Factors involved in the oligomerisation of the cyanide dihydratase from *Bacillus pumilus* C1

by

Andani Errol Mulelu

A thesis submitted in fulfilment of the requirements
for the degree of DOCTOR OF PHILOSOPHY
in Medical Biochemistry

In the Department of Integrative Biomedical Sciences
UNIVERSITY OF CAPE TOWN
December 2016

Supervisor: Prof B.T. Sewell

Co-supervisor: Dr J.D. Woodward

DECLARATION

I declare that *Factors involved in the oligomerisation of the cyanide dihydratase from Bacillus pumilus C1* is my own work, that it has not been submitted before for any degree or examination in any other university, and that all the sources I have used or quoted have been indicated and acknowledged as complete references.

Andani Errol Mulelu

Signed by candidate

Signed: ________________________________ December 2016

ABSTRACT

The cyanide dihydratase enzyme from *Bacillus pumilus* C1 (CynD$_{pum}$) is a member of the nitrilase superfamily and is known to specifically catalyse the conversion of cyanide into formic acid and ammonia. This enzyme is a good candidate for bioremediation of cyanide waste but the high alkaline pH of the cyanide waste water poses a problem in that it inactivates the wild type enzyme and therefore improvement of stability is required in order to synthesize an effective enzyme. Over the pH range of 6–8 the enzyme exists as short 18-subunit spirals which associate to form long, more stable helical fibres at pH 5.4. The reason for this pH dependent transition is not fully understood but it is hypothesized to be due to changes in the charge of histidine residues. The aim of this project is to obtain a high resolution structure of CynD$_{pum}$, relate this to its function, and investigate the role of the histidines in oligomerisation with aid of the structure. Using Cryo-electron microscopy techniques a three dimensional reconstruction structure of purified CynD$_{pum}$ was obtained at a resolution of ~5Å. By flexibly fitting a CynD$_{pum}$ homology model into this high resolution structure we were able to identify amino acid residues involved in oligomerisation and stability as well as the role of the histidines, with aid from additional mutagenesis studies. Interactions at the C-interfacial region were shown to play the most crucial role in oligomerisation and included the His71-Asp275 and Arg67-Asp275 interactions. Mutations at His128, His184, His241 and His285 were shown to affect the oligomerisation of the enzyme by indirectly disrupting interactions at the interfacial regions. The Q86R+H305K+H308K+H323K mutations were shown to increase the stability of the CynD$_{pum}$ by introducing a stronger arginine-arginine interaction at the D interfacial region and a new strong interaction at the C-terminal region.

ACKNOWLEDGEMENTS

I wish to thank, first and foremost, my supervisor Professor Trevor Sewell (UCT) for his knowledge, outstanding support and encouragement which helped me through all the obstacles in the completion of this research work.

I would like to thank my co-supervisor Dr Jeremy Woodward from the Electron Microscope Unit (UCT) who despite his heavy schedule took time to share valuable insights and help with my project and thesis write-up.

I'm highly indebted to Prof. Dr. Achilleas Frangakis from the Buchmann Institute for Molecular Life Sciences (Frankfurt, Germany) for his advice, collaboration and access to his FEI Titan Electron Microscope.

My thanks and appreciation go to Dr Michael Benedik, Mary Abou Nader and Jason Park from Texas A&M University (USA) for their advice, collaboration and providing insights that helped in the interpretation on my findings.

Special thanks to Jason van Rooyen from ICTS (UCT) for his insights and IT support that allowed us to obtain high resolution structures.

I would like to thank our overseas collaborators Dr Jean Watermeyer (Birkbeck College, London), Dr Dan Clare (Birkbeck College) and Mr Deryck Mills (The Max Planck Institute of Biophysics) for their assistance in the EM sample preparation and data collection.

I would like to thank Dr Brandon Weber from the Structural Biology Research Unit (UCT) for his shared valuable insights and advise while I was performing experiments in the lab.

I would like to thank Mr Mohammed Jaffer from the Electron Microscope Unit (UCT) for his time and patience while helping me find my way around our microscopes.

I would like to thank Dr Mariette Smart from the Centre for Bioprocess Engineering Research (UCT) for sharing her expertise and equipment during the thermostability experiments.

I would like to express my gratitude towards my family for their encouragement and support throughout the duration of my studies.

Lastly, I would like to thank the National Research Foundation and the CSIR for providing me with the funds to help me carry out my research.

NOMENCLATURE

KCN	Potassium cyanide
CynD	Cyanide dihydratase
CynD_{pum}	Cyanide dihydratase from *Bacillus pumilus*
$\text{CynD}_{\text{stut}}$	Cyanide dihydratase from *Pseudomonas stutzeri* AK61
βaS	β-alanine synthase
WT	Wild type
FSC	Fourier shell correlation
DPR	Differential phase residual
PDB	Protein data bank
DCase	*N*-carbamyl-*D*-amino acid amidohydrolase
EM	Electron microscopy/microscope
IHRSR	Iterative Helical Real Space Reconstruction
LA	Lurai agar
NAD^+	Nicotinamide adenine dinucleotide
TEM	Transmission electron microscopy/microscope
CCD	Charge coupled device
SDS-PAGE	Sodium dodecyl sulphate polyacrylamide gel electrophoresis
SNR	Signal-to-noise-ratio
IPTG	isopropyl-β-D-thiogalactopyranoside
TRIS	Tris(hydroxymethyl)aminomethane
DSF	Differential scanning fluorimetry
MDFF	Molecular dynamics flexible fitting

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER I

LITERATURE REVIEW

1.1 Nitrilase classification

Nitrilases are thiol enzymes that are mostly involved in natural product biosynthesis and post translational modification by performing a wide variety of non-peptide carbon-nitrogen hydrolysis reactions (Pace and Brenner, 2001). These nitrilases are ubiquitous in nature and are found in most species of plants, animals, fungi and many prokaryotes with more than 180 known members (Pace and Brenner, 2001; Gordon et al., 2013). The reactions performed by the nitrilases can be classified into four types which include the nitrilase, amidase, carbamylase and the amidase condensation reaction (Fig 1.1). These nitrilase reactions are generally chemo-, regio-, or enantio- selective and because of these characteristics, nitrilases have become commercially interesting in the chemical industry (Kiziak et al., 2005). Nitrilases that carry out the nitrilase reaction are particularly interesting as they can directly hydrolyse nitriles which are essential intermediates in the chemical synthesis of various products (Heinemann et al., 2003).

Nitrilases are collectively classified into one group known as the nitrilase superfamily. Members of the nitrilase superfamily are further divided into 13 branches (Fig. 1.2) as proposed by Pace and Brenner (2001). This classification of nitrilases was achieved on the basis of global and structure-based sequence analysis. Although members of the nitrilase superfamiliy have been historically classified as nitrilase-related based on sequence analysis, only branch 1 members have been identified to carry out nitrile-hydrolysing activity to form corresponding acids and ammonia (reaction a, Fig. 1.1). Members of branch 2–11 have apparent amidase, carbamylase or amide-condensation activities (reaction b, c or d; Fig. 1.1) and insufficient information is known about the last four branches of the superfamily (branch 10–13).

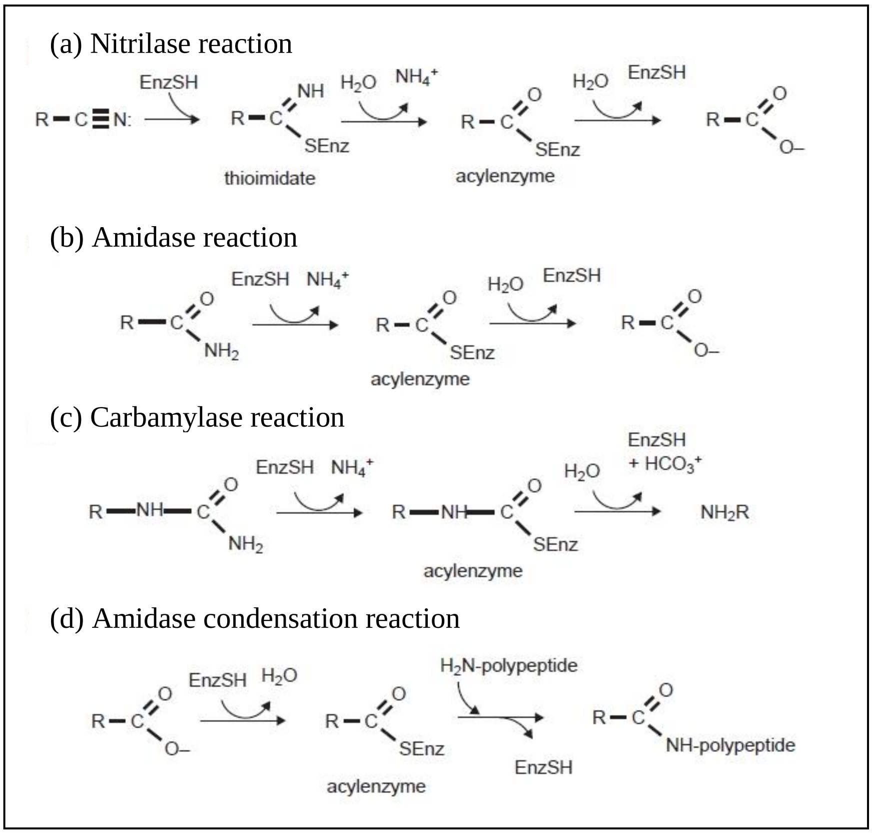


Figure 1.1. Four types of reactions carried out by the nitrilase superfamily members. Branch 1 nitrilases are known to carry out the nitrilase reaction (reaction **a)**. Branches 2–11 members have amidase activities (reaction **b**) with the exception of branch 5 and 6 members which carry out the carbamylase reaction (reaction **c**), and branch 9 members which carry out the amide-condensation reaction (reaction **d**). (Pace and Brenner, 2001)

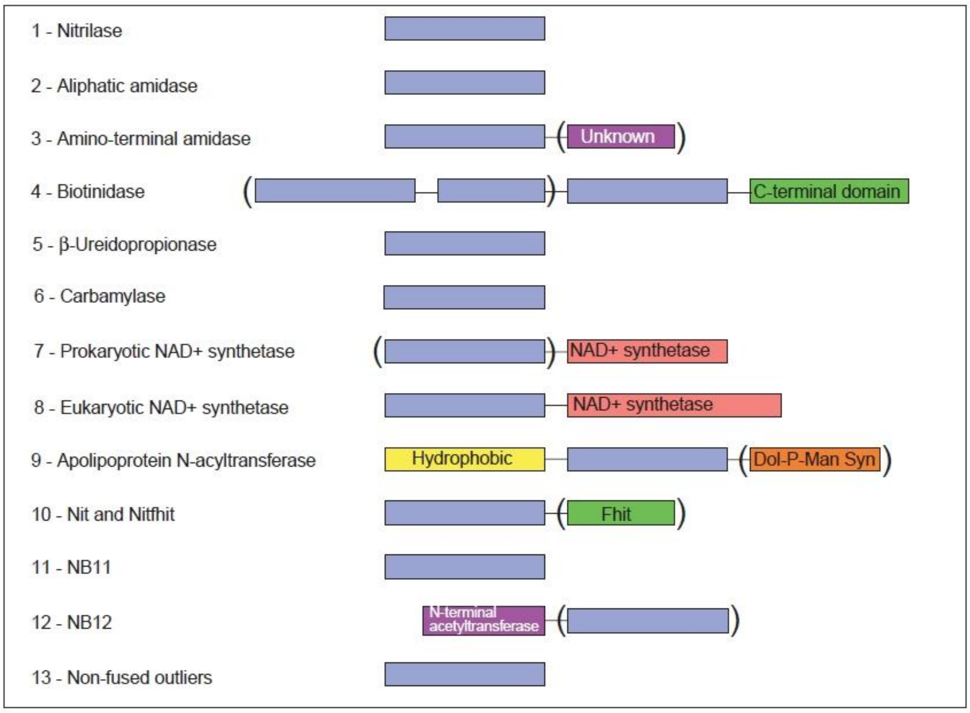


Figure 1.2. Domain structures of the 13 branches of the nitrilase superfamily. Parentheses denote domains found in only some members of the branch (Pace and Brenner, 2001).

1.2 Applications of the nitrilase enzymes

In industry, nitrilases have the potential to be used as biocatalysts due to the importance of nitrile intermediates in the chemical synthesis of various products. Like most biocatalysts nitrilases could potentially overcome some of the problems faced by traditional chemical methods of converting nitriles, like the harsh conditions required, the formation of toxic by-products and low chemical selectivity of reactions. There have been several successes in the application of nitrile-converting organisms and isolated enzymes as biocatalysts. These successes include the use of R. rhodochrous K22 to catalyse the conversion of adiponitrile to 5-cyanovaleric acid during the production of nylon-6 (Godtfredsen et al., 1985), the use of Acremonium sp. to catalyse the selective mono-hydrolysis of trans 1,4-dicyano cyclohexane during the production of Tranexamic acid, a homeostatic drug (Nishise et al., 1987) and the use of the nitrilase from Rhodococcus sp. NDB 1165 by the company Lanza in their production of nicotinic acid (vitamin B3) to allow them to use cyanopyridine as a substrate instead of the more hazardous nicotine (Prasad et al. 2007). Some industrial applications of nitrilases are still under development and these include the use of the nitrilases from Rhodococcus rhodochrous J1 (J1 nitrilase) to aid in the production of acrylic acid (Nagasawa et al. 2000), the use of a re-engineered nitrilase from Acidovorax facilis 72W in the production of glycolic acid from glycolonitrile (Wu et al. 2008), the use of nitrilases for modifying the polymer polyacrylonitrile to enhance dye uptake (Matamá et al. 2007), and the use of the nitrilase from Alcaligenes faecalis in the producuction of pure (R)-(-)-mandelic acid which is used in the production of pharmaceuticals such as penicillins, cephalosporins, antitumor drugs, and anti-obesity agents (Xue et al. 2011).

Nitrilases have the potential of being used in the bioremediation of industrial wastewater that contains synthetic nitrile compounds. The release of nitrile compounds into the environment must to be controlled due to the nitriles toxic, carcinogenic and mutagenic nature, and nitrilases could potentially offer cheap and environmentally friendly solution for the detoxification of these harmful nitriles. There have been a few successes were mixed cultures of bacteria containing different nitrile hydrolysing enzymes were used to metabolize effluent containing acrylonitrile, furmaronitrile, succinonitrile, and similar nitriles by growing the bacteria on these components of waste (Wyatt and Knowles et al., 1995). Besides industrial waste there have been accidental toxic spills of nitriles around the world. These spills include the cyanide spills of 2006 in Ghana, 2006 in Czech Republic and others in Romania, Australia, China and the

Philippines (Kovac, 2000). Other sources of environmental cyanide contamination include the marine ornamental industry where fish are stunned with sprayed cyanide to allow for easy collection of the aquarium ornamental fish (Jones and Steven, 1997).

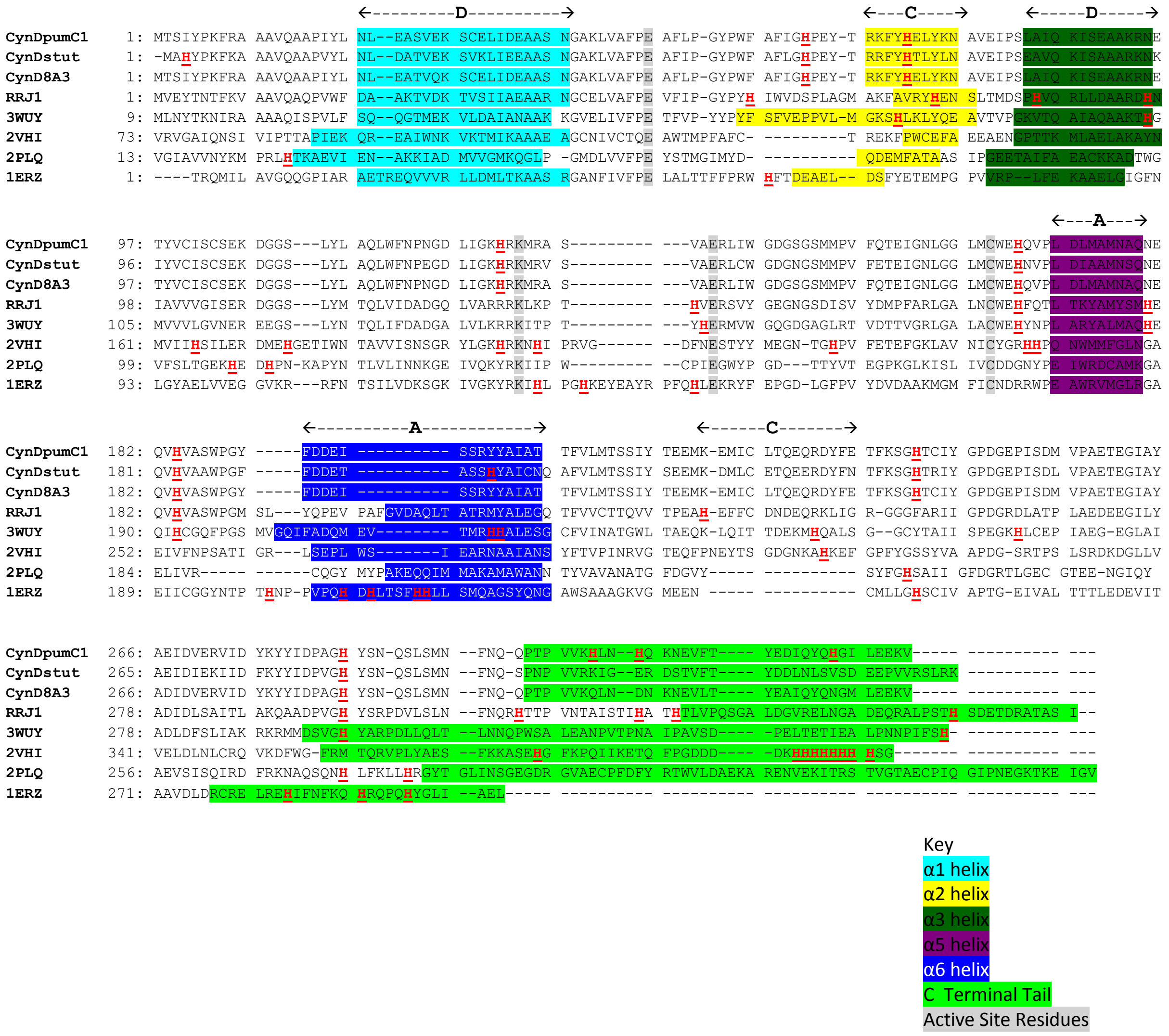
The incorporation of the herbicide degrading nitrilase gene from *Klebsiella pneumoniae* subsp. *ozaenae* into commercial plants has been one of the most successful application of a nitrilase. The incorporation of this bacterial gene has led to the successful construction of tobacco and cotton plants that are resistant to high levels of nitrile-herbicides such as bromoxynil (Stalker et al., 1988, Stalker et al., 1996). The incorporation of this and other similar nitrilase genes into plants could potentially allow the use of new nitrile-herbicides as weed killers and prevent be accumulation of hazardous nitrile-herbicides in plant based foods.

Several members of the nitrilase superfamily are found in mammalian genomes which encode eight nitrilase enzymes (Barglow et al., 2008). These include the vanins, biotinidases and pantetheinases involved in vitamin synthesis, and the glutaminase domain of glutamine-dependent NAD$^+$ synthatase which generates ammonia from glutamine for NAD$^+$ synthesis (Barglow et al., 2008). Studies done on some of these nitrilases could potentially lead to insights that will help in drug design. One potential drug target in the treatment of tuberculosis is the branch 7 glutamine-dependent NAD$^+$ synthetase. By exploiting *Mycobacterium tuberculosis's* dependence on NAD$^+$ synthetase in its NAD$^+$ recycling pathways as opposed to the human NAD+ recycling pathways which are not all dependent on NAD$^+$ synthetase, one could exclusively target *Mycobacterium tuberculosis* by starving the pathogen of NAD$^+$ (LaRonde-LeBlanc et al., 2009). Another potential drug target from the nitrilase superfamily is the two amidases from branch 2, AmiE and AmiF, found in *Helicobacter pylori*, a human gastric pathogen that colonises the mucus membrane of the stomach resulting in various stomach disorders such as ulcers (Skouloubris et al., 1997, Skouloubris et al., 2001; Bury-Mone et al., 2003, Dunne et al., 2014). AmiE and AmiF play a role in protecting the bacterium from the harsh acidic stomach conditions thus making these amidases potential drug targets in the eradication of the pathogen (Dunne et al., 2014). β-Ureidopropionases and the prominent nitrilase-1 (Nit1) and nitrilase-2 (Nit2), which have been shown to exert effects on cell growth in mice and possibly serve as tumour suppressor genes, are furthermore examples of nitrilase superfamily members found in the mammalian genome (Lin et al., 2007; Barglow et al., 2008). More insights about these nitrilases could potentially lead to the design of tumour suppressor drugs.

1.3 Structure and homology of nitrilases

The nitrilase superfamily enzymes share a significant structural homology despite varying sequence conservation (Fig. 1.3) and differing substrate affinities (Pace and Brenner, 2001). There are over 30 crystal structures of nitrilase superfamily members that have been solved to date, a number that keeps growing (Table 1.1). The first few crystal structures to be solved include two members of branch 10 enzymes; *Caenorhabditis elegans* NitFhit (PDB ID:1EMS; Pace et al., 2000) and *Saccharomyces cerevisiae* Nit3 (PDB ID:1F89; Kumaran et al., 2003), and one member from branch 6, *Agrobacterium N*-carbamyl-D-amino acid amidohydrolase (PDB ID:1ERZ; Nakai et al., 2000). These first crystal structures all displayed a consistent αββα sandwich monomer architecture, a monomer fold which was later found to be conserved among the nitrilase superfamily enzymes (Thuku et al., 2009). Across all well characterised nitrilase superfamily structures, the αββα monomers have been shown to stack together to form active dimer complexes. These dimer complexes often further associate to form larger complexes which differ across the members of the superfamily (Table 1.1).

The *C. elegans* NitFhit and *Agrobacterium N*-carbamyl-D-amino acid amidohydrolase form active tetramer while *S. cerevisiae* Nit3 protein forms dimers (Table 1.1). After the structural elucidation of the *C. elegans* NitFhit, *S. cerevisiae* Nit3 and *Agrobacterium N*-carbamyl-D-amino acid amidohydrolase structures, an increasing number of nitrilase structures began to accumulate in the Protein Data Bank (PDB) with some structures possessing bound ligands and mutations which have provided further insight into the structure and the catalytic mechanism of some of these and other nitrilases (Table 1.1). Although the number of nitrilase superfamily structures has been increasing over the years, it is only recently that the crystal structure of a Branch 1 nitrilases has been solved. Branch 1 nitrilases have proved difficult to crystallise and it is postulated to be due to the enzymes lack of symmetry that conforms to that which can be packed into crystals due to the high degree of oligomerisation of monomers which results in the formation of long complex quaternary structures. To date the Nit6803 nitrilase from *Synechocystis sp.* Strain PCC6803 is the only branch 1 crystal structures that has been solved (Zhang et al., 2014). Before this branch 1 crystal structure was solved, methods of structure prediction such as homology modelling and electron microscopy (EM) were used. Homology modelling has successfully helped in the structural model prediction of some branch 1 nitrilases such as the cyanide dihydratase from *Pseudomonas stutzeri* AK61 (Sewell et al., 2005), the nitrilase from *Rhodococcus rhodochrous J1* (Thuku et al., 2007) and the cyanide dihydratase from *Bacillus pumilus* C1 (Scheffer, 2006). These models were docked into low resolution 3D EM maps to gain key insights into the oligomeric structure of the branch 1 nitrilases. These branch 1 nitrilases were shown to have a two-fold symmetric structure which conserves a αββααββα and are homodimeric (Sewell et al., 2005; Thuku et al., 2007).



```
                                      ←---------D----------→                                                   ←---C----→       ←-----D-----→
CynDpumC1   1: MTSIYPKFRA AAVQAAPIYL NL--EASVEK SCELIDEAAS NGAKLVAFPE AFLP-GYPWF AFIGHPEY-T RKFYHELYKN AVEIPSLAIQ KISEAAKRNE
CynDstut    1: -MAHYPKFKA AAVQAAPVYL NL--DATVEK SVKLIEEAAS NGAKLVAFPE AFIP-GYPWF AFLGHPEY-T RRFYHTLYLN AVEIPSEAVQ KISAAARKNK
CynD8A3     1: MTSIYPKFRA AAVQAAPIYL NL--EATVQK SCELIDEAAS NGAKLVAFPE AFLP-GYPWF AFIGHPEY-T RKFYHELYKN AVEIPSLAIQ KISEAAKRNE
RRJ1        1: MVEYTNTFKV AAVQAQPVWF DA--AKTVDK TVSIIAEAAR NGCELVAFPE VFIP-GYPYH IWVDSPLAGM AKFAVRYHEN SLTMDSPHVQ RLLDAARDHN
3WUY        9: MLNYTKNIRA AAAQISPVLF SQ--QGTMEK VLDAIANAAK KGVELIVFPE TFVP-YYPYF SFVEPPVL-M GKSHLKLYQE AVTVPGKVTQ AIAQAAKTHG
2VHI       73: VRVGAIQNSI VIPTTAPIEK QR--EAIWNK VKTMIKAAAE AGCNIVCTQE AWTMPFAFC- ---------T REKFPWCEFA EEAENGPTTK MLAELAKAYN
2PLQ       13: VGIAVVNYKM PRLHTKAEVI EN--AKKIAD MVVGMKQGLP -GMDLVVFPE YSTMGIMYD- ---------- QDEMFATAAS IPGEETAIFA EACKKADTWG
1ERZ        1: ----TRQMIL AVGQQGPIAR AETREQVVVR LLDMLTKAAS RGANFIVFPE LALTTFFPRW HFTDEAEL-- DSFYETEMPG PVVRP--LFE KAAELGIGFN

                                                                                                                ←---A---→
CynDpumC1  97: TYVCISCSEK DGGS---LYL AQLWFNPNGD LIGKHRKMRA S--------- ---VAERLIW GDGSGSMMPV FQTEIGNLGG LMCWEHQVPL DLMAMNAQNE
CynDstut   96: IYVCISCSEK DGGS---LYL AQLWFNPEGD LIGKHRKMRV S--------- ---VAERLCW GDGNGSMMPV FETEIGNLGG LMCWEHNVPL DIAAMNSQNE
CynD8A3    97: TYVCISCSEK DGGS---LYL AQLWFNPNGD LIGKHRKMRA S--------- ---VAERLIW GDGSGSMMPV FQTEIGNLGG LMCWEHQVPL DLMAMNAQNE
RRJ1       98: IAVVVGISER DGGS---LYM TQLVIDADGQ LVARRRKLKP T--------- ---HVERSVY GEGNGSDISV YDMPFARLGA LNCWEHFQTL TKYAMYSMHE
3WUY      105: MVVVLGVNER EEGS---LYN TQLIFDADGA LVLKRRKITP T--------- ---YHERMVW GQGDGAGLRT VDTTVGRLGA LACWEHYNPL ARYALMAQHE
2VHI      161: MVIIHSILER DMEHGETIWN TAVVISNSGR YLGKHRKNHI PRVG------ --DFNESTYY MEGN-TGHPV FETEFGKLAV NICYGRHHPQ NWMMFGLNGA
2PLQ       99: VFSLTGEKHE DHPN-KAPYN TLVLINNKGE IVQKYRKIIP W--------- --CPIEGWYP GD---TTYVT EGPKGLKISL IVCDDGNYPE IWRDCAMKGA
1ERZ       93: LGYAELVVEG GVKR---RFN TSILVDKSGK IVGKYRKIHL PGHKEYEAYR PFQHLEKRYF EPGD-LGFPV YDVDAAKMGM FICNDRRWPE AWRVMGLRGA

                               ←----------A------------→                ←------C-------→
CynDpumC1 182: QVHVASWPGY -----FDDEI ---------- SSRYYAIAT  TFVLMTSSIY TEEMK-EMIC LTQEQRDYFE TFKSGHTCIY GPDGEPISDM VPAETEGIAY
CynDstut  181: QVHVAAWPGF -----FDDET ---------- ASSHYAICNQ AFVLMTSSIY SEEMK-DMLC ETQEERDYFN TFKSGHTRIY GPDGEPISDL VPAETEGIAY
CynD8A3   182: QVHVASWPGY -----FDDEI ---------- SSRYYAIAT  TFVLMTSSIY TEEMK-EMIC LTQEQRDYFE TFKSGHTCIY GPDGEPISDM VPAETEGIAY
RRJ1      182: QVHVASWPGM SL---YQPEV PAFGVDAQLT ATRMYALEGQ TFVVCTTQVV TPEAH-EFFC DNDEQRKLIG R-GGGFARII GPDGRDLATP LAEDEEGILY
3WUY      190: QIHCGQFPGS MVGQIFADQM EV-------- TMRHHALESG CFVINATGWL TAEQK-LQIT TDEKMHQALS G--GCYTAII SPEGKHLCEP IAEG-EGLAI
2VHI      252: EIVFNPSATI GR---LSEPL WS-------I EARNAAIANS YFTVPINRVG TEQFPNEYTS GDGNKAHKEF GPFYGSSYVA APDG-SRTPS LSRDKDGLLV
2PLQ      184: ELIVR----- ------CQGY MYPAKEQQIM MAKAMAWANN TYVAVANATG FDGVY----- ---------- SYFGHSAII  GFDGRTLGEC GTEE-NGIQY
1ERZ      189: EIICGGYNTP THNP-PVPQH DHLTSFHHLL SMQAGSYQNG AWSAAAGKVG MEEN ----- ---------- CMLLGHSCIV APTG-EIVAL TTTLEDEVIT

CynDpumC1 266: AEIDVERVID YKYYIDPAGH YSN-QSLSMN --FNQ-QPTP VVKHLN--HQ KNEVFT---- YEDIQYQHGI LEEKV----- ---------- ---
CynDstut  265: AEIDIEKIID FKYYIDPVGH YSN-QSLSMN --FNQ-SPNP VVRKIG--ER DSTVFT---- YDDLNLSVSD EEPVVRSLRK ---------- ---
CynD8A3   266: ADIDVERVID YKYYIDPAGH YSN-QSLSMN --FNQ-QPTP VVKQLN--DN KNEVLT---- YEAIQYQNGM LEEKV----- ---------- ---
RRJ1      278: ADIDLSAITL AKQAADPVGH YSRPDVLSLN --FNQRHTTP VNTAISTIHA THTLVPQSGA LDGVRELNGA DEQRALPSTH SDETDRATAS I--
3WUY      278: ADLDFSLIAK RKRMMDSVGH YARPDLLQLT --LNNQPWSA LEANPVTPNA IPAVSD---- -PELTETIEA LPNNPIFSH- ---------- ---
2VHI      341: VELDLNLCRQ VKDFWG-FRM TQRVPLYAES --FKKASEHG FKPQIIKETQ FPGDDD---- -DKHHHHHHH HSG------- ---------- ---
2PLQ      256: AEVSISQIRD FRKNAQSQNH LFKLLHRGYT GLINSGEGDR GVAECPFDFY RTWVLDAEKA RENVEKITRS TVGTAECPIQ GIPNEGKTKE IGV
1ERZ      271: AAVDLDRCRE LREHIFNFKQ HRQPQHYGLI --AEL----- ---------- ---------- ---------- ---------- ---------- ---
```

Figure 1.3. Sequence alignment of the cyanide dihydratases **CynD$_{\mathrm{pum}}$C1** from *Bacillus pumilus C1* (Meyers et al., 1993; Jandhyala et al., 2003), **CynD8A3** from *Bacillus pumilus* stain 8A3 (Eicher, 2007) and **CynD$_{\mathrm{stu}}$** from *Pseudomonas stutzeri AK61* (Sewell et al., 2003) with the sequences of the four homologs: **RRJ1**, the nitrilase from *Rhodococcus rhodochrous J1* (Thuku et al., 2007); **3WUY**, the extended helix forming nitrilase Nit6803 from *Synechocystis sp. Strain* PCC6803 (Zhang et al., 2014), **2VHI**, the C-shaped β-alanine-synthase (βaS) from *Drosophila melanogaster* (Lundgren et al., 2008), **2PLQ**, a non-spiral-forming aliphatic amidase from *Geobacillus pallidus* RAPc8 (Kimani et al., 2007), **1ERZ**, the *N*-carbanyl-D-amino acid amidohydrolase from *Agrobacterium* sp. KNK712 (Nakai et al., 2000). The multiple sequence alignment was constructed using ClustalW (Thompson et al., 1994). The approximate regions of the interacting surfaces based on CynD$_{\mathrm{pum}}$ are indicated in the top line (**A**, **C**, and **D**-surface). The conserved interfacial alpha helices α1, α2, α3, α5, and α6 are highlighted in turquoise, yellow, green, violet and blue respectively. The C-terminal regions are highlighted in bright green and the catalytic residue are highlighted in grey. The histidine residues are highlighted and underlined in red.

Table 1.1 List of members of the nitrilase superfamily with solved atomic structures. The atomic structures have been arranged using their designated PDB ID and structures from the same publication have been grouped together.

PDB ID	Enzyme	Species	Superfamily Branch	Resolution	Oligomeric form	Reference
1EMS	Nit-fragile histidine triad fusion protein	*Caenorhabditis elegans*	10	2.80 Å	Tetramer	Pace et al., 2000
1ERZ	*N*-carbanyl-D-amino acid amidohydrolase	*Agrobacterium* sp. KNK712	6	1.70 Å	Tetramer	Nakai et al., 2000
1F89	Nit 3	*Saccharomyces cerevisiae*	10	2.40 Å	Dimer	Kumaran et al., 2003
1FO6	*N*-carbamyl-D-amino-acid amidohydrolase	*Agrobacterium radiobacter*	10	1.95 Å	Tetramer	Wang et al., 2001
1J31	Hypothetical protein PH0642 from Pyrococcus horikoshii	*Pyrococcus horikoshii*	Unclassified	1.60 Å	Dimer	Sakai et al., 2004
1UF4 1UF5 1UF7 1UF8	*N*-carbamyl-D-amino acid amidohydrolase	*Agrobacterium* sp.	10	2.15Å 1.6 0Å 1.9 0Å 1.80Å	Tetramer	Hashimoto, H (unpublished)
2GGK 2GGL	N-carbamoyl-D-amino-acid amidohydrolase (D-NCAase)	*Agrobacterium tumefaciens*	6	2.30 Å 2.40 Å	Tetramer	Chiu et al., 2006
2DYU 2DYV 2E2K 2E2L	Formamidase (AmiF)	*Helicobacter pylori*	2	1.75 Å 2.00 Å 2.50 Å 2.29 Å	Hexamer	Hung et al., 2007
2E11	XC1258	*Xanthomonas campestris*	Unclassified	2.30 Å	Dimer	Chin et al., 2007
2PLQ	Aliphatic Amidase	*Geobacillus pallidus* RAPc8	2	1.90 Å	Hexamer	Kimani et al., 2007
2UXY	Aliphatic Amidase	*Pseudomonas aeruginosa*	2	1.25 Å	Hexamer	Andrade et al., 2007
2VHH 2VHI	Beta-alanine synthase	*Drosophila melanogaster*	5	2.80 Å 3.30 Å	Octamer	Lundgren et al., 2008
2W1V	Nitrilase-2	*Mus musculus*	10	1.49 Å	Dimer	Barglow et al., 2008
3DLA	Glutamine-dependent NAD^+ synthetase	*Mycobacterium tuberculosis*	7	2.35 Å	Octamer	LaRonde-LeBlanc et al., 2009

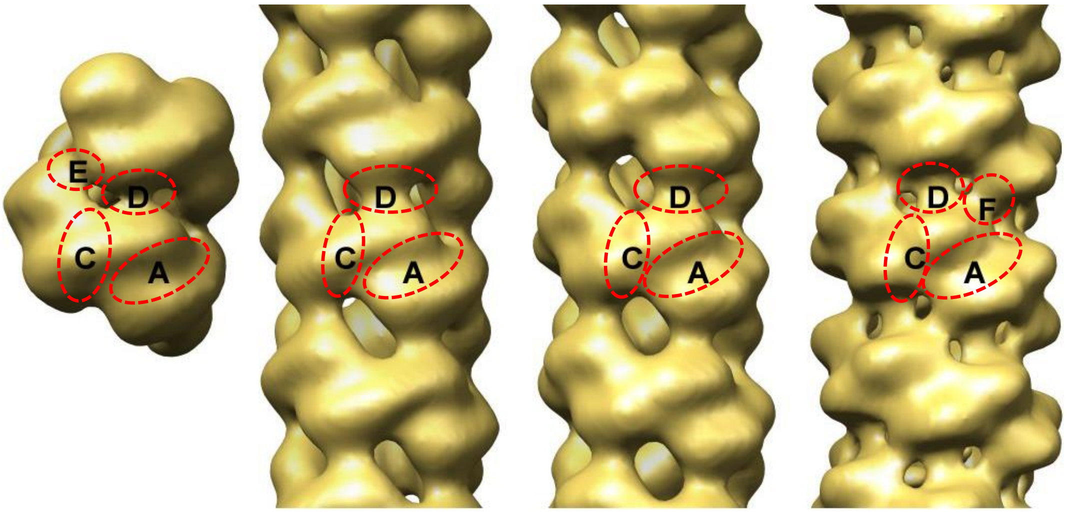
Table 1.1. (Continued)

PDB ID	Enzyme	Species	Superfamily Branch	Resolution	Oligomeric form	Reference
3IW3	PaNit	*Pyrococcus abyssi*	Unclassified	1.80 Å	Dimer	Raczynska et al., 2010
3IVZ				1.57 Å		
3KLC				1.76 Å		
3KI8				1.60 Å		
3HKX	Amidase	*Nesterenkonia sp. ANI*	2	1.66 Å	Dimer	Nel et al., 2011
3ILV	Glutamine dependent NAD$^+$ synthetase	*Cytophaga hutchinsonii*	7 or 8	1.79 Å	Dimer	Palani K. (unpublished)
3P8K	CN-hydrolase SA0302	*Staphylococcus aureus*	10	1.7 Å	Dimer	Gordon et al., 2013
4KZF	Aliphatic Amidase	*Geobacillus pallidus* RAPc8	2	1.85 Å	Hexamer	Weber et al., 2013
4LF0				1.10 Å		
4GYN				1.90 Å		
4GYL				1.90 Å		
4HG3	Nit2	*Saccharomyces cerevisiae*	10	1.93 Å	Dimer	Lui et al., 2013
4HG5				1.91 Å		
4HGD				2.04 Å		
3WUY	Nit6803	*Synechocystis sp. PCC6803*	1	3.10 Å	Decamer	Zhang et al., 2014
4IZW	Amidase	*Nesterenkonia sp. ANI*	2	1.60 Å	Dimer	Kimani S.W. (unpublished)
4IZV				1.65 Å		
4IZU				1.40 Å		
4IZT				1.92 Å		
4IZS				1.44 Å		
5JQN	Amidase	*Nesterenkonia sp. ANI*	2	1.19 Å	Dimer	Sewell B.T. (unpublished)

1.4 Oligomerisation and interfacial regions

A majority of the studied branch 1 nitrilases originate from microbial species and they have been shown to form spiral quaternary structures in solution (Thuku et al., 2009). Some of these branch 1 nitrilases that have been studied include cyanide dihydratases and cyanide hydratases which either form helices of variable length or short terminating spirals which have a specific number of subunits (Fig. 1.4). The cyanide dihydratase from *Pseudomonus stutzeri AK61* ($\text{CynD}_{\text{stut}}$) forms a 14-subunit spiral (Sewell et al., 2003) while the homologous enzymes from *Bacillus pumilus* C1 (CynD_{pum}) and *Bacillus pumilus* 8A3, form 18 and 22 subunit spirals, respectively (Jandhyala et al., 2003; Eicher, 2007). Short spirals are formed by the nitrilases from *Rhodococcus rhodochrous* J1 (Thuku et al., 2007) and *Fusarium solani* O1 (Vejvoda et al., 2008), although the exact number of subunits has not been determined. The cyanide hydratases from *Gloeocercospora sorghi* (Woodward et al., 2008) and *Neurospora crassa* (Dent et al., 2008) occur as long, variable length helices.

It is understood that the quaternary structures formed by some of the branch 1 nitrilases are due to the initial association of monomers to form dimers followed by the association of these dimers to form the complex quaternary structures. Modelling of branch 1 nitrilases based on solved nitrilase superfamiliy crystal structures, coupled with docking into 3D electron microscopy reconstructions, has aided in the identification of interfacial regions or surfaces that are involved in the association of monomers to form oligomers (Sewell et al., 2003, 2005; Scheffer, 2006; Thuku et al., 2007, Woodward et al., 2008, Dent et al., 2009, Williamson et al., 2010). Branch 1 nitrilases generally have six of these surfaces namely the A, B, C, D and F-surfaces (Fig. 1.4–1.6). The association of subunits at different interfaces results in a variety of oligomeric shapes that are observed in members of this nitrilase branch (Fig. 1.4).



N / A
variable
Cyanide dihydratase from
Pseudomonas stutzeri, pH8

D_1 $S_{4.7}$
$\Delta\phi$ = -76.6 °, Δz = 15.9 Å
Cyanide dihydratase from
Bacillus pumilus C1, pH5.4

D_1 $S_{4.9}$
$\Delta\phi$ = -73.65°, Δz = 15.85 Å
Truncated WT nitrilase from
Rhodococcus rhodochrous J1

D_1 $S_{5.4}$
$\Delta\phi$ = -65.96°, Δz = 13.48 Å
Cyanide hydratase from
Gloeocercospora sorghi

Figure 1.4. Three dimensional electron microscopic reconstuction of the microbial nitrilases studied at low resolution. The structures include the cyanide dihydratase from *Pseudomonas stutzeri* (Sewell et al., 2003), cyanide dihydratase from *Bacillus pumilus* C1 (Scheffer, 2006), the nitrilase from *Rhodococcus rhodochrous* J1 (Thuku et al., 2007) and the cyanide hydratase from *Gloeocercospora sorghi* (Woodward et al., 2008). Calculated helical symmetry and interfacial surfaces shown in bold (circled in red). The figure was produced using UCSF Chimera (Pettersen et al., 2004).

The A-surface and C-surface are the only surfaces that have been visualized at atomic resolution (Wang et al., 2001; Chin et al., 2007; Kimani et al., 2007, Lundgren et al., 2008; Zhang et al., 2014). Evidence of the existence of other surfaces was shown when homology models were docked into low resolution 3D electron microscopic reconstructions (Sewell et al., 2003, 2005; Thuku et al., 2007; Woodward et al., 2008; Dent et al., 2009; Williamson et al., 2010). Based on homology modelling using related nitrilase superfamily atomic structures it has been demonstrated that the A-surface comprises of two conserved alpha helices, α5 and α6 which correspond to residues 171–179 and 192–205 respectively, based on the CynD_{pum} homology model (Park et al., 2016)(Fig. 1.6). The association of these two α helices has been shown to be common in both crystalline and spiral structures (Thuku et al., 2009). The A-surface is located between two monomers and interactions at this surface are postulated to be the basis of dimerisation. Although all the solved nitrilase superfamily structures possess the two conserved A-surface α helices, it has been demonstrated that the intermolecular ionic and hydrophobic interactions across this surface varies in different organisms (Wang et al., 2001; Chin et al., 2007; Kimani et al., 2007).

The C-surface is located approximately at right angles to the A-surface and the interactions across this surface has been postulated to be the key to the elongation of the spiral quaternary structures (Sewell et al., 2003). Through homology modelling the C-surface has been shown to consist primarily of two insertions in the spiral forming nitrilases, relative to the non-spiral-forming members of the superfamily (Thuku et al., 2009). The first insertion comprises of residues 55–72 based on the CynD_{pum} homology model (C-surface region 1) and is shown to occur at the α2 helix region of the structure (Fig 1.5 and 1.6). The second insertion comprises of residue 222–235 based on the CynD_{pum} homology model (C-surface region 2) and forms a small loop. The importance of the C-surface interactions has been demonstrated by the structural comparison of the insertions from members of the superfamily that form different oligomers (Park et al., 2016). The crystal structure of the β-alanine synthase (βaS) from Drosophila melanogaster (PDB ID: 2VHI; Lundgren et al., 2008) which forms C-shaped spiral oligomers demonstrates a different region 2 loop conformation when compared to the corresponding loop from the Nit6803 from *Synechocystis sp.* Strain PCC6803 (PDB ID: 3WUY; Zhang et al., 2014) which forms an extended helix with six-fold symmetry. The βaS region 2 loop is more extended than the corresponding loop from Nit6803 and interacts with a two-fold related loop from an adjacent monomers, an intermolecular interaction that is absent in the Nit6803 structure (Fig.1.5). The βaS region 1 insertion is eight residues shorter than that of Nit6803 and shows no intermolecular interactions across the C-surface while the Nit6803 structure demonstrates potential interactions between two-fold related region 1 amino acids from an adjacent monomers (Fig.1.5). The difference in the C-surface interactions and oligomerisation between these crystal structures illustrates how interactions across the C-surface can dictate the overall oligomerisation of the nitrilase superfamily members. When mutations were introduced at the C-surface insertions it was demonstrated that mutations at region 1 affected the enzyme's activity and oligomerisation more than region 2 (Park et al., 2016), thus demonstrating that the interaction across the CynD_{pum} C-surface may closely resemble that of Nit6803 than βaS, although a high resolution structure of CynD_{pum} is required to confirm this.

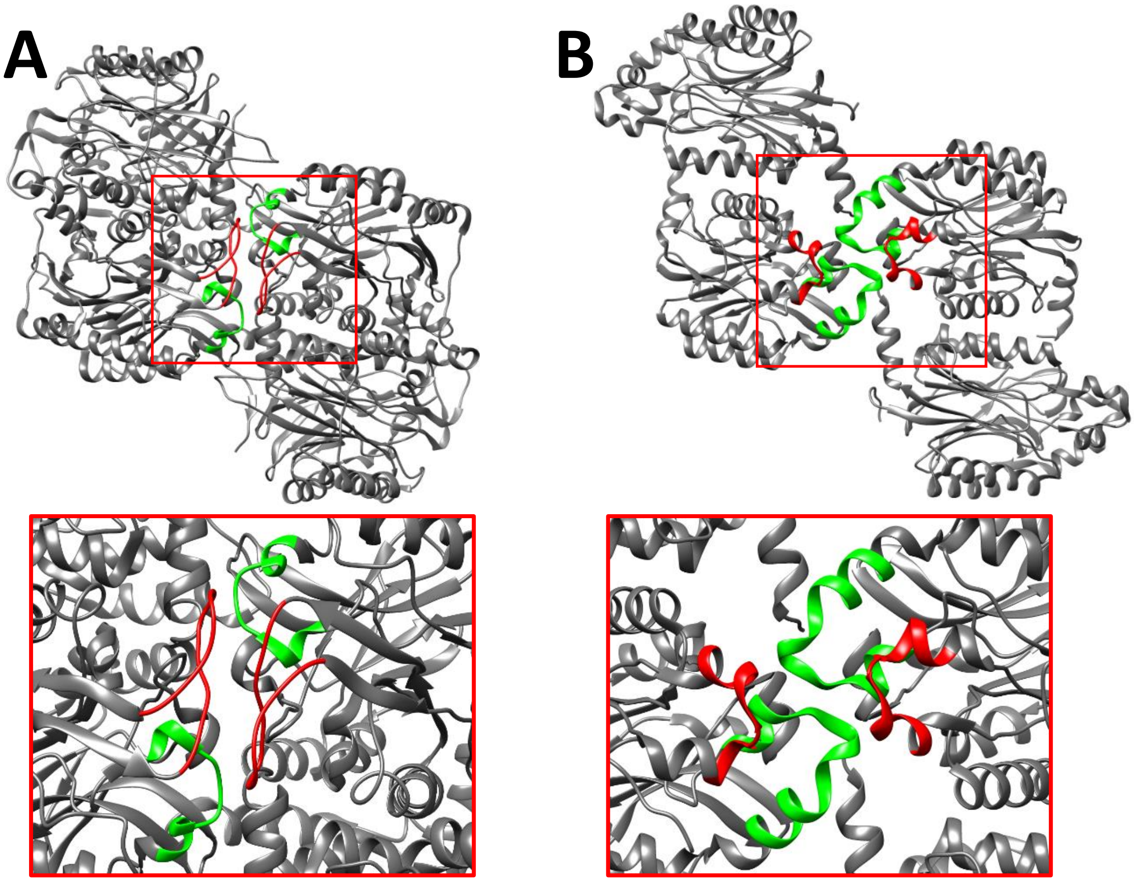


Figure 1.5. Comparison between the region 1 (green) and region 2 (red) insertions and the C-surface of the crystal structure of **(A)** the C-shaped spiral forming β-alanine synthase (βaS) from *Drosophila melanogaster* (PDB ID: 2VHI; Lundgren et al., 2007) and **(B)** the extended helix forming Nit6803 nitrilase from *Synechocystis sp.* Strain PCC6803 (PDB ID: 3WUY; Zhang et al., 2014).

The D-surface interactions occur across the groove of the spiral or helix and only exist when the spiral completes one turn (Fig. 1.6). Nitrilases have little amino acid sequence conservation around the D-surface but have a mixture of positively and negatively charged residues suggesting the existence of strong electrostatic interactions across this surface (Fig. 1.3). Based on homology modelling it is suggested that these electrostatic interactions are two-fold symmetric and help maintain the elongating assembly during the oligomerisation of helical nitrilases (Sewell et al., 2005; Thuku et al., 2007). The D-surface comprises of two nitrilase-conserved alpha helices, α1 and α3, which corresponds to residue 22–39 and residue 83–95 respectively, based on the $\mathrm{CynD_{pum}}$ homology model (Fig. 1.6).

The F-surface is not a common surface among helical nitrilases and if present, is located near the D-surface. The F-surface is postulated to be involved in the further stabilization of the two-fold symmetric interactions in enzymes that possess it such as the cyanide hydratases from

Gloeocercospora sorghi (Woodward et al., 2008). It is postulated that interactions occurring across the ‘C’ ‘D’ and ‘F’ surfaces influence the helical symmetry of helical nitrilases where the enzymes with the ‘F’ surface show an increase in helical twist (Fig 1.4). The E-surface is different from the other surfaces in that its contributors are asymmetric and arise from different regions of the subunit (Fig. 1.4). The E-surface only occurs in terminating spirals (Sewell et al., 2003).

Microbial nitrilases have been shown to possess an extended C-terminal sequence that is about 40–100 amino acid longer than that of other related homologous nitrilase structures (Fig. 1.3). For helical nitrilases, this C-terminal region has not been visualized at high resolution but it is postulated to be located in the inner surface of the helical structures based on distinct electron densities observed on the inner surface of 3D EM reconstruction models of some of the helical nitrilases (Sewell et al., 2003; Thuku et al., 2007)(Fig. 1.6C). The C-terminal tail has been shown to play an important role in oligomerisation and activity in branch 1 nitrilases. The CynD$_{\text{pum}}$ enzyme and the *Rhodococccus rhodochrous* J1 nitrilase have been shown to lose activity after the truncation of the C-terminal by 28 and 55 residues respectively (Sewell et al., 2005, Thuku et al., 2007). The removal of the C-terminal residues has also been shown to be required for the formation of long regular helices by the *Rhodococccus rhodochrous* J1 nitrilase (Thuku et al., 2007). The truncation the C-terminal region from the *Rhodococccus rhodochrous* J1 nitrilase has been demonstrated to occur naturally by autolysis, a process which has been recently shown to also occur with the Nit6803 from *Synechocystis* sp. Strain PCC6803 (Zhang et al., 2014). The reason for this post-transcriptional cleavage is not fully understood but demonstrates the importance of the C-terminal region.

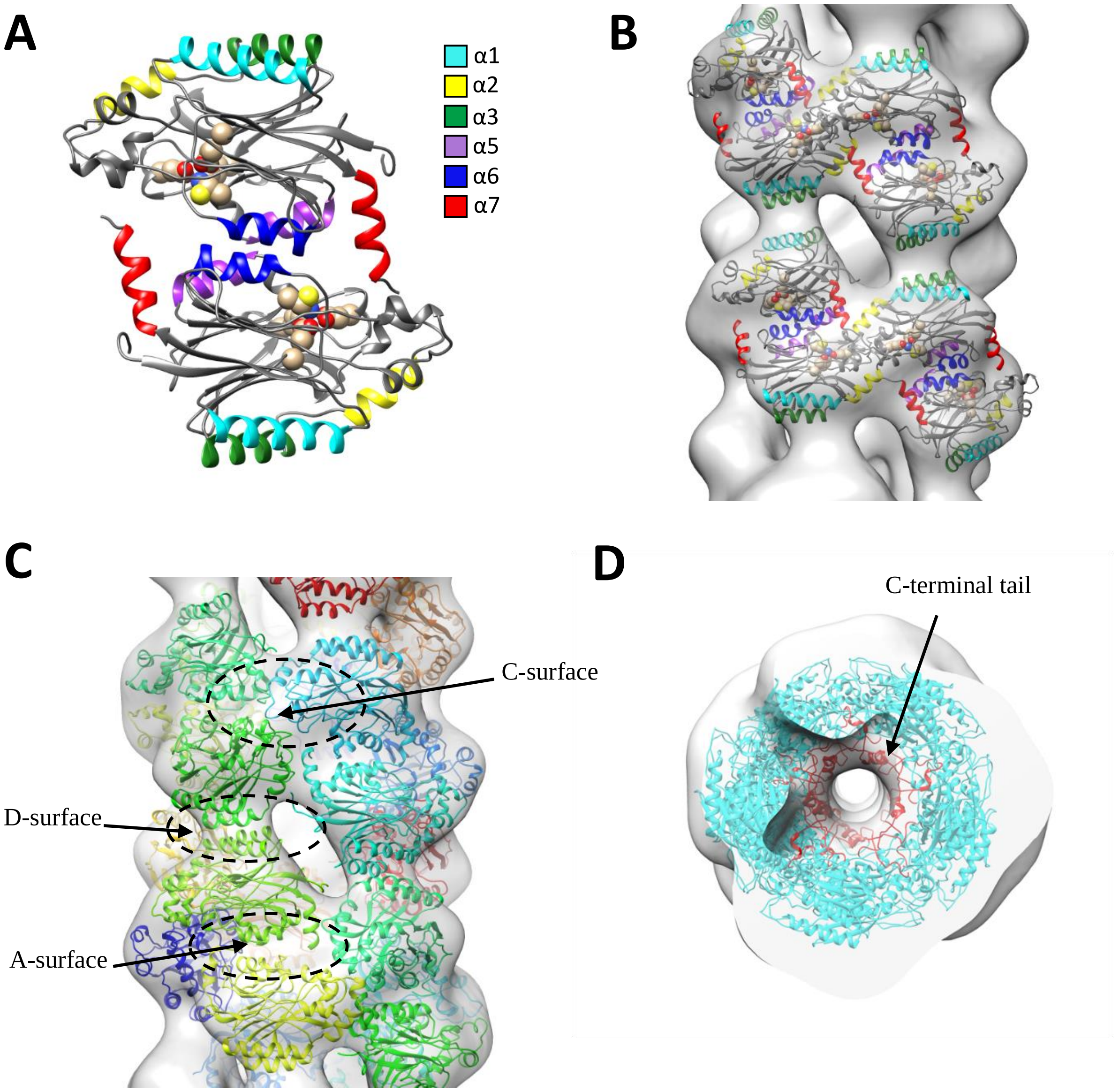


Figure 1.6. (**A**) View of a dimer homology model of the cyanide dihydratase from *B. pumilus* C1 nitrilase (Park et al., 2016) the conserved interfacial alpha helices α1, α2, α3, α5, α6, and α7 are highlighted in turquoise, yellow, green, violet, blue and red respectively. The conserved catalytic residues are shown as spheres. The homology model was built based on the structural homology of the nitrilase-related X-ray crystal structure of the nitrilases Nit6803 from *Synechocystis sp*. Strain PCC6803 (PDB ID: 3WUY). (**B**) The docking of the same homology model into the low resolution 3D electron microscopy reconstruction of the cyanide dihydratase from *B. pumilus* C1 (Mulelu, 2013). (**C**) The docking of the homology model into the low resolution 3D electron microscopy reconstruction of the cyanide dihydratase from *B. pumilus* C1 (Mulelu, 2013) showing the location of interfacial surfaces. (**D**) The docking of the homology model into the low resolution 3D electron microscopy reconstruction of the cyanide dihydratase from *B. pumilus* C1 (Mulelu, 2013) showing the proposed location of C-terminal tail.

1.5 Catalysis

All known members of the nitrilase superfamily have been shown to utilize a conserved catalytic triad consisting of a glutamate, lysine and cysteine (Pace et al., 2000; Nakai et al., 2000)(Fig. 1.5). Of the catalytic triad residues, the cysteine was the first to be discovered when it was observed that inhibition of activity by labelling with a catalytic cysteine specific iodoacetate resulted in an inactive form of the *N*-carbamyl-D-amino acid amidohydrolase (D-NCAase) from *Agrobacterium* (Grifantini et al., 1996). Evidence of this catalytic cysteine was reinforced by site directed mutagenesis experiments in which mutations of the cysteine (Cys172 in D-NCAase) produced a non-functional enzyme (Grifantini et al., 1996). Through topological mapping a solvent-accessible cleft or pocket in the D-NCAase was later identified within the vicinity of the catalytic cysteine (Grifantini et al., 1996). Around this cleft, highly conserved residues were identified by structural alignment of various nitrilase crystal structures. The structural alignment identified four highly conserved residues, namely Glu47, Lys127, Glu146 and Cys172 (based on D-NCAase sequence), which were located close to each other and clustered near the floor of the pocket (Wang et al., 2001). It was then postulated that the three conserved residues Cys172, Glu47 and Lys127 of D-NCAase (Cys164, Glu48 and Lys130 in CynD_{pum} sequence) are most likely to participate in the amidohydrolytic reaction. The essentiality of the first glutamate (Glu47 in D-NCAase) and lysine in catalysis was demonstrated by mutagenesis experiments of *Pseudomonas* aliphatic amidases where E59Q and K134N mutations of the aliphatic amidase resulted in an inactive enzyme while E59N and K134R mutations resulted in an inactive and unstable enzyme (Novo et al., 2002). This not only showed the importance of these residues in catalysis but also indicated their role in protein stability. The participation of the second glutamate (Glu146) in catalysis was later confirmed in the amidase reaction mechanism by mutational studies of the *Geobacillus pallidus* amidase (Kimani et al., 2007; Weber et al., 2013).

A plausible catalytic mechanism was later proposed in which the catalytic cysteine acts as a nucleophile, the glutamate acts as a general base catalyst and the lysine stabilizes a tetrahedral intermediate (Wang et al., 2001; Andrade et al., 2007). The two conserved glutamates (Glu47 and Glu164) have been proposed to play a role in correctly positioning the substrate. This has been demonstrated using substrate-bound crystal structures of D-NCAase, the formamidase from *Helicobacter pylori* and the amidase from *Geobacillus pallidus* (Chen et al., 2003, Hung et al., 2007; Weber et al., 2013). Based on the crystal structure of Nit6803 from *Synechocystis* *sp.* Strain PCC6803, it has been proposed that the second conserved glutamate (E142) may also play a role in stabilizing and/or activating the catalytic triad by forming hydrogen bonds with the catalytic lysine residue (Zhang et al., 2014). Although the catalysis mechanism of cyanide dihydratases remains ambiguous, a general catalysis mechanism has been proposed (Jandhyala et al., 2005; Thuku et al., 2009) (Fig. 1.7). The active site cysteine (Cys164) initiates a nucleophilic attack on the substrate to form a lysine (Lys130) stabilised tetrahedral intermediate (i) and this is followed by the increase of the cysteine's nucleophilicity and proton transfer which is initiated by one of the glutamates (Glu48), resulting in the release of ammonia and formation of a thioester intermediate (ii). One of the glutamates (Glu137) then activates a hydrolysis of the thioester intermediate leading to a second tetrahedral intermediate (iii), which decomposes to release formic acid and the restoration of the enzyme.

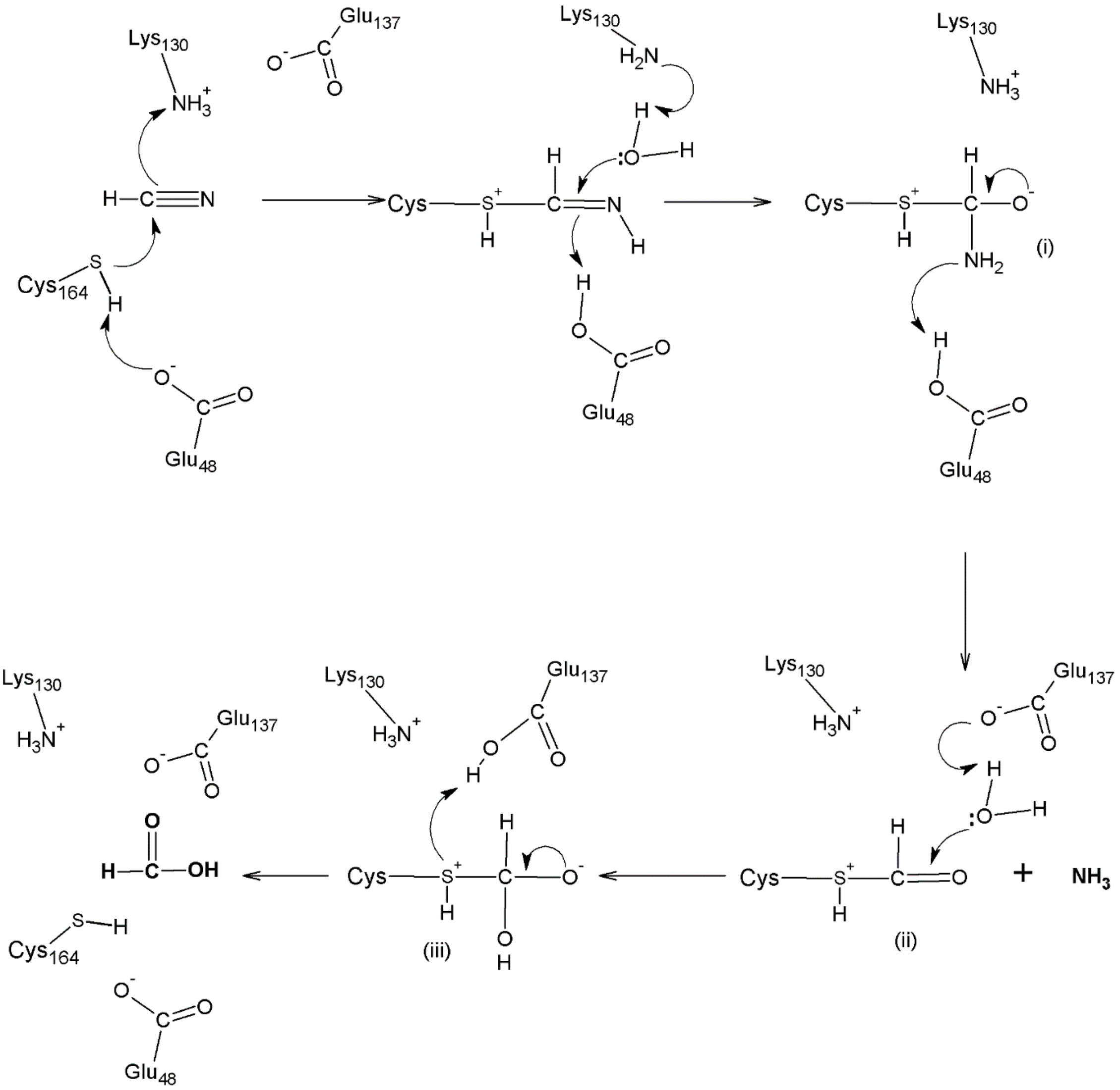


Figure 1.7. The proposed nitrilase catalytic mechanism of the *Bacillus pumilus C1* cyanide dihydratase (Jandhyala et al., 2005; Sewell, pers. commun.).

1.6 Previous studies on the cyanide dihydratase from *Bacillus pumilus* C1

1.6.1 Characterisation

This study will focus on the cyanide dihydratase from *Bacillus pumilus* C1 (CynD_{pum}), a member of the branch 1 nitrilases. The *Bacillus pumilus* C1 organism was first isolated from soil samples near a cyanide wastewater dam in South Africa (Meyers et al., 1991). CynD_{pum} is an enzyme that is known to convert cyanide directly into formate and ammonia (Fig. 1.1, reaction a). The activity of the enzyme is optimal at a temperature range of 37°C–42°C and a pH range of 7–8 (Jandhyala et al., 2005). Like most nitrilases, CynD_{pum} has a homodimeric building block with the αββα-αββα sandwich fold as previously mentioned. Over the pH range of 6–8 the CynD_{pum} enzyme exists as active short spirals approximately 18.9nm in length (Fig. 1.8). The spirals associate to form long active helical fibres at pH 5.4 (Sewell et al., 2005; Wang et al., 2011)(Fig. 1.8). These fibres have a diameter of about 13nm and length of over 100nm (Jandhyala et al., 2005, Wang et al., 2011).

The reason for the termination of the spirals at pH 8 is still unclear but it was initially hypothesized to be due to distortion of the helical symmetry, where the pH environmental change affected the helical symmetry of the overall oligomeric structure by altering the inter-subunit interactions responsible for the formation of the oligomeric structure (Sewell et al., 2003). The onset of fibre formation has been shown to correspond to a small increase in activity for the CynD_{pum} suggesting that the terminal subunits may be inactive (potentially caused by the distortion of the helical symmetry) and that the structural transformation to long helices at pH 5.4 results in more active subunits in the population thus increasing the activity (Jandhyala et al., 2005; Sewell et al., 2005). It was later demonstrated that there is a correlation between oligomerisation and stability of CynD_{pum}, when the fibre lengths of CynD_{pum} mutants were compared with their stability (Wang et al., 2011).

A number of mutagenesis experiments have been carried out to investigate the C-terminal residues and residues at the interfacial regions. A majority of single substitution mutation studies were done on residues at interfacial regions (Fig 1.9A). These experiments have helped identify substitution mutations that confer various characteristics to the native enzyme such as destabilisation and increase in thermostability (Fig 1.9B). Using this information from these experiments we were able to make hypothesis about residues involved in oligomerisation, activity and stability of CynD_{pum}.

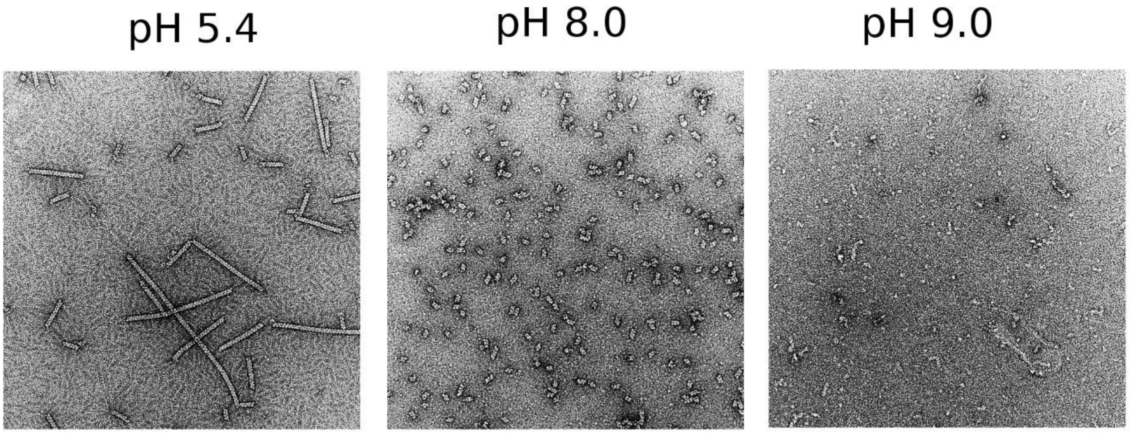


Figure 1.8. Negative stain electron micrographs of the wild-type cyanide dihydratase from *B. pumilus* C1 at pH 5.4, pH 8 and pH 9, showing the different oligomeric states at the various pH (Wang et al., 2011).

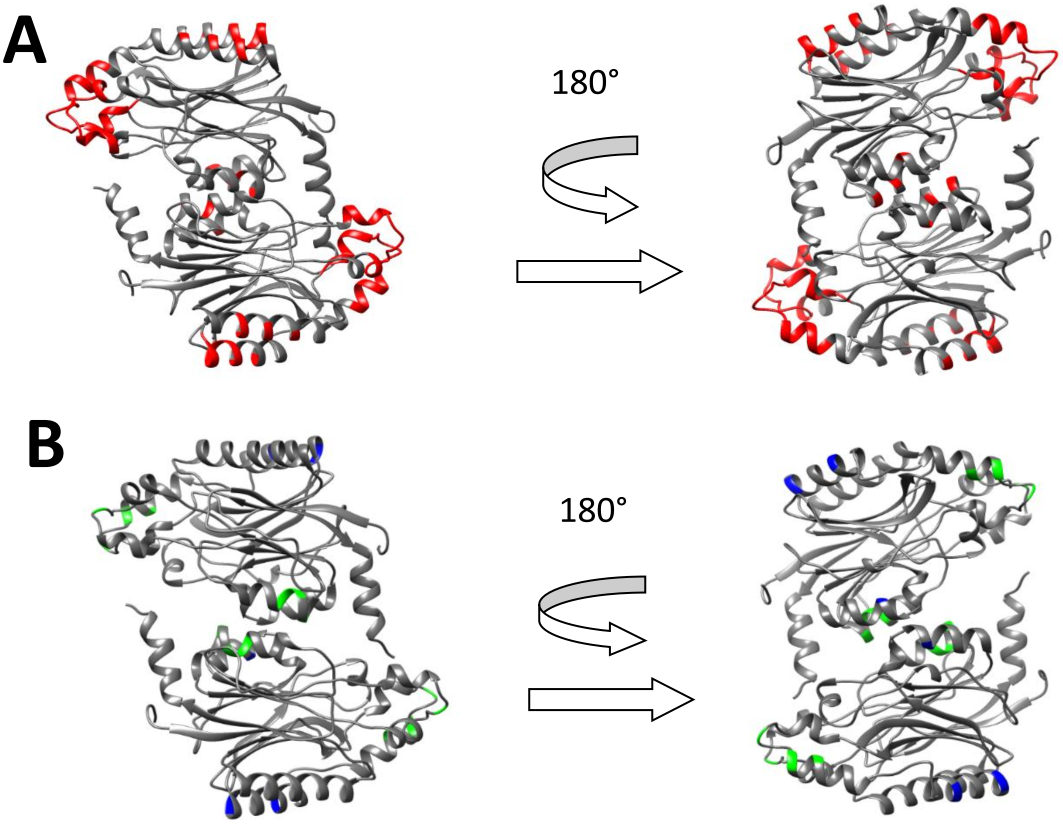


Figure 1.9. Views of a dimer homology model of the cyanide dihydratase from *B. pumilus* C1 nitrilase at 180° rotations (Park et al., 2016). **(A)** Regions that have been studied through mutagenesis experiments are highlighted in red. **(B)** Regions in which mutagenesis studies have shown a loss of activity/protein aggregation are highlighted in green while regions that demonstrated an increase in stability relative to the wild-type are highlighted in blue. The homology model was built based on the structural homology of the nitrilase-related X-ray crystal structure of the nitrilase Nit6803 from *Synechocystis sp*. Strain PCC6803 (PDB ID: 3WUY).

1.6.2 C-surface studies

Initial CynD$_{pum}$ mutagenesis experiments were done by Sewell et al. (2005) where a C-surface mutation with excised insertions which resulted in a shorter C-surface β sheet, similar to those found in non-spiral-forming homologs, showed inactivity or failure to express (Δ 219–233, Table 1.2). This suggested that the C-surface interactions may be involved in fibre formation and activation of the enzyme. In a more recent study a test on the involvement of the C-surface residues in spiral formation was carried out by Park et al. (2016) in which the proposed C-surface residues were all mutated to a cysteine with the aim of forming new disulfide bridges at one or more positions (P55C–E72C and E222–E235C,Table 1.2). Although no conclusive disulfide bridges were formed the experiment demonstrated that some of these C-surface mutations lead to the loss of activity or change in the size of the oligomers formed or both these effects. Both these C-surface studies suggested that the C-surface interactions may be involved in both oligomerisation and activation of the CynD$_{pum}$ enzyme.

1.6.3 D-surface studies

When investigating the CynD$_{pum}$ D-surface, residues 90–97 (residues that could possibly form salt bridges across the D-surface that stabilized the spirals) were mutated to residues 89–96 from the homologuous cyanide dihydratase from *Pseudomonas stutzeri* (CynD$_{stut}$) that cannot form salt bridges across the D-surface (MB3799, Table 1.2) (Sewell et al., 2005). The mutations resulted in no loss of activity suggesting that the spirals were still active in the absence of the salt bridge, thus the activity of CynD$_{pum}$ is not affected by the D-surface interactions. In the same study a series of C-terminal truncations were also generated for both CynD$_{pum}$ (Δ 279, Δ 293 and Δ 303, Table 1.2) and the homologous CynD$_{stut}$ (Δ 310) and it was demonstrated that the CynD$_{pum}$ retained activity even after a larger truncation than that of CynD$_{stut}$. Because these enzymes only differ in residues on the D-surface and C-terminus, it was suggested that the D-surface may be involved in the stabilization of the spirals which contain truncated C-terminal tail (Sewell et al., 2005). In a recent study a D-surface CynD$_{pum}$ mutation, Q86R (Table 1.2), was shown to significantly increase the length of the fibres formed at pH 5.4 compared to those of the wild-type at the same pH, thus further demonstrating the D-surface's involvement in stabilization (Wang et al., 2011). Another mutation K93R (Table 1.2) has been shown to increase the catalytic activity of CynD$_{pum}$ thus presenting evidence of a relationship between the D-surface and the catalysis of the enzyme (Crum et al., 2016).

1.6.4 A-surface studies

When CynD$_{pum}$ mutations of residues on the A-surface were created, they resulted in the inactivation of the enzyme (Y201D, A202T and A204D; Table 1.2). This provides evidence of the involvement of the A-surface in the activation of the enzyme and hence the need for dimerisation in activation of the enzyme (Sewell et al., 2005; Crum et al., 2016).

1.6.5 C-terminal region studies

The importance of the C-terminal tail in CynD$_{pum}$ was first demonstrated when removal of more than 28 C-terminal residues (Δ279, Table 1.2) resulted in the loss of activity as mentioned above (Sewell et al., 2005). Histidines have been implicated in the transition from short spirals to long helices because this structural change occurs as the pH drops to 5.4, a pH at which histidine undergoes changes in charge (Histidine pKa of 6.0). There are 3 C-terminal tail histidines in the amino acid sequence of CynD$_{pum}$ (Fig. 1.3) and these were proposed to be involved in the pH dependent transition. The involvement of these histidines in the pH dependent oligomerisation was first demonstrated by the fact that the highly homologous CynD$_{stut}$, which lacks these C-terminal histidines, does not undergo the transition from short spirals to long fibers at pH 6 (Sewell et al., 2005). Eicher (2007) demonstrated that a similar strain of CynD$_{pum}$ from *Bacillus pumilus 8A3* which also lacks these histidines on the C-terminal tail forms aggregated and irregular short fibres at pH below 6 instead of regular long fibres formed by the *Bacillus pumilus* C1 strain. To investigate this we mutated these C-terminal histidines with an aim of preventing the pH dependent transition and extending the fibre formation to a broader pH range (H305K, H308K, H323K, and CynD$_{pum\text{-}stut}$, Table 1.2). This study demonstrated that the C-terminal histidines are only partially involved in the pH dependent transition from short spirals to long helices, hence other histidines in the CynD$_{pum}$ may be involved (Mulelu, 2013).

Table 1.2. List of CynD$_{pum}$ mutants constructed during previous mutagenesis studies and a list of their conferred characteristics.

Modification	**Location**	**Oligomerisation**		**Activity**	**Stability** (pH Stability/ Thermostability)	**Reference**
		Microscopy	**Size Exclusion**			
CynD$_{pum}$ original	N/A	Formation of fibers at pH 5.4 Helical symmetry: Δφ =76.98, Δz =15.71 Å, Resolution= 17 Å. Formation of short spirals at pH 8.	Single peak observed corresponding to 18mers at pH 8.	Active at pH 8.	Optimal activity at pH 8, Optimal Temperature at 37°C.	Meyers et al., 1991 Sewell et al., 2005 Jandhyala et al., 2005
MB3799 (90EAAKRNE97 –AAARKNK)	D-surface α3	Formation of fibers at pH 5.4 (comparable to WT) Helical symmetry: Δφ =77.35, Δz =14.68 Å, Resolution= 29 Å	Single peak observed comparable to wild-type.	Active at pH 8.	ND	Sewell et al., 2005 Van der Vyver, 2007
MB3799 + E31K	D-surface (α3 **MB3799**) (α1 **E31K**)	Formation of fibers at pH 5.4 (comparable to WT) Helical symmetry: Δφ =77.12, Δz =14.77 Å, Resolution= 29 Å	Single peak observed comparable to wild-type at pH 8.	Active at pH 8.	ND	Van der Vyver, 2007
MB3799 +E27K	D-surface (α3 **MB3799**) (α1 **E27K**)	Formation of fibers at pH 5.4 (comparable to WT) Helical symmetry: Δφ =77.02, Δz =16.18 Å, Resolution= 30.65 Å	Single peak observed comparable to wild-type at pH 8.	Active at pH 8.	ND	Gray, 2008
MB3799 +E27K +E31K	D-surface (α3 **MB3799**) (α1 **E27K,E31K**)	Formation of fibers at pH 5.4 (comparable to WT) Helical symmetry: Δφ =77.17, Δz =16.29 Å, Resolution= 28.94 Å	Single peak observed comparable to wild-type at pH 8.	Active at pH 8.	ND	Gray, 2008
MB3799 + E27K +E31K +D34N	D-surface (**MB3799** α3) (α1 **E27K, E31K, D34N**)	Formation of fibers at pH 5.4 (comparable to WT).	Two peaks observed. Second smaller peak had no activity.	Active at pH 8.	ND	Gray, 2008
MB3799 + E23K E27K +E31K +D34K+E35K	D-surface (α3**MB3799**) (α1 **E27K, E31K, D34N, E35K**)	Aggregation of fibers at pH 8.	Partially insoluble and difficult to purify.	Active at pH 8.	ND	Gray, 2008

Table 1.2. (Continued)

Modification	**Location**	**Oligomerisation**		**Activity**	**Stability** (pH Stability/ Thermostability)	**Reference**
		Microscopy	**Size Exclusion**			
Q86R	D-surface α3	Formation of very long fibers at pH 5.4, Helical symmetry: $\Delta\varphi$ =77.00, Δz =15.61 Å, Resolution= 20 Å. Formation of short fibers at pH 8 Helical symmetry: $\Delta\varphi$ =76.70, Δz =15.86 Å, Resolution= 19 Å. Formation of short spirals at pH 9 (comparable to WT pH 8 spirals).	Eluted at a slightly higher molecular weight fraction compared to wild-type at pH 8.	Active at pH 8.	pH stability similar to wild-type (Optimal at pH 8). More thermostable than wild-type (higher $\Delta\Delta G$ and melting temperature).	Wang et al., 2011 Mulelu, 2013
E327G	C-terminal tail	Formation of fibers at pH 5.4 (comparable to WT). Formation of short spirals and some short fibers at pH 8. Formation of short spirals at pH 9 (comparable to WT pH 8 spirals).	Single peak observed comparable to wild-type at pH 8.	Active at pH 8.	Slightly more stable at alkali pH than wild-type (some activity observed at pH 9). Slightly more thermostable than wild-type.	Wang et al., 2011 Mulelu, 2013
Q86R+E96G+ D254E	D-surface (α3 **Q86R,E96G**)	Formation of very long fibers at pH 5.4, Helical symmetry: $\Delta\varphi$ =76.50, Δz =16.31 Å, Resolution= 22.22 Å. Formation of short fibers at pH 8. Short fibers still present at pH 9.	Eluted at a slightly higher molecular weight fraction compared to wild-type at pH 8.	Active at pH 8.	Stable at alkaline pH (activity still present at pH 9).	Wang et al., 2011 Mulelu, 2013
Q86R+E96G+ D254E+E327G	D-surface (α3 **Q86R,E96G**) C-terminal tail (**E327G**)	Formation of very long fibers at pH 5.4. Formation of short fibers at pH 8. Short fibers still present at pH 9.	Eluted at a slightly higher molecular weight fraction compared to wild-type at pH 8.	Active at pH 8.	Stable at alkaline pH (activity still present at pH 9). More thermostable than wild-type (higher $\Delta\Delta G$ and melting temperature)	Wang et al., 2011 Mulelu, 2013
K93R	D-surface α3	Formation of fibers at pH 5.4 (comparable to WT). Formation of short fibers at pH 8 (comparable to WT).	Single peak observed comparable to wild-type at pH 8.	Higher catalysis activity than wild-type at pH 7.7.	pH stability similar to wild-type (optimal at pH 8). More thermostable than wild-type.	Crum et al., 2016

Table 1.2. (Continued)

Modification	**Location**	**Oligomerisation**		**Activity**	**Stability** (pH Stability/ Thermostability)	**Reference**
		Microscopy	**Size Exclusion**			
A202T	A-surface α6	ND	Formed aggregates.	Inactive.	ND	Crum et al., 2016
D172N	A-surface α5	ND	ND	Active at pH 8.	pH stability similar to wild-type (optimal at pH 8). More thermostable than wild-type.	Crum et al., 2016
E327K	C-terminal tail	Formation of fibers at pH 5.4 (comparable to WT). Formation of short fibers at pH 8 (comparable to WT).	Single peak observed comparable to wild-type at pH 8.	Higher catalysis activity than wild-type at pH 7.7.	pH stability similar to wild-type (optimal at pH 8). Slightly more thermostable than wild-type.	Crum et al., 2016
Combinations of K93R, D172N, A202T and E327K. (**K93R+E327K**, **D172N+A202T**, **K93R+A202T**, **E327K+A202T**, **K93R+D172N +A202T**	D-surface (α3 **K93R**) A-surface (α6 **A202T,** α5 **D172N**) C-terminal tail (**E327K**)	ND	**K93R +A202T** and **E327K+A202T** formed aggregates.	Activity at pH 7.7 for **K93R+E327K**, **D127N+A202T** and **K93R+D172N +A202T**. Inactivity for **K93R +A202T** and **E327K+A202T**	**K93R+E327K** showed pH stability similar to wild-type (optimal at pH 8). **K93R+E327K** and **K93R+D172N+A202T** showed more thermostablity than wild-type. **D172N+A202T** showed 50% loss of activity at pH 8 compared to wild-type.	Crum et al., 2016
Y201D+A204D	A-surface α6	ND	ND	Inactive.	ND	Sewell et al., 2005
Q86R+E96G+ D254E+E327G+ K93R	D-surface (α3 **Q86R, E96G, K93R**) C-terminal tail (**E327G**)	Formation of very long fibers at pH 5.4. Formation of short fibers at pH 8. Short fibers still present at pH 9.	Single peak observed comparable to wild-type at pH 8.	Active at pH 8.	Stable at alkaline pH (activity still present at pH 9). More thermostable than wild-type.	Crum, 2012

Table 1.2. (Continued)

Modification	Location	Oligomerisation		Activity	Stability (pH Stability/ Thermostability)	Reference
		Microscopy	Size Exclusion			
Cysteine single substitution mutations of C-surface region 1 residue 55–72 (**P55C**, **W56C**, **F57C**, **A58C**, **F59C**, **I60C**, **G61C**, **H62C**, **P63C**, **E64C**, **Y65C**, **T66C**, **R67C**, **K68C**, **F69C**, **Y70C**, **H71C** and **E72C**)	C-surface region 1 (α2 **R67C**, **K68C**, **F69C**, **Y70C**, **H71C, E72C**)	**R67C** showed disruption of the native fiber formation at pH 5.4 and showed formation of short spirals of variable sizes at pH 8. **Y70C** showed disruption of the native fiber formation at pH 5.4 and formation of short spirals at pH 8 (comparable to WT).	Single peak observed corresponding to 18-mers for **P55C**, **A58C**, **F59C**, **I60C**, **H62C**, **P63C**, **Y65C**, **K68C** and **E72C** at pH 8. Single peak observed corresponding to 16-mers for **W56C**, **E64C**, **T66C**, **F69C**, **Y70C** and **H71C** at pH 8. Single peak observed corresponding to 14-mers for **F57C**, **G61C**, and **R67C** at pH 8.	Activity at pH 8 for **P55C**, **W56C**, **F57C**, **A58C**, **F59C**, **I60C**, **G61C**, **H62C**, **P63C**, **Y65C**, **T66C**, **K68C**, **F69C**, **H71C** and **E72C**. Weak activity at pH 8 for **E64C**, **R67C** and **Y70C**.	**H71C** was less thermostable than wild-type (Lower ΔΔG).	Park et al., 2016
Cysteine single substitution mutations of C-surface region 1 residue 222–235 (**E222C**, **M223**, **I224C**, **C225A**, **L226C**, **T227C**, **Q228C**, **E229C**, **Q230C**, **R231C**, **D232C**, **Y233C**, **F234C** and **E235C**)	C-surface region 2	**Q228C** formed aggregates.	Single peak observed corresponding to 18-mers for **E222C**, **M223**, **I224C**, **C225A**, **L226C**, **T227C**, **E229C**, **Q230C**, **R231C**, **D232C** and **E235C** at pH 8. Single peak observed corresponding to 16-mers at pH 8 for **Y233C** and **F234C**.	Activity at pH 8 for **E222C**, **M223**, **I224C**, **C225A**, **L226C**, **T227C**, **Q228C**, **E229C**, **Q230C**, **R231C**, **D232C**, **Y233C**, **F234C** and **E235C**.	ND	Park et al., 2016

Table 1.2. (Continued)

Modification	Location	Oligomerisation		Activity	Stability (pH Stability/ Thermostability)	Reference
		Microscopy	Size Exclusion			
H305K	C-terminal tail	Formation of fibers at pH 5.4 (comparable to WT). Formation of short spirals at pH 8 (comparable to WT).	Single peak observed comparable to wild-type at pH 8.	Active at pH 8.	pH stability similar to wild-type (optimal at pH 8). More thermostable than wild-type (higher $\Delta\Delta G$ and melting temperature).	Mulelu, 2013
H308K	C-terminal tail	Formation of fibers at pH 5.4 (comparable to WT). Formation of short spirals at pH 8 (comparable to WT).	Single peak observed comparable to wild-type at pH 8.	Active at pH 8.	pH stability similar to wild-type (optimal at pH 8). More thermostable than wild-type (higher $\Delta\Delta G$ and melting temperature).	Mulelu, 2013
H323K	C-terminal tail	Formation of fibers at pH 5.4 (comparable to WT). Formation of short spirals at pH 8 (comparable to WT) with a high degree of aggregation.	Single peak observed comparable to wild-type at pH 8.	Active at pH 8.	pH stability similar to wild-type (optimal at pH 8).	Mulelu, 2010
Combinations of **H305K**, **H308K** and **H323K**. (**H305K+H308K**, **H305+H323K**, **H308K+H323K**, **H305K+H308K+ H323K**)	C-terminal tail	**H305K+H308K**, **H305+H323K** and **H308K+H323K** formed fibers at pH 5.4 (comparable to WT) and formed short spirals at pH 8 (comparable to WT). **H305K+H308K+H323K** formed very long fibers at pH 5.4, Helical symmetry: $\Delta\varphi = 77.02$, $\Delta z = 15.80$ Å, Resolution= 20 Å. **H305K+H308K+H323K** formed short fibers at pH 8, Helical symmetry: $\Delta\varphi = 77.00$, $\Delta z = 15.80$ Å, Resolution = 21 Å. **H305K+H308K+H323K** formed short fibers still present at pH 9.	**H305K+H308K**, **H305+H323K** and **H308K+H323K** eluted in single peak observed comparable to wild-type at pH 8. **H305K+H308K +H323K** eluted at a slightly higher molecular weight fraction compared to wild-type at pH 8	All active at pH 8.	All had pH stability similar to wild-type (optimal at pH 8) and more thermostablity than wild-type (higher $\Delta\Delta G$ and melting temperature).	Mulelu, 2013

Table 1.2. (Continued)

Modification	**Location**	**Oligomerisation**		**Activity**	**Stability** (pH Stability/ Thermostability)	**Reference**
		Microscopy	**Size Exclusion**			
GERDST (307NHQKNE312–GERDST)	C-terminal tail	ND	ND	Active at pH 7.7.	pH stability similar to wild-type (optimal at pH 8). More thermostable than wild-type.	Crum, 2012
CynD$_{\text{pum-stut}}$ (residues 1–286 from CynD$_{\text{pum}}$ plus residues 287–end from CynD$_{\text{stut}}$)	C-terminal tail	Formation of short spirals at pH 5.4. Formation of short spirals at pH 8. Formation of short spirals at pH 9.	Single peak observed comparable to wild-type at pH 8.	Active at pH 8.	Stable at alkaline pH (activity still present at pH 9.5). More thermostable than wild-type.	Sewell et al., 2005 Crum et al., 2015a Mulelu, 2013
CynD$_{\text{pum-stut}}$ + R67C/Y70C/K93R	C-terminal tail (**pum –stut**) C-surface (α2 **R67C**, **Y70C**) D-surface (α3 **K93R**)	ND	Single peak observed comparable to wild-type at pH 8. **CynD$_{\text{pum-stut}}$ + R67C** eluted as two merging peaks at around 18-mer size.	**CynD$_{\text{pum-stut}}$ + R67C/K93R** showed activity at pH 7.7, **CynD$_{\text{pum-stut}}$ + Y70C** showed no activity at pH 7.7.	CynD$_{\text{pum-stut}}$ +K93R was stable at alkaline pH (activity still present at pH 9.5) and more thermostable than wild-type.	Park, 2014 Crum et al., 2015a
CynD$_{\text{pum-cras}}$ (residues 1–286 from CynD$_{\text{pum}}$ plus 287–end from CHT$_{\text{cras}}$)	C-terminal tail	ND	Single peak observed comparable to wild-type at pH 8.	Active at pH 7.7. 59 ± 2 % reaction rate compared to wild-type at pH 8.	ND	Crum et al., 2015b
Δ303 (Truncation from C-terminal to V303	C-terminal tail	ND	Two peaks observed. A small peak eluted at 669 kDa (18-subunit) and large peak at 150–180 kDa (tetramers) at pH 8.	Active at pH 8. 87 ± 5 % reaction rate compared to wild-type at pH 8.	Less thermostable than wild-type.	Sewell et al., 2005 Crum et al., 2015b
Δ 293 (Truncation from C- terminal to M293)	C-terminal tail	ND	ND	Partial activity at pH 8. 27 ± 2 % reaction rate compared to wild-type at pH 8.	ND	Sewell et al., 2005

Table 1.2. (Continued)

Δ 279 (Truncation from C-terminal to Y276)	C-terminal tail	ND	ND	Inactive.	ND	Sewell et al., 2005
Δ **219-233** (219MKEMICLTQEQR DYF233 was deleted)	C-surface region 2	ND	ND	Inactive.	ND	Sewell et al., 2005

1.7 Motivation and study objectives

1.7.1 Motivation

Cyanide dihydratase enzymes have potential to be used in detoxification of cyanide waste produced during industrial processes, such as leach mining, metal finishing, and electroplating, as well as detoxification of accidental cyanide spills. Cyanide is a potent inhibitor of cytochrome oxidase, making it toxic to aerobic organisms, and extended exposure of humans to cyanide has been linked to certain neuropathies (Jandhyala et al., 2003). Conventional methods used for cyanide detoxification, such as alkaline chlorination, use hazardous chemicals that are both costly and generate toxic by-products, and therefore the use of biological based methods would be desirable. Bacterial detoxification could provide a cheap, on-site enzymatic solution. $\mathrm{CynD_{pum}}$ is a good candidate because it does not need additional cofactors and has a functionality over a wide substrate concentration (Jandhyala et al., 2005). The drawback of this cyanide dihydratase enzyme is that cyanide wastes are highly alkaline and have been shown to inactivate the wild-type enzyme, therefore improvement of the stability would allow for a more effective use of this enzyme.

In previous studies we demonstrated that the stability of $\mathrm{CynD_{pum}}$ is related to its oligomerisation (Wang et al., 2011, Mulelu, 2010, Mulelu, 2013) and hence understanding the process of oligomerisation can potentially help identify interfacial residues which can confer high stability. This investigation will give us a better understanding of the mechanism for the transition from short to long helices and the understanding of factors that dictate stability which will lead to the design of useful industrial enzymes which can be used for bioremediation. This study can potentially give more insights into other nitrilases and related enzymes due to the high structural homology.

In a previous study (Mulelu, 2013) we investigated the factors involved in the oligomerisaton of $\mathrm{CynD_{pum}}$ in which we attempted to identify key amino acid residue responsible for the formation of $\mathrm{CynD_{pum}}$ oligomers. Using mutagenesis studies we investigated three C-terminal histidines which we proposed are involved in fibre formation. Based on our results we concluded that these three histidines are only partially involved in fibre formation and hence a study has to be done on the 7 remaining histidines to identify the key histidine residue involved in the pH dependent oligomerisation.

1.7.2 Study objectives

The aim of this project is to obtain a high resolution structure of CynD$_{pum}$, understand the structure, relate this to its function, and investigate the role of histidines in oligomerisation with help of the structure. This is to be accomplished by addressing three main objectives:

- **To obtain a high resolution structure of CynD$_{pum}$.**
  Cryo-Electron Microscopy 3D reconstruction techniques will be used to obtain high resolution 3D EM maps of CynD$_{pum}$. A homology model will be docked into the 3D maps to allow for interpretation.
- **Investigate the amino acid residues involved in the pH dependent oligomerisation transition from spirals to helices of CynD$_{pum}$.**
  Based on previous mutagenesis and structural studies, various CynD$_{pum}$ mutants will be designed to help us identify the key amino acids involved in oligomerisation. As a continuation of the Mulelu (2013) study the 7 remaining CynD$_{pum}$ histidines (His62, His71, His128, His167, His184, His241 and His285) will be investigated as these are likely to be the key residues involved in the observed pH dependent fibre formation. As part of the investigation various CynD$_{pum}$ histidine mutants will be constructed, purified and visualized under a transmission microscope at various pH to observe their oligomerisation at both acidic and alkali pH. Electron microscopy 3D reconstruction techniques will be used to visualise structural differences in the quaternary structure between these CynD$_{pum}$ mutants and wild-type at the various pH, with the aid of the solved high resolution CynD$_{pum}$ structure.
- **To identify CynD$_{pum}$ mutations that confer stability.**
  Thermal shift assays will be used to characterise the stability of the CynD$_{pum}$ mutants. The melting temperatures of the CynD$_{pum}$ mutants will be assayed using differential scanning fluorimetry to confirm stability.
- **To investigate the structural role of the CynD$_{pum}$ mutations**.

The refined high resolution structures obtained in the initial steps will not only allow us to structurally characterise the CynD$_{pum}$ mutants constructed in this project but will also help us characterise previous CynD$_{pum}$ mutants which were formed over the years of the CynD$_{pum}$ study. The combination of all the mutational studies and the high resolution CynD$_{pum}$ structure will provide key insights in the overall structure of the enzyme.

CHAPTER II

HIGH RESOLUTION THREE DIMENSIONAL STRUCTURES

2.1 Introduction

Obtaining a crystal structure of the CynD$_{pum}$ has proven challenging over the years thus making it difficult to obtain accurate 3D atomic coordinates of this enzyme, and the use of predicted structures such as homology models has proven unreliable. To date only one crystal structures from branch 1 nitrilase members has been solved, namely the Nit6803 from *Synechocystis* sp. *Strain PCC6803* (Zhang et al., 2014). One of the main reasons why CynD$_{pum}$ cannot be crystallized readily is its lack of symmetry that conforms to that which can be packed into crystals due to the oligomerisation of monomers that results in the formation of long complex quaternary structures. To overcome this limitation, we used Cryo-electron microscopy (Cryo-EM) techniques to acquire a high resolution structure of CynD$_{pum.}$ Previous attempts to obtain a high resolution structure of CynD$_{pum}$ have not been successful and one possible factor preventing the acquisition of high resolution 3D cryo-EM structures of CynD$_{pum}$ could be the enzymes low stability which results in structural variability of the enzyme in solution. As part of a previous investigation (Mulelu, 2013) we designed and produced significantly more stable CynD$_{pum}$ variants through mutagenesis experiments in which we manipulated some of the residues located at the interfacial regions and C-terminal tail. In the same investigation we managed to construct two variants which showed higher stability compared to the wild-type. These variants include the D-surface variant Q86R and the C-terminal histidine variant H308K+H305K+H323K. In this chapter we describe how we combined the above mutations to produce a more stable variant (Q86R+H308K+H305K+H323K) which is shown to be more stable than both the parent variants in chapter IV (section 4.4.4). This variant was expressed, purified and visualised under the transmission electron microscope. A cryo-EM image dataset of the Q86R+H308K+H305K+H323K variant was then collected on two different electron microscopes, the FEI Polara at Birkbeck College, London, and the FEI Titan Krios at the Buchmann Institute for Molecular Life Sciences, Frankfurt. An additional cryo-EM image dataset of the wild-type CynD$_{pum}$ was also collected on the FEI Titan Krios to help determine the structural changes conferred by the Q86R+H308K+H305K+H323K mutations. The cryo-EM image datasets were then processed to produce 3D reconstructions using various programs. The high structural stability of the Q86R+H308K+H305K+H323K variant was proposed to reduce structural variability in solution thus producing a more uniform structure which could potentially be resolved at higher resolution than the wild-type using the cryo-EM techniques.

2.2 Review of cryo-EM and helical three dimensional reconstruction

The majority of non-viral biological assemblies that have been determined to near atomic resolution using cryo-EM techniques are single particles although a large number of prokaryotic, eukaryotic and archaeal protein complexes with significant roles exist as helical polymers. These helical polymers include assemblies such as intracellular actin filaments, myosin filaments, microtubules, extracellular collagen, keratin and bacterial flagella. Over the past two years cryo-EM coupled with image averaging techniques has made significant strides in structure determination of singular biological assemblies at sub 3 Å resolutions. Some of these great achievements include solving the 700 kDa proteasome from *Thermoplasma acidophilum* at 2.8 Å (Campbell et al., 2015), the 440 kDa anthrax protective antigen pore at 2.9 Å (Jiang et al., 2015) and the Escherichia coli β-galactosidase at 2.2 Å (Bartesaghi et al., 2015). Despite having an advantage over single particle biological assemblies in that a single view provides sufficient information to produce a 3D reconstruction, non-viral helical complexes have only been structurally determined at resolutions around 4 Å. These include the bacterial flagellum (Yonekura, 2003) and nicotinic acetylcholine receptor (Unwin, 2005) both solved at a resolution of 4 Å. Helical complex structures have proven difficult to solve using conventional structure determination methods such as X- ray crystallography. The major drawback of the structural determination of helical complexes is their inability to form an ordered crystal lattice due to their irregularly shaped surfaces. The only filaments that have been demonstrated to pack with crystalline symmetry are those that have exactly two, three, four or six subunits per turn (Egelman, 2007). With no requirement for crystal formation, cryo-EM techniques have proven successful in elucidating the structures of helical polymers, although presenting their own problems such as flexibility and disorder of the specimen's filaments. Real space computational approaches such as the Iterative Helical Real Space Reconstruction (IHRSR) algorithm have significantly provided solutions to some of these problems (Egelman, 2007).

Besides the existence of more stable microscopes and new computational methods, one other major contributing factor to the recent ability to obtain near-atomic resolution with cryo-EM is the increase detective quantum efficiency (DQE) of new high-speed direct electron detectors (Cheng, 2015; Henderson, 2015). The DQE is the measure of the additional noise added by a detector. Detectors with a high DQE consequently have a high signal to noise ratio which results in higher image contrast. These new detectors, such as the FEI Falcon II and the Gatan K2 Summit, have high sensitivity and high readout speeds that make it possible to track specimen movement during irradiation. Beam induced drifting of specimen during irradiation has been shown to reduce image quality by causing image blurring and loss of high resolution information (Henderson and Glaeser, 1985; Dubochet et al., 1988). The tracking of specimen movement by the new detectors has made it possible to realign micrograph movie sub-frames in order to reduce image blurring thus allowing for the acquisition of high resolution structures (Brilot et al., 2012; Bai et al., 2013; Li et al., 2013). The use of higher electron energies in new microscopes has also contributed in the acquisition of high resolution structures by reducing the damage of radiation sensitive samples such as ice embedded biological samples (McMullan et al., 2014)

2.3 Iterative Helical Real Space Reconstruction (IHRSR)

IHRSR is a robust algorithm developed by Egelman (2000) and was adopted from single-particle methods for the reconstruction of helical filaments. This technique has overcome some of the difficulties previously encountered in helical image analysis such as incorrect indexing, specimen disorder and flexibility. To carry-out IHRSR, helical filaments are initially selected from micrographs and then segmented to produce a stack of individual segments of appropriate length. These segments are treated as individual particles in subsequent steps. An initial model is then inserted into the algorithm (Fig. 2.1). The source of the initial model varies as it could be generated computationally using the stack of segmented filaments, generated from previous IHRSR runs or sourced from 3D models of homologous specimens. The above sources have the potential of introducing model bias and hence a commonly used featureless cylinder is preferable for initial reconstructions. Projections of the initial model are then generated by azimuthally rotating the model at appropriate angle increments about the filament axis, generating 2D reference projections for each rotation. The 2D references are then used for the multi-reference alignment against all the segmented images of the helical filament (Fig. 2.1). Multi-reference alignment is achieved by computationally calculating the best cross-correlation of each individual segment image to each reference projection and assigning that segment image to a reference projection that gives the highest cross-correlation. Thus the multi-reference alignment will generate a list of x-shifts, y-shifts, in-plane rotation and out-of-plane rotation parameters. This list is then used to apply the translational parameters and bring the aligned images into register with their assigned reference projection in order to generate an asymmetric 3D volume by a process called back projection (Fig. 2.1).

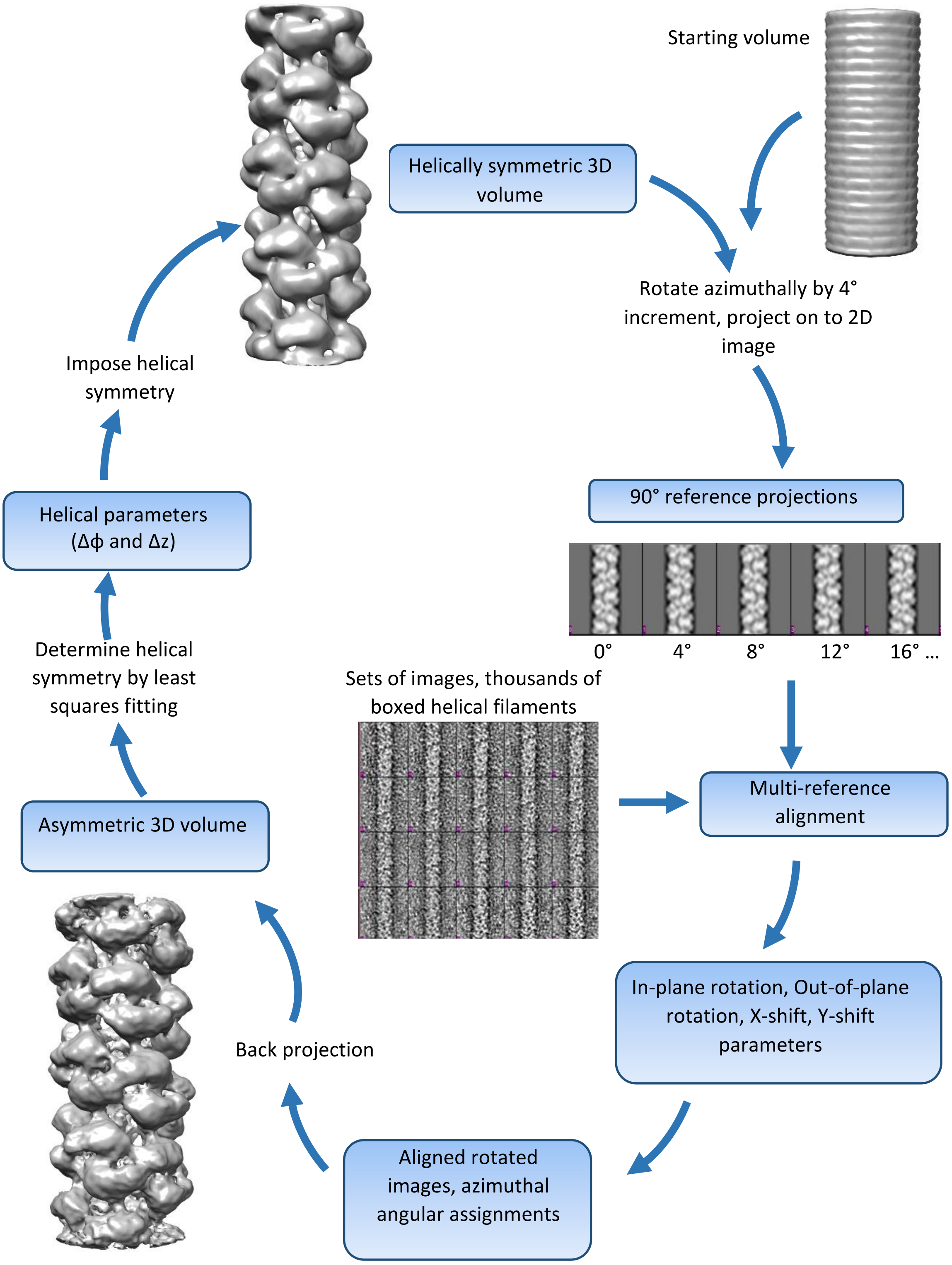


Figure 2.1. Schematic diagram of the cycle used the IHRSR algorithm (Egelman, 2000).

To further improve the reconstruction IHRSR takes advantage of the helical symmetry of the helical filament which it imposes on the asymmetric 3D model (Egelman, 2000). Helical symmetry is the symmetry of objects related by a coupled rotation about an axis with a translation parallel to the axis and is defined by two parameters, the rise between adjacent subunits (Δz) and the rotation per subunits (Δφ). In IHRSR the least-squares search algorithm calculates the helical symmetry parameters (Δz and Δφ) of the asymmetric 3D volume. The least-squares search algorithm calculates the mean-squared deviation in density between densities at symmetry-related positions in the asymmetric volume. The calculated helical symmetry parameters are then imposed to the asymmetric 3D volume to produce a symmetrized volume which is then used as a new reference for the next cycle (Fig. 2.1). During the initial cycle the IHRSR algorithm imposes a set of user-defined helical symmetry parameters to the first asymmetric 3D model. These initial helical symmetry parameters are obtained either from previous reconstruction or calculated using Fourier-Bessel methods (DeRosier and Klug, 1968). After the imposition of the helical parameters the whole process is then iterated until a stable helical symmetry solution is obtained, that is until the helical symmetry shows no further change and converges to a single solution (Fig. 2.1).

2.4 Resolution criteria

One problem with cryo-EM reconstruction is the lack of a standard technique of determining the resolution of EM models. The concept of resolution is defined as the minimal distance between two points in an image at which they can be still be distinguished from each other, and has been widely used to assess the quality of 3D reconstructions. Determining the resolution of EM model provides a measure of the extent at which reliable structural details can be interpreted (Cardone et al., 2013). Over the years a number of methods to calculate the resolution of 3D reconstructions have been developed. These methods include the Differential Phase Residual (Frank et al., 1981), Q-factor (Kessel et al., 1985), Fourier Shell Phase Residual (van Heel, 1987), Fourier Shell Correlation (FSC) (Harauz and van Heel, 1986), Spectral Signal-to-Noise Ratio 3D (SSNR3D) (Penczek, 2002; Unser et al., 2005), and the R measure method (Sousa and Grigorieff, 2007). The most commonly used method is the FSC method (Harauz and van Heel, 1986) which is based on the comparison of two independent averages from two randomly drawn subsets of equal size, in the Fourier domain. In this method the Fourier transforms of the two subsets are compared and a measure of discrepancy is calculated as a function of spatial frequency. Thus the FSC method measures the correlation as a function of resolution shells between Fourier coefficients computed from two independent reconstructions, generated from two 2D image subsets of the same specimen.

2.5 Materials and Methods

2.5.1 DNA Preparation

Site directed mutagenesis was performed using the KAPA HIFI PCR kit (Kapabiosystems, Boston, USA). The Q86R+H308K+H305K+H323K plasmid was constructed by introducing a Q86R mutation to an already constructed H308K+H305K+H323K plasmid (Mulelu, 2013). The Q86R mutation was introduced to the H308K+H305K+H323K template plasmid, consisting of the CynD$_{pum}$ mutated sequence inserted into a pet26b vector with kanamycin resistance (Novagen, USA), by using overlapping mutagenic primers listed in Table 2.1. The overlapping mutagenic primer PCR strategy is outlined in Fig. 2.2.

Table 2.1. Primer sequences used in site-directed mutagenesis. Primers written as 5'–3'. The underlined nucleotides correspond to the substituted bases,

Primer	Sequence
Q86R	CCCTAGCTTAGCCATTC<u>GC</u>AAAATAAGTGAGGCAGC GCTGCCTCACTTATTTT<u>GC</u>GAATGGCTAAGCTAGGG

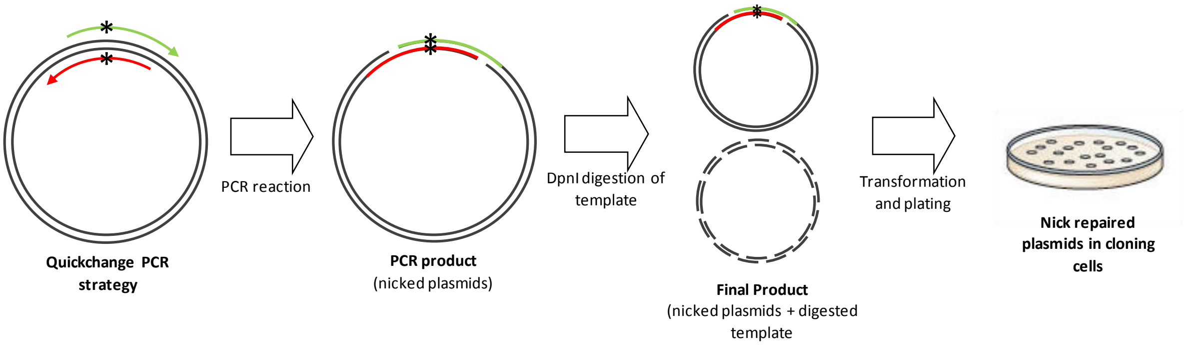


Figure 2.2. Schematic diagram of the overlapping mutagenic primer PCR strategy.

PCR reaction were carried out in a 50 µl reaction mixture which consisted of 0.3 µM forward primer, 0.3µM reverse primer, 1x KAPA HiFi Buffer, 0.3 mM KAPA dNTP Mix , 1 U of KAPA HiFi DNA Polymerase and 1 ng of the H308K+H305K+H323K template plasmid. The PCR reaction was run with the cycling conditions listed in Table 2.2. PCR product was then subjected to Dpn1 (Thermo Fisher Scientific, USA) digestion to remove the remaining template DNA and 100 ng of the Dpn1 digested PCR product was later used to transform chemically competent *Escherichia coli* (*E.coli*) DHα cells using the protocol described in section 2.3.2. Overnight cell cultures of the transformed *E. coli* DHα were grown and the plasmid DNA was then extracted. The success of the mutagenesis was validated by DNA sequencing. The wild-type CynD$_{pum}$ plasmid was obtained from a previous study (Mulelu, 2013).

Table 2.2. PCR cycling conditions used in the construction of the Q86R+H308K+H305K+H323K variant plasmid.

Step	Temperature	Duration	Cycles
Initial denaturation	95°C	3 min	1
Denaturation	98°C	20 sec	20
Annealing	62°C	20sec	
Extension	72°C	4 min	
Final extension	72°C	10 min	1

2.5.2 Transformation and Expression

The wild-type or Q86R+H308K+H305K+H323K CynD$_{pum}$ plasmid DNA (100 ng) was added to ice thawed *Escherichia coli* BL21 (DE3) chemically competent cells, and the mixture was incubated on ice for 30 min. The mixture was then heat-shocked for 45 seconds at 42°C followed by a 2 min chill on ice. Nutrient broth recovery medium was added to the mixture which was then incubated at 37°C with shaking. The cells were then plated on LA + kanamycin plates (25 µg/ ml kanamycin) and incubated at 37°C overnight. One of the *E. coli* BL21 transformed colonies was used to grow 5 ml of overnight starter culture which was then added to 1 litre of nutrient broth (25 µg/ ml kanamycin) and cultured at 37°C to an OD$_{600\ nm}$ of 0.4–0.6. Expression was then induced by adding IPTG to the culture to a final concentration of 1 mM. After the addition of IPTG the culture was incubated overnight at 30°C. The cell culture was then centrifuged at 4000 x g for 15 min and the cell pellet was stored at -20°C.

2.5.3 Cell lysis

The stored cell pellets were thawed and resuspended in 50 mM Tris-HCl, pH 8 buffer with EDTA-free protease inhibitors (Roche Ltd, USA) on ice and then sonicated for 6 min (15sec pulsar on/off) using a Sonicator® 3000 (Misonix, USA). An ethanol and ice water bath was used to cool the sample during the sonication. After sonication samples were centrifuged at

20000x g for 30 min and ammonium sulphate precipitation was immediately carried out on the supernatant.

2.5.4 Ammonium sulphate $(NH_4)_2SO_4$ precipitation

Increasing amounts of ammonium sulphate were added using 10% saturation increments from 20%–50%. After each addition of ammonium sulphate, the sample was incubated for 20 min on ice to allow proteins to precipitate out of solution. After incubation, the suspension was centrifuged at 10000 x g for 10 min at 4°C. The supernatant was then collected and its volume was measured in order to calculate the mass of ammonium sulphate needed for the next 10% saturation increase. The ammonium sulphate was then added to the supernatant while the pellet was resuspended in 50 mM Tris-HCl, pH 8 buffer. The process was repeated until there was no further apparent protein precipitation. Samples were analysed by SDS-PAGE gel (section 2.3.9) to determine which fraction contained the $CynD_{pum}$.

2.5.5 Anion exchange chromatography

The ammonium sulphate precipitation fraction with the $CynD_{pum}$ was run through a Q Sepharose XL anion exchange column (Amersham Biosciences, Sweden) for further protein separation. The proteins were eluted with a linear gradient from a low salt buffer (50 mM Tris-HCl pH 8, 150 mM NaCl) to a high salt buffer (50 mM Tris-HCl pH 8, 1M NaCl) at a flow rate of 2.5 ml/min. SDS-PAGE gels (section 2.3.9) were run for all fractions that produced a A_{280nm} peak and the results were used to identify the fraction that contained the $CynD_{pum}$.

2.5.6 Gel filtration size exclusion chromatography

The protein solution from the anion exchange chromatography was then run through a TSKgel PWXL4000, HPLC gel filtration column (Tosoh Corporation, Tokyo) pre-equilibrated and run with 150 mM NaCl in 50 mM Tris-HCl pH 8 buffer, at a flow rate of 0.5 ml/min. $CynD_{pum}$ fractions with the highest A_{280nm} peak were collected and tested for activity using the picric acid activity assay (section 2.3.7). SDS-PAGE gels (section 2.3.9) were also run for the collected fractions to access the purity on the enzyme.

2.5.7 Enzyme activity assay

The picric acid activity assay (Fisher and Brown, 1952) was used to measure cyanide degrading activity of the purified CynD_{pum} protein. 20 µl of 25 mM KCN was added to 80 µl of enzyme solution in 50 mM Tris-HCl pH 8 buffer (5 mM final [KCN] in 100 µl reaction mixture). The reaction was allowed to continue for at least 30 min at 37°C and stopped by addition of 80 µl of the picric acid/carbonate mixture (50% picric acid, 50% carbonate). The samples were then boiled for 5 minutes to allow the colour complex to develop. Distilled water at room temperature (1 ml) was added to stop the colour reaction. Absorbance of 100 µl of this solution was measured at 520 nm on a Multiskan spectrophotometer (Titertek, Helsinki). A brown to yellow colour change represented cyanide degrading activity.

2.5.8 Protein concentration determination

The concentration of protein was determined using the Bradford assay (Bradford, 1976). Bovine serum albumin was used to prepare 0.1–0.5 mg/ml protein standards. The protein samples (10 µl) were mixed with 200 µl of the Bio-Rad protein assay solution (Bio-Rad, USA) and allowed to incubate for 10 min at room temperature. The absorbance was then measured at 595 nm in a Multiskan spectrophotometer (Titertek, Helsinki).

2.5.9 SDS-PAGE

SDS-PAGE was performed on BIORAD Mini-PROTEAN® 3 Cell (BIORAD, UK) gel apparatus. 15 µl of protein samples were mixed with 5 µl of 4× Sample Loading Buffer. 10 µl of the mixture was heated to 100°C for 10 min and then run on a 10% SDS polyacrylamide gel. Page Ruler Prestained protein ladder (Thermo Fisher Scientific, USA) was used to determine the molecular weight of the protein. After the run, the SDS gels were stained with Coomassie Brilliant Blue R-250.

2.5.10 Negative stain electron microscopy

Purified samples from gel filtration chromatography were buffer exchanged in buffers of desired pH using Nanosep 10kDa molecular weight cut-off, spin columns (Microsep Pty Ltd, South Africa). The samples were buffer exchanged with 50 mM Citrate/ Na_2HPO_4 buffer at pH 5.4 or left in 50 mM Tris-HCl for pH 8 visualisation. Carbon-coated copper grids were glow-discharged at a plasma current of 25 mA for 30 seconds using a EMS100× Glow Discharge

Unit (Electron Microscopy Sciences, USA). 3 µl of protein solution was then placed on the grid and allowed to stand for 30 seconds. The grid was then washed with 2 drops of water then with 3 drops of uranyl acetate while blotting with filter paper between each wash. The grids were blotted and air dried then viewed on the Tecnai T20 transmission electron microscope. Micrographs were collected at 50 000 x calibrated magnification using a Gatan US2000 CCD camera (Gatan, USA). The particle average lengths and length ranges were calculated from 50 randomly selected oligomers per pH condition.

2.5.11 Cryo-electron microscopy sample preparation

2.5.11.1 Sample preparation for the FEI Polara microscope

For the FEI Polara dataset, purified Q86R+H305K+H308K+H323K CynD$_{pum}$ at pH 8.0 was buffer exchanged into citrate buffer pH 5.4 using a spin filter (Waters, USA) and diluted out to a concentration of 0.07 mg/ml. Samples were stored at 4 °C for 48 hours prior to being applied to the grids to allow elongation of the CynD$_{pum}$ helical filaments. Grids were prepared at 100 % humidity using a Vitrobot (FEI, USA). 4 µl of sample was applied to a glow-discharged R2/2 copper grid overlaid with a 7 nm continuous carbon film (Quantifoil, Germany), incubated for approximately 60 seconds, blotted from both sides (3.5 seconds) and plunge-frozen in liquid ethane. The sample preparation was carried out with the generous help of Dr Jean Watermeyer (Birkbeck College, London)

2.5.11.2 Sample preparation for FEI Titan Krios microscope

For the Q86R+H305K+H308K+H323K and wild-type CynD$_{pum}$ FEI Titan Krios datasets a similar cryo-EM sample preparation procedure as section 2.3.11.1 was used but with more experimentally refined conditions. The changes included the use of a diluted concentration of 0.2 mg/ml, a sample volume of 2.5 µl for the grid application and the use of R2/2 holey carbon coated grids.

2.5.12 Cryo-electron microscopy data collection

2.5.12.1 Data collection on the FEI Polara microscope

Images were collected under cryo conditions using a liquid nitrogen-cooled FEI Polara transmission microscope (FEI, The Netherlands) operated at 300 kV. Imaging was carried out using a GIF Quantum energy filter (20 e^-V slit width) and a K2 Summit® Direct Detector (Gatan, USA). Images were collected in counting mode at a dose rate of ~1 $e^-/Å^2$/sec (approximately 5 e^-/pixel per second), with the total dose of ~25 $e^-/Å^2$ fractionated into 100 x 0.25 sec sub-frames or 50 x 0.5 sec sub-frames. Serial EM software (Mastronarde, 2005) was used for low-dose data collection. Images were collected at a defocus range of -2.5 – -0.4 μm at a magnification of 21834X with a calibrated physical pixel size of 2.29 Å/pixel. The setup of the FEI Polara microscope was performed with the generous help of Dr Dan Clare (Birkbeck College, London)

2.5.12.2 Data collection on the FEI Titan Krios microscope

Images were collected under cryo conditions using a liquid nitrogen-cooled FEI Titan Krios transmission electron microscope (FEI, The Netherlands) operated at 300 kV. Imaging was carried out using a GIF Quantum energy filter (20 e^-V slit width) and a K2 Summit® Direct Detector (Gatan, USA). Images were collected in counting mode at a dose rate of ~9.9 $e^-/Å^2$/sec (approximately 7 e^-/pixel / second), with the total dose of ~59.4 $e^-/Å^2$ fractionated into 30 x 0.2 sec sub-frames. The Gatan Microscopy Suite software (Gatan, USA) was used for low-dose data collection. Images were collected at a defocus range of -2.5 – -0.5 μm at a magnification of 165000 X with a calibrated physical pixel size of 0.84 Å/pixel. The setup of the FEI Titan Krios microscope was performed with the generous help of Mr Deryck Mills (The Max Planck Institute of Biophysics, Frankfurt)

2.5.13 Cryo-electron microscopy data processing

2.5.13.1 Data processing of the FEI Polara electron microscope images

Image processing was performed using multiple software as outlined in Fig. 2.3. Using MOTIONCORR (Li et al., 2013), dose-fractionated sub-frames of each cryo-EM movie were aligned using a whole-image method then averaged to produce motion-corrected average images. The motion-corrected average images were uploaded into the Appion image processing workflow (Lander et al., 2009). CTFFIND3 (Mindell et al., 2003) was used to calculate the contrast transfer function (CTF) of each motion-corrected average image and images with significant drift and astigmatism were discarded. CTF correction by phase-flipping was carried out on entire images using ACE2 (Mallick et al., 2005). Helical particle picking was accomplished interactively using Appion's integrated Manual picker to produce helical segments of 160 x 160 pixels with 94% overlap, along the length of the fibre. The best rotationally aligned segments were then identified and extracted using the image-sort-by-statistic (Junksort) function from XMIPP (Sorzano et al., 2009), accessed within Appion. At this stage of the image processing, the stack containing the best segments was exported to SPIDER (Frank et al., 1996) where the segments were centred by applying integer x shifts only and classified based on their diameter. The stack with the centred and diameter classified segments was then subjected to iterative helical real space reconstruction (IHRSR) with IHRSR++ scripts (Egelman et al., 2000 & Parent et al., 2010) in SPIDER (Frank et al., 1996). IHRSR was carried out using initial helical parameters of $\Delta\varphi$ = -77.0° and Δz = 16.0 Å, an in-plane angle increment of 4° around the helical axis and out-of-plane tilt increments of 4° each to a maximum of 32°. The cut segments were initially divided into two defocus groups (-0.71µm – -2µm and -2µm – -3.4µm), were the images from the far from focus group (-2µm – -3.4µm) were used to generate a starting model for sequential IHRSR runs. Due to their high contrast the far from focus images generated a starting model which improved the alignment of the cut segments from the whole image stack during sequential IHRSR runs thus improving the quality of the model. The final model was further sharpened by applying a B-factor correction to the map using the EMBFACTOR (Fernandez et al., 2008) program. The resolution range for the B-factor calculation was empirically determined using a trial-and-error method were the noise level, continuity of backbone densities and features of side-chain densities were monitored.

2.5.13.2 Data processing of the FEI Titan Krios electron microscope images

The FEI Titan Krios electron microscope images of the Q86R+H305K+H308K+H323K and wild-type CynD$_{pum}$ were processed using slightly different strategies to the FEI Polara electron microscope images. The dose-fractionated sub-frames of the cryo-EM movie were initially aligned using MOTIONCORR (Li et al., 2013) with a binning factor of 2 to produce motion-corrected average images with a pixel size of 1.7 Å/pixel. CTFFIND3 (Mindell et al., 2003) was used to calculate the contrast transfer function (CTF) of each motion-corrected average image and images with significant drift and astigmatism were discarded. Segments along the length of the helical fibres were picked using E2HELIXBOXER from the EMAN suite (Ludtke et al., 1999). SEGMENT from the SPRING package (Desfosses et al., 2014) was then used to apply a complete CTF amplitude and phase correction and extract helical segments of 240 x 240 pixels with 94% overlap. The segment stacks were visually inspected using V2 of the

EMAN suite (Ludtke et al., 1999) and bad segments (bent, overlapping and broken) were manually deleted. At this stage of the image processing, the stacks containing the best segments were exported to SPIDER (Frank et al., 1996) where the segments were centred by applying integer x shifts only and classified based on their diameter. The stacks with the centred and diameter classified segments were then used for the iterative helical real space reconstruction (IHRSR) with IHRSR++ scripts (Egelman et al., 2000 & Parent et al., 2010) which run in SPIDER (Frank et al., 1996). The model generated from the FEI Polara microscope dataset was used as an initial model. This above step of using the binned images was introduced to allow for a faster optimisation of the IHRSR processing and to produce the best starting model for the following steps.

The second part of the image processing of the FEI Titan Krios data was performed as outlined in Fig. 2.3. The original sub-frames of each cryo-EM image movie were aligned again using a modified version of MOTIONCORR (Li et al., 2013, modified by Mr Michael Kunz, Buchmann Institute for Molecular Life Sciences, Frankfurt) to produce motion-corrected average images with no binning and with the original pixel size of 0.84 Å/pixel. The modified version of MOTIONCORR showed improved accuracy of frame alignment of the low contrast unbinned frames than the original program. Using a similar procedure to that of the FEI Polara data processing (Fig. 2.3), the motion-corrected average images were uploaded into the Appion image processing workflow (Lander et al., 2009). CTFFIND3 (Mindell et al., 2003) was used to calculate the contrast transfer function (CTF) of each motion-corrected average image and images with significant drift and astigmatism were discarded. Complete CTF amplitude and phase correction was carried out on entire images using ACE2 (Mallick et al., 2005). Helical particle picking was accomplished interactively using Appion's integrated Manual picker to produce helical segments of 448 x 448 pixels with 94% overlap, along the length of the fibre. The best rotationally aligned segments were then identified and extracted using the image-sort-by-statistic (Junksort) function from XMIPP (Sorzano et al., 2009), accessed within Appion. Due to the poor selection of good segments by Junksort, V2 (Ludtke et al., 1999) was used to carry out an extra step of visual inspection and manual deletion of bad segments (bent, overlapping and broken). The stack containing the best segments were exported to SPIDER (Frank et al., 1996) where the segments were centred by applying integer x shifts only and classified based on their diameter. The stacks with the centred and diameter classified segments were then subjected to iterative helical real space reconstruction (IHRSR) with IHRSR++ scripts (Egelman et al., 2000 & Parent et al., 2010) in SPIDER (Frank et al., 1996). IHRSR was carried out using the previously determined helical parameters of Δφ = -77.0° and Δz = 16.5 Å as an initial estimate. The search templates were projections of the previous reconstruction with in-plane angle increments of 1° around the helical axis and out-of-plane tilt increments of 4° each to a maximum of 20°. To further improve the multi-reference alignment for the wild-type dataset, out-of-plane tilt increments of 4° each to a maximum of 32° were used. To improve the IHRSR reconstruction, the previous models obtained using the images with a binning factor of 2 were used as a starting model. The final models were further sharpened by applying a B-factor correction to the maps and scaling amplitudes using the EMBFACTOR (Fernandez et al., 2008) program. Similar to the FEI Polara data, the resolution range for the B-factor calculations were empirically determined using a trial-and-error method in which the noise level, continuity of backbone densities and features of side-chain densities were monitored.

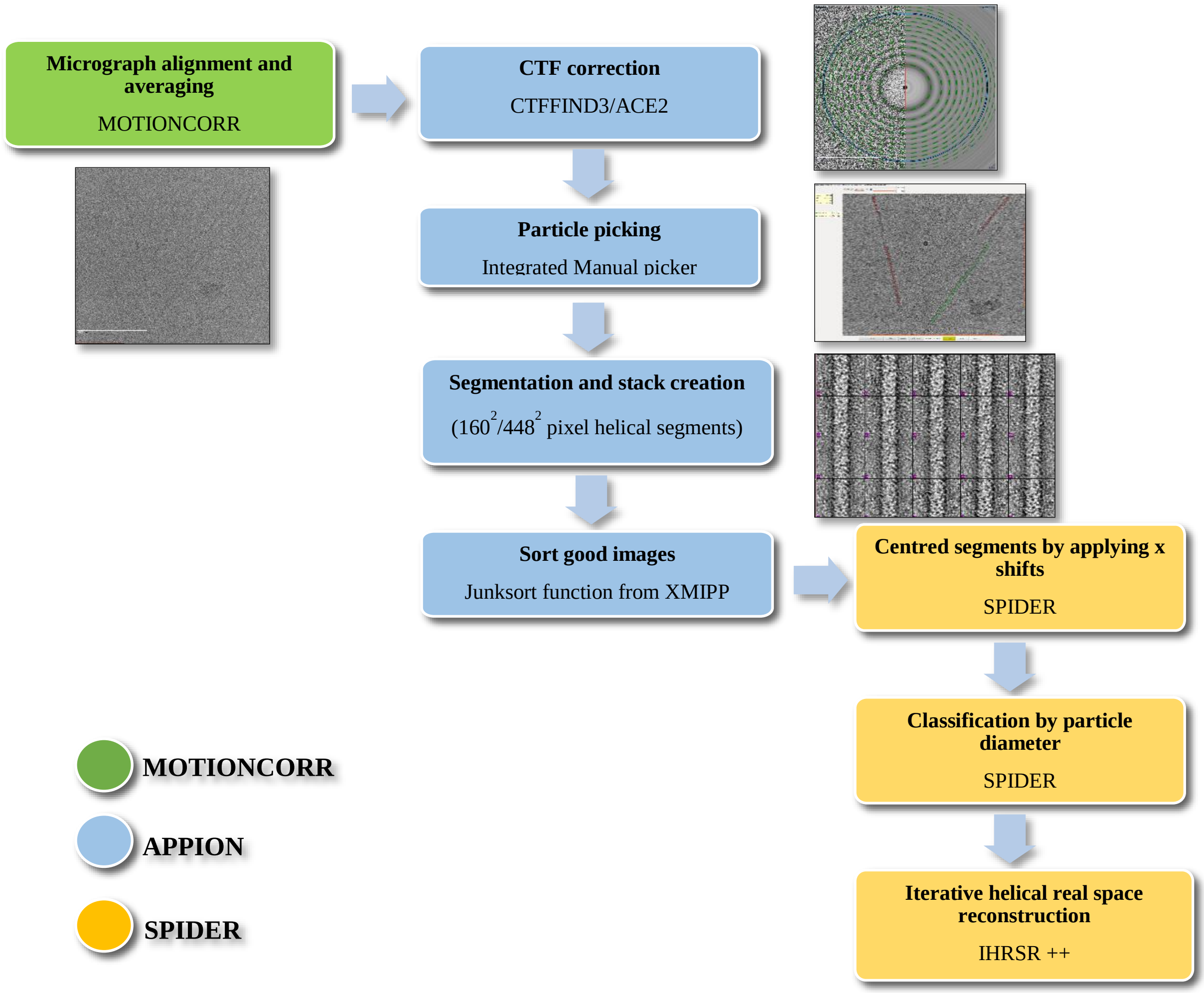


Figure 2.3. Outline of the image processing method used to obtain the 3D reconstructions by using multiple softwares.

2.5.14 Resolution determination

The Fourier Shell Correlation (FSC) method was used for resolution determination. As suggested by Yang et al. (2003), two semi-independent IHRSR 3D reconstructions were generated using different arbitrary user-defined initial helical symmetry parameters and a featureless cylinder as a starting model to prevent any false correlations at high frequencies. The 3D model reconstructions were iterated to allow convergence of the helical symmetry and the resultant models were used in subsequent steps. False corrections at high frequencies due to common mask artefacts were avoided by using different inner tube radii during symmetry imposition for each reconstruction. The two volumes were then aligned (using SPIDER) in z and φ to bring them into register with each other before calculating the FSC. The resolution-defining threshold value of 0.5 (Beck-Mann et al., 1997; Bottcher et al., 1997) and 0.143 (Orlova et al., 1997) were used to interpolate and calculate the resolution of the 3D models.

2.5.15 Visualisation and docking of homology model

Visualisations and inspections of the 2D images were performed in Appion (Lander et al., 2009) and EMAN's V2 (Ludtke et al., 1999). UCSF Chimera (Petterson et al., 2004) was used to view the 3D reconstructions and to manually dock the homology model into the final volume, aided by the 'Fit model in map' function. UCSF Chimera was also used to measure the local correlation between two maps using the 'Local correlation' function. Simulated EM maps were calculated and generated from a CynD$_{pum}$ homology model using pdb2vol function in SITUS (Wriggers, 2012).

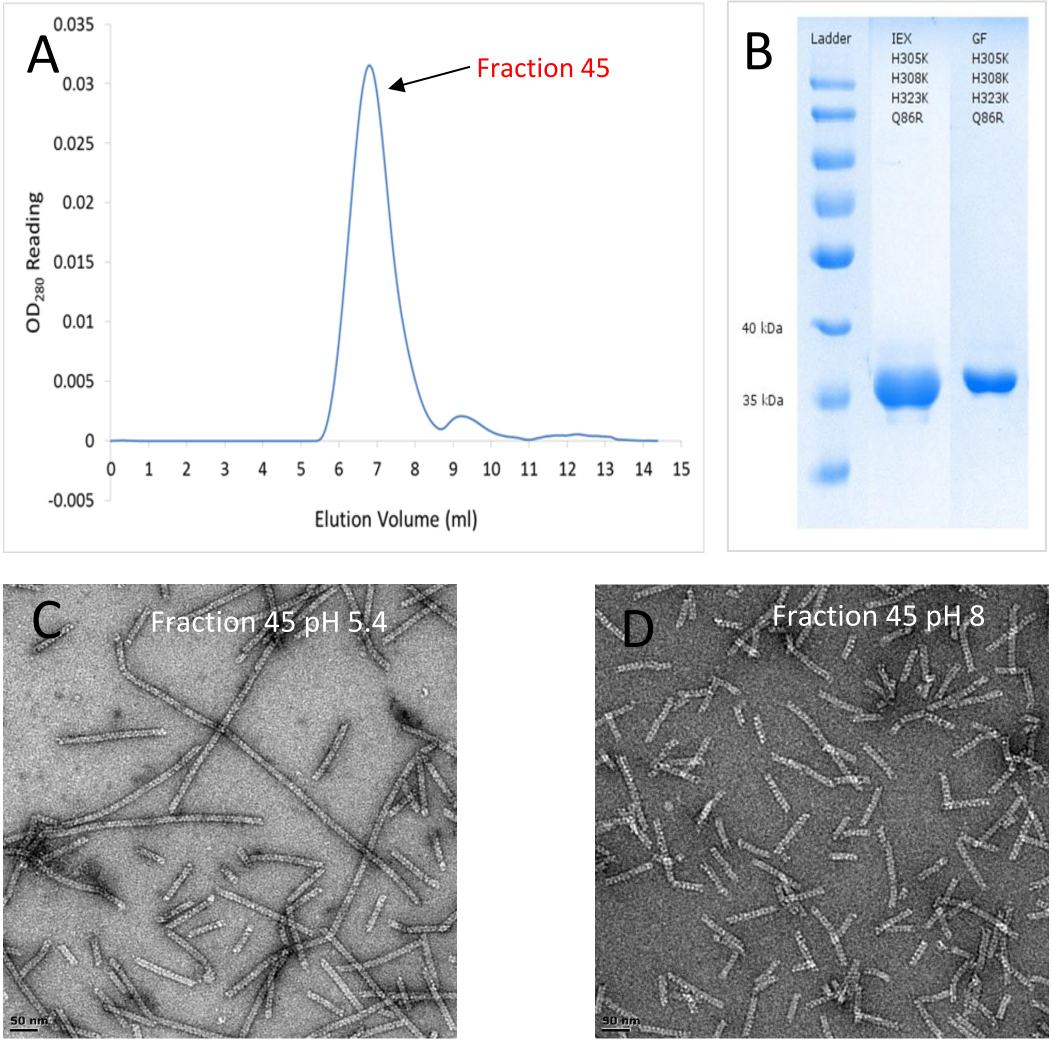
2.6 Results

2.6.1 DNA Preparation and Purification

The introduction of the Q86R mutation to the already synthesised H305K+H308K+H323K plasmid was successful and was confirmed by DNA sequencing. After protein expression a three step purification process was used to purify the $CynD_{pum}$ enzymes. The successful purification of the Q86R+H305K+H308K+H323K $CynD_{pum}$ variant was demonstrated by both the appearance of a distinct single pronounced band of about 37 kDa (corresponding to the monomeric size of $CynD_{pum}$) in the SDS-PAGE gel (Fig. 2.4B) and the distinct sharp peak around fraction 45 (7 ml elution volume) in the gel filtration chromatography elution profile (Fig. 2.4A). The Q86R+H305K+H308K+H323K variant remained active throughout the purification process and this was measured using the picric acid assay (Fisher and Brown, 1952). Wild-type $CynD_{pum}$ purification results were similar to the ones previously obtained (Mulelu, 2013).

2.6.2 Negative stain electron microscopy

Using negative stain electron microscopy, we confirmed that the Q86R+H305K+H308K+H323K variant formed significantly longer fibres (4–6 fold longer) than the wild-type at pH 5.4 (Fig. 2.4C). Although shorter than the pH 5.4 fibres (~92 nm average), the variant retained its fibre form at pH 8 (Fig. 2.4D) whereas the wild-type is known to exist as short spirals (~23 nm) at the same pH.



Average particle length = 249 nm
Range of particle length = 79 –538 nm

Average particle length = 92 nm
Range of particle length = 49 –151 nm

Figure 2.4. Protein purification data of the stable Q86R+H305K+H308K+H323K CynD$_{pum}$ variant. **(A)** Gel filtration chromatography elution profile where the elution of the CynD$_{pum}$ variant at pH 8 was monitored by measurement of absorbance at 280 nm, **(B)** 10% SDS-PAGE gel profile and negative stain electron microscopy micrograph of the Q86R+H305K+H308K+H323K CynD$_{pum}$ variant at **(C)** pH 5.4 and **(D)** pH 8.

2.6.3 Cryo-electron microscopy data processing

2.6.3.1 Data processing of the Q86R+H305K+H308K+H323K CynD$_{pum}$ FEI Polara electron microscope images

Using the purified Q86R+H305K+H308K+H323K CynD$_{pum}$ variant sample at pH 5.4 we made cryo-EM grids. We then collected an image dataset using the FEI Polara electron microscope in Birkbeck College, London. After optimisation of the enzyme concentration and blotting time using the Vitrobot (FEI, USA) an optimal concentration of the enzyme was obtained and this was demonstrated by the non-overlapping, even distribution of the helical fibres (Fig. 2.5A). The retention of contrast at low defocus levels demonstrated the presence of thin vitreous ice.

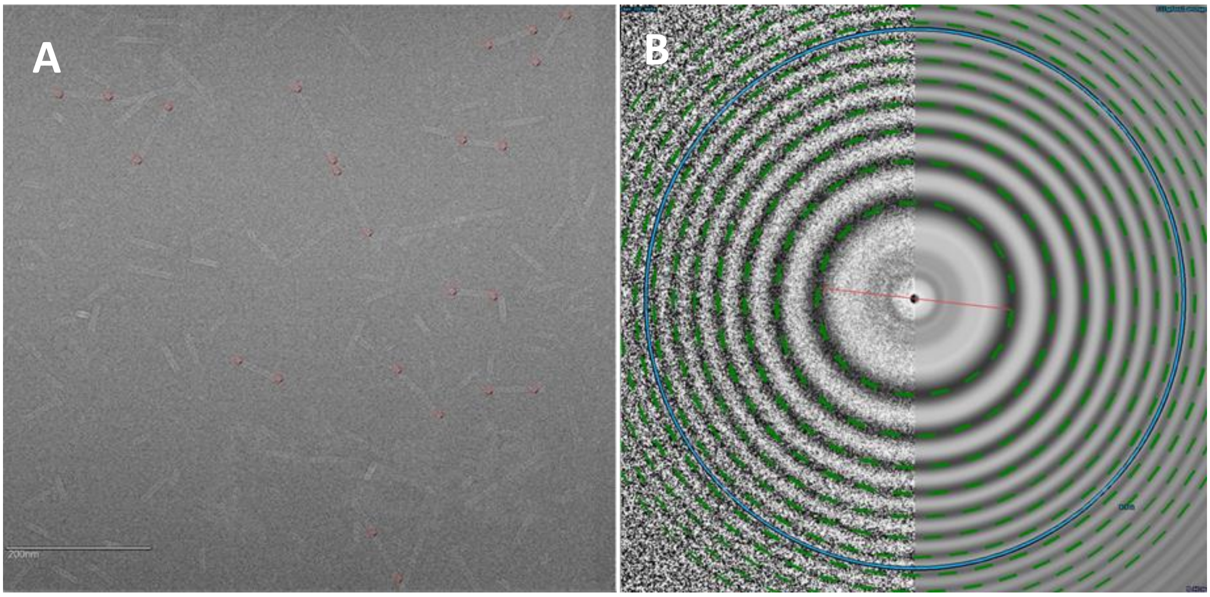


Figure 2.5. **(A)** Cryo-electron micrograph of the Q86R+H305K+H308K+H323K CynD$_{pum}$ variant at pH 5.4 collected on an FEI Polara microscope with marked helical particles picked using Appion’s integrated Manual picker program. **(B)** CTFFIND3 generated graphical output containing a divided image with one half showing the averaged power spectrum calculated from the micrograph (left) and the other half showing the simulated power spectrum of the determined CTF (right).

The FEI Polara EM images were collected as dose-fractionated movie frames which were later aligned and averaged using MOTIONCORR (Li et al., 2013). Correction of the CTF of the microscope was then carried out in order to improve the final resolution of the 3D reconstruction. CTF correction was carried out by first determining the image’s defocus and astigmatism using CTFFIND3 (Mindell et al., 2003) and then applying the calculated CTF using ACE2 (Mallick et al., 2005). For each micrograph CTFFIND3 generated a graphical output containing a divided image with one half showing the averaged power spectrum calculated from the micrograph and the other half showing the simulated power spectrum of the determined CTF (Fig. 2.5B). Using these graphical outputs, micrographs demonstrating significant drift and astigmatism were discarded and of the 147 micrographs originally collected only 82 were kept for further processing (Table 2.3).

CTFFIND3 also generated a plot of the defocus levels of each micrograph (Fig. 2.6A) with a confidence score of the calculation (Fig. 2.6B). The results from the CTFFIND3 showed an even distribution of the image defocus, ranging from -0.71µm – -3.4µm with a majority of images within the range from -1.3µm to -1.9µm. Confidence score of the CTF calculations were all above 99% demonstrating a high accuracy in the CTFFIND3 calculations.

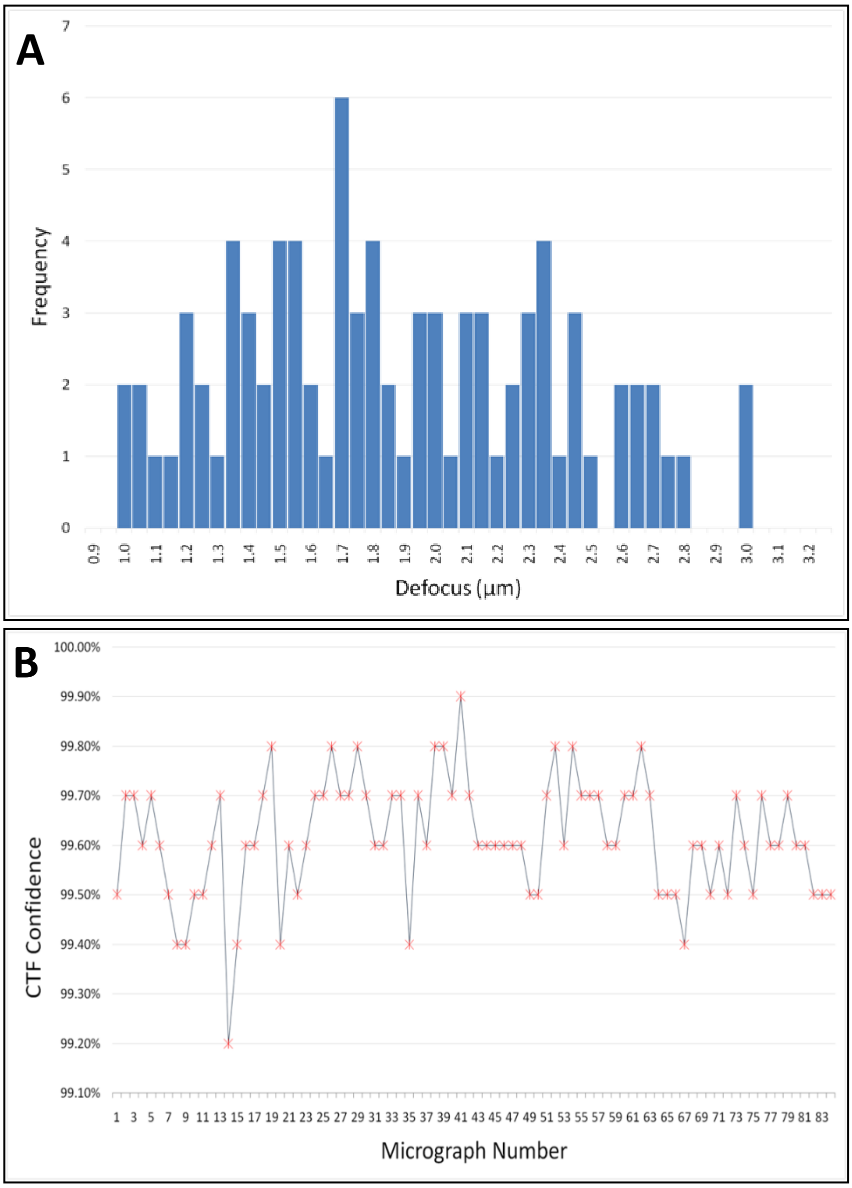


Figure 2.6. (A) Frequency distribution of defocus levels of collected micrographs calculated by CTFFIND3 during the CTF correction of the FEI Polara Q86R+H305K+H308K+H323K CynD$_{\mathrm{pum}}$ dataset. **(B)** The CTF confidence of the calculated CTF of each micrograph.

Particle picking was performed manually in which straight intact helical fibre were selected on each micrograph while avoiding the selection of terminal areas of the fibres which may introduce structural variability as shown in Fig. 2.5A. The selected fibres were then segmented into 160 x 160 pixel segments with 94% overlap. The initial segmentation produced 29099 segments of which 21099 were discarded after analysis by XMIPP’s image-sort-by-statistic function (Table 2.3) (Sorzano et al., 2009). The image-sort-by-statistic function automatically sorts images and ranks them based on how closely they resemble the average, to allow the user to determine and apply a cut-off where only the best images are selected. The helical segments were further sorted based on their diameter to reduce the degradation of the final 3D model due to the structural heterogeneity of the helical segments. The diameter size distribution of the helical segments showed that a majority (4262 segments) of fibres were between 112–128 Å and these segments were selected and used in the IHRSR (Fig. 2.7).

Table 2.3. Image pre-processing summary for the FEI Polara image dataset showing the number of micrographs or segmented helical filaments after each image pre-processing step.

Number of micrographs collected	141
Number of micrographs after discarding of bad CTF images	82
Number of extracted segments after particle picking	29099
Number of segments in stack after sorting by statistics	8000
Number of segments in stack after manual sorting	8000
Number of segmented images in stack after sorting by diameter	4262

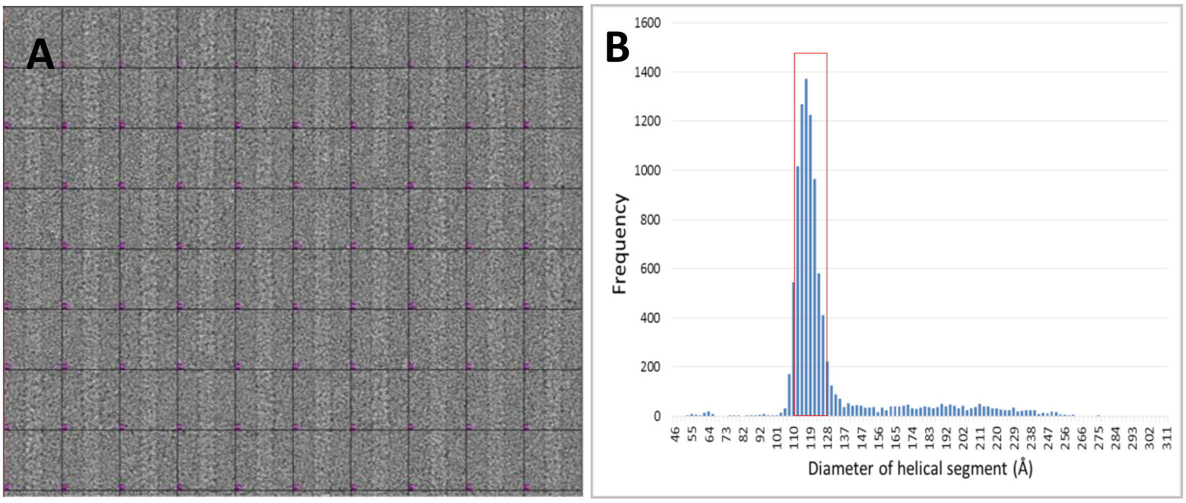


Figure 2.7. (A) Image stack of the Q86R+H305K+H308K+H323K CynD$_{pum}$ variant at pH 5.4 which consists of centred and diameter classified helical segments viewed using the V2 program. The stack was produced by cutting selected helical fibres from cryo-electron micrographs into 160 X 160 pixel segments. **(B)** Diameter size distribution of the helical segments and (in red box) selected fibre diameter (112–128 Å) of particles used in IHRSR.

The diameter sorted stack was then subjected to IHRSR to produce a 3D reconstruction of the Q86R+H305K+H308K+H323K CynD$_{pum}$ variant. During the first rounds of IHRSR distorted maps were generated and the cause of this was traced back to the presence of high out-of-plane tilt angles of some of the helical segments, tilt we initially assumed was minimal and negligible. Thus before any data processing could be carried out, an additional out-of-plane tilt optimisation was performed. The out-of-plane tilt optimisation involved carrying out multiple IHRSR runs at incrementally increased out-of-plane tilt angles until the quality of the resultant 3D models showed no further improvement and the frequency distribution of the reference projections generated from the multi-reference alignment showed a flat and even distribution. As the out-of-plane tilt angle increased the high frequency cluster of the number of helical segment images assigned to reference projections at angles at the user-defined upper and lower limits decreased, and this decrease corresponded to improvement in the quality of the generated EM maps (Fig. 2.8). From the out-of-plane tilt optimisation results we determined that an out-of-plane tilt angle of 30° was ideal for further IHRSR processing. The out-of-plane tilt was also evident when comparing the average power spectrum generated from a full dataset of the helical segments (Fig. 2.9A) and the average power spectrum from a subset of helical segments identified as not having no out-of-plane tilt during the multi-reference alignment (Fig. 2.9B). The out-of-plane tilt was demonstrated by the blurring of diffraction spots from the average power spectrum of all the helical segments when compared to the average power spectrum of the 0° tilt helical segments, thus showing a misalignment during the superimposition of the power spectra diffraction spots when helical segments with out-of-plane tilt are introduced (Fig. 2.9).

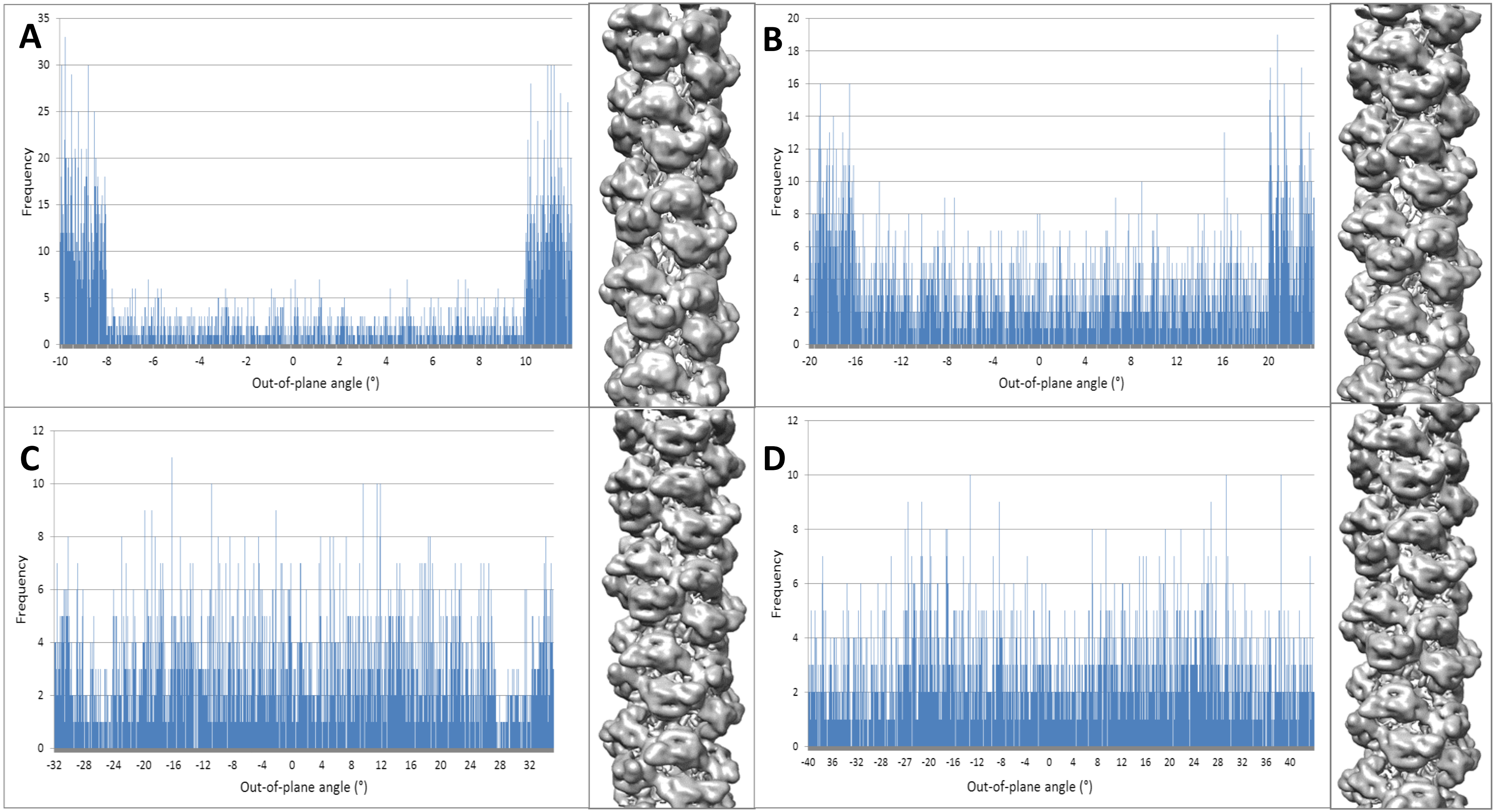


Figure 2.8. Frequency distribution of the reference projections and resultant 3D model generated during out-of-plane tilt angle optimisation. The out-of-plane tilt angle optimisation involved carrying out multiple IHRSR runs at incrementally increased out-of-plane tilt angles **(A)** 10°, **(B)** 20°, **(C)** 32° and **(D)** 40° until the quality of the resultant 3D models showed no further improvement and the frequency distribution of the reference projections generated from the multi-reference alignment showed a flat and even distribution.

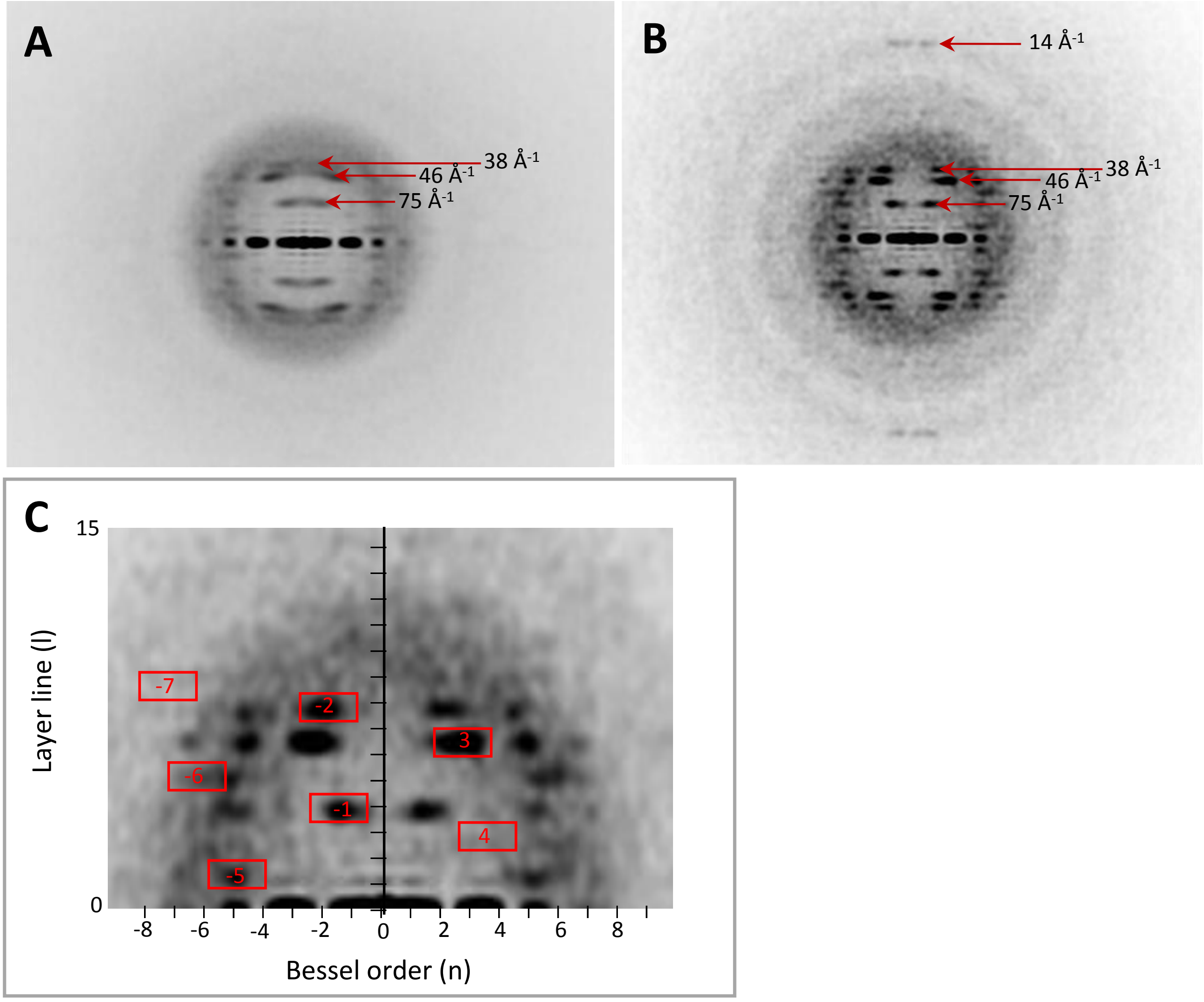


Figure 2.9. **(A)** Average power spectrum generated from a full dataset of the helical segments from the Q86R+H305K+H308K+H323K CynD$_{pum}$ variant collected on the FEI Polara microscope. **(B)** Average power spectrum generated from a subset of the helical segments identified as not having out-of-plane tilt during the multi-reference alignment. Positions of layer lines with visible diffraction spots are indicated by the red arrows. **(C)** Indexing of the power spectrum generated from the 0° out-of-plane tilt subset of the helical segments. The positions of the diffraction spots corresponding to the calculated Bessel orders are highlighted in red.

IHRSR was run for 20 cycles, to a point where no further changes in the helical symmetry were observed. The final IHRSR run converged at a helical symmetry solution with a twist of ($\Delta\varphi$) -77.0° and rise (Δz) of 16.5 Å. The same helical symmetry solution was obtained when IHRSR was run using two half image datasets and two different sets of arbitrary user-defined initial helical parameters, when calculating the resolution of the final 3D EM model (Fig. 2.10).

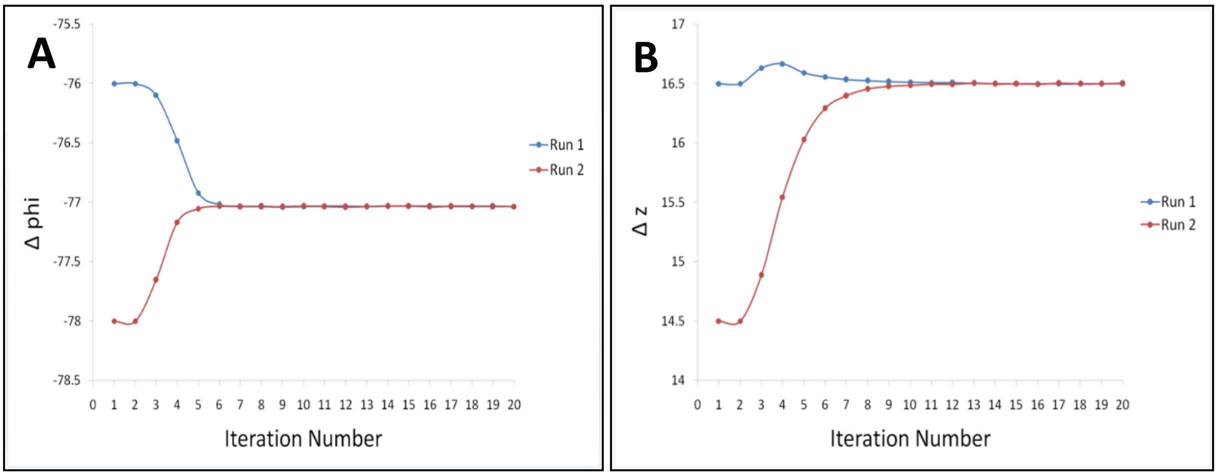


Figure 2.10. Helical symmetry solution after 20 IHRSR iterations of the FEI Polara cryo-EM images of the purified Q86R+H305K+H308K+H323K CynD$_{pum}$ variant. Two IHRSR runs were carried out using two half image datasets and two different sets of arbitrary user-defined initial helical parameters. The IHRSR run converged at a (**A**) Δphi of -77.0° and (**B**) Δz of 16.5 Å.

We obtained a detailed final model from the IHRSR (Fig. 2.11A) which we determined to be at a resolution of 11.1 Å using the FSC cut-off of 0.5 (Fig. 2.11B). Using the same FSC plot the resolution of the final model was also determined to be 7.6 Å using the FSC cut-off of 0.143. The final model was further sharpened by applying a B-factor to the map using the EMBFACTOR (Fernandez et al., 2008) program. EMBFACTOR calculated and applied a B-factor of -647.77 Å^2 (using the resolution range of 7–12 Å) which was used to scale up spherically average amplitude of the EM map. The Guinier plot produced by EM-BFACTOR after sharpening (Fig. 2.11D) showed the natural logarithm of the spherically average amplitude of the unsharpened map, the noise weighted amplitude, the restored amplitudes after sharpening, the zero scattering amplitude and the Wilson statistics as a function of the resolution (1/d^2). From the plot the amplification of the decaying spherically average amplitude of both the experimental and noise weighted (C$_{ref}$) amplitudes was observed after the application of the calculated B-factor during the sharpening of the 3D reconstruction. The sharpening improved the final model to the point where individual peptide alpha helices could be seen and differentiated (Fig. 2.11B). A CynD$_{pum}$ atomic homology model which we previously generated (Park et al., 2016) was then docked into the sharpened model to help validate the model (Fig. 2.11E and F). The homology model showed a high correlation to the sharpened model and most of the peptide backbone fitted perfectly in the EM density (2.11E and F).

Using the IHRSR calculated helical symmetry parameters, indexing of the diffraction spots from the average power spectrum, generated from the 0° out-of-plane tilt subset of the helical segments, was carried out (Fig. 2.9C). This step involved the comparing the ratios of different principal maxima of different Bessel orders with those from the experimentally generated power spectrum. The indexing scheme agreed with the left handed helical handedness which has been previously determined for the wild-type CynD$_{pum}$ by platinum-carbon shadowing of the CynD$_{pum}$ fibres (Jandhyala et al., 2003). The strong diffraction spot indexed as a Bessel function of order -1 on the power spectrum and the IHRSR calculated helical symmetry parameters both confirmed that the CynD$_{pum}$ Q86R+H305K+H308K+H323K variant exist as a helix with a pitch of ~ 75 Å and ~4.6 subunits per turn.

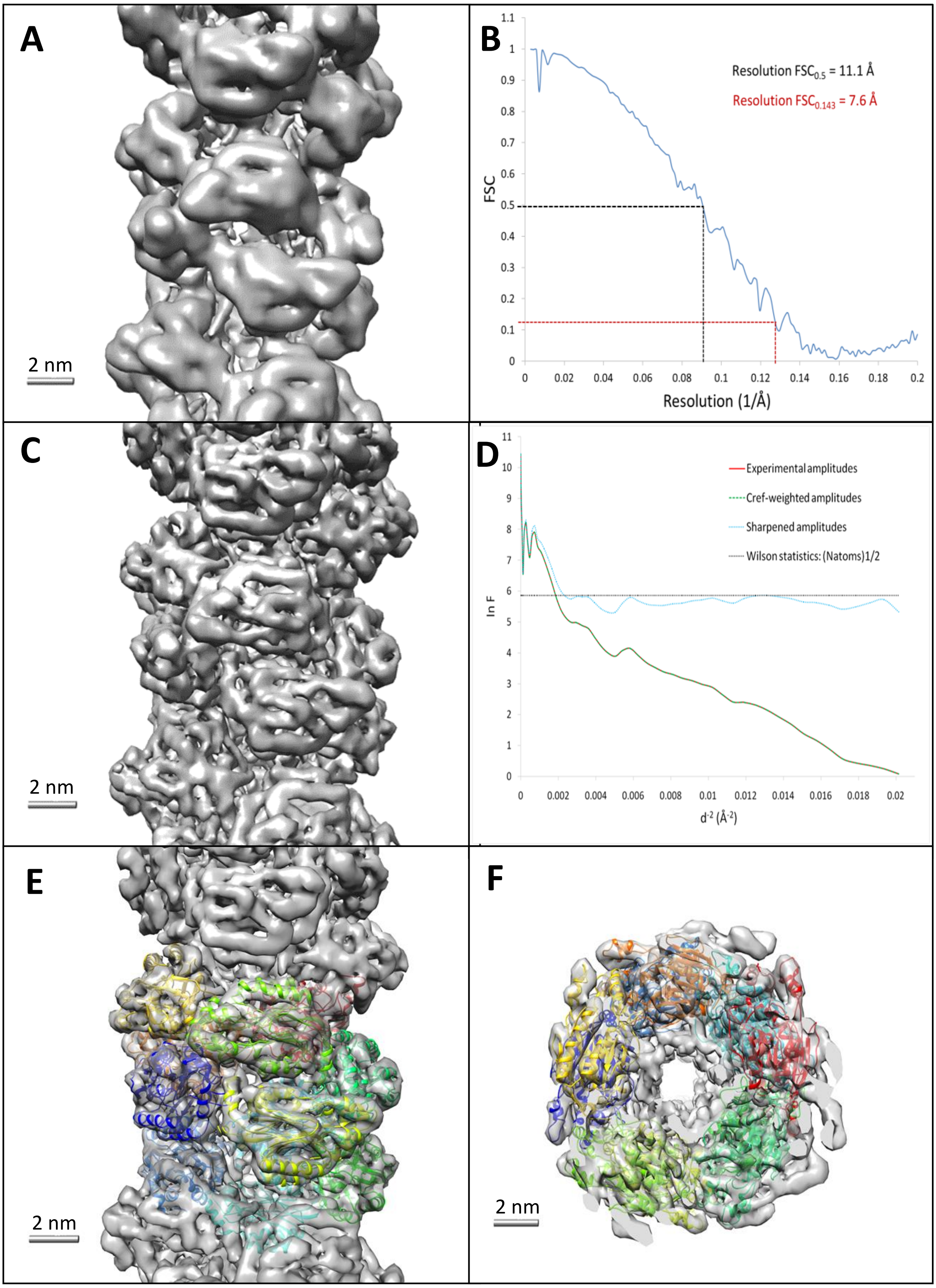


Figure 2.11. (A) The final 3D reconstruction produced from the FEI Polara cryo-EM images of the purified Q86R+H305K+H308K+H323K CynD_{pum} variant. **(B)** The Fourier Shell Correlation (FSC) curve of the final 3D reconstruction used to calculate the resolution of the model. **(C)** The EM-BFACTOR sharpened version of the same final 3D reconstruction. **(D)** The Guinier plot produced by EM-BFACTOR after sharpening of the final 3D reconstruction. **(E)** The docking of a CynD_{pum} homology model (Park et al., 2016) to the sharpened version of the final 3D reconstruction. **(F)** The same docked model viewed at a 90° rotation.

2.6.3.2 Data processing of the Q86R+H305K+H308K+H323K and wild-type CynD$_{pum}$ FEI Titan Krios electron microscope images

Using the purified Q86R+H305K+H308K+H323K and wild-type CynD$_{pum}$ sample at pH 5.4 we made cryo-EM grids. We then collected an image dataset using the FEI Titan Krios electron microscope at the Buchmann Institute for Molecular Life Sciences, Frankfurt. Using the knowledge gained when collecting and processing the FEI Polara EM data were able to improve the conditions used for data collection on the FEI Titan Krios electron microscope. Similar to the data collected on the FEI Polara microscope, the FEI Titan Krios micrographs showed an optimal enzyme concentration demonstrated by the non-overlapping, even distribution of the helical fibres (Fig. 2.12A and C). The FEI Titan Krios micrographs also demonstrated the presence of thin vitreous ice by the retention of contrast at small defocus levels as thick ice would have reduced the visibility of the fibres. On this microscope 302 and 230 micrographs were collected for the variant and the wild-type CynD$_{pum}$, respectively (Table 2.4). Both these number of micrographs were substantially higher than those collected using the FEI Polara microscope.

Table 2.4. Image pre-processing summary for the FEI Titan Krios image datasets showing the number of micrographs or segmented helical filaments after each image pre-processing step.

Enzyme variant	**Q86R+H305K+H308K +H323K**	**Wild-type**
Number of micrographs collected	230	302
Number of micrographs after discarding of bad CTF images	203	287
Number of extracted segments after particle picking	51697	60039
Number of segments in stack after sorting by statistics	45000	50000
Number of segments in stack after manual sorting	37946	35153
Number of segmented images in stack after sorting by diameter	28954	26284

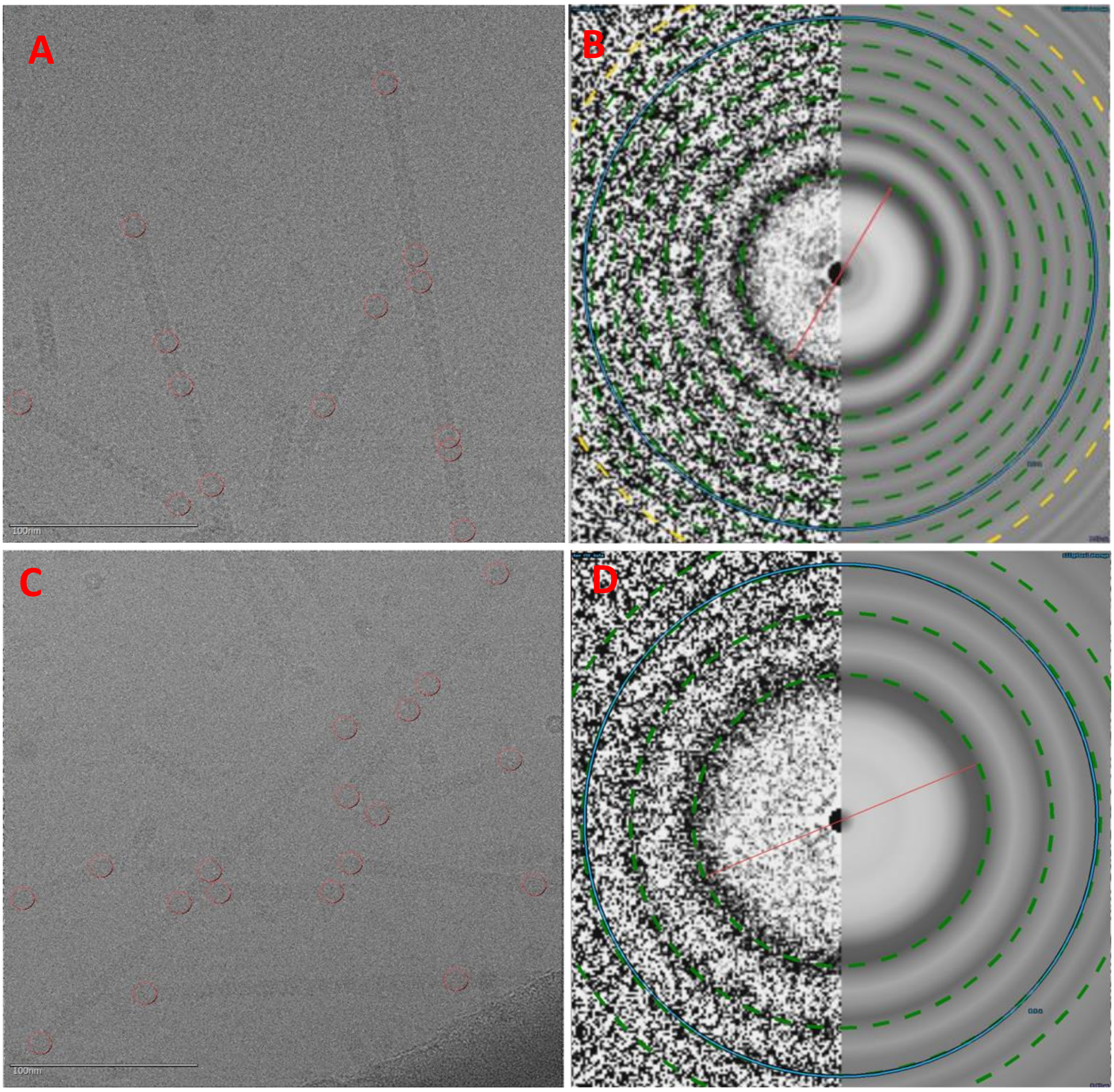


Figure 2.12. Cryo-electron micrograph of the Q86R+H305K+H308K+H323K **(A)** and wild-type **(C)** CynD$_{\text{pum}}$ at pH 5.4 collected on an FEI Titan Krios microscope with marked helical particles picked using Appion's integrated Manual picker program. CTFFIND3 generated graphical output, containing a divided image with one half showing the averaged power spectrum calculated from the micrograph (left) and the other half showing the simulated power spectrum of the determined CTF (right), of the Q86R+H305K+H308K+H323K **(B)** and wild-type **(D)** CynD$_{\text{pum}}$.

The pre-processing of the FEI Titan Krios EM data was carried out using a similar method to the one used for the FEI Polara EM dataset with slight variations. The FEI Titan Krios images were collected as dose-fractionated movie frames which were later aligned and averaged. After running CTFFIND3 on the averaged sub-frames for both the variant and wild-type CynD$_{\text{pum}}$ FEI Titan Krios micrographs, sets of averaged and simulated power spectra were generated (Fig. 2.12B and D). These power spectra generated from the FEI Titan Krios data showed less defined Thon rings than those generated by the FEI Polara data. Although the Thon rings of the averaged power spectra produced from the FEI Titan Krios were less defined, the CTFFIND3 program was still able to calculate the CTF but with lower confidence levels than those of the FEI Polara data (Fig. 2.13B and 2.14B). The Thon rings of the averaged power spectra were visible enough to show any signs of drift and astigmatism in the micrographs. Micrographs with significant drift and astigmatism were discarded. About 12% and 5% of the micrographs were discarded for variant and wild-type CynD_{pum} respectively (Table 2.4). These numbers of discarded micrographs were significantly reduced when compared to the number of micrographs discarded during the pre-processing of the FEI Polara data in which 42% of the micrographs were discarded. The defocus levels calculated by CTFFIND3 showed that a broader defocus range was used during data collection of the FEI Titan Krios micrographs. The calculated defocus of the Q86R+H305K+H308K+H323K CynD_{pum} data ranged from -0.55 µm – -2.9 µm with a majority of images within the range from -1.3 µm to -2.3 µm (Fig. 2.13A). The calculated defocus of the wild-type CynD_{pum} data ranged from -0.55 µm – -3.4 µm with a majority of images within the range from -0.8 µm to -1.8 µm (Fig. 2.14A).

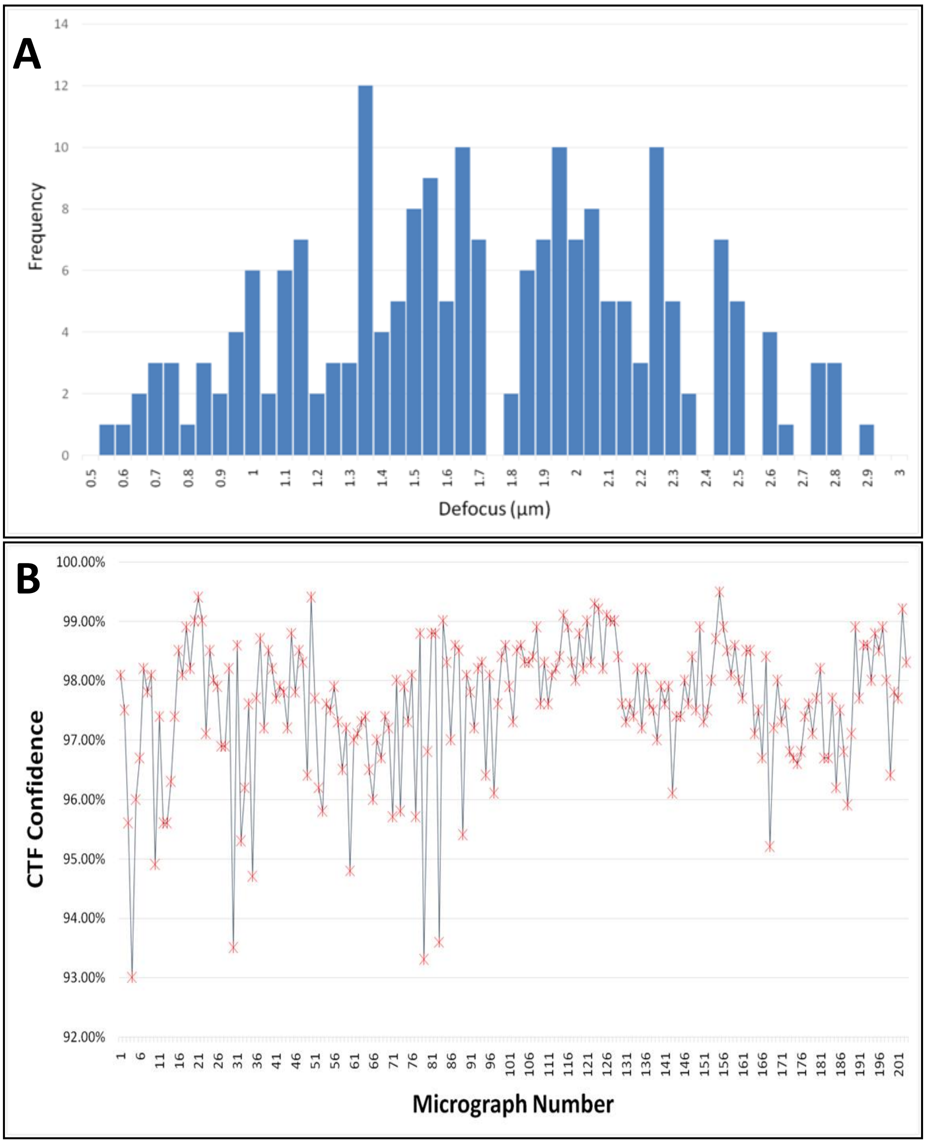


Figure 2.13. (A) Frequency distribution of defocus levels of collected micrographs calculated by CTFFIND3 during the CTF correction of the FEI Titan Krios Q86R+H305K+H308K+H323K CynD$_{\text{pum}}$ dataset. **(B)** The CTF confidence of the calculated CTF of each micrograph.

Particle picking and segmentation were carried out on the FEI Titan Krios datasets to produce 448 X 448 pixel segments with 94% overlap. The initial segmentation produced 51697 segments for the variant and 60039 for the wild-type (Table 2.4). These were further reduced to 45000 and 50000 for the variant and wild-type respectively, after analysis by XMIPP's image-sort-by-statistic function. Due to the poor selection of good segments by the image-sort-by-statistic function, V2 (Ludtke et al., 1999) was used to carry out an extra step of visual inspection and manual deletion of bad helical segments which were either bent, overlapping of broken. During this sorting process it was evident that fewer helical segment images were discarded for the FEI Titan Krios data (16% for the variant and 30% for the wild-type) when compared to the FEI Polara data (72% for the variant).

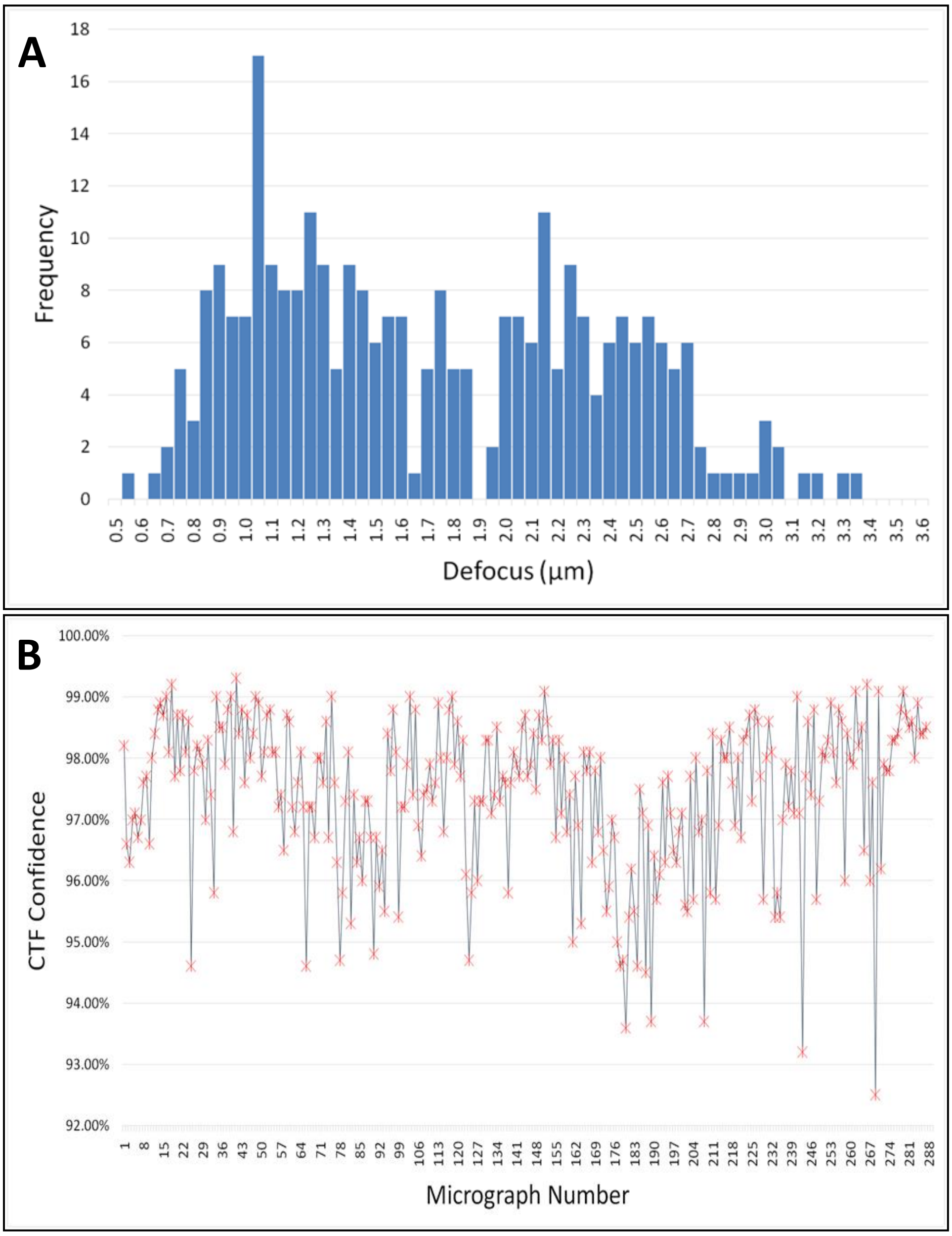


Figure 2.14. (A) Frequency distribution of defocus levels of collected micrographs calculated by CTFFIND3 during the CTF correction of the FEI Titan Krios wild-type CynD_{pum} dataset. **(B)** The CTF confidence of the calculated CTF of each micrograph.

Similar to the FEI Polara data, the helical segments from the two FEI Titan Krios datasets were further sorted based on their diameter to reduce structural heterogeneity. The diameter size distribution of the helical segments from the CynD_{pum} variant data showed that a majority (26284 segments) of fibres were between 113–129 Å in diameter (Fig. 2.15) while the wild-type CynD_{pum} data diameter size distribution showed that a majority (26284 segments) of fibres were between 112–134 Å in diameter (Fig. 2.16), and these segments were used in the following IHRSR runs. The majority of helical segments from the CynD$_{pum}$ variant FEI Titan Krios data and variant FEI Polara data had around the same diameter range while the majority of helical segments from the wild-type CynD$_{pum}$ Titan Krios data were within a broader diameter range, thus demonstrating slightly more structural heterogeneity for the wild-type fibres.

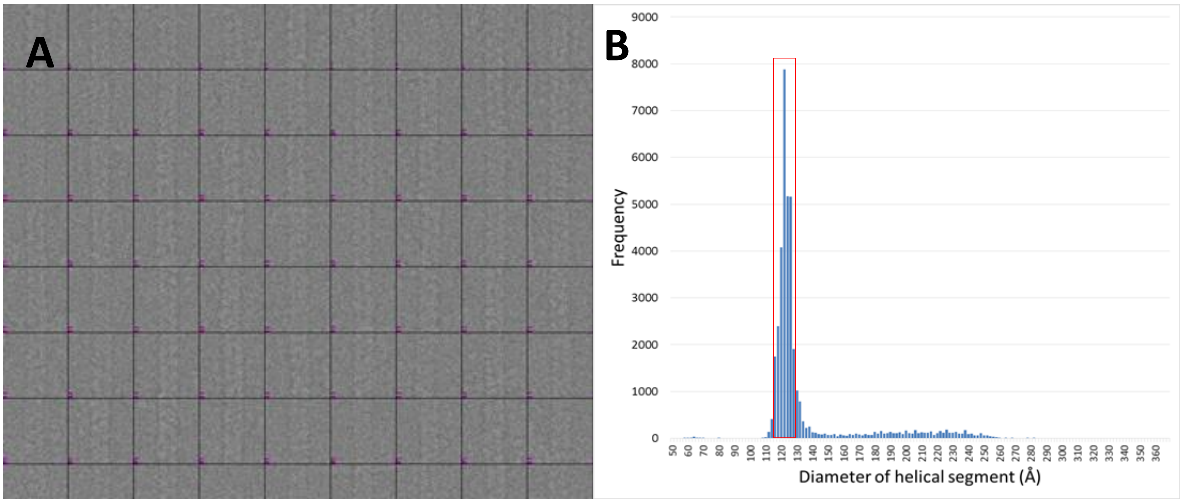


Figure 2.15 (A) Image stack of the FEI Titan Krios Q86R+H305K+H308K+H323K CynD$_{pum}$ variant at pH 5.4 which consists of centred and diameter classified helical segments viewed using the V2 program (Ludtke et al., 1999). The stack was produced by cutting selected helical fibres from cryo-electron micrographs into 448 X 448 pixel segments. **(B)** Diameter size distribution of the helical segments and (in red box) selected fibre diameter (113- 129 Å) of particles used in IHRSR.

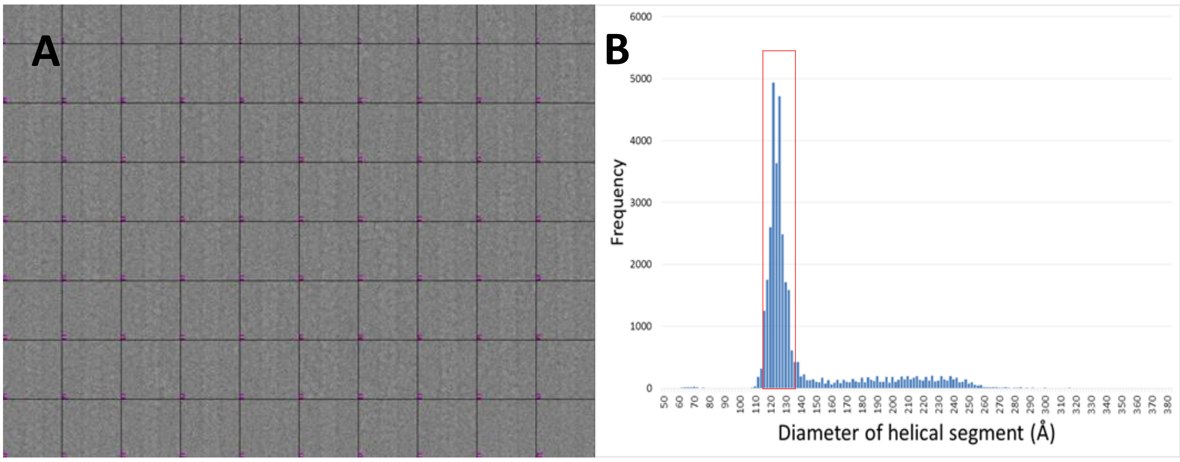


Figure 2.16 (A) Image stack of the FEI Titan Krios wild-type CynD$_{pum}$ at pH 5.4 which consists of centred and diameter classified helical segments viewed using the V2 program (Ludtke et al., 1999). The stack was produced by cutting selected helical fibres from cryo-electron micrographs into 448 X 448 pixel segments. **(B)** Diameter size distribution of the helical segments and (in red box) selected fibre diameter (112- 134 Å) of particles used in IHRSR.

Unlike the FEI Polara data no additional out-of-plane tilt optimisation step was required for the Q86R+H305K+H308K+H323K CynD$_{pum}$ FEI Titan Krios dataset due to the minimal out-of-plane tilt observed during the IHRSR (Fig. 2.17). The absence of major out-of-plane tilt was demonstrated by the absence of blurring of diffraction spots (which was demonstrated by the FEI Polara data) when the average power spectrum of all the helical segments was compared to the average power spectrum of the 0° tilt helical segments (Fig. 2.18), thus showing that most of the helical segment were within the same out-of-plane tilt angle. The minimal out-of-plane tilt was also demonstrated by the absence of high frequency cluster of the number of helical segment images assigned to reference projections at user-defined upper and lower limit angles (-20 and 20°) during the multi-reference alignment (Fig. 2.17). Interestingly the wild-type FEI Titan Krios dataset demonstrated this high frequency cluster when using the same user-defined limits (-20 and 20°)(Fig. 2.19A), although no blurring of diffraction spots was observed when the tilted and non-tilted average power spectra were compared (Fig. 2.20). The user-defined out-of-plane limits for the wild-type FEI Titan Krios dataset were then set to a higher angle (-32 and 32) to reduce the high frequency cluster at these limits (Fig. 2.19B), but this unexpectedly did not improve the quality of the resultant 3D reconstruction (results not shown).

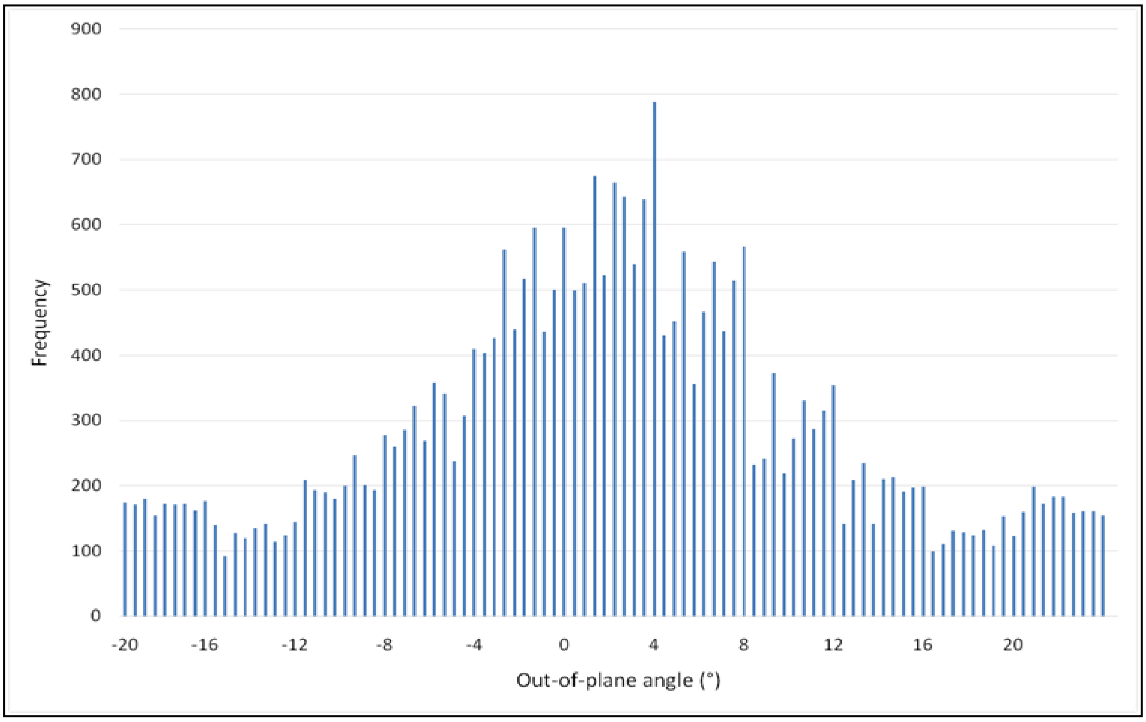


Figure 2.17. Frequency distribution of the reference projections for the Q86R+H305K+H308K+H323K FEI Titan Krios data generated from the multi-reference alignment during IHRSR. An out-of-plane tilt angle of 20° was used.

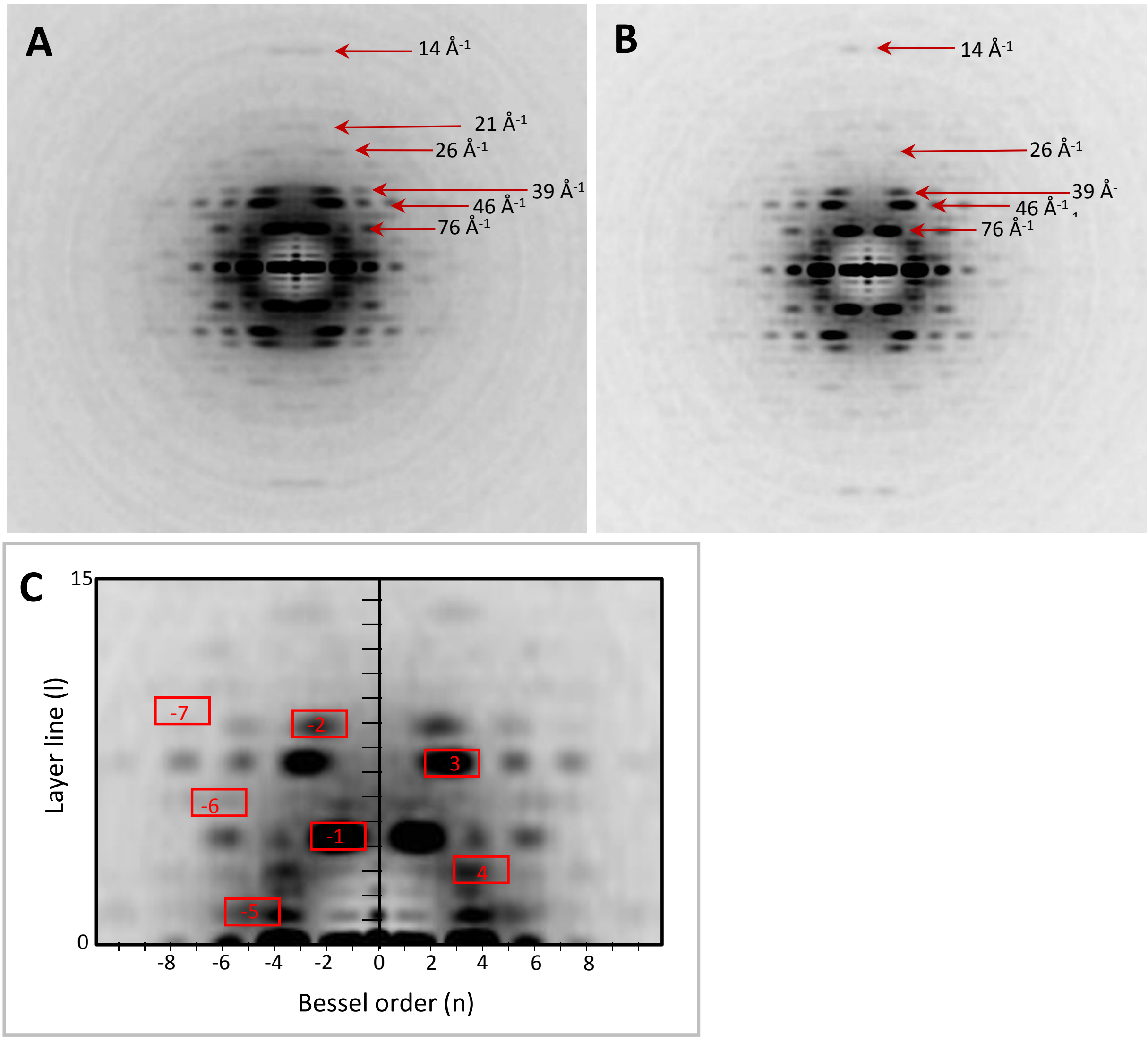


Figure 2.18. **(A)** Average power spectrum generated from a full dataset of the helical segments from the Q86R+H305K+H308K+H323K CynD$_{\text{pum}}$ variant collected on the FEI Titan Krios microscope. **(B)** Average power spectrum generated from a subset of the helical segments identified as not having out-of-plane tilt during the multi-reference alignment. Positions of layer lines with visible diffraction spots are indicated by the red arrows. **(C)** Indexing of the power spectrum generated from the 0° out-of-plane tilt subset of the helical segments. The positions of the diffraction spots corresponding to the calculated Bessel orders are highlighted in red.

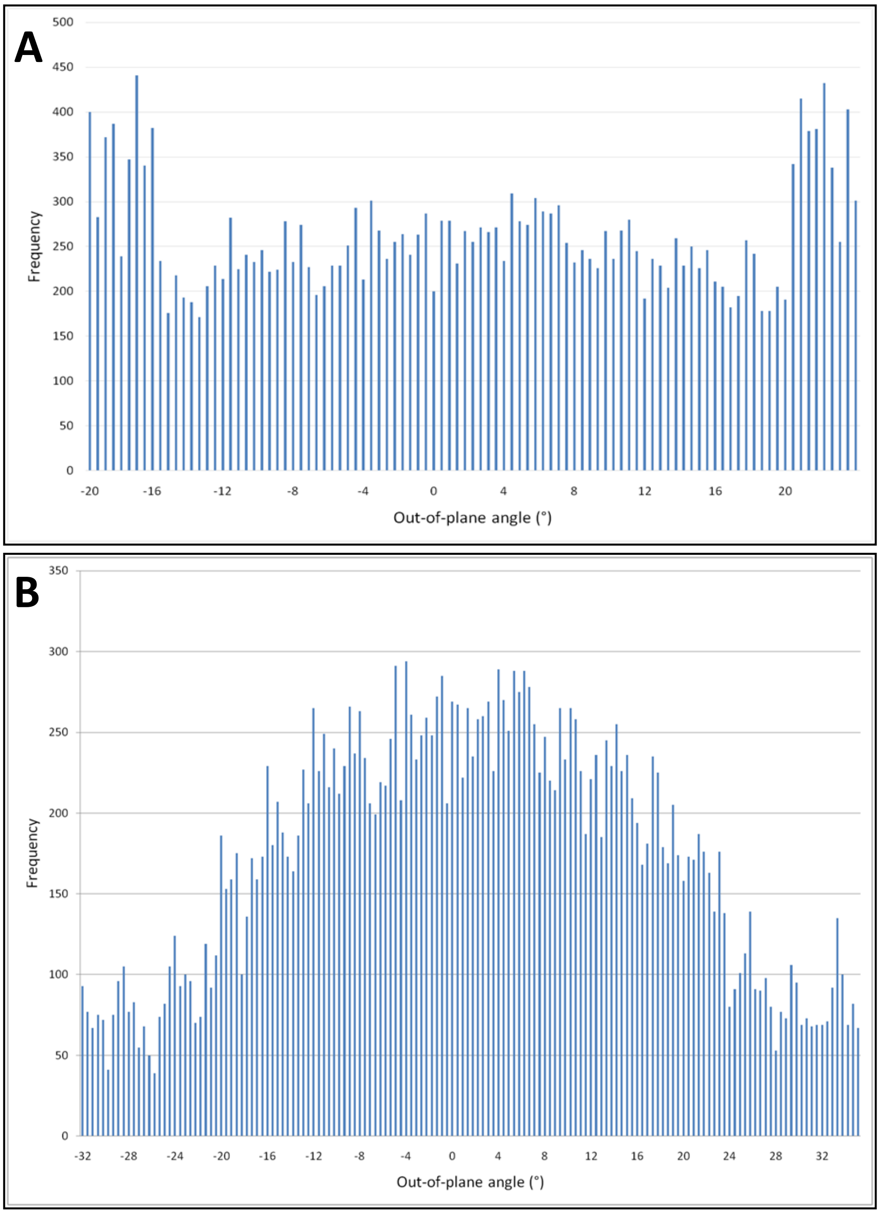


Figure 2.19. Frequency distribution of the reference projections for the wild-type CynD$_{pum}$ FEI Titan Krios data generated from the multi-reference alignment during IHRSR. Out-of-plane tilt angles of **(A)** 20° and **(B)** 32° were used.

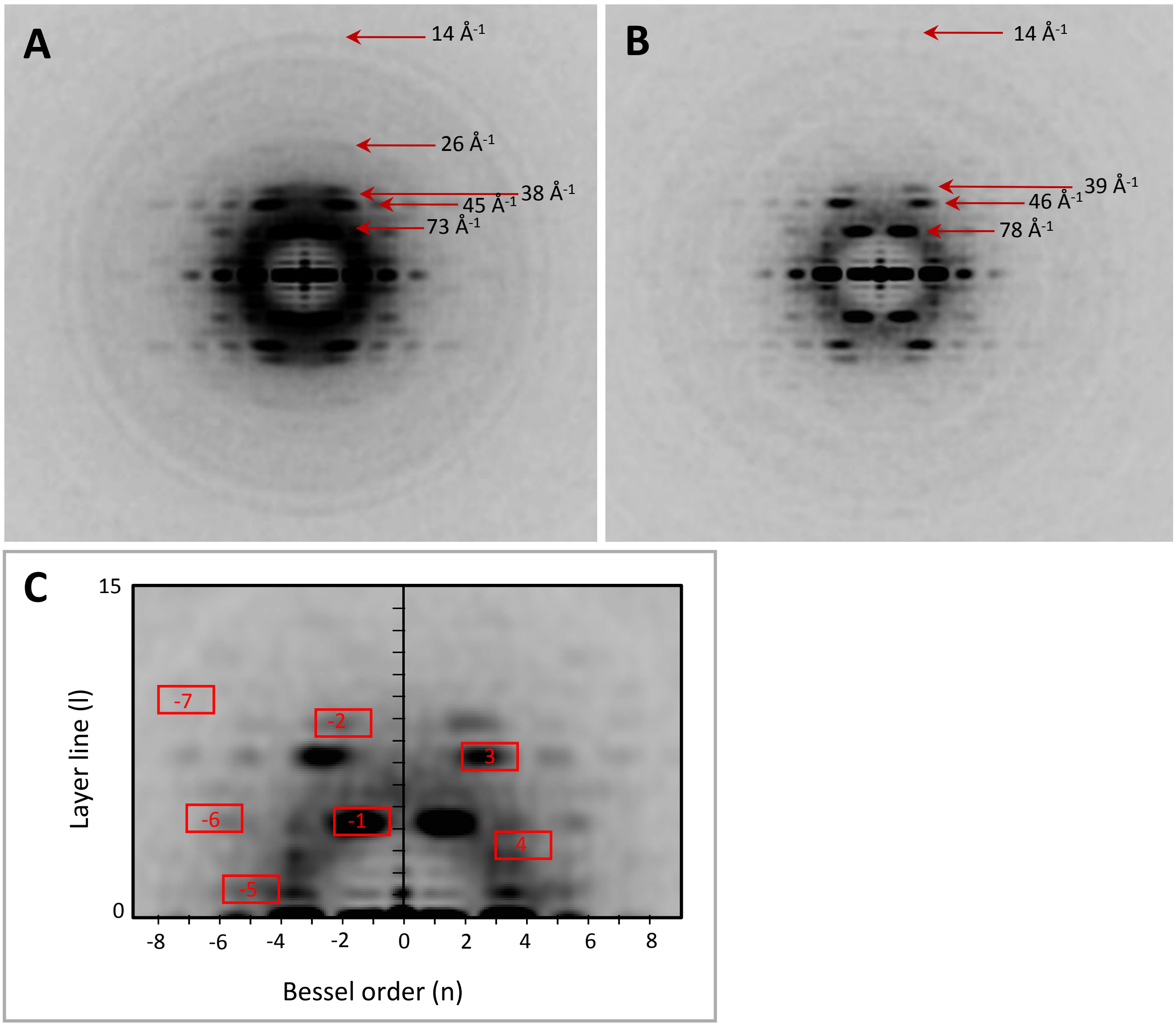


Figure 2.20. **(A)** Average power spectrum generated from a full dataset of the helical segments from the wild-type CynD$_{\text{pum}}$ collected on the FEI Polara microscope. **(B)** Average power spectrum generated from a subset of the helical segments identified as not having out-of-plane tilt during the multi-reference alignment. Positions of layer lines with visible diffraction spots are indicated by the red arrows. **(C)** Indexing of the power spectrum generated from the 0° out-of-plane tilt subset of the helical segments. The positions of the diffraction spots corresponding to the calculated Bessel orders are highlighted in red.

IHRSR was run for 15 cycles for both the helical segments from the variant and wild-type CynD$_{\text{pum}}$ FEI Titan Krios data. By 15 cycles no further changes in the helical symmetry were observed for both datasets. There was no major difference in the final converged helical symmetry values between the variant and the wild-type. The final IHRSR run converged at a helical symmetry solution with a twist of ($\Delta\varphi$) -77.0° and rise (Δz) of 16.7 Å for the variant and a twist of ($\Delta\varphi$) -77.2° and rise (Δz) of 17.0 Å for the wild-type. These helical symmetry solutions were verified when IHRSR was run using two half image datasets and two different sets of arbitrary user-defined initial helical parameters, when calculating the resolution of the final 3D EM models (Fig. 2.21 and Fig. 2.2).

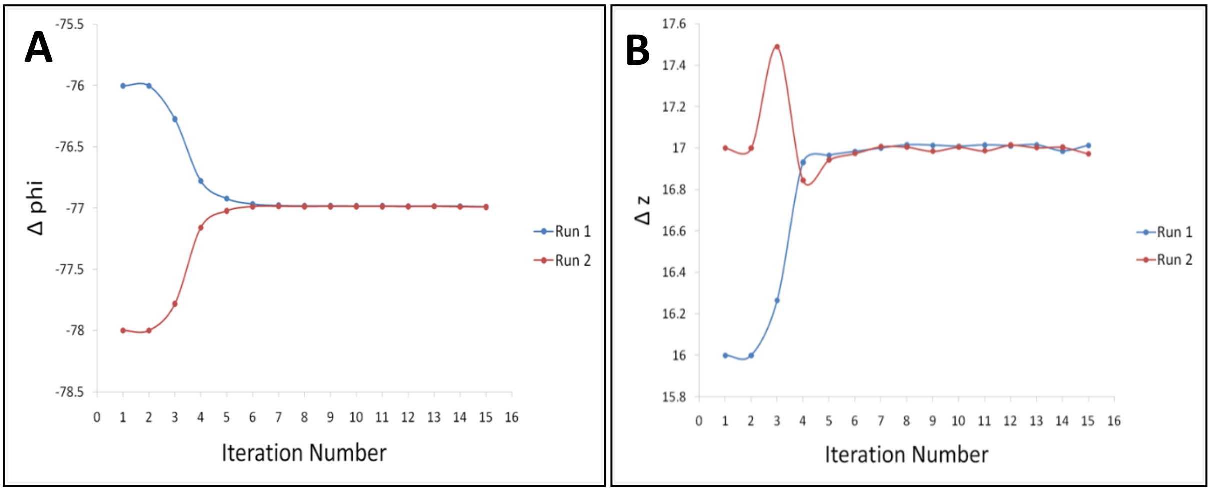


Figure 2.21. Helical symmetry solution after 15 IHRSR iterations of the FEI Titan Krios cryo-EM images of the purified Q86R+H305K+H308K+H323K CynD$_{pum}$ variant. Two IHRSR runs were carried out using two half image datasets and two different sets of arbitrary user-defined initial helical parameters. The IHRSR run converged at a **(A)** Δphi of -77.0° and **(B)** Δz of 16.7 Å.

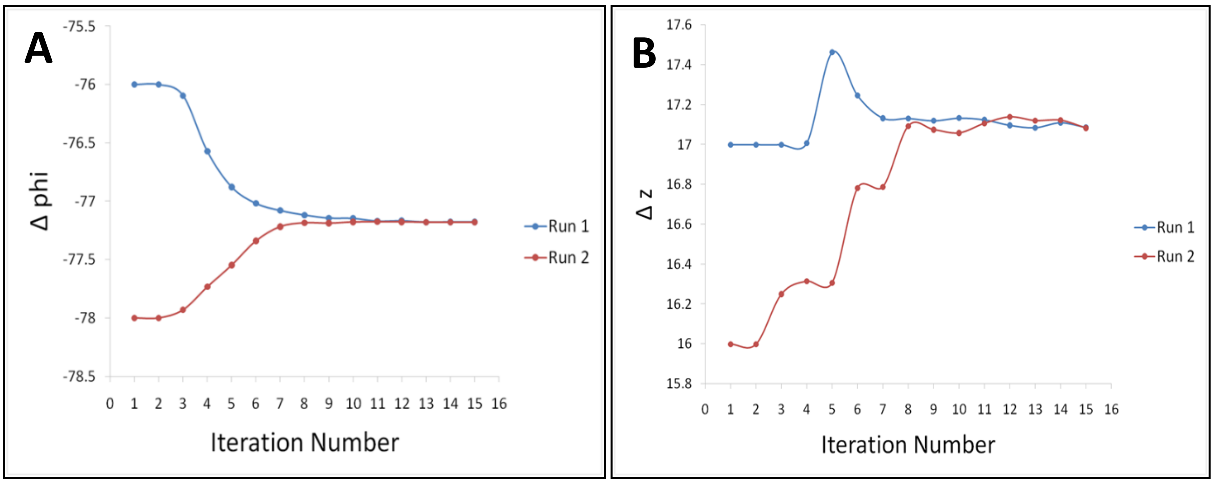


Figure 2.22. Helical symmetry solution after 15 IHRSR iterations of the FEI Titan Krios cryo-EM images of the purified wild-type CynD$_{pum}$ variant. Two IHRSR runs were carried out using two half image datasets and two different sets of arbitrary user-defined initial helical parameters. The IHRSR run converged at a (**A**) Δphi of -77.2° and (**B**) Δz of 17.0 Å.

We obtained detailed models from the IHRSR for both the Q86R+H305K+H308K+H323K CynD$_{pum}$ variant (Fig. 2.23A) and the wild-type CynD$_{pum}$ (Fig. 2.24A) which we then calculated their resolution using the Fourier Shell Correlation (FSC) method (Yang et al., 2003). Using the FSC cut-off of 0.5 the resolution of the final models was determined to be 6.8 Å (Fig. 2.23B) for the variant and 7.4 Å for the wild-type (Fig. 2.24B). Using the same FSC plot the resolution of the final models were also determined to be 5.0 Å and 4.9 Å for the variant and wild-type respectively, using the FSC cut-off of 0.143.

These final models were further sharpened by applying a B-factor to the maps using the EMBFACTOR (Fernandez et al., 2008) program. EMBFACTOR calculated and applied a B-factor of -263.57 Å^2 and -195.88 Å^2 for the Q86R+H305K+H308K+H323K variant (Fig. 2.23C) and the wild-type (Fig. 2.24C) model respectively (using the resolution range of 5-12 Å) which were used to scale amplitudes. Like with the FEI Polara data a Guinier plot was produced by EM-BFACTOR after sharpening showing the amplification of the decaying spherically average amplitude of both the experimental and noise weighted (Cref) amplitudes after the application of the calculated B-factor during the sharpening (Fig. 2.23D and Fig. 2.24D).

A previously generated CynD$_{pum}$ atomic homology model (Park et al., 2016) was then docked in to the sharpened models, to help validate the models. The homology model showed a high correlation to the sharpened models and most of the peptide backbone fitted perfectly in the EM densities (Fig. 2.23E–F and Fig. 2.24E–F). The high resolution of final sharpened models was demonstrated by not only the appearance of individual peptide alpha helices but also the appearance ridges within the individual peptide alpha helices corresponding to the helical turns of the peptide backbone (Fig. 2.23D and Fig. 2.24D).

Indexing of the diffraction spots from the average power spectrum of the wild-type and the Q86R+H305K+H308K+H323K CynD$_{pum}$, generated from the 0° out-of-plane tilt subsets of the helical segments, was carried out similarly to the FEI Polara data (Fig. 2.18C and Fig. 2.20C). The indexing of the power spectra from the Titan Krios data showed similar diffraction patterns between the wild-type and the Q86R+H305K+H308K+H323K CynD$_{pum}$, with both enzymes demonstrating a strong diffraction spot at the layer line of ~75 Å, indexed as a Bessel function of order -1 and corresponding to the pitch of the helix. The similarity between the diffraction patterns of the two variants demonstrates that the introduced Q86R+H305K+H308K+H323K mutations did not confer major differences to the structural positioning of the CynD$_{pum}$ subunits in helical fibres. The indexing of the power spectra and the IHRSR calculated helical symmetry parameters both confirmed that both the wild-type and Q86R+H305K+H308K+H323K CynD$_{pum}$ variant exist as left handed helices with a pitch of ~75 Å and ~4.6 subunits per turn.

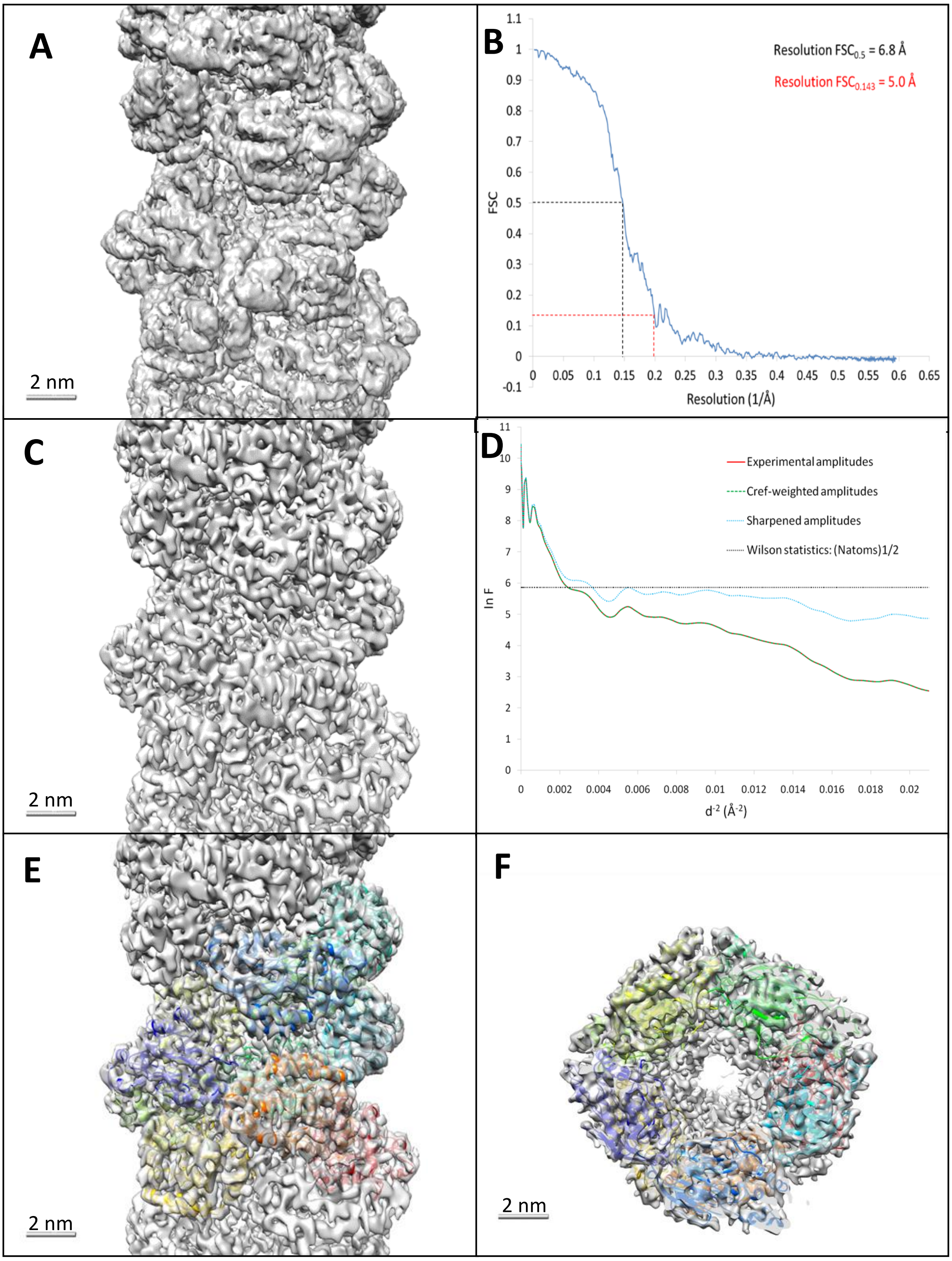


Figure 2.23. (A) The final 3D reconstruction produced from the FEI Titan Krios cryo-EM images of the purified Q86R+H305K+H308K+H323K CynD$_{pum}$ variant. **(B)** The Fourier Shell Correlation (FSC) curve of the final 3D reconstruction used to calculate the resolution of the model. **(C)** The EM-BFACTOR sharpened version of the same final 3D reconstruction. **(D)** The Guinier plot produced by EM-BFACTOR after sharpening of the final 3D reconstruction. **(E)** The docking of a CynD$_{pum}$ homology model (Park et al., 2016) to the sharpened version of the final 3D reconstruction. **(F)** The same docked model viewed at a 90° rotation.

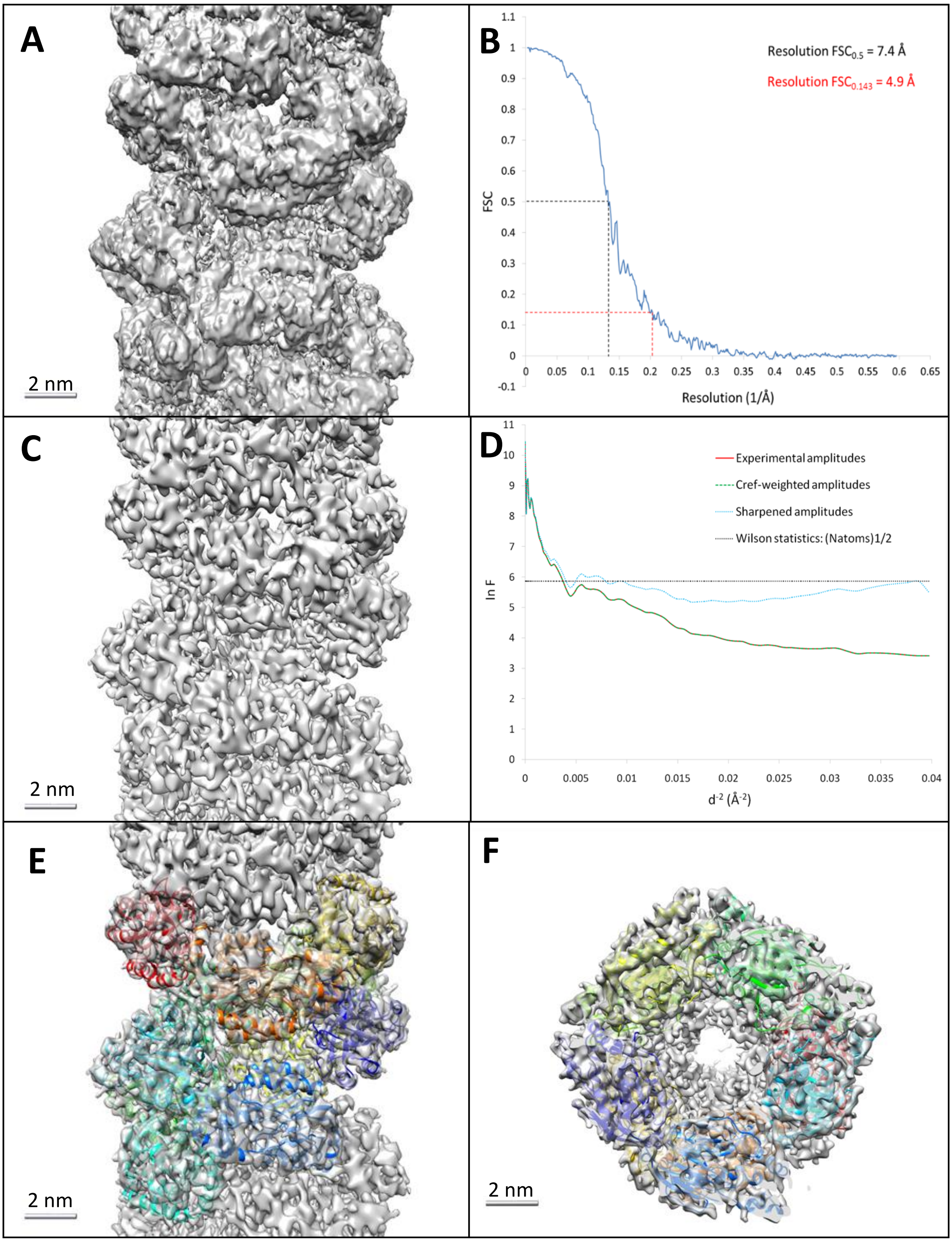


Figure 2.24. **(A)** The final 3D reconstruction produced from the FEI Titan Krios cryo-EM images of the purified wild-type CynD$_{pum}$ variants. **(B)** The Fourier Shell Correlation (FSC) curve of the final 3D reconstruction used to calculate the resolution of the model. **(C)** The EM-BFACTOR sharpened version of the same final 3D reconstruction. **(D)** The Guinier plot produced by EM-BFACTOR after sharpening of the final 3D reconstruction. **(E)** The docking of a CynD$_{pum}$ homology model (Park et al., 2016) to the sharpened version of the final 3D reconstruction. **(F)** The same docked model viewed at a 90° rotation.

To further help visually estimate the resolution of our final 3D reconstructions we constructed various simulated 3D models from a previous CynD$_{pum}$ homology model (Park et al., 2016) at different resolutions (Fig. 2.25). Using these resolution templates, we estimated the resolution of the Q86R+H305K+H308K+H323K CynD$_{pum}$ FEI Polara and Titan Krios 3D reconstruction to be at about 8 Å and 5.5 Å respectively. The FEI Titan Krios wild-type 3D reconstruction was visually estimated to be about 5.9 Å.

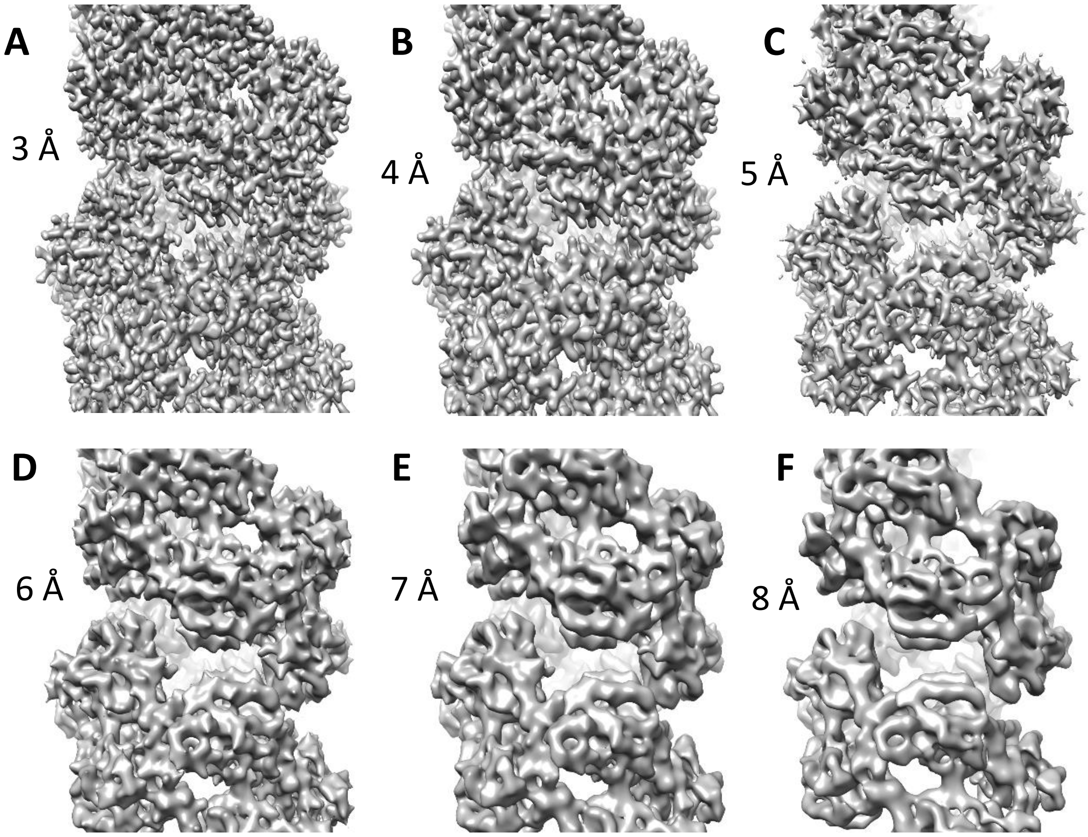


Figure 2.25. Simulated EM maps calculated and generated from a CynD$_{pum}$ atomic homology model (Park et al., 2016) at resolutions of 3 Å (**A**), 4 Å (**B**), 5 Å (**C**), 6 Å (**D**),7 Å (**E**) and 8 Å (**F**).

Using UCSF Chimera's local correlation function, the final models produced from the Q86R+H305K+H308K+H323K and wild-type CynD$_{pum}$ FEI Titan Krios final 3D models were aligned and local correlation was measured in order to assess the structural variations between the two models. Using the resultant variation map, it was observed that there was less variation around the core than the outer surface of each monomer of the two variants and a high level of variation in the inner surface of the helical fibre (Fig. 2.26).

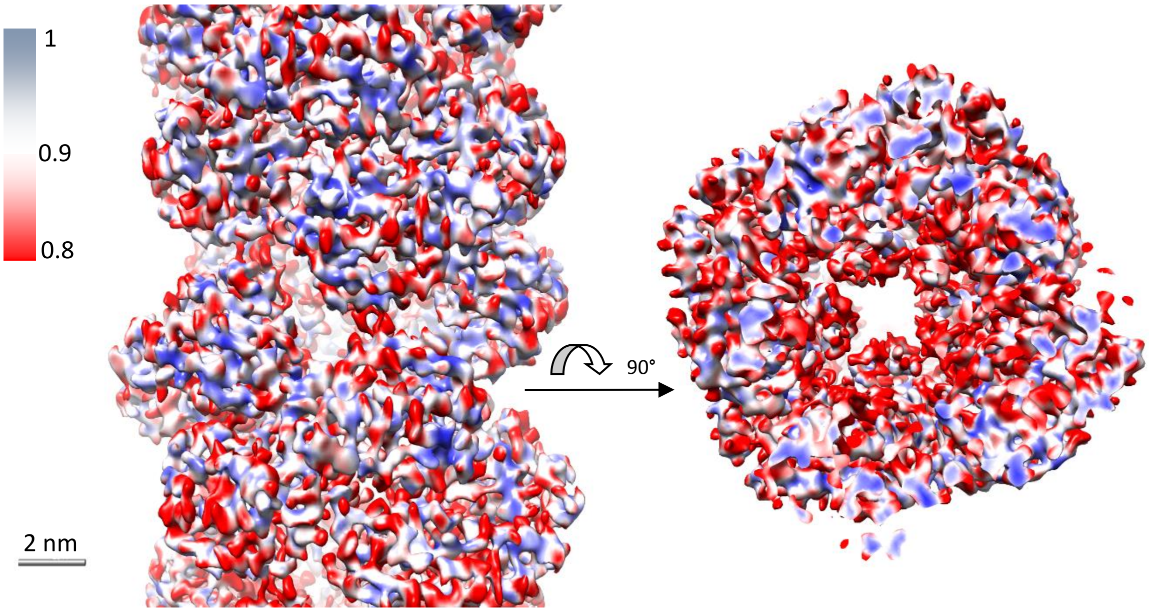


Figure 2.26. Variance map of the wild-type and Q86R+H305K+H308K+H323K CynD$_{pum}$ 3D reconstructions obtained from the FEI Titan Krios datasets. The variability of density is colour coded from blue indicating the highest (local correlation=1) to red indicating the lowest (local correlation=0.8).

2.7 Discussion

2.7.1 Purification

The CynD_{pum} wild-type and Q86R+H305K+H308K+H323K variant were successfully purified and confirmed by the presence of activity in the gel filtration peak fractions. SDS-PAGE gel also confirmed the purity of the enzyme preparations with the observed 37 kDa band which corresponds to the size of a CynD_{pum} monomer. At pH 5.4 and pH 8 the size of the fibres formed by the Q86R+H305K+H308K+H323K variant were significantly longer than those previously observed for the wild-type but were of similar size to those previously observed for the separate Q86R and H305K+H308K+H323K variants (Mulelu, 2013). This may suggest that the combination of the Q86R and H305K+H308K+H323K variants may have not resulted in a very high increase in stability compared to the separate Q86R and H305K+H308K+H323K variants, but still resulted in a variant more stable than the wild-type CynD_{pum}.

2.7.2 Data Collection

Previous attempts to obtain a high resolution cryo-EM structures by our group have been unsuccessful. We previously only had access to old generation microscope such as the Tecnai F20 equipped with a CCD camera or use of film. Access to more powerful microscope such as the FEI Polara and Titan Krios, both equipped with a K2 Summit direct electron detector, has greatly improved the quality of our cryo-EM reconstruction structures. The use of these powerful microscopes has demonstrated that the major hindrance that previously prevented us from acquiring high resolution structure was the use outdated CCD detectors as well as the poor stability of the old microscopes. The major advantage that the direct electron detectors have over the CCD or film is the ability to track specimen movement during irradiation thus reducing image blurring. This was significantly demonstrated when an initial dataset of the Q86R+H305K+H308K+H323K CynD_{pum} variant from a different study was collected using the FEI Falcon II set to single exposer mode, with no tracking of specimen movement. The resulted 3D reconstruction from this initial dataset was of low resolution (results not shown) and the following datasets collected on the K2 with tracking of specimen movement produced reconstructions of significantly higher resolution, thus demonstration the advantageous effect of specimen movement tracking during irradiation.

The models produced by the FEI Titan Krios microscope where significantly more detailed than the model obtained using the FEI Polara microscope. This improvement in the quality of the models does not necessarily demonstrate the FEI Titan Krios superiority over the FEI Polara microscope as there were other factors that contributed to the improvement of the models generated from the data from the FEI Titan Krios microscope. The first contributor to the improvement in the quality of data collected on the the FEI Titan Krios microscope was the optimisation of the cryo-grid preparation using the knowledge gained after processing the FEI Polara data. For the FEI Polara data collection, cryo-grids with an overlaid film of carbon were initially used with the aim of providing the helical filaments a flat support to ensure that all the filaments overlaid horizontally on the cryo-grid with minimal tilt. These was shown to be unsuccessful as a large out-of-plane angle was required to ensure a flat and even frequency distribution of the reference projections during the multi-reference alignment (Fig. 2.8). With this insight the next set of cryo-grids used were overlaid with a holey carbon film and these were loaded in the FEI Titan Krios. For the Q86R+H305K+H308K+H323K FEI Titan Krios data the use of holey cryo-grids showed a reduction in the amount of tilt of the helical fibres and this may due to the restricted tilt movement by the extra thin layer of ice at the carbon holes in which the filaments were embedded as opposed to the flexible carbon film present in the cryo-grids used on the FEI Polara microscope. The holey carbon cryo-grids used when collecting the wild-type FEI Titan Krios data showed the same high out-of-plane tilt as the FEI Polara non-holey carbon grid data but unlike the FEI Polara data the adjustment of the degree of tilt did not improve the quality of the final 3D reconstruction. This suggest that the observed out-of-plane tilt on the wild-type FEI Titan Krios data may be due to poor multi-reference alignment due to the higher structural variability of the wild-type CynD$_{pum}$.

Although the holey cryo-grids reduced the tilt movement of the filaments, the lack of an overlaid carbon film reduced the quality of the thong rings produced on the power spectra of the micrographs (Fig. 2.5 and 2.12), thus slightly reducing the accuracy of the CTF calculations. This was evident when comparing the percentage confidence of the CTF calculation from the micrographs recorded from cryo-grids with a holey (Fig. 2.13B and Fig. 2.14B) and non-holey (Fig. 2.6) carbon film, were the percentage confidence of the CTF calculations from the holey grids data were marginally lower than those calculated from the non-holey grids data. This slight reduction in the accuracy of the CTF calculation did not significantly affect the quality of the 3D reconstruction as the models produced from the holey grids data were of higher resolution than the one from the non-holey grids. Another factor that contributed to the resolution improvement of the models generated from the data from the FEI Titan Krios microscope over the model generated by the data from the FEI Polara was the use of a higher magnification on the FEI Titan Krios microscope. The need to use a higher magnification during data collection was first evident on the FSC plot generated for the 3D model generated from the FEI Polara image dataset (Fig. 2.11B). This FSC plot was slightly unhealthy in that it lacked the steepness generally displayed by healthy FSC plots. This less steep FSC curve demonstrated that some of the structural information is located near the nyquist limit of the data and this is due to a limit in the amount of information that can be obtained from the data as a consequence of the low magnification. Using this information, the magnification was greatly increased when micrographs were later collected on the FEI Titan Krios microscope, although this significantly reduced the field of view hence reducing the number filament per micrographs (Fig. 2.5 and 2.12), a disadvantage that was compensated by increasing the total number of micrographs collected (Table 2.3 and 2.4). The FSC plots of the 3D models from the high magnification data showed a steeper and healthier curve and demonstrated that minimal information is located near the nyquist limit thus demonstrating that less high frequency information was lost from the original data during the processing of the 3D models (Fig. 2.23B and Fig.2.24).

2.7.3 Data processing

During CTF correction two strategies were tested, one where a complete CTF amplitude and phase correction was carried out, and one where only phase-flipping was carried out. During the CTF correction of the FEI Polara micrographs it was apparent that the phase-flipping approach resulted in a better model than the complete CTF correction approach (results not shown). This is speculated to be due to the observed retention of more contrast after phase-flipping than after a complete CTF correction. This contrast is later essential in accurately calculating cross correlations between reference projections and segmented helical segments during the multi-reference alignment step of IHRSR, thus the loss of contrast may consequentially lead to misalignments which result in degradation of the quality of the 3D reconstruction. Interestingly there were no major differences in the quality of the final models produced using either CTF correction methods for the FEI Titan Krios datasets and no distinguishable loss of contrast were observed between images obtained after CTF correction by the two different approaches (results not shown). When comparing the helical segments from the FEI Polara (Fig. 2.7) and the segments from the FEI Titan Krios data (Fig. 2.15A and Fig. 2.16A) it was evident that the FEI Polara segment possessed more contrast after CTF correction. This higher level of contrast in the helical segments from the FEI Polara data may have improved the results of the multi-reference alignment during IHRSR and may explain the acquisition of a very detailed 3D reconstruction despite the use of a very low number of helical segments (Table 2.3).

Before IHRSR was run a number of pre-processing steps were carried out to ensure that the IHRSR was run using the best helical segment obtainable from the EM data. The first step in accomplishing this was to discard micrographs that showed high levels of drift and astigmatism. During this discarding, more micrographs were removed from the FEI Polara datasets than the FEI Titan Krios data (Table 2.3 and 2.4) demonstrating the higher quality of the images collected on the Titan Krios microscope which was due to an improved cryo-grid preparation and an improved data collection technique. The same trend was observed when comparing the number of helical segments discarded after segmentation of the picked helical filaments from the two microscope's datasets (Table 2.3 and 2.4). After the discarding of bad micrographs, picking of helical filaments and segmentation of the filaments was carried out. To reduce structural variability between helical segments, picking of helical filaments was carried out while avoiding broken, overlapping, bent and terminal sections of the filaments (Fig. 2.12). Structural variability was also minimised by the selection of segmented helical fibres within a certain diameter range. Both the FEI Polara and Titan Krios datasets for the Q86R+H305K+H308K+H323K CynD$_{\text{pum}}$ varaint showed that the majority of their helical segment diameters were within the same range (Fig. 2.7A and Fig. 2.15A). The majority of the wild-type CynD$_{\text{pum}}$ helical segment's diameters were within a broader range than that of the CynD$_{\text{pum}}$ variant (Fig. 2.16A) and this may be due to the previously demonstrated flexibility of the wild-type CynD$_{\text{pum}}$ which lead to the creation of the more stable Q86R+H305K+H308K+H323K CynD$_{\text{pum}}$ variant.

To cut out the long picked helical filaments into smaller boxes segmentation was carried out. One of the important parameter to consider during segmentation is the size of the boxes to be cut out. In order to obtain an optimal box size, the box should not be too long as to include structural variations from filament bending and internal disorder, which may introduce misalignment during the IHRSR. The box size should also not be too short that the cross correlation would be ambiguous during the multi-reference alignment in the IHRSR. The optimal box size was calculated and optimized empirically for our structure. The optimal box size of our CynD$_{\text{pum}}$ data was determined to be about 366 Å X 366 Å as smaller box sizes (293 Å X 293 Å) resulted in 3D reconstruction of low quality (results not shown). Another parameter to be considered during segmentation is the percentage overlap of consecutive segments. Unlike the box size this parameter was not determined empirically but rather calculated from prior knowledge of the helical rise per subunit of the CynD_{pum} (Mulelu, 2013). An overlap step size of 23 Å was determined as it would increase the number of segment images while preventing ambiguous cross correlations during the multi-reference alignment in the IHRSR.

IHRSR was run using helical segments from the datasets and the process was iterated until a stable helical symmetry was obtained. No major differences in the helical symmetry were observed for all the final reconstruction including the wild-type dataset. It was previously proposed that the cause of the termination of the elongation of helical fibres of the CynD_{pum} was due to the distortion in the helical symmetry which results in the collapse of the terminal subunits thus preventing the further addition of subunit (Sewell et al., 2003). This was not the case as both the variant and the wild-type displayed similar helical symmetries although having different oligomerisation forms at the same pH's, thus suggesting that other factors besides the distortion of the helical symmetry may be responsible for the pH dependent oligomerisation of CynD_{pum}.

2.7.4 Structure evaluation and comparison

After IHRSR the resolution of the final 3D reconstructions was evaluated using the FSC method. Determining the resolution of helical structures using the FSC method was shown to raise more challenges than the resolution determination of single particle reconstructions. These challenges include separating the full 2D image dataset into two independent subsets. Due to the nature of helical structures and the extraction of overlapped images during segmentation, it is difficult to produce two truly independent subsets as asymmetric units from the same filament are duplicated and may be classified in the same subset. As a result, two semi-independent subsets were used when calculating the resolution and this might have given slightly inaccurate result. To help us validate the resolution accurately we constructed various simulated 3D models from a previous homology model at different resolutions and used these to visually estimate the resolution of our final models (Fig. 2.25). Using these resolution templates we observed that the $\text{FSC}_{0.5}$ resolution calculation of the Q86R+H305K+H308K+H323K CynD_{pum} FEI Polara data was understated (11.1 Å) as the resolution templates suggested that the resolution was about 8 Å, an estimate which agreed with the corresponding $\text{FSC}_{0.143}$ resolution calculation of the same structure (7.6 Å)(Fig. 2.11B). A similar result was observed for the FEI Titan Krios datasets were in both cases the

$FSC_{0.5}$ resolution calculations were understated (6.8 Å for the variant and 7.4 Å for the wild-type) when compared to the visually estimated resolutions (5.5 Å for both the variant and wild-type) but unlike the FEI Polara structure the $FSC_{0.143}$ resolution calculation of the FEI Titan Krios structures were overstated (5.0 Å and 4.9 Å) as the resolution templates estimated lower resolutions (Fig. 2.23B and Fig. 2.24B). These variable resolution results demonstrate the unreliability of using one value to determine the resolution of a structure. This phenomenon is common in cryo-EM reconstructions and the variable resolution results have been proposed to be due to the specimen heterogeneity which translates into the reconstructions resulting in distinct regions showing variable levels of details (Cardone et al., 2013).

The final models produced from the Q86R+H305K+H308K+H323K and wild-type $CynD_{pum}$ FEI Titan Krios were aligned and variations between the two models were assessed (Fig. 2.26). From the variation map produced from these two models it was observed that there was less variation around the core than the outer surface of each monomer demonstrating that both the variant and wild-type $CynD_{pum}$ may possess similar interactions that stabilise the core of the monomers. A high level of variation was observed in the inner surface of the helical fibre, an area previously proposed to be the location of the C-terminal tail (Sewell et al., 2003; Thuku et al., 2007). The high variation may be due to the high structural variation between the C-terminal tail of the variant and the wild-type as a consequence of introducing the H305K+H308K+H323K mutations which are located at the C-terminal tail. The H305K+H308K+H323K mutations may be conferring stability to the C-terminal tail of the variant thus displaying high variations when compared to the more flexible C-terminal tail of the wild-type $CynD_{pum}$. The higher degree of flexibility of the wild-type's C-terminal tail is also demonstrated by the reduction of defined density around the C-terminal tail region.

2.8 Conclusion

The Q86R+H305K+H308K+H323K CynD$_{pum}$ variant was successfully constructed and it demonstrated more stability than the wild-type CynD$_{pum}$. Cryo-EM image datasets were collected on two different microscopes, the FEI Titan Krios and FEI Polara using different sets of conditions. After image processing it was demonstrated how these conditions such as the use of grids with overlaid holey carbon film and use of high magnification can help improve the quality of 3D reconstructions.

IHRSR was carried-out on using the cryo-EM image datasets and three high resolution 3D structures were generated. The 3D models include the FEI Polara Q86R+H305K+H308K+H323K CynD$_{pum}$ structure at a resolution of 11.1 Å, the FEI Titan Krios Q86R+H305K+H308K+H323K CynD$_{pum}$ structure at a resolution of 6.8 Å and the wild-type CynD$_{pum}$ structure at a resolution of 7.4 Å. The stated resolutions were obtained using the FSC$_{0.5}$ threshold but we have presented evidence that this threshold may greatly understate the resolution. Nevertheless, the structures were still the highest resolution CynD$_{pum}$ structures ever obtained.

The comparison between the high resolution wild-type and CynD$_{pum}$ variant structures demonstrated that there are no major changes in the CynD$_{pum}$ helical symmetry after introducing the stability conferring mutation, despite there being differences in the pH dependent oligomerisation. This demonstrated that other factors may be involved in oligomerisation of CynD$_{pum}$ other than the distortion of the helical symmetry. The comparison between the wild-type and CynD$_{pum}$ variant structures also showed that there is a high level of variation around the postulated C-terminal region and this variation is proposed to be due to the wild-type's more flexible C-terminal tail.

CHAPTER III

CRYO-EM MAP MODEL FITTING AND REFINEMENT

3.1 Introduction

Due to the absence of a CynD_{pum} crystal structure, methods such as homology modelling had to be implemented in order to generate a structural prediction of the enzyme. Over the past years, homology modelling of CynD_{pum} structural models has been carried out and further improvement of these models was done through a number of refinement steps which were based on experimental insights obtained about the CynD_{pum} enzymes and other related nitrilase structures. Until recently the best model of the CynD_{pum} was generated through homology modelling using six X-ray crystal structures of related but non-branch 1 nitrilases as templates and further refinement was carried out using constrained molecular dynamics simulations (Mulelu, 2013). The recent publication of the crystal structure of the branch 1 Nit6803 nitrilase from *Synechocystis* sp. Strain PCC6803 (Zhang et al., 2014) has provided a number of structural insights about the branch 1 nitrilases. This Nit6803 structure has also enabled the generation of a more accurate CynD_{pum} homology model which coupled with mutagenesis experiments has provided insights about the structure of CynD_{pum} (Park et al., 2016). In this chapter we describe the refinement of this homology model using the high resolution 3D structures obtained in chapter two and the structural insights obtained from these refined models. We further interpret the refined CynD_{pum} models by correlating these models to previous mutagenesis experimental results, thus also validating the quality of the refined models.

3.2 Review of near-atomic resolution cryo-EM map model fitting

3.2.1 Cryo-EM map model fitting

A cryo-EM map alone is limited in explaining the mechanisms of biochemical functions but an accompanying atomic model derived from the same map can significantly increase the information content of that cryo-EM map (Zhou, 2011). Although there has been an improvement in the resolutions that can be achieved by cryo-EM techniques, there is still a lack of tools for atomic model building and refinement using cryo-EM maps of near-atomic resolutions. The use of current X-ray crystallography tools has proven challenging as they fail to converge on an accurate atomic model when using cryo-EM maps at near-atomic or moderate resolutions ~4 Å resolutions (Zhou, 2011; DiMaio et al., 2015). At this resolution the maps contain regions with branched densities and only a few side chain densities which make even backbone tracing difficult. In such cases it is favourable to first construct a homology model and refine this model based on the cryo-EM map. The refinement of the atomic model can then be achieved by optimizing the fit between the cryo-EM density map and the homology model. Special methods have been developed in an attempt to improve the fitting and refinement of atomic models in cryo-EM densities. These methods include molecular dynamics flexible fitting (MDFF) and the Rosetta all-atom refinement approach (DiMaio et al., 2009). The MDFF method is the most widely used method when fitting atomic models into EM maps and although different approaches of this strategy have been developed (Volkmann et al., 2000; Chapman, 1995; Chen et al., 2001; Wriggers et al., 2000; Wriggers and Birmanns, 2001; Trabuco et al., 2008), it generally involves allowing many degrees of fitting during the refinement thus permitting the atomic models to undergo conformational changes that improve their correspondence to the EM map. The Rosetta all-atom refinement approach involves the structural refinement of atomic models based on a local measure of their fit to the corresponding EM densities, a measure which is also used to identify regions incompatible with the densities which are then targeted for extensive rebuilding (DiMaio et al., 2009).

Some current X-ray crystallography tools have recently included additional tools designed to help with the fitting of atomic models into near-atomic cryo-EM maps. The COOT program (Emsley and Cowtan, 2004) contains tools that aid in the fitting of various sections of an atomic model into a cryo-EM map such as rigid body refinement which allows movement of a single or multiple residues as one rigid body, real space refinement which allows all atoms to move and structural regularisation to correct peptide bond or side chain geometry, and apply Ramachandran restraints. The real_space_refine program (Afonine et al., 2013) from the X-ray crystallography software, Phenix (Adams et al 2010), has proven successful in the refinement of atomic models using near-atomic cryo-EM maps (Afonine et al., 2015). During the fitting the phenix.real_space_refine program implements both new real-space refinement methods and in-depth structure validation, and integrates them into a refinement workflow to produce refined models with no or minimal geometry violation such as rotamer or Ramachandran plot outliers (Fig. 3.1). This refinement procedure involves a cycle of around six steps (Fig. 3.1). The rigid body fitting step involves automated multi-zone rigid-body refinement with large convergence radius and this step is followed by the model idealisation step which involves the repacking of the model to match requested secondary structure with minimal deviation from the initial model, followed by idealisation of geometry. The model morphing step involves large rigid-body translational shifts to match the EM map. The local real-space fitting step is the most important step in the refinement step and it involves fitting with the use of molecule specific geometries such as the secondary structure, Ramachandran plots, CB-deviations and rotamer-specific restraints. This step also involves gradient-driven (LBFGS) minimisation. The model morphing step is followed by simulated annealing which is a molecular dynamic method that simulates the heating a protein followed by cooling steps that allow the structure to evolve towards the energetically favourable final structure under the influence of a force field derived from the constraints. After the simulated annealing step the model is subjected to rotamer fitting in which a valid rotamer that fits the map best is chosen and restrained to fit its current conformation. The whole refinement process is then iterated until the best fit is obtained.

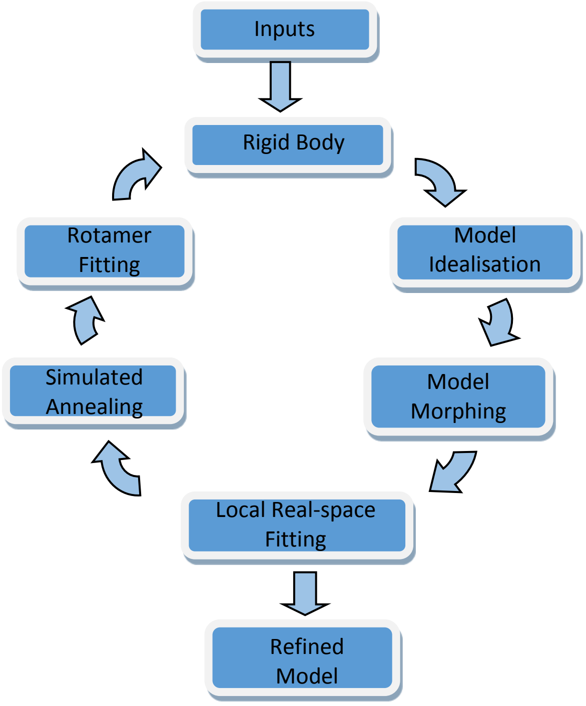


Figure 3.1. Schematic diagram of the steps carried out during the real-space refinement in the Phenix program (Afonine et al., 2015).

3.2.2 Model validation

Generated atomic models can be evaluated to assess what information can be extracted from the model. Testing of the model is carried out by running calculations based on sequence similarity, structural criteria and energetic criteria to help identify the most accurate model. Different programs can be used to validate generated atomic models and the choice of the program to use depends on the criteria being assessed. Programs such as RAMPAGE (Lovell et al., 2003), PROCHECK (Laskowski et al., 1993) or WHATCHECK (Hooft et al., 1996) test the model by identifying unreliable regions based on the models stereochemistry. Other programs such as PROSAII (Sippl, 1993) and VERIFY 3D (Eisenberg et al., 1997) predict unreliable regions of the model by generating graphs of energy distributions. PROSAII generates a Z-score, which can be used to assess the quality of the model. The Z-score evaluates the global model quality and is used to check whether the input structure is within the range of scores that are usually found for native proteins of similar size (Oany et al., 2015). VERIFY 3D measures the correctness of the model based on its 3D profile by scoring the compatibility of the model's structure with any amino acid sequence (Eisenberg et al., 1997). The fitting of the atomic models into cryo-EM maps can also be evaluated by calculating how well they fit into their corresponding EM maps. This can be achieved by calculating the cross correlation between the atomic model and the map (Map-CC) using programs such as the phenixbb function available in the Phenix software (Adams et al., 2010).

3.3 MATERIALS AND METHODS

3.3.1 Cryo-EM map model fitting and refinement

The previously generated homology model for CynD_{pum} (Park et al., 2016) was used as a starting atomic model for the model fitting. Coordinates from the Nit6803 nitrilase from *Synechocystis sp. Strain PCC6803* with an amino acid sequence identity of 32 % (PDB ID: 3WUY; Zhang et al., 2014) were used as a template to generate the homology model as this was the only confirmed branch 1 nitrilase with both the C-surface insertions and a crystal structure. The homology model was further refined by manually fitting atoms into the 3D reconstructions generated in chapter two. Different sections of the homology model were initially modified using COOT (Emsley and Cowtan, 2004) were a combination of refinement strategies were implemented. These include rigid body refinement which allows movement of a single or multiple residues as one rigid body, real space refinement which allows all atoms to move and structural regularisation to correct peptide bond or side chain geometry, and apply Ramachandran restraints.

The model's stereochemistry and fit into the cryo-EM maps were further improved in a pseudocrystallo-graphic manner using the Phenix program (Adams et al., 2010). In this method EM densities corresponding to a single monomer were cropped out using UCSF Chimera (Petterson et al., 2004) and put into an artificial P1 crystal lattices using the map_to_structure_factors program in Phenix were structural factors are also calculated. The structural factors of the models were then used for model refinement using the real-space refinement implemented by the phenix.real_space_refine program in Phenix (Fig. 3.1)(Afonine et al., 2013). The homology model was initially refined using the Q86R+H305K+H308K+H323K CryoEM map and the resultant model was used as a starting model for the wild-type EM map based refinement to help with the fitting in regions of weak or missing densities as the wild-type EM map was of a lower resolution than Q86R+H305K+H308K+H323K EM map.

3.3.2 Model validation

The quality of the generated models was validated using various programs which include RAMPAGE (Lovell et al., 2003), PROCHECK (Laskowski et al., 1993), PROSAII (Sippl, 1993) and VERIFY 3D (Eisenberg et al., 1997). The superimposition of the atomic models and calculation of their root mean standard deviation (RMSD) between their atoms was carried out using UCSF Chimera (Petterson et al., 2004). The accuracy of the model fitting was evaluated using the phenixbb function available in the Phenix software (Adams et al., 2010)

3.3.3 Visualisation and docking of atomic models

UCSF Chimera (Petterson et al., 2004) was used to view the 3D reconstructions and to manually rigid-dock the atomic models into the cryo-EM volumes using the ‘Fit model in map’ function.

3.4 Results

3.4.1 Homology model validation

Using the structure of the nitrilase Nit6803 from *Synechocystis sp. Strain PCC6803* (PDB ID: 3WUY) as a template a homology model of CynD$_{pum}$ was previously produced (Park et al., 2016). The quality of this homology model was first evaluated by the superimposition with the template and calculation of the atoms RMSD (Fig. 3.2A). The calculated RMSD of the atoms between the homolog model and the template was 0.524 Å (Table 3.1) demonstrating that the generated homology model is reasonably good and does not significantly deviate from the template structure. The overall stereochemical quality of the homology model was validated using RAMPAGE and PROCHECK. The Ramachandran plot generated from these programs (Fig. 3.2B) showed that 93.6% of the residues were in the favourable region, 4.6% in the allowed region and 1.8 % in the outlier region demonstrating that most of the amino acids are in a phi-psi distribution consistent with right handed α-helix thus showing that the homology model was of good quality. The homology model was further evaluated using PROSAII which generated Z-score of -7.47 (Fig. 3.2C). This negative Z-score indicated that the homology model was of good quality and that there is no significant deviation from the scores determined for proteins of similar size. The last test done on the model was to check the packing quality of each residue using the Verify-3D program (Fig. 3.2D). All the 3D-1D average scores were above zero (Fig. 3.2; Table 3.1), thus providing computational evidence that the residues of the homology model are in acceptable side-chain environments.

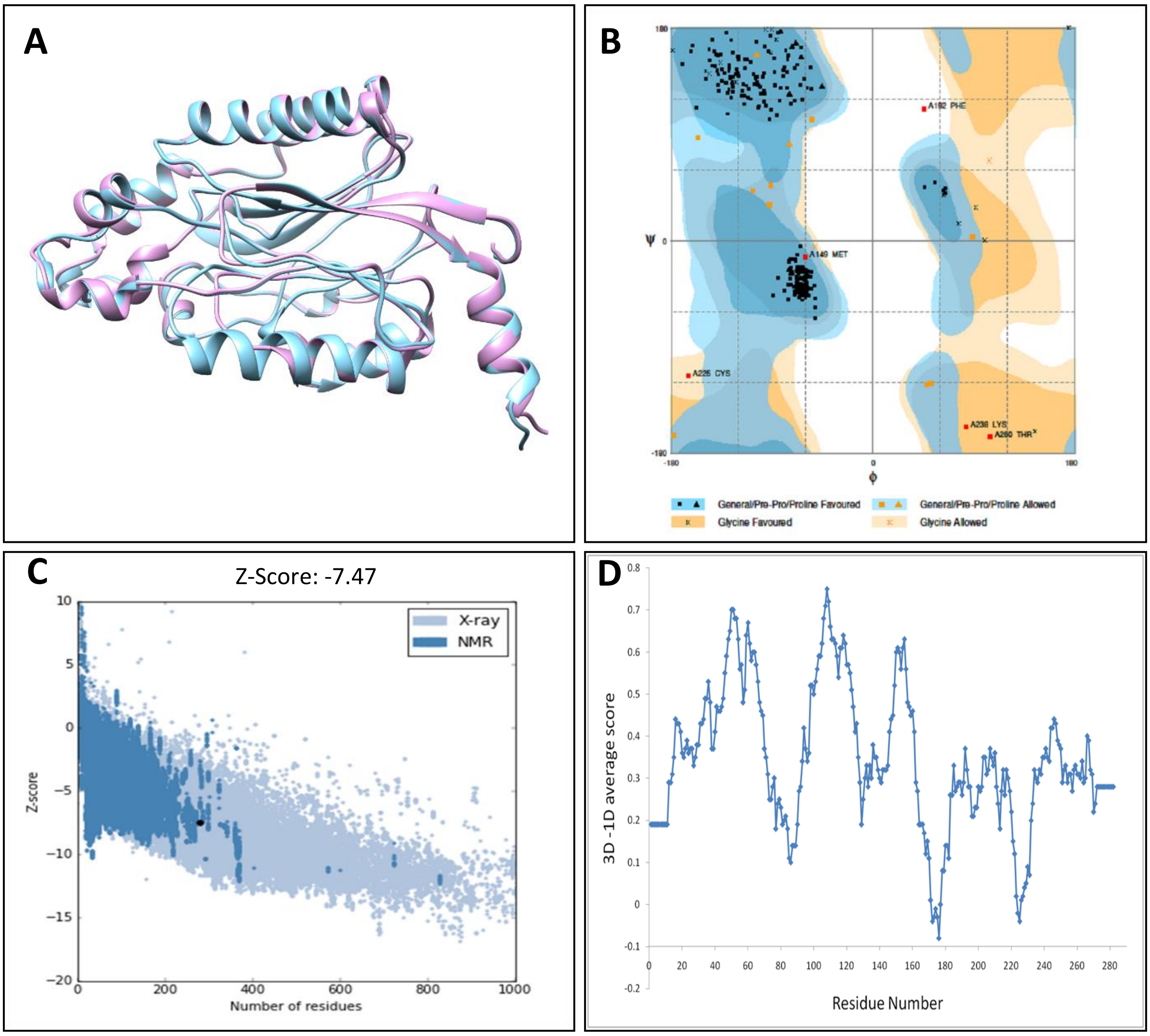


Figure 3.2. Structural evaluation of the previously generated homology model (Park et al., 2016). (**A**) The superimposition of the generated homology model and the nitrilase Nit6803 template from *Synechocystis sp. Strain PCC6803* (RMSD=0.524 Å). (**B**) The Ramachandran plot generated from the homology model using RAMPAGE. (**C**) The Z-score plot generated from the homology model using PROSAII. (**D**) The 3D-1D protein scores for each residue of the homology model calculated using Verify-3D.

3.4.2 Cryo-EM map based model fitting and refinement

Manual fitting of the atomic models into both the Q86R+H305K+H308K+H323K and wild-type cryo-EM maps was carried out using COOT (Emsley and Cowtan, 2004). The resolution of the generated EM maps was sufficient to trace the peptide backbone of the proteins (Fig. 3.3A and B) and observe well defined densities in most regions with large and aromatic residues (Fig. 3.3E). The resolution of the EM map also allowed straightforward tracing of the peptide backbone through the α-helix regions (Fig. 3.3C) but this became more challenging when tracing beta-sheets as these regions were abundant with branched and missing densities, and lacked distinguishable large or aromatic side chain residues to aid the fitting (Fig. 3.3D).

For both the variant and wild-type, only the backbone of residues 1-286 could be traced and the C-terminal residues could not be fitted due to the presence of large uninterpretable densities at the C-terminal regions. The manual fitting of the homology model into the cryo-EM maps generated new atomic models which were accessed by calculating how well they fit into their corresponding EM maps. This was achieved by calculating the cross correlation between the atomic model and the map used to generate it (Map-CC). The map-CC of the variant model greatly improved after the manual fitting (Table 3.1) and this increase in the map-CC was less for the wild-type model due to the use of the pre-refined variant model as a starting model rather than the homology model. The atomic models were further refined using the real-space refinement program in Phenix (Adams et al., 2010). The real-space refinement further improved the fitting of the variant model which was demonstrated by the increase in the map-CC. The wild-type model did not demonstrate any significant changes in the map-CC after the real-space refinement and this was expected as the pre-refined variant model was used as a starting model.

Table 3.1. Calculated scores used in the validation and fitting assessment of the homology model, Q86R+H305K+H308K+H323K and the wild-type CynD$_{pum}$ model.

Validation parameter	**Homology model**	**Q86R+H305K+H308K+H323K**	**Wild-type**
Ramachandran (Favourable region)	93.6 %	89.4 %	90.8 %
Ramachandran (Allowed region)	4.6 %	9.2 %	5.3 %
Ramachandran (Outlier region)	1.8 %	1.4 %	3.9 %
Z-score	-7.47	-6.12	-6.48
Residues with 3D-1D average score above zero	100 %	93.4 %	95.5 %
RMSD vs Nit6803 template	0.524 Å	n/a	n/a
RMSD vs Homology model	n/a	1.23 Å	1.23 Å
RMSD vs Mutant	1.23 Å	n/a	0.70 Å
RMSD vs Wild-type	1.23 Å	0.70 Å	n/a
Map vs model overall CC before fitting	n/a	0.476	0.423
Map vs model overall CC after manual fitting	n/a	0.538	0.497
Map vs model overall CC after Real-space refinement	n/a	0.570	0.491

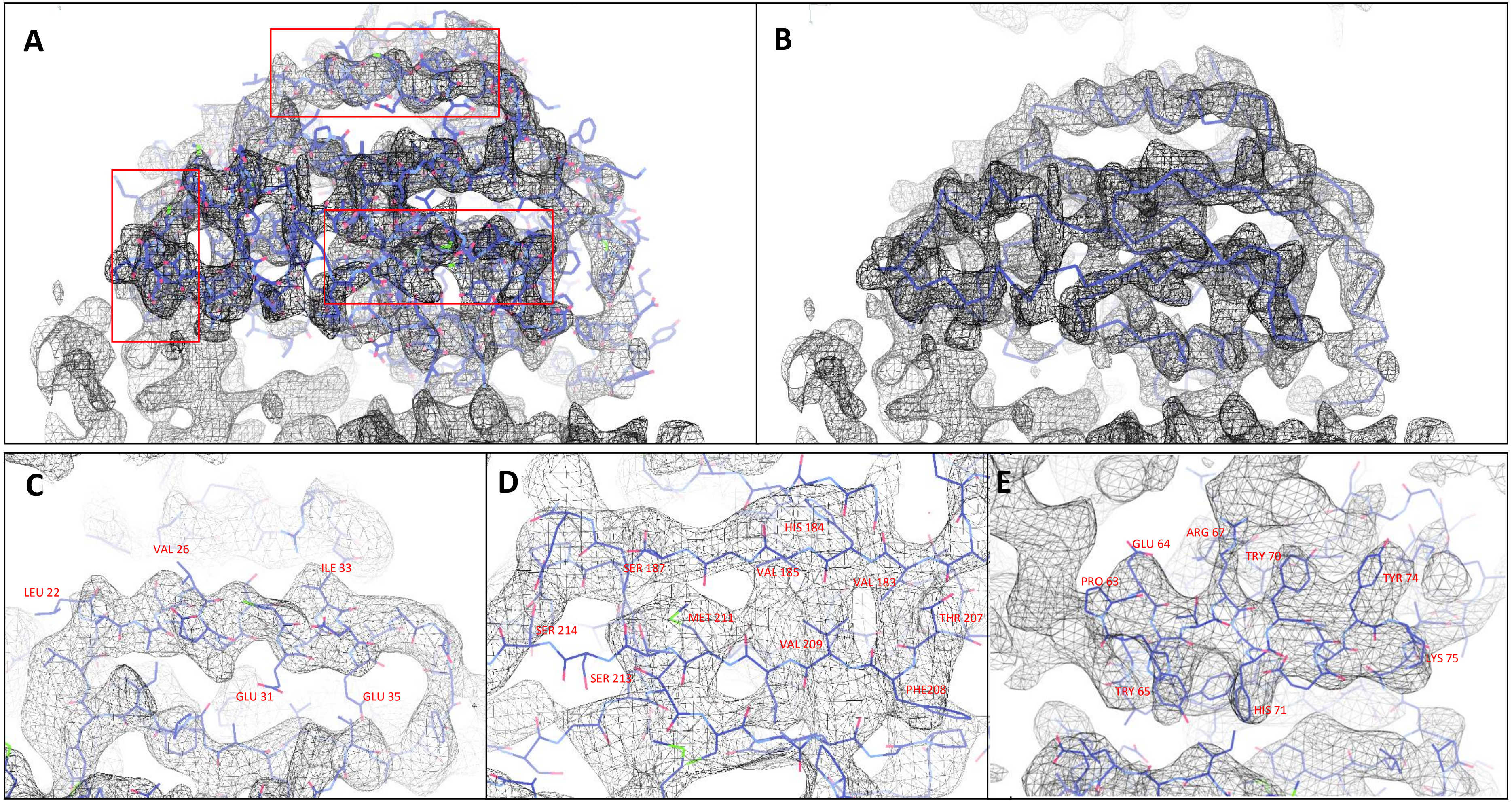


Figure 3.3. The manual fitting of the homology model into the Q86R+H305K+H308K+H323K cryo-EM map. **(A)** Overview of the fitting of the atomic model of one monomer into the Q86R+H305K+H308K+H323K cryo-EM map. **(B)** The Overview of the fitting of the peptide backbone of one monomer into the Q86R+H305K+H308K+H323K cryo-EM map. **(C)** The zoomed view of a region from image (A) demonstrating the fitting of an alpha helix. **(D)** The zoomed view of a region from image (A) demonstrating the fitting of a beta sheet. **(E)** The zoomed view of a region from image (A) demonstrating the fitting of a large/aromatic side chain residues. The manual fitting was carried out using COOT (Emsley and Cowtan, 2004).

The Q86R+H305K+H308K+H323K and wild-type cryo-EM map based refinement models (Fig. 3.4A and 3.5A) were structurally validated using a similar method to the one used for the homology model validation. The Ramachandran plot generated from the wild-type model showed a slight increase in the number of residues in the outlier region when compared to the homology model and variant model (Fig. 3.4B and 3.5B; Table 3.1). This may be due to the slightly lower quality of the wild-type cryo-EM map which may result in the incorrect fitting at region of low resolution.

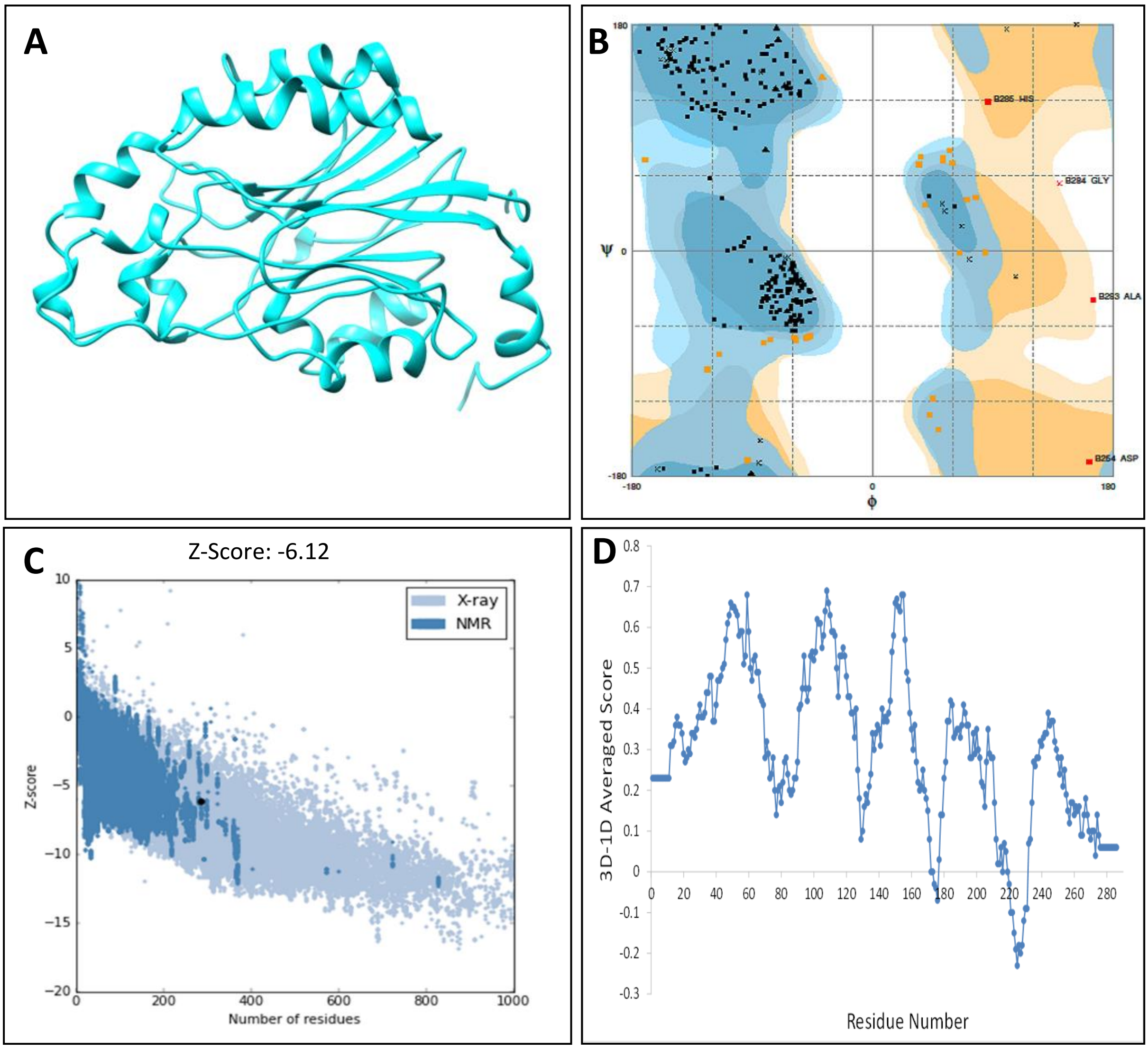


Figure 3.4. Structural evaluation of the cryo-EM map refined Q86R+H305K+H308K+H323K model. (**A**) The structure of the cryo-EM map refined Q86R+H305K+H308K+H323K model. (**B**) The Ramachandran plot generated from the Q86R+H305K+H308K+H323K model using RAMPAGE. (**C**) The Z-score plot generated from the Q86R+H305K+H308K+H323K model using PROSAII. (**D**) The 3D-1D protein scores for each residue of the Q86R+H305K+H308K+H323K model calculated using Verify-3D.

This increase in the number of residues in the outliers region hardly affected the overall stereochemical quality of the wild-type models as only ~4% of the residues were in this region. The Z-scores calculated for both the refined models were slightly greater than that of the homology model but were still negative enough to demonstrate the good quality of the models (Fig. 3.4C and 3.5C; Table 3.1). The 3D-1D average scores were calculated using the Verify-3D program to validate the packing quality of each residue in both the wild-type and mutant model. The mutant and wild-type refined models both had a majority of their residue's 3D-1D average scores above zero, with too few residues with 3D-1D average scores below zero to significantly affect the overall percentage of residues in acceptable side-chain environments (Fig. 3.4D and 3.5D; Table 3.1).

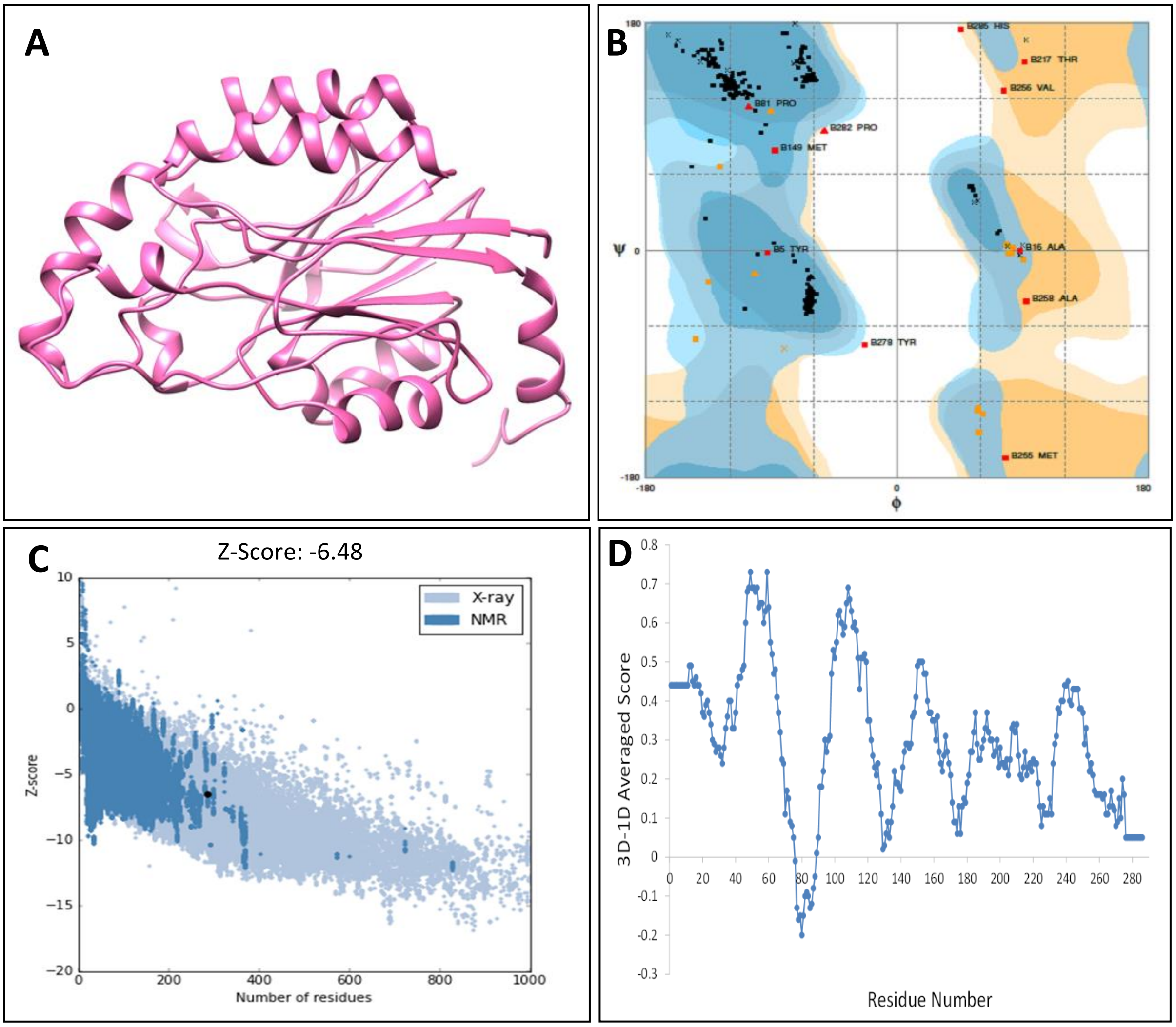


Figure 3.5. Structural evaluation of the cryo-EM map refined wild-type model. (**A**) The structure of the cryo-EM map refined wild-type model. (**B**) The Ramachandran plot generated from the wild-type model using RAMPAGE. (**C**) The Z-score plot generated from the wild-type model using PROSAII. (**D**) The 3D-1D protein scores for each residue of the wild-type model calculated using Verify-3D.

3.4.3 Overall structure comparison of the generated structures

There were no major structural differences between the homology model and the refined models, and this was demonstrated by the low RMSD value (1.23 Å) which was calculated after the superimposition of the models (Fig. 3.6A; Table 3.1). Interestingly there were also no significant difference observed between the Q86R+H305K+H308K+H323K and wild-type (Fig. 3.6B) and this was validated by an even lower RMSD (0.70 Å; Table 3.1).

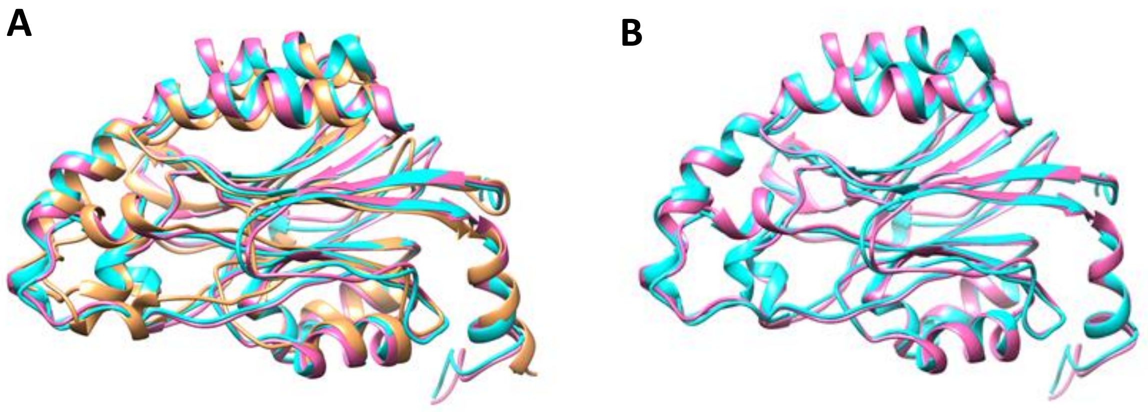


Figure 3.6. Structural comparison of the homology model, the cryo-EM map refined Q86R+H305K+H308K+H323K and wild-type models. (**A**) The superimposition of the homology model, Q86R+H305K+H308K+H323K and wild-type model. (**B**) The superimposition of the cryo-EM map refined Q86R+H305K+H308K+H323K and wild-type models. The homology model, Q86R+H305K+H308K+H323K and wild-type models are coloured in brown, purple and blue, respectively.

Multiple EM map refined monomers models of the Q86R+H305K+H308K+H323K and wild-type were docked into large EM maps to help investigate the interactions that occur at the interfacial surfaces. Interestingly there were no major differences observed between the positioning on the individual monomer after fibre formation between the mutant and the wild-type despite the previously observed differences in oligomerisation of these two variants (Fig. 3.7).

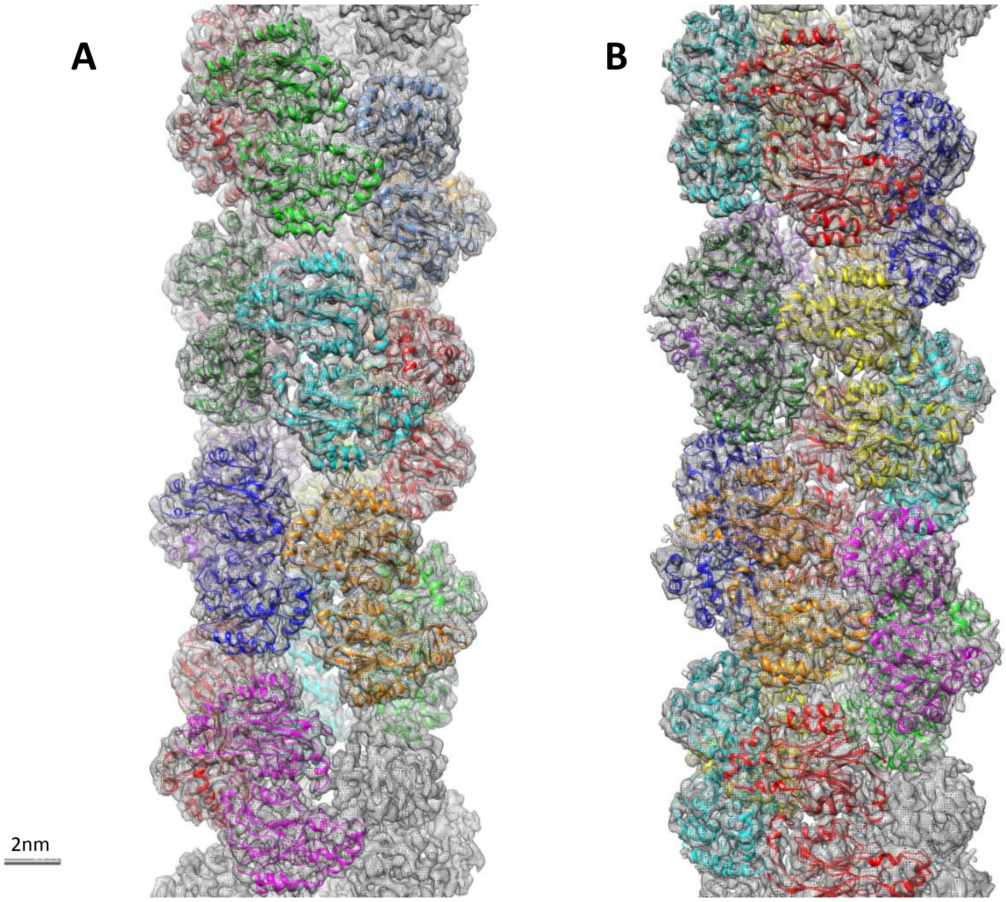


Figure 3.7. The docking of the EM refined Q86R+H305K+H308K+H323K **(A)** and wild-type **(B)** atomic models to the sharpened versions of the EM 3D reconstructions.

3.4.4 Structural insights

3.4.4.1 Interactions at the D-surface

The D-surface is the only interfacial surface which contained one of the mutations (Q86R) from the Q86R+H305K+H308K+H323K variant. Based on the comparison between the mutant and wild-type model there were no major structural changes conferred by the Q86R mutation and the interactions at this surface were similar for both models (Fig. 3.8A and B). The two alpha helices, α1 (residue 24–36) and α3 (residue 83–94), were previously postulated to be both involved in the interaction across the D-surface but based on our structures it is evident that the residues from the α1 helix may not participate in formation of this interface due to the large distance between the residues from adjacent monomers. For the wild-type, the α3 helix is the only helix involved in the formation of the D-surface (Fig. 3.8B). Based on the wild-type structure an amide bridge interaction may occur between Gln86 and the Gln86 from an adjacent monomer from another dimer. This Gln86-Gln86 interaction occurs at the Q86R mutation site for the Q86R+H305K+H308K+H323K mutant and the substitution of the Gln86 to Arg86 resulted in the introduction of a corresponding Arg-Arg interaction between two neighbouring Arg86 across the D-surface (Fig 3.8A). Based on the comparison between the electron densities from the wild-type and Q86R+H305K+H308K+H323K cryo-EM map it was observed that the mutations introduced other new interactions across the D-surface in which the C-terminal tail is involved (Fig 3.9A and B). The difference in the C-terminal tail electron density between the wild-variants is also demonstrated in the variance map constructed in chapter II (Fig. 2.26).

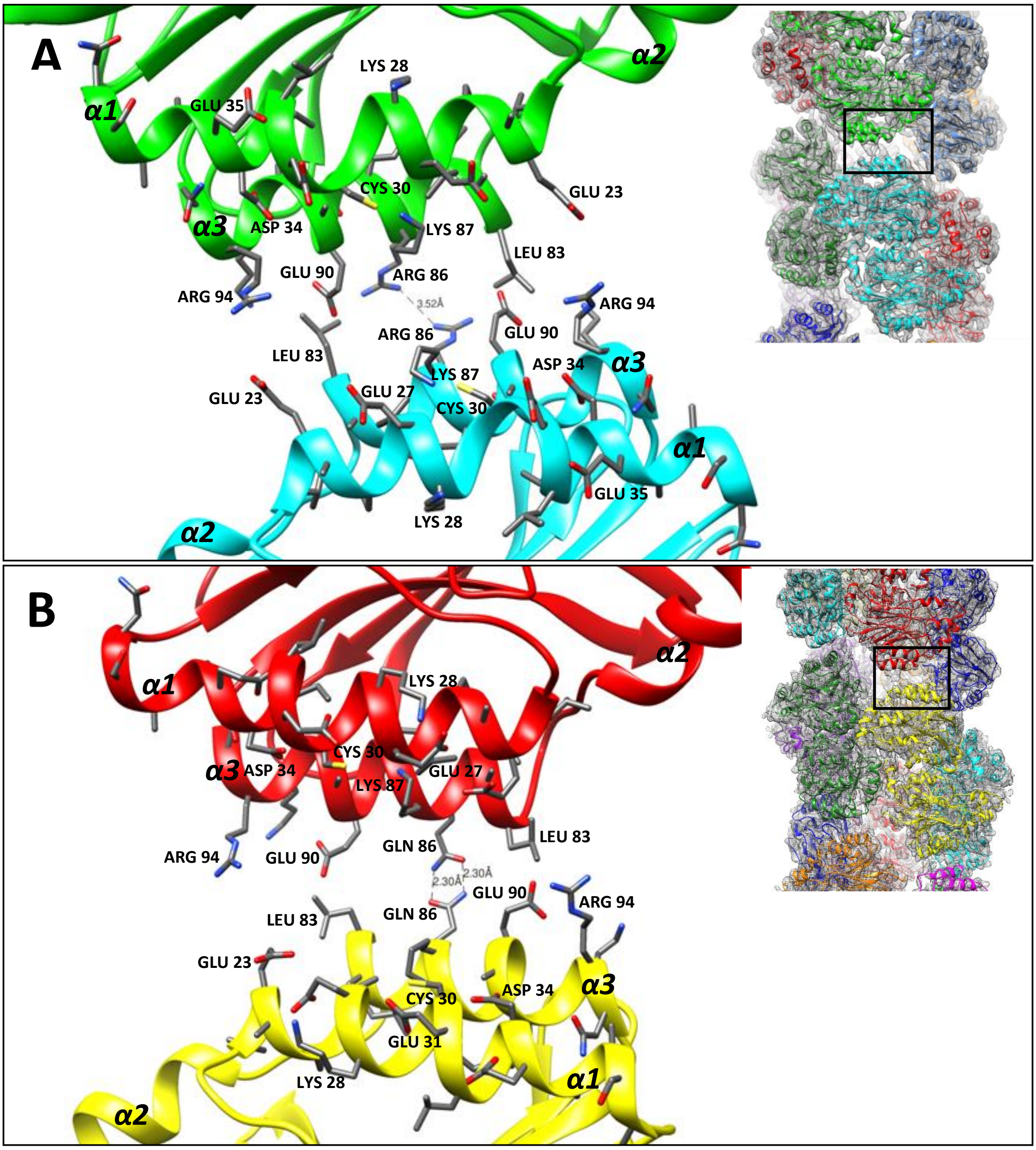


Figure 3.8. The atomic view of residues located at the D-surface of th**e (A)** Q86R+H303K+H308K+H323K CynD_{pum} and the **(B)** wild-type CynD_{pum} produced from the atomic coordinates of the cryo-EM refined models.

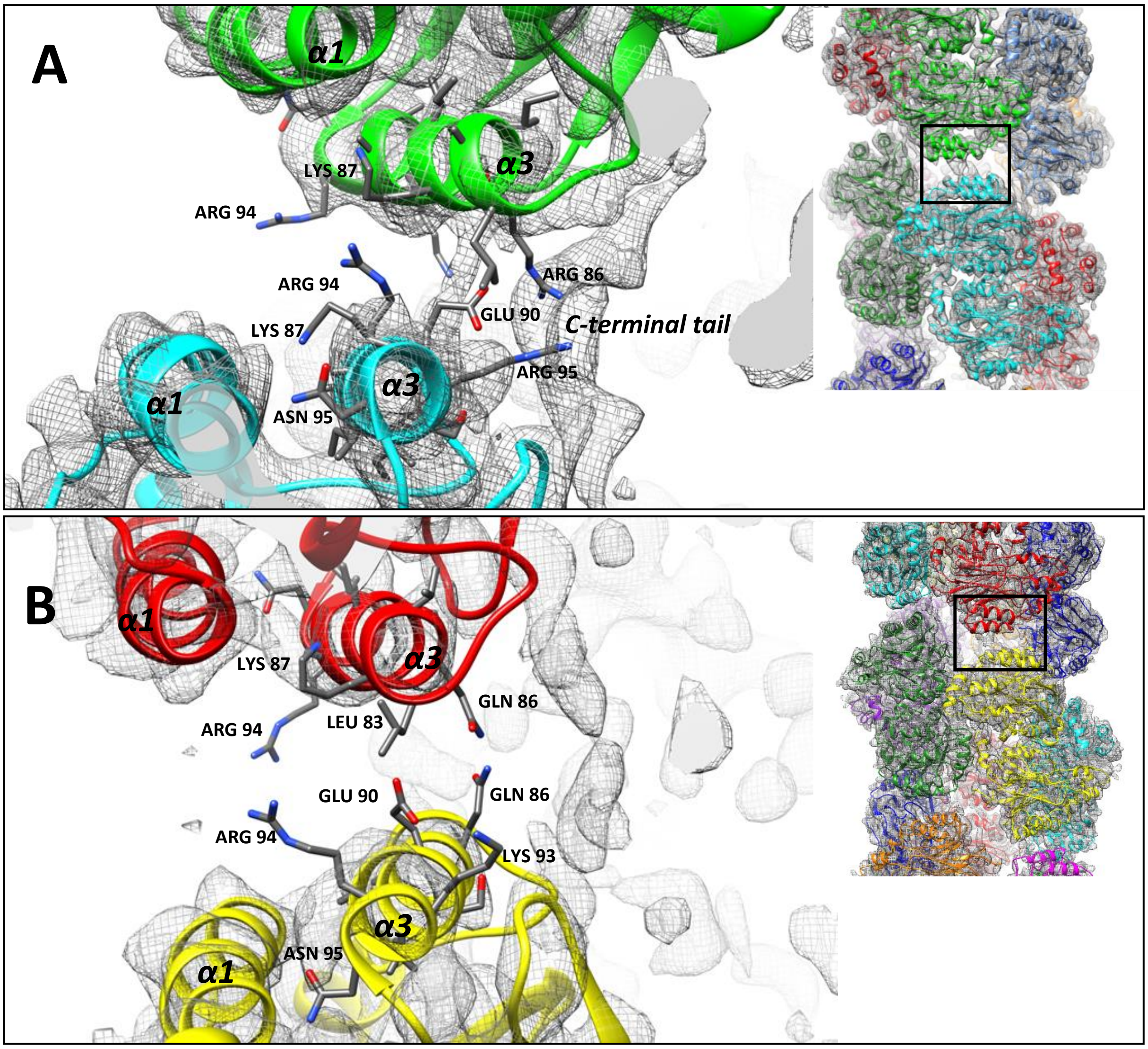


Figure 3.9. The atomic view of residues located at the D-surface of the **(A)** Q86R+H303K+H308K+H323K CynD$_{pum}$ and the **(B)** wild-type CynD$_{pum}$ produced from the atomic coordinates of the cryo-EM refined models, and their corresponding cryo-EM densities.

3.4.4.2 Interactions at the C-surface

Based on both the wild-type and the Q86R+H303K+H308K+H323K variant structure, it is demonstrated that the C-surface interactions are the predominant interactions in the stabilisation of the oligomeric structure of CynD$_{pum}$. The main C-surface interaction occurs between α2 helix (residue 65–75) and α7 helix (residue 273–278) across monomers from adjacent dimers (Fig. 3.10). The interaction between α2 and α7 include ironic interactions between Arg67 and Asp275, and between His71 and Asp 275. Hydrophobic interactions also occur between Tyr70, Tyr70 and Tyr278 at this surface. Another possible C-surface interaction may occur between the two-fold symmetry related loop regions (residue 63–59) across monomers from adjacent dimers, although identification of specific interacting amino acid is limited by the resolution of the cryo-EM map (Fig. 3.10).

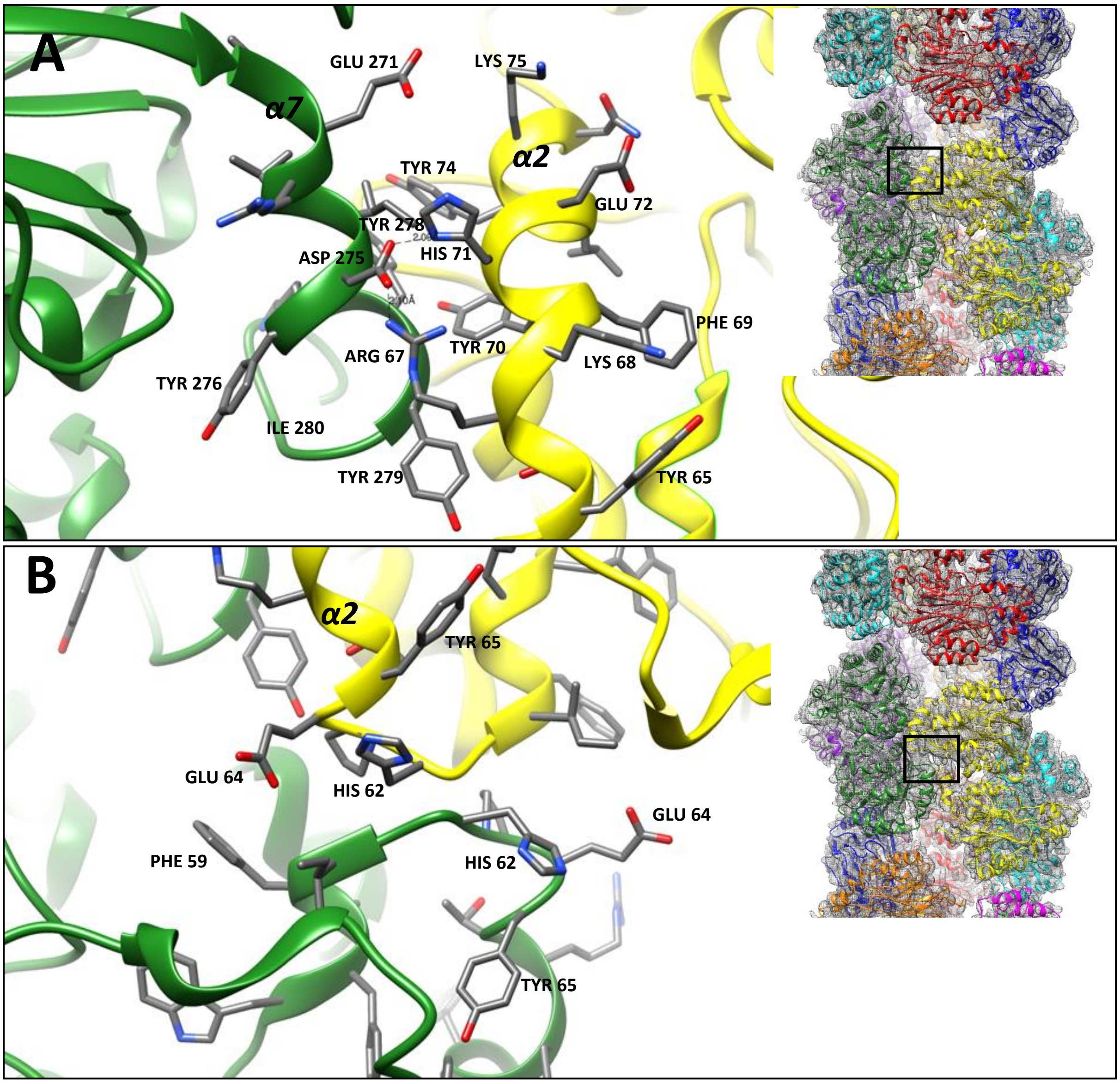


Figure 3.10. The atomic view of residues located at the C-surface of the wild-type CynD$_{pum}$ produced from the atomic coordinates of the cryo-EM refined model.

3.4.4.3 Interactions at the A-surface

The A-surface was shown to only consist of interactions between the two-fold symmetry related α6 helices (residue 196–204) across monomers from the same dimer (Fig. 3.11A), with no observable evidence of the participation of the α5 helix (residue 171–179) across this surface (Fig. 3.11B). Interactions between the α6 helices are formed by hydrophobic interaction between Tyr200 and Tyr200 from the second monomer from the same dimer (Fig. 3.11A).

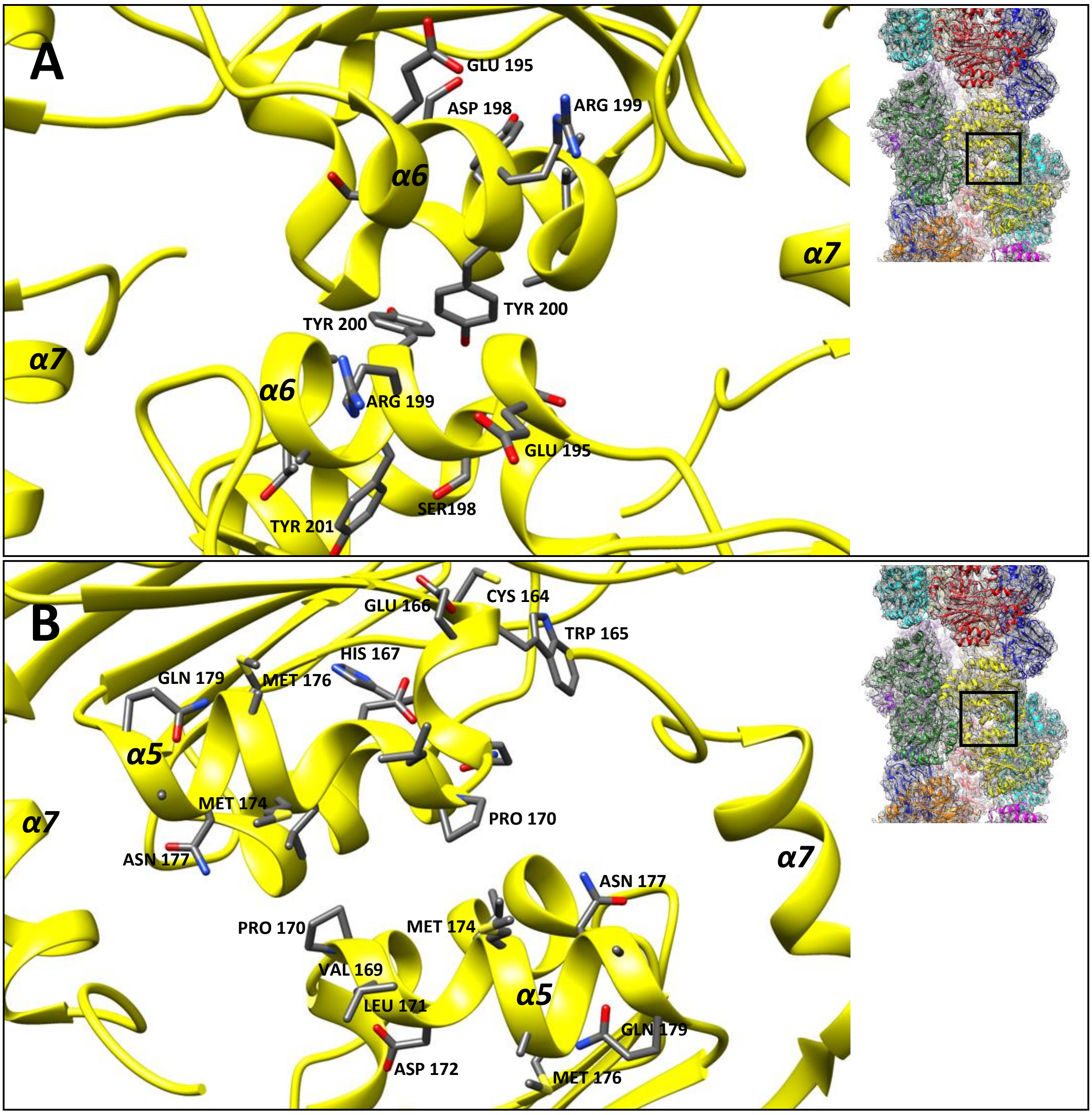


Figure 3.11. The atomic view of residues located at the A-surface of th**e** wild-type CynD$_{pum}$ produced from the atomic coordinates of the cryo-EM refined model, consisting of the α6 helix **(A)** and the α5 helix **(B)**.

3.5 Discussion

3.5.1 Model fitting and refinement

A previously generated CynD$_{pum}$ homology model was refined based on the high resolution cryo-EM maps of the Q86R+H305K+H308K+H323K and the wild-type. The low RMSD values calculated between the homology model and both the mutant and wild-type demonstrates a close relationship between CynD$_{pum}$ and the nitrilase Nit6803 from *Synechocystis sp.* Strain PCC6803 which was used as a template during the construction of the homology model. This close relationship may also exist with other branch 1 nitrilases and may allow for more accurate structural predictions of the unsolved branch 1 nitrilases. Refinement of the models based on the cryo-EM maps was challenging due to the limited resolution of the maps and the lack of good near atomic resolution fitting tools. Initial refinement of the models was attempted using molecular dynamics flexible fitting (MDFF) but this process was unsuccessful and resulted in misfitting as higher resolution constraints were applied demonstrating a low resolution limitation of MDFF based approaches. Although not ideal, X-ray crystallography tools proved adequate in the fitting of atomic models into near atomic cryo-EM maps. At near atomic resolution the location of side chains of large residues such as aromatic residues were easy to identify. In contrast, smaller residues had little or missing density making it challenging to accurately identify their locations. Some of the slightly larger residue such as the Glu and Asp residues also had missing densities and this has been reported in previous cryo-EM studies where it has been suggested that carboxyl groups are particularly susceptible to radiation damage (Grigorieff et al., 1996; Mitsuoka et al., 1999; Sachse et al., 2007). Only atomic coordinates of residues 1–286 could be fit into the wild-type and mutant cryo-EM maps as after residue 286 the density corresponding to the residue were undistinguishable. The high cluster of EM map density that prevent the accurate fitting at this C-terminal region may be due to either the flexibility of the C-terminal tail or the high number of intermolecular interactions between the C-terminal tail from neighbouring monomers which result in the large undistinguishable densities. Although the final refined models had high map-CC scores, the limited resolution of the maps and the use of a predicted homology model as a template limits the amount of structural information that can be extracted but nevertheless the models had enough structural information to help elucidated the interactions that occur at the interfacial regions of CynD$_{pum}$ enzyme.

The small RMSD value calculated for the wild-type and mutant atomic models demonstrated little structural variation between these enzymes, thus the Q86R+H305K+H308K+H323K mutations confer small structural changes that ultimately result in significant changes to the oligomerisation of $\mathrm{CynD_{pum}}$. The similarity between the mutant and the wild-type was also demonstrated by the variance map produced in chapter II (Fig. 2.26) in which a majority of the variations between the models occurred only at the C-terminal region.

3.5.2 Structural insights and effects of mutations

3.5.2.1 Interactions at the D-surface

Previous experiments have demonstrated the involvement of the D-surface in the stabilisation of the $\mathrm{CynD_{pum}}$ quaternary structure (Sewell et al., 2005; Wang et al., 2011, Mulelu, 2013). The $\mathrm{CynD_{pum}}$ D-surface was initially postulated to consist of interactions between two alpha helices (α1 and α3) from adjacent monomers and it was suggested that a majority of these interactions were salt bridges across this surface (Sewell et al., 2005). Based on our atomic structures it was observed that only the α3 helix residues could participate across the D-surface, and it was also observed that only a few salt bridges could be formed across this surface (Fig. 3.8). Although residues from the α1 helix may not be involved in the formation of the D-surface, they have been shown to be involved in the general stabilisation of $\mathrm{CynD_{pum}}$ as mutation of the α1 glutamates to lysine (E27K, E31K and E35K, Table 1.2) results in aggregation and insolubility of the enzyme (Gray, 2008).

In previous mutagenesis studies we identified the D-surface Q86R mutation which has been shown to enhance the degree of oligomerisation of the $\mathrm{CynD_{pum}}$ enzyme (Wang et al., 2011). Gln86 is located at the D-surface α3 helix and we propose that its substitution to an arginine removes the Gln86-Gln86 amide bridge interaction and introduces a more stabilising interaction between a pair of Arg86 across the D-surface, thus increasing the stability of the oligomeric structure of the enzyme and resulting in the observed increase in the degree of oligomerisation (Fig. 3.8 and Fig. 3.12). Although this Arg-Arg interaction is not electrostatically favourable it has been demonstrated that electrostatics-defying Arg-Arg interactions are common in microenvironments of polarizable residues such as the D-surface region, and act as a possible thermodynamic driving force favourable for protein-protein interaction (Pednekar et al., 2009, Zhang et al., 2013). Another interesting mutation located at the D-surface α3 helix is the K93R mutation which has been shown to enhance the catalytic activity of $\mathrm{CynD_{pum}}$ (Crum et al., 2016). Based on our structure there was no clear observed interaction that the Lys93 residue may be involved in but we postulated that the introduction of an arginine at this position may introduce either an Arg93-Asn121 hydrogen bond or an Arg93-Asp123 ionic bond which helps stabilise the catalytic residue and results in the observed increase in the catalytic activity of the enzyme (Fig. 3.12).

The comparison between the wild-type and the mutant cryo-EM density maps demonstrated that the Q86R+H305K+H308K+H323K mutations introduced a new interaction across the D-surface that involves the C-terminal tail (Fig. 3.9A and B). Only the H305K+H308K+H323K mutations, rather than the combination of Q86R+H305K+H308K+H323K, are the likely contributors to the introduction of this interaction as they are the only mutations located at the C-terminal region. The H305K+H308K+H323K mutant has been shown to enhance oligomerisation (Mulelu, 2013) and this observed characteristic may be due to the introduction of the additional interaction that further strengthens the D-surface interactions resulting in a more stable enzyme. Although we were unable to obtain detailed structural information about the C-terminal tail, comparing the difference in the density between the wild-type and the Q86R+H305K+H308K+H323K mutant at this C-terminal region has demonstrated that the C-terminal tail is highly involved in the oligomerisation of CynD$_{pum}$. The H305K+H308K+H323K mutations may introduce other new interaction besides the ones observed at the D-surface, as a high structural variation was observed at the C-terminal region when comparing the wild-type and Q86R+H305K+H308K+H323K cryo-EM maps (Fig. 2.26) but based on the high similarity between the atomic models of these two CynD$_{pum}$ variants it was demonstrated that the major interactions that confer the Q86R+H305K+H308K+H323K mutant's high stability characteristics are located at the D-surface.

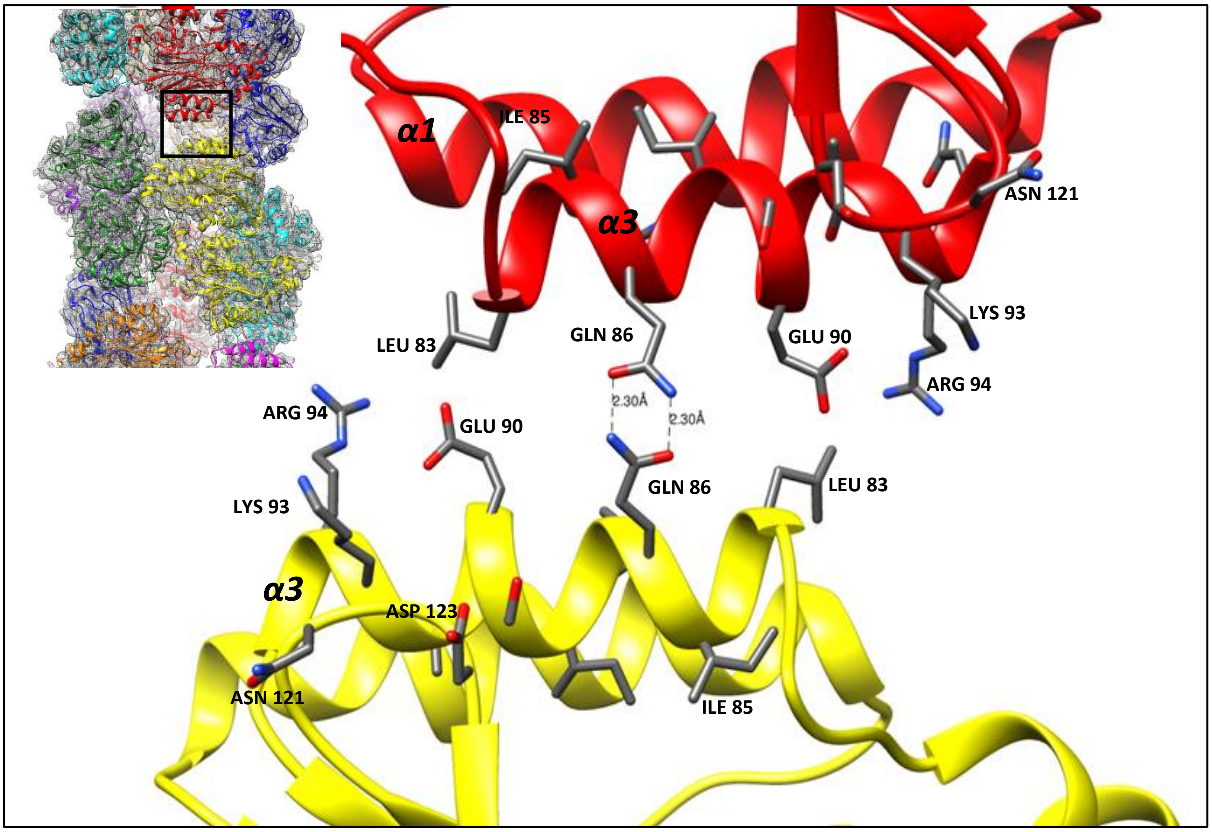


Figure 3.12. The atomic view of α3 helix residues located at the D-surface of the wild-type CynD$_{pum}$ produced from the atomic coordinates of the cryo-EM refined model.

3.5.2.2 Interactions at the C-surface

There has been evidence that the putative C-surface may be the predominant region responsible for fibre formation of CynD$_{pum}$ (Sewell et al., 2005). Using homology modelling two insertions, present in the spiral forming CynD$_{pum}$ and not in non-spiral forming nitrilases, have been identified (Thuku et al., 2009). These two insertions designated region 1 and 2 consist of residues 55–72 and residues 222–235 respectively (Fig. 3.13). Based on our cryo-EM structures we observed that the C-surface region 2 residues are less likely to interact across the two-fold symmetry as the distance between adjacent region 2 loops is shown to be greater than previously proposed (Mulelu, 2013) (Fig. 3.13). In a recently published study, each of the CynD$_{pum}$ residues from these insertions were investigated through mutagenesis studies (Park et al., 2016). It was discovered that certain residues in the insertions, when mutated to a cysteine, had an effect on the oligomerisation of CynD$_{pum}$. Substitution mutations F57C and G61C resulted in the formation of short 14-mers similar to those of the homologous cyanide dihydratase from *Pseudomonas stutzeri* (Sewell et al., 2003) at pH 8. Based on our high resolution structures the Phe57 is located at a highly hydrophobic area and may be required to stabilize the hydrophobic interaction around this zone thus an introduction of a cysteine may lead to partial disruption in the oligomerisation of CynD$_{\mathrm{pum}}$ (Fig. 3.13A). The Gly61 residue is one of a few C-surface region 1 residues that could potentially interact across the two fold symmetry axis at the C-surface (Fig. 3.13A). The substitution of the region 1 residues to cysteines was postulated to introduce new disulphide bonds which could potentially result in better stabilization of the enzyme but this was not the case as none of the cysteine mutations around this two-fold axis region (F59C, I60C, G61C and H62C) introduced new disulphide bonds (Park et al., 2016). The absence of new disulphide bonds demonstrates that none of residues at the two-fold axis region are optimally aligned to allow Cys-Cys interactions between corresponding residues from adjacent monomers, although Cys-Cys interactions between cysteines from non-corresponding position may be possible as shown in our structures (Fig. 3.13A). Other substitution mutations including W56C, E64C, T66C, F59C, Y70C, H71C, Y233C and F234C resulted in the formation of a mixture of both the 14-mers and the native 18-mers at pH 8. A majority of the C-surface mutations that conferred changes in the oligomerisation are located at the region 1 position thus confirming the unlikely interactions between adjacent region 2 loops shown in our structures. Tyr233 and Phe234 from the proposed region 2 loop at the C-surface may be responsible for stabilizing the two short α helices located at this region and these stabilising interactions may result in structural changes that affect the oligomerisation of the enzyme (Fig. 3.13B). The C-surface region 1 was shown to consist of a cluster of hydrophobic residues which form hydrophobic interactions which may be responsible for the stabilisation of this region. These hydrophobic residues include Pro55, Trp56, Phe57, Phe59 and Phe69 and mutation of some of these residues may cause the disruption of the stabilising hydrophobic interaction resulting in the observed changes in the oligomerisation of the enzyme. Interestingly the R67C mutation caused partial defects in spiral formation in which a mixture of short spirals, ring, C-shaped and smaller oligomers were observed. Based on our atomic structure we demonstrated that the predominant interactions that occur across the C-surface are the Arg67-Asp275 and His71-Asp275 ionic bond interactions. These interactions and the higher degree of disruption in oligomerisation after the introduction of the R67C mutation when compared to the introduction of the H71C mutation illustrates that the Arg67-Asp275 may be the key interaction at the C-surface responsible for maintaining the CynD$_{\mathrm{pum}}$ oligomeric structure.

In the Park et al. (2016) study, it was also demonstrated how some of the residues from the C-surface are crucial in the activity of CynD$_{\mathrm{pum}}$. A majority of the substitution mutations around the region 1 insertion resulted in ≤ 50% reduction in activity when compared to the wild-type (P55C, F57C, G61C, E64C, Y65C, T66C, R67C, F69C and Y70C) while only two region 2 mutations (Q230C and E235C) reduced activity to ≤ 50%. Based on our atomic structures we observed that some of the C-surface region 1 residues are located on or near the loop containing the catalytic Glu48 residue. We hypothesise that the disruption of the C-surface region 1 interactions by the activity reducing mutations mentioned above destabilises the catalytic active site by repositioning the catalytic Glu48 resulting in the observed reduction in the activity of the enzyme. The C-surface region 2 residues are not located near any of the catalytic residue and this explains why a majority of mutations at the C-surface region 1 resulted in reduction of activity and only two mutations at the C-surface region 2 resulted in reduction of activity at a lesser degree. In the same study it was observed that out of the 12 C-surface substitution mutations that caused changes in oligomerisation, 10 of them conferred changes in both oligomerisation and activity thus demonstrating a relationship between the oligomerisation and activation of $\mathrm{CynD}_{\mathrm{pum}}$.

In the Park et al. (2016) study on the C-surface interactions, all the C-surface region 1 and 2 residues were individually mutated to cysteines in hope of creating new disulphide bonds across this surface if any of the residues interacted across the C-surface two-fold axis. No new disulphide bonds were observed after the mutations thus validating our atomic models in which the C-surface interactions were shown to occur between the residues of the α2 and α7 helix and not across the C-surface two fold axis as previously proposed. The unlikely participation of the C-surface region 2 residues mentioned above further validates our atomic models in which region 2 residues from adjacent monomers are shown to be too far apart to interact (Fig. 3.13A).

The presence of the two C-surface insertions in spiral forming nitrilases relative to non-spiral-forming homologues has demonstrated the important role of these insertions in the oligomerisation (Sewell et al, 2003). When comparing the relationship between the C-surface insertions of two crystal structures from two spiral-forming nitrilases, the C-shaped spiral forming β-alanine synthase from *Drosophila melanogaster* (Lundgren et al., 2008) and the extended helix forming Nit6803 nitrilase from *Synechocystis sp.* Strain PCC6803 (Zhang et al., 2014), it was evident that conformations and interactions between neighbouring monomers differ at the C-surface of the two crystal structures (Park et al., 2016). Based on our high resolution structure we confirmed that the $\mathrm{CynD}_{\mathrm{pum}}$ C-surface insertions more closely resemble those of the Nit6803 nitrilase than the β-alanine synthase. The difference in conformations and interactions at the C-surface between these two crystal structures, as well as our $\mathrm{CynD}_{\mathrm{pum}}$

structures, demonstrates that although the interaction at the C-surface are predominantly responsible for oligomerisation they differ across spiral-forming nitrilase members and may contribute to the different oligomeric structures observed in the nitrilase superfamily despite the members having high monomeric structural homology.

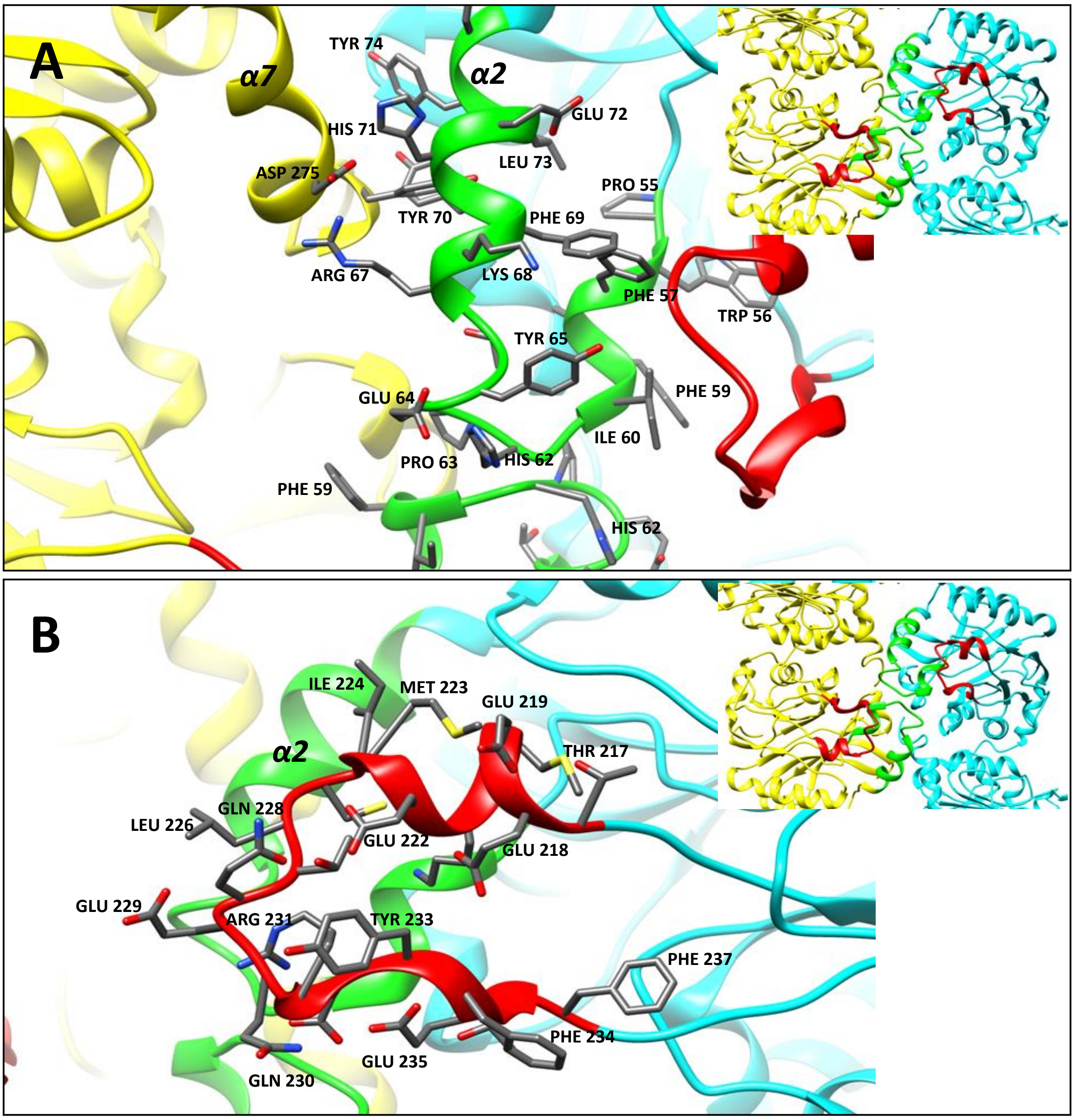


Figure 3.13. The atomic view of residues located at the C-surface region 1 **(A)** and region 2 **(B)** of the wild-type CynD$_{\text{pum}}$ produced from the atomic coordinates of the cryo-EM refined model. The residues of the C-surface region 1 are highlighted in green and the residues of the C-surface region 2 are highlighted in red.

3.5.2.3 Interactions at the A-surface

Interactions across the A-surface have been postulated to be the basis of dimerisation based on structural comparisons between nitrilase superfamily members (Thuku et al., 2009). Although the A-surface is common amongst nitrilase superfamily members, interactions across this surface has been shown to vary between the members (Sewell et al., 2003; Sewell et al., 2005; Thuku et al., 2007; Woodward et al., 2008, Dent et al., 2009; Williamson et al., 2010). The A-surface interactions have been shown to be important in activation of $\mathrm{CynD_{pum}}$ and were previously postulated to occur between two alpha helices namely α5 and α6 (Sewell et al., 2005; Crum et al., 2016). Based on our atomic structures we observed that only the α6 helix residues are involved in interactions across the A-surface with only one hydrophobic interaction between a pair of Tyr200 (Fig. 3.14A). This low number of interactions across the A-surfaces demonstrates the possibility that the dimerisation of $\mathrm{CynD_{pum}}$ may not occur at the A-surface, but rather the C-surface. Due to the close proximity of the α6 helix to the α5 helix, residues from the α6 helix may influence interactions at the α5 helix which contains the catalytic Cys164 residue thus affecting the catalysis of the enzyme (Fig. 3.14). This was evident when mutations introduced at the α6 helix (Y201D, A202T and A204D) rendered the enzyme inactive (Sewell et al., 2005; Crum et al., 2016). The A202T mutation at the α6 helix was also shown to cause the aggregation of the enzyme (Crum et al., 2016) demonstrating the importance of interactions at the A-surface region in stabilising the structure of the $\mathrm{CynD_{pum}}$. Interestingly a D172N mutation at the α5 helix was shown to increase the themostability of $\mathrm{CynD_{pum}}$ (Crum et al., 2016). Based on our atomic structures it was difficult to elucidate how the removal of the negative Asp172 charge to a polar Asn172 would result in an increase in stability but we propose that the mutation may introduce interactions across two adjacent α5 helix at the A-surface resulting in more stable enzyme (Fig. 3.14B).

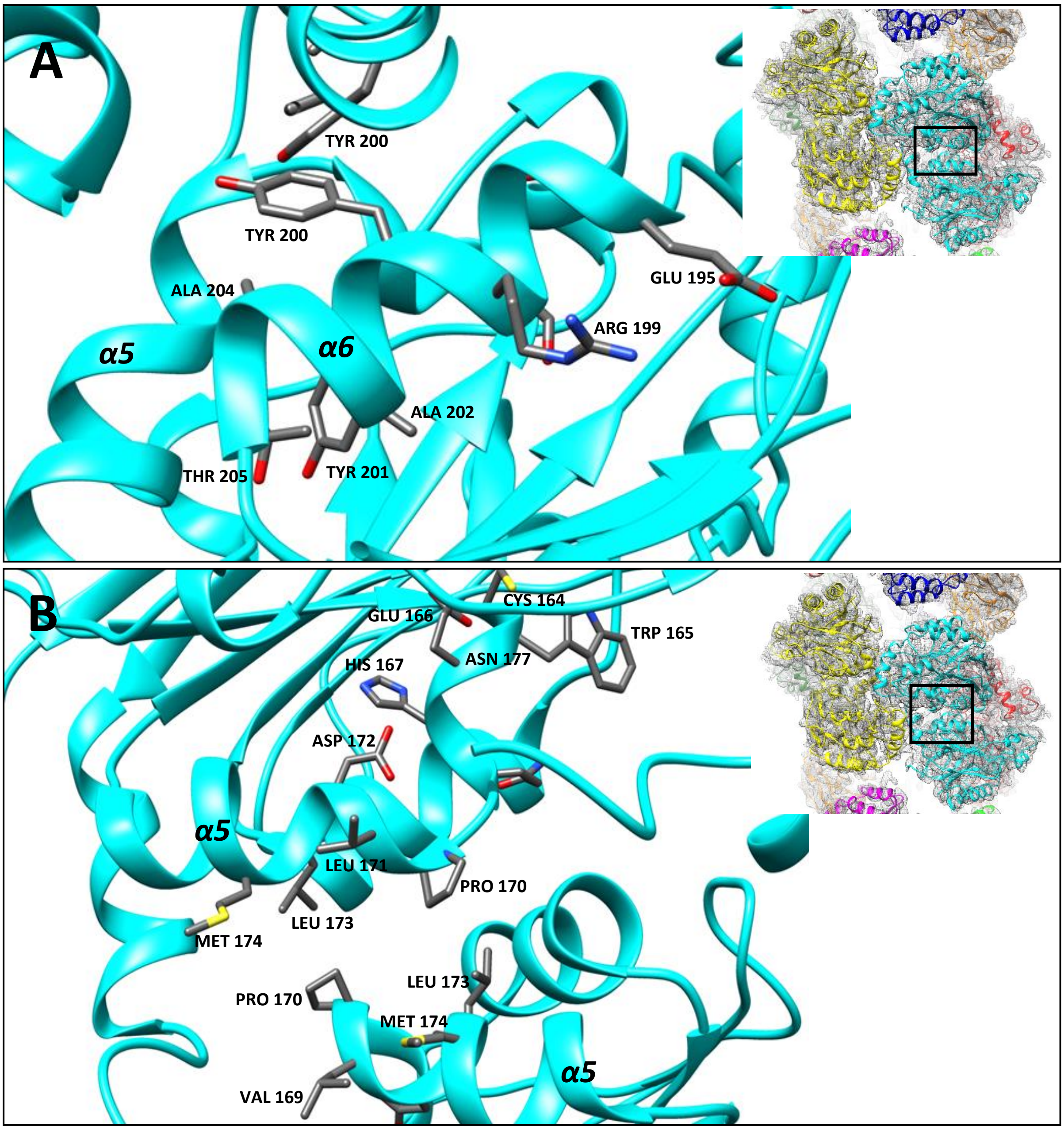


Figure 3.14. The atomic view of residues located at the A-surface of th**e** wild-type CynD$_{pum}$ produced from the atomic coordinates of the cryo-EM refined model, consisting of the α6 helix **(A)** and the α5 helix **(B)**.

3.6 Conclusion

We successfully refined a previously constructed homology model (Park et al., 2016) by fitting the model into the cryo-EM density maps obtained in chapter II. The resolution of the cryo-EM allowed for easy fitting of the peptide backbone, alpha helices and large side chain residues with the fitting of beta sheets and loops proving more difficult. Two cryo-EM map refined atomic structures were generated for the wild-type and Q86R+H305K+H308K+H323K mutant. Based on this atomic structures we observed that the high stability conferred by the Q86R+H305K+H308K+H323K mutations was due the substitution of the Gln86-Gln86 hydrogen bond to a more stabilising Arg86-Arg86 interaction across the D-surface (Q86R) and the introduction of a new C-terminal tail interaction across the D-surface as a result of the H305K+H308K+H323K mutations. The C-surface interactions were shown to be the predominant interfacial interaction responsible for the oligomerisation of $\text{CynD}_{\text{pum,}}$ were only residues from the C-surface region 1 and none from region 2 were shown to participate across this surface. The major interactions across the C-surface occur between the α7 helix and the α2 helix were the Arg67-Asp275 and His71-Asp275 ionic bond interactions play a crucial role in oligomerisation of CynD_{pum} enzyme. A relationship between the oligomerisation and activation of CynD_{pum} was observed when analysing the C-surface interactions. The A-surface interactions were demonstrated to play the role in the activation and stabilisation of the enzyme. The possibility of the A-surface's non-involvement in dimerisation was also illustrated with the possibility of dimerisation occurring at the C-surface. Using a combination of the atomic structure and previous mutagenesis studies we were not only able gain new insights into the structure of CynD_{pum} but were able to validate the accuracy of the our CynD_{pum} atomic models at interfacial regions.

CHAPTER IV

THE ROLE OF HISTIDINES IN THE OLIGOMERISATION OF CynD$_{pum}$

4.1 Introduction

CynD$_{pum}$ has been shown to exists as active short spirals over the pH range of 6–8, spirals which associate to form long active helical fibres at pH 5.4 (Fig. 1.8) (Sewell et al., 2005; Wang et al., 2011, Mulelu, 2013). The transition from short spirals to long helices occurs as the pH drops below 6.0 and this structural change points to the participation of histidine which is the only amino acid that undergoes a change in charge around pH 5.4 (histidine pKa of 6.0). There are 10 histidines in the amino acid sequence of CynD$_{pum}$, three of which are located on the C-terminal tail (Fig. 1.3). During a previous investigation of the factors involved in the oligomerisation of CynD$_{pum}$ by attempting to identify key amino acid residue responsible for the formation of CynD$_{pum}$ oligomers, using mutagenesis studies, it was demonstrated that the three C-terminal tail histidines (His305, His308 and His323) were only partially involved in fibre formation and hence a study had to be done on the 7 remaining histidines (His62, His71, His128, His167, His184, His241 and His285) to identify the key histidine residues involved in the observed fibre formation of CynD$_{pum}$. To test the involvement of these histidines, we mutated each of the histidines to either a lysine (to observe the effect of substituting with a residue with the same charge but a higher pKa of 10), an aspartate (to observe the effect of substituting with a residue with an opposite charge) or an alanine (to observe the effect of substituting with a residue with no charge). Differential scanning fluorimetry assays were also carried out on the histidine mutants to assess whether the observed changes in the oligomerisation were due to a change in stability or the change in the structural conformation of the mutants. Differential scanning fluorimetry assays were also used to help identify any histidine mutants with increased stability caused by the substitution mutations. In this chapter we describe the construction, solubility optimization, solubility validation and purification of the histidine substitution mutants. We then describe the observed changes in the pH dependent oligomerisation of the CynD$_{pum}$, conferred by the histidine substitution mutations, using the Transmission Electron Microscope.

4.2 Differential scanning fluorimetry

Differential scanning fluorimetry (DSF) is a fluorescence-based thermostability assay developed by Pantoliano et al. (2001). It is a high throughput screen method mainly used for buffer optimization and ligand-induced stabilisation to allow rapid identification of ligands of target protein from compound libraries during drug discovery (Pantoliano et al., 2001). DSF has been successfully applied in buffer screening (Phillips et al., 2011), kinetic studies (Matulis et al., 2005), ligand screening (Caver et al., 2005) and testing of the stability of different functional protein variants (Lavinder et al., 2009). The principle behind DSF is that folded and unfolded proteins can be distinguished through exposure to a hydrophobic fluorescent probe. The fluorescent probe is quenched in aqueous solution but binds to the exposed hydrophobic interior of an unfolding protein resulting in a sharp decrease in quenching which produces a readily detectable fluorescence emission that can be measured as a function of temperature (Errisson et al., 2006). During the DSF assay the temperature of the protein-probe solution is gradually increased as a function of time leading to the unfolding of the protein and as this happens the fluorescent probe binds to its exposed hydrophobic region resulting in a significant increase in fluorescence emission (Fig. 4.1). As the temperature continues to increase more hydrophobic regions are exposed and an exponential increase of the fluorescence is observed until a maximum is reached which is followed by a steady decrease of the fluorescence (Fig. 4.1). This steady decrease in fluorescence is postulated to be due to precipitation and aggregation of the protein. The temperature at which the concentration of the folded and unfolded state of the protein reaches equilibrium is considered as the melting temperature (T_m) and is defined as the midpoint temperature for the protein–unfolding transition (Errisson et al., 2006). The simplest way to calculate T_m values is to determine the maximum of the first derivative of the fluorescence versus temperature plot. T_m values are used to calculate changes in the thermostability of the protein. The most commonly used fluorescent probe is the SYPRO Orange dye, due to its high signal to noise ratio. Another property that makes SYPRO Orange a favourable dye is its relatively high wavelength of excitation (465 nm excitation, 500 nm emission) which decreases the interference of the optical properties by any small molecules which usually cause the quenching of the fluorescence intensity (Niesen et al., 2007).

DSF is an excellent method to screen for conditions or mutations that stabilise proteins, owing to the very small quantities of protein needed, reproducible results, quick simultaneous screening and no required prior knowledge of the proteins (Matulis et al., 2005).

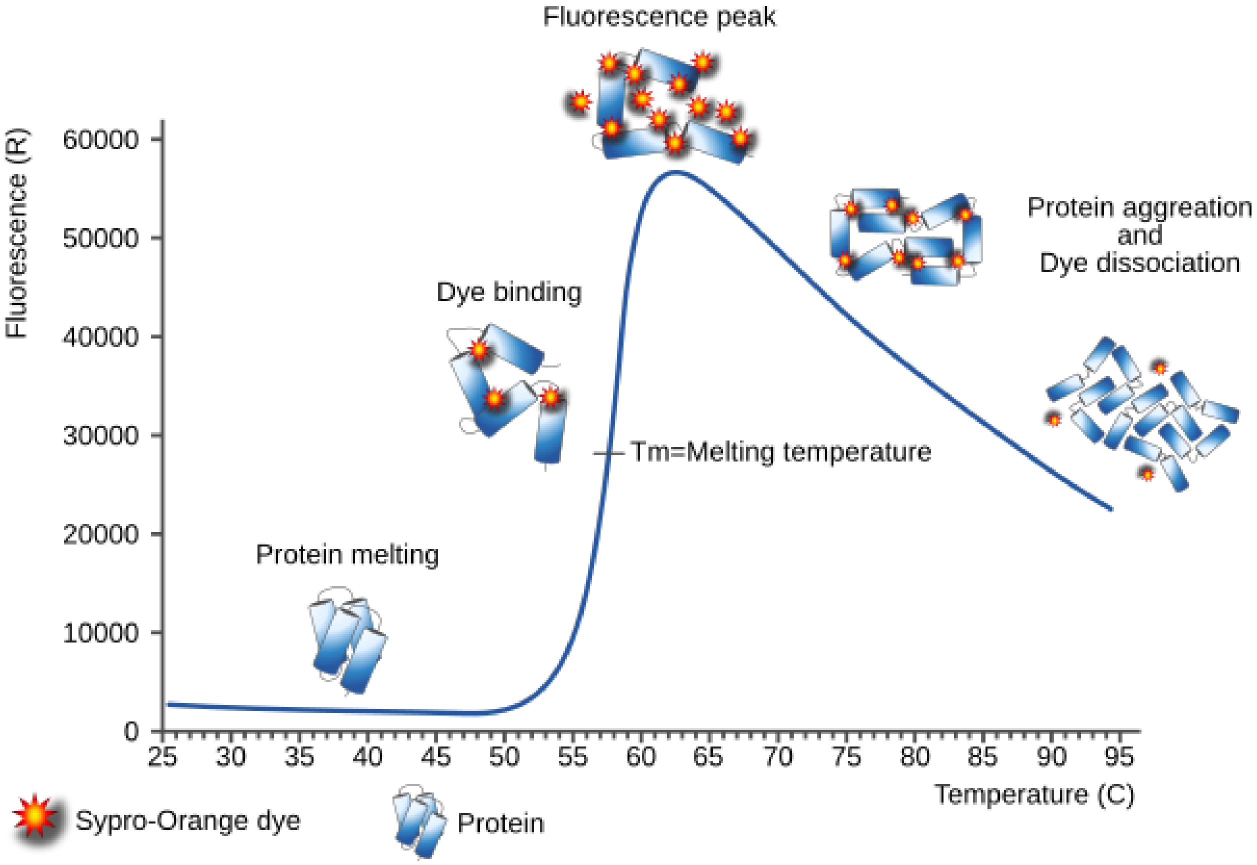


Figure 4.1. Schematic of the Differential scanning fluorimetry were there is a recording of fluorescence intensity versus temperature for the unfolding of a protein in the presence of SYPRO orange. (Goran tek-en, 2015)

4.3 Materials and Methods

4.3.1 DNA Preparation

Site directed mutagenesis was performed using the KAPA HIFI PCR kit (Kapabiosystems, Boston, USA). The histidine mutant plasmids were constructed by introducing individual histidine to lysine/aspartate/alanine substitution mutations to the CynD$_{pum}$ wild-type plasmid (Mulelu, 2013). The histidine substitution mutation were introduced to the CynD$_{pum}$ wild-type template plasmid, consisting of the CynD$_{pum}$ wild-type sequence inserted into a pet26b vector (Novagen, USA) with kanamycin resistance, by using phosphorylated non-overlapping mutagenic primers listed in Table 4.1. The non-overlapping mutagenic primer PCR strategy outlined in Fig. 4.2. was used to construct all the histidine substitution mutants except the H241K mutant. The construction of the H241K mutant plasmid using the non-overlapping mutagenic primer PCR strategy was unsuccessful and resulted in the use of the overlapping mutagenic primer PCR strategy outlined in chapter II (Fig. 2.2).

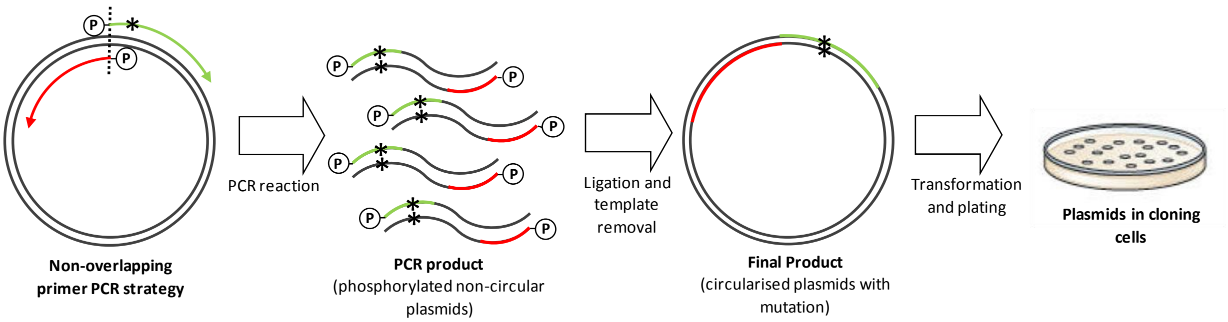


Figure 4.2. Schematic diagram of the non-overlapping mutagenic primer PCR strategy.

Table 4.2. Primer sequences used in site-directed mutagenesis of the histidine substitution mutants. Primers written as 5’–3’. The underlined nucleotides correspond to the substituted bases.

Primer	Sequence
H62A	GTTTGCTTTTATTGGTGCTCCAGAATATACGAG
	CAAGGATAACCAGGTAAAAATGCTTCTGGG
H62K	GTTTGCTTTTATTGGTAAGCCAGAATATACGAG
	CAAGGATAACCAGGTAAAAATGCTTCTGGG
H62D	GTTTGCTTTTATTGGTGATCCAGAATATACGAG
	CAAGGATAACCAGGTAAAAATGCTTCTGGG
H71A	GAAAGTTCTATGCTGAATTATATAAAAATGCCG
	TCGTATATTCTGGATGACCAATAAAAGCAAAC
H71K	GAAAGTTCTATAAGGAATTATATAAAAATGCCG
	TCGTATATTCTGGATGACCAATAAAAGCAAAC
H71D	GAAAGTTCTATGATGAATTATATAAAAATGCCG
	TCGTATATTCTGGATGACCAATAAAAGCAAAC
H128A	GATTTAATAGGAAAAGCTCGGAAAATGAGAGC
	CCCATTAGGATTAAACCAAAGCTGAGC
H128K	GATTTAATAGGAAAAAAGCGGAAAATGAGAGC
	CCCATTAGGATTAAACCAAAGCTGAGC
H128D	GATTTAATAGGAAAAGATCGGAAAATGAGAGC
	CCCATTAGGATTAAACCAAAGCTGAGC
H167A	GTGCTGGGAGGCTCAAGTCCC
	ATCAATCCGCCAAGGTTTCCAATTTC
H167K	GTGCTGGGAGAAGCAAGTCCC
	ATCAATCCGCCAAGGTTTCCAATTTC
H167D	GTGCTGGGAGGATCAAGTCCC
	ATCAATCCGCCAAGGTTTCCAATTTC
H184A	CAAAATGAGCAGGTAGCTGTAGCCTCTTG
	GGCATTCATCGCCATAAGATCAAGTG
H184K	CAAAATGAGCAGGTAAAGGTAGCCTCTTG
	GGCATTCATCGCCATAAGATCAAGTG
H184D	CAAAATGAGCAGGTAGATGTAGCCTCTTG
	GGCATTCATCGCCATAAGATCAAGTG
H241A	CATTTAAGAGCGGGGCTACGTGCATTTAC
	TTTCAAAGTAATCTCTTTGCTCCTGCG
H241K	CATTTAAGAGCGGGAAGACGTGCATTTAC
	TTTCAAAGTAATCTCTTTGCTCCTGCG
H241K *	CTTTGAAACATTTAAGAGCGGGAAGACGTGCATTTACG
	GTCCGGCCCGTAAATGCACGTCTTCCCGCTCTTAAATG
H241D	CATTTAAGAGCGGGGATACGTGCATTTAC
	TTTCAAAGTAATCTCTTTGCTCCTGCG
H285A	GATCCGGCTGGAGCTTACTCCAA
	AATATAATACTTGTAATCAATGACTCTTTCTAC
H285K	GATCCGGCTGGAAAGTACTCCAA
	AATATAATACTTGTAATCAATGACTCTTTCTAC
H285D	GATCCGGCTGGAGATTACTCCAA
	AATATAATACTTGTAATCAATGACTCTTTCTAC

*Second set of H241K non-phosphorylated overlapping mutagenic primers used after the unsuccessful non-overlapping mutagenic primers

PCR reactions were carried out in a 50 µl reaction mixture which consisted of 0.3 µM forward primer, 0.3 µM reverse primer, 1x KAPA HiFi Buffer, 0.3 mM KAPA dNTP Mix , 1 U of KAPA HiFi DNA Polymerase and 1 ng of the CynD$_{\mathrm{pum}}$ wild-type template plasmid. The PCR reaction was run with the cycling conditions listed in Table 4.2. Depending on the success of the initial PCR reactions, variable annealing temperature were used. PCR products were then subjected to ligation using T4 DNA ligase (Thermo Fisher scientific, USA) in order to circularise the phosphorylated PCR product and 100 ng of the circularised PCR product was later used to transform chemically competent *Escherichia coli* (*E.coli*) DH5α cells using the protocol described in section 2.3.2. Overnight cell cultures of the transformed *E. coli* DHα were grown and plasmid DNA was extracted in order to validate the success of the mutagenesis by DNA sequencing.

Table 4.2. PCR cycling conditions used in the construction of the of the histidine substitution mutant plasmids.

Step	**Temperature**	**Duration**	**Cycles**
Initial denaturation	95°C	3 min	1
Denaturation	98°C	20 sec	20
Annealing	62°C (± 5°C)	20sec	
Extension	72°C	4 min	
Final extension	72°C	10 min	1

4.3.2 Transformation

Transformation of *E. coli* BL21 using the histidine mutant plasmids was carried out as described in section 2.3.2.

4.3.3 Solubility validation and optimisation

To test for solubility of each of the histidine mutant enzymes a single *E. coli* BL21 transformed colony was used to grow 5 ml of overnight starter culture which were then added to 25 ml of nutrient broth (25 µg/ ml kanamycin) and cultured at 37°C to an OD$_{600\ \mathrm{nm}}$ of 0.4–0.6. Expression was then induced by adding IPTG to the culture to a final concentration of 1 mM. After the addition of IPTG the culture was incubated overnight at 30°. The culture was then centrifuged at 4000 x g for 15 min and the cell pellet sonicated as described in section 2.3.3. After centrifugation of the sonicated sample, the soluble supernatant and the insoluble pellet were separated. The pellet was then resuspended using buffer of equal volume to the sonication supernatant and SDS-PAGE gels were run on both fractions as described in section 4.3.7. To optimize the solubility one of the histidine mutant with the least solubility (H241A) was expressed under different temperatures (25°C and 30°C) and IPTG concentrations (0.05 mM, 0.1 mM, 0.2 mM, 0.5mM) to obtain the optimal solubility conditions.

4.3.4 Protein Expression and cell lysis

Protein expression was carried out for all the histidine substitution mutants in 1 litre cultures regardless of their solubility. Expression and cell lysis were carried out as described in section 2.3.2 and 2.3.3.

4.3.5 Protein purification

Purification was carried out using a three-step purification process consisting of ammonium sulphate fractional precipitation, anion exchange, and gel filtration chromatography as described in section 2.3.4–2.3.6.

4.3.6 Protein concentration determination

The concentration of protein was determined using the Bradford assay (Bradford, 1976) as described in section 2.3.8.

4.3.7 SDS-PAGE

SDS-PAGE analysis were carried out as outlined in section 2.3.9. SDS gels were stained with Coomassie Brilliant Blue R-250. To increase stain sensitivity during solubility tests, SDS gels were stained using the Acqua Stain protein staining reagent (Bulldog Bio, USA).

4.3.8 Enzyme activity assay

Enzyme activity was measured using the picric acid assay (Fisher and Brown, 1952) as described in section 2.3.7.

4.3.9 Negative stain electron microscopy

Negative stain electron microscopy sample preparation and visualisation were carried out as described in section 2.3.10. The samples of purified soluble histidine CynD_{pum} mutants were visualized using negative stain electron microscopy at various pH to observe their oligomerisation at both pH 5.4 and pH 8. Micrographs were collected using the Tecnai T20 transmission electron microscope operated at 200 kV and equipped with a Gatan US2000 CCD camera (Gatan, USA). Micrographs were collected at 25500 x and 44000 x magnification. The particle average lengths and length ranges were calculated from 50 randomly selected oligomers per pH condition.

4.3.10 Differential Scanning Fluorimetry

Solutions containing 10 µl of SYPRO Orange (Sigma-Aldrich, USA) dilution and 30 µl of the CynD_{pum} sample both in the same buffer were made up in transparent 0.2 ml PCR tubes. Previously optimised concentrations of CynD_{pum} (0.1 mg/ml) and SYPRO Orange dye (5X) were used (Mulelu, 2013). The tubes were placed in a Corbett Research Rotor Gene 6000 Real Time PCR machine (Corbett Research Ltd, UK) where they were heated from 30°C to 95°C in increments of 0.5°C with a holding time of 30 sec, while the fluorescence changes were monitored and recorded for each tube with wavelengths of excitation and emission at 490 nm and 575 nm respectively. To obtain the temperature midpoint for the protein unfolding transition (T_m), the maximum value of the first derivative of the fluorescence versus temperature curve was calculated. Two separate sets of the DSF experiment were carried out for the different CynD_{pum} strains, with one set run in the presence of 50 mM citrate/Na_2HPO_4 buffer at pH 5.4 and the other in 50 mM Tris-HCl at pH 8.

4.4 Results

4.4.1 Solubility validation and optimisation

The construction of the 21 histidine substitution mutant plasmid was successful and confirmed by DNA sequencing. Before any purification was to be carried out, the solubility of the histidine substitution mutants was assessed by SDS-PAGE gels in which the presence of the protein of interest band (37 kDa) was compared in the soluble and insoluble lysate fractions after cell lysis of induced 25 ml cultures. The SDS-PAGE gels initially showed that only 6 of the 21 histidine mutants expressed as soluble protein namely H62K, H71A, H71D, H71K, H241K and H285A (Fig. 4.3). Initial experiments failed to express soluble protein and thus expression conditions were further optimized. To optimize the solubility one of the histidine mutant which showed the least solubility (H241A) was expressed under different temperatures and IPTG concentrations to obtain the optimal solubility conditions (Fig. 4.4). The results showed that the most optimal condition for solubility was when using 0.5 mM IPTG for induction and an expression temperature of 25°C. These conditions were later used in the expression of all the histidine substitution mutants regardless of their solubility.

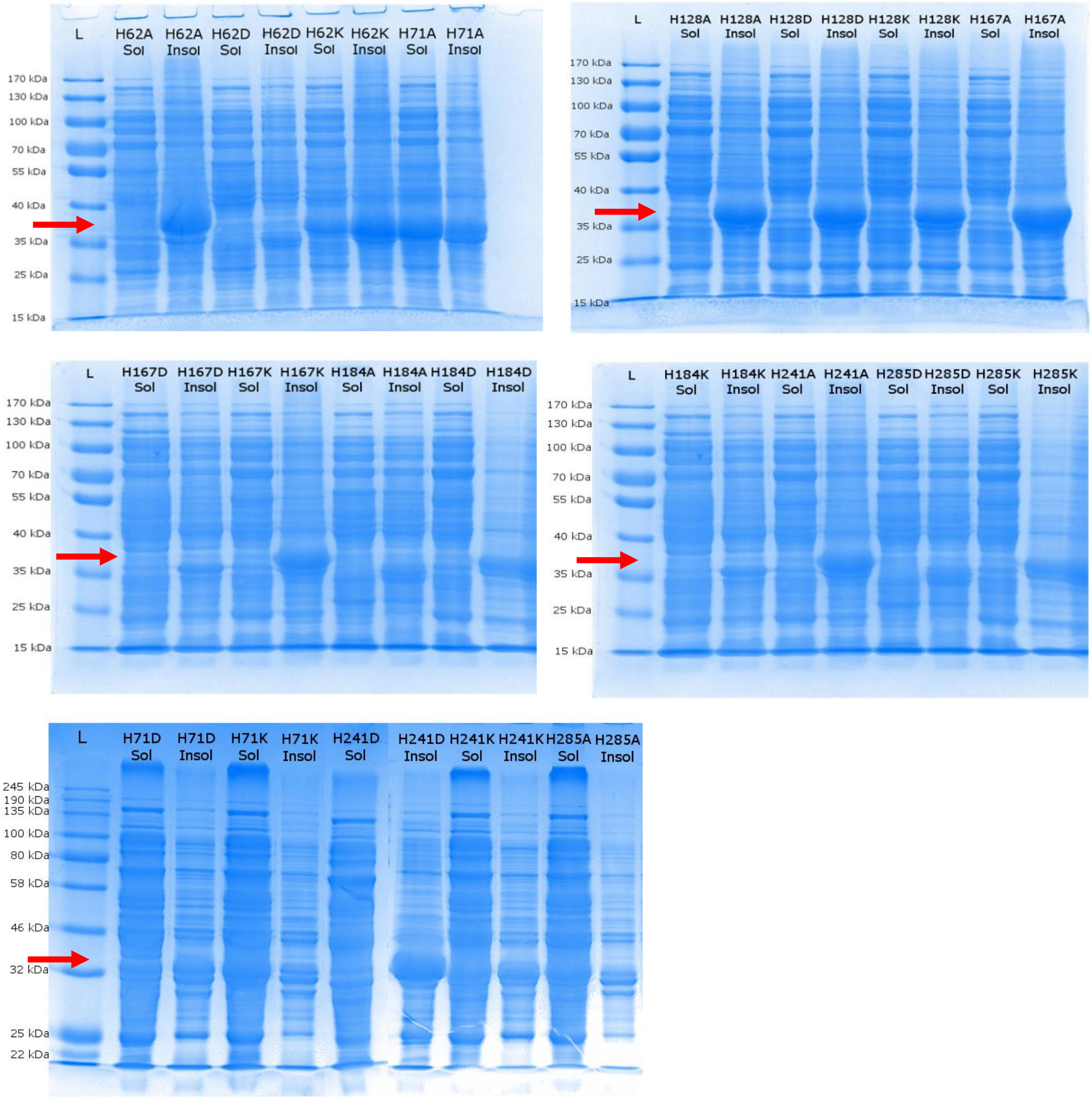


Figure 4.3. SDS-PAGE gels of insoluble and soluble lysate fractions of the histidine substitution mutants after cell lysis. The position corresponding to the size of CynD$_{pum}$ (37 kDA) is indicated by the red arrows. The sizes and positions of the protein markers are indicated on the left.

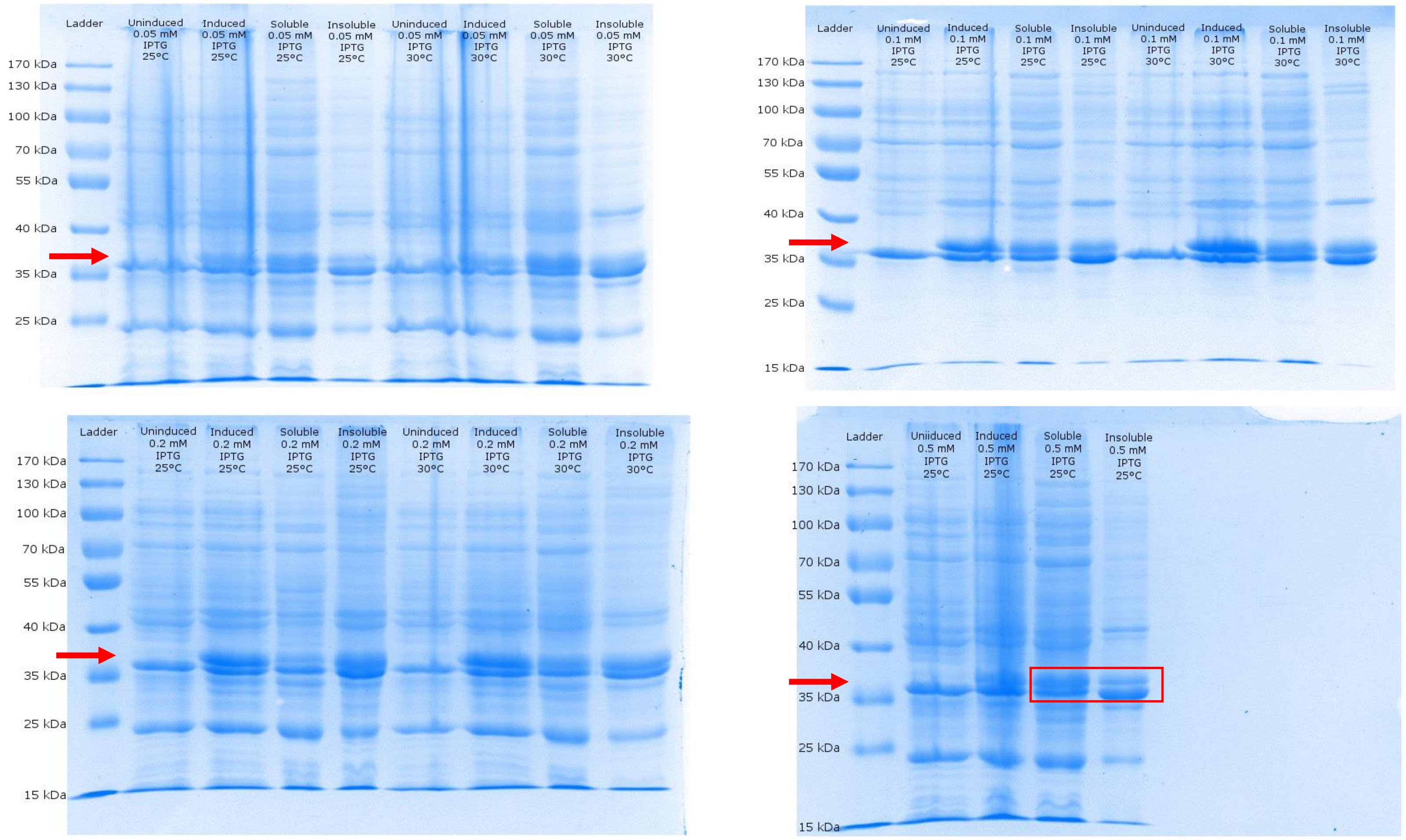


Figure 4.4. Solubility optimization SDS-PAGE gels of insoluble and soluble lysate fractions of the H241A mutant after expression at different temperatures and different IPTG concentration. The position corresponding to the size of CynD$_{pum}$ (37 kDA) is indicated by the red arrows. The sizes and positions of the protein markers are indicated on the left. The results showed that the most optimal condition for solubility were when using 0.5 mM IPTG for induction and an expression temperature of 25°C (red box).

4.4.2 Purification

After expression of all the soluble histidine mutants in litre cultures, the mutants were subjected to a three-step purification process consisting of ammonium sulphate fractional precipitation, anion exchange and gel filtration regardless of solubility. This large scale purification of the histidine mutants that demonstrated insoluble or partial solubility on the previous SDS-PAGE gels was carried out in an attempt to collect any small amounts of soluble protein that were not observed on the SDS-PAGE gels due to the sensitivity of the gel stain. All the histidine mutants precipitated out at the 40–50% ammonium sulphate saturation during the ammonium sulphate fractional precipitation. After anion exchange the presence of the protein of interest band (37 kDa) was checked using SDS-PAGE gels thereby also checking for solubility of any soluble histidine mutants that were initially not identified during the previous solubility assessment (Fig 4.5). In addition to the 6 soluble histidine mutants previously identified, 5 more histidine mutants were identified as soluble after expression in litre cultures, namely H128K, H184D, H241A, H285D and H285K. To increase purity and to separate different sized isoforms, the 11 soluble histidine mutants were further subjected to size exclusion chromatography. The gel filtration elution profiles of all the soluble histidine mutants except the H71D mutant showed a large peak around the elution volume of 8 ml, corresponding to fractions 50–55 (~670 kDa) of the elution run (Fig. 4.6–4.16) with the H62K, H128K and H285D mutants demonstrating other peaks corresponding to particles of smaller molecular size. Although the H184D, H241A, H241K, H285A and H285K also showed these small molecular size absorption peaks, these peaks did not contain protein that could be detected by either electron microscopy and the Bradford assay (results not shown). The H71D mutant eluted out as a single peak at an elution volume of 9 ml corresponding to fraction 57 (~410 kDa). Using the large gel filtration fraction samples, the purity of the mutants was assessed using SDS-PAGE. A substantial 37 kDa band, corresponding to the CynD$_{\text{pum}}$ was present for all the soluble histidine mutant samples with a few smaller molecular weight bands present in some of the mutants (H128K, H184D, H285A, H285D and H285K)(Fig. 4.5).

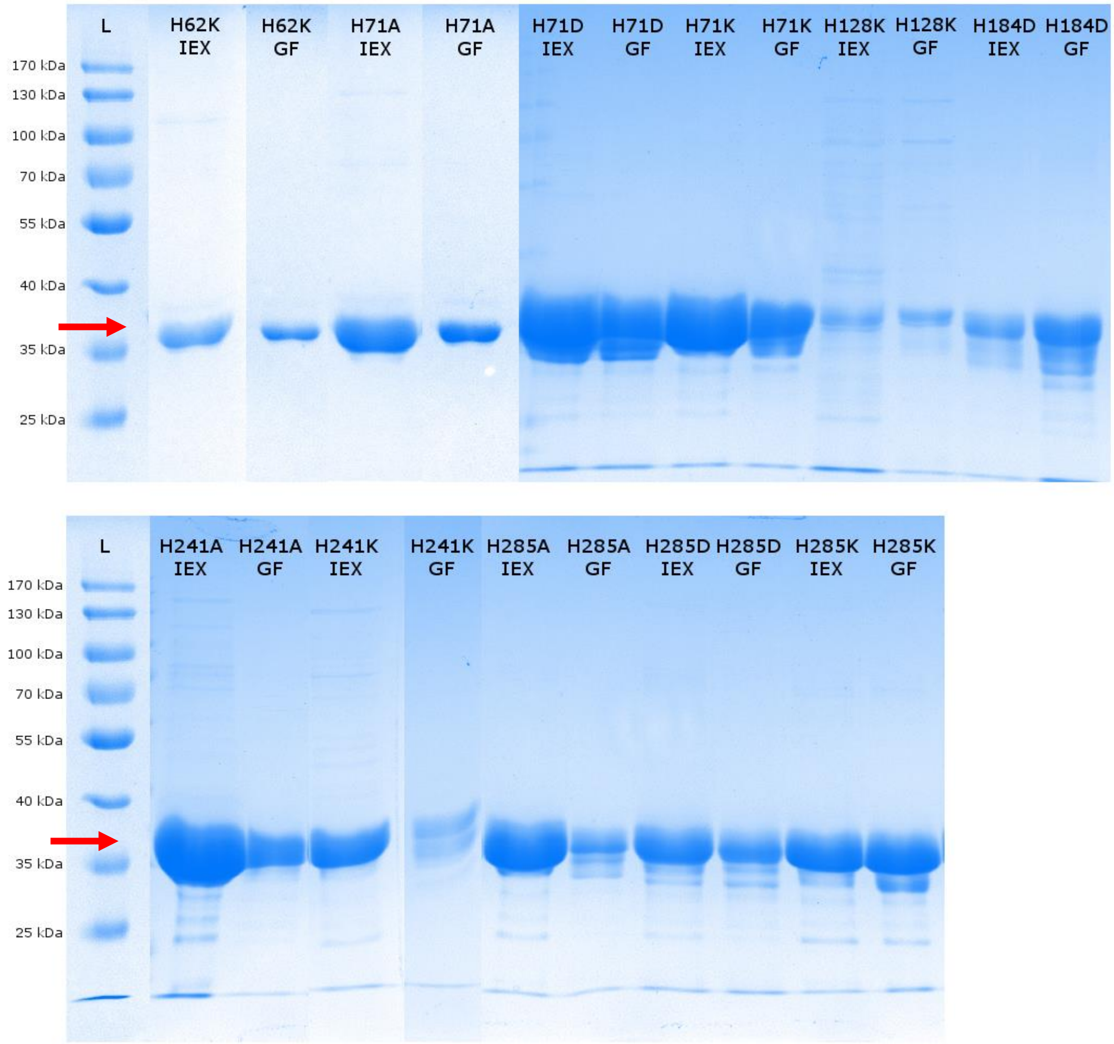
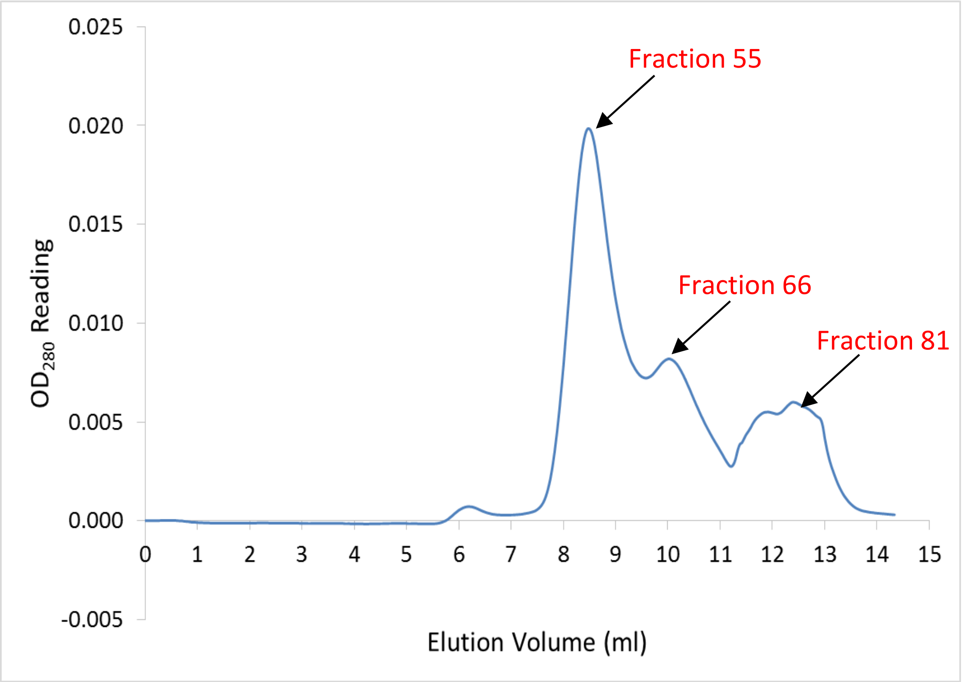


Figure 4.5. SDS-PAGE gels of anion exchange (IEX) chromatography and gel filtration (GF) fractions of the soluble histidine mutants that were successfully purified. The position corresponding to the size of CynD$_{pum}$ (37 kDA) is indicated by the red arrows. The sizes and positions of the protein markers are indicated on the left.

4.4.3 Negative stain electron microscopy

After purification the 11 soluble histidine CynD$_{pum}$ mutants were visualized using the Tecnai T20 transmission electron microscope to observe their oligomeric state at acidic and alkali pH (Fig. 4.6–4.16). At pH 8, all the soluble histidine mutants existed as short spirals of similar length to those of the wild-type, with lengths ranging from 12–42 nm, except the H285K mutant which had shorter spirals with lengths ranging from 5–11 nm (Table 4.3). All the soluble histidine mutants displayed an ability to associate from pH 8 short spirals to long fibres at pH 5.4 similar to the wild-type, with fibre lengths ranging from 29–239 nm, except the H241K mutant which formed slightly shorter fibres than the wild-type (Fig. 4.13) and the H285K mutant which formed very short filaments that are marginally classified as fibres (Fig. 4.16). The length of the short pH 5.4 fibres formed by the H241K mutant ranged from 11–67 nm while those formed by the H285K mutant ranged from 12–40 nm (Table 4.3). H128K and H285D formed significantly longer (but not straight) fibres than those observed for the wild-type at pH 5.4 (Fig 4.10 and 4.15) while the H71D mutant formed fibres which were slightly longer than the wild-type pH 5.4 fibres (Fig 4.8). The length of the long pH 5.4 fibres formed by the H128K and H285D mutants ranged from 105–853 nm and 51–658 nm, respectively, while those formed by the H71D mutant ranged from 61–357 nm (Table 4.3).



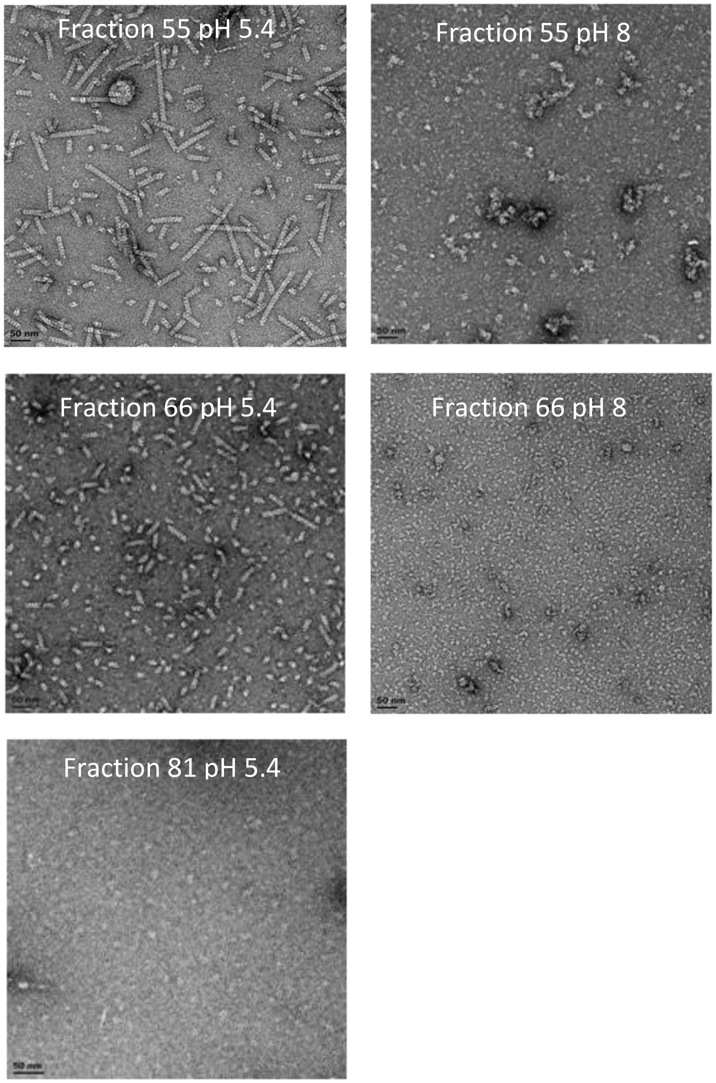
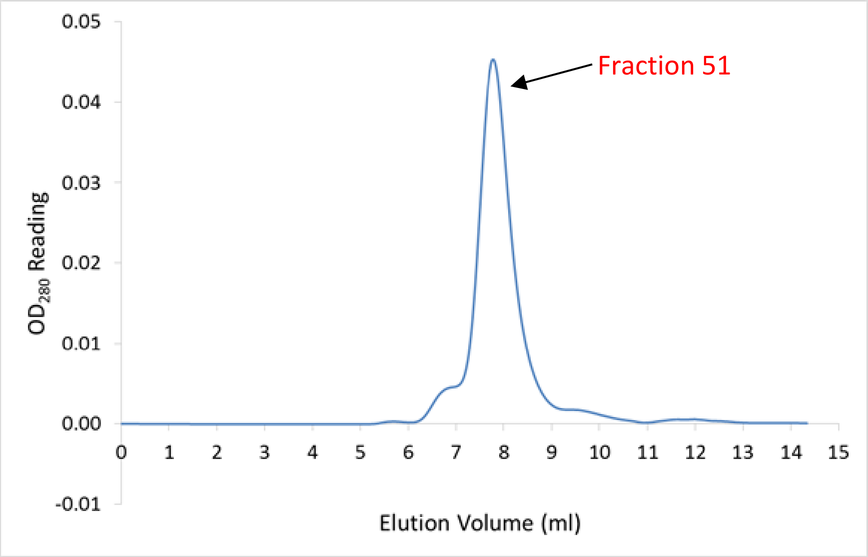


Figure 4.6. Oligomeric structure of the H62K mutant at pH 5.4 and 8. This figure shows the Transmission Electron Microscopy micrographs of samples corresponding to the gel filtration elution peaks of H62K in pH 5.4 or pH 8 buffers.



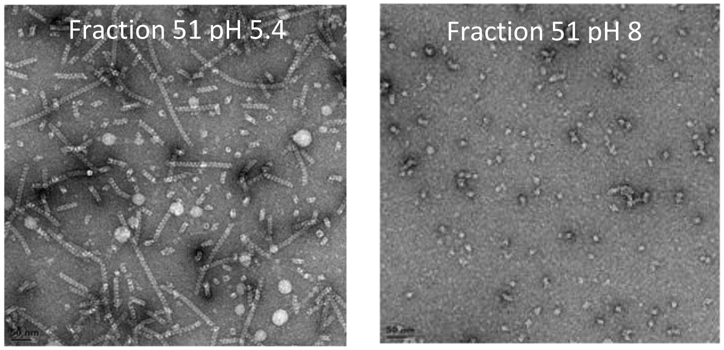
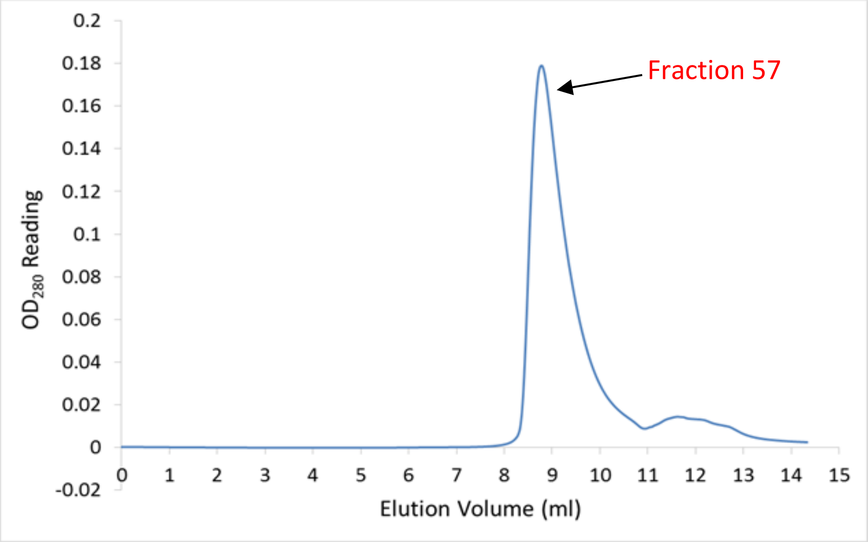


Figure 4.7. Oligomeric structure of the H71A mutant at pH 5.4 and 8. This figure shows the Transmission Electron Microscopy micrographs of samples corresponding to the gel filtration elution peaks of H71A in pH 5.4 or pH 8 buffers.



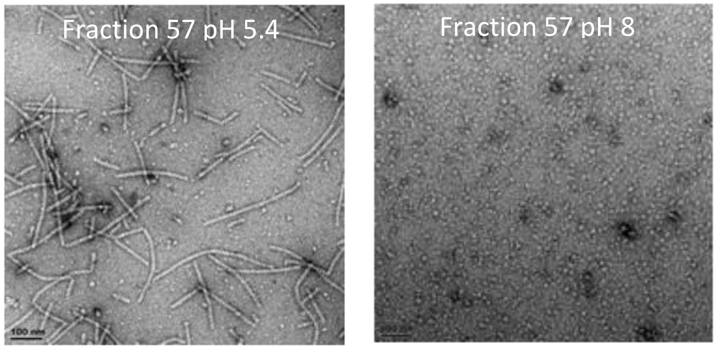


Figure 4.8. Oligomeric structure of the H71D mutant at pH 5.4 and 8. This figure shows the Transmission Electron Microscopy micrographs of samples corresponding to the gel filtration elution peaks of H71D in pH 5.4 or pH 8 buffers.

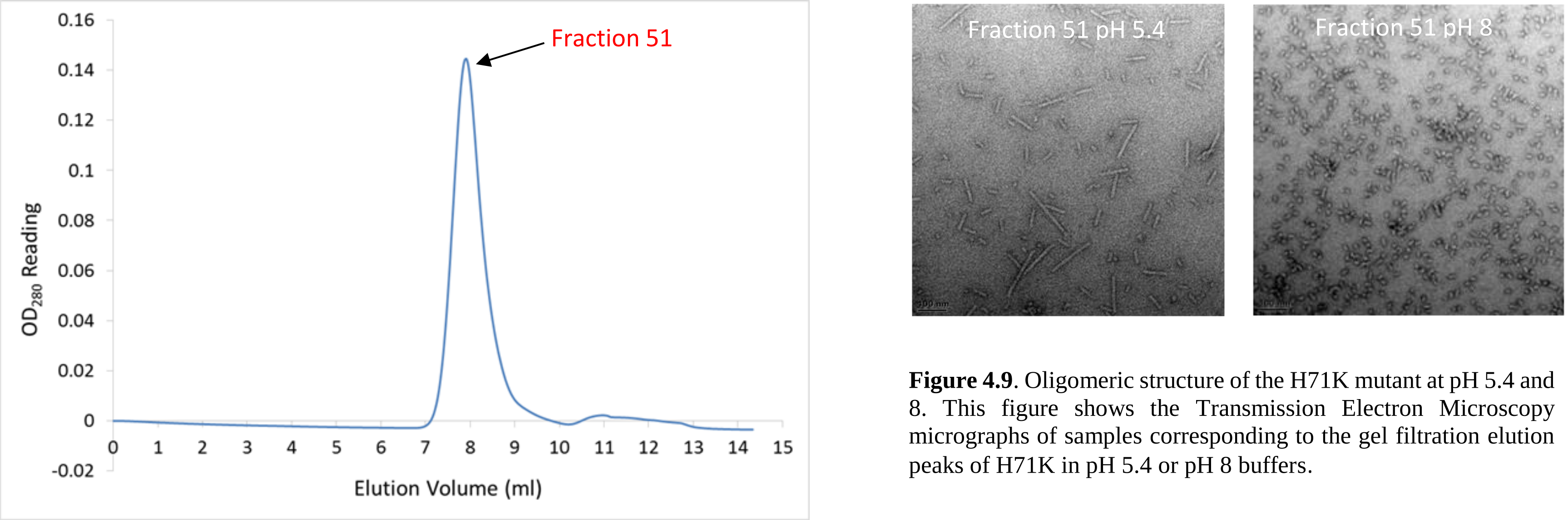
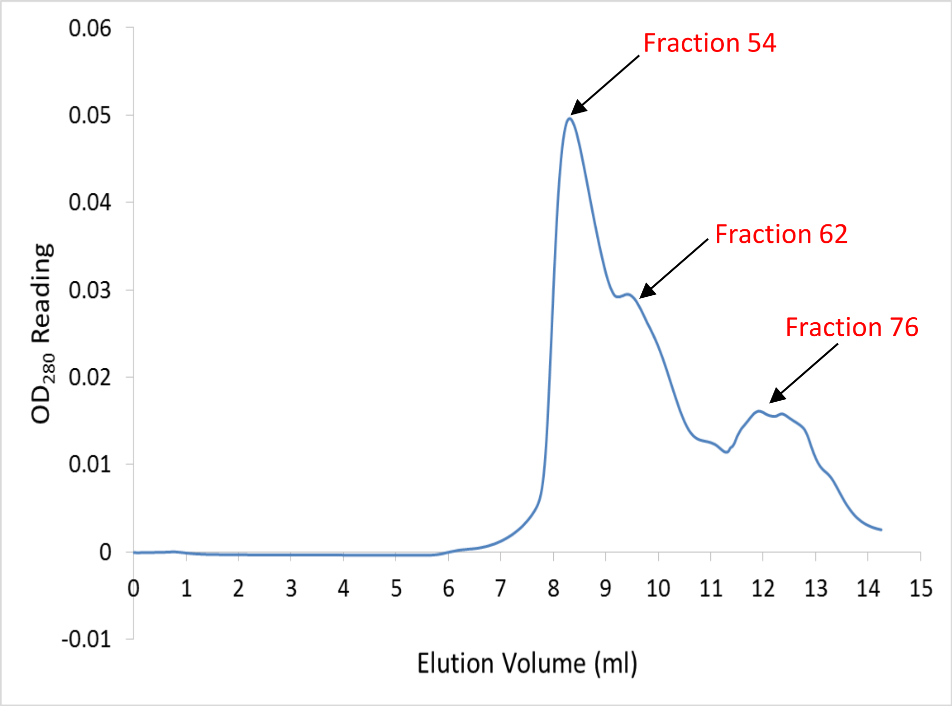


Figure 4.9. Oligomeric structure of the H71K mutant at pH 5.4 and 8. This figure shows the Transmission Electron Microscopy micrographs of samples corresponding to the gel filtration elution peaks of H71K in pH 5.4 or pH 8 buffers.



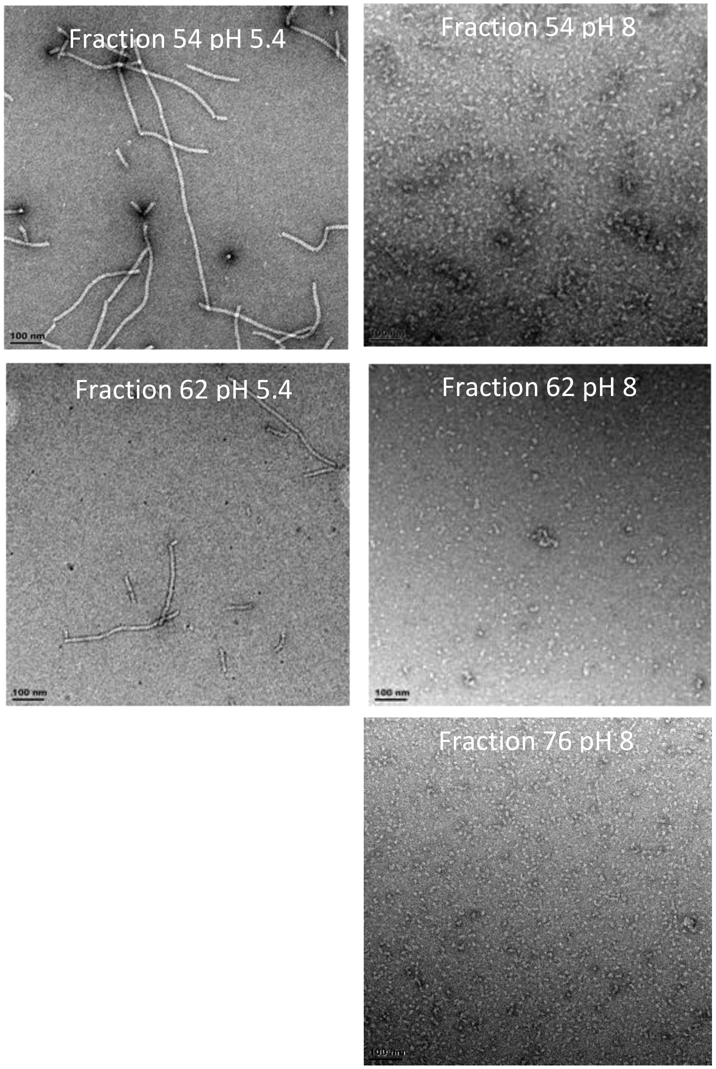
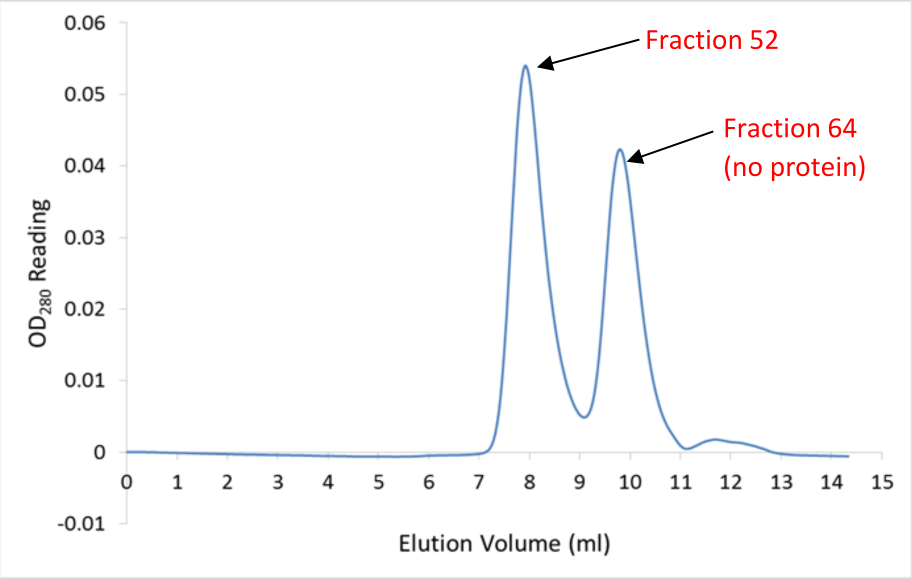


Figure 4.10. Oligomeric structure of the H128K mutant at pH 5.4 and 8. This figure shows the Transmission Electron Microscopy micrographs of samples corresponding to the gel filtration elution peaks of H128K in pH 5.4 or pH 8 buffers.



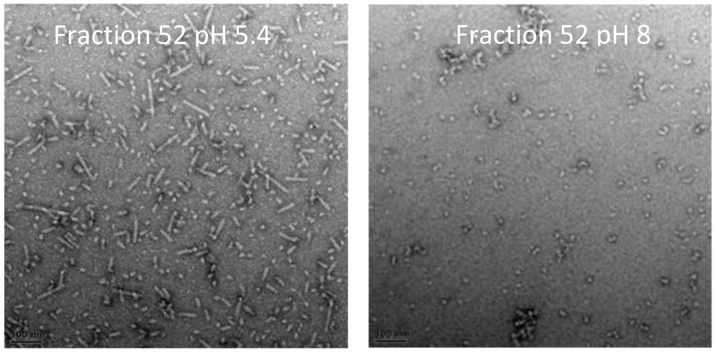
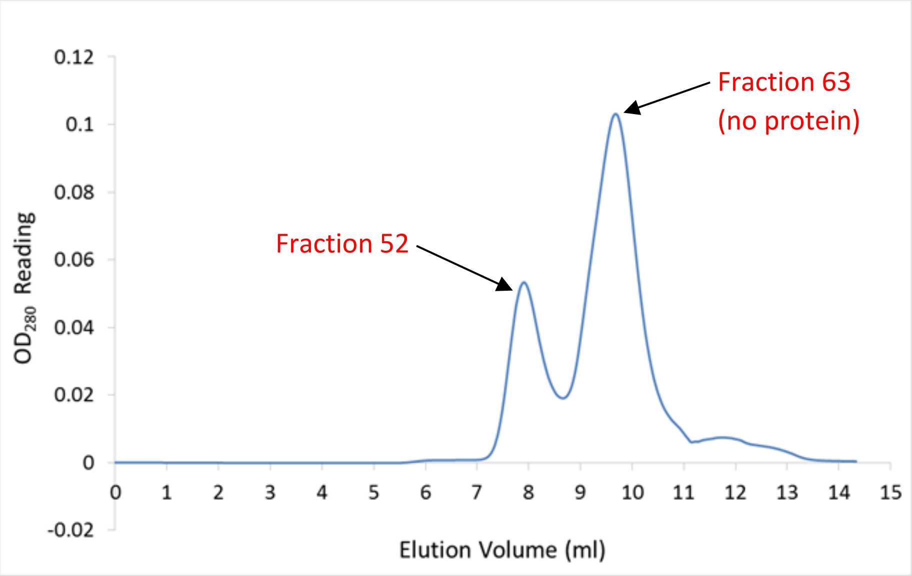


Figure 4.11. Oligomeric structure of the H184D mutant at pH 5.4 and 8. This figure shows the Transmission Electron Microscopy micrographs of samples corresponding to the gel filtration elution peaks of H184D in pH 5.4 or pH 8 buffers.



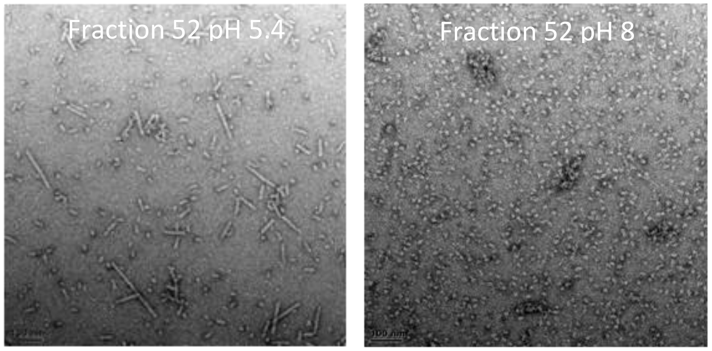
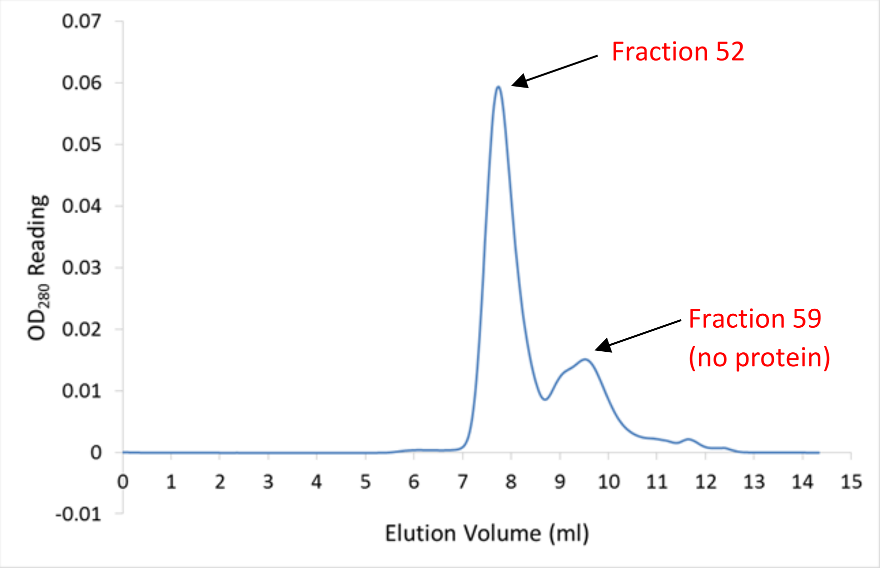


Figure 4.12. Oligomeric structure of the H241A mutant at pH 5.4 and 8. This figure shows the Transmission Electron Microscopy micrographs of samples corresponding to the gel filtration elution peaks of H241A in pH 5.4 or pH 8 buffers.



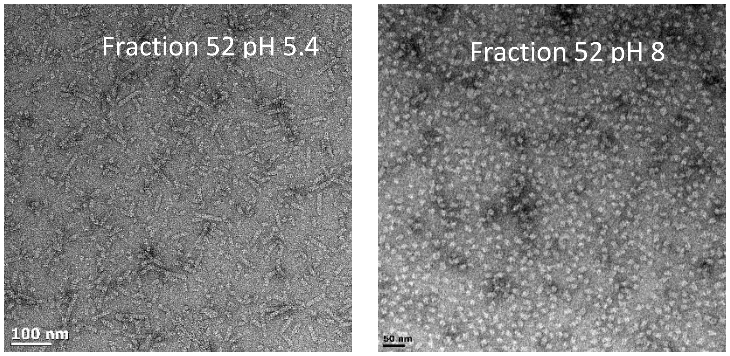
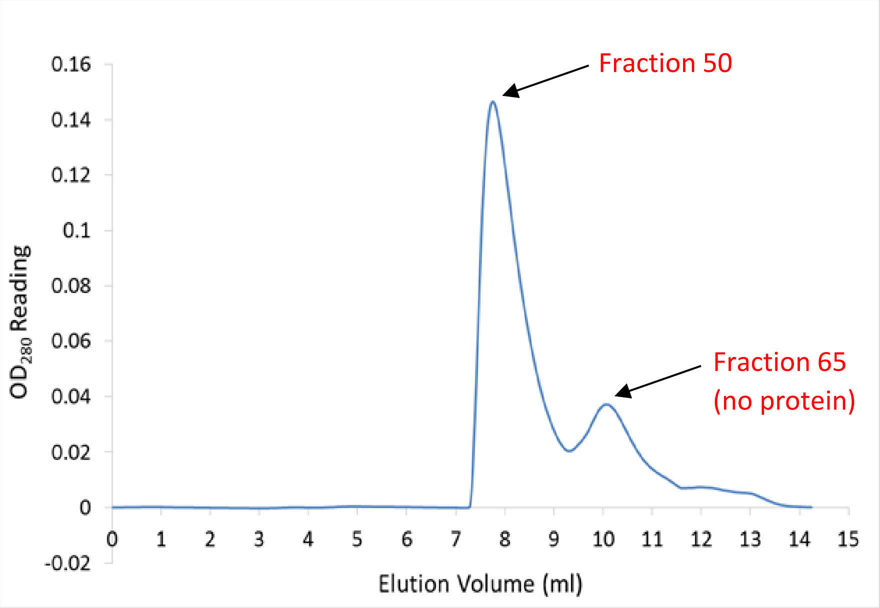


Figure 4.13. Oligomeric structure of the H241K mutant at pH 5.4 and 8. This figure shows the Transmission Electron Microscopy micrographs of samples corresponding to the gel filtration elution peaks of H241K in pH 5.4 or pH 8 buffers



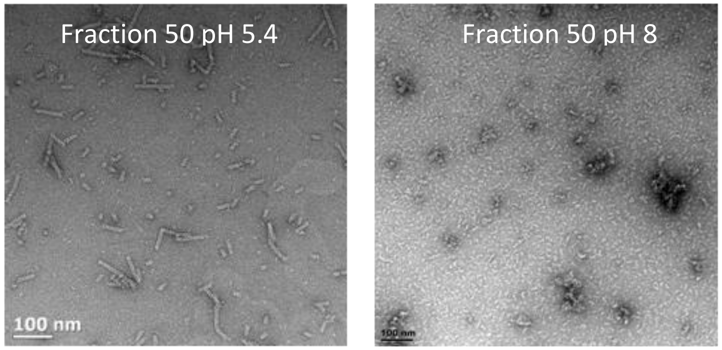


Figure 4.14. Oligomeric structure of the H285A mutant at pH 5.4 and 8. This figure shows the Transmission Electron Microscopy micrographs of samples corresponding to the gel filtration elution peaks of H285A in pH 5.4 or pH 8 buffers.

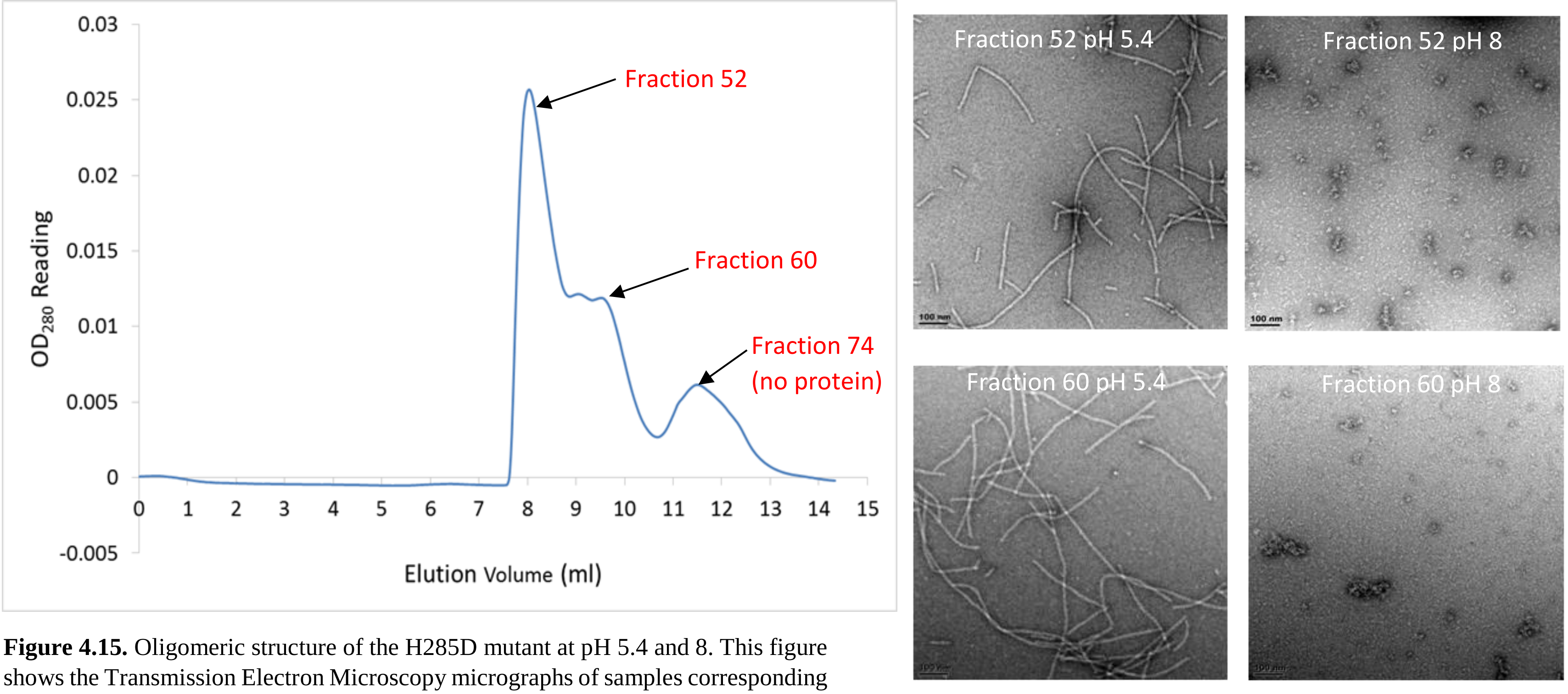


Figure 4.15. Oligomeric structure of the H285D mutant at pH 5.4 and 8. This figure shows the Transmission Electron Microscopy micrographs of samples corresponding to the gel filtration elution peaks of H285D in pH 5.4 or pH 8 buffers.

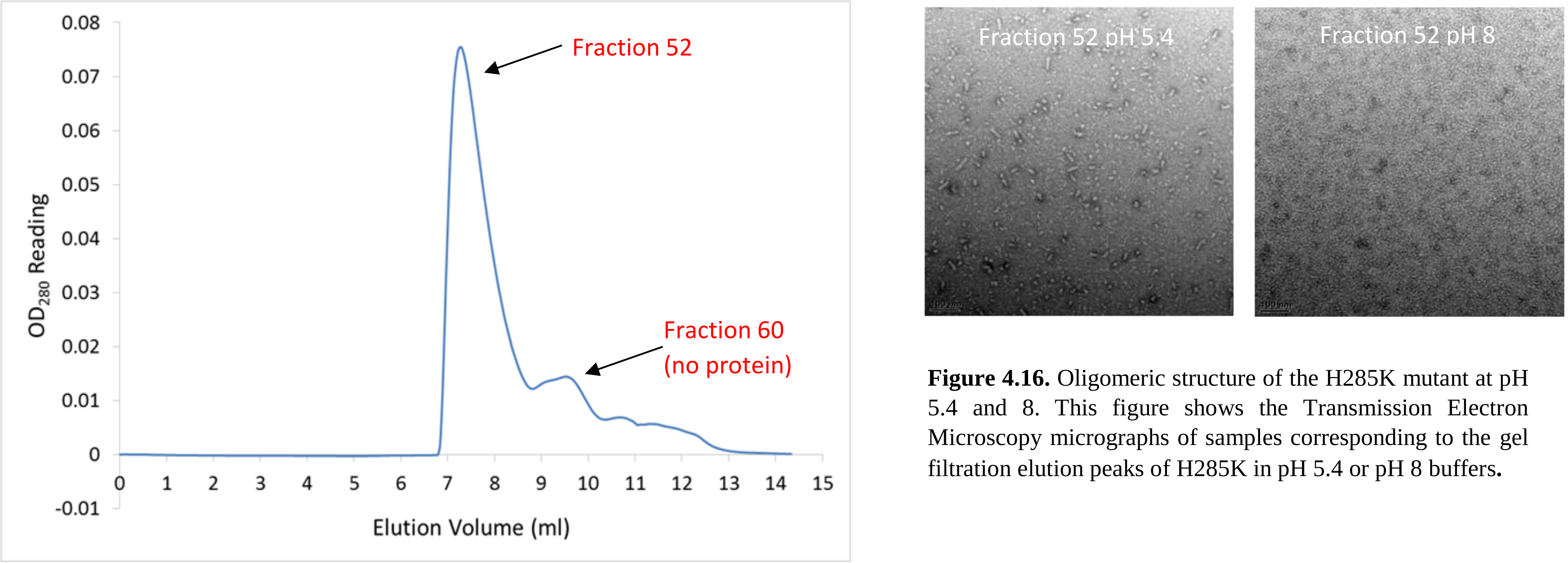


Figure 4.16. Oligomeric structure of the H285K mutant at pH 5.4 and 8. This figure shows the Transmission Electron Microscopy micrographs of samples corresponding to the gel filtration elution peaks of H285K in pH 5.4 or pH 8 buffers.

4.4.4 Differential Scanning Fluorimetry

DSF was carried out to measure the midpoint temperature of the protein–unfolding transition (T_m) of the $CynD_{pum}$ histidine substitution mutants, wild-type and the Q86R+H305K+H308K+H323K mutant in both pH 5.4 and pH 8 buffers. T_m was obtained by calculating the maximum value of the first derivative curve (Fig. 4.17B and Fig. 4.18B) of the fluorescence versus temperature curve (Fig. 4.17A and Fig. 4.18A). All the histidine mutants demonstrated an exponential increase of fluorescence as the temperature was increased except the H128K mutant which only showed a steady decrease fluorescence which may be due to early destabilization of the enzyme thus making it difficult to measure the stability of this mutant. The T_m of the mutant enzymes were calculated and they displayed a general decrease as the pH changed from pH 5.4 to pH 8 (Fig. 4.19A). The T_m values of the mutants were then compared with that of the wild-type (Fig. 4.19B). At pH 5.4 all the histidine mutants demonstrated similar T_m to that of the wild-type while the Q86R+H305K+H308K+H323K mutant showed a large increase in T_m compared to the wild-type. At pH 8 all the histidine mutants demonstrated similar T_m to that of the wild-type with the H71K showing a slight increase relative to the wild-type. The Q86R+H305K+H308K+H323K mutant showed a significant increase in T_m relative to the wild-type and all the histidine mutants at both pH 5.4 and 8. The Q86R+H305K+H308K+H323K (pH 5.4 $T_m = 68.6$; pH 8 $T_m = 64.7$) mutant demonstrated a higher thermostability than its parent variants Q86R (pH 5.4 $T_m = 66.3$; pH 8 $T_m = 61.0$) and H305K+H308K+H323K (pH 5.4 $T_m = 65.3$; pH 8 $T_m = 61.0$)(Mulelu, 2013).

4.4.5 Activity

Cyanide degrading activity was tested on all the purified and soluble histidine substitution mutants. All the histidine mutations had cyanide degrading activity except H71D, H285A, H285D and H285K (Table 4.3). Interestingly some of the soluble mutants that formed fibres at pH 5.4 showed no cyanide degrading activity namely the H71D, H285A, and H285D mutants.

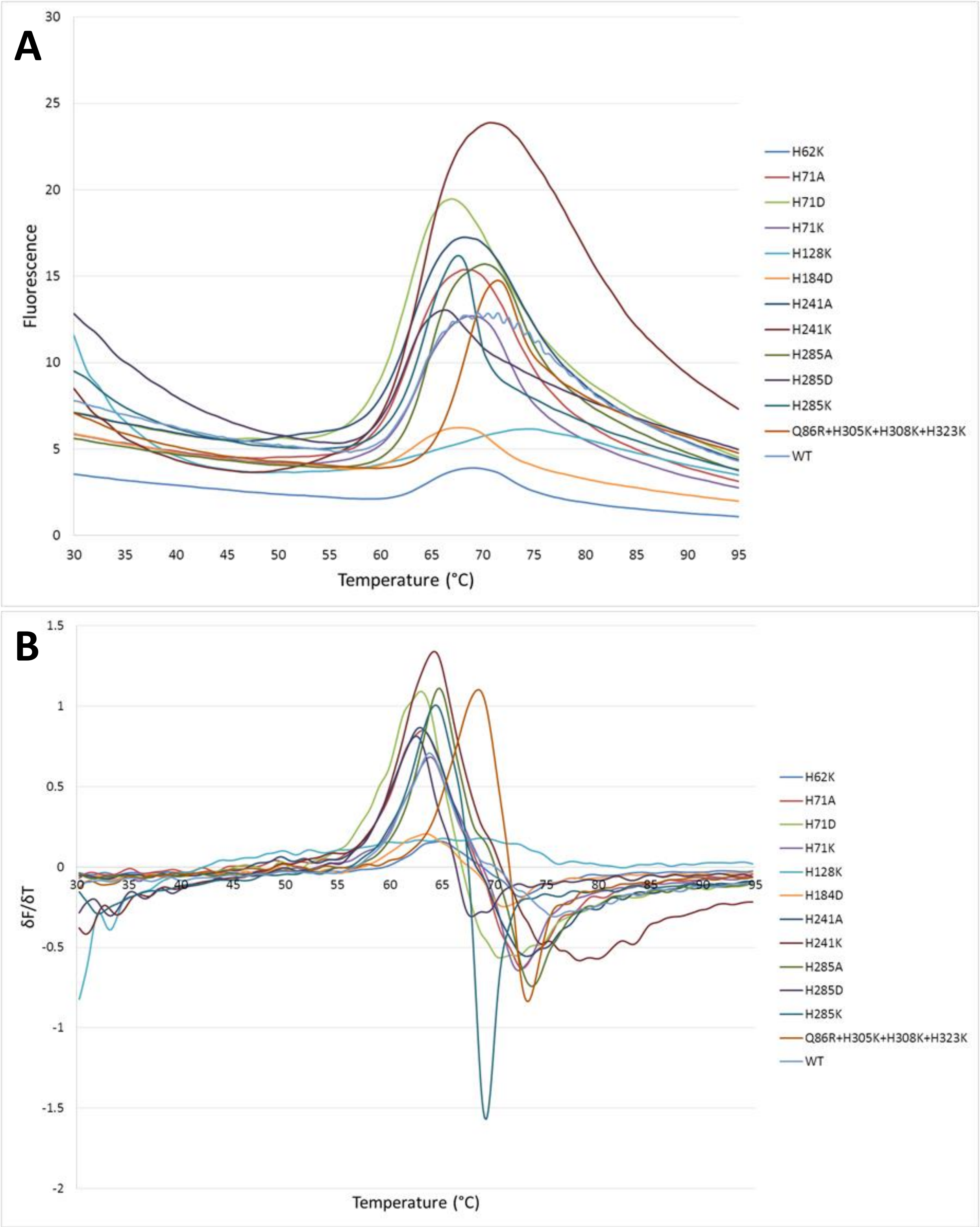


Figure 4.17. (A) DSF results showing the thermal denaturation of the CynD$_{pum}$ variants at pH 5.4. **(B)** The first derivative of graph (A) where maximum value determines T$_m$. The individual T$_m$ values were extrapolated and summarised in Fig. 4.19A.

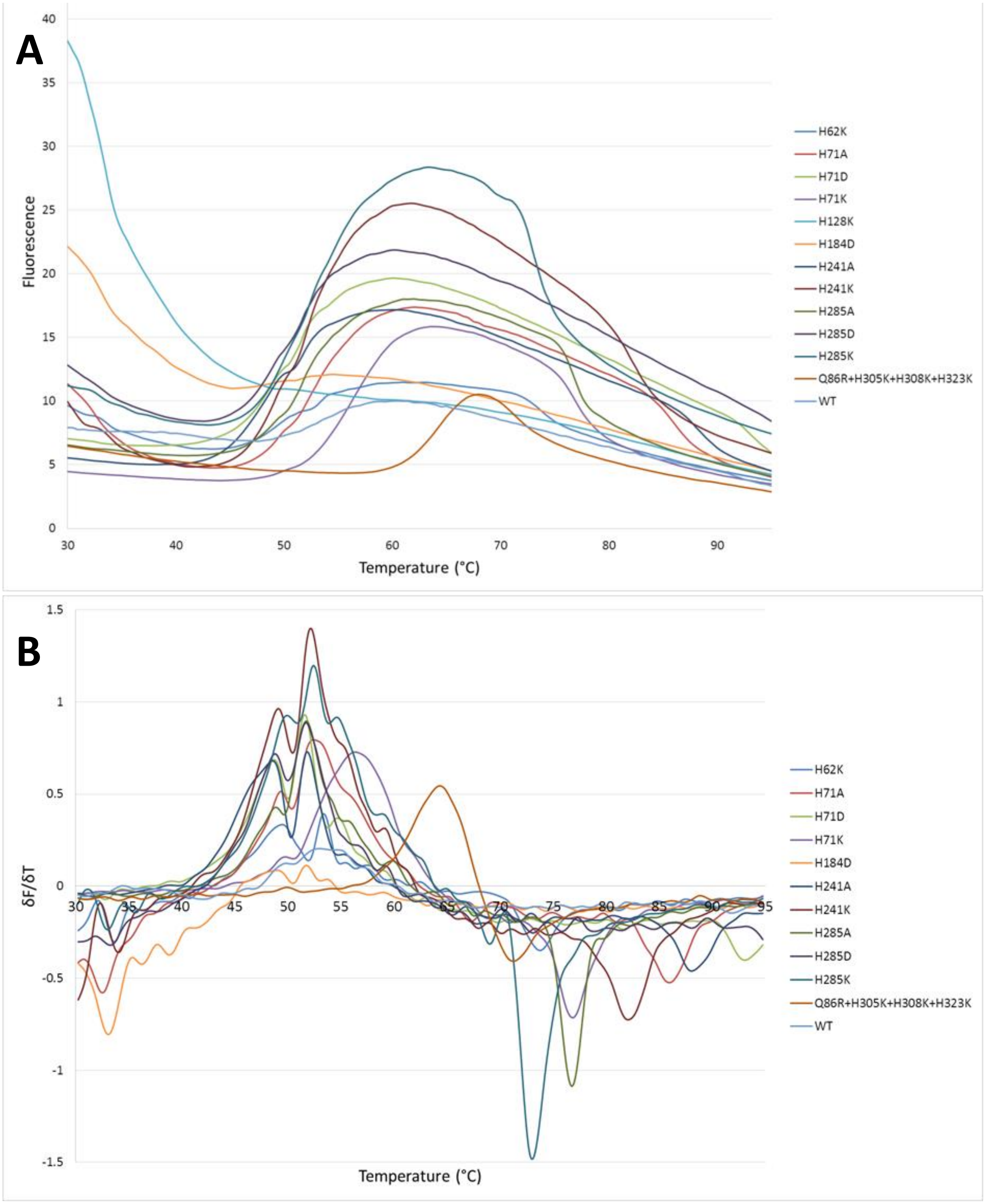


Figure 4.18. (A) DSF results showing the thermal denaturation of the CynD$_{pum}$ variants at pH 8. **(B)** The first derivative of graph (A) where maximum value determines T$_m$. The individual T$_m$ values were extrapolated and summarised in Fig. 4.19A.

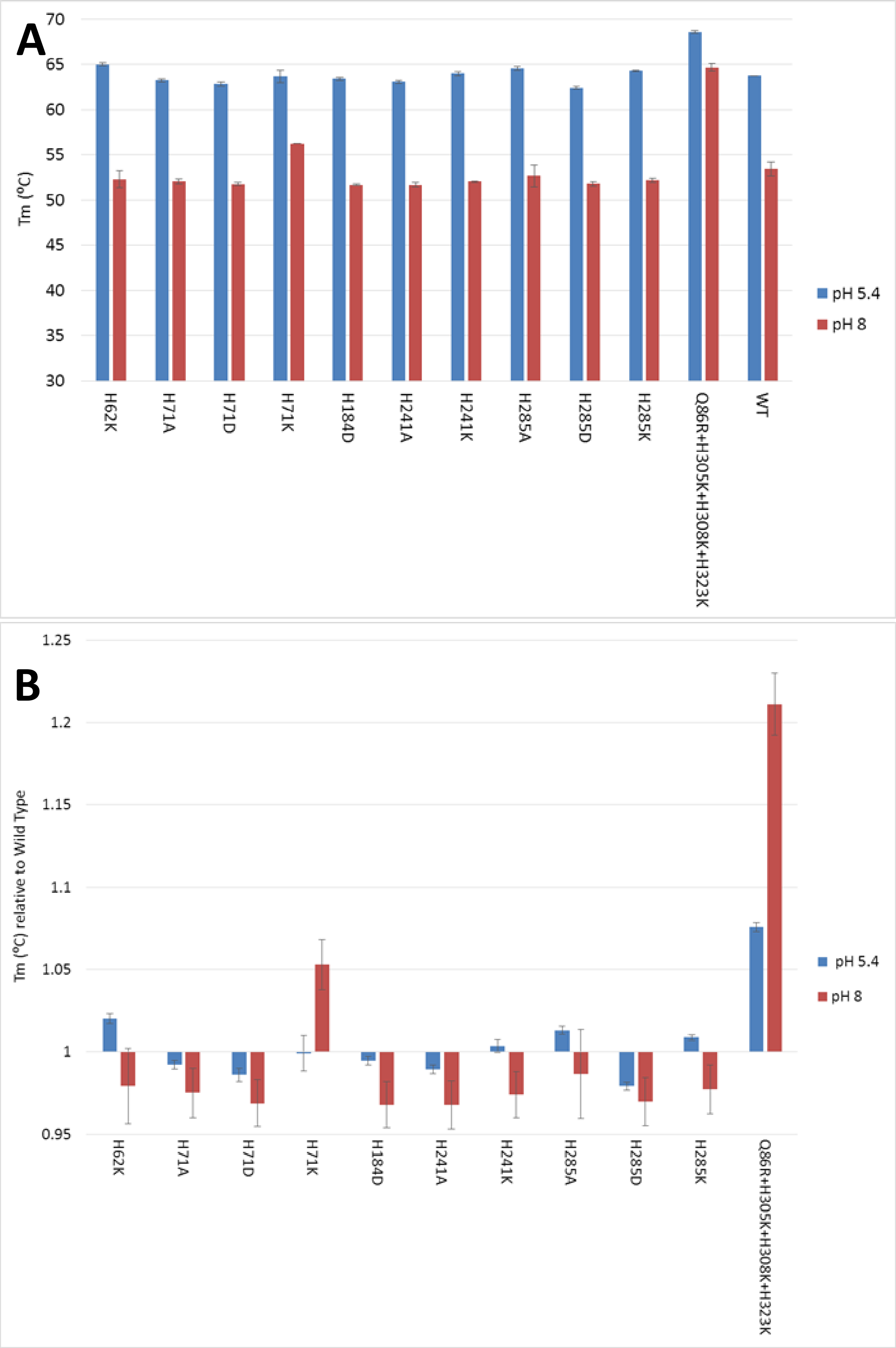


Figure 4.19. (A) Midpoint temperatures of the protein-unfolding transition (T_m) for the CynD$_{pum}$ wild-type and variants at pH 5.4 and pH 8. **(B)** The T_m of the mutants relative to the wild-type at pH 5.4 and pH 8.

Table 4.3: Summary of the characterized histidine substitution mutants. The table shows a summary of the solubility, oligomerisation, activity and thermostability of each of the histidine substitution mutants.

Enzyme variant	**H62A**	**H62D**	**H62K**	**H71A**	**H71D**	**H71K**	**H128A**	**H128D**	**H128K**	**H167A**	**H167D**
Soluble	No	No	Yes	Yes	Yes	Yes	No	No	Yes	No	No
Forms fibres at pH 5.4	n/a	n/a	Yes	Yes	Yes	Yes	n/a	n/a	Yes	n/a	n/a
Exists as short spirals at pH 8	n/a	n/a	Yes	Yes	Yes	Yes	n/a	n/a	Yes	n/a	n/a
Average length at pH 5.4 (nm) (Range)	n/a	n/a	94 (29-198)	128 (31-301)	186 (61-357)	101 (31-242)	n/a	n/a	286 (105-853)	n/a	n/a
Average length at pH 8 (nm) (Range)	n/a	n/a	17 (9-34)	16 (10-27)	12 (8-17)	18 (11-27)	n/a	n/a	11 (8-15)	n/a	n/a
Active	n/a	n/a	Yes	Yes	No	Yes	n/a	n/a	Yes	n/a	n/a
T_m at pH 5.4 (°C)	n/a	n/a	65.0	63.3	62.9	63.7	n/a	n/a	n/a	n/a	n/a
T_m at pH 8 (°C)	n/a	n/a	52.3	52.2	51.8	56.2	n/a	n/a	n/a	n/a	n/a

Enzyme variant	**H167K**	**H184A**	**H184D**	**H184K**	**H241A**	**H241D**	**H241K**	**H285A**	**H285D**	**H285K**	**WT**
Soluble	No	No	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Yes
Forms fibres at pH 5.4	n/a	n/a	Yes	n/a	Yes	n/a	Yes	Yes	Yes	No	Yes
Exists as short spirals at pH 8	n/a	n/a	Yes	n/a	Yes	n/a	Yes	Yes	Yes	Yes	Yes
Average length at pH 5.4 (nm) (Range)	n/a	n/a	56 (12-165)	n/a	54 (25-113)	n/a	39 (11-67)	71 (13-231)	230 (51-658)	23 (12-40)	89 (29-239)
Average length at pH 8 (nm) (Range)	n/a	n/a	16 (11-24)	n/a	12 (9-19)	n/a	13 (9-22)	10 (8-15)	12 (9-17)	8 (5-11)	23 (12-42)
Active	n/a	n/a	Yes	n/a	Yes	n/a	Yes	No	No	No	Yes
T_m at pH 5.4 (°C)	n/a	n/a	63.4	n/a	63.1	n/a	64.0	64.6	62.4	64.3	63.8
T_m at pH 8 (°C)	n/a	n/a	51.7	n/a	51.7	n/a	52.0	52.7	51.8	52.2	53.4

4.5 Discussion

4.5.1. Solubility of histidine substitution mutants

Histidine single substitution mutants were constructed and the changes in oligomerisation behaviour at various pH were observed. The histidine substitution mutations introduced a variety of structural changes to the CynD$_{pum}$ enzyme. The most drastic of these structural changes was the rendering of the enzyme insoluble by introduction of these single residue mutations. The assessment of solubility was initially carried out in small volume cultures and later carried out in larger volume cultures after the determination of optimum solubility expression conditions. Some histidine mutants which were initially considered insoluble demonstrated solubility when larger volume cultures were used. This demonstrates that the optimised expression condition greatly improved the solubility of some of the histidine mutants. When comparing the initial expression conditions with the optimised conditions both the temperature and IPTG concentration were reduced after optimisation and this may have resulted in lower levels of expression of the enzymes thus preventing the common intracellular accumulation of the protein as insoluble aggregates of inclusion bodies, triggered by high levels of protein expression in *E. coli* cells.

Of the 21 histidine substitution mutants expressed, only 11 showed soluble expression thus demonstrating the importance of the CynD$_{pum}$ histidines in the correct folding of the enzymes. The atomic models of CynD$_{pum}$ obtained in chapter III suggest that a majority of the substitution mutations that rendered the enzyme insoluble are located in the core of each CynD$_{pum}$ subunit and these include H128A, H128D, H167A, H167D, H167K, H184A, H184K and H241D (Fig. 4.20–4.23). This was expected as the corresponding histidines are initially buried in a cluster of charged, polar or hydrophobic side chains and any changes in any of the side chains would potentially collapse the whole fold of the enzyme’s monomers resulting in insolubility. Based on the atomic structures generated in chapter III, His128 is postulated to be located in the vicinity of a Gln115, forming a hydrogen bond which may be required for the stability of the enzyme (Fig. 4.20). The formation of this hydrogen bond between the His128 and Gln115 may be crucial in the solubility of CynD$_{pum}$ as the substitution of the His128 with an uncharged residue (H128A) rendered the enzyme insoluble. Interestingly the substitution of His128 with a negatively charged residue (H128D) rendered the enzyme insoluble while the substitution with a positively charged residue (H128K) conserved the enzymes solubility. This was not expected as both the introduced Asp128 and Lys128 were initially proposed to be capable of hydrogen bonding with Gln115 due to its polar nature, thus demonstrating that other interactions around the His128 region besides the His128-Gln115 hydrogen bonding are responsible for maintaining the enzymes stability and solubility.

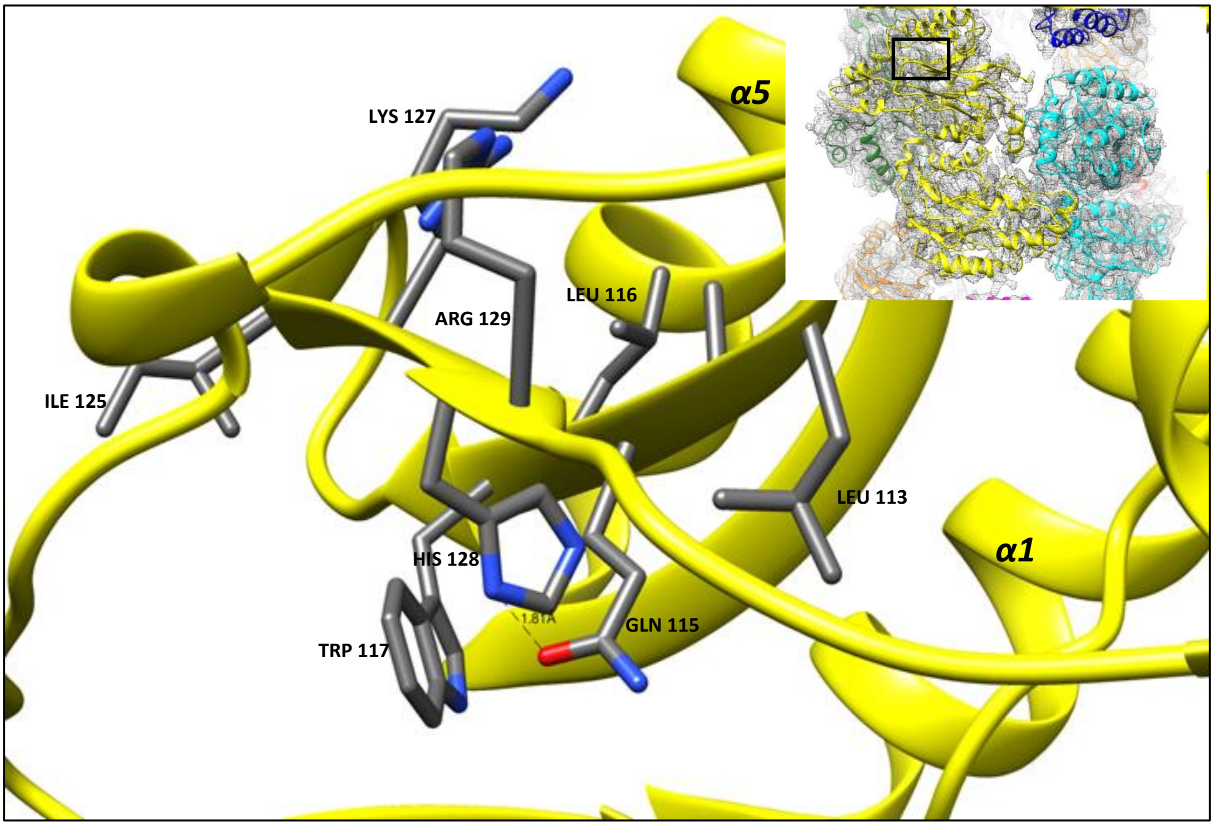


Figure 4.20. The cryo-EM map refined atomic model of the wild-type CynD$_{pum}$ showing the location of the His128 and its neighbouring residues. The His128-Gln115 hydrogen bonding is depicted by the dashed line with a bond distance of 1.81 Å.

The His167 residue is located on the conserved 3-10 helix that contains the active site Cys164 and in close proximity to His184 (Fig. 4.21). His167 and His184 interact to form a hydrogen bond which may be crucial in stabilising the enzyme and maintain its solubility as mutations of either histidine residues resulted in the insolubility of the enzyme. Substitution of the histidines resulted in the insolubility of the enzyme regardless of the charge of the substitution residue (H167A, H167D, H167K, H184A and H184K) except the H184D mutation which retained the enzymes solubility, thus demonstrating that the introduced Asp184 has the ability to maintain the hydrogen bond with His167 allowing for stabilisation of the enzyme. Interestingly the positions of His167 and His184 were shown to be conserved in the structure of the branch 1 nitrilase Nit6803 from *Synechocystis sp.* Strain PCC6803 (PDB ID: 3WUY), when the Nit6803 structure was structurally aligned and superimposed with the CynD$_{pum}$

structure (results not shown). The conservation of these histidines is also evident in the sequence alignment of CynD$_{pum}$ with other branch 1 nitrilases (Fig. 1.3), thus demonstrating the importance of the His167-His184 interaction in the stability and possibly in the positioning of the catalytic Cys164 of CynD$_{pum}$ and other related branch 1 nitrilases.

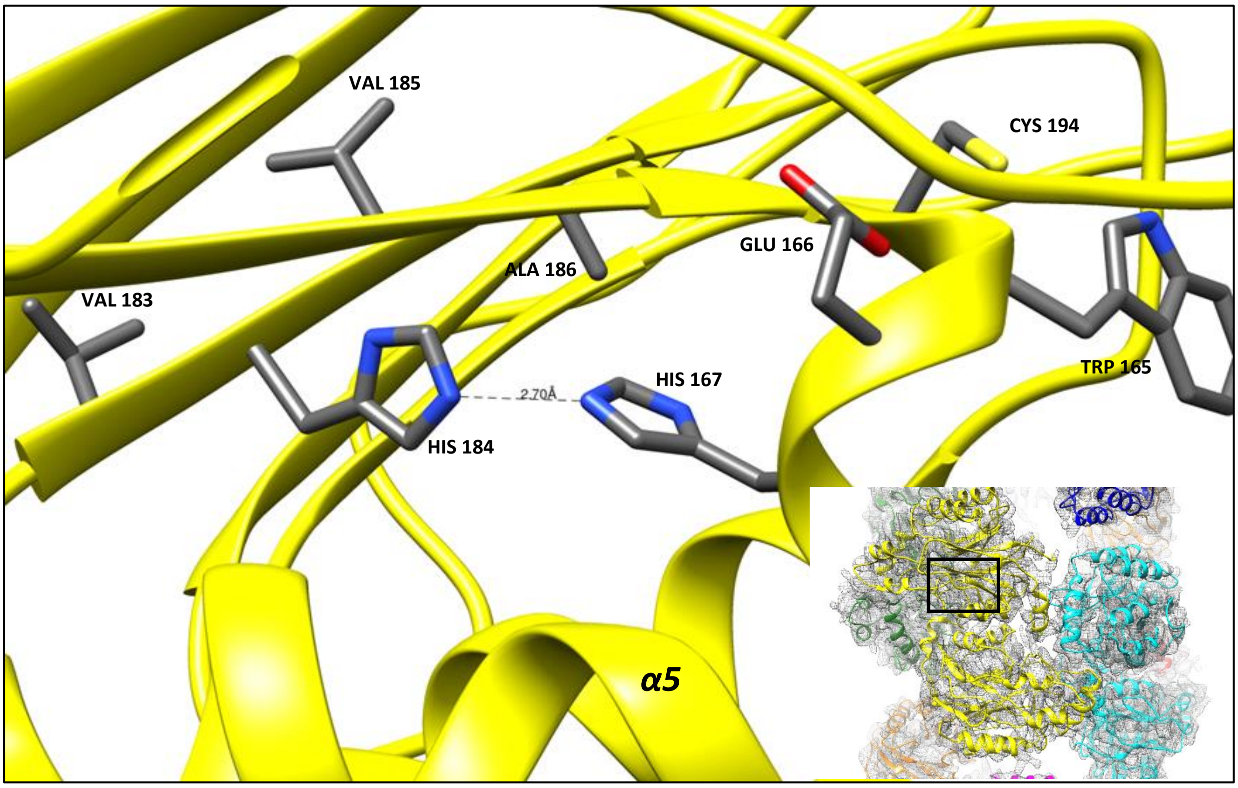


Figure 4.21. The cryo-EM map refined atomic model of the wild-type CynD$_{pum}$ showing the location of the His167, His184 and its neighbouring residues. The His167-His184 hydrogen bonding is depicted by the dashed line with a bond distance of 2.70 Å.

The His241 residue forms a hydrogen bond with a negatively charged Glu195 (Fig. 4.22). The His241-Gln195 hydrogen bond that is formed is not crucial in maintaining the enzymes solubility as removal of the pH 8 polar His241 by the H241A substitution mutation had little effect on the solubility of the enzyme. On the other hand the introduction of a negatively charged residue through the H241D substitution mutation destabilises the enzymes and results in insolubility and this may be due to the negative charge repulsion between the introduced Asp241 and Glu195, hence the substitution mutation to introduce a positively charged lysine (H241K) would retain the enzymes solubility as observed.

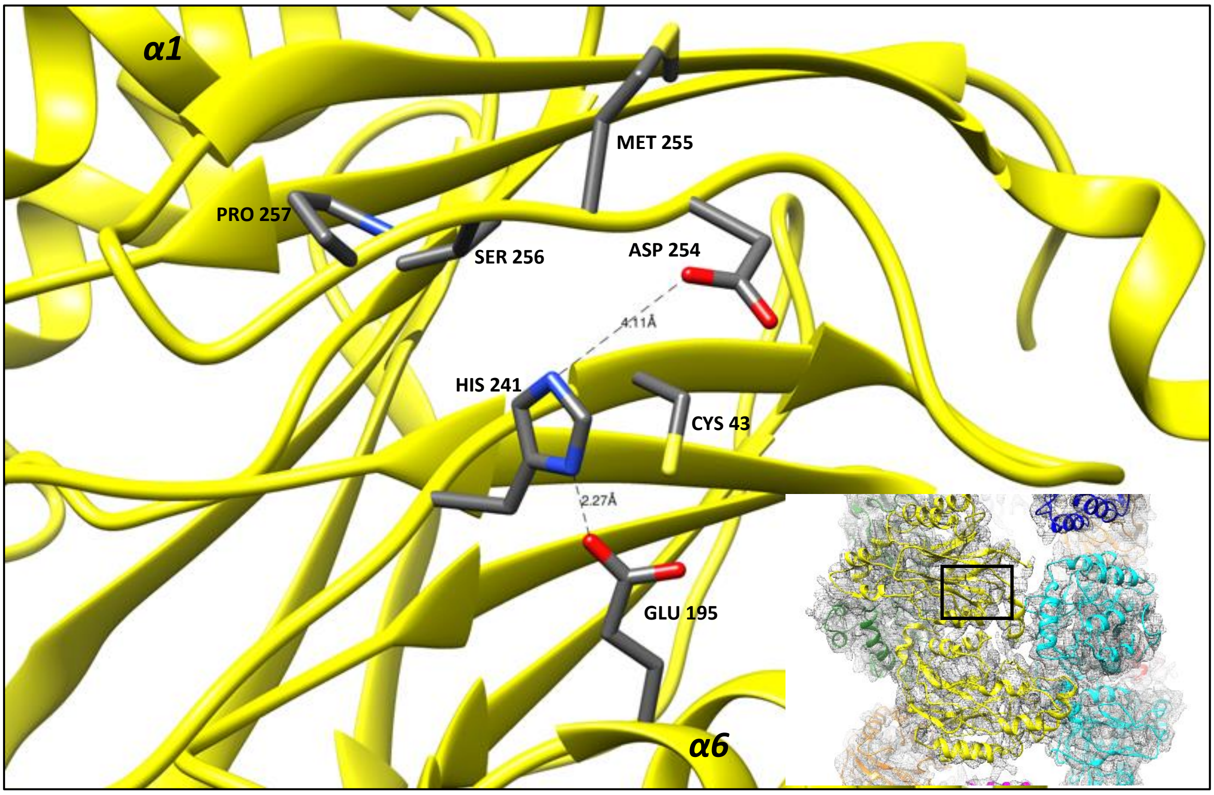


Figure 4.22. The cryo-EM map refined atomic model of the wild-type CynD$_{pum}$ showing the location of the His241 and its neighbouring residues. The His241-Glu195 hydrogen bonding is depicted by the dashed line with a bond distance of 2.27 Å.

Interestingly some histidine mutants with mutations located around the interfacial regions also resulted in insoluble protein expression. These include the H62A and H62D mutations. The His62 is located around the C-surface region in which interactions may occur between the two-fold symmetry related loop region (residue 63–59) across monomers from adjacent dimers (Fig. 4.23). Based on a previous study, no new disulphide bonds were formed when individual residues from this loop were substituted to cystines (including H62C), demonstrating that there may be no interaction across the two-fold symmetry related loop at the C-surface (Park et al., 2016). Based on our atomic models it was shown that His62 has no clear residue partner that it could interact with (Fig. 4.23), but a positive charge at this position is required for the solubility of the enzyme as a substitution with a positively charged residue (H62K) did not affect the solubility of the enzyme. The H62D mutation may introduce a Asp62-Gln230 hydrogen bond and the H62A mutation may remove the positive residue at the His62 position, which both may affect the structural conformation of the overall enzyme resulting in insolubility.

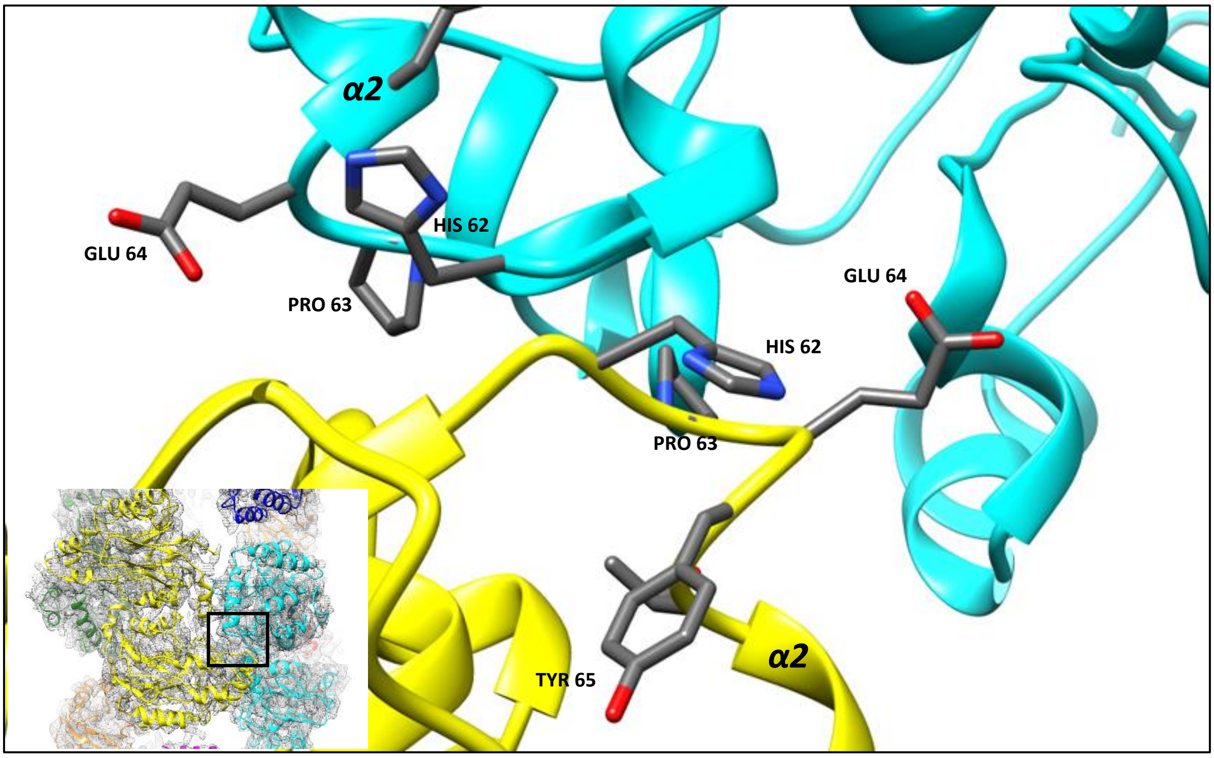


Figure 4.23. The cryo-EM map refined atomic model of the wild-type CynD$_{pum}$ showing the location of the His62 and its neighbouring residues.

Only two histidines, His71 and His285, showed no disruption of solubility when substituted with an alanine, aspartate or lysine. The His71 is located on the α2 helix at the C-surface (Fig. 4.24) while His285 is located near the C-terminal tail (Fig. 4.25). The retention of solubility regardless of the substitution mutation demonstrates that these histidines are located in very stable regions which are minimally affected by the introduction of different amino acid residues or may have compensating effects to mutations. Based on our atomic model His71 forms a hydrogen bond with Asp275 as previously described in section 3.4 4.2 (Fig. 4.24) while His285 may form a hydrogen bond interaction with Arg132 from an adjacent monomer (Fig. 4.25), which helps stabilise the enzyme and maintains its solubility.

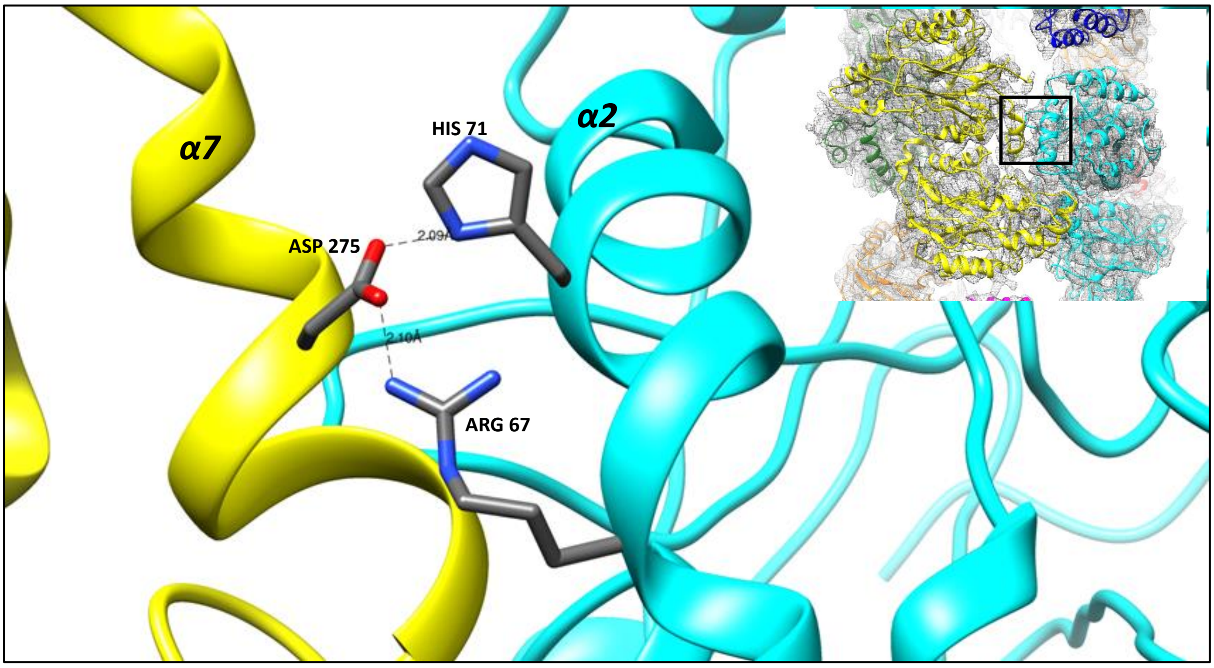


Figure 4.24. The cryo-EM map refined atomic model of the wild-type CynD$_{pum}$ showing the location of the His71 and its neighbouring residues. The His71-Asp275 hydrogen bonding is depicted by the dashed line with a bond distance of 2.09 Å.

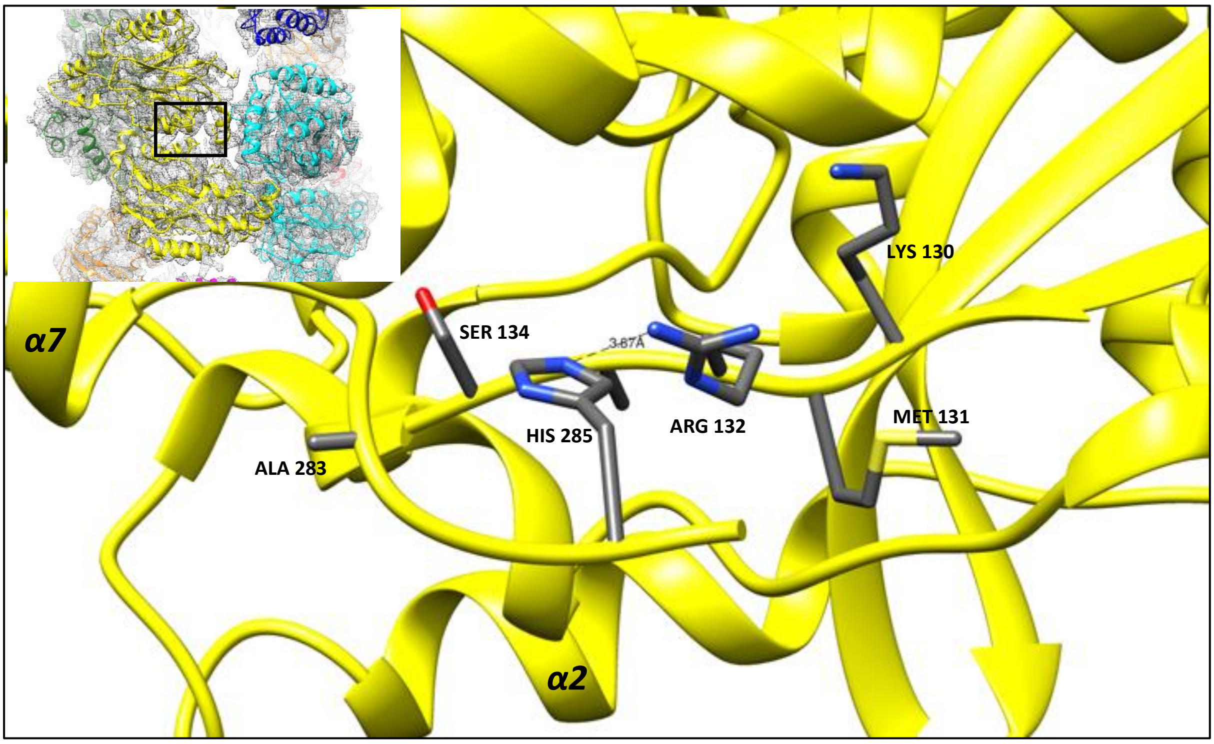


Figure 4.25. The cryo-EM map refined atomic model of the wild-type CynD$_{pum}$ showing the location of the His 285 and its neighbouring residues. The His285-Arg132 hydrogen bonding is depicted by the dashed line with a bond distance of 3.67 Å.

4.5.2. Purification of soluble histidine substitution mutants

The 11 histidine mutants that showed soluble expression were subjected to a three-step purification process consisting of ammonium sulphate fractional precipitation, anion exchange and gel filtration. All the soluble histidine mutants except H71D showed a large peak around the gel filtration elution volume of 8 ml (~670 kDa) which correspond to the previously observed 18mer molecular weight of around 666 kDa in the wild-type pH 8 gel filtration elution profile. This demonstrated that the mutations of the soluble histidine mutants did not result in any major changes in the number of subunits needed in the formation of the pH 8 short spirals observed during the microscopy. Using the samples from the large fraction 50 peak, the purity of the mutants was assessed using SDS-PAGE gels. Some of the soluble histidine mutants showed the presence of smaller sized contaminants. These contaminants were unlikely to interfere with the subsequent results as there were at low concentrations and too small in size to be observed by electron microscopy.

4.5.3. Microscopy, thermostability and activity of the soluble histidine mutants

The purified soluble histidine substitution mutants were viewed under the transmission electron microscope to observe the enzymes under different pH conditions. Thermostability of the mutants was also measured to verify if the observed structural changes were due to changes in the stability of the enzyme and also identify any mutations that confer higher stability. Of all the histidine substitution mutations the H62K, H71A, H71K and H241A displayed similar characteristics to the wild-type in terms of the pH dependent oligomerisation, thermostability and activity. The similarity in the pH dependent oligomerisation between the H62K mutant and the wild-type demonstrated that the His62 may not be directly involved in this oligomerisation. On the assumption that the His62 were directly involved in the pH dependent oligomerisation, one would expect the lysine from the H62K substitution to retain the positive charge at pH 8 (pka Lys=10) which His62 loses above pH 6 (pka His=6) and result in the retention of the fibres at pH 8 as opposed to the observed short spirals. The H62K mutant also showed a presence of additional smaller sized isomers when compared to the wild-type (Fig. 4.6) and this may be due to the conformational variations coffered by the new introduced positive charged lysine proposed to stabilise the H62K mutant.

The His71 substitution mutations are one of the two sets of mutations that resulted in mutants that had no effect on the solubility of the enzyme. Similar to the H62K mutant the introduction of the lysine at the His71 position (H71K) did not result in the retention of the fibres at pH 8, thus demonstrating no direct involvement of His71 in the pH dependent oligomerisation of CynD_{pum}. As previously observed on our atomic models, the His71 interacts with Asp275 to form a hydrogen bond that helps stabilise the C-surface (Fig. 4.24). The introduction of the H71K mutation may result in a stronger Lys71- Asp275 ionic bond interaction which may confer the slightly higher thermostability observed at pH 8 for this mutant (Fig. 4.19). As the pH was dropped to 5.4 the thermostability of the H71K mutant became similar to that of the wild-type as the His71 of the wild-type gains a stronger positive charge below pH 6 resulting in a His71- Asp275 hydrogen bond interaction of almost equal strength to the Lys71- Asp275 interaction from the H71K mutant. The removal of the polar His71 (H71A) did not result in any major changes in the oligomerisation of the enzyme and this may be due to the reinforced stabilisation conferred by the Arg67-Asp275 ionic bond which we previously identified as the predominant interaction that stabilises the C-surface (Fig. 4.24). Only a small decrease in thermostability was observed for the H71A mutant compared to the wild-type thus further demonstrating the compensating effect of the Asp275-Arg67 ionic bond in stabilising the C-surface. The introduction of an aspartate negative charge (H71D) not only improved the oligomerisation at pH 5.4 but also interestingly rendered the enzyme inactive. This is good evidence that demonstrates that the activity of the CynD_{pum} enzyme is independent of the oligomerisation of the enzyme. The slight decrease in thermostability relative to the wild-type displayed by the H71D demonstrates that the improvement in the oligomerisation of this mutant at pH 5.4 is due to the structural changes caused by the mutation as opposed to an increase in the stability of the enzyme. Based on our atomic structure there is a possibility that the introduction of an aspartate at the His71 position may result in the disruption of the native His71-Asp275 hydrogen bond interaction, causing conformational changes at the C-surface which allows for the addition of more subunits during oligomerisation. As previously observed in section 3.5.2.2, the disruption of interactions at the C-surface may lead to the disruption of the interactions between the α2 helix and the loop containing the catalytic Glu48 (Fig. 3.13) hence the disruption of the His71- Asp275 C-surface interaction results in an inactive enzyme and demonstrates the important role of His71 in the activation of the CynD_{pum} enzyme. The similar thermostability trend to that of the wild-type displayed by the H71D demonstrates that the improvement in the oligomerisation of this mutant at pH 5.4 is due to the structural changes caused by the mutation as opposed to an increase in the stability of the enzyme.

The H128K formed significantly longer but not straight fibres than those observed for the wild-type at pH 5.4. Although H128K formed long fibres at pH 5.4 it interestingly displayed signs of denaturation at pH 8 under the microscope (Fig. 4.10) which was confirmed by the continuous steady decrease in fluorescence in the differential scanning fluorimetry assay (Fig. 4.17). The long fibres formed by H128K at pH 5.4 were not straight and showed some destabilisation in their overall structure. The introduction of the lysine in the H128K mutant did not result the retention of the fibres at pH 8 but instead resulted in the destabilisation on the mutant and this may be due to the introduced lysine forming a weaker hydrogen bond interaction with Gln115 than the native His128-Gln115 interaction resulting in the partial destabilisation of the fibres (Fig. 4.20).

The H241A mutant displayed similar characteristics to the wild-type while H241K formed slightly shorter fibre than the wild-type demonstrating that the native His241 may not be crucial in the pH dependent oligomerisation of CynD_{pum}. Introduction of a lysine at this position partially disrupted the pH dependent oligomerisation at pH 5.4 instead of retaining the fibre formation at pH 8 due to the lysine's high pKa, thus further demonstrating His241 non-involvement in oligomerisation. Based on our atomic structure the hydrogen bond between the native His241 and Glu195 (Fig. 4.22) may not be crucial in oligomerisation and stability as removal of the polar His241 (H241A) did not affect the oligomerisation properties of the enzyme and only slightly reduced its stability.

The substitution mutations at the His285 position all retained the enzyme's solubility but rendered the mutant enzymes inactive. This demonstrates the importance of the native His285 in the activation of CynD_{pum}. The removal of the polar His285 (H285A) resulted in the formation of slightly shorter fibres than the wild-type demonstrating the importance of the native His285 in maintaining the structure of the fibres at pH 5.4. The H285A mutant is the first mutant that we have encountered that has shown strong evidence of the disruption of the fibre formation at pH 5.4. Based on our atomic model His285 may form a hydrogen bonding interaction with Arg132 from an adjacent monomer which helps stabilise the enzyme and maintains its solubility (Fig. 4.25). Arg132 is located on the same loop region containing the catalytic Lys130 and introducing the H285A mutation at the His285 position may disrupt the His285-Arg132 hydrogen bond which may result in the destabilisation of the active site hence the inactivation of the enzyme. The H285K mutant formed very short fibres at pH 5.4 than those formed by the wild-type and clearly showed evidence of the disruption of the pH dependent oligomerisation of CynD_{pum}, thus further demonstrating the important role of His285 in the oligomerisation of the enzyme. The enzyme inactivation and disruption of the pH dependent oligomerisation by the introduction of the H285K mutation may also be due to the disruption of the native His285- Arg132 hydrogen bond between two adjacent monomers. The formation of short oligomers by the H285A and the H285K mutant were shown to be not due to the destabilisation but rather the conformational changes of the enzyme as the thermostability assay showed only a small decrease in stability relative to the wild-type (Fig. 4.19). Although the H285D mutant was inactive it had the ability to form longer fibres than those formed by the wild-type at pH 5.4. These fibres were more flexible compared to the wild-type and were not longer as a result of an increase in stability but rather from the conformational change caused by the introduced mutation as shown by the small decrease in thermostability relative to the wild-type. The formation of the long fibres by the H285D mutant may be due to the introduction of a strong ionic bond, between the mutant's Asp285 and a neighbouring Arg132 (Fig. 4.25), which may result in the conformational change of the enzyme that allows further addition of subunits during oligomerisation. Similar to the H285A and the H285K mutants, the inactivation of the H285D may be due to the disruption of the of the loop region contain the potential catalytic Lys130 residue. The results above demonstrated that His285 may be crucial in both activation and the pH dependent oligomerisation of CynD$_{pum}$.

The thermostability assay was also able to verify the high stability of the Q86R+H305K+H308K+H323K mutant and illustrated that this mutant was a good choice for obtaining the CynD$_{pum}$ cryo-EM structure in chapter II due to its reduced structural variability. The high melting temperatures from thermostability assay results of the Q86R+H305K+H308K+H323K mutant also provided a standard to help access if any of the histidine substitution mutations conferred a significant increase in stability and from our results no histidine mutant displayed any significant increase in stability.

4.6 Conclusion

A majority of the histidine substitution mutations resulted in some change in the solubility, pH dependent oligomerisation, thermostability and activity when compared to the wild-type enzyme. Only about 50% of the histidine substitution mutants displayed soluble expression demonstrating that majority CynD$_{pum}$ histidines are required for the correct folding of the enzyme which results in solubility and small changes such as single mutations at the histidine positions may render the overall enzyme insoluble. Based on our atomic structure a majority of the CynD$_{pum}$ histidine are located around the core of each CynD$_{pum}$ subunit thus explaining the high sensitivity to mutational changes at these areas. The branch 1 nitrilase conserved His167-His184 interaction was shown to be crucial in the stability and possibly in the positioning of the catalytic Cys164 of CynD$_{pum}$ and other related branch 1 nitrilases. The disruption of the pH dependent oligomerisation was not demonstrated by the introduction of histidine to alanine mutations and no retention of the fibres at pH 8 was observed when histidine to lysine mutations were introduced. This suggests that no one CynD$_{pum}$ histidine is responsible for the observed CynD$_{pum}$ pH dependent oligomerisation but may be due to a combined effect by multiple histidines or other residues, although the H285A and H285K mutation demonstrated that His285 may be the most involved histidine in oligomerisation. From the EM micrographs and the thermostability assay of H128K we observed that the quality and not only the length of the fibres should be used to assess the stability of CynD$_{pum}$ mutants. The H71D and the H285D mutants demonstrated that the activity of the CynD$_{pum}$ enzyme is not necessarily dependent on the oligomerisation of the enzyme.

CHAPTER V

STRUCTURE OF THE FIBRE FORMING HISTIDINE SUBSTITUTION MUTANTS

5.1 Introduction

In the previous chapter we attempted to either enhance or disrupt the native pH dependent oligomerisation of the CynD$_{pum}$ enzyme by mutating the histidines. During this study we identified some histidine mutants that not only retained the native pH dependent oligomerisation but demonstrated different oligomerisation properties and activities to those of the wild-type enzyme. To understand the role of the mutations on the changes observed we collected negative stain electron microscopy datasets of all the fibre forming histidine mutants (H62K, H71A, H71D H71K, H128K H184D, H241A and H285D), processed this data to produce low resolution 3D reconstructions. In this chapter we describe the acquisition of the 3D reconstructions of the fibre forming histidine mutants and how these models were used to demonstrate variations in the helical symmetry of the mutants which helped investigate the role of the fibres' helical twist in the oligomerisation and activity of the CynD$_{pum}$ enzyme

5.2 Methods and Materials

5.2.1 Negative stain electron microscopy sample preparation

Negative stain electron microscopy samples were prepared as described in section 2.3.10 using only the pH 5.4 buffer.

5.2.2 Negative stain electron microscopy data collection

The negative stain electron microscopy method was used to collect datasets which were later used for generate three dimensional reconstructions. Negative stain micrographs datasets of the soluble histidine mutants that formed long fibres at pH 5.4 were collected on the Tecnai F20 transmission electron microscope. Datasets for the 3D reconstruction were collected at low dose at 50 000x magnification using a Gatan US4000 CCD camera (Gatan, USA). The calibrated sampling of 2.17 Å/pixel and defocus of 0.8–1.0 µm was used to manually capture datasets of micrographs for each CynD_{pum} variant.

5.2.3 Image pre-processing

The Helixboxer program of the EMAN suite (Ludtke et al., 1999) was used to pick and extract straight helical filaments from the micrographs. Power spectra were calculated for each of the boxed helices with the removal of helices that failed to diffract. All the successful power spectra were then averaged together and the resultant power spectrum was used to evaluate the data. The boxed filaments were then further cut into several overlapping segments of 160 × 160 pixels with a shift of 10 pixels between adjacent segments, and all the segments were compiled into a stack. The raw images were centred by applying x-shifts only, using SPIDER routines (Frank et al., 1996). The new stacks were further visually inspected using V2 of the EMAN suite (Ludtke et al., 1999) and bad segments (bent, overlapping or broken) were deleted from the image stacks. The helical segments were then classified based on their diameter using SPIDER routines (Frank et al., 1996) and later used in IHRSR.

5.2.4 Three-dimensional reconstruction

3D reconstruction was computed using the IHRSR++ software package (Parent et al., 2010) as described in section 2.2. A featureless cylinder (130 Å in diameter) was used as starting model for IHRSR. For each mutant two separate IHRSR++ processes were run using different arbitrary user-defined initial helical symmetry parameters to clearly observe any convergence of the helical symmetry, thus validating the final solution. Although the initial helical symmetry parameters used varied, they were kept within the range of those of the wild-type $\mathrm{CynD_{pum}}$ ($\Delta\varphi = -77.0°$, $\Delta z = 16.0$ Å).

5.2.5 Model validation

To validate the 3D reconstruction of the helical histidine mutants, simulated power spectra of the reconstructions were generated and compared to the power spectra produced from the original stack of helical segments. Visible layer lines from the pair of power spectra, corresponding to axial spacings, were compared to assess if the solved 3D structures were correct.

5.2.6 Resolution determination

The FSC method was used for resolution determination as described in section 2.3.14.

5.2.7 Visualisation

Visualisations of 2D images were done using BOXER and V2 both of the EMAN suite (Ludtke et al., 1999). UCSF Chimera (Petterson et al., 2004) was used to view the 3D reconstruction volumes and to dock in the wild-type atomic model (from Chapter III) into the generated 3D volumes, aided by the ‘Fit model in map’ function.

5.3 Results

5.3.1 Data collection and image pre-processing

Negative stain micrographs datasets of the fibre forming soluble histidine CynD_{pum} mutants were collected on the Tecnai F20 transmission electron microscope at low dose. Although 11 of the soluble CynD_{pum} histidine mutants formed fibres at pH 5.4 only 8 mutants formed fibres long enough for 3D reconstruction. Around 100 micrographs were collected for each of the datasets (Table 5.1). Straight intact fibres were selected from the micrographs (Fig. 5.1A) and later used to generate stacks of segmented helical fibres (Fig. 5.1B). The segmented fibres were then classified and selected based on their diameter, after which all the histidine mutants retained more than 50% of their original helical segments except the H184D variant which lost about 70% due to the its broad diameter range, thus demonstrating a high level of structural variability (Table 5.1).

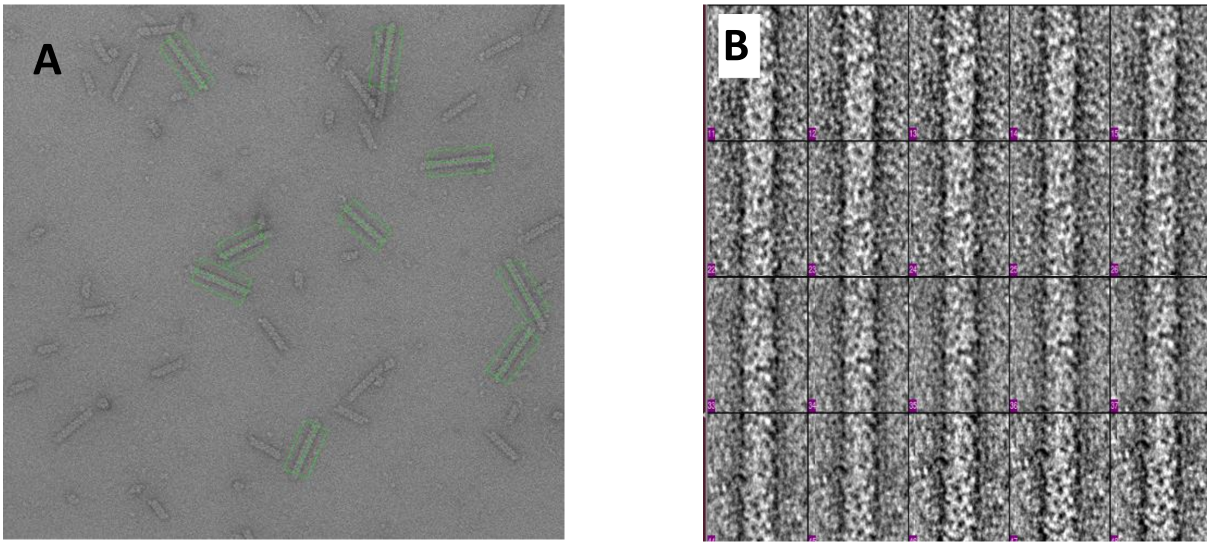


Figure 5.1. Representative image illustrating the pre-processing of EM micrographs of the CynD_{pum} H71K mutant. **(A)** Selection of helical fibres by boxing using the Helixboxer program (Ludtke et al., 1999). **(B)** Image stack which consists of centred helical segments viewed using the V2 program (Ludtke et al., 1999).

Table 5.1. Image pre-processing summary for the fibre forming CynD$_{pum}$ histidine substitution mutants.

Histidine Mutant	# of micrographs	# of extracted segments after particle picking	# of segments in stack after manual sorting	# of segments in stack after sorting by diameter
H62K	104	9768	6946	3850
H71A	102	11571	5923	5923
H71D	100	20404	18728	11390
H71K	103	23473	23327	16277
H128K	105	18319	14214	9920
H184D	130	17123	15952	4729
H241A	104	13130	13034	6840
H285D	103	21710	21256	17648

5.3.2 Electron microscopy 3D reconstruction

The stacks containing the segmented helical fibres of the histidine mutants were processed using IHRSR to produce 3D models and calculate the helical symmetry parameter of the CynD$_{pum}$ histidine mutants. The low resolution 3D reconstructions of the mutants (resolution of 15–29 Å; Table 5.2) were used to assess any structural differences (Fig. 5.2–5.9) and any differences in the helical symmetry parameters (Table 5.2). Four of the histidine mutants demonstrated the same helical symmetry parameters as the wild-type (H62K, H71K, H184D and H241A) and four mutants had different helical parameters to those of the wild-type (H71A, H71D, H128K and H285D).

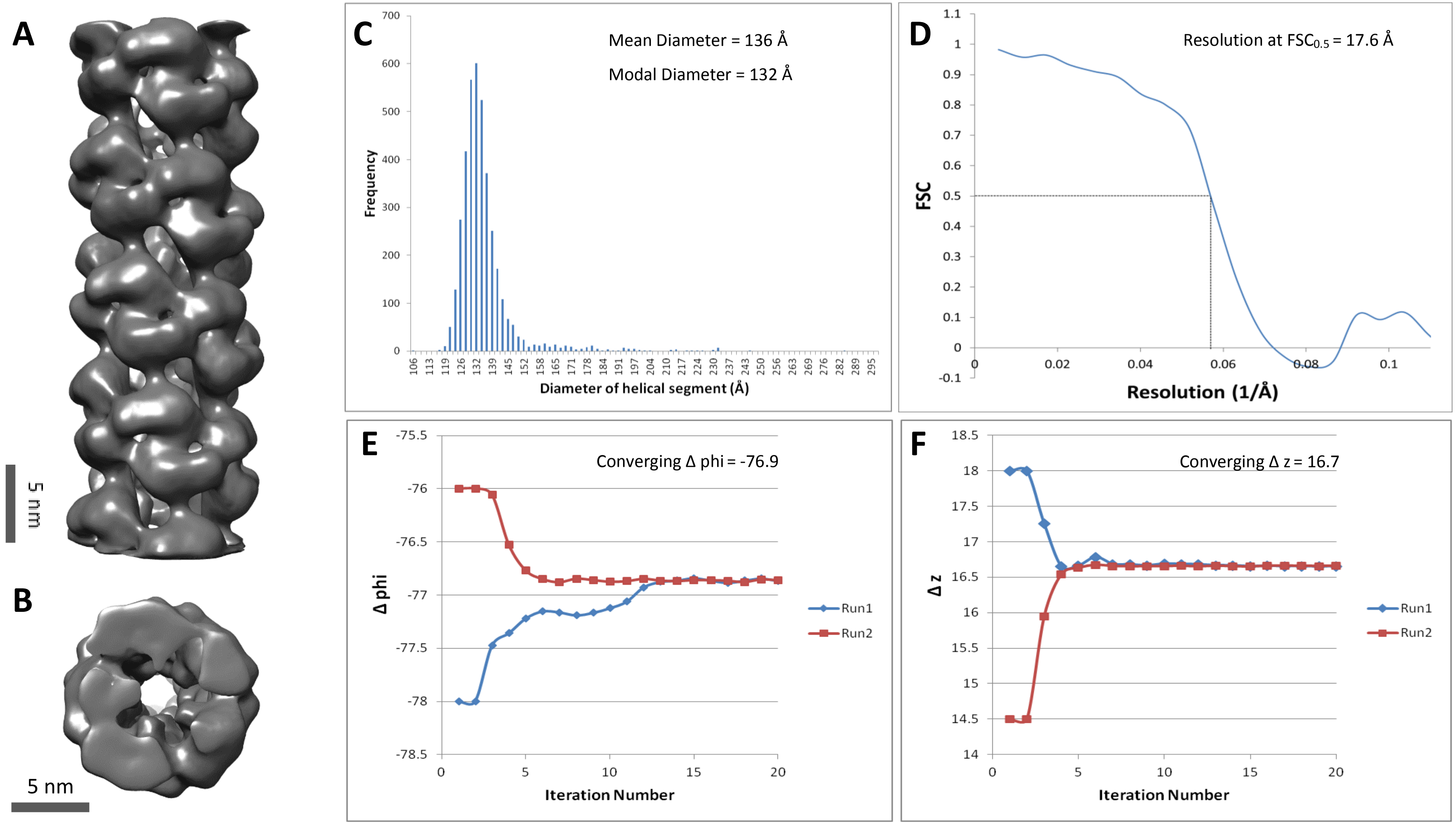


Figure 5.2. **(A)** Three-dimensional reconstruction model of the H62K CynD$_{pum}$ mutant at pH 5.4 and **(B)** 90° rotated view of the 3-D model. **(C)** Frequency distribution of boxed helical segments, used to generate the model, based on diameter. **(D)** Estimated resolution of the 3-D model determined by the 0.5 Fourier shell correlation cut-off. Converging symmetry values, **(E)** Δ phi and **(F)** Δ z, where calculated from two separate reconstructions (run 1 and 2) with different initial helical symmetry parameters.

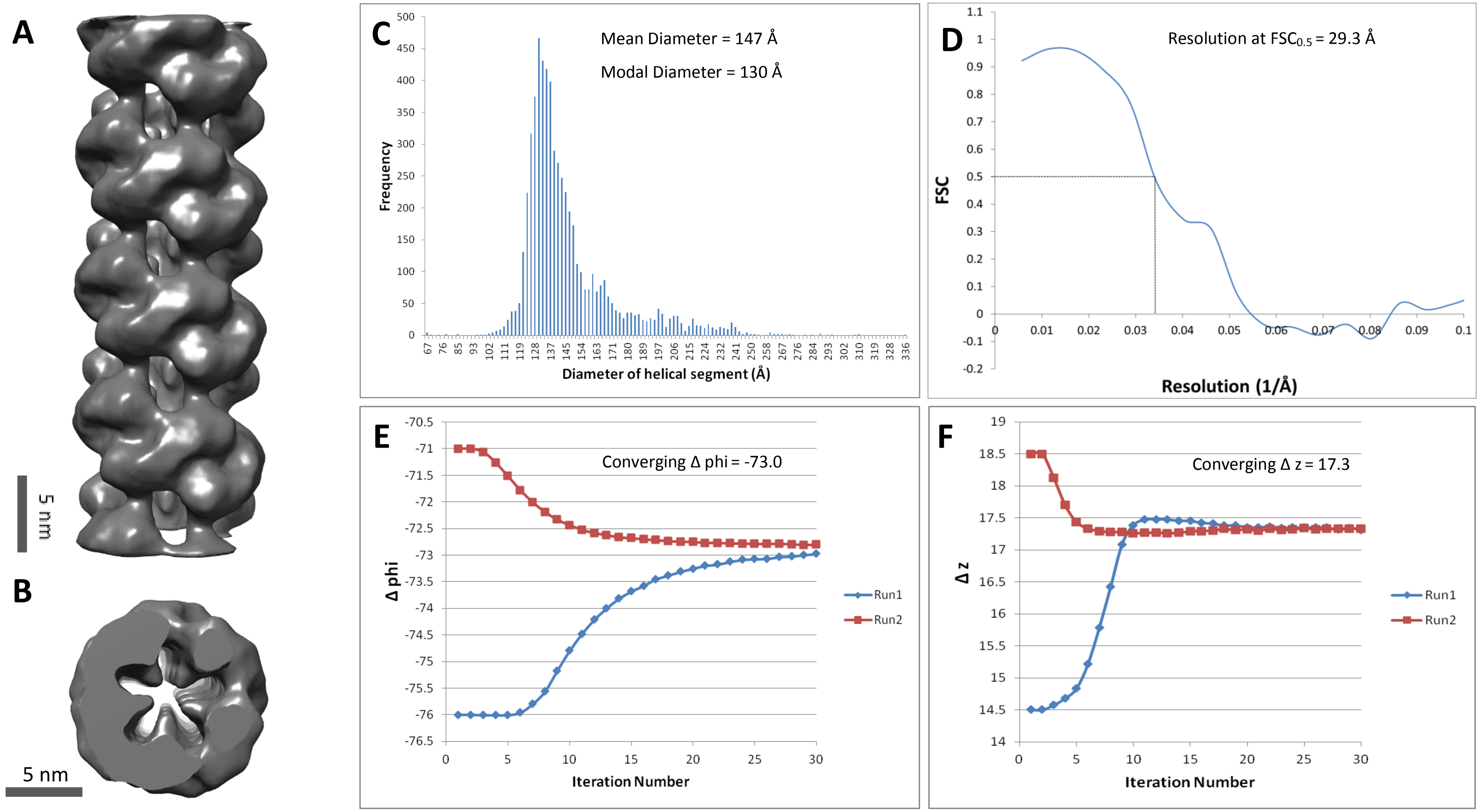


Figure 5.3. (A) Three-dimensional reconstruction model of the H71A $CynD_{pum}$ mutant at pH 5.4 and **(B)** 90° rotated view of the 3-D model. **(C)** Frequency distribution of boxed helical segments, used to generate the model, based on diameter. **(D)** Estimated resolution of the 3-D model determined by the 0.5 Fourier shell correlation cut-off. Converging symmetry values, **(E)** Δ phi and **(F)** Δ z, where calculated from two separate reconstructions (run 1 and 2) with different initial helical symmetry parameters.

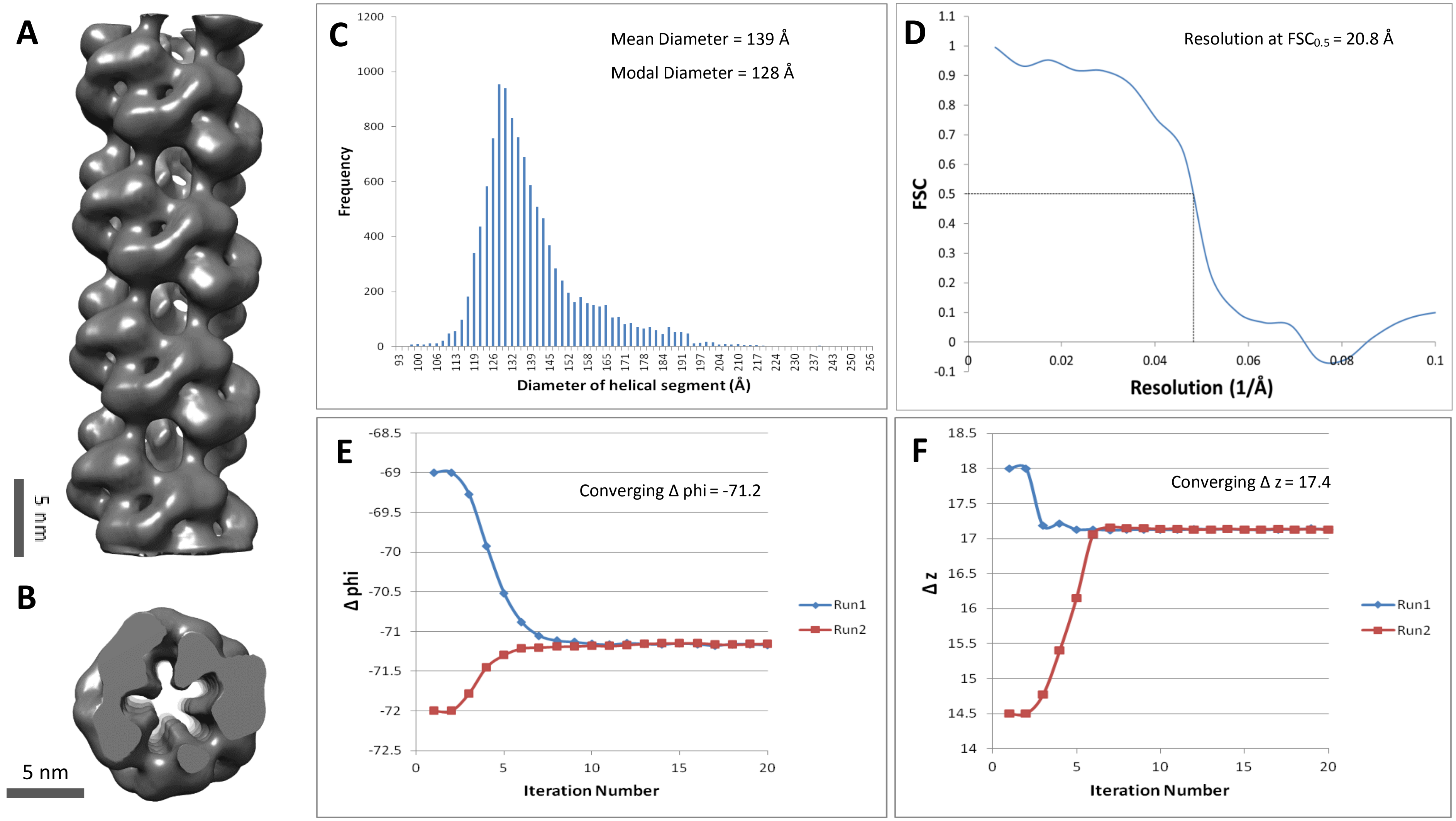


Figure 5.4. **(A)** Three-dimensional reconstruction model of the H71D CynD$_{pum}$ mutant at pH 5.4 and **(B)** 90° rotated view of the 3-D model. **(C)** Frequency distribution of boxed helical segments, used to generate the model, based on diameter. **(D)** Estimated resolution of the 3-D model determined by the 0.5 Fourier shell correlation cut-off. Converging symmetry values, **(E)** Δ phi and **(F)** Δ z, where calculated from two separate reconstructions (run 1 and 2) with different initial helical symmetry parameters.

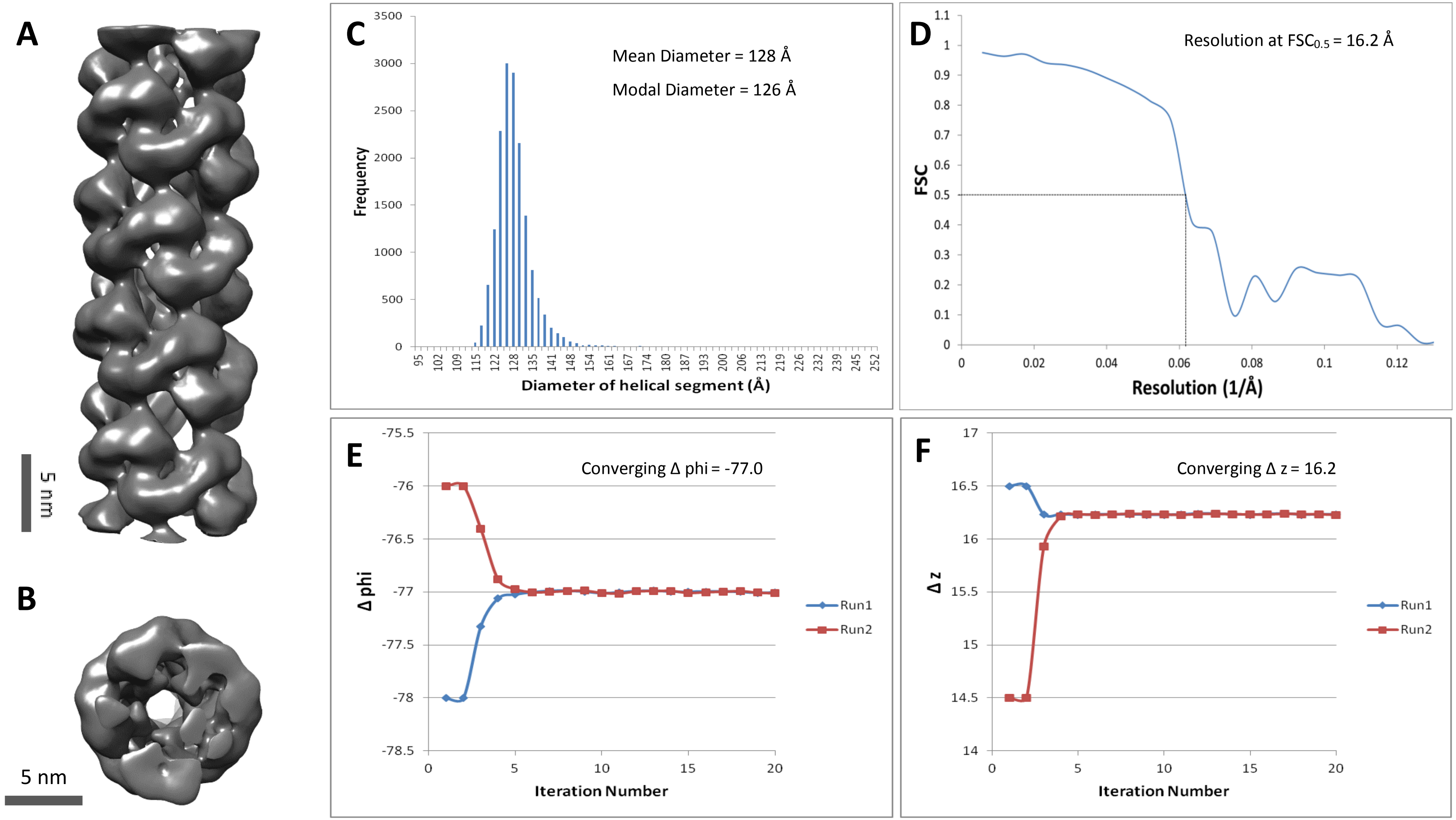


Figure 5.5. (A) Three-dimensional reconstruction model of the H71K CynD_{pum} mutant at pH 5.4 and **(B)** 90° rotated view of the 3-D model. **(C)** Frequency distribution of boxed helical segments, used to generate the model, based on diameter. **(D)** Estimated resolution of the 3-D model determined by the 0.5 Fourier shell correlation cut-off. Converging symmetry values, **(E)** Δ phi and **(F)** Δ z, where calculated from two separate reconstructions (run 1 and 2) with different initial helical symmetry parameters.

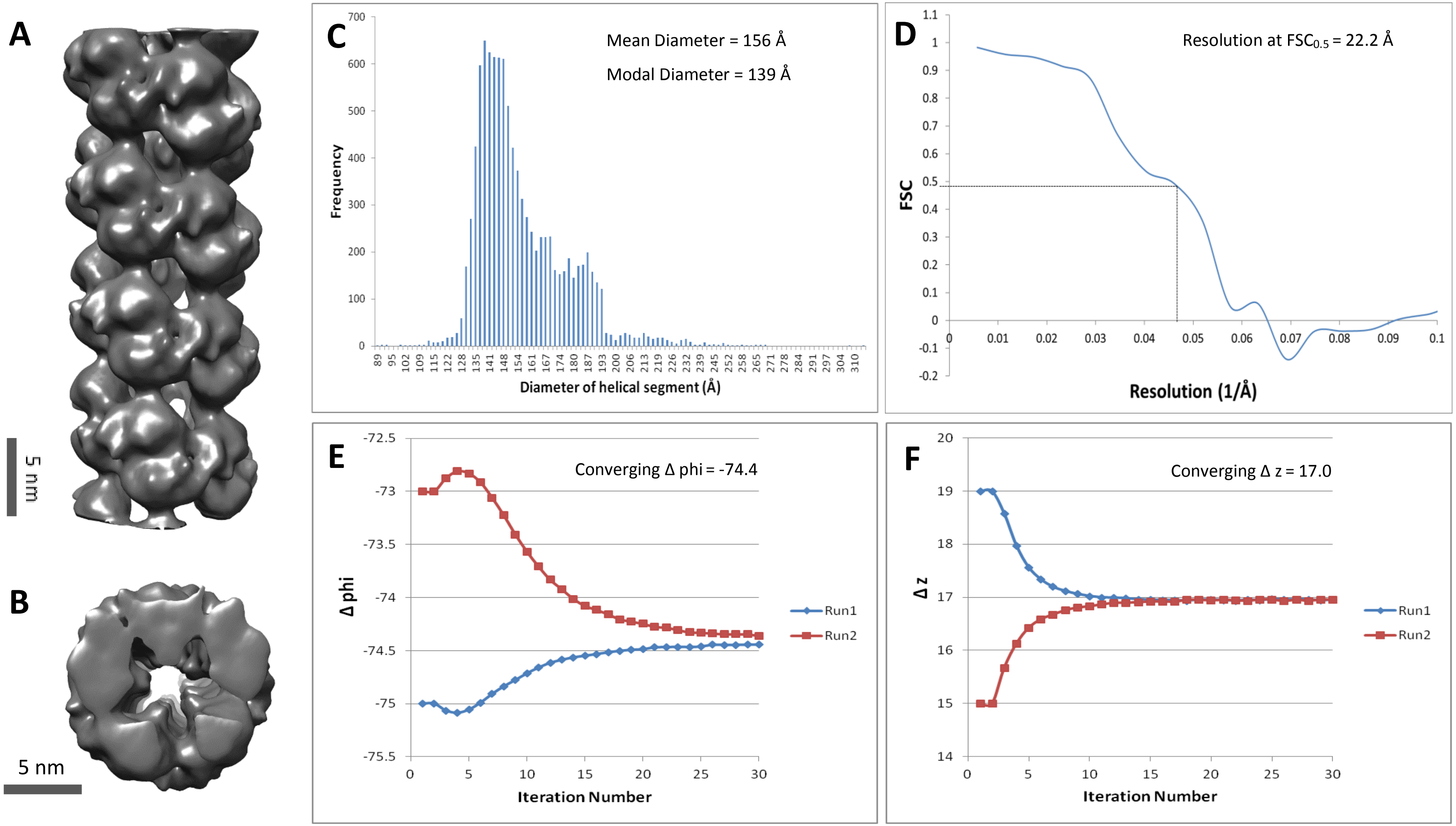


Figure 5.6. **(A)** Three-dimensional reconstruction model of the H128K CynD$_{pum}$ mutant at pH 5.4 and **(B)** 90° rotated view of the 3-D model. **(C)** Frequency distribution of boxed helical segments, used to generate the model, based on diameter. **(D)** Estimated resolution of the 3-D model determined by the 0.5 Fourier shell correlation cut-off. Converging symmetry values, **(E)** Δ phi and **(F)** Δ z, where calculated from two separate reconstructions (run 1 and 2) with different initial helical symmetry parameters.

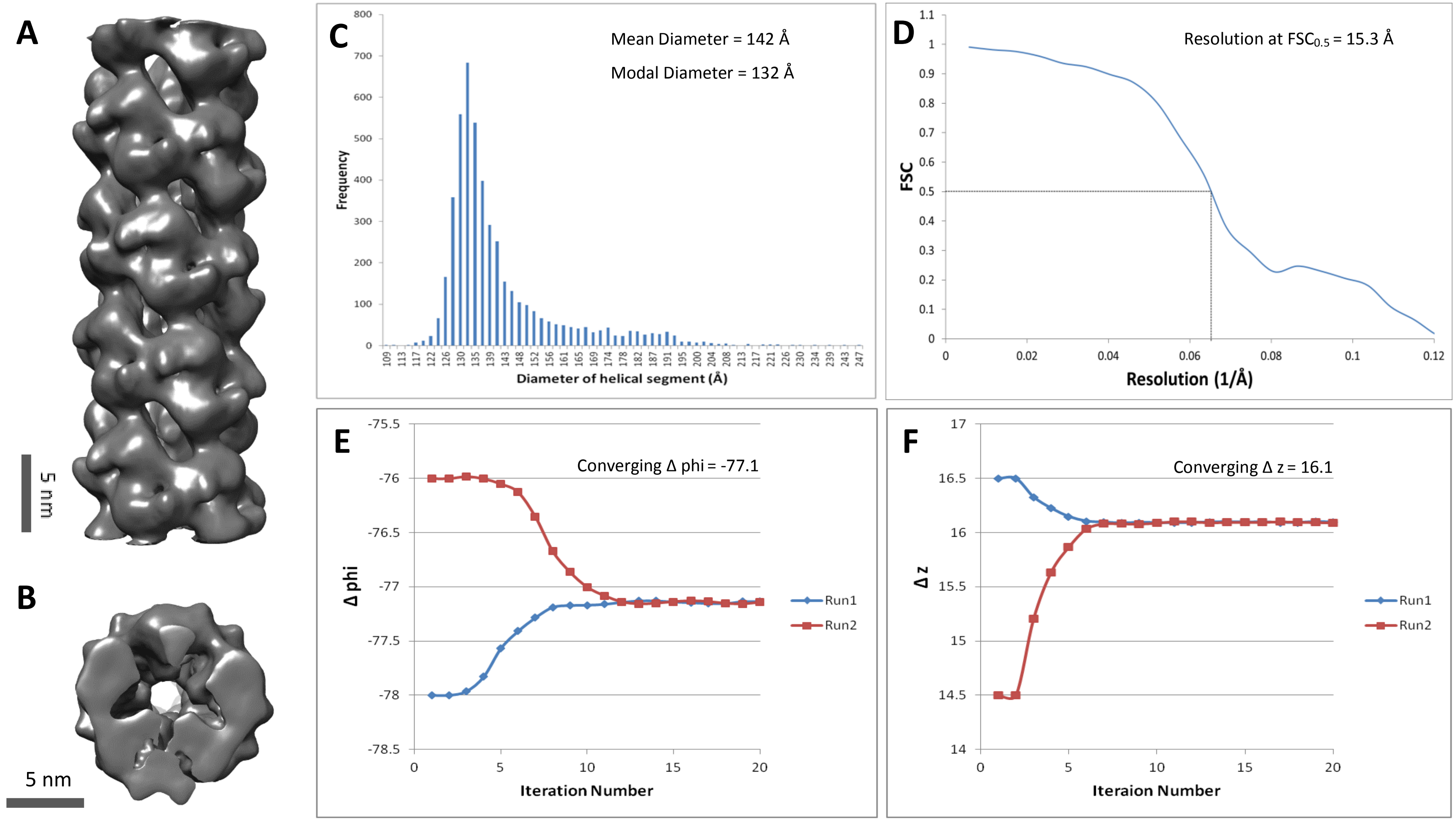


Figure 5.7. (A) Three-dimensional reconstruction model of the H184D CynD$_{pum}$ mutant at pH 5.4 and **(B)** 90° rotated view of the 3-D model. **(C)** Frequency distribution of boxed helical segments, used to generate the model, based on diameter. **(D)** Estimated resolution of the 3-D model determined by the 0.5 Fourier shell correlation cut-off. Converging symmetry values, **(E)** Δ phi and **(F)** Δ z, where calculated from two separate reconstructions (run 1 and 2) with different initial helical symmetry parameters.

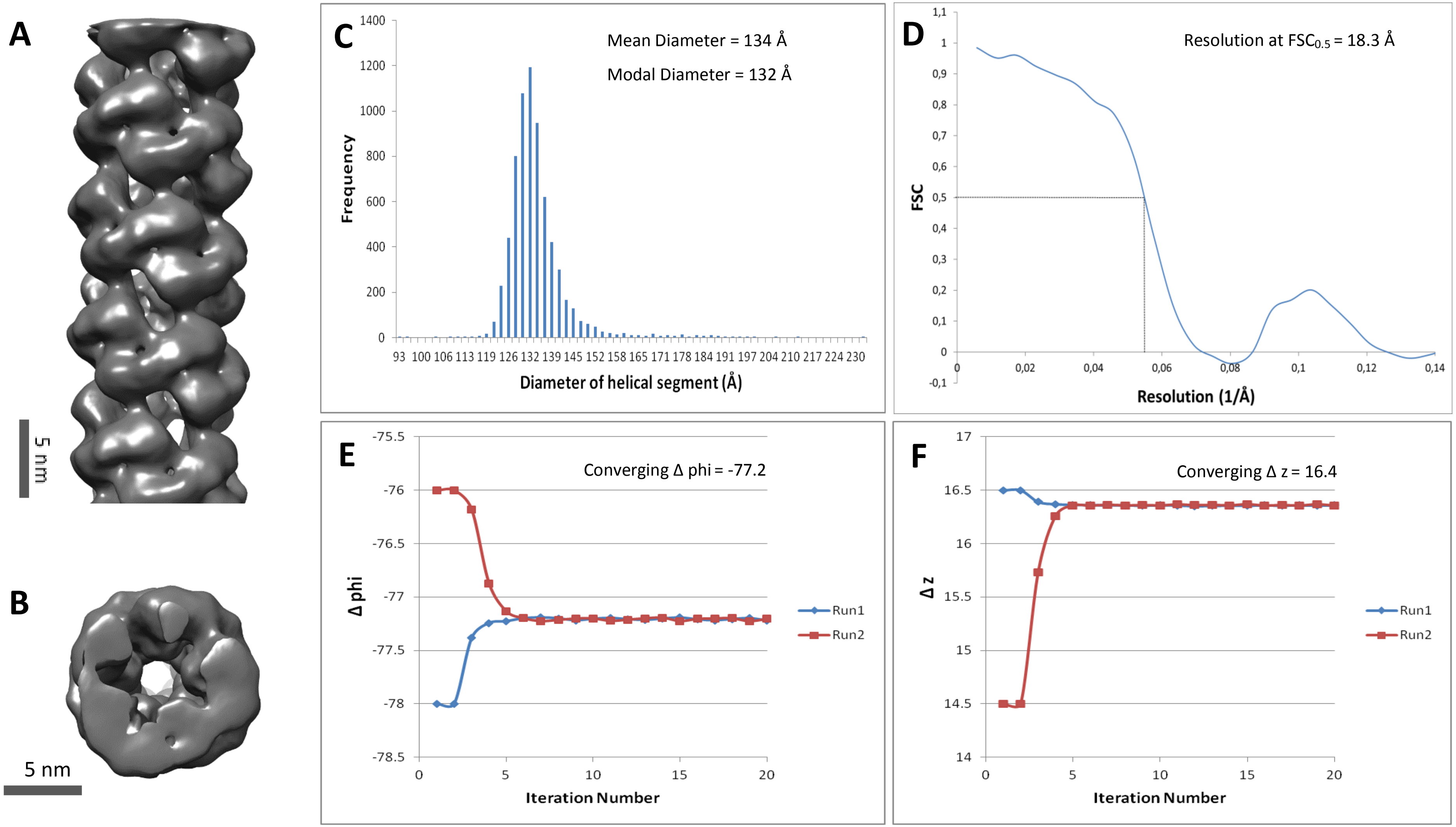


Figure 5.8. (A) Three-dimensional reconstruction model of the H241A CynD$_{pum}$ mutant at pH 5.4 and **(B)** 90° rotated view of the 3-D model. **(C)** Frequency distribution of boxed helical segments, used to generate the model, based on diameter. **(D)** Estimated resolution of the 3-D model determined by the 0.5 Fourier shell correlation cut-off. Converging symmetry values, **(E)** Δ phi and **(F)** Δ z, where calculated from two separate reconstructions (run 1 and 2) with different initial helical symmetry parameters.

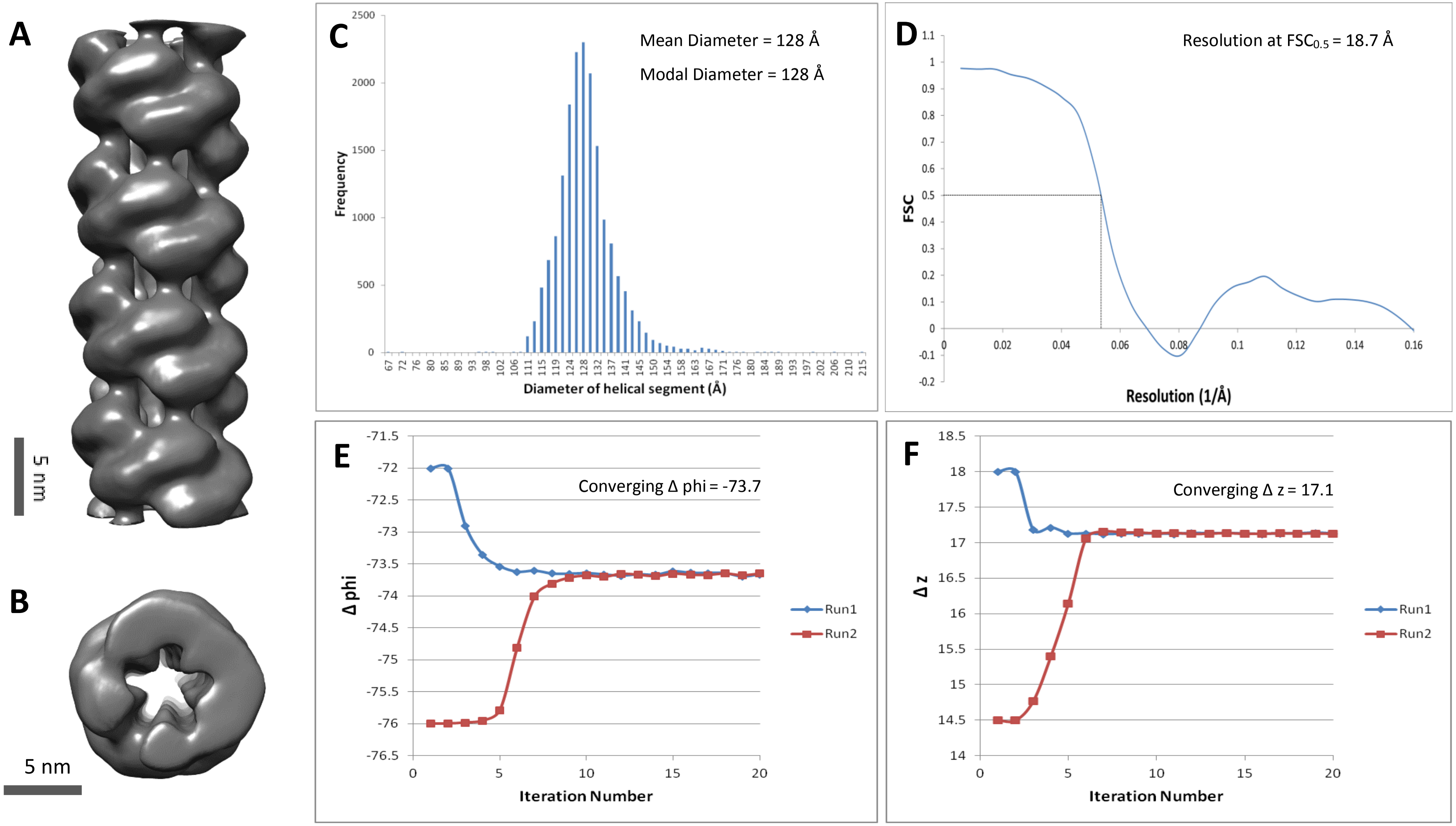


Figure 5.9. (A) Three-dimensional reconstruction model of the CynD_{pum} mutant H285D at pH 5.4 and **(B)** 90° rotated view of the 3-D model. **(C)** Frequency distribution of boxed helical segments, used to generate the model, based on diameter. **(D)** Estimated resolution of the 3-D model determined by the 0.5 Fourier shell correlation cut-off. Converging symmetry values, **(E)** Δ phi and **(F)** Δ z, where calculated from two separate reconstructions (run 1 and 2) with different initial helical symmetry parameters.

Table 5.2. IHRSR summary for the fibre forming CynD$_{pum}$ histidine substitution mutants and the wild-type CynD$_{pum.}$

Enzyme variant	**Final helical symmetry**		**Diameter size**		**Resolution at FSC$_{0.5}$ (Å)**
	Δ phi (°)	**Δ z (Å)**	**Average diameter (Å)**	**Modal diameter (Å)**	
H62K	-76.9	16.7	136	132	17.6
H71A	-73.0	17.3	147	130	29.3
H71D	-71.2	17.4	139	128	20.8
H71K	-77.0	16.2	128	126	16.2
H128K	-74.4	17.0	156	139	22.2
H184D	-77.1	16.1	142	132	15.3
H241A	-77.2	16.4	134	132	18.3
H285D	-73.7	17.1	128	126	18.7
WT	-77.0	15.7	130	130	17.0

5.3.3 Model validation

Power spectra of the resultant reconstructions were generated and compared to the power spectra produced by the segments from the raw data in order to validate the 3D models. Visible layer lines from the pair of power spectra, corresponding to axial spacings, were compared and it was observed that all the layer lines of the pair of power spectra of the different mutants were identical thus demonstrating that all the reconstructions determined are likely to be correct.

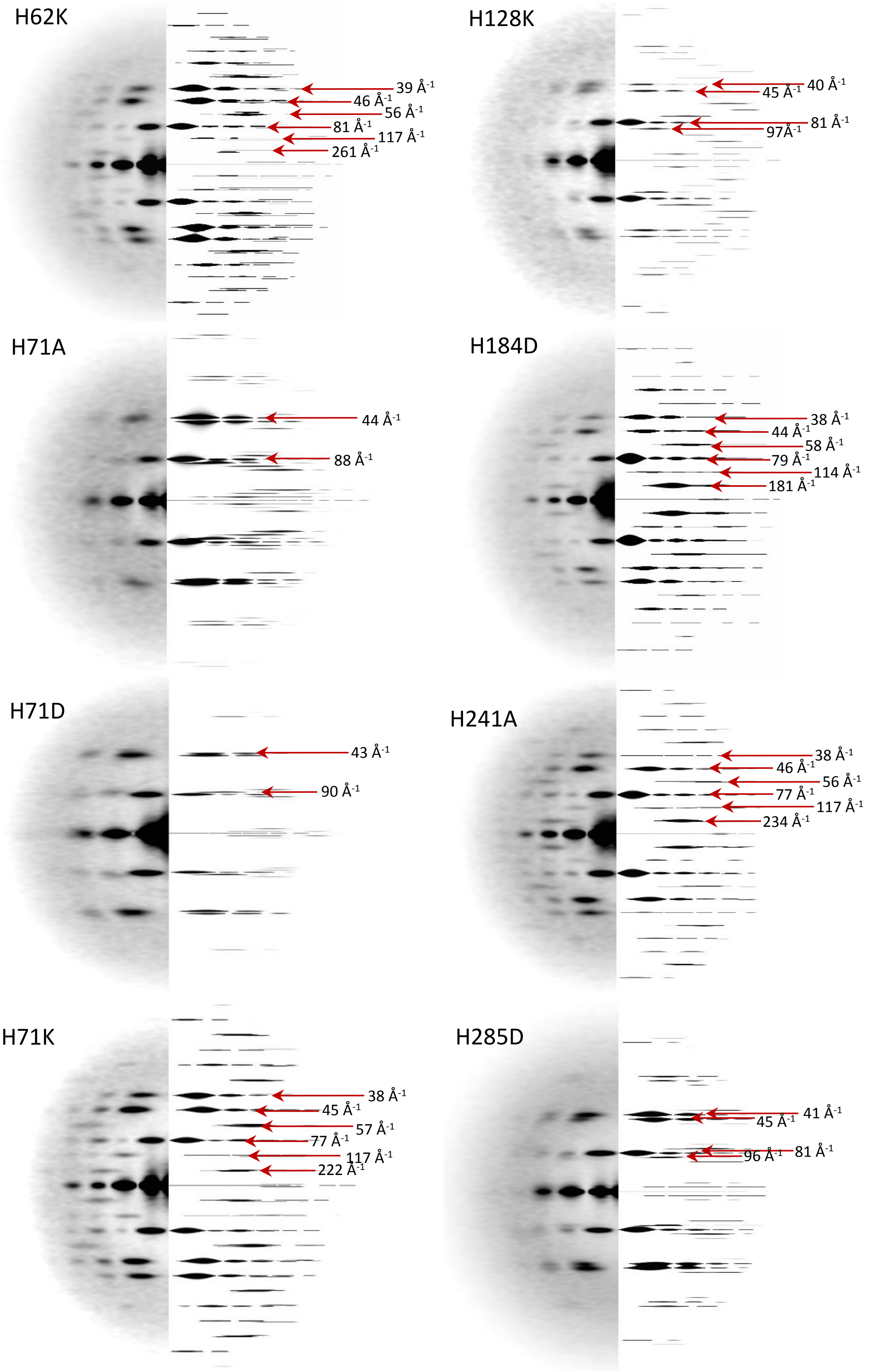


Figure 5.10. Comparison of the average power spectra of the $\mathrm{CynD}_{\mathrm{pum}}$ mutants obtained from the raw images (left) and from reconstructed maps (right) both showing layer lines corresponding to the axial spacing.

5.4 Discussion

5.4.1 Helical symmetry of the fibre forming histidine mutants

Of the 11 soluble histidine substitution mutants only 8 formed long enough fibres to generate 3D reconstructions. The low resolution 3D reconstructions of the mutants were not detailed enough to observe structural changes at a residue level but were enough to demonstrate the helical arrangement of individual monomers thus allowing us to observe any variations in the helical symmetry of these fibre forming histidine mutants. Of the 8 fibre forming histidine mutants, 4 had mutations located at the C surface namely the H62K, H71A, H71D and H71K mutant. The H62K mutant demonstrated no change in the helical symmetry when compared to the wild-type (Fig. 5.2). The His71 substitution mutants were the only set of soluble mutants that all formed helical fibres that were long enough to enable the generation of 3D reconstructions. His71 is located at the α2 helix which interacts predominantly with the α7 helix of an adjacent monomer across the C-surface. The removal of the positive charge at the His71 position (H71A) may have disrupted the α2-α7 interactions resulting in a more loose helical fibre that is demonstrated by the decrease in the helical twist (Fig. 5.3). The introduction of a negative charge at the His71 position (H71D) may have a similar but slightly greater effect as that of the H71A mutant in which the disruption of the C-surface interactions result in a looser helical fibre and this is demonstrated by the greater decrease in the helical twist of the H71D mutant (Fig. 5.4). The H71K mutant showed no change in the helical symmetry when compared to the wild-type (Fig. 5.5). The introduction of the positive lysine at the His71 position may influence the interactions at the C-surfaces but not to the extent which results in the disruption of the helical symmetry as observed for the H71A and H71D mutants. The H71K mutation demonstrates the importance of the positive charge at the His71 position and how its presence may help stabilize the positioning of the monomers thus maintaining the helical symmetry of the CynD$_{pum}$ fibre.

The H184D and H241A (Fig. 5.7 and Fig. 5.8) mutations resulted in no changes in the native helical symmetry while the H128K mutation resulted in a looser helical fibre illustrated by the decrease in the helical twist and an increase in the diameter of the helical fibre when compared to the wild-type (Fig. 5.6,Table 5.2). The decrease in the helical twist of the H128K mutant may explain the observed high level of flexibility and low level of stability demonstrated by the high variability in fibre diameter (Fig. 5.6C) and low thermostability (Fig. 4.17) observed in chapter IV. His128, His241 and His184 are both located at the core of the CynD$_{pum}$ monomer thus making it challenging to propose mechanisms that results in the observed characteristics of these mutants.

The H285D mutation resulted in the decrease of the helical twist. As mentioned in Chapter IV the His285 position is located near the C-terminal tail and the H285D mutant's change in helical twist may be due to conformational changes that result from the introduction of the Asp285-Arg132 ionic bond proposed in section 4.5.3.

It was previously hypothesised that the termination of the CynD_{pum} short spirals at pH 8 is due to the distortion of the helical symmetry which prevents the further addition of subunits to the elongating helical fibres (Sewell et al., 2005). This distortion in the helical symmetry was suggested to be due to pH dependent changes in the charge of one or more histidine residues. Our result suggest that this pH dependent oligomerisation may not be due to the distortion in the helical symmetry as some of the changes in charge at the histidine position distorted the helical symmetry but did not cause any significant changes in the native pH dependent oligomerisation of CynD_{pum} (Chapter IV). On the other hand our results do demonstrate that His71, His128K and His285 are involved in maintaining the helical symmetry of CynD_{pum}.

5.4.2 Activity of the fibre forming histidine mutants

The manipulation of some of the histidine did not only change the helical symmetry of the enzyme but also rendered the enzyme inactive such as the H71D and H285D mutations. In one of the previous studies on helical nitrilases it was discovered that changes in the helical symmetry of some nitrilases by mutagenesis can potentially change the substrate specificity of the enzyme (Woodward, 2011). It was proposed that as the helical twist of the nitrilase decreases, the helical filament opens up and the diameter of the fibre increases, a property displayed by the H128K mutant. The opening up of the helical filaments increases the size of the active-size pocket thus allowing access for larger substrates ultimately changing the substrate specificity of the nitrilase. The loss of activity by the H71D and H285D mutants may be limited to cyanide degrading activity due to the decrease in the twist of the helical filaments, thus these mutants could potentially be active against larger substrate. Interestingly the H71A and the H128K mutant were active for cyanide despite displaying a decrease in the helical twist demonstrating that the change helical twist will not necessarily affect the activity or change the substrate specificity of the enzyme thus other factors are involved in the stabilisation of the active-site pocket.

5.5 Conclusion

Low resolution structures were obtained for the soluble fibre forming histidine substitution mutants and these were used to observe variations in the helical symmetry of the fibres. The 3D structures of the His71 substitution mutants demonstrated the importance of the positive charge at the His71 position in maintaining the helical symmetry of the CynDpum fibre. Our structures demonstrated that the pH dependent oligomerisation of $\mathrm{CynD_{pum}}$ may not be due to the distortion in the helical symmetry as previously suggested. It was also demonstrated that the loss of cyanide degrading activity by some of the histidine mutants was not solely due to the changes in the helical symmetry of the helical filaments.

CHAPTER VI
GENERAL DISCUSSION AND FUTURE WORK

6.1 Overview

Nitrilases such as the cyanide dihydratase from *Bacillus pumilus C1* have potential to be used in the detoxification of cyanide waste produced during industrial processes, such as leach mining, metal finishing and electroplating. The drawback of this cyanide dihydratase enzyme is that cyanide wastes are highly alkaline and have been shown to inactivate the wild-type enzyme. We have carried out a number of mutagenesis experiments in an attempt to construct alkaline pH tolerant variants of this enzyme by increasing its stability. The $\mathrm{CynD_{pum}}$ enzyme exists as short spirals over the pH range of 6–8 and these spirals associate to form long active helical fibres at pH 5.4 (Jandhyala et al., 2003; Sewell et al., 2005; Wang et al., 2011; Mulelu, 2010; Mulelu, 2013). In previous studies we demonstrated that the stability of $\mathrm{CynD_{pum}}$ is related to its oligomerisation (Wang et al., 2011; Mulelu, 2010; Mulelu, 2013) and proposed that understanding the process of oligomerisation could potentially help identify residues which can confer higher stability.

6.2 Cryo-electron microscopy and atomic model refinement

To fully understand the mechanism of the oligomerisation of the $\mathrm{CynD_{pum}}$ a high resolution structure had to be obtained. Before this study the only available structures were low resolution negative stain 3D EM reconstructions and homology models (Mulelu, 2013). Although these models produced some good structural insights, they proved to be unreliable in certain regions. Early attempts of obtaining a crystal structure of the $\mathrm{CynD_{pum}}$ were unsuccessful thus an attempt of obtaining a high resolution structure using cryo-EM coupled with 3D reconstruction techniques was carried out. Using this cryo-EM technique, we were able to obtain three detailed 3D structures. Two Q86R+H305K+H308K+H323K $\mathrm{CynD_{pum}}$ structures of 11.1 Å and 6.8 Å resolution were obtained using the FEI Polara and FEI Titan Krios electron microscope respectively. One wild-type $\mathrm{CynD_{pum}}$ structure of 7.4 Å resolution was obtained using the FEI Titan Krios electron microscope. The structure from the FEI Polara was initially obtained and the knowledge gained during its acquisition was later used to refine and optimise the data collection method for the following FEI Titan Krios structures. The optimisation significantly improved the quality of the structures thus demonstrating that the data collection method is unlikely to be the limiting factor in the further acquisition of higher resolution structures. We propose that structural variability of the enzyme due to low stability may be the limiting factor in the acquisition a higher resolution structure. We initially attempted to reduce the structural variability by constructing the very stable Q86R+H305K+H308K+H323K CynD$_{pum}$ variant and although this improved the resolution of the final structure, it still demonstrated high structural variability at some areas such as the C-terminal region. Thus, further work needs to be carried out on improving the stability of CynD$_{pum}$ to produce a variant with limited structural variability thus allowing for higher resolution structures to be obtained. The main challenge in the construction of stable CynD$_{pum}$ variants will be the maintaining as much of the native structure of the enzyme as possible as we have demonstrated the high sensitivity of the CynD$_{pum}$ structure to minimal mutations throughout this study.

One challenge encountered during the acquisition of the high resolution CynD$_{pum}$ EM structures was the estimation of resolution. We stated the resolution of our structures using a fixed-valued FSC threshold but we have observed an understating of the resolution when visually comparing our structures with various simulated CynD$_{pum}$ EM density maps generated at different resolutions. This method of resolution determination may not be ideal for EM density maps as it was observed that different regions of our Cryo-EM map were of varying resolutions. An alternative may be to calculate the local resolution of different regions in the EM density map using programs such as Resmap (Kucukelbir et al., 2014) and report the resolution as a range rather than a single value generated by fixed-valued FSC threshold method which calculates the global resolution.

We successfully further refined a previous homology model of CynD$_{pum}$ (Park et al., 2016) by flexibly fitting it into the density maps obtained from the cryo-EM. The resolution of the cryo-EM density maps allowed for easy fitting of the peptide backbone, alpha helices and large side chain residues with the fitting of beta sheets and loops proving more difficult. Although we successfully managed to fit our homology model into our cryo-EM maps using pseudocrystallographic methods such as the use of COOT (Emsley and Cowtan, 2004) and the real_space_refine program (Afonine et al., 2013), it was evident that there is a lack of programs designed to specifically fit atomic models into near atomic resolution cryo-EM maps. At present, only a few programs are under development to specifically tackle this problem and hence one may have to wait for the development of such programs in order to obtain the best fit of atomic models into the near atomic resolution cryo-EM maps.

6.3 Structural insights

Using the FEI Titan Krios image datasets we were able to obtain two cryo-EM map refined atomic structures of both the Q86R+H305K+H308K+H323K and wild-type CynD_{pum}. Based on these atomic structures we observed that there is high structural similarity between the wild-type and the Q86R+H305K+H308K+H323K mutant although differences in oligomerisation and stability between these enzymes were observed. When comparing our two structures we observed that the high stability conferred by the Q86R+H305K+H308K+H323K mutations was due to the substitution of the Gln86-Gln86 hydrogen bond to a more stabilising Arg86-Arg86 interaction across D-surface (Q86R) and the introduction of a new C-terminal tail interaction across the D-surface as a result of the H305K+H308K+H323K mutations. We also observed that interactions at the C-surface were the predominant interfacial interactions responsible for the oligomerisation of CynD_{pum}, in which only the residues from the C-surface region 1, and none from region 2, are likely to participate across this surface. We observed that the major interactions across the C-surface occur between the α7 helix and the α2 helix in which the Asp275-Arg67 ionic bond and the Asp275-His71 hydrogen bond interactions play a crucial role in oligomerisation of the CynD_{pum} enzyme. The only significant structural difference observed between the wild-type and the Q86R+H305K+H308K+H323K mutant was located at the C-terminal region. For both enzymes it was difficult to fit in C-terminal residues into the EM densities as there were a number of large and undistinguishable densities located at the C-terminal region. Cryo-EM maps of higher resolution are needed to allow accurate fitting of the C-terminal region. Solving the structure of the C-terminal tail would provide significant insight as the C-terminal tail has been shown to play a crucial role in both the oligomerisation and activation of the CynD_{pum} enzyme (Sewell et al., 2005; Wang et al., 2011; Mulelu, 2013).

When comparing the crystal structure of the branch 1 helix-forming nitrilase Nit6803 from *Synechocystis sp. Strain PCC6803* (PDB code: 3WUY) and its distant, C shaped, non-spiral forming homolog β-alanine-synthase (βaS) from *Drosophila melanogaster* (PDB code: 2VHI), it was demonstrated that the main difference between these structures was the conformation of the C-surface region 1 and 2 (Park et al., 2016). Based on our CynD_{pum} structures we observed that there was a higher structural similarity between the CynD_{pum} and Nit6803 than the βaS, and we proposed that most of the branch 1 nitrilases share this C-surface structural conformation, which allows them to form extended helices. Further studies of structural variations between the branch 1 nitrilases and the other members of the nitrilase superfamily still have to be carried out to fully understand the factors that dictate the helix formation, but this may require more branch 1 nitrilase crystal structures to be solved, as only one is currently available.

6.4 pH dependent oligomerisation

The participation of the histidines has been proposed in the pH dependent transition from short spirals to long helices, as this transition occurs as the pH drops below 6.0 (Histidine pKa of 6.0). To investigate the role of the histidines in the oligomerisation of $\mathrm{CynD_{pum}}$ we mutated all the non-C-terminal tail histidines (His62, His71, His128, His167, His184, His241 and His285) to either an alanine, aspartate or lysine. About 50% of the histidine substitution mutants resulted in the insoluble expression demonstrating the importance of some of the histidines in the solubility and hence the correct folding of the enzyme. The branch 1 nitrilase conserved His167-His184 interaction was shown to be crucial in the stability and possibly in the positioning of the catalytic Cys164 of $\mathrm{CynD_{pum}}$ and other related branch 1 nitrilases. We observed that the residues from the interfacial regions are not the only residues involved in oligomerisation as other non-interfacial region histidine residues were shown to affect the oligomerisation of the enzyme (His128, His184, His241 and His285). We observed that no single $\mathrm{CynD_{pum}}$ histidine is responsible for the observed pH dependent oligomerisation as no one histidine mutation resulted in the complete disruption of the enzyme’s oligomerisation. We propose that multiple histidines are responsible for the pH dependent oligomerisation with the His285 being the most involved histidine in the oligomerisation. It is still unclear whether the change in the histidines’ electrostatic properties directly affects the pH dependent oligomerisation or does the change in the histidines’ charge confer conformational changes of the overall $\mathrm{CynD_{pum}}$ structure which in turn results in the changes in oligomerisation. With the help of our $\mathrm{CynD_{pum}}$ structures further mutagenesis studies could be carried out in order to elucidate the precise interactions in which these histidines are involved in.

Using low resolution negative stain EM reconstructions of the fibre forming histidine substitution mutants (H62K, H71A, H71D H71K, H128K H184D, H241A and H285D) we were able to observe variations in the helical symmetry of the mutant fibres and discovered the importance of the positive charge at the His71 position in maintaining the helical symmetry of the CynDpum fibre. The 3D structures also demonstrated that the loss of cyanide degrading activity by some of the histidine mutants was not solely due to the changes in the helical symmetry. It has been previously shown that the change in the helical symmetry of some filamentous nitrilases by mutagenesis can potentially change the substrate specificity of the enzyme (Woodward, 2011). Two of the fibre forming histidine substitution mutations (H71D and H285D) resulted in both the inactivation and the change in the helical symmetry of the enzyme. The inactivation of these mutants may only be limited to cyanide degrading activity and not other nitrile substrates, hence a substrate profiling study of these mutants may be required to access if the change in the helical symmetry affected the substrate specificity of the enzyme. Further investigation into the helical twist determining residues using a combination of mutagenesis studies and high resolution CynD$_{\mathrm{pum}}$ structures could help identify specific helical twist determining residues thus allow for the design of multiple substrate specific industrial enzyme using the CynD$_{\mathrm{pum}}$ as the sole template. A furture study on the enzyme kinetics and pH profile of the CynD$_{\mathrm{pum}}$ variants could also yield good insights and may help us understand the relationship between the activity and oligomerisation of the enzyme.

Obtaining a crystal structure of the CynD$_{\mathrm{pum}}$ has proven challenging due to the difficult production of protein crystals. One of the factors preventing the formation of crystals has been proposed to be the enzyme’s formation of complex quaternary structures. As one of our initial goal we proposed that identifying the residues responsible for oligomerisation could potentially help in manipulation of the oligomerisation, through mutations, to form isoforms of the enzyme which may pack with crystalline symmetry and crystallise readily. This may prove even more difficult than previously anticipated as we have shown that the CynD$_{\mathrm{pum}}$ is highly sensitive to minimal mutations. Disrupting the interactions at the interfacial region was proposed as a strategy for the disruption of the overall oligomerisation of the enzyme. This strategy has also been shown to be more challenging to implement as we have shown that the interaction at the interfacial regions have a compensating mechanism which reduces or prevents changes at one interfacial region from affecting the overall oligomeric structure of the CynD$_{\mathrm{pum}}$ enzyme.

REFERENCES

Abou Nader, M. (2012) Directed evolution of cyanide degrading enzymes. Doctoral dissertation, Texas A&M University, College Station, TX.

Adams, P. D., Afonine, P. V., Bunkóczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., McCoy, A.J., Moriarty, N.W., Oeffner, R., Read, R. J., Richardson, D.C, Richardson, J.S., Terwilliger T.C., Zwart, P.H., (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Cryst* **D66**, 213–221.

Afonine, P.V., Headd, J.J., Terwilliger, T.C., Adams, P.D. (2013) Computational Crystallography Newsletter Volume **4**, Part 2, 43-44. http://phenixonline.org/newsletter/CCN_2013_07.pdf.

Afonine, P.V., Dar, Y., Sobolev, O. V, Terwilliger, T.C., Urzhumtsev, A., Adams, P.D. (2015) Phenix tools for validated refinement of atomic models into Cryo-EM maps. Poster presented at: WCPCW. XXII West Coast Protein Crystallography Workshop, Monterey, United States.

Andrade, J., Karmali, A., Carrondo, M.A., Frazao, C. (2007) Structure of amidase from *Pseudomonas aeruginosa* showing a trapped acyl transfer reaction intermediate state. *J Biol Chem* **282**, 19598–19605.

Bai, X.-C., Fernandez, I.S., McMullan, G., Scheres, S.H. (2013) Ribosome structures to near-atomic resolution from thirty thousand cryo-EM particles. *eLife* **2**, e00461.

Barglow K.T., Saikatendu K.S., Bracey M.H., Huey R., Morris G.M., Olson A.J., Stevens R.C., Cravatt, B.F.(2008) Fuctional proteomic and structural insights into molecular recognition in the nitrilase family enzymes. *Biochemistry* **47**, 13514–13523.

Bartesaghi, A., Merk, A., Banerjee, S., Matthies, D., Wu, X., Milne, J.L.S., Subramaniam, S. (2015) 2.2 Å resolution cryo-EM structure of β-galactosidase in complex with a cell-permeant inhibitor. *Science* **348 (6239)**, 1147–1151.

Beckmann, R., Bubeck, D., Grassucci, R., Penczek, P., Verschoor, A., Blobel, G., Frank, J. (1997) Alignment of conduits for the nascent polypeptide chain in the ribosome-Sec61 complex. *Science* **278**, 2123–2126.

Berman, M.N. (2003) Quaternary structures of the cyanide dihydratase of *Bacillus pumilus* C1 and *Pseudomonuas stutzeri* AK61. Masters Dissertation, University of Cape Town, Rondebosch, South Africa.

Bottcher, B., Wynne, S.A., Crowther, R.A. (1997) Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature* **386**, 88–91.

Bradford, M.M. (1976) A rapid and sensitive method for quantitation of microgram quantities of protein utilizing the principle of protein-dye-binding. *Anal Biochem* **72**, 248–54.

Brilot, A.F., Chen, J.Z., Cheng, A., Pan, J., Harrison, S.C., Potter, C.S., Carragher, B., Henderson, R., Grigorieff, N. (2012) Beam-induced motion of vitrified specimen on holey carbon film. *J Struct Biol* **177**, 630–637.

Campbell, M.G., Veesler, D., Cheng, A., Potter, C.S., Carragher B. (2015) 2.8 Å resolution reconstruction of the *Thermoplasma acidophilum* 20S proteasome using cryo-electron microscopy. *eLife* **4**, e06380.

Cardone, G., Heymann, J.B., Steven, A.C. (2013) One number does not fit all: mapping local variations in resolution in cryo-EM reconstructions. *J Struct Biol* **184**, 226–36.

Carver, T.E., Bordeau, B., Cummings, M.D., Petrella, E.C., Pucci, M.J., Zawadzke, L.E. (2005) Decrypting the biochemical function of an essential gene from *Streptococcus pneumoniae* using ThermoFluor technology. *J Biol Chem* **280**, 11704–11712.

Chapman, M.S. (1995) Restrained real-space macromolecular atomic refinement using a new resolution-dependent electron-density function. *Acta Crystallogr* **A51**, 69–80.

Chen, L.F., Blanc, E., Chapman, M.S., Taylor, K.A. (2001) Real space refinement of acto-myosin structures from sectioned muscle. *J Struct Biol* **133**, 221–232.

Chen, C., Chiu, W., Liu, J., Hsu, W., Wang, W.C. (2003) Structural basis for catalysis and substrate specificity of *Agrobacterium radiobacter* N-Carbamoyl-D-amino acid amidohydrolase. *J Biol Chem* **278**, 26194–26201.

Cheng, Y. (2015) Single-particle cryo-EM at crystallographic resolution. *Cell* **161**, 450–457.

Chin, K.H., Tsai, Y.D., Chan, N.L., Huang, K.F., Wang, A.H.J., Chou, S.H. (2007) The crystal structure of XC1258 from *Xanthomonas campestris*: a putative prokaryotic Nit protein with an arsenic adduct in the active site. *Proteins: Struct Funct Bioinform* **69**, 665–671.

Chiu, W.C., You, J.Y., Liu, J.S., Hsu, S.K., Hsu, W.H., Shih, C.H., Hwang, J.K., Wang, W.C. (2006) Structure-Stability-Activity Relationship in Covalently Cross-linked N-Carbamoyl D-Amino acid Amidohydrolase and N-Acylamino acid Racemase. *J Mol Biol* **359**, 741–753.

Crum, M.A., Park, J.M., Sewell, B.T., Benedik M.J. (2015a) C-terminal hybrid mutant of *Bacillus pumilus* cyanide dihydratase dramatically enhances thermal stability and pH tolerance by reinforcing oligomerization. *J Appl Microbiol* **118**, 881–889.

Crum, M.A., Park, J.M., Mulelu, A.E., Sewell, B.T., Benedik, M.J. (2015b) Probing C-terminal interactions of the *Pseudomonas stutzeri* cyanide-degrading CynD protein. *Appl Microbiol Biotechnol* **99**, 3093–3102.

Crum, M.A., Sewell, B.T., Benedik, M.J. (2016) *Bacillus pumilus* Cyanide Dihydratase Mutants with Higher Catalytic Activity. *Front Microbiol* **7**, 1264.

Dent, K.C., Weber, B.W., Benedik, M.J., Sewell, B.T. (2009) The cyanide hydratase from *Neurospora crassa* forms a helix which has a dimeric repeat. *Appl Microbiol Biotechnol* doi: 10.1007/s00253-008-1735–4.

DeRosier, D. J. and Klug, A. (1968) Reconstruction of three dimensional structures from electron micrographs. *Nature* **217**, 130–134.

Desfosses, A., Ciuffa, R., Gutsche, I., Sachse, C. (2014) SPRING – An image processing package for single-particle based helical reconstruction from electron cryomicrographs. *J Struct Biol* **185**, 15–26.

DiMaio, F., Tyka, M.D., Baker, M.L., Chiu, W., Baker, D. (2009) Refinement of protein structures into low-resolution density maps using rosetta. *J Mol Biol* **392**, 181–90.

DiMaio, F., Song, Y., Li, X., Brunner, M. J., Xu, C., Conticello, V., Egelman. E., Marlovits, T.C., Cheng, Y., Baker, D., (2015) Atomic-accuracy models from 4.5-Å cryo-electron microscopy data with density-guided iterative local refinement. *Nature Methods* **12**, 361–365.

Dubochet, J., Adrian, M., Chang, J.J., Homo, J.C., Lepault, J., McDowall, A.W., Schultz, P., (1988) Cryo-electron microscopy of vitrified specimens. *Q Rev Biophys* **21**, 129–228.

Dunne C., Dolan B., Clyne M. (2014) Factors that mediate colonization of the human stomach by *Helicobacter pylori*. *World J Gastroenterol* **19**, 5610–24.

Egelman, E.H. (2000) A robust algorithm for the reconstruction of helical filaments using single-particle methods. *Ultramicroscopy* **85**, 225–234.

Egelman, E.H. (2007) Single-particle reconstruction from EM images of helical filaments. *Curr Opin Struct Biol* **17(5)**, 556–561.

Eicher, J. J. (2007) pH-dependence of the quaternary structure of the cyanide dihydratase from *Bacillus pumilus*. Masters dissertation, University of Cape Town, Rondebosch, South Africa.

Eisenberg, D., Luthy, R., Bowie, J.U. (1997) VERIFY3D: assessment of protein models with three-dimensional profiles. *Methods in Enzymology* **277**, 396–404.

Emsley, P., Cowtan, K. (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr* **D60**, 2126–2132.

Ericsson, U.B., Hallberg, M.B., DeTitta, G.T., Dekker, N., Nordlund, P. (2006) Thermofluor-based high-throughput stability optimization of proteins for structural studies. *Anal Biochem* **357**, 289–298.

Fernandez, J.J., Luque, D., Caston, J.R., Carrascosa, J.L. (2008) Sharpening high resolution information in single particle electron cryomicroscopy. *J Struct Biol* **164**, 170–175.

Fisher, F.B. and Brown, J.S. (1952) Colorimetric determination of cyanide in stack gas and waste water. *Anal Chem* **24**, 1440–1444.

Frank, J., Verschoor, A., Boublik, M. (1981) Computer averaging of electron micrographs of 40S ribosomal subunits. *Science* **214**, 1353–1355.

Frank, J., Radermacher, M., Penczek, P., Zhu, J., Li, Y., Ladjadj, M., Leith, A. (1996) SPIDER and WEB: processing and visualization of images in 3D electron microscopy and related fields. *J Struct Biol* **116**, 190–199.

Godtfredsen, S.E., Ingvorsen, K., Yde, B., Anderson, O. (1985) In: Tramper J, Vanderplas HC, Linko P (eds) Biocatalysis in organic syntheses. Elsevier, Amsterdam, pp 3–18.

Goran tek-en. (2015) Thermal Shift Assay diagram, Digital diagram, viewed May 2015, https://commons.wikimedia.org/wiki/File:Thermal_Shift_Assay_diagram.svg

Gordon, R,D., Qiu W., Romanov, V., Lam, K., Soloveychik, M., Benetteraj, D., Battaile, K.P., Chirgadze, Y.N., Pai, E.F., Chirgadze, N.Y. (2013) Crystal structure of the CN-hydrolase SA0302 from the pathogenic bacterium *Staphylococcus aureus* belonging to the Nit and NitFhit Branch of the nitrilase superfamily. *J Biomol Struct Dyn* **10**, 1057–1065.

Gray, A. (2008) Nitrilase fibre stability: Resistance to mutagenesis of the D-surface. Honours Thesis, University of Cape Town, Rondebosch, South Africa.

Grifantini, R., Pratesi, C., Galli, G., Grandi, G. (1996) Topological mapping of the cysteine residues of Ncarbamyl-D-amino-acid amidohydrolase and their role in enzymatic activity. *J Biol Chem* **271**, 9326–9331.

Grigorieff, N., Ceska, T.A., Downing, K.H., Baldwin, J.M., Henderson, R. (1996) Electron-crystallographic refinement of the structure of bacteriorhodopsin. *J Mol Biol* **259**, 393–421.

Harauz. G., van Heel, M. (1986) Exact filters for general geometry three-dimensional reconstruction. *Optik* **73**, 146–156.

Heinemann, U., Engels, D., Burger S., Kiziak C., Mattes, R., Stolz, A. (2003) Cloning of a Nitrilase Gene from the cyanobacterium *Synechocystis sp.* Strain PCC6803 and heterologous expression and characterization of the encoded protein. *Appl Environ Microbiol* **239**, 4359–4366.

Henderson, R., Glaeser, R.M., (1985) Quantitative analysis of image contrast in electron micrographs of beam-sensitive crystals. *Ultramicroscopy* **16**, 139–150.

Henderson, R. (2015) Overview and future of single particle electron cryomicroscopy. *Arch Biochem Biophys* **581**, 19–24.

Hooft, R.W., Vried, G., Sander, C., Abola, E.E. (1996). Errors in protein structures. *Nature* **381** (6580), 272.

Hung, C.L., Liu, J.H., Chiu, W.C., Huang, S.W., Wang, W.C. (2007) Crystal structure of *Helicobacter pylori* Formamidase AmiF reveals a cysteine-glutamate-lysine catalytic triad. *J Biol Chem* **282**, 12220–12229.

Jandhyala, D.M., Berman, M., Meyers, P.R., Sewell, B.T., Willson, R.C., Benedik, M.J. (2003) Cyn D, the cyanide dihydratase from *Bacillus pumilus*: gene cloning and structural studies. *Appl Environ Microbiol* **69**, 4794–4805.

Jandhyala, D.M., Willson, R.C., Sewell, B.T., Benedik, M.J. (2005) Comparison of cyanide degrading nitrilases. *Appl Microbiol Biotechnol* **68**, 327–335.

Jiang, J., Pentelute, B.L., Collier, R.J., Zhou, Z.H. (2015) Atomic structure of anthrax protective antigen pore elucidates toxin translocation. *Nature* **521**, 545–549.

Jones, R.J, and Steven, A.L**.** (1997) Effects of cyanide on corals in relation to cyanide fishing on reefs. *Mar. Freshwater Res* **48,** 517–522.

Kessel, M., Radermacher, M., Frank, J. (1985) The structure of the stalk surface layer of a brine pond microorganism: correlation averaging applied to a double layered lattice structure. *J Microsc* **139**, 63–74.

Kimani, S.W., Agarkar, V.B., Cowan, D.A., Sayed, M.F.R., Sewell, B.T. (2007) Structure of an aliphatic amidase from *Geobacillus pallidus* RAPc8. *Acta Crystallogr Sect D: Biol Crystallogr* **63**, 1048–1048.

Kiziak, C., Conradt D., Stolz, A., Mattes, R., Klein, J. (2005) Nitrilase from *Pseudomonas fluorescens* EBC191: cloning and heterologous expression of the gene and biochemical characterization of the recombinant enzyme. *Microbiology* **151**, 3639–3648.

Kovac, C. (2000). Cyanide spill threatens health in Hungary. *Brit. Med. J.* **320,** 536.

Kucukelbir, A., Sigworth, F.J., Tagare, H.D. (2014) Quantifying the Local Resolution of Cryo-EM Density Maps. *Nature Methods* **11**, 63-65.

Kumaran, D., Eswaramoorthy, S., Gerchman, S.E., Kycia, H., Studier, F.W., Swaminathan, S. (2003) Crystal structure of a putative CN hydrolase from yeast. Proteins 52, 283–291.

LaRonde-LeBlanc, N., Resto, M., Gerratana, B. (2009) Regulation of active site coupling in glutamine-dependent NAD+ synthetase. *Nat Struct Mol Biol* **16**, 421–429.

Lander, G.C.; Stagg, S.M., Voss, N.R., Cheng, A., Fellmann, D., Pulokas, J., Yoshioka, C., Irving, C., Mulder, A., Lau, P.W. (2009) Appion: an integrated, database-driven pipeline to facilitate EM image processing. *J Struct Biol* **166**, 95–102.

Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thomton, J.M., (1993) PROCHECK: a program to check the stereochemical quality of protein structures. *J Appl Cryst* **26**, 283–291.

Lavinder, J.J., Hari, S.B., Sullivan, B.J., Magliery, T.J. (2009) High-throughput thermal scanning: a general, rapid dye-binding thermal shift screen for protein engineering. *J Am Chem Soc* **131**, 3794–3795.

Li, X., Mooney, P., Zheng, S., Booth, C.R., Braunfeld, M.B., Gubbens, S., Agard, D.A., Cheng, Y. (2013) Electron counting and beam-induced motion correction enable near-atomic resolution single-particle cryo-EM. *Nat. Methods* **10**, 584–590.

Lin, C.H., Chung M.Y., Chen W.B., Chien C.H. (2007) Growth inhibitory effect of the human NIT2 gene and its allelic imbalance in cancers. *FEBS J.* **11**, 2946–56.

Liu. H., Gao, Y., Zhang, M., Qiu, X., Cooper, A.J., Niu, L., Teng, M. (2013) Structures of enzyme-intermediate complexes of yeast Nit2: insights into its catalytic mechanism and different substrate specificity compared with mammalian Nit2. *Acta Crystallogr* **D69**, 1470-81

Lovell, S.C., Davis, I.W., Arendall, W.B., deBakker, P.W., Word, J.M., Prisant, M.G., Richardson, J.S., Richardson, D.C., (2003) Structure validation by Cα geometry: Phi, Psi and Cβ deviation. *Proteins* **50(3)**, 437–450.

Ludtke, S.J., Baldwin, P.R., Chiu, W. (1999) EMAN: semi-automated software for high-resolution single particle reconstructions. *J Struct Biol* **128**, 82–96.

Lundgren, S., Lohkamp, B., Andersen, B., Piskur, J., Dobritzsch, D. (2008) The crystal structure of B-alanine synthase from *Drosophila melanogaster* reveals a homoactameric helical turn-like assembly. *J Mol Biol* **377**, 1544–1559.

Mallick, S.P., Carragher, B., Potter, C.S., Kriegman, D.J. (2005) ACE: automated CTF estimation. *Ultramicroscopy* **104**, 8–29.

Matamá, T., Carneiro, F., Caparrós, C., Gübitz, G.M., Cavaco-Paulo, A. (2007) Using a nitrilase for the surface modification of acrylic fibres. *Biotechnol J* **2**, 353-360.

Matulis, D., Kranz, J.K., Salemme, F.R., Todd, M.J. (2005) Thermodynamic stability of carbonic anhydrase: measurements of binding affinity and stoichiometry using ThermoFluor. *Biochemistry* **44**, 5258–5266.

McMullan, G., Faruqia, A.R., Clare, D., Henderson, R. (2014) Comparison of optimal performance at 300 keV of three direct electron detectors for use in low dose electron microscopy. *Ultramicroscopy* **147**, 56–163.

Meyers, P.R., Gokool, P., Rawlings, D.E., Woods, D.R. (1991) An efficient cyanide-degrading Bacillus pumilus strain. *J Gen Microbiol* **137**, 1397–1400.

Meyers, P.R., Rawlings, D.E., Woods, D.R., Lindsey, G.G. (1993) Isolation and characterization of a cyanide dihydratase from *Bacillus pumilus* C1. *J Bacteriol* **175**, 6105–6112.

Mitsuoka, K., Hirai, T., Murata, K., Miyazawa, A., Kidera, A., Kimura, Y., Fujiyoshi, Y. (1999) The structure of bacteriorhodopsin at 3.0 Å resolution based on electron crystallography: implication of the charge distribution. *J Mol Biol* **286**, 861–882.

Mindell, J.A, Grigorieff, N. J. (2003) Accurate determination of local defocus and specimen tilt in electron microscopy. *J Struct Biol* **142**, 334–347.

Mulelu, A.E. (2010) Factors involved in the oligomerisation of the cyanide dihydratase from *Bacillus pumilus* C1. Honours Thesis, University of Cape Town, Rondebosch, South Africa.

Mulelu, A.E. (2013) Factors involved in the oligomerisation of the cyanide dihydratase from *Bacillus pumilus* C1. Masters dissertation, University of Cape Town, Rondebosch, South Africa.

Nagasawa, T., Wieser, M., Nakamura, T., Iwahara, H., Yoshida, T., Gekko, K. (2000) Nitrilase of Rhodococcus rhodochrous J1. Conversion into the active form by subunit association. *Eur J Biochem* **267**, 138-44.

Nakai, T., Hasegawa, T., Yamashita, E., Yamamoto, M., Kumasaka, T., Ueki, T., Nanba, H., Ikenaka, Y. (2000) Crystal structure of N-carbamyl-D-amino acid amidohydrolase with a novel catalytic framework common to amidohydrolases. *Structure* **8**, 729–739.

Nel, A.J., Tuffin, I.M., Sewell, B.T., Cowan, D.A. (2011) Unique aliphatic amidase from a psychrotrophic and haloalkaliphilic *nesterenkonia* isolate. *Appl Environ Microbiol* **77**, 3696–3702.

Niesen, F.H., Berglund, H., Vedadi, M. (2007) The use of differential scanning fluorimetry to detect ligand interactions that promote protein stability. *Nature Protocols* **2**, 2212–2221.

Nishise, H., Kurihara, M., Tani, Y. (1987) Microbial synthesis of tran-examic acid intermediate from dinitrile. *Agric Biol Chem* **51**, 2613–2616.

Novo, C., Farnaud, S., Tata, R., Clemente, A., Brown, P.R. (2002) Support for a three-dimensional structure predicting a Cys-Glu-Lys catalytic triad for *Pseudomonas aeruginosa* amidase comes from site-directed mutagenesis and mutations altering substrate specificity. *Biochem J* **365**, 731–738.

Oany, A.R., Ahmed, M.S., Jahan, N., Latif, M.A., Mahmud, S., Hossain, M.A., Akter, F., Rakib H.H., Islam, M.S. (2015) Homology modeling and assigned functional annotation of an uncharacterized antitoxin protein from *Streptomyces xinghaiensis*. *Bioinformation* **11**, 493–500.

Orlova, E.V., Dube, P., Harris, J.R., Beckmann, E., Zemlin, F., Markl, J., van Heel, M. (1997) Structure of Keyhole Limpet Hemocyanin Type 1 (KLH1) at 15 Å resolution by electron cryomicroscopy and angular reconstitution. *J Mol Biol* **271**, 417–437.

Pace, H.C., Hodawadekar, S.C., Draganescu, A., Huang, J., Bieganowski, P., Pekarsky, Y., Croce, C.M., Brenner, C. (2000) Crystal structure of the worm NitFhit Rosetta Stone protein reveals a Nit tetramer binding two Fhit dimers. *Curr Biol* **10**, 907–917.

Pace, H. and Brenner, C. (2001) The nitrilase superfamily: classification, structure and function. *Genome Biol* **2**, 1–9.

Pantoliano, M.W., Petrella, E.C., Kwasnoski, J.D., Lobanov, V.S., Myslik, J., Graf, E., Carver, T., Asel, E., Springer, B.A., Lane, P., Salemme, F.R. (2001) High-density miniaturized thermal shift assays as a general strategy for drug discovery. *J Biomol Screen* **6**, 429–440.

Parent, K.N., Sinkovits, R.S., Suhanovsky, M.M., Teschke, C.M., Egelman, E.H., Baker, T.S. (2010) Cryo-reconstructions of P22 polyheads suggest that phage assembly is nucleated by trimeric interactions among coat proteins. *Phys Biol* **7**. 45004.

Park, J. (2014) Regions involved in the oligomerization and activity of the spiral forming nitrilase Cyanide Dihydratase. Doctoral dissertation, Texas A&M University, College Station, TX.

Park, J. M., Mulelu, A., Sewell, B.T., Michael J. Benedik (2016) Probing an interfacial surface in the cyanide dehydratase from *Bacillus pumillus,* a spiral forming nitrilase. *Front Microbiol* **6**, 1479.

Pednekar D1, Tendulkar A, Durani S. (2009) Electrostatics-defying interaction between arginine termini as a thermodynamic driving force in protein-protein interaction. *Proteins* **74**, 155-163.

Penczek, P.A. (2002) Three-dimensional spectral signal-to-noise ratio for a class of reconstruction algorithms. *J Struct Biol* **138**, 34–46.

Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C., Ferrin, T.E. (2004) UCSF Chimera—a visualization system for exploratory research and analysis. *J Comput Chem* **25**, 1605–1612.

Phillips, K., de la Peña A.H. (2011) The combined use of the Thermofluor assay and ThermoQ analytical software for the determination of protein stability and buffer optimization as an aid in protein crystallization. *Curr Protoc Mol Biol.* Chapter **10**, Unit10.28.

Prasad, S, Misra, A., Jangir, V.P., Awasthi, A., Raj, J., Bhalla, T.C. (2007) A propionitrile-induced nitrilase of Rhodococcus sp NDB 1165 and its application in nicotinic acid synthesis. *World J Microbiol Biotechnol* **23**, 345-353.

Raczynska, J.E., Vorgias, C.E., Antranikian, G., Rypniewski, W. (2011) Crystallographic analysis of a thermoactive nitrilase. *J Struct Biol* **173**, 294–302.

Sachse, C., Chen, J.Z., Coureux, P.D., Stroupe, M.E., Fandrich, M., Grigorieff, N. (2007) High-resolution electron microscopy of helical specimens: a fresh look at tobacco mosaic virus. *J Mol Biol* **371**, 812–835.

Sakai, N., Tajika, Y., Yao, M., Watanabe, N., Tanaka, I. (2004) Crystal structure of hypothetical protein PH0642 from *Pyrococcus horikoshii* at 1.6 Å resolution. *Proteins: Struct Funct Bioinform* **57**, 869–873.

Scheffer, M.P. (2006) Helical structures of the cyanide degrading enzymes from *Gloeocercospora sorghi* and *Bacillus pumilus* providing insights into nitrilase quaternary interactions. MSc Thesis, University of Cape Town, Rondebosch, South Africa.

Sewell, B.T., Berman, M.N., Meyers, P.R., Jandhyala, D., Benedik, M.J. (2003) The cyanide degrading nitrilase from *Pseudomonas stutzeri* AK61 is a two-fold symmetric, 14-subunit spiral. *Structure* **11**, 1–20.

Sewell, B.T., Thuku, R.N., Zhang, X., Benedik, M.J. (2005) The oligomeric structure of nitrilases: the effect of mutating interfacial residues on activity. *Ann NY Acad Sci* **1056**, 153–159.

Sippl, M.J. (1993) Recognition of errors in three-dimensional structures of proteins. *Proteins* **17(4)**, 355–362.

Skouloubris S., Labigne A., De Reuse H. (2001) The AmiE aliphatic amidase and AmiF formamidase of *Helicobacter pylori*: natural evolution of two enzyme paralogues. *Mol Microbiol* **40**, 596–609.

Skouloubris S., Labigne A., De Reuse H. (1997) Identification and characterization of an aliphatic amidase in *Helicobacter pylori*. *Mol Microbiol* **25**, 989–998.

Sorzano, C.O.S., Marabini, R., Velazquez-Muriel, J., Bilbao-Castro, J.R, Scheres, S.H.W., Carazo, J.M., Pascual-Montano, A. (2004) XMIPP: a new generation of an open-source image processing package for electron microscopy. *J Struct Biol* **148(2)**, 194–204.

Sousa, D., Grigorieff, N. (2007) Ab initio resolution measurement for single particle structures. *J Struct Biol* **157**, 201–210.

Stalker, D.M., McBride, K.E., Malyj, L.D. (1988) Herbicide resistance in transgenic plants expressing a bacterial detoxification gene. *Science* **242**, 419–423.

Stalker, D.M., Kiser, J.A., Baldwin, G., Coulombe, B., Houck, C.M. (1996) Cotton weed control using the BXN system. In *Herbicide Resistant Crops* ed. O Duke, S. pp. 93–105. Boca Raton (Florida): CRC press (Lewis Publishers).

Thompson, J.D., Higgins, D.G., Gibson, T.J. (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.

Thuku, R.N., Weber, B.W., Varsani, A., Sewell B.T. (2007) Post-translational cleavage of recombinantly expressed nitrilase from *Rhodococcus rhodochrous* J1 yields a stable, active helical form. *FEBS J* **274**, 2099-2108.

Thuku, R.N., Brady, D., Benedik, M.J., Sewell B.T. (2009) Microbial nitrilases: versatile, spiral forming, industrial enzymes. *J Appl Microbiol* **106**, 703–727.

Trabuco, L.G., Villa, E., Mitra, K., Frank, J., Schulten, K. (2008) Flexible fitting of atomic structures into electron microscopy maps using molecular dynamics. *Structure* **16**, 673–683.

Unser, M., Sorzano, C.O., Thevenaz, P., Jonic, S., El-Bez, C., De Carlo, S., Conway, J.F., Trus, B.L. (2005) Spectral signal-to-noise ratio and resolution assessment of 3D reconstructions. *J Struct Biol* **149**, 243–255.

Unwin, N. (2005). Refined structure of the nicotinic acetylcholine receptor at 4 Å resolution. *J Mol Biol* **346**, 967–989.

Van der Vyver, L. (2007) Quaternary structures of the cyanide dihydratase of Bacillus pumilus C1 is resistant to site-directed mutagenesis of the D-surface. Honours Thesis, University of Cape Town, Rondebosch, South Africa.

Van Heel, M., Keegstra, W., Schutter, W.G., van Bruggen, E.F.J. (1983). Arthropod hemocyanin studied by image analysis. The EMBO Workshop on Invertebrate Respiratory Proteins Leeds, *Structure Life Chemistry Reports*. 69–73.

Van Heel, M. (1987) Similarity measures between images. *Ultramicroscopy* **21**, 95–100.

Vejvoda, V., Kaplan, O., Bezouska, K., Pompach, P., Sulc, M., Cantarella, M., Benada, O., Uhnakova, B. et al. (2008) Purification and characterization of a nitrilase from *Fusarium solani* O1. *J Mol Catal B Enzym* **50**, 99–106.

Volkmann, N., Hanein, D., Ouyang, G., Trybus, K.M., DeRosier, D.J., Lowey, S., (2000) Evidence for cleft closure in actomyosin upon ADP release. *Nat Struct Biol* **7**, 1147–1155.

Wang, W.C., Hsu, W.H., Chien, F.T., Chen, C.Y. (2001) Crystal structure and site-directed mutagenesis studies of N-Carbamoyl-D-amino-acid amidohydrolase from Agrobacterium radiobacter reveals a homotetramer and insight into a catalytic cleft. *J Mol Biol* **306**, 251–261.

Wang, L., Watermeyer, J.M., Mulelu, A.E., Sewell, B.T., Benedik, M.J. (2011) Engineering pH-tolerant mutants of a cyanide dihydratase. *Appl Microbiol Biotechnol* **94**, 131–140.

Weber, B.W., Kimani, S.W., Cowan, D.A., Hunter, R., Venter, G.A., Gumbart, J.C., Thuku, R.N., Sewell, B.T., Varsani, A. (2013) The mechanism of the amidases: mutating the glutamate adjacent to the catalytic triad inactivates the enzyme due to substrate mispositioning. *J Biol Chem* **288**, 28514–28523.

Williamson, D.S., Dent, K.C., Weber, B.W., Varsani, A., Frederick, J., Thuku, R.N., Cameron, R.A., van Heerden, J.H., Cowan, D.A., Sewell, B.T.(2010) Structural and biochemical characterization of a nitrilase from the thermophilic bacterium, *Geobacillus pallidus* RAPc8. *Appl Microbiol Biotechnol* **88**, 143–153.

Woodward, J.D., Weber, B.W., Scheffer, M.P., Benedik, M.J., Hoenger, A. and Sewell, B.T. (2008) Helical structure of unidirectionally shadowed metal replicas of cyanide hydratase from Gloeocercospora sorghi. *J Struct Biol* **161**, 111–119.

Woodward, J. D. (2011). The Relationship between Structure and Specificity in the Plant Nitrilases. Ph.D. dissertation, University of Cape Town, Cape Town.

Wriggers, W., Agrawal, R.K., Drew, D.L., McCammon, A., Frank, J. (2000) Domain motions of EF-G bound to the 70S ribosome: Insights from a hand-shaking between multi-resolution structures. *Biophys J* **79**, 1670–1678.

Wriggers, W., Birmanns, S. (2001) Using Situs for flexible and rigid-body fitting of multiresolution single-molecule data. *J Struct Biol* **133**, 193–202.

Wriggers, W. (2012) Conventions and Workflows for Using Situs. *Acta Cryst* **D68**, 344–351.

Wu, S.J., Fogiel, A.J., Petrillo, K.L., Jackson, R.E., Parker, K.N., DiCosimo, R., Ben-Bassat, A., O'Keefe, D.P., Payne, M.S. (2008) Protein engineering of nitrilase for chemoenzymatic production of glycolic acid. *Biotechnol Bioeng* **99**, 717-720.

Wyatt, J. and Knowles, C. (1995) Microbial degradation of acrylonitrile waste effluents: the degradation of effluents and condensates from the manufacture of acrylonitrile. *Int Biodeterior Biodegrad* **35**, 227–248.

Xue, Y.P., Xu, S.Z., Liu, Z.Q., Zheng, Y.G., Shen, Y.C. (2011) Enantioselective biocatalytic hydrolysis of (R,S)-mandelonitrile for production of (R)-(-)-mandelic acid by a newly isolated mutant strain. *J Ind Microbiol Biotechnol* **38**, 337-345.

Yang, S., Yu, X., Galkin, V.E., Egelman, E.H. (2003) Issues of resolution and polymorphism in single-particle reconstruction. *J Struct Biol* **144**, 162–171.

Yonekura, K., Maki-Yonekura, S., Namba, K. (2003). Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. *Nature* **424**, 643–650.

Zhang, Z., Xu, Z., Yang, Z., Liu, Y., Wang, J., Shao, Q., Li, S., Lu, Y., Zhu, W. (2013) The stabilization effect of dielectric constant and acidic amino acids on arginine-arginine (Arg-Arg) pairings: database survey and computational studies. *J Phys Chem B* **117(17)**, 4827–35.

Zhang, L., Yin, B., Wang C., Jiang, S., Wang, H., Yuan Y.A., Wei, D. (2014) Structural insights into enzymatic activity and substrate specificity determination by a single amino acid in nitrilase from *Syechocystis sp.* PCC6803. *J Struct Biol* **188**, 93–101.

Zhou, Z.H. (2011) Atomic resolution cryo-electron microscopy of macromolecular complexes. *Adv Protein Chem Struct Biol* **82**, 1–35.