure (-)M062-Mav WT (see Fig. 20).

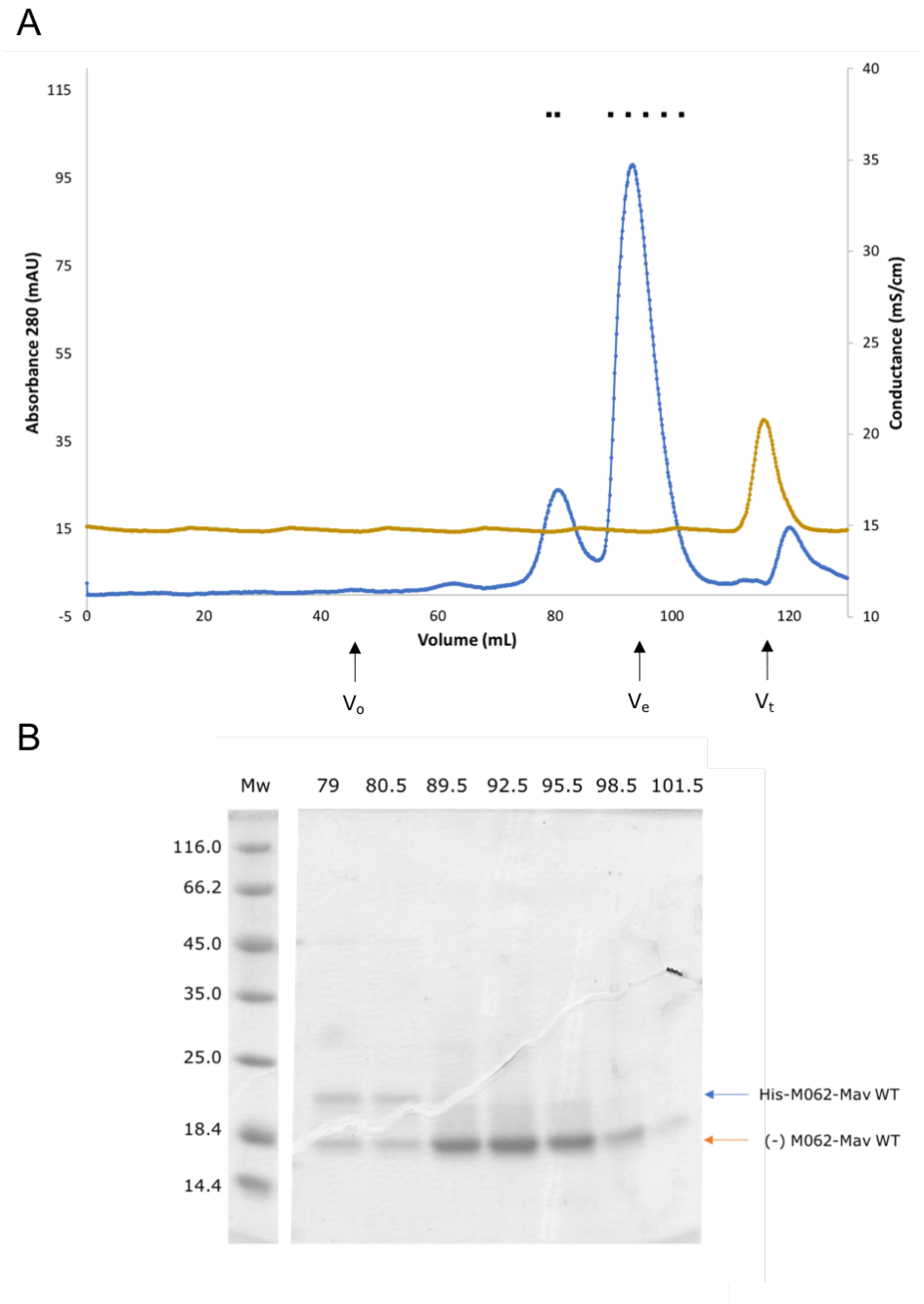


Figure 19: SEC of (-)M062-Mav WT. (A) Size-exclusion chromatograph of (-)M062-Mav WT on a superdex 200 16/60 column: A280 (blue) and conductance (gold) vs volume (mL). Void volume (VO) = 45 mL, (-)M062-Mav WT elutes with an elution volume (Ve) = 93 mL, column volume (Vt) = 120 mL. Fractions analysed by SDS-PAGE indicated by (■). (B) 15% SDS-PAGE of SEC fractions, fraction volume (mL) indicated above each lane.

3.3.3. IEX of (-)M062-Mav WT

Separation of His-M062-Mav WT from (-)M062-Mav WT was achieved by anion-exchange (see Fig. 20). (-)M062-Mav WT exhibits a broad elution profile by IEX eluting between 270 – 370 mM NaCl (see Fig. 20 (A)), indicating a non-uniform affinity between the proteins in solution and the resin. Isolation of pure (-)M062-Mav WT was achieved by IEX however (see Fig. 20 (B)), and these fractions were pooled and concentrated for crystallisation trials.

3.3.4. Crystallisation screening of (-)M062-Mav WT

Purified (-)M062-Mav WT (~98% pure, disregarding aggregates) was concentrated to 4.5 mg/mL (see Fig. 20 (C)) for crystallisation screening. Sparse matrix screening as described in section 2.2.6.2 yielded no crystals, or promising hits. The lack of favourable hits indicated alternate strategies for crystallisation were required. In parallel to modifying the purification strategy of the WT protein, construct engineering to enhance crystallisation was carried out.

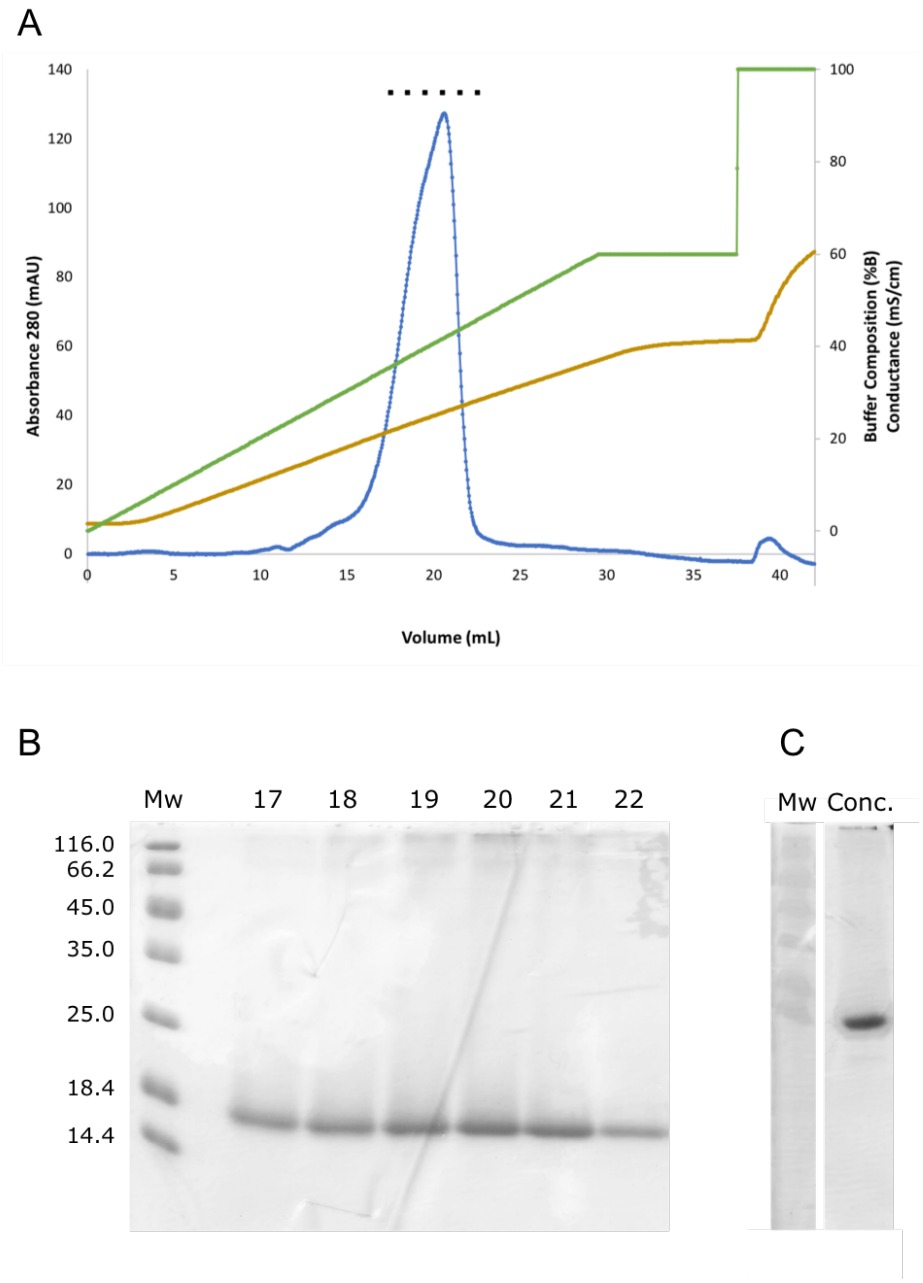


Figure 20: IEX of (-)M062-Mav WT. (A) Ion-exchange chromatograph of (-)M062-Mav WT on a monoQ 5/50 column. A_{280} (blue), buffer composition (green) and conductance (gold) vs volume (mL). Fractions analysed by SDS-PAGE annotated (■). (B) 15% SDS-PAGE of IEX fractions, fraction volume (mL) indicated above each lane. (C) 15% SDS-PAGE of pooled IEX fractions of (-)M062-Mav WT concentrated to 4.5 mg/mL for crystallisation (Conc.).

3.4. M062 construct engineering for enhancement of crystallisation

3.4.1. Multiple sequence alignment of the C7-family

M062 contains four Cys residues (see Fig. 21). Mutation of Cys to Ser residues is a strategy employed in engineering proteins to enhance their probability of crystallisation, by stabilizing the protein – removing a sulfhydryl group which can otherwise be susceptible to oxidation[122–126]. However in order to minimise the impact of mutagenesis of residues which may be important in C7 function, a multiple-sequence alignment of members of the C7 family was carried out (see Fig. 21). A selection of 25 C7 family proteins were chosen from BLASTp[127] alignment of M062 against non-redundant sequences and aligned against M062 using Clustal Omega[25,128].

The MSA indicates that the C7-family possess regions of highly conserved residues, particularly in the N terminus. Of the four Cys' of M062, only C31 is conserved among the C7 family, while C90, C107 and C143 are not conserved (see Fig. 21). A triple Cys → Ser mutant of M062-Mav was ordered (see Section 2.2.2), with M062-Lau 3CS generated by mutagenesis of T79 to Ile (see Section 2.2.2.3).

In addition to the MSA, homology modelling of M062 using the crystal structure of C7 as a template also indicated that C31 is likely buried, if occupying the same buried position as seen in the C7 and M064, whereas C90 and C107 are likely to be exposed to solvent. C143 is not shown in the model as it lies C terminally to the homology model (see Fig. 22). Mutagenesis of solvent exposed residues has been demonstrated to increase the crystallisability of proteins recalcitrant to crystallisation[122,125].

The impact of these cysteine mutants on M062's ability to interact with SAMD9 was also investigated by the Liu lab, with the 3CS mutants of Lau and Mav variants immunoprecipitating SAMD9 in an equivalent manner to the WTs (see Fig. 16).

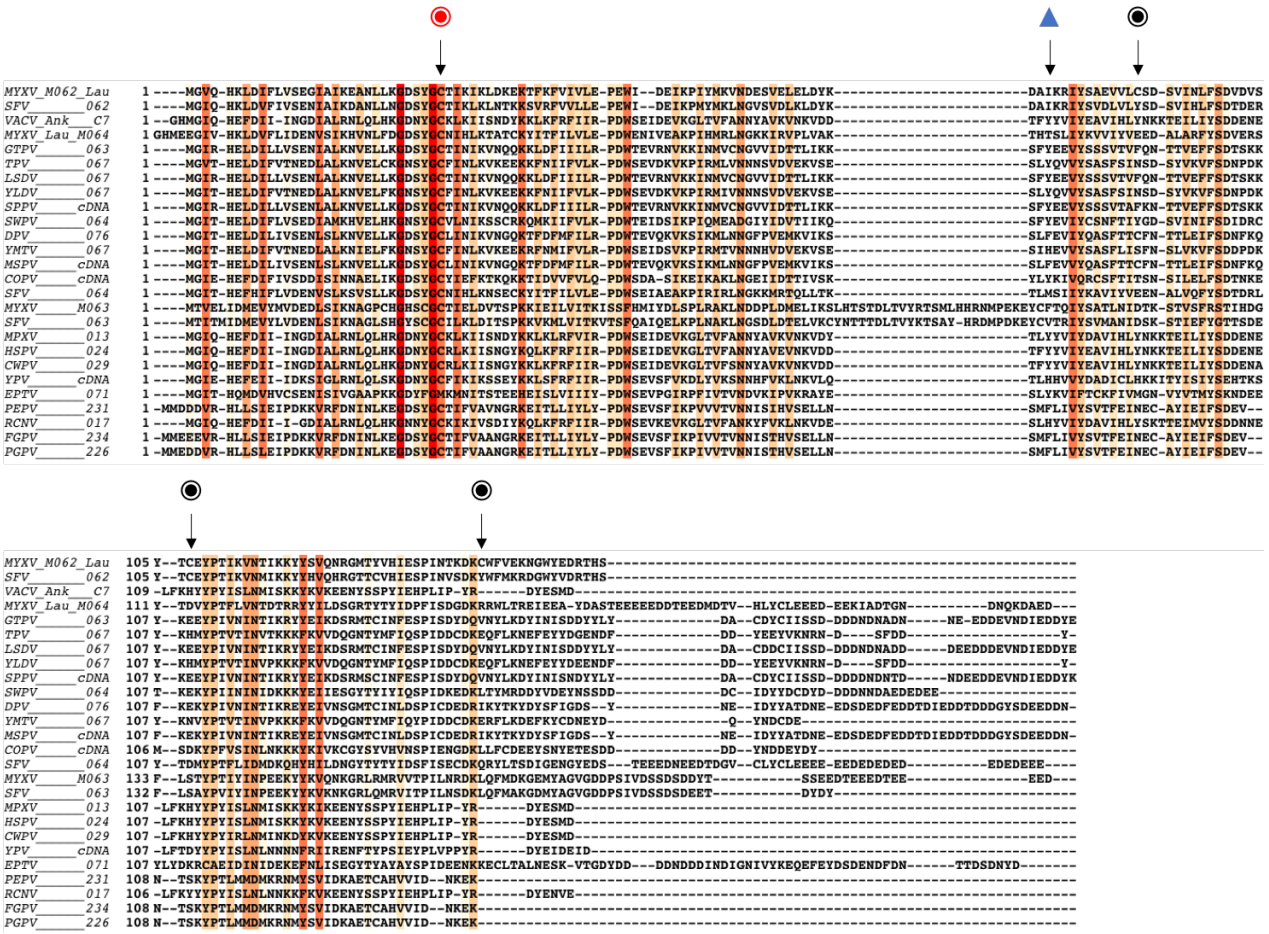


Figure 21: Multiple sequence alignment of the C7 family. Virus and protein identifier left of each sequence in full in Table III. Residues conserved at >30% are coloured in a gradient, from 100% (red) to 30% (yellow). The conservation of M062's Cys residues' are annotated: (◉) not conserved, (◉) conserved in > 30% aligned C7 proteins. The Lau/Mav variant locus 79 is annotated by (▲).

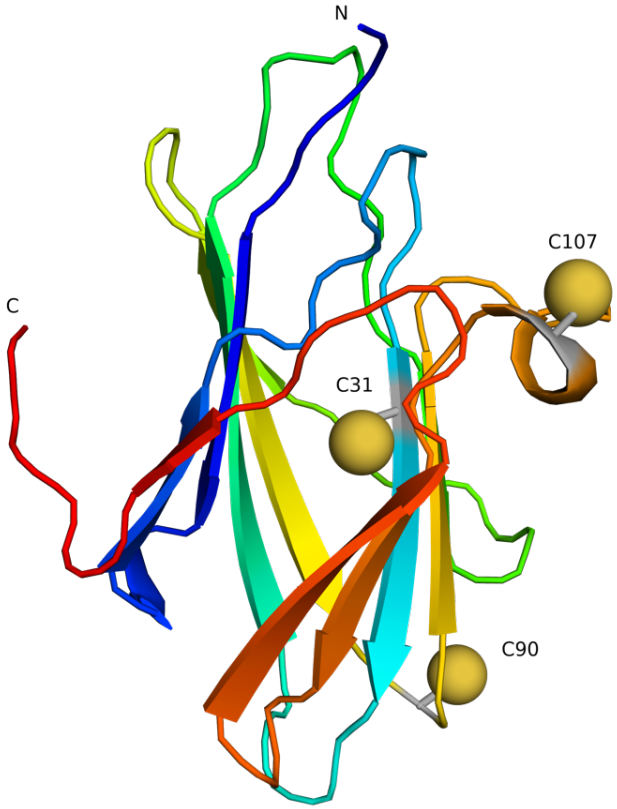


Figure 22: Structural prediction of the location of M062 Cys residues, based on homology modelling of M062 using C7 as a template. A structural model of M062 was computationally generated using SWISS-MODEL[129] by using the crystal structure of C7 [5CYW] as a template. A ribbon model of the M062 generated structure is shown in rainbow colour (N terminus (blue), C terminus (red), annotated), with the Cys residues shown as yellow spheres, and their residue number annotated. C143 is omitted from the model, as it is located downstream of the C terminus of the homology model generate by SWISS-MODEL. C31 is buried in the β sandwich, while C90 and C107 are solvent exposed.

3.5. Large-scale expression and purification of His-M062-Mav 3CS

3.5.1. IMAC and SEC of M062-Mav 3CS

His-M062-Mav 3CS was purified in a similar fashion to His-M062-Mav WT (see Section 2.2.4). In order to minimise concentration-dependent precipitation, the eluate was diluted with an equal volume of Extraction Buffer as it eluted from the Ni-NTA column. Higher proteolytic efficiency (see Fig. 18 and 23 (B)) coupled with a decrease in precipitation allowed for a higher yield of (-)M062-Mav 3CS preparations in comparison to the WT preparations. A single band (90% pure) post 2nd IMAC 'FT2' (see Fig. 23 (B)) is obtained, however when applied to SEC superdex 75 16/60 displays a non-Gaussian peak (see Fig. 23 (A)). While a single band is seen by SDS-PAGE (see Fig. 23 (C)) for the eluted peak by SEC, it was thought that distinct populations of heterogeneously folded (-)M062-Mav 3CS may be present in solution, therefore IEX was carried out on the pooled peak.

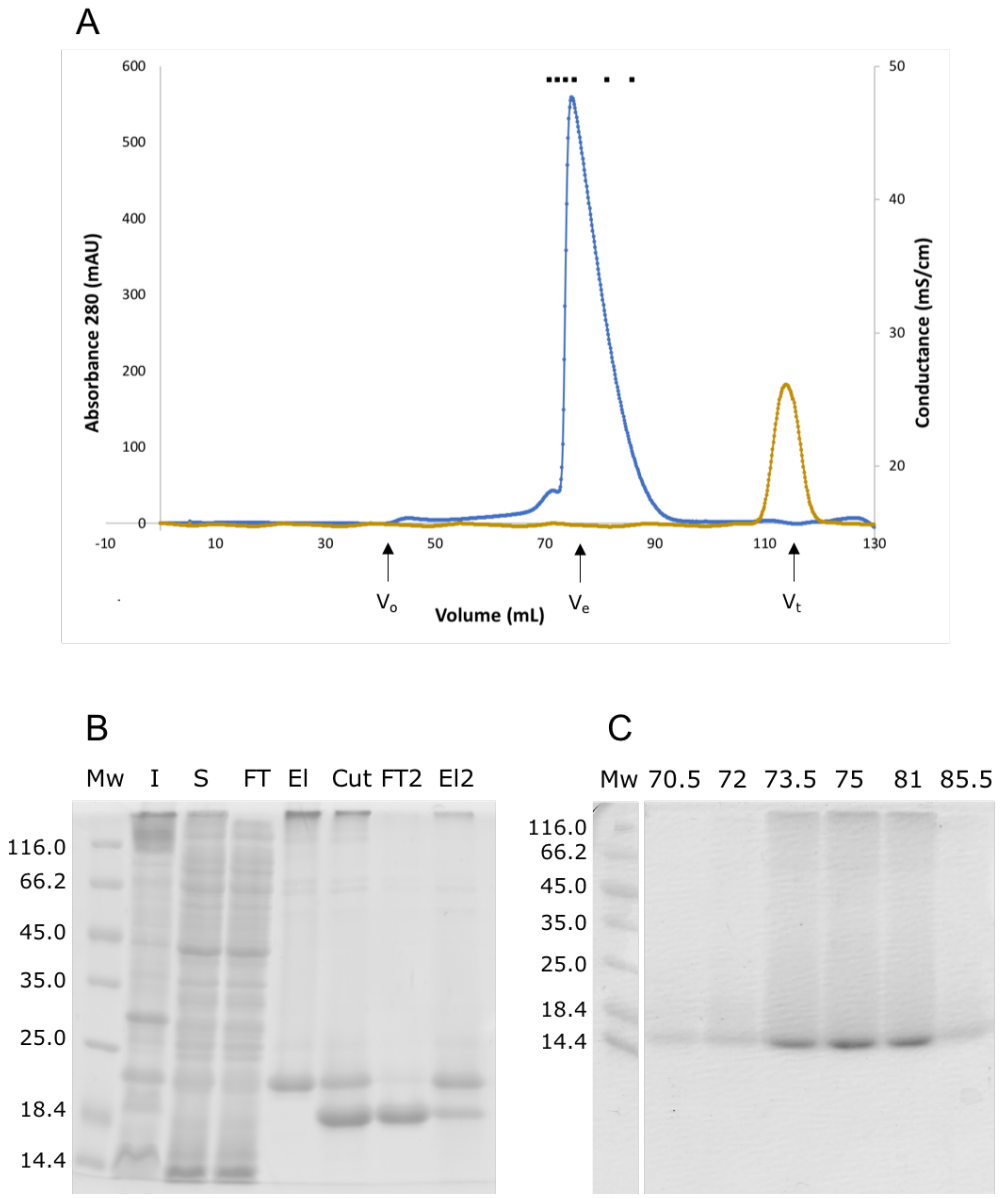


Figure 23: Purification by IMAC and SEC of M062-Mav 3CS. (A) Size-exclusion chromatograph of (-)M062-Mav 3CS on a superdex 75 16/60 column: A_{280} (blue) and conductance (gold) vs volume (mL). Void volume (V_O) = 45 mL, (-)M062-Mav 3CS elutes with an elution volume (V_e) = 75 mL, column volume (V_t) = 120 mL. Fractions analysed by SDS-PAGE indicated by (■). (B) 12% SDS-PAGE of IMAC His-M062-Mav 3CS purification by IMAC followed by proteolysis by thrombin. (C) 15% SDS-PAGE of SEC fractions from run (A), fraction volume (mL) indicated above each lane.

3.5.2. IEX of (-)M062-Mav 3CS

As per (-)M062-Mav WT, a 60% gradient on a monoQ anion exchange column was applied to (-)M062-Mav 3CS (see Fig. 20). The elution profile of (-)M062-Mav 3CS (see Fig. 24 (A)) is very similar to that of (-)M062-Mav WT (see Fig. 20 (A)), with elution of protein over a wide ionic range, from 220 – 370 mM NaCl, with the 3CS mutant eluting earlier than the WT (3CS: 320 mM NaCl, WT: 345 mM NaCl). SDS-PAGE (see Fig. 24 (B)) show all fractions produce equivalent bands, with concentration dependent streaking of the protein through the gel (enhanced from 17.5 mL – 20 mL), with high Mw aggregates at the stacking and separating gel interface as seen in the IMAC and SEC purification stages.

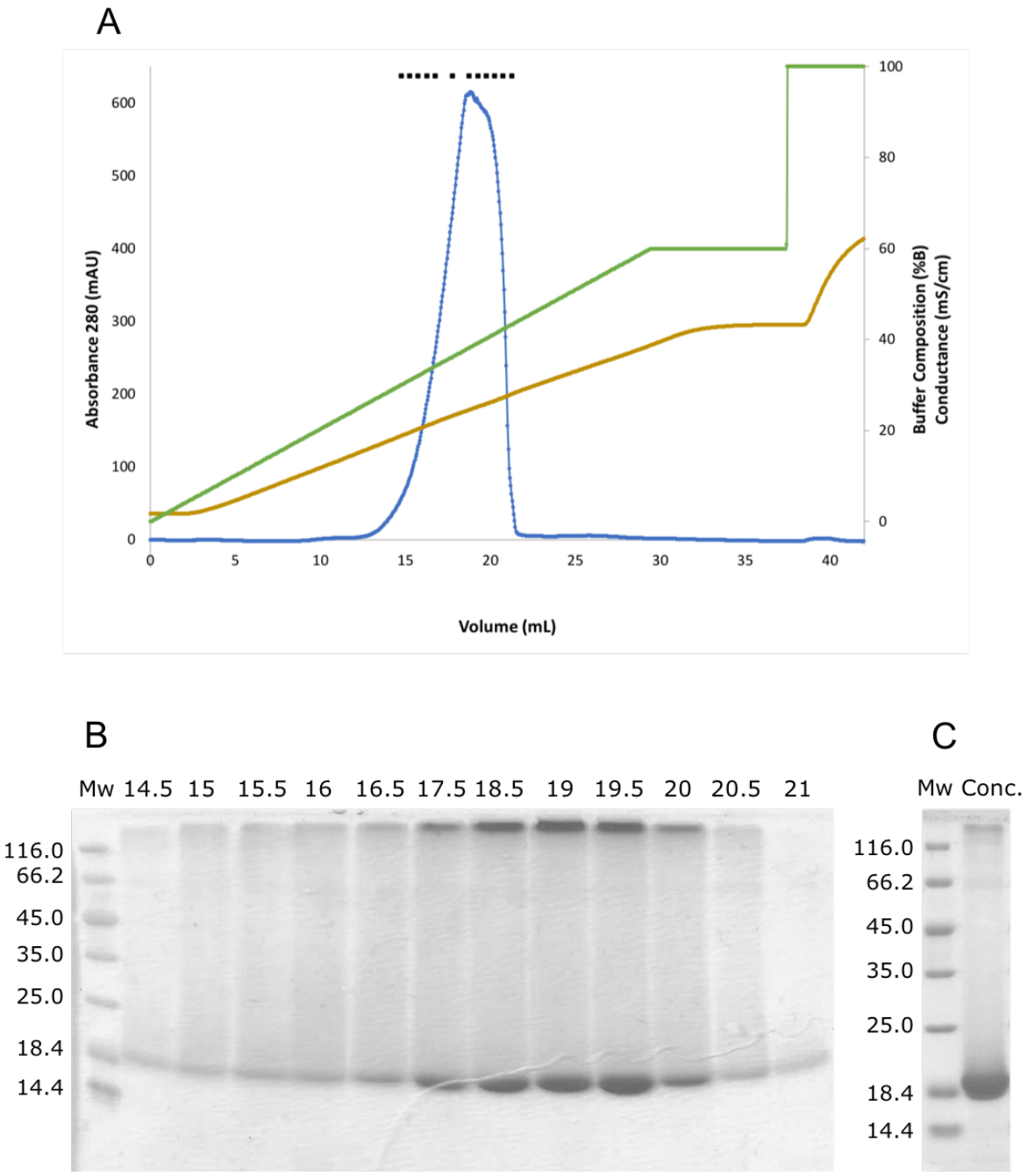


Figure 24: Ion-exchange purification of (-)M062-Mav 3CS. (A) Ion-exchange chromatograph of (-)M062-Mav 3CS on a monoQ 5/50 column: A280 (blue), buffer composition (green) and conductance (gold) vs volume (mL). Fractions analysed by SDS-PAGE annotated (■). (B) 12% SDS-PAGE of IEX fractions, fraction volume (mL) indicated above each lane. (C) 15% SDS-PAGE of pooled IEX fractions of (-)M062-Mav WT concentrated to 6.6 mg/mL for crystallisation (Conc.).

3.6. **Crystallisation of (-)M062-Mav 3CS**

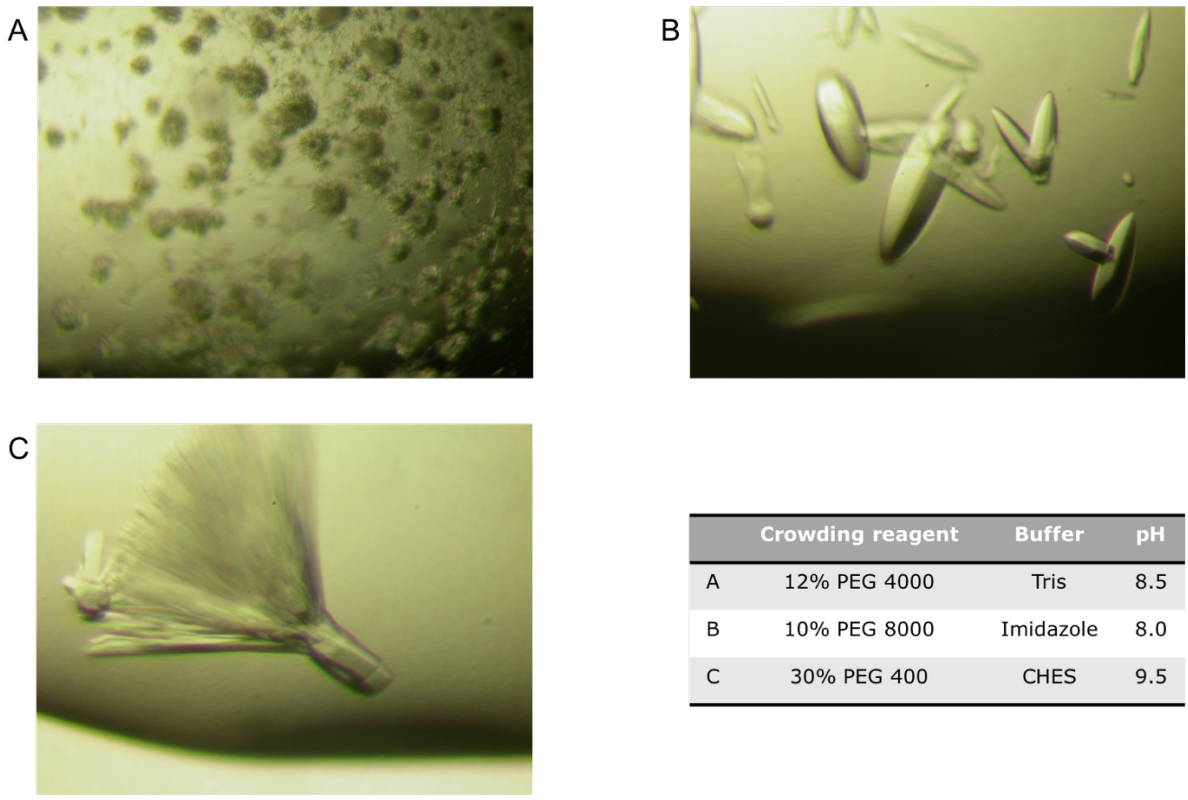
3.6.1. Sparse matrix screening of (-)M062-Mav 3CS

Purified (-)M062-Mav 3CS (~93% pure, including aggregates, streaking above the band increases background and lowers the accuracy of quantification of the band) was concentrated to 6.6 mg/mL (see Fig. 24 (C)) for crystallisation screening. Screening was carried out at a variety of final concentrations from IEX purified fractions, from 3 – 12 mg/mL. Sparse matrix screening was initially carried out, as described in section 2.2.6.2, and yielded some promising initial hits (see Fig. 25).

Crystallisation conditions were similar, comprising a high pH (8.0 – 9.5) and PEG as a crowding agent, with different buffering reagent and Mw PEGs enabling crystallisation. Individual crystals (Fig. 25 (B) (C)) grew between 6-9 days, while the clusters of microneedles (Fig. 25 (A)) were seen at 1 month. Hits varied in their morphology, with conditions producing a shower of clustered micro-needles (Fig. 25 (A)), cuboid crystals with multiple secondary nucleation sites from which long, thin needle-like crystals grew (Fig. 25 (C)), and larger, single almond-shaped crystals with smooth facets (Fig. 25 (B)).

3.6.2. Diffraction data collection from initial crystal hits of (-)M062-Mav 3CS

Crystals were harvested as per Section 2.2.6.10, and diffraction images collected at the synchrotron Soleil in Paris (see Fig. 26 (D) (E)). Unfortunately, insufficient cryoprotection and poor cryopreservation of the crystals led to the formation of ice crystals on/in the cryoprotectant (see Fig. 26 (A)). This led to powder diffraction rings (see Fig. 26 (D) (E)). Regardless of the presence of ice however, all crystals harvested from the initial hits from sparse matrix screening in 96-well trays did not diffract to resolutions greater than 5 Å. Figure 26 (D) (E) show strong reflections out to ~7 Å, with a subsequent rapid drop in signal intensity. The spot profiles are non-uniform, with Figure 26 (B) spots streaking, and (C) spots overexposed at low resolution. These factors indicated that there is disorder in the crystal lattice in the crystals obtained from initial screens. In order to obtain higher resolution data, refined screening was carried out.



	Crowding reagent	Buffer	pH
A	12% PEG 4000	Tris	8.5
B	10% PEG 8000	Imidazole	8.0
C	30% PEG 400	CHES	9.5

Figure 25: Initial crystal hits from sparse matrix screening of (-)M062-Mav 3CS. Pictured are crystal hits obtained from (-)M062-Mav 3CS by sparse matrix screening using the sitting-drop vapour-diffusion method. (A) shows a shower of microcrystals, with a 'sea-urchin' looking clustering of micro-needles. (B) produced almond-shaped individual crystals, with 4 facets. (C) shows 2° nucleation sites upon a cuboid crystal, with many long-thin needles projecting from a single base. Crystallisation conditions shown in table.

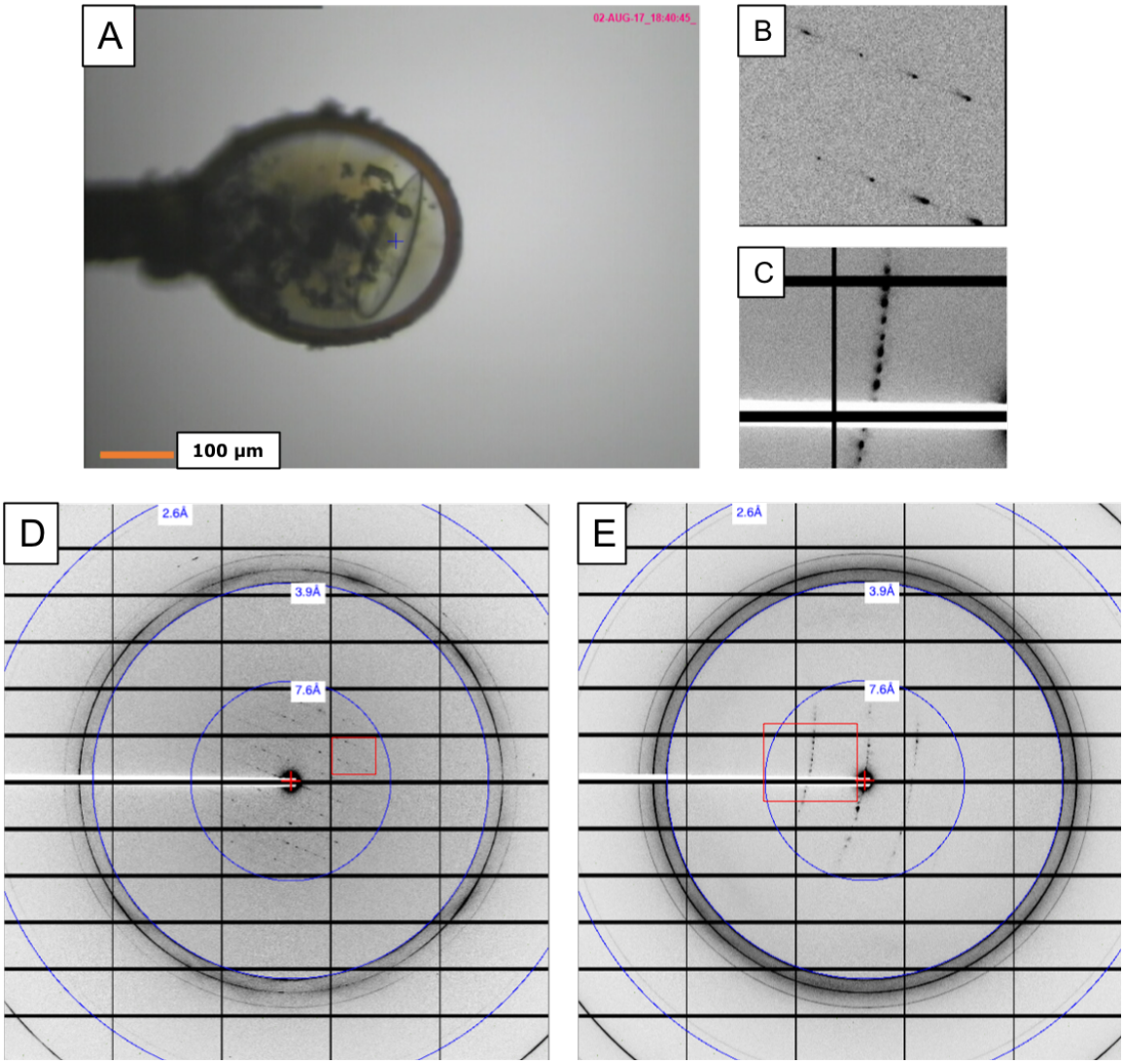


Figure 26: Synchrotron x-ray diffraction data collection images from initial crystal hits of (-)M062-Mav 3CS. (A) A single almond morphology crystal harvested from Fig. 25 (B) mounted in a 0.1 mm loop, scale bar (orange). Crystal dimensions ~200 x 50 x 25 µm. Significant ice crystals are present on the surface of the frozen mother liqor. (B) and (C) depict enhanced reflections from (D) and (E) respectively (outlined in red). (D) and (E) show 2 diffraction images collected at 0° and 90°. Resolution rings (blue) are shown at 7.6 Å, 3.9 Å, and 2.6 Å. Powder diffraction rings are present due to ice crystals in the mother liquor (black circles ~3.9 Å, 3.7 Å, 3.4 Å).

3.6.3. Refined crystallisation screening of (-)M062-Mav 3CS

Optimisation of crystal growth of (-)M062-Mav 3CS required thorough screening of a range of pH buffers, PEG Mw and concentration variation, protein concentration, and IEX fraction pooling. An example of a simple PEG, buffer screen is depicted in Figure 27.

Refined screening in 24-wells using the hanging-drop method was conducted at both 16°C and 4°C, however as (-)M062-Mav 3CS' solubility was significantly reduced at 4°C, screening trials were focussed at 16°C to allow for a sufficient concentration of protein in the drop for crystal growth. Crystallisation was largely successful, with crystals growing in a variety of buffers identified by sparse matrix screening, including Tris, Imidazole, and CHES pH 7.5 - 9.5, coupled with PEGs of Mw 400, 3,350, 4,000 and 8,000. Combinations of these buffers at varying pH values found that crystallisation was most likely between pH 7.5 – 8.0, including with CHES, while 8.0 is below its recommended buffer range 8.6 - 10.6 at 16 °C[130].

3.6.4. Alternate crystallisation techniques applied in (-)M062-Mav 3CS crystal screening

Obtaining single crystals without 2° crystals nucleating off the body of the main crystal required iterative crystallisation trials, changing the purification (such as pooling different IEX fractions) and crystallisation condition variables. Single crystals, when obtained were frequently of poor diffraction quality. In an attempt to sample more of the solubility space of (-)M062-Mav 3CS alternate crystallization techniques were employed. "Protein Crystallization: techniques, strategies, and tips" by Terese M. Bergfors[131] was utilized as a resource for alternate crystallization strategies.

	1	2	3	4	5	6
A	11.5% →		PEG 8,000 / Imidazole pH 8.0			14%
B	14.5% →					20%
C	11.5% →		PEG 3,350 / CHES pH 8.0			14%
D	14.5% ■ →		■			20%

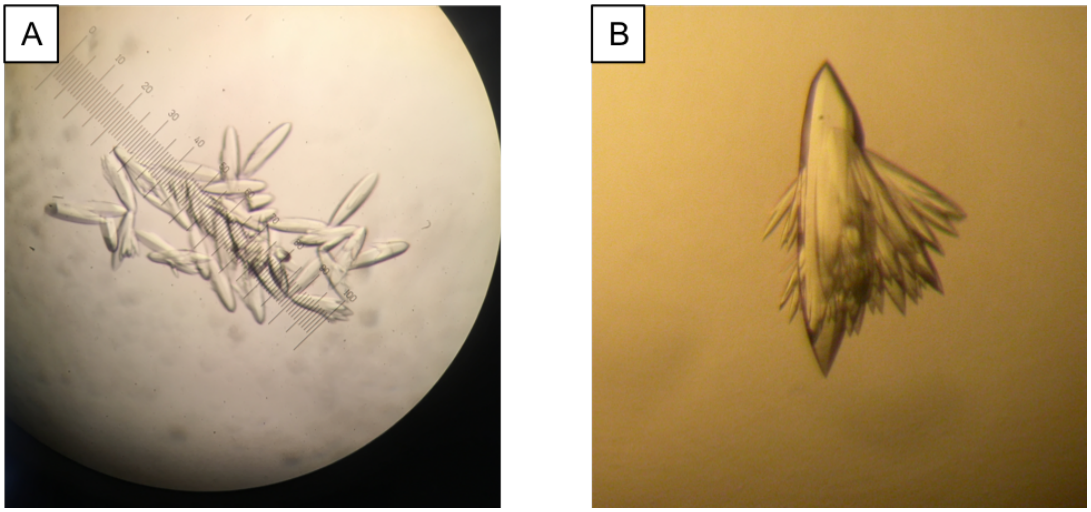


Figure 27: Refined (-)M062-Mav 3CS crystallisation screening of buffer reagent vs PEG concentration. The table depicts the sampling of an increasing (left to right) gradient of PEG concentration, against different pH buffering reagent. Crystals pictured in (A) and (B) indicated by black squares. (A), (B), images of crystals of (-)M062-Mav 3CS grown in 0.1 M CHES pH 8.0, 14.5%, 15.5% PEG 3,350. Experimental set-up: (-)M062-Mav 3CS at 10.54 mg/mL, 1:1 ratio protein:reservoir, 16°C, crystals appeared on day 5.

3.6.5.1. Reverse vapour diffusion crystallisation of (-)M062-Mav 3CS

The solubility of (-)M062-Mav 3CS at different salt concentrations was inferred by the concentration at which precipitation occurred in buffers varying only in ionic strength. It was observed that precipitation of (-)M062-Mav 3CS initiated at lower protein concentration at lower salt concentration. A corollary increase in solubility at higher NaCl concentration (200-500 mM) than low (5-150 mM) was found. Attempting to exploit this phase-change behaviour, reverse vapour diffusion was tested as a means of crystallisation. A concentrated solution of (-)M0623-Mav 3CS (6 – 18 mg/mL) at high NaCl concentration (220 – 370 mM) was set up as per Section 2.2.6.7.

Vapour diffusion proceeds with water vapour moving from a region of low osmolarity to high osmolarity, leading to the growth in size of the hanging drop, as water moves from the reservoir to the drop. This increase in volume decreases both protein and NaCl concentration. Figure 28 (A – D) shows crystallisation of (-)M062-Mav 3CS following the decrease in solubility following reverse vapour diffusion.

Crystals grew with the same almond morphology as in previous conditions, but varied in size, from microcrystals, clusters of microcrystals, to single crystals ~50 µm in length.

3.6.5.2. Additive screening of precipitants for (-)M062-3CS crystallisation

(-)M062-Mav 3CS' propensity for precipitation encouraged additive screening to investigate whether a concentrated solution of (-)M062-Mav 3CS could have its solubility lowered by mild perturbation of the buffer condition. The Hampton Research Additive Screen HR2-138 was consulted for recommended drop concentrations of common additives, and various alcohols, sugars and salts were investigated for their effect on (-)M062-Mav 3CS solubility. Figure 28 (E – H) shows crystals which were grown with mild precipitant concentrations (5% PEG 3,350, 1% EtOH).

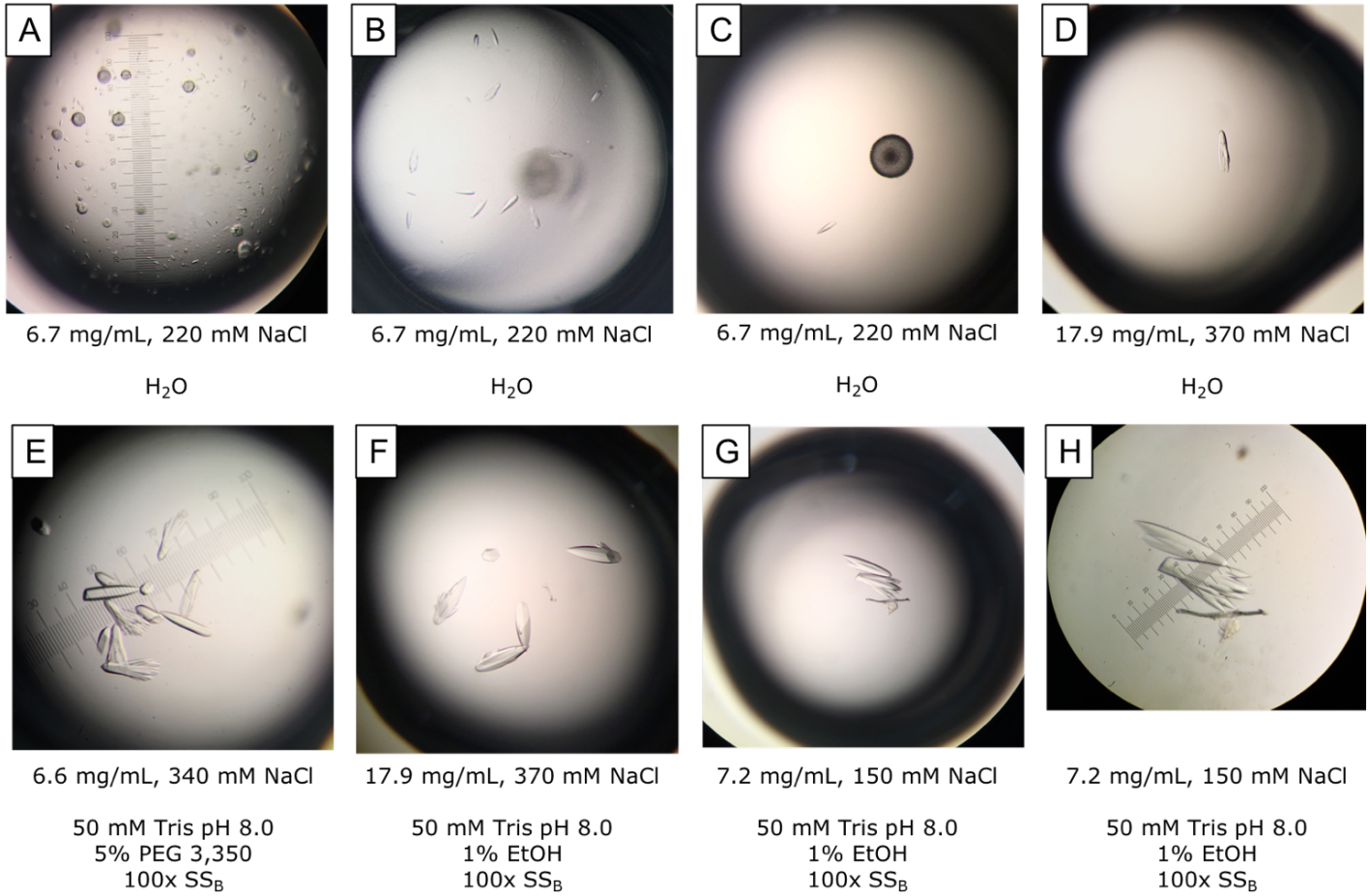
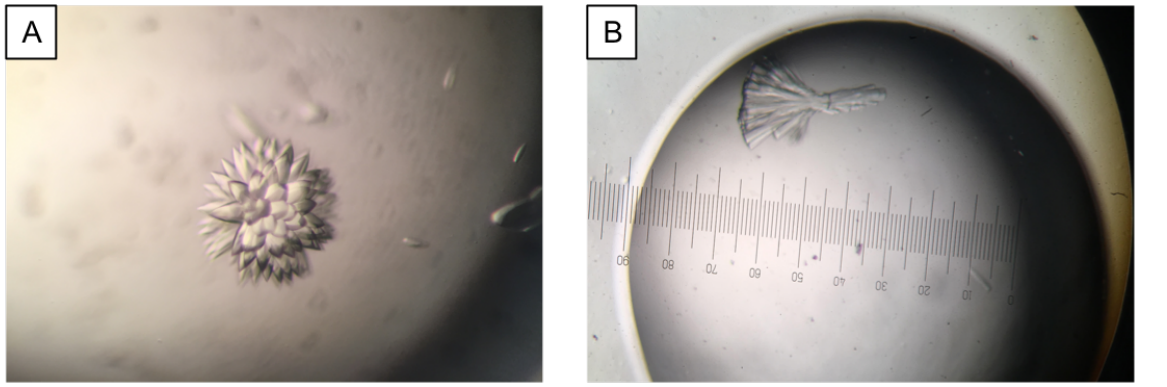


Figure 28: Crystals of (-)M062-Mav 3CS obtained utilising various crystallisation techniques. Images A – D show crystals of (-)M062-Mav 3CS obtained by reverse vapour diffusion. Crystals in E – H were obtained by seeding with seed stock B (SS_B). Crystals of SeMet-derivatised M062 shown in G, H. Protein (mg/mL) and NaCl (mM) concentrations and reservoir conditions are shown below each image.

3.6.5.3. Seeding crystallisation trials of (-)M062-Mav 3CS

Seeding is an established technique which can promote crystallisation by by-passing the sometimes elusive, narrow conditions required to induce nucleation of protein within the 'metastable zone'. Seed stocks were prepared as described in Section 2.2.6.5., from crystals shown in Fig. 28. The entire clusters shown in (A) and (B) were used in each seed stock preparation.

The seed stock prepared from (B) (SS_B), at 100x strength proved to reliably promote crystallisation of (-)M062-Mav 3CS (see Fig. 28 (E – H)), leading to fewer, larger crystals compared to drops set up in parallel, without the addition of the seed stock.



12% PEG 3350, 0.1 M CHES pH 8.0 16% PEG 3350, 0.1 M CHES pH 8.0, 0.1M A.S.

Figure 29: Crystal sources for the preparation of seed stocks. Images of crystals used to prepare seed stocks for (-)M062-Mav 3CS crystallisation are shown. (A) shows a cluster of crystals of almond morphology. (B) shows a cluster of needle-like crystals originating from a single, larger crystal. Crystallisation conditions are shown below each image.

3.6.5.4. Salting-out as a means of crystallisation of (-)M062-Mav 3CS

Salting-out was explored as a means to crystallise (-)M062 by lowering the ionic strength of the buffer, thereby decreasing protein solubility. While reverse vapour diffusion lowers salt concentration, sampling different regions of the protein solubility, changes in drop volume also simultaneously alter protein concentration. Salting-out, by dialysing out NaCl from a solution of (-)M062 as described in Section 2.2.6.6. was carried out.

Protein solubility was characterised by visual inspection of the (-)M062-Mav 3CS sample over the course of the dialysis, with no visual signs of any decrease in solubility (solution remained clear) between 0 – 48 hours, followed by significant precipitation (solution turned cloudy, white precipitate build up at bottom of dialysis tubing) in the subsequent 24 hours.

3.6.5.5. Temperature-dependent crystallisation trial of (-)M062-Mav 3CS

It was observed that a concentrated solution of pure (-)M062-Mav 3CS (≥ 3 mg/mL, ≤ 150 mM NaCl) would readily precipitate at 4°C, while remaining soluble at RT, and displaying reversible solubility at concentrations of NaCl ≥150 mM. Reducing the temperature of a 3.4 mg/mL sample at 150 mM NaCl as described in Section 2.2.6.8. was carried out to investigate if crystallization could be induced by slowly reducing the solubility of (-)M062-Mav 3CS. The sample, however remained soluble with no signs of crystal growth.

3.6.6. Diffraction data collection from optimised crystals of (-)M062-Mav 3CS

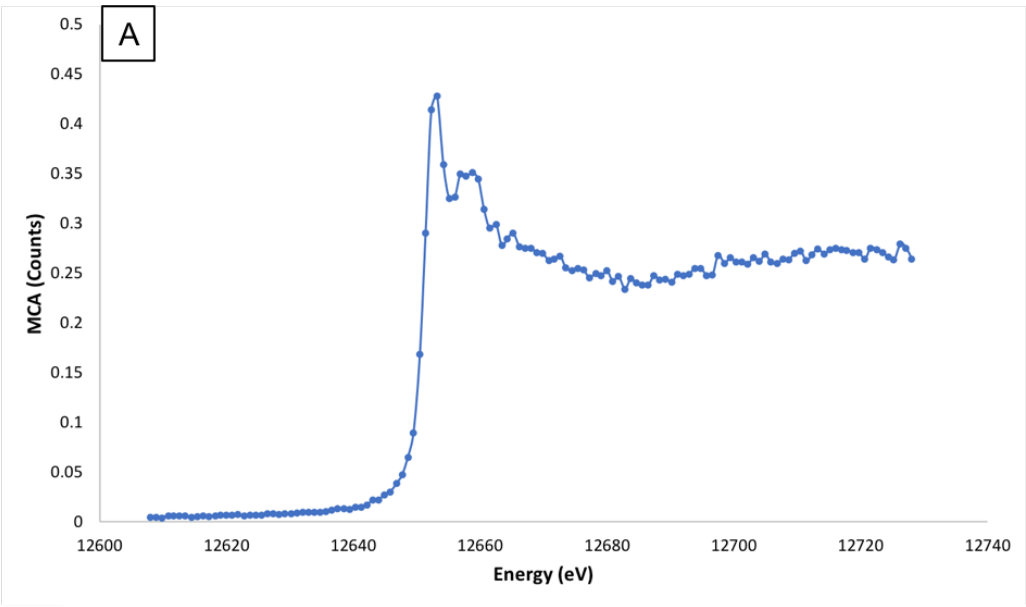
The crystals obtained through the refined screening and application of a variety of crystallization techniques to (-)M062-Mav 3CS were harvested and shipped to the synchrotron (Soleil and APS) for diffraction data collection. Over 100 crystals were tested, with initial diffraction images showing the majority of crystals did not diffract to high resolution >4 Å. A small subset of crystals with high resolution limits between 4 – 3.5 Å were obtained, however nearly all of these were twinned (data not shown).

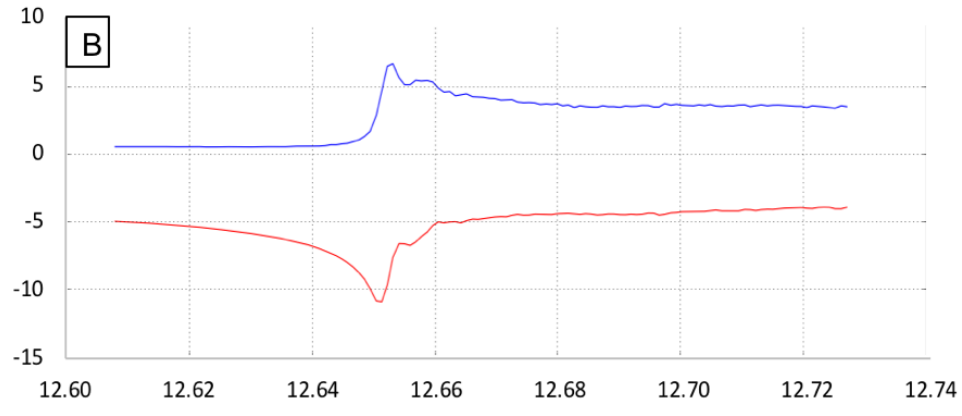
Ultimately, a 3.2 Å dataset was obtained from a crystal shown in Fig. 27 (A), grown in 0.1 M CHES pH 8.0, 12.5% PEG 3,350 at 16°C, and harvested using 0.1 M CHES pH 8.0, 14% PEG 3,350, 0.1 M ammonium sulfate, 20% v/v glycerol as cryoprotectant. Diffraction data is discussed in Chapter 4.

3.6.7. Expression, purification and crystallisation of a selenomethionine derivative of M062-Mav 3CS

Molecular replacement using either C7 or M064 as a model was attempted using the (-)M062-Mav native datasets (4.0 – 3.5 Å). However, it was not possible to produce electron density maps of sufficient quality to refine the structure of M062. In attempt to solve the phases by anomalous diffraction experimental methods, selenomethionine (SeMet) derivatised (-)M062-Mav 3CS was produced similarly to native (-)M062-Mav 3CS (see Section 2.2.6.9).

Crystals of SeMet (-)M062-Mav 3CS are shown in Figure 28 (G, H), grown in 50 mM Tris pH 8.0, 1% EtOH, seeded with SS_B (100x). Crystals were harvested and shipped to the synchrotron Soleil for x-ray diffraction data collection. Incorporation of Se was confirmed by an energy absorption x-ray fluorescence scan (see Section 2.2.7.2), as shown in Fig. 30. The reported absorption maximum (K) for Se (according to the University of Washington, using the theoretical approximation developed by Cromer and Liberman[132]) is 12.6578 keV, with the experimental K value: 12.653 keV. Unfortunately, while significant anomalous diffraction was obtained, the SeMet data were not of sufficient quality to solve the crystal structure. Data were low resolution, and twinned.





	Energy (keV)	f'	f''
Peak (K)	12.6532	-7.68	6.69
Inflection point	12.6514	-10.97	4.66

Figure 30: X-ray fluorescence energy scan of a selenomethionine derivatised crystal of (-)M062-Mav 3CS. SeMet incorporation was verified by (A) an x-ray fluorescence scan, in which the incident x-ray energy was varied from 12600 eV – 12740 eV, while photon count was measured (MCA: multichannel analyser). (B) The first and second derivative of (A) are plotted: f' (red), f'' (blue). The table shows the peak (K) and inflection point energy (keV). (B) calculated by CHOOCH[133] on-site at Px2a in Soleil.

3.7. Expression, purification and crystallisation of M062-Lau 3CS

3.7.1. Initial crystallisation trials of (-)M062-Lau 3CS

The Lausanne variant of M062, differs from Mav M062 with Ile at position 79 vs Thr. While the variants have >99% sequence identity, rescreening for crystallisation conditions by sparse matrix screening was carried out, as it has been widely reported that point mutants can have unknown consequences on protein crystallisation, including altering solubility, crystal space group, and diffraction resolution limits[134].

His-M062-Lau 3CS was expressed and purified by IMAC in the same fashion as His-M062-Mav 3CS, yielding comparable levels of pure protein (see Fig. 31 (A)), ~93% purity in 'FT2'.

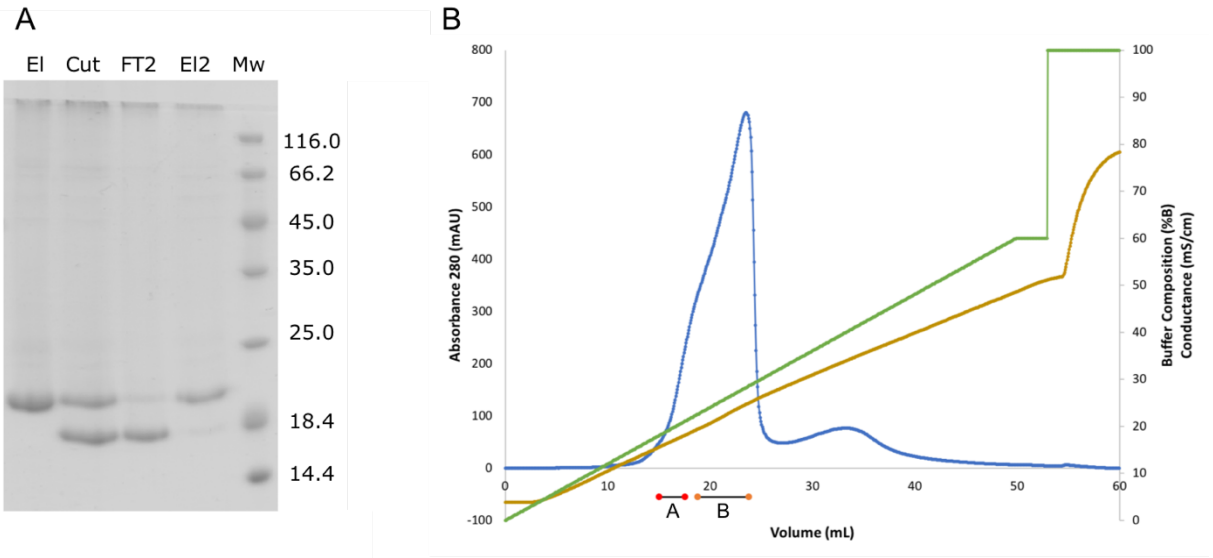


Figure 31: Purification of M062-Lau 3CS by IMAC and IEX. (A) 12% SDS-PAGE of IMAC His-M062-Lau 3CS purification by IMAC followed by proteolysis by thrombin. (B) Ion-exchange chromatograph of (-)M062-Lau 3CS on a sourceQ 15Q 4.6/100 PE column. A_{280} (blue), buffer composition (green) and conductance (gold) vs volume (mL). Pooled fractions indicated, connected by black line, A: 15 – 19 mL (red), B: 20 – 24 mL (orange).

Initial crystallisation trials of (-)M062-3CS Lau were conducted following purification using a similar strategy used for the crystallisation of (-)M062-3CS Mav. Briefly: His-M062-Lau 3CS was isolated by IMAC, followed by proteolytic

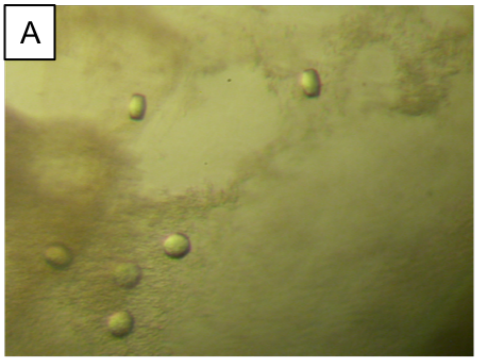
cleavage and removal of the His tag, and lastly IEX (see Fig. 31 (B)), with SEC omitted. Screening was carried out with IEX fraction A: 7.4 mg/mL, 210 mM NaCl and B: 6.3 mg/mL, 270 mM NaCl.

96-well sitting drop sparse matrix screening identified novel hit conditions which failed to produce hits with (-)M062-Mav 3CS (see Table 13). One condition produced crystals with morphology different to those seen with the Mav variant.

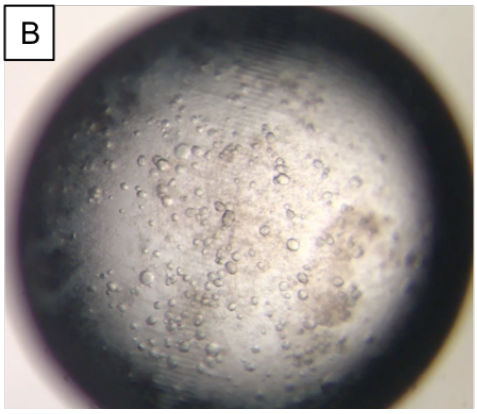
Condition	Hit description	IEX Fraction
0.2 M Li_2SO_4, 0.1 M Na.Acetate pH 4.5, 50% PEG 400	Spherulites	A, B
0.2 M KSCN, 0.1 M BTP pH 8.5, 20% PEG 3350	Round, seed-like crystals (Fig. 32 (A))	B
0.1 M CHES pH 9.5, 10% PEG 3000	Small needles.	A
0.2 M $MgCl_2$, 0.1 M Na.Cacodylate pH 6.5, 10% PEG 3000	Spherulites	A
0.1 M Imidazole pH 8.0, 10% PEG 8000	Small needles.	A
0.3 M $MgCl_2$, 0.3 M $CaCl_2$, 0.1 M HEPES/MOPS pH 7.5, 20% ethylene glycol, 10% PEG 8000	Shower of rugby ball shaped crystals	B

Table 13: Hit conditions from sparse matrix screening of (-)M062-Lau 3CS. Description of (-)M062-Lau 3CS hit morphology against crystallisation conditions are shown, with IEX pooled fractions included (as per Fig. 31).

Refined screening around these hit conditions produced some crystals of similar morphology (see Fig. 32). The rounded faces of these crystals, and high nucleation rate, leading to the growth of a high number of small crystals indicated altering the protein concentration and precipitant conditions may produce fewer, larger crystals. Therefore refined screening, with modification of the purification strategy was conducted.



	PEG 3350	Buffer	Additive	pH
A	20%	BTP	0.2 M KSCN	8.5
B	19%	BTP	0.2 M KSCN	8.5
C	18%	BTP	0.2 M KSCN	8.4



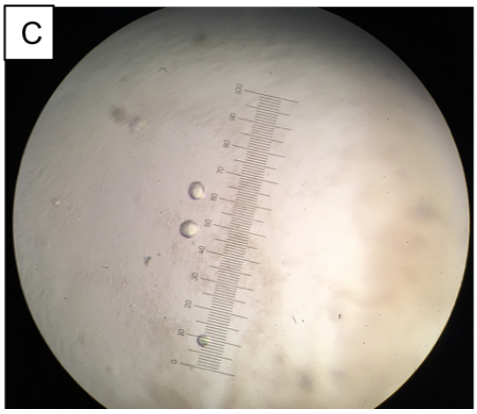


Figure 32: Initial hits from (-)M062-Lau 3CS crystallisation trials. (A) round, seed-like crystals obtained from sparse matrix screening. (B, C) reproduced crystals of similar morphology from refined screening. Crystallisation conditions shown in table.

3.7.2. Refined purification and crystallisation of (-)M062-Lau 3CS

3.7.2.1. Refined purification of (-)M062-Lau 3CS

One of the variables in the crystallisation of (-)M062-Lau 3CS investigated was if IEX was isolating different populations of (-)M062-Lau 3CS. It was common to obtain crystal hits from one of the pooled IEX fractions but only (eg. Fig. 31 (B), fraction B obtained the seed-like crystals shown in Fig. 32, while fraction A failed to), or in different crystallisation conditions. These pooled IEX fractions, while they could be concentrated to equivalent mg/mL, differ in their NaCl concentration. A revised strategy in which IEX fractions were subsequently gel filtrated was carried out, with the SEC step exchanging the buffer, eluting the fractions in identical SEC Buffer, and also allow for comparison in elution profiles.

Refinement of the concentration of NaCl required for elution of (-)M062-Lau 3CS was achieved by using a step-gradient in place of a linear concentration gradient, ultimately allowing for more precise elution of (-)M062-Lau 3CS (see Fig. 33).

Prior to SEC, the IEX fractions were pooled, with early eluting protein fractions separated from the later eluted protein. Subsequent SEC of these pooled IEX fractions shows that although these protein populations elute at different NaCl concentrations by IEX, they have effectively identical SEC elution profiles and retention, V_e A: 82.0 mL, V_e B: 82.4 mL (see Fig. 34 (A)), and are of equal purity by SDS-PAGE A fractions: (Fig. 34 (C)), B fractions: (Fig. 34 (D)). The SEC fractions, after being pooled and concentrated (Fig. 34 (B)) now vary only by the isolation of different species by IEX.

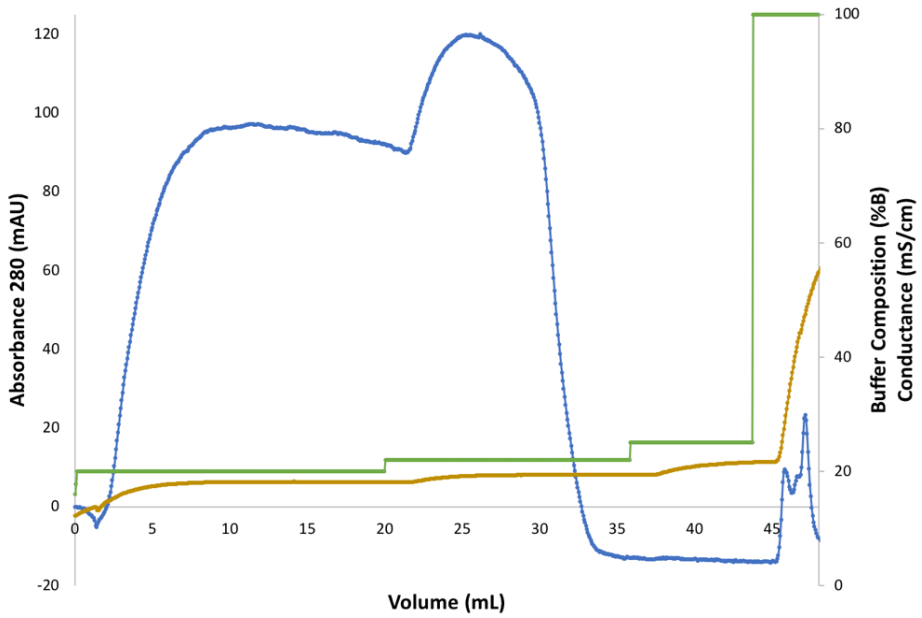


Figure 33: Refined IEX of (-)M062-Lau 3CS. Ion-exchange chromatograph of (-)M062-Lau 3CS on a sourceQ 15Q 4.6/100 PE column, with elution by step-gradient buffer composition (green) change. A_{280} (blue), (green) and conductance (gold) vs volume (mL).

3.7.2.2. Crystallisation of (-)M062-Lau 3CS

SEC purified IEX fractions were concentrated and refined screening of crystallisation conditions identified by sparse matrix screening carried out. As these samples were now in 150 mM NaCl (SEC Buffer), the overall solubility of the protein is lower than in IEX elutes. Crystallisation trials were therefore carried out at lower concentration, with optimised crystals obtained at 2.7 mg/mL final concentration (see Fig. 35). There appeared to be a relatively higher chance of crystallisation of (-)M062-Lau 3CS from later IEX elution fractions: 'B'.

Crystals were obtained using 0.1 M Tris pH 7.5 or 7.9, 50 mM $MgCl_2$, 50 mM $CaCl_2$, and 12 – 16% ethylene glycol, 6 - 8% PEG 8000 (Fig. 35 (A) – (D)) or 0.1 M Tris 7.9, 0.15 M KSCN, 14 – 16% PEG 400, +/- 1% EtOH or iProH (Fig. 35 (E) – (H)). Crystals were harvested, transferred to a variety of cryoprotective reagents, and snap frozen in LN2 and shipped to the synchrotron for diffraction data collection.

3.7.3. Diffraction data collection from optimised crystals of (-)M062-Lau 3CS

Crystals of (-)M062-Lau 3CS were harvested and shipped to the synchrotron (Soleil and APS) for diffraction data collection. Of the crystals tested, most diffracted to between 3 - 4 Å. Ultimately, a 2.3 Å dataset was obtained from a crystal shown in Fig. 35 (F), grown in 0.15 M KSCN, 0.1 M Tris pH 7.9, 14% PEG 400, 1% iPrOH, at 16°C, and transferred into reservoir supplemented to increasing glycerol concentration: 15%, 20%, 25% glycerol. Diffraction data are discussed in Chapter 4.

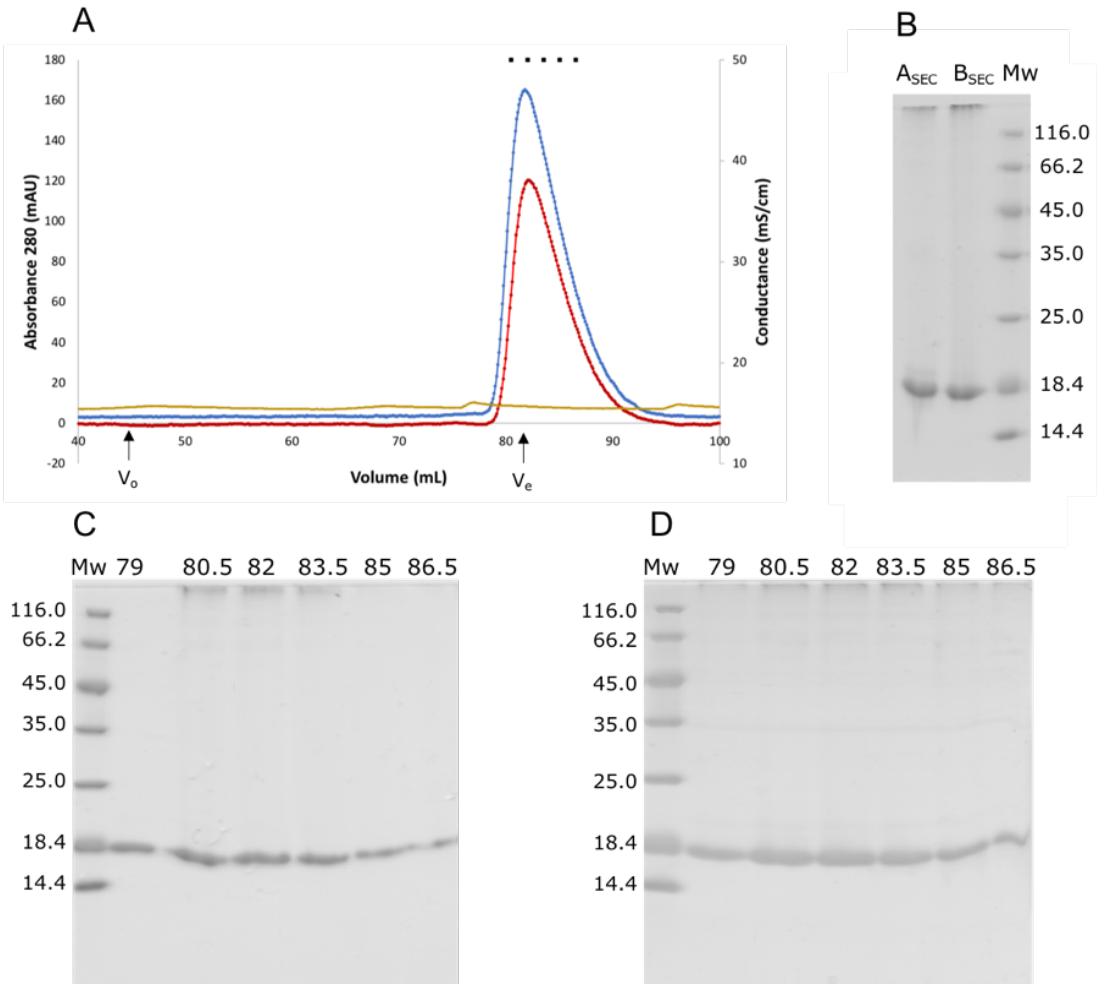


Figure 34: SEC of (-)M062-Lau 3CS.

(A) Size-exclusion chromatograph of (-)M062-Lau 3CS on a superdex 75 16/60 column: A_{280} fraction A: (blue), fraction B (red), and conductance (gold) vs volume (mL). Void volume (V_O) = 45 mL, (-)M062-Lau 3CS elutes with an elution volume A: (V_e) = 82.0 mL, B: (V_e) = 82.4 mL. Fractions analysed by SDS-PAGE indicated by (■). (B) 12% SDS-PAGE of pooled, concentration SEC fractions. (C) (D) 12% SDS-PAGE of SEC fractions from run (A), fraction volume (mL) indicated above each lane.

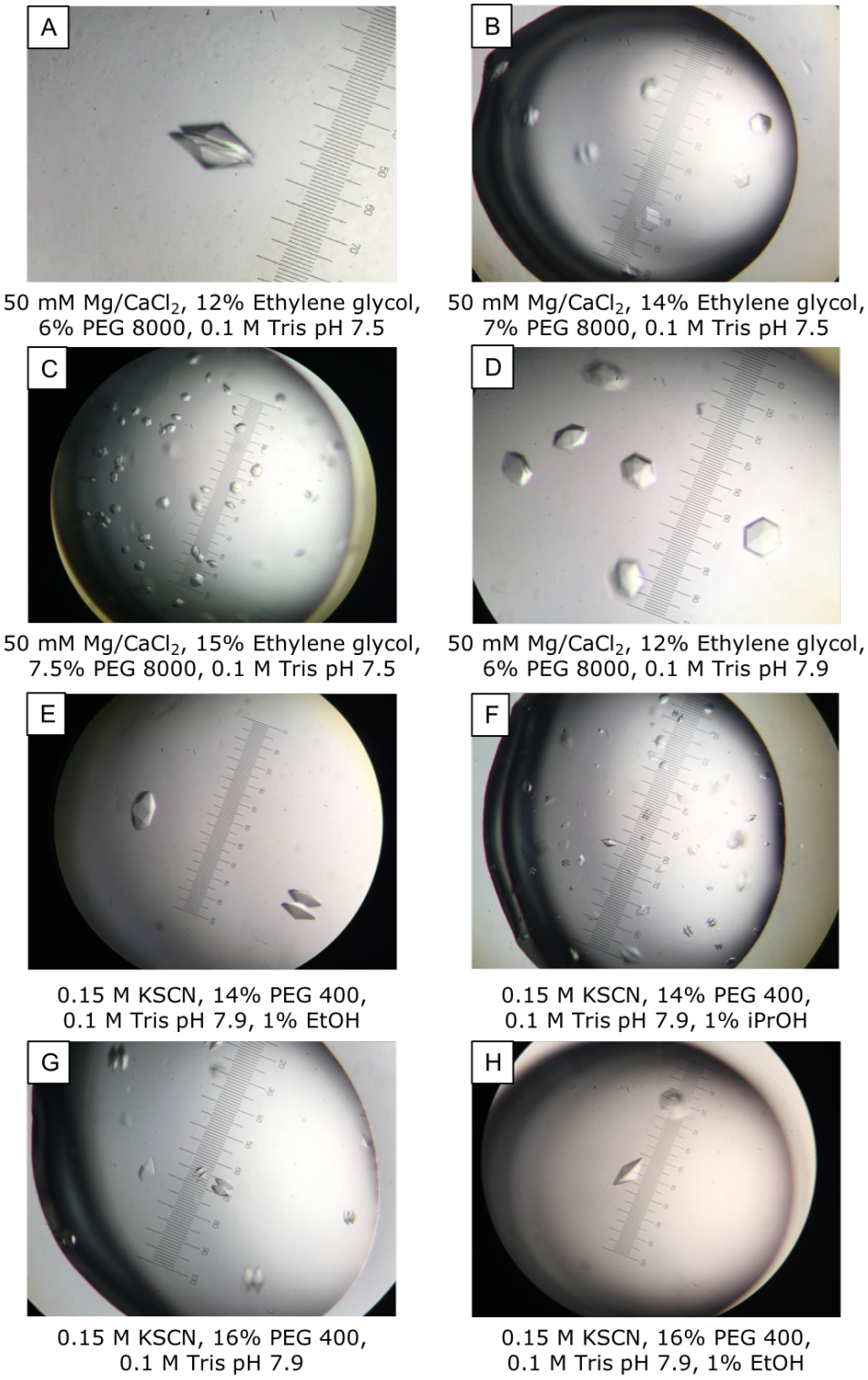


Figure 35: Crystals of (-)M062-Lau 3CS from refined screening of initial hit conditions. Crystal morphology in (A) – (D) is of two hexagonal pyramids connected base-to-base, e.g. (A). (E) – (H) vary in morphology, adopting a more classical diamond geometry, sometimes with two crystals growing end-to-end, e.g. (E).

3.8. Discussion

The crystallisation of M062 required significant optimisation and modification of the overall initial strategy, from construct engineering of a more crystallisable mutant of M062, fine-tuning of the purification procedure, iterative screening of crystallisation conditions, and testing of many crystals to obtain diffraction data sufficient to solve the structure.

Mutagenesis of non-conserved cysteine residues was necessary to procure crystals of M062. While these mutants were designed rationally to minimise the impact of M062 function based on MSA of the C7 family, the influence of these point mutants on M062's binding to SAMD9 cannot be ruled out, as a reliable *in vitro* biophysical binding assay has yet to be demonstrated. Co-IP data from our collaborators however does indicate that the 3CS binds SAMD9 equivalently to the WT by co-IP.

Crystals of the 3CS mutants of both Lau and Mav variants of M062 were obtained readily, however the overall resolution and quality of the diffraction data from initially obtained crystals was insufficient to resolve the structures. M062 was found to possess interesting characteristics during purification, such as its formation of high Mw aggregates when loaded at high concentration onto an SDS-PAGE gel (see Fig. 24 (B, C)) and its sensitivity to precipitation at concentrations >1 mg/mL at ≤4ºC, and at ≤150 mM NaCl. Optimisation of the purification strategy to prevent loss of protein due to precipitation was necessary to increase the yield of M062, as was factoring in the solubility of the protein at different concentration of NaCl for obtaining crystals for high resolution data collection (see Section 2.2.6.1).

A broad elution profile by IEX was optimised to fractionate the protein into distinct sub-populations with unique crystallisation conditions, and M062's sensitive solubility (increasing with temperature and salt concentration) was exploited in crystal trials. Iterative rescreening and modification of the purification and crystallisation strategy based on established crystallographic techniques ultimately yielded crystals which provided data sufficient to solve the structures, discussed in Chapter 4.

Chapter 4: Structural analyses of poxviral protein M062

4.1. Aims

This chapter aims to describe and provide analyses the crystal structures of the Lausanne and Mav variants of M062. This chapter analyses the collected diffraction data obtained from the crystals described in Chapter 3, and the results of phasing the M062 structures, model building and refinement, and the quality of the final solved structures obtained for each strain variant. Subsequently, these novel structures are described in detail, drawing attention to unique characteristics of the M062 structures. The Lau and Mav M062 variants are compared, their structural features examined, and contrasted against the previously known structures of the C7 family: C7 and M064.

4.2. M062-Lau 3CS diffraction data

4.2.1. Analysis of M062-Lau 3CS diffraction data

M062-Lau 3CS diffraction data were collected at the Advanced Photon Source synchrotron at Argonne National Laboratory, Illinois as described in Section 2.2.7.2. Initial diffraction images are shown in Figure 36. Strong reflections are observed to ~3 Å, and ice rings are visible at 3.67, 2.25 and 1.92 Å. Data processing and reduction was carried out by rapid automated processing of data (RAPD), NECAT's automated data processing pipeline. The dataset was initially determined to diffract to 2.2 Å, in space group P6.

The data were assessed by PHENIX.xtriage for their quality. The $P6_2$ spacegroup was assigned when reprocessed using Pointless[135]. Xtriage concluded no translational non-crystallographic symmetry (tNCS) or twinning were present by Patterson peak and Wilson's analysis. The overall Wilson's B factor calculated for the complete dataset was 84.54 $Å^2$, indicating a high degree of disorder in the crystal structure. Furthermore, Xtriage highlighted abnormalities in the data at high resolution, with a strong drop in data quality at ~2.4 Å (see Fig. 37).

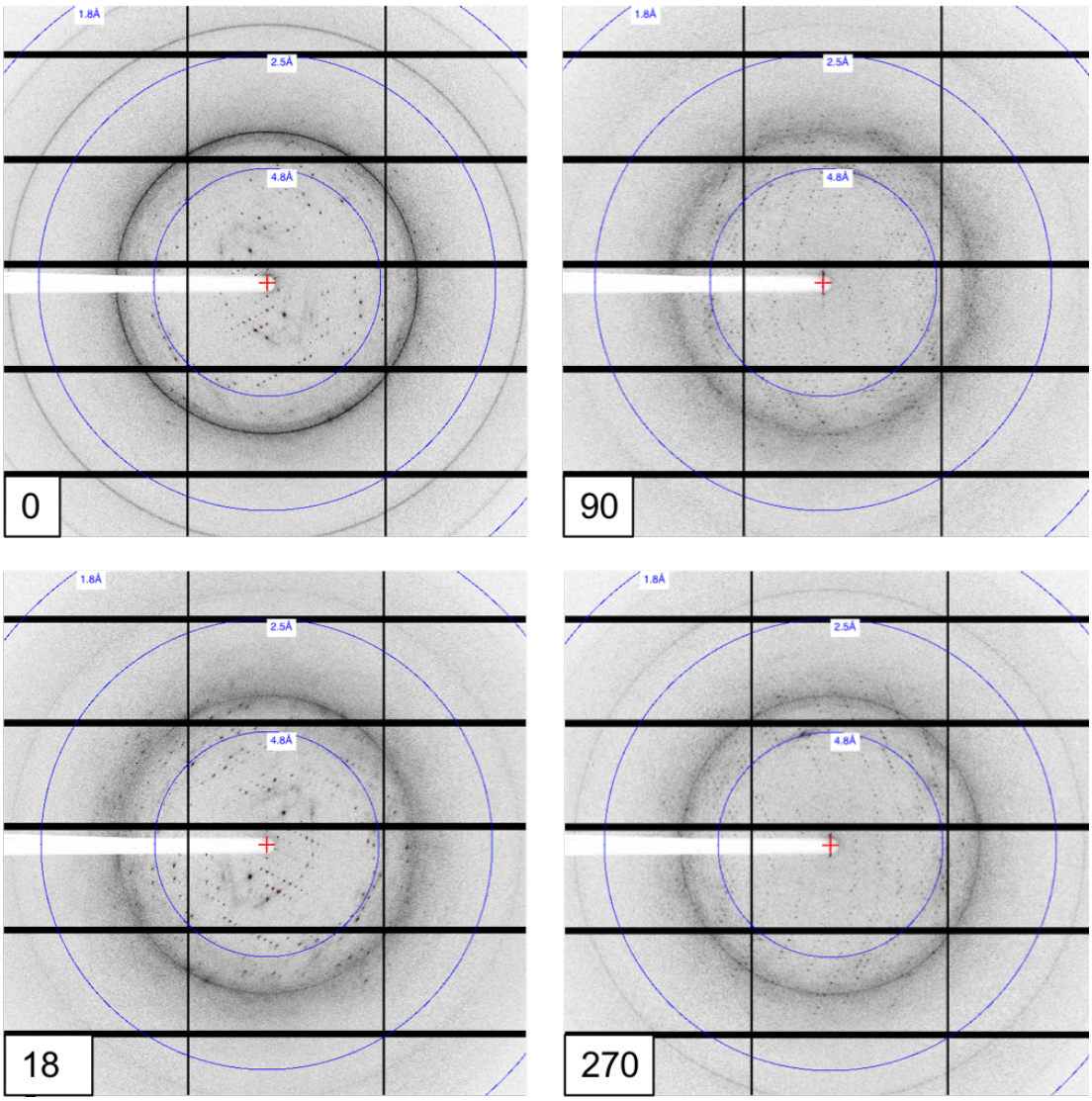


Figure 36: Diffraction data from M062-Lau 3CS. Images were taken at 90° intervals, 1° oscillation, 0.25 sec exposure, at 10% transmission, at APS, Argonne National Labs, beamline: 24-ID-C. Resolution rings (4.8, 2.5 and 1.8 Å) shown in blue. Reflections are observed to approximately 3 Å. Images generated using Albula[136].

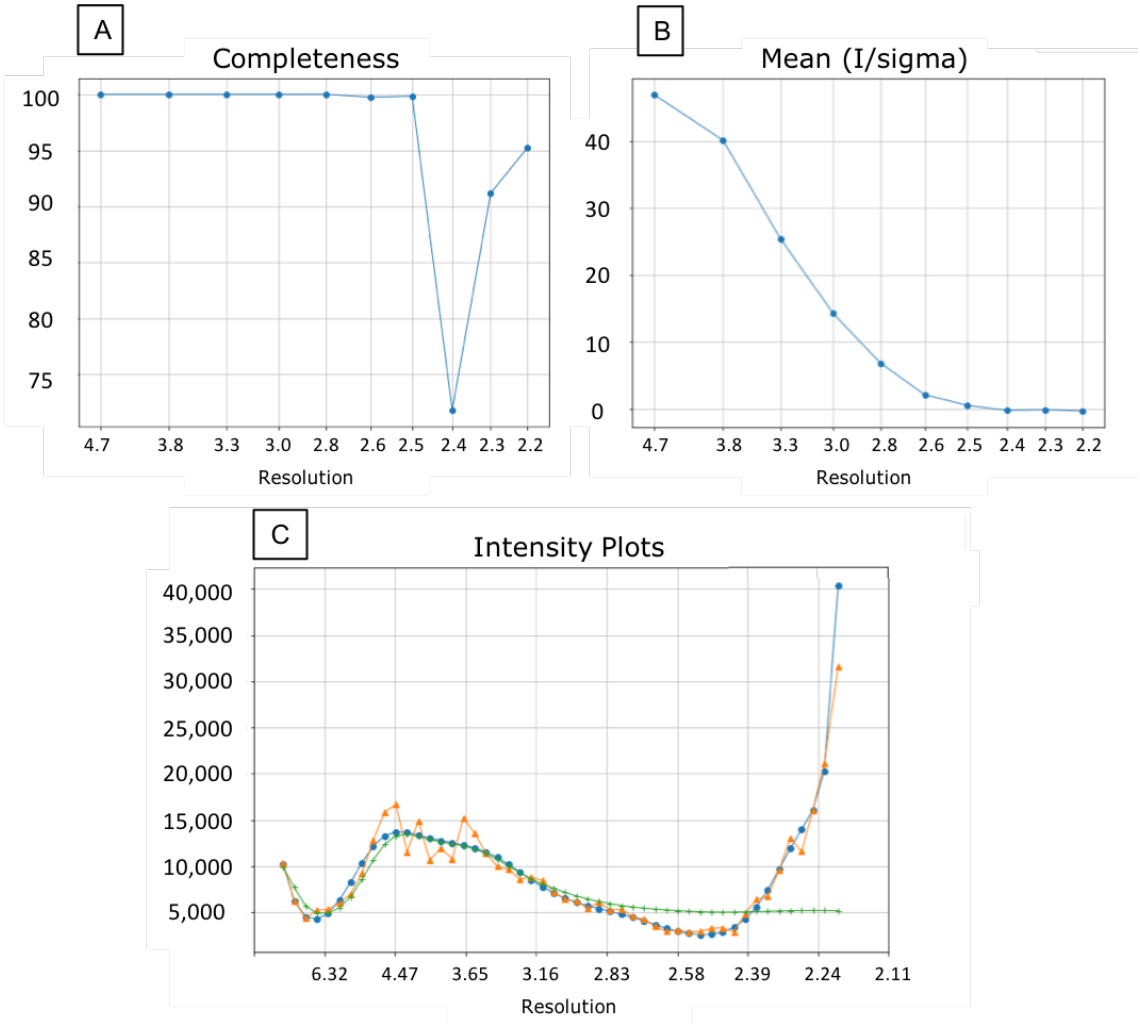


Figure 37: PHENIX: Xtriage analysis of Lau diffraction data. (A) Data completeness vs resolution. (B) Mean (I/σ) vs resolution. (C) Wilson Plot: intensity drop-off vs resolution (binned: orange, smoothed approx.: blue, predicted: green). The data analysis suggest a cut-off of 2.45 Å. Figures exported from Xtriage.

It is evident that the data quality ~2.4 Å is weaker, with the completeness dropping below 75% (Fig. 37 (A)) before returning to >90% at 2.28 Å. A similar sharp dip in signal intensity is seen from ~2.6 Å, with mean (I/σ) < 0.6 from 2.5 Å (Fig. 37 (B)), and the deviation from the predicted reflection intensities at high resolution (Fig. 37 (C)) show that the data are not reliable to the resolution reported by automated processing. Thus, the resolution limit for the Lau structure was cut at 2.45 Å for structure refinement.

4.2.2. X-ray fluorescence scan of reveals nickel in the crystal of M062-Lau 3CS

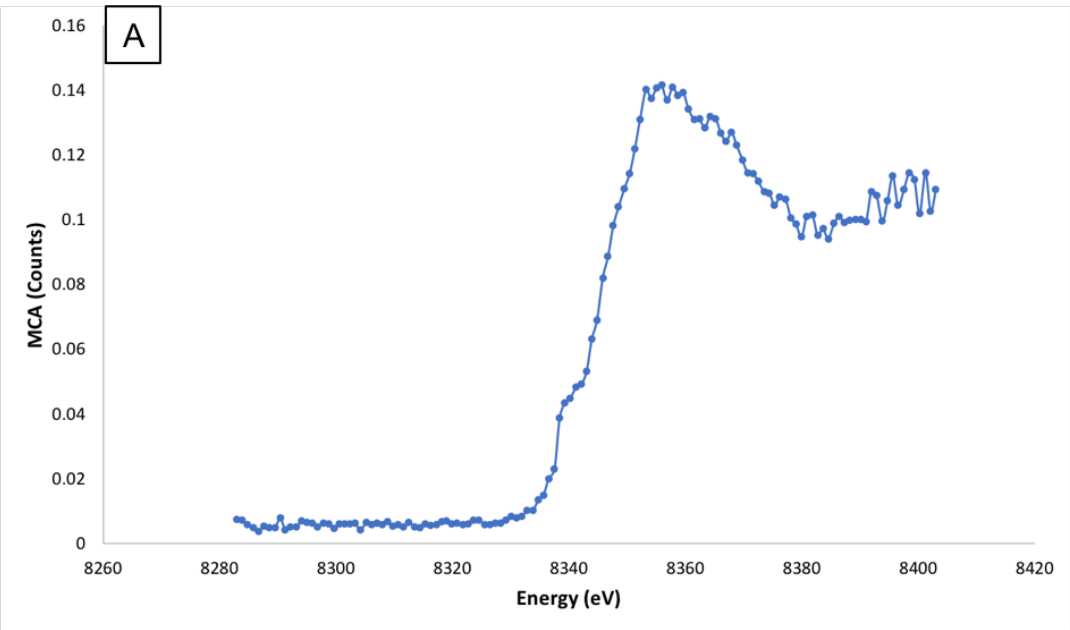
Following the advice of Martin Savko, beamline scientist at PXII Soleil, an x-ray fluorescence scan (XRF) was carried out on an isomorphous crystal, grown in the same drop as that used for structure determination, as described in Section 2.2.7.2. The x-ray fluorescence scan (Fig. 38 (A)) identified the presence of a heavy atom, with K edge absorption of 8.3458 keV, corresponding to the theoretical absorption maximum (K) for Ni (according to the University of Washington, using the theoretical approximation developed by Cromer and Liberman[132]) of 8.3328 keV.

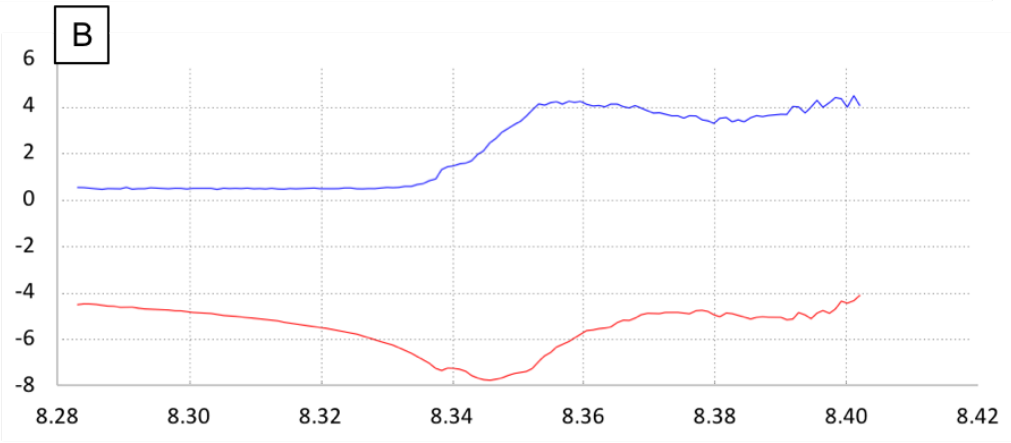
This data, indicating the presence of a heavy atom was used to attempt to solve the phase problem by experimental methods, as several molecular replacement attempts using M064 and/or C7 on previously collected data had failed.

4.2.3. Experimental phasing by multi-wavelength anomalous dispersion

Using anomalous data obtained from an isomorphous crystal at the synchrotron Soleil, it was possible to phase the M062-Lau 3CS diffraction data by direct methods. Two datasets, collected at the Ni K peak (8356 eV), and near the inflection point (8401 eV) were used for substructure determination. The anomalous signal was reasonably strong, extending to ~2.7 Å at 0.8 σ (Fig. 39 (A)), with the CC_{anom} (1/2) >30% up to ~3.6 Å (Fig. 39 (B)). Multi-wavelength anomalous dispersion (MAD) phasing using the SHELX suite of programs[137] determined that there was likely a single heavy atom with occupancy >25% (Fig. 39 (C)).

While the substructure was determined, SHELXE was unable to carry out automated structure refinement by phase modification and poly-alanine tracing. Manual refinement of the placed fragments was not feasible. However, with the higher resolution M062-Lau 3CS dataset obtained at APS, it was possible to solve the structure by molecular replacement instead, as described in Section 2.2.8.3.





	Energy (keV)	f'	f''
Peak (K)	8.3458	-7.76	2.47
Inflection point	8.4012	-4.34	4.48

Figure 38: X-ray fluorescence energy scan of (-)M062-Lau 3CS. (A) x-ray fluorescence scan, in which the incident x-ray energy varied from 8283 eV – 8403 eV, while photon count was measured (MCA: multichannel analyser). (B) The first and second derivative of (A) are plotted: f' (red), f'' (blue). The table shows the peak (K) and inflection point energy (keV). (B) calculated by CHOOCH[133] on-site at the synchrotron Soleil.

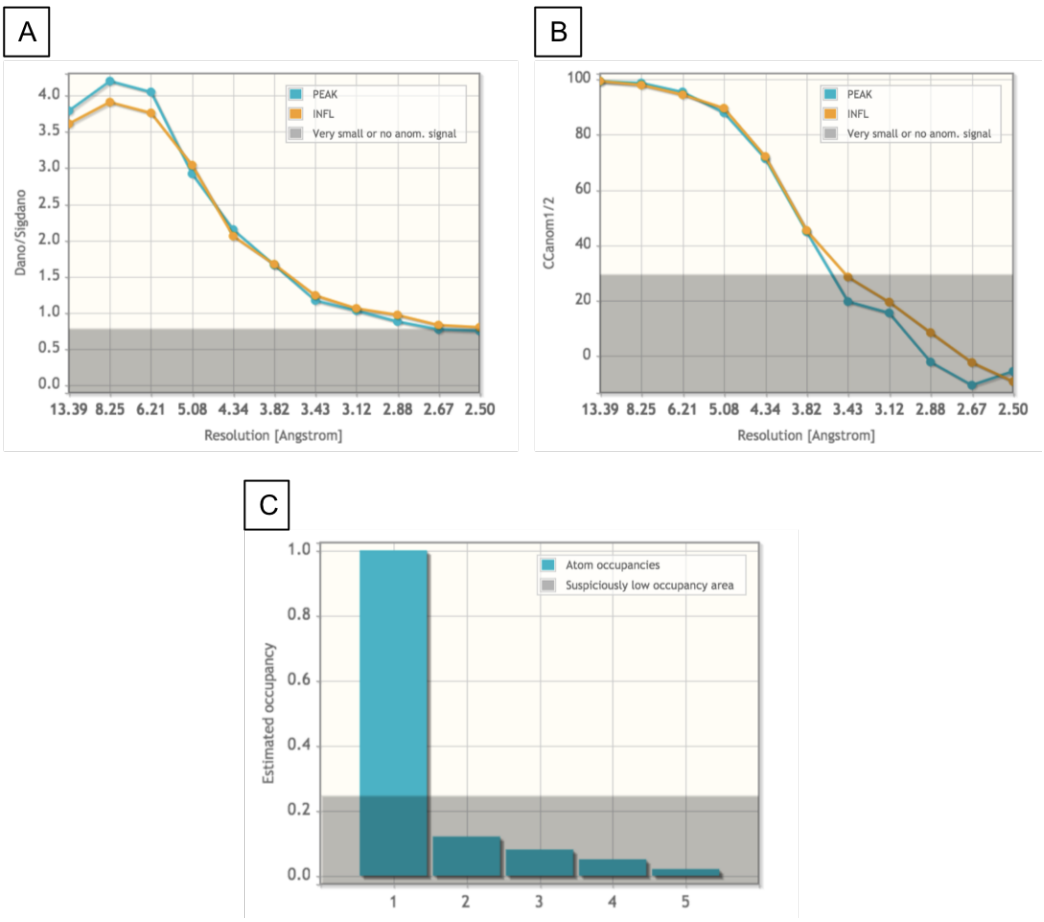


Figure 39: MAD phasing of M062-Lau 3CS. (A) The anomalous signal is shown vs resolution, signal:noise <0.8 σ (grey). (B) The CC_{anom} ½ (%) is shown vs resolution (Å), <30 (grey). (C) SHELXD prediction of the occupancy (<25% grey) and number of heavy atoms in the substructure. Figures exported from SHELX[137].

4.3. Structure solution of M062-Lau 3CS

The crystal structure of M062-Lau 3CS is accessible from the PDB via accession code 6S5I. Solving the structure of M062-Lau 3CS required extensive manual refinement and iterative build cycles, with unexpected structural features present, distinct from the known C7 structures. Presented here is an initial overview describing the final deposited structure and refinement statistics. This is followed by the refinement steps taken to arrive at the solution, with descriptions of the distinct structural features of the M062-Lau 3CS crystal structure.

4.3.1. The crystal structure of M062-Lau 3CS

M062-Lau 3CS crystallised in the $P6_4$ space group, with unit cell parameters of $a = b = 94.8$, $c = 44.1$, $\alpha = \beta = 90$, $\gamma = 120$. A single molecule of M062-Lau 3CS forms the ASU, adopting the β sandwich fold characterised by the C7 family. Table 16 summarises the crystallographic data. Through manual refinement of the initial model generated using PHENIX: Autobuild of 137 residues, the final model consists of 156 of the total 161 residues of (-)M062-Lau 3CS, with an R-work 0.26 and R-free: 0.31 (Table 16). Five residues are omitted from the structure as the electron density did not allow for their accurate modelling. A break in the main chain from residue W144 to K148 is present, with the sidechain of K148, and the final 2 carboxy residues omitted due to model disorder (see Fig. 40).

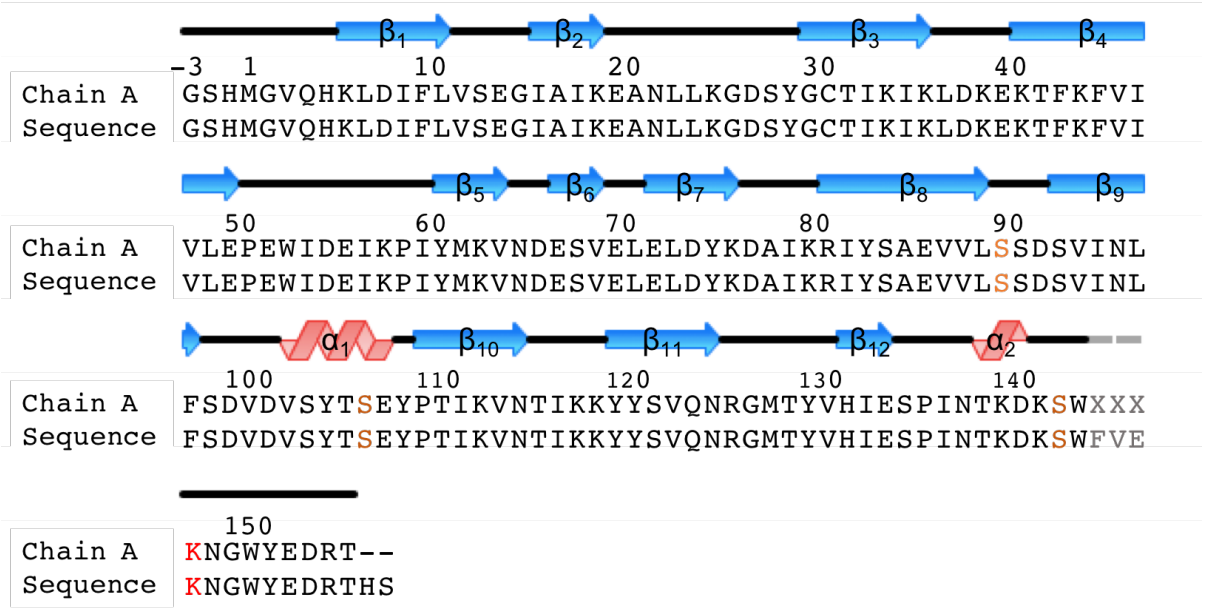


Figure 40: Alignment of the M062-Lau 3CS model and (-)M062-Lau 3CS sequence. The sequence of the refined structural model of M062-Lau 3CS aligned against the complete (-)M062-Lau 3CS construct sequence. Secondary structural features and residue number are annotated. Residues omitted from the final model shown as 'X's (grey), '-' (black), Cys → Ser mutants are shown in orange, and K148 (omitted sidechain) in red. Alignment generated by PHENIX: Molprobity[138].

The β sandwich fold of M062-Lau 3CS, shown in Figure 41, is formed by the curved interface between two sets of six antiparallel β sheets, with each strand connected by loops of varying length, and is flanked by short α helices. The N terminus, coordinating a Ni atom, extends past the surrounding loops above the β core. The C terminus is discontinuous as the density for residues 145 – 147 was weak. The C terminus extends out from the globular fold of the protein, forming interactions with a symmetry related neighbouring molecule.

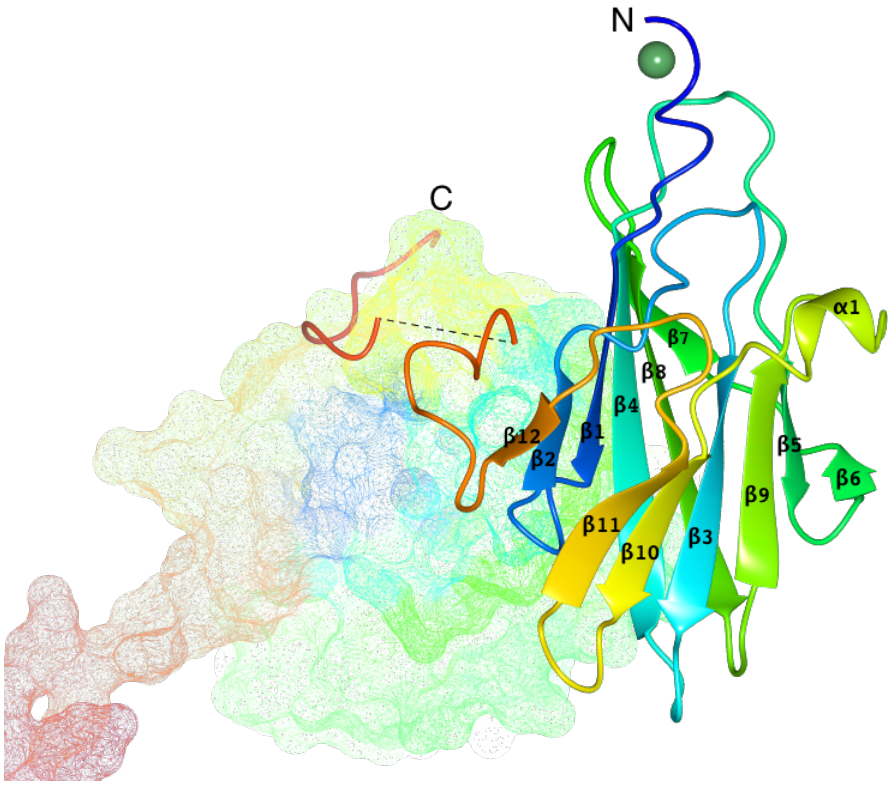


Figure 41: The crystal structure of M062-Lau 3CS. M062-Lau 3CS adopts the C7 β sandwich fold. A ribbon model is shown in rainbow colour (N terminus (blue), C terminus (red)), with the secondary structures and termini labelled. A Ni atom (green sphere) is shown chelated by the N terminus. A dashed line connects residues 144 – 148, where there is a break in the main chain density, followed by the C terminal tail which extends beyond the β sandwich fold, making contact with a symmetry mate (mesh). Figure generated with CCP4mg[139].

Table 16: Data collection and refinement statistics for the crystal structure of M062-Lau 3CS, PDB: 6S5I

Wavelength	0.979100
Resolution range	47.41 - 2.45 (2.53 - 2.45)
Space group	P 6_4
Unit cell (a, b, c) (α, β, ɣ)	(94.8, 94.8, 44.1) (90, 90, 120)
Total reflections	82181 (7776)
Unique reflections	8422 (846)
Multiplicity	9.8 (10.1)
Completeness (%)	99.55 (100.00)
Mean I/sigma(I)	23.92 (0.31)
Wilson B-factor	84.54
R-merge	0.05385 (11.77)
R-meas	0.057 (12.4)
R-pim	0.01841 (3.894)
CC1/2	0.999 (0.407)
CC*	1 (0.761)
Reflections used in refinement	8467 (846)
Reflections used for R-free	873 (83)
R-work	0.2628 (0.7027)
R-free	0.3138 (0.6054)
CC(work)	0.579 (0.254)
CC(free)	0.647 (0.013)
Number of non-hydrogen atoms	1277

macromolecules	1270
ligands	1
solvent	6
Protein residues	156
RMS(bonds)	0.014
RMS(angles)	1.91
Ramachandran favoured (%)	90.13
Ramachandran allowed (%)	7.89
Ramachandran outliers (%)	1.97
Rotamer outliers (%)	6.21
Clashscore	8.25
Average B-factor	54.62
macromolecules	54.60
ligands	94.88
solvent	52.25
Number of TLS groups	1

Statistics for the highest-resolution shell are shown in parentheses. R-free was calculated using 10% of data.

4.3.2. Phase solution of M062-Lau 3CS by molecular replacement using M064 as a search model

The quality of diffraction data obtained at APS was sufficient to phase M062-Lau 3CS by MR. The crystal structure of M064 (5CZ3)[29] was solved to 2.5 Å by the Xiang group in 2015. M062 and M064 have 41% sequence identity, and Chainsaw[109,110] was used to modify M064 as described in Section 2.2.8.2, prior to its use as a search model with PHENIX: Phaser[111]. The Log likelihood gain (LLG) was 278, and TFZ-score 19.2, indicating that a correct solution had been found. A model was then built from the electron density map generated by MR using PHENIX: Autobuild. The model, comprising 137 (of 161) residues was of adequate quality for refinement in Coot[112], with initial residuals of R-work: 0.26 and R-free: 0.31.

The initial model, displayed in Fig. 42 (A) shows M062-Lau 3CS adopts an overall fold consistent with existing C7 family structures, that of a β sandwich. A superposition of the initial M062-Lau 3CS model and C7 using PHENIX: superpose_pdbs[138] (Fig. 42 (B)) has a global RMSD of 0.3575 Å^2, and shows the overall fold is highly conserved. There are however distinct structural differences, notably at the termini which were not completed in the initial model.

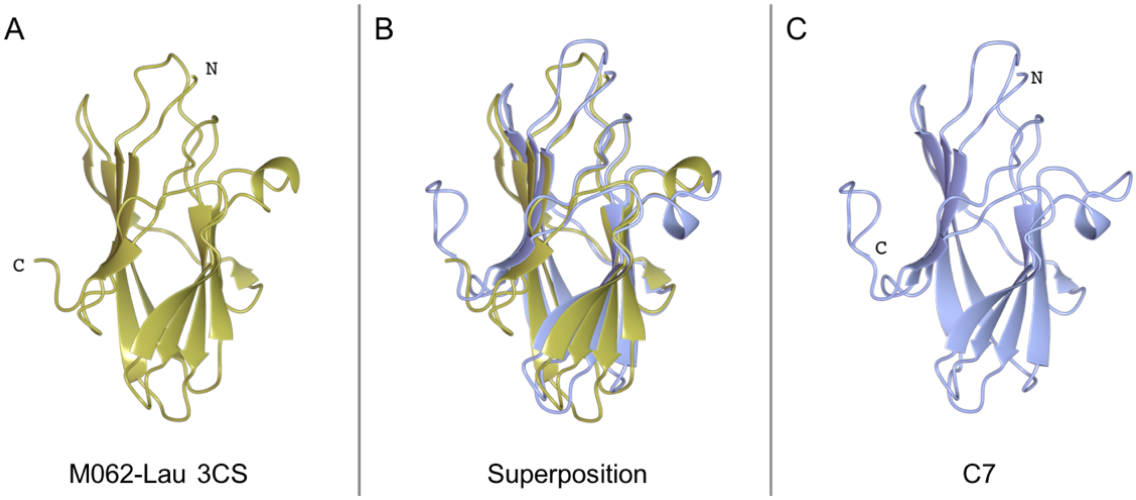


Figure 42: Superposition of initial M062-Lau 3CS and C7 structures. Ribbon models of M062-Lau 3CS, gold (A), C7, blue [PDB: 5CYW] (C) and a superposition (B) are shown. Termini annotated. Figures generated with CCP4mg132.

4.3.3. Structural refinement of the N terminus

The PHENIX: Autobuild model was used as a starting point from which the structure was refined. Model building was carried out using Coot[112], with residue side chains built into the electron density manually. The N and C termini were particularly difficult regions of the structure to resolve, as the initial model was missing 4 residues at the N terminus, and 20 at the C terminus. The initial density at the N and C termini extended marginally at the terminal residues, yet enough allowed for iterative model building, which gradually extending the chain length, with modest improvements to the refinement statistics. Progress in refinement of the polypeptide halted at both N and C termini once the electron density no longer allowed for visualisation of continuous backbone density at 1.5 σ (see Figure 43). While the electron density continued at the N terminus, the density assumed a morphology into which it was not initially possible to build the remaining post-thrombin cleavage tag residues -2 to 0 (GSH), see Fig. 43.

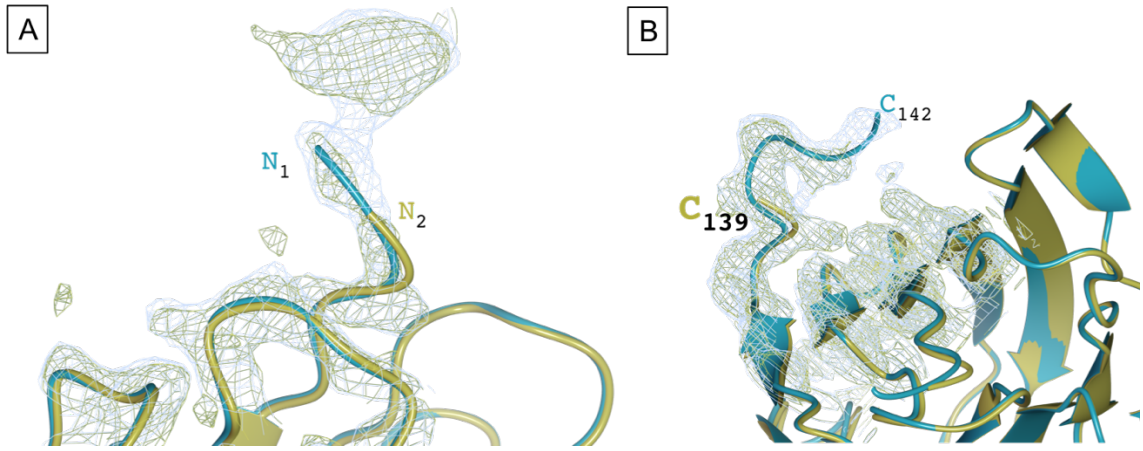


Figure 43: Initial N and C termini electron density for M062-Lau 3CS. Electron density maps are shown contoured at 1.5 σ (gold, cyan) corresponding to superimposed ribbon models: Autobuild (gold) and manual build (cyan). (A) shows the N terminus beginning at GLY 2 (Autobuild), MET 1 (manual) and (B) the C terminus, ending at ASN 139 (Autobuild) and LYS 142 (manual). Figure made with CCP4mg[139].

4.3.3.1. Nickel chelation by the N terminus

When viewing the electron density at a high contour level (>3 σ), it was evident that there was a heavy atom located in close proximity to the N terminus (see

Fig. 44). The overview of the M062 protein (Fig. 44 (A)) shows there is density (shown in grey) for the β sandwich, and the side chains of the globular region of the protein. Additionally, beyond the N terminus, electron density can be seen in Fig. 44 (B) contoured at 5 σ. The proximity of the N terminus (Gly5) to the 'blob' is shown more clearly in Fig. 44 (C).

Considering the electron density, combined with the XFS data (see Fig. 38) and the SHELXD substructure determination, it was concluded that a single Ni atom was the source of this density. After fitting the Ni atom, and considering the coordination complexes of nickel and the chemistry of the remaining N terminus residues, the final model was built with the polypeptide chelating the Ni atom. The GSH tag residues coordination the nickel atom using nitrogen lone pairs of electrons (see Fig. 45 (A)).

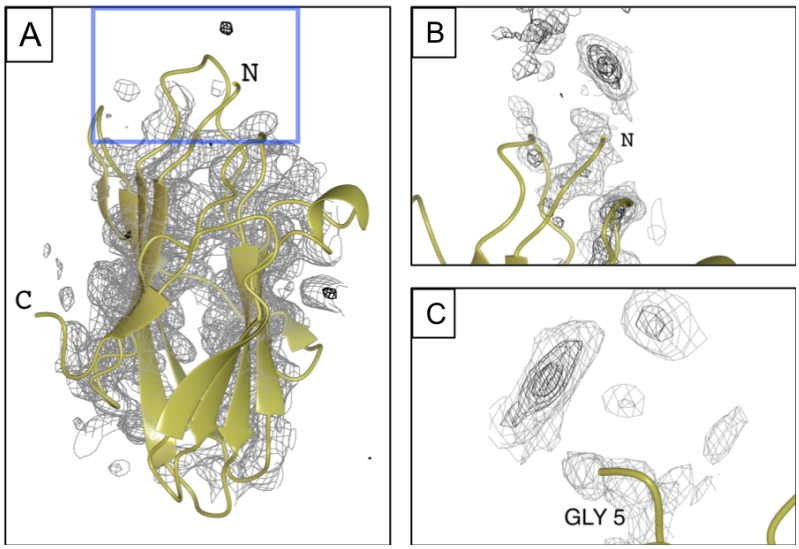


Figure 44: Initial N terminus with unknown electron density. (A) A ribbon model of the Autobuild M062-Lau 3CS model is shown (gold) with electron density map contoured at 1 σ (light grey), 3 σ (grey) and 5 σ (black), with N and C termini annotated. A region of high electron density at the N terminus (blue box) is expanded in (B). (B) A close-up of the N terminus and the unaccounted for electron density is shown. (C) View of (B) rotated 90° shows Gly5 oriented towards the unexplained electron density. Figures made with CCP4mg[139].

The N terminal chelating structure fits the electron density well, and explains how the additional electron density seen in Fig. 44 is a result of the Ni atom coordination, with the N terminus GSH residues of the structure forming a planar coordination complex, before the protein continues into the globular β sandwich fold (Fig. 45 (B)). A square planar coordination state is assumed by GSH:Ni, with the 3 backbone amines, and the imidazole of His-1's (ND1) providing 4 nitrogen atoms which act as a tetradentate chelator (Fig. 45 (C, D)).

Validation of the coordination chemistry and geometry of the nickel chelation by these residues was carried out using the CheckMyMetal (CMM) server[11]. CMM validation indicates that the coordination of the Ni modelled falls within 'Acceptable' and 'Borderline' grades for the validation parameters (Fig. 46, Table), discussed further in Section 4.7.2.1.

The coordination bonds between the Ni-N and their bond length are plotted in Fig. 46, and 3 of the 4 interactions (red bars) match the average Ni-N bond lengths from the CMM database (obtained from the Cambridge Structural Database). Overall, the CMM server validates the selection and coordination of Ni by the N terminus of M062-Lau 3CS. It is likely that the Ni atom was chelated during the IMAC Ni-NTA chromatography step and eluted with the protein from the resin, as nickel was not a component of any used in purification or crystallisation of M062.

The presence of this Ni atom does not reflect the native M062 structure, as it is coordinated solely by residual thrombin cleavage tag residues. This N terminal feature is crucially facilitating many crystal contacts between the asymmetric units (ASU), and may have been essential in crystal growth. The C terminal also plays a major role in the formation of crystal contacts, with implications regarding the understanding of M062, and potentially the C7 family's function.

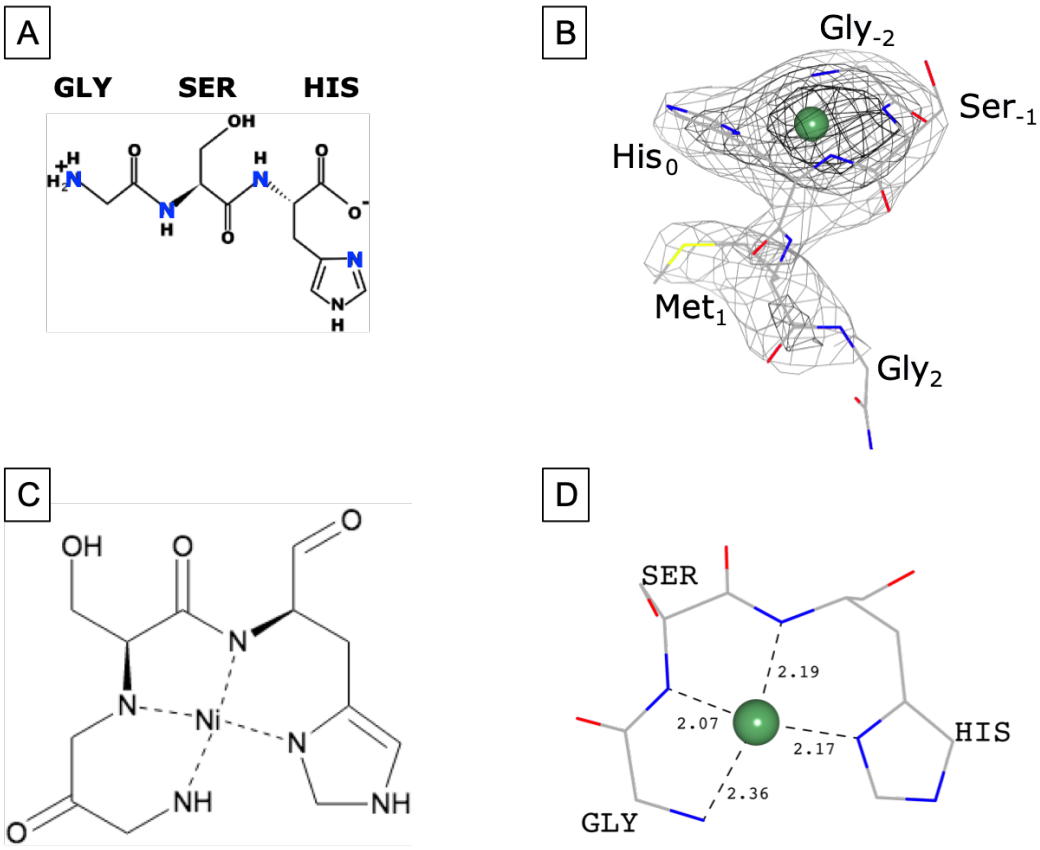
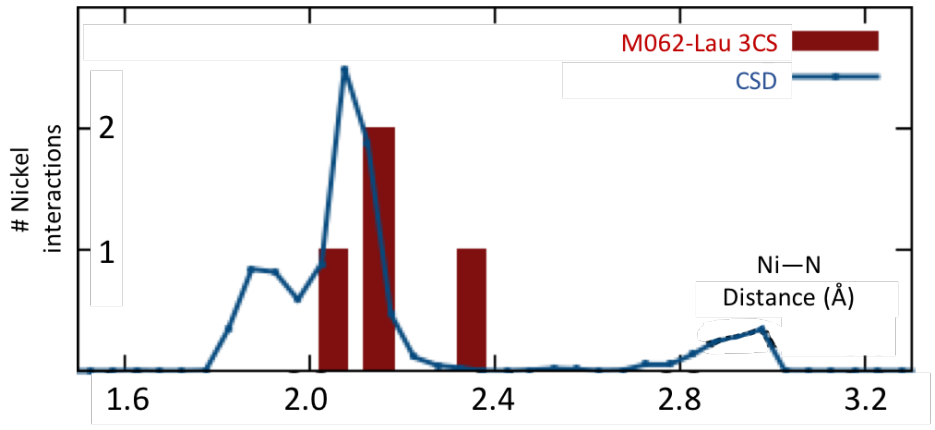


Figure 45: Chelation of nickel by M062-Lau 3CS N terminal residues. (A) The composition of resides -2 – 0 of M062-Lau 3CS shown in skeleton form. These 3 residues remain of the thrombin cleavage site after proteolysis. Nitrogen atoms which coordinate Ni shown in blue. Figure made using PepDraw[140]. (B) Coordination of Ni by the N terminus of M062-Lau 3CS showing Ni atom as a sphere (green). Structure shown coloured by element, annotated with residue 3-letter code and sequence number. Electron density map contoured at 1.5 σ (light grey), 3 σ (grey) and anomalous difference map at 5 σ (black), bonds are shown as dashes. The coordination of Ni by the N terminal residues is shown by dashed line in skeleton view (C) and in the model, with bond lengths in Å (D). Figure (C) made with ChemDraw[141], (B) and (D) generated with CCP4mg[139].



Atom	Occupancy	B factor (env.)	Ligands	Valence	nVECSUM	Geometry	gRMSD	Vacancy
Ni	1	94.9 (38.4)	N_4	1.2	0.22	Square Planar	14.3°	0

Legend	Acceptable	Borderline	Outlier

Figure 46: CheckMyMetal validation of the GSH:Ni coordination by M062-Lau 3CS. Metal binding validation parameters are shown in the table, colour coded according to the legend table. The Ni B factor is shown with the environmental B factor[117] in parentheses. nVECSUM measures the deviation of the coordination complex symmetry[142], and gRMSD gives the RMSD of the geometry angles compared to ideal geometry[117]. Graphed in blue is the average number of Ni-N interactions vs bond length (Å), as per the Cambridge Structural Database (CSD), in red, the number Ni-N interactions and their bond length Å in M062-Lau 3CS. The CMM validation server indicates that the Ni coordination complex in the M062-Lau 3CS model is within the target values, and thus is likely representative of the metal-binding site.

4.3.3.2. N terminus crystal contacts: serendipitously stuck together

The previously described structural feature: the Ni chelated by the N terminus, facilitates crystal contacts between symmetry-related molecules, as shown in Fig. 47. The N terminus:Ni complex occupies a small region proximal to helix 1 of its neighbouring symmetry mate (Fig. 47 (B)), with a local contact surface area of 146 Å^2. The contact surface area between the symmetry mates through the N terminal region contributes significantly to the total contact surface area between all crystal symmetry mates, at 655.7 Å^2 (Fig. 47 (B)) of the total 1,313.7 Å^2 (calculated by the PISA analysis server[143]).

This crystal contact is formed primarily by hydrogen bond interactions between the Gly-3 backbone amine and Tyr109's backbone carbonyl oxygen and Thr111's sidechain hydroxyl group, and His-1's imidazole ring with Ser107's backbone carbonyl oxygen (Fig. 47 (C)). Hydrogen bonding between Gly-3's carbonyl backbone oxygen and Thr 32, and Ser-2's sidechain hydroxyl with Thr106's sidechains also form crystal contacts (Fig. 47 (D)).

This N terminal mediated crystal contact comprises 50% of the total crystal contact surface area. This contact is a result of the chelation of the Ni atom by the N terminal thrombin cleavage tag residual 'GSH' residues (Fig. 40), and its interaction with its symmetry mate, which includes a hydrogen bond with S107, one of the three Cys → Ser mutagenized residues. It is reasonable to conclude that the crystallisation of M062-Lau 3CS was mediated by the interaction of non-native residues, a result of the affinity tag cleavage residuals, and the construct engineering which incorporated mutagenesis of the surface exposed cysteines to serine.

4.3.4. Structural refinement of the C terminus

4.3.4.1. Missing residues and unexplained electron density

The M062-Lau 3CS model was refined to encompass residues -2 through 144, 147 residues in total, with the N terminus completed by the Ni structure, equalling the overall length of the C7 and M064 structures. C7 is shorter than other members of the C7 family (see Fig. 61), and while the crystal structure of C7 is nearly complete (missing only the C terminal residue), the model of M064 is missing 53 residues at the C terminus, thought to be a consequence of proteolytic cleavage over the course of crystallisation[29].

The electron density for residues 145 – 161 of M062-Lau 3CS was initially thought to be absent from the crystal structure, being highly disordered and without the possibility of being accurately modelled with the crystal data available (see Fig. 48 (B)). There did, however, remained an unexplained region of electron density sitting at the interface of the β sandwich, between β strands 11 and 12 (Fig. 48 (A)). As the electron density map improved with each round of refinement, it revealed what appeared to be a polypeptide chain beside β strand 11 (Fig. 48 (C)).

A polyalanine chain was initially built into the density beside β strand 11, and using the bulky sidechain density visible adjacent to Tyr 121 (Fig. 48 (D)), it was possible to build a polypeptide, accounting for the missing C terminal residues. As only a single molecule of M062-Lau 3CS forms the ASU, it was only possible to explain the C terminus contributing to this density as a result of crystallographic symmetry. It is apparent that the distance from C terminus to β 11 is too great to allow for an intramolecular interaction with the missing residues coiling back upon the β sandwich.

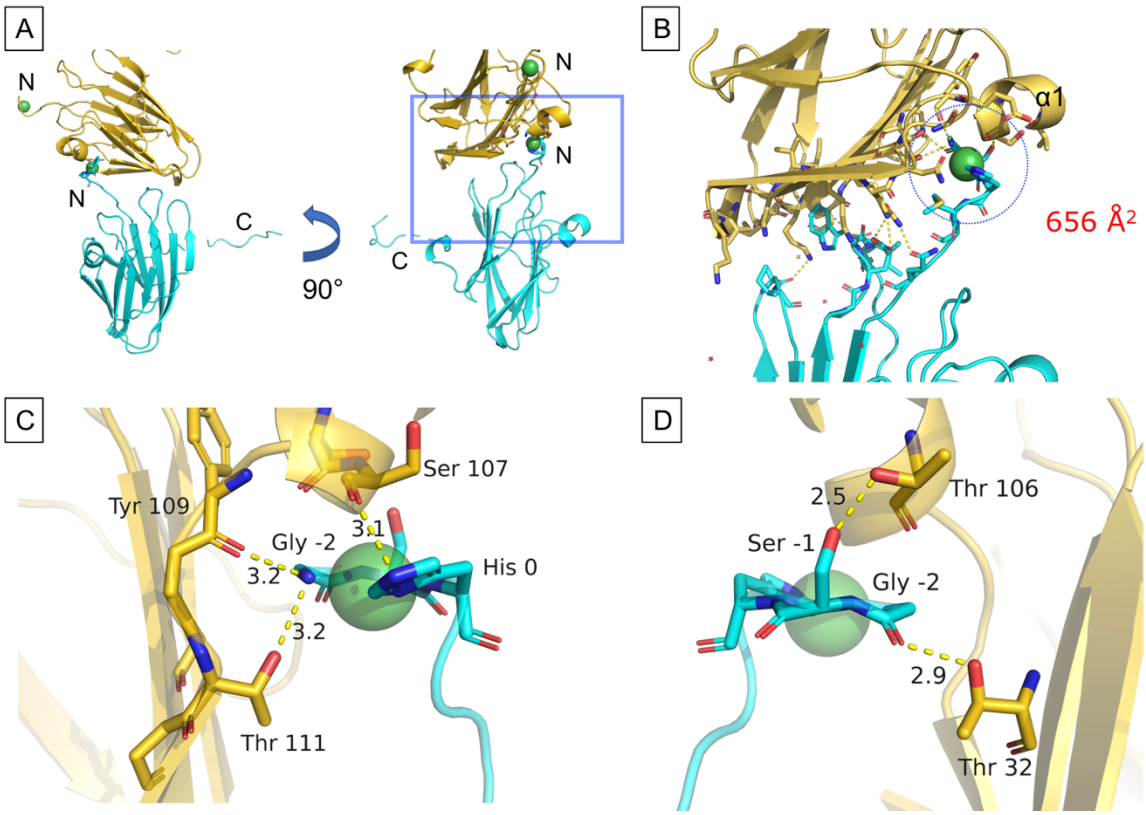


Figure 47: N terminus:Ni complex crystal contacts. (A) Overview of the N terminus crystal contacts between M062-Lau 3CS, featuring two ribbon models (gold, cyan) with Ni shown as a sphere (green), termini annotated. Boxed region (blue) expanded in (B) shows the N terminus GSH:Ni complex and the crystal contact residues (stick bonds, coloured by element). The N terminus forms crystal contacts with a symmetry mate in proximity to helix 1 (annotated) and has a large contact surface area of 656 Å^2. (C) and (D) show the crystal contacts between GSH residues (cyan) and the neighbouring molecule residues (gold), annotated (circled in (B) in blue). Hydrogen bonding interactions are shown as dashed lines with distances annotated.

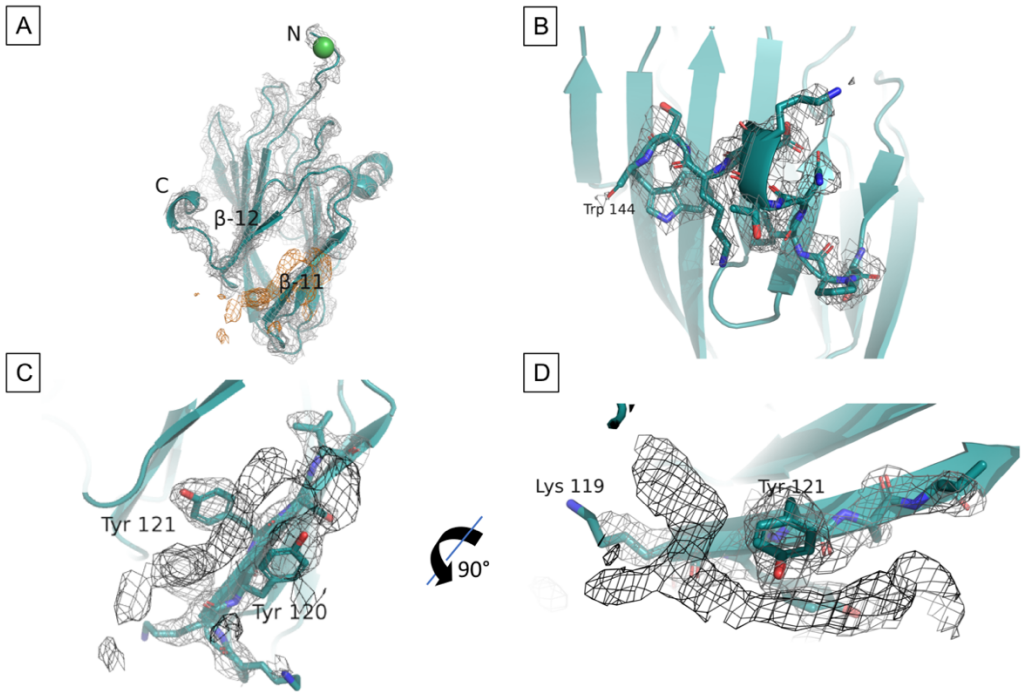


Figure 48: Unexplained electron density in M062-Lau 3CS. (A) Ribbon model overview of the M062-Lau 3CS fold with termini annotated, Ni atom displayed as a sphere (green). A region of unexplained electron density is shown (orange) between β strands 11 and 12. (B) The extent of the C terminus with electron density (1 σ) is shown, ending at Trp144. (C) Enhanced view of the electron density at β 11 (black) shows the unknown density woven between Tyr120 and Tyr121. (D) Rotating (C) 90° in the plane of the page, the density is seen to branch around Tyr121.

4.3.4.2. M062-Lau 3CS C terminal tail-mediated crystal contacts

The $P6_4$ space group defines a six-fold symmetry coupled with a right-handed screw axis about the c axis, which is shown to be formed by the C terminal tail extending out from one ASU and forming crystal contacts with the subsequent symmetry mate (see Fig. 49). Closer views of the C terminus are shown in figure 50 (B) and (C), where the tail is shown to form contact between β strands 11 and 12 of the adjacent molecule. Figure 50 (D) shows the sidechains of the tail form intimate contacts with the groove formed between β 11 and β 12. More thorough examination of these interactions is presented in the next section.

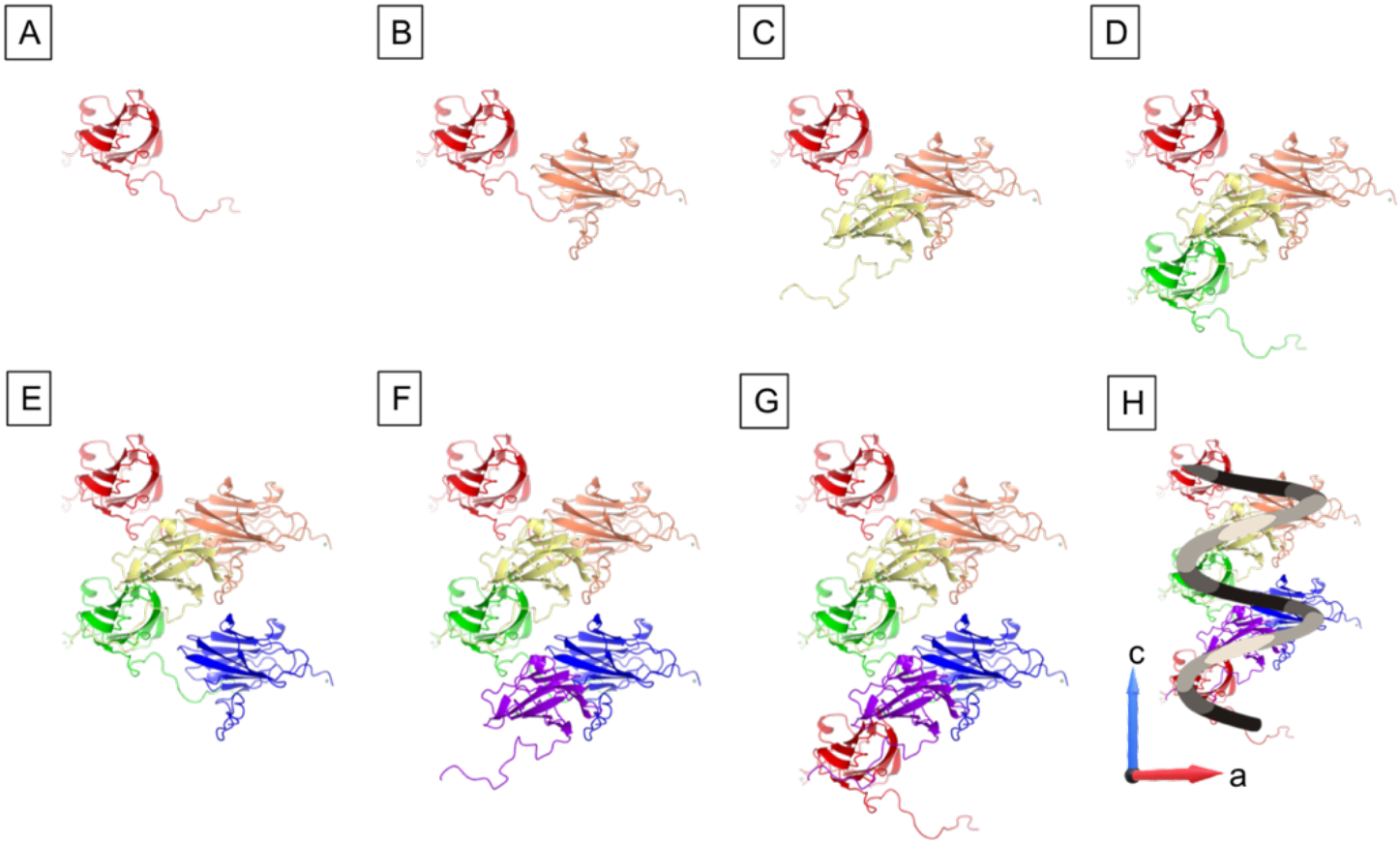


Figure 49: M062-Lau 3CS helical crystal contacts. (A) – (G) Ribbon models of M062-Lau 3CS, coloured by chain are shown, with the addition of a C terminal-mediated crystal contact symmetry mate in each subsequent figure. (H) Depicts the $P6_4$ crystal symmetry: a right-handed helix overlaid upon the C terminal tail crystal contact mates. Crystal axes labelled. Figures generated with CCP4mg[139]

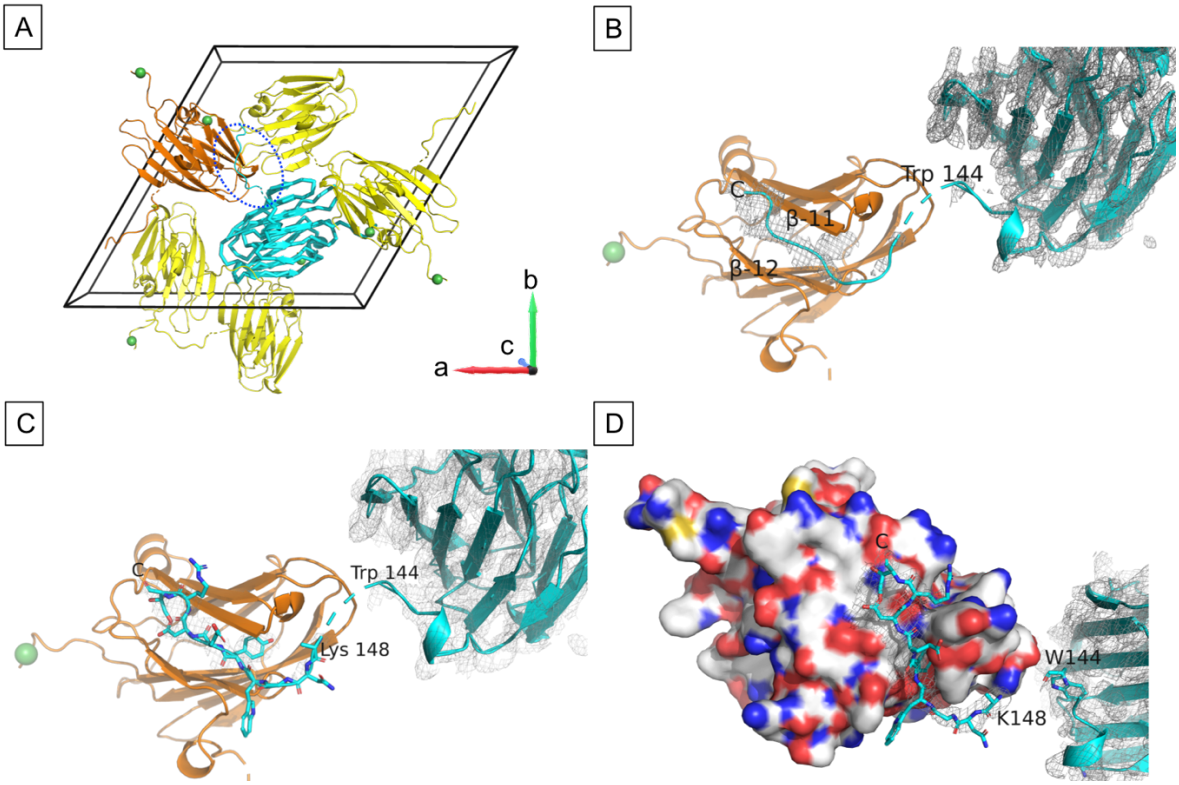


Figure 50: Interaction of the C terminus of M062-Lau 3CS with its crystal contact symmetry mate. (A) A unit cell containing crystal contact symmetry mates to M062-Lau 3CS (cyan) are shown in ribbon model (yellow, orange). Crystal axes are shown. The C terminal crystal contact mate is in orange, with the tail circled in blue. (B) Ribbon models depict the C terminal tail inserted in a groove formed by β 11- β 12. (C): ribbon model, and (D): surface view, show the sidechains of the C terminal tail interfacing with the β 11- β 12 groove. Electron density maps contoured at 1.5 σ. Figures generated using PyMOL[144].

4.3.5. C terminal tail interaction with the β11-β12 groove

A series of interactions between the C terminal tail, from residue 151 – 158, and the groove region formed at the interface of β strands 11 and 12 are identifiable, and displayed in Figure 51. The PISA analysis server[143] calculates the interface contact surface area is 555.2 Å^2. The interface is best described by three modes of interaction: i) between the backbones of β 11 and the C terminal tail, ii) at the initial point of contact between the C terminal tail and the groove, Trp151 and Pro136 of β 12, and iii) the 'lock and key' interaction between Tyr121 and the hydrophobic pocket formed at β 11.

The hydrogen bonding between the backbone of C terminal tail residues Glu 153 and Arg155 and β 11 residues Lys119 and Tyr121 (Fig. 51 (A)) is similar to the hydrogen bonding seen in the β sheets which form the β sandwich of M062-Lau 3CS. These interactions orient the tail adjacent to β 11, and β 12, effectively extending this end of the β sandwich by the insertion of a pseudo β strand from the C terminal tail of the contributing symmetry mate.

The tail structure turns as it coils into the groove (Fig. 51 (C)), and at the start of the groove, the indole sidechain of Trp151 is oriented with the ring perpendicular to Pro136 in the loop immediately after β 12. A C-H$\cdots\pi$ interaction from Pro136 C_α to the delocalized Trp151 ring is present, as seen in Fig. 51 (B). This stabilizing interaction is discussed further in Section 4.6.3.3.

The major interaction identified between each subunit at this interface is mediated by burying of the phenol ring of Tyr152 into a hydrophobic pocket which is seen clearly as a depression in the surface between β 11 and β 12 in Fig. 51 (C). This pocket is highly hydrophobic, comprising Ile, Val, Tyr, Lys, Thr, Ser and Glu residues. The sidechains of these residues are oriented to create a hydrophobic environment in which Tyr152 is inserted, visible in Fig. 51 (D), and (E) which features a hydrogen bond between Tyr152 and Glu14.

Sequence alignment of the C7 family shows conservative substitutions are present for residues I16, N115, K119, and Y121 which interact with the buried Y152 are conserved among C7 homologues, indicating a selection to maintain the biochemical properties of the residues in these positions (see Fig. 61).

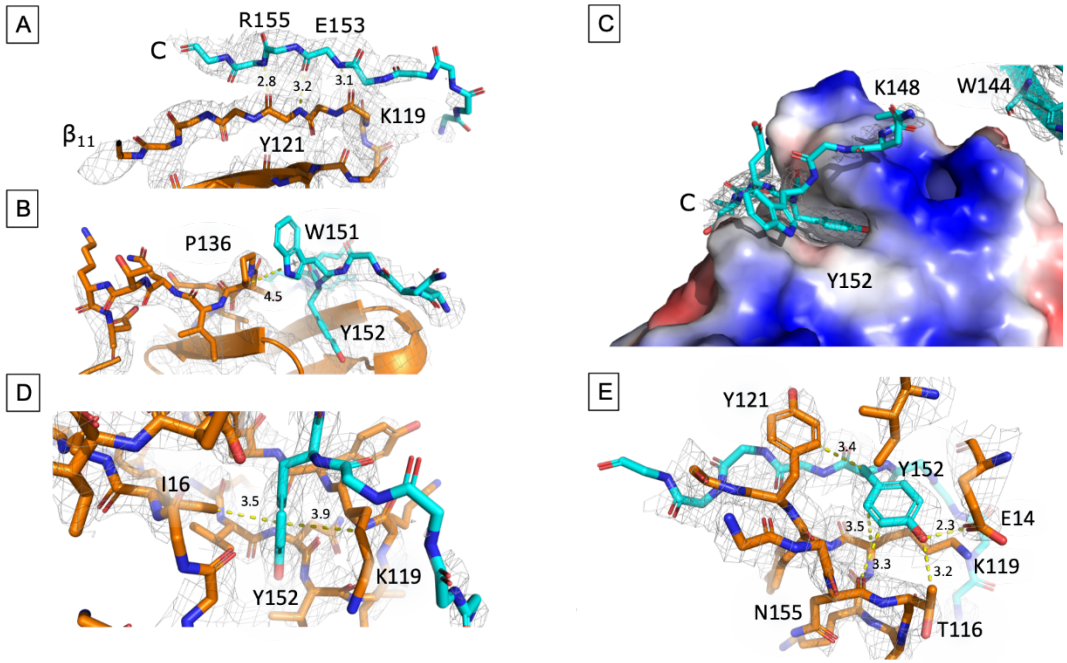


Figure 51: M062-Lau 3CS C terminal tail:groove interface bonding interactions. (A) Hydrogen bonds between the C terminal tail (cyan) and β 11 (orange) backbone are shown. (B) A C-H···π bond is formed between Pro136 of β 12 and Trp151. (C) A surface electrostatic potential map shows Tyr152 is buried in a hydrophobic pocket. (D) Hydrophobic packing interactions between Ile16 and Lys119 sidechains with the phenol ring of Tyr152 are shown. (E) Further interactions between the residues which form the pocket (Asn115, Tyr121, Thr116 and Glu14) and Tyr121 are shown. $2F_O$-F_C electron density map contoured at 1.5 σ. Bonding interactions (yellow dashed line), distances (Å) and residues are annotated. Figures generated using PyMOL[144].

4.4. M062-Mav 3CS diffraction data

M062-Mav 3CS diffraction data were collected at the Advanced Photon Source synchrotron at Argonne National Laboratory, Illinois as described in Section 2.2.7.2. A representative diffraction image (Fig. 52) show there are strong reflections to ~ 4 Å, after which there are fewer reflection. An intense, diffuse ice ring band spanning 3.3 - 3.9 Å is shown, which is present in the background of all images between 0 and 120°, with its intensity increasing with the angle. The full dataset was 600 images, 0.2° oscillation. Data processing and reduction was carried out with RAPD, NECAT's automated data processing pipeline at APS. The dataset was initially determined to diffract to 3.21 Å, in space group P6.

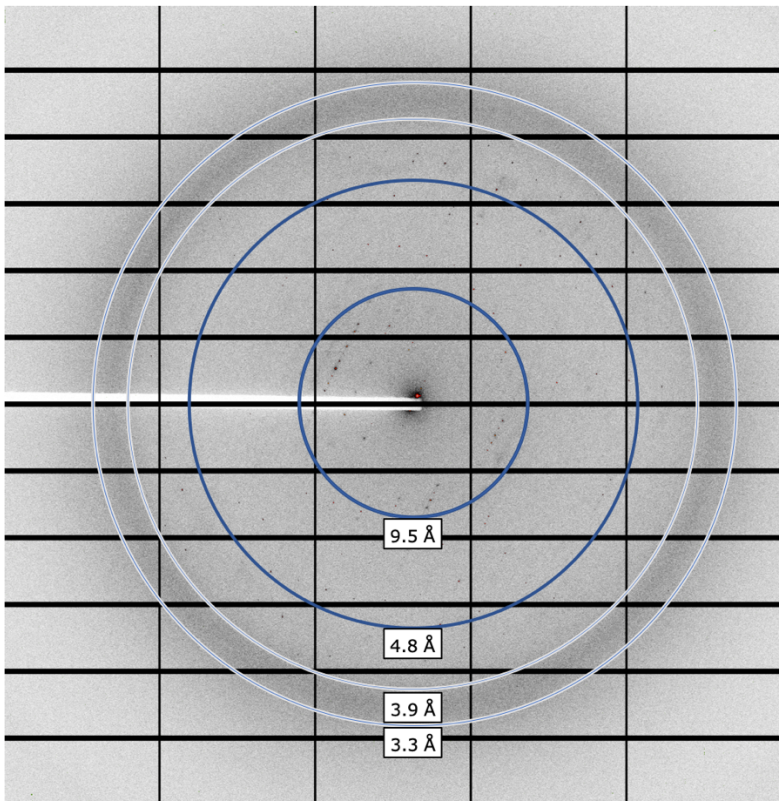
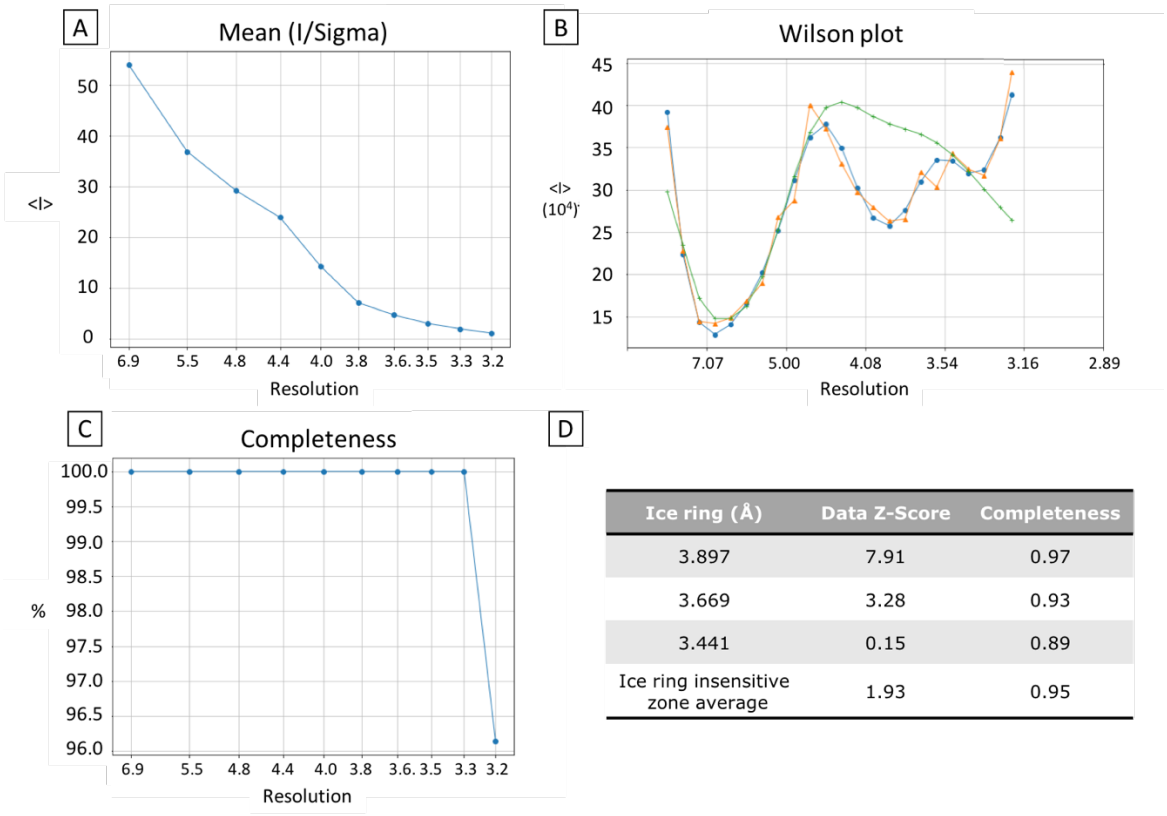


Figure 52: Diffraction data from M062-Mav 3CS. A representative diffraction image of M062-Mav 3CS is shown. Diffraction data images were collected in 0.2° intervals, 0.2 sec exposure, at 10% transmission, at the synchrotron Soleil. Resolution rings 9.5 Å and 4.8 Å shown in blue. A disperse ice ring is present between 3.9 Å and 3.3 Å, shown in grey. Images generated using Albula[136].

The data were assessed by PHENIX.xtriage for their quality, which reported it was unable to detect pathologies in the data, indicating no tNCS symmetry was present, they are not twinned, or anisotropic, the correct space group was assigned, and the resolution limit of ~3.2 Å is reasonable as per the signal:noise ratio (Fig. 53 (A)) and completeness (Fig. 53 (C)). Xtriage failed to identify the presence of ice rings, despite the reflection intensities within the ice ring sensitive shell having Z-scores significantly higher than those without (Fig. 53 (D)). Coinciding dips were also seen in the signal intensities in these resolution shells (Fig. 53 (B)).



Ice ring (Å)	Data Z-Score	Completeness
3.897	7.91	0.97
3.669	3.28	0.93
3.441	0.15	0.89
Ice ring insensitive zone average	1.93	0.95

Figure 53: Analysis of M062-Mav 3CS diffraction data. (A) Signal:noise shows the overall data quality are good. (B) The Wilson plot shows the signal intensity (binned: orange, smoothed approx.: blue) deviates from the predicted intensity (green) at the ice ring resolution shells. (C) Completeness for the data range is 100% >3.4 Å, after which it falls to 97.8%. (D) Ice ring sensitive zone completeness and standard score are shown in comparison to ice ring insensitive zone averages.

Xtriage revealed a high temperature factor for the data, with Wilson's B factor of 139.48 Å^2. This indicates the data reflect high flexibility and non-static protein and solvent position within the crystal lattice. In Figure 53 (A) the mean intensity at low resolution data is high, dropping off significantly at higher resolution, while the data intensity suffered during collection between ~4.5 – 3.6 Å (see Fig. 53 (B)). Overall analysis shows high quality data was obtained at low resolution, with higher resolution data of lower quality, while maintaining a mean (I/σ) greater than 1.5.

4.5. Structure solution of M062-Mav 3CS

The crystal structure of M062-Mav 3CS is accessible from the PDB via accession code 6S6O. The crystal structure of M062-Mav 3CS presented here contains sections in which the density allowed for modelling of the peptide backbone without sidechains, and in others regions residues are omitted, owing to the disorder inherent in the crystal lattice, seen in the diffraction data presented above (Section 4.4). This section describes the finalised model, drawing attention to the overall structural features, and evaluates the refined model statistically, including the limiting factors which prevented modelling the intact protein.

4.5.1. The crystal structure of M062-Mav 3CS

M062-Mav 3CS crystallised in the P6 space group, with unit cell parameters of $a = b = 148.8$, $c = 38.6$, $\alpha = \beta = 90$, $\gamma = 120$. Two molecules of M062-Mav 3CS arranged perpendicularly to one another form the ASU (see Fig. 56). Table 17 summarises the crystallographic data. Both proteins adopt the β sandwich fold characterised by the C7 family, with an R-work 0.2766 and R-free: 0.3087. Absences are features in the model, with Chain A comprising 154/161 residues, and chain B comprising 153/161 residues. Five residues are absent at the beginning of each chain, with the final 12 and 13 residues omitted in chain A and B respectively due to disorder in these residues (see Fig. 55). Note the high Wilson's B factor of 149.78 (see Table 17).

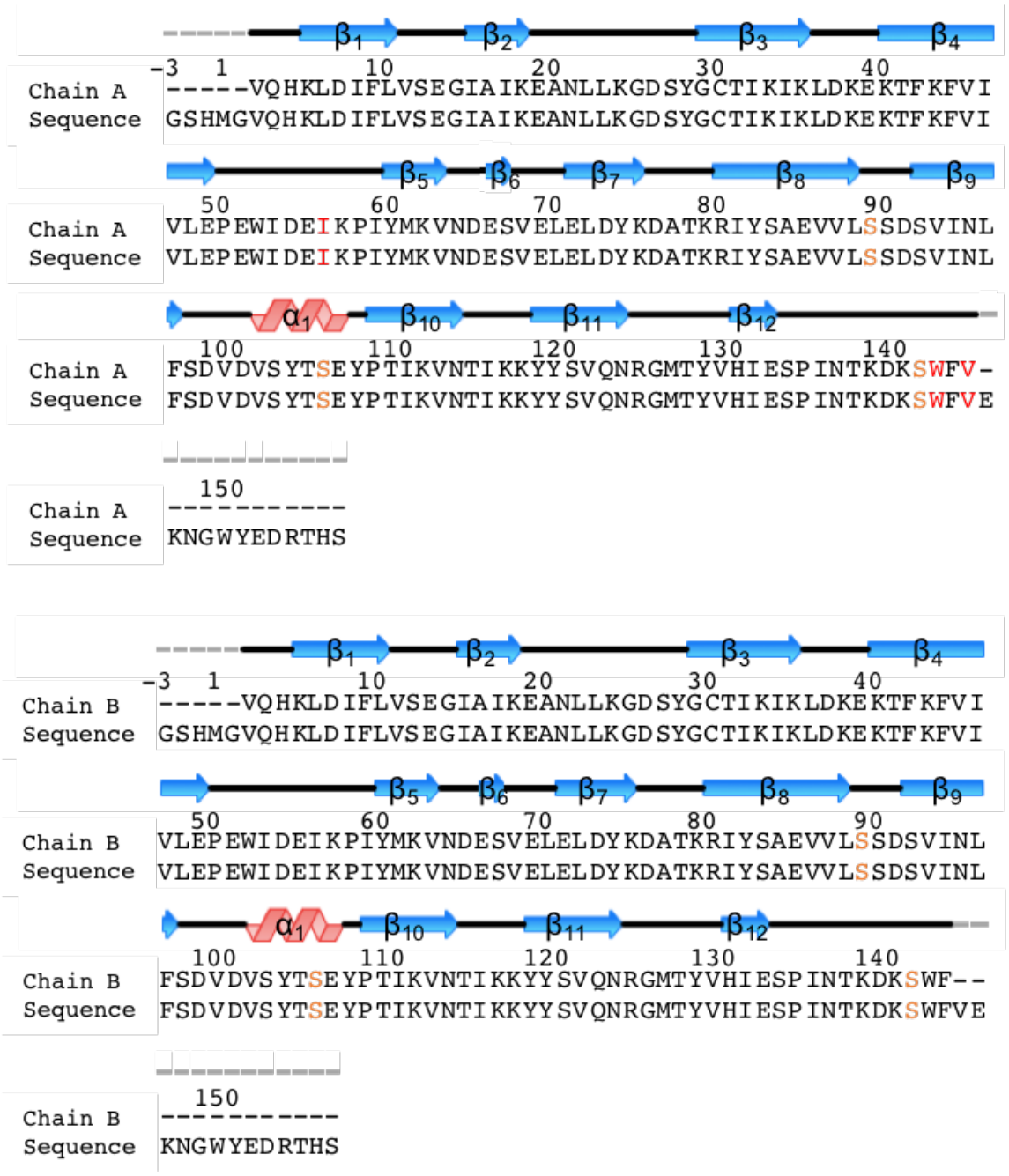


Figure 54: Alignment of the M062-Mav 3CS model and (-)M062-Mav 3CS sequence. The sequence of the refined structural model of M062-Lav 3CS aligned against the complete (-)M062-Mav 3CS construct sequence. Secondary structural features and residue number are annotated. Residues omitted from the final model shown as '-' (grey), Cys → Ser mutants are shown in orange, residues with their sidechain omitted in red. Alignment generated by PHENIX: Molprobity[138].

Figure 55: The crystal structure of M062-Mav 3CS. M062-Mav 3CS adopts the β sandwich fold of the C7 family. Two molecules form the ASU, shown as ribbon models in rainbow colour (N terminus (blue), C terminus (red)), with the secondary structures and termini labelled. Termini labels contain chain ID as subscript. Figure generated with CCP4mg[139].

Table 17: Data collection and refinement statistics for the crystal structure of M062-Mav 3CS, PDB: 6S6O

Wavelength	0.9791
Resolution range	74.42 - 3.21 (3.32 - 3.21)
Space group	P 6
Unit cell (a, b, c) (α, β, γ)	(148.8, 148.8, 38.6) (90, 90, 120)
Total reflections	56041
Unique reflections	8299 (782)
Multiplicity	6.5
Completeness (%)	99.71 (98.24)
Mean I/sigma(I)	27.0
Wilson B-factor	149.78
R-merge	0.052
R-meas	0.056
R-pim	0.022
CC1/2	1.000
Reflections used in refinement	8295 (782)
Reflections used for R-free	392 (35)
R-work	0.2766 (0.3899)
R-free	0.3087 (0.3772)
Number of non-hydrogen atoms	2325
macromolecules	2325
Protein residues	287
RMS(bonds)	0.003

RMS(angles)	0.64
Ramachandran favoured (%)	91.17
Ramachandran allowed (%)	7.07
Ramachandran outliers (%)	1.77
Rotamer outliers (%)	0.37
Clashscore	12.19
Average B-factor	160.17
macromolecules	160.17
Number of TLS groups	14

Statistics for the highest-resolution shell are shown in parentheses. R-free was calculated using 5% of data.

4.5.2. Phase solution of M062-Mav 3CS by molecular replacement using M062-Lau 3CS as a search model

M062-Mav 3CS was solved using the completed M062-Lau 3CS crystal structure obtained in this study as an ensemble for MR with PHENIX: Phaser[111]. Phaser's best solution placed 2 molecules in the ASU, with an LLG of 171, and TFZ-score 14.3, indicating that a reasonable model to fit the data had been found. The MR model placed 134 residues in chain A and B which were then modified with Coot[112], and refined using PHENIX.refine[113] to the final structure presented in Fig. 55.

4.5.3. Structural refinement of M062-Mav 3CS

4.5.3.1. Refinement of the β 4 - β 5 loop

The electron density maps initial calculated by the MR solution enabled modest alteration of the placed molecules. The overall fold of M062-Mav 3CS was evident, but the initial MR failed to place 27 residues in each chain, spanning from residue 3 – 143, with the loop between β 4 and β 5 absent. The loop density was used to build the missing 5 residues (Fig. 56).

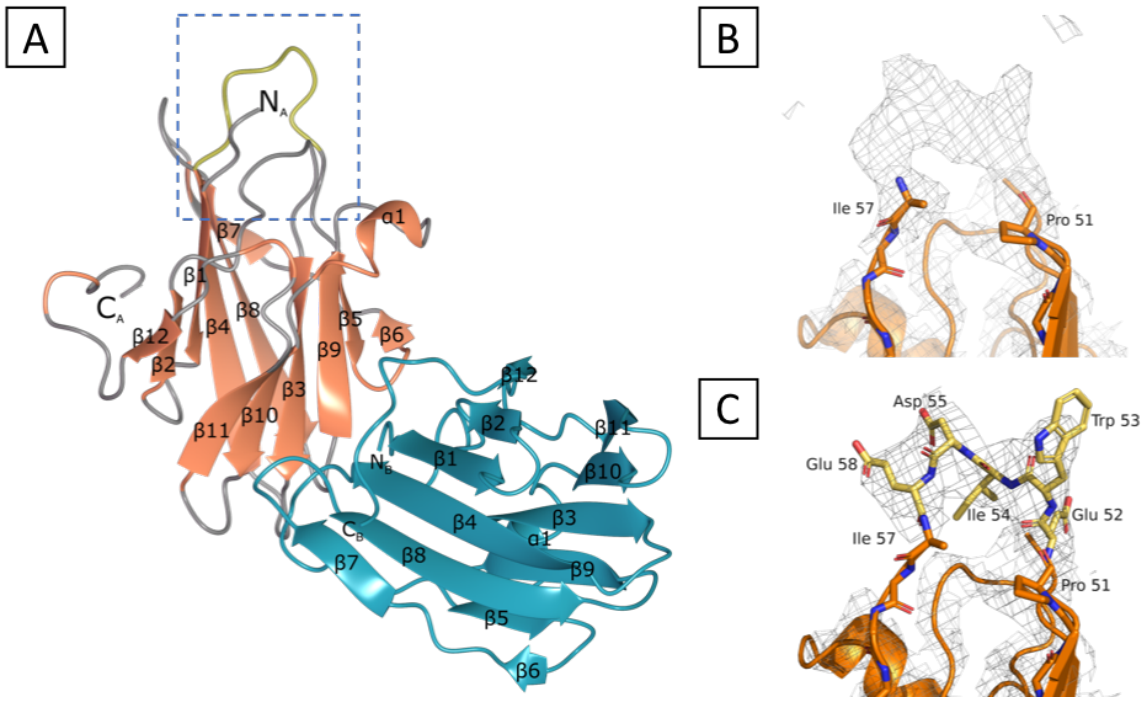


Figure 56: Refinement of missing M062-Mav 3CS loop. Refinement of chain A shown. (A) Ribbon model overview of the completed structure highlights the position of the missing loop (yellow) between β strands 4 and 5, chain A (orange), B (cyan), secondary features annotated. (B) A view of the loop region in the blue square in (A). Continuous electron density ($2F_O$-F_C: 1 σ) links Pro 51 and Ile 57 (missing sidechain). (C) Residues 52 – 56 (yellow) modelled into the electron density of the proposed model of the M062-Mav 3CS $\beta_{4.5}$ loop.

Loop$_{4,5}$ is the largest of 3 extended loops at the N terminal side of the β sandwich, dubbed the 'trimolecular claw' by Meng *et al.*[29]. The loop occupies different morphologies in chain A vs chain B, shown in Figure 57. The electron density of chain A and B (Fig. 57 (D)) at 1.5 σ indicate the chains do represent unique structures, as highlighted by the superposition (Fig. 57 (A – C)).

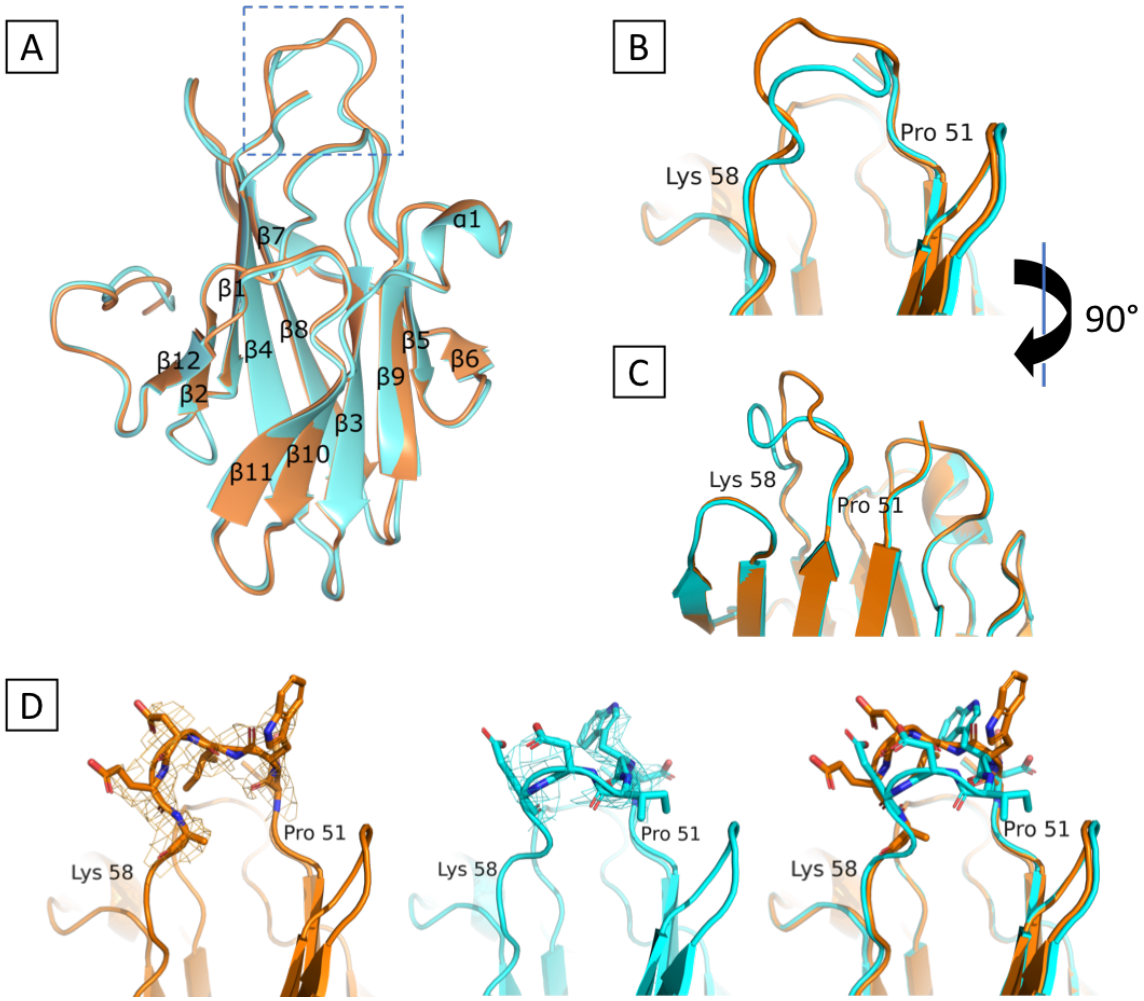


Figure 57: Comparison of loop$_{4,5}$ from chain A and B of M062-Mav 3CS. (A) An overview of the superposition of chains A (orange) and B (cyan) in ribbon view, with secondary structures annotated. Loop$_{4,5}$ connecting β 4 - β 5 boxed in blue. (B) Enhanced view of loop$_{4,5}$ from (A) show the loops are structurally distinct, (C) view of loop$_{4,5}$ from (B) rotated 90°. (D) Loop$_{4,5}$ sidechains of chain A, B and a superposition are shown, with $2F_OF_C$ map contoured at 1.5 σ. Figure generated with PyMOL[144].

4.5.3.2. N terminal electron density

Refinement of the M062-Mav 3CS structure was difficult in certain regions of the model, for example, where the electron density had limited resolving capability, especially in flexible zones or closely packed areas between β sheets. With $loop_{4,5}$ complete, there were 5 residues missing from the N terminus, and 18 from the C terminus.

The N terminus of each chain is oriented towards its ASU partner (or symmetry mate). No electron density is observed at the N terminus prior to Val3 in either chain (Fig. 58). Attempts to build into the minimal density with a polyGly chain were unsuccessful. Polder and FEM maps didn't reveal any density in the region adjacent to the N terminus which allowed for further refinement of the N terminus past the initial MR model.

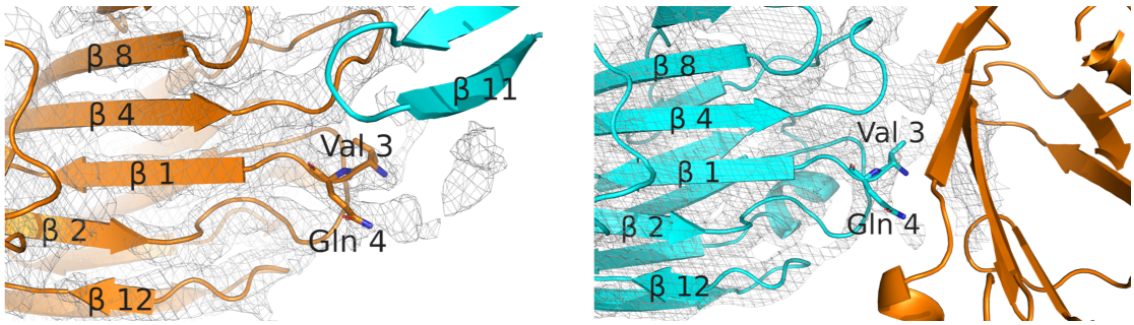


Figure 58: N terminal electron density of M062-Mav 3CS. The electron density ($2F_O$-F_C, 1.0 σ) within 10 Å of Val 3 for chain A (orange) and B (cyan) are shown. Secondary structures and residues 3-4 annotated. Figure generated with PyMOL[144].

4.5.3.3. Refinement of the C terminus

The initial model produced by MR built the C terminus to S143 of each chain. Weak electron density continued past S143, and it the C terminus was built to V146 (with its sidechain omitted) in chain A, and F145 in chain B. The structures of C7 and M064 terminate at approximately equivalent positions[29]. The electron density is weak following V146. The submitted PDB Mav chains terminate at V146 and F145 respectively. However, the weak C terminus electron density was experimentally built into, in order to account for the remaining 15 C terminal residues.

4.5.3.4. The M062-Mav 3CS C terminal tail

The crystal structure of M062-Lau 3CS revealed the C terminal tail extends to make contacts with its symmetry mate, nestled between β strands 11 and 12. While there are 2 copies of M062-Mav 3CS in the ASU, the chosen copies are oriented with their C termini at substantial distances from one another (see Fig. 55). Symmetry mates (Fig. 59 (A)) show copies of chain A and B which are oriented with the C termini aligned in close proximity to the $\beta_{11,12}$ end of the β sandwich of the partner chain are present (see Fig. 59 (B)). This arrangement closely resembles the ASU of M064 (5CZ3). Inspection of the electron density at β 11 and 12 in both chains show the C terminus may in fact occupy a position similar to that seen in the Lau crystal structure (Fig. 59 (C)).

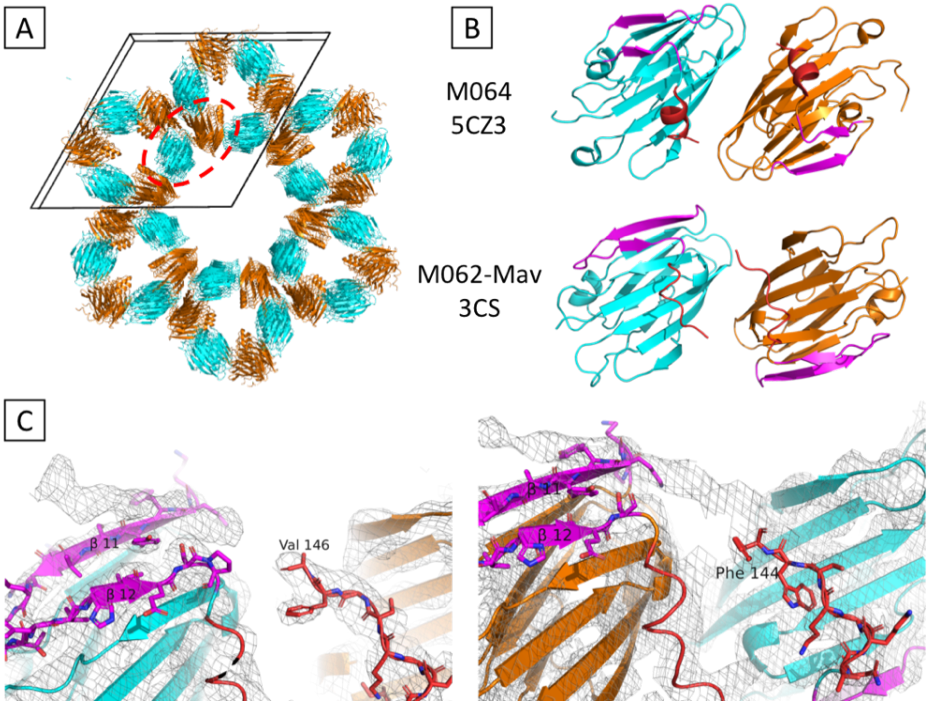


Figure 59: The C terminal tail of each M062-Mav 3CS chain is oriented towards the β11 - β12 of its complimentary chain. (A) Circled in red are symmetry mates of M062-Mav 3CS (chain A: orange, chain B: cyan) with complimentary C terminal:β11 – β12 orientation. (B) The crystal structures of M064 and M062-Mav 3CS show the C termini (red) in close proximity to the β11 - β12 strands (magenta) of the opposite chain. (C) The C terminal:β11 – β12 interface electron density ($2F_O$-F_C, 1.0 σ) shows the C termini likely extend to make contact with the β11 – β12 groove. Figure generated using PyMOL[144].

Due to the low resolution of the data and the high B factor, it wasn't possible to refine the C terminal tail to the same extent as in the M062-Lau 3CS structure, resulting in its omission in the deposited PDB model. The data clearly show that there is a peptide-shaped region of electron density sitting in the $\beta_{11,12}$ groove for both chains. Modelling of a peptide into the niche between $\beta_{11,12}$ shows the potential for the formation of β sheet style hydrogen bonds (see Fig. 60 (A)). A difference (F_O-F_C) map shows a region of positive density in close proximity to Tyr121, and Pro136, and a pocket at $\beta_{11,12}$ in which Tyr152 may reside (see Fig. 60 (B)) similar to the M062-Lau 3CS structure.

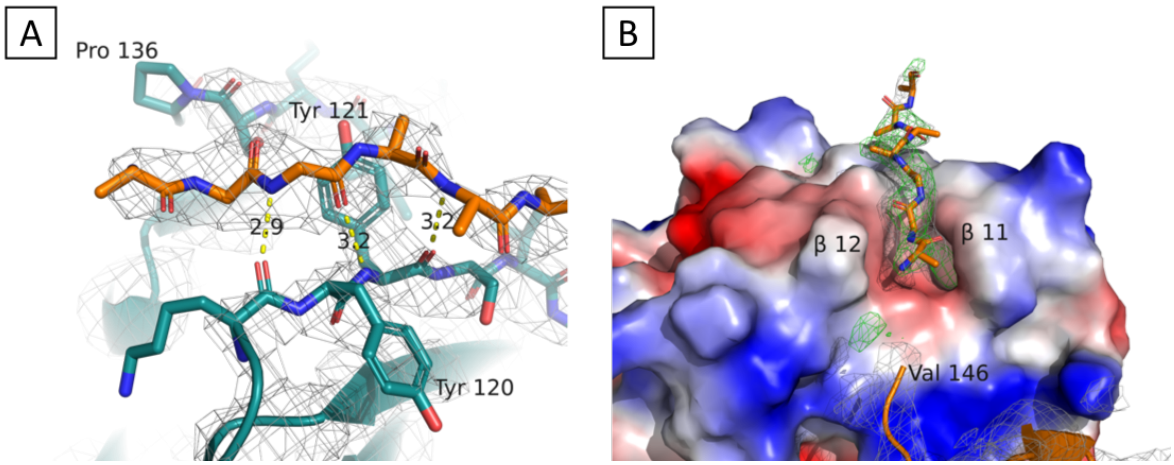


Figure 60: M062-Mav 3CS may form contacts with a neighbour via a C terminus:$\beta_{11,12}$ groove interaction. (A) Backbone hydrogen bonding interactions (yellow, distance (Å)) between a polyAla peptide (orange) built into the electron density ($2F_O$-F_C, 1.5 σ) proximal to the C terminus and β_{11} (cyan) are shown. (B) A hydrophobic pocket can be seen in the surface electrostatic potential map of chain B at the interface of the peptide and β_{11}. A difference map (F_O-F_C, 3 σ) shows regions of positive density along the peptide backbone, and in the pocket. Figures generated using PyMOL[144].

4.6. Structural comparison of members of the C7 family

In this section, the structure of the Lau and Mav variants of M062 are compared to one another, and the other members of the C7 family. Attention is drawn to the conservation of the globular fold between crystal structures; regions with high structural variance; and the possibility of a conserved dimerization interface in MYXV members. The structure of trimolecular claw, and residues which mediate interaction of the C7 family with SAMD9 are investigated by superimposing the crystal structures of the C7 family.

4.6.1. The β sandwich fold is a unique, highly conserved fold found in diverse C7 family members

Superpositions of the M062 crystal structures and the existing C7 family structures of MYXV protein M064: 5CZ3, which does not act as a host restriction factor, and VACV C7: 5CYW reveal that although the sequence identities of the C7 family may be low (the resolved M062 and C7 structures share sequence identity of 29.3%), the overall C7 β sandwich fold is highly conserved (see. Fig. 62). The RMSDs of the aligned structures are all <1.5 $Å^2$, and the globular β sandwich, which features a core of highly conserved hydrophobic residues (see Fig. 61 is structurally conserved among all of the C7 family structures (see Fig. 62).

Furthermore, submission of the C7 structures to the PDBeFold secondary structure matching (SSM) tool shows that there are no structures in the PDB which possess structural similarities of significance to any C7 family structures (see Table 18). This result supports the reports of Meng *et al*[29] and Liu *et al.*[11], which found no structural-, or sequence-related homologues of the C7 family outside of the *Chordopoxviriniae*.

While the globular core of the structure is conserved, forming the β sandwich fold, some obvious differences are present, most evident at the N and C termini, and in the loop regions at the N terminal pole of the β sandwich. The C terminus and the role of the tail interaction is a novel structural finding in the C7 family. Dissection of these variations reveals potential effects of the C terminal tail, and the 79 locus on the structure of M062 in light of previous work.

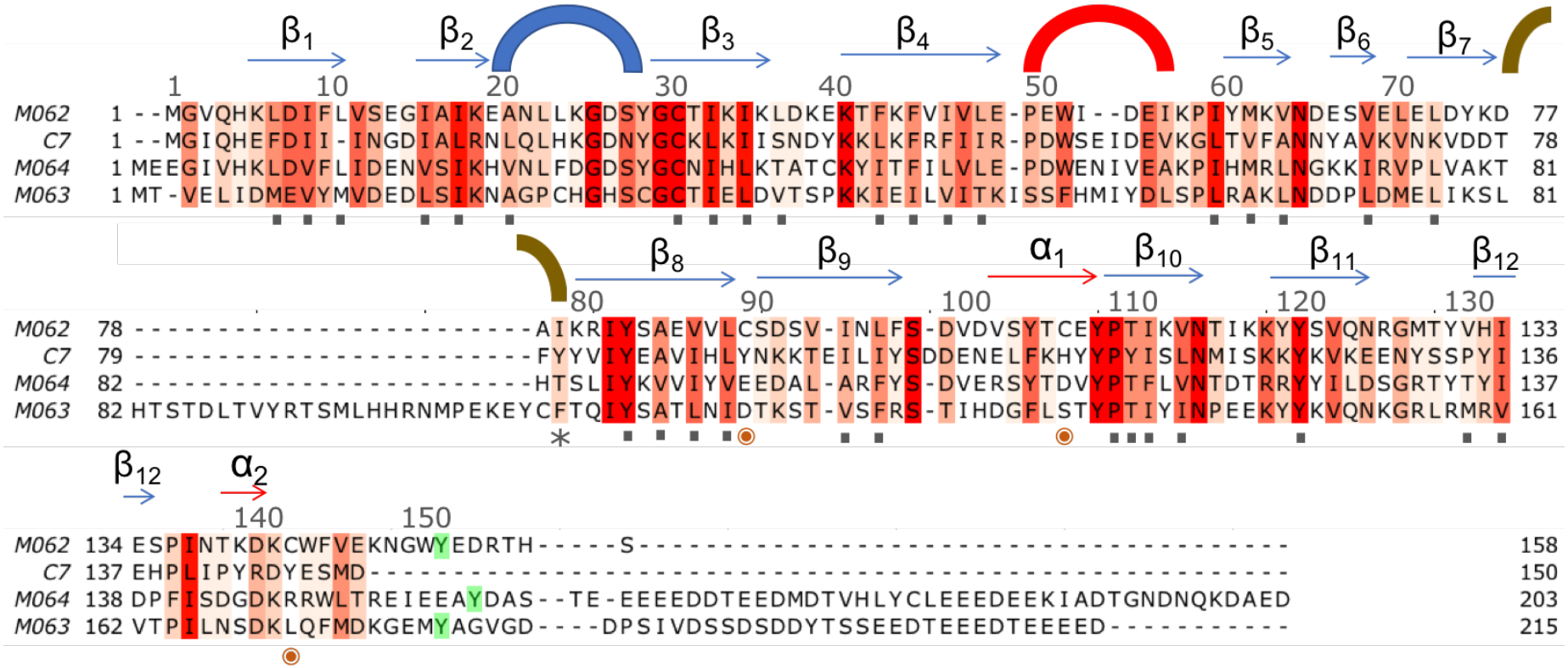
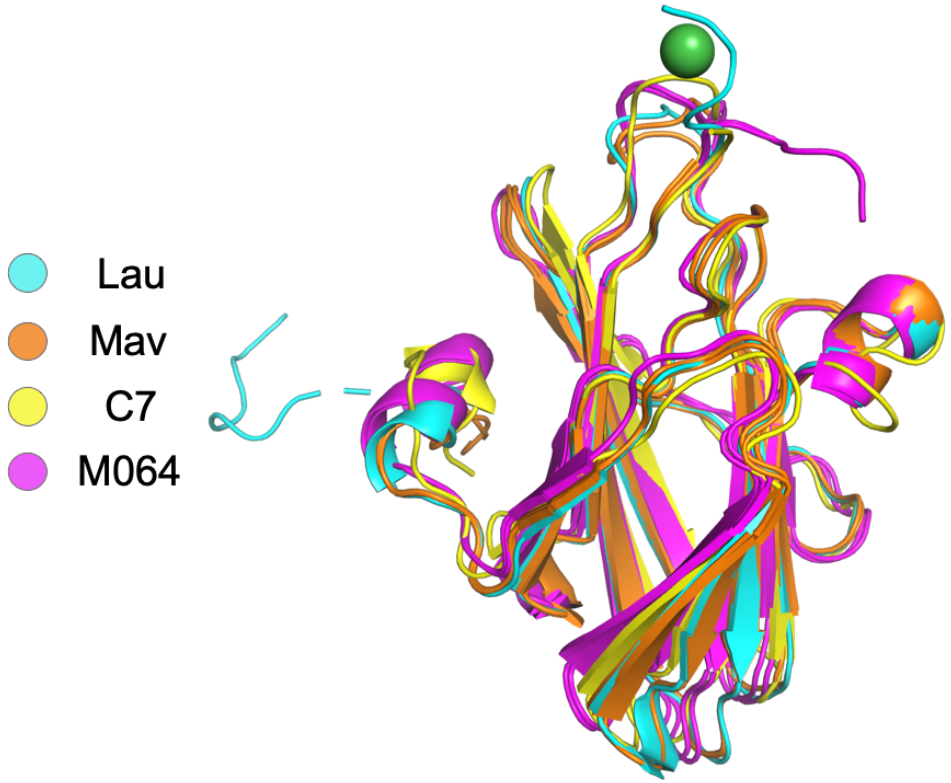


Figure 61: MYXV and VACV C7 family sequence alignment. Sequences are coloured by consensus, as calculated by ClustalOmega[25]. Y152 and its equivalent residues in M063 and M064 are coloured green. Annotated above the alignment are the M062 secondary structure, the position of the basic (blue), acidic (red) and neutral (brown) loops of the trimolecular claw, and sequence position. Annotations: residues which form the hydrophobic core of M062-Lau 3CS (■), the 79 locus (*) and the Cys → Ser mutants (◉). Figure generated with Aline[26].



Seq. ID	PDB: Chain	Protein	RMSD (Å²) Lau	Mav
29.3%	5CYW: B	C7	1.26	1.31
38.5%	5CZ3: A	M064	0.97	0.91
38.5%	5CZ3: B	M064	0.93	0.90
98.0%	6S6O: A	M062-Mav	0.65	0.00
99.0%	6S6O: B	M062-Mav	0.68	0.68

Figure 62: Superposition of the C7 family crystal structures. A structure-based alignment of ribbon models of the C7 family show the core β sandwich is highly conserved. Models are coloured by crystal structure (legend, left). The table shows the sequence (%) identity and RMSD (Å²) of the C7 family crystal structures compared to the M062 3CS structures. RMSD calculated by PDBeFold[145]. Sequence identity calculated by alignment of the deposited structure FASTA files, % sequence identity against M062-Lau 3CS, alignment by Clustal Omega[25]. Figure generated using PyMOL[144]

	Rank	P-score	RMSD	Sequence Identity (%)	Target
C7 5CYW	1	19.35	1.335	25.53	5cz3:B
	2	19.53	1.348	25.53	5cz3:A
	3	0	3.713	06.93	1cyw:A
	4	0	3.836	06.93	1cyx:A
	5	0	4.271	04.59	6eay:H
M064 5CZ3 A	1	42.38	0.211	100	5cz3:B
	2	0	3.465	02.41	2fd6:L
	3	0	3.948	07.87	3gi9:H
	4	0	3.98	07.87	3gi9:H
	5	0	3.797	08.33	2fd6:H
M062-Lau 3CS	1	24.92	0.934	41.84	5cz3:B
	2	25.67	0.97	41.84	5cz3:A
	3	18.17	1.262	32.62	5cyw:B
	4	0	3.728	08.91	1cyw:A
	5	0	3.932	08.82	1cyx:A
M062-Mav 3CS	1	26.67	0.902	42.86	5cz3:B
	2	27.56	0.914	42.86	5cz3:A
	3	21.9	1.306	30.66	5cyw:B

Table 18: Structural alignment results of the C7 family structures. The crystal structures of C7 (5CYW), M064 (5CW3, chain A), M062-Lau 3CS and M062-Mav 3CS were submitted to the PDBeFold server[145]. The top 5 results (ranked by P-score) are shown (only 3 matches are identified for M062-Mav 3CS). The target of comparison's PDB accession code and chain are listed in the 'Target' column. P-score results of 0 are obtained for any structures outside of the C7 family, indicating that any structural similarity is not significant. P-score: -Log10 P, with P-scores <3 indicating the structural similarity of the target to the query structure is statistically insignificant.

4.6.2. Structural variance of the 'Trimolecular Claw'

While the overall C7 fold is conserved in the crystal structures of C7 and its homologues, some of the key loops identified by Meng *et al.*, dubbed the 'trimolecular claw'[29] vary significantly in their conformation, and sequence identity between the homologues.

Of the 3 loops of the trimolecular claw, the Lau and Mav M062-3CS structures are notably similar in the basic loop $\beta_{2,3}$, which aligns consistently between all chains of the known C7 structures (see. Fig. 63 (A)). The acidic and neutral loops, however, particularly $\beta_{4,5}$, feature the most prominent variation between C7 structures (see Fig. 63 (B)). These 2 loops feature the highest relative B factors of the trimolecular claw loops in both C7 and M064, indicating a higher degree of flexibility in these loops[146].

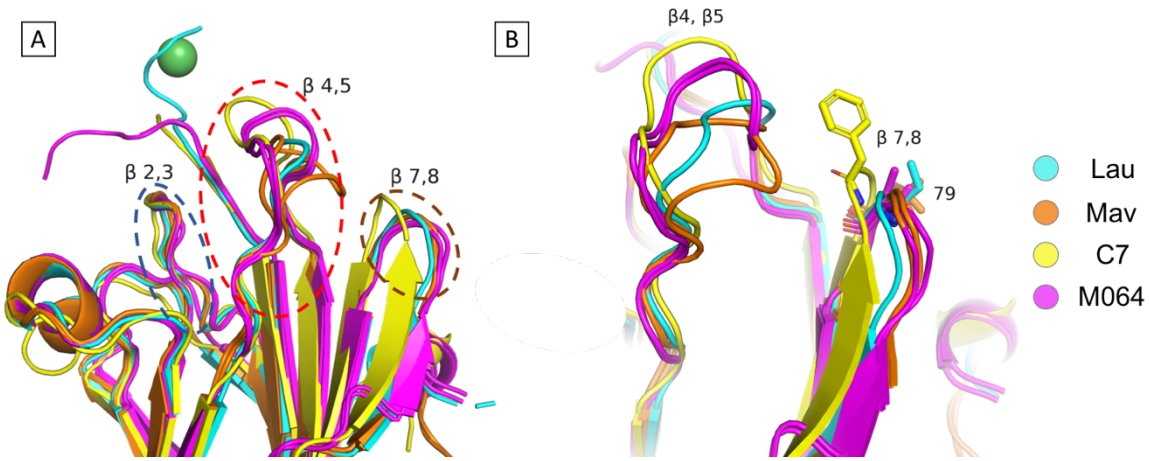


Figure 63: The 79 locus and structural variation of the loops of the trimolecular claw. (A) Ribbon models of the C7 family crystal structures (legend, right) are superimposed, showing the 3 loops of the trimolecular claw proposed to mediate interaction with SAMD9 circled: basic loop $\beta_{2,3}$ (blue), acidic loop $\beta_{4,5}$ (red) and neutral loop $\beta_{7,8}$ (orange). (B) Residues at the 79 locus on loop $\beta_{7,8}$ are shown. Figures generated using PyMOL[144].

4.6.2.1. Position 79 and the conformation of loops $\beta_{4,5}$ and $\beta_{7,8}$

Position 79 of M062 where MYXV Lau and Mav strains vary, has been identified by co-IP to act as a locus for interaction with SAMD9 (see Fig. 16). Position 79 lies on the $\beta_{7,8}$ loop, and is conserved in the C7 family, occupied by a hydrophobic residue (see Fig. 2). A triple mutant of the equivalent and adjacent residues in C7 (Tyr78, Phe79, Tyr80) lead to the loss of SAMD9 binding[29]. The structures of the Mav and Lau variants of M062 show both Ile79 (Lau) and Thr79 (Mav) co-align, and the overall conformation of the $\beta_{7,8}$ loop resembles the M064 structure more closely than that of the C7 $\beta_{7,8}$ loop (see Fig. 64 (A)).

Three of the five key residues identified by Meng et al.[29] are within the $\beta_{4,5}$ loop, which is located directly beside loop $\beta_{7,8}$ and position 79 (see Fig. 64). The conformation of loop $\beta_{4,5}$ appears to be the most variable of the 3 loops of the trimolecular claw. While the position of D51 and E57 equivalent residues are similar between structures, residues 52 – 56 are highly diverse spatially, with each structure's loop adopting a unique conformation (see Fig. 64 (B)). M062 is missing 2 residues from this loop compared to the other members of the C7 family (see Fig. 61), resulting in a shorter loop which extends less than the $\beta_{4,5}$ loop of C7 and M064 (see Fig. 64 (A)).

Position 79 in C7 acts in conjunction with its adjacent residue Y80 to form a triad of aromatics with W52, forming a bridging interaction between loops $\beta_{4,5}$ and $\beta_{7,8}$ (see Fig. 64 (C)). However, position 80 is not conserved in the C7 family (see Fig. 2), and the MYXV M062 Lau and Mav point mutation at position 79 alone is sufficient to impact SAMD9 binding (see Fig. 16).

Chain A and B of M062-Mav 3CS adopt unique conformations of Loop $\beta_{4,5}$, one of which is the most distal from C7's $\beta_{4,5}$ loop (see Fig. 64 (B)). This may be a result of the I79T mutant influencing interloop interactions, removing a hydrophobic connection between W53 and the neutral loop. Comparison of the surface electrostatic potential of C7 and M062-Mav 3CS shows the loop interaction is lost in M062-Mav 3CS, and both the acidic and neutral loops are less well defined than in C7 (see Fig. 64 (D)).

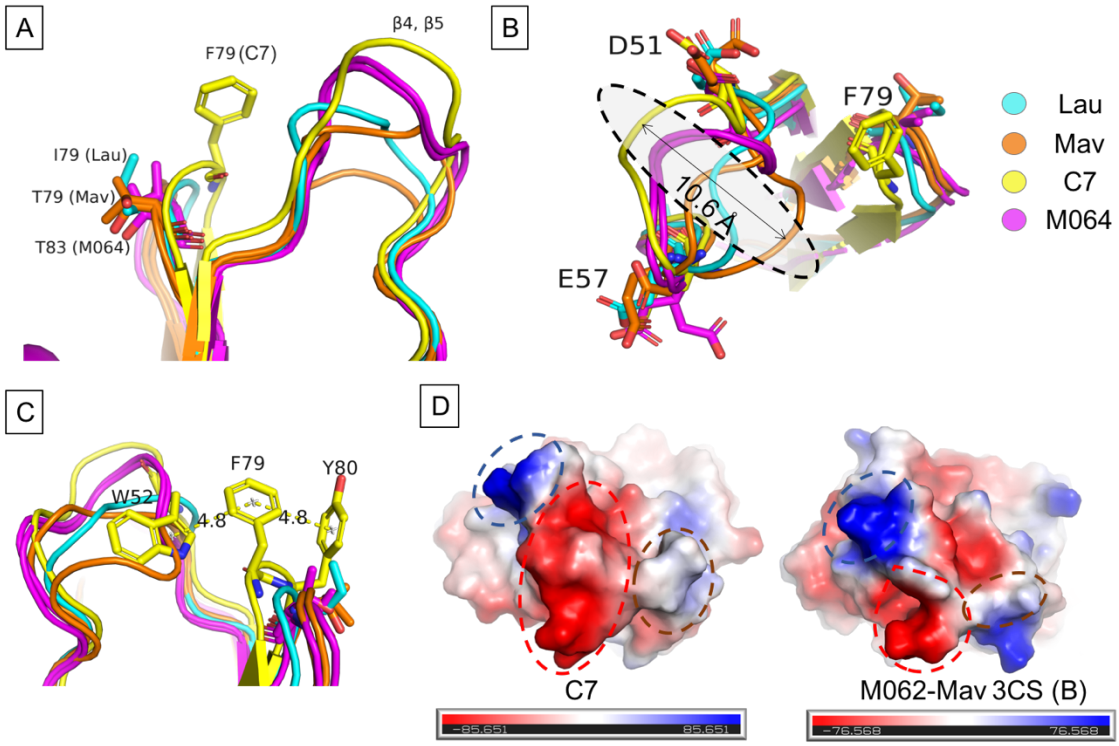


Figure 64: The trimolecular claw varies between C7 family structures most at loop $\beta_{4,5}$. (A) Superposition of the C7 family structures (legend, right) shows loop $\beta_{4,5}$ (annotated) of C7 adopts a distinct conformation in comparison to the MYXV members. Position 79 equivalent residues are shown (stick bonds, annotated) with structure annotated. (B) An overhead view of loops $\beta_{4,5}$ and $\beta_{7,8}$ is shown, with conserved residues which mediate interaction with SAMD9[29] annotated with C7 sequence identifiers. Loop $\beta_{4,5}$ is conformationally diverse between C7 structures, with a maximum variance of 10.6 Å between C7 and M062-Mav 3CS chain B. (C) Position 79 of C7 is shown to interact with loop $\beta_{4,5}$ via a triad of aromatic groups (annotated). F79, Y80 and W52 link the loops via a series of π bonds. (D) Surface electrostatic potential maps of C7 and chain B of M062-Mav 3CS show the basic (blue), acidic (red) and neutral (brown) loops vary, with the acidic and neutral loops of M062-Mav 3CS less distinct. Figures generated using PyMOL[144].

4.6.3. The conformation of the $\beta_{11,12}$ groove and C terminus may determine protein-protein interaction between C7 homologues

The crystal structures of M062 3CS in this study indicate that a hydrophobic pocket at the $\beta_{11,12}$ groove is occupied by a conserved C terminal tail residue. MSA of the C7 family shows that the C terminus of C7 proteins is not conserved among poxviruses (see Fig. 2). By investigating the conformation of the $\beta_{11,12}$ groove in the known C7 homologues, distinct structural features are identifiable which appear to correlate with a sequence variation between homologues that lack a C terminal tail.

4.6.3.1. Formation of the hydrophobic pocket is dependent on the orientation of the conserved β_{11} tyrosine side chain

The $\beta_{11,12}$ groove features a conserved Tyr residue at position 121 (see Fig. 65). The MYXV homologues orient the sidechain of this Tyr perpendicular to the β_{11} backbone, creating an open channel between β strands 11 and 12 (the first 2 residues of β_{11} are oriented away from β_{12}) (see Fig. 65 (A_1 -C_1). This Tyr adopts an alternate conformation in C7, however, positioned with the sidechain oriented towards the start of β_{11}, occupying the hydrophobic pocket location seen the MYXV homologues (Fig. 65 (D_1).

The crystal structure M062-Lau 3CS differs from that of M062-Mav 3CS, C7 (5CYW) and M064 (5CZ3) by the inclusion of a region of the C terminal tail that is difficult to model in M062-Mav 3CS (see Sections 4.3.5 and 4.5.3.4). Interestingly, a region of unaccounted for electron density is present in M064 (5CZ3). Located at the position of the hydrophobic pocket, a difference map (F_O-F_C, 3 σ) shows electron density visible between Y126 and P139, as seen in the M062 3CS structures (see Fig. 65 (C)). With the C terminus unaccounted for in the M064 structure[29], and with the M064 chains of the ASU aligned identically to M062-Mav 3CS with the C termini directly opposite the $\beta_{11,12}$ groove (see Fig. 59 (B)), it is feasible that the C terminus of M064 interacts in the same way as seen in M062 3CS.

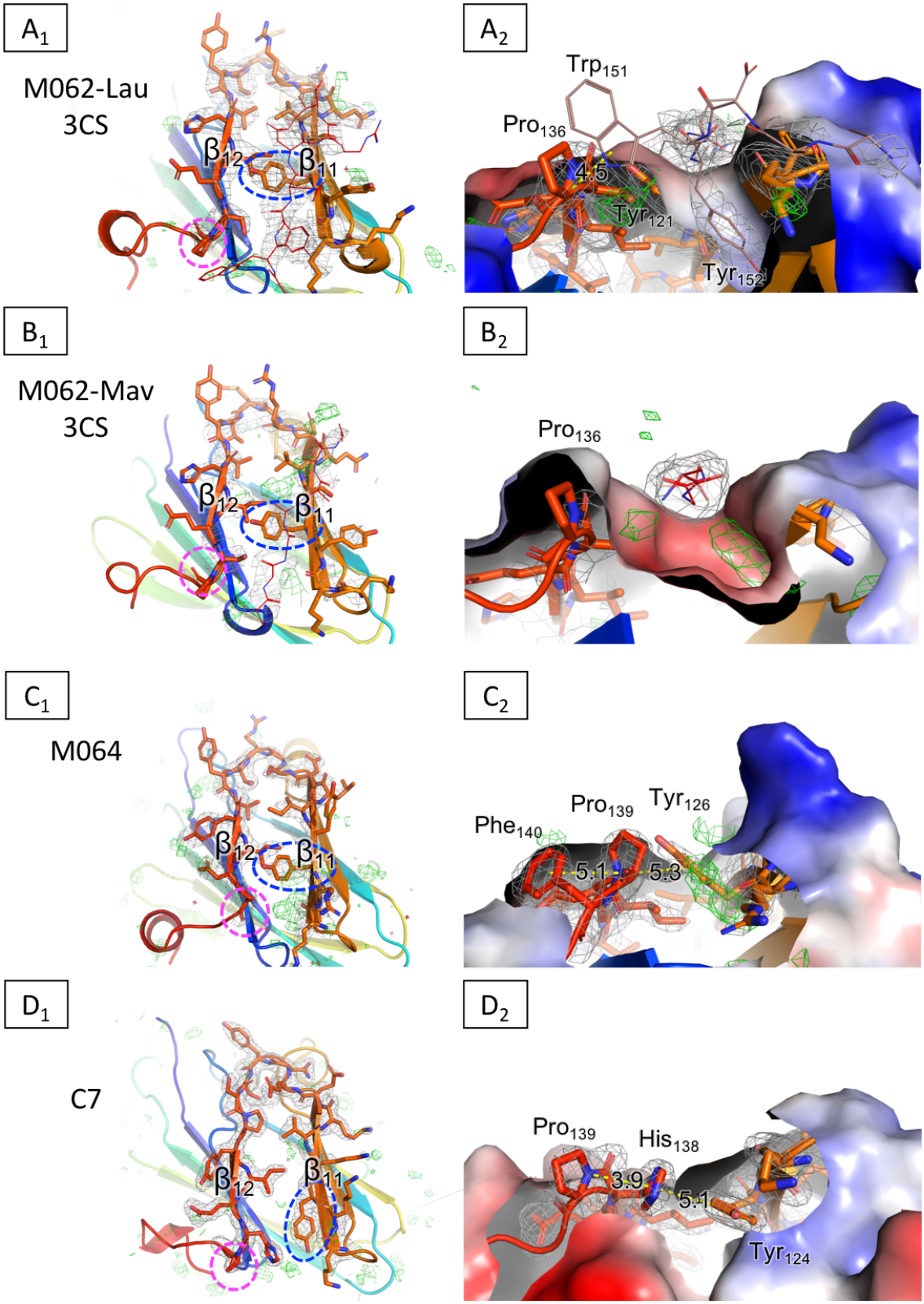
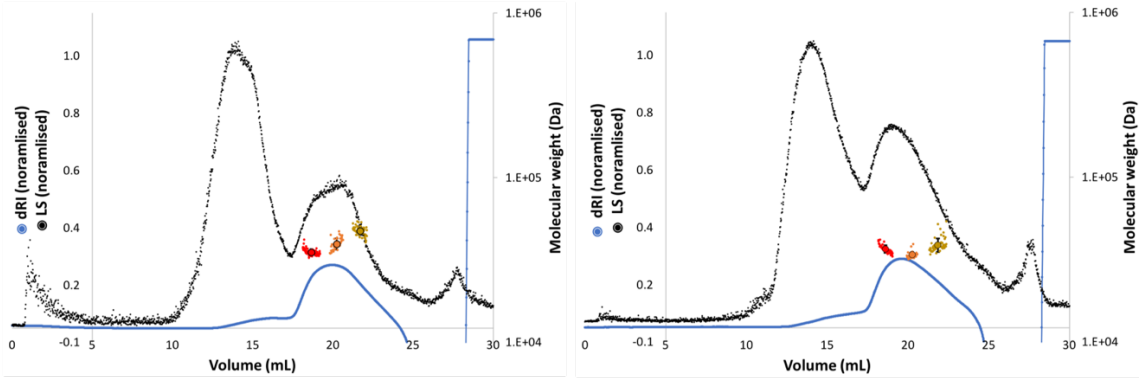

A1
M062-Lau
3CS
β12
β11
A2
Trp151
Pro136
4.5
Tyr121
Tyr152
B1
M062-Mav
3CS
β12
β11
B2
Pro136
C1
M064
β12
β11
C2
Pro139
Tyr126
Phe140
5.1
5.3
D1
C7
β12
β11
D2
Pro139
His138
3.9
5.1
Tyr124

Figure 65: A hydrophobic pocket between β strands 11 and 12 in MYXV C7 family proteins is occluded by a His residue in C7. (1) An overview of the $\beta_{11,12}$ groove is shown for the C7 homologue structures, with β_{11} and β_{12} labelled. Cartoon structures are shown with $\beta_{11,12}$ residues as sticks. Y121 (blue) and P136 (magenta) and their equivalent residues in C7 homologues are circled. (2) A side-on view shows the $\beta_{11,12}$ interface, and hydrophobic pocket, with a cutaway surface electrostatic map revealing high electron density occupying the pockets of MYXV C7 homologues (occupied by Y152 in the Lau M062 3CS pocket). C-H···π interactions between aromatic side chains with Pro Cα are shown (yellow, Å). Electron density maps: $2F_O$-F_C, 1.5 σ (grey) and F_O-F_C difference map, 3 σ (green) are shown around $\beta_{11,12}$.

4.6.3.2. The Lau and Mav variants of His-M062 WT appear dimeric in solution

Multi-angle light scattering (MALS) experiments were carried out to determine if the interaction between individual chains of the M062 3CS structures is measurable in solution. M062-Lau WT and M062-Mav WT were recombinantly expressed, and their molecular weights in solution calculated by SEC-MALS (see Fig. 66). A single molecule of His-M062 WT is 20.6 kDa, and the SEC-MALS calculated Mw indicate both the Lau and Mav variants of M062 are approximately dimeric in solution, with Mw ranges (Mw ± SEM) Lau: 33.7 – 50.6 kDa and Mav: 32.4 – 42.5 (see Fig. 66 (table)).



His-M062-Lau WT	1	2	3
Mn molecular weight (kDa)	35.3	39.7	47.6
SEM (%)	4.6	7.5	6.3

His-M062-Mav WT	1	2	3
Mn molecular weight (kDa)	36.7	33.8	38.8
SEM (%)	5.3	4.2	9.5

Figure 66: Multi-angle light scattering estimates Lau and Mav His-M062 WT are dimeric in solution. 0.5 mg His-M062 WT Lau (left) or Mav (right) were run on an sx75 10/300 column at 0.7 mL/min, in 150 mM NaCl, 20 mM HEPES pH 7.4, 1 mM DTT.. Elution profile (solid line, blue) recorded by the RI detector is presented together with mass calculations for collected peaks (individual data points). The table shows the average calculated Mw and % SEM. Lau and Mav His-M062 WT are likely dimeric in solution (monomer Mw: 20.6 kDa). The initial high LS signal (black) is due to the presence of high Mw aggregates which elute prior to the non-aggregated protein.

4.6.3.3. Formation of a C-H···π interaction with the conserved cis-Pro residue at the end of β_{12} is preserved in all C7 homologues

All of the structures of the C7 proteins consistently have an aromatic side chain located beside the Pro at the end of β_{12}. This Pro is always present as a cis isomer, and is stabilised by a C-H···π interaction between the Pro's C_α and the nearby aromatic side chain (see Fig. 65 (2)). In M062-Lau 3CS, this C-H···π interaction is formed between Pro136 and Trp151 of the C terminal tail from a neighbouring molecule of M062 (see Fig. 65 (A_2)). While not modelled explicitly in M062-Mav 3CS the electron density in the hydrophobic pocket and adjacent to Pro136 indicate both M062 variants feature the same interaction (see Fig. 65 (B_2)).

M064 is unique among the C7 family as it has an aromatic residue *after* its cis-Pro, F140 (see Fig. 65, 67), and appears to sandwich P139 between F140 and Y126 (see Fig. 65 (C_2). In contrast to the M062 structures where the aromatic residue is located in the C terminal tail, C7 has an aromatic residue *before* its cis-Pro, His138 which takes part in the C-H···π interaction, and also in an edge-to-face π-π interaction with Y124 (see Fig. 65 (D_2)).

4.6.3.4. C7 homologues that lack a C terminal tail have a conserved aromatic residue prior to the β_{12} cis-Pro

The amino acid composition of β_{12} appears to covary with the presence of a C terminal tail: *i.e.* in homologues which lack an extended C terminus, an aromatic residue is consistently found immediately prior (n-1) to the cis-Pro position (see Fig. 67, blue). C7 homologues with longer C termini consistently feature an aromatic residue followed by a hydrophobic residue 15-28 residues downstream of the cis-Pro (see Fig. 67: cis-Pro: grey; aromatic: green; hydrophobic; magenta). The M062 3CS structures hint at a conserved C terminal tail interaction with the cis-Pro and $\beta_{11,12}$ hydrophobic pocket at this location.

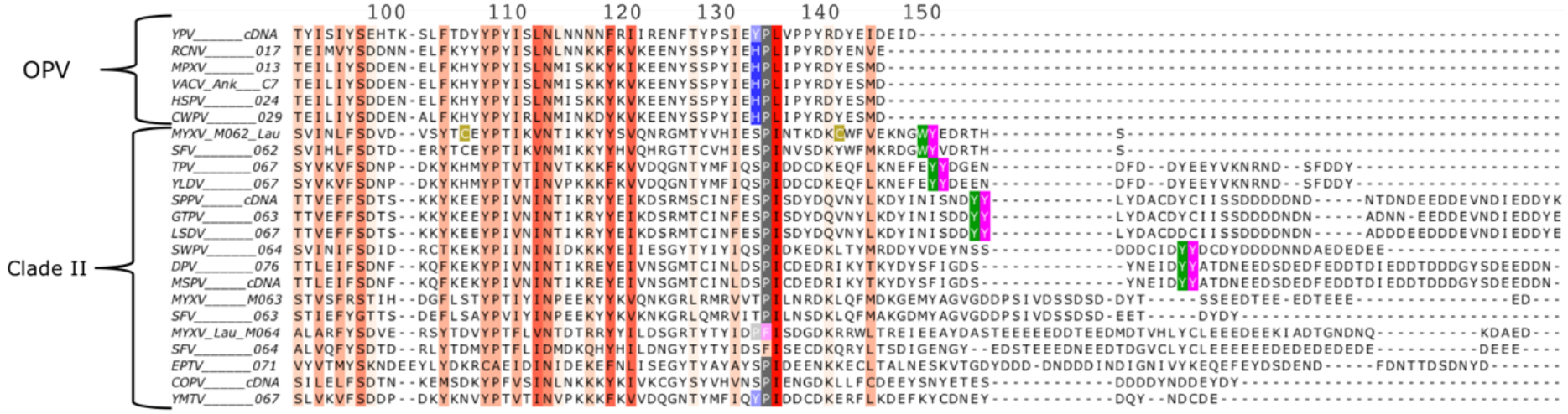


Figure 67: Multiple sequence alignment of the C terminus of the C7 family. Alignment as per Figure 2. A multiple sequence alignment was carried out using Clustal Omega[25]. Amino acids are coloured by conservation in a gradient (above). Poxvirus abbreviations described in Table III; (OPV) orthopoxviruses. The sequence position of M062 annotated above the alignment. Aromatic residues which are predicted to form C-H···π interaction with the cis-Pro (grey) coloured: C terminal tail (green), His residues of the OPV homologues and cis-Pro adjacent aromatics (blue). Hydrophobic residues predicted to occupy the hydrophobic pocket (magenta). Cys → Ser mutants (gold). Figure generated with Aline[147].

4.7. **Discussion**

4.7.1. A comment on the quality of the diffraction data

Sections 4.2 and 4.4 analyse the diffraction data of M062-Lau 3CS and M062-Mav 3CS respectively, and show that each dataset are of adequate quality to solve the structures, however there are pathologies which limit the data strength to medium resolution for both M062 variants. The M062-Lau 3CS data has visibly intense reflections to ~3.67 Å where an ice ring is visible (see Fig. 36, (0, 180). A sharp dip in the completeness, loss of signal intensity, and high B factors at higher resolution (see Fig. 37) meant the data were cut-off at 2.45 Å for refinement.

The best M062-Mav 3CS data collected were of low resolution, extending to 3.2 Å where the mean (I/σ) was 1.1 (see Fig. 53 (A)). The diffuse ice ring shadow between 3.9 and 3.3 Å (see Fig. 52) may have impacted the data quality as well, as the Z-scores were affected (see Fig. 53 (D)) along with the signal intensity, which shows high deviation from the expected intensity plot (see Fig. 53 (B)). Morris *et al.*[148] have shown that cutting resolution shells which deviate significantly from the expected Wilson's plot mean intensities can result in more successful model building, however in this circumstance it would result in throwing out the majority of the higher resolution data, and so was neglected.

Wilson's B factor of 149.78 Å^2 of the M062-Mav 3CS structure (see Table 17) is significantly higher than that of equivalent resolution structures with the average 3.2 Å structure having a B factor of ~80 Å^2, the reported B factor is closer to that of a 4 Å structure[149]. The high B factor indicates the M062-Mav 3CS crystal is not highly ordered, with the model representing best approximations of the average positions occupied by the atoms in the partially flexible structure.

The diffraction data did provided sufficient structural information to identify differences in the Lau and Mav M062 3CS variants from one another, and the existing C7 homologue crystal structures. It is said that "all models are wrong, but some are useful" – E. P. Box, an insight worth keeping in mind when the structures here are viewed.

4.7.2. Non-native features of the M062 3CS crystal structures

4.7.2.1. The serendipitous formation of the M062-Lau 3CS nickel coordination complex

The chelation of the nickel atom by the N terminus of M062-Lau 3CS, though quite interesting in its existence and coordination chemistry (see Fig. 45), is unrelated to the function of M062 and the C7 family, as it is a structural artefact produced as a result of recombinant protein expression. The Ni atom is coordinated by 4 nitrogen atoms of the primary 4 residues of (-)M062-Lau 3CS (see Fig. 45), however this includes the GSH residual amino acids that remain following cleavage by thrombin of the hexahistidine tag upstream of Met 1 of M062. The Ni atom itself is likely to have been picked up during IMAC purification using the Ni-NTA beads, as it is not included in any buffers or the crystallisation condition or cryo-protectant (see Chapter 2).

While the Ni and its coordination is an artefact, PISA analysis[150] showed it was involved in the crystal contacts with symmetry mates, having the largest intermolecular contact surface area (see Fig. 47). Contributing 43% of the total contact surface area. This may have played a role in the growth of crystals which diffracted to higher resolution than those grown previously, with no Ni observed. The coordination complex may have introduced a more ordered structure in the protein molecules, thereby increasing lattice regularity via enhanced N terminal crystal contacts (see Fig. 47). Without incorporation of Ni, it is plausible the M062 structures could not have been solved at all.

The proximity of the N terminus to the trimolecular claw should be noted (see Fig. 63 (A)). It is close to the basic loop $\beta_{2,3}$ and acidic loop $\beta_{4,5}$, and it has the potential to have locked these loops into non-native conformations by steric interference, or as a consequence of the crystal contacts formed.

The coordination complex itself is verified by the CMM server[118]. The only outlier is the Ni B factor, which falls into the 'Outlier' range, however it is worth noting that the dataset from which CMM bins values into different grades includes PDB structures $\leqslant$ 1.5 Å[117] only, coupled with the lower resolution, and high overall B factor of the M062-Lau 3CS structure, this lowers the strength of this validation parameter in this instance.

4.7.2.2. The 3CS mutation

Crystallisation of WT M062 was unsuccessful, however mutation of the non-conserved Cys residues to Ser proved effective in obtaining crystals (see Section 3.6.1). Construct engineering is a common technique used to increase the resolution at which crystals diffract[151], and was necessary for this study, as discussed in Section 3.4. Efforts were made to minimise the impact of these mutants on protein function by choosing non-conserves Cys residues of the C7 family (see Fig. 21). These residues were mutated to Ser, whose sidechain is similar to cysteine, differing only is the substitution of the S atom in Cys for O in Ser. Comparison of the binding of the WT protein and the 3CS mutant variant to SAMD9 by co-IP indicates the ability to pull-down SAMD9 is preserved in the 3CS mutant (see Fig. 16). The M062 3CS structures also appear to adopt the C7 β sandwich fold with high fidelity to those seen in the C7 and M064 structures (see Fig. 62).

The crystallisation of the Lau M062 was successfully increased by the introduction of the 3CS mutants, and investigation of the crystal contacts in the (-)M062-Lau 3CS structure reveals the interaction of C107S Ser with the 'GSH' residues forming the Ni chelating structure contributes to the crystal contact between the N terminus and surface adjacent to α_1 (see Fig. 68 (B)). Formation of the above crystal contact is direct evidence of the impact the Cys → Ser construct engineering mutants played in enhancing crystal formation.

PISA analysis of the crystal contact interfaces between the ASU's of the Mav structure revealed a contribution of the 3CS mutants to crystal formation. (-)M062-Mav 3CS S90 of chain B forms a hydrogen bond with the sidechain of K140 from chain A of the proximal symmetry mate. This surface contact area of this interaction is 41.9 Å^2 out of the total crystal contact surface area in this interface of 135.5 Å^2 (30.8%). This interface contributes 7.6% to the total crystal contact surface area.

Not revealed in the crystal structures is the potential effect the Cys → Ser mutants may have played in increasing surface charge homogeneity by substituting sulfhydryl groups with hydroxyl groups, removing the possibility of oxidation of the surface exposed Cys sidechains (see Fig. 68 (A, C, D)).

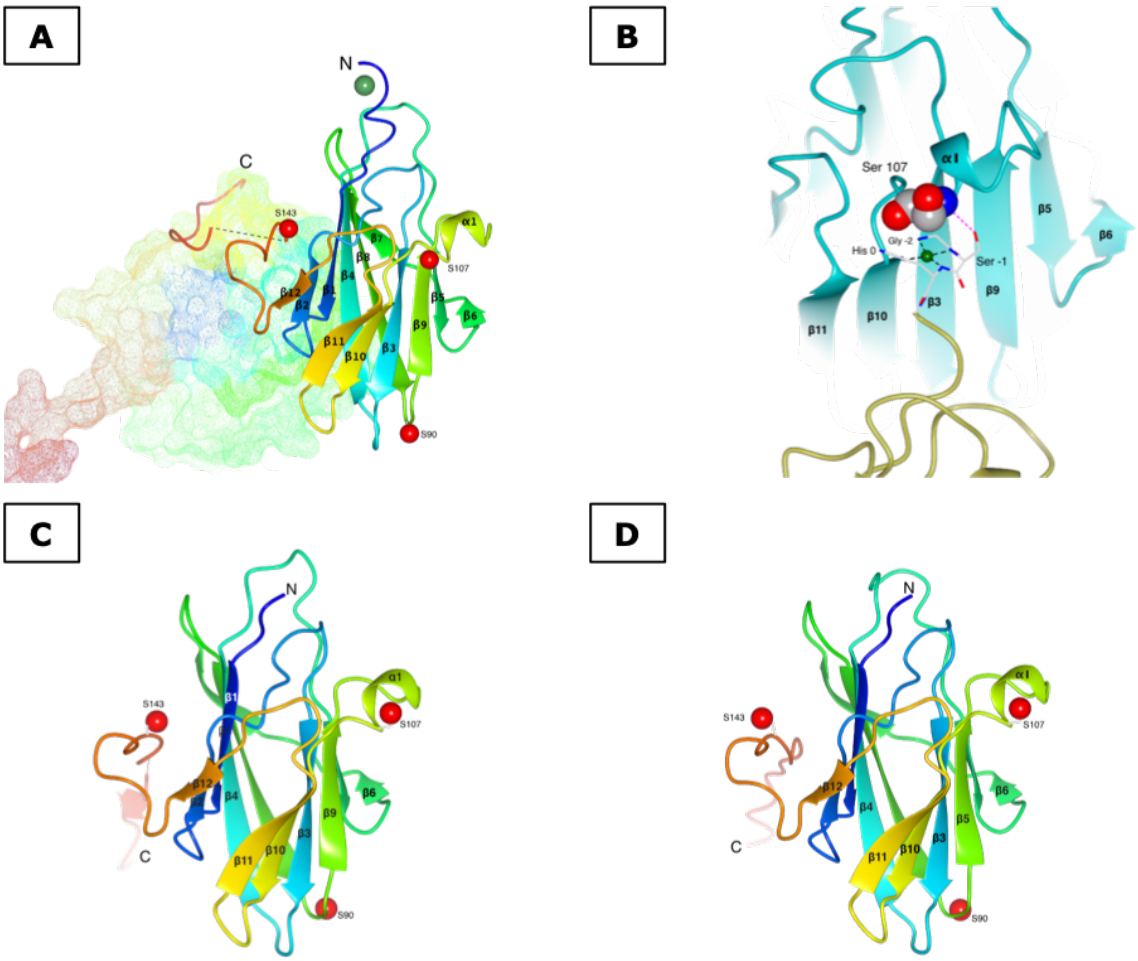


Figure 68: Location of the three Cys → Ser mutations in the crystal structures of Lau and Mav M062 3CS. Ribbon models of the crystal structures are shown in rainbow colour (N terminus blue, C terminus red), with the secondary structures, termini, and location of the mutagenized Ser residues annotated. Ser hydroxyl O atoms shown as red spheres. The trimolecular claw is located at the top of each structure, beside the N terminus. (A) (-)M062-Lau 3CS, with symmetry mate shown as mesh. (-)M062-Mav 3CS chain A (C) and chain B (D) are shown. (B) A hydrogen bond contributing to the major crystal contact between the N terminal:Ni structure and Ser 107 (shown as spheres) is shown (magenta). The location of these mutants in the structure is distal from the identified sites of interest such as the trimolecular claw and the $\beta_{11,12}$ interface.

4.7.3. Implications of a C terminal tail interaction with a hydrophobic pocket in the β sandwich

A novel intermolecular interaction observed in the crystal structure of M062-Lau 3CS between the C terminal tail and a cleft between β strands 11 and 12 has revealed a potential dimerization interface in the C7 family. Further investigation into this interface may reveal if protein-protein interaction plays a role in C7 antagonism of host immunity, or mediate host range.

Ultimately, the biological relevance of this novel crystallographic interface can only be inferred to be either a crystal contact or a quaternary structure interface. Confirmation must come through structure-based, rationally-designed experiments to test the oligomeric state and biological relevance of this putative dimerisation interface to the protein's function.

4.7.3.1. Crystal contact or specific interaction?

Determining whether a crystal contact (a contact between molecules facilitating crystal formation) is an artefact of crystallisation or represents a native oligomeric interaction has been a long-standing point of discussion among macromolecular crystallographers. Differences between crystal contacts and protein-protein interactions which mediate oligomerisation have been evaluated by different criteria such as the specificity and composition of the interaction, the contact surface area[150,152,153].

In the case of the C terminal tail interaction with the hydrophobic pocket, PISA analysis shows a buried surface area of 555 Å^2 (see Section 4.3.5), which is below the apparent lower limit of ~1000 Å^2 sufficient for specific protein-protein interaction[150]. The configuration of the Lau tail:pocket interaction in comparison to the Mav configuration is worth taking into account however.

The proposed biological assembly in the M062-Mav 3CS structures adopts a yin-yang formation with a symmetrical, reciprocal interaction between the subunits (see Fig. 59 (B)). This formation, if representative of the physiological dimer contains a reciprocal interaction which doubles the net buried surface area formed by the tail-pocket interaction, which is within the range of a specific interaction[150].

The interface specificity is discussed in Section 4.3.5, where 3 distinct modes of interaction which span the tail region are highlighted (see Fig. 51). The amino acid composition at the contact surface more closely resembles the residue pair

interactions favoured in oligomers over crystal contacts, featuring Y-I, Y-K and W-P interactions (see Fig. 51)[153]. The burying of the aromatic side chain of Y152 also excludes water from the $\beta_{11,12}$ groove (seen in the C7 and M064 structures[29]), forming an oligomer-favoured style of hydrophobic interaction[153]. The backbone interaction between the C terminal tail and β_{11} (see Fig. 51 (A)) take on a parallel β sheet hydrogen bonding formation.

4.7.3.2. Cis-Pro stabilisation by C-H···π interaction

The conserved Pro at the end of β_{12} is found in cis conformation in all C7 homologue structures[29]. Cis-Pro are found in about 5% of proline residues[116154], this rises to 12-16% when the Pro is preceded by an aromatic residue, with evidence that the aromatic residue stabilises the cis-Pro conformation[155]. Direct evidence of the stabilising effect of the C-H···π interaction has been demonstrated in the Pro-Pro-Aromatic motif[156], and the role of C-H···π bond in protein has been widely studied and expected to play a significant role in overall protein conformation[155–159]. The stabilisation of the cis-Pro isomer by local interaction with an aromatic π bond donor appears to be conserved among C7 homologues (see Fig. 65 (2)). Notably, OPV and clade II virus C7 homologues vary in the location, and conservation of the aromatic residue which acts as the π bond donor in the C-H···π bond.

The cis-Pro···π interaction distances shown in the C7 homologues (3.9 – 5.3 Å) are within the bounds experimentally determined by Raleigh *et al.*, which found "the distance from the centre of the aromatic ring to the closest of the two proline ɣ protons varies between 3.94 and 5.65 Å"[155]. These large interatomic bonding distances have been justified by considering "the sum of the van der Waals radii of the participating groups is larger than in the case of conventional H-bonds"[159].

4.7.3.3. The role of dimerization in M062 host antagonism

Biophysical measurement of the size of native M062 in solution was carried out using SEC-MALS to obtain the oligomerisation state. The average Mw for M062 WT ranged from (Mw ± SEM) Lau: 33.7 – 50.6 kDa, and Mav: 32.4 – 42.5 (see Fig. 66). Though these values vary significantly over the eluted protein peak, they indicate M062 exists as a population comprising both monomeric and dimeric states in solution,. Confirmation M062 contains the protein-protein

homodimeric interface inferred to exist from the crystal structure intermolecular interaction and sequence analysis requires further investigation. The effect of point mutations of key residues of the interface such as W151, Y152 and P136 on protein size would be valuable in determining whether these residues indeed facilitate the proposed dimerisation.

A report by Liu *et al.* proposed a heterodimeric interaction between M062 and M063, and showed M063 enhances M062 interaction with SAMD9[24]. It has also been demonstrated that M063 is required for MYXV to establish productive infection in *O. cuniculus*, and replication and transition to late stage viral gene expression in rabbit cells[36]. It is feasible that a heterodimeric interaction between M062 and M063 is possible via a similar interface as in Figure 50, and that this interaction could play a role in preventing IFN-induced SAMD9/SAMD9L antiviral responses in rabbit[160,43].

The identification of a locus which restricts the host range[31] and binding specificity of SPPV C7 homologue 063 to murine SAMD9L[32] is of interest in relation to the proposed dimeric binding site. Mutation of residues N135 and F136 of SPPV-063 (H132 I133 of M062) to the C7 consensus YI restores binding of SPPV-063 to mouse SAMD9L[32]. These residues are located in $\beta_{12,}$ distal to the trimolecular claw, but local to the hydrophobic pocket/$\beta_{11,12}$ groove, where it may disrupt the fold.

The role of dimerization in modulating antiviral activities is already known for *MYXV* protein T2, a secreted antagonist of TNF signalling. Both monomeric and disulphide-linked dimers of T2 are excreted by *MYXV* infected cells, and while their binding affinities to rabbit TNFα were similar, the dimeric form was found to inhibit TNFα-stimulated cell death more potently than the monomeric form[161]. The authors hypothesise that the dimeric T2 could more reliably prevent TNF receptor dimerization by occupying the TNF receptor binding sites more efficiently than 2 monomers successively binding the same TNF receptor. A similar mechanism of antagonism may be employed by the C7 family, where C7 binding inhibits SAMD9-mediated responses by preventing SAMD9 oligomerisation, binding to effector proteins, or detection of viral PAMPs, processes potentially enhanced by preformation of a dimer.

4.7.4. Dichotomy of the *C7L* family: OPV vs clade II

Phylogenetic analysis of the *C7L* family by Liu *et al.* groups poxviruses into distinct clades[11]. C7 homologues with C termini which extend past the end of C7 form the clade II grouping. C7 homologues roughly equal in size to C7, and without the C terminal tail group into the OPV. These phylogenetic clades was similarly identified by Meng *et al.*, who also noted that the location of the *C7L* orthologs in the viral genome is distinct between these groups[8]. The shorter *C7L* orthologs are located near the end of the OPV genomes, while the clade II orthologs are located centrally, between thymidine kinase and the polyA polymerase regulatory protein (no gene is present here in OPVs).

Another clear distinction between the OPVs and clade II viruses is that while most OPVs contain a copy of a K1L gene, it is missing in the clade II viruses[1] (see Fig. 1). K1 has been shown to complement C7 function in VACV[23], and along with CP77[162] has been shown to bind SAMD9, at distinct locations of SAMD9[33]. Presented here are 2 correlations identified in the C7 homologues, dependent on the presence of the C terminal tail:

1) Both the OPV and clade II structures appear to preserve the C-H···π stabilising interaction at the cis-Pro of β_{12} (see Fig. 65 (2)). There is a conserved motif in the OPV C7 family in which the n-1 position relative to the cis-Pro residue is occupied by an aromatic residue, primarily His (see Fig. 67: cis-Pro, grey; His, blue). The cis-Pro n-1 position is occupied by a Ser or Thr residue in clade II C7 proteins. In the clade II homologues, a pair of residues (an aromatic followed by a hydrophobic) is found 15 – 28 residues downstream of the cis-Pro (see Fig. 67: aromatic, green; hydrophobic, magenta).
2) The OPV viruses contain a copy of a *K1L* ortholog, which is absent in the clade II viruses (see Fig. 1).

4.7.5. The structural impact of the M062 mutant I79T

4.7.5.1. Orientation of the acidic loop of the trimolecular claw

Comparison of the Lau and Mav M062-3CS structures show that they are very similar, with an RMSD of 0.48 and 0.74 Å^2 (chain A and B, see table, Fig. 62). The acidic loop of the trimolecular claw[29] varies most of the 3 loops proposed to mediate interaction with SAMD9 (see Fig. 64), however the M062 Lau and Mav acidic loops are closer spatially to one another than either M064 or C7 loops. M062 uniquely has a deletion of 2 residues in the $\beta_{7,8}$ loop (see Fig. 61), corresponding with its shorter acidic loop, which does not impact M062's ability to compensate for C7 in VACV C7$^-$/K1^{-8}.

While the $\beta_{7,8}$ loop is quite flexible, with high B factors in all of the C7 crystal structures, it is possible that a hydrophobic residue at position 79 is essential during interaction with SAMD9, and may act as a stabilising point to properly orient the conserved Trp positioned at the centre of the $\beta_{7,8}$ loop (see Fig. 61), as seen in the C7 structure (see Fig. 64 (C)). Co-crystallisation of a C7 homologue with SAMD9 may be necessary to reveal the true role of the 79 locus in determining the C7:SAMD9 interface.

Chapter 5: Structural analyses of the pyrin domain of myxoma virus protein M013

5.1. Aims

This chapter aims to describe the expression, purification and structural analysis of the pyrin domain of M013 by circular dichroism and NMR spectroscopy. Previous attempts in the lab to express FL WT M013 had led to insoluble expression, and so a construct with a solubility tag and 4 mutations was developed: His-MBP-M013 1-93 4Mut. M013 was recalcitrant to crystallisation, but within the size limits for structural analysis by NMR. Here, the expression, purification and preparation of mg amounts of ^{15}N, ^{13}C double labelled M013 1-93 4Mut is presented. Bioinformatic structural analysis, circular dichroism secondary structural analysis, and NMR analysis are presented, and the structure of the pyrin domain of M013 is described, highlighting the presence of surface charge residues and their potential role in mediating protein-protein interactions.

5.2. Bioinformatic analyses of M013

Across different secondary structure prediction servers, M013 is predicted to have 6 well-defined secondary structural regions, comprising 5-6 α helices. (see Fig. 69 (A, B)). While the predicted secondary structure boundaries vary, distinct sections are identified up to approximately residue 105, beyond which the C terminus less consistently defined. Homology detection from sequence comparison tools BLASTp[127] and HHPred[163] predict the N terminus of M013 is distinctly homologous to the pyrin domain (PYD) of the death domain superfamily (see Fig. 69 (C, D))[9]. The PYD comprises 6 antiparallel α helices, spanning approximately 80 - 100 amino acids[9,74] which M013 1-90 aligns with, hence its designation as a viral 'pyrin-only protein' or vPOP[75].

The C terminal region after the homologous PYD has little sequence identity or homology to known proteins. A large charge difference is seen between the

proposed PYD and the C terminus, where residues 1-93 have a predicted pI of 4.6, and residues 94-126 have a predicted pI of 10.4.

Figure 69: Bioinformatic analyses of M013 predict homology with the pyrin domain. Secondary structure prediction servers predict 6 regions of defined secondary structure: (A) Jpred[164] predicts mostly α helical features, with a β strand at the N terminus, (B) PsiPred[165] predicts a wholly α helical protein, with 6 α helices. Prediction by (C) BLAST[127] and (D) HHPred[163] (top 25 homologues shown) identify homology between M013 with the pyrin domain, a subfamily of the death domain (DD) superfamily, localised to the N terminal 60 – 90 residues.

5.3. Expression and purification of ^{15}N, ^{13}C (-)M013 1-93 4Mut for NMR analysis

5.3.1. Expression, solubility and purification of His-MBP-M013 1-93 4Mut

^{15}N, ^{13}C double-labelled M013 was expressed and purified as a fusion construct: His-MBP-M013 1-93 4Mut, as described in Section 2.2.9. His-MBP-M013 1-93 4Mut is soluble and expresses well (see Fig. 70). IMAC purification with Ni-NTA resin yields highly pure protein in the eluate, >95% by densitometry (see Fig. 70 (El)). Isolation of (-)M013 1-93 4Mut (necessary for NMR characterisation, as the ~42 kDa tag would contribute the majority of the spectroscopic signal[166]), requires cleavage of the His-MBP fusion tag, which was accomplished with 3C PreScission protease followed by downstream purification steps.

While the expression of the MBP fusion protein was very high, proteolysis was never 100% efficient (see Fig. 70), and obtaining high quantities of pure (-)M013 1-93 4Mut required its efficient isolation from the fusion protein and cleaved His-MBP, which came through readily in second Ni-NTA (see Fig. 70 (FT2)) or amylose resin (not shown) affinity chromatography steps. An alternate purification strategy was implemented to effectively remove MBP/uncleaved contaminants while minimising the loss of (-)M013 1-93 4Mut.

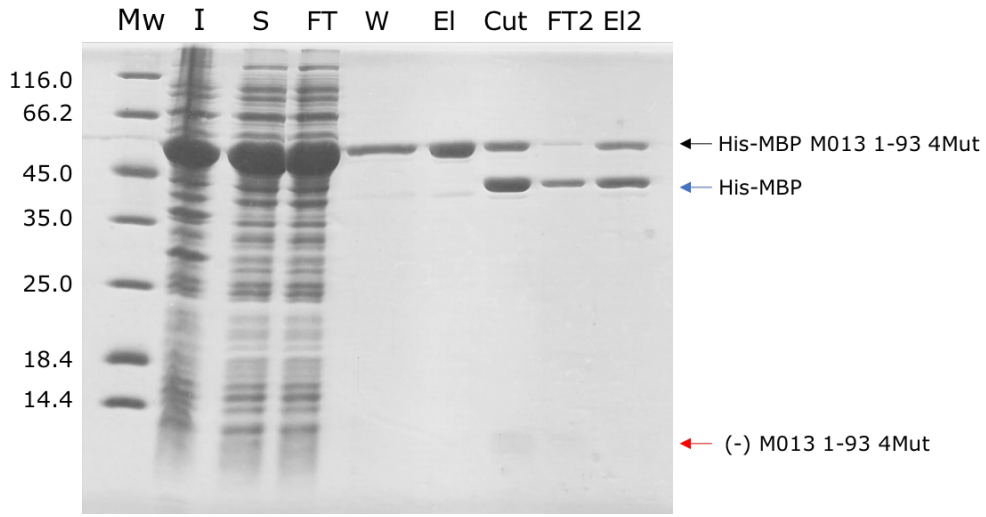


Figure 70: His-MBP-M013 1-94 4Mut expression and purification by IMAC. 12% SDS-PAGE analysis of insoluble (I) and soluble (S) fractions of BL21DE3s expressing recombinant His-MBP-M013 1-93 4Mut. IMAC of the soluble fraction saw significant loss in the flow-through (FT). Following thorough washing (W), eluted >95% pure His-MBP-M013 1-93 4Mut (El) runs as a single band at the predicted Mw of 55.4 kDa (black arrow). Separation of 3C PreScission protease-treated eluate (Cut) by second IMAC could not separate the uncleaved or His-MBP from (-)M013 1-93 4Mut (FT2), and while removing His-MBP-M013 1-93 4Mut (black arrow) it also reduced the (-)M013 1-93 4Mut obtained (red arrow, faint band).

5.3.2. Ion exchange isolation of (-)M013 1-93 4Mut from proteolyzed His-MBP-M013 1-93 4Mut

As application of protease-treated His-MBP-M013 1-93 4Mut to secondary affinity columns failed to isolate (-)M013 1-93 4Mut, IEX was used to separate the products which varied sufficiently in their charge (see Table 19).

Construct	pI (predicted)
His-MBP-M013 1-93 4Mut	5.22
(-)M013 1-93 4Mut	4.63
His-MBP	5.58

Table 19: Predicted isoelectric point (pI) of His-MBP-M013 1-93 4Mut and its 3C PreScission protease cleavage products. pI computed with ProtParam[101].

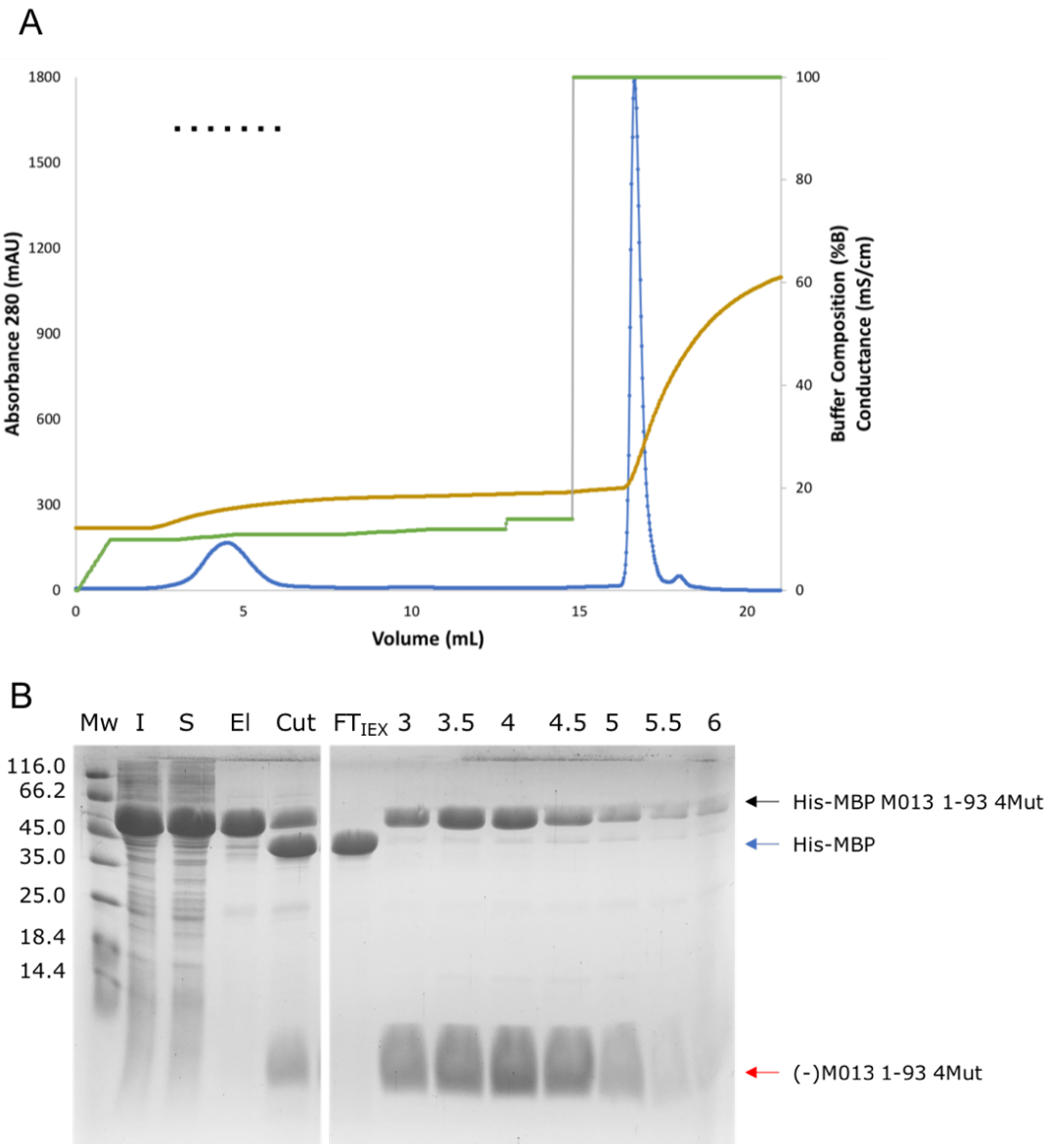


Figure 71: IEX separation of His-MBP from (-)M013 1-93 4Mut. (A) Ion-exchange chromatograph of 3C PreScission protease cleaved His-MBP-M013 1-93 4Mut on a sourceQ 15Q 4.6/100 PE column. A_{280} (blue), buffer composition (green) and conductance (gold) vs volume (mL). Fractions analysed by SDS-PAGE annotated (■). (B) 15% SDS-PAGE of His-MBP-M013 1-93 4Mut expression and purification, followed by IEX fractions, fraction volume (mL) indicated above each lane. Insolubly expressed protein (I), soluble fraction (S), IMAC eluate (El), 3C PreScission protease treated eluate (Cut). Flow-through of the applied cleaved product to the ion exchange column at 120 mM NaCl (FT_{IEX}) contains His-MBP (blue arrow).

Initial attempts to separate the different components of the cut fraction by IEX were made by binding all components at low NaCl concentration (5 mM). However, it was found that His-MBP elutes earliest of the components, at an NaCl concentration of approximately 100 mM. By dialysing the cut fraction to a final NaCl concentration of 120 mM, His-MBP binding to the sourceQ column was prevented, and it passes through in the flow-through (see Fig. 71 (FT_{IEX})). Meanwhile, (-)M013 1-93 4Mut and the uncleaved fusion protein bind the resin. A step-wise gradient was devised which optimised the separation of (-)M013 1-93 4Mut from the uncut protein. While His-MBP was removed, the uncut His-MBP-M013 1-93 4Mut contaminates the (-)M013 1-93 4Mut, and they co-elute as a broad peak from approximately 200 – 210 mM NaCl. Removal of the MBP fusion protein was carried out by sequential SEC.

5.3.3. Separation of (-)M013 1-93 4Mut from His-MBP-M013 1-93 4Mut by size exclusion chromatography

Separation of the coeluted (-)M013 1-93 4Mut and His-MBP-M013 1-93 4Mut following IEX (see Fig. 71) was achieved by their size difference allowing sufficient separation by application to a sx75 16/60 SEC column (see Fig. 72). Baseline separation was usually not achievable with a single SEC run, however by pooling the fractions of pure (-)M013 1-93 4Mut from the right-hand side of the peak (>84 mL, Fig. 72) separately from the (-)M013 1-93 4Mut with residual high Mw contaminants (79 – 84 mL, Fig. 72 (B)), it was possible to isolate the cleaved M013.

Tthe preparation of isotopically enriched protein is a costly endeavour, so, in an effort to optimize the yield of (-)M013 1-93 4Mut from each preparation, the left-hand side of the SEC M013 peak was pooled from separate runs, concentrated and re-applied to the SEC column. The pure (-)M013 1-93 4Mut fractions were then combined and concentrated, flash frozen, and analysed by Dr. Arnoult Kalverda by NMR. The final samples were ~95% pure (see Fig. 73).

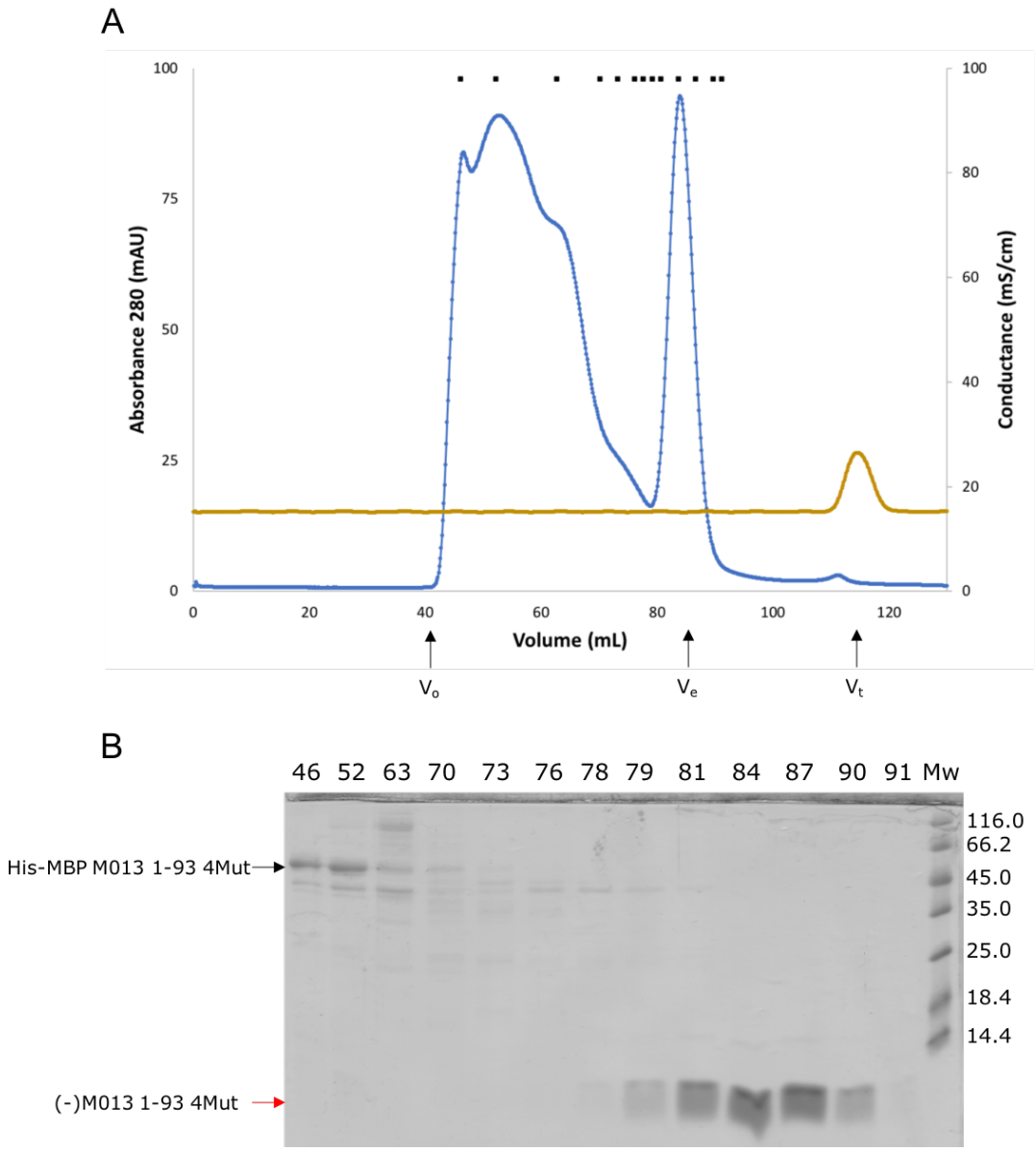


Figure 72: Separation of His-MBP-M013 1-93 4Mut from (-)M013 1-93 4Mut by SEC. (A) Size-exclusion chromatograph of IEX pooled fractions on a superdex 75 16/60 column: A_{280} (blue) and conductance (gold) vs volume (mL). Column volume (V_t) = 120 mL, His-MBP-M013 1-93 4Mut and high Mw contaminants elute in a broad peak from the void volume (V_O) = 40 mL, to ~79 mL. (-)M013 1-93 4Mut elutes in a sharp peak (V_e) = 84 mL. Fractions analysed by SDS-PAGE indicated by (■). (B) 18% SDS-PAGE of SEC fractions from run (A), fraction volume (mL) indicated above each lane.

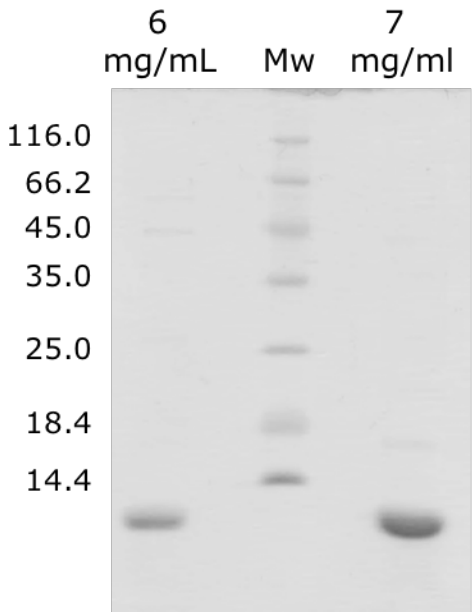


Figure 73: Purified $^{15}N,^{13}C$ (-)M013 1-93 4Mut for NMR. 15% SDS-PAGE of 2 batches of purified double-labelled (-)M013 1-93 4Mut (94%, 6mg/mL; 98%, 7 mg/mL).

5.4. Circular dichroism secondary structural analysis of M013 1-93

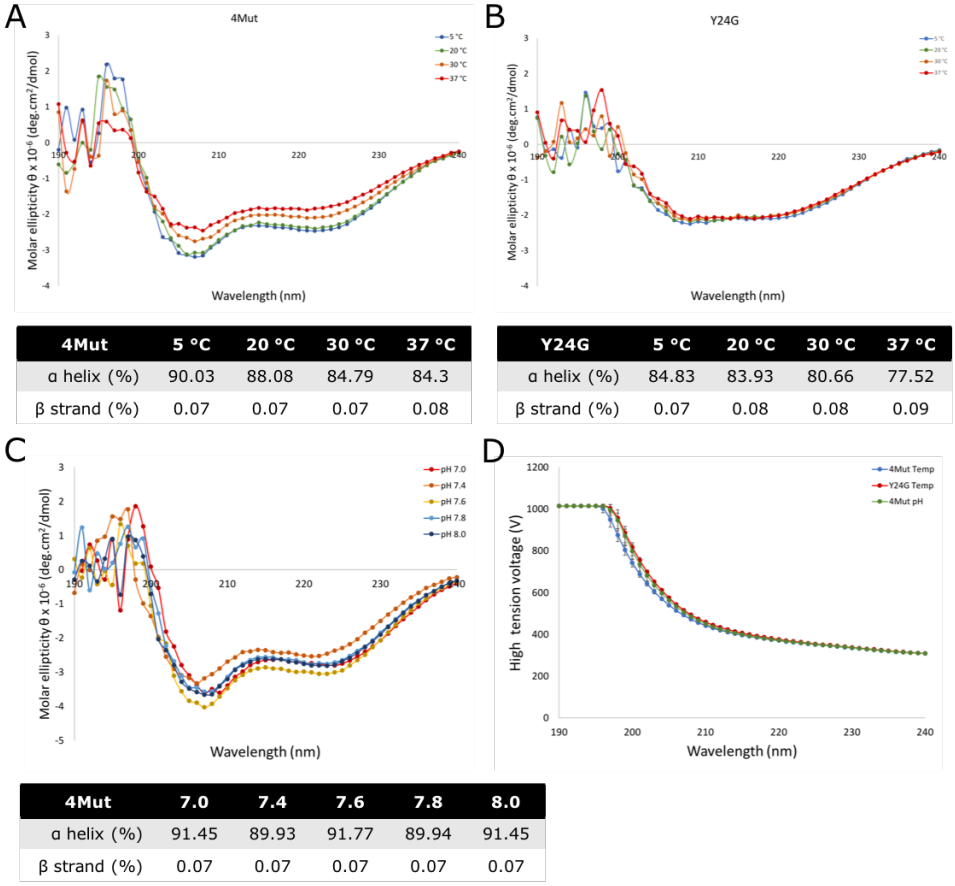
Circular dichroism (CD) biophysical characterisation of the M013 PYD was carried out in order to obtain data of the secondary structure in solution at different conditions of pH and temperature. Comparison of the CD spectrum with the single M013 PYD mutant Y24G also indicates whether the introduction of the mutants required for enhanced solubility were drastically altering the secondary structure. CD data were collected as described in Section 2.2.11.

(-)M013 1-93 4Mut, and (-)M013 1-93 Y24G (single mutant) were expressed and purified as described in section 2.2.9, and CD spectra collected at a range of temperatures and pH (see Fig. 74). Comparison of the 4Mut and Y24G single mutant CD spectra (see Fig. 74 (A, B) indicates that the 4Mut M013 is adopts a more folded α-helical secondary structure than the Y24G M013.

The estimated secondary structure, calculated with the K2D3 server[167] predict M013 is nearly all α-helical, consistent with the bioinformatic prediction of it being a POP. The 4Mut spectra displays the distinctive negative shifts at 222 and 208 nm[168], which are present but less well defined for the Y24G spectra. The secondary structural signal can be seen to decay with an increase in

temperature (see Fig. 74 (A)), with the estimated α-helical percentage dropping from 90% to 84%.

Changes in the buffer pH (see Fig. 74 (C)) have less of an impact on the secondary structure than changes in temperature, indicating the α-helical bundle of the M013 PYD is reasonably stable from pH 7 – 8. CD spectra were collected in low salt buffer to minimise far-UV absorbance (see Fig. 74 (D)).



4Mut	5 °C	20 °C	30 °C	37 °C
α helix (%)	90.03	88.08	84.79	84.3
β strand (%)	0.07	0.07	0.07	0.08

Y24G	5 °C	20 °C	30 °C	37 °C
α helix (%)	84.83	83.93	80.66	77.52
β strand (%)	0.07	0.08	0.08	0.09

4Mut	7.0	7.4	7.6	7.8	8.0
α helix (%)	91.45	89.93	91.77	89.94	91.45
β strand (%)	0.07	0.07	0.07	0.07	0.07

Figure 74: Comparison of the secondary structure of M013 1-93 4Mut vs Y24G at varying temperature and pH by circular dichroism. CD spectra from 190 – 240 nm were collected for M013 1-93 4Mut (A) or Y24G (B), at 0.5 mg/mL, in dilute PBS, pH 7.5, at 5 °C, 20 °C, 30 °C and 37 °C. Secondary structure estimates (tabulated) were calculated using the K2D3 server[167]. (C) The effect of pH on the secondary structure of M013 1-93 4Mut, with CD spectra collected at 0.5 mg/mL, 20 mM PBS pH 7.0 or 20 mM Tris pH 7.4, 7.6, 7.8 and 8.0. (D) The averaged high tension volume for each of (A) (B) and (C) are shown with error bars ± SD.

5.5. NMR analyses of M013 1-93 4Mut

The prepared isotopically-enriched samples of (-)M013 1-93 4Mut were analysed by Dr. Arnoult Kalverda at the Astbury Centre for Structural Molecular Biology at the University of Leeds. All NMR experiments and structure calculations are credited to him.

5.5.1. ^{1}H-^{15}N heteronuclear spin quantum coherence spectroscopy

A preliminary experiment was performed to discern whether NMR was a feasible method for the structural characterisation of M013. A single-labelled sample incorporating ^{15}N was prepared as above, with equivalent yield and purity, and a HSQC experiment performed in order to ascertain if distinct peaks could be assigned for (-)M013 1-93 4Mut (see Fig. 75). Most of the N-H bonds were assignable, with approximately 65 distinct peaks identified. There was significant overlap of some peaks, indicating a degree of disorder of the backbone.

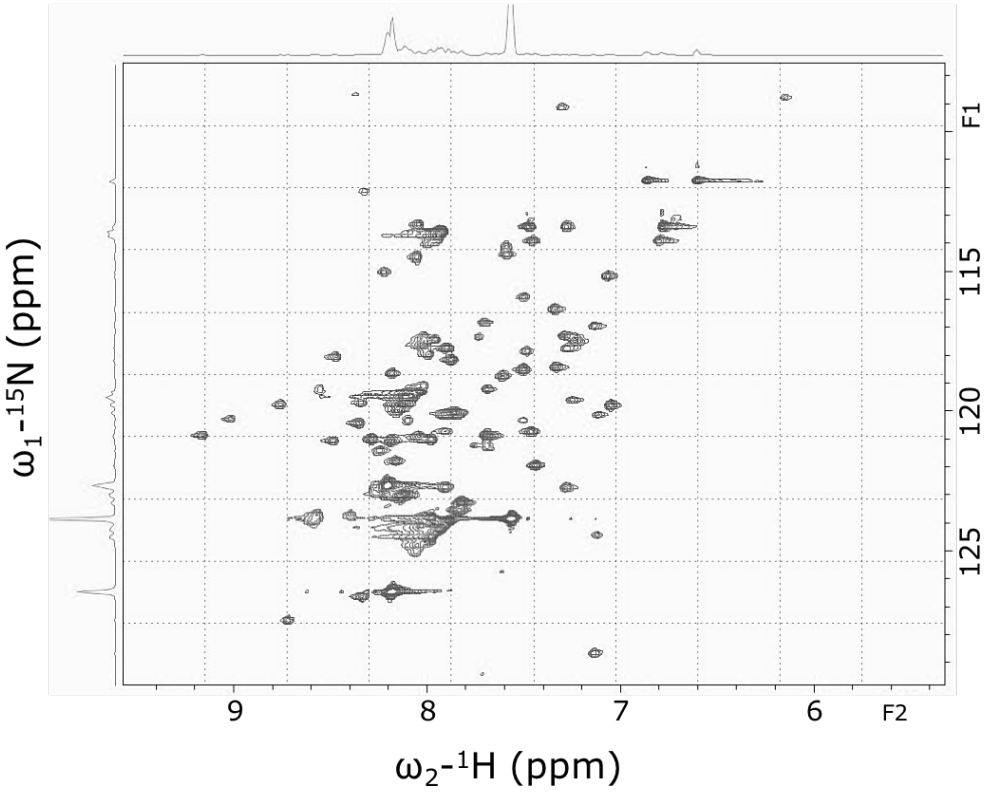


Figure 75: ^{1}H-^{15}N heteronuclear spin quantum coherence spectrum of ^{15}N labelled (-)M013 1-93 4Mut. Approximately 65 N-H bond peaks are identifiable in the 2D ^{1}H-^{15}N HSQC. Credit: Dr. Arnoult Kalverda.

5.5.2. Chemical shift assignment and CS-Rosetta structure prediction

Chemical shift (CS) assignment using ^{15}N, ^{13}C double-labelled (-)M013 1-93 4Mut was successful to trace the backbone (see Fig. 76 (A)), and sufficient to assign $^{13}C\alpha$ and sidechain ^{13}C chemical shifts, and calculate their deviation from random coil values (see Fig. 76 (B)). The chemical shift index (CSI) indicates 5 distinct α helices between 1-70 (annotated) however residues 70-93 have weaker CS values, with less defined secondary structural elements.

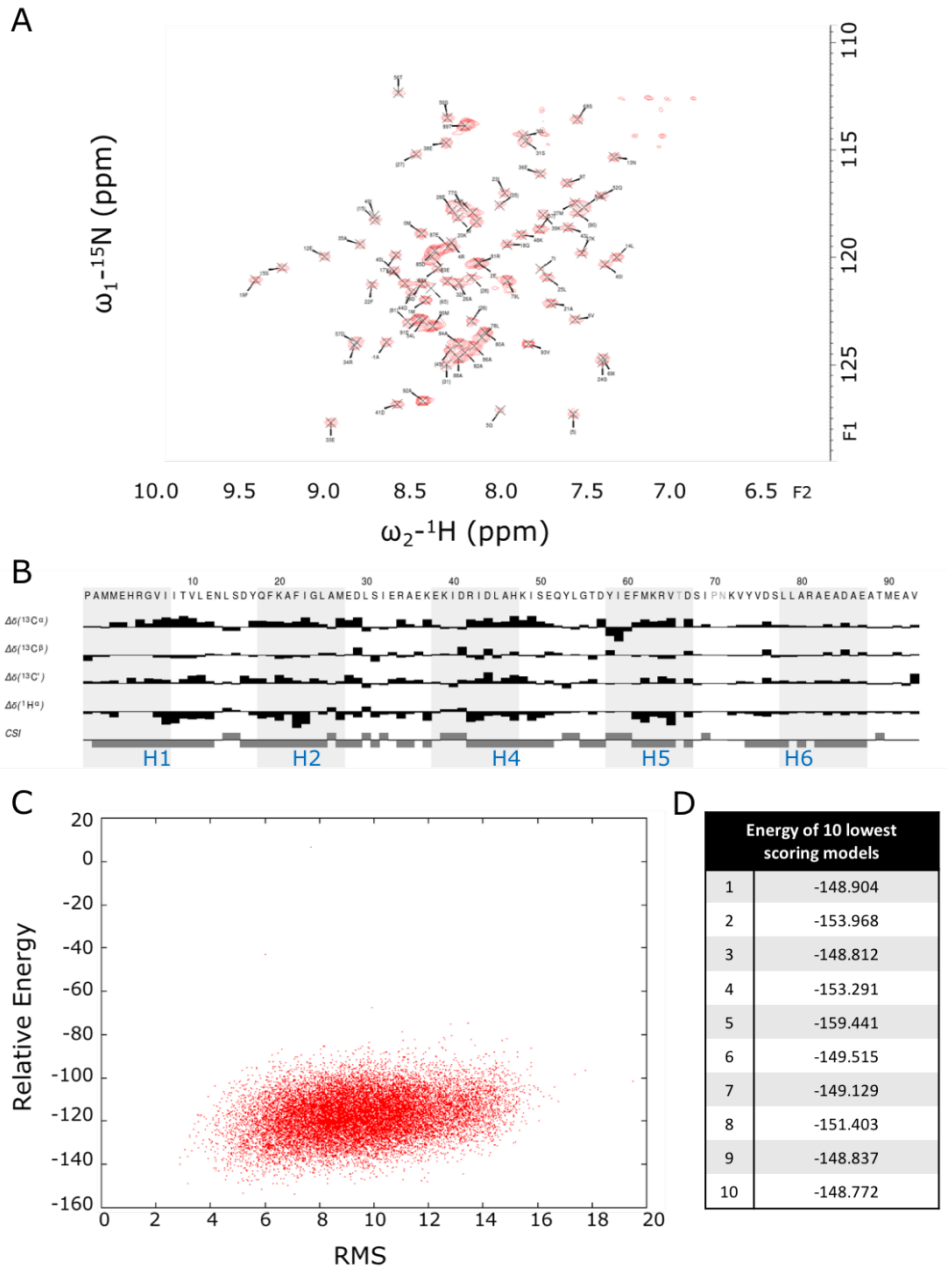
The CS values were used as constraints to generate structural models using CS-Rosetta[169,170] (see Fig. 76 (C)). CS-Rosetta was used to generate 20,000 models, of which the 10 models with the lowest relative energy scores were extracted for visualisation (see Fig. 76 (D)).

5.5.3. NMR structural models of (-)M013 1-93 4Mut

5.5.3.1. M013 1-93 4Mut adopts the pyrin domain fold

The top 10 CS-Rosetta generated (-)M013 1-93 4Mut models have 5-6 defined α helices and show M013 adopts the PYD as predicted, with the pseudo-symmetric anti-parallel α helical bundle configuration (see Fig. 77 (A)). While the models each fold as a PYD consistently, the final α helix is highly flexible, its secondary structure and its position relative to the other helices varies highly between different models (see Fig. 76). Helix 3, the smallest PYD helix also varies in its secondary structural definition, but maintains its position within the globular domain with the well-defined helices 1, 2, 4 and 5 (see Fig. 77 (A)).

The structures presented here represent a truncated, mutated construct of M013, required for its solubility in order to investigate the characteristics of a viral POP. The 4 mutants: M21A, Y24G, Y31S and Y68S remove exposed hydrophobic residues located at the DD Type I interface (see Fig. 77 (B)), a tactic previously used for the preparation of the PYD of AIM2[93]. The surface electrostatic potential map (see Fig. 79 (A)) shows the distinct locations of the acidic and basic patches which comprise the Type interfaces.



Energy of 10 lowest scoring models	
1	-148.904
2	-153.968
3	-148.812
4	-153.291
5	-159.441
6	-149.515
7	-149.129
8	-151.403
9	-148.837
10	-148.772

Figure 76: NMR chemical shift assignments and CS-Rosetta model generation from ^{15}N-^{13}C (-)M013 1-93 4Mut. (A) Residue assignment of the 2D ^{1}H-^{15}N TROSY N-H backbone chemical shifts. (B) Chemical shift deviations from random coil for assigned $^{13}C_\alpha$, $^{13}C_\beta$, $^{13}C'$ and $^1H_\alpha$ and chemical shift index (CSI) shown with sequence of (-)M013 1-93 4Mut annotated. Predicted helices annotated (blue) (C) The relative energy and RMS for the 20,000 CS-Rosetta-generated models are graphed, with the 10 models with lowest energies listed in (D).

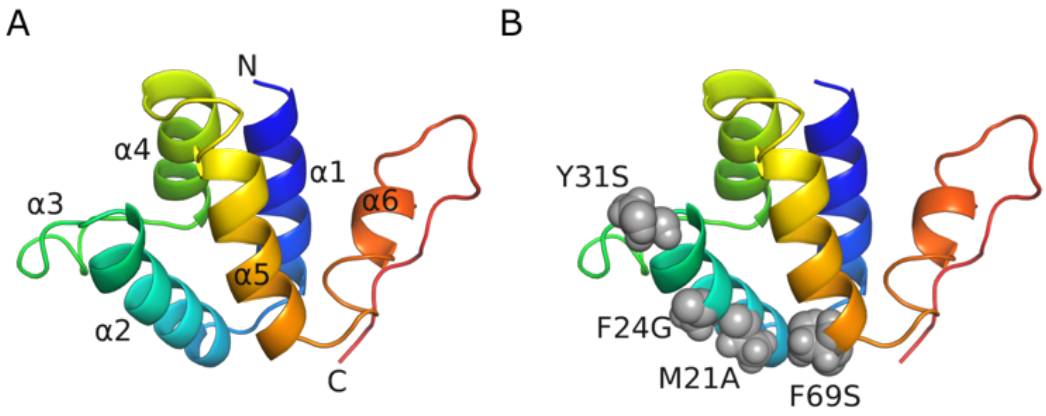


Figure 77: The NMR (-)M013 1-93 4Mut model adopts the Pyrin DD fold. The most compact PYD of the 10 lowest energy CS-Rosetta models is used to represent the M013 structure. (A) A ribbon model of (-)M013 1-93 4Mut is shown, rainbow coloured, with N and C termini and α helices annotated. The PYD 6 antiparallel α helix bundle can be seen, with helix 3 lacking defined secondary structure, and helix 6 less tightly associated with the globular domain. (B) The location of the 4 mutated residues required for soluble expression are shown as spheres (grey): M21A, Y24G, Y31S and F68S. The mutants co-localise to the Type Ia face between helix 2 and 3.

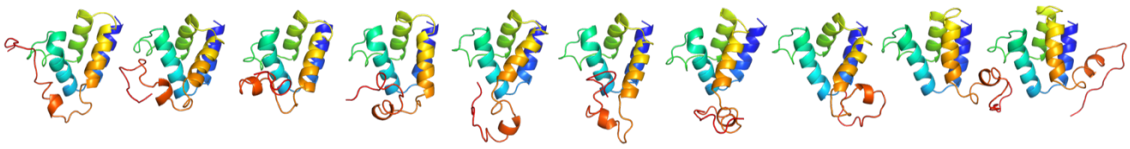


Figure 78: (-)M013 1-93 4Mut adopts the PYD fold. The top 10 NMR models of (-)M013 1-93 4Mut calculated by CS-Rosetta are shown in ribbon model, rainbow coloured. Helices 1-5 can be seen to be confined together, while C terminal α helix (H6) varies in its location and secondary structure between models. Figures made with PyMOL[144].

5.5.3.2. The Type I interfaces of M013

The DD has been shown to form homo- and heteromeric interactions via complementary charged faces in a Type I interaction, mediated by the interaction of Type Ia and Ib surfaces (Type Ia: helices 1 and 4, Type Ib: helices 2 and 3[95,171–173].

The surface electrostatic potential maps of the CS-Rosetta M013 PYD model shows the Type Ia surface resembles the ASC Type Ia surface, with a cluster of positive charge at the interface between helix 1 and 4 and a negatively charged lower face (see Fig. 79 (B)). Arg 4 of M013 is the main contributing residue to the basic patch. 4 acidic residues from helix 1: Glu 12, helix 2: Asp 16 and helix 4: Asp 41, Asp 44 form the negatively charge region beneath the Type Ia interface.

The Type Ib interfaces of the M013 PYD and ASC vary however, with M013 exhibiting a negative surface charge above the positive face seen in both ASC and M013 (see Fig. 79 (C)). Homotypic self-association of the M013 PYD, and heterotypic association between the M013 and ASC PYD via Type I appears feasible based on the Type Ia and Ib faces being complementary in the distribution and polarity of their surface charge (see Fig. 79 (B, C)).

5.6. Discussion

5.6.1. Structural characterisation of M013 confirms its assignment as a viral pyrin-only domain protein

5.6.1.1. Bioinformatic prediction of a PYD is consistent with the CD secondary structural assignment

Bioinformatic approaches have characterized M013 as containing a PYD since its description in 2001[9]. The secondary structure and homology prediction servers maintain this identification of M013 being mostly composed of a PYD, with C terminal residues after the PYD having poorly defined secondary structure and little homology to any known structures (see Fig. 69). CD spectra of M013 1-93 4Mut (see Fig. 74 (A, C)) match those of other PYDs[86]. Along with the secondary structure analysis of the CD spectra by the K2D3 server[167], this indicates it is predominantly α-helical, supporting the overall structural assignment of the PYD of M013.

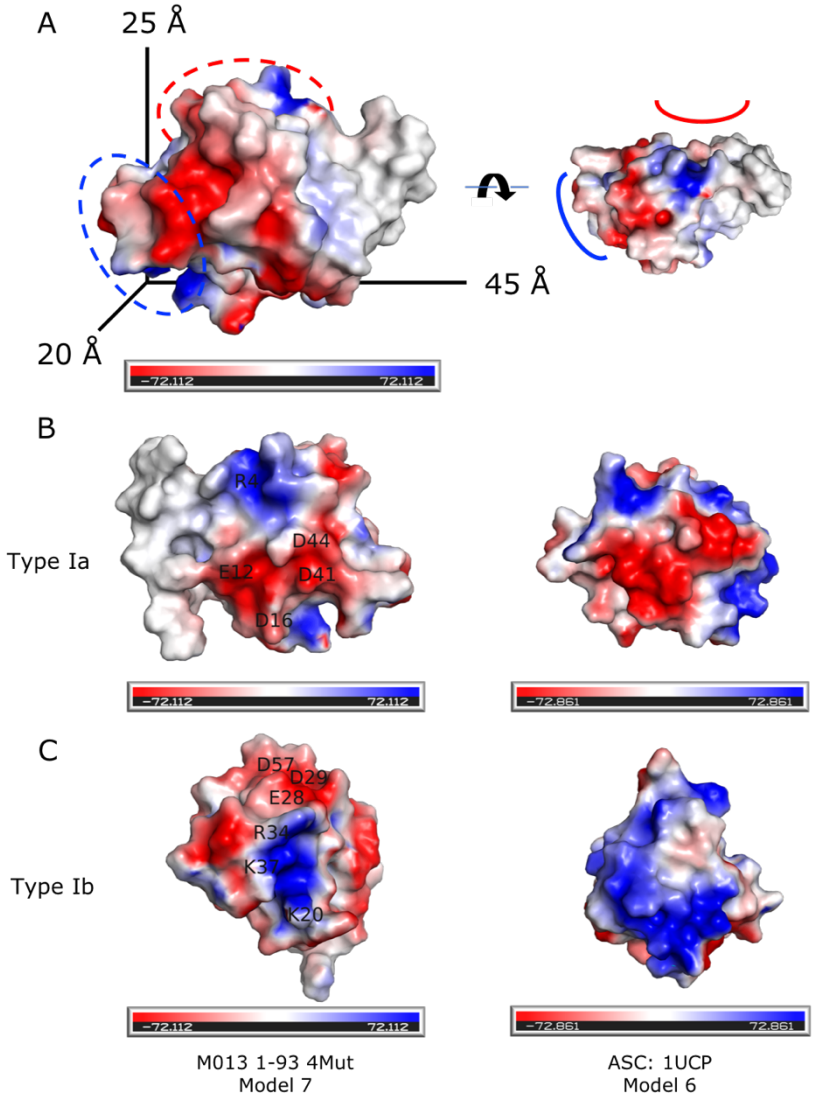


Figure 79: Comparison of the Type I interfaces of M013 1-93 4Mut and ASC. (A) An electrostatic surface map of the CS-Rosetta model of M013 1-93 4Mut in Fig. 77 is shown with the DD Type Ia (red, posterior) and Type Ib (blue) interfaces circled. The M013 PYD is approximately 45 x 25 x 20 Å. An overhead view shows the concave and convex curvature of the Type Ia and Ib faces (red and blue, respectively). Comparison of the Type I interfaces of M013 and ASC are shown in (B) and (C), with the charged surface residues of M013 annotated. (B) The Type Ia interfaces of M013 and ASC are similar in their surface charges, featuring a positively charged face contributed mainly by R4, under which is a negative region. (C) The Type Ib interfaces of M013 and ASC have different surface charge distribution, both featuring a distinctly negative face, but with M013 possessing a negative patch above. Figures made with PyMOL[144].

5.6.1.2. CS-Rosetta modelling of M013 1-93 4Mut shows the C terminal helix flexibly associates with the PYD

The NMR data provided restraints with which CS-Rosetta[169] calculated 20,000 models and obtained 10 models of the lowest free energy[174]. The general rule of thumb for whether the generated models provide a reliable structural model when a known structure against which they can be compared is unavailable, is to check the 10 lowest scoring models for their convergence on a similar model[169]. As can be seen in Figure 78, there is a lack of convergence of these models, and the terminal α helix's orientation and proximity to the PYD helices 1-5 varies dramatically between structures.

The CS index (see Fig. 76 (B)) implies helix 6 has defined secondary structure, however the chemical shifts have weaker signal than the PYD α-helical bundle, which does appear to have converged in the 10 best models (see Fig. 78). Taken together, the NMR model indicates that the overall M013 1-93 4Mut PYD is only partially stable, and there is a high degree of flexibility in the C terminus. The CS-Rosetta calculations were carried out using the CS restraints up to the end of helix 5 only (residues 1-73), and in this experimental scenario helices 1-4 did not form a stable α-helical bundle, indicating the C terminal helix, while flexible stabilises the overall PYD of the models (Dr. Arnoult Kalverda, personal correspondence).

The M013 construct is of course highly mutated from the WT protein (see Fig. 77 (B)), and the models may reflect the structural impact of the point mutations, or more likely, that the C terminal positively charged 'tail' could alter the overall fold in unknown ways. The models do however predict the PYD consistently enough to allow for inference of the role of certain residues in the Type I interfaces, which correlate with the antagonistic feature of the M013 PYD.

5.6.2. The M013 PYD models provide supporting structural evidence for the functional role of key surface residues in the Type I interface

A recent study used the structure of the AIM2 PYD to generate a homology model of the M013 PYD, and with structure-based mutagenesis investigated the role of predicted surface charged residues on M013's antagonism of ASC and NFκB[76]. Charge neutralisation of the negative surface residues Glu 28 and Asp 29 of the Type Ib interface completely disrupted co-IP with ASC, while interaction with p105 of NFκB was unimpacted. In contrast, D57N neutralisation impacted upon interaction of the PYD of M013 with NFκB while binding to ASC was unaffected. Mutation of Asp 41 of the Type Ia interface had no major effect on either ASC or NFκB binding[76].

The NMR model shows residues 28 and 29 are located at the N terminal side of the Type Ib interface, above a patch of positive surface charge (see Fig. 79 (B)). The charge distribution of the Type I faces of ASC and M013 show the convex positive M013 Type Ib surface with a region of negative density above is complemented by the ASC Type Ia face, which is positive at its concave centre, with a positive patch located above (see Fig. 79 (B, C)). The role of the acidic residues of the Type Ia face is therefore supported by the NMR structural models. The lack of impact of the D41A mutant on ASC interaction[76] may show that residues Glu 12, Asp 16 and Asp 44 may predominate the Type Ia interaction.

The role of hydrophobic residues at the centre of DD Type Ib interfaces (mutated here as Y24G, F27G in the AIM2 [4O7Q] crystal), have been proposed, and shown by mutagenesis studies to be implicated in Type I PYD-PYD interaction[92,93,95,175]. As mutation of these hydrophobic residues on the surface of the Type Ib face (see Fig. 77 (B)) was required for M013's expression, the NMR models cannot be assumed to accurately reflect the surface charge of the WT structure.

The overall PYD fold does allow for reasonable confidence in the relative position of the residues which determine the surface charge of M013. The NMR models potentially allow for more accurate identification of key residues of M013 which contribute to its dual-antagonism of ASC and NFκB.

5.6.3. The current model: dual-effector mechanisms of action of M013

The current proposed mechanism of antagonism by M013 is depicted in Fig. 80. Activation of TLRs by *MYXV* infection leads to the recruitment of intracellular signalling adaptor proteins containing TIR domains. This in turn leads to the activation of TRAF6, TAB2, TAK1 and TAB1, ultimately allowing for TAB1 activation of the IKK complex. IKKβ activation leads to the phosphorylation, and ubiquitination of p105. Following processing by proteolytic cleavage to p50, the NFκB p50:p65 heterodimer translocates to the nucleus, where it leads to the transcription of pro-inflammatory cytokines[96]. M013 antagonises NFκB signalling via its interaction with p105, preventing its processing to p50 and inhibiting NFκB-mediated expression of cytokines[176].

Cellular stress induced by viral replication can lead to activation of NLRP3[177], and result in the formation of the multiprotein complex inflammasome[178]. Homotypic interactions between the PYDs of NLRP3 and ASC facilitate interaction with the CARDs of ASC and pro-Caspase-1, leading to the recruitment and activation of Caspase-1[88]. Activated Caspase-1 processes the pro-inflammatory cytokines pro-IL-1β and pro-IL-18 to their active forms, leading to inflammatory cellular signalling. M013 disrupts the processing of pro-inflammatory cytokines IL-1β and IL-18 to their active forms via its direct interaction with the ASC, via homotypic DD interactions of the PYDs[96,176].

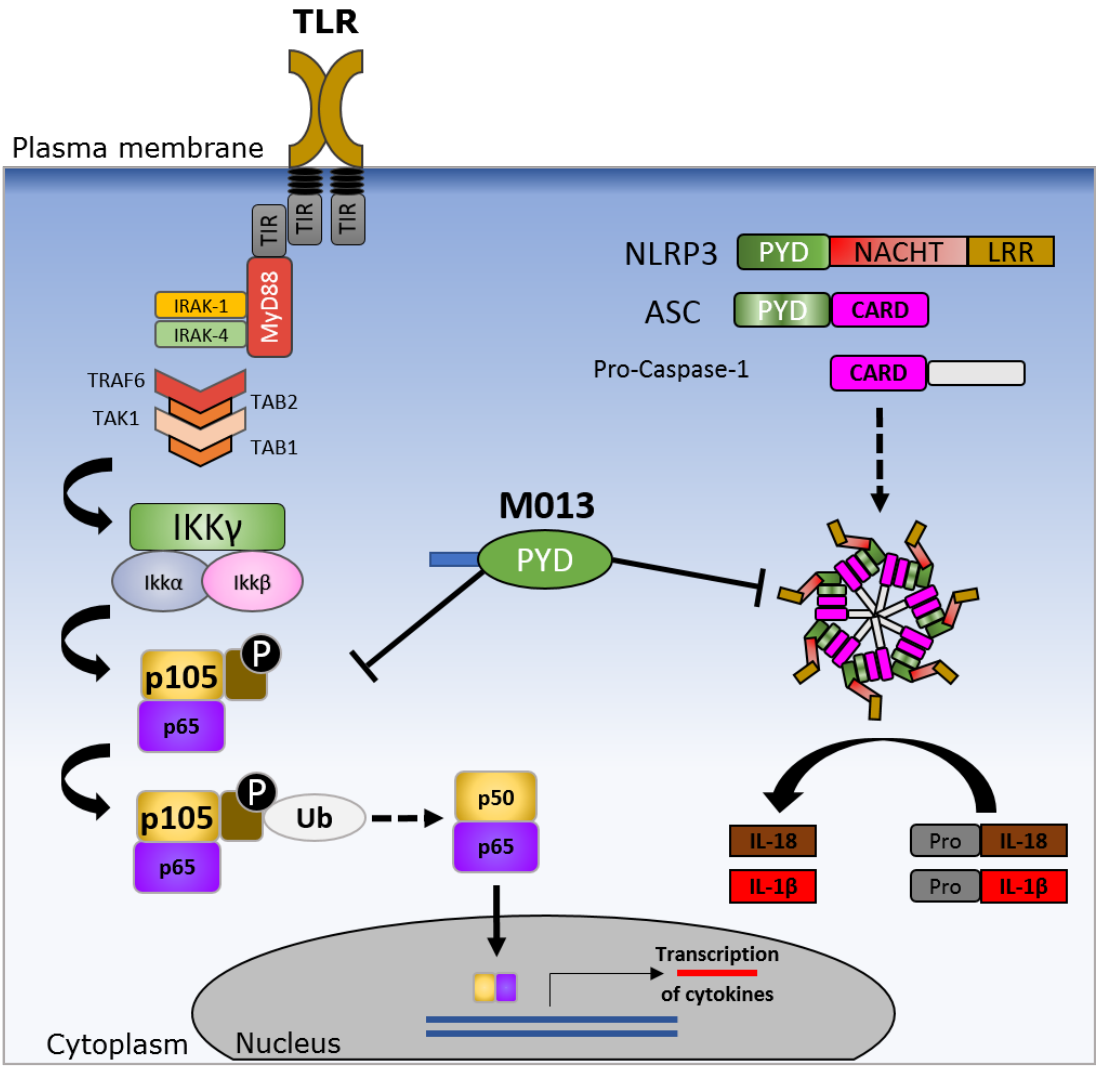


Figure 80: The mechanism of antagonism of the host immune response by M013's interaction with distinct cellular targets. Schematic showing the mechanisms of action of M013's antagonism of NFκB and ASC, targeting distinct antiviral immune signalling pathways. *MYXV* M013^{-} infection leads to the production of pro-inflammatory cytokines via both NFκB and ASC-dependent pathways[96]. M013 has been shown to antagonism NFκB via direct interaction with p105, preventing its processing to p50. This effector function appears to require the C terminal tail of M013. The M013 PYD is sufficient to antagonise inflammasome-mediated cytokine production, via the interaction of M013 with ASC, preventing processing of IL-1β and IL-18 from their pro-forms to their active forms.

Chapter 6: Conclusion

6.1. The structure of M062 presents a novel, putative oligomerisation interface of the C7 family conserved in the OPVs

Chapter 3 presents the expression, purification and crystallisation of the Lau and Mav variants of M062. Wild type M062 did not crystallise readily, and structure-based mutagenesis of non-conserved cysteine residues was conducted to engineer M062-Lau 3CS and M062-Mav 3CS, which were successfully purified and crystallised. Chapter 4 reports the process of refinement of the crystal structures, and describes the structure solutions of M062-Lau 3CS [6S5I] and M062-Mav 3CS [6S6O]. The intermolecular interface formed by the interaction of the C terminal tail residue Y152 with the hydrophobic pocket formed by interface of β strands 11 and 12, and W152 with cis-Pro 136 at the end of β12 is highlighted.

Sequence alignment of the C7 homologues encoded by orthopoxviruses allowed for the identification of a conserved dichotomy between the clade II and OPVs. OPVs were shown to encode C7 homologues that terminate at positions equivalent to C7, and also encode a K1 homologue. The clade II viruses lack K1, and encode C7 homologues with additional residues at the C terminal compared to those of C7. P136, shown to adopt the cis-conformation is stabilized by C-H$\cdots\pi$ interaction, however the aromatic residue acting as the π-bond donor in the OPV C7 homologues occupies the n-1 position, forming an intramolecular bond, while the clade II C7 homologues stabilize the cis-Pro via an intermolecular interaction involving a pair of hydrophobic residues located 15-28 downstream of the conserved proline residue.

6.2. The potential role of C7 homologue complex formation as a determinant of host range specificity

SPPV-063 restriction in murine 3T3 cells was found to be rescued by the mutation of 2 residues located in β strand 12: N134Y F135I[31], and mutation at this locus was responsible for interaction with mouse SAMD9L[32]. Interaction of SPPV-063 with human SAMD9L was also distinctly weaker than other functional C7 homologues, and could not recover *VACV* C7$^-$ K1$^-$ SAMD9$^{-/-}$ host range in human cell lines pre-treated with IFN[32].

It has been well documented that *VACV* K1$^-$ cannot replicate in rabbit cells[23], and transient C7 expression cannot rescue *VACV* C7$^-$ K1$^-$ host range in RK13 cells, while the K1 revertant mutant is permissive in RK13s[24]. It has been shown that the ankyrin repeat poxviral antagonists of SAMD9, K1 and CP77 target the central region of SAMD9[33], while the C7 homologue M062 interacts with the N terminal 385 residues of SAMD9[30]. Distinct host range is seen between *VACV* and *MYXV*, with the latter having extended tropism in both rabbit and human cell lines[36,179].

In aggregate, M062 can overcome a replication block in rabbit[24] which restricts *VACV* tropism. M063 has also been shown to be essential for *MYXV* pathogenesis in rabbits[36], and it has been proposed that a complex is formed during *MYXV* antagonism of SAMD9 which involves SAMD9 and both M062 and M063[24]. The crystal structures of M062 presented here indicate a potential dimerization interface which may affect the antagonism of SAMD9 by C7 homologues, and could influence on the host range conferred by specific C7 homologues. The proximity of the putative dimerization interface (located between β strands 11 and 12) to the SPPV-063 locus which determines its binding to SAMD9L supports the hypothesis that the $\beta_{11,12}$ interface may be of structural significance in determining the specificity of C7 homologue interactions.

The absence of the C terminal tail in the OPV homologues of C7 correlates with a strictly conserved His residue located adjacent to the conserved cis-Pro at the end of β_{12}. It is possible that OPV C7 homologues accommodated the loss of their C terminal tail by the compensatory mutation of the cis-Pro n-1 residue to an aromatic residue, which can stabilise the cis-Pro orientation, and bolster the overall C7 structure, enabling effective SAMD9 interaction. In the clade II C7 homologues, a pair of aromatic residues located downstream of the C

terminal α helix is semi-conserved, and may supply the π-bond to stabilise the cis-Pro, maintained in position by burying of the secondary aromatic in the hydrophobic pocket.

A corresponding genetic variation between the OPV and clade II poxviruses is that the OPV viruses, potentially as a result of evolutionary compensation, all possess a copy of a *K1L* ortholog, which is known to target SAMD9/SAMD9L in a separate region to the C7 homologues. It is tempting to postulate that loss of the C7 tail in the OPVs incurs a host restriction cost, perhaps limiting the variety of SAMD9/SAMD9L variants which can be successfully antagonised by the compact OPV C7 homologues. However, possession of a secondary SAMD9 antagonist in K1, which acts via specific interaction with a distinct region to C7 may compensate for whatever restriction the truncated C7 homologues may incur.

Comparison of the sequence of human and *O. cuniculus* SAMD9 shows a 38 residue deletion in rabbit SAMD9 compared to human SAMD9, of equivalent residues 186-223 (see Fig. 81). This deletion lies in the predicted region of SAMD9 which shows homology to the Alba 2 domain[53], in the N terminal region which interacts with M062[30]. It is possible that the deletion in rabbit SAMD9 alters the structure sufficiently to restrict C7 host range by preventing its antagonism. Persistence of 3 *MYXV* C7 homologues may have emerged as *MYXV* adapted to interact with the N terminally truncated rabbit SAMD9, unable to be compensated by the OPV secondary SAMD9 antagonist K1.

Figure 81: Sequence alignment of human and rabbit (*O. cuniculus*) SAMD9 shows a deletion in rabbit SAMD9 of human SAMD9 equivalent residues 186 – 223. Sequence alignment was carried out using Clustal Omega[25]. Amino acids are coloured by conservation in a gradient (50 - 100% conservation: pink to red). Sequence limits shown. Figure generated with Aline[26].

6.2.1. Oligomerisation of the *MYXV* C7 homologues as a mechanism of immune modulation

The identification and description of a putative oligomerisation site within the C7 family has implications for an adaptive function of the C7 family for modulation of the immune response, specifically by the *Leporipoxviruses*, via oligomerisation of the different C7 family members they encode. With three copies of C7 proteins, *MYXV* has the potential to generate six distinct dimeric species: homodimers of M062, M063 and M064, or heterodimers – M062:M063, M062:M064 or M063:M064. These dimeric variations have the potential to exhibit profoundly different immunomodulatory effects within the infected cell.

Given the studies characterising the functional impact of the individual deletion of M062[24], M063[36] and M064[37] on *MYXV* replication, conclusions were drawn which indicate M062 is essential for *MYXV* replication and pathogenesis; M063 is essential for *MYXV* replication in rabbit cells, but has no impact upon *MYXV* replication in human cancer cell lines which allow permissive replication of wild-type *MYXV*; and M064 impacts upon the kinetics of myxomatosis disease progression, while "not contribut[ing] to *MYXV* host range functions"[37]. These knockout studies provide significant insight into the broad characteristics of the *MYXV* C7 homologues, however the mechanisms of action of M062, M063 and

M064 remain poorly understood. The finding that M062, M063 and M064 may form homo-, and more interestingly heterodimers may indicate that the function of these homologues may rely on their interaction during infection.

The Bcl-2 family of proteins modulate apoptosis by regulating pore formation in the outer mitochondrial membrane (OMM), a critical step controlling the induction of cell death. The family of Bcl-2 proteins are categorised based on their function, and comprises 3 classes: i) anti-apoptotic, ii) pro-apoptotic *effectors*, and iii) pro-apoptotic *activators*[180]. Apoptosis is controlled by the Bcl-2 family by the interaction between these different classes, which are composed of different Bcl-2 homology domains. Apoptotic determination is controlled by homo- and heterodimer formation of members of the Bcl-2 family. Pro-apoptotic family members become activated by cellular stress increasing leading to an increase in pro-apoptotic signalling. The activated pro-inflammatory Bcl-2 family members are initially sequestered by the formation of heterodimers with anti-apoptotic Bcl-2 family members, however, requiring sufficient pro-apoptotic signalling to overcome the dynamic switch to lead to apoptosis[181].

A similar mechanism of action may be at play during *MYXV* infection. Viral success is dependent on the survival of the host for a period sufficient to allow for viral replication, and transmission to another permissive host. The attenuation of a virus via lineage-specific inactivation of host range genes has been proposed as a method of increased virulence through increasing host survival, and thus viral transmission[1]. It is possible that the *Leporipoxviruses* have achieved a similar outcome at the infected cell level, not through inactivation of a single host range gene, but in this case via the adaptation of a gene duplicate to act similarly to the anti-apoptotic Bcl-2 family members. The proposal here is that *MYXV* has attenuated its pathogenesis at a cellular level by modulating its antagonism of the induced cellular immune response in order to allow for successful replication and dissemination of virion particles through dynamic interactions of its encoded C7 proteins.

Based on the independent KO studies, M062 and M063 are essential in the antagonism of the rabbit antiviral response, and have been theorised to form heterodimers[24]. A plausible mechanism of action of M064's impact on the kinetics of myxomatosis disease progression, is via its sequestration of both M062 and M063 via the formation of heterodimers which limit their antagonism of SAMD9 and/or SAMD9L, prolonging individual cell survival by allowing the

formation of SAMD9:RNA stress granule during viral infection[182]. During infection with *VACV*, C7 or K1, or with *MYXV*, M062, inhibit SAMD9-mediated translational arrest, and prevent the formation of SAMD9:RNA stress granules (shown to form during infection of HeLa cells with *VACV* C7$^-$ K1$^-$)[182183].

The influence of M064 and M063 on the formation of these SAMD9 stress granules has not yet been investigated. It may be that heterodimers that incorporate M064 enhance viral replication by paradoxically reducing the ability of *MYXV* to prevent SAMD9 stress granule formation. This would lead to a reduced rate of viral gene synthesis, and reduced viral replication[184]. Indeed, it has been shown that infection with *MYXV* M064$^-$ leads to reduced levels of viral early and late gene products[37]. While counterintuitive, M064 decreasing the viral load in the host by tuning *MYXV* antagonism may increase overall viral fitness, and explain the conservation of M064 among diverse *MYXV* strain isolates[22].

6.3. **Future prospects**

6.3.1. Proposed investigations on the effect of structure-based mutagenesis of M062 on host range, and interaction with SAMD9

Analysis of the M062 crystal structures has indicated the presence of a previously unidentified oligomerization site, and proposed a role for certain conserved residues at this site. In order to determine whether these residues impact on the functional role of the C7 family, structure-based mutagenesis studies may determine if the putative residues effect host range, or interaction with SAMD9.

A set of such experiments are currently underway at the time of writing, investigating the ability of the following M062-Lau C terminal mutants to interact with SAMD9 by co-IP: M062-Cdel: C terminal truncation at K142; M062 W151A; M062 Y152A; and double mutation of W151 and Y152, M062 W151A Y152A. It would also be interesting to investigate the effect of mutation of H138 of C7, to see if loss of the aromatic residue at the n-1 position to cis-Pro effects C8 protein stability, or interaction with SAMD9. It would be interesting to determine whether the n-1 aromatic present in the OPV C7 homologues prevents oligomerisation of M062 by introducing mutant S135H and measuring the Mw by SEC-MALS, and also investigating the effect on host range and binding to SAMD9.

This study provides a structural description of a novel conserved binding interface which can be used to investigate the interactions of the poorly understood mechanism of action of the C7 family of proteins. First, mutational screening of the residues of the proposed binding interface described in this study which result in disruption of the formation of the M062 homodimer (see Section 4.6.3.2) should be carried out to challenge the validity of this postulation. Independently, the molecular weight of M063 and M064 in solution should be measured by SEC-MALS.

Experiments testing for heterodimerisation of the *MYXV* C7 proteins can be carried out by mixing recombinantly expressed combinations of these proteins. Analytical SEC of the individual proteins and mixtures of the homologues may be sufficient to detect heterodimer formation, by observing a shift in the chromatogram of the combined proteins in comparison to the chromatogram when run individually. This method is dependent on the formation of a stable heterodimer, with a significantly shifted elution profile relative to the that of the individually run proteins, however. Calculation of the molecular weights of mixtures of the homologues by analysing these mixtures by SEC-MALS[185] or observation of bands of apparent higher molecular weight by native PAGE[186] are plausible ways to infer the formation of heterodimers, or higher order oligomers of the C7 family proteins. Analytical ultracentrifugation could also be applied to accurately determine the molecular weight, and possible differentiate between homo- and hetero-dimers of the *MYXV* C7 homologues[187].

Should residues whose mutagenesis successfully prevents the oligomerization of M062, M063 or M064 be identified, as confirmed by the biophysical methods presented above, *in vivo* and *in vitro* studies could be carried out which investigate the effect of these mutants on *MYXV* virulence. The permissibility of *MYXV* replication in usually susceptible rabbits, RK-13s, or human cell lines with these mutated homologues would indicate the impact oligomer formation has on innate immune antagonism.

As of yet, no structure of SAMD9 has been published, and predictions of its function at a structural level can only be inferred from its homology to distantly related proteins, or the location of mutants which cause disease states. The solution of the x-ray structure of SAMD9, or from co-crystallisation or a cryo-EM structure of a complex between SAMD9 and any C7 homologue and SAMD9 would provide direct evidence of the mechanism of interaction.

Furthermore, while initial surface plasmon resonance experiments investigating the binding affinity of SAMD9 and M062 have been carried out, development of a more robust biophysical assay to measure the interaction of SAMD9 and the C7 family would support evidence from *in vivo* studies.

6.4. Structural characterisation of the pyrin domain of M013

Chapter 5 describes the results of structural analyses of the PYD of M013 at the primary, secondary and tertiary structural levels. Bioinformatic analysis of M013 align with initial structural predictions that identify M013 as encoding a pyrin domain[9,74], here shown to encompass the first ~95 residues. The C terminus of M013 past the PYD is shown to be less well characterised at a secondary structural level, and possess a distinctly high predicted pI of 10.4, while the predicted pI of the M013 PYD is 4.6 (see Section 5.2). Section 5.3 describes the expression and purification of the the construct M013 1-93 4Mut, comprising the PYD. Section 5.4 presents circular dichroism spectroscopy which analysed the secondary structure of M013 1-93 4Mut and M013 1-93 Y24G, and found M013 1-93 4Mut was approximately 87% alpha helical, and appeared to have a more defined secondary structure than the single Y24G mutant.

A structural model of the M013 PYD was generated using CS-Rosetta, using chemical shift assignments gathered by NMR spectroscopy. The M013 1-93 4Mut CS-Rosetta model provides the first structural model of a vPOP generated using structural data. The first five of the six alpha helicies of the M013 PYD are well defined, while the sixth helix varies in its localisation relative to the globular, compact five alpha helical bundle formed by helicies 1-5 (see Fig. 78). The predicted surface electrostatic potential of the M013 PYD is described, with comparison to its molecular target of host antagonism: ASC. The influence of surface charged residues of the Type Ia and Type Ib DD interfaces on M013's mechanism of antagonism are described, showing the structural model supports mutagenesis studies that identified Glu28 and Asp29 as determinants of M013 antagonism of ASC[76]. Furthermore, Glu12, Asp16 and Asp44 are proposed to act as key surface charged residues of the Type Ia interface, and Lys20, Arg34 and Lys37 as key residues of forming the charged surface of the Type Ib interface of the M013 PYD.

6.4.1. Elucidation of the molecular basis of interaction between M013 and ASC/NFκB

The expression of soluble M013 required extensive mutagenesis however, and alternative structural studies with a protein closer to the native M013 may in the future advance understanding of this vPOP, and the role of the PYD and C terminal tail in their distinct modes of antagonism of host antiviral signalling. The results of the CD, NMR and bioinformatic analyses of M013 1-93 4Mut contributes corroborating structural data which supports evidence of M013 representing a *MYXV* vPOP. The NMR structural models also mostly agree with the location of the charged surface residues predicted to mediate interaction of the M013 PYD with ASC presenting on the Type Ib, and with NFκB on the Type Ia interfaces which were identified by homology modelling.

The current model of M013 antagonism of the host antiviral immune response indicates that M013 targets ASC and NFκB via structurally distinct motifs (see Fig. 82). The prediction of the specific mode of binding of M013 to its effector proteins has thus far relied on structural models of the M013 PYD by homology modelling using the PYD cellular proteins as a model[76]. With the experimental structural analyses of the M013 PYD in this study, a more comprehensive structural model is now available, which has identified novel residues which may mediate M013's antagonism.

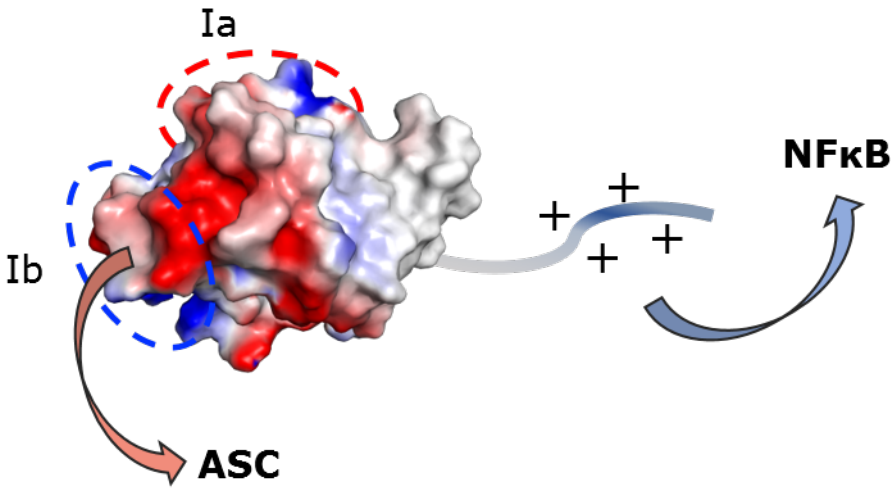


Figure 82: M013 antagonises NFκB and ASC via interaction with distinct structural domains. An electrostatic surface map of the CS-Rosetta model of M013 1-93 4Mut is shown with the DD Type Ia (red, posterior) and Type Ib (blue) interfaces circled. The positively charged C terminal tail is depicted. Interaction of M013 with ASC is mediated by the PYD of M013. Mutagenesis of residues of the Type Ib interface have been shown to disrupt the interaction of M013 and ASC. Interaction of M013 with NFκB has been shown to be dependent on the C terminal tail of M013[76]. Figure adapted from Garg *et al*[76].

6.5. **Concluding remarks**

The projects presented in this thesis have provided novel insights into the structural basis of the antagonism of host immunity by 2 distinct immunomodulatory proteins of MYXV. The M062 Lau and Mav crystal structures reveal a putative oligomerisation site conserved in the clade II C7 homologues. NMR modelling of the PYD of M013 provides structural evidence to support the prediction of residues essential for M013's antagonism of its distinct cellular targets. Based on the structural analyses presented here, further investigations which may uncover the role of the regions predicted to influence the function of M062 and M013 can be carried out. These studies may ultimately advance the understanding of the antiviral mechanisms employed by these pathogenic viral proteins, and reveal methods of immunomodulation of molecular targets and signalling pathways which prevent the mounting of an effective immune response, ultimately facilitating the spread of disease.

References

1. Haller, S. L., Peng, C., McFadden, G. & Rothenburg, S. Poxviruses and the evolution of host range and virulence. *Infect. Genet. Evol.* **21**, 15–40 (2014).
2. Buller, R. M. & Palumbo, G. J. Poxvirus pathogenesis. *Microbiol. Rev.* **55**, 80–122 (1991).
3. McFadden, G. Poxvirus tropism. *Nature Reviews Microbiology* **3**, 201–213 (2005).
4. Morens, D. M., Folkers, G. K. & Fauci, A. S. *Emerging infections in historical context*. (2004).
5. Kerr, P. J. & Best, S. M. Myxoma virus in rabbits. *Rev. sci. tech. Off. int. Epiz* **17**, 256–268 (1998).
6. Seet, B. T. *et al.* Poxviruses and immune evasion. *Annu. Rev. Immunol.* **21**, 377–423 (2003).
7. Alcami, A. & Koszinowski, U. H. Viral mechanisms of immune evasion. *Trends Microbiol.* **8**, 410–418 (2000).
8. Meng, X., Chao, J. & Xiang, Y. Identification from diverse mammalian poxviruses of host-range regulatory genes functioning equivalently to vaccinia virus C7L. *Virology* **372**, 372–383 (2008).
9. Pawlowski, K., Pio, F., Chu, Z.-L., Reed, J. C. & Godzik, A. PAAD – a new protein domain associated with apoptosis, cancer and autoimmune diseases. *Trends Biochem. Sci.* **26**, 85–87 (2001).
10. Barrett, J. W., Cao, J. X., Hota-Mitchell, S. & McFadden, G. Immunomodulatory proteins of myxoma virus. *Semin. Immunol.* **13**, 73–84 (2001).
11. Liu, J., Rothenburg, S. & McFadden, G. The poxvirus C7L host range factor superfamily. *Curr. Opin. Virol.* **2**, 758–766 (2012).
12. Bratke, K. A., McLysaght, A. & Rothenburg, S. A survey of host range genes in poxvirus genomes. *Infect. Genet. Evol.* **14**, 406–425 (2013).
13. Bratke, K. A. & Mclysaght, A. Identification of multiple independent horizontal gene transfers into poxviruses using a comparative genomics approach. *BMC Evol. Biol.* **8**, 1–13 (2008).
14. Hughes, A. L. & Friedman, R. Poxvirus genome evolution by gene gain and loss. *Mol. Phylogenet. Evol.* **35**, 186–195 (2005).
15. Matthee, C. A., Jansen Van Vuuren, B., Bell, D. & Robinson, T. J. A

Molecular Supermatrix of the Rabbits and Hares (Leporidae) Allows for the Identification of Five Intercontinental Exchanges During the Miocene. *Syst. Biol* **53**, 433–447 (2004).

16. Cameron, C. *et al.* The complete DNA sequence of myxoma virus. *Virology* **264**, 298–318 (1999).
17. Fenner, F., Day, M. F. & Woodroofe, G. M. Epidemiological Consequences of the Mechanical Transmission of Myxomatosis by Mosquitoes. *J. Hyg. (Lond).* **54**, 284–303 (1956).
18. Fenner, F., Marshall, I. D. & Woodroofe, G. M. Studies in the Epidemiology of Infectious Myxomatosis of Rabbits I . Recovery of Australian Wild Rabbits (Oryctolagus cuniculus) from Myxomatosis under Field Conditions. *Cambridge Univ. Press* **51**, 225–244 (1953).
19. Muller, A. *et al.* Partial sequencing of recent Portuguese myxoma virus field isolates exhibits a high degree of genetic stability. *Vet. Microbiol.* **140**, 161–166 (2010).
20. Morales, M. *et al.* Genome Comparison of a Nonpathogenic Myxoma Virus Field Strain with Its Ancestor, the Virulent Lausanne Strain. *J. Virol.* **83**, 2397–2403 (2009).
21. McKercher, D. G. & Saito, J. K. An Attenuated Live Virus Vaccine for Myxomatosis. *Nature* **4935**, 933–934 (1964).
22. Braun, C. *et al.* Genetic Variability of Myxoma Virus Genomes. *J. Virol.* **91**, 1–19 (2017).
23. Perkus, M. E. *et al.* Vaccinia virus host range genes. *Virology* **179**, 276–286 (1990).
24. Liu, J., Wennier, S., Zhang, L. & McFadden, G. M062 is a host range factor essential for myxoma virus pathogenesis and functions as an antagonist of host SAMD9 in human cells. *J. Virol.* **85**, 3270–82 (2011).
25. Sievers, F. & Higgins, D. G. Clustal Omega for making accurate alignments of many protein sequences. *Protein Sci.* **27**, 135–145 (2018).
26. Bond, C. & Schüttelkopf, A. ALINE: a WYSIWYG protein-sequence alignment editor for publication-quality alignments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **65**, (2009).
27. Meng, X. *et al.* Vaccinia Virus K1L and C7L Inhibit Antiviral Activities Induced by Type I Interferons. *J. Virol.* **83**, 10627–10636 (2009).
28. Sivan, G., Ormanoglu, P. R., Buehler, E. C., Martin, S. E. & Moss, B.

Identification of restriction factors by human genome-wide RNA interference screening of viral host range mutants exemplified by discovery of SAMD9 and WDR6 as inhibitors of the vaccinia virus k1L$^-$C7L$^-$ mutant. *MBio* **6**, 1–9 (2015).

29. Meng, X., Krumm, B., Li, Y., Deng, J. & Xiang, Y. Structural basis for antagonizing a host restriction factor by C7 family of poxvirus host-range proteins. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 14858–63 (2015).
30. Nounamo, B. *et al.* An interaction domain in human SAMD9 is essential for myxoma virus host-range determinant M062 antagonism of host anti-viral function. *Virology* **503**, 94–102 (2017).
31. Meng, X. *et al.* C7L family of poxvirus host range genes inhibits antiviral activities induced by type I interferons and interferon regulatory factor 1. *J. Virol.* **86**, 4538–47 (2012).
32. Meng, X. *et al.* A paralogous pair of mammalian host restriction factors form a critical host barrier against poxvirus infection. *PLoS Pathog.* **14**, 1–19 (2018).
33. Zhang, F., Meng, X., Townsend, M. B., Subbian Satheshkumar, P. & Xiang, Y. Identification of CP77 as the Third Orthopoxvirus SAMD9 and SAMD9L Inhibitor with Unique Specificity for a Rodent SAMD9L. *J. Virol.* **93**, 225–244 (2019).
34. Willer, D. O., McFadden, G. & Evans, D. H. The complete genome sequence of shope (rabbit) fibroma virus. *Virology* **264**, 319–343 (1999).
35. Needleman, S. B. & Wunsch, C. D. A general method applicable to the search for similarities in the amino acid sequence of two proteins. *J. Mol. Biol.* **48**, 443–453 (1970).
36. Barrett, J. W. *et al.* Myxoma virus M063R is a host range gene essential for virus replication in rabbit cells. *Virology* **361**, 123–132 (2007).
37. Liu, J. *et al.* Myxoma Virus M064 Is a Novel Member of the Poxvirus C7L Superfamily of Host Range Factors That Controls the Kinetics of Myxomatosis in European Rabbits. *J. Virol.* **86**, 5371–5375 (2012).
38. Asou, H. *et al.* Identification of a common microdeletion cluster in 7q21.3 subband among patients with myeloid leukemia and myelodysplastic syndrome. *Biochem. Biophys. Res. Commun.* **383**, 245–251 (2009).
39. Ma, Q., Yu, T., Ren, Y. Y., Gong, T. & Zhong, D. S. Overexpression of SAMD9 suppresses tumorigenesis and progression during non small cell

lung cancer. *Biochem. Biophys. Res. Commun.* **454**, 157–161 (2014).

40. Topaz, O. *et al. A Deleterious Mutation in SAMD9 Causes Normophosphatemic Familial Tumoral Calcinosis*. *The American Journal of Human Genetics* **79**, (2006).
41. Li, C. F. *et al.* Human sterile alpha motif domain 9, a novel gene identified as down-regulated in aggressive fibromatosis, is absent in the mouse. *BMC Genomics* **8**, (2007).
42. Sanda, C. *et al.* Differential Gene Induction by Type I and Type II Interferons and Their Combination. *J. Interf. Cytokine Res.* **26**, 462–472 (2006).
43. Hershkovitz, D. *et al.* Functional Characterization of SAMD9, a Protein Deficient in Normophosphatemic Familial Tumoral Calcinosis. *J. Invest. Dermatol.* **131**, 662–669 (2011).
44. Chefetz, I. *et al.* Normophosphatemic Familial Tumoral Calcinosis Is Caused by Deleterious Mutations in SAMD9, Encoding a TNF-α Responsive Protein. *J Invest Dermatol* **128**, 1423–1429 (2008).
45. McMahon, S. B. & Monroe, J. G. The role of early growth response gene 1 (egr-1) in regulation of the immune response. *J. Leukoc. Biol.* **60**, 159–166 (1996).
46. Wu, L. *et al.* miR-96 induces cisplatin chemoresistance in non-small cell lung cancer cells by downregulating SAMD9. *Oncol. Lett.* **11**, 945–952 (2016).
47. Nagamachi, A. *et al.* Haploinsufficiency of SAMD9L, an Endosome Fusion Facilitator, Causes Myeloid Malignancies in Mice Mimicking Human Diseases with Monosomy 7. *Cancer Cell* **24**, 305–317 (2013).
48. Narumi, S. *et al.* SAMD9 mutations cause a novel multisystem disorder, MIRAGE syndrome, and are associated with loss of chromosome 7. *Nat. Genet.* **48**, 792–797 (2016).
49. Jeffries, L. *et al.* A novel SAMD9 mutation causing MIRAGE syndrome: An expansion and review of phenotype, dysmorphology, and natural history. *Am. J. Med. Genet. Part A* **176**, 415–420 (2018).
50. Shima, H. *et al.* Two patients with MIRAGE syndrome lacking haematological features: role of somatic second-site reversion SAMD9 mutations. *J. Med. Genet.* **55**, 81–85 (2018).
51. Tesi, B. *et al.* Gain-of-function SAMD9L mutations cause a syndrome of

cytopenia, immunodeficiency, MDS, and neurological symptoms. *Blood* **129**, 2266–2279 (2017).

52. Lemos de Matos, A., Liu, J., McFadden, G. & Esteves, P. J. Evolution and divergence of the mammalian SAMD9/SAMD9L gene family. *BMC Evol. Biol.* **13**, 121 (2013).
53. Mekhedov, S. L., Makarova, K. S. & Koonin, E. V. The complex domain architecture of SAMD9 family proteins, predicted STAND-like NTPases, suggests new links to inflammation and apoptosis. *Biol. Direct* **12**, 13 (2017).
54. Qiao, F. & Bowie, J. U. The many faces of SAM. *Sci. STKE* **286**, 1–10 (2005).
55. Slupsky, C. M. *et al.* Structure of the Ets-1 pointed domain and mitogen-activated protein kinase phosphorylation site. *Proc. Natl. Acad. Sci.* **95**, 12129–12134 (1998).
56. Goroncy, A. *et al.* NMR Structure of Sterile alpha motif (SAM) domain of ephrin type-A receptor 8. *To be Publ.* doi:10.2210/PDB1UCV/PDB
57. Yang, S. Solution structure of the SAM (sterile alpha motif) domain of DLC1 (deleted in liver cancer 1). *doi.org* doi:10.2210/pdb2gyt/pdb
58. Johnson, P. E. & Donaldson, L. W. RNA recognition by the Vts1p SAM domain. *Nat. Struct. Mol. Biol.* **13**, 177–178 (2006).
59. Harada, B. T. *et al.* Regulation of Enzyme Localization by Polymerization: Polymer Formation by the SAM Domain of Diacylglycerol Kinase δ1. *Structure* **16**, 380–387 (2008).
60. Zhong, D. *et al.* The SAM domain of the RhoGAP DLC1 binds EF1A1 to regulate cell migration. *J. Cell Sci.* **122**, 414–424 (2009).
61. Knight, M. J., Leettola, C., Gingery, M., Li, H. & Bowie, J. U. A human sterile alpha motif domain polymerizome. *Protein Sci.* **20**, 1697–1706 (2011).
62. Green, J. B., Gardner, C. D., Wharton, R. P. & Aggarwal, A. K. RNA Recognition via the SAM Domain of Smaug. *Mol. Cell* **11**, 1537–1548 (2003).
63. Aviv, T. *et al.* The RNA-binding SAM domain of Smaug defines a new family of post-transcriptional regulators. *Nat. Struct. Biol.* **10**, (2003).
64. Bell, S. D., Botting, C. H., Wardleworth, B. N., Jackson, S. P. & White, M. F. The Interaction of Alba, a Conserved Archaeal Chromatin Protein, with

Sir2 and Its Regulation by Acetylation. *Science (80-.).* **296**, 148–151 (2002).

65. Goyal, M., Banerjee, C., Nag, S. & Bandyopadhyay, U. The Alba protein family: Structure and function. *Biochim. Biophys. Acta - Proteins Proteomics* **1864**, 570–583 (2016).
66. Brachmann CB *et al.* The SIR2 gene family, conserved from bacteria to humans, functions in silencing, cell cycle progression, and chromosome stability. *Genes Dev.* **9**, 2888-2902. (1995).
67. Landry, J. *et al.* The silencing protein SIR2 and its homologs are NAD-dependent protein deacetylases. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 5807–11 (2000).
68. Snider, J., Thibault, G. & Houry, W. A. The AAA+ superfamily of functionally diverse proteins. *Genome Biol.* **9**, 216 (2008).
69. Iyer, L. M., Leipe, D. D., Koonin, E. V. & Aravind, L. Evolutionary history and higher order classification of AAA+ ATPases. *J. Struct. Biol.* **146**, 11–31 (2004).
70. Das, A. K. *et al.* The structure of the tetratricopeptide repeats of protein phosphatase 5: implications for TPR-mediated protein-protein interactions. *EMBO J.* **17**, 1192–9 (1998).
71. Blatch, G. L. & Lässle, M. The tetratricopeptide repeat: A structural motif mediating protein-protein interactions. *BioEssays* **21**, 932–939 (1999).
72. Lu, A. & Wu, H. Structural mechanisms of inflammasome assembly. *FEBS J.* **282**, 435–444 (2015).
73. Kuehl, F. A. & Egan, R. W. Prostaglandins, arachidonic acid, and inflammation. *Science (80-.).* **210**, 978 LP – 984 (1980).
74. Staub, E., Dahl, E. & Rosenthal, A. The DAPIN family: a novel domain links apoptotic and interferon response proteins. *Trends Biochem. Sci.* **26**, 83–85 (2001).
75. Johnston, J. B. *et al.* A poxvirus-encoded pyrin domain protein interacts with ASC-1 to inhibit host inflammatory and apoptotic responses to infection. *Immunity* **23**, 587–598 (2005).
76. Garg, R. R. *et al.* Myxoma virus M013 protein antagonizes NF-B and inflammasome pathways via distinct structural motifs. *J. Biol. Chem.* **294**, 8480–8489 (2019).
77. Dorfleutner, A. *et al.* A Shope Fibroma virus PYRIN-only protein modulates

the host immune response. *Virus Genes* **35**, 685–694 (2007).

78. Masumoto, J. *et al.* ASC, a Novel 22-kDa Protein, Aggregates during Apoptosis of Human Promyelocytic Leukemia HL-60 Cells*. *J. Biol. Chem.* **274**, 33835–33838 (1999).
79. Masumoto, J. *et al.* Expression of Apoptosis-associated Speck-like Protein Containing a Caspase Recruitment Domain, a Pyrin N-terminal Homology Domain-containing Protein, in Normal Human Tissues. *J. Histochem. Cytochem.* **49**, 1269–1275 (2001).
80. Richards, N. *et al.* Interaction between Pyrin and the Apoptotic Speck Protein (ASC) Modulates ASC-induced Apoptosis*. *J. Biol. Chem.* **276**, 39320–39329 (2001).
81. Srinivasula, S. M. *et al.* The PYRIN-CARD protein ASC is an activating adaptor for caspase-1. *J. Biol. Chem.* **277**, 21119–21122 (2002).
82. Fernandes-Alnemri, T., Yu, J.-W., Wu, J., Datta, P. & Alnemri, E. S. AIM2 activates the inflammasome and cell death in response to cytoplasmic DNA. *Nature* **458**, 509–513 (2009).
83. Ohtsuka, T. *et al.* ASC is a Bax adaptor and regulates the p53-Bax mitochondrial apoptosis pathway. *Nat. Cell Biol.* **6**, 121 (2004).
84. De Alba, E. Structure and Interdomain Dynamics of Apoptosis-associated Speck-like Protein Containing a CARD (ASC). *J. Biol. Chem.* **284**, 32932–32941 (2009).
85. Martinon, F., Hofmann, K. & Tschopp, J. The pyrin domain: a possible member of the death domain-fold family implicated in apoptosis and inflammation. *Curr. Biol.* **10**, (2001).
86. Fairbrother, W. J. *et al.* The PYRIN domain: A member of the death domain-fold superfamily. *Protein Sci.* **10**, 1911–1918 (2002).
87. Liepinsh, E. *et al.* The Death-domain Fold of the ASC PYRIN Domain, Presenting a Basis for PYRIN/PYRIN Recognition. *J. Mol. Biol.* **332**, 1155–1163 (2003).
88. Nanson, J. D., Kobe, B. & Ve, T. Death, TIR, and RHIM: Self-assembling domains involved in innate immunity and cell-death signaling. *J. Leukoc. Biol.* **105**, 363–375 (2019).
89. Weber, C. H. & Vincenz, C. The death domain superfamily: a tale of two interfaces? *Trends Biochem. Sci.* **26**, 475–481 (2001).
90. Park, H. H. *et al.* Death Domain Assembly Mechanism Revealed by Crystal

Structure of the Oligomeric PIDDosome Core Complex. *Cell* **128**, 533–546 (2007).
91. Lu, A. *et al.* Unified Polymerization Mechanism for the Assembly of ASC-Dependent Inflammasomes. *Cell* **156**, 1193–1206 (2014).
92. Jin, T., Perry, A., Smith, P., Jiang, J. & Xiao, T. S. Structure of the Absent in Melanoma 2 (AIM2) Pyrin Domain Provides Insights into the Mechanisms of AIM2 Autoinhibition and Inflammasome Assembly. *J. Biol. Chem.* **288**, 13225–13235 (2013).
93. Lu, A. *et al.* Crystal Structure of the F27G AIM2 PYD Mutant and Similarities of its Self-association to DED/DED Interactions. *J Mol Biol* **426**, 1420–1427 (2014).
94. Dorfleutner, A., Chu, L. & Stehlik, C. Inhibiting the inflammasome: One domain at a time. *Immunol. Rev.* **265**, 205–216 (2015).
95. Qin, H. *et al.* Structural basis of procaspase-9 recruitment by the apoptotic protease-activating factor 1. *Nature* **399**, 549–557 (1999).
96. Rahman, M. M. & McFadden, G. Myxoma Virus Lacking the Pyrin-Like Protein M013 Is Sensed in Human Myeloid Cells by both NLRP3 and Multiple Toll-Like Receptors, Which Independently Activate the Inflammasome and NF- B Innate Response Pathways. *J. Virol.* **85**, 12505–12517 (2011).
97. Rahman, M. M., Mohamed, M. R., Kim, M., Smallwood, S. & McFadden, G. Co-regulation of NF-kB and inflammasome-mediated inflammatory responses by myxoma virus pyrin domain-containing protein M013. *PLoS Pathog.* **5**, (2009).
98. Stehlik, C. *et al.* The PAAD/PYRIN-Family Protein ASC Is a Dual Regulator of a Conserved Step in Nuclear Factor B Activation Pathways. *J. Exp. Med.* **196**, 1605–1615 (2002).
99. Dubendorf, J. W. & Studier, F. W. Controlling basal expression in an inducible T7 expression system by blocking the target T7 promoter with lac repressor. *J. Mol. Biol.* **219**, 45–59 (1991).
100. GSL Biotech. SnapGene software. Available at: www.snapgene.com.
101. Gasteiger, E. *et al.* Protein identification and analysis tools on the ExPASy server. in *The proteomics protocols handbook* 571–607 (Springer, 2005).
102. Agilent. *QuikChange II Site-Directed Mutagenesis Kit Instruction Manual*.
103. Mlynek, G. *et al.* The Center for Optimized Structural Studies (COSS)

platform for automation in cloning, expression, and purification of single proteins and protein-protein complexes. *Amino Acids* **46**, 1565–1582 (2014).

104. Bradford, M. M. A rapid and sensitive for the quantitation of microgram quantities of protein-dye binding. *Anal Biochem* **72**, 248–254 (1976).
105. Jarmila Jancarik, B. & Kim, S. *Sparse matrix sampling: a screening method for crystallization of proteins*. *J. Appl. Cryst* **24**, (1991).
106. Van Duyne, G. D., Standaert, R. F., Karplus, A., Schreiber, S. L. & Clardy, J. Atomic Structures of the Human Immunophilin FKBP-12 Complexes with FK506 and Rapamycin. *J. Mol. Biol.* **229**, 105–124 (1993).
107. M9 salt solution. *Cold Spring Harb. Protoc.* **2009**, pdb.rec11973-pdb.rec11973 (2009).
108. Winn, M. D. *et al.* Overview of the CCP4 suite and current developments. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 235–242 (2011).
109. Stein, N. CHAINSAW: a program for mutating pdb files used as templates in molecular replacement. *J. Appl. Crystallogr.* **41**, 641–643 (2008).
110. Schwarzenbacher, R., Godzik, A., Grzechnik, S. K., Jaroszewski, L. & IUCr. The importance of alignment accuracy for molecular replacement. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 1229–1236 (2004).
111. McCoy, A. J. *et al.* Phaser Crystallographic Software. *J. Appl. Crystallogr. Cryst* **40**, 658–674 (2007).
112. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr.* **D**, 486–501 (2010).
113. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **68**, 352–367 (2012).
114. Painter, J., Merritt, E. A. & IUCr. TLSMD web server for the generation of multi-group TLS models. *J. Appl. Crystallogr.* **39**, 109–111 (2006).
115. Painter, J., Merritt, E. A. & IUCr. Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **62**, 439–450 (2006).
116. Chen, V. B. *et al.* MolProbity : all-atom structure validation for macromolecular crystallography. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 12–21 (2010).
117. Zheng, H. *et al.* Validation of metal-binding sites in macromolecular

structures with the CheckMyMetal web server. *Nat. Protoc.* **9**, 156–170 (2014).
118. Zheng, H. *et al.* CheckMyMetal: a macromolecular metal-binding validation tool. *Acta Crystallogr. Sect. D Struct. Biol.* **73**, 223–233 (2017).
119. Cavanagh, J., Fairbrother, W. J., Palmer, I. & Skelton, N. J. *Protein NMR spectroscopy : principles and practice.* (Elsevier Science, 1995).
120. Simpson, R. J. SDS-PAGE of Proteins. *CSH Protoc.* (2006). doi:10.1101/pdb.prot4313
121. Rasband, W. ImageJ [PC program], version 1.25 s. *Natl. Institutes Heal. Bethesda, Maryl.* (2002).
122. Derewenda, Z. S. The use of recombinant methods and molecular engineering in protein crystallization. *Methods* **34**, 354–363 (2004).
123. Derewenda, Z. S. Application of protein engineering to enhance crystallizability and improve crystal properties. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 604–615 (2010).
124. Derewenda, Z. S. Rational Protein Crystallization by Mutational Surface Engineering. *Structure* **12**, 529–535 (2004).
125. Schwede, T. F., Bädeker, M., Langer, M., Rétey, J. & Schulz, G. E. Homogenization and crystallization of histidine ammonia-lyase by exchange of a surface cysteine residue. *Protein Eng.* **12**, 151–153 (1999).
126. Glover, J. N. M. & Harrison, S. C. Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA. *Nature* **373**, 257–261 (1995).
127. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
128. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7**, 539–539 (2014).
129. Waterhouse, A. *et al.* SWISS-MODEL: Homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
130. Roy, R. N. *et al. Buffers for the Physiological pH Range: Acidic Dissociation Constants of Zwitterionic Compounds (ACES and CHES) in Water from 5 to 55 °C*. (1997).
131. Bergfors, T. *Protein crystallization: techniques, strategies, and tips*. (International University Line, La Jolla California USA, 1999).

132. Merritt, E. A. X-ray Absorption Edges. Available at: http://skuld.bmsc.washington.edu/scatter/AS_periodic.html. (Accessed: 3rd May 2019)
133. Evans, G., Pettifer, R. F. & IUCr. CHOOCH : a program for deriving anomalous-scattering factors from X-ray fluorescence spectra. *J. Appl. Crystallogr.* **34**, 82–86 (2001).
134. Dale, G. E., Oefner, C. & Dõarcy, A. The protein as a variable in protein crystallization. doi:10.1016/S1047-8477(03)00041-8
135. Evans, P. R. & IUCr. An introduction to data reduction: space-group determination, scaling and intensity statistics. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 282–292 (2011).
136. SBGrid. Albula.
137. Sheldrick, G. M. SHELX applications to macromolecules. in *Direct methods for solving macromolecular structures* 401–411 (Springer, 1998).
138. Adams, P. D. *et al.* PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **66**, 213–221 (2010).
139. McNicholas, S., Potterton, E., Wilson, K. S., Noble, M. E. M. & IUCr. Presenting your structures: the CCP4mg molecular-graphics software. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **67**, 386–394 (2011).
140. Freeman, T. C. PepDraw. *Tulane university* (2015).
141. PerkinElmer. ChemDraw.
142. Müller, P., Köpke, S. & Sheldrick, G. M. Is the bond-valence method able to identify metal atoms in protein structures? *Acta Crystallogr. D. Biol. Crystallogr.* **59**, 32–7 (2003).
143. Krissinel, E. & Henrick, K. Inference of Macromolecular Assemblies from Crystalline State. *J. Mol. Biol.* **372**, 774–797 (2007).
144. Schrödinger, L. PyMOL The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC.
145. Krissinel, E. & Henrick, K. Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. *Acta Crystallogr. Sect. D Biol. Crystallogr.* **60**, 2256–2268 (2004).
146. Sun, Z., Liu, Q., Qu, G., Feng, Y. & Reetz, M. T. Utility of B-Factors in Protein Science: Interpreting Rigidity, Flexibility, and Internal Motion and Engineering Thermostability. *Chem. Rev.* **119**, 1626–1665 (2019).

147. Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M. & Barton, G. J. Jalview Version 2--a multiple sequence alignment editor and analysis workbench. *Bioinformatics* **25**, 1189–1191 (2009).
148. Morris, R. J. *et al. Breaking good resolutions with ARP/wARP*. *J. Synchrotron Rad* **11**, (2004).
149. Carugo, O. How large B-factors can be in protein crystal structures. *BMC Bioinformatics* **19**, 61 (2018).
150. Bahadur, R. P. & Zacharias, M. The interface of protein-protein complexes: Analysis of contacts and prediction of interactions. *Cell. Mol. Life Sci. Rev.* **65**, (2008).
151. Price, S. R. & Nagai, K. *Protein engineering as a tool for crystallography Introduction Use of protein engineering to improve crystals*. *Current Opinion in Biotechnology* **6**, (1995).
152. Janin, J. Specific versus non-specific contacts in protein crystals. *Nat. Struct. Biol.* **4**, 973–974 (1997).
153. Dasgupta, S., Iyer, G. H., Bryant, S. H., Lawrence, C. E. & Bell, J. A. Extent and nature of contacts between protein molecules in crystal lattices and between subunits of protein oligomers. *Proteins Struct. Funct. Genet.* **28**, 494–514 (1997).
154. Stewart, D. E., Sarkar, A. & Wampler, J. E. Occurrence and role of cis peptide bonds in protein structures. *J. Mol. Biol.* **214**, 253–260 (1990).
155. Wu, W. J. & Raleigh, D. P. Local control of peptide conformation: Stabilization of cis proline peptide bonds by aromatic proline interactions. *Biopolymers* **45**, 381–394 (1998).
156. Ganguly, H. K., Majumder, B., Chattopadhyay, S., Chakrabarti, P. & Basu, G. Direct Evidence for CH···π Interaction Mediated Stabilization of Pro-cisPro Bond in Peptides with Pro-Pro-Aromatic motifs. *J. Am. Chem. Soc.* **134**, 4661–4669 (2012).
157. Wu, W.-J. & Raleigh, D. P. *Conformation: Stabilization of cis Proline Peptide Bonds by Aromatic Proline Interactions*. *Biopolymers* **45**, (John Wiley & Sons, Inc, 1998).
158. Nishio, M., Umezawa, Y., Fantini, J., Weiss, M. S. & Chakrabarti, P. CH–p hydrogen bonds in biological macromolecules. *Phys. Chem. Chem. Phys* **16**, 12648 (2014).
159. Brandl, M., Weiss, M. S., Jabs, A., Sühnel, J. & Hilgenfeld, R. C-h···π-

interactions in proteins. *J. Mol. Biol.* **307**, 357–377 (2001).

160. Pappas, D. J. *et al.* Longitudinal system-based analysis of transcriptional responses to type I interferons. *Physiol. Genomics* **38**, 362–371 (2009).
161. Schreiber, M., Rajarathnam, K. & Mcfadden, G. *Myxoma Virus T2 Protein, a Tumor Necrosis Factor (TNF) Receptor Homolog, Is Secreted as a Monomer and Dimer That Each Bind Rabbit TNF, but the Dimer Is a More Potent TNF Inhibitor**. (1996).
162. Ramsey-Ewing, A. L. & Moss, B. Complementation of a Vaccinia Virus Host-Range K1L Gene Deletion by the Nonhomologous CP77 Gene. *Virology* **222**, 75–86 (1996).
163. Sö, J., Biegert, A. & Lupas, A. N. The HHpred interactive server for protein homology detection and structure prediction. doi:10.1093/nar/gki408
164. Drozdetskiy, A., Cole, C., Procter, J. & Barton, G. J. JPred4: a protein secondary structure prediction server. *Nucleic Acids Res.* **43**, W389–W394 (2015).
165. Buchan, D. W. A. & Jones, D. T. The PSIPRED Protein Analysis Workbench: 20 years on. *Nucleic Acids Res.* **47**, W402–W407 (2019).
166. Kelly, S. M., Jess, T. J. & Price, N. C. How to study proteins by circular dichroism. *Biochim. Biophys. Acta - Proteins Proteomics* **1751**, 119–139 (2005).
167. Louis-Jeune, C., Andrade-Navarro, M. A. & Perez-Iratxeta, C. Prediction of protein secondary structure from circular dichroism using theoretically derived spectra. *Proteins Struct. Funct. Bioinforma.* **80**, 374–381 (2012).
168. Holzwarth2, G. & Doty, P. The Ultraviolet Circular Dichroism of Polypeptides1. *1Journal Am. Chem. Soc.* **87**, 218–228 (1965).
169. Shen, Y. *et al.* Consistent blind protein structure generation from NMR chemical shift data. *PNAS* **105**, 4685–4690 (2008).
170. Shen, Y., Robert, A. E., Ae, V. & Baker, D. De novo protein structure generation from incomplete chemical shift assignments. *J. Biomol. NMR* **43**, 63–78 (2009).
171. Weber, C. H. & Vincenz, C. A docking model of key components of the DISC complex: Death domain superfamily interactions redefined. *FEBS Lett.* **492**, 171–176 (2001).
172. Lin, S.-C., Lo, Y.-C. & Wu, H. Helical assembly in the MyD88:IRAK4:IRAK2 complex in TLR/ IL-1R signaling. *Nature* **465**, 885–890 (2010).

173. Wang, L. *et al.* The Fas–FADD death domain complex structure reveals the basis of DISC assembly and disease mutations. *Nat Struct Mol Biol* **17**, 1324–1329 (2010).
174. Bradley, P. Toward High-Resolution de Novo Structure Prediction for Small Proteins. *Science (80-.).* **309**, 1868–1871 (2005).
175. Morrone, S. R. *et al.* Cooperative assembly of IFI16 filaments on dsDNA provides insights into host defense strategy. *PNAS* E26–E71 (2013). doi:10.1073/pnas.1313577111
176. Rahman, M. M., Mohamed, M. R., Kim, M., Smallwood, S. & Mcfadden, G. Co-Regulation of NF-kB and Inflammasome-Mediated Inflammatory Responses by Myxoma Virus Pyrin Domain-Containing Protein M013. *PLOS Pathog.* **5**, (2009).
177. Le, H. T. & Harton, J. A. Pyrin- and CARD-only Proteins as Regulators of NLR Functions. *Front. Immunol.* **4**, 275 (2013).
178. Research, P. *et al.* The PYRIN domain-only protein POP3 inhibits ALR inflammasomes and regulates responses to infection with DNA viruses. *Nat. Immunol. Vol.* **15**, (2014).
179. Sypula, J., Wang, F., Ma, Y., Bell, J. & McFadden, G. Myxoma virus tropism in human tumor cells. *Gene Ther. Mol. Biol.* **8**, 103–114 (2004).
180. Llambi, F. *et al.* A Unified Model of Mammalian BCL-2 Protein Family Interactions at the Mitochondria. *Mol. Cell* **44**, 517–531 (2011).
181. Shamas-Din, A., Kale, J., Leber, B. & Andrews, D. W. Mechanisms of action of Bcl-2 family proteins. *Cold Spring Harb. Perspect. Biol.* **5**, 1–21 (2013).
182. Meng, X. & Xiang, Y. RNA granules associated with SAMD9-mediated poxvirus restriction are similar to antiviral granules in composition but do not require TIA1 for poxvirus restriction. *Virology* **529**, 16–22 (2019).
183. Liu, J. & McFadden, G. SAMD9 is an innate antiviral host factor with stress response properties that can be antagonized by poxviruses. *J. Virol.* **89**, 1925–31 (2015).
184. Miller, C. L. Stress granules and virus replication. *Future Virol.* **6**, 1329–1338 (2011).
185. Carroll, S. F., Barbieri, J. T. & Collier, R. J. Dimeric Form of Diphtheria Toxin: Purification and Characterization. *Biochemistry* **25**, 2425–2430 (1986).
186. Braz, V. A. & Howard, K. J. Separation of Protein Oligomers by Blue Native

Gel Electrophoresis. *Anal Biochem* **388**, 170–172 (2009).

187. Cole, J. L., Lary, J. W., P. Moody, T. & Laue, T. M. Analytical Ultracentrifugation: Sedimentation Velocity and Sedimentation Equilibrium. *Methods Cell Biol.* **84**, 143–179 (2008).